Differential Geometry I and II, 2021–2022, HU Berlin

Chris Wendl

Contents

These notes will be expanded gradually over the course of the semester. If you notice any typos or mathematical errors, please send e-mail about them to `wendl@math.hu-berlin.de` and they will be corrected.

While the notes are written in English, I make an effort to include the German translations (*geschrieben in dieser Schriftart*) of important terms wherever they are introduced. I will occasionally omit these translations in cases where the English and German words are identical, or if the word has already appeared before with its translation in a different context (e.g. the word "smooth" needs to be defined many times in different contexts, and its German translation is always the same), and also in cases where I can't reliably figure out what the German word is. The latter will happen more often as the course goes on, because the deeper one gets into advanced mathematics, the harder it becomes to find authoritative German sources for clarifying the terminology (and I am not linguistically qualified to invent terms in German myself).

First semester (Differentialgeometrie I)

1. Introduction

Before diving in with definitions, theorems and proofs, I want to give an informal taste of what differential geometry is all about. The word "informal" means, in this case, that you should try not to worry too much about the precise definitions or rigorous arguments behind what we are discussing, but focus instead on the big picture. Before the first lecture is finished, I will revert to being a proper mathematician and give some actual definitions.

1.1. A foretaste of Riemannian geometry. Let's assume for the moment that we all understand what a "smooth surface" is, e.g. you can picture it as a subset[1] of $\mathbb{R}^3$ such that every point has a neighborhood parametrized by some injective[2] C^∞-map

$$\mathbb{R}^2 \overset{\text{open}}{\supset} \mathcal{U} \hookrightarrow \mathbb{R}^3.$$

With this understood, assume

$$\Sigma \subset \mathbb{R}^3$$

is a smooth surface.

1.1.1. *Distances and geodesics.* We could view Σ as a metric space by defining the distance between two points $x, y \in \Sigma$ via the Euclidean metric, but this is not necessarily the most natural thing to do. A more natural notion of distance in the surface Σ would be one that tells you something about the actual distance that an ant has to travel if it walks a path *along the surface* between x and y. If that path is parametrized by a smooth map $\gamma : [a, b] \to \mathbb{R}^3$ satisfying $\gamma([a, b]) \subset \Sigma$, $\gamma(a) = x$ and $\gamma(b) = y$, then the distance travelled is

$$(1.1) \qquad \ell(\gamma) := \int_a^b |\dot{\gamma}(t)|\, dt = \int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle}\, dt,$$

where $\dot{\gamma}(t)$ denotes the time derivative of $\gamma(t)$, $\langle v, w \rangle$ denotes the Euclidean inner product of two vectors $v, w \in \mathbb{R}^3$, and $|v| := \sqrt{\langle v, v \rangle}$ denotes the Euclidean norm. If we denote by $\mathcal{P}(x, y)$ the set of all smooth paths in Σ connecting x to y, then a natural notion of distance on Σ can now be defined by

$$(1.2) \qquad d(x, y) := \inf_{\gamma \in \mathcal{P}(x,y)} \ell(\gamma).$$

The infimum needs to be taken since, in general, there are many distinct paths from x to y that will have different lengths. In principle we are interested in the *shortest* such path, though it is not obvious in general whether such a shortest path must exist:

---

[1]We will soon improve this definition so that surfaces do not need to be regarded as subsets of $\mathbb{R}^3$. In fact, there are some important examples of surfaces that *cannot* be embedded in $\mathbb{R}^3$; a famous example is the Klein bottle, see https://en.wikipedia.org/wiki/Klein_bottle.

[2]We will need to add a condition concerning the derivative of the map $\mathcal{U} \hookrightarrow \mathbb{R}^3$ before this becomes an adequate definition, but let's worry about that later.

QUESTION 1.1. *Given a smooth surface Σ and two distinct points $x, y \in \Sigma$, does there exist a smooth path on Σ from x to y that has the* shortest *possible length? Is it unique?*

We will see later in this semester that the answer to both questions is always yes if x and y are close enough to each other, and the shortest path can then be characterized by a second-order ordinary differential equation. Such a path is called a **geodesic** (*Geodäte* or *geodätische Linie*), and it serves as the best possible substitute for a "straight line" on Σ, even in cases where no actual *straight* paths on Σ exist. The canonical example you should picture is the unit sphere $\Sigma := S^2 \subset \mathbb{R}^3$, whose geodesics are the so-called *great circles*, namely the subsets $S^2 \cap P$ defined via 2-dimensional linear subspaces $P \subset \mathbb{R}^3$. These are the paths that all airplanes would traverse along the Earth if there were no additional factors such as weather conditions or no-fly zones to consider.

1.1.2. *Angles, isometries, and curvature.* The fundamental piece of data that makes the above definition of distance on Σ possible is the Euclidean inner product $\langle\ ,\ \rangle$. In fact, $\langle\ ,\ \rangle$ contains strictly more information than is actually needed for defining distances on Σ; if you look again at the formula (1.1), you'll notice that it doesn't really require knowing what $\langle v, w \rangle$ is for every $v, w \in \mathbb{R}^3$, but is already well-defined if we know how to define this for every pair of vectors v, w that are *tangent* to Σ at any given point. (Indeed, $\dot{\gamma}(t) \in \mathbb{R}^3$ is always tangent to Σ at $\gamma(t)$.) In fact, it would suffice to know what $\langle v, v \rangle$ is for every individual tangent vector v, but knowing $\langle v, w \rangle$ for two distinct vectors provides some additional information that is of geometric interest: it allows us to compute the *angle* between any two tangent vectors. Indeed, the angle θ between two vectors $v, w \in \mathbb{R}^3$ can always be deduced from the formula

$$\langle v, w \rangle = |v| \cdot |w| \cdot \cos \theta.$$

We can therefore define not only the length of any smooth path along Σ, but also the angle between two smooth paths wherever they intersect. This information makes Σ into what we will later call a (2-dimensional) **Riemannian manifold** (*Riemannsche Mannigfaltigkeit*), and the restriction of the inner product to the tangent spaces on Σ, which determines all lengths and angles, is called a **Riemannian metric** (*Riemannsche Metrik*).[3]

Here is a natural question one can ask about Riemannian manifolds. Suppose $\Sigma_1, \Sigma_2 \subset \mathbb{R}^3$ are two smooth surfaces, and $\varphi : \Sigma_1 \to \Sigma_2$ is a smooth bijective map between them whose inverse is also smooth.[4] We call φ in this case a **diffeomorphism** (*Diffeomorphismus*), and say that Σ_1 and Σ_2 are **diffeomorphic** (*diffeomorph*). We say that φ is additionally an **isometry** (*Isometrie*) if it preserves all distances and angles, and in this case, Σ_1 and Σ_2 are said to be **isometric** (*isometrisch*).

QUESTION 1.2. *Given two diffeomorphic surfaces, how can we measure whether they are isometric?*

In simple examples, it is often easy to recognize when two surfaces are diffeomorphic: an example is shown in Figures 1 and 2, where we can compare the standard unit sphere $S^2 \subset \mathbb{R}^3$ with a "nonstandard" embedding of S^2 into $\mathbb{R}^3$ that elongates a portion of the sphere into something more closely resembling a cylinder. It is surely not hard to imagine that these two surfaces in $\mathbb{R}^3$ are diffeomorphic; writing down an explicit example of a diffeomorphism would be a pain in the neck,

---

[3]Caution: there is a potential for confusion in this terminology, because a Riemannian metric is not a particular kind of metric in the sense of metric spaces, though it does determine one via formulas such as (1.2). A Riemannian metric carries strictly more information, since it determines angles in addition to distances.

[4]For the purposes of this discussion, you may assume that a function on a smooth surface $\Sigma \subset \mathbb{R}^3$ is smooth if it can be extended to a smooth function on a neighborhood of Σ; the latter notion is familiar from your first-year Analysis class since the neighborhood is an open subset of $\mathbb{R}^3$. We will later give an equivalent but more elegant definition of smoothness for functions on manifolds.

but we will content ourselves with the intuitive understanding that in the process of "stretching" the standard sphere into its nonstandard counterpart, one could if necessary come up with a smooth bijection between the two. The much deeper observation is that they are *not* isometric, and we will need to develop some technology before we can prove this rigorously. One of the key ideas behind the proof is shown in Figures 1 and 2: on any surface Σ, one can draw a closed piecewise-smooth path along Σ, choose a starting point p_0 on the path and a tangent vector v_0 at p_0, then translate the vector v_0 along the path via a process known as **parallel transport**. We will have to give a careful definition later of what is meant by parallel transport, but Figures 1 and 2 will hopefully give you some intuition about this. The interesting question is now: if we parallel transport the vector v_0 once around our chosen closed path, does it return to the same starting vector? As you can see in the pictures, the answer is no for the triangular path in Figure 1, but yes for the rectangular path in Figure 2. It will turn out that this observation encodes a fundamental difference between these two Riemannian manifolds: the standard sphere has positive **curvature** (*Krümmung*) at every point, but the elongated sphere does not—if fact, the surface in Figure 2 has zero curvature everywhere on the elongated region where our rectangle is drawn.

A major portion of the second half of this semester will be devoted to the precise definition of curvature and its important properties. One of these is that it completely characterizes the notion of *local flatness*:

QUESTION 1.3. *Given a smooth surface $\Sigma \subset \mathbb{R}^3$ and a point $p \in \Sigma$, does p have any neighborhood that is isometric to an open subset of the "flat" surface $\mathbb{R}^2 \times \{0\} \subset \mathbb{R}^3$?*

A surface $\Sigma \subset \mathbb{R}^3$ is called **locally flat** (*lokal flach*) if the answer to Question 1.3 is yes for every point $p \in \Sigma$. Figure 3 shows an example of a surface that is locally flat, even though it does not look flat in the picture: you know it is locally flat because you know that an ordinary piece of paper can be bent into this cylindrical shape without breaking or stretching it. This is *not* true of the standard unit sphere in $\mathbb{R}^3$. Perhaps you've never held in your hand a piece of paper that's shaped like part of a globe[5], but you can surely imagine that if you did, you could never make it *flat* without breaking or stretching it. This is another symptom of the positive curvature of the round sphere.[6] By contrast, the cylindrical surface in Figure 3 has zero curvature everywhere. The statement that a cylinder is in some sense "not curved" may seem jarring at first, but you'll get used to it: the point is that the quantity we're calling curvature should depend only on the Riemannian metric, and not on the specific way we've chosen to embed our Riemannian manifold in $\mathbb{R}^3$. If two surfaces are isometric, then their curvatures at corresponding points will always be the same.

The positive curvature of the round sphere is not unrelated to the fact that the angles of the "triangle" in Figure 1 add up to considerably more than 180 degrees. We will later also see examples of surfaces with *negative* curvature: the basic picture to have in mind is the shape of a *saddle*. In these surfaces, the angles in a triangle will add up to *less* than 180 degrees. The elongated sphere in Figure 2 has zero curvature in the shaded region, but not everywhere; since it is diffeomorphic to S^2, one could reinterpret this as the statement that S^2 admits a Riemannian metric that is locally flat in some region. That is not a deep or surprising statement, as *every* Riemannian metric on an arbitrary manifold can in fact be modified to make it flat in some small region. A more interesting question is whether it can be modified to make it locally flat *everywhere*, like the cylindrical surface in Figure 3. Let us take this opportunity to state a standard corollary of a rather deep theorem:

---

[5]If you know where to buy one, please let me know!

[6]This is also the mathematical reason why it is impossible to create a flat map of the Earth without distorting distances and angles in some regions.
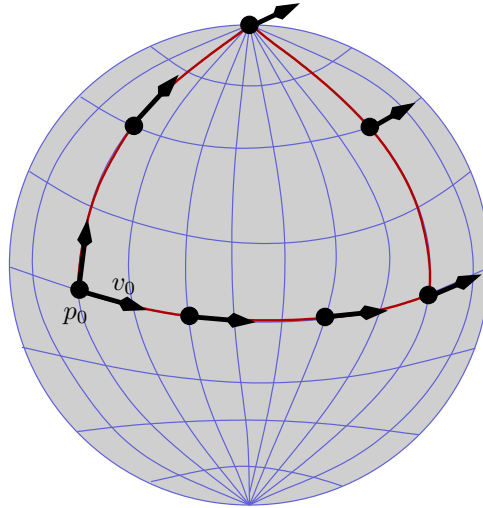
FIGURE 1.   The "round" sphere $S^2 \subset \mathbb{R}^3$. Parallel transport of a vector along a closed path leads to a different vector upon return.
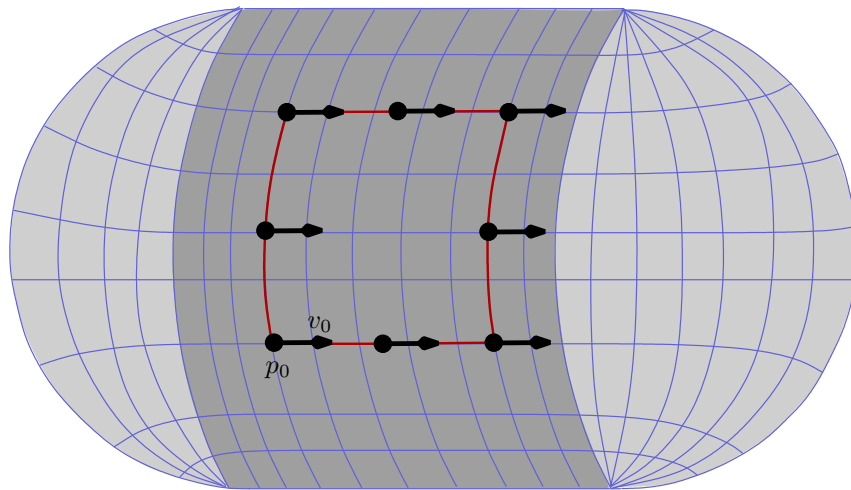
FIGURE 2.   A different embedding of S^2 in $\mathbb{R}^3$, so that the darkly shaded region is locally flat. Parallel transport of a vector around a closed path in this region always leads back to the same initial vector.

THEOREM. *There is no Riemannian metric on the sphere S^2 that is everywhere locally flat.*

This will follow from the beautiful *Gauss-Bonnet theorem* for surfaces, to be proved near the end of this semester. It relates the integral of the curvature over a compact surface to a topological quantity, its *Euler characteristic*, which in the case of S^2 is positive. This is the reason why Figure 2 could not have been drawn so that *every* part of the sphere had zero curvature. We will also use a variant of this theorem to understand what the various observations above about sums of angles of triangles have to do with curvature.
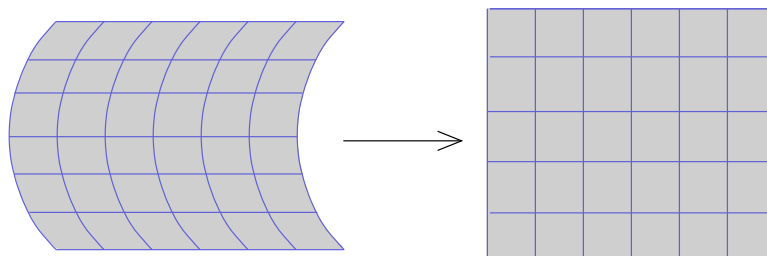
FIGURE 3.   A piece of a cylinder can be flattened to a plane without changing any lengths or angles on the surface.

1.1.3. *Spacetime as a pseudo-Riemannian 4-manifold.* Differential geometry is not only about surfaces, and it also plays an important role in subjects that cannot accurately be called "pure" mathematics. This is true especially in several areas of theoretical physics, the most famous of which is Einstein's theory of gravitation, known as the general theory of relativity *(allgemeine Relativitätstheorie)*. We will not directly discuss gravitation in this course, but several of the mathematical concepts we will cover are essential for understanding Einstein's picture of the universe.

The paradigm introduced by Einstein for an understanding of space and time can be summarized as follows:

(1) There are three spatial dimensions, but time adds a fourth. Locally, an "event" occurring in a particular place at a particular time thus requires four coordinates for its description, defining a point in $\mathbb{R}^4$.

(2) The picture in item (1) is only local, i.e. it is sufficient for describing interactions between events on a small or medium scale, but one should not assume that the set of all events in the universe (known as **spacetime** or *Raumzeit*) is in bijective correspondence with $\mathbb{R}^4$. In general, spacetime could be any smooth 4-dimensional manifold.

(3) Spacetime is endowed with a (pseudo-)Riemannian metric, which determines a notion of geodesics. In the absence of forces other than gravity, all objects move along geodesics in spacetime.

(4) The presence of mass affects the curvature of spacetime and thus changes the geodesics. A precise relationship between mass and curvature is given by the Einstein equation, the fundamental field equation of general relativity.

In this paradigm, gravity is not a force: it is just a geometric effect produced by the interaction between mass and curvature. In other words, the reason a brick falls toward the Earth if you drop it is that as soon as you let go, it starts following a geodesic in spacetime, and the Earth's mass causes curvature that determines the shape of that geodesic: moving forward in time while moving closer to the Earth in space.

I should say a word about the appearance of the prefix "pseudo-" in the above paradigm, which places Einstein's theory slightly outside the realm of standard Riemannian geometry. As sketched above, a Riemannian metric on a manifold M is a choice for each point $p \in M$ of an inner product on the space of tangent vectors to M at p. As you know, an inner product $\langle\ ,\ \rangle$ on a real vector space V is a positive-definite bilinear form, implying in particular that it is

- **symmetric**: $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$;
- **nondegenerate**: For every $v \in V \backslash \{0\}$, there exists $w \in V$ such that $\langle v, w \rangle \neq 0$.

To define a pseudo-Riemannian metric on M, one adopts these two assumptions for the inner product $\langle \ , \ \rangle$ on the space of tangent vectors at every point $p \in M$, but without assuming any positivity, i.e. we do not require $\langle v, v \rangle$ to be positive whenever $v \neq 0$. The classification of quadratic forms (or equivalently the spectral theorem for symmetric linear maps) implies that any n-dimensional vector space V with a symmetric nondegenerate bilinear form $\langle \ , \ \rangle$ can be split into two orthogonal (with respect to $\langle \ , \ \rangle$) subspaces

$$V = V_+ \oplus V_-$$

such that $\langle \ , \ \rangle$ is positive-definite on V_+ and negative-definite on V_-. (Note that if both subspaces are nontrivial, then there always also exist nonzero vectors $v \in V$ such that $\langle v, v \rangle$ is zero—this does not contradict nondegeneracy!) The pseudo-Riemannian metrics used in general relativity have the property that on every tangent space, $\dim V_+ = 3$ and $\dim V_- = 1$.[7] Pseudo-Riemannian metrics with this property are also sometimes called **Lorentzian metrics**, and said to have **Lorentz signature**.

The canonical example of a Lorentzian inner product is what is called the **Minkowski metric** on $\mathbb{R}^4$: we define it by

$$(1.3) \qquad \langle x, y \rangle = -x^0 y^0 + \sum_{j=1}^{3} x^j y^j,$$

where we are following the physicists' convention of labeling vectors $v \in \mathbb{R}^4$ by their coordinates v^μ with $\mu = 0, 1, 2, 3$. It is actually crucial for Einstein's theory that the metric on spacetime is *not* positive-definite, because the Lorentzian signature is precisely what produces qualitative physical distinctions between the three spatial dimensions and the fourth one, time. In the convention used above to write down the Minkowski metric, time is labelled as the zeroth coordinate, and is thus distinguished by the minus sign appearing in (1.3). More generally, a vector v in a vector space V with a Lorentzian inner product $\langle \ , \ \rangle$ is called **time-like** if $\langle v, v \rangle < 0$, **space-like** if $\langle v, v \rangle > 0$, and **light-like** if $\langle v, v \rangle = 0$. With a bit of linear algebra, one can see that the set of all space-like vectors is connected, but the set of vectors that are time-like or light-like splits into two connected components, which we think of as representing motion *forward* or *backward in time*. Similarly, on a Lorentzian manifold, a geodesic can be either time-like, light-like or space-like, and in the first two categories one can distinguish between parametrizations of the geodesic that are oriented forward or backward in time, while for space-like geodesics there is no such distinction. The physical significance of these observations is the following: in general relativity, all particles with mass travel through spacetime along time-like geodesics, while particles with no mass travel along light-like geodesics—the latter are the particles that observers perceive as travelling at the speed of light. As far as we know, *nothing* travels along space-like geodesics, which is equivalent to saying that nothing travels faster than light. According to the geometry of spacetime, anything that *could* do this would also sometimes be observed to travel backward in time. Naturally, the non-existence of such particles according to the known laws of physics has not stopped physicists from giving them a name—tachyons—and they are mentioned frequently in science fiction, as a clearly necessary ingredient in time travel.

While we will probably not say anything further about general relativity in this course, we will prove some results about pseudo-Riemannian manifolds, and will try to avoid assuming that inner products are positive-definite unless that assumption is absolutely necessary.

1.1.4. *Gauge theory.* To round out this motivational introduction, I want to mention briefly another area of physics beyond general relativity where differential geometry plays a key role. The last half-century has witnessed intense and fruitful interactions between geometry and quantum

---

[7]Or possibly the other way around—the literature is not unanimous on this convention.

field theory (on which the theory of elementary particles is based), along with its more exotic and controversial cousin, string theory. Each of the classical fields underlying the various types of elementary particles can be described mathematically as a geometric object, namely a *section* of a *smooth fiber bundle*. The particles that mediate the electromagnetic, strong and weak nuclear forces, in particular, are described via so-called *gauge fields*, which are known to mathematicians as *connections*: these are a fundamental piece of geometric data on a fiber bundle, analogous to the Lorentzian metrics on the spacetime manifold of general relativity. This subject as a whole is known as *gauge theory*, a term which means slightly different things in the two fields: physicists understand it as the basis of their understanding of the forces of nature, while for mathematicians, it is a powerful framework for developing geometric and topological invariants based on spaces of solutions to nonlinear PDEs. In the big picture, gauge theory is both, and it has served as one of the most exciting sources of interactions between theoretical physics and pure mathematics during the past few decades. We will lay a few of the basic foundations for this subject via the study of vector bundles in the second half of this semester.

1.2. Charts and transition maps. We now begin the study of differential geometry in earnest.

The fundamental objects of study in this subject are called *smooth, finite-dimensional manifolds*. We will spend most of the first two lectures explaining the definition of this term and giving some basic examples.

We start with the intuition that a 1-dimensional manifold is what you have previously called a "curve" (*Kurve*), and a 2-dimensional manifold is a "surface" (*Fläche*). For arbitrary $n \in \mathbb{N}$, an elementary example of an n-dimensional manifold will be the so-called n-**sphere**

$$S^n := \left\{ x \in \mathbb{R}^{n+1} \mid |x| = 1 \right\},$$

where $|\cdot|$ again denotes the Euclidean norm. The word "sphere" (*Sphäre*) on its own normally refers to the familiar case $n = 2$, though it can also refer to the general case if the value of n is clear from context. The 1-sphere has been known to you since Kindergarten under a different name: the **circle** (*Kreis*). Let us examine this example a bit more closely, and clarify in particular the following point: S^1 is defined as a subset of $\mathbb{R}^2$, so why do we consider it a "one-dimensional" object?

The answer can be explained via an intelligent choice of coordinates. Consider the standard polar coordinates (r, θ) on $\mathbb{R}^2$, which are related to the Cartesian coordinates (x, y) by

$$x = r \cos \theta, \qquad y = r \sin \theta.$$

For concreteness, we assume (and will *always* assume) the angle θ is measured in radians, so the range $\theta \in [0, 2\pi]$ describes a full rotation. In polar coordinates, S^1 is the subset $\{r = 1\} \subset \mathbb{R}^2$, thus one of the coordinates becomes irrelevant, and having one coordinate left makes S^1 a one-dimensional object.

The above discussion of polar coordinates glossed over an important point: one cannot simultaneously describe *every* point in S^1 via a unique value of the angular coordinate $\theta \in \mathbb{R}$, at least not if we want the values of θ to be unambiguously defined and continuously dependent on the points that they describe. One could e.g. require θ to take values only in a half-open interval like $[0, 2\pi)$ or $(-\pi, \pi]$: this creates a one-to-one correspondence between points on S^1 and values of the coordinate, but the function one defines in this way from S^1 to $[0, 2\pi)$ or $(-\pi, \pi]$ has a jump discontinuity at the point where the coordinate reaches either end of the allowed interval. If you want to avoid such discontinuities, then the only option is to give up on the notion of describing *all* of S^1 in a single coordinate system, and instead use multiple coordinate systems defined on

different subsets. For instance, we could define two subsets of the circle by

$$\mathcal{U} := S^1 \backslash \{(1,0)\}, \qquad \mathcal{V} := S^1 \backslash \{(-1,0)\},$$

and associate to these two subsets two potentially different angular coordinates θ and ϕ respectively, each taking values in an appropriate open interval, thus defining *continuous* functions

$$\theta : \mathcal{U} \to (0, 2\pi), \qquad \phi : \mathcal{V} \to (-\pi, \pi).$$

Since $S^1 = \mathcal{U} \cup \mathcal{V}$, these two coordinate systems together can be used to describe every point in S^1. Moreover, there is a large region on which both coordinates θ and ϕ are defined: it consists of the two semi-circles $S^1_+ := \{(x,y) \in S^1 \mid y > 0\}$ and $S^1_- := \{(x,y) \in S^1 \mid y < 0\}$, and on each of these one can easily derive a relationship between θ and ϕ, namely

$$(1.4) \qquad \phi = \begin{cases} \theta & \text{on } S^1_+, \\ \theta - 2\pi & \text{on } S^1_-. \end{cases}$$

The pairs $(\mathcal{U}, \theta)$ and $(\mathcal{V}, \phi)$ are our first examples of what we will call *charts* on the 1-dimensional manifold S^1, and together they form a *smooth atlas* that determines a *smooth structure* on S^1. Let us now begin giving precise definitions to these terms.

In the following, assume M is a set, and $n \geqslant 0$ is an integer. For the sake of intuition, you may picture M as a surface (in which case $n = 2$), and picture the subsets $\mathcal{U}, \mathcal{V} \subset M$ as open subsets of that surface.[8] Recall that a continuous map defined on an open subset of Euclidean space is called **smooth** (*glatt*) if it admits derivatives of all orders.

DEFINITION 1.4. An n-dimensional **chart** (*Karte*)[9] $(\mathcal{U}, x)$ on M consists of a subset $\mathcal{U} \subset M$ and an injective map $x : \mathcal{U} \hookrightarrow \mathbb{R}^n$ whose image $x(\mathcal{U}) \subset \mathbb{R}^n$ is an open set.

Any two charts $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ determine a pair of **transition maps** (*Kartenübergänge*)

$$(1.5) \qquad \begin{aligned} \mathbb{R}^n \supset x(\mathcal{U} \cap \mathcal{V}) \xrightarrow{y \circ x^{-1}} y(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{R}^n, \\ \mathbb{R}^n \supset y(\mathcal{U} \cap \mathcal{V}) \xrightarrow{x \circ y^{-1}} x(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{R}^n, \end{aligned}$$

which are inverse to each other, and are thus bijections between subsets of $\mathbb{R}^n$. We say that the two charts are C^k-**compatible** (*verträglich*) for some $k \in \mathbb{N} \cup \{0, \infty\}$ if the sets $x(\mathcal{U} \cap \mathcal{V})$ and $y(\mathcal{U} \cap \mathcal{V})$ are both open and the transition maps $y \circ x^{-1}$ and $x \circ y^{-1}$ are both of class C^k. If $k = \infty$, we say the charts are **smoothly compatible** (*glatt verträglich*).

A picture of what a pair of overlapping charts on a surface might look like is shown in Figure 4. An individual chart $(\mathcal{U}, x)$ should be understood as defining a *coordinate system* for describing all points in the subset $\mathcal{U} \subset M$, where the individual **coordinates** (*Koordinaten*) are the n real-valued functions

$$x^1, \ldots, x^n : \mathcal{U} \to \mathbb{R}$$

defined as the component functions of the map $x = (x^1, \ldots, x^n) : \mathcal{U} \to \mathbb{R}^n$. Note that in Definition 1.4, it is permissible for the domains $\mathcal{U}$ and $\mathcal{V}$ of the two charts to be disjoint, in which case

---

[8]Saying the word "open" presumes that M has some structure beyond merely being an arbitrary set, e.g. it could be a subset of some Euclidean space $\mathbb{R}^n$, or more generally, a metric or topological space. We will address this point properly in the next lecture, but since we have not addressed it yet, Definition 1.4 refers to $\mathcal{U}$ and $\mathcal{V}$ simply as "subsets" of M, without saying they are open. In practice, they always will be.

[9]A word of caution for German speakers: the mathematical word *Abbildung* (as in "eine injektive Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^m$") can be translated into English as either "map" or "mapping", but do not be tempted to translate "map" into mathematical German as *Karte*. In mathematical English, a "chart" and a "map" are not exactly the same thing.

PSfrag replacements
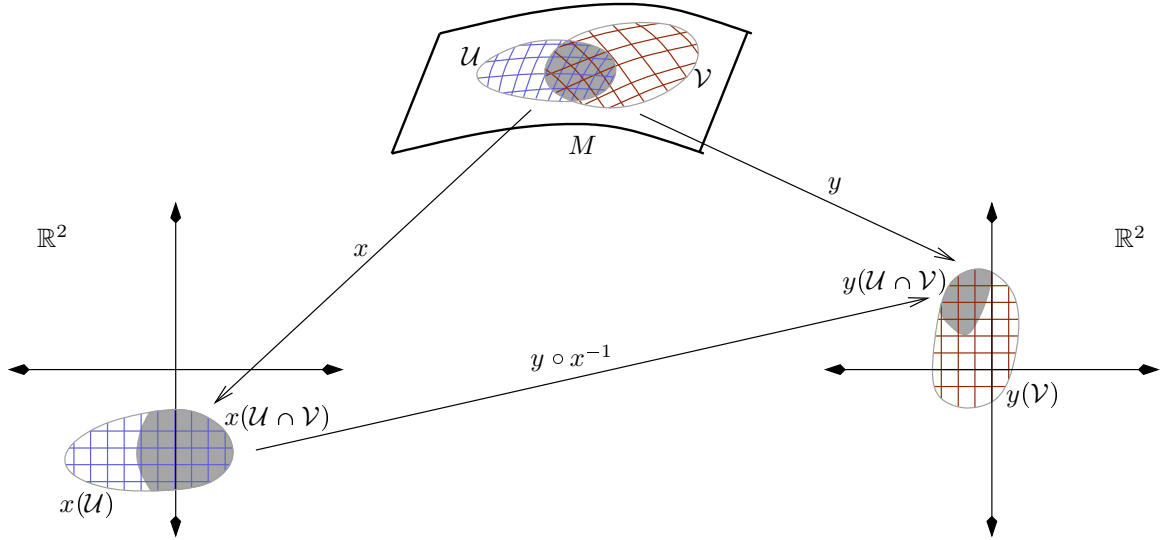
FIGURE 4.   Two charts $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ on a surface M, with an associated transition map $y \circ x^{-1}$ defining a bijection between two open sets (the shaded regions) in $\mathbb{R}^2$.

the transition maps $y \circ x^{-1}$ and $x \circ y^{-1}$ are both just the trivial map from the empty set to itself. But if $\mathcal{U} \cap \mathcal{V} \neq \varnothing$, then the transition map

$$\mathcal{U} \cap \mathcal{V}$$

$$\mathbb{R}^n \overset{\text{open}}{\supset} x(\mathcal{U} \cap \mathcal{V}) \xrightarrow{\ y \circ x^{-1}\ } y(\mathcal{U} \cap \mathcal{V}) \overset{\text{open}}{\subset} \mathbb{R}^n$$

defines a coordinate transformation, e.g. for any point $p \in \mathcal{U} \cap \mathcal{V}$, $y \circ x^{-1}$ sends the vector $(x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n$ that represents p in "x-coordinates" to the vector that represents the same point in "y-coordinates", namely $(y^1(p), \ldots, y^n(p)) \in \mathbb{R}^n$. It is often convenient in this situation to write the y-coordinates on the overlap region as functions of the x-coordinates, i.e. if we identify each point in $\mathcal{U} \cap \mathcal{V}$ with the vector in $\mathbb{R}^n$ determined by its x-coordinates, then the y-coordinates can be viewed as functions of n variables, which are naturally labelled $x^1, \ldots, x^n$, producing a transformation

(1.6) $$(x^1, \ldots, x^n) \mapsto (y^1(x^1, \ldots, x^n), \ldots, (y^n(x^1, \ldots, x^n)).$$

This is a slight abuse of notation, because in this expression, the variables $x^1, \ldots, x^n$ are no longer interpreted as real-valued functions on $\mathcal{U} \subset M$, but simply as the usual Cartesian coordinates on the open subset $x(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{R}^n$. With this understood, (1.6) is just another expression for the transition map $y \circ x^{-1}$, and the inverse transition map $x \circ y^{-1}$ can similarly be written as

(1.7) $$(y^1, \ldots, y^n) \mapsto (x^1(y^1, \ldots, y^n), \ldots, (x^n(y^1, \ldots, y^n)),$$

with the variables $y^1, \ldots, y^n$ now understood to represent Cartesian coordinates on $y(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{R}^n$. If the two charts are C^k-compatible, then both of the transformations in (1.6) and (1.7) are of

class C^k. If $k \geq 1$, then since the two transformations are inverse to each other, it follows that the n-by-n matrix with entries

$$\frac{\partial y^i}{\partial x^j}(x^1, \ldots, x^n), \qquad i, j \in \{1, \ldots, n\}$$

is invertible for every $(x^1, \ldots, x^n) \in x(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{R}^n$.

REMARK 1.5. You may have been accustomed to using *subscripts* $x_1, \ldots, x_n$ for coordinates on $\mathbb{R}^n$ in your studies up to this point, and will thus wonder why I am instead using superscripts in all the expressions above. This is not an arbitrary choice—it is a convention that is widespread in differential geometry, and especially popular among physicists, and we will try to use it consistently throughout this course. Subscripts will at some point also appear, but they will have a different meaning.

EXAMPLE 1.6. In the discussion of the unit circle S^1 above, we defined two charts $(\mathcal{U}, \theta)$ and $(\mathcal{V}, \phi)$, with images $\theta(\mathcal{U}) = (0, 2\pi) \subset \mathbb{R}$ and $\phi(\mathcal{V}) = (-\pi, \pi) \subset \mathbb{R}$. The overlap region $\mathcal{U} \cap \mathcal{V}$ of these two charts is the union of two disjoint open sets that we denoted by S^1_+ and S^1_-, the upper and lower semicircle (disjoint from the x-axis). The transition map $\phi \circ \theta^{-1} : \theta(S^1_+ \cup S^1_-) \to \phi(S^1_+ \cup S^1_-)$ is then found by writing ϕ as a function of θ as in (1.4), which gives

$$\phi(\theta) = \begin{cases} \theta & \text{for } 0 < \theta < \pi, \\ \theta - 2\pi & \text{for } \pi < \theta < 2\pi. \end{cases}$$

Observe that while this map appears at first glance to have a jump discontinuity, its actual domain is $\theta(S^1_+ \cup S^1_-) = (0, \pi) \cup (\pi, 2\pi)$, i.e. it excludes the point π at which the discontinuity would occur. As a result, this transition map is smooth, and so is its inverse; the two charts $(\mathcal{U}, \theta)$ and $(\mathcal{V}, \phi)$ are therefore smoothly compatible.

EXERCISE 1.7. The standard **spherical coordinates** (*Kugelkoordinaten*) on $\mathbb{R}^3$ are defined via the transformation

$$(1.8) \qquad (r, \theta, \phi) \mapsto (x, y, z), \qquad \begin{cases} x & := r \cos\theta \cos\phi, \\ y & := r \sin\theta \cos\phi, \\ z & := r \sin\phi, \end{cases}$$

where θ plays the role of an angle in the xy-plane, and $\phi \in [-\pi/2, \pi/2]$ is the angle between the vector $(x, y, z) \in \mathbb{R}^3$ and the xy-plane.[10] Restricting to $r = 1$, the other two coordinates (θ, ϕ) can be used to describe points on the unit sphere $S^2 \subset \mathbb{R}^3$, though there are choices to be made since θ is only defined up to multiples of 2π (and it is not defined at all at the north and south poles $p_\pm := (0, 0, \pm 1) \in S^2$, where $\phi = \pm\pi/2$.)

   (a) Find two subsets $\mathcal{U}_1, \mathcal{U}_2 \subset S^2$ with $\mathcal{U}_1 \cup \mathcal{U}_2 = S^2 \backslash \{p_+, p_-\}$ such that for $i = 1, 2$, there are 2-dimensional charts of the form $(\mathcal{U}_i, \alpha_i)$ with $\alpha_i = (\theta_i, \phi_i)$, where the coordinate functions $\theta_i, \phi_i : \mathcal{U}_i \to \mathbb{R}$ are continuous and satisfy the spherical coordinate relations (1.8), and have images $\alpha_1(\mathcal{U}_1) = (0, 2\pi) \times (-\pi/2, \pi/2) \subset \mathbb{R}^2$ and $\alpha_2(\mathcal{U}_2) = (-\pi, \pi) \times (-\pi/2, \pi/2) \subset \mathbb{R}^2$.
   (b) One cannot use spherical coordinates to construct a chart on S^2 that contains either of the poles $p_\pm = (0, 0, \pm 1)$. Can you think of another way to construct charts on open subsets of S^2 that contain these two points?

---

[10]Achtung: there are various conventions for spherical coordinates in use. I'm told that this is the standard convention learned by mathematics students in Germany. I learned a different convention as a physics student in the U.S.: $x = r \cos\phi \sin\theta$, $y = r \sin\phi \sin\theta$, $z = r \cos\theta$. Here ϕ plays the role of the angle in the xy-plane, and $\theta \in [0, \pi]$ is the angle between $(x, y, z) \in \mathbb{R}^3$ and the positive z-axis.

Hint: On any sufficiently small neighborhood of p_+ or p_- in S^2, every point has its z-coordinate determined by the x and y-coordinates.

(c) Now that you've constructed charts that cover every point on S^2, write down the associated transition maps and show that your charts are all smoothly compatible with each other.

2. Smooth manifolds

In this lecture we give the definition of the term *smooth manifold* and look at a few more examples.

2.1. Atlases and smooth structures. We concluded Lecture 1 by defining the notion of a *chart* on a set M, and C^k-compatibility between two charts. A chart $(\mathcal{U}, x)$ should be interpreted as a "local" coordinate system, which can be used to label points in the subset $\mathcal{U} \subset M$. We saw in the example of the circle S^1 that while one cannot apparently describe *all* points in S^1 via a single chart, it was easy to find two smoothly compatible charts such that every point is in at least one or the other. Exercise 1.7 similarly outlines how to cover S^2 with four charts using spherical coordinates. These were the first examples of the following general concept.

DEFINITION 2.1. An **atlas of class** C^k for the set M (or **smooth atlas** in the case $k = \infty$) is a collection of charts $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ that are all C^k-compatible with each other, such that $\bigcup_{\alpha \in I} \mathcal{U}_\alpha = M$.[11]

In first-year analysis, you learned what it means for a real-valued function on an open subset of $\mathbb{R}^n$ to be differentiable; it was important in that definition that the domain of the function should be *open*, as differentiation at a point p involves limits that are not well defined unless f itself is defined on some ball around p. In differential geometry, we would also like to be able to differentiate functions

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

defined on a manifold M, such as the circle S^1 or sphere S^2. This is a nontrivial problem, even in simple examples such as S^n that are given as subsets of Euclidean space, since they are not generally *open* subsets. But if M is a set equipped with an atlas, then M is covered by subsets that have coordinate systems, so for each chart $(\mathcal{U}, x)$ we can write down f "in local coordinates", meaning we identify each point $p \in \mathcal{U}$ with its coordinate vector $(x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n$, so that $f|_\mathcal{U} : \mathcal{U} \to \mathbb{R}$ becomes a function of n real variables

$$(2.1) \qquad\qquad (x^1, \ldots, x^n) \mapsto f(x^1, \ldots, x^n),$$

with $x^1, \ldots, x^n$ interpreted as the standard Cartesian coordinates on the open set $x(\mathcal{U}) \subset \mathbb{R}^n$. This is another slight abuse of notation, similar to the coordinate expressions for transition maps described in (1.6) and (1.7); in fact, the function that is literally described in (2.1) is not $f : M \to \mathbb{R}$ but rather

$$x(\mathcal{U}) \xrightarrow{f \circ x^{-1}} \mathbb{R}.$$

It now seems natural to say that f is differentiable at $p \in \mathcal{U} \subset M$ if and only if its coordinate expression $f \circ x^{-1}$ is differentiable (in the sense of first-year analysis) at the corresponding point $x(p) \in x(\mathcal{U}) \subset \mathbb{R}^n$. For this to be a reasonable definition, we need to know that it does not depend on the *choice* of the chart $(\mathcal{U}, x)$, as our atlas may indeed contain multiple distinct charts that contain the point p. This issue is precisely what the compatibility condition in Definition 1.4 was designed to settle:

---

[11]In this definition, I may be any set, finite, countable or uncountable. We refer to it as an **index set** since it is only used for labelling purposes and is otherwise unimportant in itself.

LEMMA 2.2. *Suppose $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ are two C^k-compatible charts on M, and $f : M \to \mathbb{R}$ is a function. Then for each nonnegative integer $r \leqslant k$, the function $x(\mathcal{U} \cap \mathcal{V}) \xrightarrow{f \circ x^{-1}} \mathbb{R}$ is of class C^r if and only if the function $y(\mathcal{U} \cap \mathcal{V}) \xrightarrow{f \circ y^{-1}} \mathbb{R}$ is of class C^r.*

PROOF. The statement follows from the chain rule, since $f \circ y^{-1} = (f \circ x^{-1}) \circ (x \circ y^{-1})$ and $f \circ x^{-1} = (f \circ y^{-1}) \circ (y \circ x^{-1})$. □

DEFINITION 2.3. *For a set M with an atlas $\mathcal{A}$ of class C^k and $r \in \mathbb{N} \cup \{0, \infty\}$ with $r \leqslant k$, a function $f : M \to \mathbb{R}$ is said to be **of class** C^r if and only if the function $x(\mathcal{U}) \xrightarrow{f \circ x^{-1}} \mathbb{R}$ is of class C^r for every chart $(\mathcal{U}, x) \in \mathcal{A}$.*

EXERCISE 2.4. Convince yourself that Lemma 2.2 becomes false in general if one allows $r > k$. (See also Example 2.7 below for a concrete special case.) This has the following consequence: if we want to define what it means for a function on a manifold to be of class C^k, then we need to have an atlas of class C^k *or better* to test it with. In particular, the notion of smooth functions on M cannot be defined unless M is equipped with a *smooth* atlas.

The examples of smooth atlases we saw in Lecture 1 on S^1 and S^2 were finite, and this will turn out to be a general pattern: we will see that almost all manifolds we are interested in admit finite atlases, though it is not often important to know this. On the other hand, a general atlas can be uncountably infinite, and one can always enlarge a finite atlas $\{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ in trivial ways, e.g. by choosing subsets $\mathcal{U}'_\alpha \subset \mathcal{U}_\alpha$ for which $x_\alpha(\mathcal{U}'_\alpha) \subset \mathbb{R}^n$ is open and adding in the restricted charts $(\mathcal{U}'_\alpha, x_\alpha|_{\mathcal{U}'_\alpha})$, which are obviously still compatible with all the others. We say that an atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ of class C^k is **maximal** if it cannot be enlarged any further without sacrificing compatibility, i.e. every chart that is C^k-compatible with all of the charts in $\mathcal{A}$ already belongs to $\mathcal{A}$.

LEMMA 2.5. *Given an atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ of class C^k on M, let $\mathcal{A}'$ denote the collection of all charts on M that are C^k-compatible with all the charts in $\mathcal{A}$. Then $\mathcal{A}'$ is a maximal atlas of class C^k, and it is the only one containing $\mathcal{A}$.*

PROOF. To show that $\mathcal{A}'$ is an atlas, we need to show that any two charts $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ that are C^k-compatible with every $(\mathcal{U}_\alpha, x_\alpha)$ are also C^k-compatible with each other. Given a point $p \in \mathcal{U} \cap \mathcal{V}$, pick $\alpha \in I$ so that $p \in \mathcal{U}_\alpha$. The set $x(\mathcal{U} \cap \mathcal{V} \cap \mathcal{U}_\alpha) \subset \mathbb{R}^n$ is then the intersection of the two open sets $x(\mathcal{U} \cap \mathcal{U}_\alpha)$ and $x(\mathcal{V} \cap \mathcal{U}_\alpha)$ and is thus an open neighborhood of $x(p)$, so on this neighborhood, the transition map $y \circ x^{-1}$ can then be written as

$$y \circ x^{-1} = (y \circ x_\alpha^{-1}) \circ (x_\alpha \circ x^{-1}),$$

which is a composition of two C^k-maps and is therefore of class C^k on the neighborhood of $x(p)$ in question. This trick works (possibly with different choices of α) for any point $p \in \mathcal{U} \cap \mathcal{V}$, and it also works for the inverse transition map $x \circ y^{-1}$, thus it implies that both of the transition maps relating x and y are everywhere of class C^k, and $\mathcal{A}'$ is therefore an atlas. It clearly also contains $\mathcal{A}$, and it is maximal, since any chart compatible with every chart in $\mathcal{A}'$ is also compatible with every chart in $\mathcal{A}$, and thus belongs to $\mathcal{A}'$ by definition. Finally, if $\mathcal{A}''$ is any other atlas containing $\mathcal{A}$, then every chart in $\mathcal{A}''$ is compatible with every chart in $\mathcal{A} \subset \mathcal{A}''$ and therefore belongs to $\mathcal{A}'$ by definition, proving $\mathcal{A}'' \subset \mathcal{A}'$. If $\mathcal{A}''$ is also maximal, it follows that $\mathcal{A}'' = \mathcal{A}'$. □

DEFINITION 2.6. *For $k \in \mathbb{N} \cup \{\infty\}$, a C^k-**structure** (C^k-*Struktur*) or **differentiable structure of class** C^k (*differenzierbare Struktur von der Klasse* C^k) on a set M is a maximal atlas $\mathcal{A}$ of class C^k on M. In the case $k = \infty$, we also call this a **smooth structure** (*glatte Struktur*) on M. If M has been endowed with a C^k-structure $\mathcal{A}$, then a chart $(\mathcal{U}, x)$ on M will be referred to as a C^k-**chart** (or a **smooth chart** in the case $k = \infty$) if it belongs to the maximal atlas $\mathcal{A}$.*

The maximality condition in Definition 2.6 is convenient for bookkeeping purposes (see Remark 2.8 below), but Lemma 2.5 shows that it is not a meaningful restriction. In practice, one typically specifies a smooth structure by first describing the smallest atlas one is able to construct, and then replacing it with its unique maximal extension. We will usually carry out the latter step without even mentioning it.

EXAMPLE 2.7. The following defines an atlas of class C^0 but not C^1 on $\mathbb{R}$: consider two charts $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ with

$$\mathcal{U} := (-\infty, 1), \qquad x(t) := t,$$
$$\mathcal{V} := (-1, \infty), \qquad y(t) := t^3.$$

The resulting transition maps both send $(-1, 1) \to (-1, 1)$ and are given by

$$y(x) = x^3, \qquad x(y) = \sqrt[3]{y},$$

so both are continuous, but $x \circ y^{-1}$ is not differentiable. This has the consequence that functions $\mathbb{R} \to \mathbb{R}$ that look differentiable in the x-coordinate might not look differentiable in the y-coordinate. An easy example is the identity map $f(t) = t$, which looks like $f(x) = x$ and is thus smooth in the x-coordinate, but its expression in the y-coordinate is $f(y) = \sqrt[3]{y}$, which fails to be differentiable at the point $0 \in y(\mathcal{V}) = (-1, \infty)$.

Note that if we enlarge both $\mathcal{U}$ and $\mathcal{V}$ to $\mathbb{R}$, then while the two charts $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ together do not determine any smooth structure on $\mathbb{R}$, each of these charts individually forms a smooth atlas—an atlas with only one chart is always smooth since it has no nontrivial transition maps whose differentiability would need to be checked. Each therefore determines a smooth structure via Lemma 2.5, and in this way, one obtains two *different* smooth structures on $\mathbb{R}$.

REMARK 2.8. The advantage of requiring maximality in Definition 2.6 is the following: if $\mathcal{A}$ and $\mathcal{A}'$ are two atlases on M for which every chart in $\mathcal{A}$ is compatible with every chart in $\mathcal{A}'$, then the two notions of differentiability for functions on M defined via these two atlases will be the same, and we would therefore prefer to think of them is defining the *same* smooth structure, even if they are different atlases, strictly speaking. In this scenario, it is easy to check that both atlases do in fact have the same maximal extension.

2.2. Some topological notions. With the concept of a smooth atlas in hand, a reasonable guess for the "right" definition of a smooth manifold would be that it is any set endowed with the additional structure of a smooth atlas. In practice, however, doing anything interesting with manifolds requires imposing one or two further restrictions on what is allowed to be a manifold and what is not.

I do not want to assume previous knowledge of topology in this course, but a few basic notions of the subject now need to be discussed before we can give the precise definition of a manifold. Most of them will play a negligible role in this course, and in fact, the intuition you already have about metric spaces is fully sufficient for understanding the definition of a manifold (cf. Remark 2.20 below)—nonetheless, you will not be able to understand *why* that definition is what it is unless we first discuss the alternatives.

Since you have seen metric spaces before, you know how to define fundamental notions such as **continuity** (*Stetigkeit*), **convergence** of a sequence to a point (*Konvergenz einer Folge gegen einen Punkt*) and **closed** sets (*abgeschlossene Teilmengen*) in metric spaces. You will also have seen important concepts such as that of a **neighborhood** (*Umgebung*) of a point $x \in X$, meaning any subset $\mathcal{U} \subset X$ that contains an open subset containing x, and probably also a **homeomorphism** (*Homöomorphismus*), which is a continuous bijection whose inverse is also continuous. One detail you may or may not already be aware of is that all of these notions can be defined without any explicit reference to a metric, so long as one knows what an "open set" is. In particular:

PROPOSITION 2.9 (first-year analysis). *Assume X and Y are metric spaces.*

(1) *A sequence $x_n \in X$ converges to a point $x \in X$ if and only if for every neighborhood $\mathcal{U} \subset X$ of x, $x_n \in \mathcal{U}$ for all sufficiently large n.*
(2) *A subset $\mathcal{U} \subset X$ is closed if and only if its complement $X \backslash \mathcal{U} \subset X$ is open.*
(3) *A map $f : X \to Y$ is continuous if and only if for every open subset $\mathcal{U} \subset Y$, $f^{-1}(\mathcal{U}) := \{x \in X \mid f(x) \in \mathcal{U}\}$ is an open subset of X.*
(4) *A bijective map $f : X \to Y$ is a homeomorphism if and only if it defines a bijective correspondence between the open subsets of X and the open subsets of Y, i.e. for all subsets $\mathcal{U} \subset X$, $\mathcal{U}$ is open if and only if $f(\mathcal{U}) \subset Y$ is open.*

$\square$

EXERCISE 2.10. If you do not already find Proposition 2.9 obvious, prove it.

Topology begins with the observation that it is sometimes convenient to define what an open set is without the aid of a metric. For this idea to be useful, we just need open sets to satisfy a few properties that are already familiar from the theory of metric spaces:

DEFINITION 2.11. A **topology** (*Topologie*) on a set X is a collection $\mathcal{T}$ of subsets of X satisfying the following axioms:

(i) $\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$;
(ii) For every subcollection $I \subset \mathcal{T}$, $\bigcup_{\mathcal{U} \in I} \mathcal{U} \in \mathcal{T}$;
(iii) For every pair $\mathcal{U}_1, \mathcal{U}_2 \in \mathcal{T}$, $\mathcal{U}_1 \cap \mathcal{U}_2 \in \mathcal{T}$.

The pair $(X, \mathcal{T})$ is then called a **topological space** (*topologischer Raum*), and we call the sets $\mathcal{U} \in \mathcal{T}$ the **open** subsets (*offene Teilmengen*) in $(X, \mathcal{T})$.

We will usually not give an actual label to the topology when discussing a topological space, so e.g. instead of talking about $(X, \mathcal{T})$, we will talk about "the topological space X" with the understanding that a subset $\mathcal{U} \subset X$ is called "open" if and only if it belongs to the topology that has been specified on X. For topological spaces X and Y, one now takes the statements in Proposition 2.9 as *definitions* of the notions of convergence, closed subsets, continuity and homeomorphisms.

We call a topological space X **metrizable** (*metrisierbar*) if it admits a metric for which the given topology of X consists of all sets that are unions of open balls, i.e. the metrizable spaces are the topological spaces that you already saw (but without using the word "topology") when you studied metric spaces. Two things about this notion are important to understand:

(1) If X is metrizable, then the metric that defines its topology is typically far from being unique. For example, $d(x, y) := c|x - y|$ for any constant $c > 0$ defines a "nonstandard" metric on $\mathbb{R}$ that nonetheless induces the same topology as the standard one.
(2) Many topological spaces are not metrizable, and they can easily have properties that are counterintuitive. (We will see an example in a moment.)

We saw in §2.1 that an atlas of class C^k on a set M determines a natural way to define what it means for a function $f : M \to \mathbb{R}$ to be of class C^r for any $r \leqslant k$. This holds in particular for $r = 0$, so that continuity of functions can be defined in a certain sense, even though we never explicitly endowed M with a topology. But actually, we did, we just didn't notice:

PROPOSITION 2.12. *Given an atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ of class C^0 on a set M, there exists a unique topology on M such that the sets $\mathcal{U}_\alpha \subset M$ are all open and the maps x_α are all homeomorphisms onto their images.[12] Moreover, for every other chart $(\mathcal{U}, x)$ that is C^0-compatible with the charts in $\mathcal{A}$, $\mathcal{U} \subset M$ is also open and x is also a homeomorphism onto its image.*

PROOF. Suppose M carries a topology with the properties described, and $\mathcal{O} \subset M$ is an open subset. Then each of the sets $\mathcal{O}_\alpha := \mathcal{O} \cap \mathcal{U}_\alpha$ is open, and $\mathcal{O} = \bigcup_{\alpha \in I} \mathcal{O}_\alpha$. Since each x_α is a homeomorphism onto its image in $\mathbb{R}^n$, $x_\alpha(\mathcal{O}_\alpha)$ is then also an open subset of $\mathbb{R}^n$. Conversely, if $\mathcal{O} \subset M$ is *any* subset such the sets $\Omega_\alpha := x_\alpha(\mathcal{O} \cap \mathcal{U}_\alpha) \subset \mathbb{R}^n$ are all open, then each $\mathcal{O}_\alpha := \mathcal{O} \cap \mathcal{U}_\alpha = x_\alpha^{-1}(\Omega_\alpha) \subset M$ must also be open since x_α is a homeomorphism, and therefore so is the union $\mathcal{O} = \bigcup_{\alpha \in I} \mathcal{O}_\alpha$. This proves that a topology with the stated properties is unique: if it exists, then it is precisely the collection of all subsets $\mathcal{O} \subset M$ such that $x_\alpha(\mathcal{O} \cap \mathcal{U}_\alpha) \subset \mathbb{R}^n$ is open for every $\alpha \in I$.

To prove existence, one now has to prove that the collection of subsets of M described above satisfies the axioms of a topology, i.e. it contains M and $\varnothing$ and is closed under arbitrary unions and finite intersections. This is a straightforward exercise.

Finally, let us fix the topology on M described above and suppose $(\mathcal{U}, x)$ is another chart that is C^0-compatible with $(\mathcal{U}_\alpha, x_\alpha)$ for every $\alpha \in I$. We need to show that $\mathcal{U} \subset M$ is open and $x : \mathcal{U} \to \mathbb{R}^n$ is a homeomorphism onto its image, which is equivalent to showing that for subsets $\mathcal{O} \subset \mathcal{U}$, $\mathcal{O}$ is open in M if and only if $x(\mathcal{O})$ is open in $\mathbb{R}^n$. For this, we make use of the transition maps relating $(\mathcal{U}, x)$ and $(\mathcal{U}_\alpha, x_\alpha)$ for an arbitrary choice of $\alpha \in I$:

$$
\begin{array}{ccc}
& \mathcal{U} \cap \mathcal{U}_\alpha & \\
& \overset{x}{\swarrow} \qquad \overset{x_\alpha}{\searrow} & \\
\mathbb{R}^n \overset{\text{open}}{\supset} x(\mathcal{U} \cap \mathcal{U}_\alpha) & \underset{x \circ x_\alpha^{-1}}{\overset{x_\alpha \circ x^{-1}}{\rightleftarrows}} & x_\alpha(\mathcal{U} \cap \mathcal{U}_\alpha) \overset{\text{open}}{\subset} \mathbb{R}^n
\end{array}
$$

By the assumption of C^0-compatibility, the two maps in the bottom row of this diagram are both continuous, and since they are inverse to each other, they are homeomorphisms, meaning they define a bijection between the open subsets of $x(\mathcal{U} \cap \mathcal{U}_\alpha)$ and $x_\alpha(\mathcal{U} \cap \mathcal{U}_\alpha)$. Now suppose $\mathcal{O} \subset M$ is open, which means $x_\alpha(\mathcal{O} \cap \mathcal{U}_\alpha) \subset x_\alpha(\mathcal{U} \cap \mathcal{U}_\alpha) \subset \mathbb{R}^n$ is open for every α. Feeding this set into the homeomorphism $x \circ x_\alpha^{-1}$ gives $x(\mathcal{O} \cap \mathcal{U}_\alpha)$, proving that the latter is an open set, and therefore so is $x(\mathcal{O}) = \bigcup_{\alpha \in I} x(\mathcal{O} \cap \mathcal{U}_\alpha)$. Conversely, if $\mathcal{O} \subset M$ is an arbitrary subset such that $x(\mathcal{O})$ is open, then for every $\alpha \in I$, $x(\mathcal{O} \cap \mathcal{U}_\alpha)$ is the intersection of two open sets $x(\mathcal{O})$ and $x(\mathcal{U} \cap \mathcal{U}_\alpha)$, and is thus also open. Feeding it into $x_\alpha \circ x^{-1}$ then shows that $x_\alpha(\mathcal{O} \cap \mathcal{U}_\alpha)$ is also open, proving that $\mathcal{O} \subset M$ is open. $\square$

Whenever we discuss a set M with an atlas $\mathcal{A}$ from now on, we will assume that M is endowed with the topology described in Proposition 2.12.

REMARK 2.13. Notice that according to the last statement in Proposition 2.12, the topologies induced on M by $\mathcal{A}$ or any extension of $\mathcal{A}$ to a larger (e.g. maximal) atlas are the same.

REMARK 2.14. It is rarely actually necessary to apply Proposition 2.12 for defining a topology on a manifold. The much more common situation is that our manifold M comes equipped with some natural topology that is clear from the context (e.g. because M is a subset or quotient of $\mathbb{R}^n$ or some other manifold that we already understand), and when specifying an atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ for M,

---

[12]Recall that $x_\alpha(\mathcal{U}_\alpha)$ is an open subset of a Euclidean space $\mathbb{R}^n$, so it is understood in this statement to carry the obvious topology that it inherits from the Euclidean metric on $\mathbb{R}^n$.

we just need to check that the topology determined by the atlas is the same as the natural topology. In other words, we need to check that the sets $\mathcal{U}_\alpha$ are open and the maps $x_\alpha : \mathcal{U}_\alpha \to x_\alpha(\mathcal{U}_\alpha) \subset \mathbb{R}^n$ are all homeomorphisms with respect to the natural topology. In most situations, this will be obvious.

EXERCISE 2.15. We now have two ways of defining what it means for a function $f : M \to \mathbb{R}$ to be continuous: one is the case $k = 0$ of Definition 2.3, in terms of the atlas $\mathcal{A}$, and the other is the standard notion of continuity in topological spaces, using the topology determined by $\mathcal{A}$ according to Proposition 2.12. Convince yourself that these two definitions are equivalent.

Since the atlas identifies small neighborhoods in M with neighborhoods in Euclidean space, and the topology of Euclidean space is pleasantly familiar to us, one might intuitively expect the topology induced on M by $\mathcal{A}$ to have similarly pleasant properties. The next example shows that this intuition is wrong.

EXAMPLE 2.16. Define an equivalence relation $\sim$ on the set $\widetilde{M} := \mathbb{R} \times \{0, 1\}$ such that every element is equivalent to itself and $(t, 0) \sim (t, 1)$ for all $t \in \mathbb{R}\backslash\{0\}$, but not for $t = 0$. Let
$$M := \widetilde{M}/\sim$$
denote the set of equivalence classes. We can think of M intuitively as a "real line with two zeroes", because it mostly looks just the same as $\mathbb{R}$ (each number $t \neq 0$ corresponding to the equivalence class of $(t, 0)$ and $(t, 1)$), but $t = 0$ is an exception, where there really are *two* distinct points $[(0, 0)]$ and $[(0, 1)]$ in M. The following pair of 1-dimensional charts define a smooth atlas on M: let
$$\mathcal{U}_\alpha := \left\{ [(t, 0)] \in M \mid t \in \mathbb{R} \right\}, \qquad \mathcal{U}_\beta := \left\{ [(t, 1)] \in M \mid t \in \mathbb{R} \right\},$$
and define both $x_\alpha : \mathcal{U}_\alpha \to \mathbb{R}$ and $x_\beta : \mathcal{U}_\beta \to \mathbb{R}$ by $[(t, k)] \mapsto t$ for $k = 0, 1$. The transition maps relating these two charts are both the identity map on $\mathbb{R}\backslash\{0\}$, thus the charts are smoothly compatible, and clearly $M = \mathcal{U}_\alpha \cup \mathcal{U}_\beta$.

Now consider the sequence
$$p_j := [(1/j, 0)] \in M.$$
Does it converge? We need to think for a moment about what convergence means in the topology induced by an atlas: if $p \in \mathcal{U}_\alpha$, then since x_α is a homeomorphism onto its image, p_j will converge to p if and only if $x_\alpha(p_j)$ converges to $x_\alpha(p)$ in $\mathbb{R}$, and a moment's thought reveals that that condition holds for $p := [(0, 0)]$. However, if we use the *other* chart x_β, then since $(1/j, 0) \sim (1/j, 1)$ for every j, the same condition also holds for the point $p' := [(0, 1)] \in \mathcal{U}_\beta$, and we have thus found two *distinct* points $p \neq p'$ such that $p_j \to p$ and $p_j \to p'$.

This seems like a contradiction if you have not seen any topology before, but it is not: it merely shows that M is a much stranger topological space than our intuition about metric spaces had led us to expect. In fact, the points p and p' have the peculiar property that every neighborhood of p intersects every neighborhood of p', so even though they are distinct points, the topology of M does not "separate" them; the technical term for this is that the topology of M is not **Hausdorff**.[13]

We do not want our notion of manifolds to include pathological examples in which a sequence can converge to two distinct points at once. Among other issues, it would clearly be impossible to define a metric compatible with that notion of convergence, as the triangle inequality ensures that limits of sequences are unique in metric spaces. Since the notion of distance on manifolds is one of the main things we plan to study when we get further into this subject, we would like to have a guarantee that every manifold *admits* a metric that is compatible with its natural topology,

---

[13]Or, as my topology professor in grad school once put it, the points p and p' are not "housed off" from each other. The proper delivery of this joke requires a Brooklyn accent.

i.e. we will insist that all manifolds be metrizable. This condition will turn out to have many advantages beyond the study of distance, though we will rarely need to make explicit use of it: it will only become important when we discuss the construction of global geometric structures (such as Riemannian metrics) via partitions of unity.

Although it will play no significant role in this course, we need one more topological notion in order to understand the main definition: a topological space is called **separable** (*separabel*) if it contains a countable dense subset. Euclidean spaces, for example, are separable, because $\mathbb{Q}^n \subset \mathbb{R}^n$ is a countable dense subset. Every space of interest in this course will be separable, and one can often use the result of the following exercise to prove it.

EXERCISE 2.17. Show that every subset of a separable metric space (X, d) is also a separable metric space.
Hint: Given a countable dense subset $E \subset X$ and another subset $Y \subset X$, show first that every open set in X is a union of open balls of the form $B_r(x) := \{y \in X \mid d(y, x) < r\}$ for $x \in E$ and $r \in \mathbb{Q}$. (This depends on the density of E.) Then define $E_0 \subset Y$ to consist of exactly one element from each of the sets $B_r(x) \cap Y$ for $x \in E$ and $r \in \mathbb{Q}$, whenever those sets are nonempty. Show that E_0 is countable and dense in Y.

2.3. The definition of a manifold. Hopefully you now have sufficient motivation to accept the following definition.

DEFINITION 2.18. Assume $k \in \mathbb{N} \cup \{\infty\}$. A **differentiable manifold of class** C^k (*differenzierbare Mannigfaltigkeit von der Klasse C^k*) or C^k-**manifold** (*C^k-Mannigfaltigkeit*) is a set M endowed with a C^k-structure (see Definition 2.6) such that the induced topology on M is metrizable and separable. In the case $k = \infty$, we also call M a **smooth manifold** (*glatte Mannigfaltigkeit*). We say that M is n-**dimensional** and refer to M as an n-**manifold**, written

$$\dim M = n,$$

if every chart in its differentiable structure is n-dimensional.[14]

REMARK 2.19. For the purposes of this course, you are essentially free to ignore the separability condition in Definition 2.18, as nothing in our study of differential geometry will truly depend on it. An example of something that satisfies every condition in the definition except separability would be the disjoint union of *uncountably* many copies of a manifold (see §2.4.3 below for more on disjoint unions); in fact, one can show that the condition on separability in our definition is equivalent to requiring M to have at most countably many connected components. One does sometimes need to know this for important results in differential *topology*, e.g. there is a theorem guaranteeing that every smooth n-manifold M can be embedded as a smooth submanifold of $\mathbb{R}^{2n+1}$, and this would clearly contradict Exercise 2.17 if M were not separable. (This issue is related to the second countability axiom—see Remark 2.21.)

REMARK 2.20. If you prefer never to think about topological spaces, then you can read Definition 2.18 as saying that a manifold M is a separable metric space endowed with an atlas $\{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ for which the sets $\mathcal{U}_\alpha \subset M$ are open and the bijections $x_\alpha : \mathcal{U}_\alpha \to x_\alpha(\mathcal{U}_\alpha) \subset \mathbb{R}^n$ are continuous with continuous inverses. Calling M a "metric space" comes however with the following caveat: while the *existence* of a suitable metric on M is an important condition, the *choice* of metric on M is not considered a part of its intrinsic structure, i.e. you are free to replace it with any other metric that has the above properties with respect to the atlas. This is why we have used the word "metrizable" in Definition 2.18 instead of just calling M a "metric space".

---

[14]Note that in our general definition of a manifold, M might admit multiple charts of different dimensions. One can show however that each individual connected component of M is itself a manifold with a uniquely defined dimension. For this reason we will usually only consider manifolds that have a well-defined dimension.

REMARK 2.21. For students who have seen some topology, the more standard definition of a manifold found in many textbooks would replace the conditions of metrizability and separability with the conditions that M is *Hausdorff* and *second countable*. This gives an equivalent definition, though proving this equivalence would require more of a digression into point-set topology than we have space for here; the details can (mostly) be found in [**Lee11**, Chapter 2].

REMARK 2.22. Another reasonable guess for a good definition of a manifold would be to drop metrizability and separability from Definition 2.18 but still require M to be Hausdorff (thus excluding things like Example 2.16). It turns out that this also does not include enough conditions to rule out some pathological behavior. The issue here is that a locally Euclidean Hausdorff space may fail to be *paracompact*, in which case the construction of basic geometric objects like Riemannian metrics becomes impossible. (We will discuss paracompactness and its applications later in the course.) If you have some topological background and would like to see some examples of the kinds of pathological behavior I'm talking about, see the discussion of the *long line* and *Prüfer surface* in [**Wen18**, Lecture 18].

In this course, we will almost always consider only the case $k = \infty$ of Definition 2.18, so that we speak of *smooth* manifolds. Actually, a large portion of differential geometry still makes sense for C^1-manifolds, though the important notion of *curvature* on a Riemannian manifold depends on second derivatives of the metric, and thus only makes sense on manifolds of class C^2. In either case, one has to be very careful in every proof so as not to differentiate anything more times than is allowed, and since the most important examples of manifolds are of class C^∞, it is conventional to avoid this annoyance by restricting attention to the smooth case. There is an additional reason to allow this restriction: according to a standard theorem in differential topology (see [**Hir94**, Theorem 2.9]), every manifold of class C^1 can be made into a *smooth* manifold by removing some of the charts in its maximal C^1-atlas. In this sense, one does not lose any significant generality by ignoring manifolds that are differentiable but not smooth.

You may have noticed on the other hand that Definition 2.18 also makes sense for $k = 0$, though in this case one cannot use the word "differentiable"; manifolds of class C^0 are called **topological manifolds** (*topologische Mannigfaltigkeiten*). These really are a different beast than differentiable manifolds: for every $n \geqslant 4$, there exist topological n-manifolds that *do not admit* any differentiable structure, i.e. their topology is not compatible with any atlas of class C^k for $k \geqslant 1$. Proving such things typically requires very advanced techniques, e.g. from mathematical gauge theory, which uses nonlinear PDEs to derive topological restrictions on smooth manifolds. (The classic introduction to this subject is [**DK90**].) In any case, the study of topological manifolds as such belongs squarely to the subject of topology, not differential geometry, so we will say no more about it here.

2.4. Some basic examples.

2.4.1. *Vector spaces.* For each integer $n \geqslant 0$, $\mathbb{R}^n$ admits a canonical smooth atlas consisting of a single n-dimensional chart, namely $(\mathbb{R}^n, \mathrm{Id})$. The smoothness of this atlas is a triviality: since there is only one chart, there is only one transition map to consider, which is the identity map and is therefore smooth. The unique extension of this atlas to a maximal smooth atlas on $\mathbb{R}^n$ defines what we will call the **standard smooth structure** on $\mathbb{R}^n$. The topology induced by this atlas is the standard one, which can also be defined in terms of the standard Euclidean metric; this follows via Remark 2.14 from the observations that $\mathbb{R}^n \subset \mathbb{R}^n$ is an open subset and $\mathrm{Id} : \mathbb{R}^n \to \mathbb{R}^n$ is a homeomorphism. It follows that $\mathbb{R}^n$ with its standard smooth structure is metrizable and (in light of the countable dense subset $\mathbb{Q}^n \subset \mathbb{R}^n$) separable. We conclude that $\mathbb{R}^n$ is, in a natural way, a smooth n-dimensional manifold. Note that it is *possible* to define different smooth structures on

$\mathbb{R}^n$, as shown by Example 2.7 in the case $n = 1$, but whenever we discuss $\mathbb{R}^n$ as a manifold in this course, we will always assume unless stated otherwise that it carries its standard smooth structure.

Since every real n-dimensional vector space V is isomorphic to $\mathbb{R}^n$, one can always choose such an isomorphism $\Phi : V \to \mathbb{R}^n$ and similarly regard V as a smooth n-manifold with an atlas consisting of the global chart (V, Φ). While the choice of isomorphism Φ here is typically not canonical, the resulting smooth structure on V is, since any other choice of isomorphism $\Psi : V \to \mathbb{R}^n$ would produce a chart (V, Ψ) that is related to (V, Φ) by the transition map $\Phi \circ \Psi^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. The latter is a vector space isomorphism, and thus a smooth map with a smooth inverse. In this way, we can regard every real n-dimensional vector space naturally as a smooth n-manifold.

2.4.2. *Open subsets.* If M is an n-dimensional C^k-manifold with atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$, then any open subset $\mathcal{O} \subset M$ admits a natural atlas

$$\mathcal{A}_\mathcal{O} := \{(\mathcal{U}_\alpha \cap \mathcal{O}, x_\alpha|_{\mathcal{U}_\alpha \cap \mathcal{O}})\}_{\alpha \in I},$$

which is also of class C^k since its transition maps are all restrictions of transition maps from $\mathcal{A}$ to open subsets. The key point here is that since $\mathcal{O} \subset M$ is open, each $\mathcal{U}_\alpha \cap \mathcal{O}$ is an open subset of $\mathcal{U}_\alpha$ and is thus mapped homeomorphically by x_α to another open subset of $\mathbb{R}^n$, making it an n-dimensional chart on $\mathcal{O}$. This atlas endows $\mathcal{O}$ with a natural C^k-structure, and since it is a subset of a separable metrizable space, Exercise 2.17 implies that it is also separable and metrizable, and is thus an n-dimensional C^k-manifold. Combining this with §2.4.1, we can now regard every open subset of $\mathbb{R}^n$ as a smooth n-manifold in a natural way.

2.4.3. *Disjoint unions.* The **disjoint union** (*disjunkte Vereinigung*) of a collection of sets $\{M_j\}_{j \in J}$ can be defined as the set

$$\coprod_{j \in J} M_j := \left\{ (j, t) \mid j \in J, \ t \in M_j \right\}.$$

Here J can be an arbitrary index set, finite, countable or uncountable. In the special case where J is finite, e.g. if $J = \{1, \ldots, N\}$, we also use the notation

$$M_1 \sqcup \ldots \sqcup M_N := \coprod_{j=1}^N M_j := \coprod_{j \in \{1, \ldots, N\}} M_j.$$

Identifying each of the individual sets M_j with the subset $\{j\} \times M_j \subset \coprod_{j \in J} M_j$, we can think of $\coprod_{j \in J} M_j$ as literally a union of all the sets M_j, with the caveat that for $j \neq k$, M_j and M_k are always *disjoint* as subsets of $\coprod_{j \in J} M_j$, even if as abstract sets they have elements in common. For example, the set $S^1 \sqcup S^1$ contains two copies of every point on the circle, and is thus not the same set as $S^1 \cup S^1 = S^1$. If you think of S^1 as the unit circle in $\mathbb{R}^2$, then the definition above gives $S^1 \sqcup S^1 = \{1, 2\} \times S^1 \subset \mathbb{R}^3$, so the disjoint union consists of two copies of the circle that live in disjoint planes in $\mathbb{R}^3$.

Suppose now that each of the sets M_j is a C^k-manifold with atlas $\mathcal{A}^{(j)} = \{(\mathcal{U}_\alpha^{(j)}, x_\alpha^{(j)})\}_{\alpha \in I_j}$. Regarding each set M_j as a subset of $\coprod_{j \in J} M_j$ makes each of the sets $\mathcal{U}_\alpha^{(j)}$ also into subsets of $\coprod_{j \in J} M_j$, such that $\mathcal{U}_\alpha^{(j)} \cap \mathcal{U}_\beta^{(k)} = \varnothing$ whenever $j \neq k$. It follows that the union

$$\mathcal{A} := \bigcup_{j \in J} \mathcal{A}^{(j)}$$

defines an atlas of class C^k on $\coprod_{j \in J} M_j$, whose set of transition maps is just the union of the sets of transition maps for all the atlases $\mathcal{A}^{(j)}$. (Transition maps relating two charts $(\mathcal{U}_\alpha^{(j)}, x_\alpha^{(j)})$ with $(\mathcal{U}_\beta^{(k)}, x_\beta^{(k)})$ with $j \neq k$ do not arise here since their overlap is always empty.)

It does *not* follow however that every disjoint union of a collection of C^k-manifolds is naturally a C^k-manifold—this is one of the few situations where we have to pay attention to the condition of separability. The topology induced by the atlas $\mathcal{A}$ on $\coprod_{j \in J} M_j$ is the so-called **disjoint union topology**, in which a subset $\mathcal{O} \subset \coprod_{j \in J} M_j$ is open if and only if $\mathcal{O} \cap M_j$ is an open subset of M_j for every $j \in J$. If the sets M_j are nonempty for uncountably many distinct values of $j \in J$, then no *countable* subset $E \subset \coprod_{j \in J} M_j$ can have an element in every one of the subsets M_j, and it follows that E cannot be dense, so the disjoint union cannot be separable. On the other hand, one can show (see Exercise 2.23 below) that every *finite* or *countable* disjoint union of separable metrizable spaces is also separable and metrizable. We conclude that for any $N \in \mathbb{N} \cup \{\infty\}$ and any finite or countable collection $\{M_j\}_{j=1}^N$ of C^k-manifolds, the disjoint union $\coprod_{j=1}^N M_j$ is also a C^k-manifold in a natural way. Moreover, if $\dim M_j = n$ for every j, then the disjoint union is also n-dimensional.

EXERCISE 2.23.

(a) Show that for any metric space (X, d), the formula
$$d'(x,y) := \begin{cases} d(x,y) & \text{if } d(x,y) < 1, \\ 1 & \text{if } d(x,y) \geqslant 1 \end{cases}$$
defines another metric d' on X that induces the same topology as d.

(b) Show that for any collection of metric spaces $\{(X_j, d_j)\}_{j \in J}$ with $d_j(x,y) \leqslant 1$ for all $j \in J$ and $x, y \in X_j$, the formula
$$d(x,y) := \begin{cases} d_j(x,y) & \text{if } x, y \in X_j \text{ for some } j \in J, \\ 2 & \text{if } x \in X_j \text{ and } y \in X_k \text{ for some } j, k \in J \text{ with } j \neq k \end{cases}$$
defines a metric on $\coprod_{j \in J} X_j$ that induces the disjoint union topology.

(c) Show that the metric d on $\coprod_{j \in J} X_j$ in part (b) is separable if J is a finite or countable set and all of the metric spaces (X_j, d_j) are separable.

EXERCISE 2.24. Recall that a metrizable space[15] is called **compact** (*kompakt*) if every open covering has a finite subcover. Show that a disjoint union $\coprod_{j \in J} M_j$ is compact if and only if J is finite and M_j is compact for every $j \in J$.

2.4.4. *Dimension zero.* You may not have thought about the case $n = 0$ when we defined the notion of an n-dimensional chart, but the definition in that case does make sense: $\mathbb{R}^0$ consists of a single point, and its only nontrivial open subset is itself, so if $(\mathcal{U}, x)$ is a 0-dimensional chart on M, then $\mathcal{U} \subset M$ is a single point. It follows that if M is a 0-dimensional manifold with atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$, then every point of M is its own open set, implying that *every* subset of M is open. This is known as the **discrete topology**, and it is always metrizable; a suitable metric is the **discrete metric**, defined by
$$d(x,y) := \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

The only dense subset of M in this topology is M itself, so separability requires M to be finite or countable. We conclude: a 0-dimensional manifold is simply a *finite or countable discrete set*, and it is compact if and only if it is finite. Equivalently, every 0-dimensional manifold can be identified with the disjoint union of at most countably many copies of the manifold $\mathbb{R}^0$, which is a single

---

[15]In fact this definition is also valid for arbitrary topological spaces.

point. Notice that since every map from $\mathbb{R}^0$ to itself is trivially smooth, every atlas on a 0-manifold is automatically a smooth atlas.

2.4.5. *Dimension one.* We have seen two explicit examples thus far of 1-dimensional manifolds: $\mathbb{R}$ and S^1, where the former carries its standard smooth structure as defined in §2.4.1, and the latter has a smooth structure that we defined using two charts based on polar coordinates in Lecture 1. We can now add to this list arbitrary open subsets of each, and arbitrary finite or countable disjoint unions of such open subsets. In this entire list, the only actual *compact* examples are S^1 and its finite disjoint unions; the compactness of the circle $S^1 \subset \mathbb{R}^2$ follows from the general fact that closed and bounded subsets of Euclidean space are compact. Up to a natural notion of equivalence for smooth manifolds that we will discuss in the next lecture, it turns out that these really are the *only* examples: in particular, every compact and *connected* 1-manifold is "diffeomorphic" to S^1. Later when we discuss manifolds with boundary, we will have to add the compact interval $[0, 1]$ to the list of compact 1-manifolds up to diffeomorphism. Similarly, it turns out that every noncompact connected 1-manifold is diffeomorphic to $\mathbb{R}$. We will not prove such classification results in this course, nor make use of them, but the curious reader will find a sketch of the corresponding result about connected topological 1-manifolds up to homeomorphism in [**Wen18**, Lecture 18]. Note that this is one of the important results that becomes false if one drops the metrizability condition from the definition of a manifold; we already saw one peculiar counterexample in Example 2.16, and another is the so-called "long line", which is essentially a union of *uncountably* many compact intervals glued together at their end points (see [**Wen18**, Lecture 18] or [**Spi99a**, Appendix to Chapter 1]).

2.4.6. *Cartesian products.* Since we have no plans to discuss infinite-dimensional manifolds in this course, we will not talk about infinite products, but finite products still provide a useful way of producing new manifolds from old ones. Assume M and N are C^k-manifolds of dimensions m and n respectively, with atlases $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ on M and $\mathcal{B} = \{(\mathcal{V}_\beta, y_\beta)\}_{\beta \in J}$ on N. For each $(\alpha, \beta) \in I \times J$, one can then define a **product chart** on $M \times N$ with domain $\mathcal{U}_\alpha \times \mathcal{V}_\beta$ by

$$\mathcal{U}_\alpha \times \mathcal{V}_\beta \to \mathbb{R}^{m \times n} : (p, q) \mapsto (x_\alpha(p), y_\beta(q)).$$

Each of the transition maps relating two product charts is just the Cartesian product of a transition map from $\mathcal{A}$ with one from $\mathcal{B}$, thus they are all of class C^k, and the collection of all product charts therefore defines an atlas of class C^k and makes $M \times N$ into a C^k-manifold of dimension $m + n$.[16] One can of course repeat this construction finitely many times to make any finite product of manifolds $M_1 \times \ldots \times M_N$ into a manifold.

An important special case of this construction is the compact smooth n-manifold known as the n-**torus**, defined by

$$\mathbb{T}^n := \underbrace{S^1 \times \ldots \times S^1}_{n}.$$

In the case $n = 1$, this is just another name for the circle, but the most popular torus is the case $n = 2$: as we've defined it, $\mathbb{T}^2$ is literally a subset of $\mathbb{R}^4$, but for visualization purposes there is also a favorite way of embedding it in $\mathbb{R}^3$, as shown in Figure 5.

---

[16] Note that even if $\mathcal{A}$ and $\mathcal{B}$ are maximal atlases, the set of all product charts is generally not maximal, but this is immaterial since it has a unique maximal extension.
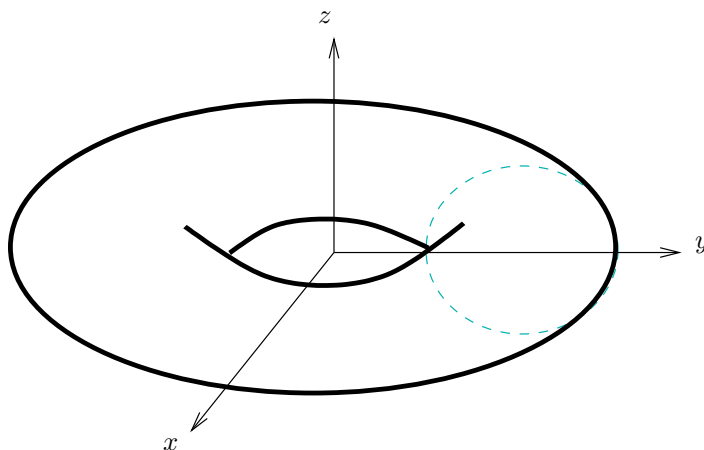
FIGURE 5.   A representation of the torus $\mathbb{T}^2$ as a submanifold of $\mathbb{R}^3$.

The n-torus for $n \geqslant 3$ is less straightforward to visualize, but it is often useful to think of it[17] as the quotient of $\mathbb{R}^n$ by the lattice $\mathbb{Z}^n$, using the bijection

$$\mathbb{R}^n/\mathbb{Z}^n \to \underbrace{S^1 \times \ldots \times S^1}_{n} : [(\theta_1, \ldots, \theta_n)] \mapsto (e^{2\pi i \theta_1}, \ldots, e^{2\pi i \theta_n}),$$

where for computational convenience we have replaced $\mathbb{R}^2$ with $\mathbb{C}$ in order to describe points in the unit circle S^1 as complex exponentials. Under this identification, a point in $\mathbb{T}^n$ is represented by a vector in $\mathbb{R}^n$, with the understanding that two vectors represent the same point in the torus if and only if they differ by a vector with integer coordinates. This perspective is especially useful in the study of Fourier series, as a function $f : \mathbb{R}^n \to \mathbb{C}$ that is 1-periodic in each of the n variables can now be regarded equivalently as a function $f : \mathbb{T}^n \to \mathbb{C}$.

EXERCISE 2.25. Convince yourself that the natural smooth structure on $\underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n}$ derived from the standard smooth structure of $\mathbb{R}$ is the same as the standard smooth structure of $\mathbb{R}^n$.

2.4.7. *The projective plane and the Klein bottle.* We conclude with two explicit examples of surfaces (i.e. smooth 2-manifolds) that are somewhat harder to visualize, because they cannot be embedded in $\mathbb{R}^3$.[18]

The **projective plane** (*projektive Ebene*) is the set of equivalence classes

$$\mathbb{RP}^2 := S^2/\sim,$$

where the equivalence relation is defined by $p \sim p$ and $p \sim -p$ for all $p \in S^2 \subset \mathbb{R}^3$, meaning that every point p in the unit sphere gets identified with its *antipodal* point $-p$. (For more on why this might be a natural object to define, see Exercise 2.26 below.) If you have ever been on a long-haul international flight, then you are familiar with the notion of traversing a continuous path along S^2. In order to picture a continuous path on $\mathbb{RP}^2$, you should imagine that there are

---

[17]In fact, many sources in the literature prefer to *define* $\mathbb{T}^n$ as the quotient group $\mathbb{R}^n/\mathbb{Z}^n$, in which case its smooth structure can be derived from the standard smooth structure of $\mathbb{R}^n$ using a general result about quotients by discrete group actions.

[18]The claim that embedding them into $\mathbb{R}^3$ is *impossible* is something I expect you to find plausible, but not obvious. Proving it would require some methods from topology which we do not yet have at our disposal in this course, though we may come back to this later.

always two identical and interchangeable airplanes, containing identical copies of the same crews and passengers, constrained to fly at exact antipodal points over the Earth. If one of those airplanes flies from Shanghai to Buenos Aires while the other one flies along the antipodal path,[19] then since the two planes are completely interchangeable, they can be understood to describe a *closed loop* on $\mathbb{RP}^2$. Got it? Good.

It is relatively easy to see that $\mathbb{RP}^2$ is a smooth 2-manifold in a natural way. First, it has a natural metric, in which one can describe each point of $\mathbb{RP}^2$ as a set consisting of two points in S^2 and define the distance between two points in $\mathbb{RP}^2$ as the distance between those two sets. The fact that S^2 is separable (as a subset of the separable metric space $\mathbb{R}^3$) implies easily that $\mathbb{RP}^2$ is also separable. One can also derive a smooth atlas on $\mathbb{RP}^2$ from the one that we already constructed on S^2 in Exercise 1.7: the only issue is that some of the charts need to have their domains shrunk so that they no longer contain any pairs of antipodal points, as the coordinate map will otherwise fail to be injective, but this can easily be done.

The second example is the **Klein bottle** (*Kleinsche Flasche*), a picture of which is shown in Figure 6. The picture must be interpreted with caution, since what it shows is not really a manifold in the usual sense, but if you imagine perturbing part of it in an unseen fourth dimension so that part of the surface no longer has to pass through another part, then you get the right intuition. The picture also shows a "grid" structure similar to the coordinate grid one would obtain on $\mathbb{T}^2$ after identifying it with $\mathbb{R}^2/\mathbb{Z}^2$, but the Klein bottle is not the same thing as the torus. The latter can be identified with the quotient

$$\left(\mathbb{R} \times (\mathbb{R}/\mathbb{Z})\right) \Big/ \sim$$

by the smallest equivalence relation on $\mathbb{R} \times (\mathbb{R}/\mathbb{Z})$ such that $(s, [t]) \sim (s+1, [t])$ for all $s, t \in \mathbb{R}$. One obtains a rigorous definition of the Klein bottle from this via a reversal of orientation: instead of $(s, [t]) \sim (s+1, [t])$, one takes the smallest equivalence relation on $\mathbb{R} \times (\mathbb{R}/\mathbb{Z})$ such that

$$(s, [t]) \sim (s+1, [-t])$$

for all $s, t \in \mathbb{R}$. If you think about what grid lines of the form $\{s = \text{const}\}$ and $\{t = \text{const}\}$ look like in the set of equivalence classes defined via this relation, you will end up with something resembling Figure 6. It is not difficult to construct an atlas of smoothly compatible 2-dimensional charts on this quotient: the basic idea is to view it as a quotient of $\mathbb{R}^2$, and restrict the canonical global chart of $\mathbb{R}^2$ to neighborhoods that are sufficiently small so as to contain at most one element from every equivalence class.

EXERCISE 2.26. The projective plane is the $n = 2$ case of the **real projective n-space** (*reeller projektiver Raum*)

$$\mathbb{RP}^n := S^n \big/ \sim,$$

where here again the equivalence relation identifies antipodal points $x \sim -x \in S^n \subset \mathbb{R}^{n+1}$. A useful interpretation of this definition comes from the observation that there is a unique line through the origin passing through each pair of points $\{x, -x\} \subset \mathbb{R}^{n+1}$. One can therefore view $\mathbb{RP}^n$ equivalently as the space of all *lines through the origin in* $\mathbb{R}^{n+1}$, which can be defined more precisely as the quotient

$$\mathbb{RP}^n = (\mathbb{R}^{n+1}\backslash\{0\})\big/ \sim$$

where two nontrivial vectors $v, w \in \mathbb{R}^{n+1}$ are now considered equivalent if and only if $v = \lambda w$ for some $\lambda \in \mathbb{R}$. From this perspective, it is convenient to denote points in $\mathbb{RP}^n$ via so-called

---

[19] According to the British science fiction TV series *Torchwood*, Buenos Aires and Shanghai are at exact antipodal points on the Earth. Wikipedia says this is true up to an error of about 400km. Let's just pretend it's true.
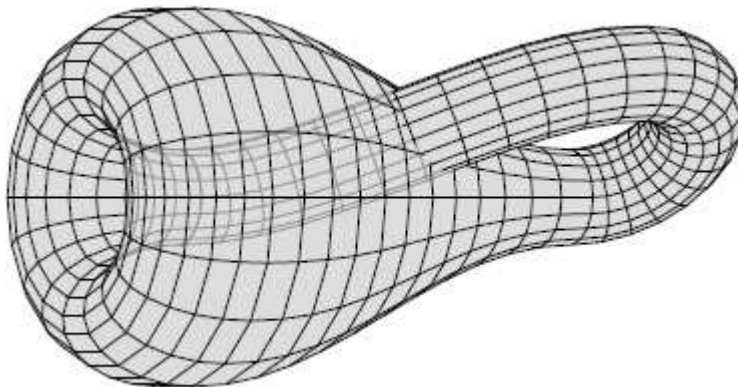
FIGURE 6.   An immersion of the Klein bottle into $\mathbb{R}^3$. It is not an embedding because it intersects itself. (We will discuss the precise meanings of the words "immersion" and "embedding" in Lecture 4.)

homogeneous coordinates, in which the symbol

$$[x_0 : \ldots : x_n] \in \mathbb{RP}^n$$

means the equivalence class containing the vector $(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \backslash \{0\}$.

The homogeneous coordinates can be used to define an explicit smooth atlas on $\mathbb{RP}^n$. For $j = 0, \ldots, n$, define

$$\mathcal{U}_j := \left\{ [x_0 : \ldots : x_n] \in \mathbb{RP}^n \mid x_j \neq 0 \right\}$$

and a map $\varphi_j : \mathbb{R}^n \to \mathbb{RP}^n$ by

$$\varphi_j(t_1, \ldots, t_n) := [t_1 : \ldots : t_j : 1 : t_{j+1} : \ldots : t_n].$$

Show that φ_j is an injective map onto $\mathcal{U}_j$, so $(\mathcal{U}_j, \varphi_j^{-1})$ is a chart, and compute the transition maps relating any two of the charts constructed in this way for different values of $j = 0, \ldots, n$. Show that these $n + 1$ charts together form a smooth atlas.

3. Smooth maps and tangent vectors

We have several more definitions to get through before the subject of differential geometry gets seriously underway. In this lecture we clarify what it means for a map between two manifolds to be differentiable, and what kind of object its derivative is.

3.1. Smooth maps between manifolds. We defined in §2.1 what it means for a real-valued function on a smooth manifold to be smooth (see Definition 2.3). The following is based on the same idea.

DEFINITION 3.1. Assume M and N are manifolds of dimensions m and n respectively, with differentiable structures $\mathcal{A}_M$ and $\mathcal{A}_N$ of class C^k. A continuous map $f : M \to N$ is said to be **of class C^r** for some $r \leqslant k$ (or **smooth** in the case $r = k = \infty$) if for every pair of charts $(\mathcal{U}, x) \in \mathcal{A}_M$ and $(\mathcal{V}, y) \in \mathcal{A}_N$, the map

$$\mathbb{R}^m \overset{\text{open}}{\supset} x(\mathcal{U} \cap f^{-1}(\mathcal{V})) \overset{y \circ f \circ x^{-1}}{\longrightarrow} y(\mathcal{V}) \overset{\text{open}}{\subset} \mathbb{R}^n$$

is of class C^r.

In other words, a map $f : M \to N$ is of class C^r if it looks like a map of class C^r when expressed in local coordinates on both the domain and the target. The assumption $r \leqslant k$ is again crucial here, and guarantees that for any given point $p \in M$, the question of whether f is of class C^r near p does not depend on the charts one has to choose near $p \in M$ and $f(p) \in N$. Note that we had to explicitly assume f was continuous in this definition: this assumption guarantees that $f^{-1}(\mathcal{V}) \subset M$ is an open set, so that $x(\mathcal{U} \cap f^{-1}(\mathcal{V}))$ is open in $\mathbb{R}^n$, and differentiability on this domain can therefore be checked.

The set of C^k maps from M to N is often denoted by

$$C^k(M, N) = \left\{ f : M \to N \mid f \text{ is of class } C^k \right\}.$$

One can endow this space with various natural topologies to make it into a topological (and sometimes also metrizable) space, though you should be aware that it is generally not a vector space, since N is not. On the other hand, the special case $N = \mathbb{R}$ is quite important, and is often abbreviated

$$C^k(M) := C^k(M, \mathbb{R}).$$

This *is* a vector space in a natural way, i.e. real-valued functions on a manifold M can be added and multiplied by constants.

EXERCISE 3.2. Show that for the standard smooth structure on $\mathbb{R}$ defined in §2.4.1, the notion of differentiability for a map $f : M \to \mathbb{R}$ as given in Definition 3.1 matches our previous definition for real-valued functions (Definition 2.3).

Up until this point I have been including non-smooth manifolds in the picture. I could continue doing this, but it would require frequently including slightly annoying extra hypotheses (like $r \leqslant k$) in statements of results, and the generality one gains by doing this does not fully compensate for the annoyance, so I will mostly assume $k = \infty$ from now on.

We can now define the natural notion of equivalence for smooth manifolds.

DEFINITION 3.3. For two smooth manifolds M and N, a smooth map $f : M \to N$ is called a **diffeomorphism** (*Diffeomorphismus*) if it is bijective and its inverse $f^{-1} : N \to M$ is also smooth. Two smooth manifolds are called **diffeomorphic** (*diffeomorph*) if there exists a diffeomorphism between them.

EXERCISE 3.4. Viewing S^1 as the unit circle in $\mathbb{C}$, the quotient group $\mathbb{R}^n / \mathbb{Z}^n$ admits a natural bijection to the n-torus $\mathbb{T}^n = S^1 \times \ldots \times S^1$, given by

$$\mathbb{R}^n / \mathbb{Z}^n \to \mathbb{T}^n : [(\theta_1, \ldots, \theta_n)] \mapsto (e^{2\pi i \theta_1}, \ldots, e^{2\pi i \theta_n}).$$

For each $v \in \mathbb{R}^n$, choose a neighborhood $\widetilde{\mathcal{U}}_v \subset \mathbb{R}^n$ of v that is small enough to contain at most one element from each equivalence class in $\mathbb{R}^n / \mathbb{Z}^n$, and use this to define an n-dimensional chart $(\mathcal{U}_v, x_v)$ of the form

$$\mathcal{U}_v = \left\{ [w] \in \mathbb{R}^n / \mathbb{Z}^n \mid w \in \widetilde{\mathcal{U}}_v \right\}, \qquad x_v([w]) = w.$$

Show that the collection of all charts of this form determines a smooth atlas on $\mathbb{R}^n / \mathbb{Z}^n$ such that the bijection to $\mathbb{T}^n$ described above is a diffeomorphism.

3.2. Tangent and cotangent spaces. Let us start this discussion with a concrete example: on the unit sphere $S^2 \subset \mathbb{R}^3$, a *tangent vector* to S^2 at a point $p \in S^2$ is by definition any vector of the form

$$\gamma'(0) \in \mathbb{R}^3,$$

where $\gamma : (-\epsilon, \epsilon) \to S^2$ is any choice of smooth path in $\mathbb{R}^3$ whose image is in S^2 and satisfies $\gamma(0) = p$. It should be easy to convince yourself that the set of all vectors of this form is a linear subspace of $\mathbb{R}^3$, namely, it is the orthogonal complement of p. We would now like to generalize

this notion to an arbitrary smooth manifold, without needing to assume that is a subset of some Euclidean space.

For the rest of this subsection, assume M is a smooth manifold and $p \in M$. Having defined what a smooth map between manifolds is, we can fix the standard smooth structure on small intervals such as $(-\epsilon, \epsilon) \subset \mathbb{R}$ and talk about smooth maps $\gamma : (-\epsilon, \epsilon) \to M$. If $\gamma(0) = p \in M$, then we will refer to any such smooth map as a **path through** p **in** M. Note that the value of $\epsilon > 0$ here is not fixed, so it is allowed to be arbitrarily small.

Let us say that two paths α, β through p in M are **tangent** if for some some chart $(\mathcal{U}, x)$ with $p \in \mathcal{U}$,

$$\frac{d}{dt}(x \circ \alpha)\bigg|_{t=0} = \frac{d}{dt}(x \circ \beta)\bigg|_{t=0}.$$

It is easy to show that this condition does not depend on the choice of chart: indeed, if $(\mathcal{V}, y)$ is another chart with $p \in \mathcal{V}$, then for all t close enough to 0 so that $\alpha(t) \in \mathcal{U} \cap \mathcal{V}$, we have $(y \circ \alpha)(t) = (y \circ x^{-1}) \circ (x \circ \alpha)(t)$ and thus by the chain rule,

$$(3.1) \qquad (y \circ \alpha)'(0) = D(y \circ x^{-1})(x(p))(x \circ \alpha)'(0),$$

where $D(y \circ x^{-1})(x(p)) : \mathbb{R}^n \to \mathbb{R}^n$ denotes the derivative of the transition map $y \circ x^{-1}$ at $x(p)$, which is an invertible linear map since $y \circ x^{-1}$ is smooth and has a smooth inverse. Since $(y \circ \beta)'(0)$ is related to $(x \circ \beta)'(0)$ in the same way, it is equal to $(y \circ \alpha)'(0)$ if and only if $(x \circ \beta)'(0) = (x \circ \alpha)'(0)$.

DEFINITION 3.5. A **tangent vector** (*Tangentialvektor*) to M at p is an equivalence class $[\gamma]$ of paths γ through p in M, where two paths are considered equivalent if and only if they are tangent. The set of all tangent vectors to M at p is called the **tangent space** (*Tangentialraum*) to M at p, and is denoted by

$$T_p M = \left\{ [\gamma] \mid \gamma \text{ a path through } p \text{ in } M \right\}.$$

This definition of $T_p M$ has many intuitive advantages, but it leaves several details unclear, foremost among them the fact that $T_p M$ is a vector space. In order to see this, we'll need to make more use of coordinates.

PROPOSITION 3.6. *The tangent space $T_p M$ has a unique vector space structure such that for any smooth n-dimensional chart $(\mathcal{U}, x)$ with $p \in \mathcal{U}$, the map*

$$(3.2) \qquad d_p x : T_p M \to \mathbb{R}^n : [\gamma] \mapsto (x \circ \gamma)'(0)$$

is a vector space isomorphism. In particular, every tangent space of a smooth n-manifold is naturally an n-dimensional vector space.

PROOF. The map (3.2) is a bijection by definition, so one can clearly always *choose* a chart $(\mathcal{U}, x)$ and define a vector space structure on $T_p M$ so as to make this map an isomorphism. The point is then to show that any other choice of chart $(\mathcal{V}, y)$ would have given the same vector space structure on $T_p M$. This follows from the formula

$$d_p y \circ (d_p x)^{-1} = D(y \circ x)(x(p)) : \mathbb{R}^n \to \mathbb{R}^n,$$

which follows from (3.1) and shows that this transformation is itself a vector space isomorphism. $\square$

EXAMPLE 3.7. If M is an open subset of an n-dimensional vector space V, then the derivative $\gamma'(0)$ for a smooth path $\gamma : (-\epsilon, \epsilon) \to V$ can be defined in the classical way as a vector in V, giving rise to a canonical map

$$T_p M \to V : [\gamma] \mapsto \gamma'(0)$$

for every $p \in M$. It is a straightforward exercise to show that this map is a vector space isomorphism.

In the future, we shall always use this isomorphism to identify tangent spaces on open subsets of a vector space V with V itself, so that we do not need to talk about equivalence classes of paths. In particular, every tangent space on an open subset of $\mathbb{R}^n$ is in this way canonically identified with $\mathbb{R}^n$. We will see in §4.3 below that whenever N is a submanifold of M, one can also naturally regard T_pN for each $p \in N$ as a linear subspace of T_pM, so in the special case where N is a submanifold of $\mathbb{R}^n$, its tangent spaces will all naturally be subspaces of $\mathbb{R}^n$. This means that for the vast majority of examples we are interested in, it will not be necessary to use the original definition in terms of equivalence classes of paths for describing a tangent space.

EXERCISE 3.8. Show that for two smooth manifolds M, N and any two points $p \in M$ and $q \in N$, there is a canonical vector space isomorphism $T_{(p,q)}(M \times N) = T_pM \times T_qN$.

In linear algebra, it is often useful to associate to any vector space V its **dual space** (*Dualraum*), which is the space of all scalar-valued linear maps on V. Assuming V is a *real* (rather than complex) vector space, this can be denoted by

$$V^* := \mathrm{Hom}(V, \mathbb{R}),$$

where for two real vector spaces V, W in general we denote by $\mathrm{Hom}(V, W)$ the vector space of linear maps $V \to W$. When V is a tangent space T_pM on a manifold M, its dual space is called the **cotangent space** (*Kotangentialraum*) to M at p and denoted by

$$T_p^*M := \mathrm{Hom}(T_pM, \mathbb{R}).$$

Its elements are called **cotangent vectors** (*Kotangentialvektoren*), or sometimes also **covectors**.

REMARK 3.9. Among physicists, covectors are often called "covariant vectors", while ordinary tangent vectors are called "contravariant vectors". I will not use this terminology.

3.3. The tangent bundle. The usefulness of the following definition will probably not be obvious to you at first glance, but it will become more apparent when we start differentiating smooth maps.

DEFINITION 3.10. The **tangent bundle** (*Tangentialbündel*) TM of a smooth manifold M is the union of all its tangent spaces:

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

The map $\pi : TM \to M$ such that $\pi^{-1}(p) = T_pM \subset TM$ for each $p \in M$ is called the **tangent projection**, and the subset in TM consisting of the zero vectors $0 \in T_pM$ for all $p \in M$ is called the **zero-section** (*Nullschnitt*) of TM. As subsets of TM, the individual tangent spaces $T_pM \subset TM$ for each $p \in M$ are sometimes referred to as the **fibers** (*Fasern*) of the tangent bundle.

Note that for distinct points $p \neq q \in M$, the tangent spaces T_pM and T_qM are by definition disjoint sets. Do not be tempted to think that the zero vector in T_pM is the same point as the zero vector in T_qM for $p \neq q$; in fact, there is a natural identification of the zero-section with M, giving rise to a natural inclusion

$$(3.3) \qquad\qquad\qquad i : M \hookrightarrow TM : p \mapsto 0 \in T_pM.$$

At the level of set theory, we could just as well have used the disjoint union notation $\coprod_{p \in M} T_pM$ in Definition 3.10, but we did not do this because it would give a misleading impression about the topology and smooth structure we intend to define on TM.

LEMMA 3.11. *On a manifold M, any n-dimensional chart $(\mathcal{U}, x)$ determines a $2n$-dimensional chart $(T\mathcal{U}, Tx)$ on the tangent bundle TM, where $T\mathcal{U} = \bigcup_{p \in \mathcal{U}} T_pM$ is the tangent bundle of the open*

subset $\mathcal{U} \subset M$, *and* $Tx : T\mathcal{U} \to \mathbb{R}^{2n}$ *is defined in terms of the linear isomorphism* $d_p x : T_p M \to \mathbb{R}^n$
of (3.2) *by*

$$T\mathcal{U} \supset T_p M \ni X \mapsto (x(p), d_p x(X)) \in \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}.$$

If $(\mathcal{V}, y)$ *is another chart on* M, *then transition maps relating the charts* $(T\mathcal{V}, Ty)$ *and* $(T\mathcal{U}, Tx)$
on TM *are given by*

$$Ty \circ (Tx)^{-1}(q, v) = \left(y \circ x^{-1}(q), D(y \circ x^{-1})(q)v \right).$$

PROOF. The map $Tx : T\mathcal{U} \to \mathbb{R}^{2n}$ is clearly injective, and its image is $x(\mathcal{U}) \times \mathbb{R}^n$, which is open. The stated formula for the transition map $Ty \circ (Tx)^{-1}$ follows from (3.1). $\qquad \square$

COROLLARY 3.12. *For any smooth manifold* M, *the tangent bundle* TM *can be endowed naturally with the structure of a smooth manifold such that the tangent projection* $\pi : TM \to M$, *the inclusion* $i : M \hookrightarrow TM$ *of the zero-section* (3.3) *and the natural inclusions* $T_p M \hookrightarrow TM$ *for all* $p \in M$ *are smooth maps.*[20] *If* $\dim M = n$, *then* $\dim TM = 2n$.

PROOF. We endow TM with the unique maximal smooth atlas containing all charts of the form $(T\mathcal{U}, Tx)$ determined via Lemma 3.11 from smooth charts $(\mathcal{U}, x)$ on M.

To check that $\pi : TM \to M$ is a smooth map, one can now write its coordinate expression with respect to any chart $(\mathcal{U}, x)$ on M and the corresponding chart $(T\mathcal{U}, Tx)$ on TM: the resulting map from an open subset of $\mathbb{R}^{2n}$ to $\mathbb{R}^n$ takes the form $(q, v) \mapsto q$, and is thus clearly smooth. Writing down the inclusion of the zero-section $M \hookrightarrow TM$ in similar coordinates produces $q \mapsto (q, 0)$, and for the inclusion $T_p M \hookrightarrow TM$, one obtains $v \mapsto (q, v)$. All of these maps are smooth.

I hope you find it plausible that TM with the atlas constructed above is metrizable and separable. Separability is easy to prove, e.g. one can take the union of countable dense subsets of individual fibers $T_p M$ for all p in some countable dense subset of M, thus forming a countable dense subset of TM. The easiest way I can think of to prove metrizability is by constructing a Riemannian metric on TM, which we will do in Lecture 15. That construction will rely on the assumption that M is metrizable; we will not need to assume this about TM. $\qquad \square$

EXERCISE 3.13. Find a diffeomorphism from the tangent bundle TS^1 to the product manifold $S^1 \times \mathbb{R}$.

One can similarly define a **cotangent bundle** (*Kotangentialbündel*)

$$T^* M := \bigcup_{p \in M} T_p^* M,$$

which satisfies a result analogous to Corollary 3.12. We will postpone the proof of this fact, since it follows from more general results about vector bundles to be discussed later in the course, and we will not really have use for it until then.

3.4. Tangent maps. We can now answer a question you may have wondered about: we know how to define whether a map $f : M \to N$ between manifolds is differentiable, but how does one actually *differentiate* it, i.e. what is its derivative at a point? In the special case $M \overset{\text{open}}{\subset} \mathbb{R}^m$ and $N = \mathbb{R}^n$, the answer you learned from first-year analysis is to view the derivative $Df(p)$ at a point $p \in M$ as a linear map $\mathbb{R}^m \to \mathbb{R}^n$, and according to the chain rule, it satisfies the relation

$$(f \circ \gamma)'(0) = Df(p)\gamma'(0)$$

for any smooth path γ through p. In fact, since any vector in $\mathbb{R}^m$ can be the derivative of some smooth path through p, this formula uniquely characterizes the linear map $Df(p) : \mathbb{R}^m \to \mathbb{R}^n$. It

---

[20]Here we are using the vector space structure of $T_p M$ to regard it as a smooth manifold as in §2.4.1.

also admits an obvious generalization to the setting of smooth manifolds, using the fact that if $\gamma : (-\epsilon, \epsilon) \to M$ is a path through $p \in M$, then $f \circ \gamma : (-\epsilon, \epsilon) \to N$ is a path through $f(p) \in N$.

DEFINITION 3.14. For two smooth manifolds M, N and a smooth map $f : M \to N$, the **tangent map** (*Tangentialabbildung*) of f is the map

$$Tf : TM \to TN : [\gamma] \mapsto [f \circ \gamma].$$

Its restriction to the tangent space at a specific point $p \in M$ can be denoted by

$$T_p f : T_p M \to T_{f(p)} N,$$

and is also called the **derivative** of f at p.[21]

LEMMA 3.15. *The map $T_p f : T_p M \to T_{f(p)} N$ defined above for a smooth map $f : M \to N$ and a point $p \in M$ is independent of choices, and it is linear. Moreover, if $f : M \to N$ is smooth, then $Tf : TM \to TN$ is also smooth.*

PROOF. All of these statements will become obvious if we write down a local coordinate expression for the map $Tf : TM \to TN$. Choose charts $(\mathcal{U}, x)$ on M and $(\mathcal{V}, y)$ on N with $p \in \mathcal{U}$ and $f(p) \in \mathcal{V}$. These give rise to charts $(T\mathcal{U}, Tx)$ on TM and $(T\mathcal{V}, Ty)$ on TN as in Lemma 3.11, so that given any $[\gamma] \in T_p M$, $Tx([\gamma]) = (x(p), (x \circ \gamma)'(0)) \in \mathbb{R}^m \times \mathbb{R}^m$, and according to the definition of Tf,

$$Ty(Tf([\gamma])) = (y(f(p)), (y \circ (f \circ \gamma))'(0)) \in \mathbb{R}^n \times \mathbb{R}^n.$$

The assumption that f is smooth means that $y \circ f \circ x^{-1}$ is smooth on its domain of definition, which is a neighborhood of $x(p)$ in $\mathbb{R}^m$. On this neighborhood, we can then write $y \circ (f \circ \gamma) = (y \circ f \circ x^{-1}) \circ (x \circ \gamma)$ and apply the chain rule to derive from the above expression,

$$Ty \circ Tf \circ (Tx)^{-1}(x(p), (x \circ \gamma)'(0)) = (y \circ f \circ x^{-1}(x(p)), D(y \circ f \circ x^{-1})(x(p))(x \circ \gamma)'(0)),$$

or if we simplify by writing $q := x(p) \in \mathbb{R}^m$ and $v := (x \circ \gamma)'(0) \in \mathbb{R}^m$,

$$Ty \circ Tf \circ (Tx)^{-1}(q, v) = (y \circ f \circ x^{-1}(q), D(y \circ f \circ x^{-1})(q)v).$$

This formula does not depend on any choice of path γ to represent the tangent vector $[\gamma] \in T_p M$, thus it proves that $Tf([\gamma])$ also does not depend on this choice, and moreover, it defines a smooth map $TM \to TN$ with a linear restriction $T_p M \to T_{f(p)} N$. $\square$

The tangent bundle provides a more elegant language for talking about derivatives than was available in your first-year analysis course. As justification for this claim, I offer the following reformulation of the chain rule in the language of manifolds; it follows directly from the definitions of tangent spaces and tangent maps (which are in themselves crucially dependent on the chain rule from first-year analysis).

PROPOSITION 3.16 (chain rule). *For any pair of smooth maps $f : M \to N$ and $g : N \to Q$ between smooth manifolds, $T(g \circ f) = Tg \circ Tf : TM \to TQ$.* $\square$

COROLLARY 3.17. *If $f : M \to N$ is a diffeomorphism, then so is $Tf : TM \to TN$, and $(Tf)^{-1} = T(f^{-1}) : TN \to TM$.*

PROOF. Observe first that the tangent map to the identity map on M is the identity map on TM. The chain rule then implies $\mathrm{Id}_{TM} = T(f \circ f^{-1}) = Tf \circ T(f^{-1})$. $\square$

---

[21]You will find a variety of alternative notation in the literature for what I am calling $T_p f$, e.g. *df*(*p*) and *Df*(*p*) are also popular choices. In these notes, I will try to consistently reserve $Df(p)$ for the notion of derivatives defined in first-year analysis, where one only considers maps between open subsets of Euclidean spaces. The notation *df* will be reserved for the *differential* of a function valued in $\mathbb{R}$ or another vector space, to be defined in the next lecture.

REMARK 3.18. Since $T_q\mathbb{R}^n$ is canonically isomorphic to $\mathbb{R}^n$ for every $q \in \mathbb{R}^n$, the tangent bundle $T\mathbb{R}^n$ has a canonical identification with $\mathbb{R}^n \times \mathbb{R}^n$ in which $T_q\mathbb{R}^n = \{q\} \times \mathbb{R}^n$. Under this identification, the chart $Tx : T\mathcal{U} \to \mathbb{R}^n \times \mathbb{R}^n$ on TM derived in Lemma 3.11 from a chart $x : \mathcal{U} \to \mathbb{R}^n$ on M is simply the tangent map of x.

REMARK 3.19. If you are familiar with the language of categories and functors, then you might appreciate the following interpretation of Proposition 3.16. One can define a category Diff whose objects are the smooth manifolds, with morphisms $M \to N$ defined to be smooth maps, hence the isomorphisms in this category are the diffeomorphisms. The construction of the tangent bundle now gives rise to a functor $T : \mathsf{Diff} \to \mathsf{Diff}$ which sends each manifold M to TM and associates to any morphism $f : M \to N$ its tangent map $Tf : TM \to TN$. The formula $T(g \circ f) = Tg \circ Tf$ is the main step required for proving that T is a functor.

REMARK 3.20. If M is a manifold of class C^k for some finite $k \in \mathbb{N}$, then the definition of tangent spaces requires a slight adjustment since the notion of *smooth* paths in M might not make sense; it is good enough however (and gives an equivalent definition) if we consider all paths $\gamma : (-\epsilon, \epsilon) \to M$ of class C^1. Inspecting the proof of Corollary 3.12 now reveals that TM is naturally a manifold of class C^{k-1}; one derivative is lost because the transition maps for TM involve derivatives of the transition maps for M. Similarly, if $f : M \to N$ is of class C^r with $1 \leqslant r \leqslant k$, then the tangent map $Tf : TM \to TN$ can be defined as a map of class C^{r-1}.

4. Submanifolds

The overarching message of this lecture will be that sometimes, understanding what is happening in a manifold is just a matter of finding the right coordinates.

4.1. Partial derivatives and differentials. There are two special situations in which the tangent map of $f : M \to N$ can be expressed in slightly more convenient forms. First, if $\mathcal{U} \subset \mathbb{R}^n$ is an open subset of Euclidean space, M is a manifold and $f : \mathcal{U} \to M$ is smooth, then f can be regarded (without needing to make a choice of coordinates) as an M-valued function of n variables, $f(x^1, \ldots, x^n)$. For each point $x_0 = (x_0^1, \ldots, x_0^n) \in \mathcal{U}$, f now determines n smooth paths through $f(x_0)$, namely

$$\gamma_j(t) := f(x_0^1, \ldots, x_0^{j-1}, x_0^j + t, x_0^{j+1}, \ldots, x_0^n), \qquad j = 1, \ldots, n.$$

The equivalence classes of these paths are called the **partial derivatives** of f at x_0,

$$\partial_j f(x_0) := \frac{\partial f}{\partial x^j}(x_0) := [\gamma_j] \in T_{f(x_0)}M.$$

They are actually just particular values of the tangent map, i.e. $\partial_j f(x_0) = T_{x_0}f(e_j)$, where we are using the fact that $T_{x_0}\mathcal{U}$ is canonically isomorphic to $\mathbb{R}^n$ (see Example 3.7) and thus comes with a canonical basis $e_1, \ldots, e_n$. The n tangent vectors $\partial_1 f(x_0), \ldots, \partial_n f(x_0) \in T_{f(x_0)}M$ all together thus contain the same information as the tangent map $T_{x_0}f : T_{x_0}\mathcal{U} \to T_{f(x_0)}M$.

The second special situation is in some dense dual to the first: we consider a smooth function on a smooth manifold M with values in a finite-dimensional vector space V,

$$f : M \to V.$$

The most important special case of this is when $V = \mathbb{R}$, so that f is a real-valued function. Taking advantage again of the canonical isomorphisms $T_{f(p)}V = V$ from Example 3.7, we can rewrite $Tf(X) \in T_{f(p)}V$ for each $p \in M$ and $X \in T_pM$ as a vector in V, denoted by $df(X) \in V$. This associates to every smooth function $f : M \to V$ a smooth function

$$df : TM \to V,$$

called the **differential** (*Differential*) of f. We will denote its restriction to each individual tangent space $T_p M$ for $p \in M$ by
$$d_p f : T_p M \to V.$$
In terms of equivalence classes of paths through p, a direct formula for $d_p f$ is given by

(4.1) $$d_p f([\gamma]) = (f \circ \gamma)'(0),$$

and one can deduce from Lemma 3.15 that this is independent of the choice of path γ in the equivalence class, and moreover, $d_p f : T_p M \to V$ is a linear map. In particular, for a smooth real-valued function $f : M \to \mathbb{R}$, $d_p f$ is an element of the cotangent space at p,
$$d_p f \in T_p^* M \qquad (\text{for } f : M \to \mathbb{R}).$$
This makes the differentials df of smooth real-valued functions $f : M \to \mathbb{R}$ into our first examples of *differential forms*; we will have a lot more to say about them when we discuss integration in a few weeks.

EXAMPLE 4.1. The differentials defined above directly generalize the linear map $d_p x : T_p M \to \mathbb{R}^n$ in (3.2), which can be associated to any smooth chart $(\mathcal{U}, x)$ on M and a point $p \in \mathcal{U}$. This map can also be constructed out of the differentials of the coordinate functions $x^1, \ldots, x^n : \mathcal{U} \to \mathbb{R}$; it is given by
$$d_p x(X) = (d_p x^1(X), \ldots, d_p x^n(X)) \in \mathbb{R}^n.$$

4.2. The inverse function theorem. In the examples of manifolds we have dealt with so far, we have always had charts that were explicitly constructed, but such explicit constructions are not always convenient in more general situations. A nice tool for obtaining less explicit but often more useful constructions of charts is provided by the inverse function theorem from first-year analysis. Let us recall the statement:

THEOREM (inverse function theorem). *Suppose $\mathcal{U} \subset \mathbb{R}^n$ is open, $f : \mathcal{U} \to \mathbb{R}^n$ is a map of class C^k for some $k \in \mathbb{N} \cup \{\infty\}$, and $x_0 \in \mathcal{U}$ is a point at which the derivative $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism. Then there exist open neighborhoods $x_0 \in \Omega \subset \mathcal{U}$ and $f(x_0) \in \Omega' \subset \mathbb{R}^n$ such that f maps Ω bijectively onto Ω' and the inverse $(f|_\Omega)^{-1} : \Omega' \to \Omega$ is also of class C^k.* $\square$

We will now turn this standard analytical result into a pair of criteria for proving that certain maps we construct define smooth charts.

LEMMA 4.2. *Suppose M is a smooth n-manifold, $\mathcal{U} \subset \mathbb{R}^n$ is an open set, $\varphi : \mathcal{U} \to M$ is a smooth map and $x_0 \in \mathcal{U}$ is a point at which the partial derivatives $\partial_1 \varphi(x_0), \ldots, \partial_n \varphi(x_0)$ form a basis of $T_{\varphi(x_0)} M$. Then there exist open neighborhoods $x_0 \in \Omega \subset \mathcal{U}$ and $p := \varphi(x_0) \in \mathcal{O} \subset M$ such that φ maps Ω bijectively onto $\mathcal{O}$ and $(\mathcal{O}, (\varphi|_\Omega)^{-1})$ defines a smooth chart on M.*

PROOF. Choose any smooth chart $(\mathcal{V}, y)$ on M with $p = \varphi(x_0) \in \mathcal{V}$, and observe that $d_p y(\partial_j \varphi(x_0)) = \partial_j (y \circ \varphi)(x_0)$In for each $j = 1, \ldots, n$. Since $d_p y : T_p M \to \mathbb{R}^n$ is an isomorphism, our assumption on the basis $\partial_1 \varphi(x_0), \ldots, \partial_n \varphi(x_0) \in T_p M$ means that $\partial_1 (y \circ \varphi)(x_0), \ldots, \partial_n (y \circ \varphi)(x_0)$ is similarly a basis of $\mathbb{R}^n$, which is equivalent to saying that the linear map $D(y \circ \varphi)(x_0) : \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism. The inverse function theorem thus provides open neighborhoods $x_0 \in \Omega \subset \mathcal{U}$ and $y(p) \in \Omega' \subset \mathbb{R}^n$ such that $y \circ \varphi$ is a diffeomorphism between Ω and Ω', implying that $\varphi = y^{-1} \circ (y \circ \varphi)$ sends Ω bijectively to an open neighborhood $\mathcal{O} := y^{-1}(\Omega')$ of p. Denoting the inverse of this bijection by $x : \mathcal{O} \to \Omega \subset \mathbb{R}^n$, the transition map $y \circ x^{-1}$ is now just $y \circ \varphi|_\Omega$, so it is smooth and has a smooth inverse. $\square$

LEMMA 4.3. *Suppose M is a smooth n-manifold, $\mathcal{U} \subset M$ is an open set, $x^1, \ldots, x^n : \mathcal{U} \to \mathbb{R}$ are smooth functions and $p \in \mathcal{U}$ is a point such that the differentials $d_p x^1, \ldots, d_p x^n$ form a basis of*

$T_p^* M$. Then there exists an open neighborhood $p \in \mathcal{O} \subset \mathcal{U}$ such that $(\mathcal{O}, x)$ with $x := (x^1, \ldots, x^n) : \mathcal{O} \to \mathbb{R}^n$ defines a smooth chart on M.

PROOF. Since $d_p x^1, \ldots, d_p x^n$ is a basis of $T_p^* M$, it is dual to a unique basis $X_1, \ldots, X_n$ of $T_p M$, meaning the two bases are related by

$$d_p x^i (X_j) = \delta_j^i := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Define the linear map $d_p x := (d_p x^1, \ldots, d_p x^n) : T_p M \to \mathbb{R}^n$ as in Example 4.1, so $d_p x$ is the tangent map $T_p x : T_p M \to T_{x(p)} \mathbb{R}^n$ after identifying $T_{x(p)} \mathbb{R}^n = \mathbb{R}^n$. Since $d_p x$ sends the basis $X_1, \ldots, X_n$ to the standard basis of $\mathbb{R}^n$, it is an isomorphism. Now if $(\mathcal{V}, y)$ is any smooth chart with $p \in \mathcal{V}$, the map $x \circ y^{-1}$ is smooth on a neighborhood of p, and the chain rule gives

$$D(x \circ y^{-1})(y(p)) = d_p x \circ (d_p y)^{-1},$$

hence the latter is also an isomorphism $\mathbb{R}^n \to \mathbb{R}^n$. The inverse function theorem now provides open neighborhoods $y(p) \in \Omega \subset \mathbb{R}^n$ and $x(p) \in \Omega' \subset \mathbb{R}^n$ such that $x \circ y^{-1}$ is a diffeomorphism from Ω onto Ω', so $\mathcal{O} := y^{-1}(\Omega) = x^{-1}(\Omega')$ is then a neighborhood of p on which the restriction of x defines a chart that is smoothly compatible with $(\mathcal{V}, y)$. □

4.3. Slice charts. We have used the word "submanifold" already a few times in an informal way, e.g. the unit circle S^1 is a manifold that lives inside the manifold $\mathbb{R}^2$, so we called it a submanifold. It is now time to clarify more precisely what this word means.

The archetypal example of a submanifold is a linear subspace of a vector space, for instance

$$\mathbb{R}^\ell \times \{0\} = \left\{ (x^1, \ldots, x^\ell, 0, \ldots, 0) \in \mathbb{R}^n \mid (x^1, \ldots, x^\ell) \in \mathbb{R}^\ell \right\} \subset \mathbb{R}^n.$$

Basic results in linear algebra imply that *any* ℓ-dimensional subspace of an n-dimensional vector space looks like this example after a suitable linear change of coordinates. The notion of a smooth submanifold generalizes this by allowing nonlinear (but smooth) changes of coordinates.

DEFINITION 4.4. A chart $(\mathcal{U}, x)$ on an n-manifold M is called an ℓ-dimensional **slice chart** (*Bügelkarte*) for a subset $L \subset M$ if

$$L \cap \mathcal{U} = x^{-1}(\mathbb{R}^\ell \times \{0\}),$$

i.e. the points in $\mathcal{U}$ belong to L if and only if their coordinates $x^{\ell+1}, \ldots, x^n$ vanish.

DEFINITION 4.5. Suppose M is a smooth n-manifold. A subset $L \subset M$ is called an ℓ-**dimensional smooth submanifold** (*Untermannigfaltigkeit*) of M if M admits a collection of smooth slice charts for L whose domains cover every point of L.

REMARK 4.6. More generally, if M is a manifold of class C^k but not necessarily smooth, one can speak of *submanifolds of class C^k*, in which the transition maps between slice charts are required to be of class C^k. Note that a C^k-manifold can also be regarded as a C^r-manifold for any $r \leqslant k$, so under this condition it makes sense to talk about C^r-submanifolds, but e.g. there is no such thing as a *smooth* submanifold of M if the latter is of class C^k for some $k < \infty$ but not equipped with a smooth structure.

EXAMPLE 4.7. The smooth structure we constructed on $S^1 \subset \mathbb{R}^2$ in Lecture 1 was obtained from polar coordinates by restricting to the unit circle $\{r = 1\}$; this gave rise to two charts $(\mathcal{U}, \theta)$ and $(\mathcal{V}, \phi)$, where θ and ϕ both had the meaning of an angle in polar coordinates, but with different ranges of values, namely $\theta(\mathcal{U}) = (0, 2\pi)$ and $\phi(\mathcal{V}) := (-\pi, \pi)$. These two coordinates were defined on open subsets of S^1, but they also have natural extensions to open subsets of $\mathbb{R}^2$, namely

$$\mathcal{U}' := \left\{ tv \in \mathbb{R}^2 \mid v \in \mathcal{U}, \ t > 0 \right\}, \qquad \mathcal{V}' := \left\{ tv \in \mathbb{R}^2 \mid v \in \mathcal{V}, \ t > 0 \right\}.$$

The radial coordinate r is defined on $\mathbb{R}^2\backslash\{0\}$ and takes all positive values; if we now set $\rho := r - 1$ so that $\{r = 1\} = \{\rho = 0\}$, we obtain a pair of smoothly compatible slice charts $(\mathcal{U}', (\theta, \rho))$ and $(\mathcal{V}', (\phi, \rho))$ for S^1 such that $S^1 \subset \mathcal{U}' \cup \mathcal{V}'$. This means that S^1 is a smooth submanifold of $\mathbb{R}^2$.

One can similarly turn the atlas for S^2 in Exercise 1.7 into a family of slice charts to prove that S^2 is a submanifold of $\mathbb{R}^3$. In practice, however, constructing slice charts by hand is not usually necessary, as we will see in §4.4 that some much more general and powerful tools for this purpose are provided by the inverse function theorem.

Let us first clarify the fact that a submanifold of a manifold is also a manifold in its own right.

PROPOSITION 4.8. *If L is an ℓ-dimensional C^k-submanifold of an n-dimensional C^k-manifold M, then L inherits naturally from M the structure of an ℓ-dimensional C^k-manifold such that the inclusion map $L \hookrightarrow M$ is of class C^k. Moreover, for each $p \in L$, the tangent space T_pL is naturally an ℓ-dimensional linear subspace of T_pM.*

PROOF. We associate to every slice chart $(\mathcal{U}, x)$ for $L \subset M$ a chart of the form $(\mathcal{U} \cap L, x_L)$ on L, where we use the coordinate projection $\pi_\ell(x^1, \ldots, x^n) := (x^1, \ldots, x^\ell)$ to define

$$x_L = \pi_\ell \circ x|_{\mathcal{U} \cap L} : \mathcal{U} \cap L \to \mathbb{R}^\ell.$$

By assumption, L can be covered by slice charts, so the collection of all charts of this form defines an atlas on L. Given two such charts $(\mathcal{U} \cap L, x_L)$ and $(\mathcal{V} \cap L, y_L)$ derived from two C^k-compatible slice charts $(x, \mathcal{U})$ and $(y, \mathcal{V})$, the transition map $y \circ x^{-1}$ preserves the subspace $\mathbb{R}^\ell \times \{0\} \subset \mathbb{R}^n$, and its restriction to the intersection of its domain with this subspace is the transition map $y_L \circ x_L^{-1}$, which is therefore of class C^k. Moreover, the fact that M is metrizable and separable implies the same for L by Exercise 2.17, thus L is a C^k-manifold. The local coordinate expression for the inclusion $i : L \hookrightarrow M$ with respect to any slice chart $(\mathcal{U}, x)$ and the associated chart $(\mathcal{U} \cap L, x_L)$ on L is $(x^1, \ldots, x^\ell) \mapsto (x^1, \ldots, x^\ell, 0, \ldots, 0)$, which is clearly smooth, thus the inclusion is of class C^k.[22] For each $p \in L$, the tangent map $T_pi : T_pL \to T_pM$ is simply the canonical inclusion $T_pL \hookrightarrow T_pM$ defined by regarding each path in L as a path in M. Since its image is a linear subspace, it gives a canonical isomorphism of T_pL to a linear subspce of T_pM. $\square$

Whenever we speak of a submanifold $L \subset M$ from now on, we will assume that L is endowed with the differentiable structure described in Proposition 4.8, so that it can also be regarded as a manifold in its own right. We will often make use of the canonical identification of tangent spaces T_pL with subspaces of T_pM, especially in the case $M = \mathbb{R}^n$, where (in light of Example 3.7) this identification allows us to view each tangent space T_pL as a subspace of $\mathbb{R}^n$.

EXERCISE 4.9. Assume in the following that M and N are both C^k-manifolds and $f : M \to N$ is a map of class C^k. Prove:

(a) For any C^k-submanifold $L \subset M$, the restriction $f|_L : L \to N$ is also a map of class C^k.
(b) If $L \subset N$ is a C^k-submanifold such that $f(M) \subset L$, then the resulting map $f : M \to L$ is also of class C^k.

4.4. Immersions and submersions.

DEFINITION 4.10. A smooth map $f : M \to N$ is called an **immersion** at $p \in M$ if the linear map $T_pf : T_pM \to T_{f(p)}N$ is injective, and similarly, f is a **submersion** at p if $T_pf : T_pM \to$

---

[22]Recall that if both L and M are manifolds of class C^k but $k < \infty$, then it does not make sense to say that the inclusion $L \hookrightarrow M$ is smooth, even though it looks smooth in the particular local coordinates we chose. The point is that one could also choose different coordinates in which it would still appear to be a map of class C^k, but not necessarily C^∞.

$T_{f(p)}N$ is surjective. If one says that f is an immersion/submersion without specifying a point p, the meaning is that it is true for *all* points in M. One sometimes uses the notation

$$f : M \hookrightarrow N$$

to indicate when f is an immersion.

Recall that for any two finite-dimensional vector spaces V, W, the sets of linear maps $V \to W$ that are injective or surjective are open. It follows that if f is an immersion or submersion at some point $p \in M$, then this is also true on a *neighborhood* of p; equivalently, the set of points at which f is an immersion or submersion is open.

There is a good reason to single out these two particular classes of smooth maps between manifolds: it turns out that up to choices of smooth coordinates near $p \in M$ and $f(p) \in N$, all immersions look the same, and similarly for all submersions. This fact will give us a new user-friendly tool for identifying smooth submanifolds. The main tool required in its proof is the inverse function theorem, or more precisely, the two lemmas in §4.2 that used the inverse function theorem to construct charts.

THEOREM 4.11. *Assume M is a smooth m-manifold, N is a smooth n-manifold, $f : M \to N$ is a smooth map, $p \in M$ and $q = f(p) \in N$. If f is either an immersion or a submersion at p, then there exist smooth charts $(\mathcal{U}, x)$ on M with $x(p) = 0 \in \mathbb{R}^m$ and $(\mathcal{V}, y)$ on N with $y(q) = 0 \in \mathbb{R}^n$ such that the coordinate expression $y \circ f \circ x^{-1}$ for f is given by*

$$\mathbb{R}^m \ni (x^1, \ldots, x^m) \mapsto \begin{cases} (x^1, \ldots, x^n) \in \mathbb{R}^n & \text{if } m \geqslant n \text{ (submersion case)}, \\ (x^1, \ldots, x^m, 0, \ldots, 0) \in \mathbb{R}^n & \text{if } m < n \text{ (immersion case)}. \end{cases}$$

PROOF. Assume first that $T_p f : T_p M \to T_{f(p)} N$ is injective, so $n \geqslant m$, and set $\ell := n - m$. Choose a smooth chart $(\mathcal{U}, x)$ on M with $p \in \mathcal{U}$ and $x(p) = 0 \in \mathbb{R}^m$; note that the latter can be assumed without loss of generality by taking any chart with $p \in \mathcal{U}$ and composing the map $\mathcal{U} \to \mathbb{R}^n$ with a translation on $\mathbb{R}^n$ sending the image of p to the origin. With this understood, $\Omega := x(\mathcal{U}) \subset \mathbb{R}^m$ is an open neighborhood of the origin, and we observe that $F := f \circ x^{-1} : \Omega \to N$ is now a smooth map such that $F(0) = q$ and $T_0 F = T_p f \circ (d_p x)^{-1} : \mathbb{R}^m \to T_q N$ is injective. The latter is equivalent to the condition that the partial derivatives $\partial_1 F(0), \ldots, \partial_m F(0) \in T_q N$ are linearly independent.

We claim that after possibly shrinking Ω to a smaller neighborhood of $0 \in \mathbb{R}^m$, and choosing $\epsilon > 0$ sufficiently small, $F : \Omega \to N$ can be extended to a smooth map

$$\widetilde{F} : \Omega \times (-\epsilon, \epsilon)^\ell \to N$$

such that $\widetilde{F}(x^1, \ldots, x^m, 0, \ldots, 0) = F(x^1, \ldots, x^m)$ and the partial derivatives $\partial_1 \widetilde{F}, \ldots, \partial_n \widetilde{F}$ at the origin form a basis of $T_q N$. This extension is not canonical, but it is also not difficult: if N were simply $\mathbb{R}^n$, we could define it by choosing any extension of the linearly independent set $\partial_1 F(0), \ldots, \partial_m F(0)$ to a basis $\partial_1 F(0), \ldots, \partial_m F(0), Y_{m+1}, \ldots, Y_n$ of $T_q N$ and then defining

$$\widetilde{F}(x^1, \ldots, x^n) := F(x^1, \ldots, x^m) + \sum_{j=m+1}^{n} x^j Y_j.$$

This formula does not make sense in general if N is not a vector space, but one could more generally choose a chart on N near q in order to express F in local coordinates, and define the extension in this way in coordinates. Lemma 4.2 now implies that on a sufficiently small neighborhood of $0 \in \mathbb{R}^n$, $\widetilde{F}$ can be inverted to define a chart $(\mathcal{V}, y)$ on N with the stated properties.

Next suppose $T_p f : T_p M \to T_{f(p)} N$ is surjective, thus $m \geqslant n$, and we can set $\ell := m - n$. The idea now is to choose any chart $(\mathcal{V}, y)$ on N with $y(q) = 0$ and define the first n coordinates over

the neighborhood $f^{-1}(\mathcal{V}) \subset M$ of p by

$$x^i := y^i \circ f, \qquad i = 1, \ldots, n.$$

Writing $\hat{x} := (x^1, \ldots, x^n) : f^{-1}(\mathcal{V}) \to \mathbb{R}^n$, we have $d_p\hat{x} = d_q y \circ T_p f$, thus $d_p\hat{x} : T_p M \to \mathbb{R}^n$ is surjective, which is equivalent to the condition that the n covectors $d_p x^1, \ldots, d_p x^n \in T_p^* M$ are linearly independent.

To define the remaining ℓ coordinates on M near p, first choose an extension of the linearly-independent set $d_p x^1, \ldots, d_p x^n$ to a basis $d_p x^1, \ldots, d_p x^n, \Lambda^{n+1}, \ldots, \Lambda^m$ of $T_p^* M$. For each $i = n+1, \ldots, m$, we can then define a smooth function x^i on a neighborhood of p such that $x^i(p) = 0$ and $d_p x^i = \Lambda^i$; this is another step that would be trivial to carry out if M were the vector space $\mathbb{R}^m$, so the idea is to choose a chart near p and write down suitable functions in local coordinates. With this done, Lemma 4.3 implies that after possibly shrinking to a smaller neighborhood $\mathcal{U} \subset M$ of p, $x = (x^1, \ldots, x^m)$ becomes a smooth chart with the desired properties. $\qquad\square$

REMARK 4.12. For a continuous map $f : M \to N$ between topological manifolds, one can define f to be a *topological immersion* or *topological submersion* at $p \in M$ if there exist continuous charts near p and $q := f(p)$ in which f satisfies the coordinate formula in Theorem 4.11. Note that without having at least one continuous derivative at our disposal, there is no alternative way to characterize either of these conditions in terms of a tangent map being injective or surjective, nor is there any inverse function theorem available for proving such statements. On the other hand, Theorem 4.11 does make sense in the setting of C^k-manifolds for any $k \in \mathbb{N}$; in this case one must assume that $f : M \to N$ is of class C^k, and the resulting charts will be as well. (One should not be fooled by the fact that f will then *look* like a smooth map with respect to those charts—if $k < \infty$, it will not look smooth after arbitrary changes of C^k-coordinates.)

4.5. Embeddings and regular level sets. We now have enough technology to produce many more examples of submanifolds.

DEFINITION 4.13. A smooth map $f : M \to N$ is called an **embedding** (*Einbettung*) if it is an injective immersion whose inverse $f(M) \xrightarrow{f^{-1}} M$ is also continuous. The notation

$$f : M \hookrightarrow N$$

is sometimes used to indicate that f is an embedding.

The typical example of an embedding is the natural inclusion $M \hookrightarrow N$ that exists whenever M is a submanifold of N. The next result states that, up to diffeomorphism, all examples are this one.

THEOREM 4.14. *If $f : M \to N$ is an embedding, then its image $f(M)$ is a smooth submanifold of N.*

PROOF. Suppose $q \in f(M)$. By injectivity, there is a unique point $p \in M$ such that $f(p) = q$, and Theorem 4.11 provides charts $(\mathcal{U}, x)$ on M and $(\mathcal{V}, y)$ on N with $x(p) = 0$ and $y(q) = 0$ such that $y \circ f \circ x^{-1}$ takes the form $(x^1, \ldots, x^m) \mapsto (x^1, \ldots, x^m, 0, \ldots, 0)$. Since the inverse $f(M) \to M$ is also continuous, we are free to assume after possibly shrinking $\mathcal{V} \subset N$ to a smaller neighborhood of q that

$$f^{-1}(\mathcal{V} \cap f(M)) \subset \mathcal{U},$$

or in other words, $\mathcal{V} \cap f(M) = f(\mathcal{U})$. This proves that $(\mathcal{V}, y)$ is a slice chart for the subset $f(M)$. $\quad\square$

The following consequence appears in some books as an alternative definition of the notion of a submanifold:

COROLLARY 4.15. *A subset $L \subset M$ of a smooth manifold M is a smooth submanifold if and only if it admits a smooth structure for which the inclusion map $L \hookrightarrow M$ is a smooth embedding.* $\square$

It is worth pausing a moment to consider what an immersion $f : M \looparrowright N$ can look like if it is *not* an embedding. Theorem 4.11 implies that every immersion is *locally* an embedding, i.e. for every $p \in M$, one can find a neighborhood $\mathcal{U} \subset M$ of p such that $f|_\mathcal{U} : \mathcal{U} \hookrightarrow N$ is an embedding and $f(\mathcal{U}) \subset N$ is therefore a submanifold. On the other hand, f may fail to be an embedding *globally* because it is not injective, meaning it has self-intersections $f(p) = f(p')$ with $p \neq p'$. The notation "$f : M \looparrowright N$" is meant to evoke this possibility by allowing the arrow to loop around and intersect itself. A classic example of a non-injective immersion is the picture of the Klein bottle in Figure 6, which shows the image of an immersion of a compact smooth 2-manifold into $\mathbb{R}^3$. Images of immersions are sometimes called **immersed submanifolds** in the literature, though I am personally not fond of this terminology,[23] so I will not use it.

For slightly subtler reasons, an injective immersion can also fail to be an embedding:

EXAMPLE 4.16. Let $N = \mathbb{R}^2$ and $M = \mathbb{R} \sqcup (0, \pi)$, and defne the immersion $f : M \looparrowright \mathbb{R}^2$ by

$$f(t) := (t, 0) \qquad \text{for } t \in \mathbb{R},$$
$$f(\theta) := (\cos\theta, \sin\theta) \qquad \text{for } \theta \in (0, \pi).$$

Omitting the points 0 and π from the interval $(0, \pi)$ makes this map an injective immersion, but the inverse $f(M) \xrightarrow{f^{-1}} M$ is discontinuous at the two points $(\pm 1, 0)$, which are precisely the points at which it fails to be a submanifold.

Turning our attention to submersions, we can now state a popular corollary of the implicit function theorem that you may have heard referred to before as the "regular value theorem".

DEFINITION 4.17. For a smooth map $f : M \to N$, $p \in M$ is called a **regular point** (*regulärer Wert*) of f if f is a submersion at p, and a **critical point** (*kritischer Wert*) otherwise. A point $q \in N$ is a **critical value** (*kritischer Wert*) of f if $q = f(p)$ for some critical point p, and q is otherwise called a **regular value** (*regulärer Wert*) of f.

THEOREM 4.18 (implicit function theorem). *For any smooth map $f : M \to N$ with regular value $q \in N$, $L := f^{-1}(q) \subset N$ is a smooth submanifold with $\dim L = \dim M - \dim N$, and its tangent space at any point $p \in L$ is $T_p L = \ker T_p f \subset T_p M$.*

PROOF. For each $p \in L = f^{-1}(q)$, f is by assumption a submersion at p, so Theorem 4.11 provides charts x near p and y near q such that $x(p)$ and $y(q)$ are both the origin in their respective Euclidean spaces and $y \circ f \circ x^{-1}$ becomes the map $(x^1, \ldots, x^m) \mapsto (x^1, \ldots, x^n)$. The zero-set of this map is a neighborhood of p in $f^{-1}(q)$ as seen in the x-coordinates, thus x is a slice chart. To see that $T_p L = \ker T_p f$, observe first that for any path γ in L through p, $f \circ \gamma$ is a constant path at $q \in N$, thus $T_p f([\gamma]) = 0 \in T_q N$, proving $T_p L \subset \ker T_p f$. The rest is dimension counting, as the surjectivity of $T_p f : T_p M \to T_q N$ implies

$$\dim T_p L = \dim L = \dim M - \dim N = \dim T_p M - \dim T_q N = \dim \ker T_p f.$$

$\square$

---

[23] I have two objections to the term "immersed submanifold": first, it sounds as if it should be a type of submanifold, but it isn't. Second, one cannot always uniquely recover the manifold M from the image of an immersion $M \looparrowright N$. For example (the following is only for readers with a background in topology), a closed surface Σ_g of genus $g \geqslant 2$ admits smooth covering maps $\Sigma_h \to \Sigma_g$ by surfaces of arbitrarily large genus h (the degree of the cover will be correspondingly large). If one chooses an embedding of Σ_g into $\mathbb{R}^3$, one obtains a submanifold that is also the image of an immersion $\Sigma_h \looparrowright \mathbb{R}^3$ for arbitrarily large values of h.

Submanifolds of the form $f^{-1}(q) \subset M$ for regular values $q \in N$ are sometimes called **regular level sets** of f. In particular, a submersion $f : M \to N$ is distinguished by the property that *all* of its level sets are regular, and are thus smooth submanifolds.

4.6. Examples. We now have a *very* easy way of proving that simple examples like the unit spheres $S^n \subset \mathbb{R}^{n+1}$ really are smooth submanifolds.

EXAMPLE 4.19. Define $f : \mathbb{R}^{n+1} \to \mathbb{R}$ in terms of the standard Euclidean inner product by $f(x) = |x|^2 = \langle x, x \rangle$. This is a smooth map, with differential at any point $x \in \mathbb{R}^{n+1}$ given by $d_x f(v) = 2\langle x, v \rangle$, so it is a submersion everywhere except at the origin. This makes $S^n = f^{-1}(1)$ into a smooth submanifold of dimension $(n+1) - 1 = n$, so in particular, S^n inherits a natural smooth structure for which the inclusion $S^n \hookrightarrow \mathbb{R}^{n+1}$ is a smooth embedding. The kernel of $d_x f$ at a point $x \in S^n$ is the orthogonal complement of x, hence

$$T_x S^n = x^\perp \subset \mathbb{R}^{n+1}.$$

EXAMPLE 4.20. The smooth map $f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto xy$ has only one critical point, at $(x, y) = (0, 0)$, thus $f^{-1}(t)$ is a smooth submanifold (a hyperbola) for every $t \neq 0$, and so is $f^{-1}(0) \backslash \{(0, 0)\}$, but $f^{-1}(0)$ fails to be a submanifold at the origin.

EXERCISE 4.21. Identifying the torus $\mathbb{T}^2$ with $\mathbb{R}^2 / \mathbb{Z}^2$ via Exercise 3.4, find an explicit formula for an embedding $\mathbb{T}^2 \hookrightarrow \mathbb{R}^3$ whose image looks like Figure 5.

For the next set of exercises, the symbol $\mathbb{F}$ always denotes either the real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$, and we denote the vector space of m-by-n matrices over $\mathbb{F}$ by

$$\mathbb{F}^{m \times n} := \{m\text{-by-}n \text{ matrices over } \mathbb{F}\}.$$

If $\mathbb{F} = \mathbb{R}$, this is a real vector space of dimension mn. In the case $\mathbb{F} = \mathbb{C}$, it is a complex vector space of this same dimension, which means it can also be regarded as a *real* vector space of dimension $2mn$. (Indeed, if V is any complex vector space with complex basis $v_1, \ldots, v_k$, then a basis of V as a *real* vector space is given by $v_1, iv_1, \ldots, v_k, iv_k$.) Since they are vector spaces, $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ carry natural smooth structures and are thus smooth manifolds of dimensions mn and $2mn$ respectively. For $m = n$, there is a distinguished open subset

$$\mathrm{GL}(n, \mathbb{F}) = \left\{ \mathbf{A} \in \mathbb{F}^{n \times n} \mid \mathbf{A} \text{ is invertible} \right\},$$

which is therefore also naturally a smooth manifold of dimension n^2 or (in the complex case) $2n^2$. That $\mathrm{GL}(n, \mathbb{F}) \subset \mathbb{F}^{n \times n}$ is open can be deduced easily from the observation that the determinant

$$\det : \mathbb{F}^{n \times n} \to \mathbb{F}$$

defines a continuous function for which $\mathrm{GL}(n, \mathbb{F}) = \det^{-1}(\mathbb{F} \backslash \{0\})$. In fact, $\det(\mathbf{A})$ is a polynomial in the entries of $\mathbf{A}$, which are all linear functions of $\mathbf{A}$, thus $\det : \mathbb{F}^{n \times n} \to \mathbb{F}$ is a smooth real- or complex-valued function. By Cramer's rule, the function

$$\mathrm{GL}(n, \mathbb{F}) \to \mathrm{GL}(n, \mathbb{F}) : \mathbf{A} \mapsto \mathbf{A}^{-1}$$

is also smooth.

EXERCISE 4.22. The n-dimensional **orthogonal group** $\mathrm{O}(n) \subset \mathbb{R}^{n \times n}$ is the set of all real n-by-n matrices $\mathbf{A}$ with the property

$$\mathbf{A}^T \mathbf{A} = \mathbb{1},$$

where $\mathbb{1}$ is the n-by-n identity matrix and $\mathbf{A}^T$ denotes the *transpose* of $\mathbf{A}$, i.e. if $\mathbf{A}$ has entries A_{ij}, then the corresponding entries of $\mathbf{A}^T$ are A_{ji}. This is precisely the set of all linear transformations $\mathbb{R}^n \to \mathbb{R}^n$ which preserve the Euclidean inner product, which means geometrically that they preserve lengths of vectors and angles between them. We will show in this exercise that $\mathrm{O}(n)$ is a smooth submanifold of $\mathbb{R}^{n \times n}$.

(a) Define the linear subspace consisting of all symmetric matrices,

$$\Sigma(n) := \left\{ \mathbf{A} \in \mathbb{R}^{n \times n} \mid \mathbf{A} = \mathbf{A}^T \right\} \subset \mathbb{R}^{n \times n}.$$

There is a map

$$f : \mathbb{R}^{n \times n} \to \Sigma(n) : \mathbf{A} \mapsto \mathbf{A}^T \mathbf{A},$$

such that the orthogonal group is the level set $\mathrm{O}(n) = f^{-1}(\mathbb{1})$. The entries of $f(\mathbf{A})$ are quadratic functions of the entries of $\mathbf{A}$, thus f is clearly a smooth map. Show that its derivative at any $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the linear map

$$Df(\mathbf{A}) : \mathbb{R}^{n \times n} \to \Sigma(n) : \mathbf{H} \mapsto \mathbf{A}^T \mathbf{H} + \mathbf{H}^T \mathbf{A}.$$

Hint: In theory you can do this by computing all the partial derivatives of f with respect to the entries of $\mathbf{A}$, but it's much, much easier to use the definition of the derivative, i.e. regarding $\mathbb{R}^{n \times n}$ and $\Sigma(n)$ simply as vector spaces, show that a "remainder" formula of the form

$$f(\mathbf{A} + \mathbf{H}) = f(\mathbf{A}) + Df(\mathbf{A})\mathbf{H} + R(\mathbf{H}) \cdot |\mathbf{H}|$$

with $\lim_{\mathbf{H} \to 0} R(\mathbf{H}) = 0$ is satisfied. One useful thing you may want to assume: for a reasonable choice of norm on $\mathbb{R}^{n \times n}$, matrix products satisfy $|\mathbf{AB}| \leqslant |\mathbf{A}||\mathbf{B}|$.

(b) Show that $Df(\mathbf{A})$ is surjective if $\mathbf{A} \in \mathrm{O}(n)$. In fact, you won't even need to assume $\mathbf{A} \in \mathrm{O}(n)$, but it *is* useful to assume that $\mathbf{A}$ is invertible (which is automatically true for orthogonal matrices). It is also *crucial* that the target space is $\Sigma(n)$ rather than the entirety of $\mathbb{R}^{n \times n}$—$Df(\mathbf{A})$ is certainly *not* surjective onto $\mathbb{R}^{n \times n}$.

(c) It follows now from the implicit function theorem that $\mathrm{O}(n)$ is a smooth submanifold of $\mathbb{R}^{n \times n}$. What is its dimension? (For a sanity check I will tell you: $\dim \mathrm{O}(2) = 1$ and $\dim \mathrm{O}(3) = 3$.)

(d) Show that $T_{\mathbb{1}} \mathrm{O}(n) \subset T_{\mathbb{1}} \mathbb{R}^{n \times n} = \mathbb{R}^{n \times n}$ is the space of all *antisymmetric* matrices $\mathbf{H}$, i.e. those which satisfy $\mathbf{H}^T = -\mathbf{H}$.

EXERCISE 4.23. The complex analogue of Exercise 4.22 involves the **unitary group**

$$\mathrm{U}(n) = \left\{ \mathbf{A} \in \mathbb{C}^{n \times n} \mid \mathbf{A}^\dagger \mathbf{A} = \mathbb{1} \right\},$$

where $\mathbf{A}^\dagger$ denotes the Hermitian adjoint of $\mathbf{A}$, defined as the complex conjugate of its transpose. Prove that $\mathrm{U}(n)$ is a smooth submanifold of $\mathbb{C}^{n \times n}$, compute its dimension, and show

$$T_{\mathbb{1}} \mathrm{U}(n) = \left\{ \mathbf{H} \in \mathbb{C}^{n \times n} \mid \mathbf{H}^\dagger = -\mathbf{H} \right\}.$$

EXERCISE 4.24. The **special linear group** over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ is defined by

$$\mathrm{SL}(n, \mathbb{F}) = \left\{ \mathbf{A} \in \mathbb{F}^{n \times n} \mid \det(\mathbf{A}) = 1 \right\}.$$

(a) Show that the derivative of $\det : \mathbb{F}^{n \times n} \to \mathbb{F}$ at $\mathbb{1}$ is given by the **trace** (*Spur*):

$$D(\det)(\mathbb{1})\mathbf{H} = \mathrm{tr}(\mathbf{H}).$$

Hint: Write $\mathbf{H}$ in terms of n column vectors as $\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}$, so

$$\det(\mathbb{1} + t\mathbf{H}) = \det \begin{pmatrix} \mathbf{e}_1 + t\mathbf{v}_1 & \cdots & \mathbf{e}_n + t\mathbf{v}_n \end{pmatrix},$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denotes the standard basis of $\mathbb{F}^n$. Differentiate this expression with respect to t at $t = 0$, using the fact that the determinant of a matrix is a multilinear function of its columns.

(b) Use the relation $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$ to generalize the formula in part (a) to

$$D(\det)(\mathbf{A})\mathbf{H} = \det(\mathbf{A}) \cdot \mathrm{tr}(\mathbf{A}^{-1}\mathbf{H}) \qquad \text{for any} \qquad \mathbf{A} \in \mathrm{GL}(n, \mathbb{F}).$$

(c) Prove that $\mathrm{SL}(n, \mathbb{F})$ is a smooth submanifold of $\mathbb{F}^{n \times n}$, compute its dimension, and show

$$T_{\mathbb{1}} \,\mathrm{SL}(n, \mathbb{F}) = \left\{ \mathbf{H} \in \mathbb{F}^{n \times n} \;\middle|\; \mathrm{tr}(\mathbf{H}) = 0 \right\}.$$

(d) Consider the set of *non-invertible* n-by-n matrices,

$$M := \left\{ \mathbf{A} \in \mathbb{F}^{n \times n} \;\middle|\; \det(\mathbf{A}) = 0 \right\}.$$

Is 0 a regular value of $\det : \mathbb{F}^{n \times n} \to \mathbb{F}$? Is M a submanifold of $\mathbb{F}^{n \times n}$?
Hint: Clearly M contains the trivial matrix $0 \in \mathbb{F}^{n \times n}$. If M is a submanifold, what can you say about the tangent space $T_0 M \subset \mathbb{F}^{n \times n}$? In how many different directions can you find smooth paths $\gamma : (-\epsilon, \epsilon) \to \mathbb{F}^{n \times n}$ through 0 that are contained in M?

EXERCISE 4.25. The **special orthogonal** and **special unitary** groups are defined as

$$\mathrm{SO}(n) = \mathrm{O}(n) \cap \mathrm{SL}(n, \mathbb{R}), \qquad \text{and} \qquad \mathrm{SU}(n) = \mathrm{U}(n) \cap \mathrm{SL}(n, \mathbb{C})$$

respectively. Prove:

(a) $\mathrm{SO}(n)$ is an open (and also closed) subset of $\mathrm{O}(n)$, hence it is a smooth submanifold with the same dimension and $T_{\mathbb{1}} \,\mathrm{SO}(n) = T_{\mathbb{1}} \,\mathrm{O}(n)$.
(b) $\mathrm{SU}(n)$ is a smooth submanifold of $\mathrm{U}(n)$ with $\dim \mathrm{SU}(n) = \dim \mathrm{U}(n) - 1$, and

$$T_{\mathbb{1}} \,\mathrm{SU}(n) = \left\{ \mathbf{H} \in \mathbb{C}^{n \times n} \;\middle|\; \mathbf{H}^{\dagger} = -\mathbf{H} \text{ and } \mathrm{tr}(\mathbf{H}) = 0 \right\}.$$

Hint: Use Exercise 4.9 to show that the determinant defines a smooth map $\det : \mathrm{U}(n) \to S^1$, where S^1 in this case denotes the unit circle in $\mathbb{C}$. Prove that 1 is a regular value of this map.

Finally, we consider an interesting space of matrices that does not form a group, but is nonetheless a manifold.

EXERCISE 4.26. For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and nonnegative integers m, n and $r \leqslant \min\{m, n\}$, let

$$V_r(m, n, \mathbb{F}) := \left\{ \mathbf{A} \in \mathbb{F}^{m \times n} \;\middle|\; \mathrm{rank}(\mathbf{A}) = r \right\}.$$

By the standard formula relating ranks and kernels, $V_r(m, n, \mathbb{F})$ is the set of all m-by-n matrices $\mathbf{A}$ over $\mathbb{F}$ such that $\dim_{\mathbb{F}} \ker \mathbf{A} = n - r$, and the latter condition is also equivalent to $\dim_{\mathbb{F}} \mathrm{coker}\,\mathbf{A} = m - r$, where the **cokernel** of $\mathbf{A}$ is defined from its image $\mathrm{im}(\mathbf{A}) \subset \mathbb{F}^m$ as the quotient space $\mathbb{F}^m / \mathrm{im}(\mathbf{A})$.

Given any $\mathbf{M}_0 \in V_r(m, n, \mathbb{F})$, one can find splittings $\mathbb{F}^n = V \oplus K$ and $\mathbb{F}^m = W \oplus C$ such that $K = \ker \mathbf{M}_0$ and $W = \mathrm{im}\,\mathbf{M}_0$. Regarding any other matrix $\mathbf{M} \in \mathbb{F}^{m \times n}$ as a linear map $\mathbb{F}^n \to \mathbb{F}^m$, these splittings of $\mathbb{F}^n$ and $\mathbb{F}^m$ give rise to a block decomposition

$$\mathbf{M} = \begin{pmatrix} \mathbf{A}(\mathbf{M}) & \mathbf{B}(\mathbf{M}) \\ \mathbf{C}(\mathbf{M}) & \mathbf{D}(\mathbf{M}) \end{pmatrix} : V \oplus K \to W \oplus C,$$

thus defining linear (and therefore smooth) maps $\mathbf{A} : \mathbb{F}^{m \times n} \to \mathrm{Hom}(V, W)$, $\mathbf{B} : \mathbb{F}^{m \times n} \to \mathrm{Hom}(K, W)$, $\mathbf{C} : \mathbb{F}^{m \times n} \to \mathrm{Hom}(V, C)$ and $\mathbf{D} : \mathbb{F}^{m \times n} \to \mathrm{Hom}(K, C)$. By construction, the functions $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ all vanish at $\mathbf{M}_0$, while $\mathbf{A}(\mathbf{M}_0) : V \to W$ is invertible. Observe that the invertible maps in $\mathrm{Hom}(V, W)$ form an open subset; this is true for the same reason that $\mathrm{GL}(n, \mathbb{F})$ is an open subset of $\mathbb{F}^{n \times n}$. We can therefore fix an open neighborhood $\mathcal{O} \subset \mathbb{F}^{m \times n}$ of $\mathbf{M}_0$ such that $\mathbf{A}(\mathbf{M}) : V \to W$ is invertible for all $\mathbf{M} \in \mathcal{O}$, and use this to define two smooth maps $\Phi : \mathcal{O} \to \mathrm{Hom}(K, C)$ and $\Psi : \mathcal{O} \to \mathbb{F}^{n \times n}$ by

$$\Phi(\mathbf{M}) := \mathbf{D}(\mathbf{M}) - \mathbf{C}(\mathbf{M}) \mathbf{A}(\mathbf{M})^{-1} \mathbf{B}(\mathbf{M}), \quad \text{and} \quad \Psi(\mathbf{M}) := \begin{pmatrix} \mathbb{1} & -\mathbf{A}(\mathbf{M})^{-1} \mathbf{B}(\mathbf{M}) \\ 0 & \mathbb{1} \end{pmatrix},$$

where in the latter expression we are regarding $\Psi(\mathbf{M})$ as a linear map $\mathbb{F}^n \to \mathbb{F}^n$ and writing its block decomposition with respect to the splitting $\mathbb{F}^n = V \oplus K$.

(a) Show that $\Psi(\mathbf{M}) \in \mathbb{F}^{n \times n}$ is invertible for every $\mathbf{M} \in \mathcal{O}$.
(b) Show that for every $\mathbf{M} \in \mathcal{O}$, the kernel of the matrix product $\mathbf{M}\Psi(\mathbf{M}) : \mathbb{F}^n \to \mathbb{F}^m$ is
$\{0\} \oplus \ker \Phi(\mathbf{M}) \subset V \oplus K = \mathbb{F}^n$.
(c) Deduce from parts (a) and (b) that $\mathcal{O} \cap V_r(m,n,\mathbb{F}) = \Phi^{-1}(0)$.
Hint: What is the largest dimension that $\ker \mathbf{M}$ *can have for* $\mathbf{M} \in \mathcal{O}$?
(d) Show that $\mathbf{M}_0$ is a regular point of Φ, and deduce from this that $V_r(m,n,\mathbb{F}) \subset \mathbb{F}^{m \times n}$ is a smooth submanifold with

$$T_{\mathbf{M}} V_r(m,n,\mathbb{F}) = \left\{ \mathbf{H} \in \mathbb{F}^{m \times n} \mid \mathbf{H}(\ker \mathbf{M}) \subset \operatorname{im} \mathbf{M} \right\}$$

for every $\mathbf{M} \in V_r(m,n,\mathbb{F})$, and

$$\dim V_r(m,n,\mathbb{R}) = mn - (m-r)(n-r), \qquad \dim V_r(m,n,\mathbb{C}) = 2 \dim V_r(m,n,\mathbb{R}).$$

(e) A matrix $\mathbf{M} \in \mathbb{F}^{m \times n}$ is said to have **maximal rank** if its rank is $\min\{m,n\}$, which means it is either injective or surjective. Deduce from the result of part (d) that the set of maximal rank matrices is open and dense in $\mathbb{F}^{m \times n}$.

The result of this exercise produces what is called a **stratification** *of* $\mathbb{F}^{m \times n}$, *meaning that it decomposes* $\mathbb{F}^{m \times n}$ *into a collection of smooth submanifolds of various dimensions such that every matrix belongs to exactly one of them.*

5. Vector fields

A **vector field** (*Vektorfeld*) X on a smooth manifold M associates to every point $p \in M$ a vector in the corresponding tangent space,

$$X(p) \in T_p M.$$

For example, on $S^2 \subset \mathbb{R}^3$, the tangent space $T_p S^2$ is the orthogonal complement of the vector $p \in S^2 \subset \mathbb{R}^3$, thus a vector field associates to each such point another vector that is orthogonal to it. We say that a vector field X is **smooth** if the map

$$M \to TM : p \mapsto X(p)$$

is smooth. The set of all smooth vector fields on M forms a vector space, which we will denote by

$$\mathfrak{X}(M) := \left\{ X \in C^\infty(M, TM) \mid X(p) \in T_p M \text{ for every } p \in M \right\}.$$

As with real-valued functions, one can define the **support** (*Träger*) of a vector field X as the closure in M of the set $\{p \in M \mid X(p) \neq 0\}$.

5.1. The flow of a vector field. The most important fact about vector fields on manifolds is that they determine dynamical systems. For a smooth path $\gamma : (a,b) \to M$, the derivative

$$\dot{\gamma}(t) := \frac{d\gamma}{dt}(t) \in T_{\gamma(t)} M$$

can be defined for each $t \in (a,b)$ as a special case of our definition of *partial* derivatives in §3.4. In important special cases such as when M is a submanifold of $\mathbb{R}^n$, $\dot{\gamma}(t)$ means exactly what you think it should; more generally, it is the equivalence class $[\gamma_t]$ represented by the reparametrized path $\gamma_t(s) := \gamma(t+s)$ that passes through $\gamma(t)$ at $s = 0$. Given $X \in \mathfrak{X}(M)$, a path $\gamma : (a,b) \to M$ is called a **flow line** or **orbit** of X if it satisfies

$$\dot{\gamma}(t) = X(\gamma(t)).$$

The following fundamental result translates most of the basic existence/uniqueness theory for ordinary differential equations into the language of differential geometry.

THEOREM 5.1. *For any smooth vector field $X \in \mathfrak{X}(M)$ on a manifold M, there exists a unique open subset $\mathcal{O} \subset \mathbb{R} \times M$ and smooth map*

$$\mathcal{O} \to M : (t, p) \mapsto \varphi_X^t(p),$$

*called the **flow** (Fluss) of X, such that for every $p \in M$, the set*

$$\ell_p := \left\{ t \in \mathbb{R} \mid (t, p) \in \mathcal{O} \right\} \subset \mathbb{R}$$

is an open interval containing 0, and

$$\gamma_p : \ell_p \to M : t \mapsto \varphi_X^t(p)$$

is the maximal solution to the initial value problem

$$\dot{\gamma}(t) = X(\gamma(t)), \qquad \gamma(0) = p.$$

Moreover, if X has compact support, then $\mathcal{O} = \mathbb{R} \times M$.

PROOF. For the most part, this result is proved by choosing local coordinates so as to rewrite the initial value problem in $\mathbb{R}^n$ and then applying standard results from the theory of ODEs. We will merely add a few observations in order to see how this works. First, given $p_0 \in M$, choose a smooth chart $(\mathcal{U}, x)$ with $p_0 \in \mathcal{U}$, which gives rise to a smooth chart $(T\mathcal{U}, Tx)$ on TM. The smoothness of X means that $p \mapsto Tx(X(p)) = (x(p), d_p x(X(p)))$ is a smooth function $\mathcal{U} \to \mathbb{R}^{2n}$, thus in particular, so is the function

$$\Phi : \mathcal{U} \to \mathbb{R}^n : p \mapsto d_p x(X(p)).$$

A path $\gamma : (-\epsilon, \epsilon) \to \mathcal{U}$ with $\gamma(0) = p_0$ will now satisfy $\dot{\gamma}(t) = X(\gamma(t))$ if and only if

$$(x \circ \gamma)'(t) = d_{\gamma(t)} x(\dot{\gamma}(t)) = d_{\gamma(t)} x(X(\gamma(t))),$$

meaning that $\alpha := x \circ \gamma : (-\epsilon, \epsilon) \to x(\mathcal{U}) \subset \mathbb{R}^n$ must be a solution to the initial value problem

(5.1) $$\dot{\alpha}(t) = F(\alpha(t)), \qquad \alpha(0) = x(p_0),$$

where we define $F : x(\mathcal{U}) \to \mathbb{R}^n$ by

$$F(q) := d_{x^{-1}(q)} x(X(x^{-1}(q))) = \Phi \circ x^{-1}(q).$$

This last expression shows that F is a smooth function, so in particular it is Lipschitz, and the Picard-Lindelöf theorem therefore applies, telling us that a solution $\alpha : (-\epsilon, \epsilon) \to x(\mathcal{U})$ to (5.1) exists for some $\epsilon > 0$ and is unique. Since F is smooth, this solution also depends smoothly on the initial point $x(p_0)$. Replacing α with $\gamma = x^{-1} \circ \alpha : (-\epsilon, \epsilon) \to \mathcal{U}$, we similarly obtain existence and uniqueness of a solution to $\dot{\gamma}(t) = X(\gamma(t))$ with $\gamma(0) = p_0$, along with smooth dependence on the point p_0. This uniquely defines the flow map $(t, p) \mapsto \varphi_X^t(p)$ for all (t, p) in some neighborhood of $\{0\} \times M \subset \mathbb{R} \times M$.

It remains to establish that the flow map has a unique extension to a maximal domain which is an open subset $\mathcal{O} \subset \mathbb{R} \times M$, and is all of $\mathbb{R} \times M$ if X has compact support. This follows via the same tricks that are used to prove the corresponding statement in $\mathbb{R}^n$, e.g. whenever a flow line $\gamma : [0, T] \to M$ with $\gamma(0) = p_0$ exists, one can find a finite partition $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$ such that the subintervals $[t_{j-1}, t_j]$ are each sufficiently small for $\gamma([t_{j-1}, t_j])$ to lie within the domain of a single chart. One can then make use of the formula

$$\gamma(T) = \varphi_X^T(p_0) = \varphi_X^{t_N - t_{N-1}} \circ \ldots \circ \varphi_X^{t_2 - t_1} \circ \varphi_X^{t_1}(p_0),$$

in which each map in the composition is already known to be smooth and defined on an open neighborhood of the relevant point as long as the increments $t_j - t_{j-1}$ are small enough. This establishes that $\mathcal{O} \subset \mathbb{R} \times M$ is open and $(t, p) \mapsto \varphi_X^t(p)$ is smooth. Finally, if the support $K \subset M$ of X is a compact subset, then clearly every flow line through a point $p_0 \in M \backslash K$ is constant, so that

$(t, p_0) \in \mathcal{O}$ for all $t \in \mathbb{R}$. For the same reason, uniqueness of solutions implies that a flow line with initial value at a point $p_0 \in K$ can never escape from K; if it did, then it would become constant outside of K, and must therefore have *always* been a constant path outside of K. We claim now that for every $p_0 \in K$, the maximal solution to $\dot{\gamma}(t) = X(\gamma(t))$ with $\gamma(0) = p_0$ is defined for all $t \in \mathbb{R}$. If not, then suppose $\gamma : (a, b) \to M$ is the maximal solution and either $a > -\infty$ or $b < \infty$; for concreteness we will assume the latter since there is no substantial difference between the two cases. Then (a, b) contains a sequence t_j with $t_j \to b$, and after restricting to a subsequence, the compactness of K implies that we can assume $\gamma(t_j)$ converges to some point $p_1 \in K$. But solutions to the initial value problem starting at points near p_1 also exist and are unique on some sufficiently small interval, so for j large enough, $\gamma(t_j)$ must eventually lie on one of these solutions. The only way to have $\gamma(t_j) \to p_1$ is then if γ eventually matches (up to parametrization) the unique flow line through p_1, in which case it must reach that point at time $t = b$ and can be continued past it; this contradicts the assumption that γ could not be extended beyond the interval (a, b). $\qquad\square$

We say that $X \in \mathfrak{X}(M)$ admits a **global flow** if the domain $\mathcal{O} \subset \mathbb{R} \times M$ of the flow map $(t, p) \mapsto \varphi_X^t(p)$ is $\mathbb{R} \times M$. This can sometimes be true even if X does not have compact support, e.g. it is easy to show that every C^0-bounded smooth vector field on $\mathbb{R}^n$ has a global flow. (There are also easy counterexamples if X is allowed to be unbounded, such as $X(x) := x^2$ on $\mathbb{R}$.) In the general case, φ_X^t defines for each $t \in \mathbb{R}$ a smooth map $\mathcal{O}_X^t \to M$ on the open set

$$\mathcal{O}_X^t := \left\{ p \in M \mid (t, p) \in \mathcal{O} \right\},$$

and in fact, φ_X^t is a diffeomorphism from $\mathcal{O}_X^t$ to $\mathcal{O}_X^{-t}$, with inverse

$$(\varphi_X^t)^{-1} = \varphi_X^{-t}.$$

In particular, if the flow is global, then $\mathcal{O}_X^t = M$ for each $t \in \mathbb{R}$, and φ_X^t is therefore a diffeomorphism from M to itself. It is also possible however to have $\mathcal{O}_X^t = \varnothing$ for $t \neq 0$, though this cannot happen when t is close to 0. Indeed, it follows directly from the definition that

$$\mathcal{O}_X^s \supset \mathcal{O}_X^t \qquad \text{whenever} \quad 0 \leqslant s \leqslant t \quad \text{or} \quad t \leqslant s \leqslant 0,$$

and short-time existence of solutions also implies

$$\mathcal{O}_X^0 = \bigcup_{t>0} \mathcal{O}_X^t = \bigcup_{t<0} \mathcal{O}_X^t = M.$$

The most important properties of the flow are perhaps

$$\varphi_X^0 = \mathrm{Id}, \qquad \text{and} \qquad \varphi_X^{s+t} = \varphi_X^s \circ \varphi_X^t \quad \text{on} \quad \mathcal{O}_X^s \cap \mathcal{O}_X^t \cap \mathcal{O}_X^{s+t} \text{ for every } s, t \in \mathbb{R},$$

which follow from the uniqueness of solutions to the initial value problem. Whenever the flow is global, this means that the map $t \mapsto \varphi_X^t$ defines a group homomorphism from $\mathbb{R}$ to the group $\mathrm{Diff}(M)$ of diffeomorphisms $M \to M$. This is, in practice, the single easiest way to produce a diffeomorphism on a manifold: one need not write it down explicitly, but can instead often write down an appropriate vector field more-or-less explicitly and deduce the existence of a suitable diffeomorphism via its flow. The following exercise is a demonstration of this technique:

EXERCISE 5.2. A manifold M is called **connected** (*zusammenhängend*)[24] if for every pair of points $p, q \in M$, there exists a continuous path $\gamma : [0, 1] \to M$ from $\gamma(0) = p$ to $\gamma(1) = q$. Show that under this assumption, there exists a diffeomorphism $\varphi : M \to M$ that is the identity map outside of a compact subset and satisfies $\varphi(p) = q$.

---

[24]If you know some topology, you may notice that what we are defining here is actually the notion of a **path-connected** space, and connectedness (without mentioning paths) usually means something else. However, every manifold is *locally* path-connected, so a general theorem from point-set topology (see [**Wen18**, Theorem 7.17]) implies that connectedness and path-connectedness on a manifold are equivalent conditions.

Hint: You should first convince yourself that the path $\gamma : [0,1] \to M$ can be assumed to be a smooth embedding without loss of generality. (This is obvious if γ happens to lie in the domain of a chart $(\mathcal{U}, x)$ such that $x(\mathcal{U}) \subset \mathbb{R}^n$ is convex, and notice that $\gamma([0,1]) \subset M$ can always be covered by finitely many such charts.) Then choose a vector field that has a flow line containing this path.

REMARK 5.3. If the vector field X is not smooth but is of class C^k for some $k \in \mathbb{N}$, then the proof of Theorem 5.1 above can be adapted to produce a flow map $(t,p) \mapsto \varphi_X^t(p)$ that is also of class C^k. As you may recall from your analysis courses, all bets are off if X is continuous but not C^1: in this case local solutions exist but may not be unique, so the flow cannot be defined.

5.2. Pullbacks and pushforwards. A diffeomorphism

$$\psi : M \to N$$

between two manifolds can be viewed as a way of "translating" all geometric data from M into equivalent geometric data on N or vice versa. The exact mechanism for the translation depends on the kind of data we are talking about: for points $p \in M$, the translation in N is simply $\psi(p) \in N$. For a function $f \in C^\infty(M)$, the equivalent data on N is a function

$$\psi_* f \in C^\infty(N)$$

that has the same value at the equivalent point $\psi(p)$ that f has at the original point p, thus

$$\psi_* f \circ \psi = f, \qquad \text{or equivalently} \qquad \psi_* f = f \circ \psi^{-1}.$$

We call $\psi_* f$ the **pushforward** of f via the diffeomorphism ψ. This process is invertible: one can associate to any $f \in C^\infty(N)$ a **pullback**

$$\psi^* f \in C^\infty(M)$$

via ψ, which takes the same value at p that f takes at $\psi(p)$; the definition is thus

$$\psi^* f = f \circ \psi.$$

To do the same trick with tangent vectors, we need to recall that the tangent map of a diffeomorphism $\psi : M \to N$ is also a diffeomorphism $T\psi : TM \to TN$, one which sends $T_p M$ isomorphically to $T_{\psi(p)} N$ for each $p \in M$. This gives the natural way of "translating" tangent vectors between M and N, so for each $X \in TM$ and $Y \in TN$, we denote

$$\psi_* X := T\psi(X) \in TN, \qquad \psi^* Y := T\psi^{-1}(Y) \in TM.$$

The pushforward of a vector field $X \in \mathfrak{X}(M)$ should then be a vector field

$$\psi_* X \in \mathfrak{X}(N)$$

whose value at $\psi(p)$ for each $p \in M$ is the corresponding translation of the tangent vector $X(p)$, namely $\psi_*(X(p))$. This gives

$$(\psi_* X) \circ \psi = T\psi \circ X, \qquad \text{or equivalently} \qquad \psi_* X = T\psi \circ X \circ \psi^{-1}.$$

The pullback of a vector field $Y \in \mathfrak{X}(N)$ is obtained by inverting this procedure, thus

$$\psi^* Y := T\psi^{-1} \circ Y \circ \psi \in \mathfrak{X}(M).$$

PROPOSITION 5.4. *Suppose $\psi : M \to N$ is a diffeomorphism, $X \in \mathfrak{X}(N)$ is a vector field, and $t \in \mathbb{R}$. Then a point $p \in M$ is in the domain of the flow $\varphi_{\psi_* X}^t$ if and only if $\psi(p)$ belongs to the domain of φ_X^t, and $\psi \circ \varphi_{\psi_* X}^t = \varphi_X^t \circ \psi$.*

PROOF. The result follows from the observation that ψ provides a natural bijective correspondence between the flow lines of X on N and flow lines of ψ^*X on M. Indeed, suppose $a < 0 < b$ and $\gamma : (a,b) \to N$ is a flow line of X, satisfying $\dot{\gamma}(t) = X(\gamma(t))$ and $\gamma(0) = q := \psi(p)$. Then $\alpha := \psi^{-1} \circ \gamma : (a,b) \to M$ satisfies $\alpha(0) = p$ and

$$\dot{\alpha}(t) = T\psi^{-1}\big(\dot{\gamma}(t)\big) = T\psi^{-1}\big(X(\gamma(t))\big) = T\psi^{-1} \circ X \circ \psi(\alpha(t)) = (\psi^*X)(\alpha(t)).$$

Conversely, the same computation implies that if α is a flow line of ψ^*X, then $\gamma := \psi \circ \alpha$ is a flow line of X. $\square$

EXERCISE 5.5. For two diffeomorphisms $\psi : M \to N$ and $\varphi : N \to Q$, prove the following relations:

(a) $(\varphi \circ \psi)_* f = \varphi_*(\psi_* f) \in C^\infty(Q)$ for $f \in C^\infty(M)$.
(b) $(\varphi \circ \psi)^* g = \psi^*(\varphi^* g) \in C^\infty(M)$ for $g \in C^\infty(Q)$.
(c) $(\varphi \circ \psi)_* X = \varphi_*(\psi_* X) \in \mathfrak{X}(Q)$ for $X \in \mathfrak{X}(M)$.
(d) $(\varphi \circ \psi)^* Y = \psi^*(\varphi^* Y) \in \mathfrak{X}(M)$ for $Y \in \mathfrak{X}(Q)$.

We will see later that when $\psi : M \to N$ is a diffeomorphism, pullbacks and pushforwards can be defined for any meaningful geometric data one might want to consider on M or N. A special case that arises quite often is where $M = N$ and $\psi : M \to M$ is defined by the flow of a vector field; we will see an example of this in the next lecture when we discuss the Lie derivative of a vector field. It will also be important to know that for certain types (but not all types) of data, either the pushforward *or* the pullback (but not both) can be defined via arbitrary smooth maps $\psi : M \to N$, not only for diffeomorphisms. One example of this is already apparent: for $f \in C^\infty(N)$, the pullback

$$\psi^* f := f \circ \psi \in C^\infty(M)$$

makes sense for any smooth map $\psi : M \to N$, so M and N need not be diffeomorphic. One cannot similarly define pushforwards of functions in this context, since ψ^{-1} might not be defined. We will see many more examples of this phenomenon when we discuss tensors and differential forms.

5.3. Derivations. For real-valued functions $f : M \to \mathbb{R}$, there is no natural notion of "partial derivatives" of f, unless M happens to be an open subset of $\mathbb{R}^n$. It is still natural however to talk about the **directional derivative** (*Richtungsableitung*) of f at a point $p \in M$ with respect to a tangent vector $X \in T_pM$: this is known as the **Lie derivative** (*Lie-Ableitung*) $\mathcal{L}_X f \in \mathbb{R}$ of f with respect to X, and can be evaluated using the differential df, i.e.

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

If X is not just a tangent vector at a single point but a smooth vector field, then the Lie derivative defines another smooth function $M \to \mathbb{R}$, written

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

The differential operator $\mathcal{L}_X$ associated to any $X \in \mathfrak{X}(M)$ thus defines a map

$$\mathcal{L}_X : C^\infty(M) \to C^\infty(M),$$

and one can check using the usual rules of differentiation that this map is linear:

$$\mathcal{L}_X(f + g) = \mathcal{L}_X f + \mathcal{L}_X g, \qquad \mathcal{L}_X(cf) = c\mathcal{L}_X f, \qquad \text{for all } f, g \in C^\infty(M), c \in \mathbb{R}.$$

Moreover, the product rule for differentiation translates into the following so-called **Leibniz rule**:

$$\mathcal{L}_X(fg) = (\mathcal{L}_X f)g + f\mathcal{L}_X g.$$

This formula motivates a short digression on algebras and Lie algebras.

DEFINITION 5.6. An **algebra** is a vector space $\mathcal{A}$ that is endowed with the additional structure of a bilinear multiplication operation

$$\mathcal{A} \times \mathcal{A} \to \mathcal{A} : (x, y) \mapsto xy$$

that is also associative, i.e. $(xy)z = x(yz)$ for all $x, y, z \in \mathcal{A}$.[25] A **derivation** on $\mathcal{A}$ is a linear map $L : \mathcal{A} \to \mathcal{A}$ that satisfies the Leibniz rule

$$L(xy) = (Lx)y + x(Ly) \qquad \text{for all } x, y \in \mathcal{A}.$$

An algebra endowed with a derivation is called a **differential algebra** (*Differentialalgebra*).

DEFINITION 5.7. A **Lie algebra** (*Lie-Algebra*) is a vector space V that is endowed with the additional structure of a bilinear operation

$$[\cdot, \cdot] : V \times V \to V,$$

its so-called **Lie bracket** (*Lie-Klammer*), which satisfies:
- **antisymmetry**: $[u, v] = -[v, u]$ for all $u, v \in V$;
- the **Jacobi identity**: $[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0$ for all $u, v, w \in V$.

EXERCISE 5.8. Show that on any algebra $\mathcal{A}$, the space $\mathcal{D}$ of all derivations on $\mathcal{A}$ can be made into a Lie algebra by defining the bracket

$$[L_1, L_2] := L_1 \circ L_2 - L_2 \circ L_1.$$

In this course, the most important example of an algebra is the space of smooth real-valued functions $C^\infty(M)$ on a manifold M, in which multiplication is defined pointwise by $(fg)(p) := f(p)g(p)$. The previous remarks show that for any smooth vector field $X \in \mathfrak{X}(M)$, the associated Lie derivative operator $\mathcal{L}_X$ defines a derivation on $C^\infty(M)$. A somewhat less obvious class of examples comes from the observation in Exercise 5.8 that the **commutator bracket** of any two derivations is also a derivation, so in particular, any pair of vector fields $X, Y \in \mathfrak{X}(M)$ gives rise to a derivation on $C^\infty(M)$ defined by

$$[\mathcal{L}_X, \mathcal{L}_Y]f = \mathcal{L}_X \mathcal{L}_Y f - \mathcal{L}_Y \mathcal{L}_X f.$$

One says that the vector fields X and Y **commute** (*kommutieren*) whenever this bracket vanishes. This will turn out to be an important condition, but its meaning will take some effort to unpack. We first need to make the surprising and useful observation that the examples we have seen so far of derivations on $C^\infty(M)$ are the *only* examples that exist:

THEOREM 5.9. *Every derivation* $L : C^\infty(M) \to C^\infty(M)$ *is of the form* $L = \mathcal{L}_X$ *for some (unique) smooth vector field* $X \in \mathfrak{X}(M)$.

PROOF. The uniqueness of X is clear, since different vector fields define different derivations. The proof of existence follows from a series of claims.

Claim 1: *If* $f : M \to \mathbb{R}$ *is a constant function, then* $Lf = 0$ *for every derivation* L *on* $C^\infty(M)$.

Indeed, if f is constant, then multiplication of an arbitrary function $g \in C^\infty(M)$ by f is the same as scalar multiplication, so linearity implies $L(fg) = f Lg$, and combining this with the Leibniz rule gives $(Lf)g = 0$. Plugging in the function $g \equiv 1$, we conclude $Lf \equiv 0$.

---

[25]If you're into algebra, you may notice that the definition of an algebra is quite similar to that of a ring. The difference is that while a ring is also an abelian group with respect to its "+" operation and has a distributive product operation, it does not generally come with any notion of scalar multiplication and is thus not a vector space. One can however define the notion of an algebra more generally, so that it is a module over a commutative ring R instead of a vector space. The case where R is a field then agrees with the definition we've given, but one can also speak of an algebra over $\mathbb{Z}$, which is the same thing as a ring since modules over $\mathbb{Z}$ are the same thing as abelian groups.

Claim 2: *The stated result is true in the special case where M is a convex open subset of Euclidean space, $\Omega \subset \mathbb{R}^n$.*

This is the heart of the proof, and it depends on an important fact in first-year analysis that follows from the fundamental theorem of calculus. Assume $\Omega \subset \mathbb{R}^n$ is open and convex, and fix a point $x_0 = (x_0^1, \ldots, x_0^n) \in \Omega$. For any other point $x = (x^1, \ldots, x^n) \in \Omega$, the convexity of Ω implies that it contains the line segment between x_0 and x, so using the fundamental theorem of calculus and the chain rule, we find that any smooth function $f : \Omega \to \mathbb{R}$ satisfies

$$\begin{aligned}
(5.2) \quad f(x) &= f(x_0) + \int_0^1 \frac{d}{d\tau} f(x_0 + \tau(x - x_0)) \, d\tau = f(x_0) + \int_0^1 Df(x_0 + \tau(x - x_0))(x - x_0) \, d\tau \\
&= f(x_0) + \sum_{j=1}^n \left(\int_0^1 \partial_j f(x_0 + \tau(x - x_0)) \, d\tau \right) (x^j - x_0^j) =: f(x_0) + \sum_{j=1}^n h_j(x)(x^j - x_0^j),
\end{aligned}$$

where we've defined smooth functions $h_j : \Omega \to \mathbb{R}$ by $h_j(x) := \int_0^1 \partial_j f(x_0 + \tau(x - x_0)) \, d\tau$. To make use of this formula, we can regard each of the coordinates $x^1, \ldots, x^n$ as smooth real-valued functions on Ω and associate to these the smooth functions

$$X^j := L(x^j) \in C^\infty(\Omega), \qquad j = 1, \ldots, n.$$

Linearity and the Leibniz rule, together with Claim 1, now produce from (5.2) the formula $Lf(x) = \sum_{j=1}^n \left[Lh_j(x) \cdot (x^j - x_0^j) + h_j(x) X^j(x) \right]$, so in particular,

$$Lf(x_0) = \sum_{j=1}^n h_j(x_0) X^j(x_0) = \sum_{j=1}^n X^j(x_0) \partial_j f(x_0).$$

The definition of the functions $X^j \in C^\infty(\Omega)$ did not depend on the choice of point $x_0 \in \Omega$, thus this formula is valid for every such point, giving an equality of functions

$$Lf = \sum_{j=1}^n X^j \partial_j f = \mathcal{L}_X f \qquad \text{on } \Omega,$$

where we define the smooth vector field $X \in \mathfrak{X}(\Omega)$ by $X(x) = (X^1(x), \ldots, X^n(x)) \in \mathbb{R}^n = T_x\Omega$.

Claim 3: *If the theorem holds for a particular manifold M, then it also holds for every manifold that is diffeomorphic to M.*

Assume $\psi : N \to M$ is a diffeomorphism between two manifolds, and the theorem is already known to hold for M. Any derivation L on $C^\infty(N)$ then determines a "pushforward" derivation $\psi_* L$ on $C^\infty(M)$ via the formula

$$(5.3) \qquad (\psi_* L)f := L(f \circ \psi) \circ \psi^{-1}.$$

By assumption, the latter is $\mathcal{L}_X$ for some vector field $X \in \mathfrak{X}(M)$, and it is reasonable to guess that L will therefore correspond to the pullback vector field $\psi^* X \in \mathfrak{X}(N)$ as defined in §5.2. Let's check this: $\psi^* X$ is defined by

$$\psi^* X(p) = T\psi^{-1}(X(\psi(p))).$$

For $g \in C^\infty(N)$ and $p \in N$, we define $f := g \circ \psi^{-1} \in C^\infty(M)$ and use (5.3) to write

$$\begin{aligned}
(Lg)(p) &= L(f \circ \psi)(p) = [(\psi_* L)f](\psi(p)) = (\mathcal{L}_X f)(\psi(p)) = df(X(\psi(p))) \\
&= d(g \circ \psi^{-1})(X(\psi(p))) = dg \circ T\psi^{-1}(X(\psi(p))) = dg(\psi^* X(p)) = \mathcal{L}_{\psi^* X} g(p),
\end{aligned}$$

so the guess is correct!

For the remaining claims, assume M is a fixed manifold and $L : C^\infty(M) \to C^\infty(M)$ is a derivation.

Claim 4: *If $f \in C^\infty(M)$ vanishes on a neighborhood of some point $p \in M$, then $Lf(p) = 0$.*

To see this, suppose $\mathcal{U} \subset M$ is a neighborhood of p on which $f \in C^\infty(M)$ vanishes, and choose any $g \in C^\infty(M)$ so that $g(p) = 1$ but g has compact support in $\mathcal{U}$.[26] Then $fg = 0$, thus $0 = (Lf)g + f(Lg)$, and evaluating the right hand side at p gives $0 = Lf(p) \cdot g(p) = Lf(p)$.

In light of linearity, a corollary of Claim 4 is that for any $f \in C^\infty(M)$, the value of $Lf(p)$ at any given point $p \in M$ depends only on the values of f on an arbitrarily small neighborhood of p.

Claim 5: *For any open subset $\mathcal{U} \subset M$, L determines a unique derivation $L_\mathcal{U} : C^\infty(\mathcal{U}) \to C^\infty(\mathcal{U})$ such that for every $f \in C^\infty(M)$, $L_\mathcal{U}(f|_\mathcal{U}) = (Lf)|_\mathcal{U}$.*

This follows from the observation at the end of Claim 4 that $Lf(p)$ depends on f only in a neighborhood of p. Indeed, for any $f \in C^\infty(\mathcal{U})$, there is a unique function $L_\mathcal{U} f \in C^\infty(\mathcal{U})$ characterized by the property that for each $p \in \mathcal{U}$ and $f_p \in C^\infty(M)$ with $f_p \equiv f$ near p, $L_p f \equiv L_\mathcal{U} f$ near p. It is straightforward to verify that $L_\mathcal{U}$ defined in this way is a derivation.

Conclusion: Choose an open cover $M = \bigcup_{\alpha \in I} \mathcal{U}_\alpha$ such that for every α, there is a chart $(\mathcal{U}_\alpha, x_\alpha)$ whose image $x(\mathcal{U}_\alpha) \subset \mathbb{R}^n$ is convex. Claims 2 and 3 imply that the theorem holds for each of the open subsets $\mathcal{U}_\alpha \subset M$, thus for the derivation L_α determined on $C^\infty(\mathcal{U}_\alpha)$ by Claim 5, we have $L_\alpha = \mathcal{L}_{X_\alpha}$ for some vector field $X_\alpha \in \mathfrak{X}(\mathcal{U}_\alpha)$. We claim that for every pair $\alpha, \beta \in I$, X_α and X_β match on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$. Indeed, if $X_\alpha(p) \neq X_\beta(p)$ for some point p, then we can find a function $f \in C^\infty(M)$ with compact support in $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ such that $\mathcal{L}_{X_\alpha} f(p) \neq \mathcal{L}_{X_\beta} f(p)$, which is a contradiction since $L_\alpha(f|_{\mathcal{U}_\alpha})$ and $L_\beta(f|_{\mathcal{U}_\beta})$ should both have the same restriction as Lf on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$. The claim now implies that the vector fields X_α can be patched together to form a smooth vector field $X \in \mathfrak{X}(M)$, and in light of Claim 4, the relation $Lf = \mathcal{L}_X f$ now follows on each $\mathcal{U}_\alpha$ from $L(f|_{\mathcal{U}_\alpha}) = \mathcal{L}_{X_\alpha}(f|_{\mathcal{U}_\alpha})$. $\square$

REMARK 5.10. In light of Theorem 5.9, it is common in differential geometry to blur the distinction between smooth vector fields on M and derivations on $C^\infty(M)$, and many books even use exactly the same notation for both, thus writing

$$Xf := \mathcal{L}_X f \in C^\infty(M)$$

so as to view the vector field $X \in \mathfrak{X}(M)$ as a differential operator acting on the function $f \in C^\infty(M)$. I personally prefer not to do this, and will thus continue writing $\mathcal{L}_X$ to distinguish the derivation defined by a vector field $X \in \mathfrak{X}(M)$ from the vector field itself; the sole exception to this will be the coordinate vector fields discussed in the next subsection. Many authors would probably call this practice overly pedantic, and I cannot say with confidence that they are wrong.

EXERCISE 5.11. For a diffeomorphism $\psi : M \to N$, vector field $X \in \mathfrak{X}(M)$ and function $f \in C^\infty(M)$, prove $\mathcal{L}_{\psi_* X}(\psi_* f) = \psi_*(\mathcal{L}_X f) \in C^\infty(N)$.

5.4. Coordinate vector fields. Given a smooth chart $(\mathcal{U}, x)$ on a manifold M, the coordinate functions $x^1, \ldots, x^n : \mathcal{U} \to \mathbb{R}$ define a natural family of derivations on $C^\infty(\mathcal{U})$, namely the n partial derivative operators

$$\partial_j := \frac{\partial}{\partial x^j} : C^\infty(\mathcal{U}) \to C^\infty(\mathcal{U}), \qquad j = 1, \ldots, n,$$

which are defined by writing any function $f \in C^\infty(\mathcal{U})$ in its local coordinate representation $(x^1, \ldots, x^n) \mapsto f(x^1, \ldots, x^n)$ and differentiating the resulting function of n variables as one would in first-year analysis. The more precise way to say this is that for each $f \in C^\infty(\mathcal{U})$ and $p \in \mathcal{U}$, the function $\partial_j f \in C^\infty(\mathcal{U})$ is given by

$$(\partial_j f)(p) := \partial_j(f \circ x^{-1})(x(p)),$$

---

[26] Such a function can be constructed in local coordinates our of functions of the form $\mathbb{R}^n \to [0, 1] : x \mapsto \beta(|x|^2)$, where $\beta : \mathbb{R} \to [0, 1]$ is a smooth function with $\beta(t) = 0$ for all $t \geqslant \epsilon > 0$ and $\beta(0) = 1$. The construction of β is an easy exercise once you've seen examples like $h(t) := e^{-1/t^2}$, a smooth function on $(0, \infty)$ admitting a smooth extension to $\mathbb{R}$ that vanishes on $(-\infty, 0]$.

where the right-hand side is a perfectly ordinary partial derivative of a real-valued function of n real variables. The fact that the operators $\partial_1, \ldots, \partial_n$ define derivations on $C^\infty(\mathcal{U})$ follows immediately from the usual product rule. The corresponding vector fields in $\mathfrak{X}(\mathcal{U})$ are also easy to identify: they come from the standard basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$ as transferred over to $\mathcal{U}$ by the chart, i.e. the derivation ∂_j corresponds to the vector field

$$v_j(p) := (d_p x)^{-1}(e_j), \qquad p \in \mathcal{U}.$$

Since this notation is bit clumsy, it has become conventional in differential geometry to use the notation

$$\partial_1, \ldots, \partial_n \quad \text{or equivalently} \quad \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \in \mathfrak{X}(\mathcal{U})$$

not just for the derivations but also for the corresponding vector fields on $\mathcal{U}$, and I will follow that convention in these notes, in spite of what I said in Remark 5.10 above. We call these the **coordinate vector fields** determined on $\mathcal{U}$ by the chart $(\mathcal{U}, x)$. Two issues are very important to understand:

(1) The vector fields $\frac{\partial}{\partial x^j}$ are *only* defined on $\mathcal{U} \subset M$; it does not make sense to write down formulas involving ∂_j everywhere on M unless $(\mathcal{U}, x)$ happens to be a *global* chart, meaning $\mathcal{U} = M$.

(2) For each individual $j \in \{1, \ldots, n\}$, the vector field $\frac{\partial}{\partial x^j}$ depends not only on the coordinate function $x^j : \mathcal{U} \to \mathbb{R}$ but on all n of the coordinates $x^1, \ldots, x^n$. Indeed, the vector $\frac{\partial}{\partial x^j}$ points in the unique direction where x^j increases but all the other coordinates are constant. The issue is easy to see in simple examples, e.g. using the standard polar coordinates (r, θ) and Cartesian coordinates (x, y) on suitable regions in $\mathbb{R}^2$, one can define both (r, θ) and (r, y) as smooth charts on the open right half-plane $\{x > 0\} \subset \mathbb{R}^2$. But the partial derivative operator $\frac{\partial}{\partial r}$ has different meanings in these two coordinate systems, because differentiating in a direction where r increases but θ is constant does not typically give the same result as differentiating in a direction where r increases but y is constant.

6. The Lie algebra of vector fields

We saw in the last lecture that there is a natural equivalence between the space of smooth vector fields $\mathfrak{X}(M)$ on a smooth manifold M and the space of all derivations $L : C^\infty(M) \to C^\infty(M)$ on the algebra of smooth functions. It was also observed in Exercise 5.8 that the latter has a natural Lie algebra structure defined via the commutator bracket

$$[L_1, L_2] := L_1 L_2 - L_2 L_1,$$

which is antisymmetric and satisfies the Jacobi identity (see Definition 5.7). Lie algebras are a large topic that we will discuss in more detail next semester; if you have not seen them at all before, then I would not expect you to have any intuition as to why a bilinear bracket satisfying antisymmetry and the Jacobi identity might be an interesting or useful object to study. But we will see a first example of the answer to that question in this lecture: the Lie algebra structure on the space of vector fields characterizes the commutativity (or lack thereof) of their respective flows. This will be easily the deepest result we have proved so far in this course, and it will serve as a foundation for several later results involving curvature and integrability.

6.1. Components and the summation convention. Recall that any smooth chart $(\mathcal{U}, x = (x^1, \ldots, x^n))$ on a manifold M defines an associated set of *coordinate vector fields* $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \in$

$\mathfrak{X}(\mathcal{U})$. These form a basis of T_pM at each point $p \in \mathcal{U}$, so any $X \in \mathfrak{X}(M)$ restricted to $\mathcal{U} \subset M$ can now be written uniquely in the form

$$(6.1) \qquad X = \sum_{i=1}^{n} X^i \partial_i = \sum_{i=1}^{n} X^i \frac{\partial}{\partial x^i}$$

for uniquely defined smooth functions $X^1, \ldots, X^n \in C^\infty(\mathcal{U})$, called the **components** of X with respect to the chart $(\mathcal{U}, x)$. This observation will be useful for computations, but it becomes more so if we can make the notation a bit less cumbersome. Einstein introduced a nice trick for this, which is known as the **Einstein summation convention**: the trick is to omit the summation symbol, but assume that whenever a matching pair of "upper" and "lower" indices appears, a summation of that index over all coordinates (in this case from 1 to n) is implied. Using this convention, (6.1) becomes

$$X = X^i \partial_i = X^i \frac{\partial}{\partial x^i},$$

where the convention is also to interpret the upper index in $\frac{\partial}{\partial x^i}$ as a lower index because it appears in the denominator. (I advise you not to search for any deeper meaning behind this—just take it as a definition for now, and you will see presently why it is useful.) The simplicity of this expression in comparison with (6.1) is perhaps not so dramatic, but the Einstein convention becomes especially useful in situations where multiple indices need to be summed over at the same time, which will happen a lot once we start talking about tensors next week.

Let us derive a coordinate transformation formula: suppose $(\widetilde{\mathcal{U}}, \widetilde{x})$ is a second chart with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$, and the components of X in these alternative coordinates over $\widetilde{\mathcal{U}}$ are denoted by $\widetilde{X}^i$, so $X = \widetilde{X}^i \frac{\partial}{\partial \widetilde{x}^i}$ on $\widetilde{\mathcal{U}}$. How do the components X^i and $\widetilde{X}^i$ relate to each other on the region $\mathcal{U} \cap \widetilde{\mathcal{U}}$ where their domains overlap?

To answer this, we start with the observation that for any $f \in C^\infty(\mathcal{U} \cap \widetilde{\mathcal{U}})$, the chain rule relates the partial derivatives of f with respect to the two different coordinate systems by

$$(6.2) \qquad \frac{\partial f}{\partial x^i} = \frac{\partial f}{\partial \widetilde{x}^j} \frac{\partial \widetilde{x}^j}{\partial x^i},$$

where the Einstein convention gives an implied summation $\sum_{j=1}^{n}$ on the right hand side. This formula is hopefully familiar to you from analysis, at least when applied to functions on open subsets of $\mathbb{R}^n$; in the present setting, the partial derivatives on both sides are interpreted as derivations applied to smooth functions on $\mathcal{U} \cap \widetilde{\mathcal{U}} \subset M$, but these have been defined in terms of ordinary partial derivatives of functions on $\mathbb{R}^n$. In that context, the left hand side is the ith component of the *gradient* ∇f of f in coordinates $(x^1, \ldots, x^n)$, interpreted as a row vector, while the right hand side is the ith component of the product of the row vector $\widetilde{\nabla} f$ (the gradient of f is coordinates $(\widetilde{x}^1, \ldots, \widetilde{x}^n)$) with the Jacobian matrix $\frac{\partial \widetilde{x}}{\partial x}$ of the transition map $(x^1, \ldots, x^n) \mapsto (\widetilde{x}^1(x^1, \ldots, x^n), \ldots, \widetilde{x}^n(x^1, \ldots, x^n))$. Equation (6.2) is thus equivalent to the relation

$$D(f \circ x^{-1})(x(p)) = D(f \circ \widetilde{x}^{-1})(\widetilde{x}(p)) \circ D(\widetilde{x} \circ x^{-1})(x(p)),$$

which follows directly from the chain rule. Now, the function f was not actually important in this discussion at all: what we are really interested in is a formula relating derivations, namely

$$(6.3) \qquad \frac{\partial}{\partial x^i} = \frac{\partial \widetilde{x}^j}{\partial x^i} \frac{\partial}{\partial \widetilde{x}^j},$$

which can now equally well be interpreted as a formula for the coordinate vector field $\frac{\partial}{\partial x^i}$ as a linear combination of the other set of coordinate vector fields $\frac{\partial}{\partial \widetilde{x}^j}$ where they overlap. This implies

$$X = X^i \frac{\partial}{\partial x^i} = X^i \frac{\partial \widetilde{x}^j}{\partial x^i} \frac{\partial}{\partial \widetilde{x}^j} = \widetilde{X}^j \frac{\partial}{\partial \widetilde{x}^j},$$

from which we derive (after interchanging the indices i and j just for good measure) the transformation formula

$$(6.4) \qquad\qquad \widetilde{X}^i = \frac{\partial \widetilde{x}^i}{\partial x^j} X^j.$$

You may agree that if we'd had to write summation symbols in all of these expressions, we would be slightly more tired now. Notice that this formula has an easy interpretation in terms of matrix-vector multiplication: if we package the components together into $\mathbb{R}^n$-valued functions $\xi := (X^1, \ldots, X^n) : \mathcal{U} \to \mathbb{R}^n$ and $\widetilde{\xi} := (\widetilde{X}^1, \ldots, \widetilde{X}^n) : \widetilde{\mathcal{U}} \to \mathbb{R}^n$, then (6.4) relates these two functions to each other via multiplication with the Jacobian matrix $\frac{\partial \widetilde{x}}{\partial x}$:

$$\widetilde{\xi} = \frac{\partial \widetilde{x}}{\partial x} \xi.$$

The Einstein convention has nothing intrinsically to do with differential geometry—it is actually just linear algebra. Once you get used to it, you may begin to wish you had always been doing linear algebra this way.

We will use the Einstein convention consistently throughout the rest of this course, and only include explicitly written summation symbols in situations where their omission might cause confusion.

REMARK 6.1. Using the summation convention requires being very careful and consistent about the distinction between upper and lower indices: coordinates and components of vector fields are *always* written with upper indices, while partial derivative operators (and their associated coordinate vector fields) always carry lower indices. Forgetting these conventions can cause grave confusion and should be avoided at all costs. Unfortunately, not all differential geometry books written by mathematicians are completely consistent about this, though books by physicists are—Einstein was one of them, after all, so his mathematical innovations are taken as gospel.

6.2. The Lie bracket. The **Lie bracket** (*Lie-Klammer*) of two vector fields $X, Y \in \mathfrak{X}(M)$ on a manifold M is defined to be the unique vector field

$$[X, Y] \in \mathfrak{X}(M) \qquad \text{such that} \qquad \mathcal{L}_{[X,Y]} = \mathcal{L}_X \mathcal{L}_Y - \mathcal{L}_Y \mathcal{L}_X.$$

This definition makes sense as a consequence of Exercise 5.8 and Theorem 5.9. In particular, we say that X and Y **commute** if $[X, Y] \equiv 0$.

EXERCISE 6.2. Suppose $(\mathcal{U}, x)$ is a chart on M and we express two vector fields $X, Y \in \mathfrak{X}(M)$ over $\mathcal{U}$ in this chart as $X = X^i \partial_i$ and $Y = Y^i \partial_i$.

(a) Show that the components $[X, Y]^i$ of $[X, Y]$ with respect to the same chart are given by

$$(6.5) \qquad\qquad [X, Y]^i = X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}.$$

(b) Use the coordinate transformation formulas (6.3) and (6.4) to give a direct computational proof (without using the result of part (a)) that the vector field defined on $\mathcal{U}$ via the right hand side of (6.5) depends only on $X, Y \in \mathfrak{X}(\mathcal{U})$ and not on the choice of chart $(\mathcal{U}, x)$. In other words, show that for any other chart $(\widetilde{\mathcal{U}}, \widetilde{x})$,

$$\left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \right) \frac{\partial}{\partial x^i} = \left(\widetilde{X}^j \frac{\partial \widetilde{Y}^i}{\partial \widetilde{x}^j} - \widetilde{Y}^j \frac{\partial \widetilde{X}^i}{\partial \widetilde{x}^j} \right) \frac{\partial}{\partial \widetilde{x}^i} \quad \text{on} \quad \mathcal{U} \cap \widetilde{\mathcal{U}}.$$

Hint: The matrices with entries $\frac{\partial \tilde{x}^i}{\partial x^j}$ and $\frac{\partial x^i}{\partial \tilde{x}^j}$ are Jacobi matrices for transformations that are inverse to each other, thus they satisfy

$$\frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial x^j}{\partial \tilde{x}^k} = \delta^i_k := \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

REMARK 6.3. Physicists like being able to do explicit computations, so they tend to emphasize coordinate-based formulas in this subject much more than mathematicians do. For example, some physics books take the formula (6.5) as a *definition* of the Lie bracket $[X, Y]$, without first talking about commutators of derivations. The price for doing this is that one must prove that switching to a different local coordinate system would not change the definition, i.e. one must do Exercise 6.2(b). The exercise is tedious, but I recommend doing it exactly once in your life, as it may give you some useful insight into the way that physicists do mathematics, and in any case, it is never bad to get better at explicit computations. As a cautionary tale, I also recommend convincing yourself that the simpler formula

$$X^j \frac{\partial Y^i}{\partial x^j} \frac{\partial}{\partial x^i} = \tilde{X}^j \frac{\partial \tilde{Y}^i}{\partial \tilde{x}^j} \frac{\partial}{\partial \tilde{x}^i} \quad \text{on} \quad \mathcal{U} \cap \tilde{\mathcal{U}}$$

is *false* in general, thus one cannot define a vector field $Z = Z^i \partial_i$ by $Z^i := X^j \partial_j Y^i$ and expect the definition to be independent of the choice of coordinates.

EXERCISE 6.4. For $X, Y \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$, give two proofs of the formulas

$$[fX, Y] = f[X, Y] - (\mathcal{L}_Y f)X, \qquad [X, fY] = f[X, Y] + (\mathcal{L}_X f)Y,$$

using different methods:
  (a) Directly from the definition of the Lie bracket via Theorem 5.9;
  (b) Using the coordinate formula (6.5).

EXERCISE 6.5. For a diffeomorphism $\psi : M \to N$ and two vector fields $X, Y \in \mathfrak{X}(M)$, prove $\psi_*[X, Y] = [\psi_* X, \psi_* Y] \in \mathfrak{X}(N)$.

EXAMPLE 6.6. The coordinate vector fields $\partial_1, \ldots, \partial_n$ defined from any chart on an open subset all commute with each other. One can deduce this either from the fact that $\partial_i \partial_j f = \partial_j \partial_i f$ for all smooth functions f,[27] or as a trivial application of the formula in Exercise 6.2.

My goal for the rest of this lecture is to explain not just what the Lie bracket of two vector fields *is*, but what it *means*. The discussion starts with the following observation related to Example 6.6 above. Consider the manifold $M = \mathbb{R}^n$ with the standard Cartesian coordinates $x^1, \ldots, x^n$ regarded as a global chart on M; this chart is actually just the identity map $\mathbb{R}^n \to \mathbb{R}^n$. The resulting coordinate vector fields $\partial_1, \ldots, \partial_n$ produce the standard basis of the tangent space $T_p \mathbb{R}^n = \mathbb{R}^n$ at every point $p \in \mathbb{R}^n$. It is easy to write down the flow of ∂_j for each $j = 1, \ldots, n$: it is

$$\varphi^t_{\partial_j}(x^1, \ldots, x^n) = (x^1, \ldots, x^{j-1}, x^j + t, x^{j+1}, \ldots, x^n).$$

We see from this that for any two $i, j \in \{1, \ldots, n\}$ and $s, t \in \mathbb{R}$, the corresponding flows commute:

$$\varphi^s_{\partial_i} \circ \varphi^t_{\partial_j} = \varphi^t_{\partial_j} \circ \varphi^s_{\partial_i}.$$

This is a generalization of the basic observation that if you start from some point (x, y) in the plane $\mathbb{R}^2$, move a distance s to the right and then a distance t upward, you'll end up at the same point as if you had made those two moves in the reverse order, namely $(x + s, y + t)$. In other words,

---

[27] And since this is not an analysis course, there is no need to worry about the fact that $\partial_i \partial_j f = \partial_j \partial_i f$ does not generally hold for functions whose second-order derivatives exist but are discontinuous. With very few exceptions, all functions that we choose to worry about in the remainder of this course will be of class C^∞.

the two paths, each consisting of two straight line segments, combine to form a closed rectangle. This observation is not as trivial as it may seem: in particular, it becomes false in general if you replace ∂_i and ∂_j by different vector fields, e.g. in the example of $\mathbb{R}^2$, one could replace the "horizontal" coordinate vector field ∂_1 with one that still points in the x-direction but flows at different speeds along the lower and upper segments of the rectangle, in which case the rectangle fails to close up. There is no reason in general why the flows of two vector fields should always commute. They do commute in the case of coordinate vector fields on $\mathbb{R}^n$, and it follows easily that flows of coordinate vector fields determined by a chart $(\mathcal{U}, x)$ on a manifold M will generally commute as long as one keeps s and t close enough to 0 so that the flow lines do not escape from $\mathcal{U}$. But pairs of coordinate vector fields are special, and one symptom of this is the fact that their Lie brackets vanish. We will show in §6.4 that this is a general phenomenon: in particular, for any two vector fields $X, Y \in \mathfrak{X}(M)$ whose flows exist globally, one has $\varphi_X^s \circ \varphi_Y^t = \varphi_Y^t \circ \varphi_X^s$ for all $s, t \in \mathbb{R}$ if and only if $[X, Y] \equiv 0$.

6.3. The Lie derivative of a vector field. Before we can prove a result on commuting flows, we need a short digression to address the following question: What might it mean to differentiate a vector field $Y \in \mathfrak{X}(M)$ at a point $p \in M$ in the direction $X \in T_pM$? A naive attempt to define this would proceed as follows: choose any smooth path $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\dot\gamma(0) = X$, and set

$$\mathcal{L}_X Y(p) \overset{?}{:=} \frac{d}{dt} Y(\gamma(t))\Big|_{t=0} = \lim_{t \to 0} \frac{Y(\gamma(t)) - Y(p)}{t} \ ?$$

If Y were a real-valued function instead of a vector field, then we would be on solid ground with this definition, but for a vector field the right hand side does not make sense: outside of the uninteresting special case where γ is a constant path, $Y(\gamma(t)) \in T_{\gamma(t)}M$ and $Y(p) \in T_pM$ generally belong to different vector spaces, so there is no well-defined way of subtracting one from the other.

A solution to this conundrum arises if one allows X to be a vector *field* on M, rather than just a single tangent vector. In this case, the flow of X gives a natural choice of the path

$$\gamma(t) = \varphi_X^t(p),$$

which is defined for t in a sufficiently small interval $(-\epsilon, \epsilon)$ even if the flow does not globally exist. More importantly, the tangent map of the flow gives rise to natural isomorphisms,

$$T_p\varphi_X^t : T_pM \to T_{\varphi_X^t(p)}M = T_{\gamma(t)}M$$

for t close to 0, which gives us a way of identifying with each other the distinct tangent spaces in which $Y(p)$ and $Y(\gamma(t))$ live. Since the inverse of $T\varphi_X^t$ is $T\varphi_X^{-t}$, it now makes sense to define the **Lie derivative** (*Lie-Ableitung*) of $Y \in \mathfrak{X}(M)$ with respect to $X \in \mathfrak{X}(M)$ as the vector field

$$\mathcal{L}_X Y \in \mathfrak{X}(M), \qquad \mathcal{L}_X Y(p) := \frac{d}{dt} T\varphi_X^{-t}\big(Y(\varphi_X^t(p))\big)\Big|_{t=0} = \lim_{t \to 0} \frac{T\varphi_X^{-t}\big(Y(\varphi_X^t(p))\big) - Y(p)}{t}.$$

Recalling the definition of the *pullback* of a vector field in §5.2, we can abbreviate this formula as

$$\mathcal{L}_X Y = \frac{d}{dt}(\varphi_X^t)^* Y\Big|_{t=0} \ .$$

It turns out that $\mathcal{L}_X Y$ is just a new perspective on the Lie bracket:

PROPOSITION 6.7. *For any $X, Y \in \mathfrak{X}(M)$, $\mathcal{L}_X Y = [X, Y]$.*

PROOF. We need to show that for every $f \in C^\infty(M)$,

(6.6) $$\mathcal{L}_{\mathcal{L}_X Y} f = \mathcal{L}_X \mathcal{L}_Y f - \mathcal{L}_Y \mathcal{L}_X f.$$

In the following, when writing expressions such as $\varphi_X^t(p)$, we always assume that t is close enough to 0 for this flow to be defined. With this understood, we claim that

$$f \circ \varphi_X^t = f + tg_t$$

for some smooth family of smooth real-valued functions g_t on M with $g_0 = \mathcal{L}_X f \in C^\infty(M)$.[28] This follows from the fundamental theorem of calculus: for $p \in M$ and $t \in \mathbb{R}$ close to 0, we write

$$f(\varphi_X^t(p)) - f(p) = \int_0^1 \frac{d}{ds} f(\varphi_X^{st}(p)) \, ds = \int_0^1 df\big(\partial_s \varphi_X^{st}(p)\big) \, ds$$
$$= \int_0^1 df\big(tX(\varphi_X^{st}(p))\big) \, ds = t \int_0^1 df\big(X(\varphi_X^{st}(p))\big) \, ds,$$

define $g_t(p)$ to be the integral on the right, and compute

$$g_0(p) = \int_0^1 df\big(X(\varphi_X^0(p))\big) \, ds = \int_0^1 df\big(X(p)\big) \, ds = df(X(p)) = \mathcal{L}_X f(p),$$

proving the claim. Using this formula, we find

$$df\big([(\varphi_X^t)^* Y](p)\big) = df\big(T\varphi_X^{-t}(Y(\varphi_X^t(p)))\big) = d(f \circ \varphi_X^{-t})(Y(\varphi_X^t(p)))$$
$$= d(f - tg_t)\big(Y(\varphi_X^t(p))\big) = df\big(Y(\varphi_X^t(p))\big) - t \, dg_t\big(Y(\varphi_X^t(p))\big)$$
$$= \mathcal{L}_Y f(\varphi_X^t(p)) - t \, \mathcal{L}_Y g_t(\varphi_X^t(p)).$$

If we now differentiate this relation with respect to t and set $t = 0$, the left hand side becomes $df(\mathcal{L}_X Y(p)) = \mathcal{L}_{\mathcal{L}_X Y} f(p)$, while the right hand side becomes

$$d(\mathcal{L}_Y f)(X(p)) - \mathcal{L}_Y g_0(p) = \mathcal{L}_X \mathcal{L}_Y f(p) - \mathcal{L}_Y \mathcal{L}_X f(p),$$

proving (6.6). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

REMARK 6.8. The formula $\mathcal{L}_X Y = [X, Y]$ reveals that the Lie derivative of a vector field does not quite admit the interpretation we were hoping for: if $\mathcal{L}_X Y(p)$ were merely the directional derivative of $Y \in \mathfrak{X}(M)$ at p in the direction of $X \in T_p M$, then it should only depend on Y and the specific value $X(p)$, but as we see in (6.5), $[X, Y](p)$ also depends on the *first derivatives* of X at p in coordinates, not just on its value. We will see later that a straightforward directional derivative of anything more complicated than a real-valued function cannot typically be defined without making additional choices, e.g. the definition of $\mathcal{L}_X Y(p)$ requires extending $X(p)$ to a vector field that takes that value at p, and the resulting derivative depends on that choice. We will see a different and in some sense simpler way to define directional derivatives of vector fields when we study *connections* later in the semester, but a connection is also a choice that is not canonically defined in general.

6.4. Commuting flows. We can now discuss the relationship between the Lie bracket $[X, Y]$ and the question of whether the flows of X and Y commute. To understand the statement, recall from §5.1 that for each $X \in \mathfrak{X}(M)$ and $s \in \mathbb{R}$, the flow defines a diffeomorphism

$$\varphi_X^s : \mathcal{O}_X^s \to \mathcal{O}_X^{-s}$$

between two open subsets $\mathcal{O}_X^s, \mathcal{O}_X^{-s} \subset M$, which may in general be empty, but are guaranteed to be nonempty if s is close enough to 0; in fact, we have $\mathcal{O}_X^0 = \bigcup_{s>0} \mathcal{O}_X^s = \bigcup_{s<0} \mathcal{O}_X^s = M$.

---

[28]Saying that g_t is a "smooth family" of functions on M means literally that the function $(t, p) \mapsto g_t(p)$ for (t, p) in some open subset of $\mathbb{R} \times M$ is smooth. A slightly subtle point here is that we do not need the function $g_t : M \to M$ to be well-defined *everywhere* on M for some $t \neq 0$; for our purposes, it will suffice if $g_t(p)$ is defined for all (t, p) in some *neighborhood* of the set $\{0\} \times M$. If M is not compact, it may happen that the domain of $(t, p) \mapsto g_t(p)$ does not contain any set of the form $\{t\} \times M$ for $t \neq 0$, but is still an open neighborhood of $\{0\} \times M$.

For another vector field $Y \in \mathfrak{X}(M)$ and another $t \in \mathbb{R}$, the composition $\varphi_Y^t \circ \varphi_X^s$ is defined on $(\varphi_X^s)^{-1}(\mathcal{O}_Y^t) \subset M$, which is also open and could be empty, but is definitely not empty if both $|s|$ and $|t|$ are sufficiently small. The domain of $\varphi_X^s \circ \varphi_Y^t$ may be a different open subset of M, but is also guaranteed to overlap the domain of $\varphi_Y^t \circ \varphi_X^s$ if $|s|$ and $|t|$ are sufficiently small; in fact for every $p \in M$, there exists ϵ such that both $\varphi_X^s \circ \varphi_Y^t(p)$ and $\varphi_Y^t \circ \varphi_X^s(p)$ are defined whenever $|s|, |t| < \epsilon$.

THEOREM 6.9. *For two smooth vector fields $X, Y \in \mathfrak{X}(M)$ on a manifold M, the following conditions are equivalent:*

(i) *$[X, Y] \equiv 0$;*

(ii) *Suppose $p \in M$ and $s, t \in \mathbb{R}$ are such that $\varphi_X^\sigma \circ \varphi_Y^\tau(p)$ is defined for all σ between 0 and s and all τ between 0 and t. Then $\varphi_Y^\tau \circ \varphi_X^\sigma(p)$ is also defined for all such σ and τ, and it equals $\varphi_X^\sigma \circ \varphi_Y^\tau(p)$. In particular, if X and Y both have global flows, then they define commuting diffeomorphisms*

$$\varphi_X^s \circ \varphi_Y^t = \varphi_Y^t \circ \varphi_X^s \in \text{Diff}(M)$$

for all $s, t \in \mathbb{R}$.

PROOF. We prove first that (ii) $\Rightarrow$ (i), so suppose X and Y are two vector fields whose flows commute in the sense described in the statement. For each $p \in M$, one can find a neighborhood $\mathcal{U} \subset \mathbb{R}^2$ of $(0, 0)$ small enough so that the smooth map

$$\alpha : \mathcal{U} \to M : (s, t) \mapsto \varphi_X^s \circ \varphi_Y^t(p) = \varphi_Y^t \circ \varphi_X^s(p)$$

is well-defined via either of the compositions on the right hand side. This map satisfies $\partial_s \alpha(s, t) = X(\alpha(s, t))$ and $\partial_t \alpha(s, t) = Y(\alpha(s, t))$, where the proof of the first identity requires the first version of the composition, and the second requires the second. Given $f \in C^\infty(M)$, we now define $g := f \circ \alpha : \mathcal{U} \to \mathbb{R}$ and observe that

$$\mathcal{L}_X f(\alpha(s, t)) = \partial_s g(s, t) \qquad \text{and} \qquad \mathcal{L}_Y f(\alpha(s, t)) = \partial_t g(s, t),$$

and similarly,

$$\mathcal{L}_X \mathcal{L}_Y f(\alpha(s, t)) = \partial_s \partial_t g(s, t) = \partial_t \partial_s g(s, t) = \mathcal{L}_Y \mathcal{L}_X f(\alpha(s, t)).$$

This proves in particular that $(\mathcal{L}_X \mathcal{L}_Y - \mathcal{L}_Y \mathcal{L}_X) f(p) = 0$, hence $[X, Y](p) = 0$ for all $p \in M$.

To prove (i) $\Rightarrow$ (ii), assume $[X, Y] \equiv 0$, and fix $p \in M$ and $s, t \in \mathbb{R}$ satisfying the condition specified in (ii). Then for each σ in the interval between 0 and s, φ_X^σ defines a diffeomorphism

$$M \overset{\text{open}}{\supset} \mathcal{O}_X^\sigma \xrightarrow{\varphi_X^\sigma} \mathcal{O}_X^{-\sigma} \overset{\text{open}}{\subset} M$$

whose domain and target satisfy $\mathcal{O}_X^\sigma \supset \mathcal{O}_X^s$ and $\mathcal{O}_X^{-\sigma} \supset \mathcal{O}_X^{-s}$ respectively, and moreover, the flow line $\gamma(\tau) := \varphi_Y^\tau(p)$ exists and has image in $\mathcal{O}_X^s$ for τ in the interval between 0 and t. The main step in the proof will be to show that for every σ between 0 and s, the pullback of the vector field Y from $\mathcal{O}_X^{-\sigma}$ to $\mathcal{O}_X^\sigma$ via φ_X^σ matches Y itself on $\mathcal{O}_X^s$, i.e.

(6.7) $$Y = (\varphi_X^\sigma)^* Y \quad \text{on} \quad \mathcal{O}_X^s.$$

Assuming this for the moment, it then follows from Proposition 5.4 and (6.7) that the path $\tau \mapsto \varphi_X^\sigma \circ \gamma(\tau)$ for τ between 0 and t is also a flow line of Y, namely the unique one beginning at $\varphi_X^\sigma(p)$, which proves

$$\varphi_Y^\tau(\varphi_X^\sigma(p)) = \varphi_X^\sigma(\gamma(\tau)) = \varphi_X^\sigma(\varphi_Y^\tau(p)).$$

It remains only to prove (6.7). Since the statement is clearly true for $\sigma = 0$, it will suffice to prove that the derivative of the family of vector fields $(\varphi_X^\sigma)^* Y$ with respect to the parameter σ vanishes at every point on $\mathcal{O}_X^s$ for all σ between 0 and s. To see this, we use the identities $[X, Y] = \mathcal{L}_X Y = 0$ and $\varphi_X^{\sigma+\tau} = \varphi_X^\tau \circ \varphi_X^\sigma$, which gives $(\varphi_X^{\sigma+\tau})^* = (\varphi_X^\sigma)^*(\varphi_X^\tau)^*$ by Exercise 5.5.

In the following, we will only need the latter relation for values of $\tau \in \mathbb{R}$ that are arbitrarily close to 0, thus we will be free to assume that any given point in the domain of φ_X^σ is also in the domain of $\varphi_X^{\sigma+\tau}$. Working over the open set $\mathcal{O}_X^s$, we now compute,

$$\frac{d}{d\sigma}(\varphi_X^\sigma)^*Y = \frac{d}{d\tau}(\varphi_X^{\sigma+\tau})^*Y\bigg|_{\tau=0} = \frac{d}{d\tau}(\varphi_X^\sigma)^*(\varphi_X^\tau)^*Y\bigg|_{\tau=0} = (\varphi_X^\sigma)^*\left(\frac{d}{d\tau}(\varphi_X^\tau)^*Y\bigg|_{\tau=0}\right)$$
$$= (\varphi_X^\sigma)^*(\mathcal{L}_X Y) = 0.$$

$\square$

7. Tensors

It will turn out that many types of "geometric structure" on manifolds can be expressed in terms of multilinear maps on tangent and cotangent spaces, known collectively as *tensor fields*. Before beginning with the contents of this lecture, I should remind you that the Einstein summation convention (see §5.4) is in effect from now on—we are going to be needing it a lot. We will also need the following convenient notational device: for any pair of indices $i, j \in \{1, \ldots, n\}$, we define

$$\delta^{ij} = \delta_{ij} = \delta_j^i := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The choice of whether each index is an upper or lower index will depend on the context, but the meaning will always be the same. So for example, if $\mathbf{A} \in \mathrm{GL}(n, \mathbb{R})$ is a matrix with entries $A^i{}_j$, the matrix-multiplication relation $\mathbf{A}\mathbf{A}^{-1} = \mathbb{1}$ becomes

$$A^i{}_j (A^{-1})^j{}_k = \delta_k^i.$$

Here it is very important to remember that by the summation convention, the symbol "$\sum_{j=1}^n$" has been omitted from the left hand side; we chose to write the first index of $A^i{}_j$ as an upper index and the second as a lower index mainly so that this use of the summation convention would work. Here is another example that already came up in our discussion of vector fields (cf. Exercise 6.2): if $(\mathcal{U}, x)$ and $(\widetilde{\mathcal{U}}, \widetilde{x})$ are two overlapping charts on a manifold M, then at every point in $\mathcal{U} \cap \widetilde{\mathcal{U}}$, the matrices with entries $\frac{\partial \widetilde{x}^i}{\partial x^j}$ and $\frac{\partial x^i}{\partial \widetilde{x}^j}$ are inverse to each other, as they are Jacobi matrices of inverse transition maps, thus

$$\frac{\partial \widetilde{x}^i}{\partial x^j}\frac{\partial x^j}{\partial \widetilde{x}^k} = \delta_k^i.$$

Other versions of δ will sometimes arise with the indices placed in various ways in order to make the summation convention work. This symbol is known as the **Kronecker delta**, and maybe it would have been called something different if it had been invented in the age of Covid-19, but here we are.

7.1. Motivational examples. In order to motivate the idea of a tensor field on a manifold, it's best to start with a few examples that are already somewhat familiar.

7.1.1. *One-forms.* Any smooth function $f : M \to \mathbb{R}$ has a differential

$$df : TM \to \mathbb{R},$$

whose restriction to each individual tangent space T_pM is a linear map $T_pM \to \mathbb{R}$ and thus an element of the cotangent space T_p^*M. In this sense, df is analogous to a vector field, but instead of associating a tangent vector $X(p) \in T_pM$ to every point $p \in M$, it associates a cotangent vector $d_pf \in T_p^*M$, thus defining a map

$$M \to T^*M : p \mapsto d_pf.$$

In general, a map

$$\lambda : TM \to \mathbb{R}$$

whose restriction to each individual tangent space is linear is called a 1-**form** on M, or sometimes also a **dual vector field** or **covector field**. For each $p \in M$, it is common to denote the restriction $\lambda|_{T_pM} : T_pM \to \mathbb{R}$ by

$$\lambda_p \in T_p^*M = \operatorname{Hom}(T_pM, \mathbb{R}),$$

hence one can equivalently view a 1-form λ as associating to each point $p \in M$ a cotangent vector $\lambda_p \in T_p^*M$. For the special case where λ is the differential of a function f, we have been writing $d_pf \in T_p^*M$ for the restriction to T_pM, but the notation $(df)_p$ would also be sensible, and is preferred by many authors.[29]

Since we have not yet endowed the cotangent bundle T^*M with a smooth structure, we need to put some thought into defining what it means for a 1-form to be "smooth". The easiest way to do this is by writing it in local coordinates. Any chart $(\mathcal{U}, x)$ on M gives rise to coordinate functions $x^i : \mathcal{U} \to \mathbb{R}$ for $i = 1, \ldots, n$, whose differentials dx^i are 1-forms on $\mathcal{U}$.

PROPOSITION 7.1. *For each $p \in \mathcal{U}$, every element $\lambda \in T_p^*M$ can be expressed as a linear combination $\lambda = \lambda_i \, d_px^i$ for unique real numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. In other words, the differentials $d_px^1, \ldots, d_px^n$ form a basis of T_p^*M.*

PROOF. What's actually happening here is that $d_px^1, \ldots, d_px^n$ is the dual basis to the basis of coordinate vector fields $\partial_1, \ldots, \partial_n$ defined by the chart $(\mathcal{U}, x)$ at p; indeed, for each $i, j \in \{1, \ldots, n\}$,

$$dx^i(\partial_j) = dx^i\left(\frac{\partial}{\partial x^j}\right) = \mathcal{L}_{\frac{\partial}{\partial x^j}} x^i = \frac{\partial x^i}{\partial x^j} = \delta_j^i.$$

The coefficients λ_i are thus given by $\lambda_i = \lambda(\partial_i)$. $\qquad\qquad\square$

The 1-forms $dx^1, \ldots, dx^n$ on $\mathcal{U}$ defined by a chart $(\mathcal{U}, x)$ are known as the **coordinate differentials**, and Proposition 7.1 implies that every 1-form λ can be written over the region $\mathcal{U}$ as

$$\lambda = \lambda_i \, dx^i,$$

where its uniquely determined **component** functions $\lambda_i : \mathcal{U} \to \mathbb{R}$ are given by

$$\lambda_i(p) := \lambda\left(\frac{\partial}{\partial x^i}(p)\right), \qquad p \in \mathcal{U}.$$

For example, the component functions of the differential df are precisely the partial derivatives of f, namely $(df)_i = df(\partial_i) = \partial_i f : \mathcal{U} \to \mathbb{R}$, giving rise to the formula

$$df = \partial_i f \, dx^i \qquad \text{on } \mathcal{U},$$

which was understood for at least two centuries in terms of "infinitessimal quantities" before it was given a mathematically rigorous meaning in terms of 1-forms.

REMARK 7.2. Notice that while components of vector fields are written with upper indices, components of 1-forms get lower indices. This is necessary in order for the summation convention to work properly, since coordinate differentials come with upper indices.

---

[29]Or if one prefers to think of df as a function $M \to T^*M$, one can write $df(p)$ instead of d_pf or $(df)_p$. I have done that in some of my research papers, but will avoid it in these notes for the sake of consistency, as we have defined df as a function $TM \to \mathbb{R}$ rather than $M \to T^*M$.

EXERCISE 7.3. Suppose $(\mathcal{U}, x)$ and $(\widetilde{\mathcal{U}}, \widetilde{x})$ are two smooth charts with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$, so any 1-form λ can be written as both $\lambda_i \, dx^i$ and $\widetilde{\lambda}_i \, d\widetilde{x}^i$ in the overlap region. Prove the following coordinate transformation formulas on $\mathcal{U} \cap \widetilde{\mathcal{U}}$, analogous to the formulas (6.3) and (6.4) for vector fields:

$$(7.1) \qquad\qquad dx^i = \frac{\partial x^i}{\partial \widetilde{x}^j} \, d\widetilde{x}^j \qquad \text{and} \qquad \widetilde{\lambda}_i = \lambda_j \frac{\partial x^j}{\partial \widetilde{x}^i}.$$

The formula (7.1) shows that if a 1-form has smooth component functions with respect to any given chart, its component functions in any other chart defined on the same domain will also be smooth, due to the fact that transition maps (and therefore also their derivatives $\frac{\partial x^i}{\partial \widetilde{x}^j}$) are smooth. The following definition therefore makes sense.

DEFINITION 7.4. A 1-form on M is said to be **smooth** if and only if its component functions with respect to every chart are smooth. The set of all smooth 1-forms on M forms a vector space, which we denote by

$$\Omega^1(M) := \{\text{smooth 1-forms on } M\}.$$

EXERCISE 7.5. Show that a 1-form λ on M is smooth if and only if the function $M \to \mathbb{R} : p \mapsto \lambda(X(p))$ is smooth for every smooth vector field $X \in \mathfrak{X}(M)$.

From now on, we will assume that all 1-forms we consider are smooth unless stated otherwise.

7.1.2. *Vector fields.* Recall that every finite-dimensional vector space V is naturally isomorphic to the dual of its dual space, with a canonical isomorphism $\Phi : V \to V^{**}$ given by

$$\Phi(v)\lambda := \lambda(v).$$

If we choose to, we can therefore also think of every tangent space T_pM as a dual space, namely $(T_p^*M)^*$, meaning that every vector field $X \in \mathfrak{X}(M)$ can equivalently be viewed as associating to each $p \in M$ a linear map $\tau_p : T_p^*M \to \mathbb{R}$, defined by $\tau_p(\lambda) := \lambda(X(p))$. I'm sure you can imagine why we didn't define vector fields this way in the first place, but we could have done so if we'd wanted to. From this perspective, the notion of smoothness for a vector field can also be characterized analogously to Exercise 7.5:

EXERCISE 7.6. Show that a vector field X on M is smooth if and only if the function $M \to \mathbb{R} : p \mapsto \lambda(X(p))$ is smooth for every smooth 1-form $\lambda \in \Omega^1(M)$.

7.1.3. *Riemannian metrics.* A Riemannian metric g on a manifold M associates to every point $p \in M$ an inner product g_p on T_pM, so in particular, g_p is a bilinear map

$$g_p : T_pM \times T_pM \to \mathbb{R}$$

that is also symmetric and positive-definite. We can think of g itself as a function

$$g : TM \oplus TM \to \mathbb{R},$$

where $TM \oplus TM := \bigcup_{p \in M} (T_pM \times T_pM)$. As a provisional notion of smoothness for Riemannian metrics, we can define g to be **smooth** if and only if the function

$$M \to \mathbb{R} : p \mapsto g(X(p), Y(p))$$

is smooth for every pair of smooth vector fields $X, Y \in \mathfrak{X}(M)$. Under this condition, g is an example of something we will shortly define as a "smooth covariant tensor field of rank 2" on M.

7.1.4. *Almost complex structures.* Here is an example you may not have heard of before. One can make any $2n$-dimensional real vector space V into an n-dimensional complex vector space by choosing a linear map $J : V \to V$ with $J^2 = -\mathbb{1}$ and defining complex scalar multiplication on V by $(a + ib)v := av + bJv$. Such a linear map J is therefore called a **complex structure** on V. It is sometimes useful to introduce such a structure on the tangent spaces of an even-dimensional manifold M. An **almost complex structure** (*fast komplexe Struktur*) on M is a map

$$J : TM \to TM$$

whose restriction to each individual tangent space is a complex structure $J_p : T_pM \to T_pM$. We can define J to be **smooth** if and only if the vector field $p \mapsto JX(p)$ is smooth for all smooth vector fields $X \in \mathfrak{X}(M)$. The following lemma gives an alternative algebraic way of understanding what an almost complex structure is.

LEMMA 7.7. *For a finite-dimensional real vector space V, let $\mathrm{End}(V) = \mathrm{Hom}(V, V)$ denote the vector space of all linear maps $V \to V$, $V^* = \mathrm{Hom}(V, \mathbb{R})$ the dual space of V, and $\mathrm{Hom}(V^* \otimes V, \mathbb{R})$ the vector space of all bilinear maps $V^* \times V \to \mathbb{R}$. There exists a canonical isomorphism*

$$\Phi : \mathrm{End}(V) \to \mathrm{Hom}(V^* \otimes V, \mathbb{R}), \qquad \Phi(A)(\lambda, v) := \lambda(Av).$$

PROOF. It is easy to check that Φ is a linear injection, and if $\dim V = n$, then $\dim \mathrm{End}(V) = \dim \mathrm{Hom}(V^* \otimes V, \mathbb{R}) = n^2$, thus Φ is also surjective. □

For an almost complex structure J on M, Lemma 7.7 allows us to view $J_p : T_pM \to T_pM$ equivalently as a bilinear map $T_p^*M \times T_pM \to \mathbb{R}$, and from this perspective, one can check that J is smooth (according to our previous definition) if and only if the function $M \to \mathbb{R} : p \mapsto J(\lambda_p, X(p))$ is smooth for all choices of smooth vector field $X \in \mathfrak{X}(M)$ and smooth 1-form $\lambda \in \Omega^1(M)$.

7.2. Tensor fields in general. We now describe a more general notion that encompasses all of the examples in §7.1 as special cases.

Recall that for vector spaces $V_1, \dots, V_n$ and W, a map

$$T : V_1 \times \dots \times V_n \to W$$

is called **multilinear** if it is linear with respect to each variable individually, i.e. for every $i = 1, \dots, n$ and every fixed tuple of vectors $v_j \in V_j$ for $j = 1, \dots, i-1, i+1, \dots, n$, the map

$$V_i \to W : v_i \mapsto T(v_1, \dots, v_n)$$

is linear. Observe that the space of all multilinear maps $V_1 \times \dots \times V_n \to W$ is naturally also a finite-dimensional vector space. We will sometimes denote it by[30]

$$\mathrm{Hom}(V_1 \otimes \dots \otimes V_n, W).$$

DEFINITION 7.8. For integers $k, \ell \geq 0$ with $k + \ell > 0$ and a finite-dimensional real vector space V, we will denote by V_ℓ^k the vector space of multilinear maps

$$\underbrace{V^* \times \dots \times V^*}_{k} \times \underbrace{V \times \dots \times V}_{\ell} \to \mathbb{R},$$

where V^* as usual denotes the dual space $\mathrm{Hom}(V, \mathbb{R})$. In the case $k = \ell = 0$, we define $V_0^0 = \mathbb{R}$.

---

[30]We will not make use of the abstract algebraic notion of the tensor product of vector spaces in this lecture, but readers already familiar with that notion may want to pause and consider why our definition of the symbol "$\mathrm{Hom}(V_1 \otimes \dots \otimes V_n, W)$" is equivalent to the one they've seen before. It is important that we are explicitly assuming all vector spaces to be finite dimensional in this discussion; if we did not assume this, then some more serious digressions into the meaning of the symbol "$\otimes$" would be necessary.

REMARK 7.9. To motivate the convention $V_0^0 = \mathbb{R}$, you can imagine perhaps that a "real-valued multilinear function of *zero* variables" is the same thing as a real number. If that doesn't convince you, the convention will at least begin to seem more natural when we discuss tensor products (cf. Remark 7.19).

DEFINITION 7.10. For a smooth manifold M and integers $k, \ell \geqslant 0$, a **tensor field** (*Tensorfeld*) S of type (k, ℓ) associates to each point $p \in M$ an element

$$S_p \in (T_p M)_\ell^k.$$

If $k + \ell > 0$, then the tensor field S is said to be **smooth** if and only if the function $M \to \mathbb{R}$: $p \mapsto S_p(\lambda_p^1, \ldots, \lambda_p^k, X_1(p), \ldots, X_\ell(p))$ is smooth for every tuple of smooth vector fields $X_1, \ldots, X_\ell \in \mathfrak{X}(M)$ and smooth 1-forms $\lambda^1, \ldots, \lambda^k \in \Omega^1(M)$. We will denote the vector space of smooth tensor fields by

$$\Gamma(T_\ell^k M) := \{\text{smooth tensor fields of type } (k, \ell)\}.$$

For $k = \ell = 0$, a tensor field is just a real-valued function on M, so we define $\Gamma(T_0^0 M) := C^\infty(M)$.

The **support** (*Träger*) of a tensor field $S \in \Gamma(T_\ell^k M)$ is defined as the closure in M of the set $\{p \in M \mid S_p \neq 0\}$.

EXAMPLE 7.11. A smooth 1-form is equivalently a smooth tensor field of type $(0, 1)$:

$$\Omega^1(M) = \Gamma(T_1^0 M).$$

Just as 1-forms $\lambda \in \Omega^1(M)$ are regarded as functions $TM \to \mathbb{R}$, it will often be useful to regard a tensor field $S \in \Gamma(T_\ell^k M)$ in the case $k + \ell > 0$ as a function

$$S : T^* M^{\oplus k} \oplus TM^{\oplus \ell} \to \mathbb{R},$$

where we introduce the notation

$$T^* M^{\oplus k} \oplus TM^{\oplus \ell} := \bigcup_{p \in M} \left(\underbrace{T_p^* M \times \ldots \times T_p^* M}_{k} \times \underbrace{T_p M \times \ldots \times T_p M}_{\ell} \right).$$

The key property of S is then that its restriction S_p to $T_p^* M \times \ldots \times T_p^* M \times T_p M \times \ldots \times T_p M \subset T^* M^{\oplus k} \oplus TM^{\oplus \ell}$ for each $p \in M$ is a multilinear map.

In the setting of smooth manifolds, the term "tensor field" is often abbreviated simply as **tensor**. The terminology for tensors of type (k, ℓ) can also vary among different sources, e.g. one sometimes says that a tensor $S \in \Gamma(T_\ell^k M)$ is **contravariant of rank k** and **covariant of rank ℓ**. The latter terminology is especially favored among physicists.

EXAMPLE 7.12. Under the canonical isomorphism identifying each tangent space $T_p M$ with $\mathrm{Hom}(T_p^* M, \mathbb{R})$, a smooth vector field becomes the same thing as a smooth tensor field of type $(1, 0)$, hence

$$\mathfrak{X}(M) = \Gamma(T_0^1 M).$$

Here the function $T^* M \to \mathbb{R}$ corresponding to a given vector field $X \in \mathfrak{X}(M)$ sends $\lambda \in T_p^* M$ to $\lambda(X(p))$.

EXAMPLE 7.13. Every Riemannian metric (see §7.1.3) is an example of a tensor field of type $(0, 2)$.

EXAMPLE 7.14. Every almost complex structure (see §7.1.4) is an example of a tensor field of type $(1, 1)$.

EXERCISE 7.15. Generalize Lemma 7.7 to show the following: for any finite-dimensional real vector spaces $V_1, \ldots, V_n, W$, there exists a canonical isomorphism

$$\mathrm{Hom}(V_1 \otimes \ldots \otimes V_n, W) \xrightarrow{\Phi} \mathrm{Hom}(W^* \otimes V_1 \otimes \ldots \otimes V_n, \mathbb{R}),$$
$$\Phi(A)(\lambda, v_1, \ldots, v_n) := \lambda(A(v_1, \ldots, v_n)).$$

EXAMPLE 7.16. For arbitrary integers $\ell \geqslant 1$, Exercise 7.15 identifies any tensor field S of type $(1, \ell)$ with a map

$$\bigcup_{p \in M} \left(\underbrace{T_p M \times \ldots \times T_p M}_{\ell} \right) =: TM^{\oplus \ell} \xrightarrow{\widehat{S}} TM$$

whose restriction $\widehat{S}_p$ to $T_p M \times \ldots \times T_p M$ for each $p \in M$ is a multilinear map $T_p M \times \ldots \times T_p M \to T_p M$. The precise correspondence between S and $\widehat{S}$ is given by

$$S(\lambda, X_1, \ldots, X_\ell) = \lambda\big(\widehat{S}(X_1, \ldots, X_\ell)\big),$$

and it is straightforward to show that S is smooth if and only if $\widehat{S}(X_1, \ldots, X_\ell)$ defines a smooth vector field for all choices of smooth vector fields $X_1, \ldots, X_\ell \in \mathfrak{X}(M)$. The case $\ell = 0$ also fits into this picture if one adopts the perspective that a "$T_p M$-valued function of zero variables" just means an element of $T_p M$: this reproduces the observation in Example 7.12 that tensor fields of type $(1, 0)$ are equivalent to vector fields.

REMARK 7.17. The alternative perspective on tensors of type $(1, \ell)$ in Example 7.16 will generally be quite useful, and from now on we will typically use the same notation for the objects that are called S and $\widehat{S}$ in that example. We have already adopted this convention in our discussion of vector fields and almost complex structures as tensors of type $(1, 0)$ and $(1, 1)$ respectively.

DEFINITION 7.18. For $S \in \Gamma(T_\ell^k M)$ and $T \in \Gamma(T_s^r M)$, the **tensor product** (*Tensorprodukt*) of S and T is the tensor field $S \otimes T \in \Gamma(T_{\ell+s}^{k+r} M)$ defined at each point $p \in M$ by

$$(S \otimes T)_p(\lambda^1, \ldots, \lambda^k, \mu^1, \ldots, \mu^r, X_1, \ldots, X_\ell, Y_1, \ldots, Y_s) :=$$
$$S_p(\lambda^1, \ldots, \lambda^k, X_1, \ldots, X_\ell) \cdot T_p(\mu^1, \ldots, \mu^r, Y_1, \ldots, Y_s).$$

REMARK 7.19. For $f \in C^\infty(M) = \Gamma(T_0^0 M)$, the tensor product of f with $S \in \Gamma(T_\ell^k M)$ is just the ordinary point-wise product of S with a scalar-valued function, i.e. $(f \otimes S)_p = (S \otimes f)_p = f(p) S_p$.

7.3. Coordinate representations. We've seen that a chart $(\mathcal{U}, x)$ on M gives rise to coordinate vector fields $\partial_1, \ldots, \partial_n \in \mathfrak{X}(\mathcal{U})$ and coordinate differentials $dx^1, \ldots, dx^n \in \Omega^1(\mathcal{U})$ which define bases of $T_p M$ and $T_p^* M$ respectively at each point $p \in \mathcal{U}$. Regarding vector fields as tensors of type $(1, 0)$, it turns out that a natural basis of $(T_p M)_\ell^k$ can then be constructed by taking all possible tensor products of k coordinate vector fields with ℓ coordinate differentials. Indeed:

PROPOSITION 7.20. *Given a chart $(\mathcal{U}, x)$ on an n-manifold M, every tensor field S of type (k, ℓ) can be written uniquely over $\mathcal{U}$ as*

$$(7.2) \qquad S = S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} \, \frac{\partial}{\partial x^{i_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{i_k}} \otimes dx^{j_1} \otimes \ldots \otimes dx^{j_\ell},$$

*where the $n^{k+\ell}$ **component** functions $S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} : \mathcal{U} \to \mathbb{R}$ are given by*

$$S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} := S(dx^{i_1}, \ldots, dx^{i_k}, \partial_{j_1}, \ldots, \partial_{j_\ell}).$$

REMARK 7.21. Writing down (7.2) without the Einstein summation convention would have required inserting the symbols

$$\sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \sum_{j_1=1}^{n} \cdots \sum_{j_\ell=1}^{n}$$

just to the right of the equal sign, so the right hand side is actually a sum of $n^{k+\ell}$ terms.

PROOF OF PROPOSITION 7.20. Any ℓ vector fields can be written over $\mathcal{U}$ as $X_a = X_a^i \, \partial_i$ for $a = 1, \ldots, \ell$ with unique component functions $X_a^i : \mathcal{U} \to \mathbb{R}$, and similarly, any k 1-forms can be written as $\lambda^b = \lambda_j^b \, dx^j$ with unique components $\lambda_j^b : \mathcal{U} \to \mathbb{R}$. By multilinearity, we then have

$$
\begin{aligned}
(7.3) \qquad S(\lambda^1, \ldots, \lambda^k, X_1, \ldots, X_\ell) &= S(\lambda_{i_1}^1 \, dx^{i_1}, \ldots, \lambda_{i_k}^k \, dx^{i_k}, X_1^{j_1} \, \partial_{j_1}, \ldots, X_\ell^{j_\ell} \, \partial_{j_\ell}) \\
&= S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} \lambda_{i_1}^1 \ldots \lambda_{i_k}^k X_1^{j_1} \ldots X_\ell^{j_\ell}.
\end{aligned}
$$

It is straightforward to check that the tensor field on the right hand side of (7.2) gives the same result when evaluated on the same tuple of vector fields and 1-forms. $\qquad\square$

EXERCISE 7.22. Show that a tensor field of type (k, ℓ) is smooth if and only if for every smooth chart, the corresponding component functions are all smooth.

EXERCISE 7.23. Show that in local coordinates, the components of two tensor fields $S \in \Gamma(T_\ell^k M)$, $T \in \Gamma(T_s^r M)$ and their tensor product $S \otimes T \in \Gamma(T_{\ell+s}^{k+r} M)$ are related by

$$(S \otimes T)^{i_1 \ldots i_k a_1 \ldots a_r}{}_{j_1 \ldots j_\ell b_1 \ldots b_s} = S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} T^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}.$$

EXERCISE 7.24. Suppose $(\mathcal{U}, x)$ and $(\widetilde{\mathcal{U}}, \widetilde{x})$ are two smooth charts with $\mathcal{U} \cap \widetilde{\mathcal{U}} \neq \varnothing$, and denote the component functions of a tensor field $S \in \Gamma(T_\ell^k M)$ with respect to each chart by $S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell}$ and $\widetilde{S}^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell}$ respectively. Prove that on the overlap region $\mathcal{U} \cap \widetilde{\mathcal{U}}$,

$$(7.4) \qquad \widetilde{S}^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} = \frac{\partial \widetilde{x}^{i_1}}{\partial x^{a_1}} \cdots \frac{\partial \widetilde{x}^{i_k}}{\partial x^{a_k}} S^{a_1 \ldots a_k}{}_{b_1 \ldots b_\ell} \frac{\partial x^{b_1}}{\partial \widetilde{x}^{j_1}} \cdots \frac{\partial x^{b_\ell}}{\partial \widetilde{x}^{j_\ell}}.$$

Hint: Use (6.3) and (7.1).

REMARK 7.25. We have been writing all tensor fields so far as functions that take covectors $\lambda^1, \ldots, \lambda^k$ followed by vectors $X_1, \ldots, X_\ell$, but in some circumstances, one may want to be more flexible with the ordering, so that e.g. a tensor of type $(1, 2)$ could be written as a multilinear function

$$TM \oplus T^*M \oplus TM \to \mathbb{R} : (X, \lambda, Y) \mapsto S(X, \lambda, Y).$$

The component functions of such a tensor would then be written as $S_i{}^j{}_k$, with evaluation on $X = X^i \, \partial_i$, $\lambda = \lambda^j \, dx^j$ and $Y = Y^k \, \partial_k$ defined by the rule

$$S(X, \lambda, Y) = S_i{}^j{}_k X^i \lambda_j Y^k.$$

EXAMPLE 7.26. Suppose $J : TM \to TM$ is an almost complex structure, so $J_p : T_p M \to T_p M$ is a linear map satisfying $J_p^2 = -\mathbb{1}$ for every $p \in M$. As we've seen, J can be regarded as a tensor field of type $(1, 1)$ and thus defines a function $T^*M \oplus TM \to \mathbb{R}$, with component functions with respect to a chart $(\mathcal{U}, x)$ written as

$$J^i{}_j = J(dx^i, \partial_j) := dx^i(J\partial_j), \qquad i, j \in \{1, \ldots, n\}.$$

In this line, the second expression views J_p as a bilinear map $T_p^* M \times T_p M \to \mathbb{R}$, while the third views it as a linear map $T_p M \to T_p M$. This means that for two tangent vectors $X = X^i \, \partial_i$ and $Y = Y^i \, \partial_i$ at a point $p \in \mathcal{U}$, we have

$$JX = Y \qquad \Longleftrightarrow \qquad Y^i = dx^i(Y) = dx^i(JX) = dx^i(J(X^j \, \partial_j)) = X^j \, dx^i(J\partial_j) = J^i{}_j X^j,$$

so in other words, the linear map $J_p : T_pM \to T_pM$ is represented in coordinates by matrix-vector multiplication: the n-by-n matrix with entries $J^i{}_j$ gets multiplied by the n-dimensional row vector with entries X^j to produce the row vector with entries $(JX)^i$. The condition $J^2 = -\mathbb{1}$ can thus be expressed in local coordinates on $\mathcal{U}$ as

$$J^i{}_j J^j{}_k \equiv -\delta^i_k \qquad \text{on } \mathcal{U}.$$

From this perspective, the transformation formula (7.4) also ends up looking like something familiar from linear algebra: the component functions $J^i{}_j$ and $\widetilde{J}^i{}_j$ for two overlapping charts $(\mathcal{U}, x)$ and $(\widetilde{\mathcal{U}}, \widetilde{x})$ are related by

$$\widetilde{J}^i{}_j = \frac{\partial \widetilde{x}^i}{\partial x^k} J^k{}_\ell \frac{\partial x^\ell}{\partial \widetilde{x}^j}.$$

In terms of matrices, this just says

$$\widetilde{\mathbf{J}} = \left(\frac{\partial \widetilde{x}}{\partial x} \right) \mathbf{J} \left(\frac{\partial \widetilde{x}}{\partial x} \right)^{-1},$$

where $\mathbf{J}$ and $\widetilde{\mathbf{J}}$ denote the n-by-n matrices with entries $J^i{}_j$ and $\widetilde{J}^i{}_j$ respectively, while $\frac{\partial \widetilde{x}}{\partial x}$ is the n-by-n Jacobian matrix with entries $\frac{\partial \widetilde{x}^i}{\partial x^j}$.

8. Derivatives of tensors and differential forms

We motivate this lecture with the following question: for a smooth tensor field $S \in \Gamma(T^k_\ell M)$, can one define a "directional derivative" of S at a point $p \in M$ in the direction $X \in T_pM$? We considered this question for the special case of vector fields $Y \in \mathfrak{X}(M) = \Gamma(T^1_0 M)$ in §6.3, and the answer we came up with there was not entirely satisfactory: a vector field Y can be differentiated with respect to another vector field X, producing the Lie derivative $\mathcal{L}_X Y \in \mathfrak{X}(M)$, but $\mathcal{L}_X Y(p)$ depends on X as a vector field, not just on the value $X(p)$ (see Remark 6.8). Naively, one might hope for instance that if $S \in \Gamma(T^k_\ell M)$ has components $S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell}$ with respect to some chart $(\mathcal{U}, x)$, then one could define a tensor "dS" of type $(k, \ell + 1)$ whose components are

(8.1)   "$(dS)^{i_1 \ldots i_k}{}_{j_0 \ldots j_\ell} = \partial_{j_0} S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell}$",

so that for any $p \in M$ and $X \in T_pM$, the multilinear map $(dS)(\ldots, X, \ldots) : (T^*_p M)^{\times k} \times (T_p M)^{\times \ell} \to \mathbb{R}$ could be interpreted as the derivative of S in the direction X. But I put that expression in quotation marks because, indeed, it doesn't work: outside of the special case $k = \ell = 0$ where the objects we are differentiating are just real-valued functions, one cannot define from $S \in \Gamma(T^k_\ell M)$ any tensor field $dS \in \Gamma(T^k_{\ell+1} M)$ whose components are given in *all* choices of local coordinates by (8.1). (Exercise 8.1(b) below asks you to prove this in the case $(k, \ell) = (0, 1)$.) In other words, the formula (8.1) is not *coordinate invariant*.

Before discussing directional derivatives further, we should talk about a sticky issue that arose in the previous paragraph: what *practical* methods do we have for writing down the definition of a tensor field? What we attempted above could be called the *physicists' method*: it starts by choosing a chart $(\mathcal{U}, x)$ and writing down a formula for the component functions of the tensor with respect to those local coordinates. That is fine if one only needs a tensor field defined on the subset $\mathcal{U} \subset M$, but the hope of course is that the formula we write down might be valid in *arbitrary* local coordinates, in which case it gives a well-defined tensor field everywhere on M. The important step is therefore to check, using the transformation formula (7.4), that the definition we've written is coordinate invariant, and that is what fails in the case of (8.1). On the other hand, sometimes it succeeds, for instance:

EXERCISE 8.1. Prove:

(a) For any $\lambda \in \Gamma(T_1^0 M)$, there exists a tensor field $S \in \Gamma(T_2^0 M)$ whose components S_{ij} with respect to arbitrary charts $(\mathcal{U}, x)$ are related to the corresponding components λ_i of λ by

$$S_{ij} = \partial_i \lambda_j - \partial_j \lambda_i.$$

(b) For general choices of λ, one cannot similarly define $S \in \Gamma(T_2^0 M)$ so that its relation to λ in arbitrary local coordinates is $S_{ij} = \partial_i \lambda_j$.

Physicists like to summarize the result of Exercise 8.1(a) by saying that the expression $\partial_i \lambda_j - \partial_j \lambda_i$ "defines a tensor" of type $(0, 2)$. In fact, many textbooks on general relativity give a definition of tensors that is cosmetically quite different from ours: without mentioning multilinear maps, they define a tensor S of type (k, ℓ) as an association to each chart $(\mathcal{U}, x)$ of a collection of real-valued functions $S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} : \mathcal{U} \to \mathbb{R}$ that satisfy the transformation formula (7.4). There are good theoretical reasons why mathematicians do not usually give that as the definition of a tensor field, and contrary to what many physicists may tell you, it is also not true that defining a tensor or computing something from it always requires choosing local coordinates.

8.1. C^∞-linearity. Here is a trick for writing down tensor fields that mathematicians tend to prefer, because it does not require local coordinates. For example, let us regard a tensor field S of type $(1, \ell)$ as associating to each point $p \in M$ an ℓ-fold multilinear map $S_p : T_p M \times \ldots \times T_p M \to T_p M$, as described in Example 7.16. It therefore also defines a multilinear map

$$(8.2) \qquad S : \underbrace{\mathfrak{X}(M) \times \ldots \times \mathfrak{X}(M)}_{\ell} \to \mathfrak{X}(M),$$

by interpreting $S(X_1, \ldots, X_\ell)$ for any tuple of smooth vector fields $X_1, \ldots, X_\ell$ as the vector field

$$p \mapsto S_p(X_1(p), \ldots, X_\ell(p)).$$

We already know one important concrete example of multilinear map of this type: the Lie bracket is a bilinear map

$$[\cdot, \cdot] : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M).$$

But does the Lie bracket therefore define a tensor field of type $(1, 2)$? It would be surprising if this were true, because being a tensor field would imply that the value $[X, Y](p)$ for each $p \in M$ depends only on the values $X(p)$ and $Y(p)$, whereas we saw in Exercise 6.2 that in local coordinates, $[X, Y](p)$ also depends on the first derivatives of X and Y at p. An easy way to make this intuition more precise is via the following observation: if S is a tensor field, then the map in (8.2) is not just multilinear, it also satisfies

$$(8.3) \qquad S(X_1, \ldots, X_{j-1}, fX_j, X_{j+1}, \ldots, X_\ell) = fS(X_1, \ldots, X_\ell) \quad \text{for all } f \in C^\infty(M)$$

for every $j = 1, \ldots, \ell$. The key point here is that the function f does not need to be constant, so this is a much stronger statement than just saying that (8.2) respects scalar multiplication (as every multilinear map must). A multilinear map on the space of vector fields is said to be C^∞-**linear** in its jth argument if it satisfies (8.3). In general, the notion of C^∞-linearity can be defined for multilinear maps between any vector spaces on which there is a natural notion of multiplication by smooth functions[31], e.g. we had $\mathfrak{X}(M)$ in the above example because the product of a smooth vector field with a smooth function is also a smooth vector field, but for similar reasons, one could just as well work with $\Omega^1(M)$, the other spaces of smooth tensor fields $\Gamma(T_\ell^k M)$, or $C^\infty(M)$ itself. From this perspective, the obvious reason why the Lie bracket does not define a tensor field is that it is not C^∞-linear: according to Exercise 6.4, it satisfies

$$[fX, Y] = f[X, Y] - (\mathcal{L}_Y f)X, \qquad [X, fY] = f[X, Y] + (\mathcal{L}_X f)Y,$$

---

[31]in other words, spaces that are naturally *modules* over $C^\infty(M)$

for $f \in C^\infty(M)$, which is not the desired relation except in the special case where f is constant.

It will be exceedingly useful to observe that C^∞-linearity is not only necessary for a multilinear map on vector fields or 1-forms to define a tensor field—it is also sufficient.

PROPOSITION 8.2. *For a multilinear map*

$$S : \underbrace{\Omega^1(M) \times \ldots \times \Omega^1(M)}_{k} \times \underbrace{\mathfrak{X}(M) \times \ldots \times \mathfrak{X}(M)}_{\ell} \to C^\infty(M)$$

that is C^∞-linear in every argument, there exists a unique tensor field $\widehat{S} \in \Gamma(T^k_\ell M)$ such that for every $p \in M$, $X_1, \ldots, X_\ell \in \mathfrak{X}(M)$ and $\lambda^1, \ldots, \lambda^k \in \Omega^1(M)$,

$$\widehat{S}(\lambda^1_p, \ldots, \lambda^k_p, X_1(p), \ldots, X_\ell(p)) = S(\lambda^1, \ldots, \lambda^k, X_1, \ldots, X_\ell)(p).$$

Before proving the theorem, let us observe that it can be adapted easily for the slightly different situation in (8.2), where our multilinear map takes values in $\mathfrak{X}(M)$ instead of $C^\infty(M)$:

EXERCISE 8.3. Deduce from Proposition 8.2 that for any multilinear map

$$S : \underbrace{\mathfrak{X}(M) \times \ldots \times \mathfrak{X}(M)}_{\ell} \to \mathfrak{X}(M)$$

that is C^∞-linear in every argument, there exists a unique tensor field $\widehat{S} \in \Gamma(T^1_\ell M)$ such that for every $p \in M$, $X_1, \ldots, X_\ell \in \mathfrak{X}(M)$, the multilinear map $\widehat{S}_p : T_pM \times \ldots \times T_pM \to T_pM$ satisfies

$$\widehat{S}(X_1(p), \ldots, X_\ell(p)) = S(X_1, \ldots, X_\ell)(p).$$

PROOF OF PROPOSITION 8.2. Let us consider only the case $\ell = 1$ and $k = 0$, as there is no substantial difference in the general case beyond requiring more complicated notation. We therefore assume $\Lambda : \mathfrak{X}(M) \to C^\infty(M)$ is a linear map satisfying $\Lambda(fX) = f\Lambda(X)$ for all $f \in C^\infty(M)$ and $X \in \mathfrak{X}(M)$, and we need to find a smooth 1-form $\lambda \in \Omega^1(M)$ such that $\lambda(X(p)) = \Lambda(X)(p)$ for all $p \in M$ and $X \in \mathfrak{X}(M)$. The uniqueness of λ is clear, since every tangent vector at a point $p \in M$ can be the value at that point of a smooth vector field (just write it down in local coordinates, multiply by a smooth cutoff function and extend outside of the coordinate neighborhood as 0).

To prove existence, it suffices to show that for any point $p \in M$, the value of $\Lambda(X)(p)$ is completely determined by $X(p)$ and does not otherwise depend on the choice of vector field X having this particular value at p. This will follow from linearity after proving two claims:

Claim 1: *If $X \in \mathfrak{X}(M)$ vanishes in a neighborhood of p, then $\Lambda(X)(p) = 0$.*

Indeed, if $\mathcal{U} \subset M$ is an open neighborhood on which X vanishes, choose a smooth function $\beta : M \to [0,1]$ with compact support in $\mathcal{U}$ satisfying $\beta(p) = 1$. Then $\beta X \equiv 0$, thus by C^∞-linearity,

$$0 = \Lambda(\beta X) = \beta \Lambda(X) \in C^\infty(M),$$

implying in particular that $\Lambda(X)(p) = \beta(p)\Lambda(X)(p) = 0$.

Claim 2: *If $X \in \mathfrak{X}(M)$ satisfies $X(p) = 0$, then $\Lambda(X)(p) = 0$.*

To see this, choose a chart $(\mathcal{U}, x)$ with $p \in \mathcal{U}$, and write $X = X^i\partial_i$ on $\mathcal{U}$, so the functions $X^i \in C^\infty(\mathcal{U})$ satisfy $X^1(p) = \ldots = X^n(p) = 0$. Using smooth cutoff functions, we can also choose global vector fields $e_1, \ldots, e_n \in \mathfrak{X}(M)$ and functions $f^1, \ldots, f^n \in C^\infty(M)$ such that

$$f^i = X^i \quad \text{and} \quad e_i = \partial_i \qquad \text{near } p, \text{ for all } i = 1, \ldots, n,$$

producing another vector field $Y := f^i e_i \in \mathfrak{X}(M)$ which matches X on some small neighborhood of p within $\mathcal{U}$. Claim 1 then implies $\Lambda(Y - X)(p) = \Lambda(Y)(p) - \Lambda(X)(p) = 0$. In light of C^∞-linearity and the condition $f^i(p) = X^i(p) = 0$ for $i = 1, \ldots, n$, we then have

$$\Lambda(X)(p) = \Lambda(Y)(p) = \Lambda(f^i e_i)(p) = f^i(p)\Lambda(e_i)(p) = 0.$$

$\square$

From now on, we will say that a multilinear map on the spaces of vector fields and/or 1-forms **defines a tensor** whenever it is C^∞-linear in every argument, so that Proposition 8.2 or its obvious corollaries such as Exercise 8.3 apply. We can now carry out the "coordinate free" version of Exercise 8.1:

EXERCISE 8.4. Show that for any given 1-form $\lambda \in \Omega^1(M)$, the tensor of type $(0, 2)$ that was defined via coordinates in Exercise 8.1 can also be defined via the bilinear map
$$\mathfrak{X}(M) \times \mathfrak{X}(M) \to C^\infty(M) : (X, Y) \mapsto \mathcal{L}_X\left[\lambda(Y)\right] - \mathcal{L}_Y\left[\lambda(X)\right] - \lambda([X, Y]),$$
which is C^∞-linear in both arguments. (In this expression, we associate to each vector field $Z \in \mathfrak{X}(M)$ the smooth real-valued function $\lambda(Z) \in C^\infty(M)$ whose value at $p \in M$ is $\lambda(Z(p))$.)

EXERCISE 8.5. Suppose $J \in \Gamma(T^1_1 M)$ is a smooth almost complex structure, which we will regard as a smooth map $J : TM \to TM$ whose restriction to each tangent space $T_p M$ is a linear map $J_p : T_p M \to T_p M$ with $J_p^2 = -\mathbb{1}$. The **Nijenhuis tensor**[32] is defined from J via the map
$$N : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M), \qquad N(X, Y) := [JX, JY] - J[JX, Y] - J[X, JY] - [X, Y].$$
   (a) Use Exercise 8.3 to prove that this formula defines a tensor field of type $(1, 2)$.
   (b) Show that in local coordinates, the components of N and J are related by
   $$N^i_{jk} = J^\ell_j\, \partial_\ell J^i_k - J^\ell_k\, \partial_\ell J^i_j + J^i_\ell\left(\partial_k J^\ell_j - \partial_j J^\ell_k\right).$$
   (c) Show that N vanishes identically if $\dim M = 2$.
   Hint: Notice that $N(X, Y)$ is antisymmetric in X and Y. What is $N(X, JX)$?
   (d) An almost complex structure J is called *integrable* if near every point $p \in M$ there exists a chart $(\mathcal{U}, x)$ in which the components J^i_j become the entries of the constant matrix
   $$\mathbf{J}_0 := \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \in \mathbb{R}^{2n \times 2n},$$
   where each of the four blocks is an n-by-n matrix and $\dim M = 2n$. Show that if J is integrable, then $N \equiv 0$.
   Advice: One can use the formula in part (b) for this, but an argument based directly on the definition of N via Lie brackets is also possible.

Remark: The matrix $\mathbf{J}_0$ represents the linear transformation $\mathbb{C}^n \to \mathbb{C}^n : \mathbf{z} \mapsto i\mathbf{z}$ if one identifies $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ via the correspondence $\mathbb{C}^n \ni \mathbf{x} + i\mathbf{y} \leftrightarrow (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$, thus an integrable almost complex structure makes M into a "complex manifold". By a deep theorem of Newlander and Nirenberg from 1957, the converse of part (d) is also true: if the Nijenhuis tensor vanishes, then J is integrable.

8.2. Differential forms and the exterior derivative. In Exercises 8.1 and 8.4, we saw that if we "antisymmetrize" the partial derivatives of the components of a 1-form, the result is a well-defined tensor field of type $(0, 2)$. We shall now generalize this observation, and in the process, introduce an important special class of tensor fields that will play a major role when we discuss integration on manifolds.

A multilinear map $T : V \times \ldots \times V \to W$ is called **antisymmetric** (*antisymmetrisch*) or **skew-symmetric** (*schiefsymmetrisch*) or **alternating** if the value $T(v_1, \ldots, v_n)$ changes by a sign whenever any two of its arguments are interchanged. One can express this condition equivalently in terms of arbitrary permutations: let S_n denote the **symmetric group** on n elements, which consists of all bijections from the set $\{1, \ldots, n\}$ to itself, also known as **permutations** (*Permutationen*). There are exactly $n!$ elements in S_n, and the group is generated by the so-called *flips*,

---

[32]Approximate pronounciation: "NIGH-en-house", where "nigh" rhymes with English "sigh".

which satisfy $\sigma(i) = j$ and $\sigma(j) = i$ for two distinct elements $i, j \in \{1, \ldots, n\}$ while leaving every other element fixed. Every permutation can therefore be expressed as a composition of flips, and while a given permutation will generally admit many distinct decompositions into varying numbers of flips, one can show that for any fixed $\sigma \in S_n$, the number of flips required is always either even or odd, i.e. a composition of evenly many flips cannot also be expressed as a composition of an odd number of flips, or vice versa. We call each permutation $\sigma \in S_n$ **even** (*gerade*) or **odd** (*ungerade*) accordingly, and define its **parity** by[33]

$$|\sigma| := \begin{cases} 0 & \text{if } \sigma \text{ is even,} \\ 1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

In applications, the parity usually appears in the form $(-1)^{|\sigma|}$, thus one sometimes also refers to odd or even permutations as *negative* or *positive* respectively. With this notion in place, a multilinear map $T : \underbrace{V \times \ldots \times V}_{n} \to W$ is antisymmetric if and only if it satisfies

$$T(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = (-1)^{|\sigma|} T(v_1, \ldots, v_n)$$

for all $v_1, \ldots, v_n \in V$ and $\sigma \in S_n$. One can turn any multilinear map $T : V \times \ldots \times V \to W$ into one that is antisymmetric by defining

$$(\text{Alt}\, T)(v_1, \ldots, v_n) := \frac{1}{n!} \sum_{\sigma \in S_n} (-1)^{|\sigma|} T(v_{\sigma(1)}, \ldots, v_{\sigma(n)}).$$

We observe that $\text{Alt}(T) = T$ if and only if T is antisymmetric, thus Alt defines a linear projection map $\text{Hom}(\bigotimes^n V, W) \to \text{Hom}(\bigotimes^n V, W)$ onto the subspace of antisymmetric maps.

DEFINITION 8.6. For any integer $k \geqslant 0$, an antisymmetric tensor field of type $(0, k)$ on M is called a **differential k-form** (or just **k-form** for short). The vector space of smooth k-forms on M is denoted by

$$\Omega^k(M) := \{\text{smooth } k\text{-forms on } M\}.$$

Note that antisymmetry is a vacuous condition in the cases $k = 0, 1$, which is why $\Omega^1(M) = \Gamma(T^0_1 M)$ and $\Omega^0(M) = \Gamma(T^0_0 M) = C^\infty(M)$. Given a chart $(\mathcal{U}, x)$, a k-form $\omega \in \Omega^k(M)$ can be written in local coordinates as

$$\omega = \omega_{i_1 \ldots i_k}\, dx^{i_1} \otimes \ldots \otimes dx^{i_k} \qquad \text{on } \mathcal{U},$$

where antisymmetry means that the component functions $\omega_{i_1 \ldots i_k} : \mathcal{U} \to \mathbb{R}$ change by a sign whenever two of the indices are interchanged. In this context, the following notational device is often useful. Suppose $T_{i_1 \ldots i_k}$ is a collection of symbols associating to each k-tuple of integers $i_1, \ldots, i_k \in \{1, \ldots, n\}$ an element of some vector space, e.g. $C^\infty(\mathcal{U})$ in the example above. We can then **antisymmetrize** these symbols to define

$$T_{[i_1 \ldots i_k]} := \frac{1}{k!} \sum_{\sigma \in S_k} (-1)^{|\sigma|} T_{i_{\sigma(1)} \ldots i_{\sigma(k)}},$$

so the symbols $T_{[i_1 \ldots i_k]}$ are antisymmetric with respect to interchanging pairs of indices, and one has $T_{[i_1 \ldots i_k]} = T_{i_1 \ldots i_k}$ if and only if $T_{i_1 \ldots i_k}$ already has this property. Note that in this definition, there is no need to assume that $T_{i_1 \ldots i_k}$ are the components of a well-defined tensor, but usefully, it may nonetheless happen that $T_{[i_1 \ldots i_k]}$ *does* define a tensor. We saw an example of this already

---

[33]One easy way to see that the parity is well defined is by associating to each permutation $\sigma \in S_n$ the unique linear map $\mathbf{A}_\sigma : \mathbb{R}^n \to \mathbb{R}^n$ that permutes the standard basis vectors by σ. The matrix of $\mathbf{A}_\sigma$ is obtained from the identity matrix by permuting its columns, and $\det \mathbf{A}_\sigma = (-1)^{|\sigma|}$.

in Exercise 8.1, where the tensor $S \in \Gamma(T_2^0 M)$ defined from any 1-form $\lambda \in \Omega^1(M)$ can now be abbreviated in local coordinates by

$$S_{ij} = 2 \, \partial_{[i} \lambda_{j]}$$

PROPOSITION 8.7. *For every smooth differential form $\omega \in \Omega^k(M)$, $k \geqslant 0$, there exists a unique $(k+1)$-form $d\omega \in \Omega^{k+1}(M)$ determined by the formula*

$$(8.4) \quad d\omega(X_0, \ldots, X_k) = \sum_{i=0}^{k} (-1)^i \mathcal{L}_{X_i} \left[\omega(X_0, \ldots, \widehat{X}_i, \ldots, X_k) \right]$$
$$+ \sum_{0 \leqslant i < j \leqslant k} (-1)^{i+j} \omega \big([X_i, X_j], X_0, \ldots, \widehat{X}_i, \ldots, \widehat{X}_j, \ldots, X_k\big)$$

for $X_0, \ldots, X_k \in \mathfrak{X}(M)$, where the hats over certain terms in sequences like "$X_0, \ldots, \widehat{X}_i, \ldots, X_k$" mean that those terms do not appear in the sequence but every other term does. For any chart $(\mathcal{U}, x)$, the components of $d\omega$ in local coordinates over $\mathcal{U} \subset M$ are given by

$$(d\omega)_{i_0 \ldots i_k} = (k+1) \partial_{[i_0} \omega_{i_1 \ldots i_k]}.$$

PROOF. We claim first that both terms on the right hand side of (8.4) are antisymmetric functions of the vector fields $X_0, \ldots, X_k$. In fact, the first term satisfies

$$(8.5) \quad \sum_{i=0}^{k} (-1)^i \mathcal{L}_{X_i} \left[\omega(X_0, \ldots, \widehat{X}_i, \ldots, X_k) \right] = \frac{1}{k!} \sum_{\sigma \in S_{k+1}} (-1)^{|\sigma|} \mathcal{L}_{X_{\sigma(0)}} \left[\omega(X_{\sigma(1)}, \ldots, X_{\sigma(k)}) \right],$$

where the right hand side is manifestly antisymmetric, and in this setting S_{k+1} means the group of permutations of the elements $\{0, \ldots, k\}$. This can be seen by considering separately for each $i = 0, \ldots, k$ the permutations σ with $\sigma(0) = i$, and then exploiting the antisymmetry of ω to place $X_{\sigma(1)}, \ldots, X_{\sigma(k)}$ in a canonical order. A similar approach shows that the second term is a constant multiple of the antisymmetric expression $\sum_{\sigma \in S_{k+1}} (-1)^{|\sigma|} \omega \big([X_{\sigma(0)}, X_{\sigma(1)}], X_{\sigma(2)}, \ldots, X_{\sigma(k)}\big)$.

We claim next that the right hand side of (8.4) is C^∞-linear in X_i for every $i = 0, \ldots, k$. By antisymmetry, it suffices to prove this for $i = 0$, and the proof is then a straightforward computation based on Exercise 6.4. We can now conclude from Proposition 8.2 that $d\omega$ is a well-defined $(k+1)$-form. Finally, the coordinate formula for $d\omega$ follows from (8.5) since $[\partial_i, \partial_j] \equiv 0$ for all i, j. $\qquad \square$

DEFINITION 8.8. For a smooth k-form on ω, the $(k+1)$-form $d\omega$ defined in Proposition 8.7 is called the **exterior derivative** (*äußere Ableitung*) of ω.

EXAMPLE 8.9. For a 0-form $f \in C^\infty(M) = \Omega^0(M)$, the definition above makes $df \in \Omega^1(M)$ the usual differential of f.

For $k > 0$, the exterior derivative $d\omega$ of $\omega \in \Omega^k(M)$ does not contain *all* information about the first derivative of ω at each point, e.g. in local coordinates, the individual partial derivatives $\partial_j \omega_{i_1 \ldots i_k}$ cannot be deduced from $(d\omega)_{i_0 \ldots i_k}$, nor can ω be recovered from $d\omega$ up to addition of a constant. We will see more comprehensive (though non-canonical) ways of defining derivatives of ω when we discuss connections. The exterior derivative will be essential, however, due to the role it plays in Stokes' theorem, the n-dimensional generalization of the fundamental theorem of calculus.

8.3. Pullbacks and pushforwards. For a diffeomorphism $\psi : M \to N$, pushforwards and pullbacks of tensor fields can be defined in much the same way as for functions and vector fields in §5.2. Recalling the notation

$$\psi_* := T\psi : TM \to TN, \qquad \psi^* := (T\psi)^{-1} : TN \to TM,$$

we can dualize to define

$$\psi^* : T^*N \to T^*M, \qquad \psi_* : T^*M \to T^*N$$

by

$$(\psi^*\lambda)(X) := \lambda(\psi_* X), \qquad (\psi_*\lambda)(X) := \lambda(\psi^* X).$$

Every $S \in \Gamma(T_\ell^k M)$ with $k > 0$ or $\ell > 0$ then has a **pushforard** $\psi_* S \in \Gamma(T_\ell^k N)$ defined by

$$(\psi_* S)(\lambda^1, \ldots, \lambda^k, X_1, \ldots, X_\ell) := S(\psi^*\lambda^1, \ldots, \psi^*\lambda^k, \psi^* X_1, \ldots, \psi^* X_\ell),$$

and similarly, $S \in \Gamma(T_\ell^k N)$ has a **pullback** $\psi^* S \in \Gamma(T_\ell^k M)$ defined by

$$(\psi^* S)(\lambda^1, \ldots, \lambda^k, X_1, \ldots, X_\ell) := S(\psi_*\lambda^1, \ldots \psi_*\lambda^k, \psi_* X_1, \ldots, \psi_* X_\ell).$$

The reader should take a moment to check that under the canonical identification $\mathfrak{X}(M) = \Gamma(T_0^1 M)$, this definition of the pushforward and pullback for tensor fields of type $(1, 0)$ matches what we defined in §5.2 for vector fields. The maps

$$\psi_* : \Gamma(T_\ell^k M) \to \Gamma(T_\ell^k N), \qquad \psi^* : \Gamma(T_\ell^k N) \to \Gamma(T_\ell^k M)$$

are vector space isomorphisms, and are inverse to each other. It is straightforward to show that if $\varphi : N \to Q$ is another diffeomorphism, the composition $\varphi \circ \psi : M \to Q$ satisfies

$$(8.6) \qquad\qquad (\varphi \circ \psi)_* = \varphi_*\psi_*, \qquad (\varphi \circ \psi)^* = \psi^*\varphi^*.$$

Notice that the pushforward $\psi_* X = T\psi(X) \in TN$ of a tangent vector $X \in TM$ is defined without reference to the inverse ψ^{-1}, and can therefore also be defined when $\psi : M \to N$ is any smooth map, not necessarily a diffeomorphism. The same thus holds for the *pullback* of a fully covariant tensor field $S \in \Gamma(T_k^0 N)$: the definition of $\psi^* S \in \Gamma(T_k^0 M)$ as

$$\psi^* S(X_1, \ldots, X_\ell) = S(\psi_* X_1, \ldots, \psi_* X_\ell) = S(T\psi(X_1), \ldots, T\psi(X_\ell))$$

makes sense for any smooth map $\psi : M \to N$, though the resulting linear map $\psi^* : \Gamma(T_k^0 N) \to \Gamma(T_k^0 M)$ need not be invertible if ψ is not a diffeomorphism. This applies in particular for differential forms: they can always be pulled back via smooth maps.

EXERCISE 8.10. Assume $\psi : M \to N$ is a smooth map and $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ are charts on M and N respectively such that $\mathcal{U} \cap \psi^{-1}(\mathcal{V}) \neq \varnothing$. Abbreviating $\psi^i := y^i \circ \psi : \psi^{-1}(\mathcal{V}) \to \mathbb{R}$ for the component functions of ψ written in coordinates, show that the components of a k-form $\omega \in \Omega^k(N)$ in the coordinates $y^1, \ldots, y^n$ are related to those of its pullback $\psi^*\omega \in \Omega^k(M)$ in coordinates $x^1, \ldots, x^m$ by

$$(\psi^*\omega)_{i_1 \ldots i_k} = \frac{\partial \psi^{j_1}}{\partial x^{i_1}} \ldots \frac{\partial \psi^{j_k}}{\partial x^{i_k}}(\omega_{j_1 \ldots j_k} \circ \psi) \qquad \text{on } \mathcal{U} \cap \psi^{-1}(\mathcal{V}).$$

8.4. The Lie derivative of a tensor field. As with vector fields in §6.3, there is a natural way to differentiate any tensor field $S \in \Gamma(T_\ell^k M)$ with respect to a vector field $X \in \mathfrak{X}(M)$, giving the most general version of the **Lie derivative**

$$\mathcal{L}_X S := \left. \frac{d}{dt}(\varphi_X^t)^* X \right|_{t=0} \in \Gamma(T_\ell^k M).$$

This is well defined even if none of the flow maps φ_X^t are globally defined on M for $t \neq 0$, since for any point $p \in M$, φ_X^t is at least defined on a neighborhood of p for every t close enough to 0.

As with the Lie derivative of vector fields, one should keep in mind that for each $p \in M$, $(\mathcal{L}_X S)_p$ depends on more than just S and the value of X at p, due to the fact that pulling back via the flow requires differentiating it, and this derivative will also depend on the derivatives of X at p. The only exception is the case $k = \ell = 0$, in which S is just a function $f : M \to \mathbb{R}$ and $\mathcal{L}_X f = df(X)$ as before.

The Lie derivative has important applications to questions of *invariance*, e.g. if $\dim M = n$, we will see that one can use a differential form $\omega \in \Omega^n(M)$ to define a notion of *volume* for regions in M, and the condition $\mathcal{L}_X \omega \equiv 0$ will then characterize vector fields whose flows are volume preserving. We will need to develop the technology somewhat further before we can do nontrivial things with this, as it is typically quite difficult to compute $\mathcal{L}_X S$ directly from the definition, due to the fact that the flow of a vector field is typically not easy to write down. Let us mention however that there is a very user-friendly formula for the Lie derivative of a differential form:

THEOREM 8.11 (Cartan's formula). *For any $\omega \in \Omega^k(M)$ and $X \in \mathfrak{X}(M)$,*
$$\mathcal{L}_X \omega = d(\iota_X \omega) + \iota_X(d\omega),$$
*where the **interior product** $\iota_X \alpha \in \Omega^{q-1}(M)$ of a differential form $\alpha \in \Omega^q(M)$ with a vector field $X \in \mathfrak{X}(M)$ is defined by*
$$(\iota_X \alpha)(Y_1, \ldots, Y_{q-1}) := \alpha(X, Y_1, \ldots, Y_{q-1}).$$

We will prove this in Lecture 11, after we have discussed the algebra of differential forms in more detail.

9. The algebra of differential forms

Our goal for the next two lectures is to make sense of symbols like $\int_M f$ when M is a manifold. The naive hope would be that one could associate a real number $\int_M f \in \mathbb{R}$ to every (let's say continuous and compactly supported) function $f : M \to \mathbb{R}$, one that weights the values of f in proportion to the amount of volume covered. We will see that this notion does not make sense in general for real-valued *functions*, but if $\dim M = n$, it does make sense when f is replaced by a differential n-form.

9.1. Measure and volume on manifolds. The basic problem with defining $\int_M f$ for a function $f : M \to \mathbb{R}$ is that we have not specified any measure on M with which to define what "volume" means. Certain special classes of manifolds admit canonical measures, e.g. if M is a k-dimensional submanifold of $\mathbb{R}^n$, then one can derive a notion of "k-dimensional volume" on subsets of M from the Euclidean geometry of $\mathbb{R}^n$. But this measure on M will depend on the precise embedding $M \hookrightarrow \mathbb{R}^n$, e.g. the volume of any given region in M will change by a factor of L^k if we modify the embedding by multiplication with a scalar $L > 0$. And in any case, not all manifolds are presented as submanifolds of Euclidean space.

Another idea would be to use local coordinates, meaning that for any chart $(x, \mathcal{U})$ on M, the measure of a subset $\mathcal{O} \subset \mathcal{U}$ could be defined as the Lebesgue measure of $x(\mathcal{O}) \subset \mathbb{R}^n$. This definition, however, clealy depends on the choice of chart: according to the change of variables formula, the Lebesgue measure of $y(\mathcal{O}) \subset \mathbb{R}^n$ for another chart $(\mathcal{V}, y)$ with $\mathcal{O} \subset \mathcal{V}$ will be the Lebesgue integral of $|\det D(y \circ x^{-1})|$ over $x(\mathcal{O})$, and this integral is not typically the same as the measure of $x(\mathcal{O})$.

Let us drop the question of whether M carries a canonical measure (usually it doesn't), and ask instead how one might go about *choosing* a measure on M, i.e. what kinds of properties should a notion of n-dimensional volume on M have? Heuristically, one useful way to approach this question is by thinking of the tangent space $T_p M$ at a point $p \in M$ is an "approximation" of a neighborhood of p in M, so if we can define volumes of regions in that neighborhood, we should

also be able to define volumes of regions in the vector space T_pM. How does one define volume in an n-dimensional vector space? For example, given vectors $X_1, \ldots, X_n \in T_pM$, consider the so-called **parallelepiped** spanned by $X_1, \ldots, X_n$, meaning the set

$$P(X_1, \ldots, X_n) := \left\{ t^i X_i \in T_pM \mid t^1, \ldots, t^n \in [0,1] \right\} \subset T_pM,$$

where as usual there is an implied summation in the expression $t^i X_i$. Suppose $\mu : T_pM \times \ldots \times T_pM \to [0, \infty)$ is a function that associates to each n-tuple $(X_1, \ldots, X_n)$ the n-dimensional volume of $P(X_1, \ldots, X_n)$. What kind of function is μ? Basic geometric considerations dictate the following:

(1) If one of the vectors X_i is multiplied by a nonnegative constant, the volume scales by the same constant, i.e.

$$\mu(X_1, \ldots, cX_i, \ldots, X_n) = c\mu(X_1, \ldots, X_i, \ldots, X_n)$$

for $c \geq 0$.

(2) The volume is additive[34] with respect to each variable, i.e.

$$\mu(X_1, \ldots, X_i + X_i', \ldots, X_n) = \mu(X_1, \ldots, X_i, \ldots, X_n) + \mu(X_1, \ldots, X_i', \ldots, X_n).$$

An elementary geometric justification of this relation in the case $n = 2$ is shown in Figure 7. Using the letters A through E to denote the areas of the various regions in this picture, one has $\mu(X_1, X_2) = A + B$, $\mu(X_1', X_2) = C + D$, and $\mu(X_1 + X_1', X_2) = A + C + E = A + C + B + D = \mu(X_1, X_2) + \mu(X_1', X_2)$.

(3) If any two of the vectors $X_1, \ldots, X_n$ match, then $P(X_1, \ldots, X_n)$ is contained in an $(n-1)$-dimensional subspace and thus has zero n-dimensional volume, so

$$\mu(X_1, \ldots, X_n) = 0 \quad \text{whenever} \quad X_i = X_j \text{ for some } i \neq j.$$

The first two properties suggest multilinearity, though μ itself cannot be multilinear since it only takes nonnegative values, and the scalar multiplication property only involves nonnegative scalars. On the other hand, a good way to find functions μ that satisfy these two properties is by choosing an actual multilinear function $\omega : T_pM \times \ldots \times T_pM \to \mathbb{R}$ and setting

$$\mu(X_1, \ldots, X_n) := |\omega(X_1, \ldots, X_n)|.$$

The third property now imposes a serious restriction on ω:

PROPOSITION 9.1. *If V is a vector space and $\omega : V \times \ldots \times V \to \mathbb{R}$ is an n-fold multilinear function that vanishes whenever two of its arguments are identical, then ω is alternating.*

PROOF. In the case $n = 2$, it suffices to choose any $v, w \in V$ and use multilinearity to observe

$$0 = \omega(v + w, v + w) = \omega(v, v) + \omega(w, w) + \omega(v, w) + \omega(w, v) = \omega(v, w) + \omega(w, v).$$

The general case works similarly. $\square$

The upshot of this discussion is that a reasonable notion of volume for paralelepipeds in a tangent space T_pM can be defined by choosing an alternating n-fold multilinear form ω on T_pM and taking its absolute value. If the gaps in the discussion leading to this conclusion made you uncomfortable, one could alternatively derive it from a basic result in measure theory: every translation-invariant measure on $\mathbb{R}^n$ is a scalar $c \geq 0$ multiplied by the Lebesgue measure (see e.g. [**Sal16**, Chapter 2]). Moreover, the Lebesgue measure of the parallelepiped spanned by n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^n$ is given by $\left| \det \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix} \right|$. As you learned in linear algebra, the

---

[34]Strictly speaking, some extra condition on the vectors $X_1, \ldots, X_n$ is needed in order for the additivity property to hold, as not all possible configurations (even in the case $n = 2$) can be described by something like Figure 7. Since this is only meant to be a heuristic discussion, let's not worry about this for now.
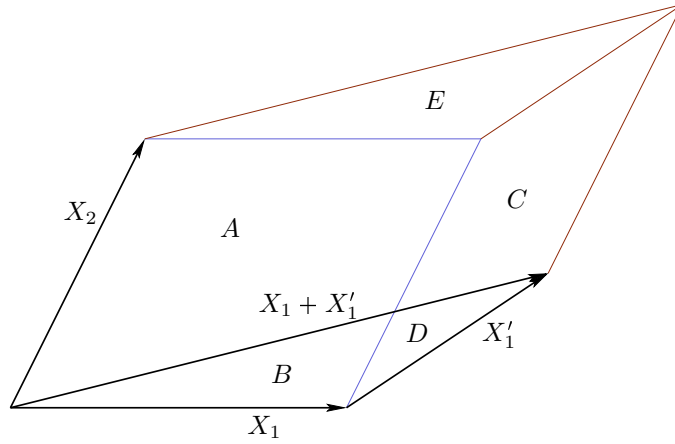
FIGURE 7.   A geometric "proof" that volumes of parallelepipeds are determined by *multilinear* functions of their spanning vectors.

determinant of a matrix is an alternating multilinear function of its columns, thus we can now write $\mu = |\omega|$ where $\omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) := c \det \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}$ defines an alternating multilinear form.

Since everything in this course is smooth, it will also make sense to assume that for reasonable notions of volume on regions in M, the associated notions of volume on the tangent spaces T_pM depend smoothly on the point p. We can now say precisely what kind of geometric object defines a smoothly varying notion of volume on tangent spaces: it is a smooth n-form $\omega \in \Omega^n(M)$.

9.2. Exterior algebra. The previous section provided some motivation to believe that differential forms are the right objects with which to define integration on manifolds. Before we can fully unpack this idea, we need to develop the algebra of differential forms a bit further.

The tasks of this section are fundamentally algebraic, so there will be no manifolds, only an n-dimensional vector space V with basis $e_1, \ldots, e_n \in V$. Let $e_*^1, \ldots, e_*^n \in V^*$ denote the corresponding **dual basis**, determined by the condition

$$e_*^i(e_j) = \delta_j^i.$$

Recall from §7.2 that V_ℓ^k denotes the space of multilinear functions $V^* \times \ldots \times V^* \times V \times \ldots \times V \to \mathbb{R}$ that take k dual vectors in V^* and ℓ vectors in V as arguments; in particular, $V_1^0 = V^*$ and V_0^1 is the "double dual" $(V^*)^*$ of V, which is canonically isomorphic to V itself. The tensor product $\otimes : V_\ell^k \times V_s^r \to V_{\ell+s}^{k+r}$ can be defined in the same way as for tensor fields, and it is associative, so in particular, the tensor product of k dual vectors $\alpha^1, \ldots, \alpha^k$ is a k-fold multilinear map $\alpha^1 \otimes \ldots \otimes \alpha^k : V \times \ldots \times V \to \mathbb{R}$ defined by

$$(\alpha^1 \otimes \ldots \otimes \alpha^k)(v_1, \ldots, v_k) = \alpha^1(v_1) \cdot \ldots \cdot \alpha^k(v_k).$$

The vector space of real-valued alternating k-fold multilinear maps on V is denoted by

$$\Lambda^k V^* := \left\{ \omega \in V_k^0 \mid \omega(\ldots, v, \ldots, w, \ldots) = -\omega(\ldots, w, \ldots, v, \ldots) \text{ for all } v, w \in V \right\},$$

and we often refer to its elements as **alternating k-forms on V**. The antisymmetry condition is vacuous for $k \leqslant 1$, thus $\Lambda^0 V^* = \mathbb{R}$ and $\Lambda^1 V^* = V^*$. Using multilinearity as in Proposition 7.20, any $\omega \in \Lambda^k V^*$ for $k \geqslant 1$ can be written in terms of the basis $e_*^1, \ldots, e_*^n \in V^*$ as

$$\omega = \omega_{i_1 \ldots i_k} \, e_*^{i_1} \otimes \ldots \otimes e_*^{i_k},$$

with unique coefficients

$$(9.1) \qquad \omega_{i_1 \ldots i_k} := \omega(e_{i_1}, \ldots, e_{i_k}) \in \mathbb{R}.$$

These coefficients are not all independent of each other: the antisymmetry of ω dictates that they satisfy

$$\omega_{i_1 \ldots j \ldots \ell \ldots i_k} = -\omega_{i_1 \ldots \ell \ldots j \ldots i_k},$$

i.e. there is a sign change whenever two distinct indices are interchanged, and $\omega_{i_1 \ldots i_k}$ can only be nontrivial when all of its indices $i_1, \ldots, i_k \in \{1, \ldots, n\}$ have distinct values. It follows that $\omega_{i_1 \ldots i_k}$ must always vanish if $k > n$, and otherwise, the number of distinct components that can be specified independently before the rest are determined is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, hence

$$\dim \Lambda^k V^* = \begin{cases} \binom{n}{k} = \frac{n!}{k!(n-k)!} & \text{for } k \leqslant n, \\ 0 & \text{for } k > n. \end{cases}$$

Observe that while the case $k = 0$ was excluded from the discussion above, the formula $\dim \mathbb{R} = \dim \Lambda^0 V^* = \binom{n}{0} = 1$ is also correct in that case. The most interesting case is $k = n$: the elements of $\Lambda^n V^*$ are sometimes called **top-dimensional** forms, since n is the largest value of k for which $\Lambda^k V^*$ is a nontrivial space. The space is 1-dimensional in this case, due to the fact that all nontrivial components of $\omega \in \Lambda^n V^*$ are obtained by permuting the indices of $\omega_{1 \ldots n}$. This elementary observation has nontrivial consequences that will be concretely useful to us, such as:

PROPOSITION 9.2. *For any basis $v_1, \ldots, v_n \in V$ of a vector space V, every $\omega \in \Lambda^n V^*$ is uniquely determined by the number $\omega(v_1, \ldots, v_n) \in \mathbb{R}$; in particular, this number vanishes if and only if $\omega = 0$.* $\qquad \square$

EXAMPLE 9.3. The **determinant** $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$ can be characterized by the property that $\mathbb{R}^n \times \ldots \times \mathbb{R}^n \to \mathbb{R} : (\mathbf{v}_1, \ldots, \mathbf{v}_1) \mapsto \det \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}$ is the unique element of $\Lambda^n(\mathbb{R}^n)^*$ satisfying $\det \begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{pmatrix} = 1$ for the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$. Using the dual basis $\mathbf{e}_*^1, \ldots, \mathbf{e}_*^n \in (\mathbb{R}^n)^*$ to the standard basis, one can write down a concrete element of $\Lambda^n(\mathbb{R}^n)^*$ with this property in the form

$$\sum_{\sigma \in S_n} (-1)^{|\sigma|} \mathbf{e}_*^{\sigma(1)} \otimes \ldots \otimes \mathbf{e}_*^{\sigma(n)} \in \Lambda^n(\mathbb{R}^n)^*.$$

Plugging in the columns of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with entries $A^i{}_j$, an explicit formula for the determinant is thus given by

$$(9.2) \qquad \det(\mathbf{A}) = \sum_{\sigma \in S_n} (-1)^{|\sigma|} A^{\sigma(1)}{}_1 \cdot \ldots \cdot A^{\sigma(n)}{}_n.$$

Proposition 9.2 now implies that every $\omega \in \Lambda^n(\mathbb{R}^n)^*$ can be written as

$$\omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) = c \cdot \det \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix},$$

with a constant given by $c := \omega(\mathbf{e}_1, \ldots, \mathbf{e}_n) \in \mathbb{R}$.

For $k \geqslant 1$, a natural linear projection $\mathrm{Alt} : V_k^0 \to V_k^0$ onto the subspace $\Lambda^k V^* \subset V_k^0$ is defined by

$$\mathrm{Alt}(\omega)(v_1, \ldots, v_k) := \frac{1}{k!} \sum_{\sigma \in S_k} (-1)^{|\sigma|} \omega(v_{\sigma(1)}, \ldots, v_{\sigma(k)}).$$

Indeed, one readily checks that $\mathrm{Alt}(\omega)$ is alternating for every $\omega \in V_k^0$, and ω itself is alternating if and only if $\mathrm{Alt}(\omega) = \omega$. If we write $\omega = \omega_{i_1 \ldots i_k} e_*^{i_1} \otimes \ldots \otimes e_*^{i_k}$ for a general $\omega \in V_k^0$, applying

Alt changes the components via the antisymmetrization operation introduced in §8.2, which can
be written succinctly as
$$\mathrm{Alt}(\omega)_{i_1 \ldots i_k} = \omega_{[i_1 \ldots i_k]}.$$
Note that for $k = 1$, Alt is simply the identity map $V^* \to V^*$. It will be a useful convention to
extend this definition to $k = 0$ so that Alt is also the identity map on $V_0^0 = \mathbb{R}$.

We would now like to define a product operation on alternating forms that has geometric
meaning. Let us regard each of the chosen basis 1-forms $e_*^i \in \Lambda^1 V^*$ as defining a notion of *length*
(also known as "1-dimensional volume") for vectors in the 1-dimensional subspace $V_i := \mathbb{R}e_i \subset V$,
so by this definition, the basis vectors $e_i \in V_i$ have unit length. The fact that each e_*^i vanishes on all
the other subspaces $V_j \subset V$ for $j \neq i$ can be interpreted moreover as an "orthogonality" condition,
so that we regard all the subspaces $V_1, \ldots, V_n \subset V$ as orthogonal to each other. Geometrically, the
paralelepiped in V spanned by $e_1, \ldots, e_n$ should then have volume 1, and we would like to define
the product n-form $e_*^1 \wedge \ldots \wedge e_*^n \in \Lambda^n V^*$ to reproduce this notion of volume, i.e. it should satisfy
$$(e_*^1 \wedge \ldots \wedge e_*^n)(e_1, \ldots, e_n) = 1.$$
Since $\dim \Lambda^n V^* = 1$, there is exactly one element of $\Lambda^n V^*$ that satisfies this condition, and it is
given by
$$e_*^1 \wedge \ldots \wedge e_*^n = n! \, \mathrm{Alt}(e_*^1 \otimes \ldots \otimes e_*^n) = \sum_{\sigma \in S_n} (-1)^{|\sigma|} e_*^{\sigma(1)} \otimes \ldots \otimes e_*^{\sigma(n)}.$$

We take this observation as motivation for the general definition of the **wedge product**, which is
contained in the theorem below. To state it properly, we define the vector space
$$\Lambda^* V^* := \bigoplus_{k=0}^{\infty} \Lambda^k V^*,$$

which is finite dimensional since $\Lambda^k V^* = \{0\}$ for $k > n$, hence $\Lambda^* V^*$ is equivalent to the finite
product $\Lambda^0 V^* \times \ldots \times \Lambda^n V^*$. We can regard each of the spaces $\Lambda^k V^*$ as subspaces of $\Lambda^* V^*$ in
the obvious way. A nontrivial element $\alpha \in \Lambda^* V^*$ is said to be **homogeneous of degree** k if it
belongs to the subspace $\Lambda^k V^* \subset \Lambda^* V^*$, in which case we also sometimes write its degree as
$$\deg(\alpha) = |\alpha| := k \quad \text{for} \quad \alpha \in \Lambda^k V^*.$$

One should keep in mind that not all elements of $\Lambda^* V^*$ are homogeneous, but this is of little im-
portance in practice because every nontrivial element is a sum of a unique finite set of homogeneous
elements of various degrees.

THEOREM 9.4. *There exists a unique bilinear map $\Lambda^* V^* \times \Lambda^* V^* \to \Lambda^* V^* : (\alpha, \beta) \mapsto \alpha \wedge \beta$
that satisfies*
$$c \wedge \alpha = \alpha \wedge c := c\alpha \qquad \text{for all } \alpha \in \Lambda^* V^* \text{ and } c \in \Lambda^0 V^* = \mathbb{R},$$
the associativity property
$$(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma) \qquad \text{for all } \alpha, \beta, \gamma \in \Lambda^* V^*,$$
and
$$(9.3) \qquad \alpha^1 \wedge \ldots \wedge \alpha^k = \sum_{\sigma \in S_k} (-1)^{|\sigma|} \alpha^{\sigma(1)} \otimes \ldots \otimes \alpha^{\sigma(k)} \qquad \text{for all } k \in \mathbb{N}, \ \alpha^1, \ldots, \alpha^k \in \Lambda^1 V^*,$$
*where the k-fold product on the left hand side is defined by arbitrarily inserting parentheses to
produce a sequence of binary operations. Moreover, the following conditions are satisfied:*

(1) For any integers $k, \ell \geqslant 0$ and $\alpha \in \Lambda^k V^*$, $\beta \in \Lambda^\ell V^*$,
$$(9.4) \qquad \alpha \wedge \beta = \frac{(k+\ell)!}{k! \ell!} \mathrm{Alt}(\alpha \otimes \beta) \in \Lambda^{k+\ell} V^*.$$

(2) *The wedge product is* graded commutative, *i.e. for homogeneous elements* $\alpha, \beta \in \Lambda^* V^*$,

$$\alpha \wedge \beta = (-1)^{|\alpha| \cdot |\beta|} \beta \wedge \alpha.$$

Before proving the theorem, we make the useful observation that if one defines k-fold wedge products of 1-forms via the right hand side of (9.3), then they can be used to turn any basis of V^* into a basis of $\Lambda^k V^*$:

PROPOSITION 9.5. *Given the basis* $e_1, \ldots, e_n \in V$ *and its dual basis* $e_*^1, \ldots, e_*^n \in V^*$, *every* $\omega \in \Lambda^k V^*$ *can be written as*

(9.5)
$$\omega = \sum_{i_1 < \ldots < i_k} \omega_{i_1 \ldots i_k} \, e_*^{i_1} \wedge \ldots \wedge e_*^{i_k}$$

for unique coefficients $\omega_{i_1 \ldots i_k} \in \mathbb{R}$, *which are given by*[35]

$$\omega_{i_1 \ldots i_k} = \omega(e_{i_1}, \ldots, e_{i_k}) \in \mathbb{R}.$$

PROOF. One uses the formula (9.3) to show that both sides of (9.5) match when evaluated on any tuple of basis vectors $(e_{i_1}, \ldots, e_{i_k})$ with $i_1 < \ldots < i_k$, and by antisymmetry, it follows that they also match when evaluated on *any* tuple of basis vectors. Multilinearity then implies that they match when evaluated on arbitrary k-tuples of vectors. $\square$

REMARK 9.6. Proposition 9.5 is one of the few places where we are *not* using the Einstein summation convention. The reason is that the summation here does not cover all choices of tuples $i_1, \ldots, i_k \in \{1, \ldots, n\}$, as the summation convention would dictate, but rather only those for which the $i_1, \ldots, i_k$ are in strictly increasing order. Including all permutations of such tuples would produce extra terms that (due to the antisymmetry of both $\omega_{i_1 \ldots i_k}$ and $e_*^{i_1} \wedge \ldots \wedge e_*^{i_k}$) match the terms already present in the sum, i.e. exactly $k!$ copies of each term, plus some trivial terms for tuples in which some of the indices $i_1, \ldots, i_k$ match. This overcounting results in the formula

$$\omega = \frac{1}{k!} \omega_{i_1 \ldots i_k} \, e_*^{i_1} \wedge \ldots \wedge e_*^{i_k},$$

in which the coefficients are defined the same as before but the summation convention *is* in effect.

EXAMPLE 9.7. The following case of (9.3) is worth drawing special attention to: for two 1-forms $\alpha, \beta \in \Lambda^1 V^*$, $\alpha \wedge \beta \in \Lambda^2 V^*$ is given by $\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha$, thus

$$(\alpha \wedge \beta)(v, w) = \alpha(v)\beta(w) - \beta(v)\alpha(w).$$

One sees easily from this formula that the wedge product of 1-forms is *anticommutative*, i.e. it satisfies $\alpha \wedge \beta = -\beta \wedge \alpha$, and in particular, $\alpha \wedge \alpha = 0$.

PROOF OF THEOREM 9.4. By Proposition 9.5, every $\alpha \in \Lambda^k V^*$ and $\beta \in \Lambda^\ell V^*$ for $k, \ell \geqslant 1$ can be expressed as sums of wedge products of the basis 1-forms $e_*^1, \ldots, e_*^n \in V^*$ as determined by (9.3), so bilinearity and associativity together with (9.3) then uniquely determine $\alpha \wedge \beta \in \Lambda^{k+\ell} V^*$. The only problem with taking the resulting formula as a general *definition* of $\alpha \wedge \beta$ is that it may a priori depend on the choice of the basis $e_*^1, \ldots, e_*^n$. In order to dismiss this concern, we will show that this definition of $\alpha \wedge \beta$ also satisfies the formula (9.4), and observe that the right hand side of this expression is clearly independent of choices. By bilinearity and Proposition 9.5, it suffices to check that this is true when α and β are themselves products of the form

$$\alpha = e_*^{i_1} \wedge \ldots \wedge e_*^{i_k}, \qquad \beta = e_*^{j_1} \wedge \ldots \wedge e_*^{j_\ell}$$

---

[35]Notice that the coefficients in Proposition 9.5 are the same ones that appeared in (9.1).

for some choice of $i_1, \ldots, i_k, j_1, \ldots, j_\ell \in \{1, \ldots, n\}$, and to show this, it is enough to evaluate both $\alpha \wedge \beta$ (as defined via (9.3)) and the right hand side of (9.4) on the ordered tuple of basis vectors

$$e_{a_1}, \ldots, e_{a_k}, e_{b_1}, \ldots, e_{b_\ell} \in V$$

for an arbitrary choice of $a_1, \ldots, a_k, b_1, \ldots, b_\ell \in \{1, \ldots, n\}$. By antisymmetry, both clearly vanish unless the integers $a_1, \ldots, a_k, b_1, \ldots, b_\ell$ are all distinct, so let us assume this. Both will also vanish if any of those numbers are not contained in the set $\{i_1, \ldots, i_k, j_1, \ldots, j_\ell\}$, so assume this as well from now on, which implies that the numbers $i_1, \ldots, i_k, j_1, \ldots, j_\ell$ must also be all distinct, and thus

$$\{a_1, \ldots, a_k, b_1, \ldots, b_\ell\} = \{i_1, \ldots, i_k, j_1, \ldots, j_\ell\}.$$

Using antisymmetry, we can now apply a permutation and assume without loss of generality that the two ordered tuples are exactly the same, i.e. $a_m = i_m$ and $b_m = j_m$ for all m, so we need only evaluate both $\alpha \wedge \beta$ and $\frac{(k+\ell)!}{k!\ell!} \mathrm{Alt}(\alpha \otimes \beta)$ on the ordered tuple

$$(v_1, \ldots, v_{k+\ell}) := e_{i_1}, \ldots, e_{i_k}, e_{j_1}, \ldots, e_{j_\ell}.$$

The result for $\alpha \wedge \beta$ is immediate from (9.3): only the trivial permutation produces a nontrivial term, and the answer is 1. Now consider

$$\frac{(k+\ell)!}{k!\ell!} \mathrm{Alt}(\alpha \otimes \beta)(v_1, \ldots, v_{k+\ell}) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} (-1)^{|\sigma|} \alpha(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) \cdot \beta(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}).$$

Since the sets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_\ell\}$ are disjoint, the only permutations that contribute nontrivially to the right hand side of this expression are those which preserve the subsets $\{1, \ldots, k\}$ and $\{k+1, \ldots, k+\ell\}$, and the sign of such a permutation is the product of the signs of the permutations of these two subsets, so the sum can be rewritten as

$$\frac{1}{k!\ell!} \sum_{(\sigma_1, \sigma_2) \in S_k \times S_\ell} (-1)^{|\sigma_1|} \alpha(e_{i_{\sigma_1(1)}}, \ldots, e_{i_{\sigma_1(k)}}) \cdot (-1)^{|\sigma_2|} \beta(e_{j_{\sigma_2(1)}}, \ldots, e_{j_{\sigma_2(\ell)}}).$$

Finally, observe that since α and β are both antisymmetric, every term in this last sum is identical, and there are exactly $k!\ell!$ of them, so we can restrict to the trivial permutation and simplify the expression to

$$\alpha(e_{i_1}, \ldots, e_{i_k}) \cdot \beta(e_{j_1}, \ldots, e_{j_\ell}) = 1,$$

since both terms in the product equal 1 by (9.3). This establishes the existence of the associative product $\wedge : \Lambda^* V^* \times \Lambda^* V^* \to \Lambda^* V^*$ and the formula (9.4). One still has to show that it also satisfies (9.3), i.e. not just for the basis 1-forms e_*^i but for arbitrary tuples of 1-forms $\alpha^1, \ldots, \alpha^k \in \Lambda^1 V^*$. This can be derived from (9.4) by induction on k and a bit of combinatorics; we leave the details as an exercise.

To prove graded commutativity, it suffices again to consider the case where α and β are both products of 1-forms, and the relation then follows from the case $k = \ell = 1$ which was observed in Example 9.7. The key observation is that the number of flips required for permuting $i_1, \ldots, i_k, j_1, \ldots, j_\ell$ to $j_1, \ldots, j_\ell, i_1, \ldots, i_k$ is $k\ell$. $\qquad\square$

The wedge product turns the vector space $\Lambda^* V^*$ into an algebra; it is called the **exterior algebra** (*äußere Algebra*) over V^*.[36]

---

[36]You may at this point be wondering what the "exterior algebra over V", presumably denoted by $\Lambda^* V$, might be. Since V is finite dimensional, the cheap way to define it is by identifying V with the dual space of V^*, so that homogeneous elements of $\Lambda^* V$ are antisymmetric multilinear maps $V^* \times \ldots \times V^* \to \mathbb{R}$. That is a correct definition, but not the most elegant formulation possible, and it also does not generalize to the case where V is infinite-dimensional since it may then fail to be isomorphic to its double dual. One can define $\Lambda^* V$ in terms of the abstract tensor product of vector spaces, and the details can be found in many standard algebra textbooks, but we will not need them here.

EXERCISE 9.8. Prove that a set of dual vectors $\alpha^1, \ldots, \alpha^k \in V^*$ is linearly independent if and only if its wedge product $\alpha^1 \wedge \ldots \wedge \alpha^k \in \Lambda^k V^*$ is nonzero.

Hint: Consider products of the form $\left(\sum_{i=1}^k c_i \alpha^i \right) \wedge \alpha^2 \wedge \ldots \wedge \alpha^k$.

EXERCISE 9.9. Show that if $\alpha \in \Lambda^k V^*$ and $\beta \in \Lambda^\ell V^*$ are written in terms of the basis $e^1_*, \ldots, e^n_* \in V^*$ as $\alpha = \alpha_{i_1 \ldots i_k} e^{i_1}_* \otimes \ldots \otimes e^{i_k}_*$ and $\beta = \beta_{i_1 \ldots i_\ell} e^{i_1}_* \otimes \ldots \otimes e^{i_\ell}_*$, then $\alpha \wedge \beta = (\alpha \wedge \beta)_{i_1 \ldots i_{k+\ell}} e^{i_1}_* \otimes \ldots \otimes e^{i_{k+\ell}}_*$ where

$$(\alpha \wedge \beta)_{i_1 \ldots i_{k+\ell}} = \frac{(k+\ell)!}{k! \ell!} \alpha_{[i_1 \ldots i_k} \beta_{i_{k+1} \ldots i_{k+\ell}]}.$$

The following formula for top-dimensional forms will have many useful applications:

PROPOSITION 9.10. *Given a basis $e_1, \ldots, e_n \in V$ with dual basis $e^1_*, \ldots, e^n_* \in V^*$, we have*

$$\lambda^1 \wedge \ldots \wedge \lambda^n = \det \begin{pmatrix} \lambda^1(e_1) & \cdots & \lambda^1(e_n) \\ \vdots & \ddots & \vdots \\ \lambda^n(e_1) & \cdots & \lambda^n(e_n) \end{pmatrix} e^1_* \wedge \ldots \wedge e^n_*$$

for any $\lambda^1, \ldots, \lambda^n \in \Lambda^1 V^*$.

PROOF. Use (9.3) to evaluate $(\lambda^1 \wedge \ldots \wedge \lambda^n)(e_1, \ldots, e_n)$, then plug in the formula (9.2) for the determinant. $\square$

EXERCISE 9.11. Find a second proof of Proposition 9.10 using the following idea. Associate to each $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ the 1-form $\mathbf{v}_\flat := v_i e^i_* \in \Lambda^1 V^*$. What can you say about the multilinear function $\omega : \mathbb{R}^n \times \ldots \times \mathbb{R}^n \to \mathbb{R}$ defined by $\omega(\mathbf{v}^1, \ldots, \mathbf{v}^n) := (\mathbf{v}^1_\flat \wedge \ldots \wedge \mathbf{v}^n_\flat)(e_1, \ldots, e_n)$?

9.3. The differential graded algebra of forms.

Everything stated in the previous section implies a statement about differential forms on a manifold M, simply by replacing the vector space V with a tangent space $T_p M$ and then letting $p \in M$ vary. In particular, a k-form $\omega \in \Omega^k(M)$ can now be understood as a function that associates to each $p \in M$ an element

$$\omega_p \in \Lambda^k T_p^* M := \Lambda^k (T_p M)^*.$$

It follows that if $\dim M = n$, then k-forms for $k > n$ are identically 0, hence the direct sum

$$\Omega^*(M) := \bigoplus_{k=0}^{\infty} \Omega^k(M)$$

has only finitely many nontrivial summands. (It is an infinite-dimensional space nonetheless, since each $\Omega^k(M)$ for $k = 0, \ldots, n$ is infinite dimensional.) The wedge product of differential forms is now defined pointwise, i.e. given $\alpha \in \Omega^k(M)$ and $\beta \in \Omega^\ell(M)$, we define $\alpha \wedge \beta \in \Omega^{k+\ell}(M)$ by

$$(\alpha \wedge \beta)_p = \alpha_p \wedge \beta_p \in \Lambda^{k+\ell} T_p^* M.$$

The smoothness of $\alpha \wedge \beta$ by this definition will become clear momentarily when we write it down in local coordinates. Given a chart $(\mathcal{U}, x)$, the natural basis of $T_p M$ to use at points $p \in \mathcal{U}$ is given by the coordinate vector fields $\partial_1, \ldots, \partial_n$, and its dual basis consists of the coordinate differentials $dx^1, \ldots, dx^n$. Any smooth k-form $\omega \in \Omega^k(M)$ can thus be written over $\mathcal{U}$ as

$$(9.6) \quad \begin{aligned} \omega &= \omega_{i_1 \ldots i_k} \, dx^{i_1} \otimes \ldots \otimes dx^{i_k} = \frac{1}{k!} \omega_{i_1 \ldots i_k} \, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \\ &= \sum_{i_1 < \ldots < i_k} \omega_{i_1 \ldots i_k} \, dx^{i_1} \wedge \ldots \wedge dx^{i_k}, \end{aligned}$$

where the Einstein summation convention is in effect for the first line but (in order to eliminate redundancy caused by antisymmetry) not for the second, and the smooth component functions are given by

$$\omega_{i_1 \ldots i_k} = \omega(\partial_{i_1}, \ldots, \partial_{i_k}) \in C^\infty(\mathcal{U}).$$

A coordinate formula for the wedge product can then be extracted from Exercise 9.9, namely

$$(\alpha \wedge \beta)_{i_1 \ldots i_{k+\ell}} = \frac{(k+\ell)!}{k!\ell!} \alpha_{[i_1 \ldots i_k} \beta_{i_{k+1} \ldots i_{k+\ell}]},$$

so assuming that α and β have smooth components, the same is clearly true for $\alpha \wedge \beta$. Theorem 9.4 now carries over to the statement that $\wedge$ defines a bilinear map

$$\Omega^*(M) \times \Omega^*(M) \to \Omega^*(M) : (\alpha, \beta) \mapsto \alpha \wedge \beta$$

that is associative and graded commutative, where the latter again means that for homogeneous elements $\alpha \in \Omega^k(M)$ and $\beta \in \Omega^\ell(M)$, $\alpha \wedge \beta = \pm \beta \wedge \alpha$, with the minus sign appearing if and only if k and ℓ are both odd.

EXAMPLE 9.12. Using Cartesian coordinates (x, y, z) on $\mathbb{R}^3$, the second line of (9.6) says that every $\omega \in \Omega^2(\mathbb{R}^3)$ has a unique presentation in the form

$$\omega = \omega_{xy}\, dx \wedge dy + \omega_{xz}\, dx \wedge dz + \omega_{yz}\, dy \wedge dz,$$

determined by three smooth functions $\omega_{xy}, \omega_{xz}, \omega_{yz} : \mathbb{R}^3 \to \mathbb{R}$.

EXAMPLE 9.13. For $k = n$, the summation in the second line of (9.6) contains only one term. It follows that on an n-manifold M with smooth chart $(\mathcal{U}, x)$, every $\omega \in \Omega^n(M)$ can be written in local coordinates as

$$\omega = f\, dx^1 \wedge \ldots \wedge dx^n \qquad \text{on } \mathcal{U},$$

where the real-valued function $f \in C^\infty(\mathcal{U})$ is given by $f = \omega(\partial_1, \ldots, \partial_n)$.

EXERCISE 9.14. Beginners sometimes fixate on the antisymmetry of the wedge product for 1-forms and thus expect $\omega \wedge \omega = 0$ to hold always, but graded commutativity only implies this when ω has odd degree. Find a concrete example of a 2-form ω on $\mathbb{R}^4$ such that $\omega \wedge \omega \neq 0$.

We can now give a more practically useful characterization of the exterior derivative $d : \Omega^k(M) \to \Omega^{k+1}(M)$, which was defined in §8.2 via C^∞-linearity. A quick word about signs: you've already noticed that in the wedge product, a minus sign gets introduced whenever the order of two elements with odd degree is changed. One can use this same rule to remember the sign in the Leibniz rule below if one thinks of the operator d itself as an object with odd degree; it makes sense in fact to define its degree as 1, since that is the amount by which it raises the degree of any homogeneous element of $\Omega^*(M)$ fed into it.

PROPOSITION 9.15. *The exterior derivative* $d : \Omega^*(M) \to \Omega^*(M)$ *is the unique linear map that satisfies the following conditions:*

(1) *d is* local, *meaning that for every form $\omega \in \Omega^*(M)$ and every $p \in M$, $(d\omega)_p \in \Lambda^* T_p^* M$ depends only on the restriction of ω to a neighborhood of p.*

(2) *For each $f \in \Omega^0(M) = C^\infty(M)$, $df \in \Omega^1(M)$ is the differential of f.*

(3) *For any homogeneous elements $\alpha, \beta \in \Omega^*(M)$, d satisfies the "graded Leibniz rule"*

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^{|\alpha|} \alpha \wedge d\beta.$$

(4) *$d \circ d = 0$.*

COROLLARY 9.16. *For any chart* $(\mathcal{U}, x)$ *and any smooth function* $f : \mathcal{U} \to \mathbb{R}$,

$$(9.7) \quad d\left(f\,dx^{i_1} \wedge \ldots \wedge dx^{i_k}\right) = df \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} = \partial_j f\,dx^j \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} \qquad on\ \mathcal{U}.$$

$$\square$$

PROOF OF PROPOSITION 9.15. Let us start by ignoring the definition of $d : \Omega^*(M) \to \Omega^*(M)$ given in §8.2 and showing that a map satisfying the four properties stated above exists and is unique. The uniqueness follows from the observation that for any chart $(\mathcal{U}, x)$, every k-form on $\mathcal{U}$ is a sum of terms of the form $f\,dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, and if d satisfies properties (2)–(4) then its action on this particular product is given by (9.7). To prove existence, suppose first that $M = \mathcal{U}$ is the domain of a global chart x, in which case the only possible definition of d satisfying the required properties is again via (9.7). It is immediate that d by this definition satisfies properties (1) and (2); let us verify that it also satisfies (3) and (4). To prove the graded Leibniz rule, we observe first that it is true for a pair of 0-forms $f, g \in \Omega^0(M) = C^\infty(M)$, as the product rule from first-year analysis implies

$$d(fg) = df \cdot g + f \cdot dg.$$

For the general case, bilinearity allows us to restrict attention to a pair $\alpha, \beta \in \Omega^*(\mathcal{U})$ of the form $\alpha = f\,dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ and $\beta = g\,dx^{j_1} \wedge \ldots \wedge dx^{j_\ell}$. To make the notation more manageable, let us abbreviate $dx^I := dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ and $dx^J := dx^{j_1} \wedge \ldots \wedge dx^{j_\ell}$; then

$$d(\alpha \wedge \beta) = d\left(fg\,dx^I \wedge dx^J\right) = d(fg) \wedge dx^I \wedge dx^J = (df \cdot g + f \cdot dg) \wedge dx^I \wedge dx^J$$

$$= \left(df \wedge dx^I\right) \wedge \left(g\,dx^J\right) + (-1)^k \left(f\,dx^I\right) \wedge \left(dg \wedge dx^J\right) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta,$$

where the sign $(-1)^k$ arose when we changed the order of $dg \in \Omega^1(\mathcal{U})$ and $dx^I \in \Omega^k(\mathcal{U})$. To prove $d \circ d = 0$, we can similarly consider $\alpha = f\,dx^I$ and compute

$$d(d\alpha) = d(df \wedge dx^I) = d(\partial_j f\,dx^j \wedge dx^I) = d(\partial_j f) \wedge dx^j \wedge dx^I = \partial_k \partial_j f\,dx^k \wedge dx^j \wedge dx^I.$$

This last expression contains implied summations over both k and j, and we observe that while exchanging the roles of k and j leaves $\partial_k \partial_j f$ unchanged, it switches the sign of $dx^k \wedge dx^j$, so that every term in this sum is balanced by a cancelling term, and the sum if therefore 0.

Observe next that while our definition of $d : \Omega^*(\mathcal{U}) \to \Omega^*(\mathcal{U})$ above was expressed in terms of the specific coordinates $x^1, \ldots, x^n$, the fact that it satisfies properties (1)–(4) implies that any other choice of coordinates would have given the same result, as it would also have given a definition satisfying properties (1)–(4). On a general manifold M, one can now define $d : \Omega^*(M) \to \Omega^*(M)$ on small neighborhoods using local coordinates and appeal to the fact that the definition is independent of coordinates, producing a global definition.

It remains only to prove that our definition of d via properties (1)–(4) matches the definition in §8.2. We will prove this by showing that (9.7) implies the same local coordinate formula that was derived in Proposition 8.7. Recall that in a local chart $(\mathcal{U}, x)$, an arbitrary k-form with components $\omega_{i_1 \ldots i_k} = \omega(\partial_{i_1}, \ldots, \partial_{i_k})$ can be written as

$$\omega = \sum_{i_1 < \ldots < i_k} \omega_{i_1 \ldots i_k}\,dx^{i_1} \wedge \ldots \wedge dx^{i_k} = \frac{1}{k!}\omega_{i_1 \ldots i_k}\,dx^{i_1} \wedge \ldots \wedge dx^{i_k},$$

where the summation convention is in effect only in the second expression, in which the combinatorial factor accounts for the fact that each term in the implied summation appears in $k!$ identical copies arising from permutations of the indices $i_1, \ldots, i_k$. The formula (9.7) then implies

$$d\omega = \frac{1}{k!}d\omega_{i_1 \ldots i_k} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} = \frac{1}{k!}\partial_{i_0}\omega_{i_1 \ldots i_k}\,dx^{i_0} \wedge \ldots \wedge dx^{i_k}.$$

In this last sum, nonzero contributions come only from terms in which the numbers $i_0, \ldots, i_k \in \{1, \ldots, n\}$ are all distinct, and if we write S_{k+1} for the group of bijections on $\{0, \ldots, k\}$, each

of these terms can be permuted by some $\sigma \in S_{k+1}$ to produce a product $dx^{i_0} \wedge \ldots \wedge dx^{i_k}$ with $i_0 < \ldots < i_k$, at the cost of applying the inverse permutation to the indices of $\partial_{i_0} \omega_{i_1 \ldots i_k}$ and multiplying by the sign $(-1)^{|\sigma|}$. The expression therefore becomes

$$\frac{1}{k!} \sum_{i_0 < \ldots < i_k} \sum_{\sigma \in S_{k+1}} (-1)^{|\sigma|} \partial_{i_{\sigma(0)}} \omega_{i_{\sigma(1)} \ldots i_{\sigma(k)}} \, dx^{i_0} \wedge \ldots \wedge dx^{i_k}$$

$$= \frac{(k+1)!}{k!} \sum_{i_0 < \ldots < i_k} \partial_{[i_0} \omega_{i_1 \ldots i_k]} \, dx^{i_0} \wedge \ldots \wedge dx^{i_k} = (k+1) \sum_{i_0 < \ldots < i_k} \partial_{[i_0} \omega_{i_1 \ldots i_k]} \, dx^{i_0} \wedge \ldots \wedge dx^{i_k},$$

which matches Proposition 8.7. $\qquad \square$

The wedge product and exterior derivative make $\Omega^*(M)$ into an example of a (commutative) **differential graded algebra** (*graduierte Differentialalgebra*), or "DGA" for short. The inclusion of the word "graded" refers in the first place to the direct sum decomposition $\Omega^*(M) = \bigoplus_{k \geqslant 0} \Omega^k(M)$, but more importantly it refers to the sign appearing in the Leibniz rule of Proposition 9.15. A similar sign prevents $\Omega^*(M)$ from satisfying the commutativity relation $\alpha \wedge \beta = \beta \wedge \alpha$ in general, but the convention is nonetheless to call it a "commutative DGA" if it satisfies the graded commutativity relation $\alpha \wedge \beta = (-1)^{|\alpha||\beta|} \beta \wedge \alpha$.

Recall from §8.3 that pullbacks of differential forms can be defined for arbitrary smooth maps $\varphi : M \to N$, not just diffeomorphisms.

PROPOSITION 9.17. *For any smooth map $\varphi : M \to N$:*

(1) $\varphi^*(\alpha \wedge \beta) = \varphi^* \alpha \wedge \varphi^* \beta$ for all $\alpha, \beta \in \Omega^*(N)$;

(2) $\varphi^*(d\omega) = d(\varphi^* \omega)$ for all $\omega \in \Omega^*(N)$.

PROOF. The first statement follows directly from the definitions. For the second, we start with the case $\omega = f \in C^\infty(N) = \Omega^0(N)$ and use the chain rule: $\varphi^*(df) := df \circ T\varphi = d(f \circ \varphi) =: d(\varphi^* f)$. Since every differential form is locally a finite sum of wedge products of functions and differentials, the graded Leibniz rule then extends this result to all $\omega \in \Omega^k(N)$. $\qquad \square$

10. Oriented manifolds and the integral

10.1. Change of variables. One of the messages of the previous lecture was that on an n-manifold M, one can use differential n-forms to define sensible notions of "n-dimensional volume" and thus measures, from which a notion of integration should emerge. Let's consider first how this might work when M is an open subset $\mathcal{U} \subset \mathbb{R}^n$ in Euclidean space.

There is a canonical choice of coordinates $x^1, \ldots, x^n$ on $\mathcal{U} \subset \mathbb{R}^n$, leading us naturally to consider the n-form $dx^1 \wedge \ldots \wedge dx^n \in \Omega^n(\mathcal{U})$. It has the desirable property that at every point $p \in \mathcal{U}$, if one feeds into it the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\mathbb{R}^n = T_p\mathcal{U}$, the result (by (9.3)) is 1, which happens also to be the Lebesgue measure of the paralelepiped spanned by these vectors, i.e. the n-dimensional unit cube. It follows that if one interprets $dx^1 \wedge \ldots \wedge dx^n$ as a way of computing volumes on tangent spaces $T_p\mathcal{U} = \mathbb{R}^n$, the volume it computes is the *standard* notion of volume, i.e. the Lebesgue measure.

This observation motivates the following definition, which (in light of Example 9.13) tells us how to integrate an arbitrary compactly supported n-form on $\mathcal{U} \subset \mathbb{R}^n$.

DEFINITION 10.1. For any integer $n \geqslant 1$, any compactly supported smooth function $f : \mathcal{U} \to \mathbb{R}$ on an open subset $\mathcal{U} \subset \mathbb{R}^n$ and any Lebesgue-measurable subset $A \subset \mathcal{U}$, the **integral** of the n-form $\omega := f \, dx^1 \wedge \ldots \wedge dx^n$ over A is defined to be the Lebesgue integral of f on A with respect to the standard Lebesgue measure m on $\mathbb{R}^n$, i.e.

$$\int_A \omega = \int_A f \, dx^1 \wedge \ldots \wedge dx^n := \int_A f \, dm \in \mathbb{R}.$$

REMARK 10.2. If you prefer to think in terms of Riemann integrals rather than Lebesgue integrals, you are free to do so in Definition 10.1 at the cost of being slightly more restrictive about the subset $A \subset \mathcal{U}$, e.g. for almost all[37] applications it suffices to imagine that A is an open or closed subset. Nothing in our discussion of integration will depend in any serious way on the distinction between the Riemann and Lebesgue integrals. We will continue to use the language of Lebesgue integration because it seems the most natural.

Analysis conventions sometimes denote the Lebesgue measure on $\mathbb{R}^n$ more suggestively as "$dx^1 \ldots dx^n$", so that Definition 10.1 becomes the easy-to-remember formula

$$\int_A f \, dx^1 \wedge \ldots \wedge dx^n := \int_A f(x^1, \ldots, x^n) \, dx^1 \ldots dx^n.$$

Let's get a bit more ambitious now: suppose M is a more general n-manifold and $\omega \in \Omega^n(M)$ is a compactly supported top-dimensional differential form that happens to have its support contained in the domain $\mathcal{U} \subset M$ of some chart $(\mathcal{U}, x)$. In the corresponding local coordinates, ω can therefore also be written within $\mathcal{U}$ as $f \, dx^1 \wedge \ldots \wedge dx^n$ for a smooth compactly supported function $f : \mathcal{U} \to \mathbb{R}$. Expressing f as a function of the coordinates $x^1, \ldots, x^n$ on $\mathcal{U}$, it now seems natural to define

$$(10.1) \qquad \int_A \omega := \int_{x(A)} f(x^1, \ldots, x^n) \, dx^1 \ldots dx^n$$

for any subset $A \subset \mathcal{U}$ such that $x(A) \subset x(\mathcal{U}) \subset \mathbb{R}^n$ is measurable, i.e. the function whose Lebesgue integral we are actually computing is $f \circ x^{-1} : x(\mathcal{U}) \to \mathbb{R}$. To see why this might be a sensible definition, write the standard Cartesian coordinates on $\mathbb{R}^n$ as $t^1, \ldots, t^n$ so as to distinguish them from the coordinates $x^1, \ldots, x^n$ on $\mathcal{U}$; regarding both sets of coordinates as functions on their respective domains, they are related by

$$(10.2) \qquad t^i \circ x = x^i \quad \text{on } \mathcal{U}, \qquad i = 1, \ldots, n.$$

Definition 10.1 now identifies the Lebesgue integral we just described with the integral of the n-form $(f \circ x^{-1}) \, dt^1 \wedge \ldots \wedge dt^n$ over $x(A) \subset x(\mathcal{U}) \subset \mathbb{R}^n$. According to Proposition 9.17 and (10.2), the diffeomorphism $M \supset \mathcal{U} \xrightarrow{x} x(\mathcal{U}) \subset \mathbb{R}^n$ pulls this n-form back to $\mathcal{U}$ as

$$x^* \left((f \circ x^{-1}) \, dt^1 \wedge \ldots \wedge dt^n \right) = f \cdot \left(x^* dt^1 \wedge \ldots \wedge x^* dt^n \right) = f \cdot \left(d(x^* t^1) \wedge \ldots \wedge d(x^* t^n) \right)$$

$$= f \, dx^1 \wedge \ldots \wedge dx^n = \omega,$$

so (10.1) follows from Definition 10.1 if we stipulate that the integral should satisfy

$$\int_A x^* \alpha = \int_{x(A)} \alpha$$

for all compactly supported n-forms α on $x(\mathcal{U}) \subset \mathbb{R}^n$. This identity is consistent with our intuition about pullbacks via diffeomorphisms: x^* gives a bijection allowing geometric data on $x(\mathcal{U}) \subset \mathbb{R}^n$ to be identified with geometric data on $\mathcal{U} \subset M$, and it would make sense for our definition of the integral to respect such identifications.

But there is still a crucial question to be answered: does our definition of $\int_A \omega$ as described above depend on the choice of chart $x : \mathcal{U} \to \mathbb{R}^n$?

Suppose $y : \mathcal{U} \to \mathbb{R}^n$ is a second chart defined on the same domain, so ω can also be written as $\omega = g \, dy^1 \wedge \ldots \wedge dy^n$ for some function $g : \mathcal{U} \to \mathbb{R}$, and $\int_A \omega$ according to this chart should be $\int_{y(A)} g \circ y^{-1} \, dm$, so we need to know whether this is the same as $\int_{x(A)} f \circ x^{-1} \, dm$. To clarify

---

[37]no pun intended

this, let us abbreviate $\psi := y \circ x^{-1} : x(\mathcal{U}) \to y(\mathcal{U})$ for the transition map relating x and y, and use Proposition 9.10 to write

$$dy^1 \wedge \ldots \wedge dy^n = \det\left(\frac{\partial y}{\partial x}\right) dx^1 \wedge \ldots \wedge dx^n \qquad \text{on } \mathcal{U},$$

where we abbreviate the matrix-valued function

$$\frac{\partial y}{\partial x} := \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y^n}{\partial x^1} & \cdots & \frac{\partial y^n}{\partial x^n} \end{pmatrix} : \mathcal{U} \to \mathbb{R}^{n \times n}.$$

The identity $f \, dx^1 \wedge \ldots \wedge dx^n = \omega = g \, dy^1 \wedge \ldots \wedge dy^n$ thus implies $f = g \cdot \det\left(\frac{\partial y}{\partial x}\right)$. At any point $p \in \mathcal{U}$, $\frac{\partial y}{\partial x}(p)$ is just the Jacobian matrix of the transition map ψ at $x(p)$, and this last identity thus implies

$$f \circ x^{-1} = \left(g \circ x^{-1}\right) \cdot \det D\psi.$$

If we now write $G := g \circ y^{-1}$, then $f \circ x^{-1}$ becomes $(G \circ \psi) \cdot \det D\psi$, and the identity we were hoping for becomes

$$(10.3) \qquad \int_{y(A)} g \circ y^{-1} \, dm = \boxed{\int_{\psi(x(A))} G \, dm \overset{?}{=} \int_{x(A)} (G \circ \psi) \cdot \det D\psi \, dm} = \int_{x(A)} f \circ x^{-1} \, dm.$$

This should look familiar, as it is *almost* the classical change-of-variables formula, except for one detail: in the classical formula, the Jacobian determinant $\det(D\psi)$ is replaced by its absolute value. That is fine if $\det(D\psi)$ happens to be positive—we do of course know that it can never be 0, since ψ is a diffeomorphism and $D\psi(q) : \mathbb{R}^n \to \mathbb{R}^n$ is therefore an isomorphism for all $q \in x(\mathcal{U})$. But nothing in our discussion so far has ruled out the possibility that $\det(D\psi)$ may sometimes be *negative*, and there certainly do exist diffeomorphisms between regions in $\mathbb{R}^n$ that have negative Jacobian determinant, e.g. the reflection $(x, y) \mapsto (x, -y)$ on $\mathbb{R}^2$. The answer to the crucial question about (10.1) is therefore a resounding *sometimes*:

PROPOSITION 10.3. *In the setting of* (10.1), *two charts defined on* $\mathcal{U}$ *give matching definitions of* $\int_A \omega$ *if the Jacobian determinant of their transition map is everywhere positive.* $\square$

10.2. Orientations. The upshot of our change-of-variables discussion is that integrating an n-form $\omega \in \Omega^n(M)$ by writing it in local coordinates as $\omega = f \, dx^1 \wedge \ldots \wedge dx^n$ and then integrating the function f in coordinates does not give a fully coordinate-invariant result, but it will *become* coordinate-invariant if for some reason we never have to worry about transition maps whose Jacobian determinant is negative. This is our first encounter in this course with the notion of *orientation*.

DEFINITION 10.4. Given open subsets $\mathcal{U}, \mathcal{V} \subset \mathbb{R}^n$ for $n \geq 1$, a diffeomorphism $\psi : \mathcal{U} \to \mathcal{V}$ is called **orientation preserving** (*orientierungserhaltend*) if the Jacobian matrix $D\psi(p) \in \mathrm{GL}(n, \mathbb{R})$ at every point $p \in \mathcal{U}$ has positive determinant. It is called **orientation reversing** (*orientierungsumkehrend*) if $\det D\psi(p) < 0$ for all p.

We will say more about the intuitive meaning of this definition in a moment, but for now, you may want to keep the following linear examples in mind:

(1) Every rotation $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ defines an orientation-preserving diffeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$. More generally, every element of the special orthogonal group $\mathrm{SO}(n)$ (cf. Exercise 4.25) defines an orientation-preserving diffeomorphism $\mathbb{R}^n \to \mathbb{R}^n$.

(2) The reflection $(x, y) \mapsto (x, -y)$ is an orientation-reversing diffeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$, and more generally, every element of $\mathrm{O}(n) \backslash \mathrm{SO}(n)$ defines an orientation-reserving diffeomorphism $\mathbb{R}^n \to \mathbb{R}^n$. In particular, this includes every linear transformation on $\mathbb{R}^n$ that is defined by reflecting across an $(n-1)$-dimensional subspace.

DEFINITION 10.5. A smooth atlas $\mathcal{A} = \{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ on a manifold M of dimension $n \geqslant 1$ is called **oriented** (*orientiert*) if all of its transition maps $x_\alpha \circ x_\beta^{-1}$ are orientation preserving. An **orientation** (*Orientierung*) of a manifold M with maximal smooth atlas $\mathcal{A}$ is a subset $\mathcal{A}^+ \subset \mathcal{A}$ that forms a maximal oriented atlas for M. A smooth manifold that has been equipped with an orientation $\mathcal{A}^+$ is called an **oriented manifold** (*orientierte Mannigfaltigkeit*), and the smooth charts in $\mathcal{A}^+$ are then called the **oriented charts**. A manifold is called **orientable** (*orientierbar*) if it admits an orientation.

One can argue as in Lemma 2.5 that given a smooth structure $\mathcal{A}$, every oriented atlas $\mathcal{A}^+ \subset \mathcal{A}$ has a unique extension to a maximal one and thus determines an orientation. In practice, we will see that there are usually more convenient ways to specify an orientation than by explicitly finding an oriented atlas, but here are a few examples where the latter can easily be done:

EXERCISE 10.6. Show that the atlas we defined on S^1 in Lecture 1 is oriented.

EXERCISE 10.7. Use the atlas from Exercise 1.7 to show that S^2 is orientable. (Depending on how you constructed the charts in that exercise, you might now have to modify them slightly for the sake of orientations.)

EXAMPLE 10.8. The manifold $\mathbb{R}^n$ carries a canonical global chart defined by the identity map, so this chart forms an oriented atlas and thus endows $\mathbb{R}^n$ with a canonical orientation.

EXAMPLE 10.9. If M has an oriented atlas $\mathcal{A}^+$ and $\mathcal{O} \subset M$ is an open subset, then the atlas $\mathcal{A}_\mathcal{O}^+$ on $\mathcal{O}$ constructed as in §2.4.2 is automatically also oriented, thus open subsets of oriented manifolds inherit natural orientations. In light of the previous example, this applies in particular to open subsets of $\mathbb{R}^n$.

EXERCISE 10.10. Show that if M and N are both orientable, then so is $M \times N$.

EXERCISE 10.11. Convince yourself that the atlases on the projective plane and Klein bottle described in §2.4.7 are not oriented. (This does not yet prove that these manifolds are not orientable, since one might imagine that there are other ways to construct an oriented atlas. But we will see below that this is impossible.)

DEFINITION 10.12. For two oriented smooth manifolds M and N, a diffeomorphism $f : M \to N$ is called **orientation preserving** or **orientation reversing** if the map $y \circ f \circ x^{-1}$ is orientation preserving / reversing respectively for every choice of oriented smooth charts $(\mathcal{U}, x)$ on M and $(\mathcal{V}, y)$ on N.

EXERCISE 10.13. Show that for the orientations of S^1 and S^2 defined in Exercises 10.6 and 10.7, the antipodal map $S^n \to S^n : p \mapsto -p$ is orientation preserving for $n = 1$ but orientation reversing for $n = 2$.

REMARK 10.14. In light of Definition 10.12 and the canonical orientations of $\mathbb{R}^n$ and open subsets specified by Examples 10.8 and 10.9, a smooth chart $(\mathcal{U}, x)$ on an oriented manifold M is an oriented chart if and only if the diffeomorphism $M \supset \mathcal{U} \xrightarrow{x} x(\mathcal{U}) \subset \mathbb{R}^n$ is orientation preserving.

Let's discuss next some useful alternative perspectives on the notion of orientation. We recall first the basic notion from topology of *connected components*. In topology one distinguishes between two slightly different notions of connectedness, but we will not need to worry about this distinction since for manifolds, they are equivalent.

DEFINITION 10.15. A manifold M is **connected** (*zusammenhängend*) if for every pair of points $p, q \in M$, there exists a continuous path $\gamma : [0, 1] \to M$ with $\gamma(0) = p$ and $\gamma(1) = q$. The **connected components** (*Zusammenhangskomponenten*) of M are the maximal connected subsets.

It should be easy to convince yourself that each connected component of a manifold is both closed and open as a subset, hence it is also a manifold. In fact, if M has connected components $\{M_\alpha\}_{\alpha \in I}$, then there is a natural diffeomorphism $\coprod_{\alpha \in I} M_\alpha \cong M$.

Returning to the subject of orientations, consider a 2-dimensional subspace $P \subset \mathbb{R}^3$, i.e. a plane. One common way of characterizing what it should mean intuitively for P to be "oriented" in one way or the other is to decide which side of P is the "top" and which is the "bottom"; in other words, we draw a distinction between the two connected components of $\mathbb{R}^3 \backslash P$, labelling one component as "above" the plane and the other as "below" it. An equivalent way to say this is that one makes a choice of a unit vector $\mathbf{n} \in \mathbb{R}^3$ orthogonal to P, so that one can then decide to call the direction indicated by $\mathbf{n}$ "above" and the opposite direction "below". There are obviously two possible choices of the vector $\mathbf{n}$, and for an arbitrary plane $P \subset \mathbb{R}^3$, neither choice can be considered canonical.

Now, the case of a plane $P \subset \mathbb{R}^3$ is rather special since it is a submanifold of $\mathbb{R}^3$, and we do not want to have to assume all manifolds we consider are presented to us as submanifolds of Euclidean space. But actually, there is another way to characterize the choice of normal vector $\mathbf{n}$ in terms of vectors that are tangent to P. You may have learned it as the "right hand rule" when you first encountered vectors and the cross product in school: imagine positioning your right hand along the plane $P \subset \mathbb{R}^3$ so that your thumb points orthogonal to it in the direction of $\mathbf{n}$, but your other four fingers are tangent to P. Those four fingers will want to curl in a particular manner, defining a direction of rotation on the plane that one might choose to label "counterclockwise". (This is exactly what one does—at least in the northern hemisphere—when one visualizes the Earth "from above" and says that it rotates counterclockwise. In that situation, "from above" means that one chooses to view the Earth from a vantage point that is centered on the north pole; if one centered the picture on the south pole instead, the rotation would look clockwise! For the same reason, it is important to consistently use the *right* hand rather than the left hand when implementing the right hand rule, as switching hands would indicate a rotation in the other direction.)

The upshot of this heuristic discussion is this: our intuitive notion of what it means to orient a plane $P \subset \mathbb{R}^3$ is equivalent to making a choice of which direction of rotation on P should be labelled as counterclockwise instead of clockwise. This notion can be defined on *any* surface Σ by talking about rotations in the tangent spaces $T_p\Sigma$, and there is no longer any need to discuss normal vectors or assume an embedding $\Sigma \hookrightarrow \mathbb{R}^3$ is given. Moreover, we will see presently that instead of specifying a preferred direction of rotation in $T_p\Sigma$, it is equivalent to specify a preferred class of ordered bases.

DEFINITION 10.16. For a vector space V of dimension $n \geqslant 1$, let

$$\mathcal{B}(V) \subset V^{\times n} := \underbrace{V \times \ldots \times V}_{n}$$

denote the set of all ordered n-tuples $(v_1, \ldots, v_n)$ that form bases of V.

Observe that $\mathcal{B}(V)$ is an open subset of $V^{\times n}$ since linear independence cannot be destroyed by small perturbations. In fact, after choosing any isomorphism $V \to \mathbb{R}^n$, the vectors in any tuple $(v_1, \ldots, v_n) \in \mathcal{B}(V)$ can be put together as columns of an n-by-n matrix, thus identifying $\mathcal{B}(V)$ with the general linear group $\mathrm{GL}(n, \mathbb{R})$, which is indeed an open subset of the space of matrices $\mathbb{R}^{n \times n}$.

Now consider the case $V = \mathbb{R}^2$. Given any $(v_1, v_2) \in \mathcal{B}(\mathbb{R}^2)$, moving from the direction of v_1 to that of v_2 requires a rotation of less than 180 degrees that is either counterclockwise or clockwise;

for example, a counterclockwise rotation is required in order to move from the first standard basis vector $e_1 = (1,0)$ to the second one $e_2 = (0,1)$, but if we exchange their roles and order the standard basis as $(e_2, e_1) \in \mathcal{B}(\mathbb{R}^2)$, then getting from e_2 to e_1 requires a clockwise rotation. For a tangent space $T_p\Sigma$ to a surface Σ, the implication is that if one has chosen which rotations to call counterclockwise as opposed to clockwise, then one has also chosen a preferred class of ordered bases $(X_1, X_2) \in \mathcal{B}(T_p\Sigma)$, i.e. we call (X_1, X_2) a *positively oriented* basis of the rotation moving from X_1 to X_2 is counterclockwise, and *negatively oriented* if that rotation is clockwise. The following facts should now be apparent:

(1) If $(X_1, X_2) \in \mathcal{B}(T_p\Sigma)$ is positively oriented, then every $(X_1', X_2') \in \mathcal{B}(T_p\Sigma)$ that can be connected to (X_1, X_2) by a continuous path in $\mathcal{B}(T_p\Sigma)$ is also positively oriented. Conversely, any two choices of positively oriented basis are related to each other by a continuous deformation of ordered bases, meaning they are connected by a continuous path in $\mathcal{B}(T_p\Sigma)$. Both statements also apply of course to negatively oriented bases.

(2) Any choice of basis $(X_1, X_2) \in \mathcal{B}(T_p\Sigma)$ can be used to *define* the distinction between clockwise and counterclockwise rotation in $T_p\Sigma$: one simply chooses it so that (X_1, X_2) is a positively oriented basis.

(3) An ordered basis (X_1, X_2) is positively oriented if and only if (X_2, X_1) is negatively oriented.

There is a basic fact about $\mathrm{GL}(2, \mathbb{R})$ in the background of the first observation above: it has exactly two connected components, characterized by the conditions $\det(\mathbf{A}) > 0$ and $\det(\mathbf{A}) < 0$. This turns out to be true in every dimension:

PROPOSITION 10.17. *For every $n \in \mathbb{N}$, the sets of* $\mathrm{GL}_+(n, \mathbb{R}) := \big\{ \mathbf{A} \in \mathrm{GL}(n, \mathbb{R}) \;\big|\; \det(\mathbf{A}) > 0 \big\}$ *and* $\mathrm{GL}_-(n, \mathbb{R}) := \big\{ \mathbf{A} \in \mathrm{GL}(n, \mathbb{R}) \;\big|\; \det(\mathbf{A}) < 0 \big\}$ *are both connected.*

PROOF. Since $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$, it suffices to prove that $\mathrm{GL}_+(n, \mathbb{R})$ is connected. To start with, we use polar decomposition to reduce this to a statement about the special orthogonal group $\mathrm{SO}(n)$. Given $\mathbf{A} \in \mathrm{GL}_+(n, \mathbb{R})$, the matrix $\mathbf{A}^T\mathbf{A}$ is symmetric and positive definite, thus it is diagonalizable with only positive eigenvalues, and therefore admits a "square root"

$$\mathbf{P} := \sqrt{\mathbf{A}^T\mathbf{A}},$$

defined in the same orthogonal basis by taking the square roots of the eigenvalues. Clearly $\mathbf{P}$ is also symmetric and positive definite, and it is now straightforward to check that $\mathbf{R} := \mathbf{A}\mathbf{P}^{-1}$ satisfies $\mathbf{R}^T\mathbf{R} = \mathbb{1}$, i.e. it is orthogonal; moreover, $\mathbf{R} \in \mathrm{SO}(n)$ since $\mathbf{A}$ and $\mathbf{P}^{-1}$ each have positive determinant. Now choose a continuous path of symmetric positive-definite matrices $\{\mathbf{P}_t\}_{t \in [0,1]}$ such that $\mathbf{P}_1 = \mathbf{P}$ and $\mathbf{P}_0 = \mathbb{1}$; such a path can be found by fixing the orthonormal eigenbasis of $\mathbf{P}$ while deforming all its (positive!) eigenvalues to 1. The path $\mathbf{A}_t := \mathbf{R}\mathbf{P}_t$ then connects $\mathbf{A}_1 = \mathbf{A}$ to $\mathbf{A}_0 = \mathbf{R} \in \mathrm{SO}(n)$, so we will be done if we can show that $\mathrm{SO}(n)$ is connected.

We argue the latter by induction: the case $n = 1$ is already clear since $\mathrm{SO}(1) = \{1\}$. Assuming $\mathrm{SO}(n-1)$ is already known to be connected, suppose $\mathbf{A} \in \mathrm{SO}(n)$ is given. We claim that there exists a continuous path $\{\mathbf{A}_t \in \mathrm{SO}(n)\}_{t \in [0,1]}$ such that $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_0$ is a matrix of the form

$$\mathbf{A}_0 = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{B} \end{pmatrix}, \qquad \text{for some } \mathbf{B} \in \mathrm{SO}(n-1).$$

Observe that this claim implies the inductive step, as $\mathrm{SO}(n-1)$ is already known to be connected. To prove the claim, first choose any continuous path of unit vectors $v_1(t) \in \mathbb{R}^n$ such that $v_1(1)$ is the first column of $\mathbf{A}$ and $v_1(0)$ is the first standard basis vector $e_1 = (1, 0, \ldots, 0)$; this is possible since the unit sphere S^{n-1} is connected. For any $t_0 \in [0,1]$, one can complete $v_1(t_0)$ to an orthonormal basis $v_1(t_0), \ldots, v_n(t_0) \in \mathbb{R}^n$, and then find a connected neighborhood $J \subset [0,1]$ of t_0

such that the set of vectors $v_1(t), v_2(t_0), \ldots, v_n(t_0)$ remains linearly independent for every $t \in J$. Now define a continuous family of orthonormal bases $v_1(t), v_2(t), \ldots, v_n(t)$ for $t \in J$ by applying the Gram-Schmidt algorithm to $v_1(t), v_2(t_0), \ldots, v_n(t_0)$; regarding these as columns of a matrix, we have in this way constructed a continuous family of orthogonal matrices $\{\widehat{\mathbf{A}}_t \in \mathrm{O}(n)\}_{t \in J}$ whose first columns are $v_1(t)$. Their determinants depend continuously on t and are thus either $+1$ or -1 for all $t \in J$; in the latter case, we can replace $v_n(t)$ by $-v_n(t)$ in order to assume $\widehat{\mathbf{A}}_t \in \mathrm{SO}(n)$ without loss of generality. Since $[0, 1]$ is compact, we can cover it with finitely many neighborhoods J as described above, and in this way construct a family of matrices $\{\widehat{\mathbf{A}}_t \in \mathrm{SO}(n)\}_{t \in [0,1]}$ that satisfy $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_0 = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{B} \end{pmatrix}$, and such that the first column of $\mathbf{A}_t$ depends continuously on t, while the other columns are continuous except at finitely many points $0 < t_1 < \ldots t_N < 1$, where there are jump discontinuities. At any of these points t_j, the two matrices

$$\widehat{\mathbf{A}}_{t_j}^- := \lim_{t \to t_j^-} \widehat{\mathbf{A}}_t, \qquad \widehat{\mathbf{A}}_{t_j}^+ := \lim_{t \to t_j^+} \widehat{\mathbf{A}}_t$$

may differ, but they have the same first column, namely $v_1(t_j)$. But expressing these matrices in any orthonormal basis that starts with $v_1(t_j)$ puts both of them in the form $\begin{pmatrix} 1 & 0 \\ 0 & \mathbf{B}_\pm \end{pmatrix}$ for some $\mathbf{B}_\pm \in \mathrm{SO}(n-1)$, and by the inductive hypothesis, there exists a continuous path in $\mathrm{SO}(n-1)$ from $\mathbf{B}_-$ to $\mathbf{B}_+$. In this way, we can insert extra intervals at each of the points t_j and fill in the discontinuities, then reparametrize the interval to construct the continuous family $\mathbf{A}_t$ in the claim. $\qquad \square$

COROLLARY 10.18. *For any vector space V of dimension $n \geqslant 1$, the set of ordered bases $\mathcal{B}(V)$ has exactly two connected components.* $\qquad \square$

REMARK 10.19. It is very important in this entire discussion that we are talking about *real* vector spaces, not complex. In particular, the analogous set of ordered complex bases on a complex vector space is *connected*, due to the fact that $\mathrm{GL}(n, \mathbb{C})$ is connected. A hint of this is provided by the fact that the determinant on $\mathrm{GL}(n, \mathbb{C})$ takes values in $\mathbb{C}\backslash\{0\}$, which is connected, unlike $\mathbb{R}\backslash\{0\}$. As a consequence, there is no meaningful notion of orientations for *complex* manifolds; actually, every complex manifold can also be regarded as a real manifold and is orientable as a real manifold, but the orientation is *canonically* determined by its complex structure. The reason for the latter is that if we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ via the correspondence $\mathbb{C}^n \ni \mathbf{x} + i\mathbf{y} \leftrightarrow (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$, then every complex-linear isomorphism $\mathbf{A} \in \mathrm{GL}(n, \mathbb{C})$ becomes an element of $\mathrm{GL}(2n, \mathbb{R})$ with *positive* determinant.

EXERCISE 10.20 (just for fun). Adapt the proof of Proposition 10.17 to prove that $\mathrm{GL}(n, \mathbb{C})$ is connected for every $n \in \mathbb{N}$.
Hint: $\mathrm{O}(1)$ is not connected, but $\mathrm{U}(1)$ is.

We can now give a general definition of orientations of vector spaces and relate it to the previously defined notion of oriented manifolds.

DEFINITION 10.21. An **orientation** $\mathfrak{o}_V$ of an n-dimensional vector space V for $n \geqslant 1$ is a labelling of the two connected components of $\mathcal{B}(V)$ as $\mathcal{B}^+(V)$ and $\mathcal{B}^-(V)$, which are then said to consist of the **positively oriented** and **negatively oriented** bases respectively. An **oriented vector space** is a vector space that has been equipped with an orientation. A linear isomorphism $A : V \to W$ between two oriented vector spaces is called **orientation preserving** if for every positively-oriented basis $(v_1, \ldots, v_n)$ of V, $(Av_1, \ldots, Av_n)$ is a positively-oriented basis of W, and A is otherwise called **orientation reversing**.

Notice that unlike manifolds, vector spaces always admit orientations, and there are always exactly two possible choices of orientation.

EXAMPLE 10.22. As a vector space, $\mathbb{R}^n$ carries a canonical orientation for which the standard basis is regarded as positively oriented.

EXERCISE 10.23. Show that for the vector space $\mathbb{R}^n$ with its canonical orientation, an invertible linear map $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^n$ is orientation preserving if and only if $\det(\mathbf{A}) > 0$.
Hint: The identity map $\mathbb{R}^n \to \mathbb{R}^n$ is clearly orientation preserving.

In light of Exercise 10.23, a diffeomorphism $\psi : \mathcal{U} \to \mathcal{V}$ between two open subsets $\mathcal{U}, \mathcal{V} \subset \mathbb{R}^n$ is orientation preserving as in Definition 10.4 if and only if its derivative at every point is an orientation-preserving isomorphism $\mathbb{R}^n \to \mathbb{R}^n$ in the sense of Definition 10.21. We only need one more notion before we can set up a precise correspondence between orientations of manifolds and of their tangent spaces:

DEFINITION 10.24. Suppose M is an n-manifold with $n \geqslant 1$, P is a topological space, $\phi : P \to M$ is a continuous map, and we consider the family of tangent spaces $\{T_{\phi(s)}M\}_{s \in P}$ at points parametrized by the map ϕ. A **continuous family of orientations along** $\phi : P \to M$ is a family $\{\mathfrak{o}_s\}_{s \in P}$, where $\mathfrak{o}_s$ is an orientation of $T_{\phi(s)}M$ for each $s \in P$, such that for every $s_0 \in P$, there exists a neighborhood $\mathcal{O} \subset P$ of s_0 and a collection of continuous maps $X_1, \ldots, X_n : \mathcal{O} \to TM$ for which $(X_1(s), \ldots, X_n(s))$ is a positively-oriented basis of $T_{\phi(s)}M$ with respect to $\mathfrak{o}_s$ for each $s \in \mathcal{O}$. In the case $P = M$ with ϕ chosen to be the identity map, we will simply refer to this as a **continuous family of orientations of the tangent spaces** of M.

PROPOSITION 10.25. *On smooth manifolds M of dimension $n \geqslant 1$, there is a natural bijective correspondence between orientations of M and continuous families of orientations of the tangent spaces of M, and it is uniquely determined by the condition that for any diffeomorphism $f : M \to N$ between two smooth oriented manifolds, f is orientation preserving if and only if the isomorphism $T_p f : T_p M \to T_{f(p)} N$ is orientation preserving for every $p \in M$. Equivalently, a chart $(\mathcal{U}, x)$ is oriented if and only if the corresponding basis of coordinate vector fields $(\partial_1, \ldots, \partial_n)$ is positively oriented for every $p \in \mathcal{U}$.*

PROOF. If M is oriented, one defines the orientation of $T_p M$ for any $p \in M$ such that for any oriented chart $(\mathcal{U}, x)$ with $p \in \mathcal{U}$, the isomorphism $d_p x : T_p M \to \mathbb{R}^n$ is orientation preserving (for the canonical orientation of $\mathbb{R}^n$). This is equivalent to the condition stated above involving coordinate vector fields, and the definition is independent of the choice of oriented chart since if $(\mathcal{V}, y)$ is a different choice, then $d_p y$ is the composition of $d_p x$ with an isomorphism $\mathbb{R}^n \to \mathbb{R}^n$ (defined by differentiating a transition map) that is orientation preserving. Conversely, given a continuous family of orientations of the tangent spaces $T_p M$, one defines the corresponding orientation of M such that a chart $(\mathcal{U}, x)$ is oriented if and only if $d_p x : T_p x \to \mathbb{R}^n$ is orientation preserving for every $p \in \mathcal{U}$. We leave it as an exercise to check that these definitions satisfy all of the stated properties. $\square$

The fact that the orientations of the tangent spaces $T_p M$ vary *continuously* with p is crucial, and it provides the easiest means of proving statements about orientations in many concrete examples.

EXERCISE 10.26. For a smooth n-manifold M with $n \geqslant 1$, prove:
(1) If M is connected and orientable, then it admits exactly two choices of orientation.
(2) M is orientable if and only if for every continuous path $\gamma : [0,1] \to M$ with $\gamma(0) = \gamma(1)$ and every continuous family of orientations $\{\mathfrak{o}_t\}_{t \in [0,1]}$ along γ, $\mathfrak{o}_0 = \mathfrak{o}_1$.

EXERCISE 10.27. Show that S^n is orientable for every $n \in \mathbb{N}$.
Hint: For every $p \in S^n$ and any basis $X_1, \ldots, X_n$ of $T_p S^n$, $(X_1, \ldots, X_n, p)$ forms a basis of $\mathbb{R}^{n+1}$. Use the fact that $\mathbb{R}^{n+1}$ is orientable.

EXERCISE 10.28. Use Exercise 10.26 to show that the projective plane $\mathbb{RP}^2$ and the Klein bottle are not orientable.

EXAMPLE 10.29. The physical universe is a 3-manifold, as you can plainly see by looking around you; from your local perspective it looks like $\mathbb{R}^3$, but since you cannot see the whole thing, it could in theory be diffeomorphic to any 3-manifold, even one that is not orientable. If indeed it is not orientable, then it is possible in theory for an astronaut to return from a long journey through space and find that what she used to call her right hand is now on the left side, and vice versa. She would not see it that way since her right and left eyes would also have been interchanged, but she would think that all writing now appears backwards, and the Earth (when viewed from the north pole) is now rotating clockwise. I am not aware of any law of physics that would rule out this scenario.

10.3. Definition of the integral. We are now in a position to define the integral of a compactly supported n-form on an oriented n-manifold for each $n \geqslant 1$. Denote the **support** (*Träger*) of a k-form $\omega \in \Omega^k(M)$ by

$$\operatorname{supp}(\omega) := \overline{\left\{ p \in M \mid \omega_p \neq 0 \right\}} \subset M,$$

and define the vector space

$$\Omega_c^k(M) := \left\{ \omega \in \Omega^k(M) \mid \operatorname{supp}(\omega) \subset M \text{ is compact} \right\} \subset \Omega^k(M).$$

In the most interesting examples for our purposes, M will often be a compact manifold, in which case $\Omega_c^k(M) = \Omega^k(M)$. We will call a subset $A \subset M$ **measurable** if for every smooth chart $(\mathcal{U}, x)$ on M, the set $x(\mathcal{U} \cap A) \subset \mathbb{R}^n$ is Lebesgue measurable. The following theorem serves simultaneously as a definition.

THEOREM 10.30. *For $n \in \mathbb{N}$, one can uniquely associate to every smooth oriented n-manifold M and measurable subset $A \subset M$ a linear map*

$$\Omega_c^n(M) \to \mathbb{R} : \omega \mapsto \int_A \omega$$

such that the following conditions are satisfied:

(1) *If $\mathcal{U} \subset M$ is an open subset containing $\operatorname{supp}(\omega) \cap A$, then then $\int_{\mathcal{U} \cap A} \omega = \int_A \omega$.*
(2) *For $M = \mathcal{U} \subset \mathbb{R}^n$ an open subset of Euclidean space with its canonical orientation and the standard Cartesian coordinates $x^1, \ldots, x^n$,*

$$\int_A f \, dx^1 \wedge \ldots \wedge dx^n = \int_A f \, dm$$

   for all smooth compactly supported functions $f : \mathcal{U} \to \mathbb{R}$, where the right hand side is the standard Lebesgue integral of f.
(3) *For any orientation-preserving diffeomorphism $\psi : M \to N$ between a pair of oriented n-manifolds,*

$$\int_A \psi^* \omega = \int_{\psi(A)} \omega$$

   holds for all $\omega \in \Omega_c^n(N)$ and measurable subsets $A \subset M$.

To summarize, the integral on arbitrary oriented manifolds is uniquely determined by its definition on open subsets of $\mathbb{R}^n$ and the change-of-variables formula, which now appears as the condition that integrals are invariant under pullbacks via orientation-preserving diffeomorphisms. We will prove this in the next lecture, but it is already easy to explain the idea. For forms $\omega \in \Omega_c^n(M)$ with $\mathrm{supp}(\omega)$ contained in the domain of a single oriented chart $(\mathcal{U}, x)$, one can write

$$\omega = f\, dx^1 \wedge \ldots \wedge dx^n = x^* \left((f \circ x^{-1})\, dt^1 \wedge \ldots \wedge dt^n \right) \qquad \text{on } \mathcal{U}$$

in terms of the standard Cartesian coordinates $t^1, \ldots, t^n$ on $x(\mathcal{U}) \subset \mathbb{R}^n$ and a uniquely determined function $f : \mathcal{U} \to \mathbb{R}$. The three properties in the statement above then reproduce the definition of $\int_A \omega$ that we saw in §10.1, namely

$$\int_A \omega = \int_{\mathcal{U} \cap A} \omega = \int_{\mathcal{U} \cap A} x^* ((f \circ x^{-1})\, dt^1 \wedge \ldots \wedge dt^n) = \int_{x(\mathcal{U} \cap A)} (f \circ x^{-1})\, dt^1 \wedge \ldots \wedge dt^n$$

$$= \int_{x(\mathcal{U} \cap A)} f \circ x^{-1}\, dm.$$

The restriction to oriented charts guarantees moreover in light of Proposition 10.3 that this result does not depend on the choice of the chart $(\mathcal{U}, x)$, though it does depend on the orientation. Linearity will then determine $\int_A \omega$ uniquely for every $\omega \in \Omega_c^n(M)$ if we can be assured that every such form is a finite sum of forms that each have compact support in the domain of some oriented chart. This is true, but not completely obvious—it will require a brief digression on the topic of *partitions of unity*, which will have many further uses as we move forward.

11. Integration and volume

11.1. Existence of the integral. I owe you a proof of Theorem 10.30 on the existence and properties of the linear map $\Omega_c^n \to \mathbb{R} : \omega \mapsto \int_A \omega$ for all oriented n-manifolds M and measurable subsets $A \subset M$. The following will serve as a useful tool for "localizing" such constructions:

LEMMA 11.1. *Given a smooth manifold M, a compact subset $K \subset M$ and a finite collection of open sets $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ that cover K, there exists a collection of smooth functions $\{\varphi_\alpha : M \to [0,1]\}_{\alpha \in I}$ satisfying the following two conditions:*

(1) For each $\alpha \in I$, φ_α has compact support contained in $\mathcal{U}_\alpha$;
(2) $\sum_{\alpha \in I} \varphi_\alpha \equiv 1$ on K.

PROOF. For each $p \in K$, choose any $\alpha_p \in I$ such that $p \in \mathcal{U}_{\alpha_p}$, and choose also a smooth function $\psi_p : M \to [0,1]$ with compact support in $\mathcal{U}_{\alpha_p}$ such that $\psi_p > 0$ on some open neighborhood $\mathcal{V}_p \subset \mathcal{U}_{\alpha_p}$ of p. The sets $\{\mathcal{V}_p\}_{p \in K}$ then form an open cover of the compact set K and therefore admit a finite subcover, i.e. there is a finite subset $K_0 \subset K$ such that $K \subset \bigcup_{p \in K_0} \mathcal{V}_p$. Now for each $\alpha \in I$, define a smooth function $\psi_\alpha : M \to [0, \infty)$ by

$$\psi_\alpha := \sum_{\{p \in K_0 \;|\; \alpha_p = \alpha\}} \psi_p.$$

By construction, ψ_α has compact support in $\mathcal{U}_\alpha$, and for each $q \in K$, there exists $p \in K_0$ such that $q \in \mathcal{V}_p$ and thus $\psi_p(q) > 0$, implying $\psi_{\alpha_p}(q) > 0$. It follows that $\sum_{\alpha \in I} \psi_\alpha > 0$ everywhere on K, and therefore also on some neighborhood $\mathcal{V} \subset M$ of K. On the neighborhood $\mathcal{V}$, we define

$$\varphi_\alpha := \frac{\psi_\alpha}{\sum_{\beta \in I} \psi_\beta}, \qquad \text{for each } \alpha \in I,$$

so that each φ_α takes values in $[0,1]$ and $\sum_{\alpha \in I} \varphi_\alpha \equiv 1$ by construction. Now choose any smooth function $f : M \to [0,1]$ that equals 1 on K and has compact support in $\mathcal{V}$, modify each φ_α by

multiplying it by f, and extend the modified function to the rest of M so that it vanishes outside of $\mathcal{V}$. $\qquad\square$

The collection of functions $\{\varphi_\alpha\}_{\alpha\in I}$ in this lemma is a special case of a general construction called a **partition of unity** subordinate to the cover $\{\mathcal{U}_\alpha\}_{\alpha\in I}$ (*eine der Überdeckung untergeordnete Zerlegung der Eins*). We will extend this notion later, when we discuss more general existence theorems for geometric structures such as Riemannian metrics.

PROOF OF THEOREM 10.30. Given an oriented n-manifold M with measurable subset $A \subset M$ and $\omega \in \Omega_c^n(M)$, choose an open subset $M_0 \subset M$ that contains $\mathrm{supp}(\omega) \cap A$ but has compact closure $\overline{M}_0 \subset M$. By compactness, we can cover $\overline{M}_0$ with a finite collection of open sets $\{\mathcal{U}_\alpha \subset M\}_{\alpha\in I}$ that are domains of oriented charts $(\mathcal{U}_\alpha, x_\alpha)$, and Lemma 11.1 provides a partition of unity $\{\varphi_\alpha : M \to [0,1]\}_{\alpha\in I}$ such that

(i) φ_α has compact support contained in $\mathcal{U}_\alpha$ for each $\alpha \in I$;
(ii) $\sum_{\alpha\in I} \varphi_\alpha \equiv 1$ on M_0.

We can now write
$$\omega = \sum_{\alpha\in I} \varphi_\alpha \omega \qquad \text{on } M_0,$$

and observe that $\varphi_\alpha \omega \in \Omega_c^n(\mathcal{U}_\alpha)$, so if the integral satisfies the properties stated in the theorem, then

$$(11.1) \qquad \int_A \omega = \int_{M_0 \cap A} \omega = \sum_{\alpha\in I} \int_{M_0 \cap A} \varphi_\alpha \omega = \sum_{\alpha\in I} \int_{\mathcal{U}_\alpha \cap A} \varphi_\alpha \omega = \sum_{\alpha\in I} \int_{x_\alpha(\mathcal{U}_\alpha \cap A)} f_\alpha \, dm,$$

where $f_\alpha : x_\alpha(\mathcal{U}_\alpha) \to \mathbb{R}$ is the unique function such that $\varphi_\alpha \omega = x_\alpha^*(f_\alpha \, dx^1 \wedge \ldots \wedge dx^n)$ on $\mathcal{U}_\alpha$. This specifies the integral uniquely.

We claim next that if $\int_A \omega \in \mathbb{R}$ is defined via (11.1), then the result is independent of all choices, namely the open subset $M_0 \subset M$ containing $\mathrm{supp}(\omega) \cap A$, the finite collection of oriented charts $\{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha\in I}$ and the functions $\{\varphi_\alpha\}_{\alpha\in I}$ satisfying (i) and (ii) above. Independence of the choice of charts follows from the discussion in §10.1, in particular Proposition 10.3. This is the step at which it is crucial that M comes with an orientation, so the transition maps that we feed into Proposition 10.3 are all orientation preserving. With this out of the way, suppose $\{(\mathcal{V}_\beta, y_\beta)\}_{\beta\in J}$ is another finite collection of oriented charts and $\{\psi_\beta : M \to [0,1]\}_{\beta\in J}$ a collection of smooth functions that each have compact support in the corresponding subsets $\mathcal{V}_\beta$ and satisfy $\sum_{\beta\in J} \psi_\beta \equiv 1$ on some open set $M_1 \subset M$ containing $\mathrm{supp}(\omega) \cap A$. The open set $M_0 \cap M_1 \subset M$ then also contains $\mathrm{supp}(\omega) \cap A$, and is covered by the finite collection of open sets

$$\{\mathcal{U}_\alpha \cap \mathcal{V}_\beta\}_{(\alpha,\beta)\in I\times J},$$

with the functions $\{\varphi_\alpha \psi_\beta : M \to [0,1]\}_{(\alpha,\beta)\in I\times J}$ having compact support in $\mathcal{U}_\alpha \cap \mathcal{V}_\beta$ and satisfying $\sum_{(\alpha,\beta)\in I\times J} \varphi_\alpha \psi_\beta \equiv 1$ on $M_0 \cap M_1$. Any oriented chart x_α defined on $\mathcal{U}_\alpha$ is also defined on $\mathcal{U}_\alpha \cap \mathcal{V}_\beta$ for each $\beta \in J$, so we can use it to compute $\int_{\mathcal{U}_\alpha \cap \mathcal{V}_\beta \cap A} \varphi_\alpha \psi_\beta \omega$ as a Lebesgue integral over $x_\alpha(\mathcal{U}_\alpha \cap A) \subset \mathbb{R}^n$ of a function with compact support in the region $x_\alpha(\mathcal{U}_\alpha \cap \mathcal{V}_\beta)$, and the additivity of the Legesgue integral then implies

$$\int_{\mathcal{U}_\alpha \cap A} \varphi_\alpha \omega = \sum_{\beta\in J} \int_{\mathcal{U}_\alpha \cap \mathcal{V}_\beta \cap A} \varphi_\alpha \psi_\beta \omega,$$

and therefore also

$$\sum_{\alpha\in I} \int_{\mathcal{U}_\alpha \cap A} \varphi_\alpha \omega = \sum_{(\alpha,\beta)\in I\times J} \int_{\mathcal{U}_\alpha \cap \mathcal{V}_\beta \cap A} \varphi_\alpha \psi_\beta \omega.$$

But if we carry out the same argument instead with the charts $(\mathcal{V}_\beta, y_\beta)$ and write $\psi_\beta \omega = \sum_{\alpha \in I} \varphi_\alpha \psi_\beta \omega$, we find that the right hand side is also equal to $\sum_{\beta \in J} \int_{\mathcal{V}_\beta \cap A} \psi_\beta \omega$, proving that the two definitions of $\int_A \omega$ obtained from these different partitions of unity match.

It remains to check that our general definition of $\int_A \omega$ satisfies the three properties stated in the theorem, but this is easy, so we will leave it as an exercise with the following hint: the freedom to choose any convenient collection of oriented charts makes the formula $\int_A \psi^* \omega = \int_{\psi(A)} \omega$ for orientation-preserving diffeomorphisms $\psi : M \to N$ virtually a tautology. □

11.2. Computational tools. The notion of integration defined in Theorem 10.30 has several useful properties that were not mentioned yet, some of which can be applied to make actual calculations considerably easier, e.g. it is rarely actually necessary in practice to choose a partition of unity. We begin with two properties whose proofs are easy exercises.

EXERCISE 11.2. Prove that for an oriented n-manifold M and $\omega \in \Omega_c^n(M)$, the following properties hold:
(1) If $A, B \subset M$ are two disjoint measurable subsets, then $\int_{A \cup B} \omega = \int_A \omega + \int_B \omega$.
(2) If $A \subset M$ has the property that $x(\mathcal{U} \cap A) \subset \mathbb{R}^n$ has Lebesgue measure zero[38] for all smooth charts $(\mathcal{U}, x)$, then $\int_A \omega = 0$.

One frequently occurring situation in simple examples is that the domain $A \subset M$ where we want to integrate lies almost entirely inside the domain of a single chart, where the word "almost" in this case carries its usual measure-theoretic meaning, i.e. "outside of a set of measure zero". In combination with the exercise above, the next result will then allow us to dispense entirely with partitions of unity and compute the integral in a single chart:

PROPOSITION 11.3. *Suppose M is an oriented n-manifold and $(\mathcal{U}, x)$ is an oriented chart on M. Then for any measurable subset $A \subset \mathcal{U}$ and $\omega \in \Omega_c^n(M)$ taking the form $f \, dx^1 \wedge \ldots \wedge dx^n$ in $\mathcal{U}$, the function $f \circ x^{-1}$ is Lebesgue integrable on $x(A) \subset \mathbb{R}^n$ and*

$$\int_A \omega = \int_{x(A)} f \circ x^{-1} \, dm.$$

PROOF. Let $K \subset M$ denote the closure of $\mathrm{supp}(\omega) \cap A \subset M$, and observe that this set is compact since it is a closed subset of $\mathrm{supp}(\omega)$, and it is also contained in the closure of $\mathcal{U}$ since $A \subset \mathcal{U}$. In particular, the set

$$\partial K := K \cap (M \backslash \mathcal{U})$$

is contained in the boundary of the closure of $\mathcal{U}$, and by assumption it is disjoint from A. Next choose a finite collection of oriented charts $\{(\mathcal{O}_\alpha, x_\alpha)\}_{\alpha \in I}$ such that

$$K \subset \mathcal{U} \cup \bigcup_{\alpha \in I} \mathcal{O}_\alpha,$$

and for each $N \in \mathbb{N}$ and $\alpha \in I$, let

$$\mathcal{O}_\alpha^N := \left\{ p \in \mathcal{O} \mid |x_\alpha(p) - x_\alpha(q)| < 1/N \text{ for some } q \in \partial K \cap \mathcal{O}_\alpha \right\}.$$

We observe the following:
(1) $K \subset \mathcal{U} \cup \bigcup_{\alpha \in I} \mathcal{O}_\alpha^N$ for every $N \in \mathbb{N}$.
(2) For each $\alpha \in I$, $\mathcal{O}_\alpha^1 \supset \mathcal{O}_\alpha^2 \supset \mathcal{O}_\alpha^3 \supset \ldots$, and, since $A \cap \partial K = \varnothing$,

$$(11.2) \qquad\qquad A \cap \bigcap_{N \in \mathbb{N}} \mathcal{O}_\alpha^N = \varnothing.$$

---

[38]We say in this case that $A \subset M$ has **measure zero**. Note that it is not actually necessary to define a measure on M in order to define this notion.

For each $N \in \mathbb{N}$, we can choose a partition of unity consisting of functions $\varphi^N, \varphi_\alpha^N : M \to [0, 1]$ for each $\alpha \in I$ with compact supports $\operatorname{supp}(\varphi^N) \subset \mathcal{U}$ and $\operatorname{supp}(\varphi_\alpha^N) \subset \mathcal{O}_\alpha^N$ such that $\varphi^N + \sum_{\alpha \in I} \varphi_\alpha^N \equiv 1$ on K. Since K contains $A \cap \operatorname{supp}(\omega)$, we then have

$$\int_A \omega = \int_A \varphi^N \omega + \sum_{\alpha \in I} \varphi_\alpha^N \omega$$

for every $N \in \mathbb{N}$. But for each $\alpha \in I$, (11.2) implies that the Lebesgue measure of $x_\alpha(\mathcal{O}_\alpha^N \cap A)$ converges to 0 as $N \to \infty$, thus

$$\lim_{N \to \infty} \int_A \varphi_\alpha^N \omega = 0,$$

from which follows

$$\int_A \varphi^N \omega \to \int_A \omega \qquad \text{as } N \to \infty.$$

Writing $\omega = x^*(f \, dx^1 \wedge \ldots \wedge dx^n)$ on $\mathcal{U}$ for a suitable function $f : x(\mathcal{U}) \to \mathbb{R}$, $\int_A \varphi^N \omega$ becomes the Lebesgue integral

$$\int_{x(A)} (\varphi^N \circ x^{-1}) f \, dm,$$

in which the integrand converges pointwise to f since each point in A is outside the support of all the φ_α^N for N sufficiently large. If you already believe that f is Lebesgue integrable on $x(A)$, then since $|(\varphi^N \circ x^{-1}) f| \leq |f|$, the dominated convergence theorem now implies that this integral converges to $\int_{x(A)} f \, dm$ as $N \to \infty$, and the latter is therefore $\int_A \omega$.

Here is a quick sketch of the proof that f really is Lebesgue integrable on $x(A)$: suppose ω is replaced by a *continuous* n-form $|\omega|$ on M that equals $-\omega$ at any point where ω evaluates negatively on some positive basis, but is otherwise identical to ω. In general $|\omega|$ will not be smooth—just as $|f|$ need not be smooth when f is a smooth function—but continuity is good enough for defining the integral $\int_A |\omega|$ as in Theorem 10.30. Changing ω to $|\omega|$ has the effect of replacing f with $|f|$ in the calculation above, and similarly in all other oriented charts. The same argument as above then proves

$$\int_{x(A)} (\varphi^N \circ x^{-1}) |f| \, dm \to \int_A |\omega| \qquad \text{as } N \to \infty.$$

Since φ^N equals 1 on subsets that exhaust all of A as $N \to \infty$, this implies a uniform upper bound for the integral of $|f|$ over arbitrary compact subsets of $x(A)$, and thus $\int_{x(A)} |f| \, dm < \infty$. □

EXERCISE 11.4. For every oriented n-manifold M with $n \geq 1$, there exists another oriented manifold $-M$ that is defined as the same manifold with the "reversed" orientation, meaning that one changes the orientation of every tangent space $T_p M$. Show that for every $\omega \in \Omega_c^n(M)$,

$$\int_{-M} \omega = -\int_M \omega.$$

Hint: If you fix the reflection map $r(t^1, t^2, \ldots, t^n) := (-t^1, t^2, \ldots, t^n)$ on $\mathbb{R}^n$ and take any oriented chart $(\mathcal{U}, x)$ on M, then $(\mathcal{U}, r \circ x)$ will be an oriented chart on $-M$.

REMARK 11.5. At long last, we can now clarify a notational issue that often bothers newcomers to integral calculus: what does $\int_b^a f(x) \, dx$ actually mean when $a < b$? It is traditional to define this as a synonym for $-\int_a^b f(x) \, dx := -\int_{[a,b]} f \, dm$ and regard it as a meaningless but useful convention, but now we can assign a deeper meaning to it: for the 1-manifold $M := (a, b) \subset \mathbb{R}$ with its canonical

orientation and the 1-form $f\,dx \in \Omega^1_c(M)$ defined via the canonical coordinate x and a compactly supported[39] function $f : (a,b) \to \mathbb{R}$, the correct definition is

$$\int_b^a f(x)\,dx := \int_{-(a,b)} f\,dx,$$

where $-(a,b)$, denotes the manifold (a,b) with the opposite of its canonical orientation. This is consistent with the way that substitution is typically applied in calculations of 1-dimensional integrals: orientation-reversing diffeomorphisms are sometimes used for substitution, but they produce integrals over intervals with reversed orientation.

11.3. Volume forms. We now consider the first true geometric application of integration: how does one compute volumes of subsets in a manifold?

In an ordinary measure space X with measure μ, the volume of $A \subset X$ is simply $\int_A d\mu$. We have seen that in n-dimensional oriented manifolds, the role of measures is played by differential n-forms; however, not all of these define geometrically appropriate notions of volume. Indeed, a form $\omega \in \Omega^n(M)$ gives a way to define volumes of paralelepipeds in each tangent space T_pM, but it can happen that $\omega_p = 0$ at some point $p \in M$, implying that all regions in T_pM have volume zero, which is not very reasonable geometrically. The objects that we will refer to as "volume forms" specifically exclude this possibility:

DEFINITION 11.6. A **volume form** (*Volumenform*) on an n-manifold M is an n-form $\omega \in \Omega^n(M)$ such that $\omega_p \neq 0$ for all $p \in M$.

NOTATION. In these notes, we will usually denote volume forms by

$$d\mathrm{vol} \in \Omega^n(M),$$

or sometimes $d\mathrm{vol}_M$ if there are various manifolds in the picture and we want to specify which one $d\mathrm{vol}$ is defined on. The notation is slightly misleading since in many cases, our volume form will not actually be the exterior derivative of anything; nonetheless, the presence of the symbol "d" is consistent with the way that measures are often written in integrals, and that is the role that we intend for $d\mathrm{vol}$ to play.

Observe that since $\dim \Lambda^n T_p^* M = 1$ for every $p \in M$, $d\mathrm{vol} := \omega \in \Omega^n(M)$ is a volume form if and only if ω_p is a basis of $\Lambda^n T_p^* M$ for every p, and it follows in this case that any other n-form $\alpha \in \Omega^n(M)$ can be written as

$$\alpha = f\,d\mathrm{vol}$$

for a unique function $f \in C^\infty(M)$. In this situation, α is also a volume form if and only if the function f is nowhere zero.

PROPOSITION 11.7. *Any volume form* $d\mathrm{vol} \in \Omega^n(M)$ *on a manifold* M *determines a unique orientation of* M *such that for each* $p \in M$, *an ordered basis* $(X_1, \ldots, X_n) \in T_pM$ *is positively oriented if and only if* $d\mathrm{vol}(X_1, \ldots, X_n) > 0$.

PROOF. Assuming $d\mathrm{vol}_p \neq 0$, Proposition 9.2 implies that $d\mathrm{vol}(X_1, \ldots, X_n) \neq 0$ for every basis $X_1, \ldots, X_n$ of T_pM. It follows that $d\mathrm{vol}$ determines a continuous map $\mathcal{B}(T_pM) \to \mathbb{R} :$ $(X_1, \ldots, X_n) \mapsto d\mathrm{vol}(X_1, \ldots, X_n)$ that is never zero, and since it clearly can take values of both signs, it must take positive values on one connected component of $\mathcal{B}(T_pM)$ and negative values on the other. Since its values also vary continuously with the point p, this distinction between the signs of $d\mathrm{vol}(X_1, \ldots, X_n)$ determines a continuous family of orientations of the tangent spaces T_pM. $\square$

---

[39]We are assuming compact support in (a,b) here because we have not yet defined manifolds *with boundary*, and thus cannot define an integral over the *closed* interval $[a,b]$. This will come in the next lecture, however.

If M is equipped with the orientation determined by a volume form $d$vol via Proposition 11.7, then it is common to write this condition as

$$d\text{vol} > 0,$$

meaning literally that $d\text{vol}(X_1, \ldots, X_n) > 0$ for every $p \in M$ and every *positively-oriented* basis $(X_1, \ldots, X_n)$ of T_pM, and $d$vol is in this case called a **positive volume form** on the oriented manifold M. Another n-form $\alpha = f\,d$vol is then also a positive volume form if and only if $f > 0$ everywhere. In particular, for any oriented chart $(\mathcal{U}, x)$, $dx^1 \wedge \ldots \wedge dx^n$ is a positive volume form on $\mathcal{U}$ since $(dx^1 \wedge \ldots \wedge dx^n)(\partial_1, \ldots, \partial_n) = 1$, thus a positive volume form $d\text{vol} \in \Omega^n(M)$ always locally takes the form

$$(11.3) \qquad\qquad d\text{vol} = f\,dx^1 \wedge \ldots \wedge dx^n, \qquad f : \mathcal{U} \to (0, \infty).$$

If $(M, d\text{vol})$ is an oriented manifold equipped with a positive volume form, the volume of a measurable subset $A \subset M$ is now defined simply as

$$\text{Vol}(A) := \int_A d\text{vol},$$

which is always nonnegative due to (11.3).

The definition of volume in M clearly depends on a choice of volume form, and for arbitrary manifolds there is generally no canonical choice—this reflects the fact that volumes of regions can appear very different when viewed in different coordinate systems. However, there are situations in which extra data determines a natural choice of volume form.

Suppose for instance that $M \subset \mathbb{R}^n$ is a k-dimensional submanifold of Euclidean space. Each tangent space T_pM is then a k-dimensional linear subspace of $T_p\mathbb{R}^n = \mathbb{R}^n$, and can thus be assigned the standard Euclidean inner product $\langle\ ,\ \rangle$, which we can then use to define lengths of vectors in T_pM and angles between them. In particular, this defines the notion of an *orthonormal* basis of T_pM. The paralelepiped spanned by an orthonormal basis of a k-dimensional subspace in $\mathbb{R}^n$ has the same dimensions as the k-dimensional unit cube, so its k-dimensional volume is 1, and it would therefore be natural to choose a volume form $d\text{vol} \in \Omega^k(M)$ that evaluates to 1 on some orthonormal basis.

To bring this discussion into its most natural setting, recall that a **Riemannian metric** (*Riemannsche Metrik*) on a manifold M is a smooth type $(0, 2)$ tensor field $g \in \Gamma(T_2^0 M)$ such that $g_p : T_pM \times T_pM \to \mathbb{R}$ defines an inner product on T_pM for every $p \in M$. The pair (M, g) is in this case called a **Riemannian manifold** (*Riemannsche Mannigfaltigkeit*). The data of a Riemannian metric makes it possible to define norms of tangent vectors and angles between them, so in particular, every tangent space T_pM acquires a well-defined notion of orthonormality.

DEFINITION 11.8. On a Riemannian manifold (M, g), a volume form $d\text{vol} \in \Omega^n(M)$ is said to be **compatible** with the metric g if for every $p \in M$ and every orthonormal basis $X_1, \ldots, X_n \in T_pM$, $|d\text{vol}(X_1, \ldots, X_n)| = 1$.

Since $\dim \Lambda^n T_p^* M = 1$ for an n-manifold M, there are clearly at most two volume forms compatible with a given metric g at any given point $p \in M$. The following algebraic lemma guarantees that there are, in fact, exactly two, corresponding to the two possible orientations of T_pM.

LEMMA 11.9. *Suppose V is an n-dimensional oriented vector space equipped with an inner product $\langle\ ,\ \rangle$, $v_1, \ldots, v_n \in V$ is a positively-oriented orthonormal basis and $v_*^1, \ldots, v_*^n \in V^*$ denotes its dual basis. Then the top-dimensional form*

$$\omega := v_*^1 \wedge \ldots \wedge v_*^n \in \Lambda^n V^*$$

satisfies $\omega(w_1, \ldots, w_n) = 1$ for every positively-oriented orthonormal basis $w_1, \ldots, w_n \in V$.

PROOF. By (9.3), it will suffice to establish that if $w_*^1, \ldots, w_*^n \in V^*$ is the dual basis of another positively-oriented orthonormal basis $w_1, \ldots, w_n \in V$, then

$$v_*^1 \wedge \ldots \wedge v_*^n = w_*^1 \wedge \ldots \wedge w_*^n.$$

By Proposition 9.10, the scaling factor relating these two n-forms is the determinant of the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with entries $A^i{}_j := w_*^i(v_j)$. Writing v_k as a linear combination of the w_i gives $v_k = w_*^i(v_k)w_i$, and orthonormality then implies

$$\delta_{k\ell} = \langle v_k, v_\ell \rangle = \langle w_*^i(v_k)w_i, w_*^j(v_\ell)w_j \rangle = w_*^i(v_k)w_*^j(v_\ell)\langle w_i, w_j \rangle = w_*^i(v_k)w_*^j(v_\ell)\delta_{ij}$$

$$= \sum_{i=1}^n w_*^i(v_k)w_*^i(v_\ell) = \sum_{i=1}^n A^i{}_k A^i{}_\ell,$$

where in the second line we can no longer use the summation convention since the index to be summed does not appear in an upper-lower pair. This calculation implies that the rows of $\mathbf{A}$ form an orthonormal set, meaning $\mathbf{A} \in \mathrm{O}(n)$ and thus $\det(\mathbf{A}) = \pm 1$. Since both bases are also positively oriented, there exists a continuous family of orthonormal bases connecting one to the other, implying that there is also a continuous family of orthogonal matrices connecting $\mathbf{A}$ to $\mathbb{1}$, thus $\det(\mathbf{A}) = 1$. $\qquad\square$

COROLLARY 11.10. *Every oriented Riemannian n-manifold (M,g) admits a unique so-called* **Riemannian volume form** *$d\mathrm{vol} \in \Omega^n(M)$ that is positive and compatible with g.*

PROOF. The existence and uniqueness of $d\mathrm{vol}_p \in \Lambda^n T_p^* M$ for each $p \in M$ follows from Lemma 11.9, so it remains only to check that the n-form defined in this way is smooth. To see this, note that for any $p \in M$, one can find a neighborhood $\mathcal{U} \subset M$ of p and smooth vector fields $X_1, \ldots, X_n \in \mathfrak{X}(\mathcal{U})$ that form a positively-oriented orthonormal basis at every point in $\mathcal{U}$; simply start e.g. with a basis of coordinate vector fields near p and then use the Gram-Schmidt process to make them orthonormal at each point. Now if $\lambda^1, \ldots, \lambda^n \in \Omega^1(\mathcal{U})$ are defined so that $\lambda_q^1, \ldots, \lambda_q^n \in T_q^* M$ is the dual basis to $X_1(q), \ldots, X_n(q) \in T_q M$ for every $q \in \mathcal{U}$, then $\lambda^1 \wedge \ldots \wedge \lambda^n$ is a smooth n-form on $\mathcal{U}$ that matches $d\mathrm{vol}$ according to Lemma 11.9. $\qquad\square$

EXAMPLE 11.11. On $\mathbb{R}^n$, there is a standard choice of Riemannian metric defined by assigning to each $T_p\mathbb{R}^n = \mathbb{R}^n$ the Euclidean inner product. This makes the standard coordinate vector fields $\partial_1, \ldots, \partial_n$ into a positively-oriented orthonormal basis at every point, and the unique positive volume form compatible with the standard metric is thus the so-called **standard volume form** $dx^1 \wedge \ldots \wedge dx^n$. The notion of volume defined by integrating it is of course just the Lebesgue measure.

EXERCISE 11.12. In local coordinates with respect to an oriented n-dimensional chart $(\mathcal{U}, x)$, a Riemannian metric $g \in \Gamma(T_2^0 M)$ is described in terms of its components $g_{ij} := g(\partial_i, \partial_j)$, so that vectors $X, Y \in T_p M$ at points $p \in \mathcal{U}$ satisfy $g(X,Y) = g_{ij}X^i Y^j$. The goal of this exercise is to prove that the Riemannian volume form is then given by

$$(11.4) \qquad\qquad d\mathrm{vol} = \sqrt{\det \mathbf{g}}\, dx^1 \wedge \ldots \wedge dx^n \qquad \text{on } \mathcal{U},$$

where $\mathbf{g} : \mathcal{U} \to \mathbb{R}^{n \times n}$ denotes the matrix-valued function whose ith row and jth column is g_{ij}. Note that this matrix necessarily has positive determinant since g is positive definite. Fix a point $p \in \mathcal{U}$ and a positively-oriented orthonormal basis $(X_1, \ldots, X_n)$ of $T_p M$, whose dual basis we will denote by $\lambda^1, \ldots, \lambda^n \in T_p^* M$. According to Lemma 11.9, $d\mathrm{vol}_p = \lambda^1 \wedge \ldots \wedge \lambda^n$. Define matrices $\mathbf{X}, \boldsymbol{\lambda} \in \mathbb{R}^{n \times n}$ whose ith row and jth column are $dx^i(X_j)$ and $\lambda^i(\partial_j)$ respectively. By Proposition 9.10, $(\lambda^1 \wedge \ldots \wedge \lambda^n)(\partial_1, \ldots, \partial_n) = \det \boldsymbol{\lambda}$.

    (1) Prove $\boldsymbol{\lambda} = \mathbf{X}^{-1}$.

(2) Prove $\mathbf{X}^T \mathbf{g} \mathbf{X} = \mathbb{1}$.
(3) Deduce (11.4).

Most people's favorite manifolds are submanifolds of Euclidean space—especially surfaces in $\mathbb{R}^3$. Generalizing this notion slightly, an $(n-1)$-dimensional submanifold M of an n-manifold N is called a **hypersurface** (*Hyperfläche*) in N. Any Riemannian metric g on N induces a Riemannian metric on any submanifold $M \subset N$, defined simply by restricting each of the inner products g_p on tangent spaces $T_p N$ to the subspaces $T_p M \subset T_p N$. To put this another way, one can denote the inclusion map of M into N by $i : M \hookrightarrow N$ and observe that for every $p \in M$, $i_* : T_p M \hookrightarrow T_p N$ is the corresponding inclusion map of vector spaces, so the Riemannian metric induced by $g \in \Gamma(T_2^0 N)$ on M is the pullback $i^* g \in \Gamma(T_2^0 M)$. With this understood, we will show next that there is an easy way to derive from the compatible volume form on an oriented Riemannian manifold the compatible volume form on any oriented hypersurface.

DEFINITION 11.13. For an n-dimensional vector space V and an integer $k = 1, \ldots, n$, the **interior product** is the bilinear map

$$V \times \Lambda^k V^* \to \Lambda^{k-1} V^* : (v, \alpha) \mapsto \iota_v \alpha$$

defined by $\iota_v \alpha(w_1, \ldots, w_{k-1}) := \alpha(v, w_1, \ldots, w_{k-1})$. On a manifold M, the map

$$\mathfrak{X}(M) \times \Omega^k(M) \to \Omega^{k-1}(M) : (X, \omega) \mapsto \iota_X \omega$$

is defined similarly by $(\iota_X \omega)_p := \iota_{X(p)} \omega_p$ for all $p \in M$.

PROPOSITION 11.14. *Assume (N, g) is a Riemannian manifold, $M \subset N$ is a hypersurface with inclusion map $i : M \hookrightarrow N$, and $\nu : M \to TN$ is a continuous map*[40] *such that for every $p \in M$, $\nu(p) \in T_p N$ is a unit vector orthogonal to $T_p M$. (In this situation we call ν a **unit normal vector field** for M.) Then if $d\mathrm{vol}_N \in \Omega^n(N)$ is a volume form on N compatible with g,*

$$d\mathrm{vol}_M := (\iota_\nu d\mathrm{vol}_N)|_{TM} \in \Omega^{n-1}(M)$$

is a volume form on M compatible with the induced metric $i^* g$.

PROOF. For any $p \in M$ and an orthonormal basis $X_1, \ldots, X_{n-1}$ of $T_p M$, the n-tuple $\nu(p), X_1, \ldots, X_{n-1}$ forms an orthonormal basis of $T_p N$, thus

$$|\iota_\nu d\mathrm{vol}_N(X_1, \ldots, X_{n-1})| = |d\mathrm{vol}(\nu(p), X_1, \ldots, X_{n-1})| = 1.$$

$\square$

EXERCISE 11.15. Using Cartesian coordinates (x, y, z) on $\mathbb{R}^3$, let $\omega := x \, dy \wedge dz + y \, dz \wedge dx + z \, dx \wedge dy \in \Omega^2(\mathbb{R}^3)$, and let $i : S^2 \hookrightarrow \mathbb{R}^3$ denote the inclusion of the unit sphere.
  (a) Show that $d\mathrm{vol}_{S^2} := i^* \omega \in \Omega^2(S^2)$ is a volume form compatible with the Riemannian metric on S^2 induced by the Euclidean inner product.
      Hint: Pick a good vector field $X \in \mathfrak{X}(\mathbb{R}^3)$ with which to write ω as $\iota_X(dx \wedge dy \wedge dz)$.
  (b) Show that in the spherical coordinates (θ, ϕ) of Exercise 1.7, $d\mathrm{vol}_{S^2} = \cos \phi \, d\theta \wedge d\phi$.
  (c) On the open upper hemisphere $\mathcal{U}_+ := \{z > 0\} \subset S^2 \subset \mathbb{R}^3$, one can define a chart $(x, y) : \mathcal{U}_+ \to \mathbb{R}^2$ by restricting to $\mathcal{U}_+$ the usual Cartesian coordinates x and y, which are then related to the z-coordinate on this set by $z = \sqrt{1 - x^2 - y^2}$. Show that $d\mathrm{vol}_{S^2} = \frac{1}{z} dx \wedge dy$ on $\mathcal{U}_+$.

---

[40]In fact it will follow from these assumptions that ν is also smooth, but one does not need to know that in advance.

(d) Compute the surface area of $S^2 \subset \mathbb{R}^3$ in two ways: once using the formula for $d\mathrm{vol}_{S^2}$ in part (b), and once using part (c) instead. In both cases, the results of §11.2 will allow you to express the answer in terms of a single Lebesgue integral over a region in $\mathbb{R}^2$, and there will be no need for any partition of unity.

11.4. Densities. [41]

You may have wondered: what if M is non-orientable, but I still want to compute its volume?

There are two problems in this situation: one is that according to Proposition 11.7, M cannot admit a volume form if it does not also admit an orientation, but there is also the more fundamental issue that the integral of an n-form over an n-manifold is not defined unless M comes with an orientation. Recall from §10.1: the trouble was that if $\omega = f\,dx^1 \wedge \ldots \wedge dx^n = g\,dy^1 \wedge \ldots \wedge dy^n$ for two different local coordinate systems $x, y : \mathcal{U} \to \mathbb{R}^n$ on the same region, then the Legesgue integrals $\int_{x(\mathcal{U} \cap A)} f \circ x^{-1}\,dm$ and $\int_{y(\mathcal{U} \cap A)} g \circ y^{-1}\,dm$ cannot generally be assumed to match unless the transition map $\psi := y \circ x^{-1} : x(\mathcal{U}) \to y(\mathcal{U})$ is orientation preserving. This problem is summarized by Equation (10.3), which resembles the classical change-of-variables formula, but does not match it exactly unless $\det(D\psi)$ is everywhere positive.

One way to circumvent this problem is to give up on intergrating the real-valued functions f and g and instead integrate their absolute values, so that (10.3) gives rise to the completely true statement

$$\int_{y(A)} \left|g \circ y^{-1}\right|\,dm = \int_{\psi(x(A))} |G|\,dm = \int_{x(A)} |(G \circ \psi)| \cdot |\det D\psi|\,dm = \int_{x(A)} \left|f \circ x^{-1}\right|\,dm,$$

in which we are again writing $G := g \circ y^{-1}$. The message of this calculation is that if we are willing to ignore the sign of an n-form and pay attention only to its magnitude, then we will no longer need to restrict ourselves to orientation-preserving transition maps.

DEFINITION 11.16. A (nonnegative) **density** on a smooth n-manifold M is a map

$$\mu : (TM)^{\oplus n} \to [0, \infty)$$

whose restriction to $T_pM \times \ldots \times T_pM$ for each $p \in M$ takes the form $\mu_p(X_1, \ldots, X_n) = |\omega_p(X_1, \ldots, X_n)|$ for some $\omega_p \in \Lambda^n T_p^* M$. In a smooth chart $(\mathcal{U}, x)$, every density can thus be written in terms of the standard volume form $dx^1 \wedge \ldots \wedge dx^n \in \Omega^n(\mathcal{U})$ as

$$\mu = f \cdot \left|dx^1 \wedge \ldots \wedge dx^n\right|$$

for a unique function $f : \mathcal{U} \to [0, \infty)$. We call μ a **smooth density** if the function f defined in this way is smooth for all choices of smooth chart on M.

REMARK 11.17. It is also possible to define densities with negative values (see e.g. [**Lee13a**]), but we will not need this. Our refusal to define negative densities means that the space

$$\mathscr{D}(M) := \{\text{smooth densities on } M\}$$

is not a vector space, but it does admit natural notions of addition and multiplication by nonnegative scalars.

The **support** of a density $\mu \in \mathscr{D}(M)$ is of course the closure of the set $\{p \in M \mid \mu_p \neq 0\} \subset M$, and we will denote

$$\mathscr{D}_c(M) := \left\{\mu \in \mathscr{D}(M) \mid \mu \text{ has compact support}\right\}.$$

---

[41]The contents of §11.4 were not covered in the lecture and will not be referred to again in this course, at least not in any serious way. This section of the notes is provided only for your information.

For smooth maps $\varphi : M \to N$, there is a natural **pullback** operation $\varphi^* : \mathscr{D}(N) \to \mathscr{D}(M)$ defined by

$$(\varphi^* \mu)(X_1, \ldots, X_n) := \mu(\varphi_* X_1, \ldots, \varphi_* X_n).$$

If we revise the discussion of §10.1 to work with densities instead of n-forms, then the key fact is that for any two charts x and y defined on the same domain $\mathcal{U}$, we have

$$\left| dy^1 \wedge \ldots \wedge dy^n \right| = \left| \det\left(\frac{\partial y}{\partial x} \right) \right| \cdot \left| dx^1 \wedge \ldots \wedge dx^n \right| \qquad \text{on } \mathcal{U},$$

thus if $\mu = f \left| dx^1 \wedge \ldots \wedge dx^n \right| = g \left| dy^1 \wedge \ldots \wedge dy^n \right|$ on this region, the nonnegative functions f and g are related by $f = g \cdot \left| \det\left(\frac{\partial y}{\partial x} \right) \right|$. The presence of the absolute value in this expression repairs our previous problem with orientations, and it now follows that the integrals $\int_{x(A)} f \circ x^{-1} \, dm$ and $\int_{y(A)} g \circ y^{-1} \, dm$ will *always* match, even if $y \circ x^{-1}$ is orientation reversing. The proof of Theorem 10.30 can now easily be adapted to establish the following:

THEOREM 11.18. *For $n \in \mathbb{N}$, one can uniquely associate to every smooth n-manifold M and measurable subset $A \subset M$ a map*

$$\mathscr{D}_c(M) \to [0, \infty) : \mu \mapsto \int_A \mu$$

such that the following conditions are satisfied:

(1) *$\int_A (\mu_1 + \mu_2) = \int_A \mu_1 + \int_A \mu_2$ for any $\mu_1, \mu_2 \in \mathscr{D}_c(M)$.*
(2) *If $\mathcal{U} \subset M$ is an open subset containing $\operatorname{supp}(\mu) \cap A$, then then $\int_{\mathcal{U} \cap A} \mu = \int_A \mu$.*
(3) *For $M = \mathcal{U} \subset \mathbb{R}^n$ an open subset of Euclidean space and the standard Cartesian coordinates $x^1, \ldots, x^n$,*

$$\int_A f \left| dx^1 \wedge \ldots \wedge dx^n \right| = \int_A f \, dm$$

for all smooth compactly supported functions $f : \mathcal{U} \to [0, \infty)$, where the right hand side is the standard Lebesgue integral of f.
(4) *For any diffeomorphism $\psi : M \to N$ between a pair of n-manifolds,*

$$\int_A \psi^* \mu = \int_{\psi(A)} \mu$$

holds for all $\mu \in \mathscr{D}_c(N)$ and measurable subsets $A \subset M$.

$\square$

The freedom in this theorem to allow non-orientable manifolds and diffeomorphisms that are not orientation preserving is paid for by the fact that integrals of nonnegative densities are always nonnegative, and thus tend to deliver less information than the *real*-valued integrals of differential forms. As mentioned in Remark 11.17 above, one can also allow densities with negative values and thus obtain negative integrals, but this does not add very much generality: it is tantamount to defining a measure μ via integrals of a positive density and then computing integrals $\int_A f \, d\mu$ of functions f that are also allowed to have negative values. Integration of densities is a somewhat less elegant and less useful construction on the whole than integration of forms; in particular, there are many more beautiful theorems involving the latter. Nonetheless, there are of course geometric situations in which an integral that is guaranteed to be nonnegative is exactly what one wants:

DEFINITION 11.19. A **volume element** on a smooth n-manifold M is a density $d\mathrm{vol}$ such that $d\mathrm{vol}_p \neq 0$ for every $p \in M$. If M is equipped with a volume element $d\mathrm{vol}$, one defines the **volume**

of measurable sets $A \subset M$ by

$$\mathrm{Vol}(A) := \int_A d\mathrm{vol} \geqslant 0.$$

We can now state a version of Corollary 11.10 that does not depend on orientability; its proof is an easy adaptation of arguments in the previous section.

PROPOSITION 11.20. *Every Riemannian manifold (M, g) admits a unique volume element $d$vol such that for all $p \in M$ and every orthonormal basis $X_1, \ldots, X_n$ of $T_p M$, $d$vol$(X_1, \ldots, X_n) = 1$.* $\qquad \square$

We will not have any more occasions to talk about densities and volume elements in this course, but it is good to be aware that a theory of integration exists for non-orientable manifolds, even if it is less versatile and less powerful than the orientable case.

12. Stokes' theorem

It is finally time to tell you the true reason why the exterior derivative is important: it is "dual" in some sense to the operation of replacing a manifold by its boundary. First we will have to discuss what is meant by the *boundary* of a manifold, and we will have to be fairly careful with orientations if we want to get all the signs right.

12.1. A word about dimension zero. You may or may not have noticed that manifolds of dimension zero have been explicitly excluded from all discussion of orientations and integration so far. You probably didn't miss it, because in truth, integrals of 0-forms on 0-manifolds are not very interesting. But we have to define them now, because as soon as we start talking about manifolds with boundary, 0-manifolds will inevitably arise, namely as boundaries of 1-manifolds.

A 0-manifold M, you may recall, is simply a discrete set, and it can have at most countably many elements; it is compact if and only if it is finite. A 0-form on M is then an arbitrary function $f : M \to \mathbb{R}$. There is no need to worry about continuity or smoothness since M is discrete, and the support of f is just the set of all points p where $f(p) \neq 0$, so $f : M \to \mathbb{R}$ has compact support if and only if it is zero outside of a finite set.

Since there is no such thing as a "basis" of a 0-dimensional vector space and no meaningful sense in which one can say that a (the) map $\mathbb{R}^0 \to \mathbb{R}^0$ preserves or reverses orientation, the entire discussion of orientations in §10.2 is useless for $n = 0$. What we will use instead looks terribly naive at first glance, but we will see that it works:

DEFINITION 12.1. An **orientation** of a 0-manifold M is a function $\varepsilon : M \to \{1, -1\}$, i.e. it a assigns to each point of M a label as either "positive" or "negative". A bijection $\varphi : M \to N$ between two oriented 0-manifolds is **orientation preserving** if it maps all positive points to positive points and all negative points to negative points, and it is **orientation reversing** if it exchanges the sets of positive and negative points.

DEFINITION 12.2. For M a 0-manifold with orientation $\varepsilon : M \to \{1, -1\}$ and $f \in \Omega_c^0(M)$, the integral of f on a subset $A \subset M$ is defined by

$$\int_A f := \sum_{p \in A} \varepsilon(p) f(p),$$

where the sum is necessarily finite since f has compact support.

The only other thing worth saying for now about this definition is that it trivially satisfies the usual change-of-variables formula

$$\int_A \varphi^* f = \int_{\varphi(A)} f, \qquad f \in \Omega_c^0(N)$$

whenever $\varphi : M \to N$ is an orientation-preserving bijection of oriented 0-manifolds.

12.2. Manifolds with boundary. The definitions from Lectures 1 and 2 need to be generalized if we want to accommodate examples like the unit n-**disk**

$$\mathbb{D}^n := \left\{ x \in \mathbb{R}^n \ \big| \ |x| \leqslant 1 \right\},$$

whose interior is accurately described as a smooth n-manifold, but there are no n-dimensional charts (by our current definition) describing neighborhoods in $\mathbb{D}^n$ of points on the boundary

$$\partial \mathbb{D}^n := S^{n-1} \subset \mathbb{D}^n.$$

An even simpler example is the **half-plane**

$$\mathbb{H}^n := (-\infty, 0] \times \mathbb{R}^{n-1} \subset \mathbb{R}^n,$$

whose boundary is the linear subspace

$$\partial \mathbb{H}^n := \{0\} \times \mathbb{R}^{n-1} \subset \mathbb{R}^n.$$

Just as subspaces of this form serve as local models of submanifolds as seen through slice charts, the half-plane will serve as our local model for a manifold with boundary.

DEFINITION 12.3. An n-dimensional **boundary chart** $(\mathcal{U}, x)$ on a set M consists of a subset $\mathcal{U} \subset M$ and an injective map $x : \mathcal{U} \hookrightarrow \mathbb{H}^n$ whose image $x(\mathcal{U}) \subset \mathbb{H}^n$ is an open set.[42]

The only difference between this and Definition 1.4 is the replacement of $\mathbb{R}^n$ by the half-space $\mathbb{H}^n$. A boundary chart $(\mathcal{U}, x)$ will sometimes also be a chart according to our original definition, because an open subset $x(\mathcal{U}) \subset \mathbb{H}^n$ might also be an open subset of $\mathbb{R}^n$; indeed, it will be so if $x(\mathcal{U}) \cap \partial \mathbb{H}^n = \varnothing$. For this reason, any set that is covered by charts can equally well be covered by boundary charts: one need only modify each chart $(\mathcal{U}, x)$ by a translation so that its image lies in the interior of the half-plane, or if this is impossible because $x(\mathcal{U})$ is unbounded in the x^1-direction, first break it up into countably many open subsets so that this can be done. However, if $x(\mathcal{U})$ does contain points in the boundary $\partial \mathbb{H}^n$, then it is not open in $\mathbb{R}^n$. A typical example is the "open" half-disk

$$\mathring{\mathbb{D}}^n_- := \left\{ (x^1, \ldots, x^n) \in \mathbb{R}^n \ \big| \ (x^1)^2 + \ldots + (x^n)^2 < 1 \text{ and } x^1 \leqslant 0 \right\},$$

which is open in $\mathbb{H}^n$ but not open in $\mathbb{R}^n$ since it does not contain any ball around points in $\mathring{\mathbb{D}}^n_- \cap \partial \mathbb{H}^n$. In this sense, Definition 12.3 is strictly more general than our original definition of a chart.

The notion of **transition maps** between two charts $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ generalizes in an obvious way to boundary charts,

$$(12.1) \qquad \begin{aligned} \mathbb{H}^n \supset x(\mathcal{U} \cap \mathcal{V}) \xrightarrow{y \circ x^{-1}} y(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{H}^n, \\ \mathbb{H}^n \supset y(\mathcal{U} \cap \mathcal{V}) \xrightarrow{x \circ y^{-1}} x(\mathcal{U} \cap \mathcal{V}) \subset \mathbb{H}^n, \end{aligned}$$

though since $x(\mathcal{U} \cap \mathcal{V})$ and $y(\mathcal{U} \cap \mathcal{V})$ may be open in $\mathbb{H}^n$ but not in $\mathbb{R}^n$, the notion of smooth compatibility requires a bit of clarification. The quickest approach is to say that a map $f : \mathcal{O} \to \mathbb{R}^m$ defined on some (not necessarily open) subset $\mathcal{O} \subset \mathbb{R}^n$ is of class C^k if and only if it admits an extension of class C^k to some open neighborhood of $\mathcal{O}$ in $\mathbb{R}^n$. With this understood, we will call $(\mathcal{U}, x)$ and $(\mathcal{V}, y)$ **smoothly compatible** if both of the transition maps in (12.1) admit smooth extensions over open (in $\mathbb{R}^n$) neighborhoods of their domains.

---

[42]One finds a few variations on this definition in the literature, in which the half-space $\mathbb{H}^n = (-\infty, 0] \times \mathbb{R}^{n-1}$ gets replaced by different half-spaces such as $[0, \infty) \times \mathbb{R}^{n-1}$ or $\mathbb{R}^{n-1} \times [0, \infty)$. This detail makes no meaningful difference for the definition of a smooth manifold with boundary, but it starts to matter as soon as one has to think about orientations. The definition in the form we've given here leads to the simplest possible definition of boundary orientations, and a relatively straightforward proof of Stokes' theorem.

REMARK 12.4. For open subsets $\mathcal{O} \subset \mathbb{H}^n$ in half-space, the notion of a C^k-map $f : \mathcal{O} \to \mathbb{R}^m$ admits various alternative characterizations that do not require extending f over a larger neighborhood in $\mathbb{R}^n$. Denote $\partial\mathcal{O} := \mathcal{O} \cap \partial\mathbb{H}^n$ and $\mathring{\mathcal{O}} := \mathcal{O}\backslash\partial\mathcal{O}$. Then $f : \mathcal{O} \to \mathbb{R}^m$ is of class C^k if and only if its restriction $f|_{\mathring{\mathcal{O}}} : \mathring{\mathcal{O}} \to \mathbb{R}^m$ is of class C^k and either of the following equivalent conditions are satisfied:

- All partial derivatives of $f|_{\mathring{\mathcal{O}}} : \mathring{\mathcal{O}} \to \mathbb{R}^m$ up to order k admit continuous extensions over $\mathcal{O}$;
- All partial derivatives of $f|_{\mathring{\mathcal{O}}} : \mathring{\mathcal{O}} \to \mathbb{R}^m$ up to order k are uniformly continuous on bounded subsets of $\mathring{\mathcal{O}}$.

It is an easy analysis exercise to show that these two conditions are equivalent, and they clearly also follow from the assumption that $f : \mathcal{O} \to \mathbb{R}^m$ admits a C^k-extension to a neighborhood, but the converse takes more effort to prove. We will not do so here since we will never need to use this fact, but the details can be found e.g. in [**AF03**, §5.19–§5.21].

A **smooth n-manifold with boundary** can now be defined by generalizing our previous definition of a smooth n-manifold so that all charts in its maximal smooth atlas are allowed to be boundary charts. Implicit in this definition is the fact that an atlas of boundary charts on M determines a natural topology on M such that the domains of boundary charts are also open sets in M and the charts themselves are homeomorphisms onto their images. This definition is *strictly* more general than what we have been working with so far: a manifold with boundary can sometimes also be a manifold in our previous sense, because its atlas might consist only of regular charts whose images are open subsets of $\mathbb{R}^n$. But if M is a manifold with boundary, it contains a distinuished subset

$$\partial M := \left\{ p \in M \mid x(p) \in \partial\mathbb{H}^n \text{ for some smooth boundary chart } (\mathcal{U}, x) \right\},$$

called its **boundary** (*Rand*). It should be easy to convince yourself that if $x(p) \in \partial\mathbb{H}^n$ for some particular boundary chart $(\mathcal{U}, x)$, then this also holds for every other boundary chart $(\mathcal{V}, y)$ with $p \in \mathcal{V}$; this is because by the inverse function theorem, the transition maps in (12.1) necessarily preserve the interior of $\mathbb{H}^n$, and therefore also preserve its boundary $\partial\mathbb{H}^n$. Moreover, every boundary chart whose domain intersects ∂M can be viewed as a slice chart for ∂M, so that it is appropriate to call ∂M a smooth $(n-1)$-dimensional submanifold of M. In particular, ∂M inherits from M a natural smooth structure and becomes a smooth $(n-1)$-manifold. We observe that M itself is a manifold in our previous sense if and only if $\partial M = \varnothing$; one sometimes says in this case that M is a *manifold without boundary*. Since $x(\mathcal{U}) \cap \partial\mathbb{H}^n$ is *always* an open subset of $\partial\mathbb{H}^n = \{0\} \times \mathbb{R}^{n-1}$ for a boundary chart $(\mathcal{U}, x)$, the manifold ∂M never has boundary, i.e.

$$\partial(\partial M) = \varnothing.$$

REMARK 12.5. One can define even more general notions such as a "manifold with boundary and corners," in which images of charts are allowed to be open subsets of quadrants like $(-\infty, 0] \times (-\infty, 0] \times \mathbb{R}^{n-2}$, in which case ∂M may also be a manifold with nonempty boundary (and possibly corners). The literature on these objects seems however to be not entirely unanimous on what the correct definitions are. In this course, we will occasionally mention corners in heuristic discussions, but we will not study them in any serious way.

REMARK 12.6. From now on, you must pay careful attention whenever you see the word "manifold" without further modifiers, as its default meaning may be either "manifold without boundary" or "manifold with boundary" depending on the context. Keep in mind also that these categories are not mutually exclusive: a "manifold with boundary" may have $\partial M = \varnothing$. I generally make a point of saying "manifold with nonempty boundary" if I want to explicitly assume $\partial M \neq \varnothing$. I also will often refer to boundary charts simply as "charts" when working in the context of manifolds with boundary.

EXAMPLE 12.7. Suppose N is an n-manifold without boundary and $M \subset N$ is an open subset such that $\overline{M} \backslash M \subset N$ is a smooth $(n-1)$-dimensional submanifold, i.e. a hypersurface. Then the closure $\overline{M} \subset N$ is naturally a smooth n-manifold with boundary and

$$\partial \overline{M} = \overline{M} \backslash M,$$

because every slice chart for $\overline{M} \backslash M$ can be modified in straightforward ways so as to be interpreted as a boundary chart for $\overline{M}$. Most interesting examples of manifolds with boundary arise in this way, and it can be shown that *all* manifolds with boundary are diffeomorphic to examples of this type, though the ambient manifold N might not always be a natural part of the picture. As an important special case, if $f : N \to \mathbb{R}$ is a smooth function with $c \in \mathbb{R}$ as a regular value, then $f^{-1}((-\infty, c])$ and $f^{-1}([c, \infty))$ are naturally manifolds with boundary, the boundary in each case being the regular level set $f^{-1}(c) \subset N$. Examples of this type include the n-disk $\mathbb{D}^n \subset \mathbb{R}^n$ mentioned at the beginning of this section.

Almost all of the notions we have discussed in this course so far—tangent vectors and tangent maps, vector fields, tensors, forms, orientations—can be generalized in straightforward ways for manifolds with boundary so long as one remembers what smoothness means on open sets in half-space. The tangent spaces $T_p M$ are defined exactly as before for $p \in M \backslash \partial M$, though it takes a bit more thought to arrive at the right definition for $p \in \partial M$. Here it is useful to keep Example 12.7 in mind and imagine M as a closed subset of a larger manifold N without boundary such that $\partial M \subset N$ is a smooth hypersurface: the correct definition for $p \in \partial M$ is then $T_p M := T_p N$, so that $T_p M$ is still a vector space of the same dimension as M. If there is no ambient manifold N in the picture, then one can instead modify the original definition of $T_p M$ in terms of paths through p by allowing paths of the form $\gamma : (-\epsilon, 0] \to M$ or $\gamma : [0, \epsilon) \to M$ that run "out of" or "into" M through its boundary. The crucial thing to remember is that for any chart $(\mathcal{U}, x)$ with $p \in \mathcal{U}$, $d_p x : T_p M \to \mathbb{R}^n$ is still a linear isomorphism, even if $p \in \partial M$. Since $\partial M \subset M$ is an $(n-1)$-dimensional submanifold, $T_p(\partial M) \subset T_p M$ is an $(n-1)$-dimensional subspace. The complement $T_p M \backslash T_p(\partial M)$ has two connected components: one consists of all vectors that point **outward**, meaning they are derivatives of "departing" paths $\gamma : (-\epsilon, 0] \to M$, and the other contains vectors that point **inward**, which are derivatives of "entering" paths $\gamma : [0, \epsilon) \to M$. It should go without saying that flows of vector fields $X \in \mathfrak{X}(M)$ require extra care when $\partial M \neq \emptyset$, because e.g. if $p \in \partial M$ and $X(p)$ points outward/inward, then there is no forward/backward flow line starting at p for any nonzero time. There is no problem however if $X|_{\partial M}$ is everywhere tangent to the boundary, since it then also defines a flow on ∂M, and Theorem 5.1 in this case goes through without changes.

The notion of a submanifold also requires slight modification when boundaries are involved: the appropriate definition is to call $M \subset N$ a **submanifold** (with boundary) whenever it is the image of an embedding of some manifold with boundary. This allows a few possibilities that were not covered by our original definition in terms of slice charts: one of them was already mentioned above, namely the natural embedding of the boundary $\partial M \hookrightarrow M$. Another is Example 12.7: if N is an n-manifold and $M \subset N$ is an open subset such that $\partial \overline{M} := \overline{M} \backslash M$ is a smooth hypersurface in M, then $\overline{M}$ is a smooth n-dimensional submanifold with boundary in N. This opens the previously excluded possibility that a manifold and submanifold may have the same dimension without one being an open subset of the other.

PROPOSITION 12.8. *If M is an oriented manifold of dimension $n \geqslant 2$ with boundary, then the $(n-1)$-manifold ∂M inherits a natural orientation such that for every oriented boundary chart $(\mathcal{U}, x)$ on M, $(\mathcal{U} \cap \partial M, x|_{\mathcal{U} \cap \partial M})$ is an oriented chart on ∂M. This orientation can also be characterized as follows: for every point $p \in \partial M$ and any tangent vector $\nu \in T_p M \backslash T_p(\partial M)$ that*

points outward, a basis $(X_1, \ldots, X_{n-1})$ of $T_p(\partial M)$ is positively oriented if and only if the basis $(\nu, X_1, \ldots, X_{n-1})$ of T_pM is positively oriented.

The orientation defined on ∂M from an orientation of M via this proposition is called the **boundary orientation**. We will always assume unless otherwise specified that when M is oriented, ∂M is endowed with the boundary orientation.

PROOF OF PROPOSITION 12.8. The main point is that any orientation-preserving transition map $\psi := y \circ x^{-1} : x(\mathcal{U} \cap \mathcal{V}) \to y(\mathcal{U} \cap \mathcal{V})$ not only preserves the subset $\partial \mathbb{H}$ but is also orientation preserving on this subset. To see this, observe that the derivative $D\psi(q) : \mathbb{R}^n \to \mathbb{R}^n$ at any point q must be an isomorphism that preserves each of the subsets $\mathbb{H}^n$ and $\partial \mathbb{H}^n$, thus it is represented by a matrix of the form

$$D\psi(q) = \begin{pmatrix} a & 0 \\ \mathbf{v} & \mathbf{B} \end{pmatrix}, \qquad a > 0, \ \mathbf{v} \in \mathbb{R}^{n-1}, \ \mathbf{B} \in \mathbb{R}^{(n-1) \times (n-1)},$$

where $\mathbf{B}$ is the derivative at q of the restricted transition map on $\partial \mathbb{H}$. Clearly $\det D\psi(q) > 0$ if and only if $\det \mathbf{B} > 0$. This shows that the restriction of an oriented atlas of M to ∂M is an oriented atlas of ∂M.

To characterize the boundary orientation in terms of bases, choose any oriented chart $(\mathcal{U}, x)$ near a point $p \in \partial M$, so the coordinate vector fields $\partial_1, \ldots, \partial_n$ define a positively-oriented basis of T_pM. The restriction of $(\mathcal{U}, x)$ to ∂M now defines an oriented chart for ∂M near p, and the coordinate vector fields for this restricted chart are $(\partial_2, \ldots, \partial_n)$, which therefore form a positively-oriented basis of $T_p(\partial M)$, and this can then be deformed continuously through bases to any other positively-oriented basis $(X_1, \ldots, X_{n-1})$ of $T_p(\partial M)$. Since ∂_1 points outward at p, it follows that for any other vector $\nu \in T_pM \backslash T_p(\partial M)$ pointing outward, the basis $(\nu, X_1, \ldots, X_{n-1})$ of T_pM can be deformed continuously through bases to $(\partial_1, \ldots, \partial_n)$, simply by deforming $(X_1, \ldots, X_{n-1})$ through bases of $T_p(\partial M)$ to $(\partial_2, \ldots, \partial_n)$ and simultaneously deforming ν through outward-pointing vectors to ∂_1. This proves that $(\nu, X_1, \ldots, X_{n-1})$ is a positively-oriented basis of T_pM, and conversely, if $(X_1, \ldots, X_{n-1})$ had been negatively oriented, we could apply the same argument to the positively-oriented basis $(-X_1, X_2, \ldots, X_{n-1})$ and thus conclude that $(\nu, X_1, \ldots, X_{n-1})$ is also negatively oriented. $\qquad \square$

We had to exclude the case $\dim M = 1$ from Proposition 12.8 because orientations of 0-manifolds cannot be described in terms of charts or bases.

DEFINITION 12.9. If M is an oriented 1-manifold with boundary, the **boundary orientation** of the 0-manifold ∂M is defined by calling a point $p \in \partial M$ positive if the basis of T_pM formed by an outward-pointing vector $\nu \in T_pM$ is positively oriented, and negative otherwise.

EXAMPLE 12.10. Any nontrivial compact interval $[a, b] \subset \mathbb{R}$ is a 1-manifold with boundary, and if we assign it the canonical orientation of $\mathbb{R}$ then the boundary orientation of $\partial[a, b] = \{a, b\}$ makes b a positive point and a a negative point. Informally, we write

$$\partial[a, b] = -\{a\} \sqcup \{b\}.$$

A slightly different example is

$$\partial(-\infty, 0] = \{0\},$$

in which the point 0 is assigned a positive orientation; this will be relevant in the proof of Stokes' theorem below.

12.3. The boundary operator is a graded derivation. I want to point out something about boundary orientations that is not an essential part of this discussion, but it may help you to understand more intuitively why graded Leibniz rules keep showing up.

In the previous section we defined an operator "∂" that takes an oriented n-manifold M (with boundary) and returns an oriented $(n-1)$-manifold ∂M. It satisfies $\partial(\partial M) = \varnothing$ for all M, which seems formally similar to the relation $d \circ d = 0$ satisfied by the exterior derivative. We will see in the next section that the operators ∂ and d are in fact dual to each other in a sense that can be made precise, thus it should not be surprising that they have formally similar properties. We claim in particular that ∂ also satisfies a graded Leibniz rule.

To understand what this means, suppose M and N are two oriented manifolds with boundary, with $\dim M = m$ and $\dim N = n$. This discussion will be heuristic, so we will choose not to worry about the fact that $M \times N$ might not actually be a smooth manifold with boundary: in particular, the neighborhood of a point $(p,q) \in \partial M \times \partial N \subset M \times N$ cannot be described smoothly via our usual notion of a boundary chart, and a completely correct description would require the notion of manifolds with boundary *and corners* (cf. Remark 12.5). Nonetheless, it seems sensible to write

(12.2) $$\partial(M \times N) = (\partial M \times N) \cup (M \times \partial N),$$

and outside of the exceptional subset $\partial M \times \partial N$, it is literally true that $M \times N$ is a smooth manifold whose boundary is the union of these two pieces. Formally, $M \times N$ is a smooth manifold with boundary and corners, and its boundary consists of two smooth *faces* $\partial M \times N$ and $M \times \partial N$, each of which are smooth manifolds with boundary, and they are attached to each other at their common boundary $\partial M \times \partial N$.

Now, let's say all that again but pay attention to orientations. The product of two oriented manifolds M and N carries a natural **product orientation** such that for any $(p,q) \in M \times N$ and any pair of positively oriented bases $(X_1, \ldots, X_m)$ of T_pM and $(Y_1, \ldots, Y_n)$ of T_qN, $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ is a positively-oriented basis of $T_{(p,q)}(M \times N) = T_pM \times T_qN$; here we identify each $X_i \in T_pM$ with $(X_i, 0) \in T_pM \times T_qN = T_{(p,q)}(M \times N)$ and similarly identify $Y_j \in T_qN$ with $(0, Y_j) \in T_pM \times T_qN = T_{(p,q)}(M \times N)$. Now, if ∂M and ∂N are each endowed with their natural boundary orientations, then the two faces $\partial M \times N$ and $M \times \partial N$ of the boundary of $M \times N$ inherit product orientations, but these may or may not match the boundary orientation of $\partial(M \times N)$. Indeed, at a point $(p,q) \in \partial M \times N$, if we choose a positively-oriented basis $(X_2, \ldots, X_m)$ of $T_p(\partial M)$ and an outward-pointing vector $\nu \in T_pM \backslash T_p(\partial M)$, then $(\nu, 0) \in T_{(p,q)}(M \times N)$ also points outward through $\partial M \times N$ and $(\nu, X_2, \ldots, X_m, Y_1, \ldots, Y_n)$ forms a positively-oriented basis of $T_{(p,q)}(M \times N)$, implying that the boundary orientation of $\partial(M \times N)$ does match the product orientation of $\partial M \times N$. But things are different at a point $(p,q) \in M \times \partial N$. Choosing a positively-oriented basis $(Y_2, \ldots, Y_n)$ of $T_q(\partial N)$ and an outward-pointing vector $\nu \in T_qY \backslash T_q(\partial Y)$, a positively-oriented basis of $M \times N$ is given by $(X_1, \ldots, X_m, \nu, Y_2, \ldots, Y_n)$, but m flips are required in order to permute this basis to $(\nu, X_1, \ldots, X_m, Y_2, \ldots, Y_n)$, in which ν serves as an outward-pointing vector in $T_{(p,q)}(M \times N) \backslash T_{(p,q)}(\partial(M \times N))$ and $(X_1, \ldots, X_m, Y_2, \ldots, Y_n)$ as a positively-oriented basis for the product orientation on $M \times \partial N$. This means that the product orientation of $M \times \partial N$ matches the boundary orientation of $\partial(M \times N)$ if and only if $(-1)^m = 1$, i.e. if m is even. The oriented version of (12.2) can thus be written as

(12.3) $$\partial(M \times N) = (\partial M \times N) \cup ((-1)^m (M \times \partial N)),$$

where we define $-(M \times \partial N)$ to mean the oriented manifold obtained from $M \times \partial N$ by assigning it the *opposite* of the product orientation. The formal resemblance of this formula to a graded Leibniz rule is difficult to ignore, though we cannot make this notion precise in the present context since we have not defined any algebraic structure on the "set" of manifolds with boundary and corners. The easiest way to make such notions precise is probably by defining homology theory,

which is a topic for a topology course and not for this one, but I wanted in any case to provide (12.3) as further evidence of a formal similarity between the operators ∂ and d.

12.4. The main result. We can now define precisely what is meant by the informal statement that the operators d and ∂ are "dual" to each other. To understand the following statement, note that a k-form $\omega \in \Omega^k(M)$ induces a k-form $\Omega^k(L)$ on every submanifold $L \subset M$ by restriction, and this applies in particular to the boundary $\partial M \subset M$. Strictly speaking, the induced k-form on ∂M in this situation is $i^*\omega \in \Omega^k(\partial M)$ for the inclusion map $i : \partial M \hookrightarrow M$, but in the following we will also denote it by $\omega \in \Omega^k(\partial M)$ instead of $i^*\omega$.

THEOREM 12.11 (Stokes). *Assume M is an oriented n-manifold with boundary, where $n \geqslant 1$, and ∂M is equipped with its natural boundary orientation. Then for every $\omega \in \Omega^{n-1}_c(M)$,*

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

PROOF. As in the proof of Theorem 10.30, we can choose an open subset $M_0 \subset M$ with compact closure $\overline{M}_0$ such that $\operatorname{supp}(\omega) \subset M_0$, and then choose a finite covering of $\overline{M}_0$ by oriented charts $\{(\mathcal{U}_\alpha, x_\alpha)\}_{\alpha \in I}$ and a partition of unity $\{\varphi_\alpha : M \to [0,1]\}$ such that each φ_α has compact support in $\mathcal{U}_\alpha$ and $\sum_{\alpha \in I} \varphi_\alpha \equiv 1$ on M_0. Then each $\omega_\alpha := \varphi_\alpha \omega$ belongs to $\Omega^{n-1}_c(\mathcal{U}_\alpha)$, and we have $\omega = \sum_{\alpha \in I} \omega_\alpha$ and $d\omega = \sum_{\alpha \in I} d\omega_\alpha$ on M_0. If we can then prove $\int_{\mathcal{U}_\alpha} d\omega_\alpha = \int_{\partial \mathcal{U}_\alpha} \omega_\alpha$ for each α, we will have

$$\int_M d\omega = \int_{M_0} d\omega = \sum_{\alpha \in I} \int_{M_0} d\omega_\alpha = \sum_{\alpha \in I} \int_{\mathcal{U}_\alpha} d\omega_\alpha = \sum_{\alpha \in I} \int_{\partial \mathcal{U}_\alpha} \omega_\alpha = \sum_{\alpha \in I} \int_{\partial M} \omega_\alpha = \int_{\partial M} \omega.$$

In this way, the problem has been reduced to the special case in which M is covered by a single chart.

Next, observe that if the theorem has been proven to hold on another oriented manifold N and there is an orientation-preserving diffeomorphism $\psi : M \to N$, then we can write $\omega = \psi^*\alpha$ for $\alpha := \psi_*\omega \in \Omega^{n-1}_c(N)$ and use Proposition 9.17 along with the invariance of the integral under pullbacks to conclude

$$\int_M d\omega = \int_M d(\psi^*\alpha) = \int_M \psi^*(d\alpha) = \int_N d\alpha = \int_{\partial N} \alpha = \int_{\partial M} \psi^*\alpha = \int_{\partial M} \omega,$$

where we have also used the fact that a diffeomorphism $M \to N$ necessarily maps ∂M to ∂N. The latter is true since diffeomorphisms between regions in $\mathbb{R}^n$ map open sets to open sets, and neighborhoods of boundary points in $\mathbb{H}^n$ are not open in $\mathbb{R}^n$.

The combined result of the previous two paragraphs is that it will suffice to prove Stokes' theorem in the case where M is an open subset $\mathcal{U} \subset \mathbb{H}^n$ in half-space; in fact, since we are going to assume $\omega \in \Omega^{n-1}_c(\mathcal{U})$ has compact support, we may as well also assume M is the whole half-space $\mathbb{H}^n$. The proof now becomes a simple computation based on Fubini's theorem and the fundamental theorem of calculus. We can write ω in terms of n compactly supported smooth functions $f_1, \ldots, f_n : \mathbb{H}^n \to \mathbb{R}$ as

$$\omega = f_i \, \alpha^i, \qquad \text{where} \qquad \alpha^i := dx^1 \wedge \ldots \wedge \widehat{dx}^i \wedge \ldots \wedge dx^n \in \Omega^{n-1}(\mathbb{H}^n),$$

and the hat indicates again that the corresponding term does *not* appear. Then $d\alpha^i = 0$ for each i, and $dx^j \wedge \alpha^i = 0$ for every $j \neq i$, thus

$$d\omega = df_i \wedge \alpha^i = \sum_{i=1}^n \partial_i f_i \, dx^i \wedge \alpha^i = \sum_{i=1}^n (-1)^{i-1} \partial_i f_i \, dx^1 \wedge \ldots \wedge dx^n,$$

where we have refrained from using the summation convention in the last two expressions in order to avert confusion. Of the n terms in this sum, we claim that $n-1$ of them vanish when integrated

over $\mathbb{H}^n$. Let us check this specifically for $i = n$: choosing $N > 0$ large enough for the supports of the functions $f_1, \ldots, f_n$ to be contained in $[-N/2, 0] \times [-N/2, N/2]^{n-1}$, we use Fubini and the fundamental theorem of calculus to compute

$$\int_{\mathbb{H}^n} \partial_n f_n(x^1, \ldots, x^n) \, dx^1 \ldots dx^n = \int_{(-\infty, 0] \times \mathbb{R}^{n-2}} \left(\int_{\mathbb{R}} \partial_n f_n(x^1, \ldots, x^n) \, dx^n \right) \, dx^1 \ldots dx^{n-1} = 0$$

since the assumption on the support of f_n implies

$$\int_{\mathbb{R}} \partial_n f_n(x^1, \ldots, x_n) \, dx^n = \int_{-N}^{N} \partial_n f_n(x^1, \ldots, x_n) \, dx^n$$
$$= f_n(x^1, \ldots, x^{n-1}, N) - f_n(x^1, \ldots, x^{n-1}, -N) = 0.$$

This calculation works out the same way for each $i = 2, \ldots, n$, thus we find

$$\int_{\mathbb{H}^n} \omega = \int_{\mathbb{H}^n} \partial_1 f_1(x^1, \ldots, x^n) \, dx^1 \ldots dx^n = \int_{\mathbb{R}^{n-1}} \left(\int_{(-\infty, 0]} \partial_1 f_1(x^1, \ldots, x^n) \, dx^1 \right) dx^2 \ldots dx^n$$
$$= \int_{\mathbb{R}^{n-1}} \left(\int_{-N}^{0} \partial_1 f_1(x^1, \ldots, x^n) \, dx^1 \right) dx^2 \ldots dx^n$$
$$= \int_{\mathbb{R}^{n-1}} \left(f_1(0, x^2, \ldots, x^n) - f_1(-N, x^2, \ldots, x^n) \right) dx^2 \ldots dx^n$$
$$= \int_{\mathbb{R}^{n-1}} f_1(0, x^2, \ldots, x^n) \, dx^2 \ldots dx^n = \int_{\partial \mathbb{H}^n} f_1 \, dx^2 \wedge \ldots \wedge dx^n.$$

This last expression is $\int_{\partial \mathbb{H}^n} \omega$, as all other terms in ω contain dx^1, which vanishes when restricted to $\partial \mathbb{H}^n$. $\qquad \square$

EXAMPLE 12.12. For a smooth function $f : [a, b] \to \mathbb{R}$ on a nontrivial compact interval, we can denote the standard coordinate on $\mathbb{R}$ by x and write $df = f' \, dx$. The fundamental theorem of calculus then amounts to the following special case of Stokes' theorem,

$$\int_a^b f'(x) \, dx = \int_{[a,b]} df = \int_{-\{a\} \sqcup \{b\}} f = f(b) - f(a).$$

With this example in mind, Stokes' theorem is considered to be the natural n-dimensional generalization of the fundamental theorem of calculus.

EXERCISE 12.13. Prove the following version of *integration by parts*: if M is a compact oriented n-manifold with boundary, $\alpha \in \Omega^k(M)$ and $\beta \in \Omega^\ell(M)$ with $k + \ell = n - 1$, then

$$\int_M d\alpha \wedge \beta = \int_{\partial M} \alpha \wedge \beta - (-1)^k \int_M \alpha \wedge d\beta.$$

EXAMPLE 12.14. Heuristically, the discussion of §12.3 suggests that if M and N are compact manifolds with boundary having dimensions m and n respectively, then for any $\omega \in \Omega^{m+n-1}(M \times N)$, one should have

(12.4) $$\int_{M \times N} d\omega = \int_{\partial M \times N} \omega + (-1)^m \int_{M \times \partial N} \omega.$$

Here the right hand side is obtained from the integral of ω over $\partial(M \times N)$ by splitting the latter into the two *almost* disjoint subsets $\partial M \times N$ and $M \times \partial N$ (whose intersection $\partial M \times \partial N$ is a set of measure zero in either one), and then including a sign (cf. Exercise 11.4) to account for the fact that the product orientation of $M \times \partial N$ only matches the boundary orientation of $\partial(M \times N)$ when m is odd. As it stands, the left hand side of (12.4) does not immediately make sense unless

either ∂M or ∂N is empty (in which case (12.4) follows from Stokes' theorem), because $M \times N$ is otherwise not a smooth manifold with boundary. There are at least two ways that one could nonetheless make sense of (12.4):

(1) Define the notion of an oriented manifold *with boundary and corners* by allowing open subsets of $(-\infty, 0]^2 \times \mathbb{R}^{n-2}$ as local coordinate models, generalize the definition of the integral to this wider class of manifolds and prove that Stokes' theorem still holds if $\partial(M \times N)$ is understood in the sense of §12.3. This requires a bit of extra bookkeeping, but is not fundamentally more difficult than what we have already done.

(2) Choose a nested sequence of closed subsets $A_1 \subset A_2 \subset \ldots \bigcup_{j \in \mathbb{N}} A_j = M \times N$ such that each A_j is a smooth manifold with boundary (obtained by "smoothing the corner" of $M \times N$ in progressively small neighborhoods of $\partial M \times \partial N$), then define $\int_{M \times N} d\omega$ to mean $\lim_{j \to \infty} \int_{A_j} d\omega$ and deduce (12.4) from $\int_{A_j} d\omega = \int_{\partial A_j} \omega$.

REMARK 12.15. Much time and effort has been wasted by well-intentioned mathematicians trying to determine whether the correct orthography should be "Stokes' theorem" or "Stokes's theorem". After a years-long struggle I came to the conclusion that it is, essentially, a matter of personal taste. What I can say with absolute certainty is that it is not "Stoke's theorem".

12.5. The classical integration theorems. Various results that are considered central in classical vector calculus are easy consequences of Stokes' theorem.

12.5.1. *Divergence.* The **divergence** (*Divergenz*) of a vector field $X \in \mathfrak{X}(M)$ with respect to a volume form $d\mathrm{vol} \in \Omega^n(M)$ is defined as the unique real-valued function $\mathrm{div}(X) : M \to \mathbb{R}$ such that

$$(12.5) \qquad\qquad d(\iota_X d\mathrm{vol}) = \mathrm{div}(X) \cdot d\mathrm{vol}.$$

The definition makes sense because $\iota_X d\mathrm{vol}$ is an $(n-1)$-form and thus $d(\iota_X d\mathrm{vol})$ is an n-form, and every n-form is at each point a scalar multiple of the given volume form. It may not seem obvious at this stage why $\mathrm{div}(X)$ is a natural thing to define—we will address this question more thoroughly next week—but the following exercise should at least make it look familiar.

EXERCISE 12.16. Assume M is an n-manifold with a fixed volume form $d\mathrm{vol} \in \Omega^n(M)$, $(\mathcal{U}, x)$ is a chart on M and $f : \mathcal{U} \to \mathbb{R}$ is the unique function such that $d\mathrm{vol} = f\, dx^1 \wedge \ldots \wedge dx^n$ on $\mathcal{U}$. Show that for any $X \in \mathfrak{X}(M)$,

$$\mathrm{div}(X) = \frac{1}{f} \partial_i (f X^i) \qquad \text{on } \mathcal{U}.$$

In particular for the standard volume form $d\mathrm{vol} = dx^1 \wedge \ldots \wedge dx^n$ on $\mathbb{R}^n$, this reduces to the standard definition of divergence in vector calculus.

If M is a compact oriented n-manifold with boundary carrying a positive volume form $d\mathrm{vol}_M \in \Omega^n(M)$ and $X \in \mathfrak{X}(M)$ is a vector field, Stokes' theorem now implies

$$(12.6) \qquad\qquad \int_M \mathrm{div}(X)\, d\mathrm{vol}_M = \int_M d(\iota_X d\mathrm{vol}_M) = \int_{\partial M} \iota_X d\mathrm{vol}_M.$$

The geometric meaning of this last integral is best understood in the special case where $d\mathrm{vol}_M$ is the Riemannian volume form compatible with a Riemannian metric g on M, which we shall write in the following using the usual notation for inner products,

$$\langle X, Y \rangle := g(X, Y) \qquad \text{for } X, Y \in T_p M,\, p \in M.$$

By Proposition 11.14, the Riemannian volume form $d\mathrm{vol}_{\partial M}$ on ∂M is then

$$d\mathrm{vol}_{\partial M} := \iota_\nu d\mathrm{vol}_M|_{T(\partial M)} \in \Omega^{n-1}(\partial M),$$

where ν is the unique outward-pointing normal vector field to ∂M. (You should take a moment to convince yourself that we are getting the orientations right, i.e. $d\mathrm{vol}_{\partial M}$ really is a *positive* volume form with respect to the boundary orientation of ∂M.) To relate this to $\iota_X d\mathrm{vol}_M$, observe that along ∂M, $X = \langle X, \nu \rangle \nu + Y$ for a unique vector field $Y \in \mathfrak{X}(\partial M)$, but $\iota_Y d\mathrm{vol}_M$ vanishes when restricted to the boundary because feeding it any $(n-1)$-tuple of vectors $Y_1, \ldots, Y_{n-1}$ tangent to ∂M means evaluating $d\mathrm{vol}_M$ on $(Y, Y_1, \ldots, Y_{n-1})$, and those are *all* tangent to the $(n-1)$-dimensional boundary and thus cannot be linearly independent. We conclude

$$\iota_X d\mathrm{vol}_M |_{T(\partial M)} = \langle X, \nu \rangle \; \iota_\nu d\mathrm{vol}_M |_{T(\partial M)} = \langle X, \nu \rangle \, d\mathrm{vol}_{\partial M},$$

and the implication of (12.6) is thus

$$(12.7) \qquad \boxed{\int_M \mathrm{div}(X) \, d\mathrm{vol}_M = \int_{\partial M} \langle X, \nu \rangle \, d\mathrm{vol}_{\partial M}}.$$

This is a mild generalization of the classical result known as *Gauss's divergence theorem*.[43] Physics textbooks like to write their favorite special case of this result in some form such as

$$(12.8) \qquad \iiint_\Omega (\nabla \cdot \mathbf{X}) \, dV = \oiint_{\partial \Omega} \mathbf{X} \cdot d\mathbf{a},$$

where $\Omega \subset \mathbb{R}^3$ is assumed to be a compact region bounded by a smooth surface $\partial \Omega \subset \mathbb{R}^3$, $\nabla \cdot \mathbf{X}$ is the divergence of a vector field $\mathbf{X} \in \mathfrak{X}(\Omega)$ with respect to the standard volume form $d\mathrm{vol}_{\mathbb{R}^3} := dx \wedge dy \wedge dz$, the "V" in $dV := d\mathrm{vol}_{\mathbb{R}^3}$ stands for "volume" and the "a" in $\mathbf{X} \cdot d\mathbf{a} := \langle \mathbf{X}, \nu \rangle \, d\mathrm{vol}_{\partial \Omega}$ stands for "area". (The symbol $d\mathbf{a}$ in this situation is thought of as a "vector-valued measure" that encodes not only the 2-dimensional measure on $\partial \Omega$ but also its normal vector field.) The repetition of the integral signs corresponds to the dimension of the manifold and can be seen as a reference to Fubini's theorem; the additional loop in $\oiint$ merely refers to the fact that $\partial \Omega$ is a "closed" surface (the 2-dimensional analogue of a closed loop), i.e. it is compact and has no boundary. Gauss's theorem has an important interpretation in electrostatics: if $\mathbf{X}$ represents the electric field on a region $\Omega \subset \mathbb{R}^3$, then its divergence is the electrical charge density, and (12.8) thus says that the total electrical charge in the region Ω is equal to the total *flux* of the electric field through the boundary of Ω.

12.5.2. *Curl.* The next example only makes sense in the case

$$\dim M = 3.$$

It relies on the observation that for any n-dimensional vector space V with a nontrivial top-dimensional form $\omega \in \Lambda^n V^*$, the map

$$V \to \Lambda^{n-1} V^* : v \mapsto \iota_v \omega$$

is an isomorphism. Indeed, it is clearly injective since $\omega \neq 0$ and any $v \neq 0$ can be extended to a basis of V, so surjectivity then follows from the fact that $\dim \Lambda^{n-1} V^* = \binom{n}{n-1} = n = \dim V$. With this understood, any volume form $d\mathrm{vol}_M$ on a 3-manifold M determines an isomorphism

$$\mathfrak{X}(M) \xrightarrow{\cong} \Omega^2(M) : X \mapsto \iota_X d\mathrm{vol}_M.$$

Let us now assume (M, g) is an oriented Riemannian 3-manifold and $d\mathrm{vol}_M$ is its Riemannian volume form. The metric $\langle \, , \, \rangle := g$ also determines an isomorphism

$$\mathfrak{X}(M) \xrightarrow{\cong} \Omega^1(M) : X \mapsto X_\flat := \langle X, \cdot \rangle.$$

---

[43]or possibly "Gauss' divergence theorem", I don't know

The **curl** (*Rotation*) of $X \in \mathfrak{X}(M)$ is then defined as the unique vector field $\mathrm{curl}(X) \in \mathfrak{X}(M)$ such that
$$\iota_{\mathrm{curl}(X)} d\mathrm{vol}_M = d(X_\flat).$$

EXERCISE 12.17. Convince yourself that on $M := \mathbb{R}^3$ with its standard Riemannian metric defined via the Euclidean inner product, the curl of a vector field is the same thing that you learned about once upon a time in vector calculus.

Now if $\Sigma \subset M$ is an oriented 2-dimensional submanifold with boundary, Σ and $\partial \Sigma$ each inherit Riemannian metrics as submanifolds of M, and thus have canonical Riemannian volume forms $d\mathrm{vol}_\Sigma$ and $d\mathrm{vol}_{\partial\Sigma}$ respectively. For an appropriate choice[44] of normal vector field ν along Σ, Proposition 11.14 implies
$$d\mathrm{vol}_\Sigma = \iota_\nu d\mathrm{vol}_M|_{T\Sigma} \in \Omega^2(\Sigma),$$
and a repeat of the same argument we used for the divergence theorem then implies that for any $Y \in \mathfrak{X}(M)$,
$$\iota_Y d\mathrm{vol}_M|_{T\Sigma} = \langle Y, \nu \rangle \, d\mathrm{vol}_\Sigma.$$
If $Y = \mathrm{curl}(X)$ for some $X \in \mathfrak{X}(M)$, Stokes' theorem now implies
$$\int_\Sigma \langle \mathrm{curl}(X), \nu \rangle \, d\mathrm{vol}_\Sigma = \int_\Sigma d(X_\flat) = \int_{\partial\Sigma} X_\flat.$$

To understand the integral on the right, let $\tau \in \mathfrak{X}(\partial\Sigma)$ denote the unique positively-oriented unit vector field on $\partial\Sigma$, so $d\mathrm{vol}_{\partial\Sigma}(\tau) = 1$, and $X_\flat(\tau) = \langle X, \tau \rangle$ thus implies $X_\flat|_{T(\partial\Sigma)} = \langle X, \tau \rangle \, d\mathrm{vol}_{\partial\Sigma}$, and we obtain

(12.9)
$$\boxed{\int_\Sigma \langle \mathrm{curl}(X), \nu \rangle \, d\mathrm{vol}_\Sigma = \int_{\partial\Sigma} \langle X, \tau \rangle \, d\mathrm{vol}_{\partial\Sigma}}$$

This generalizes what is usually called the "classical" Stokes' theorem in vector calculus. In physics textbooks, one finds it written for the case $\Sigma \subset \mathbb{R}^3$ with the standard metric as
$$\iint_\Sigma (\nabla \times \mathbf{X}) \cdot d\mathbf{a} = \oint_{\partial\Sigma} \mathbf{X} \cdot d\mathbf{l},$$

where $\nabla \times \mathbf{X}$ denotes the curl of $\mathbf{X} \in \mathfrak{X}(\mathbb{R}^3)$, $d\mathbf{a}$ is the same "vector-valued measure" that appeared in (12.8), and $d\mathbf{l}$ similarly denotes a 1-dimensional vector-valued measure that encodes both the volume form $d\mathrm{vol}_{\partial\Sigma}$ and the tangent vector field τ.

13. Closed and exact forms

13.1. Some easy applications of Stokes. The following terminology is used consistently throughout differential geometry.

DEFINITION 13.1. A manifold M is **closed** (*geschlossen*) if it is compact and $\partial M = \varnothing$. We say that M is **open** (*offen*) if none of its connected components are closed, i.e. they all are noncompact and/or have nonempty boundary.[45]

---

[44]One can deduce from the assumption that both M and Σ are oriented that a normal vector field ν along Σ exists, and there are multiple choices—if Σ is connected, then there are exactly two choices, differing by a sign. The *appropriate* choice is the one that makes the volume form $\iota_\nu d\mathrm{vol}_M$ on Σ positive.

[45]Be aware that the word "closed" has a different meaning when referring to a manifold than it does when referring to a subset of a topological space. For instance, if M is a manifold, then a compact submanifold $\Sigma \subset M$ with boundary is a closed subset of M, but it is not a closed manifold if $\partial\Sigma \neq \varnothing$. The German language uses two different words for these separate meanings of "closed": a subset in a topological space can be *abgeschlossen*, but a manifold can be *geschlossen*.

EXAMPLE 13.2. Manifolds of dimension 0 never have boundary, so a 0-manifold is closed if and only if it is compact, i.e. it is a discrete finite set.

EXAMPLE 13.3. If M is a compact manifold with boundary, then ∂M is a closed manifold.

DEFINITION 13.4. A differential form $\omega \in \Omega^k(M)$ is called **closed** (*geschlossen*) if $d\omega = 0$, and it is called **exact** (*exakt*) of $\omega = d\alpha$ for some $\alpha \in \Omega^{k-1}(M)$. In the latter situation, the form α is called a **primitive** of ω.

EXAMPLE 13.5. A closed 0-form is the same thing as a locally constant function, and an exact 1-form is the same thing as a differential. There are no exact 0-forms since there is no such thing as a (-1)-form.

EXAMPLE 13.6. On an n-manifold, every n-form is closed since there are no nontrivial $(n+1)$-forms.

EXAMPLE 13.7. Given a volume form $d\mathrm{vol} \in \Omega^n(M)$, a vector field $X \in \mathfrak{X}(M)$ has vanishing divergence if and only if the $(n-1)$-form $\iota_X d\mathrm{vol}$ is closed. Similarly, if (M, g) is an oriented Riemannian 3-manifold, $X \in \mathfrak{X}(M)$ has vanishing curl if and only if the 1-form $X_\flat := g(X, \cdot)$ is closed.

Here is a bit of low-hanging fruit that can be picked as soon as one understands the above definitions and the statement of Stokes' theorem.

PROPOSITION 13.8. *If M is a closed oriented n-manifold and $\omega \in \Omega^n(M)$ is exact, then $\int_M \omega = 0$. Similarly, if M is a compact oriented n-manifold with boundary and $\alpha \in \Omega^{n-1}(M)$ is closed, then $\int_{\partial M} \alpha = 0$.*

PROOF. If you review the proof of Stokes' theorem, you will find that it is valid in the case $\partial M = \varnothing$ so long as one understands every integral over $\varnothing$ to be 0 by definition. Thus $\partial M = \varnothing$ and $\omega = d\beta$ for some $\beta \in \Omega^{n-1}(M)$ implies

$$\int_M \omega = \int_M d\beta = \int_\varnothing \beta = 0,$$

and if ∂M is not assumed empty but $\alpha \in \Omega^{n-1}(M)$ is closed,

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

$\square$

COROLLARY 13.9. *On a closed oriented n-manifold M, every n-form $\omega \in \Omega^n(M)$ with $\int_M \omega \neq 0$ is closed but not exact. In particular, this is true whenever ω is a volume form.* $\square$

REMARK 13.10. One can show that Corollary 13.9 fails whenever either $\partial M \neq \varnothing$ or M is noncompact. In the former case, $\int_M \omega \neq 0$ for an exact form $\omega = d\alpha$ is not a contradiction, since $\int_{\partial M} \alpha$ might also be nonzero. There is a different problem if M has empty boundary but is noncompact: the use of Stokes' theorem to derive the contradiction $0 \neq \int_M d\alpha = \int_{\partial M} \alpha = 0$ is not valid unless α has compact support, so it can happen for instance that $\omega \in \Omega^n_c(M)$ satisfies $\int_M \omega \neq 0$ and is the exterior derivative of an $(n-1)$-form whose support is noncompact. We will see shortly that, indeed, *every* n-form on $\mathbb{R}^n$ for $n \geqslant 1$ is exact (see Corollary 13.34 below).

EXERCISE 13.11. Show that for each $k \geqslant 0$, a k-form $\omega \in \Omega^k(M)$ is closed if and only for every compact oriented $(k+1)$-dimensional submanifold $L \subset M$ with boundary, $\int_{\partial L} \omega = 0$.
Hint: For any point $p \in M$ and linearly-independent vectors $X_1, \ldots, X_{k+1} \in T_p M$, you could choose $L \subset M$ to be a small $(k+1)$-disk through p tangent to the space spanned by $X_1, \ldots, X_{k+1}$.

13.2. The Poincaré lemma and simple connectedness. The observation in Example 13.3 that boundaries of compact manifolds are closed has a dual statement for differential forms: since $d^2 := d \circ d = 0$, every exact differential form is also closed. Corollary 13.9 reveals however that the converse is generally false. Here is a more concrete example.

EXAMPLE 13.12. On $\mathbb{R}^2\backslash\{0\}$, one can define a smooth 1-form in Cartesian coordinates (x, y) by

$$\lambda := \frac{1}{x^2 + y^2}\left(x\,dy - y\,dx\right).$$

This expression takes a more revealing form of one rewrites it in polar coordinates: assume $\mathcal{U} \subset \mathbb{R}^2\backslash\{0\}$ is a subset on which there is a well-defined chart of the form $(r, \theta) : \mathcal{U} \to \mathbb{R}^2$ such that r takes positive values and the relations $x = r\cos\theta$ and $y = r\sin\theta$ hold; concretely, we can take $\mathcal{U}$ to be the complement of a ray $\{tv \in \mathbb{R}^2 \mid t \in [0, \infty)\}$ for some $v \in \mathbb{R}^2\backslash\{0\}$, and the image of θ is then an open interval of the form $(c, c + 2\pi)$. In terms of r and θ, we have $dx = (\cos\theta)\,dr - (r\sin\theta)\,d\theta$ and $dy = (\sin\theta)\,dr + (r\cos\theta)\,d\theta$, thus

$$\lambda = \frac{1}{r^2}\left[r\cos\theta\left(\sin\theta\,dr + r\cos\theta\,d\theta\right) - r\sin\theta\left(\cos\theta\,dr - r\sin\theta\,d\theta\right)\right] = d\theta,$$

so λ is exact on $\mathcal{U}$. Since this computation holds independently of the choice of domain $\mathcal{U} \subset \mathbb{R}^2\backslash\{0\}$, it follows that $d\lambda = 0$ everywhere. But the restriction of $(\mathcal{U}, (r, \theta))$ to $\{r = 1\}$ now defines a chart on $S^1 \subset \mathbb{R}^2\backslash\{0\}$ in the form $(S^1\backslash\{q\}, \theta)$ for some point $q \in S^1$, which is a set of measure zero, thus $\int_{S^1} \lambda$ can be computed using the methods of §11.2, and the answer is

$$\int_{S^1} \lambda = \int_{(c, c+2\pi)} d\theta = 2\pi \neq 0.$$

This clearly could not happen if λ were df for some $f \in \Omega^0(\mathbb{R}^2\backslash\{0\}) = C^\infty(\mathbb{R}^2\backslash\{0\})$, as the restriction of λ to S^1 would then be $d(f|_{S^1})$ and we would have a contradiction to Proposition 13.8.

REMARK 13.13. It is conventional to denote the 1-form in Example 13.12 by

$$d\theta \in \Omega^1(\mathbb{R}^2\backslash\{0\})$$

even though, strictly speaking, it is not the differential of any smooth function $\theta : \mathbb{R}^2\backslash\{0\} \to \mathbb{R}$. One reasonable way to think about it is that while θ cannot be defined on this domain as a smooth *real*-valued function, it can be defined to take values in the quotient $\mathbb{R}/2\pi\mathbb{Z}$, which is a smooth manifold and $\theta : \mathbb{R}^2\backslash\{0\} \to \mathbb{R}/2\pi\mathbb{Z}$ in this sense is a smooth map. The latter means in practice that any point $p \in \mathbb{R}^2\backslash\{0\}$ admits a neighborhood $\mathcal{U} \subset \mathbb{R}^2\backslash\{0\}$ on which the smooth function $\theta : \mathcal{U} \to \mathbb{R}$ can be defined, though this function is not unique, as it can equally well be replaced by $\theta + 2\pi m$ for any $m \in \mathbb{Z}$. But modifying θ by addition of a constant does not change its differential, thus $d\theta$ is uniquely defined.

Remark 13.13 illustrates a phenomenon that is generalized in the following result: every closed differential form is "locally" exact.

THEOREM 13.14 (the Poincaré Lemma). *If $\omega \in \Omega^k(M)$ is closed and $k \geqslant 1$, then for every $p \in M$ there exists a neighborhood $\mathcal{U} \subset M$ of p and a $(k-1)$-form $\alpha \in \Omega^{k-1}(\mathcal{U})$ such that $d\alpha = \omega$ on $\mathcal{U}$.*

A proof of the Poincaré lemma will be given at the end of this lecture. The next two results are easier to prove, but imply a stronger statement for the case $k = 1$.

LEMMA 13.15. *A 1-form $\lambda \in \Omega^1(M)$ is exact if and only if $\int_{S^1} \gamma^*\lambda = 0$ for all smooth maps $\gamma : S^1 \to M$.*

PROOF. If $\lambda = df$ for some $f \in C^\infty(M)$, then Proposition 13.8 implies $\int_{S^1} \gamma^*\lambda = \int_{S^1} \gamma^* df = \int_{S^1} d(\gamma^* f) = 0$ for every smooth map $\gamma : S^1 \to M$. Conversely, assume $\int_{S^1} \gamma^*\lambda$ always vanishes. The following recipe for constructing a function $f : M \to \mathbb{R}$ with $df = \lambda$ can be applied on every connected component of M separately, so we may as well assume M is connected. We claim that if we fix a reference point $p_0 \in M$, then $f : M \to \mathbb{R}$ can be defined by

$$(13.1) \qquad f(p) := \int_0^a \lambda(\dot{\gamma}(t)) \, dt \quad \text{ for any } a > 0, \, \gamma \in C^\infty([0,a], M) \text{ with } \gamma(0) = p_0, \, \gamma(a) = p.$$

We must first show that $f(p)$ is independent of the choice of the path $\gamma : [0,a] \to M$ from p_0 to p. To this end, here are two useful observations: first, by the substitution rule, the integral in (13.1) does not change if we replace $\gamma : [0,a] \to M$ with $\gamma \circ \psi : [0,1] \to M$ for any smooth map $\psi : [0,1] \to [0,a]$ with $\psi(0) = 0$ and $\psi(1) = a$. As a consequence, we lose no generality by restricting our attention to paths $\gamma : [0,1] \to M$ that are constant on neighborhoods of 0 and 1, with values p_0 and p respectively. The second observation is that if t denotes the standard coordinate on the 1-manifold $[0,1] \subset \mathbb{R}$, then $(\gamma^*\lambda)_t(\partial_t) = \lambda_{\gamma(t)}(\gamma_* \partial_t) = \lambda_{\gamma(t)}(\dot{\gamma}(t))$, thus we can also write

$$f(p) = \int_{[0,1]} \gamma^*\lambda.$$

Now if $\gamma_1, \gamma_2 : [0,1] \to M$ are two smooth paths from p_0 to p that are both constant near 0 and 1, we can concatenate γ_1 with the reversal of γ_2 to form a smooth loop $\varphi : S^1 \to M$ in the form

$$\varphi(e^{\pi i t}) = \begin{cases} \gamma_1(t) & \text{for } 0 \leqslant t \leqslant 1, \\ \gamma_2(2 - t) & \text{for } 1 \leqslant t \leqslant 2, \end{cases}$$

where for convenience we are identifying $\mathbb{R}^2$ in the obvious way with $\mathbb{C}$ so that $S^1 \subset \mathbb{C}$. If we now split S^1 into its upper and lower semicircles $S^1_\pm$ with parametrizations $\psi_\pm : [0,1] \to S^1_\pm : t \mapsto e^{\pi i t}$, we have $\gamma_1 = \varphi \circ \psi_+$ and $\gamma_2 = \varphi \circ \psi_-$, but ψ_+ is orientation preserving while ψ_- is orientation reversing, thus

$$0 = \int_{S^1} \varphi^*\lambda = \int_{S^1_+} \varphi^*\lambda + \int_{S^1_-} \varphi^*\lambda = \int_{\psi_+([0,1])} \varphi^*\lambda + \int_{\psi_-([0,1])} \varphi^*\lambda$$

$$= \int_{[0,1]} \psi_+^* \varphi^*\lambda - \int_{[0,1]} \psi_-^* \varphi^*\lambda = \int_{[0,1]} (\varphi \circ \psi_+)^*\lambda - \int_{[0,1]} (\varphi \circ \psi_-)^*\lambda = \int_{[0,1]} \gamma_1^*\lambda - \int_{[0,1]} \gamma_2^*\lambda.$$

With independence of the choice of γ established, we observe that (13.1) implies $\frac{d}{dt} f(\gamma(t)) = \lambda(\dot{\gamma}(t))$ for every t and every smooth path γ starting at p_0, thus $df = \lambda$. $\qquad\square$

EXERCISE 13.16. Use a slight modification of the proof of Lemma 13.15 to show that on S^1, a 1-form $\lambda \in \Omega^1(S^1)$ is exact if and only if $\int_{S^1} \lambda = 0$.

DEFINITION 13.17. A smooth manifold M is **simply connected** (*einfach zusammenhängend*) if it is connected and every smooth map $\gamma : S^1 \to M$ admits a smooth extension over the 2-disk, i.e. a map $u : \mathbb{D}^2 \to M$ such that $u|_{\partial \mathbb{D}^2} = \gamma$.

REMARK 13.18. In algebraic topology, a topological space is called simply connected if it is path-connected and its fundamental group vanishes, but for smooth manifolds, Definition 13.17 is equivalent to this condition. In particular, one could replace the word "smooth" by "continuous" without changing anything, because by general perturbation results in differential topology (see e.g. [Hir94]), continuous maps between smooth manifolds always admit smooth approximations.

THEOREM 13.19. *If M is a simply connected manifold, then every closed 1-form $\lambda \in \Omega^1(M)$ is exact.*

PROOF. If $\lambda \in \Omega^1(M)$ is closed and every smooth map $\gamma : S^1 \to M$ admits a smooth extension $u : \mathbb{D}^2 \to M$, then

$$\int_{S^1} \gamma^* \lambda = \int_{\partial \mathbb{D}^2} u^* \lambda = \int_{\mathbb{D}^2} d(u^* \lambda) = \int_{\mathbb{D}^2} u^*(d\lambda) = 0,$$

hence λ satisfies the criterion of Lemma 13.15 and is therefore exact. □

It should be easy to convince yourself that every convex subset of $\mathbb{R}^n$ is simply connected, and every point in a manifold has a neighborhood that looks like a convex subset of $\mathbb{R}^n$ in local coordinates, implying in turn that that neighborhood is simply connected. Theorem 13.19 thus implies the $k = 1$ case of the Poincaré lemma. But it also implies more, because there are many simply connected manifolds that are more interesting than convex sets.

EXAMPLE 13.20. For each $n \geqslant 2$, the sphere S^n is simply connected. Here is an incomplete but (maybe?) believable proof: since $\dim S^n > \dim S^1$, no smooth map $\gamma : S^1 \to S^n$ can be surjective,[46] i.e. it must miss at least one point $p \in S^n$ and can thus be viewed as a map $S^1 \to S^n \backslash \{p\}$. But by stereographic projection, one can also find a diffeomorphism of $S^n \backslash \{p\}$ to $\mathbb{R}^n$ and then appeal to the fact that $\mathbb{R}^n$ (as a convex set) is simply connected. It follows that closed 1-forms on S^n for $n \geqslant 2$ are always exact.

REMARK 13.21. You may have noticed that in Theorem 13.19, it would have sufficed to assume that every smooth map $\gamma : S^1 \to M$ admits a smooth extension $u : \Sigma \to M$ over *some* compact, smooth, oriented surface Σ with boundary $\partial \Sigma = S^1$, i.e. not necessarily the disk, but any surface whose boundary is a circle. (An easy example would be obtained by cutting a hole out of the 2-torus $\mathbb{T}^2$.) This means that Theorem 13.19 is true under a somewhat more general hypothesis than simple connectedness. The natural language for this generalization is homology, i.e. the theorem holds for any manifold M whose first homology group with real coefficients vanishes. A full explanation of this statement would require a major digression into algebraic topology, so we will not discuss it any further here, but suffice it to say that in dimension 2, there are no examples for which this distinction makes a difference, but in dimension 3 there are. Poincaré famously conjectured that every closed 3-manifold with vanishing first homology group is homeomorphic to S^3, but later found an example—now known as the *Poincaré homology sphere*—that satisfies this hypothesis but (unlike S^3) is not simply connected, and thus had to revise his conjecture. The revised conjecture was proved over 100 years later.

EXAMPLE 13.22. On a Riemannian manifold (M, g), the inner product $\langle \, , \, \rangle := g$ determines an isomorphism $T_p M \to T_p^* M : X \mapsto X_\flat := \langle X, \cdot \rangle$ at every point $p \in M$, which can be used to associate to any smooth function $f : M \to \mathbb{R}$ its **gradient** vector field $\nabla f \in \mathfrak{X}(M)$, uniquely determined by

$$df = \langle \nabla f, \cdot \rangle.$$

A vector field $X \in \mathfrak{X}(M)$ cannot be the gradient of a function unless the 1-form $X_\flat \in \Omega^1(M)$ is closed, and conversely, the Poincaré lemma implies that every vector field satisfying this condition is *locally* the gradient of a function, though perhaps not globally (unless M is simply connected).

---

[46]I'm pretty sure that you cannot visualize any surjective smooth map $f : M \to N$ when $\dim M < \dim N$, though actually proving they don't exist is not completely trivial. It follows easily from Sard's theorem, a fundamental result in differential topology stating that the set of critical values of a smooth map $f : M \to N$ always has measure zero. This means that for almost every $q \in N$, $T_p f : T_p M \to T_q N$ is surjective for every $p \in f^{-1}(q)$; the only way for this to hold when $\dim M < \dim N$ is if $f^{-1}(q) = \varnothing$. The much more surprising fact is that *continuous* maps $f : M \to N$ can be surjective, even when $\dim N > \dim M$; look up the term "space-filling curve". Such maps can never be smooth.

If M is oriented and 3-dimensional, then this result can also be expressed in terms of the curl (cf. §12.5.2): any gradient $X = \nabla f$ satisfies $\iota_{\mathrm{curl}(X)} d\mathrm{vol}_M = d(df) = 0$, implying

$$\mathrm{curl}(\nabla f) \equiv 0,$$

and conversely, any vector field $X \in \mathfrak{X}(M)$ with $\mathrm{curl}(X) \equiv 0$ is locally the gradient of a function.

In the same context, the curl of any vector field $X \in \mathfrak{X}(M)$ satisfies $\iota_{\mathrm{curl}(X)} d\mathrm{vol}_M = d(X_\flat)$ and thus $d(\iota_{\mathrm{curl}(X)} d\mathrm{vol}_M) = d^2(X_\flat) = 0$, implying

$$\mathrm{div}(\mathrm{curl}(X)) \equiv 0.$$

Conversely, any divergenceless vector field $Y \in \mathfrak{X}(M)$ satisfies $d(\iota_Y d\mathrm{vol}_M) = 0$, so that by the Poincaré lemma, $\iota_Y d\mathrm{vol}_M \in \Omega^2(M)$ can be written on any sufficienly small neighborhood $\mathcal{U}$ as $d\lambda$ for some $\lambda \in \Omega^1(\mathcal{U})$. The latter is also $X_\flat$ for a unique vector field $X \in \mathfrak{X}(\mathcal{U})$, whose curl is therefore Y: in other words, any divergenceless vector field is locally the curl of another vector field.

While (13.1) provides a fairly straightforward recipe to find a local primitive of any closed 1-form, it is not as easy to derive local primitives for closed k-forms when $k \geqslant 2$. One possible approach is to work on "boxes" of the form $M := (a_1, b_1) \times \ldots \times (a_n, b_n)$ and proceed by induction on the number of dimensions, showing that if one can already find primitives for closed k-forms on the hypersurface $\Sigma_c := (a_1, b_1) \times \ldots \times (a_{n-1}, b_{n-1}) \times \{c\}$ for some constant $c \in (a_n, b_n)$, then primitives on Σ_c can be extended to primitives on M by integrating in the nth direction. I have proved the Poincaré lemma in this way when I've taught analysis courses (see [**Wen19**]), but the idea behind the argument has a tendency to get lost behind computational details. We will adopt a different approach in these notes, and deduce the Poincaré lemma from a deeper theorem about the homotopy-invariance of de Rham cohomology. We will see at the end that this approach does lead to an explicit formula generalizing (13.1) to produce local primitives of closed k-forms (see in partiular Remark 13.39), but in contrast with (13.1), one would be very unlikely to find this formula from an educated guess.

13.3. De Rham cohomology. By now we have gathered some evidence that the distinction between closed and exact forms on a manifold M has something to do with the topology of M. We shall now formalize this relation by defining an algebraic invariant of smooth manifolds.

DEFINITION 13.23. For a smooth n-manifold M and each integer $k \in \mathbb{Z}$, let $d_k : \Omega^k(M) \to \Omega^{k+1}(M)$ denote the restriction of the exterior derivative $d : \Omega^*(M) \to \Omega^*(M)$ to the subspace $\Omega^k(M) \subset \Omega^*(M)$, with the convention that for $k < 0$, $\Omega^k(M)$ is the trivial subspace (hence d_{-1} is the trivial map into $\Omega^0(M)$). The k**th de Rham cohomology** of M is the vector space

$$H_{\mathrm{dR}}^k(M) := \ker(d_k) \big/ \mathrm{im}(d_{k-1}),$$

i.e. it is the quotient of the space of closed k-forms by the subspace of exact k-forms. We write

$$H_{\mathrm{dR}}^*(M) := \bigoplus_{k \in \mathbb{Z}} H_{\mathrm{dR}}^k(M).$$

REMARK 13.24. The case $k < 0$ was included in Definition 13.23 only in order to make sure that the definition of $H_{\mathrm{dR}}^0(M)$ makes sense, but $H_{\mathrm{dR}}^k(M)$ for $k < 0$ is just the trivial vector space, and we will have no need to mention it again. It is similarly easy to see that $H_{\mathrm{dR}}^k(M) = 0$ whenever $k > \dim M$, since the space of k-forms is already trivial in this case. Thus in practice, $H_{\mathrm{dR}}^k(M)$ is potentially interesting only for k in the range $0 \leqslant k \leqslant \dim M$.

It may seem surprising at first glance that $H_{\mathrm{dR}}^k(M)$ is useful or computable: in typical cases both $\ker(d_k)$ and $\mathrm{im}(d_{k-1})$ are infinite-dimensional vector spaces, and one would not normally expect the quotient of one infinite-dimensional space by another one to carry interesting information.

It turns out however that in almost all interesting cases, the quotient is finite dimensional, and its dimension is a useful numerical invariant of manifolds. Let us first clarify what is meant by the word "invariant".

PROPOSITION 13.25. *For smooth maps* $f : M \to N$, *the linear map* $f^* : \Omega^k(N) \to \Omega^k(M)$ *sends closed forms on* N *to closed forms on* M, *and it also descends*[47] *to the quotients to define a linear map* $f^* : H_{\mathrm{dR}}^k(N) \to H_{\mathrm{dR}}^k(M)$ *that satisfies the following properties:*

(1) For another smooth map $g : N \to Q$, $(g \circ f) = f^* g^* : H_{\mathrm{dR}}^k(Q) \to H_{\mathrm{dR}}^k(M)$;

(2) For the identity map $\mathrm{Id} : M \to M$, $\mathrm{Id}^* : H_{\mathrm{dR}}^k(M) \to H_{\mathrm{dR}}^k(M)$ *is the identity map.*

It follows in particular that whenever $f : M \to N$ *is a diffeomorphism,* $f^* : H_{\mathrm{dR}}^k(N) \to H_{\mathrm{dR}}^k(M)$ *is a vector space isomorphism for each* k.

PROOF. The relation $f^*(d\omega) = d(f^*\omega)$ implies that f^* preserves both the spaces of closed forms and exact forms, and thus descends to their quotient. The rest of the statement follows immediately from the basic properties of pullbacks. $\square$

REMARK 13.26. For those who enjoy this kind of language, Proposition 13.25 says that H_{dR}^k for each $k \in \mathbb{Z}$ defines a contravariant functor from the category of smooth manifolds and smooth maps to the category of real vector spaces and linear maps.

EXAMPLE 13.27. The closed 0-forms on M are the locally constant functions, which can take independent but constant values on each connected component of M, while the subspace of exact 0-forms is trivial, thus if M has $N \in \mathbb{N}$ connected components, $H_{\mathrm{dR}}^0(M) \cong \mathbb{R}^N$.

EXAMPLE 13.28. If $M := \{\mathrm{pt}\}$ is the 0-manifold consisting of a single point, then $\Omega^0(\{\mathrm{pt}\}) \cong \mathbb{R}$, $\Omega^k(\{\mathrm{pt}\}) = 0$ for each $k > 0$, and the exterior derivative is the trivial map, implying

$$H_{\mathrm{dR}}^k(\{\mathrm{pt}\}) \cong \begin{cases} \mathbb{R} & \text{for } k = 0, \\ 0 & \text{for } k > 0. \end{cases}$$

EXAMPLE 13.29. Theorem 13.19 implies that $H_{\mathrm{dR}}^1(M) = 0$ whenever M is simply connected.

EXAMPLE 13.30. Corollary 13.9 implies that $H_{\mathrm{dR}}^n(M) \neq 0$ whenever M is a closed oriented n-manifold.

Diffeomorphism-invariance is a nice property, but de Rham cohomology also satisfies a stronger invariance property that makes it much easier to compute.

DEFINITION 13.31. Two smooth maps $f_0, f_1 : M \to N$ are called **smoothly homotopic** (*glatt homotop*) if there exists a smooth map $h : [0, 1] \times M \to N$ such that $h(0, \cdot) = f_0$ and $h(1, \cdot) = f_1$.

THEOREM 13.32. *If* $f_0, f_1 : M \to N$ *are smoothly homotopic maps, then for each* k, *the linear maps* $H_{\mathrm{dR}}^k(N) \to H_{\mathrm{dR}}^k(M)$ *defined by* f_0^* *and* f_1^* *are identical.*

Before proving this, let's think through some of the consequences. A map $f : M \to N$ is called **a smooth homotopy equivalence** (*glatte Homotopieäquivalenz*) if there exists another smooth map $g : N \to M$ such that $f \circ g : N \to N$ and $g \circ f : M \to M$ are each smoothly homotopic to the identity map. Combining Proposition 13.25 with Theorem 13.32 in this situation implies that $f^* : H_{\mathrm{dR}}^*(N) \to H_{\mathrm{dR}}^*(M)$ and $g^* : H_{\mathrm{dR}}^*(M) \to H_{\mathrm{dR}}^*(N)$ are inverses; in particular, f^* is an isomorphism:

---

[47] Recall that if $A : V \to W$ is a linear map between vector spaces and $X \subset V$ and $Y \subset W$ are linear subspaces such that $A(X) \subset Y$, then there is a well-defined linear map $V/X \to W/Y$ sending the equivalence class $[x] \in V/X$ of each $x \in V$ to the equivalence class $[Ax] \in W/Y$ of $Ax \in W$. One says in this situation that $A : V \to W$ **descends** to a map $V/X \to W/Y$.

Corollary 13.33. *If two manifolds M and N are smoothly homotopy equivalent, then their de Rham cohomologies are isomorphic.*   □

The power of Corollary 13.33 lies in the fact that two manifolds can easily be homotopy equivalent without being diffeomorphic; in fact, homotopy equivalence does not even imply that they have the same dimension. Here is an extreme example: a manifold M is called **smoothly contractible** (*glatt zusammenziehbar*) if there exists a smooth homotopy of the identity map $M \to M$ to a constant map. It is easy to see for instance that $\mathbb{R}^n$ is smoothly contractible, and so is any convex subset of $\mathbb{R}^n$. Given a smooth homotopy $h : [0,1] \times M \to M$ with $h(1,\cdot) = \mathrm{Id}_M$ and $h(0,\cdot) \equiv p \in M$ for some fixed point $p \in M$, consider the maps

$$\pi : M \to \{p\}, \qquad i : \{p\} \hookrightarrow M,$$

where π is the unique map and i is the natural inclusion. Now $\pi \circ i$ is the identity map on $\{p\}$, and $i \circ \pi : M \to M$ is $h(0,\cdot)$, which is therefore smoothly homotopic to Id_M. This proves that M is smoothly homotopy equivalent to the one-point manifold $\{p\}$, so combining Corollary 13.33 with Example 13.28 gives:

Corollary 13.34. *If M is smoothly contractible, then $H_{\mathrm{dR}}^k(M) = 0$ for all $k > 0$ and $H_{\mathrm{dR}}^0(M) \cong \mathbb{R}$.*

Proof of the Poincaré lemma. Every point $p \in M$ has a neighborhood $\mathcal{U} \subset M$ that looks like a convex set in some coordinate chart and is thus smoothly contractible. For $k > 0$, it now follows from $H_{\mathrm{dR}}^k(\mathcal{U}) = 0$ that the spaces of closed and exact k-forms on $\mathcal{U}$ are identical.   □

Proof of Theorem 13.32. We assume $h : [0,1] \times M \to N$ satisfies $h(0,\cdot) = f_0$ and $h(1,\cdot) = f_1$. Given $\omega \in \Omega^k(N)$, let us assume $L \subset M$ is a compact oriented k-dimensional submanifold with boundary and consider the integral of $h^* d\omega \in \Omega^{k+1}([0,1] \times M)$ over the domain $[0,1] \times L$. Note that the latter is not a smooth manifold with boundary unless $\partial L = \varnothing$; in general $[0,1] \times L$ can be understood as a manifold with boundary *and corners*. Nonetheless, one can make sense of Stokes' theorem on this domain as described in Example 12.14, leading to the relation

$$\int_{[0,1] \times L} h^*(d\omega) = \int_{[0,1] \times L} d(h^*\omega) = \int_{\partial([0,1] \times L)} h^*\omega := \int_{\partial[0,1] \times L} h^*\omega - \int_{[0,1] \times \partial L} h^*\omega$$

(13.2)
$$= \int_{\{1\} \times L} h^*\omega - \int_{\{0\} \times L} h^*\omega - \int_{[0,1] \times \partial L} h^*\omega$$

$$= \int_L f_1^*\omega - \int_L f_0^*\omega - \int_{[0,1] \times \partial L} h^*\omega,$$

where in the last line we have used the obvious identifications of $\{1\} \times L$ and $\{0\} \times L$ with L, so that the restrictions of $h^*\omega$ to these two submanifolds become $f_1^*\omega$ and $f_0^*\omega$ respectively. Now observe that for any compact oriented m-dimensional submanifold $Q \subset M$ and an $(m+1)$-form $\alpha \in \Omega^{m+1}(N)$, there is a natural way of presenting $\int_{[0,1] \times Q} h^*\alpha$ as the integral of an m-form over Q: we define $P\alpha \in \Omega^m(M)$ namely via the formula

$$(P\alpha)_p(X_1, \ldots, X_m) := \int_0^1 (h^*\alpha)_{(t,p)}(\partial_t, X_1, \ldots, X_m)\, dt \in \mathbb{R},$$

where ∂_t here denotes the obvious unit vector field on $[0,1] \times M$ pointing in the positive direction on the first factor, and each $X_1, \ldots, X_m \in T_p M$ is regarded as living in the subspace $\{0\} \times T_p M \subset T_t[0,1] \times T_p M = T_{(t,p)}([0,1] \times M)$. In this way we have defined a linear operator

$$P : \Omega^{m+1}(N) \to \Omega^m(M) \qquad \text{such that} \qquad \int_{[0,1] \times Q} h^*\alpha = \int_Q P\alpha$$

for all $\alpha \in \Omega^{m+1}(N)$ and compact oriented m-dimensional submanifolds $Q \subset M$. We can use this to transform (13.2) into the relation

$$\int_L (f_1^* \omega - f_0^* \omega) = \int_L P(d\omega) + \int_{\partial L} P\omega = \int_L [P(d\omega) + d(P\omega)],$$

where we have again applied Stokes' theorem to transform the integral over ∂L into one over L. We now have an equality of the integrals of two k-forms over an arbitrary compact oriented k-dimensional submanifold with boundary: in particular, one could pick any point $p \in M$ and any vectors $X_1, \ldots, X_k \in T_p M$ and then approximate the evaluation of both k-forms on $(X_1, \ldots, X_k)$ arbitrarily well by integrating them over a submanifold L that is chosen to be a small k-disk through p tangent to the space spanned by $X_1, \ldots, X_k$. The conclusion is that these two k-forms must be identical, so we have proved that $f_1^* \omega - f_0^* \omega = P(d\omega) + d(P\omega)$, or rewriting it as an equality between two linear maps $H_{\mathrm{dR}}^k(N) \to H_{\mathrm{dR}}^k(M)$,

$$(13.3) \qquad\qquad\qquad\qquad f_1^* - f_0^* = P \circ d + d \circ P.$$

This formula is well known in homological algebra: it is called the **chain homotopy relation**, and the operator $P : \Omega^*(N) \to \Omega^*(M)$ of degree -1 is consequently called a **chain homotopy** (*Kettenhomotopie*). Its existence has the following consequence: if $\omega \in \Omega^k(N)$ is closed, then

$$f_1^* \omega = f_0^* \omega + d(P\omega),$$

implying that $f_1^* \omega$ and $f_0^* \omega$ represent the same element in the quotient $H_{\mathrm{dR}}^k(M)$. $\square$

EXERCISE 13.35. Suppose $\mathcal{O}$ is an open subset of either $\mathbb{H}^n$ or $\mathbb{R}^n$. We call $\mathcal{O}$ a **star-shaped** domain if for every $p \in \mathcal{O}$, it also contains the points $tp \in \mathbb{R}^n$ for all $t \in [0, 1]$. It follows that $h(t, p) := tp$ defines a smooth homotopy $h : [0, 1] \times \mathcal{O} \to \mathcal{O}$ between the identity and the constant map whose value is the origin, making $\mathcal{O}$ smoothly contractible. Use this homotopy to extract from the proof of Theorem 13.32 an explicit formula for a linear operator $P : \Omega^k(\mathcal{O}) \to \Omega^{k-1}(\mathcal{O})$ for each $k \geq 1$ satisfying

$$\omega = P(d\omega) + d(P\omega)$$

for all $\omega \in \Omega^k(\mathcal{O})$. In particular, whenever ω is a closed k-form, $P\omega$ is a primitive of ω. (As a sanity check, a formula for P is given in Remark 13.39 at the end of this lecture, but try to derive it without knowing it in advance.)

One further property of $H_{\mathrm{dR}}^*(M)$ deserves to be mentioned, though a full explanation of it would fall far outside the scope of this course. By a result known as *de Rham's theorem*, $H_{\mathrm{dR}}^k(M)$ is naturally isomorphic to another invariant that is a standard topic in algebraic topology, namely the kth *singular cohomology* with real coefficients:

$$H_{\mathrm{dR}}^k(M) \cong H^k(M; \mathbb{R}).$$

The latter is defined for all topological spaces, not just smooth manifolds, and spaces that are *homeomorphic* always have isomorphic singular cohomologies, implying that $H_{\mathrm{dR}}^k(M)$ is actually a topological invariant. The topological invariance of $H_{\mathrm{dR}}^*(M)$ cannot be seen directly from its definition, since pullbacks of differential forms via maps $f : M \to N$ do not make sense in general when f is continuous but not differentiable. As one learns in algebraic topology, $H^k(M; \mathbb{R})$ is often surprisingly easy to compute, and for instance when M is compact, it can be derived from a finite-dimensional chain complex, implying the highly non-obvious fact that

$$\dim H_{\mathrm{dR}}^k(M) < \infty$$

whenever M is compact.

EXERCISE 13.36. Here is the most basic computation of $H^*_{\text{dR}}(M)$ for a non-contractible manifold: we will show in this exercise that for every $n \in \mathbb{N}$ and $k \in \{0, \dots, n\}$,

$$(13.4) \qquad\qquad \dim H^k_{\text{dR}}(S^n) = \begin{cases} 1 & \text{if } k = 0 \text{ or } k = n, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly every sphere S^n for $n \geqslant 1$ is connected,[48] so Example 13.27 establishes $H^0_{\text{dR}}(S^n) \cong \mathbb{R}$. For the computation of $H^k_{\text{dR}}(S^n)$ when $k \geqslant 1$, we proceed by induction on n.

(a) Show that if M is a closed oriented n-manifold, then there is a well-defined linear map

$$(13.5) \qquad\qquad H^n_{\text{dR}}(M) \to \mathbb{R} : [\omega] \mapsto \int_M \omega,$$

and the following conditions are equivalent:
  (i) $H^n_{\text{dR}}(M) \cong \mathbb{R}$;
  (ii) The map (13.5) is an isomorphism;
  (iii) Every $\omega \in \Omega^n(M)$ satisfying $\int_M \omega = 0$ is exact.

(b) Deduce via Exercise 13.16 that (13.4) is correct for $n = 1$.

(c) Suppose M is a closed n-manifold and ω_+, ω_- is a pair of k-forms on $M \times [-1, 1]$ such that $d\omega_+ = d\omega_-$. Show that the following conditions are equivalent:
  (i) $\omega_+ - \omega_-$ is exact;
  (ii) $i_t^* \omega_+ - i_t^* \omega_-$ is an exact k-form on M for every $t \in [-1, 1]$, where $i_t : M \hookrightarrow M \times [-1, 1]$ denotes the inclusion $p \mapsto (p, t)$.
  (iii) There exists a k-form ω on $M \times [-1, 1]$ which matches $\omega_\pm$ near $M \times \{\pm 1\}$ and satisfies $d\omega = d\omega_+ = d\omega_-$.
  Hint: First prove the equivalence of (i) and (ii), after convincing yourself that $i_t : M \hookrightarrow M \times [-1, 1]$ is a smooth homotopy equivalence for each t.

(d) Under the same assumptions as in part (c), suppose also that M is oriented and $k = n$. Show that the number $\int_{M \times \{t\}} \omega_+ - \int_{M \times \{t\}} \omega_- \in \mathbb{R}$ is the same for any choice of $t \in [-1, 1]$.
  Hint: Given $-1 \leqslant t_- < t_+ \leqslant 1$, integrate something over $M \times [t_-, t_+]$ and apply Stokes' theorem.

(e) Now given an integer $n \geqslant 2$, assume (13.4) is true for S^{n-1}, and fix $k \in \{1, \dots, n\}$. Regarding S^n as the unit sphere in $\mathbb{R}^{n+1}$ with standard coordinates $(x^1, \dots, x^{n+1})$, we can decompose it into two overlapping n-dimensional disks $S^n = D_+ \cup D_-$ whose intersection looks like $S^{n-1} \times [-1, 1]$; specifically, define
  $$D_+ := \{x^1 \geqslant -1/2\} \cap S^n, \qquad D_- := \{x^1 \leqslant 1/2\} \cap S^n.$$

  Take a moment to convince yourself that there is a diffeomorphism $D_+ \cap D_- \cong S^{n-1} \times [-1, 1]$. Observe next that D_+ and D_- are each smoothly contractible, thus any closed k-form ω on S^n will then by exact over each of D_+ and D_-, giving $\alpha_\pm \in \Omega^{k-1}(D_\pm)$ such that $d\alpha_\pm = \omega$ on $D_\pm$. The difficulty is that α_+ and α_- need not match on $D_+ \cap D_-$. Use the inductive hypothesis and the previous steps in this problem to show that if either $1 \leqslant k \leqslant n-1$ or $k = n$ with $\int_{S^n} \omega = 0$, then there exists $\alpha \in \Omega^{k-1}(S^n)$ satisfying $d\alpha = \omega$; show in fact that α can be chosen to match $\alpha_\pm$ on the portions of $D_\pm$ where D_+ and D_- do not overlap. This completes the inductive proof of (13.4).
  Hint: The case $k = n$ is trickiest, as you need to use the hypothesis $\int_{S^n} \omega = 0$ to deduce something about α_+ and α_-. What can you say about the integrals of $\alpha_\pm$ over the "equator" $S^{n-1} \cong \{x^1 = 0\} \subset S^n$? Try Stokes' theorem, but be careful with orientations!

---

[48]The 0-sphere is a discrete set of two points $S^0 = \{1, -1\} \subset \mathbb{R}$, and is thus not connected. That's why we excluded the case $n = 0$ from (13.4).

EXERCISE 13.37. Show that the wedge product descends to an associative and graded-commutative product $\cup : H_{\mathrm{dR}}^k(M) \times H_{\mathrm{dR}}^\ell(M) \to H_{\mathrm{dR}}^{k+\ell}(M)$, defined by

$$[\alpha] \cup [\beta] := [\alpha \wedge \beta].$$

This is called the **cup product** on de Rham cohomology.

Remark: There is similarly a cup product on singular cohomology, to which this one is isomorphic via de Rham's theorem. But this one is easier to define, and is thus often used in practice as a surrogate for the singular cup product.

EXERCISE 13.38. For this exercise, identify the n-torus $\mathbb{T}^n$ with the quotient $\mathbb{R}^n/\mathbb{Z}^n$ (recall from Exercise 3.4 that there is a natural diffeomorphism). For any sufficiently small open set $\widetilde{\mathcal{U}} \subset \mathbb{R}^n$, the usual Cartesian coordinates $x^1, \ldots, x^n : \widetilde{\mathcal{U}} \to \mathbb{R}$ can be used to define a smooth chart $(\mathcal{U}, x)$ on $\mathbb{T}^n$ where

$$\mathcal{U} := \left\{ [p] \in \mathbb{T}^n \;\middle|\; p \in \widetilde{\mathcal{U}} \right\}, \qquad x([p]) := (x^1(p), \ldots, x^n(p)) \text{ for } p \in \widetilde{\mathcal{U}}.$$

(a) Show that the coordinate differentials $dx^1, \ldots, dx^n \in \Omega^1(\mathcal{U})$ arising from the chart $(\mathcal{U}, x)$ described above are independent of the choice of the set $\widetilde{\mathcal{U}} \subset \mathbb{R}^n$, i.e. the definitions of the coordinate differentials obtained from two different choices $\widetilde{\mathcal{U}}_1, \widetilde{\mathcal{U}}_2 \subset \mathbb{R}^n$ coincide on the region $\mathcal{U}_1 \cap \mathcal{U}_2 \subset \mathbb{T}^n$ where they overlap.

(b) As a consequence of part (a), the 1-forms $dx^1, \ldots, dx^n \in \Omega^1(\mathbb{T}^n)$ are well-defined on the entire torus, and they are obviously locally exact and therefore closed, but they might not actually be exact since none of the coordinates $x^1, \ldots, x^n$ admit smooth definitions globally on $\mathbb{T}^n$. (This is another example of the phenomenon we saw with $d\theta \in \Omega^1(\mathbb{R}^2 \backslash \{0\})$ in Remark 13.13.) Show in fact that for any vector $(a_1, \ldots, a_n) \in \mathbb{R}^n \backslash \{0\}$, the 1-form

$$\lambda := a_i \, dx^i \in \Omega^1(\mathbb{T}^n)$$

is closed but not exact.
Hint: You only need to find one smooth map $\gamma : S^1 \to \mathbb{T}^n$ such that $\int_{S^1} \gamma^* \lambda \neq 0$.

(c) One can similarly produce closed k-forms $\omega \in \Omega^k(\mathbb{T}^n)$ for any $k \leqslant n$ by choosing constants $a_{i_1 \ldots i_k} \in \mathbb{R}$ and writing

$$(13.6) \qquad \omega = \sum_{i_1 < \ldots < i_k} a_{i_1 \ldots i_k} \, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \in \Omega^k(\mathbb{T}^n).$$

Show that for every nontrivial k-form of this type, one can find a cohomology class $[\alpha] \in H_{\mathrm{dR}}^{n-k}(\mathbb{T}^n)$ such that the cup product $[\omega] \cup [\alpha] \in H_{\mathrm{dR}}^n(\mathbb{T}^n)$ defined in Exercise 13.37 is nontrivial, and deduce from this that ω is not exact.
Hint: Can you choose $\alpha \in \Omega^{n-k}(\mathbb{T}^n)$ so that $\omega \wedge \alpha$ is a volume form?

Remark: One can show that all cohomology classes in $H_{\mathrm{dR}}^k(\mathbb{T}^n)$ are representable by k-forms with constant coefficients as in (13.6), thus $\dim H_{\mathrm{dR}}^k(\mathbb{T}^n) = \binom{n}{k}$.

REMARK 13.39. Here is a formula for the operator $P : \Omega^k(\mathcal{O}) \to \Omega^{k-1}(\mathcal{O})$ promised in Exercise 13.35 on a star-shaped domain $\mathcal{O}$ in $\mathbb{H}^n$ or $\mathbb{R}^n$:

$$(P\omega)_p(X_1, \ldots, X_{k-1}) := \int_0^1 t^{k-1} \omega_{tp}(p, X_1, \ldots, X_{k-1}) \, dt,$$

where since $\mathcal{O}$ is a subset of $\mathbb{R}^n$, we are using the natural isomorphisms $T_p\mathcal{O} = \mathbb{R}^n$ at every point. (Otherwise the expression $\omega_{tp}(p, X_1, \ldots, X_{k-1})$ would not generally make sense because $X_1, \ldots, X_{k-1} \in T_p\mathcal{O} \neq T_{tp}\mathcal{O}$.) In applications, it is occasionally useful to observe that $P\omega$ depends *continuously* on ω, i.e. one obtains in this way a continuous right-inverse of the operator $d_{k-1} : \Omega^{k-1}(\mathcal{O}) \to \mathrm{im}(d_{k-1}) \subset \Omega^k(\mathcal{O})$.

14. Volume-preserving and symplectic maps

14.1. Volume-preserving flows. Assume M is an oriented n-manifold with a fixed positive volume form $d\text{vol} \in \Omega^n(M)$. In §12.5, we defined the divergence of a vector field $X \in \mathfrak{X}(M)$ in this context as the unique function $\text{div}(X) : M \to \mathbb{R}$ such that

$$d(\iota_X d\text{vol}) = \text{div}(X) \cdot d\text{vol}.$$

A partial justification for this definition was furnished by the Gauss divergence theorem,

$$(14.1) \qquad \int_M \text{div}(X) \, d\text{vol}_M = \int_{\partial M} \langle X, \nu \rangle \, d\text{vol}_{\partial M},$$

a corollary of Stokes' theorem that equates the total divergence of a vector field on a Riemannian manifold with boundary to its total *flux* through the boundary (see §12.5.1). We would now like to explain a more fundamental interpretation of the divergence: it measures the extent to which the flow of X changes volume.

Writing $\text{Vol}(A) := \int_A d\text{vol}$, a diffeomorphism $\varphi : M \to M$ is called **volume preserving** if

$$\text{Vol}(\varphi(A)) = \text{Vol}(A) \qquad \text{for all measurable sets } A \subset M.$$

For a vector field $X \in \mathfrak{X}(M)$ admitting a global flow, we say that its flow is volume preserving if φ_X^t is volume preserving for every $t \in \mathbb{R}$. Without assuming there is a global flow, this condition can still be generalized as follows: for every measurable set $A \subset M$ and every $t \in \mathbb{R}$ for which the domain of φ_X^t contains A, $\text{Vol}(\varphi_X^t(A)) = \text{Vol}(A)$. Note that if A has compact closure, then this condition always makes sense at least for t close to 0. For simplicity we will assume in the following discussion that there is always a global flow, but this condition can be lifted by paying more careful attention to the domains of the flow maps φ_X^t.

The diffeomorphisms $\varphi_X^t : M \to M$ defined via the flow of a vector field are always orientation preserving—this results from the fact that $\varphi_X^0 : M \to M$ is the identity map, so for any $p \in M$, any positively oriented basis $Y_1, \ldots, Y_n$ of $T_p M$ gives rise to a continuous 1-parameter family of bases

$$(T\varphi_X^t(Y_1), \ldots, T\varphi_X^t(Y_n))$$

for the tangent spaces $T_{\varphi_X^t(p)} M$, and continuity dictates that they must all be positively oriented. We therefore have

$$\text{Vol}(\varphi_X^t(A)) = \int_{\varphi_X^t(A)} d\text{vol} = \int_A (\varphi_X^t)^* d\text{vol}$$

for every $A \subset M$, and the rate of change of this volume is

$$(14.2) \qquad \frac{d}{dt} \text{Vol}(\varphi_X^t(A)) = \frac{d}{dt} \int_A (\varphi_X^t)^* d\text{vol} = \int_A \partial_t (\varphi_X^t)^* d\text{vol}.$$

The next step in the calculation works in more general contexts: in place of the volume form $d\text{vol}$, we can consider an arbitrary tensor field $S \in \Gamma(T_\ell^k M)$. Recall that $\varphi_X^{s+t} = \varphi_X^s \circ \varphi_X^t$, thus $(\varphi_X^{s+t})^* = (\varphi_X^t)^* (\varphi_X^s)^*$, and

$$(14.3) \qquad \begin{aligned} \partial_t (\varphi_X^t)^* S &= \partial_s (\varphi_X^{s+t})^* S\big|_{s=0} = \partial_s (\varphi_X^t)^* (\varphi_X^s)^* S\big|_{s=0} \\ &= (\varphi_X^t)^* \left(\partial_s (\varphi_X^s)^* S\big|_{s=0} \right) = (\varphi_X^t)^* (\mathcal{L}_X S). \end{aligned}$$

Applying this to (14.2) gives

$$\frac{d}{dt} \text{Vol}(\varphi_X^t(A)) = \int_A (\varphi_X^t)^* (\mathcal{L}_X d\text{vol}) = \int_{\varphi_X^t(A)} \mathcal{L}_X d\text{vol}.$$

It follows that the flow is volume preserving if the Lie derivative of the volume form $d\text{vol}$ with respect to X vanishes, and conversely, the derivative of $\text{Vol}(\varphi_X^t(A))$ can only vanish for every

measurable set $A \subset M$ if the n-form $(\varphi_X^t)^*(\mathcal{L}_X d\mathrm{vol})$ vanishes identically for every t, which is equivalent to the condition $\mathcal{L}_X d\mathrm{vol} \equiv 0$ since $(\varphi_X^t)^* : \Omega^n(M) \to \Omega^n(M)$ is a bijection.

LEMMA 14.1. *For any volume form $d\mathrm{vol} \in \Omega^n(M)$ and vector field $X \in \mathfrak{X}(M)$,*

$$\mathcal{L}_X d\mathrm{vol} = d(\iota_X d\mathrm{vol}).$$

This relation will follow from the more general formula of Cartan for Lie derivatives of differential forms, to be proved in the next section. We can now alternatively characterize the divergence of X as the unique function such that

$$(14.4) \qquad\qquad \mathcal{L}_X d\mathrm{vol} = \mathrm{div}(X) \cdot d\mathrm{vol},$$

and the discussion above implies:

THEOREM 14.2. *On a manifold M with volume form $d\mathrm{vol}$, a vector field $X \in \mathfrak{X}(M)$ has a volume-preserving flow if and only if $\mathrm{div}(X) \equiv 0$.* $\qquad\square$

The divergence theorem (14.1) now admits a new geometric interpretation whenever M is a compact submanifold with boundary in a larger n-manifold N on which the vector field X and volume form $d\mathrm{vol}$ are defined. In this case, the flow φ_X^t of X is well defined on M for all t sufficiently close to zero, and the left hand side of (14.1) then becomes

$$\int_M \mathrm{div}(X)\, d\mathrm{vol}_N = \int_M \mathcal{L}_X(d\mathrm{vol}_N) = \frac{d}{dt} \int_M (\varphi_X^t)^* d\mathrm{vol}_N \Big|_{t=0} = \frac{d}{dt} \int_{\varphi_X^t(M)} d\mathrm{vol}_N \Big|_{t=0}$$

$$= \frac{d}{dt} \mathrm{Vol}(\varphi_X^t(M)) \Big|_{t=0}.$$

The divergence theorem thus relates the rate of change of the volume of M under the flow of X to the average of $\langle X, \nu \rangle$ along ∂M, which measures the extent to which X flows out of M vs. into M through its boundary.

14.2. Cartan's formula for the Lie derivative. The following practical tool for computing Lie derivatives of forms is sometimes called *Cartan's magic formula*.

THEOREM 14.3. *For any $X \in \mathfrak{X}(M)$ and $\omega \in \Omega^k(M)$,*

$$\mathcal{L}_X \omega = d(\iota_X \omega) + \iota_X(d\omega).$$

An immediate application is Lemma 14.1 above: if $d\mathrm{vol} \in \Omega^n(M)$ is a volume form, then

$$\mathcal{L}_X d\mathrm{vol} = d(\iota_X d\mathrm{vol}) + \iota_X d(d\mathrm{vol}) = d(\iota_X d\mathrm{vol})$$

since $d(d\mathrm{vol})$ is an $(n+1)$-form on an n-manifold and therefore vanishes.[49]

The following sequence of exercises sums up to a proof of Cartan's formula, the idea behind it being to show that for any given $X \in \mathfrak{X}(M)$, both of the operators $\mathcal{L}_X$ and $d\iota_X + \iota_X d$ define derivations on the exterior algebra $\Omega^*(M)$ that match when applied to functions or differentials of functions. This is sufficient for the same reason that a few formal properties centered around the graded Leibniz rule sufficed in Proposition 9.15 for characterizing the exterior derivaive: both are clearly local operators, and locally, every differential form is a finite sum of wedge products of functions and differentials.

EXERCISE 14.4 (easy). Show that Theorem 14.3 holds for all $\omega = f \in C^\infty(M) = \Omega^0(M)$.

---

[49]Here is another cautionary reminder about the oddity of our notation for volume forms: we have not defined any $(n-1)$-form "vol $\in \Omega^{n-1}(M)$" for $d\mathrm{vol}$ to be the exterior derivative of, and we have seen for instance that when M is a closed manifold, $d\mathrm{vol}$ is definitely not the exterior derivative of anything. The vanishing of $d(d\mathrm{vol})$ thus has nothing to do with the relation $d \circ d = 0$; it vanishes for a completely different reason.

LEMMA 14.5. *Theorem 14.3 holds for all $\omega = df \in \Omega^1(M)$ with $f \in C^\infty(M)$.*

PROOF. Since $d^2 = 0$, $d\iota_X df + \iota_X d(df) = d(\iota_X df)$, where $\iota_X df$ is the real-valued function $p \mapsto df(X(p))$. To evaluate $\mathcal{L}_X(df) \in \Omega^1(M)$ on some $Y \in T_p M$ at a point $p \in M$, choose a smooth path $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = Y$. Then using Proposition 9.17,

$$\mathcal{L}_X(df)(Y) = \partial_t(\varphi_X^t)^*(df)(Y)\big|_{t=0} = \partial_t d(f \circ \varphi_X^t)(Y)\big|_{t=0} = \partial_t \partial_s f(\varphi_X^t(\gamma(s)))\big|_{s=t=0}$$
$$= \partial_s \partial_t f(\varphi_X^t(\gamma(s)))\big|_{s=t=0} = \partial_s df(X(\gamma(s)))\big|_{s=0} = \partial_s \iota_X(df)(\gamma(s))\big|_{s=0} = d(\iota_X df)(Y).$$
$\square$

The next exercise follows also quite easily from the definition of the Lie derivative, plus Proposition 9.17 and the fact that the wedge product is bilinear. Notice that in contrast to the exterior derivative, no annoying sign appears in the Leibniz rule for $\mathcal{L}_X$. Formally, the reason is because $\mathcal{L}_X$ sends k-forms to k-forms for each $k \geq 0$, and is thus an operator of "degree 0", i.e. it is even, while the exterior derivative is odd.

EXERCISE 14.6. Show that $\mathcal{L}_X : \Omega^*(M) \to \Omega^*(M)$ is a derivation with respect to the wedge product, meaning

$$\mathcal{L}_X(\alpha \wedge \beta) = \mathcal{L}_X \alpha \wedge \beta + \alpha \wedge \mathcal{L}_X \beta.$$

We now turn our attention fully to the operator

$$(14.5) \qquad P_X := d\iota_X + \iota_X d : \Omega^*(M) \to \Omega^*(M),$$

in which each term is a composition of operators with degrees 1 and -1, so P_X itself also has degree 0. We've seen already that d satisfies a graded Leibniz rule; it turns out that ι_X does as well:

EXERCISE 14.7. For V an n-dimensional vector space, the goal of this exercise is to show that for every $v \in V$, the operator $\iota_v : \Lambda^* V^* \to \Lambda^* V^*$ satisfies the graded Leibniz rule

$$(14.6) \qquad \iota_v(\alpha \wedge \beta) = (\iota_v \alpha) \wedge \beta + (-1)^k \alpha \wedge (\iota_v \beta)$$

for all $\alpha \in \Lambda^k V^*$ and $\beta \in \Lambda^\ell V^*$. The statement is trivial if $v = 0$, so assume otherwise, in which case we may as well assume v is the first element e_1 of a basis $e_1, \ldots, e_n \in V$, whose dual basis we can denote by $e_*^1, \ldots, e_*^n \in V^* = \Lambda^1 V^*$.

    (a) Prove that (14.6) holds whenever α and β are both products of the form $\alpha = e_*^{i_1} \wedge \ldots \wedge e_*^{i_k}$ and $\beta = e_*^{j_1} \wedge \ldots \wedge e_*^{j_\ell}$ with $i_1 < \ldots < i_k$ and $j_1 < \ldots < j_\ell$.
        Hint: Consider separately a short list of cases depending on whether each of i_1 and j_1 are 1 and whether the sets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_\ell\}$ are disjoint.
    (b) Deduce via linearity that (14.6) holds always.

EXERCISE 14.8. Prove that the operator P_X in (14.5) is also a derivation on $\Omega^*(M)$, and deduce that $P_X = \mathcal{L}_X$, thus proving Theorem 14.3.

14.3. Symplectic manifolds and Hamiltonian systems. Volume-preserving flows arise naturally in the context of Hamiltonian systems, a special class of dynamical systems that originate in classical mechanics. From a mathematical perspective, the most natural language for this discussion is that of *symplectic* geometry.

DEFINITION 14.9. Assume M is a smooth manifold of even dimension $2n$ for some $n \in \mathbb{N}$. A 2-form $\omega \in \Omega^2(M)$ is called **symplectic** (*symplektisch*) if every point $x \in M$ admits a neighborhood $x \in \mathcal{U} \subset M$ with a coordinate chart of the form $(\mathcal{U}, (p^1, q^1, \ldots, p^n, q^n))$ such that

$$(14.7) \qquad \omega = \sum_{j=1}^n dp^j \wedge dq^j \qquad \text{on } \mathcal{U}.$$

A 2-form with this property is also sometimes called a **symplectic structure** (*symplektische Struktur*) on M, and the pair (M, ω) in this situation is called a **symplectic manifold** (*symplektische Mannigfaltigkeit*).

Observe that the coordinates $(p^1, q^1, \ldots, p^n, q^n)$ appearing in (14.7) are special; it would certainly be impossible to demand that any 2-form satisfy (14.7) for *every* choice of chart, but the definition only requires the existence of *some* chart near every point so that ω takes this form. In this sense, a symplectic structure is somewhat analogous to an orientation: it is equivalent in fact to a maximal atlas of compatible charts in which the word "compatible" has been given a new and much stricter definition, requiring all transition maps to not only be smooth but also to preserve the relation (14.7). Physicists sometimes refer to coordinates $(p^1, q^1, \ldots, p^n, q^n)$ of this type as *canonical coordinates* and call the corresponding transition maps *canonical transformations*. Mathematicians prefer to call them *Darboux coordinates*, after Darboux's theorem (see Remark 14.11 below).

EXERCISE 14.10. Show that a symplectic form $\omega \in \Omega^2(M)$ always has the following properties:
  (a) ω is closed: $d\omega = 0$.
  (b) For every $x \in M$, the linear map $T_x M \to T_x^* M : X \mapsto \omega(X, \cdot)$ is an isomorphism. (Bilinear forms with this property are called **nondegenerate**).
  (c) The "top" exterior power of ω,

$$\omega^n := \underbrace{\omega \wedge \ldots \wedge \omega}_{n} \in \Omega^{2n}(M)$$

  is a volume form on M. It follows in particular that M is orientable.
  (d) Is M is closed, then ω represents a nontrivial cohomology class $[\omega] \in H_{\mathrm{dR}}^2(M)$.
  Hint: Recall the cup product from Exercise 13.37. What can you say about the n-fold cup product of $[\omega]$ with itself?

REMARK 14.11. A fundamental result known as *Darboux's theorem* says that symplectic forms can in fact be characterized fully in terms of the first two properties in Exercise 14.10, i.e. every 2-form that is both closed and nondegenerate admits an atlas of charts satisfying (14.7) and is thus a symplectic form. This reveals for instance that every volume form on a surface[50] is a symplectic form. We will not make use of these facts here, but it is important to be aware of them since most textbooks prefer to *define* the term "symplectic form" to mean a closed and nondegenerate 2-form.

Given a smooth function $H : M \to \mathbb{R}$ on a symplectic manifold (M, ω), the nondegeneracy of ω implies that there is a unique vector field $X_H \in \mathfrak{X}(M)$ satisfying

$$(14.8) \qquad \omega(X_H, \cdot) = -dH \in \Omega^1(M).$$

We call X_H the **Hamiltonian vector field** determined by H, and in this context, the function H itself is often called a **Hamiltonian**. In Darboux coordinates, it is not hard to derive an explicit formula for the Hamiltonian vector field: writing $X_H = A^j \frac{\partial}{\partial q^j} + B^j \frac{\partial}{\partial p^j}$, we find

$$dH = \frac{\partial H}{\partial q^j} dq^j + \frac{\partial H}{\partial p^j} dp^j = -\omega(X_H, \cdot) = -\sum_{i=1}^{n} (dp^i \wedge dq^i) \left(A^j \frac{\partial}{\partial q^j} + B^j \frac{\partial}{\partial p^j}, \cdot \right)$$

$$= \sum_{i=1}^{n} \left(-B^i \, dq^i + A^i \, dp^i \right),$$

---

[50]On a manifold of dimension 2, it is also common to refer to volume forms as **area forms**.

implying

$$(14.9) \qquad X_H = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i} \right).$$

In other words, if $x(t) \in M$ denotes a smooth path passing through the domain of a Darboux chart and its coordinates in this chart at time t are written as $(p^1(t), q^1(t), \ldots, p^n(t), q^n(t))$, then x is an orbit of X_H if and only if its coordinates satisfy the following system of $2n$ first-order ODEs:

$$(14.10) \qquad \dot{q}^i(t) = \frac{\partial H}{\partial p^i}(x(t)), \qquad \dot{p}^i(t) = -\frac{\partial H}{\partial q^i}(x(t)) \qquad i = 1, \ldots, n.$$

This system is known as *Hamilton's equations*, and the dynamical system defined by the flow of X_H is called a *Hamiltonian system*.

The study of Hamiltonian systems originates with the following example.

EXAMPLE 14.12. In classical mechanics, the motion in $\mathbb{R}^3$ of a single particle with mass $m > 0$ under the influence of a force is described by a path $\mathbf{q}(t) = (q^1(t), q^2(t), q^3(t)) \in \mathbb{R}^3$ that obeys Newton's second law,

$$\mathbf{F}(\mathbf{q}(t)) = m\ddot{\mathbf{q}}(t),$$

where $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field representing the force. In standard examples, $\mathbf{F}$ is determined by a *potential* $V : \mathbb{R}^3 \to \mathbb{R}$ via the relation

$$\mathbf{F} = -\nabla V,$$

hence the individual coordinates satisfy $m\ddot{q}^i(t) = -\frac{\partial V}{\partial q^i}(\mathbf{q}(t))$. There is a popular trick for turning second-order systems of ODEs like this one into first-order systems with twice as many degrees of freedom: we associate to the position variables q^1, q^2, q^3 the corresponding *momentum* variables

$$p^i(t) := m\dot{q}^i(t), \qquad \mathbf{p} := (p^1, p^2, p^3)$$

and observe that the path $(\mathbf{q}(t), \mathbf{p}(t)) \in \mathbb{R}^6$ now satisfies the first-order system of equations

$$\dot{q}^i(t) = \frac{1}{m}p^i(t), \qquad \dot{p}^i(t) = -\frac{\partial V}{\partial q^i}(\mathbf{q}(t)), \qquad i = 1, 2, 3.$$

As it happens, this is the Hamiltonian system determined by the function $H : \mathbb{R}^6 \to \mathbb{R}$ given by

$$H(\mathbf{q}, \mathbf{p}) := \frac{|\mathbf{p}|^2}{2m} + V(\mathbf{q}).$$

Rewriting this as a function of $\mathbf{q}$ and $\dot{\mathbf{q}} := \frac{1}{m}\mathbf{p}$, the first term becomes $\frac{1}{2}m|\dot{\mathbf{q}}|^2$, which physicists call the *kinetic energy* of the moving particle. This is summed with the potential energy $V(\mathbf{q})$ to produce the Hamiltonian, which therefore has an interpretation as the *total energy* of the particle.

The Hamiltonian formalism lends itself to generalization: to turn the example above into a system of $N > 1$ moving particles, one can package the coordinates of all particles together to form a path in $\mathbb{R}^{3N}$, define corresponding momenta to produce a path in the so-called **phase space** $\mathbb{R}^{6N}$, write the total energy of the system as a function of all its position and momentum variables, and then write down Hamilton's equations (14.10). More generally, one can consider systems with constraints that prevent their positions from moving freely in Euclidean space, but confine them instead to a submanifold. In this situation there might not exist any global coordinate system in which Hamilton's equations (14.10) make sense, but if we have a symplectic form and a Hamiltonian function, then (14.8) defines the Hamiltonian vector field in a way that is independent of coordinates. We will see for instance that on any n-dimensional Riemannian manifold, the geodesic equation can be identified with a Hamiltonian system on a manifold of dimension $2n$.

If you've wondered why we are discussing symplectic manifolds in the same lecture with volume-preserving flows, here is the reasons:

THEOREM 14.13 (Liouville's theorem). *For any symplectic manifold (M, ω) and Hamiltonian $H \in C^\infty(M)$, the flow of the Hamiltonian vector field X_H is volume preserving with respect to the volume form $\omega^n \in \Omega^{2n}(M)$.*

PROOF. Let's do two proofs. The first is a coordinate-based computation: in any Darboux chart on some region in M, ω^n becomes a constant multiple of the standard volume form

$$\omega^n = \left(\sum_{i_1=1}^n dp^{i_1} \wedge dq^{i_1} \right) \wedge \ldots \wedge \left(\sum_{i_n=1}^n dp^{i_n} \wedge dq^{i_n} \right) = n \, dp^1 \wedge dq^1 \wedge \ldots \wedge dp^n \wedge dq^n,$$

and according to Exercise 12.16 and (14.9), the divergence of X_H is thus

$$\operatorname{div}(X_H) = \sum_{i=1}^n \left(\frac{\partial}{\partial q^i} \frac{\partial H}{\partial p^i} - \frac{\partial}{\partial p^i} \frac{\partial H}{\partial q^i} \right) = 0.$$

The result now follows from Theorem 14.2.

The second proof is more elegant, because it does not require coordinates, and it also proves a stronger result. Using Cartan's formula and the defining property of the vector field X_H, we find

$$\mathcal{L}_{X_H} \omega = d(\iota_{X_H} \omega) + \iota_{X_H}(d\omega) = -d(dH) = 0.$$

It follows via (14.3) that the 2-forms $(\varphi_{X_H}^t)^* \omega$ are independent of t, and thus

(14.11) $$(\varphi_{X_H}^t)^* \omega = \omega \text{ for all } t.$$

It follows that for each t, $\varphi := \varphi_{X_H}^t$ also preserves the volume form ω^n, since

(14.12) $$\varphi^*(\omega \wedge \ldots \wedge \omega) = \varphi^* \omega \wedge \ldots \wedge \varphi^* \omega = \omega \wedge \ldots \wedge \omega.$$

$\square$

I mentioned that our second proof of Liouville's theorem actually proves a stronger result. On a symplectic manifold (M, ω), a diffeomorphism $\psi : M \to M$ that satisfies

$$\psi^* \omega = \omega$$

is called a **symplectomorphism** (*Symplektomorphismus*), which can be viewed as an abbreviation for **symplectic diffeomorphism**. We see from (14.11) that Hamiltonian flows $\varphi_{X_H}^t$ have this property for every t, and by (14.12), all symplectomorphisms are also volume preserving.

While the subject of symplectic geometry has existed since the beginning of the 20th century, it was unknown for many decades whether the condition of being a symplectomorphism is truly more restrictive than being volume preserving. The following answer to this question emerged in 1985 and opened up a whole new subfield of geometry, known as *symplectic topology*:

THEOREM (Gromov's non-squeezing theorem [**Gro85**]). *Fix the global coordinates $(p^1, q^1, \ldots, p^n, q^n)$ on $\mathbb{R}^{2n}$ with the "standard" symplectic form $\omega = \sum_{i=1}^n dp^i \wedge dq^i$, and let $B_r^k \subset \mathbb{R}^k$ denote the open ball of radius r. Then for two constants $r, R > 0$, the $2n$-ball $B_r^{2n} \subset \mathbb{R}^{2n}$ is symplectomorphic to a subset of the "cylinder"*

$$Z_R^{2n} := B_R^2 \times \mathbb{R}^{2n-2} \subset \mathbb{R}^{2n}$$

if and only if $r \leqslant R$.

This is a hard theorem; various proofs are known, but all of them require a substantial amount of analytical machinery which cannot be fit into an introductory course. The significance of the non-squeezing theorem is that if $n \geqslant 2$, then no matter how small $R > 0$ may be, the cylinder Z_R^{2n} contains unlimited space in $2n - 2$ of its $2n$ dimensions, and it is never difficult to find a volume-preserving embedding $B_r^{2n} \hookrightarrow Z_R^{2n}$ that compresses the first two dimensions as much as needed while expanding the others to compensate. The fact that *symplectic* embeddings cannot do this when $R < r$ means that there are meaningful restrictions on symplectic maps beyond the requirement that they must preserve volume. That subject is still an active area of research today.

EXERCISE 14.14. In 1915, Emmy Noether established a beautiful correspondence between the conserved quantities of a mechanical system and its symmetries. A simple version of this theorem in the Hamiltonian context takes the following form. Assume (M, ω) is a symplectic manifold, and $H : M \to \mathbb{R}$ and $F : M \to \mathbb{R}$ are two functions such that the corresponding Hamiltonian vector fields X_H and X_F have global flows. We say that F is *conserved* under the flow of X_H if F is constant along every orbit of X_H.

(a) Show that F is conversed under the flow of X_H if and only if H is conserved under the flow of X_F.

(b) In some settings, there is a converse to the result proved in part (a). Suppose M is simply connected, and $Y \in \mathfrak{X}(M)$ is a vector field with a global flow that is symplectic and preserves H, i.e.

(14.13) $$(\varphi_Y^t)^* \omega = \omega \qquad \text{and} \qquad H \circ \varphi_Y^t = H$$

for all t. One says in this situation that Y determines a *symmetry* of the Hamiltonian system on (M, ω) defined by H. Under these assumptions, show that there exists a function $F : M \to \mathbb{R}$, uniquely defined up to addition of a constant, such that $Y = X_F$, and F is then conserved under the flow of X_H.

Let's work out a concrete example. Let $M = \mathbb{R}^4$ with coordinates (p_x, x, p_y, y) and the standard symplectic form

$$\omega = dp_x \wedge dx + dp_y \wedge dy \in \Omega^2(\mathbb{R}^4).$$

We can think of $\mathbb{R}^4$ as the "position-momentum space" (also known as *phase space*) representing the motion of a single particle of mass $m > 0$ in a plane: its position is given by $\mathbf{q} := (x, y) \in \mathbb{R}^2$, and $\mathbf{p} := (p_x, p_y) \in \mathbb{R}^2$ are the corresponding momentum variables. Given a "potential" function $V : \mathbb{R}^2 \to \mathbb{R}$, the total energy of the system is given by the function $H : \mathbb{R}^4 \to \mathbb{R}$,

$$H = \frac{|\mathbf{p}|^2}{2m} + V(\mathbf{q}).$$

Suppose now that the potential V is chosen to be *rotationally symmetric*, e.g. this is the case if $\mathbf{q}$ represents the position of the Earth moving around the sun (with the latter positioned at the origin). To express this condition succinctly, one can transform to polar coordinates (r, θ) on a suitable subset of $\mathbb{R}^2$, related to the (x, y)-coordinates as usual by $x = r\cos\theta$ and $y = r\sin\theta$. The condition imposed on V is then $\partial_\theta V \equiv 0$.

(c) Regarding r and θ as real-valued functions on (a suitable subdomain of) $\mathbb{R}^4$ that depend on the coordinates x and y but not on p_x and p_y, define two additional functions on the same domain by

$$p_r := \frac{x}{r} p_x + \frac{y}{r} p_y, \qquad p_\theta := y p_x - x p_y.$$

Show that $(p_r, r, p_\theta, \theta)$ is then a Darboux chart for the symplectic form ω.

Hint: It suffices to compute ω in the new coordinates and show that it satisfies the right

formula, but this computation is a bit long. You could make your life easier by observing that $\omega = d\lambda$ for $\lambda := p_x \, dx + p_y \, dy$, and then computing λ in the new coordinates.

(d) Write down H as a function of $(p_r, r, p_\theta, \theta)$ and show that the vector field $Y := \partial_\theta$ defined in these coordinates on $\mathbb{R}^4 \backslash \{r = 0\}$ satisfies (14.13). Derive a formula for the corresponding conserved quantity F as promised by part (b). It is called the angular momentum of the system.

15. Partitions of unity

In Lecture 11, we constructed partitions of unity subordinate to finite open covers of compact manifolds: more precisely, if $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is a finite collection of open sets in a manifold M whose union contains the compact subset $K \subset M$, then there exists an associated collection of smooth functions $\{\varphi_\alpha : M \to [0,1]\}_{\alpha \in I}$ such that

$$\sum_{\alpha \in I} \varphi_\alpha \equiv 1 \text{ on } K, \qquad \text{and} \qquad \text{supp}(\varphi_\alpha) \subset \mathcal{U}_\alpha \text{ is compact for every } \alpha \in I.$$

This was used in order to "localize" the problem of defining integrals $\int_A \omega$, and we used the same localization trick again to prove Stokes' theorem in Lecture 12. In this lecture, we will use a more general localization trick to prove that Riemannian metrics exist on all smooth manifolds M. Unless M happens to be compact, we will not be able to get away with considering only finite open covers or functions with compact support. We will therefore need a more general notion of partitions of unity and an extension of the previous construction. This turns out to be the point where one must finally make explicit use of the assumption that manifolds are metrizable.

15.1. Local finiteness. A collection of subsets $\{\mathcal{U}_\alpha \subset X\}_{\alpha \in I}$ in a topological space X is called **locally finite** if every point $p \in X$ has a neighborhood that intersects at most finitely many of the sets $\mathcal{U}_\alpha$. Similarly, a collection of functions $\{f_\alpha : X \to \mathbb{R}\}_{\alpha \in I}$ is called **locally finite** if the sets $\{f_\alpha^{-1}(\mathbb{R} \backslash \{0\}) \subset X\}_{\alpha \in I}$ form a locally finite collection. This condition has the following advantage: if $\{f_\alpha : M \to \mathbb{R}\}_{\alpha \in I}$ is a locally finite collection of *smooth* functions on a manifold M, then one can make sense of the sum

$$\sum_{\alpha \in I} f_\alpha(p) \in \mathbb{R}$$

for every $p \in M$ since, even if I is an uncountably infinite set, at most finitely many terms in this sum are nonzero. Even better, p admits a neighborhood $\mathcal{V} \subset M$ that intersects at most finitely many of the sets $f_\alpha^{-1}(\mathbb{R} \backslash \{0\})$, implying that at most finitely many of the functions f_α can have nonzero values anywhere on $\mathcal{V}$, and $\sum_{\alpha \in I} f_\alpha$ therefore makes sense as a *smooth* function on $\mathcal{V}$. We therefore obtain a global smooth function

$$\sum_{\alpha \in I} f_\alpha \in C^\infty(M),$$

even if the sum contains uncountably many terms that are (somewhere) nontrivial functions on M.

EXERCISE 15.1. Show that if X is a topological space with open subset $\mathcal{U} \subset X$ and a locally finite collection of continuous functions $\{f_\alpha : X \to \mathbb{R}\}_{\alpha \in I}$ satisfying $\text{supp}(f_\alpha) \subset \mathcal{U}$ for every $\alpha \in \mathcal{U}$, then $\sum_{\alpha \in I} f_\alpha$ also has support in $\mathcal{U}$.

DEFINITION 15.2. Given an open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of a smooth manifold M, a **partition of unity** subordinate to $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is a locally finite collection of smooth functions $\{\varphi_\alpha : M \to [0,1]\}_{\alpha \in I}$ which satisfy the following assumptions:

(1) For each $\alpha \in I$, $\text{supp}(\varphi_\alpha) \subset \mathcal{U}_\alpha$;
(2) $\sum_{\alpha \in I} \varphi_\alpha \equiv 1$.

Note that in Definition 15.2, the condition $\sum_{\alpha \in I} \varphi_\alpha \equiv 1$ makes sense due to the local finiteness assumption; this condition was automatic in Lecture 11 because we were considering only a finite collection of functions, but here we are not assuming the collection is finite, nor that the functions have compact support. This relaxation of assumptions makes it possible to prove the following without assuming M is compact:

THEOREM 15.3. *Every open cover of a smooth manifold admits a subordinate partition of unity.*

This theorem will be proved in §15.4.

15.2. Existence of Riemannian metrics and volume forms. Before proving that partitions of unity always exist, we shall demonstrate their usefulness by proving the following:

THEOREM 15.4. *Every smooth manifold admits a Riemannian metric.*

As a preliminary remark relevant to the proof, we observe that on any vector space V, the set of inner products on V forms a *convex* subset of the vector space of bilinear maps $V \times V \to \mathbb{R}$. Indeed, the symmetric bilinear maps form a linear subspace, and whenever $\langle\ ,\ \rangle_0$ and $\langle\ ,\ \rangle_1$ are two inner products on V, the interpolation $\langle\ ,\ \rangle_t := t\langle\ ,\ \rangle_1 + (1-t)\langle\ ,\ \rangle_0$ for $t \in [0,1]$ also satisfies

$$\langle v, v \rangle_t = t\langle v, v \rangle_1 + (1-t)\langle v, v \rangle_0 > 0$$

for every nonzero $v \in V$. More generally, any *convex combination* of finitely many inner products on V is also an inner product, i.e. for any finite collection of inner products $\langle\ ,\ \rangle_i$ and numbers $\tau_i \in [0,1]$ for $i = 1, \ldots, k$ with $\sum_{i=1}^{k} \tau_i = 1$,

$$\sum_{i=1}^{k} \tau_i \langle\ ,\ \rangle_i$$

is an inner product.

LEMMA 15.5. *Suppose $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is an open cover of a smooth manifold M with subordinate partition of unity $\{\varphi_\alpha\}_{\alpha \in I}$, and for each $\alpha \in I$, $g_\alpha \in \Gamma(T_2^0 \mathcal{U}_\alpha)$ is a Riemannian metric on the open subset $\mathcal{U}_\alpha$. Then the formula*

$$g := \sum_{\alpha \in I} \varphi_\alpha g_\alpha$$

defines a Riemannian metric on M, where in this sum, the term $\varphi_\alpha g_\alpha$ is interpreted as an element of $\Gamma(T_2^0 M)$ that vanishes outside of $\mathcal{U}_\alpha$.

PROOF. Since $\mathrm{supp}(\varphi_\alpha) \subset \mathcal{U}_\alpha$, the tensor field $\varphi_\alpha g_\alpha \in \Gamma(T_2^0 \mathcal{U}_\alpha)$ can be extended to a smooth tensor field on M that vanishes outside of $\mathcal{U}_\alpha$, and we will continue to denote the extension by $\varphi_\alpha g_\alpha \in \Gamma(T_2^0 M)$. The sum then makes sense and is smooth due to local finiteness, as every point is contained in a neighborhood on which only finitely many terms of the sum are nontrivial. Moreover, at each individual point $p \in M$, $g_p : T_p M \times T_p M \to \mathbb{R}$ is a convex combination of inner products, and is therefore also an inner product. $\square$

PROOF OF THEOREM 15.4. Choose an open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of M such that each $\mathcal{U}_\alpha$ is the domain of a chart x_α, and define a Riemannian metric g_α on $\mathcal{U}_\alpha$ that looks like the standard Euclidean inner product in the chosen coordinates. A global Riemannian metric $g \in \Gamma(T_2^0 M)$ can then be defined via Lemma 15.5 after choosing a subordinate partition of unity. $\square$

In light of Corollary 11.10 on the Riemannian volume form associated to a Riemannian metric, Theorem 15.4 implies:

COROLLARY 15.6. *Every smooth oriented manifold admits a volume form.* $\square$

EXERCISE 15.7. Use a partition of unity to prove Corollary 15.6 without mentioning Theorem 15.4 or Riemannian metrics. Use instead the fact that for any oriented n-dimensional vector space V, the set

$$\left\{ \omega \in \Lambda^n V^* \mid \omega(v_1, \ldots, v_n) > 0 \text{ for some positively-oriented basis } v_1, \ldots, v_n \in V \right\}$$

is convex.

REMARK 15.8. Without assuming M is oriented, Theorem 15.4 also implies that every smooth manifold admits a volume element (see §11.4).

15.3. Paracompactness. Any Riemannian manifold (M, g) is also a metric space in a natural way, at least if it is connected, because one can define the distance between two points $p, q \in M$ by

$$(15.1) \qquad\qquad \operatorname{dist}(p, q) := \inf_{\gamma} \int_a^b \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))}\, dt,$$

where the infimum is over all intervals $[a, b] \subset \mathbb{R}$ and smooth paths $\gamma : [a, b] \to M$ with $\gamma(a) = p$ and $\gamma(b) = q$. For a Riemannian manifold with multiple connected components, each component has a natural metric defined in this way, and there are standard tricks for defining metrics on any disjoint union of metric spaces (see e.g. Exercise 2.23). The point is: if you hadn't already assumed that smooth manifolds are metrizable but you assumed that Theorem 15.4 is true, then the theorem would imply metrizability.

EXERCISE 15.9. Take a moment to convince yourself that (15.1) really does define a metric, in particular that it satisfies the triangle inequality.
Hint: One can reparametrize the path $\gamma : [a, b] \to M$ quite freely without changing the integral. If you take advantage of this freedom, then a path from p to q and a path from q to r can always be concatenated smoothly.

The existence of the metric (15.1) is a dead giveaway that something about Theorem 15.4 depends on our assumption that all manifolds are metrizable. We haven't used that assumption in this course until now. But we will need it for constructing the partition of unity.

Recall that a **refinement** of an open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is another open cover $\{\mathcal{V}_\beta\}_{\beta \in J}$ such that for every $\beta \in J$, $\mathcal{V}_\beta$ is contained in $\mathcal{U}_\alpha$ for some $\alpha \in I$.

DEFINITION 15.10. A topological space X is **paracompact** if every open cover of X admits a locally finite refinement.

Compact topological spaces are obviously paracompact since a finite subcover can also be viewed as a locally finite refinement. I can now tell you the true reason why we need to assume manifolds are metrizable: *all metrizable spaces are paracompact*. We will not prove quite such a general statement here, but we will make use of the metrizability assumption in the following to prove that manifolds are always paracompact.

LEMMA 15.11. *Every manifold M is σ-compact, i.e. it is the union of countably many compact subsets.*

PROOF. The result is true for every connected locally compact metric space (see e.g. [**Spi99a**, Theorem 1.2]), but for our purposes it will be more convenient to drop connectedness and instead assume separability, which holds in any case on all manifolds. Fix a metric d on M that is compatible with its topology. The term "locally compact" refers to the following observation: for every $p \in M$, the closed ball

$$\bar{B}_r(p) := \left\{ q \in M \mid d(p, q) \leqslant r \right\}$$

is compact for every $r > 0$ sufficiently small. This holds because whenever r is sufficiently small, $\bar{B}_r(p)$ lies in the domain of a chart that identifies it with a closed and bounded (and therefore compact) subset of Euclidean space. On the other hand, closed and bounded subsets of arbitrary metric spaces are not always compact, so we cannot assume $\bar{B}_r(p)$ is compact for every $r > 0$, but there is a positive (if not infinite) upper bound

$$\kappa(p) := \sup\left\{r > 0 \mid \bar{B}_r(p) \text{ is compact}\right\} \in (0, \infty].$$

If $\kappa(p) = \infty$ at any point p, then M is exhausted by the sequence of compact sets $\bar{B}_k(p)$ for $k = 1, 2, 3, \ldots$ and we are therefore done. Otherwise, observe that by the triangle inequality, every $q \in \bar{B}_{\frac{1}{3}\kappa(p)}(p)$ satisfies

$$\bar{B}_{\frac{1}{3}\kappa(p)}(q) \subset \bar{B}_{\frac{2}{3}\kappa(p)}(p),$$

implying that $\bar{B}_{\frac{1}{3}\kappa(p)}(q)$ is also compact and thus

$$(15.2) \qquad\qquad \kappa(q) \geqslant \frac{\kappa(p)}{3} \quad \text{for all} \quad q \in \bar{B}_{\frac{1}{3}\kappa(p)}(p).$$

Now for any dense sequence $p_1, p_2, p_3, \ldots \in M$, we claim that

$$M = \bigcup_{k=1}^{\infty} \bar{B}_{\frac{2}{3}\kappa(p_k)}(p_k),$$

where the sets on the right hand side are clearly all compact. Indeed, for any $p \in M$, we can replace $p_1, p_2, p_3, \ldots$ with a subsequence such that $p_k \to p$ as $k \to \infty$, and it follows from (15.2) that $\kappa(p_k) \geqslant \kappa(p)/3$ for all k sufficiently large, so that eventually $p \in \bar{B}_{\frac{2}{3}\kappa(p_k)}$. $\qquad\square$

EXERCISE 15.12. Show that if X is a topological space with a locally finite open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ and $K \subset X$ is a compact subset, then K intersects only finitely many of the sets $\mathcal{U}_\alpha$. (It follows from this that if X is σ-compact, then the set I cannot be uncountable, i.e. all locally finite open covers are at most countable. By Lemma 15.11, this applies in particular to all manifolds.

THEOREM 15.13. *Every smooth manifold is paracompact.*

PROOF. Assume $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is an open cover of M, and using Lemma 15.11, write $M = \bigcup_{j=1}^{\infty} K_j$ for compact subsets $K_1, K_2, K_3, \ldots$. Choose an open neighborhood $\mathcal{V}_1 \subset M$ of K_1 whose closure is compact, so the set $\bar{\mathcal{V}}_1 \cup K_2$ is also compact. Next, choose $\mathcal{V}_2 \subset M$ to be an open neighborhood of $\bar{\mathcal{V}}_1 \cup K_2$ whose closure is compact, so $\bar{\mathcal{V}}_2 \cup K_3$ is compact. Continuing in this way, one obtains a nested sequence

$$\varnothing =: \mathcal{V}_0 \subset \mathcal{V}_1 \subset \bar{\mathcal{V}}_1 \subset \mathcal{V}_2 \subset \bar{\mathcal{V}}_2 \subset \mathcal{V}_3 \subset \bar{\mathcal{V}}_3 \subset \ldots \subset \bigcup_{j=1}^{\infty} \mathcal{V}_j = M$$

such that each $\mathcal{V}_j$ is open and each $\bar{\mathcal{V}}_j$ is compact. We will now construct a locally finite refinement of $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ by using the "annular" regions

$$A_j := \bar{\mathcal{V}}_j \backslash \mathcal{V}_{j-1} \subset M, \qquad j = 1, 2, 3, \ldots,$$

which are all compact, and their union is also M. For each $j \in \mathbb{N}$, pick a finite open covering $\{\mathcal{O}_\beta^j \subset M\}_{\beta \in I_j}$ of A_j such that each of the open sets $\mathcal{O}_\beta^j$ is contained in $\mathcal{U}_\alpha$ for some $\alpha \in I$ and is also contained in $\mathcal{V}_{j+1} \backslash \mathcal{V}_{j-2}$. The union of these finite collections for $j = 1, 2, 3, \ldots$ forms an open cover of M that refines $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ and is also locally finite. $\qquad\square$

EXERCISE 15.14. Show that without loss of generality, one can assume that all of the open sets in the locally finite refinement given by Theorem 15.13 are diffeomorphic to open balls in Euclidean space.

Remark: This fact is frequently used in proofs that smooth manifolds admit partitions of unity, see for example [**Lee13a**, §II.3]. *It is not strictly necessary, however, and we will not use it. The proof given below is conceived to be as close as possible in spirit to proofs of similar results on more general topological spaces, which need not look locally like Euclidean space.*

15.4. Existence of partitions of unity. Now that we know that locally finite refinements can always be found, we need two further ingredients in order to construct partitions of unity. The first is purely topological.

A topological space X is called **normal** if every pair of disjoint closed subsets $A, B \subset X$ have neighborhoods in X that are also disjoint from each other.

EXERCISE 15.15. Show that all metric spaces are normal.

LEMMA 15.16 (the "shrinking lemma"). *Given a locally finite open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of a normal topological space X, there exists another open cover $\{\mathcal{V}_\alpha\}_{\alpha \in I}$ such that $\bar{\mathcal{V}}_\alpha \subset \mathcal{U}_\alpha$ for every $\alpha \in I$.*

PROOF. We shall give a proof under the extra assumption that the set I is at most countable, which is always true on manifolds due to Exercise 15.12. A proof without this assumption is possible using Zorn's lemma, see e.g. [**nLa**].

Since I is at most countable, we can relable the open cover as $\{\mathcal{U}_i\}_{i=1}^N$ where $N \in \mathbb{N} \cup \{\infty\}$. The sets $A_1 := X \backslash \bigcup_{i=2}^\infty \mathcal{U}_2$ and $X \backslash \mathcal{U}_1$ are closed and disjoint, so we can choose $\mathcal{V}_1 \subset X$ to be any open neighborhood of A_1 that is also disjoint from some neighborhood of $X \backslash \mathcal{U}_1$, implying $\bar{\mathcal{V}}_1 \subset \mathcal{U}_1$. Since $X = \mathcal{V}_1 \cup \bigcup_{i=2}^N \mathcal{U}_i$, we can next take the latter as another open cover on X, and perform the same trick on $\mathcal{U}_2$, producing an open set $\mathcal{V}_2 \subset \bar{\mathcal{V}}_2 \subset \mathcal{U}_2$ such that $X = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \bigcup_{i=3}^N \mathcal{U}_i$. Now repeat this procedure for $i = 3, 4, \ldots, N$, producing a sequence of shrunken open sets $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \ldots \subset X$ such that for each $m \in \mathbb{N}$,

$$(15.3) \qquad\qquad X = \bigcup_{i=1}^m \mathcal{V}_i \cup \bigcup_{i=m+1}^N \mathcal{U}_i.$$

If $N < \infty$ then we are done. If $N = \infty$, we now appeal to local finiteness and observe that for every $p \in M$, there exists a largest $m \in \mathbb{N}$ for which $p \in \mathcal{U}_m$, hence (15.3) implies $p \in \bigcup_{i=1}^m \mathcal{V}_i$ and thus $X = \bigcup_{i=1}^\infty \mathcal{V}_i$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

LEMMA 15.17 (the smooth Urysohn lemma). *Given a smooth manifold M with subsets $A \subset \mathcal{U} \subset M$ such that A is closed and $\mathcal{U}$ is open, there exists a smooth function $f : M \to [0,1]$ with support in $\mathcal{U}$ such that $f|_A \equiv 1$.*

PROOF, PART 1. For this first of two steps in the proof, we add the assumption that $A \subset M$ is compact. Since the open sets $\mathcal{U}$ and $M \backslash A$ form a finite open cover of M, the compact case of our existence result for partitions of unity (Lemma 11.1) provides a pair of smooth functions $\varphi, \psi : M \to [0,1]$ that have compact support in $\mathcal{U}$ and $M \backslash A$ respectively such that $\varphi + \psi \equiv 1$ on A. Since $\psi|_A \equiv 0$, the function we were looking for is φ. $\qquad\qquad\qquad\qquad\square$

Before finishing the proof of Lemma 15.17, it will be convenient to forge ahead and show how these results imply the existence of partitions of unity.

PROOF OF THEOREM 15.3, WITH A CAVEAT. Starting from an arbitrary open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of M, we can first replace $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ by a locally finite refinement $\{\mathcal{O}_\beta\}_{\beta \in J}$. The latter has the property that for every $\beta \in J$, we can choose some $\alpha(\beta) \in I$ satisfying

$$\mathcal{O}_\beta \subset \mathcal{U}_{\alpha(\beta)}.$$

Next apply the shrinking lemma to find another open cover $\{\mathcal{V}_\beta\}_{\beta \in J}$ such that $\bar{\mathcal{V}}_\beta \subset \mathcal{O}_\beta$ for each $\beta \in J$. By Lemma 15.17, we can choose for each $\beta \in J$ a smooth function $f_\beta : M \to [0,1]$ with

support in $\mathcal{O}_\beta$ such that $f_\beta|_{\bar{\mathcal{V}}_\beta} \equiv 1$. Local finiteness implies that the sum $\sum_{\beta \in J} f_\beta$ is a well-defined smooth function on M, and since every point is contained in at least one of the sets $\mathcal{V}_\beta$, this sum is everywhere positive. Now for each $\alpha \in I$, define $\psi_\alpha : M \to \mathbb{R}$ by

$$\psi_\alpha := \sum_{\{\beta \in J \ | \ \alpha(\beta) = \alpha\}} f_\beta.$$

Local finiteness implies that these are also smooth functions on M and satisfy

$$\sum_{\alpha \in I} \psi_\alpha = \sum_{\beta \in J} f_\beta > 0,$$

and moreover, since each f_β in the sum for $\alpha(\beta) = \alpha$ has support in $\mathcal{O}_\beta \subset \mathcal{U}_\alpha$, ψ_α itself has support in $\mathcal{U}_\alpha$ (see Exercise 15.1). The desired functions φ_α can now be defined by

$$\varphi_\alpha := \frac{\psi_\alpha}{\sum_{\beta \in I} \psi_\beta}.$$

$\square$

Since we did not yet finish the proof of Lemma 15.17, let's pause now to consider what actually has been proved. Lemma 15.17 was used in the above proof to choose the functions f_β with support in $\mathcal{O}_\beta$ that equal 1 on $\bar{\mathcal{V}}_\beta \subset \mathcal{O}_\beta$. If we add to the hypotheses of Theorem 15.3 that each of the open sets $\mathcal{U}_\alpha \subset M$ has compact closure, then it guarantees that the sets $\bar{\mathcal{V}}_\beta$ are also compact, so that we only need to use the case of Lemma 15.17 that has already been proved. In summary, Theorem 15.3 has now been established under the extra hypothesis that each set $\bar{\mathcal{U}}_\alpha \subset M$ is compact. We can use this observation to complete the proof of Lemma 15.17 and thus establish Theorem 15.3 in full generality.

PROOF OF LEMMA 15.17, PART 2. Choose open coverings $\{\mathcal{U}_\alpha \subset M\}_{\alpha \in I}$ of A and $\{\mathcal{O}_\beta \subset M\}_{\beta \in J}$ of $M \backslash A$ such that all of the sets $\mathcal{U}_\alpha, \mathcal{O}_\beta$ have compact closure and

$$\mathcal{U}_\alpha \subset \mathcal{U} \text{ for all } \alpha \in I, \qquad \mathcal{O}_\beta \subset M \backslash A \text{ for all } \beta \in J.$$

Then $M = \bigcup_{\alpha \in I} \mathcal{U}_\alpha \cup \bigcup_{\beta \in J} \mathcal{O}_\beta$, and we can apply the case of Theorem 15.3 that has been proved already to find a locally finite partition of unity subordinate to this cover: it consists of smooth functions $\{\varphi_\alpha\}_{\alpha \in I}$ and $\{\psi_\beta\}_{\beta \in J}$ such that $\text{supp}(\varphi_\alpha) \subset \mathcal{U}_\alpha$ and $\text{supp}(\psi_\beta) \subset \mathcal{O}_\beta$ for all $(\alpha, \beta) \in I \times J$, while $\sum_{\alpha \in I} \varphi_\alpha + \sum_{\beta \in J} \psi_\beta \equiv 1$. Since every $\mathcal{O}_\beta$ is disjoint from A, it follows that $f := \sum_{\alpha \in I} \varphi_\alpha \equiv 1$ on A, and by Exercise 15.1, $\text{supp}(f) \subset \mathcal{U}$. $\square$

The proof of Theorem 15.3 is now complete.

REMARK 15.18. We made use of separability at one step in this lecture—namely in Lemma 15.11 on σ-compactness—because doing so was more convenient than the alternative, but it was not strictly necessary. As mentioned in the proof of Lemma 15.11, the lemma also holds for arbitrary connected and locally compact metric spaces, so if one works on only one connected component at a time, one obtains a proof of paracompactness for "manifolds" that are assumed metrizable but not necessarily separable. Some authors prefer in fact to define a manifold in a slightly more general way than we have, requiring them to be Hausdorff and paracompact but not necessarily separable or second countable—this shows you how highly the existence of partitions of unity is valued by differential geometers. The only difference this makes in reality is that by the more general definition, manifolds can have uncountably many connected components; in the connected case there is no difference. In any case, I have never seen an example of a non-separable "manifold" that I cared about, not even in infinite dimensions.

REMARK 15.19. On a topological space X, there is generally no well-defined notion of smooth functions, but one can still speak of partitions of unity in which the functions $\varphi_\alpha : X \to [0,1]$ are only required to be continuous. Such constructions are similarly useful in topology for proving existence results, e.g. the fact that every finite-dimensional *topological* manifold admits a proper topological embedding into $\mathbb{R}^N$ for N sufficiently large (see [**Lee11**, Chapter 4]). To prove that partitions of unity exist on a given space X, one obviously needs to know that X is paracompact, and the other major ingredients are the shrinking lemma (Lemma 15.16) and the continuous variant of Urysohn's lemma (Lemma 15.17), both of which hold whenever X is normal. It turns out that paracompact Hausdorff spaces are automatically normal, thus they admit continuous partitions of unity—in fact for Hausdorff spaces in general, the existence of partitions of unity is *equivalent* to paracompactness.

In nonlinear functional analysis, one sometimes also works with infinite-dimensional smooth manifolds that are locally modelled on Banach spaces. These are not locally compact, so our proof of paracompactness via σ-compactness does not adapt well to the infinite-dimensional setting, but one can nonetheless appeal to the fact that metric spaces are *always* paracompact. The simplest (or at least the shortest) proof of this is due to Mary Ellen Rudin [**Rud69**]. If one considers *arbitrary* metric spaces, then the proof makes slightly mysterious use of the axiom of choice, in the form of the well-ordering theorem: in particular, it uses the fact that for any open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$, the index set I can be endowed with a total order for which every subset has a smallest element. This is less mysterious however in the case of *separable* metric spaces, because every open cover in the separable case admits a finite subcover (exercise!), so one is free without loss of generality to assume the index set is $\mathbb{N}$. As a consequence, infinite-dimensional Banach manifolds are also paracompact, so long as we still agree that anything carrying the name "manifold" should be metrizable and separable. That is the convention that I adopt when I use these objects in my research, and it is not the only possible convention that people might consider reasonable, but it is relatively uncontroversial.

The existence of smooth partitions of unity in the infinite-dimensional setting is nonetheless a subtle question, because smooth compactly-supported "bump" functions do not always exist on Banach spaces—the basic problem here is that on a Banach space E with norm $\| \cdot \|$, the function $E \to \mathbb{R} : x \mapsto \|x\|^p$ is not generally differentiable at $0 \in E$ for any power $p > 0$, even for $p = 2$. As a result, the smooth Urysohn lemma is not true in this context, so smooth partitions of unity do not exist, and many popular constructions from differential geometry are simply not available on infinite-dimensional Banach manifolds. The exception is the case of Hilbert manifolds, which are locally modelled on Hilbert spaces—the inner product on a Hilbert space $\mathcal{H}$ has the convenient property that $\mathcal{H} \to \mathbb{R} : x \mapsto \|x\|^2 := \langle x, x \rangle$ is a smooth function, thus making smooth bump functions and smooth partitions of unity possible.

EXERCISE 15.20. Given a smooth manifold M, use an open cover and subordinate partition of unity on M to construct a Riemannian metric on the tangent bundle TM. Do *not* assume that Theorem 15.3 holds for TM.
Remark: This exercise ties up a loose end from early in the course: in Corollary 3.12, we defined a smooth structure on the tangent bundle TM of any smooth manifold M, but we never proved that the topology on TM induced by its maximal smooth atlas is metrizable. The existence of a Riemannian metric implies this, and if you follow the instructions in the exercise, its construction does not need to assume that TM is metrizable—it assumes only that M is.

EXERCISE 15.21. Here is an example of something that satisfies all of the conditions for being a connected smooth 2-manifold except metrizability. It is a variation due to Calabi and Rosenlicht [**CR53**] on a construction known as the **Prüfer surface**, and can be visualized as a an uncountable collection of planes that have been glued together along their open upper and lower halves, but

not along the x-axis, so that the result is a single plane in which the x-axis has been replaced by uncountably many copies of itself. Here is a precise definition: denote the open upper and lower half-planes by $\mathbb{H}_{\pm} := \{(x,y) \in \mathbb{R}^2 \mid \pm y > 0\}$, and associate to each $a \in \mathbb{R}$ a copy of the full plane $X_a := \mathbb{R}^2$. As a set, the Prüfer surface is

$$\Sigma := \mathbb{H}_{+} \cup \mathbb{H}_{-} \sqcup \left(\coprod_{a \in \mathbb{R}} X_a \right) \Big/ \sim$$

where the equivalence relation identifies each point $(x,y) \in X_a$ for $y \neq 0$ with the point $(a+yx, y) \in \mathbb{H}_{+} \cup \mathbb{H}_{-}$. Notice that $\mathbb{H}_{\pm}$ and X_a for each $a \in \mathbb{R}$ can be regarded naturally as subsets of Σ. Let us denote points $(x,y) \in X_a \subset \Sigma$ by

$$(x,y)_a \in \Sigma,$$

so by definition, $(x,y)_a = (x',y')_b$ whenever $y = y' \neq 0$ and $xy + a = x'y' + b$, but $(x,0)_a$ and $(x',0)_b$ are never equal when $a \neq b$. Prove:

(a) Σ admits a unique smooth structure for which the natural inclusions $\mathbb{H}_{\pm} \hookrightarrow \Sigma$ and $X_a \hookrightarrow \Sigma$ for each $a \in \mathbb{R}$ are diffeomorphisms onto their images. *Assume for the remaining parts of this exercise that Σ is equipped with the topology uniquely determined by this smooth structure (cf. Prop. 2.12).*

(b) For any two points $p, q \in \Sigma$, there exist neighborhoods $p \in \mathcal{U} \subset \Sigma$ and $q \in \mathcal{V} \subset \Sigma$ such that $\mathcal{U} \cap \mathcal{V} = \varnothing$. (In topological terminology, Σ is Hausdorff.)
   Hint: The only case where it is not so obvious is when p and q are both of the form $(x,0)_a$ and $(x',0)_b$. Try drawing pictures of the intersections of neighborhoods of those points with $\mathbb{H}_{+} \cup \mathbb{H}_{-}$.

(c) Σ is connected.

(d) Σ is separable.
   Hint: Show that any dense subset of $\mathbb{H}_{+} \cup \mathbb{H}_{-} \subset \Sigma$ is also dense in Σ.

(e) Here's where things get weird: the subset $\{(0,0)_a \in \Sigma \mid a \in \mathbb{R}\} \subset \Sigma$ is discrete, i.e. each of its points has a neighborhood that contains none of the others. In particular, all subsets of this set are closed.

(f) Σ is not σ-compact (no pun intended).
   Hint: According to part (e), it contains an uncountable discrete subset.

We can now deduce that Σ is not metrizable, as we would otherwise have a contradiction to the proof of Lemma 15.11. Here is an even stranger indication: recall from Exercise 15.15 that all metric spaces are normal.

(g) Suppose we have associated to each $a \in \mathbb{R}$ a "wedge-shaped" region in $\mathbb{H}_{+}$ of the form

$$W_a := \left\{ (r\cos\theta, r\sin\theta) \in \mathbb{H}_{+} \mid r \in (0, r(a)) \text{ and } \theta \in (\pi/2 - \epsilon(a), \pi/2 + \epsilon(a)) \right\}$$

for constants $r(a) > 0$ and $\epsilon(a) > 0$ that are allowed to vary arbitrarily with $a \in \mathbb{R}$. Show that there exists some $a_\infty \in \mathbb{Q}$ and a sequence $a_j \in \mathbb{R}\backslash\mathbb{Q}$ that converges to a_∞ such that $r(a_j)$ and $\epsilon(a_j)$ are both bounded from below.
   Big hint: $\mathbb{R} = \mathbb{Q} \cup \bigcup_{N \in \mathbb{N}} A_N$ where

$$A_N := \left\{ a \in \mathbb{R}\backslash\mathbb{Q} \mid r(a) \geqslant 1/N \text{ and } \epsilon(a) \geqslant 1/N \right\}.$$

According to the Baire category theorem, a nonempty complete metric space can never be the countable union of subsets that are nowhere dense, meaning sets whose closures have empty interior. Deduce from this that at least one of the sets A_N contains an open interval in its closure.

(h) Deduce that the disjoint subsets

$$Q := \big\{(0,0)_a \in \Sigma \mid a \in \mathbb{Q}\big\} \subset \Sigma \qquad \text{and} \qquad I := \big\{(0,0)_a \in \Sigma \mid a \in \mathbb{R} \backslash \mathbb{Q}\big\} \subset \Sigma$$

are both closed but do not admit disjoint neighborhoods, i.e. Σ is not normal.
(i) Show that the open cover $\{X_a \subset \Sigma\}_{a \in \mathbb{R}}$ of Σ has no locally finite refinement.
Hint: In any refinement of $\{X_a\}_{a \in \mathbb{R}}$, points of the form $(0,0)_a$ and $(0,0)_b$ for $a \neq b$ must always belong to different sets in the open cover. Show that for the point $a_\infty \in \mathbb{R}$ in part (g), every neighborhood of $(0,0)_{a_\infty}$ intersects infinitely many such sets.

The original Prüfer surface is slightly different from the variation by Calabi and Rosenlicht described above, and can be defined as

$$\Sigma' := \mathbb{H}_+ \sqcup \left(\coprod_{a \in \mathbb{R}} X_a\right) \Big/ \sim,$$

where the equivalence relation identifies points $(x,y) \in X_a$ with $(a + yx, y) \in \mathbb{H}_+$ only for $y > 0$. We can visualize Σ' as an uncountable collection of planes that have been glued together along their upper halves, leaving the lower halves separate.

(j) Show that Σ' has all the same properties we proved above for Σ, except that Σ' is not separable.

16. Vector bundles

We have already seen several examples in this course of sets of the form

$$E = \bigcup_{p \in M} E_p,$$

where M is a manifold and E_p is a vector space associated to each point $p \in M$. The obvious example is the tangent bundle TM, but we have also considered the cotangent bundle T^*M, which is the union of the dual spaces to the tangent spaces, and further examples arise in natural ways by thinking of tensor fields $S \in \Gamma(T_\ell^k M)$ as objects that associate to each point $p \in M$ an element S_p of a certain vector space of multilinear maps. For all of these examples, one can regard the vector spaces E_p as "varying smoothly" with respect to p, but this is an intuitive notion that we have not yet made precise except in the special case of TM, on which we defined a smooth structure so that the natural projection $\pi : TM \to M$ sending T_pM to p is a smooth map.

We will now start defining such notions in greater generality.

16.1. Main Definition. We begin with a few more observations about the motivating example of a vector bundle, namely the tangent bundle TM of a smooth n-manifold M. Recall that each chart $(\mathcal{U}, x)$ on M determines a family of vector space isomorphisms

$$d_p x : T_p M \to \mathbb{R}^n, \qquad p \in \mathcal{U}.$$

This information can be repackaged as a bijective map

$$\Phi_x : T\mathcal{U} \to \mathcal{U} \times \mathbb{R}^n$$

whose restriction to each of the individual vector spaces $T_pM \subset T\mathcal{U}$ for $p \in \mathcal{U}$ is $X \mapsto (p, d_p x(X)) \in \mathcal{U} \times \mathbb{R}^n$, and the smooth chart $(T\mathcal{U}, Tx)$ for TM can be derived from this by writing

$$Tx(X) = (x(p), d_p x(X)) = (x \times \mathbb{1}) \circ \Phi_x(X) \in \mathbb{R}^n \times \mathbb{R}^n \qquad \text{for } X \in T_pM, \, p \in \mathcal{U}.$$

Since $x \times \mathbb{1} : \mathcal{U} \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ is clearly a smooth map, the way that we defined the smooth structure on TM makes Φ_x not just a bijection, but also a diffeomorphism. Now, if $(\mathcal{V}, y)$ is another chart with $\mathcal{U} \cap \mathcal{V} \neq \varnothing$, there is a similar diffeomorphism

$$\Phi_y : T\mathcal{V} \to \mathcal{V} \times \mathbb{R}^n,$$

and both Φ_x and Φ_y restrict to diffeomorphisms $T(\mathcal{U} \cap \mathcal{V}) \to (\mathcal{U} \cap \mathcal{V}) \times \mathbb{R}^n$, giving rise to a map

$$\Phi_y \circ \Phi_x^{-1} : (\mathcal{U} \cap \mathcal{V}) \times \mathbb{R}^n \to (\mathcal{U} \cap \mathcal{V}) \times \mathbb{R}^n$$
$$(p, v) \mapsto (p, g(p)v),$$

where

$$g(p) := d_p y \circ (d_p x)^{-1} = D(y \circ x^{-1})(x(p)) \in \mathrm{GL}(n, \mathbb{R}) \subset \mathbb{R}^{n \times n}.$$

The smooth compatibility of x and y implies that $g : \mathcal{U} \cap \mathcal{V} \to \mathrm{GL}(n, \mathbb{R})$ is also a smooth function. The existence of maps such as Φ_x and Φ_y is one way of making precise the notion that the tangent spaces $T_p M$ vary smoothly with $p \in M$. We take this as motivation for the definition below.

NOTATION. In everything that follows, we choose a field

$$\mathbb{F} = \text{either } \mathbb{R} \text{ or } \mathbb{C},$$

and assume unless otherwise noted that all vector spaces and linear maps are $\mathbb{F}$-linear. In this way the real and complex cases can be handled simultaneously.

DEFINITION 16.1. Assume M is a smooth n-manifold, E_p is an m-dimensional vector space over $\mathbb{F}$ associated to each point $p \in M$, and define the set

$$E := \bigcup_{p \in M} E_p,$$

where E_p and E_q are regarded as disjoint sets for $p \neq q$.[51] For any subset $\mathcal{U} \subset M$, denote

$$E|_{\mathcal{U}} := \bigcup_{p \in \mathcal{U}} E_p \subset E.$$

A **local trivialization** (*lokale Trivialisierung*) of E is a pair $(\mathcal{U}_\alpha, \Phi_\alpha)$ consisting of an open subset $\mathcal{U}_\alpha \subset M$ and a bijection

$$E|_{\mathcal{U}_\alpha} \xrightarrow{\Phi_\alpha} \mathcal{U}_\alpha \times \mathbb{F}^m$$

such that for each $p \in \mathcal{U}_\alpha$, Φ_α restricts to E_p as a map of the form $v \mapsto (p, \phi_p v)$ for some vector space isomorphism $\phi_p : E_p \to \mathbb{F}^m$.

Any two local trivializations $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ determine **transition functions** (*Übergangsfunktionen*) $g_{\beta\alpha}, g_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ such that the map $\Phi_\beta \circ \Phi_\alpha^{-1} : (\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathbb{F}^m \to (\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathbb{F}^m$ and its inverse take the form

$$(16.1) \qquad \begin{aligned} \Phi_\beta \circ \Phi_\alpha^{-1}(p, v) &= (p, g_{\beta\alpha}(p)v), \\ \Phi_\alpha \circ \Phi_\beta^{-1}(p, v) &= (p, g_{\alpha\beta}(p)v). \end{aligned}$$

We say that $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ are C^k-compatible for $k \in \mathbb{N} \cup \{0, \infty\}$ (or **smoothly compatible** in the case $k = \infty$) if the transition functions $g_{\beta\alpha}$ and $g_{\alpha\beta}$ are of class C^k.

EXERCISE 16.2. Show that the two transition functions $g_{\alpha\beta}, g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ in Definition 16.1 are related to each other by $g_{\beta\alpha}(p) = [g_{\alpha\beta}(p)]^{-1} \in \mathrm{GL}(m, \mathbb{F})$ for all $p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$, and conclude that $g_{\alpha\beta}$ is of class C^k if and only if $g_{\beta\alpha}$ is.

---

[51]In set-theoretic terms, this means we are defining E as the disjoint union of all the sets E_p, so we could also have written $E = \coprod_{p \in M} E_p$. We prefer however to avoid the use of the symbol "$\coprod$" here, because we will soon define a topology on E, and it will not be the disjoint union topology.

REMARK 16.3. The notion of C^k-compatibility for transition functions is based on the premise that we know what it means to say that a real or complex matrix-valued function on a smooth manifold is of class C^k. This is fine because $\mathbb{R}^{n \times n}$ and $\mathbb{C}^{n \times n}$ can both be regarded as finite-dimensional real vector spaces (every complex vector space is also a real vector space), and the notion of smoothness for functions $f : M \to V$ is well defined whenever M is a smooth manifold and V is a real vector space. The notion of smoothness would be much less clear if we replaced $\mathbb{F}$ with a different field such as $\mathbb{Z}_2$ or $\mathbb{Q}$; there is no theory of differential calculus for functions on open subsets of $\mathbb{R}^n$ with values only in $\mathbb{Z}_2$ or $\mathbb{Q}$. That is one of a few reasons why we will never consider such generalizations in this course.

DEFINITION 16.4. Assume M is a manifold. A **vector bundle of class C^k with rank m over** M (*ein Vektorbündel von der Klasse C^k mit Rang m über M*) is a collection of m-dimensional vector spaces $E = \bigcup_{p \in M} E_p$ as in Definition 16.1, equipped with a maximal collection of C^k-compatible local trivializations $\{(\mathcal{U}_\alpha, \Phi_\alpha)\}_{\alpha \in I}$ such that $M = \bigcup_{\alpha \in I} \mathcal{U}_\alpha$. The vector spaces E_p for $p \in M$ are called the **fibers** (*Fasern*) of the vector bundle E, M is called the **base** (*Basis*) of E, and the set E itself is called the **total space** (*Totalraum*). The surjective map

$$\pi : E \to M$$

sending each fiber $E_p \subset E$ to the point $p \in M$ is sometimes called the **bundle projection**. We will denote the rank of E by

$$\mathrm{rank}_{\mathbb{F}}(E) := m \geqslant 0,$$

or simply $\mathrm{rank}(E)$ whenever the field $\mathbb{F}$ is clear from context.

EXERCISE 16.5. By identifying $\mathbb{C}^m$ with $\mathbb{R}^{2m}$, show that every complex vector bundle E of class C^k can also be regarded as a real vector vector bundle of class C^k with

$$\mathrm{rank}_{\mathbb{R}}(E) = 2\,\mathrm{rank}_{\mathbb{C}}(E).$$

REMARK 16.6. A vector bundle of rank m is also sometimes called an m-**plane bundle** or an "m-dimensional" vector bundle, and in the case $m = 1$, a **line bundle** (*Geradenbündel*). The latter terminology is quite intuitive when $\mathbb{F} = \mathbb{R}$, but one must keep in mind that in the complex case, the fibers should be visualized as *planes* rather than lines.

NOTATION. We will often refer to the vector bundle in Definition 16.4 simply as E, but doing so ignores quite a lot of important information, such as the base manifold M, fibers E_p, their vector space structures and the local trivializations. It is common in the literature to abbreviate all this data in terms of the projection map and thus refer to $\pi : E \to M$ or (E, π) as a vector bundle, sometimes also omitting the symbol π and writing

$$E \to M.$$

This is an imperfect convention, but we will sometimes also follow it: the projection map has the advantage that it determines the fibers

$$E_p = \pi^{-1}(p),$$

even though it does not determine their vector space structures or the local trivializations.

Observe that if M is a manifold of class C^ℓ for some finite ℓ, then vector bundles of class C^k make sense for every $k \leqslant \ell$, but cannot be defined for $k > \ell$. As usual, we will mostly only consider the case $k = \ell = \infty$, and then refer to E as a **smooth vector bundle**. We also call E a *real* vector bundle if $\mathbb{F} = \mathbb{R}$, and a *complex* vector bundle if $\mathbb{F} = \mathbb{C}$.

REMARK 16.7. The maximal collection of local trivializations $\{(\mathcal{U}_\alpha, \Phi_\alpha)\}_{\alpha \in I}$ in Definition 16.4 plays a similar role to the maximal atlas on a smooth manifold; maximality serves as a bookkeeping device to make sure in this setting that whenever $\{(\mathcal{U}_\alpha, \Phi_\alpha)\}_{\alpha \in I}$ and $\{(\mathcal{V}_\beta, \Psi_\beta)\}_{\beta \in J}$ are two coverings of E by smoothly compatible local trivializations such that every $(\mathcal{U}_\alpha, \Phi_\alpha)$ is smoothly compatible with every $(\mathcal{V}_\beta, \Psi_\beta)$, both can be understood as defining *the same* smooth vector bundle. As with manifolds, one never actually needs to specify a maximal collection of local trivializations, as a maximal collection is uniquely determined by *any* collection $\{(\mathcal{U}_\alpha, \Phi_\alpha)\}$ for which the sets $\mathcal{U}_\alpha$ cover M. When E is a smooth vector bundle, a local trivialization will be called **smooth** whenever it belongs to the associated maximal collection.

REMARK 16.8. Vector bundles of class C^0, also known as *topological* vector bundles, can be defined without assuming the base M is a manifold—the definition makes sense with an arbitrary topological space in place of M, and one can show that E then admits a natural topology such that the bundle projection $\pi : E \to M$ is continuous and the local trivializations are homeomorphisms. (The definition that appears in topology books usually assumes that E is *given* with a topology such that $\pi : E \to M$ is continuous and the fibers $E_p = \pi^{-1}(p)$ are vector spaces; one then calls $\pi : E \to M$ a vector bundle if and only if every $p \in M$ admits a neighborhood $\mathcal{U}$ for which there exists a homeomorphism $\Phi : \pi^{-1}(\mathcal{U}) \to \mathcal{U} \times \mathbb{F}^m$ that is a local trivialization.) For many applications, it is also advisable to assume that M is a paracompact Hausdorff space, so that partitions of unity can be used for various constructions, e.g. one can endow the fibers E_p with inner products that depend continuously on p, analogous to a Riemannian metric.

REMARK 16.9. The notion of C^k-compatibility between two local trivializations $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ could have been defined without mentioning the transition functions $g_{\beta\alpha}, g_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$, as it would be equivalent to require that the maps $\Phi_\beta \circ \Phi_\alpha^{-1}$ and $\Phi_\alpha \circ \Phi_\beta^{-1}$ are of class C^k on $(\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathbb{F}^m$. In more general contexts, in particular when we talk next semester about *fiber bundles* whose fibers are smooth manifolds rather than vector spaces, it will be necessary to reformulate the notion of smooth compatibility without the transition functions $g_{\alpha\beta}$ and $g_{\beta\alpha}$, as these naturally take values in the diffeomorphism group $\mathrm{Diff}(F)$ of some manifold F, and defining "smoothness" for maps with values in $\mathrm{Diff}(F)$ is something of a technical minefield. We do not have this problem with vector bundles, due to the fact that $\mathrm{GL}(m, \mathbb{F})$ is naturally a smooth finite-dimensional manifold, and (16.1) shows moreover that the transition functions encode all of the essential information in this setting. It will be especially useful to focus on them when we start talking about vector bundles with extra geometric structure such as bundle metrics or volume forms. In reality, this is also true for most fiber bundles that are of interest, because instead of considering $g_{\alpha\beta}$ and $g_{\beta\alpha}$ with values in $\mathrm{Diff}(F)$, one can often take them to have values in some finite-dimensional smooth Lie group G that acts smoothly on the manifold F. We will see many examples of this next semester.

Here is a generalization of the fact that tangent bundles are smooth manifolds.

PROPOSITION 16.10. *For any smooth vector bundle $\pi : E \to M$ over a smooth manifold M, the total space E naturally has the structure of a smooth manifold of dimension*

$$\dim E = \begin{cases} \dim M + \mathrm{rank}(E) & \text{if } \mathbb{F} = \mathbb{R}, \\ \dim M + 2\,\mathrm{rank}(E) & \text{if } \mathbb{F} = \mathbb{C}, \end{cases}$$

such that the projection map π and the inclusions $E_p \hookrightarrow E$ for $p \in M$ and

$$i : M \hookrightarrow E : p \mapsto 0 \in E_p$$

are all smooth maps.

PROOF. The proof is analogous to that of Corollary 3.12, which was the case $E = TM$. The key point is that M can be covered by open sets $\mathcal{U}_\alpha \subset M$ which are domains of charts $x_\alpha : \mathcal{U}_\alpha \to \mathbb{R}^n$ and also appear in local trivializations $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$. The map

$$(16.2) \qquad \phi_\alpha := (x_\alpha \times \mathbb{1}) \circ \Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathbb{R}^n \times \mathbb{F}^m$$

is then an $(n + m)$-dimensional chart for E on the domain $E|_{\mathcal{U}_\alpha} \subset E$ if $\mathbb{F} = \mathbb{R}$, or if $\mathbb{F} = \mathbb{C}$, an $(n + 2m)$-dimensional chart after identifying $\mathbb{C}^m$ with $\mathbb{R}^{2m}$. The smooth compatibility of the charts $(\mathcal{U}_\alpha, x_\alpha)$ and local trivializations $(\mathcal{U}_\alpha, \Phi_\alpha)$ implies (exercise!) that all charts of this form on E are likewise smoothly compatible. The topology defined on E via these charts is metrizable and separable for the same reasons as in the case $E = TM$; in particular, one can use a partition of unity on M to construct a Riemannian metric on the total space E as in Exercise 15.20, proving that E is metrizable. $\qquad\square$

DEFINITION 16.11. A **section** (*Schnitt*) of a vector bundle $\pi : E \to M$ is a map $s : M \to E$ such that $\pi \circ s = \mathrm{Id}_M$. In other words, s assigns to each point $p \in M$ a vector in the corresponding fiber $s(p) \in E_p$. We say s is a section **of class** C^k if it is a C^k-map $M \to E$ with respect to the smooth structure on E defined in Proposition 16.10. The vector space of smooth sections is denoted by

$$\Gamma(E) := \left\{ s \in C^\infty(M, E) \mid \pi \circ s = \mathrm{Id}_M \right\},$$

with addition and scalar multiplication in $\Gamma(E)$ defined pointwise, e.g. $s + t \in \Gamma(E)$ is defined for $s, t \in \Gamma(E)$ by $(s + t)(p) = s(p) + t(p) \in E_p$.

You might find it unsurprising but not completely obvious that $s + t$ is always a *smooth* section whenever s and t are. To make this obvious, we need to reformulate slightly the meaning of smoothness for a section $s : M \to E$. We observe first that for any local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$, every section $s : M \to E$ uniquely determines a vector-valued function

$$s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$$

such that

$$\Phi_\alpha(s(p)) = (p, s_\alpha(p)) \qquad \text{for all } p \in \mathcal{U}_\alpha.$$

We will call this the **local representation** of s with respect to the trivialization $(\mathcal{U}_\alpha, \Phi_\alpha)$. After shrinking $\mathcal{U}_\alpha$ if necessary to a smaller neighborhood of any given point in $\mathcal{U}_\alpha$, we are free to assume that it is also the domain of a chart $x_\alpha : \mathcal{U}_\alpha \to \mathbb{R}^n$, in which case (16.2) defines a corresponding chart $\phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathbb{R}^n \times \mathbb{F}^m$ for E with the convenient property that its domain contains $s(p)$ for every $p \in \mathcal{U}_\alpha$. Using the charts x_α on M and ϕ_α on E, we obtain a local coordinate representation for the map $s : M \to E$, in the form

$$\phi_\alpha \circ s \circ x_\alpha^{-1} : x(\mathcal{U}_\alpha) \to x(\mathcal{U}_\alpha) \times \mathbb{F}^m : q \mapsto (q, s_\alpha \circ x_\alpha^{-1}(q)).$$

By definition, $s : M \to E$ is a smooth map if and only if this local coordinate representation is smooth for every choice of smooth chart $(\mathcal{U}_\alpha, x_\alpha)$ on M and smooth local trivialization $(\mathcal{U}_\alpha, \Phi_\alpha)$ of E. The latter is clearly true if and only if s_α is a smooth function, so we've proved:

PROPOSITION 16.12. *A section $s : M \to E$ is smooth if and only if its local coordinate representations $s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$ with respect to arbitrary smooth local trivializations $(\mathcal{U}_\alpha, \Phi_\alpha)$ are all smooth.* $\qquad\square$

Since $C^\infty(\mathcal{U}_\alpha, \mathbb{F}^m)$ is a vector space for every open set $\mathcal{U}_\alpha$, Proposition 16.12 implies that $\Gamma(E)$ is also a vector space.

EXERCISE 16.13. Show that if $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ are two local trivializations of E and $s : M \to E$ is a section, then the local representations $s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$ and $s_\beta : \mathcal{U}_\beta \to \mathbb{F}^m$ are related to each other on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ in terms of the transition function $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ by

$$s_\beta(p) = g_{\beta\alpha}(p)s_\alpha(p) \qquad \text{for } p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta.$$

Remark: Since the transition functions on a smooth vector bundle are all smooth, this exercise implies that the condition in Proposition 16.12 does not need to be checked for all possible smooth local trivializations—it suffices to check it for a family of trivializations that cover M.

DEFINITION 16.14. Assume $E \to M$ and $F \to M$ are two smooth vector bundles over the same manifold M. A smooth map $\Psi : E \to F$ is called a **smooth linear bundle map** if for every $p \in M$, the restriction $\Psi|_{E_p}$ is a linear map

$$\Psi_p : E_p \to F_p.$$

We call Ψ **fiberwise injective / surjective** if Ψ_p is injective / surjective for every $p \in M$, and Ψ is a **bundle isomorphism** if Ψ_p is a vector space isomorphism for every $p \in M$. The bundles E and F are called **isomorphic** if and only if there exists a bundle isomorphism $E \to F$.

REMARK 16.15. Definition 16.14 presumes that E and F are both bundles over the same field $\mathbb{F}$. If one is a real vector bundle and the other is complex, then one can always regard the complex bundle as a real bundle with twice the rank (see Exercise 16.5) and thus interpret $\Psi : E \to F$ as a smooth *real*-linear bundle map.

EXERCISE 16.16. Suppose $E, F \to M$ are smooth vector bundles and $\Psi : E \to F$ is a map whose restriction to E_p for each p is a linear map $\Psi_p : E_p \to F_p$.

(a) Show that for every pair of smooth local trivializations $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ and $\Psi_\beta : E|_{\mathcal{U}_\beta} \to \mathcal{U}_\beta \times \mathbb{F}^k$, there exists a unique function

$$\Psi_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathbb{F}^{k \times m}$$

such that

$$\Phi_\beta \circ \Psi \circ \Phi_\alpha^{-1} : (\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathbb{F}^m \to (\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \times \mathbb{F}^k : (p, v) \mapsto (p, \Psi_{\beta\alpha}(p)v).$$

(b) Show that Ψ is a smooth linear bundle map if and only if for all choices of the two smooth local trivializations in part (a), the function $\Psi_{\beta\alpha}$ is smooth.

DEFINITION 16.17. Given a smooth vector bundle $E \to M$, a **smooth subbundle** (*Unterbündel*) of E is a vector bundle $F \to M$ such that for each $p \in M$, F_p is a linear subspace of E_p, and the inclusion $F \hookrightarrow E$ is a smooth linear bundle map.

16.2. Some basic examples. We now relate the definitions above to various examples that have already appeared in this course. For several of them, there is still some work to be done in showing that they naturally admit coverings by families of smoothly compatible local trivializations, and this work will be postponed until the next lecture.

EXAMPLE 16.18 (**tangent bundle**). If M is an n-manifold, its tangent bundle TM is a smooth real vector bundle of rank n, where each smooth chart $(\mathcal{U}, x)$ determines a local trivialization $\Phi : TM|_{\mathcal{U}} = T\mathcal{U} \to \mathcal{U} \times \mathbb{R}^n$ by $\Phi(X) = (p, d_px(X))$ for $X \in T_pM$. A smooth section of TM is nothing other than a smooth vector field on M,

$$\Gamma(TM) = \mathfrak{X}(M).$$

EXAMPLE 16.19 (**cotangent bundle**). The cotangent bundle T^*M of a smooth n-manifold M has fibers $T_p^*M = \text{Hom}(T_pM, \mathbb{R})$ for $p \in M$. We will construct smoothly compatible local trivializations for T^*M in the next lecture—it is a special case of the fact that every smooth vector bundle has a *dual bundle* which is also a smooth vector bundle in a natural way. The smooth sections of T^*M will then be the smooth 1-forms on M,

$$\Gamma(T^*M) = \Omega^1(M).$$

EXAMPLE 16.20 (**tensor and exterior bundles**). For each $k, \ell \geqslant 0$, there is a natural smooth real vector bundle $T_\ell^k M \to M$ of rank $n^{k+\ell}$ whose fiber at a point p is the vector space $(T_pM)_\ell^k$ of $(k+\ell)$-fold multilinear maps $T_p^*M \times \ldots \times T_p^*M \times T_pM \times \ldots \times T_pM \to \mathbb{R}$. The smooth local trivializations on $T_\ell^k M$ will also arise from more general constructions to be discussed in the next lecture. Consistently with our previous notation, the space of smooth sections $\Gamma(T_\ell^k M)$ will then be precisely the space of smooth tensor fields of type (k, ℓ).

For each $k \geqslant 0$, there is an important subbundle

$$\Lambda^k T^*M \subset T_k^0 M$$

of rank $\binom{n}{k}$ whose fiber over $p \in M$ is the vector space of alternating k-forms $\Lambda^k T_p^*M \subset (T_pM)_k^0$. The sections of $\Lambda^k T^*M$ will of course be the smooth differential k-forms,

$$\Gamma(\Lambda^k T^*M) = \Omega^k(M).$$

Note that by definition,

$$T_1^0 M = T^*M = \Lambda^1 T^*M,$$

and since $(T_pM)_0^0$ is defined simply as $\mathbb{R}$ for every p, $T_0^0 M = \Lambda^0 T^*M$ is simply the *trivial* line bundle $M \times \mathbb{R}$ (cf. Example 16.21 below).

EXAMPLE 16.21 (**trivial bundle**). For any manifold M, the **trivial m-plane bundle** over M is the product $E = M \times \mathbb{F}^m$, with fibers

$$E_p := \{p\} \times \mathbb{F}^m,$$

understood in the obvious way as m-dimensional vector spaces. This is a smooth vector bundle because (M, Id) is a local trivialization that covers the entirety of M, so the associated maximal collection of local trivializations consists of all that are smoothly compatible with this one. Smooth sections $s : M \to M \times \mathbb{F}^m$ are smooth maps of the form $p \mapsto (p, f(p))$ and are thus equivalent to smooth functions $f : M \to \mathbb{F}^m$.

DEFINITION 16.22. A vector bundle $\pi : E \to M$ of rank m is (globally) **trivial**[52] if it admits a bundle isomorphism to the trivial m-plane bundle over M.

A local trivialization $\Phi : E|_\mathcal{U} \to \mathcal{U} \times \mathbb{F}^m$ of a vector bundle E can be understood as a bundle isomorphism between the restriction $E|_\mathcal{U} \to \mathcal{U}$ and the trivial m-plane bundle over $\mathcal{U}$. By definition, every vector bundle is therefore *locally* trivial, meaning that its restriction to any sufficiently small open subset must be trivial. The next example shows that globally, this need not be true.

EXAMPLE 16.23 (**a nontrivial real line bundle**). Identify S^1 with the unit circle in $\mathbb{C}$, and define $\ell \subset S^1 \times \mathbb{R}^2$ as the union of the sets $\{e^{i\theta}\} \times \ell_{e^{i\theta}} \subset S^1 \times \mathbb{R}^2$ for all $\theta \in \mathbb{R}$, where $\ell_{e^{i\theta}} \subset \mathbb{R}^2$ is the 1-dimensional subspace

$$\ell_{e^{i\theta}} = \mathbb{R}\begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \subset \mathbb{R}^2.$$

---

[52]If we were being more pedantic, we would say **globally trivializable** in Definition 16.22 instead of "trivial", and reserve the latter for any vector bundle that is literally presented as a product $M \times \mathbb{F}^m$ with the identity map as a smooth trivialization, rather than just being isomorphic to one. But the looser use of the word "trivial" to mean "isomorphic to a trivial bundle" is widespread, so you should get used to it.

Exercise 16.24 below shows that ℓ can be regarded as a smooth line bundle over S^1 with fibers $\ell_{e^{i\theta}}$ for $e^{i\theta} \in S^1$. Observe that if we consider the subset

$$\{(e^{i\theta}, v) \in \ell \mid \theta \in \mathbb{R}, |v| \leqslant 1\}$$

consisting only of vectors of length at most 1, we obtain a Möbius strip. Local trivializations of $\ell \to S^1$ can be constructed as follows: for any $\theta_0 \in \mathbb{R}$, set $p := e^{i\theta_0} \in S^1$, and define

$$(16.3) \qquad \Phi : \ell|_{S^1 \setminus \{p\}} \to (S^1 \setminus \{p\}) \times \mathbb{R} : \left(e^{i\theta}, c \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \right) \mapsto (e^{i\theta}, c),$$

with θ assumed to vary in the interval $(\theta_0, \theta_0 + 2\pi)$.

EXERCISE 16.24. For the line bundle $\ell \to S^1$ in Example 16.23, prove:
  (a) Any two local trivializations defined as in (16.3) with different choices of $\theta_0 \in \mathbb{R}$ are smoothly compatible.
  (b) ℓ is a smooth subbundle of the trivial bundle $S^1 \times \mathbb{R}^2$.
  (c) There exists no continuous section of ℓ that is nowhere zero.
  (d) ℓ is not globally trivial.

17. Constructions of vector bundles

17.1. Local frames and components. Local trivializations of a vector bundle are generally not very easy to visualize, which makes them tricky in practice to construct. We now introduce a notion that is equivalent, but arguably easier to work with. Recall that if $E \to M$ is a smooth vector bundle and $\mathcal{U} \subset M$ is a subset, we denote the union of all the fibers over points in $\mathcal{U}$ by

$$E|_{\mathcal{U}} := \bigcup_{p \in \mathcal{U}} E_p,$$

and call this the **restriction** of E to $\mathcal{U}$. It should be clear that if $\mathcal{U} \subset M$ is an *open* subset, then $E|_{\mathcal{U}}$ is a smooth vector bundle over $\mathcal{U}$ in a natural way. (We will generalize this below to the case where $\mathcal{U} \subset M$ is an arbitrary submanifold.) The space $\Gamma(E|_{\mathcal{U}})$ of smooth sections of $E|_{\mathcal{U}}$ thus consists of all smooth maps $\mathcal{U} \to E$ that send each point $p \in \mathcal{U}$ to an element of the corresponding fiber E_p. It often happens with bundles that a section $s \in \Gamma(E)$ with certain desirable properties cannot be assumed to exist *globally*, but does exist *locally*, meaning that for any sufficiently small open subset $\mathcal{U} \subset M$, a section of $E|_{\mathcal{U}}$ with those properties can be found. We sometimes refer to sections of the restricted bundle $E|_{\mathcal{U}}$ as **local sections** of E over the subset $\mathcal{U} \subset M$.

DEFINITION 17.1. For a vector bundle $E \to M$ and open set $\mathcal{U} \subset M$, a **frame** for E over $\mathcal{U}$ is a tuple of local sections $e_1, \ldots, e_m : \mathcal{U} \to E$ of E over $\mathcal{U}$ such that for every $p \in \mathcal{U}$, the vectors $e_1(p), \ldots, e_m(p)$ form a basis of E_p. We call $e_1, \ldots, e_m$ a **smooth frame** if the sections are smooth.

Having a basis $e_1(p), \ldots, e_m(p)$ for each fiber E_p means that in the region where the frame is defined, we can talk about **components**: every $v \in E_p$ for $p \in \mathcal{U}$ is of the form

$$(17.1) \qquad\qquad\qquad v = v^j e_j$$

for unique real or complex numbers $v^1, \ldots, v^m \in \mathbb{F}$. Note that the Einstein summation convention is in effect in (17.1), and we will continue using it in similar expressions wherever possible: since the possible values of j on the right hand side are $1, \ldots, m$, it means in this case that there is an implied summation $\sum_{j=1}^m$ but the summation symbol has been omitted. Any section $s : M \to E$ is now uniquely expressible over $\mathcal{U}$ in terms of component functions $s^1, \ldots, s^m : \mathcal{U} \to \mathbb{F}$, namely as

$$s(p) = s^j(p)e_j(p).$$

The proof of the following statement is more-or-less immediate:

PROPOSITION 17.2. *Over any open set $\mathcal{U} \subset M$, there is a natural bijective correspondence between frames $e_1, \ldots, e_m : \mathcal{U} \to E$ and local trivializations $\Phi : E|_{\mathcal{U}} \to \mathcal{U} \times \mathbb{F}^m$, such that Φ is defined in terms of $e_1, \ldots, e_m$ by*

$$\Phi(v^i e_i(p)) = (p, (v^1, \ldots, v^m)).$$

Conversely, Φ determines $e_1, \ldots, e_m$ by

$$e_i(p) = \Phi^{-1}(p, \mathbf{e}_i),$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_m$ denotes the standard basis of $\mathbb{F}^m$. $\square$

EXAMPLE 17.3. On the tangent bundle $TM \to M$, the local trivialization determined by a chart $(\mathcal{U}, x)$ on M corresponds to the frame over $\mathcal{U}$ defined via the coordinate vector fields $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \in \Gamma(TM|_{\mathcal{U}})$.

Recall from the previous lecture that every local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ associates to each section $s : M \to E$ a function $s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$ such that $\Phi_\alpha(s(p)) = (p, s_\alpha(p))$. If $e_1^\alpha, \ldots, e_m^\alpha$ denotes the local frame corresponding to Φ_α, then we can also write $s(p) = s^i(p)e_i^\alpha(p)$ for unique component functions $s^i : \mathcal{U}_\alpha \to \mathbb{F}$, and the correspondence in Proposition 17.2 gives

$$\Phi_\alpha(s(p)) = \Phi_\alpha(s^i(p)e_i^\alpha(p)) = (p, (s^1(p), \ldots, s^m(p))) = (p, s_\alpha(p)).$$

This shows that the vector-valued local representation $s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$ is made up of the component functions $s^1, \ldots, s^m$ with respect to the frame:

$$s_\alpha(p) = (s^1(p), \ldots, s^m(p)) \in \mathbb{F}^m.$$

If E is a smooth vector bundle, we conclude from this and Proposition 16.12 that a section $s : M \to E$ is smooth if and only if for every smooth local trivialization $(\mathcal{U}, \Phi)$, the component functions $s^1, \ldots, s^m : \mathcal{U} \to \mathbb{F}$ with respect to the corresponding local frame are smooth.

Now let's think about smooth compatibility: suppose $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ are two local trivializations related by the transition functions $g_{\beta\alpha}, g_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$, and denote the corresponding local frames by $e_1^\alpha, \ldots, e_m^\alpha : \mathcal{U}_\alpha \to E$ and $e_1^\beta, \ldots, e_m^\beta : \mathcal{U}_\beta \to E$. On $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$, each of the sections e_i^α has uniquely-defined components with respect to the other frame $e_1^\beta, \ldots, e_m^\beta$, giving functions $h_i{}^j : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathbb{F}$ such that

$$e_i^\alpha = h_i{}^j e_j^\beta \qquad \text{on } \mathcal{U}_\alpha \cap \mathcal{U}_\beta.$$

For $p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$, let us denote by $\mathbf{h}(p) \in \mathbb{F}^{m \times m}$ the matrix whose ith row and jth column is $h_i{}^j(p)$. For any $\mathbf{v} = (v^1, \ldots, v^m) \in \mathbb{F}^m$, the correspondence in Proposition 17.2 then gives

$$\Phi_\beta \circ \Phi_\alpha^{-1}(p, \mathbf{v}) = \Phi_\beta(v^i e_i^\alpha(p)) = \Phi_\beta(v^i h_i{}^j(p)e_j^\beta(p)) = \Phi_\beta(v^j h_j{}^i(p)e_i^\beta(p)) = (p, \mathbf{h}(p)^T \mathbf{v}),$$

implying that the transition function $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ is given by

$$g_{\beta\alpha}(p) = \mathbf{h}(p)^T.$$

In particular, the matrix-valued function $g_{\beta\alpha}$ is exactly as smooth as the least smooth among the scalar-valued functions $h_i{}^j$, which are simply the components of the sections $e_1^\alpha, \ldots, e_m^\alpha$ with respect to the other frame. We've proved:

PROPOSITION 17.4. *Two local frames correspond to smoothly compatible local trivializations if and only if the component functions for the sections in each frame with respect to the other frame are all smooth.* $\square$

COROLLARY 17.5. *On a smooth vector bundle $E \to M$, a local trivialization $\Phi : E|_{\mathcal{U}} \to \mathcal{U} \times \mathbb{F}^m$ is smooth if and only if the sections forming the corresponding local frame $e_1, \ldots, e_m : \mathcal{U} \to E$ are smooth.*

PROOF. If Φ is a smooth local trivialization, then the local representations of the sections $e_1, \ldots, e_m$ with respect to Φ are constant, and thus smooth, implying via Proposition 16.12 that the sections are smooth. Conversely, smoothness of the sections $e_1, \ldots, e_m$ means that their components with respect to the frame corresponding to any smooth local trivialization are all smooth, which implies via Proposition 17.4 that Φ is smoothly compatible with any smooth local trivialization, and is therefore also smooth. $\qquad\square$

17.2. Pullbacks and restrictions. Suppose $f : M \to N$ is a smooth map and $E \to N$ is a smooth vector bundle. The **pullback** of $E \to N$ via f, also known as the **induced bundle**, is a smooth vector bundle
$$f^*E \to M$$
whose fiber over the point $p \in M$ is
$$(f^*E)_p := E_{f(p)}.$$
To see that this is naturally a smooth vector bundle, suppose $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is an open covering of N with smoothly compatible local trivializations $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$, and write
$$\Phi_\alpha(v) = (p, \Phi_{\alpha, p}v) \qquad \text{for } p \in \mathcal{U}_\alpha,\ v \in E_p,$$
defining vector space isomorphisms $\Phi_{\alpha, p} : E_p \to \mathbb{F}^m$. The sets $\{f^{-1}(\mathcal{U}_\alpha) \subset M\}_{\alpha \in I}$ then form an open covering of M, and for each $\alpha \in I$, we can define a local trivialization $f^*\Phi_\alpha$ of f^*E by
$$f^*\Phi_\alpha : (f^*E)|_{f^{-1}(\mathcal{U}_\alpha)} \to f^{-1}(\mathcal{U}_\alpha) \times \mathbb{F}^m,$$
$$v \mapsto (p, \Phi_{\alpha, f(p)}v) \qquad \text{for } p \in \mathcal{U}_\alpha,\ v \in (f^*E)_p = E_{f(p)}.$$
The transition function $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ relating Φ_α and Φ_β takes the form $g_{\beta\alpha}(p) = \Phi_{\beta, p}\Phi_{\alpha, p}^{-1}$ for $p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$, thus the map $(f^*\Phi_\beta) \circ (f^*\Phi_\alpha)^{-1}$ from $(f^{-1}(\mathcal{U}_\alpha) \cap f^{-1}(\mathcal{U}_\beta)) \times \mathbb{F}^m$ to itself is given by
$$(f^*\Phi_\beta) \circ (f^*\Phi_\alpha)^{-1}(p, v) = (p, \Phi_{\beta, f(p)}\Phi_{\alpha, f(p)}^{-1}) = (p, g_{\beta\alpha}(f(p))), \qquad p \in f^{-1}(\mathcal{U}_\alpha) \cap f^{-1}(\mathcal{U}_\beta),$$
and the resulting transition function $f^{-1}(\mathcal{U}_\alpha) \cap f^{-1}(\mathcal{U}_\beta) \to \mathrm{GL}(m, \mathbb{F})$ is therefore $g_{\beta\alpha} \circ f$. This is smooth, so f^*E is a smooth bundle.

REMARK 17.6. The above argument shows more generally that if $E \to N$ is a vector bundle of class C^k and $f : M \to N$ a map of class C^ℓ, then $f^*E \to M$ is a bundle of class $C^{\min\{k, \ell\}}$.

EXERCISE 17.7. In the situation above, show that the canonical map $f^*E \to E$ that sends $(f^*E)_p$ to $E_{f(p)}$ as the identity map for each $p \in M$ is smooth.

The map $f^*E \to E$ is a "fiberwise isomorphism" in the sense that it maps each fiber of f^*E isomorphically to a fiber of E, but it is not a *bundle map* in the sense defined in the previous lecture since f^*E and E are bundles over different manifolds. It is instead an example of the following more general notion:

DEFINITION 17.8. Assume $E \to M$ and $F \to N$ are two smooth vector bundles and $\psi : M \to N$ is a smooth map. A smooth map $\Psi : E \to F$ that sends each fiber E_p linearly to the fiber $F_{\psi(p)}$ is called **smooth linear bundle map covering** ψ.

Our previous notion of smooth linear bundle maps was the special case of Definition 17.8 in which $M = N$ and $\psi : M \to N$ is the identity map. For a bundle $E \to N$ and map $f : M \to N$, we can now understand the canonical map $f^*E \to E$ as a smooth linear bundle map covering f.

REMARK 17.9. Actually, a smooth linear bundle map $\Phi : E \to F$ covering a map $\psi : M \to N$ is more-or-less equivalent to a smooth linear bundle map from E to ψ^*F; the former is just the latter composed with the canonical map $\psi^*F \to F$.

EXAMPLE 17.10. For a smooth map $f : M \to N$, the fiber of the pullback bundle f^*TN over a point $p \in M$ is the tangent space $T_{f(p)}N$, and a section $X \in \Gamma(f^*TN)$ therefore associates to each $p \in M$ a tangent vector $X(p) \in T_{f(p)}N$. Sections of this type are called **vector fields along** f; they generalize the usual notion of a vector field on M, which is the special case where $M = N$ and f is the identity map. These objects arise naturally in the following context: suppose $\{f_t : M \to N\}_{t \in (-\epsilon, \epsilon)}$ is a smooth 1-parameter family of maps with $f := f_0$, where "smooth family" in this situation means that the map $(-\epsilon, \epsilon) \times M \to N : (t, p) \mapsto f_t(p)$ is smooth. Then

$$X(p) := \partial_t f_t(p)|_{t=0} \in T_{f(p)}N$$

defines a vector field along f. Informally, if one thinks of $C^\infty(M, N)$ as an infinite-dimensional manifold, this means that its tangent space at $f \in C^\infty(M, N)$ is $\Gamma(f^*TM)$. (With minor modifications, this statement can be made precise in the language of smooth Banach manifolds.)

EXAMPLE 17.11. If $N \subset M$ is a smooth submanifold and $i : N \hookrightarrow M$ denotes the inclusion map, then any smooth vector bundle $E \to M$ admits a **restriction** to N,

$$E|_N = i^*E \to N,$$

which is also a smooth vector bundle. (Its transition functions are just the restrictions of the transition functions of E to the submanifold.)

17.3. Subbundles, quotients, and normal bundles. The following result puts subbundles on a similar footing with sub*manifolds* by constructing the analogue of slice charts for local trivializations:

PROPOSITION 17.12. *Suppose $E \to M$ is a smooth vector bundle of rank m, $F \subset E$ is a subset, and denote for a point $p \in M$ or subset $\mathcal{U} \subset M$*

$$F_p := E_p \cap F, \qquad F|_{\mathcal{U}} := E|_{\mathcal{U}} \cap F.$$

The following statements are equivalent:

(1) *F is a smooth subbundle of rank k in the sense of Definition 16.17, i.e. it admits the structure of a smooth vector bundle of rank k such that the inclusion $F \hookrightarrow E$ is a smooth linear bundle map.*
(2) *For every $p \in M$, there exists a neighborhood $\mathcal{U} \subset M$ of p and a smooth local trivialization $\Phi : E|_{\mathcal{U}} \to \mathcal{U} \times \mathbb{F}^m$ of E such that*

$$\Phi(F|_{\mathcal{U}}) = \mathcal{U} \times (\mathbb{F}^k \times \{0\}) \subset \mathcal{U} \times \mathbb{F}^m.$$

PROOF. Suppose first that F is a smooth subbundle of rank k in the sense of Definition 16.17. Given $p \in M$, choose a smooth local trivialization $F|_{\mathcal{U}} \to \mathcal{U} \times \mathbb{F}^k$ of F with $p \in \mathcal{U}$ and let $e_1, \ldots, e_k \in \Gamma(F|_{\mathcal{U}})$ denote the corresponding smooth local frame. Since the inclusion $F \hookrightarrow E$ is a smooth linear bundle map, the $e_1, \ldots, e_k$ can equally well be regarded as smooth sections of $E|_{\mathcal{U}}$, and they are linearly independent at every point. After shrinking $\mathcal{U}$ if necessary, we can then use a local trivialization of E over $\mathcal{U}$ to find additional smooth sections $e_{k+1}, \ldots, e_m \in \Gamma(E|_{\mathcal{U}})$ for which $e_1, \ldots, e_m$ remain linearly independent and therefore form a basis of the fiber of E at every point in $\mathcal{U}$; the idea here is that in a local trivialization of E over $\mathcal{U}$, each of the sections $e_1, \ldots, e_k$ is identified with a smooth function $\mathcal{U} \to \mathbb{F}^m$ that can be assumed *nearly* constant after shrinking $\mathcal{U}$, so that it is easy to find $m - k$ constant functions that complete the basis at every point. With this understood, we now have a smooth local frame $e_1, \ldots, e_m \in \Gamma(E|_{\mathcal{U}})$ such that the sections $e_1, \ldots, e_k$ span the fiber of F over every point in $\mathcal{U}$. The corresponding local trivialization then has the desired property.

Conversely, if local trivializations of E with this property always exist, then it is clear that the sets $F_p \subset E_p$ are linear subspaces and the trivializations of E determine smoothly compatible local

trivializations of F by restriction. It is easy to check that the inclusion $F \hookrightarrow E$ is then a smooth map. (This step is analogous to the way that slice charts for a smooth submanifold $N \subset M$ are used to define a smooth structure on N so that the inclusion $N \hookrightarrow M$ is smooth.) $\qquad\square$

REMARK 17.13. It will be useful in the following to allow a mild generalization of our previous notion of a local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$. Specifically, nothing important changes if we replace the "standard" vector space $\mathbb{F}^m$ with any other m-dimensional vector space V and thus consider bijections of the form

$$\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times V$$

that map each fiber E_p isomorphically to $\{p\} \times V$. The transition functions relating two trivializations of this form take values in the group $\mathrm{GL}(V)$ of invertible $\mathbb{F}$-linear maps $V \to V$, which is an open subset of the (real or complex) vector space $\mathrm{End}(V)$. To reduce this to our previous notion, one only has to choose an isomorphism $\Psi : V \to \mathbb{F}^m$ and use it consistently, so that transition functions $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(V)$ become

$$\widetilde{g}_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F}) : p \mapsto \Psi g_{\beta\alpha}(p)\Psi^{-1};$$

clearly $g_{\beta\alpha}$ is smooth if and only if $\widetilde{g}_{\beta\alpha}$ is smooth.

Given a smooth vector bundle $E \to M$ of rank m and smooth subbundle $F \subset E$ of rank k, the **quotient bundle**

$$E/F \to M$$

is a smooth vector bundle of rank $m - k$ whose fiber over a point $p \in M$ is the quotient vector space

$$(E/F)_p := E_p/F_p.$$

One defines suitable local trivializations on E/F as follows: according to Proposition 17.12, we can find an open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of M and local trivializations $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ such that $\Phi_\alpha(F|_{\mathcal{U}_\alpha}) = \mathcal{U}_\alpha \times \mathbb{F}^k$, where $\mathbb{F}^k$ is identified with the linear subspace

$$\mathbb{F}^k := \mathbb{F}^k \times \{0\} \subset \mathbb{F}^m.$$

Writing $\Phi_\alpha(v) = (p, \Phi_{\alpha,p}v)$ for $p \in \mathcal{U}_\alpha$ and $v \in E_p$, it follows that the vector space isomorphism $\Phi_{\alpha,p} : E_p \to \mathbb{F}^m$ identifies the subspaces $F_p \subset E_p$ and $\mathbb{F}^k \subset \mathbb{F}^m$, thus it descends to an isomorphism of the quotient spaces,

$$\Phi_{\alpha,p} : E_p/F_p \to \mathbb{F}^m/\mathbb{F}^k : [v] \mapsto [\Phi_{\alpha,p}v].$$

A local trivialization

$$(E/F)|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times (\mathbb{F}^m/\mathbb{F}^k)$$

in the generalized sense of Remark 17.13 can thus be defined by sending $[v] \in E_p/F_p$ for $p \in \mathcal{U}_\alpha$ to $(p, [\Phi_{\alpha,p}v])$. Covering E/F with local trivializations defined in this way, the resulting transition functions are derived from the transition functions $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ of E by observing that since we chose the Φ_α to respect the subbundle $F \subset E$ as in Proposition 17.12, the linear map $g_{\beta\alpha}(p) : \mathbb{F}^m \to \mathbb{F}^m$ preserves the subspace $\mathbb{F}^k \subset \mathbb{F}^m$ for each $p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$ and thus descends to an isomorphism on the quotient $\mathbb{F}^m/\mathbb{F}^k$, determining a smooth function

$$\mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(\mathbb{F}^m/\mathbb{F}^k).$$

This is the transition function for the local trivializations we defined above on E/F from Φ_α and Φ_β.

EXERCISE 17.14. Show that for a smooth subbundle $F \subset E$ of $E \to M$, the natural surjective map $E \to E/F$ that restricts to the fiber over each point $p \in M$ as the quotient projection $E_p \to E_p/F_p$ is a smooth linear bundle map.

EXAMPLE 17.15. Suppose $N \subset M$ is a smooth k-dimensional submanifold in an m-manifold M. Any slice chart for N determines a local trivialization of TM that also has the property in Proposition 17.12 for the subset $TN \subset TM|_N$, thus producing a smooth subbundle

$$TN \subset TM|_N.$$

The quotient

$$\nu N := (TM|_N)/TN \to N$$

is called the **normal bundle** of the submanifold $N \subset M$.

One can gain a better intuitive picture of the normal bundle of a submanifold $N \subset M$ by choosing a Riemannian metric g on M and looking at the orthogonal complements

$$(T_pN)^\perp := \left\{ X \in T_pM \mid g(X, \cdot)|_{T_pN} = 0 \right\}$$

at points $p \in N$.

EXERCISE 17.16. Given a smooth submanifold N in a Riemannian manifold (M, g), prove:
  (a) $TN^\perp := \bigcup_{p \in N}(T_pN)^\perp$ is a smooth subbundle of $TM|_N$.
     Hint: Construct smooth local frames $X_1, \ldots, X_n$ for $TM|_N$ such that $X_1, \ldots, X_k$ are tangent to N and $X_{k+1}, \ldots, X_n$ lie in $(TN)^\perp$.
  (b) The composition of the inclusion $TN^\perp \hookrightarrow TM|_N$ with the fiberwise quotient projection $TM|_N \to TM|_N/TN$ from Exercise 17.14 defines a bundle isomorphism $TN^\perp \to \nu N$.

The "normal vector fields" along hypersurfaces $N \subset M$ we considered in Lectures 11 and 12 can now be understood as smooth sections of the bundle $(TN)^\perp$, which according to Exercise 17.16, is equivalent to the normal bundle of N.

17.4. Algebraic operations. Several natural operations that produce new vector spaces from old ones can now be generalized to the setting of vector bundles. In the following list, the smoothness of the bundles we construct can be verified easily by constructing local frames; we will leave the details as exercises.

17.4.1. *Direct sums.* The direct sum of two vector spaces V and W is the same thing as their Cartesian product,

$$V \oplus W := V \times W,$$

in which V and W can be identified naturally with the subspaces $V \times \{0\}$ and $\{0\} \times W$ respectively. When extending this notion to vector bundles, it becomes especially useful to distinguish between the symbols "$\oplus$" and "$\times$": in particular, the direct sum of two smooth vector bundles $E, F \to M$ of ranks m and k respectively is a bundle $E \oplus F \to M$ of rank $m + k$ with fibers

$$(E \oplus F)_p := E_p \oplus F_p = E_p \times F_p.$$

Notice that at the level of sets, the total space $E \oplus F = \bigcup_{p \in M}(E_p \times F_p)$ is not *at all* the same thing as the product $E \times F$. Any local trivializations of E and F over the same region can be combined in a natural way to produce a local trivialization of $E \oplus F$ over that region, and if one covers $E \oplus F$ with local trivializations constructed in this way with respect to an open covering $\{\mathcal{U}_\alpha\}_{\alpha \in I}$, one finds that the resulting transition functions $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m + k, \mathbb{F})$ take the form of block matrices,

$$g_{\beta\alpha}(p) = \begin{pmatrix} g_{\beta\alpha}^E(p) & 0 \\ 0 & g_{\beta\alpha}^F(p) \end{pmatrix},$$

where $g_{\beta\alpha}^E(p) \in \mathrm{GL}(m, \mathbb{F})$ and $g_{\beta\alpha}^F(p) \in \mathrm{GL}(k, \mathbb{F})$ are the transition functions for E and F respectively. Clearly $g_{\beta\alpha}$ is smooth if $g_{\beta\alpha}^E$ and $g_{\beta\alpha}^F$ are.

REMARK 17.17. One *can* define a "product" bundle $E \times F$ whose total space is the Cartesian product of E and F, but it is naturally a bundle over $M \times M$ rather than M. More generally, two bundles $E \to M$ and $F \to N$ over potentially different manifolds have a product which is a bundle over $M \times N$

17.4.2. *The dual bundle.* Any smooth vector bundle $E \to M$ has a **dual bundle**
$$E^* \to M$$
whose fiber over a point $p \in M$ is the dual space $E_p^* = \mathrm{Hom}(E_p, \mathbb{F})$. Any local frame $e_1, \ldots, e_m$ for E over an open subset $\mathcal{U} \subset M$ then determines a dual frame $e_*^1, \ldots, e_*^m$ for E^* via the usual notion of a dual basis, i.e. for each $p \in \mathcal{U}$,
$$e_*^i(p)\,(e_j(p)) = \delta_j^i.$$
It is a straightforward algebraic exercise to verify that whenever two frames for E on overlapping regions are smoothly compatible in the sense of Proposition 17.4, their dual frames are also smoothly compatible, so in this way one can cover E^* with smoothly compatible local trivializations, making it a smooth vector bundle. This establishes in particular that for any smooth n-manifold M, the **cotangent bundle**
$$T^*M \to M$$
is a smooth real vector bundle of rank n.

17.4.3. *The complex conjugate.* [53]

One can associate to any complex vector space V another complex vector space $\bar{V}$, which is defined as the same set with the same notion of vector addition but a different notion of scalar multiplication, defined as follows. Since V and $\bar{V}$ are identical sets, the identity map defines a canonical map between them, which we shall denote by

(17.2) $$V \to \bar{V} : v \mapsto \bar{v}.$$

In other words, $\bar{v}$ is our notation for the vector $v \in V$ when it is regarded as an element of $\bar{V}$. With this understood, multiplication of a scalar $\lambda \in \mathbb{C}$ by a vector $\bar{v} \in \bar{V}$ is defined by
$$\lambda \bar{v} := \overline{\bar{\lambda} v},$$
where for $\lambda = a + ib$ with $a, b \in \mathbb{R}$, we denote the complex conjugate by $\bar{\lambda} := a - ib$. Another way to say this is that multiplication of a *real* scalar by a vector in $\bar{V}$ is defined exactly the same as in V, so that V and $\bar{V}$ are identical as real vector spaces, but multiplication by i is defined in $\bar{V}$ with a sign change, i.e. $i\bar{v} = -\overline{iv}$. This makes the bijection in (17.2) an isomorphism of real vector spaces, but *not* an isomorphism of complex vector spaces, as it is not even a complex-linear map; it is instead complex *antilinear*.

The **conjugate** of a complex vector bundle $E \to M$ of rank m is now defined as another complex vector bundle
$$\bar{E} \to M$$
of rank m whose fiber over a point $p \in M$ is $\bar{E}_p$. Strictly speaking, E and $\bar{E}$ are the same set, and the identity map thus defines a canonical bijection between them which we will again denote by
$$E \to \bar{E} : v \mapsto \bar{v}.$$
They are different complex vector bundles because one cannot use the same local trivializations for both—any local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}^m$ of E defines a complex vector space isomorphism $\Phi_{\alpha,p} : E_p \to \mathbb{C}^m$ for every $p \in \mathcal{U}_\alpha$, but this map is not complex-linear when regarded as a bijection $\bar{E}_p \to \mathbb{C}^m$, it is antilinear. The solution is to compose it with a complex-antilinear

---

[53]This section is inessential and was skipped in the lecture, but is provided here for your information.

isomorphism $\mathbb{C}^m \to \mathbb{C}^m$ such as the complex conjugation map $z \mapsto \bar{z}$, and this produces a local trivialization $\bar{\Phi}_\alpha$ of $\bar{E}$ over the same set $\mathcal{U}_\alpha$, namely

$$\bar{E}|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}^m : \bar{v} \mapsto (p, \overline{\Phi_{\alpha,p} v})$$

for $p \in \mathcal{U}_\alpha$ and $\bar{v} \in \bar{E}_p$. The next exercise shows that the collection of all local trivializations of $\bar{E}$ constructed in this way makes $\bar{E} \to M$ a smooth vector bundle.

EXERCISE 17.18. Show that if $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ are two local trivializations of a complex vector bundle $E \to M$ related by a transition function $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{C})$, then the transition function relating the corresponding local trivializations $(\mathcal{U}_\alpha, \bar{\Phi}_\alpha)$ and $(\mathcal{U}_\beta, \bar{\Phi}_\beta)$ of $\bar{E}$ is given by

$$\mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{C}) : p \mapsto \overline{g_{\beta\alpha}(p)},$$

where the bar on the right hand side means the usual notion of complex conjugation for m-by-m complex matrices. In particular, this transition function is smooth if and only if $g_{\beta\alpha}$ is smooth.

A finite-dimensional complex vector space V is always isomorphic to its conjugate space $\bar{V}$ since the two spaces have the same dimension, but on the other hand, there is no *canonical* choice of isomorphism (remember that the map $v \mapsto \bar{v}$ does not count because it is not complex linear). Vector bundles provide a means for measuring in some precise way the non-existence of a canonical choice: we will see that in general, a complex vector bundle $E \to M$ and its conjugate bundle $\bar{E} \to M$ need not be isomorphic.

REMARK 17.19. In the setting of a complex manifold M, whose transition maps are holomorphic maps between open subsets of $\mathbb{C}^n$ so that the notion of holomorphic complex-valued functions on open subsets of M can be defined, one can also consider so-called *holomorphic vector bundles*, which are required to admit coverings by local trivializations such that all transition functions are holomorphic, and the notion of holomorphic sections can therefore be defined. Exercise 17.18 shows that if $E \to M$ is a holomorphic vector bundle, then its conjugate $\bar{E} \to M$ is naturally a smooth complex vector bundle but is not a *holomorphic* vector bundle in any natural way, as its transition functions are not holomorphic, they are antiholomorphic (i.e. they are complex conjugates of holomorphic functions).

17.4.4. *Tensor products.* We will not discuss in this course the abstract definition of the tensor product of two vector spaces V and W, but if both spaces are finite dimensional, one obtains an easy equivalent definition using the canonical identification of V and W with the duals of their dual spaces. For our purposes, if V and W have dual spaces V^* and W^*, then $V^* \otimes W^*$ can be defined as the vector space

$$V^* \otimes W^* := \{\text{bilinear maps } V \times W \to \mathbb{F}\},$$

with the tensor product $\lambda \otimes \mu \in V^* \otimes W^*$ of two elements $\lambda^* \in V^*$ and $\mu^* \in W^*$ defined by

$$(\lambda \otimes \mu)(v, w) := \lambda(v)\mu(w) \qquad \text{for } v \in V,\, w \in W.$$

Our definition of $V \otimes W$ is then actually a definition of $V^{**} \otimes W^{**}$, i.e.

$$V \otimes W := \{\text{bilinear maps } V^* \times W^* \to \mathbb{F}\},$$

and $v \otimes w \in V \otimes W$ for $v \in V$ and $w \in W$ will be the bilinear map $V^* \times W^* \to \mathbb{F}$ given by

$$(v \otimes w)(\lambda, \mu) := v(\lambda)w(\mu) := \lambda(v)\mu(w) \qquad \text{for } \lambda \in V^*,\, \mu \in W^*.$$

Given bases $v_1, \ldots, v_m \in V$ and $w_1, \ldots, w_k \in W$, one can easily check via evaluation on the corresponding dual bases of V^* and W^* that the mk distinct tensor products $v_i \otimes w_j$ form a basis of $V \otimes W$.

While it is a bit tedious from this perspective, one can also check that the tensor product is an associative operation, i.e. for any three finite-dimensional vector spaces V, W, X, there is a natural isomorphism

$$(V \otimes W) \otimes X \cong V \otimes (W \otimes X)$$

that identifies $(v \otimes w) \otimes x$ with $v \otimes (w \otimes x)$ for every $v \in V$, $w \in W$ and $x \in X$. For this reason we will usually not write the parentheses in such expressions, and arbitrary tensor products of finitely many finite-dimensional vector spaces $V_1, \ldots, V_k$ can also be defined without parentheses; in fact, there is a natural isomorphism of $V_1 \otimes \ldots \otimes V_k$ with the space of multilinear maps $V_1^* \times \ldots \times V_k^* \to \mathbb{F}$.

All of this extends to the context of smooth vector bundles E over a manifold M, after observing that the canonical isomorphisms $E_p \to E_p^{**}$ give rise to canonical bundle isomorphisms $E \to E^{**} := (E^*)^*$. For two smooth vector bundles $E, F \to M$ of ranks m and k respectively, the tensor product $E \otimes F \to M$ is thus a bundle of rank mk with fibers

$$(E \otimes F)_p := E_p \otimes F_p.$$

Given local frames $e_1, \ldots, e_m$ for E and $f_1, \ldots, f_k$ for F over $\mathcal{U}$, a local frame for $E \otimes F$ over $\mathcal{U}$ is given by the sections

$$e_i \otimes f_j : \mathcal{U} \to (E \otimes F)|_{\mathcal{U}}, \qquad i = 1, \ldots, m, \ j = 1, \ldots, k,$$

where the tensor product of sections is defined pointwise, meaning $(e_i \otimes f_j)(p) := e_i(p) \otimes f_j(p) \in E_p \otimes F_p$. It is again a straightforward exercise to check that for any smoothly compatible choices of frames for E and F on overlapping regions, the resulting frames for $E \otimes F$ are also smoothly compatible.

This discussion extends in an obvious way to arbitrary finite tensor products of vector bundles. In particular, we can now generalize the bundles $T_\ell^k M$ mentioned in Example 16.20 to

$$E_\ell^k := \underbrace{E \otimes \ldots \otimes E}_{k} \otimes \underbrace{E^* \otimes \ldots \otimes E^*}_{\ell},$$

with the convention that for $k = \ell = 0$, the fibers are just $(E_p)_0^0 := \mathbb{R}$ and E_0^0 is thus the trivial real line bundle $M \times \mathbb{R}$ over M. We also have $E_0^1 = E$ and $E_1^0 = E^*$.

For each $k \geqslant 0$, there is an important subbundle

$$\Lambda^k E \subset \underbrace{E \otimes \ldots \otimes E}_{k},$$

whose fibers are the spaces $\Lambda^k E_p$ of antisymmetric k-fold multilinear maps $E_p^* \times \ldots \times E_p^* \to \mathbb{R}$, i.e. in terms of our notation from Lecture 9, we are defining $\Lambda^k E_p := \Lambda^k V^*$ for $V := E_p^*$ after identifying E_p with its double dual. That $\Lambda^k E \subset E^{\otimes k}$ is a smooth subbundle follows mainly from the observation that any local frame $e_1, \ldots, e_m$ for E gives rise to a local frame for $\Lambda^k E$ on the same region, consisting of the k-fold wedge products

$$e_{i_1} \wedge \ldots \wedge e_{i_k}, \qquad i_1 < \ldots < i_k.$$

In particular, this makes $\Lambda^k T^* M$ into a smooth vector bundle with $\Gamma(\Lambda^k T^* M) = \Omega^k(M)$.

17.4.5. *Bundles of linear maps.* [54]

It is often useful to notice that for two vector spaces V, W, the space of linear maps $\mathrm{Hom}(V, W)$ is naturally isomorphic to the tensor product $V^* \otimes W$. Indeed:

---

[54]This section is provided for your information and will occasionally be referred to later, but it was not covered in the lecture due to lack of time.

EXERCISE 17.20. Show that for finite-dimensional vector spaces V and W, the identifying $\lambda \otimes w \in V^* \otimes W$ for each $\lambda \in V^*$ and $w \in W$ with the linear map $V \to W$ given by

$$(\lambda \otimes w)(v) := \lambda(v)w$$

uniquely determines an isomorphism $V^* \otimes W \to \mathrm{Hom}(V, W)$.

This gives us the quickest way to see that for any two smooth vector bundles $E, F \to M$ with rank m and k respectively, there exists a smooth vector bundle

$$\mathrm{Hom}(E, F) \to M$$

with rank km, having fibers $\mathrm{Hom}(E, F)_p := \mathrm{Hom}(E_p, F_p)$. In fact, $\mathrm{Hom}(E, F)$ is canonically isomorphic to the tensor product bundle $E^* \otimes F$, but without worrying about this, one can also just take Exercise 17.20 as a hint on how to define local frames for $\mathrm{Hom}(E, F)$: given frames $e_1, \dots, e_m$ for E and $f_1, \dots, f_k$ for F over a region $\mathcal{U} \subset M$, one takes the dual frame $e_*^1, \dots, e_*^m$ for E^* and defines a frame for $\mathrm{Hom}(E, F)$ over $\mathcal{U}$ consisting of the products $e_*^i \otimes f_j$, each interpreted at any point $p \in \mathcal{U}$ as the linear map $E_p \to F_p : v \mapsto e_*^i(p)(v)f_j(p)$. It is another straightforward exercise to show that any two local frames for $\mathrm{Hom}(E, F)$ constructed from smooth frames on E and F will be smoothly compatible.

We can now state a much more succinct version of one of the definitions in the previous lecture: given two smooth vector bundles $E, F \to M$, a *smooth linear bundle map* $E \to F$ is a smooth section of the bundle $\mathrm{Hom}(E, F)$.

In the case $\mathbb{F} = \mathbb{C}$, it is sometimes also useful to include complex *anti*-linear maps in the discussion, where a map $A : V \to W$ between two complex vector spaces is called **antilinear** if it satisfies

$$A(v + w) = Av + Aw, \qquad A(\lambda v) = \bar{\lambda}Av$$

for all $v, w \in V$ and $\lambda \in \mathbb{C}$. The space

$$\overline{\mathrm{Hom}}(V, W) := \left\{ A : V \to W \mid A \text{ complex antilinear} \right\}$$

is a complex vector space in a natural way, and the following exercise yields a useful alternative perspective on it in terms of the conjugate vector space (see §17.4.3).

EXERCISE 17.21. Assume V, W are finite-dimensional complex vector spaces with dual space V^*, W^* and conjugates $\bar{V}, \bar{W}$. Find natural isomorphisms between the following pairs of complex vector spaces.
    (a) $\mathrm{Hom}(\bar{V}, W)$ and $\overline{\mathrm{Hom}}(V, W)$.
    (b) $(\bar{V})^*$ and $\overline{V^*}$.
    (c) $\bar{V}^* \otimes W^*$ and the space of real-bilinear maps $V \times W \to \mathbb{C}$ that are complex antilinear in the first factor and complex linear in the second factor.

It follows from Exercise 17.21 that for smooth complex vector bundles $E, F \to M$, one can also define a smooth bundle

$$\overline{\mathrm{Hom}}(E, F) \to M$$

whose fiber at a point $p \in M$ is the space of complex-antilinear maps $E_p \to F_p$; this bundle is canonically isomorphic to $\bar{E}^* \otimes F$.

18. Vector bundles with extra structure

In this lecture we discuss various types of geometric structure that can be added to the fibers of a vector bundle, such as orientations and inner products. There is a useful way to incorporate all possible types of structures under a single umbrella in terms of the so-called *structure group* of a bundle, and this discussion requires an initial digression on the topic of Lie groups.

18.1. Some basic Lie groups. Roughly speaking, a Lie group is a group that is also a smooth manifold. I intend to discuss this subject in earnest in the followup to this course next semester, but for now, we need to become acquainted with a few of the basic examples and their properties. The first one is $\mathrm{GL}(m, \mathbb{F})$, which is naturally a manifold because it is an open subset of the (real or complex) vector space $\mathbb{F}^{m \times m}$.

DEFINITION 18.1. A **Lie subgroup of** $\mathrm{GL}(m, \mathbb{F})$ is a subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$ that is also a smooth submanifold. Its associated **Lie algebra** is the tangent space

$$\mathfrak{g} := T_{\mathbb{1}} G \subset T_{\mathbb{1}} \mathrm{GL}(m, \mathbb{F}) = \mathbb{F}^{m \times m}.$$

The discussion of why $\mathfrak{g} = T_{\mathbb{1}} G$ is called a "Lie algebra" will have to wait for next semester; for our immediate purposes, it will be enough to notice that $\mathfrak{g}$ is a linear subspace of $\mathbb{F}^{m \times m}$.

It will sometimes be useful to observe that the natural maps defined by matrix multiplication

$$\mathrm{GL}(m, \mathbb{F}) \times \mathrm{GL}(m, \mathbb{F}) \to \mathrm{GL}(m, \mathbb{F}) : (\mathbf{A}, \mathbf{B}) \mapsto \mathbf{A}\mathbf{B}$$

and inversion

$$\mathrm{GL}(m, \mathbb{F}) \to \mathrm{GL}(m, \mathbb{F}) : \mathbf{A} \mapsto \mathbf{A}^{-1}$$

are both smooth. Indeed, the first is simply a quadratic function of the entries of $\mathbf{A}$ and $\mathbf{B}$, and by Cramer's rule, the second is $1/\det(\mathbf{A})$ times a polynomial function of the entries, where $\det(\mathbf{A})$ is itself a polynomial function of the entries and is nonzero as long as we restrict to the open subset $\mathrm{GL}(m, \mathbb{F}) \subset \mathbb{F}^{m \times m}$. Since restrictions of smooth maps to smooth submanifolds are also smooth, it follows that for every Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$, the maps

$$G \times G \to G : (\mathbf{A}, \mathbf{B}) \mapsto \mathbf{A}\mathbf{B}, \qquad G \to G : \mathbf{A} \mapsto \mathbf{A}^{-1}$$

are both smooth.

EXAMPLE 18.2. The **orthogonal** group $\mathrm{O}(m) \subset \mathrm{GL}(m, \mathbb{R})$ consists of all linear transformations $\mathbb{R}^n \to \mathbb{R}^n$ that preserve the standard Euclidean inner product $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbb{R}^m} := \mathbf{v}^T \mathbf{w} = \sum_{j=1}^m v^j w^j$. It is a Lie subgroup according to Exercise 4.22, and its Lie algebra is the space of real antisymmetric matrices

$$\mathfrak{o}(m) := \left\{ \mathbf{A} \in \mathbb{R}^{m \times m} \mid \mathbf{A}^T = -\mathbf{A} \right\}.$$

EXAMPLE 18.3. The **unitary** group $\mathrm{U}(m) \subset \mathrm{GL}(m, \mathbb{C})$ consists of all linear transformations $\mathbb{C}^n \to \mathbb{C}^n$ that preserve the standard Hermitian inner product $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbb{C}^m} := \mathbf{v}^\dagger \mathbf{w} = \sum_{j=1}^m \bar{v}^j w^j$. It is a Lie subgroup according to Exercise 4.23, and its Lie algebra is the space of complex anti-Hermitian matrices

$$\mathfrak{u}(m) := \left\{ \mathbf{A} \in \mathbb{C}^{m \times m} \mid \mathbf{A}^\dagger = -\mathbf{A} \right\}.$$

EXAMPLE 18.4. The group of orientation-preserving linear transformations on $\mathbb{R}^n$ is

$$\mathrm{GL}_+(m, \mathbb{R}) := \left\{ \mathbf{A} \in \mathrm{GL}(m, \mathbb{R}) \mid \det(\mathbf{A}) > 0 \right\},$$

which is both a subgroup and an open subset of $\mathrm{GL}(m, \mathbb{R})$, and therefore a Lie subgroup. Since it is also an open subset of $\mathbb{R}^{m \times m}$, its Lie algebra is

$$\mathfrak{gl}_+(m, \mathbb{R}) = \mathfrak{gl}(m, \mathbb{R}) := \mathbb{R}^{m \times m}.$$

EXAMPLE 18.5. Exercise 4.24 shows that for $\mathbb{F}$ equal to either $\mathbb{R}$ or $\mathbb{C}$, the **special linear** group $\mathrm{SL}(m, \mathbb{F}) := \left\{ \mathbf{A} \in \mathrm{GL}(m, \mathbb{F}) \mid \det(\mathbf{A}) = 1 \right\}$ is a Lie subgroup whose Lie algebra consists of the traceless matrices,

$$\mathfrak{sl}(m, \mathbb{F}) := \left\{ \mathbf{A} \in \mathbb{F}^{m \times m} \mid \mathrm{tr}(\mathbf{A}) = 0 \right\}.$$

The special linear group consists of all linear transformations $\mathbb{F}^m \to \mathbb{F}^m$ that preserve the "standard" alternating m-form

(18.1)
$$\mu(\mathbf{v}_1, \ldots, \mathbf{v}_m) := \det \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{pmatrix}.$$

In the real case, one obtains a useful geometric interpretation by relating $\mathrm{SL}(m, \mathbb{R})$ to the larger group

$$\widehat{\mathrm{SL}}(m, \mathbb{R}) := \left\{ \mathbf{A} \in \mathrm{GL}(m, \mathbb{R}) \mid \det(\mathbf{A}) \in \{1, -1\} \right\},$$

which is the group of all volume-preserving linear transformations on $\mathbb{R}^n$. Since

$$\mathrm{SL}(m, \mathbb{R}) = \widehat{\mathrm{SL}}(m, \mathbb{R}) \cap \mathrm{GL}_+(m, \mathbb{R}),$$

$\mathrm{SL}(m, \mathbb{R})$ therefore consists of all linear transformations that preserve both orientation and volume.

EXAMPLE 18.6. The **special orthogonal** group $\mathrm{SO}(m) := \mathrm{O}(m) \cap \mathrm{SL}(m, \mathbb{R})$ is an open subset of $\mathrm{O}(m)$, and thus has the same Lie algebra,

$$\mathfrak{so}(m) = \mathfrak{o}(m),$$

which is contained in $\mathfrak{sl}(m, \mathbb{R})$ since real antisymmetric matrices vanish along the diagonal. Since every $\mathbf{A} \in \mathrm{O}(m)$ has $\det(\mathbf{A}) = \pm 1$, one could equally well write

$$\mathrm{SO}(m) = \mathrm{O}(m) \cap \mathrm{GL}_+(m, \mathbb{R}),$$

and thus interpret $\mathrm{SO}(m)$ as the group of all orientation-preserving orthogonal transformations.

EXAMPLE 18.7. The complex analogue of $\mathrm{SO}(m)$ is the **special unitary** group $\mathrm{SU}(m) := \mathrm{U}(m) \cap \mathrm{SL}(m, \mathbb{C})$, but there is a qualitative difference from the real case: according to Exercise 4.25, $\mathrm{SU}(m)$ is also a Lie subgroup, but its dimension is one less than that of $\mathrm{U}(m)$, and its Lie algebra

$$\mathfrak{su}(m) := \mathfrak{u}(m) \cap \mathfrak{sl}(m, \mathbb{C})$$

is the space of matricies that are both anti-Hermitian *and* traceless, which is not identical to $\mathfrak{u}(m)$ since anti-Hermitian matrices can have arbitrary imaginary entries on the diagonal. One can interpret $\mathrm{SU}(m)$ as the group of linear transformations on $\mathbb{C}^m$ that preserve both the standard Hermitian inner product and the alternating m-form μ in (18.1).

EXAMPLE 18.8. The following generalization of the orthogonal group is important in physics: given integers $k, \ell \geqslant 0$ with $k + \ell = m$, the **indefinite orthogonal group**

$$\mathrm{O}(k, \ell) \subset \mathrm{GL}(m, \mathbb{R})$$

consists of all linear transformations $\mathbf{A} : \mathbb{R}^m \to \mathbb{R}^m$ that satisfy $\langle \mathbf{Av}, \mathbf{Aw} \rangle_{k,\ell} = \langle \mathbf{v}, \mathbf{w} \rangle_{k,\ell}$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, where

$$\langle \mathbf{v}, \mathbf{w} \rangle_{k,\ell} := \sum_{j=1}^{k} v^j w^j - \sum_{j=k+1}^{m} v^j w^j.$$

The case $\mathrm{O}(1, 3)$ is known as the **Lorentz group** and plays a fundamental role in relativity, where the sign difference in $\langle \ , \ \rangle_{1,3}$ between the first and the other three coordinates gives a qualitative distinction between the three dimensions of physical space and a fourth dimension, interpreted as time. There is also a complex analogue, the **indefinite unitary group** $\mathrm{U}(k, \ell) \subset \mathrm{GL}(m, \mathbb{C})$.

EXERCISE 18.9. For integers $k, \ell \geqslant 0$ with $k + \ell = m$, define the block matrix $\boldsymbol{\eta} := \begin{pmatrix} \mathbb{1}_k & 0 \\ 0 & -\mathbb{1}_\ell \end{pmatrix} \in$ $\mathrm{GL}(m, \mathbb{R})$, where for any $q \geqslant 0$ we write $\mathbb{1}_q$ for the q-by-q identity matrix. Prove:
  (a) A matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ belongs to $\mathrm{O}(k, \ell)$ if and only if $\mathbf{A} \boldsymbol{\eta} \mathbf{A}^T \boldsymbol{\eta} = \mathbb{1}$.
  (b) Every $\mathbf{A} \in \mathrm{O}(k, \ell)$ satisfies $\det(\mathbf{A}) = \pm 1$.

(c) $\mathrm{O}(k,\ell)$ is a smooth submanifold and thus a Lie subgroup of $\mathrm{GL}(m,\mathbb{R})$, with Lie algebra
$$\mathfrak{o}(k,\ell) := \left\{ \mathbf{A} \in \mathbb{R}^{m \times m} \;\big|\; \mathbf{A}^* = -\mathbf{A} \right\},$$
where $\mathbf{A}^* := \boldsymbol{\eta}\mathbf{A}^T\boldsymbol{\eta}$.
Hint: For every $\mathbf{A} \in \mathrm{GL}(m,\mathbb{R})$, $\mathbf{A}\boldsymbol{\eta}\mathbf{A}^T\boldsymbol{\eta}$ *belongs to the vector space* $\left\{ \mathbf{H} \in \mathbb{R}^{n \times n} \;\big|\; \mathbf{H}^* = \mathbf{H} \right\}$.

EXAMPLE 18.10. [55] There is a natural way of regarding $\mathrm{GL}(m,\mathbb{C})$ as a Lie subgroup of $\mathrm{GL}(2m,\mathbb{R})$. The idea is to identify $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ via the real-linear isomorphism
$$\mathbb{C}^m \to \mathbb{R}^{2m} = \mathbb{R}^m \times \mathbb{R}^m : \mathbf{x} + i\mathbf{y} \mapsto (\mathbf{x},\mathbf{y}),$$
so that scalar multiplication by i becomes the linear transformation $\mathbb{R}^{2m} \to \mathbb{R}^{2m}$ defined by the matrix
$$(18.2) \qquad\qquad \mathbf{J}_0 := \begin{pmatrix} 0 & -\mathbb{1}_m \\ \mathbb{1}_m & 0 \end{pmatrix}.$$
A linear transformation $\mathbf{A} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ then represents a complex-linear transformation on $\mathbb{C}^n$ if and only if it commutes with the matrix $\mathbf{J}_0$, giving an identification of $\mathbb{C}^{m \times m}$ with the linear subspace
$$(18.3) \qquad\qquad \mathrm{End}_{\mathbb{C}}(\mathbb{R}^{2m}) := \left\{ \mathbf{A} \in \mathbb{R}^{2m \times 2m} \;\big|\; \mathbf{A}\mathbf{J}_0 = \mathbf{J}_0\mathbf{A} \right\} \subset \mathbb{R}^{2m \times 2m}.$$
In this way, the group $\mathrm{GL}(m,\mathbb{C})$ gets identified with the open subset of $\mathrm{End}_{\mathbb{C}}(\mathbb{R}^{2m})$ consisting of invertible transformations, making it a smooth submanifold and thus a Lie subgroup of $\mathrm{GL}(2m,\mathbb{R})$, with Lie algebra $\mathfrak{gl}(m,\mathbb{C}) = \mathrm{End}_{\mathbb{C}}(\mathbb{R}^{2m})$.

EXERCISE 18.11. Show that under the identification of $\mathrm{GL}(m,\mathbb{C})$ with a subgroup of $\mathrm{GL}(2m,\mathbb{R})$ explained in Example 18.10, $\mathrm{O}(2m) \cap \mathrm{GL}(m,\mathbb{C}) = \mathrm{U}(m) \subset \mathrm{GL}(2m,\mathbb{R})$.
Hint: Using the identification $\mathbb{C}^m = \mathbb{R}^{2m}$, *write down a formula for the Hermitian inner product of* $\mathbb{C}^m$ *in terms of the Euclidean inner product of* $\mathbb{R}^{2m}$ *and the matrix* $\mathbf{J}_0$ *in* (18.2).

18.2. The structure group of a vector bundle. Assume in the following that $G \subset \mathrm{GL}(m,\mathbb{F})$ is a Lie subgroup.

DEFINITION 18.12. A G-**structure** on a smooth vector bundle $E \to M$ of rank m is a maximal collection of smoothly compatible local trivializations $\{(\mathcal{U}_\alpha, \Phi_\alpha)\}_{\alpha \in I}$ of E with the property that $M = \bigcup_{\alpha \in I} \mathcal{U}_\alpha$ and the associated transition functions $g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m,\mathbb{F})$ take their values in G. When a G-structure on $E \to M$ has been given, we call G the **structure group** of the bundle, and the local trivializations that belong to the G-structure will be called G-**compatible trivializations**.

For a given group G and bundle $E \to M$, a G-structure may or may not exist, and it will typically not be unique. *Every* vector bundle of rank m has structure group $\mathrm{GL}(m,\mathbb{F})$ by default, but for a given subgroup $G \subset \mathrm{GL}(m,\mathbb{F})$, it may or may not be possible to *reduce* the structure group to G by deleting a subcollection of its smooth local trivializations, so that those which remain are related to each other by G-valued transition functions. Note also that if G is a subgroup of some larger Lie subgroup $H \subset \mathrm{GL}(m,\mathbb{R})$, then a G-structure on $E \to M$ determines an H-structure, obtained by including all local trivializations that are related by H-valued transition functions to the G-compatible trivializations. A G-structure should be thought of as a *preferred* class of local trivializations that cover M, or equivalently, a preferred class of local frames, which we will also refer to in the following as G-**compatible frames**. Our first definition of *orientations* in §10.2 was somewhat analogous to this: choosing an orientation on a manifold M means selecting a preferred

---

[55]This example was not mentioned in the lecture but is provided here for your information. The same applies to one or two other things in Lecture 18 regarding the relationship between real and complex bundles.

class of charts to be called "oriented" charts, and deleting those which are not compatible with them via orientation-preserving transition maps. A G-structure on a bundle E is also sometimes called a **reduction of the structure group** of E to G.

There is almost always a useful alternative way to interpret G-structures without mentioning transition functions, but the alternative interpretation varies depending on the specific group G. We will look next at several examples.

18.3. Global trivializations: $G = \{\mathbb{1}\}$. The trivial group $G := \{\mathbb{1}\} \subset \mathrm{GL}(m, \mathbb{F})$ is a 0-dimensional Lie subgroup of $\mathrm{GL}(m, \mathbb{F})$, and a G-structure on a bundle $E \to M$ then consists of a covering of M by a collection of local trivializations $\{(\mathcal{U}_\alpha, \Phi_\alpha)\}_{\alpha \in I}$ that are all *identical* wherever they overlap. If such a collection exists, then all of them are restrictions to the subsets $\mathcal{U}_\alpha \subset M$ of some global trivialization $\Phi : E \to M \times \mathbb{F}^m$, meaning a bundle isomorphism to the trivial m-plane bundle, so E is globally trivial. Conversely, any global trivialization $\Phi : E \to M \times \mathbb{F}^m$ determines a G-structure for $G := \{\mathbb{1}\}$ consisting of the restrictions of Φ to all possible open subsets $\mathcal{U}_\alpha \subset M$.

18.4. Orientations: $G = \mathrm{GL}_+(m, \mathbb{R})$. An **orientation** of a real vector bundle $E \to M$ is a choice of orientations for the fibers $\{E_p\}_{p \in M}$ that depend continuously on p, meaning that any collection of continuous sections $s_1, \ldots, s_m : \mathcal{U} \to E$ on a neighborhood $\mathcal{U} \subset M$ of p that form a positively-oriented basis of E_p also form positively oriented bases of E_q for all q near p. Note that an orientation of a general vector bundle $E \to M$ need not have anything to do with an orientation of the base M, which may or may not be orientable—according to Proposition 10.25, an orientation of M is equivalent to an orientation of the specific bundle $TM \to M$.

An orientation of $E \to M$ determines a preferred class of local frames for E, namely those which are positively oriented at every point. Equivalently, the preferred class of local trivializations consists of those which define orientation-preserving isomorphisms between $\mathbb{R}^m$ and the fibers E_p. The transition functions that relate two such trivializations to each other must therefore take values in the group of orientation-preserving transformations of $\mathbb{R}^m$, that is, $\mathrm{GL}_+(m, \mathbb{R})$. An orientation of $E \to M$ thus determines a $\mathrm{GL}_+(m, \mathbb{R})$-structure. Conversely, any $\mathrm{GL}_+(m, \mathbb{R})$-structure on $E \to M$ determines an orientation of the fibers via the condition that an ordered basis of a fiber E_p is positively oriented if and only if some $\mathrm{GL}_+(m, \mathbb{R})$-compatible local trivialization identifies it with a positively-oriented basis of $\mathbb{R}^m$. If this holds for one of the preferred trivializations defined at p, then it holds for all the others as well, because the transition functions that relate them act by orientation-preserving transformations on $\mathbb{R}^m$.

We've proved:

PROPOSITION 18.13. *On a real vector bundle $E \to M$, there is a natural bijective correspondence between orientations and $\mathrm{GL}_+(m, \mathbb{R})$-structures.* $\qquad\square$

Though you might already find it obvious that every trivial real vector bundle is orientable, we can now give a quick new proof of this fact in the language of structure groups: if $E \to M$ is trivial, then it admits a G-structure for $G := \{\mathbb{1}\}$, which is a subgroup of $\mathrm{GL}_+(m, \mathbb{R})$, so $E \to M$ therefore also admits a $\mathrm{GL}_+(m, \mathbb{R})$-structure, meaning an orientation. In fact, this argument shows that any global trivialization of a vector bundle determines a G-structure for every Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$.

EXERCISE 18.14. Prove that the line bundle $\ell \to S^1$ in Example 16.23 is not orientable.

18.5. Bundle metrics: $G = \mathrm{O}(m)$, $\mathrm{U}(m)$, $\mathrm{O}(k, \ell)$. The following definition generalizes the notion of a Riemannian metric in two respects: it is defined on an arbitrary vector bundle instead of a tangent bundle $TM \to M$, and it requires the weaker condition of *nondegeneracy* in place of positive-definiteness.

DEFINITION 18.15. A **bundle metric** on a real vector bundle $E \to M$ is a smooth function

$$\langle \ , \ \rangle : E \oplus E \to \mathbb{R}$$

whose restriction to the fiber $E_p \times E_p$ for each $p \in M$ is all of the following:

(i) (bilinear) $E_p \times E_p \to \mathbb{R} : (v, w) \mapsto \langle v, w \rangle$ is a bilinear map
(ii) (symmetric) $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in E_p$
(iii) (nondegenerate) The map $E_p \to E_p^* : v \mapsto \langle v, \cdot \rangle$ is injective.

We will say additionally that $\langle \ , \ \rangle$ is **positive** if the third condition is strengthened to:

(iii) (positive) $\langle v, v \rangle > 0$ for all nonzero $v \in E_p$.

For a complex vector bundle $E \to M$, we modify the above definition as follows: $\langle \ , \ \rangle$ is a smooth function

$$\langle \ , \ \rangle : E \oplus E \to \mathbb{C}$$

whose restriction to $E_p \times E_p$ is:

(i) (sesquilinear) $E_p \times E_p \to \mathbb{C} : (v, w) \mapsto \langle v, w \rangle$ is linear in the second factor and antilinear in the first
(ii) (Hermitian) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for all $v, w \in E_p$
(iii) (nondegenerate) Same as in the real case.

The positivity condition in the complex case is also the same as in the real case.

REMARK 18.16. A bundle metric is called **indefinite** if it is nondegenerate but not positive. In most of the literature, bundle metrics are assumed to be positive by default, and it is generally wise to assume this unless the word "indefinite" is included. A large portion of what we have to say about them will however be valid without assuming positivity, so in these notes, we will use "bundle metric" as a general term that includes the indefinite case. Bundle metrics are also often referred to as "Euclidean" bundle metrics in the real case and "Hermitian" bundle metrics in the complex case.

EXAMPLE 18.17. A Riemannian metric on a manifold M is a positive bundle metric on its tangent bundle $TM \to M$. If g is instead an indefinite bundle metric on $TM \to M$, it is called a **pseudo-Riemannian** (or also **semi-Riemannian**) metric on M, and the pair (M, g) is called a **pseudo-Riemannian manifold**.

In the real case, a bundle metric on $E \to M$ can also be regarded as a smooth section of the vector bundle $E^* \otimes E^* \to M$, whose fiber at $p \in M$ is the space of bilinear maps $E_p \times E_p \to \mathbb{R}$. Sesquilinearity modifies this statement in the complex case and replaces $E^* \otimes E^*$ with $\bar{E}^* \otimes E^*$, whose fiber at $p \in M$ is (according to Exercise 17.21) naturally isomorphic to the space of maps $E_p \times E_p \to \mathbb{C}$ that are antilinear in the first and linear in the second factor.

A positive bundle metric assigns to each fiber what is conventionally called an inner product, and as we observed in §15.2, the set of positive-definite inner products on any vector space is convex. Our previous existence result for Riemannian metrics therefore generalizes in a straightforward way:

THEOREM 18.18. *Every vector bundle admits a positive bundle metric.*

PROOF. Trivial bundles obviously admit positive bundle metrics since one can simply choose the standard Euclidean inner product of $\mathbb{R}^m$ or (for the case $\mathbb{F} = \mathbb{C}$) the standard Hermitian inner product of $\mathbb{C}^m$ on every fiber. It follows that on any vector bundle $E \to M$ with a collection of local trivializations $\{\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m\}_{\alpha \in I}$ covering M, one can choose bundle metrics on each $E|_{\mathcal{U}_\alpha}$, and then piece these together using a partition of unity on M subordinate to the cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$. $\square$

It is interesting to note that if we'd been allowed to assume in Theorem 18.18 that the bundle $E \to M$ has a G-structure for $G = \mathrm{O}(m)$ or (in the case $\mathbb{F} = \mathbb{C}$) $G = \mathrm{U}(m)$, then the proof would not have required a partition of unity. Indeed, if one defines $\langle \ , \ \rangle$ over regions $\mathcal{U}_\alpha \subset M$ so that it matches the standard inner product of $\mathbb{F}^m$ in some choice of G-*compatible* local trivialization over $\mathcal{U}_\alpha$, then this definition is independent of that choice: having transition functions valued in $G \in \{\mathrm{O}(m), \mathrm{U}(m)\}$ means that they preserve the standard inner product on $\mathbb{F}^m$, so any other G-compatible local trivialization on an overlapping region produces *the same* inner product on the fibers. This means that if an $\mathrm{O}(m)$- or $\mathrm{U}(m)$-structure is given, then it determines a unique positive bundle metric on $E \to M$ that looks like the standard inner product of $\mathbb{F}^m$ in any compatible local trivialization. There is also a converse to this: if a positive bundle metric $\langle \ , \ \rangle$ is given, then every smooth local frame on a region $\mathcal{U}_\alpha \subset M$ can be modified via the Gram-Schmidt algorithm to produce one that is orthonormal at every point $p \in \mathcal{U}_\alpha$, so that the corresponding local trivialization identifies $\langle \ , \ \rangle$ with the standard inner product of $\mathbb{F}^m$. Any two trivializations produced in this way will then be related by a transition function whose values preserve this inner product, meaning they are in $\mathrm{O}(m)$ or $\mathrm{U}(m)$. We've proved:

PROPOSITION 18.19. *On a vector bundle $E \to M$ of rank m, there is a natural bijective correspondence between positive bundle metrics and $\mathrm{O}(m)$-structures if E is real, or $\mathrm{U}(m)$-structures if E is complex.* $\qquad\square$

REMARK 18.20. In light of Proposition 18.19, an $\mathrm{O}(m)$-structure on a vector bundle is also sometimes called a **Euclidean structure**, and a $\mathrm{U}(m)$-structure is called a **Hermitian structure**.

Extending this discussion to the case of an indefinite bundle metric $\langle \ , \ \rangle$ on $E \to M$ requires a suitable generalization of the notion of orthonormal frames. In the following, we confine our attention to *real* vector bundles since that is the case that arises most often in applications, but there are no substantial differences in the complex case. We will say that a local frame $e_1, \ldots, e_m$ for E on some region $\mathcal{U} \subset M$ is **orthonormal** with respect to $\langle \ , \ \rangle$ if for some $k \in \{0, \ldots, m\}$, it satisfies

$$
\begin{aligned}
&\langle e_j, e_j \rangle = 1 \quad \text{for } j = 1, \ldots, k \\
(18.4) \qquad &\langle e_j, e_j \rangle = -1 \quad \text{for } j = k+1, \ldots, m \\
&\langle e_i, e_j \rangle = 0 \quad \text{for } i \neq j.
\end{aligned}
$$

The integers k and $\ell := m - k$ are determined by the bundle metric, and can be characterized as the dimensions of the largest subspace of any fiber on which $\langle \ , \ \rangle$ is positive-definite or negative-definite respectively. In general, these numbers need not be the same everywhere on M, though it should be clear that they are *locally* constant and thus constant on each connected component. As a rule, the only interesting examples are those in which k and ℓ are constant everywhere; in this case, the pair (k, ℓ) is called the **signature** of the bundle metric $\langle \ , \ \rangle$. The spacetime manifolds of general relativity are 4-manifolds with pseudo-Riemannian metrics of signature $(1, 3)$; these are known as **Lorentzian manifolds**.

LEMMA 18.21. *For any real vector bundle $E \to M$ with an indefinite bundle metric $\langle \ , \ \rangle$, every point $p \in M$ has a neighborhood $\mathcal{U} \subset M$ on which E admits an orthonormal frame.*

Since Lemma 18.21 is a local statement and all vector bundles are locally trivial, it suffices to prove it for the special case of a trivial m-plane bundle

$$
E := \mathcal{U} \times \mathbb{R}^m
$$

over some open subset $\mathcal{U} \subset M$ of a manifold. The restriction of $\langle \ , \ \rangle$ to the fiber over a point $p \in \mathcal{U}$ is in this case a bilinear form on $E_p = \mathbb{R}^m$ that can be written as

$$
\langle \mathbf{v}, \mathbf{w} \rangle_p = \langle \mathbf{v}, \mathbf{H}(p)\mathbf{w} \rangle_{\mathbb{R}^m} \qquad \text{for } \mathbf{v}, \mathbf{w} \in \mathbb{R}^m,
$$

where $\langle \ , \ \rangle_{\mathbb{R}^m}$ denotes the standard Euclidean inner product on $\mathbb{R}^m$ and $\mathbf{H}(p) \in \mathbb{R}^{m \times m}$ is a uniquely determined matrix that depends smoothly on $p \in \mathcal{U}$. Symmetry and nondegeneracy imply moreover that $\mathbf{H}(p)$ is always both symmetric and invertible respectively. It follows then from the spectral theorem that at every point $p \in \mathcal{U}$, $\mathbb{R}^m$ splits uniquely

$$\mathbb{R}^m = E_p^+ \oplus E_p^-$$

into the subspaces $E_p^+, E_p^- \subset \mathbb{R}^m$ spanned by the positive and negative eigenvalues respectively of $\mathbf{H}(p)$; on these two subspaces, $\langle \ , \ \rangle_p$ is positive- or negative-definite respectively. Notice that E_p^+ and E_p^- are orthogonal to each other with respect to both the Euclidean inner product and the given bundle metric $\langle \ , \ \rangle$. We will see below that these subspaces vary smoothly with p, but since that fact is not so obvious, let us first give a proof of Lemma 18.21 that does not require it.

FIRST PROOF OF LEMMA 18.21. For a given point $p \in \mathcal{U}$, let $k := \dim E_p^+$ and $\ell := \dim E_p^-$, choose orthonormal bases of E_p^+ and E_p^- and choose a smooth frame $\widehat{e}_1, \ldots, \widehat{e}_m$ for E on a neighborhood of p such that at the point p itself, $\widehat{e}_1, \ldots, \widehat{e}_k$ matches the chosen orthonormal basis of E_p^+ while $\widehat{e}_{k+1}, \ldots, \widehat{e}_m$ matches the chosen orthonormal basis of E_p^-. Since E_p^+ and E_p^- are orthogonal with respect to $\langle \ , \ \rangle$, it follows that $\widehat{e}_1, \ldots, \widehat{e}_m$ satisfy the orthonormality condition (18.4) at p, and we will now use a minor variation on the Gram-Schmidt algorithm to produce from this an orthonormal frame $e_1, \ldots, e_m$ that is defined on a neighborhood of p and matches $\widehat{e}_1, \ldots, \widehat{e}_m$ at p. The key observation making this possible is that since $\langle \ , \ \rangle$ is positive on E_p^+ and negative on E_p^-, it is also positive / negative on the subbundles spanned by $\widehat{e}_1, \ldots, \widehat{e}_k$ and $\widehat{e}_{k+1}, \ldots, \widehat{e}_m$ respectively. Now, define $e_1, \ldots, e_k$ simply by applying the usual Gram-Schmidt procedure to $\widehat{e}_1, \ldots, \widehat{e}_k$. Since $\langle \widehat{e}_{k+1}, \widehat{e}_{k+1} \rangle < 0$, the correct definition of e_{k+1} is slightly different: we set

$$e_{k+1} := f_1 \cdot \left(\widehat{e}_{k+1} - \sum_{j=1}^{k} \langle \widehat{e}_{k+1}, e_j \rangle e_j \right),$$

with a positive function f_1 chosen to ensure that $\langle \widehat{e}_{k+1}, \widehat{e}_{k+1} \rangle \equiv -1$ on a neighborhood of p, which is possible because the expression in parentheses matches $\widehat{e}_{k+1}$ at p, so that its product with itself is negative. Continuing in this way inductively, the new section e_{k+i} is defined out of $e_1, \ldots, e_{k+i-1}$ for each $i = 1, \ldots, \ell$ by

$$e_{k+i} := f_i \cdot \left(\widehat{e}_{k+i} - \sum_{j=1}^{k} \langle \widehat{e}_{k+i}, e_j \rangle e_j + \sum_{j=1}^{i-1} \langle \widehat{e}_{k+i}, e_{k+j} \rangle e_{k+j} \right),$$

with the positive function f_i again chosen to achieve the normalization $\langle e_{k+i}, e_{k+i} \rangle \equiv -1$. $\qquad \square$

I mentioned above that the subspaces $E_p^{\pm} \subset \mathbb{R}^m$ vary smoothly with p, which will give rise to a slightly simpler proof of Lemma 18.21. These subspaces are defined as direct sums of certain eigenspaces of the matrix $\mathbf{H}(p)$, but we have to be a bit careful here, because in general, individual eigenspaces cannot be assumed to depend smoothly on the matrix—one can show that they do whenever the corresponding eigenvalue is simple, but in our situation, eigenvalues with multiplicity may occur and there is no general way to avoid them. What we are interested in however is not an individual eigenspace, but direct sums of several eigenspaces corresponding to eigenvalues in fixed open subsets of $\mathbb{R}$, namely $(-\infty, 0)$ and $(0, \infty)$. In this situation, Cauchy's integration theory from complex analysis provides a useful trick:

LEMMA 18.22. *Suppose $\mathbf{A} \in \mathbb{C}^{m \times m}$ is a diagonalizable matrix,*

$$\sigma(\mathbf{A}) = \sigma_0 \sqcup \sigma_1 \subset \mathbb{C}$$

is a decomposition of its spectrum $\sigma(\mathbf{A})$ into two disjoint subsets, and write

$$\mathbb{C}^m = V_0 \oplus V_1, \qquad where \qquad V_j := \bigoplus_{\lambda \in \sigma_j} \ker(\lambda\mathbb{1} - \mathbf{A}), \quad j = 0, 1$$

for the corresponding splitting of $\mathbb{C}^m$ into direct sums of eigenspaces. Then for any smoothly embedded oriented circle $\gamma \subset \mathbb{C}$ that does not intersect $\sigma(\mathbf{A})$ and has winding number j around each eigenvalue in σ_j for $j = 0, 1$, the matrix-valued Cauchy integral

$$\mathbf{P} := \frac{1}{2\pi i} \int_\gamma (z\mathbb{1} - \mathbf{A})^{-1} \, dz \in \mathbb{C}^{m \times m}$$

defines the linear projection to V_1 along V_0.

PROOF. The function $\mathbb{C}\backslash\sigma(\mathbf{A}) \to \mathbb{C}^{m \times m} : z \mapsto (z\mathbb{1} - \mathbf{A})^{-1}$ is holomorphic since $z\mathbb{1} - \mathbf{A}$ is an affine function of z and, for arbitrary invertible matrices $\mathbf{B} \in \mathrm{GL}(m, \mathbb{C})$, the entries of $\mathbf{B}^{-1}$ are rational functions of the entries in $\mathbf{B}$. Cauchy's theorem thus implies that the integral will not change if γ is replaced by a disjoint union of small circles around the specific eigenvalues in σ_1, and it suffices therefore to consider the case where σ_1 consists of only one eigenvalue $\lambda_1 \in \mathbb{C}$ and γ is parametrized by the boundary of the ϵ-disk around λ_1 for $\epsilon > 0$ small. Since $\mathbf{A}$ is diagonalizable we can also assume after a change of basis on $\mathbb{C}^m$ that $\mathbf{A}$ is diagonal; let us write its diagonal entries as $\Lambda_1, \ldots, \Lambda_m \in \mathbb{C}$, keeping in mind that these are all elements of $\sigma(\mathbf{A})$ and some of them may be repeated. The values of the function $(z\mathbb{1} - \mathbf{A})^{-1}$ are then also diagonal matricies, whose diagonal entries are the complex-valued functions $\frac{1}{z - \Lambda_j}$ for $j = 1 \ldots, m$. For any j such that $\Lambda_j \neq \lambda_1$, we can assume $\frac{1}{z - \Lambda_j}$ is a holomorphic function on the disk enclosed by γ, so its integral is 0. On the other hand, whenever $\Lambda_j = \lambda_1$, integration makes the corresponding diagonal element into

$$\frac{1}{2\pi i} \int_\gamma \frac{dz}{z - \lambda_1} = 1.$$

We conclude that in our chosen basis of eigenvectors for $\mathbf{A}$, $\mathbf{P}$ is a diagonal matrix whose entries are all 1 or 0, with 1 appearing only in the places where the corresponding entry of $\mathbf{A}$ is λ_1. In other words, $\mathbf{P}$ acts as the identity on the eigenspace of λ_1 and as 0 on all the other eigenspaces. $\square$

COROLLARY 18.23. *The subspaces $E_p^+, E_p^- \subset \mathbb{R}^m$ defined as the direct sums of the positive and negative eigenspaces respectively of $\mathbf{H}(p)$ vary smoothly with the point $p \in \mathcal{U}$.*

PROOF. Given $p \in \mathcal{U}$, choose an embedded oriented circle $\gamma \subset \mathbb{C}$ that surrounds the positive eigenvalues of $\mathbf{H}(p)$ but stays in the right half-plane, so its winding around every negative eigenvalue is 0. Then according to Lemma 18.22, the matrix $\frac{1}{2\pi i} \int_\gamma (z\mathbb{1} - \mathbf{H}(p))^{-1} \, dz$ defines the orthogonal projection to E_p^+ along E_p^-, and this remains true if p is moved within a small enough region so that the eigenvalues of $\mathbf{H}(p)$ never touch γ. This matrix-valued integral clearly depends smoothly on p, and therefore so does the complementary projection to E_p^-. $\square$

SECOND PROOF OF LEMMA 18.21. Corollary 18.23 implies that the subspaces $E_p^+, E_p^- \subset \mathbb{R}^m$ form the fibers of smooth subbundles $E^\pm \subset E$, giving a splitting

$$(18.5) \qquad\qquad \mathcal{U} \times \mathbb{R}^m = E = E^+ \oplus E^-$$

such that $\pm\langle \ , \ \rangle$ restricts to a positive bundle metric on $E^\pm$, and moreover, the fibers of E^+ and E^- are mutually orthogonal with respect to $\langle \ , \ \rangle$. An orthonormal frame for E is then constructed by combining orthonormal frames of E^+ and E^-, and this can be done on a sufficiently small neighborhood of any given point. $\square$

For a bundle metric of signature (k, ℓ), the local trivialization corresponding to an orthonormal frame identifies $\langle\ ,\ \rangle$ on each fiber with the "standard" indefinite inner product

$$\langle \mathbf{v}, \mathbf{w} \rangle_{k,\ell} = \sum_{j=1}^{k} v^j w^j - \sum_{j=k+1}^{m} v^j w^j,$$

thus any two local trivializations constructed in this way are related by a transition function with values in the group $\mathrm{O}(k, \ell)$ from Example 18.8. We summarize:

PROPOSITION 18.24. *On a real vector bundle $E \to M$ of rank m with integers $k, \ell \geqslant 0$ satisfying $k + \ell = m$, there is a natural bijective correspondence between bundle metrics of signature (k, ℓ) and $\mathrm{O}(k, \ell)$-structures.* $\square$

EXERCISE 18.25. Show that for any real vector bundle $E \to M$ with a bundle metric $\langle\ ,\ \rangle$ of signature (k, ℓ) there exist smooth subbundles $E^+ \subset E$ and $E^- \subset E$ of ranks k and ℓ respectively and a bundle of isomorphism $E \cong E^+ \oplus E^-$.
Caution: This does not follow immediately from the splitting in (18.5), because that splitting was defined specifically for a trivial bundle; it can always be done locally since all bundles are locally trivial, but the result will depend on the choice of local trivialization. Obtaining such a splitting globally will require another choice, but it is a choice that can always be made.

REMARK 18.26. When $k, \ell \geqslant 1$, the existence of the splitting $E = E^+ \oplus E^-$ in Exercise 18.25 is a nontrivial condition that is not satisfied for all bundles, thus unlike the positive case, bundle metrics of arbitrary signature do not always exist. We will later see for instance that S^2 does not admit any pseudo-Riemannian metric of signature $(1, 1)$.

18.6. Volume forms: $G = \mathrm{SL}(m, \mathbb{F})$. A **volume form** on a vector bundle $E \to M$ of rank m is a section $\mu \in \Gamma(\Lambda^m E^*)$ that satisfies $\mu(p) \neq 0$ for all $p \in M$. In other words, for every $p \in M$, $\mu(p)$ is an alternating m-fold multilinear form $E_p \times \ldots \times E_p \to \mathbb{F}$ that evaluates to something nonzero on some (and therefore any) basis $v_1, \ldots, v_m \in E_p$. The terminology has a geometric motivation in the case $\mathbb{F} = \mathbb{R}$, as one can then use μ to define the notion of volume in every fiber by saying that $|\mu(p)(v_1, \ldots, v_m)|$ is the volume of the parallelepiped in E_p spanned by $v_1, \ldots, v_m$. No such geometric interpretation is available in the complex case, but the definition makes sense algebraically.

Given a volume form $\mu \in \Gamma(\Lambda^m E^*)$ and a local frame $e_1, \ldots, e_m$ for E over an open set $\mathcal{U}_\alpha \subset M$, one can always modify e_1 by multiplication with a scalar-valued function to arrange

$$(18.6) \qquad\qquad \mu(e_1, \ldots, e_m) \equiv 1 \quad \text{on } \mathcal{U}_\alpha.$$

The corresponding local trivialization then identifies μ over $\mathcal{U}_\alpha$ with the "standard" volume form on $\mathbb{F}^m$, given by

$$\mu_{\mathrm{std}}(\mathbf{v}_1, \ldots, \mathbf{v}_m) := \det \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{pmatrix} \in \mathbb{F}$$

for $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{F}^m$. The group of linear transformations $\mathbb{F}^m \to \mathbb{F}^m$ that preserve μ_{std} is the special linear group, $\mathrm{SL}(m, \mathbb{F})$, thus covering M with local frames that satisfy (18.6) determines an $\mathrm{SL}(m, \mathbb{F})$-structure on E. Conversely, if an $\mathrm{SL}(m, \mathbb{F})$-structure is given, then there is a unique volume form $\mu \in \Gamma(\Lambda^m E^*)$ that looks like μ_{std} in every $\mathrm{SL}(m, \mathbb{F})$-compatible local trivialization, proving:

PROPOSITION 18.27. *On any vector bundle $E \to M$ of rank m over the field $\mathbb{F}$, there is a natural bijective correspondence between volume forms and $\mathrm{SL}(m, \mathbb{F})$-structures.* $\square$

Several fundamental facts about volume forms on a manifold can now be generalized and proved as easy corollaries of basic observations about specific subgroups of $\mathrm{GL}(m, \mathbb{R})$:

PROPOSITION 18.28. *Every real vector bundle admitting a volume form is orientable.*

PROOF. If $E \to M$ has an $\mathrm{SL}(m, \mathbb{R})$-structure, then this determines a $\mathrm{GL}_+(m, \mathbb{R})$-structure since $\mathrm{SL}(m, \mathbb{R}) \subset \mathrm{GL}_+(m, \mathbb{R})$. $\qquad\square$

PROPOSITION 18.29. *On any oriented real vector bundle $E \to M$, any bundle metric determines a unique volume form μ such that $\mu(v_1, \ldots, v_m) = 1$ for every positively-oriented orthonormal basis of a fiber.*

PROOF. As an oriented bundle, E has structure group $\mathrm{GL}_+(m, \mathbb{R})$, and introducing a bundle metric of signature (k, ℓ) reduces its structure group further to

$$\mathrm{SO}(k, \ell) := \mathrm{O}(k, \ell) \cap \mathrm{SL}(m, \mathbb{R}) = \mathrm{O}(k, \ell) \cap \mathrm{GL}_+(m, \mathbb{R}),$$

where the equality of these two intersections results from the fact that every $\mathbf{A} \in \mathrm{O}(k, \ell)$ has determinant ± 1. Since $\mathrm{SO}(k, \ell) \subset \mathrm{SL}(m, \mathbb{R})$, we therefore also have an $\mathrm{SL}(m, \mathbb{R})$-structure and thus a volume form, which evaluates to 1 on the standard basis whenever it is viewed in an $\mathrm{SO}(k, \ell)$-compatible trivialization; in particular, this means bases that are positively oriented and orthonormal. $\qquad\square$

EXERCISE 18.30. Suppose $E \to M$ is an oriented real vector bundle of rank m with a bundle metric $\langle\,,\,\rangle$.
  (a) Reprove Proposition 18.29 by an argument analogous to Corollary 11.10 on the Riemannian volume form $d\mathrm{vol} \in \Gamma(\Lambda^n T^*M)$ for an oriented Riemannian manifold (M, g), i.e. show that the volume form $\mu \in \Gamma(\Lambda^m E^*)$ determined by the orientation and bundle metric on E can be written locally in the form $e_*^1 \wedge \ldots \wedge e_*^m$ using the dual frame to any positively-oriented orthonormal local frame $e_1, \ldots, e_m$.
  (b) Generalize the local coordinate formula for the Riemannian volume form in Exercise 11.12 as follows. Assume $e_1, \ldots, e_m$ is a positively-oriented but not necessarily orthonormal local frame over some open set $\mathcal{U} \subset M$, and write $g_{ij} := \langle e_i, e_j \rangle : \mathcal{U} \to \mathbb{R}$ for the resulting component functions of the bundle metric. Show

$$\mu = \sqrt{\pm \det \mathbf{g}}\, e_*^1 \wedge \ldots \wedge e_*^m \qquad \text{on } \mathcal{U},$$

where $e_*^1, \ldots, e_*^m$ is the dual frame to $e_1, \ldots, e_m$, $\mathbf{g} : \mathcal{U} \to \mathbb{R}^{m \times m}$ is the matrix-valued function whose entries are g_{ij}, and the sign $\pm$ is chosen to make the expression under the square root positive (this will depend on the signature of the bundle metric).

18.7. Complex structures: $G = \mathrm{GL}(m, \mathbb{C}) \subset \mathrm{GL}(2m, \mathbb{R})$. [56]
Along the lines of Example 18.10, identifying $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ makes any complex vector bundle $E \to M$ of rank m into a real vector bundle of rank $2m$ that is endowed with a G-structure for $G \cong \mathrm{GL}(m, \mathbb{C})$ defined as the subgroup of $\mathrm{GL}(2m, \mathbb{R})$ consisting of all invertible linear transformations $\mathbb{R}^{2m} \to \mathbb{R}^{2m}$ that commute with the matrix

$$\mathbf{J}_0 := \begin{pmatrix} 0 & -\mathbb{1}_m \\ \mathbb{1}_m & 0 \end{pmatrix}.$$

Recall from §7.1.4 that on any even-dimensional vector space V, a linear map $J : V \to V$ satisfying $J^2 = -\mathbb{1}$ is called a **complex structure**, thus $\mathbf{J}_0$ is an example of a complex structure on $\mathbb{R}^{2m}$. Any complex structure $J : V \to V$ makes V into a complex vector space by defining complex scalar multiplication to mean

$$(a + ib)v := av + bJv, \qquad a, b \in \mathbb{R},\ v \in V.$$

---

[56]Like Example 18.10, this section was not covered in the lecture but is provided here for your information.

If $v_1, \ldots, v_m \in V$ is any complex basis of this vector space, then $v_1, \ldots, v_m, Jv_1, \ldots, Jv_m$ is a real basis in which the matrix representing the transformation J is $\mathbf{J}_0$; this proves in particular that *every* complex structure on $\mathbb{R}^{2m}$ is equivalent to $\mathbf{J}_0$ via a change of basis.

More generally, a **complex structure** on a real vector bundle $E \to M$ of rank $2m$ is a smooth section J of the bundle

$$\operatorname{End}(E) := \operatorname{Hom}(E, E)$$

such that $J(p) : E_p \to E_p$ is a complex structure on E_p for every $p \in M$. Choosing a complex structure on E makes every fiber into a complex vector space of dimension m, and on a sufficiently small neighborhood $\mathcal{U} \subset M$ of any point p, one can choose a complex basis $v_1, \ldots, v_m$ of E_p and find a tuple of smooth sections $e_1, \ldots, e_m : \mathcal{U} \to E$ such that $e_j(p) = v_j$ for every $j = 1, \ldots, m$; after shrinking the neighborhood $\mathcal{U}$, we can then assume without loss of generality that the vectors $e_1, \ldots, e_m$ remain complex-linearly independent and thus form a basis of every fiber over points in $\mathcal{U}$. It follows that $e_1, \ldots, e_m, Je_1, \ldots, Je_m$ then forms a smooth frame for E over $\mathcal{U}$, and it defines a local trivialization $E|_{\mathcal{U}} \to \mathcal{U} \times \mathbb{R}^{2m}$ that identifies J on each fiber over points in $\mathcal{U}$ with the standard complex structure $\mathbf{J}_0 : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$. The transition functions relating any two local trivializations constructed in this way must then take values in the subgroup $\operatorname{GL}(m, \mathbb{C}) \subset \operatorname{GL}(2m, \mathbb{R})$, so we have constructed a $\operatorname{GL}(m, \mathbb{C})$-structure on E, and if we replace $\mathbb{R}^{2m}$ by $\mathbb{C}^m$, $E \to M$ can now be understood as a complex vector bundle of rank m. Conversely, any $\operatorname{GL}(m, \mathbb{C})$-structure on a real bundle $E \to M$ of rank $2m$ determines a complex structure $J \in \Gamma(\operatorname{End}(E))$ that is identified with $\mathbf{J}_0 : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$ by any $\operatorname{GL}(m, \mathbb{C})$-compatible local trivialization. This proves:

PROPOSITION 18.31. *There is a natural bijective correspondence between complex structures $J \in \Gamma(\operatorname{End}(E))$ on a real vector bundle $E \to M$ of rank $2m$ and $\operatorname{GL}(m, \mathbb{C})$-structures on E, where $\operatorname{GL}(m, \mathbb{C})$ is identified with a subgroup of $\operatorname{GL}(2m, \mathbb{R})$ as in Example 18.10. Moreover, any smooth real vector bundle E of rank $2m$ with complex structure J can be regarded naturally as a smooth complex vector bundle of rank m whose fibers over points $p \in M$ are the vector spaces E_p with complex scalar multiplication defined by $(a + ib)v := av + bJ(p)v$.* □

19. Connections on vector bundles

By way of motivation for what we will do in the next few lectures, I'd like to take a second look at a thought-experiment that was mentioned in Lecture 1. Figure 8 shows a closed path on S^2 that is made up of three smooth paths intersecting at right angles: one moving along the equator, and two that connect the equator to the north pole via longitudes. In this scenario, we pick a starting point p_0 for this path and a tangent vector $v_0 \in T_{p_0}S^2$, and then ask: if we move v_0 in a "parallel" manner along the path, keeping it tangent to the sphere as we go, will it come back to the same starting vector when the path returns to p_0? The question is imprecisely stated, because I have not said what "parallel" in this situation should mean, and that is a detail we will need to discuss. Nonetheless, the scenario in Figure 8 looks as if v_0 is being moved along the path in the most natural way possible, and the answer is clearly no: the vector it comes back to at the end of the closed loop is different.

This is not something that happens in *Euclidean* geometry. If our manifold were $\mathbb{R}^2$ instead of S^2, then there would be an obvious way to define what moving a tangent vector in a "parallel" manner along a path should mean: it means that the vector is constant, and it will therefore always return to itself when the path comes back to its starting point. On S^2, on the other hand, there is no obvious way to define what it should mean for a vector field to be *constant*, due to the fact that the tangent spaces T_pS^2 themselves are not constant as the point p moves. We will see nonetheless that if we endow the tangent spaces T_pS^2 with the Euclidean inner product and thus regard S^2 as a Riemannian manifold, then there is a natural way to define what it means for a vector field along a path to be *parallel*—that is what we will call the natural generalization of the word "constant" in
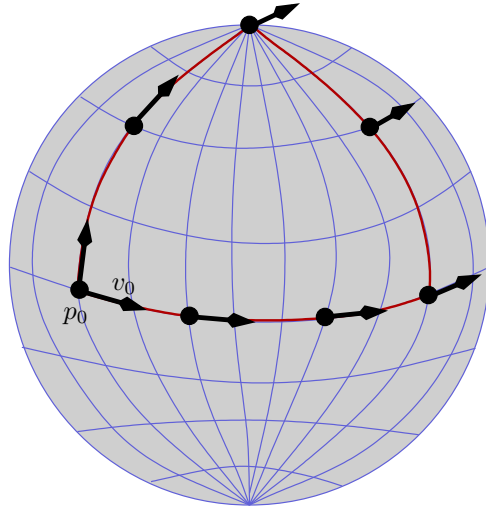
PSfrag replacements

FIGURE 8.   Parallel transport of a tangent vector around a closed path in S^2.

this context—but this notion will have some counterintuitive properties, e.g. that no vector field can ever be parallel on an entire open subset, no matter how small. Such properties are symptoms of the fact that S^2 has nontrivial *curvature*, while $\mathbb{R}^2$ with its Euclidean inner product does not.

In order to clarify what this means, we will first consider a general vector bundle $E \to M$ and ask what it might mean to say that a section $s \in \Gamma(E)$ is "constant" along a path. Such a notion can be defined, but the definition is not canonical: it depends on an extra piece of geometric data that must be chosen, and that data is called a *connection*. Several distinct definitions of the term "connection" can be found in various textbooks, and all of them are equivalent but look cosmetically quite different. Our first task is thus to understand why these particular definitions are the ones we need, and why they are equivalent.

19.1. Parallel transport and horizontal lifts. We assume for the rest of this lecture that

$$\pi : E \to M$$

is a smooth real or complex ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) vector bundle of rank m over an n-manifold M. For a section $s \in \Gamma(E)$, we have defined what it means for s to be *differentiable*, but we have not yet talked about actually differentiating it. If one wants to define, say, the derivative of s at a point $p \in M$ in the direction $X \in T_pM$, one quickly encounters a problem that we have seen before when talking about vector and tensor fields: choosing a path $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$, one cannot simply define[57]

$$(19.1) \qquad ds(X) \overset{?}{:=} \frac{d}{dt}s(\gamma(t))\Big|_{t=0} = \lim_{t \to 0} \frac{s(\gamma(t)) - s(p)}{t}$$

since $s(\gamma(t))$ and $s(p)$ belong to different vector spaces $E_{\gamma(t)} \neq E_p$. Before one can make sense of such an expression, one needs a way of identifying these vector spaces so that $s(\gamma(t)) - s(p)$ can be defined, e.g. one needs a smooth family of vector space isomorphisms

$$(19.2) \qquad P_\gamma^t : E_{\gamma(0)} \overset{\cong}{\longrightarrow} E_{\gamma(t)}, \qquad \text{such that} \qquad P_\gamma^0 = \mathbb{1}.$$

----

[57]The question mark over the equal sign in (19.1) is meant to convey a sense of confusion—because the definition does not really make sense.

Under suitable conditions to be clarified below, we will refer to families of isomorphisms of this form as **parallel transport** (*Parallelverschiebung*) (or also **parallel translation**) maps along the path γ. If such a family is given, then one can use it to turn (19.1) into a sensible definition, namely

$$(19.3) \qquad \nabla_X s := \nabla_t s(\gamma(t))|_{t=0} := \left. \frac{d}{dt}(P_\gamma^t)^{-1}(s(\gamma(t))) \right|_{t=0} = \lim_{t \to 0} \frac{(P_\gamma^t)^{-1}(s(\gamma(t))) - s(p)}{t} \in E_p.$$

This is called the **covariant derivative** (*kovariante Ableitung*) of s at p **in the direction X**, and also the **covariant derivative** of s **along the path** γ at $t = 0$.

Once parallel transport and covariant derivatives have been defined, one can also say what it means for s to be "constant" along the path γ: it means simply that

$$s(\gamma(t)) = P_\gamma^t(s(p))$$

for all t, or in terms of the covariant derivative, $\nabla_t s(\gamma(t)) \equiv 0$. Since "constant" is not really an appropriate term when the vector spaces $E_{\gamma(t)}$ vary with t, a section with this property is said to be **parallel** (or also **covariantly constant**) along the path γ. This notion clearly depends on the parallel transport isomorphisms P_γ^t, i.e. if one chose these isomorphisms differently, then a section that is parallel for one choice might not be parallel for another.

So, how does one actually go about defining parallel transport isomorphisms as in (19.2)? In the special case $E = TM$, we found one conceivable answer to this question in §6.3: one can assume that γ is a flow line of a vector field X and obtain a family of isomorphisms from the linearized flow,

$$T_p \varphi_X^t : T_p M \xrightarrow{\cong} T_{\gamma(t)} M.$$

This approach gave us the definition of the Lie derivative $\mathcal{L}_X Y$ of a vector field $Y \in \Gamma(TM) = \mathfrak{X}(M)$. The first obvious problem is that this approach only makes sense on tangent bundles, though one can perhaps imagine generalizing it to the various tensor bundles that are defined in terms of tangent bundles, leading to the Lie derivatives of tensor fields that were defined in Lecture 8. But there is a more basic problem here: a vector field $X \in \mathfrak{X}(M)$ does not define isomorphisms as in (19.2) along *arbitrary* paths γ, it only defines them along flow lines, and the derivative $\mathcal{L}_X Y$ that one ends up defining in this way is not just a derivative of Y, it also depends on the first derivative of X. (This is apparent from the local coordinate formula for $[X, Y]$ in Exercise 6.2, which matches $\mathcal{L}_X Y$ by Proposition 6.7.) For this reason, $\mathcal{L}_X Y(p)$ cannot accurately be interpreted as a directional derivative of Y at p in the direction $X(p)$.

It turns out that on a general vector bundle $E \to M$, there is no *canonical* way to define parallel transport along arbitrary paths, so instead of looking for a unique "correct" definition, it is more useful to consider what properties a reasonable definition of parallel transport should be required to satisfy. In particular, we would like the covariant derivative to behave in certain respects the way that derivatives are expected to behave, for instance:

(i) $\nabla_X s := \nabla_t s(\gamma(t))|_{t=0}$ should depend on the section $s \in \Gamma(E)$ near $p \in M$ and the tangent vector $X = \dot{\gamma}(0) \in T_p M$, but not otherwise on the path γ;

(ii) The map $T_p M \to E_p : X \mapsto \nabla_X s$ should be linear.[58]

We will see below that these two conditions lead more-or-less inevitably to the correct definition of a connection on a vector bundle.

---

[58] If $\mathbb{F} = \mathbb{C}$, then since $T_p M$ is *not* naturally a complex vector space, we ignore the complex structure of E_p in order to talk about linearity of the map $T_p M \to E_p$, i.e. "linear" just means "real-linear".

Recall that sections of E are by definition smooth maps $s : M \to E$ that satisfy $\pi \circ s = \mathrm{Id}_M$. Similarly, if γ is a smooth path in M, then a smooth path $t \mapsto s(t) \in E$ satisfying

$$\pi(s(t)) = \gamma(t)$$

for all t is called a **lift** of γ to E; equivalently, $s(t)$ belongs to the fiber $E_{\gamma(t)}$ for every t and thus defines a section of the pullback bundle $\gamma^* E$, also known as a section of E *along* γ. A family of parallel transport maps $P_\gamma^t : E_p \to E_{\gamma(t)}$ associates to every $v \in E_p$ a lift $s(t) := P_\gamma^t(v)$ of γ such that $s(0) = v$. In order to fully understand this perspective on parallel transport, it may be helpful for a while to forget that E is a vector bundle, and think of it merely as a smooth manifold that happens to be presented as a union of a smooth family of disjoint submanifolds $E_p \subset E$, its fibers. (Objects of this kind—in which the fibers are all disjoint but diffeomorphic submanifolds that need not necessarily be vector spaces—are called *fiber bundles*, and we will study them more seriously next semester.) The fibers are, in particular, the level sets of a smooth submersion $\pi : E \to M$, so differentiating π at $v \in E$ and taking its kernel gives the tangent space to the fiber containing v, which we will call the **vertical subspace** of $T_v E$:

$$V_v E := \ker\left(T_v E \xrightarrow{\pi_*} T_{\pi(v)} M \right) = T_v(E_{\pi(v)}) \subset T_v E.$$

All together, these subspaces define a distinsuished subbundle of $TE \to E$, called the **vertical subbundle**

$$V E = \ker(\pi_*) = \bigcup_{v \in E} V_v E \subset TE.$$

If we now choose to "unforget" the fact that each fiber E_p is also a vector space, then we notice that since tangent spaces to a vector space can be identified with the vector space itself, there is a canonical isomorphism

$$\mathrm{Vert}_v : E_p \xrightarrow{\cong} V_v E \qquad \text{for every } v \in E_p,\, p \in M,$$

sending $w \in E_p$ to $\frac{d}{dt}(v + tw)\big|_{t=0} \in V_v E$. A fancier way to say this is that the isomorphisms Vert_v define a canonical vector bundle isomorphism between $VE \to E$ and the pullback $\pi^* E \to E$ of $E \to M$ via its own bundle projection $\pi : E \to M$. In the lemma below, the vector space structure of the fibers E_p will be relevant for this one detail *only*, and for the most part, it will be more helpful to forget that the fibers E_p are vector spaces and think of them merely as the regular level sets of a smooth submersion $\pi : E \to M$.

LEMMA 19.1. *Suppose there is a smooth family of vector space isomorphisms*[59] $P_\gamma^t : E_{\gamma(0)} \to E_{\gamma(t)}$ *associated to every smooth path* $\gamma : (-\epsilon, \epsilon) \to M$ *such that the covariant derivative defined in* (19.3) *satisfies properties (i) and (ii) above. Then for each* $p \in M$ *and* $v \in E_p$, *there is a unique linear injection*

$$\mathrm{Hor}_v : T_p M \to T_v E$$

such that $\mathrm{Hor}_v(\dot\gamma(0)) = \frac{d}{dt} P_\gamma^t(v)\big|_{t=0}$ *for all paths* γ *with* $\gamma(0) = p$. *Moreover,* $\pi_* \circ \mathrm{Hor}_v : T_p M \to T_p M$ *is the identity map, and the image*

$$H_v E := \mathrm{im}\,\mathrm{Hor}_v \subset T_v E$$

---

[59]With very minor modifications, Lemma 19.1 is also valid on an arbitrary fiber bundle, assuming only that the maps $P_\gamma^t : E_{\gamma(0)} \to E_{\gamma(t)}$ are diffeomorphisms. In this case there may not be canonical isomorphisms $V_v E \cong E_{\pi(v)}$ since fibers need not be vector spaces, so the covariant derivative $\nabla_X s$ of a section $s \in \Gamma(E)$ in direction $X \in T_p M$ naturally takes its value in $V_{s(p)} E$ instead of E_p.

is complementary to the vertical subspace $V_v E \subset T_v E$, so it determines a splitting of TE into a direct sum of smooth subbundles,

$$TE = VE \oplus HE, \qquad where \qquad HE := \bigcup_{v \in E} H_v E.$$

PROOF. Fix $p \in M$. For any smooth path $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and any $v \in E_p$, we can think of $t \mapsto P_\gamma^t(v)$ as a smooth path in the total space of the pullback bundle $\gamma^* E = \bigcup_{t \in (-\epsilon, \epsilon)} E_{\gamma(t)}$, i.e. we regard $P_\gamma^t(v)$ is living in the fiber $(\gamma^* E)_t$. We can therefore define a smooth vector field Y on the total space of $\gamma^* E$ such that for any $v \in E_p$ and any $t \in (-\epsilon, \epsilon)$,

$$Y(P_\gamma^t(v)) = \frac{d}{dt} P_\gamma^t(v),$$

and the parallel transport maps P_γ^t can now be written in terms of the flow φ_Y^t of Y as

$$P_\gamma^t = \varphi_Y^t |_{E_p} : E_p = (\gamma^* E)_0 \to (\gamma^* E)_t = E_{\gamma(t)}.$$

The inverse of P_γ^t is then given by reversing the flow of Y, so for a section $s : M \to E$ with $s(p) = v$,

$$\mathrm{Vert}_v\left(\nabla_{\dot{\gamma}(0)} s\right) = \frac{d}{dt} \varphi_Y^{-t}(s(\gamma(t)))\Big|_{t=0} \in E_p.$$

Note that $t \mapsto \varphi_Y^{-t}(s(\gamma(t)))$ is a path through the point $\varphi_Y^0(s(p)) = s(p) = v$ in the submanifold $E_p \subset E$, so we are regarding its derivative as an element of $V_v E \subset T_v E$, so that the canonical isomorphism $\mathrm{Vert}_v^{-1} : V_v E \to E_p$ identifies it with the covariant derivative as defined in (19.3). To compute it, we write $F(t_1, t_2) = \varphi_Y^{t_1}(s(\gamma(t_2)))$ and apply the chain rule, giving

$$\mathrm{Vert}_v\left(\nabla_{\dot{\gamma}(0)} s\right) = \frac{d}{dt} F(-t, t)\Big|_{t=0} = -\frac{\partial F}{\partial t_1}(0,0) + \frac{\partial F}{\partial t_2}(0,0) = -Y(v) + Ts(\dot{\gamma}(0)) \in T_v E,$$

and thus

(19.4) $$\mathrm{Hor}_v(\dot{\gamma}(0)) := \frac{d}{dt} P_\gamma^t(v)\Big|_{t=0} = Y(v) = Ts(\dot{\gamma}(0)) - \mathrm{Vert}_v\left(\nabla_{\dot{\gamma}(0)} s\right) \in T_v E.$$

Properties (i) and (ii) above imply that this expression is a linear function of $\dot{\gamma}(0)$ and does not otherwise depend on the choice of path γ. Writing $X := \dot{\gamma}(0)$, Hor_v also satisfies

$$\pi_* \circ \mathrm{Hor}_v(X) = T\pi\left(\frac{d}{dt} P_\gamma^t(v)\Big|_{t=0}\right) = \frac{d}{dt}\left(\pi \circ P_\gamma^t(v)\right)\Big|_{t=0} = \frac{d}{dt}\gamma(t)\Big|_{t=0} = X,$$

thus $\mathrm{Hor}_v : T_p M \to T_v E$ is injective and its image $H_v E := \mathrm{im}\,\mathrm{Hor}_v$ has trivial intersection with $\ker \pi_* = V_v E$. Finally, we observe that any non-vertical vector $\xi \in T_v E \backslash V_v E$ can be written as $Ts(\dot{\gamma}(0))$ for some path γ and section s, and we then have $\xi = \mathrm{Hor}_v(\dot{\gamma}(0)) + \mathrm{Vert}_v\left(\nabla_{\dot{\gamma}(0)} s\right) \in H_v E + V_v E$, thus

$$H_v E \oplus V_v E = T_v E.$$

$\square$

Any subbundle

$$HE \subset TE$$

that satisfies $TE = VE \oplus HE$ as in Lemma 19.1 is called a **horizontal subbundle** of TE. Unlike VE, horizontal subbundles are not unique or canonical, but a choice of horizontal subbundle is equivalent via the formula $H_v E = \mathrm{im}\,\mathrm{Hor}_v$ to a choice of a smoothly varying family of linear **horizontal lift** maps,

$$\mathrm{Hor}_v : T_p M \to T_v E \qquad such\ that \qquad \pi_* \circ \mathrm{Hor}_v(X) = X \text{ for all } X \in T_p M.$$

The lemma tells us that any sensible choice of parallel transport maps for E along smooth paths in M determines a horizontal subbundle in a natural way. Conversely, any horizontal subbundle $HE \subset TE$ uniquely determines parallel transport maps by requiring all parallel lifts $s(t) := P_\gamma^t(v) \in E$ of paths $\gamma(t) \in M$ to be tangent to HE, i.e. the derivative $\dot{s}(t) \in T_{s(t)}E$ should always be horizontal, which means it is the horizontal lift of the derivative of γ:

$$(19.5) \qquad\qquad \dot{s}(t) = \mathrm{Hor}_{s(t)}(\dot{\gamma}(t)).$$

This is a first-order ordinary differential equation, so $s(t)$ is uniquely determined by the initial condition $s(0) = v$. One can also see this by using horizontal lifts to define a vector field on the total space of γ^*E as in the proof of Lemma 19.1; the parallel transport maps are then given by the flow of that vector field.

This is as far as we can go without paying attention to the fact that fibers E_p are vector spaces, and you may notice that a problem has arisen from this relaxation of assumptions. Indeed, for an arbitrary choice of horizontal subbundle $HE \subset TE$, there is no guarantee that the ODE in (19.5) with any given initial condition $s(0) = v$ will have solutions beyond an arbitrarily small interval around $t = 0$, and if it does, then the resulting family of maps $P_\gamma^t : E_p \to E_{\gamma(t)}$ will be diffeomorphisms, but they need not be linear. The following useful characterization of linearity will provide an easy remedy for this.

LEMMA 19.2. *Let V and W be normed vector spaces over $\mathbb{F}$. Then any map $F : V \to W$ that is differentiable[60] at 0 and satisfies $F(\lambda v) = \lambda F(v)$ for all scalars $\lambda \in \mathbb{F}$ and all $v \in V$ is linear.*

PROOF. The key is to show that under this assumption, F is actually equal to its derivative at zero, $DF(0) : V \to W$. Clearly $F(0) = 0$, so we can write

$$F(v) = DF(0)v + \|v\| \cdot R(v)$$

for some function $R : V \to W$ such that $\lim_{v \to 0} R(v) = 0$. Then taking $\lambda > 0$,

$$F(v) = \lim_{\lambda \to 0^+} \frac{1}{\lambda} F(\lambda v) = \lim_{\lambda \to 0^+} \frac{DF(0)\lambda v + \lambda \|v\| \cdot R(\lambda v)}{\lambda} = DF(0)v + \lim_{\lambda \to 0^+} \|v\| \cdot R(\lambda v) = dF(0)v.$$

Since $DF(0) : V \to W$ is real linear, this proves that F also respects vector addition. $\qquad\square$

On our vector bundle $E \to M$, each scalar $\lambda \in \mathbb{F}$ defines a smooth map

$$m_\lambda : E \to E : v \mapsto \lambda v,$$

which is a diffeomorphism for $\lambda \neq 0$, and its tangent map $(m_\lambda)_* : TE \to TE$ then defines vector space isomorphisms $(m_\lambda)_* : T_v E \to T_{\lambda v} E$ for every $v \in E$.

LEMMA 19.3. *For a given horizontal subbundle $HE \subset TE$, the following conditions are equivalent:*

  (i) *The parallel transport maps $P_\gamma^t : E_p \to E_{\gamma(t)}$ defined via (19.5) exist for every t in the domain of an arbitrary smooth path γ, and are linear.*
  (ii) *For all $v \in E$ and $\lambda \in \mathbb{F}$, $H_{\lambda v}E = (m_\lambda)_*(H_v E)$.*

PROOF. For any $p \in M$, $v \in E_p$, $\lambda \in \mathbb{F}$ and a smooth path γ in M through $\gamma(0) = p$, assume first that P_γ^t exists and is linear for every t. Writing $s(t) := P_\gamma^t(v)$, we have $\dot{s}(0) \in H_v E$ by definition. The corresponding lift with initial condition $\lambda v \in E_p$ is then $P_\gamma^t(\lambda v) = \lambda s(t) = m_\lambda(s(t))$, implying

$$\left. \frac{d}{dt} m_\lambda(s(t)) \right|_{t=0} = (m_\lambda)_* \dot{s}(0) \in H_{\lambda v}E,$$

---

[60]If $\mathbb{F} = \mathbb{C}$, then differentiability of $F : V \to W$ can be taken to mean the same thing as in the real case, i.e. we simply regard V and W as real vector spaces, so the derivative $DF(0) : V \to W$ is a real-linear map. It is not necessary to assume that $DF(0)$ is complex linear, which would be a holomorphicity condition on F.

hence $(m_\lambda)_*$ maps $H_v E$ to $H_{\lambda v} E$. Conversely, if this condition on HE holds, then for $\lambda \neq 0$, the same calculation implies that $s(t)$ is a horizontal lift of $\gamma(t)$ if and only if $\lambda s(t)$ is. Here it is convenient to assume $\lambda \neq 0$ so that the map $T_v E \xrightarrow{(m_\lambda)_*} T_{\lambda v} E$ is an isomorphism, but since the fibers of HE vary continuously, one can also take $\lambda \to 0$ and conclude that at points along the zero-section

$$Z := \bigcup_{p \in M} \{0\} \subset E,$$

the horizontal subspaces are uniquely determined, namely $H_v E = T_v Z$ whenever $v \in Z$. It follows that solutions to (19.5) with initial condition $s(0) = 0$ exist for all t and are identically zero. This implies in turn that for any t, solutions also exist with initial condition in some sufficiently small neighborhood of 0, but the ability to find further solutions via multiplication with arbitrary scalars now produces solutions for all t with arbitrary initial conditions. Moreover, the resulting diffeomorphisms $P_\gamma^t : E_p \to E_{\gamma(t)}$ are smooth and respect scalar multiplication, so by Lemma 19.2, they are linear. $\square$

19.2. Two equivalent definitions. The point of the previous section was to motivate the following definition.

DEFINITION 19.4 (Connections, version 1). A **connection** (*Zusammenhang*) on the vector bundle $\pi : E \to M$ is a choice of subbundle

$$HE \subset TE$$

that is complementary to the vertical subbundle $VE \subset TE$ and satisfies $(m_\lambda)_* (HE) = HE$ for every scalar $\lambda \in \mathbb{F}$.

It should not be obvious to you at this stage whether connections always exist—they do, but this is something we will have to prove,, and the proof unsurprisingly requires a partition of unity. If a connection is chosen, then it determines the notions of parallel transport, horizontal lifts and covariant derivatives as we defined them in the previous section.

There are at least two other popular ways to reformulate Definition 19.4. One of them uses the fact that a connection determines a splitting $TE = VE \oplus HE$, and splittings of vector spaces (or bundles) can be characterized in terms of linear projection maps. Indeed, let

$$\widehat{K} : TE \to VE$$

be the unique smooth linear bundle map that restricts to the identity on VE and vanishes on HE, so $HE = \ker \widehat{K}$. Since each vertical subspace $V_v E$ is canonically isomorphic to the fiber $E_{\pi(p)}$, we can compose $\widehat{K}$ with the resulting canonical map $VE \to E$ to produce a map $K : TE \to E$ as in the following definition.

DEFINITION 19.5 (Connections, version 2). A **connection** (*Zusammenhang*) on the vector bundle $\pi : E \to M$ is a smooth map $K : TE \to E$ such that

(1) For each $v \in E$, K defines a real-linear map $T_v E \to E_{\pi(v)}$.[61]
(2) $K(\text{Vert}_v(w)) = w$ for all $v, w \in E_p$, $p \in M$.
(3) For all scalars $\lambda \in \mathbb{F}$, $K \circ (m_\lambda)_* = m_\lambda \circ K$.

---

[61]We emphasize that if $\mathbb{F} = \mathbb{C}$, then E must be treated as a real vector space for the purposes of this condition, as the tangent spaces $T_v E$ are not complex in any natural way. This is because M is only a real manifold, not complex.

To get from Definition 19.5 back to Definition 19.4, one defines a hoirzontal subbundle $HE \subset TE$ from $K : TE \to E$ by

$$HE := \ker(K) \subset TE,$$

meaning $H_v E$ is the kernel of the linear map $T_v E \xrightarrow{K} E_{\pi(v)}$.

EXERCISE 19.6. Show that under the correspondence described above between horizontal subbundles $HE \subset TE$ and maps $K : TE \to E$, the condition $(m_*)(HE) = HE$ is equivalent to $K \circ (m_\lambda)_* = m_\lambda \circ K$ for all $\lambda \in \mathbb{F}$.

The projection $K : TE \to E$ provides a simpler formula for the covariant derivative of a section $s \in \Gamma(E)$ in the direction of a tangent vector $X \in T_p M$ at a point $p \in M$. Recall from (19.4) the relation

$$Ts(X) = \mathrm{Hor}_{s(p)}(X) + \mathrm{Vert}_{s(p)}(\nabla_X s).$$

Since K annihilates horizontal vectors, applying it to both sides of this relation gives

(19.6) $\boxed{\nabla_X s = K \circ Ts(X)}$,

so the covariant derivative is actually just the "vertical part" of the tangent map of $s : M \to E$ in the direction of X, obtained by removing from $Ts(X)$ its horizontal part and then identifying the resulting vertical vector with an element of E_p. Note that although the vertical subbundle is independent of any choices, the notion of a "vertical part" of a vector in TE does depend on the choice of the complementary subbundle $HE \subset TE$ along which to project it. The covariant derivative thus depends on the choice of connection, except in certain special situations such as the following.

EXERCISE 19.7. Assume $E \to M$ is a vector bundle and $s \in \Gamma(E)$. The **zero set** (*Nullstelle*) of s, sometimes denoted by $s^{-1}(0) \subset M$, is the set of all points $p \in M$ such that $s(p)$ is the zero vector in its respective fiber.

(a) Show that for any $p \in s^{-1}(0)$ the linear map

$$Ds(p) : T_p M \to E_p : X \mapsto \nabla_X s$$

is independent of the choice of connection (needed for defining $\nabla_X s$). We call this the **linearization** (*Linearisierung*) of s at p.

(b) We say that $s \in \Gamma(E)$ is **transverse to the zero-section** (*transversal zum Nullschnitt*) if the linearization $Ds(p) : T_p M \to E_p$ is surjective for every $p \in s^{-1}(0)$. Show that whenever this holds, $s^{-1}(0)$ is a smooth submanifold of M, with dimension $\dim(M) - \mathrm{rank}(E)$ in the case $\mathbb{F} = \mathbb{R}$, or $\dim(M) - 2\,\mathrm{rank}(E)$ in the case $\mathbb{F} = \mathbb{C}$.
Hint: Any local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ *determines a connection on* $E|_{\mathcal{U}_\alpha}$ *such that covariantly differentiating s over $\mathcal{U}_\alpha$ becomes equivalent to differentiating its local repesentative* $s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}$.

20. More on connections

As in the previous lecture, we fix a smooth vector bundle $\pi : E \to M$ of rank m over the field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, where the base M is a smooth n-manifold.

20.1. The Leibniz rule (a third definition). In our discussion so far, choosing a connection on $\pi : E \to M$ means choosing a horizontal subbundle $HE \subset TE$ that satisfies the conditions of Definition 19.4, or equivalently, a map $K : TE \to E$ satisfying the conditions of Definition 19.5. (We will sometimes refer to K as the **vertical projection** defining the connection.) We have two ways of writing down the covariant derivative operator determined by this connection: for a section

$s \in \Gamma(E)$, point $p \in M$ and tangent vector $X \in T_p M$, we can choose a smooth path $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\dot\gamma(0) = X$, and write

$$(20.1) \qquad \nabla_X s = \frac{d}{dt}(P_\gamma^t)^{-1}\left(s(\gamma(t))\right)\Big|_{t=0} \in E_p.$$

Alternatively, we saw in (19.6) that $\nabla_X s$ can be written in terms of $K : TE \to E$ and the tangent map $Ts : TM \to TE$ of $s : M \to E$ as

$$\nabla_X s = K(Ts(X)).$$

While this formula looks simpler, (20.1) is often more useful for proving basic properties of the covariant derivative, for instance the Leibniz rule in Exercise 20.3 below.

EXAMPLE 20.1. On the trivial bundle $E = M \times \mathbb{F}^m$, there is a natural **trivial connection**, defined by viewing the two factors in the obvious splitting $T_{(p,v)}E = T_p M \oplus T_v \mathbb{F}^m = T_p M \oplus \mathbb{F}^m$ as horizontal and vertical subspaces respectively. The vertical projection $K : TE \to E$ is then given by $K(X, w) = w \in \mathbb{F}^m = E_v$ for $(X, w) \in T_p M \oplus \mathbb{F}^m = T_{(p,v)}E$, and the parallel transport maps $P_\gamma^t : E_{\gamma(0)} = \mathbb{F}^m \to \mathbb{F}^m = E_{\gamma(t)}$ are all the identity map on $\mathbb{F}^m$. Under the obvious identification of sections $s \in \Gamma(E)$ with functions $f : M \to \mathbb{F}^m$, the covariant derivative $\nabla_X s$ is then simply the differential $df(X)$.

Since $\nabla_X s$ depends linearly on X, the covariant derivative of $s \in \Gamma(E)$ in all possible directions can be packaged as a section

$$\nabla s \in \Gamma(\mathrm{Hom}(TM, E))$$

defined by $\nabla s(p)(X) := \nabla_X s$. There is a clear analogy here with differentials: a real-valued function $f : M \to \mathbb{R}$ is the same thing as a section of the trivial real line bundle $M \times \mathbb{R} \to M$, and its differential assigns to every point $p \in M$ the linear map $d_p f : T_p M \to \mathbb{R}$. The covariant derivative of $s \in \Gamma(E)$ similarly assigns to each point $p \in M$ a linear map $\nabla s(p) : T_p M \to E_p : X \mapsto \nabla_X s$, defining what is sometimes called a "bundle-valued" 1-form $\nabla s \in \Omega^1(M, E) := \Gamma(\mathrm{Hom}(TM, E))$. Note that if E is a complex vector bundle, then $\mathrm{Hom}(TM, E)$ means the bundle of *real*-linear maps from TM to E, since TM is not naturally a complex bundle. On the other hand, $\mathrm{Hom}(TM, E)$ does have a natural complex structure if E is complex, in which case $\Gamma(\mathrm{Hom}(TM, E))$ is also a complex vector space and one can therefore speak (as in Exercise 20.2 below) of complex-linear maps $\Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$.

The next two exercises are both easy applications of (20.1), and depend crucially on the fact that parallel transport maps $P_\gamma^t : E_{\gamma(0)} \to E_{\gamma(t)}$ are linear.

EXERCISE 20.2. Show that the map $\nabla : \Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$ is linear.

EXERCISE 20.3. Show that for any $s \in \Gamma(E)$ and $f \in C^\infty(M, \mathbb{F})$, the covariant derivative of $fs \in \Gamma(E)$ in a direction $X \in T_p M$ at $p \in M$ satisfies

$$\nabla_X(fs) = df(X)s(p) + f(p)\nabla_X s.$$

This Leibniz rule is often abbreviated in the form

$$(20.2) \qquad \nabla(fs) = df \cdot s + f\,\nabla s.$$

Exercises 20.2 and 20.3 have a converse of sorts, which leads to yet another equivalent version of the definition of a connection.

DEFINITION 20.4 (Connections, version 3). A **connection** (*Zusammenhang*) on the vector bundle $\pi : E \to M$ is a linear operator

$$\nabla : \Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$$

that satisfies the Leibniz rule (20.2) for all $f \in C^\infty(M, \mathbb{F})$ and $s \in \Gamma(E)$.

To see that this is equivalent to our previous two definitions, we need to show that every linear operator $\nabla : \Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$ satisfying the Leibniz rule (20.2) is in fact the covariant derivative operator determined by a unique connection in the sense of Definitions 19.4 and 19.5. The uniqueness is an easy consequence of (19.4), since for any $p \in M$, $X \in T_p M$ and $v \in E_p$, one can choose any section $s \in \Gamma(E)$ with $s(p) = v$ and write

$$(20.3) \qquad \mathrm{Hor}_v(X) = Ts(X) - \mathrm{Vert}_v(\nabla s(X)),$$

thus using the operator ∇ to determine the horizontal lift maps Hor_v, and in this way the horizontal subbundle $HE \subset TE$. Existence will follow similarly if we can show that the right hand side of this expression does not depend on the choice of section s with $s(p) = v$. The following result will help, and is important for other reasons as well.

PROPOSITION 20.5. *For any two connections $\nabla, \widehat{\nabla}$ on $\pi : E \to M$ in the sense of Definition 20.4, there exists a smooth linear bundle map $A : E \to \mathrm{Hom}(TM, E)$ such that $\widehat{\nabla}s = \nabla s + As$ for all $s \in \Gamma(E)$.*

PROOF. We can use a minor adaptation of the notion of C^∞-linearity from §8.1. For any two vector bundles E and F, a smooth linear bundle map $A : E \to F$ defines a linear map $\Gamma(E) \to \Gamma(F) : s \mapsto As$ that is also C^∞-linear in the sense that fs is sent to $f \cdot As$ for any $f \in C^\infty(M, \mathbb{F})$. Conversely, any C^∞-linear map $\widehat{A} : \Gamma(E) \to \Gamma(F)$ arises in this way from a smooth linear bundle map $A : E \to F$, meaning in particular that for any $s \in \Gamma(E)$, the value at any given point $p \in M$ of the section $\widehat{A}s \in \Gamma(F)$ is determined by $s(p) \in E_p$ and is otherwise independent of the section s. The proof of this statement is almost identical to that of Proposition 8.2.

With this understood, we observe that while the term $df \cdot s$ in the Leibniz rule prevents either of ∇ or $\widehat{\nabla}$ from being a C^∞-linear map $\Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$, this term is identical for both connections and thus cancels when we consider $A := \widehat{\nabla} - \nabla : \Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$. It follows that the latter is C^∞-linear, and thus arises from a bundle map $E \to \mathrm{Hom}(TM, E)$. $\qquad \square$

REMARK 20.6. By Definition 20.4, the set of all connections on $\pi : E \to M$ can be regarded as a subset of the infinite-dimensional vector space $\mathrm{Hom}(\Gamma(E), \Gamma(\mathrm{Hom}(TM, E)))$, but it is not a linear subspace, e.g. it does not contain the zero element of this space. Proposition 20.5 shows however that it is an *affine* space over the vector space $\Gamma(\mathrm{Hom}(E, \mathrm{Hom}(TM, E)))$, which sits naturally inside $\mathrm{Hom}(\Gamma(E), \Gamma(\mathrm{Hom}(TM, E)))$ as the space of all C^∞-linear maps $\Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$. This shows in particular that the set of connections is convex.

Returning to (20.3), the right hand side clearly depends on $s \in \Gamma(E)$ only in a neighborhood of p, thus we are free to restrict our attention to a small neighborhood $\mathcal{U} \subset M$ of p on which E is a trivial bundle. Choosing a trivialization $\Phi : E|_\mathcal{U} \to \mathcal{U} \times \mathbb{F}^m$ yields a corresponding choice of parallel transport isomorphisms, for which sections are parallel if and only if their representations in the local trivialization are constant—this defines a connection in the sense of our previous two definitions, and it matches the "trivial" connection of Example 20.1 if we use Φ to identify $E|_\mathcal{U}$ with the trivial bundle $\mathcal{U} \times \mathbb{F}^m$. Let us denote the horizontal lift and covariant derivative operators for this connection by $\widehat{\mathrm{Hor}}_v$ and $\widehat{\nabla}$ respectively. According to Proposition 20.5, $\widehat{\nabla} = \nabla + A$ for a bundle map $A : E \to \mathrm{Hom}(TM, E)$, and (20.3) therefore becomes

$$\mathrm{Hor}_v(X) = Ts(X) - \mathrm{Vert}_v(\nabla s(X)) = Ts(X) - \mathrm{Vert}_v\left(\widehat{\nabla}s(X) - Av(X)\right)$$

$$= Ts(X) - \mathrm{Vert}_v\left(\widehat{\nabla}s(X)\right) + \mathrm{Vert}_v\left(Av(X)\right) = \widehat{\mathrm{Hor}}_v(X) + \mathrm{Vert}_v\left(Av(X)\right).$$

Now it is clear that the right hand side does not depend on the choice of section s satisfying $s(p) = v$, thus proving that any operator ∇ as in Definition 20.4 uniquely determines a horizontal subbundle

$HE \subset TE$ whose covariant derivative operator is ∇. That the parallel transport maps arising from HE are linear can then be deduced from the assumption that $\nabla : \Gamma(E) \to \Gamma(\operatorname{Hom}(TM, E))$ is linear: indeed, for any path γ with $\dot{\gamma}(0) = X \neq 0 \in T_pM$, sections $s \in \Gamma(E)$ that are parallel along γ are characterized by the condition

$$\nabla_{\dot{\gamma}(t)} s = 0 \quad \text{for all } t,$$

and the set of solutions to this equation is a vector space. It follows via Lemma 19.3 that HE satisfies the conditions of Definition 19.4, and all three of our definitions of a connection are therefore equivalent.[62]

20.2. Local coordinates and Christoffel symbols.
There are two standard ways to present a connection in local coordinates, and both rely mainly on the same two facts: (1) every trivialization determines a corresponding *trivial* connection as in Example 20.1, and (2) by Proposition 20.5, every other connection differs from that one by a bundle map. This bundle map always appears in coordinates as a so-called "zeroth-order" term, meaning that unlike the covariant derivative itself, it is not a *differential* operator.

Fix a local trivialization

$$\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$$

over some open subset $\mathcal{U}_\alpha \subset M$, and let D denote the covariant derivative operator for the resulting trivial connection on $E|_{\mathcal{U}_\alpha}$, i.e. the one for which sections are parallel if and only if Φ_α identifies them with constant functions. Given any other connection ∇ on E, $\nabla - D$ then defines a smooth linear bundle map $A : E|_{\mathcal{U}_\alpha} \to \operatorname{Hom}(TM, E)|_{\mathcal{U}_\alpha}$, which we shall write in the form

$$(Av)X = \Gamma_\alpha(X, v) \in E_p \quad \text{for} \quad p \in \mathcal{U}_\alpha, \ X \in T_pM, \ v \in E_p,$$

thus defining a smooth *bilinear* bundle map

$$\Gamma_\alpha : (TM \oplus E)|_{\mathcal{U}_\alpha} \to E|_{\mathcal{U}_\alpha}.$$

For any section $s \in \Gamma(E)$, ∇s can now be expressed over $\mathcal{U}_\alpha$ in the form

$$(20.4) \qquad\qquad \nabla_X s = D_X s + \Gamma_\alpha(X, s(p)) \qquad \text{for } p \in \mathcal{U}_\alpha, \ X \in T_pM.$$

Note that Γ_α is real linear in the first factor and $\mathbb{F}$-linear in the second. It must be emphasized that Γ_α is not globally defined, and it depends on the choice of trivialization.

One sees Γ_α expressed more often in local coordinates as a set of locally defined functions with three indices. Assume $\mathcal{U}_\alpha$ admits a coordinate chart $(x^1, \ldots, x^n)$; this then determines a frame $(\partial_1, \ldots, \partial_n)$ for the tangent bundle $TM|_{\mathcal{U}_\alpha}$. There is similarly a frame $(e_1, \ldots, e_m)$ for $E|_{\mathcal{U}_\alpha}$ corresponding to the trivialization Φ_α. Then there are smooth functions

$$\Gamma^a_{ib} : \mathcal{U}_\alpha \to \mathbb{F}, \qquad i \in \{1, \ldots, n\}, \ a, b \in \{1, \ldots, m\}$$

uniquely determined by the condition

$$\Gamma_\alpha(\partial_i, e_b) = \Gamma^a_{ib} e_a.$$

For any $X = X^i \partial_i \in T_pM$ and $v = v^b e_b \in E_p$ at a point $p \in \mathcal{U}_\alpha$, we then have

$$\Gamma_\alpha(X, v) = \Gamma_\alpha(X^i \partial_i, v^b e_b) = X^i v^b \Gamma_\alpha(\partial_i, e_b) = \Gamma^a_{ib} X^i v^b e_a,$$

so the ath component of $\Gamma_\alpha(X, v) \in E_p$ with respect to the frame $e_1, \ldots, e_m$ is

$$(\Gamma_\alpha(X, v))^a = \Gamma^a_{ib} X^i v^b.$$

The functions Γ^a_{ib} are called the **Christoffel symbols** determined by the connection.

---

[62] The argument for why Definition 20.4 implies Definition 19.4 unfortunately got skipped in the lecture, due mainly to absent-mindedness.

Recall that any section $s \in \Gamma(E)$ can be expressed over $\mathcal{U}_\alpha$ in terms of its component functions $s^1, \ldots, s^m : \mathcal{U}_\alpha \to \mathbb{F}$ as $s = s^a e_a$. Let us write

$$\nabla_i := \nabla_{\partial_i} = \nabla_{\frac{\partial}{\partial x^i}}$$

for the covariant derivative operator in the ith coordinate direction. One now obtains another formula for the Christoffel symbols from (20.4), using the observation that the frame sections $e_1, \ldots, e_m$ all satisfy $De_a \equiv 0$ by the definition of the trivial connection. Indeed, this together with (20.4) implies $\nabla_i e_b = \Gamma_\alpha(\partial_i, e_b) = \Gamma_{ib}^a e_a$, and thus

$$(20.5) \qquad\qquad\qquad\qquad \Gamma_{ib}^a = (\nabla_i e_b)^a.$$

For a general section $s = s^a e_a$ over $\mathcal{U}_\alpha$, we then apply the Leibniz rule to compute

$$\nabla_i s = \nabla_i(s^b e_b) = (\partial_i s^b) e_b + s^b \nabla_i e_b = \left(\partial_i s^a + \Gamma_{ib}^a s^b \right) e_a,$$

where we've relabelled the summed index in the first term and used (20.5) in the second term, giving rise to the formula

$$(20.6) \qquad\qquad\qquad\qquad (\nabla_i s)^a = \partial_i s^a + \Gamma_{ib}^a s^b.$$

This is of practical use for coordinate computations of covariant derivatives.

You can see from (20.6) that the Christoffel symbols Γ_{ib}^a fully determine the covariant derivative operator, and therefore the connection, at least over the region $\mathcal{U}_\alpha$. This observation gives rise to yet another variant of the definition of a connection, one that is not very elegant, but is favored by physicists: a connection is an association to every open set $\mathcal{U}_\alpha \subset M$ with a chart $x^1, \ldots, x^n$ and local trivialization $(\mathcal{U}_\alpha, \Phi_\alpha)$ of a set of smooth functions $\Gamma_{ib}^a : \mathcal{U}_\alpha \to \mathbb{F}$, which are then fed into (20.6) to define the covariant derivative. Of course, the functions Γ_{ib}^a cannot be chosen arbitrarily for all possible local trivializations and charts: once they have been chosen for one particular chart and trivialization over a set $\mathcal{U}_\alpha$, the connection over $\mathcal{U}_\alpha$ is fully determined, and any choice on a different region $\mathcal{U}_\beta$ (with different coordinates and trivialization) that overlaps $\mathcal{U}_\alpha$ had better give the same result on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$. One must therefore derive a suitable transformation formula for Christoffel symbols under changes of coordinates and local trivializations, and make sure that that formula is always satisfied. The unfortunate fact is that the correct transformation formula does not follow from anything we've already done, because Christoffel symbols do *not* define a tensor, i.e. since Γ_α always depends on the choice of trivialization Φ_α and is defined only on $\mathcal{U}_\alpha$, there is generally no globally defined tensor field or section of any vector bundle whose locally-defined components are the functions Γ_{ib}^a. This does not make the situation impossible, it only means there is still some work to be done if you want to use Christoffel symbols as a complete characterization of a connection. We leave the details as an exercise:

EXERCISE 20.7. Given a bundle $\pi : E \to M$ and a sufficiently small open set $\mathcal{U} \subset M$, let us use a coordinate chart $(x^1, \ldots, x^n)$ and a frame $(e_1, \ldots, e_m)$ to identify $E|_\mathcal{U}$ with the trivial bundle $\mathcal{V} \times \mathbb{F}^m$, where $\mathcal{V}$ is an open subset of $\mathbb{R}^n$. Suppose Γ_{ib}^a are the corresponding Christoffel symbols for some connection ∇ on E. Then another choice of coordinates and frame over the same region can be expressed via smooth functions

$$\mathcal{V} \to \mathbb{R}^n : (x^1, \ldots, x^n) \mapsto (\tilde{x}^1, \ldots, \tilde{x}^n)$$
$$\mathcal{V} \to \mathbb{R}^m : (x^1, \ldots, x^n) \mapsto \tilde{e}_1 = (\tilde{e}_1^1, \ldots, \tilde{e}_1^m)$$
$$\vdots$$
$$\mathcal{V} \to \mathbb{R}^m : (x^1, \ldots, x^n) \mapsto \tilde{e}_m = (\tilde{e}_m^1, \ldots, \tilde{e}_m^m)$$

Let $\widetilde{\Gamma}^a_{ib}$ denote the Christoffel symbols of ∇ with respect to the coordinates $(\tilde{x}^1, \ldots, \tilde{x}^n)$ and frame $(\tilde{e}_1, \ldots, \tilde{e}_m)$. Derive the transformation formula

$$\widetilde{\Gamma}^a_{ib} = \frac{\partial x^j}{\partial \tilde{x}^i} \tilde{e}^c_b \Gamma^a_{jc} + \frac{\partial x^j}{\partial \tilde{x}^i} \frac{\partial}{\partial x^j} \tilde{e}^a_b.$$

As a special case when $E = TM$, show that this becomes

$$\widetilde{\Gamma}^i_{jk} = \frac{\partial x^p}{\partial \tilde{x}^j} \frac{\partial x^q}{\partial \tilde{x}^k} \Gamma^i_{pq} + \frac{\partial x^p}{\partial \tilde{x}^j} \frac{\partial}{\partial x^p} \left(\frac{\partial x^i}{\partial \tilde{x}^k} \right).$$

Remark: I have to be honest—I don't actually recommend doing this exercise. But a physicist would consider it essential.

20.3. Connection 1-forms and G-structures. As an alternative to the Christoffel symbols, one can express covariant derivatives in local trivializations via matrix-valued 1-forms. Suppose again that $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ is a trivialization over some open subset $\mathcal{U}_\alpha \subset M$, and write

$$\Phi_\alpha(v) = (p, v_\alpha) \qquad \text{for } p \in \mathcal{U}_\alpha, \, v \in E_p,$$

thus defining $v_\alpha \in \mathbb{F}^m$. This is just a pointwise version of our usual "local representation" of sections $s \in \Gamma(E)$ over $\mathcal{U}_\alpha$ as functions $s_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$, defined such that

$$\Phi_\alpha \circ s(p) = (p, s_\alpha(p)) \qquad \text{for } p \in \mathcal{U}_\alpha.$$

In terms of local repesentatives, the trivial connection D determined by the trivialization Φ_α becomes (cf. Example 20.1) the standard differential, meaning

$$(D_X s)_\alpha = ds_\alpha(X) \qquad \text{for } p \in \mathcal{U}_\alpha, \, X \in T_p M.$$

Any other connection ∇ is related to this one by a bundle map $E \to \mathrm{Hom}(TM, E)$ over $\mathcal{U}_\alpha$. We defined the Christoffel symbols by reinterpreting this as a bilinear bundle map $TM \oplus E \to E$, but we could also choose to interpret it instead as a bundle map $TM \to \mathrm{End}(E)$ over $\mathcal{U}_\alpha$. Using the trivialization to identify fibers of E with $\mathbb{F}^m$, the fibers of $\mathrm{End}(E)$ then become the space of matrices $\mathbb{F}^{m \times m}$, and we deduce the existence of a unique m-by-m matrix-valued 1-form

$$A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$$

such that the covariant derivative ∇ is given in the local trivialization over $\mathcal{U}_\alpha$ by the formula

$$(20.7) \qquad (\nabla_X s)_\alpha(p) = ds_\alpha(X) + A_\alpha(X) s_\alpha(p), \qquad \text{for } p \in \mathcal{U}_\alpha, \, X \in T_p M,$$

often abbreviated as

$$(\nabla s)_\alpha = ds_\alpha + A_\alpha s_\alpha.$$

A word on notation: for any manifold M and any finite-dimensional (real or complex) vector space V, we will from now on denote by

$$\Omega^1(M, V) = \{\text{smooth "}V\text{-valued" 1-forms on } M\}$$

the vector space of smooth maps $\omega : TM \to V$ whose restrictions $\omega_p : T_p M \to V$ to the tangent space over each point $p \in M$ are real-linear maps. In the case $V = \mathbb{F}^{m \times m}$ seen above, elements of $\Omega^1(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$ can also be imagined as m-by-m matrices whose individual entries are smooth $\mathbb{F}$-valued 1-forms on $\mathcal{U}_\alpha$.

The existence and uniqueness of the **connection 1-form** $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$ satisfying (20.7) was deduced above from Proposition 20.5, but if you prefer, you could also derive a precise formula for A_α from the Christoffel symbols:

EXERCISE 20.8. Given a coordinate chart $(x^1, \ldots, x^n)$ on $\mathcal{U}_\alpha$, show that at each point in $\mathcal{U}_\alpha$ and for each $i = 1, \ldots, n$, the entries $A_\alpha(\partial_i)^a{}_b$ of the matrix $A_\alpha(\partial_i) \in \mathbb{F}^{m \times m}$ are the Christoffel symbols Γ^a_{ib}.

Equation 20.7 leads to yet another somewhat untidy definition of connections that is nonetheless popular in the physics world: a connection is a choice of m-by-m matrix-valued 1-forms A_α over open subsets $\mathcal{U}_\alpha$, one for each local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$, such that a certain transformation property with respect to change of trivializations on overlap regions is satisfied (see the exercise below).

EXERCISE 20.9. If $g = g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{GL}(m, \mathbb{F})$ is the transition map relating two trivializations Φ_α and Φ_β, show that the connection 1-forms A_α and A_β are related on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ by

$$A_\alpha(X) = g(p)^{-1} A_\beta(X) g(p) + g(p)^{-1}\, dg(X), \qquad \text{for } p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta,\, X \in T_p M.$$

This transformation formula is often abbreviated by

$$(20.8) \qquad\qquad A_\alpha = g^{-1} A_\beta g + g^{-1}\, dg \qquad \text{on } \mathcal{U}_\alpha \cap \mathcal{U}_\beta.$$

Physicists refer to (20.8) as a **gauge transformation** (*Eichtransformation*), alluding to the important role that connection 1-forms play in quantum field theory: in that context they are called *gauge fields*, and they serve to model elementary particles such as photons and other "gauge bosons" that mediate the fundamental forces of nature. The choice of the letter A to denote a connection form is in fact motivated by physics, where the *vector potential* of classical electromagnetic field theory (conventionally denoted by $\mathbf{A}$) can be interpreted as a connection form for a trivial Hermitian line bundle.

There is another reason to use connection 1-forms rather than Christoffel symbols when the vector bundle has extra structure. In this case it's appropriate to restrict attention to a particular class of connections, and it turns out that this restriction can be expressed elegantly via the connection forms.

DEFINITION 20.10. Let $\pi : E \to M$ be a vector bundle with a G-structure, for some Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$. Then a connection ∇ on E is called G-**compatible** if all parallel transport isomorphisms respect the G-structure: this means that for any path $\gamma(t) \in \mathcal{U}_\alpha$ in the domain of a G-compatible local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$, the maps $P_\gamma^t : E_{\gamma(0)} \to E_{\gamma(t)}$ satisfy

$$\Phi_\alpha \circ P_\gamma^t \circ \Phi_\alpha^{-1}(\gamma(0), v) = (\gamma(t), g(t)v) \qquad \text{for all } t \text{ and } v \in E_{\gamma(0)},$$

where $g(t) \in G$ is a smooth path in G through $g(0) = \mathbb{1}$.

The definition seems less abstract when we apply it to particular structures: e.g. for $G = \mathrm{O}(m)$ or $\mathrm{U}(m)$, the structure in question is a bundle metric, and the condition above means that parallel transport maps preserve the inner products on the fibers, i.e. they are isometries. In this situation we call ∇ a **metric connection**.

EXAMPLE 20.11. Recall from §18.7 that $\mathrm{GL}(m, \mathbb{C})$ can be regarded as a subgroup of $\mathrm{GL}(2m, \mathbb{R})$ so that $\mathrm{GL}(m, \mathbb{C})$-structures on a real vector bundle $E \to M$ of rank $2m$ are equivalent to complex structures on E, which make all fibers into complex m-dimensional vector spaces. A connection on the *real* bundle E is then $\mathrm{GL}(m, \mathbb{C})$-compatible if and only if the parallel transport maps are complex linear for this complex structure, and ∇ is then called a **complex connection**. Note: if we had regarded E as a complex vector bundle in the first place, then choosing a connection ∇ on that bundle would have *automatically* meant that parallel transport is complex linear, so you may be wondering why it is useful to single out a special class of "complex connections" on a real vector bundle. One answer to this question is as follows: as we will soon see, every Riemannian manifold (M, g) has a canonical connection on its tangent bundle $TM \to M$, called the *Levi-Cività connection*, which is used for defining the standard Riemannian notions of parallel vector fields and curvature. In certain situations, especially if M is also a *symplectic* manifold, it is also useful to endow M with an almost complex structure (cf. §7.1.4), meaning a bundle map $J : TM \to TM$

that satisfies $J^2 \equiv -\mathbb{1}$ everywhere, thus making $TM \to M$ into a complex vector bundle. While complex connections on TM always exist, there is no guarantee in general that the Levi-Cività connection is one—this turns out to be true if and only if g and J satisfy a very rigid compatibility condition, guaranteeing that J is integrable (cf. Exercise 8.5), hence M in this situation is a complex manifold with a special type of Riemannian metric, called a *Kähler metric*.

We will prove in the next lecture that G-compatible connections always exist. The real strength of connection 1-forms is that they give an easy characterization of the G-compatibility condition. Recall from §18.1 that the *Lie algebra* of a Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$ is the tangent space $\mathfrak{g} := T_{\mathbb{1}} G \subset T_{\mathbb{1}} \mathrm{GL}(m, \mathbb{F}) = \mathbb{F}^{m \times m}$.

THEOREM 20.12. *If $E \to M$ is a vector bundle with a G-structure and ∇ is a connection on E, then ∇ is G-compatible if and only if for every G-compatible trivialization Φ_α, the corresponding connection 1-form takes values in the Lie algebra $\mathfrak{g} \subset \mathbb{F}^{m \times m}$ of G, i.e.*

$$A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{g}).$$

Before proving the theorem, it will be helpful to deal with a minor technical point. We have occasionally mentioned the notion of a **section along a path**, meaning the following: given a path $\gamma(t) \in M$, we associate to each t in its domain a vector

$$s(t) \in E_{\gamma(t)},$$

so strictly speaking, s is a section of the pullback bundle $\gamma^* E$. While s is not quite the same thing as a section of E, there is a straightforward way to define the covariant derivative of s with respect to the parameter t: at $t = 0$ it is

$$\nabla_t s|_{t=0} := \left. \frac{d}{dt}(P_\gamma^t)^{-1}(s(t)) \right|_{t=0} \in E_{\gamma(0)},$$

and one can similarly define $\nabla_t s(t_0) \in E_{\gamma(t_0)}$ for arbitrary t_0 in the domain of γ by reprarametrizing the path to $t \mapsto \gamma(t_0 + t)$, so that it passes through $\gamma(t_0)$ at time $t = 0$. The section s is then parallel along γ if and only if $\nabla_t s \equiv 0$. A slightly subtle distinction between this and the usual covariant derivative of a section of E is that s depends directly on t, not just on $\gamma(t) \in M$, so $\nabla_t s$ can be nonzero even if the path $\gamma(t)$ is stationary. For example, if γ is a constant path at a point $p \in M$, then parallel transport P_γ^t defines the identity map $T_p M \to T_p M$ for every t, and $\nabla_t s \in T_p M$ is then just the ordinary derivative of the path $s(t)$ in the vector space $T_p M$. On the other hand, if $\dot{\gamma}(0) \neq 0$, then γ is an embedding near $t = 0$ and thus traces out a smooth 1-dimensional submanifold of M; restricting γ to a suitably small neighborhood of 0, it is then easy to see that any section s along γ can be "extended" to a global section $\widehat{s} \in \Gamma(E)$ such that

$$\widehat{s}(\gamma(t)) = s(t) \qquad \text{and} \qquad \nabla_{\dot{\gamma}(t)}\widehat{s} = \nabla_t s(t) \qquad \text{for all } t.$$

(Indeed, first write down the extension $\widehat{s}$ on a neighborhood of $\gamma(0)$ in a slice chart for the image of γ, then extend it arbitrarily to the rest of M.) In this situation, various useful things we've proven about $\nabla \widehat{s}$ will apply to $\nabla_t s$ as well: one is the formula $\nabla_X \widehat{s} = K(T\widehat{s}(X))$, which becomes

$$(20.9) \qquad \nabla_t s(t) = K(\dot{s}(t)),$$

where we are regarding $s(t) = \widehat{s}(\gamma(t))$ as a smooth path in the total space E, whose derivative is thus $\dot{s}(t) = T\widehat{s}(\dot{\gamma}(t))$. Another is the coordinate formula (20.7) for $\nabla_X \widehat{s}$ in terms of a connection 1-form: assuming $\gamma(t)$ lies in the domain $\mathcal{U}_\alpha$ of a trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ and writing $\Phi_\alpha(s(t)) = (\gamma(t), s_\alpha(t))$, we obtain

$$(20.10) \qquad (\nabla_t s)_\alpha(t) = \dot{s}_\alpha(t) + A_\alpha(\dot{\gamma}(t))s_\alpha(t) \in \mathbb{F}^m.$$

An easy continuity argument now shows that (20.9) and (20.10) are not only valid under the condition $\dot{\gamma} \neq 0$: they are valid for *all* smooth paths γ, since a path with $\dot{\gamma}(t_0) = 0$ at some point t_0 admits arbitrarily small perturbations to one with $\dot{\gamma}(t_0) \neq 0$, and the section s can be perturbed along with it.

PROOF OF THEOREM 20.12. Suppose $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ is a local trivialization and γ is a path in $\mathcal{U}_\alpha$ with $p := \gamma(0)$ and $X := \dot{\gamma}(0)$. Since parallel transport maps are linear, there exists a unique function $g(t) \in \mathrm{GL}(m, \mathbb{F})$ with $g(0) = \mathbb{1}$ such that for any parallel section $s(t) \in E_{\gamma(t)}$ along γ, the local representative $s_\alpha(t) \in \mathbb{F}^m$ defined by $\Phi_\alpha(s(t)) = (\gamma(t), s_\alpha(t))$ satisfies

$$s_\alpha(t) = g(t)s_\alpha(0).$$

By (20.10), $\nabla_t s \equiv 0$ implies that $g(t)$ is the unique solution with $g(0) = \mathbb{1}$ to the linear ODE

$$\dot{g}(t) = -A_\alpha(\dot{\gamma}(t))g(t).$$

It suffices then to show that g takes values in the subgroup G (implying ∇ is G-compatible over $\mathcal{U}_\alpha$) if and only if A_α takes values in its Lie algebra $\mathfrak{g}$. Assuming the former, we can plug $t = 0$ into the above equation to conclude $A_\alpha(X) = -\dot{g}(0) \in \mathfrak{g}$, which completes the proof that $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{g})$ since $p \in \mathcal{U}_\alpha$ and $X \in T_pM$ were chosen arbitrarily. The converse follows from Exercise 20.14 below. $\qquad\square$

EXERCISE 20.13. A smooth **time-dependent vector field** on a manifold M is a family of vector fields $\{X_t \in \mathfrak{X}(M)\}_{t \in I}$ parametrized by an interval $I \subset \mathbb{R}$ such that the map $I \times M \to TM :$ $(t, p) \mapsto X_t(p)$ is smooth. A path $\gamma(t) \in M$ is called an **orbit** or **flow line** of the time-dependent vector field $\{X_t\}_{t \in I}$ if it satisfies $\dot{\gamma}(t) = X_t(\gamma(t))$ for every t. One can develop the theory of flows for time-dependent vector fields analogously to the time-independent case, defining in particular a smooth map φ_X^t on suitable open subsets of M such that $\gamma(t) := \varphi_X^t(p)$ is the unique orbit of $\{X_t\}_{t \in I}$ satisfying $\gamma(0) = p$. If you prefer not to redo work that has already been done, you can instead do this:

(a) Given a time-dependent vector field $\{X_t\}_{t \in I}$ on M, define a time-independent vector field $Y \in \mathfrak{X}(I \times M)$ by

$$Y(t, p) := (1, X_t(p)) \in \mathbb{R} \times T_pM = T_tI \times T_pM = T_{(t,p)}(I \times Y).$$

Use the flow of Y to deduce everything you might possibly want to know about the flow of $\{X_t\}$, e.g. that φ_X^t exists and is unique and is a diffeomorphism $M \to M$ for every $t \in \mathbb{R}$ if M is compact.
Caution: Do not try to prove the relations $\varphi_X^{s+t} = \varphi_X^s \circ \varphi_X^t$ or $\varphi_X^{-t} = (\varphi_X^t)^{-1}$, which are valid in general only for time-independent vector fields.

With these basics understood, the following observation will be helpful for Exercise 20.14 below:

(b) Suppose $N \subset M$ is a smooth submanifold and $\{X_t\}_{t \in I}$ is a time-dependent vector field on M such that $X_t(p) \in T_pN$ for every $p \in N$ and $t \in I$. Show that every flow line of $\{X_t\}$ is either contained in N or disjoint from it.

EXERCISE 20.14. For any Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$ and a smooth path of matrices $\mathbf{A}(t) \in \mathfrak{g} = T_{\mathbb{1}}G$, show that the unique solution $\boldsymbol{\Phi}(t) \in \mathbb{F}^{m \times m}$ to the initial value problem

$$\begin{cases} \dot{\boldsymbol{\Phi}}(t) &= \mathbf{A}(t)\boldsymbol{\Phi}(t), \\ \boldsymbol{\Phi}(0) &= \mathbb{1} \end{cases}$$

satisfies $\boldsymbol{\Phi}(t) \in G$ for all t.
Hint: Show that for any $\mathbf{A} \in \mathfrak{g}$, $X(\mathbf{B}) := \mathbf{A}\mathbf{B} \in \mathbb{F}^{m \times m} = T_{\mathbf{B}} \mathrm{GL}(m, \mathbb{F})$ defines a smooth vector field on $\mathrm{GL}(m, \mathbb{F})$ that satisfies $X(\mathbf{B}) \in T_{\mathbf{B}}G$ for all $\mathbf{B} \in G$.

REMARK 20.15. The notion of a G-structure on a vector bundle makes sense for any subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$, i.e. the definition itself does not require the additional condition that $G \subset \mathrm{GL}(m, \mathbb{F})$ is a smooth submanifold. However, by applying Exercise 20.14, Theorem 20.12 makes crucial use of this assumption, along with the fact (used in Exercise 20.14) that the matrix multiplication map $\mathbb{F}^{m \times m} \times \mathbb{F}^{m \times m} \to \mathbb{F}^{m \times m} : (\mathbf{A}, \mathbf{B}) \mapsto \mathbf{AB}$ is smooth. In other words, while G-structures on bundles can be defined for arbitrary subgroups $G \subset \mathrm{GL}(m, \mathbb{F})$, making connections compatible with these structures requires the group G to be smooth.

21. Constructions of connections

21.1. A general existence result. Let's get this out of the way first:

THEOREM 21.1. *Every vector bundle $E \to M$ with a G-structure for some Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$ admits a G-compatible connection.*

PROOF. Choose an open covering $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of M with a subordinate partition of unity $\{\varphi_\alpha : M \to [0,1]\}_{\alpha \in I}$ such that there are also local G-compatible trivializations $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ for each $\alpha \in I$. Each of these determines a trivial connection D^α on $E|_{\mathcal{U}_\alpha}$, which is G-compatible since its parallel transport maps look like the identity $\mathbb{F}^m \to \mathbb{F}^m$ in the trivialization Φ_α. We can then define a global connection ∇ on E by

$$\nabla_X s := \sum_{\alpha \in I} \varphi_\alpha(p) D_X^\alpha s, \qquad \text{for } p \in M, \, X \in T_p M,$$

where it is understood that at each point p, the sum contains only the finitely many terms for which $p \in \mathrm{supp}(\varphi_\alpha) \subset \mathcal{U}_\alpha$, and D^α is thus well defined near p. That the resulting operator $\nabla : \Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$ is a G-compatible connection follows now by writing it down in local trivializations. Indeed, each $p \in M$ has a neighborhood $\mathcal{U} \subset M$ that is contained in all of the sets $\mathcal{U}_\alpha$ for which $p \in \mathrm{supp}(\varphi_\alpha)$, and choosing any G-compatible trivialization of $E|_\mathcal{U}$ identifies each of the relevant operators D^α with an operator of the form $d_X^\alpha f := df(X) + B_\alpha(X)f$ on functions $f \in C^\infty(\mathcal{U}, \mathbb{F}^m)$, where $B_\alpha \in \Omega^1(\mathcal{U}, \mathfrak{g})$ since D^α is G-compatible. In this same trivialization, ∇ then becomes

$$d_X f = \sum_\alpha \left[\varphi_\alpha \, df(X) + \varphi_\alpha B_\alpha(X) f \right] = df(X) + B(X)f, \quad \text{where} \quad B := \sum_\alpha \varphi_\alpha B_\alpha \in \Omega(\mathcal{U}, \mathfrak{g}),$$

and is thus G-compatible by Theorem 20.12. $\qquad \square$

It is important to understand that Theorem 21.1 says nothing about uniqueness, and indeed, connections are in general neither unique nor canonical: in the case $G = \mathrm{GL}(m, \mathbb{F})$ for instance, one can produce an infinite-dimensional family of connections by choosing any specific connection ∇ and defining other connections by $\nabla + A$ for arbitrary bundle maps $A : E \to \mathrm{Hom}(TM, E)$. (Something similar is true for G-compatible connections with arbitrary Lie subgroups $G \subset \mathrm{GL}(m, \mathbb{F})$ if one considers only bundle maps $E \to \mathrm{Hom}(TM, E)$ that preserve the relevant structure—the space of such bundle maps is typically still infinite dimensional.) The major exception is the tangent bundle $TM \to M$ of a Riemannian or pseudo-Riemannian manifold: this bundle has structure group $\mathrm{O}(k, \ell)$ determined by the signature (k, ℓ) of its bundle metric, and while there is an infinite-dimensional family of $\mathrm{O}(k, \ell)$-compatible connections on TM, we will see in the next lecture that a canonical one can be singled out, due to the fact that $TM \to M$ is not just any vector bundle but specifically a *tangent* bundle.

21.2. Pullbacks. The next two sections will be concerned with the following question: given a finite collection of vector bundles $E^1, \ldots, E^m$ with connections and a natural operation that produces a new bundle E out of $E^1, \ldots, E^m$, how do the connections on $E^1, \ldots, E^m$ determine one on E? It will usually be obvious how the connection on E should be defined—only a little bit of effort is then required to check that the result really is a connection.

We start with pullbacks: suppose $E \to M$ is a smooth vector bundle, N is a manifold and $f : N \to M$ is a smooth map. A section

$$s \in \Gamma(f^*E)$$

of the pullback bundle $f^*E \to N$ associates to each $p \in N$ a vector $s(p) \in E_{f(p)}$, and is therefore sometimes called a **section of E along f**. If a connection ∇ on $E \to M$ with parallel transport maps $P_\gamma^t : E_{\gamma(0)} \to E_{\gamma(t)}$ is given, then there is an obvious way to define parallel transport maps $P_\gamma^t : (f^*E)_{\gamma(0)} \to (f^*E)_{\gamma(t)}$ for $f^*E \to N$ along a path γ in N, namely

$$(21.1) \qquad P_\gamma^t := P_{f \circ \gamma}^t : (f^*E)_{\gamma(0)} = E_{f(\gamma(0))} \to E_{f(\gamma(t))} = (f^*E)_{\gamma(t)}.$$

To see that this really defines a connection on $f^*E \to N$, let us translate (21.1) into a definition of a horizontal subbundle. Confusion can sometimes arise from the fact that fibers of f^*E are also fibers of E, so it will be helpful to distinguish them by adopting the following slightly verbose notation: elements of E can be written as pairs

$$(p, v) \in E, \qquad \text{for } p \in M, \, v \in E_p,$$

while elements of f^*E are written similarly as

$$(p, v) \in f^*E, \qquad \text{for } p \in N, \, v \in E_{f(p)} = (f^*E)_p.$$

The canonical smooth linear bundle map $f^*E \to E$ covering $f : N \to M$ then takes the form

$$\Psi : f^*E \to E : (p, v) \mapsto (f(p), v).$$

Equation (21.1) can now be interpreted as saying that a section $s(t) \in (f^*E)_{\gamma(t)}$ of f^*E along a path γ in N is parallel if and only if the section $\Psi \circ s(t) \in E_{f(\gamma(t))}$ of E along $f \circ \gamma$ is parallel. Differentiating this relation with respect to t gives a corresponding relation between horizontal subbundles: $\dot{s}(t) \in T_{s(t)}(f^*E)$ should be horizontal if and only if $\partial_t(\Psi \circ s)(t) = T\Psi(\dot{s}(t)) \in T_{\Psi(s(t))}E$ is horizontal, so that $H(f^*E) \subset T(f^*E)$ must be defined by

$$H_{(p,v)}(f^*E) := (T\Psi)^{-1}\left(H_{(f(p),v)}E\right) \subset T_{(p,v)}(f^*E).$$

To see that this really does define a complement to the vertical subbundle $V_{(p,v)}(f^*E)$, notice that since Ψ defines isomorphisms between fibers of f^*E and fibers of E, its derivative $T\Psi$ defines isomorphisms between the corresponding vertical subspaces. The condition $T(f^*E) = V(f^*E) \oplus H(f^*E)$ then follows from $TE = VE \oplus HE$ via a simple linear-algebraic exercise:

EXERCISE 21.2. Suppose X, X' are vector spaces, $V \subset X$ and $V', H' \subset X'$ are linear subspaces such that $X' = V' \oplus H'$, and $A : X \to X'$ is a linear map that restricts to $V \subset X$ as an isomorphism onto V'. Show that the subspace $H := A^{-1}(H') \subset X$ is then complementary to V, i.e. $X = V \oplus H$.

Having shown that there is a well-defined horizontal subbundle $H(f^*E) \subset T(f^*E)$ corresponding to the parallel transport maps in (21.1), it follows from Lemma 19.3 that $H(f^*E)$ is a connection on $f^*E \to N$ in the sense of Definition 19.4, as the parallel transport maps are manifestly linear. It is also clear from this definition that if the connection on E is compatible with some structure group G on E, then the pullback connection is compatible with the induced G-structure on f^*E.

EXAMPLE 21.3. For a smooth path $\gamma : I \to M$ defined on an open interval $I \subset \mathbb{R}$, a section s of $E \to M$ along γ is the same thing as a section of the pullback bundle $\gamma^*E \to I$. Let ∂_t denote the standard basis vector on $T_t I = \mathbb{R}$ for each $t \in I$. Now if ∇ is a connection on $E \to M$ and we equip $\gamma^*E \to I$ with the resulting pullback connection, writing the covariant derivative of s in terms of parallel transport gives

$$\nabla_{\partial_t} s|_{t=0} = \left. \frac{d}{dt} (P_\gamma^t)^{-1} \left(s(t) \right) \right|_{t=0}.$$

This is of course exactly the same thing as what we have previously denoted by $\nabla_t s(0)$; in other words, the covariant derivative with respect to t of a section of E along a path $\gamma(t) \in M$ is the same thing as the covariant derivative (using the pullback connection) of the corresponding section of γ^*E in the direction of the canonical unit vector field on the interval. This should not be surprising—if it had not been true, we would have concluded that we have the wrong definition of the pullback connection and then searched for a different one.

EXAMPLE 21.4. The following generalization of Example 21.3 is sometimes useful for computations: consider an open subset $\mathcal{V} \subset \mathbb{R}^d$ and a smooth map $f : \mathcal{V} \to M$. A section $s \in \Gamma(f^*E)$ of $E \to M$ along f then assigns to each tuple $(t^1, \ldots, t^d) \in \mathcal{V}$ a vector $s(t^1, \ldots, t^d) \in E_{f(t^1,\ldots,t^d)}$, and we can define a covariant derivative with respect to each of the variables $t^1, \ldots, t^d$,

$$\nabla_i s(t^1, \ldots, t^d) \in E_{f(t^1,\ldots,t^d)}, \qquad i = 1, \ldots, d,$$

which literally means the covariant derivative (via the pullback connection) of $s \in \Gamma(f^*E)$ with respect to the standard basis vector $\partial_i \in T_{(t^1,\ldots,t^d)} \mathcal{V} = \mathbb{R}^d$. This makes $\nabla_i s$ another section of f^*E, so it can be differentiated again, defining iterated covariant derivatives $\nabla_i \nabla_j s$, $\nabla_i \nabla_j \nabla_k s$ and so forth. For example, the partial derivatives of $f : \mathcal{V} \to M$ as defined in §4.1 are vector fields along f,

$$\partial_i f \in \Gamma(f^*TM), \qquad i = 1, \ldots, d,$$

so if a connection on $TM \to M$ has been chosen, we can now use the pullback connection on $f^*TM \to \mathcal{V}$ to define higher (covariant) derivatives of f in the form $\nabla_j \partial_i f$, $\nabla_k \nabla_j \partial_i f$ and so forth.

Let us derive a local coordinate formula for $\nabla_t s(t)$ when $s(t) \in E_{\gamma(t)}$ is a section along a path γ in M. Assume the image of γ lies in an open subset $\mathcal{U}_\alpha \subset M$ for which there is a local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$, and let $e_1, \ldots, e_m$ denote the corresponding frame for E over $\mathcal{U}_\alpha$. Assume also that $\mathcal{U}_\alpha \subset M$ admits coordinates $x^1, \ldots, x^n$, so that the Christoffel symbols Γ_{ib}^a characterizing ∇ on $E|_{\mathcal{U}_\alpha}$ are defined. Writing $s(t) = s^a(t) e_a(\gamma(t)) \in E_{\gamma(t)}$ and $\dot\gamma(t) = \dot\gamma^i(t) \partial_i \in T_{\gamma(t)} M$, we claim

$$(\nabla_t s)^a(t) = \dot{s}^a(t) + \Gamma_{ib}^a(\gamma(t)) \dot\gamma^i(t) s^b(t).$$

or more succinctly,

(21.2) $$(\nabla_t s)^a = \dot{s}^a + (\Gamma_{ib}^a \circ \gamma) \dot\gamma^i s^b.$$

It suffices to prove that this holds at $t = 0$, since the parametrization of γ can always be shifted, and in fact, we are also free to assume $\dot\gamma(0) \neq 0$, for the following reason. Unless $\dim M = 0$ (in which case the statement is trivial and there is nothing to prove), any path γ in M can be perturbed if necessary to ensure $\dot\gamma(0) \neq 0$, and the section s can be perturbed with it to a section along the perturbation of γ. If the relation (21.2) is satisfied after this perturbation, then it must have been satisfied beforehand as well, simply because both sides are continuous with respect to C^1-small perturbations of γ and s. With this understood, the condition $\dot\gamma(0) \neq 0$ allows us to assume after restricting to a suitably small neighborhood of 0 that γ is an embedding, so its image is a smooth 1-dimensional submanifold of M. One can then use a slice chart on M for this submanifold in order to construct a section $\eta \in \Gamma(E)$ such that $s(t) = \eta(\gamma(t))$ for all t, and it then follows from the

definition of the covariant derivative via parallel transport that $\nabla_t s(0) = \nabla_{\dot\gamma(0)}\eta$. Writing $\dot\gamma = \dot\gamma^i \partial_i$ and $\nabla_{\dot\gamma}\eta = \dot\gamma^i \nabla_i \eta$, (20.6) now implies

$$(\nabla_t s(0))^a = \dot\gamma^i(0) \left[\partial_i \eta^a(\gamma(0)) + \Gamma^a_{ib}(\gamma(0))\eta^b(\gamma(0)) \right] = \partial_t(\eta^a \circ \gamma)(0) + \Gamma^a_{ib}(\gamma(0))(\eta^b \circ \gamma)(0),$$

which justifies (21.2).

Recall from Exercise 20.8 that the connection 1-form $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$ can be derived from the Christoffel symbols by $A_\alpha(\partial_i)^a{}_b = \Gamma^a_{ib}$, so for $X = X^i \partial_i$, $A_\alpha(X)^a{}_b = \Gamma^a_{ib} X^i$. The expression $\Gamma^a_{ib}(\gamma(t))\dot\gamma(t)$ in (21.2) can therefore be reinterpreted as $A_\alpha(\dot\gamma(t))^a{}_b$, and the formula thus reproduces (20.10), i.e.

$$(\nabla_t s)_\alpha = \dot s_\alpha + A_\alpha(\dot\gamma)s_\alpha.$$

In the general situation where N is an arbitrary manifold with a smooth map $f : N \to M$ and $s \in \Gamma(f^*E)$, we can compute the covariant derivative $\nabla_X s$ in any direction $X \in TN$ by choosing a path γ in N with $\dot\gamma(0) = X$ and computing $\nabla_t(s \circ \gamma)(0)$, i.e. $\nabla_X s$ is the covariant derivative at $t = 0$ of $s \circ \gamma(t) \in (f^*E)_{\gamma(t)} = E_{f \circ \gamma(t)}$, which is a section of E along the path $f \circ \gamma$. Writing $s(p) = s^a(p)e_a(f(p))$ for $p \in f^{-1}(\mathcal{U}_\alpha)$, (21.2) thus implies

$$(21.3) \qquad (\nabla_X s)^a = ds^a(X) + \Gamma^a_{ib}(f(p))(f_* X)^i s^b(p), \qquad \text{for } p \in f^{-1}(\mathcal{U}_\alpha),\ X \in T_p M.$$

To rewrite this in terms of a connection 1-form, observe that the frame $e_1, \ldots, e_m$ corresponding to our trivialization Φ_α on $\mathcal{U}_\alpha$ determines a local frame for f^*E over the open set $f^{-1}(\mathcal{U}_\alpha) \subset N$, consisting of the sections $e_1 \circ f, \ldots, e_m \circ f$, and the local trivialization of f^*E corresponding to this is the one that we called

$$f^*\Phi_\alpha : (f^*E)|_{f^{-1}(\mathcal{U}_\alpha)} \to f^{-1}(\mathcal{U}_\alpha) \times \mathbb{F}^m$$

in §17.2. For $s \in \Gamma(f^*E)$, let $s_\alpha : f^{-1}(\mathcal{U}_\alpha) \to \mathbb{F}^m$ denote the local representation of s in this trivialization; then (21.3) becomes

$$(\nabla_X s)_\alpha = ds_\alpha(X) + A_\alpha(f_* X)s_\alpha(p), \qquad \text{for } p \in f^{-1}(\mathcal{U}_\alpha),\ X \in T_p M.$$

In other words, the connection 1-form for the pullback connection on f^*E with respect to the trivialization $f^*\Phi_\alpha$ is exactly what one would hope for: it is the pullback of A_α,

$$f^*A_\alpha \in \Omega^1(f^{-1}(\mathcal{U}_\alpha), \mathbb{F}^{m \times m}).$$

EXERCISE 21.5. Any section $s \in \Gamma(E)$ gives rise to a section along $f : N \to M$ in the form $s \circ f \in \Gamma(f^*E)$. Prove

$$\nabla_X(s \circ f) = \nabla_{f_* X} s \qquad \text{for all } X \in TN.$$

21.3. Algebraic operations. We shall now run through the essential items on the list of algebraic constructions of vector bundles in §17.4, and outline how to construct connections on each of them. Assume throughout that $E, F \to M$ are fixed vector bundles on which connections (both labelled ∇) have already been chosen.

21.3.1. *Direct sums.* The natural way to define parallel transport for $E \oplus F$ out of the parallel transport on E and F along a path γ is

$$P^t_\gamma(v, w) := (P^t_\gamma(v), P^t_\gamma(w)) \in E_{\gamma(t)} \times F_{\gamma(t)}, \qquad \text{for} \qquad (v, w) \in E_{\gamma(0)} \times F_{\gamma(0)}.$$

The notion of covariant differentiation on $\Gamma(E \oplus F)$ that arises from this definition is quite straightforward: under the obvious identification of $\Gamma(E \oplus F)$ with $\Gamma(E) \times \Gamma(F)$, we have

$$\nabla_X(\eta, \xi) = (\nabla_X \eta, \nabla_X \xi).$$

It is trivial to check that $\nabla : \Gamma(E \oplus F) \to \Gamma(\mathrm{Hom}(TM, E \oplus F))$ by this definition satisfies the required Leibniz rule and thus defines a connection on $E \oplus F$.

21.3.2. *The dual bundle.* The isomorphisms $P_\gamma^t : E_{\gamma(0)} \to E_{\gamma(t)}$ determine isomorphisms $P_\gamma^t :$ $E_{\gamma(0)}^* \to E_{\gamma(t)}^*$ by dualization, i.e. for $\lambda \in E_{\gamma(0)}^*$ and $v \in E_{\gamma(t)}$, we define

$$P_\gamma^t(\lambda)v := \lambda\left((P_\gamma^t)^{-1}v\right).$$

Equivalently, this means that if $\lambda(t) \in E_{\gamma(t)}^*$ and $v(t) \in E_{\gamma(t)}$ are parallel sections along γ, then the natural pairing between them is constant, so

(21.4) $$P_\gamma^t(\lambda)\left(P_\gamma^t(v)\right) = \lambda(v) \qquad \text{for all } t.$$

It follows for instance that if $e_1, \ldots, e_m$ is a frame for E near $\gamma(0)$ consisting of sections that are parallel along γ, then the sections in the dual frame $e_*^1, \ldots, e_*^m$ are also parallel along γ. From (21.4), one deduces that the covariant derivative satisfies a Leibniz rule for the pairing of E^* and E: for any sections $\lambda(t) \in E_{\gamma(t)}^*$ and $v(t) \in E_{\gamma(t)}$ along γ, we have

$$\left.\frac{d}{dt}\left[\lambda(t)\left(v(t)\right)\right]\right|_{t=0} = \left.\frac{d}{dt}(P_\gamma^t)^{-1}(\lambda(t))\left((P_\gamma^t)^{-1}(v(t))\right)\right|_{t=0} = \nabla_t\lambda(0)\left(v(0)\right) + \lambda(0)\left(\nabla_t v(0)\right),$$

which implies a statement about directional derivatives of the function $\lambda(v) \in C^\infty(M, \mathbb{F})$ for $\lambda \in$ $\Gamma(E^*)$ and $v \in \Gamma(E)$ with respect to a vector field $X \in \mathfrak{X}(M)$, namely

(21.5) $$\mathcal{L}_X\left[\lambda(v)\right] = \left(\nabla_X \lambda\right)(v) + \lambda\left(\nabla_X v\right).$$

This relation uniquely characterizes the operator $\nabla : \Gamma(E^*) \to \Gamma(\text{Hom}(TM, E^*))$, and can thus be used to give an easy proof that it really does define a connection:

EXERCISE 21.6. Deduce from (21.5) that the operator $\nabla : \Gamma(E^*) \to \Gamma(\text{Hom}(TM, E^*))$ satisfies the Leibniz rule required by Definition 20.4 and thus determines a connection on $E^* \to M$.

EXERCISE 21.7. In terms of the Christoffel symbols $\Gamma_{ib}^a = (\nabla_i e_b)^a$ defined with respect to a coordinate chart $(x^1, \ldots, x^n)$ and frame $e_1, \ldots, e_m$ for E over an open set $\mathcal{U} \subset M$, show that the induced connection on E^* acts on the dual frame $e_*^1, \ldots, e_*^m$ by

$$(\nabla_i e_*^b)_a = -\Gamma_{ia}^b,$$

and deduce the general coordinate formula

$$(\nabla_i \lambda)_a = \partial_i \lambda_a - \Gamma_{ia}^b \lambda_b$$

for $\lambda \in \Gamma(E^*)$.

A few extra comments about the special case $E = TM$ are in order. Here a chart $(x^1, \ldots, x^n) :$ $\mathcal{U} \to \mathbb{R}^n$ also defines a natural frame over $\mathcal{U}$, consisting of the coordinate vector fields $\partial_1, \ldots, \partial_n$, and the resulting Christoffel symbols consist of n^3 real-valued functions

$$\Gamma_{jk}^i : \mathcal{U} \to \mathbb{R}, \qquad i, j, k \in \{1, \ldots, n\}$$

given by

$$\Gamma_{jk}^i = (\nabla_j \partial_k)^i = dx^i\left(\nabla_j \partial_k\right).$$

In local coordinates, the covariant derivative of a vector field $X = X^i \partial_i$ is thus given (cf. Equation (20.6)) by

(21.6) $$(\nabla_j X)^i = \partial_j X^i + \Gamma_{jk}^i X^k.$$

For the induced connection on the dual bundle T^*M, we observe that the covariant derivative $\nabla\lambda$ of a 1-form $\lambda \in \Gamma(T^*M) = \Omega^1(M)$ is a section of $\text{Hom}(TM, T^*M)$ and can thus be identified in a natural way with a type $(0, 2)$ tensor field $\nabla\lambda \in \Gamma(T_2^0 M)$, i.e. we define

$$(\nabla\lambda)(X, Y) := (\nabla_X \lambda)(Y).$$

The components of $\nabla\lambda$ in local coordinates thus take the form $(\nabla\lambda)_{ij} = (\nabla\lambda)(\partial_i, \partial_j) := (\nabla_i\lambda)(\partial_j) =:$ $(\nabla_i\lambda)_j$, and Exercise 21.7 gives

$$(21.7) \qquad\qquad (\nabla\lambda)_{ij} = \partial_i\lambda_j - \Gamma_{ij}^k \lambda_k.$$

The tensor $\nabla\lambda \in \Gamma(T_2^0 M)$ is our newest and best answer to the question first posed in Lecture 8 concerning how one should go about defining the "derivative" of a tensor field, in this case specifically a 1-form. One of the answers we came up with in Lecture 8 was the exterior derivative $d\lambda \in \Omega^2(M)$, which is also a type $(0,2)$ tensor, but it carries less information: if you compare the local coordinate formulas we have for $\nabla\lambda$ and $d\lambda$, you'll notice that the individual partial derivatives $\partial_i\lambda_j$ cannot all be derived from $d\lambda$, but from $\nabla\lambda$ they can. In that sense, $\nabla\lambda$ is a better way of defining the derivative of λ, but it has the comparative disadvantage that it depends on a choice, since connections can always be chosen but are not unique.

21.3.3. *Tensor bundles.* If $A : V \to V'$ and $B : W \to W'$ are linear maps, there is a unique linear map

$$A \otimes B : V \otimes W \to V' \otimes W'$$

defined via the condition $(A \otimes B)(v \otimes w) = Av \otimes Bw$ for all $v \in V$ and $w \in W$. This determines a natural definition for parallel transport maps $P_\gamma^t : E_{\gamma(0)} \otimes F_{\gamma(0)} \to E_{\gamma(t)} \otimes F_{\gamma(t)}$, via the condition

$$P_\gamma^t(\eta \otimes \xi) := P_\gamma^t(\eta) \otimes P_\gamma^t(\xi) \qquad \text{for all } \eta \in E_{\gamma(0)}, \xi \in F_{\gamma(0)}.$$

In particular, the pointwise tensor product of any parallel sections of E and F along γ then becomes a parallel section of $E \otimes F$. As in §21.3.2, this gives rise to a Leibniz rule for the covariant derivative:

$$\nabla_t\left(\eta(t) \otimes \xi(t)\right)|_{t=0} = \frac{d}{dt}(P_\gamma^t)^{-1}(\eta(t)) \otimes (P_\gamma^t)^{-1}(\xi(t))\Big|_{t=0} = \nabla_t\eta(0) \otimes \xi(0) + \eta(0) \otimes \nabla_t\xi(0),$$

implying that for any $\eta \in \Gamma(E)$, $\xi \in \Gamma(F)$ and $X \in \mathfrak{X}(M)$,

$$(21.8) \qquad\qquad \nabla_X(\eta \otimes \xi) = \nabla_X\eta \otimes \xi + \eta \otimes \nabla_X\xi.$$

Once again, this uniquely characterizes the covariant derivative and can be used to prove that what we have defined really is a connection on $E \otimes F$:

EXERCISE 21.8. Deduce from (21.8) that the operator $\nabla : \Gamma(E \otimes F) \to \Gamma(\text{Hom}(TM, E \otimes F))$ defined above is a connection on $E \otimes F$.

By finite iterations, one can extract from the constructions in this section and §21.3.2 a definition of a connection on any of the tensor bundles $E_\ell^k \cong E^{\otimes k} \otimes (E^*)^{\otimes \ell}$ that is uniquely determined by any choice of connection on E. Moreover, it is uniquely characterized by the property that all Leibniz rules one can reasonably think of to write down are satisfied. For example, the induced connection on E_2^1 is related to the chosen connection on E and the induced connection on E^* by

$$\mathcal{L}_X\left(S(\lambda, \eta, \xi)\right) = (\nabla_X S)(\lambda, \eta, \xi) + S(\nabla_X\lambda, \eta, \xi) + S(\lambda, \nabla_X\eta, \xi) + S(\lambda, \eta, \nabla_X\xi)$$

for all $S \in \Gamma(E_2^1)$, $\lambda \in \Gamma(E^*)$ and $\eta, \xi \in \Gamma(E)$, and this relation uniquely determines ∇S.

In the case $E = TM$, the covariant derivative of a type (k, ℓ) tensor field $S \in \Gamma(T_\ell^k M)$ can be understood as a type $(k, \ell + 1)$ tensor field $\nabla S \in \Gamma(T_{\ell+1}^k M)$ by defining

$$(\nabla S)(\lambda^1, \ldots, \lambda^k, X_0, \ldots, X_\ell) := (\nabla_{X_0} S)(\lambda^1, \ldots, \lambda^k, X_1, \ldots, X_\ell).$$

In local coordinates, the components of ∇S thus take the form

$$(\nabla S)^{i_1 \ldots i_k}{}_{j_0 \ldots j_\ell} = (\nabla_{j_0} S)^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} = (\nabla_{\partial_{j_0}} S)\left(dx^{i_1}, \ldots, dx^{i_k}, \partial_{j_1}, \ldots, \partial_{j_\ell}\right).$$

EXERCISE 21.9. For a connection on TM with Christoffel symbols Γ^i_{jk} in some choice of local coordinates, show that the induced connection on $T^k_\ell M$ is given in the same coordinates by

$$(\nabla S)^{i_1 \ldots i_k}{}_{j_0 \ldots j_\ell} = \partial_{j_0} S^{i_1 \ldots i_k}{}_{j_1 \ldots j_\ell} + \Gamma^{i_1}_{j_0 a} S^{a i_2 \ldots i_k}{}_{j_1 \ldots j_\ell} + \ldots + \Gamma^{i_k}_{j_0 a} S^{i_1 \ldots i_{k-1} a}{}_{j_1 \ldots j_\ell}$$
$$- \Gamma^a_{j_0 j_1} S^{i_1 \ldots i_k}{}_{a j_2 \ldots j_\ell} - \ldots - \Gamma^a_{j_0 j_\ell} S^{i_1 \ldots i_k}{}_{j_1 \ldots j_{\ell-1} a}$$

for $S \in \Gamma(T^k_\ell S)$. Notice that this formula generalizes both (21.6) and (21.7).

21.3.4. *Bundles of linear maps.* Since $\mathrm{Hom}(E, F)$ is canonically isomorphic to $E^* \otimes F$, the constructions in §21.3.2 and §21.3.3 determine a natural connection on $\mathrm{Hom}(E, F)$.

EXERCISE 21.10. Show that the connection on $\mathrm{Hom}(E, F)$ is uniquely determined from the connections on E and F via the Leibniz rule

$$\nabla_X(A\eta) = (\nabla_X A)(\eta) + A(\nabla_X \eta) \qquad \text{for all } A \in \Gamma(\mathrm{Hom}(E, F)), \eta \in \Gamma(E) \text{ and } X \in \mathfrak{X}(M).$$

Hint: It suffices to consider bundle maps $A : E \to F$ of the form $A\eta = \lambda(\eta)\xi$ for fixed sections $\lambda \in \Gamma(E^*)$ and $\xi \in \Gamma(F)$. (Why?)

21.4. Tangent bundles, torsion and symmetry. For the rest of this lecture we specialize to the rank n real vector bundle

$$TM \to M$$

over a smooth n-manifold M. A connection on $TM \to M$ is also often referred to as a **connection on the manifold** M: it defines in particular the notion of parallel vector fields. Using the constructions in §21.2 and §21.3, it also determines connections on all of the tensor bundles $T^k_\ell M \to M$ and the pullback $f^*TM \to N$ for any smooth map $f : N \to M$. We will always assume when a connection ∇ on TM has been specified that the bundles $T^k_\ell M$ and f^*TM are endowed with the connections determined by ∇ in this way.

For covariant derivatives of vector fields, a natural question arises that would not make sense on an arbitrary vector bundle. Suppose $\mathcal{V} \subset \mathbb{R}^d$ is an open set and $f : \mathcal{V} \to M$ is a smooth map as discussed in Example 21.4, so that we can define partial derivatives $\partial_j f \in \Gamma(f^*TM)$ and then covariantly differentiate to define second derivatives $\nabla_i \partial_j f \in \Gamma(f^*TM)$. Do mixed partial derivatives in this sense commute, i.e. we do we have

$$\nabla_i \partial_j f = \nabla_j \partial_i f \qquad \text{for all } i, j?$$

The question can easily be answered via a local coordinate computation: choose a chart $(\mathcal{U}, x)$ on M with coordinates $x = (x^1, \ldots, x^n)$ and, on the subset in $\mathcal{V}$ where f has image in $\mathcal{U}$, write $f^k := x^k \circ f$ for each $k = 1, \ldots, n$ so that $\partial_i f = \partial_i f^k \, \partial_k$. Applying (21.3) then gives

$$(\nabla_i \partial_j f - \nabla_j \partial_i f)^k = \partial_i \partial_j f^k - \partial_j \partial_i f^k + \Gamma^k_{ab}(\partial_i f^a)(\partial_j f^b) - \Gamma^k_{ab}(\partial_j f^a)(\partial_i f^b)$$
$$= (\Gamma^k_{ab} - \Gamma^k_{ba})(\partial_i f^a)(\partial_j f^b).$$

The only way to make sure this vanishes for arbitrary maps $f : \mathcal{V} \to M$ is if the Christoffel symbols satisfy the relation

$$\Gamma^k_{ab} = \Gamma^k_{ba} \qquad \text{for all } k, a, b \in \{1, \ldots, n\}.$$

There is no reason why an arbitrary connection on $TM \to M$ should satisfy this; in fact, on the domain of a single chart one can always define a connection whose Christoffel symbols are any desired set of n^3 functions, which need not be related to each other in any way. But differential geometers have a favorite trick for situations like this: when we see a quantity that doesn't always vanish even though we wish it would, we make it into a tensor.

EXERCISE 21.11. Given a connection ∇ on the manifold M, prove that the bilinear map $T : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ given by

$$T(X,Y) := \nabla_X Y - \nabla_Y X - [X,Y]$$

defines a type $(1,2)$ tensor field on M, whose components in any local coordinate system are given by

$$T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj}.$$

The tensor $T \in \Gamma(T^1_2 M)$ in Exercise 21.11 is called the **torsion** of the connection ∇, and ∇ is called **symmetric** if its torsion tensor vanishes. Note that since the Christoffel symbols Γ^i_{jk} are not the components of any globally-defined tensor field, it is at first glance far from obvious that $\Gamma^i_{jk} - \Gamma^i_{kj}$ should be. One can check using the transformation formula in Exercise 20.7 that the functions $\Gamma^i_{jk} - \Gamma^i_{kj}$ do indeed transform as a tensor, but this is tedious; we should be very grateful in Exercise 21.11 that we can instead use C^∞-linearity to write down a coordinate-invariant definition of the torsion tensor.

As soon as one knows that connections on $TM \to M$ exist and have a well-defined torsion tensor, it is not hard to see that *symmetric* connections also exist:

EXERCISE 21.12. Given any connection ∇ on M with torsion tensor $T \in \Gamma(T^1_2 M)$, show that $\widehat{\nabla}_X Y := \nabla_X Y - \frac{1}{2}T(X,Y)$ defines a *symmetric* connection on M.

A second proof that symmetric connections always exist will emerge in the next lecture when we discuss connections on Riemannian manifolds. Let us conclude for now by stating the most useful property of symmetric connections, which follows immediately from the calculations above:

PROPOSITION 21.13. *A connection ∇ on a manifold M is symmetric if and only if for every open set $\mathcal{V} \subset \mathbb{R}^d$ and smooth map $f : \mathcal{V} \to M$, the relation $\nabla_i \partial_j f \equiv \nabla_j \partial_i f$ holds for all $i,j \in \{1,\dots,d\}$.* $\qquad\square$

REMARK 21.14. Symmetry of a connection does *not* imply that one can also exchange the order of the operators ∇_i and ∇_j in higher covariant derivatives. That is not true in general, and we will come back to this subject when we discuss curvature.

22. Pseudo-Riemannian manifolds and geodesics

22.1. Geodesics and the exponential map. We will be assuming in most of this lecture that M is a Riemannian or pseudo-Riemannian manifold, but the general definition of a geodesic does not actually require so much structure; it only requires a connection ∇ on M, by which we mean a connection on the tangent bundle $TM \to M$. The defining property of a straight line $\gamma : (a,b) \to \mathbb{R}^n$ in Euclidean space is that its velocity $\dot{\gamma}(t) \in \mathbb{R}^n$ is constant. The obvious analogue of this condition for a path $\gamma : (a,b) \to M$ is that its velocity $\dot{\gamma} \in \Gamma(\gamma^*TM)$ should be *parallel* along γ, leading to the **geodesic equation**

$$\nabla_t \dot{\gamma} \equiv 0.$$

Paths $\gamma : (a,b) \to M$ that satisfy this condition are called **geodesics** (*Geodäten* or *geodätische Linien*) in M. It should be emphasized that the notion of a geodesic depends on the choice of connection, though we will see shortly that if a pseudo-Riemannian metric g is given, then the connection can be chosen canonically, so that the notion of a geodesic depends only on g.

When $\gamma : (a,b) \to M$ passes through the domain $\mathcal{U} \subset M$ of a chart $(x^1,\dots,x^n)$, its coordinates define a path $(\gamma^1(t),\dots,\gamma^n(t))$ in $\mathbb{R}^n$, and using (21.2), the geodesic equation then becomes a system of n second-order nonlinear differential equations for the functions $\gamma^i(t) \in \mathbb{R}$, namely

$$\ddot{\gamma}^i(t) + \Gamma^i_{jk}(\gamma(t))\dot{\gamma}^j(t)\dot{\gamma}^k(t) = 0 \qquad \text{for all } t,$$

or in succinct form,

$$(22.1) \qquad \ddot{\gamma}^i + \Gamma^i_{jk}\dot{\gamma}^j\dot{\gamma}^k \equiv 0.$$

As a second-order system on an open set in $\mathbb{R}^n$, (22.1) has a unique solution near any point $t = t_0$ with any given initial position $\gamma(t_0)$ and velocity $\dot{\gamma}(t_0)$. It follows that for every $p \in M$ and $X \in T_pM$, there exists a unique geodesic

$$(a, b) \to M : t \mapsto \gamma_X(t), \qquad \text{such that} \qquad \nabla_t \dot{\gamma}_X \equiv 0, \ \gamma_X(0) = p \text{ and } \dot{\gamma}_X(0) = X.$$

Here $-\infty \leqslant a < 0 < b \leqslant \infty$, and (a, b) is assumed to be the largest possible interval on which the solution γ_X exists. The point $\gamma_X(t) \in M$ is defined for all pairs (t, X) belonging to some open subset of $\mathbb{R} \times TM$, and it depends smoothly on both t and X; this follows from the standard theorem about smooth dependence on initial conditions for ODEs.

EXERCISE 22.1. Show that for any geodesic $\gamma : (a, b) \to M$ and any constant $c \in \mathbb{R}$, the path defined by $\hat{\gamma}(t) := \gamma(ct)$ on the appropriate interval is also a geodesic.

An interesting consequence of Exercise 22.1 is that the point $\gamma_X(t)$ doesn't just depend smoothly on t and X, it depends in fact only on their product $tX \in TM$. Indeed, consider a pair of colinear vectors $X_1, X_2 \in T_pM$ with $X_2 = cX_1$ for some $c \in \mathbb{R}$. If γ_1 and γ_2 are the unique geodesics through p with $\dot{\gamma}_1(0) = X_1$ and $\dot{\gamma}_2(0) = X_2$, then Exercise 22.1 implies $\gamma_2(t) = \gamma_1(ct)$ for all t, hence $\gamma_1(t_1) = \gamma_2(t_2)$ whenever $t_1 = ct_2$, which means $t_2X_2 = ct_2X_1 = t_1X_1$. To put this observation in its most useful form, we define the open set

$$\mathcal{O} := \{X \in TM \mid 1 \text{ is in the domain of } \gamma_X\}$$

and the smooth function

$$\exp : \mathcal{O} \to M : X \mapsto \gamma_X(1).$$

Note that the domain of exp contains the zero-section of TM since geodesics with $\dot{\gamma}(0) = 0$ can be defined for all time (they are constant). The discussion above proves:

PROPOSITION 22.2. *For each* $p \in M$ *and* $X \in T_pM$, $I_X := \{t \in \mathbb{R} \mid tX \in \mathcal{O}\}$ *is an open interval containing 0, and* $\gamma : I_X \to M : t \mapsto \exp(tX)$ *is the maximal geodesic through* $\gamma(0) = p$ *with* $\dot{\gamma}(0) = X$. $\qquad\square$

We call $\exp : TM \supset \mathcal{O} \to M$ the **exponential map**. For a point $p \in M$, its restriction to an individual tangent space $\mathcal{O}_p := \mathcal{O} \cap T_pM$ is sometimes denoted by

$$\exp_p : \mathcal{O}_p \to M,$$

and it satisfies $\exp_p(0) = p$ since the unique geodesic γ with $\dot{\gamma}(0) = 0 \in T_pM$ is constant. Moreover, Proposition 22.2 implies that the derivative of $\exp_p$ at $0 \in T_pM$ is the identity map,

$$T_0(\exp_p) : T_0(T_pM) = T_pM \to T_pM : X \mapsto \frac{d}{dt}\exp(tX)\Big|_{t=0} = X,$$

so that by the inverse function theorem, $\exp_p$ maps a sufficiently small neighborhood of 0 in T_pM diffeomorphically onto a neighborhood of p in M.

The terminology "exponential map" can be motivated in part by the following example: if S^1 is regarded as the unit circle in $\mathbb{C} = \mathbb{R}^2$, then there is a natural connection on S^1 for which the geodesics passing through 1 at time $t = 0$ are precisely the paths of the form $\gamma(t) = e^{i\theta t} =: \exp(ti\theta)$ for $\theta \in \mathbb{R}$, which satisfy $\dot{\gamma}(0) = i\theta \in i\mathbb{R} = T_1S^1$.

22.2. The Levi-Cività connection. For the rest of this lecture, assume (M, g) is a pseudo-Riemannian manifold, which means the tangent bundle $TM \to M$ is equipped with a (possibly indefinite) bundle metric and thus has structure group $O(k, \ell)$ for some integers $k, \ell \geqslant 0$ with $k + \ell = n = \dim M$. If $(k, \ell) = (n, 0)$, then g is positive and (M, g) is called a Riemannian manifold (without the "pseudo-"). We will sometimes need to assume this, but most of what we do in the present lecture will be equally valid for indefinite metrics. We will often use inner product notation as a synonym for g,

$$\langle \, , \, \rangle := g(\cdot, \cdot),$$

reserving the notation $g \in \Gamma(T_2^0 M)$ mainly for situations where its role as a tensor field needs to be emphasized. The bundle metric gives $TM \to M$ structure group $O(k, \ell)$, so we can speak of $O(k, \ell)$-compatible connections, also known as *metric* connections.

EXERCISE 22.3. Show that the following conditions for a connection ∇ on a vector bundle $E \to M$ with bundle metric $g = \langle \, , \, \rangle \in \Gamma(E_2^0)$ are equivalent:

  (i) ∇ is a metric connection;
  (ii) For all $X \in \mathfrak{X}(M)$ and $\eta, \xi \in \Gamma(E)$, $\mathcal{L}_X \langle \eta, \xi \rangle = \langle \nabla_X \eta, \xi \rangle + \langle \eta, \nabla_X \xi \rangle$;
  (iii) The induced connection on E_2^0 satisfies $\nabla g \equiv 0$.

A connection on a real vector bundle with bundle metric g is said to be **compatible with** g if it is a metric connection, or equivalently, if it satisfies any of the conditions in Exercise 22.3.

The next result is sometimes called the *fundamental theorem of (pseudo-)Riemannian geometry*, because almost every other result in the subject depends on it. It is independent of our previous proof that connections on vector bundles always exist, so if you combine it with the theorem that every manifold admits a Riemannian metric, it implies a second proof of the fact that every manifold admits a symmetric connection.

THEOREM 22.4. *For any pseudo-Riemannian manifold (M, g), there exists a unique connection on $TM \to M$ that is symmetric and compatible with g.*

PROOF. We first show uniqueness: assuming ∇ is such a connection, Exercise 22.3 implies that for any vector fields X, Y and Z, we have the three relations

$$\mathcal{L}_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle,$$
$$\mathcal{L}_Y \langle Z, X \rangle = \langle \nabla_Y Z, X \rangle + \langle Z, \nabla_Y X \rangle,$$
$$\mathcal{L}_Z \langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle.$$

Adding the first two, subtracting the third and using the assumption $T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] \equiv 0$, we find

$$\mathcal{L}_X \langle Y, Z \rangle + \mathcal{L}_Y \langle Z, X \rangle - \mathcal{L}_Z \langle X, Y \rangle$$
$$= \langle \nabla_X Y + \nabla_Y X, Z \rangle + \langle Y, \nabla_X Z - \nabla_Z X \rangle + \langle X, \nabla_Y Z - \nabla_Z Y \rangle$$
$$= \langle 2\nabla_X Y, Z \rangle - \langle [X, Y], Z \rangle + \langle [X, Z], Y \rangle + \langle [Y, Z], X \rangle,$$

thus

$$(22.2) \quad \langle \nabla_X Y, Z \rangle = \frac{1}{2} \Big(\mathcal{L}_X \langle Y, Z \rangle + \mathcal{L}_Y \langle Z, X \rangle - \mathcal{L}_Z \langle X, Y \rangle$$
$$+ \langle [X, Y], Z \rangle - \langle [X, Z], Y \rangle - \langle [Y, Z], X \rangle \Big).$$

A straightforward (though slightly tedious) calculation shows that the right hand side of this expression is C^∞-linear with respect to X and Z. It therefore associates to every $Y \in \mathfrak{X}(M)$ a tensor field $S_Y \in \Gamma(T_2^0 M)$ such that (22.2) can be rewritten in succinct form as

$$\langle \nabla_X Y, \cdot \rangle = S_Y(X, \cdot).$$

This uniquely determines $\nabla_X Y$, since by the nondegeneracy of g, the map $T_p M \to T_p^* M : Z \mapsto g_p(Z, \cdot)$ is an isomorphism for every $p \in M$, implying that $\mathfrak{X}(M) \to \Omega^1(M) : Z \mapsto \langle Z, \cdot \rangle$ is also an isomorphism. If one now *defines* $\nabla_X Y$ in terms of $S_Y(X, \cdot)$ via this relation for every $Y \in \mathfrak{X}(M)$ and $X \in TM$, one can check that it satisfies the required Leibniz rule and is thus a connection on M, in addition to being symmetric and compatible with g. $\qquad\square$

The connection in Theorem 22.4 is called the **Levi-Cività connection** on (M, g). Whenever we discuss pseudo-Riemannian manifolds from now on, we will always use the Levi-Cività connection for computations on its tangent bundle, along with the various induced connections that it determines on associated bundles such as $T^* M$ and $T_\ell^k M$. The first hint that this might be the "right" thing to do comes from the fact that the Levi-Cività connection does not depend on any choices other than the metric; this is the first time we have seen a connection that is not some kind of arbitrary choice. The real justification for using this in preference to any other connection will come from the multitude of geometrically-motivated theorems that we can use it to prove, e.g. the fact (to be proved in the next lecture) that for the Levi-Cività connection on a Riemannian manifold, geodesics are not only the natural generalization of the notion of a "straight line" but also define *shortest* paths between nearby points.

22.3. Musical isomorphisms and coordinates. We would like to write down an explicit local coordinate formula for the Levi-Cività connection. The following algebraic remarks serve as preparation for this.

On a real vector bundle $E \to M$, any bundle metric $\langle \, , \, \rangle$ determines a natural smooth linear bundle map

$$\flat : E \to E^* : v \mapsto v_\flat := \langle v, \cdot \rangle.$$

The nondegeneracy of $\langle \, , \, \rangle$ implies that $\flat$ is injective on every fiber, and it is therefore a bundle isomorphism; note that this is true for any *nondegenerate* bilinear form, so in particular $\langle \, , \, \rangle$ may be an indefinite bundle metric, it need not be positive. The inverse of $\flat$ is denoted by

$$\sharp : E^* \to E : v \mapsto v^\sharp,$$

and notation motivates terminology: we call $\flat$ and $\sharp$ the **musical isomorphisms** determined by $\langle \, , \, \rangle$.

As an isomorphism, $\flat$ can be used to transfer all data from E to E^*, e.g. it gives a natural definition of a bundle metric on E^*, namely

$$(22.3) \qquad\qquad \langle \lambda, \mu \rangle := \langle \lambda^\sharp, \mu^\sharp \rangle \qquad \text{for} \qquad \lambda, \mu \in E^* \oplus E^*.$$

EXERCISE 22.5. Assume ∇ is a metric connection on E. Show:
   (a) For the induced connections on $\mathrm{Hom}(E, E^*)$ and $\mathrm{Hom}(E^*, E)$, $\nabla(\flat) \equiv 0$ and $\nabla(\sharp) \equiv 0$.
   (b) The induced connection on E^* is compatible with the bundle metric (22.3).

Choose a frame $e_1, \ldots, e_m$ for E over some open set $\mathcal{U} \subset M$, let $e_*^1, \ldots, e_*^m$ denote the dual frame, and denote the resulting components of the bundle metrics on E and E^* by

$$g_{ij} := \langle e_i, e_j \rangle, \qquad g^{ij} := \langle e_*^i, e_*^j \rangle.$$

For $v = v^i e_i, w = w^i e_i \in E_p$ and $\lambda = \lambda_i e_*^i, \mu = \mu_i e_*^i \in E_p^*$ at a point $p \in \mathcal{U}$, one then has

$$(22.4) \qquad\qquad \langle v, w \rangle = g_{ij} v^i w^j, \qquad \langle \lambda, \mu \rangle = g^{ij} \lambda_i \mu_j.$$

The convention for the musical isomorphisms is that for $v \in E_p$ or $\lambda \in E_p^*$, one writes the components of $v_\flat$ and $\lambda^\sharp$ with the *same* symbol but with the index raised or lowered, thus

$$\eta = \eta^i e_i \quad \Leftrightarrow \quad \eta_\flat = \eta_i e_*^i, \qquad \text{and} \qquad \lambda = \lambda_i e_*^i \quad \Leftrightarrow \quad \lambda^\sharp = \lambda^i e_i.$$

Philosophically, this means in some sense that we are considering vectors in E and dual vectors in E^* to be two distinct presentations of the same fundamental object. Since $\langle v, w \rangle = v_\flat(w) = w_\flat(v)$ and $\langle \lambda, \mu \rangle = \langle \lambda^\sharp, \mu^\sharp \rangle = \lambda(\mu^\sharp) = \mu(\lambda^\sharp)$, the bundle metrics on E and E^* can now be written in the appealing form

$$\langle v, w \rangle = v^i w_i = v_i w^i, \qquad \langle \lambda, \mu \rangle = \lambda_i \mu^i = \lambda^i \mu_i.$$

Comparing this with (22.4), you may notice that it implies explicit coordinate formulas for the maps $\flat$ and $\sharp$, namely

$$v_i = g_{ij} v^j, \qquad \text{and} \qquad \lambda^i = g^{ij} \lambda_j.$$

Since $\sharp = \flat^{-1}$, it follows that the m-by-m matrices with entries g_{ij} and g^{ij} are inverse to each other, i.e.

$$(22.5) \qquad\qquad\qquad\qquad g_{ij} g^{jk} = \delta_i^k.$$

One can always deduce the components g^{ij} from this fact once the g_{ij} are known.

REMARK 22.6. It was important throughout this discussion that E is a *real* vector bundle, not complex. Several details would need to modified if E were a complex bundle, starting with the observation that $\flat$ and $\sharp$ as we defined them are no longer bundle isomorphisms, as they are complex antilinear, not complex linear.

Specializing to the case where $E = TM$ for a pseudo-Riemannian manifold (M, g), we can define the musical isomorphisms $\flat : TM \to T^*M$ and $\sharp : T^*M \to TM$ as above, use them to define a bundle metric $\langle \ , \ \rangle$ on T^*M, then fix a chart $(\mathcal{U}, x)$ and write

$$g_{ij} = \langle \partial_i, \partial_j \rangle, \qquad g^{ij} = \langle dx^i, dx^j \rangle \qquad \text{on } \mathcal{U}.$$

Setting $X := \partial_i$, $Y := \partial_j$ and $Z := \partial_k$, $\nabla_X Y$ can be expressed in terms of the Christoffel symbols using (20.5), and we thus have

$$\langle \nabla_X Y, Z \rangle = \langle \nabla_i \partial_j, \partial_k \rangle = \langle \Gamma_{ij}^a \partial_a, \partial_k \rangle = g_{ak} \Gamma_{ij}^a.$$

If ∇ is the Levi-Cività connection, then (22.2) equates this with $\frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}\right)$, in which the Lie bracket terms from (22.2) do not appear since coordinate vector fields always commute with each other. Applying (22.5) now gives a formula for explicitly computing the Levi-Cività connection: its Christoffel symbols are

$$(22.6) \qquad\qquad\qquad \Gamma_{ij}^\ell = \frac{1}{2} g^{k\ell}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right).$$

EXAMPLE 22.7. Consider $\mathbb{R}^n$ with what we will henceforth call the **standard Euclidean metric**, meaning the Riemannian metric defined via the Euclidean inner product. The Levi-Cività connection is in this case exactly what you would expect: since the components $g_{ij} = \delta_{ij}$ of the metric are all constant, the Christoffel symbols computed via (22.6) all vanish identically, and ∇ is therefore the trivial connection on the trivial bundle $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$. Note that this is true for *any* choice of pseudo-Riemannian metric on $\mathbb{R}^n$ whose components are constant, including indefinite metrics such as the Minkowski metric of special relativity. The geodesic equation for paths $\gamma : (a, b) \to \mathbb{R}^n$ is thus $\ddot{\gamma} = 0$, and its solutions are straight lines with constant speed.

EXERCISE 22.8. The **Poincaré half-plane** $(\mathbb{H}, h)$ is the 2-manifold

$$\mathbb{H} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\} \subset \mathbb{R}^2$$

with Riemannian metric

$$h_{(x,y)}(X, Y) = \frac{1}{y^2} \langle X, Y \rangle_E \qquad \text{for } X, Y \in T_{(x,y)}\mathbb{H} = \mathbb{R}^2,$$

where $\langle \ , \ \rangle_E$ denotes the Euclidean inner product on $\mathbb{R}^2$. As we will later see, this is an example of a surface with constant negative curvature.

(a) Using the obvious global coordinates, derive the Christoffel symbols for the Levi-Cività connection on $(\mathbb{H}, h)$ and show that for this connection, the geodesic equation can be written as
$$\ddot{x} - \frac{2}{y}\dot{x}\dot{y} = 0, \qquad \ddot{y} + \frac{1}{y}(\dot{x}^2 - \dot{y}^2) = 0$$
for a smooth path $\gamma(t) = (x(t), y(t))$.

(b) Show that for any constants $x_0 \in \mathbb{R}$ and $r > 0$, the geodesic equation in part (a) has solutions of the form
$$\gamma(t) = (x_0, y(t)), \qquad \text{or} \qquad \gamma(t) = (x_0 + r\cos\theta(t), r\sin\theta(t))$$
for appropriately chosen functions $y(t) > 0$ and $\theta(t) \in (0, \pi)$.

(c) Show that any two points in $(\mathbb{H}, h)$ are connected by a unique geodesic segment $\gamma : [a, b] \to M$, and compute the length of this segment, meaning the integral
$$\ell_a^b(\gamma) := \int_a^b |\dot{\gamma}(t)| \, dt := \int_a^b \sqrt{\langle\dot{\gamma}(t), \dot{\gamma}(t)\rangle} \, dt.$$

22.4. Arc length and the energy functional. We shall now begin exploring the relationship between the geodesics of the Levi-Cività connection and the problem of finding paths of minimal length between fixed points. This uses some basic concepts from the *calculus of variations*, which deals with optimization problems on infinite dimensional spaces. Fix two points $p, q \in M$ and real numbers $a < b$. We denote by
$$C^\infty([a, b], M \, ; p, q)$$
the set of all smooth paths $\gamma : [a, b] \to M$ such that $\gamma(a) = p$ and $\gamma(b) = q$. Given a Riemannian metric $g = \langle \ , \ \rangle$, we denote
$$|X| := \sqrt{\langle X, X\rangle}$$
and define the **length functional** on $C^\infty([a, b], M \, ; p, q)$ by
$$\ell_a^b(\gamma) = \int_a^b |\dot{\gamma}(t)| \, dt.$$

A related functional is the **energy functional**,
$$E_a^b(\gamma) = \int_a^b \langle\dot{\gamma}(t), \dot{\gamma}(t)\rangle \, dt,$$

which also makes sense for an arbitrary pseudo-Riemannian metric, i.e. there is no need to assume g is positive. The geometric meaning of ℓ_a^b is clear: $\ell_a^b(\gamma)$ is the length of the path traced out by γ, as measured with respect to the Riemannian metric g. As such, it depends only on the image of γ, and is thus invariant under reparametrizations, i.e. for any diffeomorphism $\varphi : [a, b] \to [a', b']$ and smooth path $\gamma \in C^\infty([a', b'], M; p, q)$, we have
$$\ell_a^b(\gamma \circ \varphi) = \ell_{a'}^{b'}(\gamma).$$

It is less obvious what geometric meaning the energy functional may have, especially in the indefinite case, but we will find it convenient as a computational tool in order to understand the length functional better.

We wish to view $C^\infty([a, b], M \, ; p, q)$ informally as an infinite-dimensional manifold, and E_a^b and ℓ_a^b as "smooth functions" on this manifold which can be differentiated. There will be no need to define this in formal terms, because for the type of optimization problem we have in mind,

it always suffices to consider the values of each functional along paths in $C^\infty([a,b], M\,;p,q)$. In general, for a given functional

$$F : C^\infty([a,b], M\,;p,q) \to \mathbb{R},$$

the first goal of the calculus of variations is to find necessary conditions on a smooth path $\gamma \in C^\infty([a,b], M\,;p,q)$ so that $F(\gamma)$ may attain a minimal or maximal value among all paths $\gamma_\epsilon \in C^\infty([a,b], M\,;p,q)$ close to γ. This condition will typically take the form of a differential equation that γ must satisfy. To make this precise, we say that a **smooth 1-parameter family** of paths from p to q is a collection $\gamma_s \in C^\infty([a,b], M\,;p,q)$ for $s \in (-\epsilon, \epsilon)$ such that the map $(s,t) \mapsto \gamma_s(t)$ is smooth. Informally, we think of this as a smooth path in $C^\infty([a,b], M\,;p,q)$ through γ_0, and its "velocity vector" at $s = 0$ is then given by the partial derivatives $\partial_s \gamma_s(t)|_{s=0} \in T_{\gamma_0(t)} M$ for all t, which define a vector field along γ_0,

$$\eta := \partial_s \gamma_s|_{s=0} \in \Gamma(\gamma_0^* TM),$$

such that $\eta(a) = 0$ and $\eta(b) = 0$. We therefore think of the vector space

$$\{\eta \in \Gamma(\gamma^* TM) \mid \eta(a) = 0 \text{ and } \eta(b) = 0\}$$

as the "tangent space" to $C^\infty([a,b], M\,;p,q)$ at γ. It is now clear how one should define a "directional derivative" of F in a direction defined by a section of $\gamma^* TM$. This motivates the following definition, which generalizes the notion of a critical point of a real-valued function in finite dimensions.

DEFINITION 22.9. The path $\gamma \in C^\infty([a,b], M\,;p,q)$ is called **stationary** for the functional $F : C^\infty([a,b], M\,;p,q) \to \mathbb{R}$ if for every smooth 1-parameter family $\gamma_s \in C^\infty([a,b], M\,;p,q)$ with $\gamma_0 = \gamma$,

$$(22.7) \qquad\qquad \frac{d}{ds} F(\gamma_s)\bigg|_{s=0} = 0.$$

Note that for an arbitrary functional, it is not *a priori* clear that the derivatives in (22.7) will always exist. This is however true in many cases of interest, and in such a situation, it's easy to see that (22.7) is a necessary condition for F to attain an extremal value at γ.

PROPOSITION 22.10. *The energy functional E_a^b is stationary at γ if and only if γ satisfies the geodesic equation for the Levi-Cività connection.*

PROOF. Pick any smooth 1-parameter family $\gamma_s \in C^\infty([a,b], M\,;p,q)$ with $\gamma_0 = \gamma$ and denote $\eta = \partial_s \gamma_s|_{s=0} \in \Gamma(\gamma^* TM)$. In the following calculation, we regard $\partial_s \gamma_s(t)$ and $\dot\gamma_s(t) := \partial_t \gamma_s(t)$ as defining vector fields along the map $(s,t) \mapsto \gamma_s(t) \in M$, which can then be covariantly differentiated using the pullback connection. Differentiating under the integral sign and using the properties of the Levi-Cività connection, we have

$$\begin{aligned}
\frac{d}{ds} E_a^b(\gamma_s)\bigg|_{s=0} &= \int_a^b \frac{\partial}{\partial s} \langle \partial_t \gamma_s(t), \partial_t \gamma_s(t) \rangle \bigg|_{s=0} dt \\
&= \int_a^b \left(\langle \nabla_s \partial_t \gamma_s(t)|_{s=0}, \dot\gamma(t) \rangle + \langle \dot\gamma(t), \nabla_s \partial_t \gamma_s(t)|_{s=0} \rangle \right) dt \\
&= 2 \int_a^b \langle \dot\gamma(t), \nabla_t \, \partial_s \gamma_s(t)|_{s=0} \rangle \, dt = 2 \int_a^b \langle \dot\gamma(t), \nabla_t \eta(t) \rangle \, dt,
\end{aligned}$$

where in the last line we've used the symmetry of the connection to replace $\nabla_s \partial_t$ with $\nabla_t \partial_s$. We now perform a geometric version of integration by parts, using the fact that $\eta(t)$ vanishes at the

end points. It follows indeed from the fundamental theorem of calculus that

$$0 = \langle \dot{\gamma}(b), \eta(b) \rangle - \langle \dot{\gamma}(a), \eta(a) \rangle = \int_a^b \frac{d}{dt} \langle \dot{\gamma}(t), \eta(t) \rangle \, dt$$
$$= \int_a^b \langle \nabla_t \dot{\gamma}(t), \eta(t) \rangle \, dt + \int_a^b \langle \dot{\gamma}(t), \nabla_t \eta(t) \rangle \, dt,$$

thus

$$\frac{d}{ds} E_a^b(\gamma_s) \Big|_{s=0} = -2 \int_a^b \langle \nabla_t \dot{\gamma}(t), \eta(t) \rangle \, dt.$$

Since choosing arbitrary 1-parameter families γ_s leads to arbitrary sections $\eta \in \Gamma(\gamma^* TM)$ with $\eta(a) = 0$ and $\eta(b) = 0$, this expression will vanish for all such choices if and only if $\nabla_t \dot{\gamma} \equiv 0$. $\qquad \square$

To see what this tells us about the length functional, suppose now that the metric $\langle \, , \, \rangle$ is positive, so that $|X| = \sqrt{\langle X, X \rangle}$ can be defined and interpreted as the length of any tangent vector $X \in TM$. The **speed** of a path $\gamma : (a, b) \to M$ at time t is then the length of its velocity vector $\dot{\gamma}(t)$, and another easy observation about geodesics follows from the fact that ∇ is a metric connection: we have

$$\partial_t |\dot{\gamma}(t)|^2 = \partial_t \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle = 2 \langle \nabla_t \dot{\gamma}(t), \dot{\gamma}(t) \rangle,$$

so if the geodesic equation is satisfied, the speed $|\dot{\gamma}(t)|$ is constant. We claim: every immersed path $\gamma \in C^\infty([a, b], M \,; p, q)$ has a unique reparametrization $\beta = \gamma \circ \varphi \in C^\infty([a, b], M \,; p, q)$ that has constant speed. Indeed, to derive β, we can first figure out which constant $v := |\dot{\beta}(t)|$ needs to be: since $\ell_a^b(\gamma) = \ell_a^b(\beta) = \int_a^b v \, dt = v(b - a)$, we must have $v = \ell_a^b(\gamma)/(b - a)$. If we then assume $\beta = \gamma \circ \varphi : [a, b'] \to [a, b]$ for some $b' > a$ and a strictly increasing diffeomorphism $\varphi : [a, b] \to [a, b']$, the condition $|\dot{\beta}(t)| = v$ is satisfied if and only if φ satisfies the differential equation $\dot{\varphi}(t) = v/|\dot{\gamma}(\varphi(t))|$. The right hand side of this equation is positive and bounded away from 0, so after imposing the initial condition $\varphi(a) = a$, there will be a unique solution φ on some interval $[a, b']$ with $b' > a$ uniquely determined by the condition $\varphi(b') = b$. Appealing again to reparametrization invariance, we then find

$$\ell_a^b(\gamma) = \ell_a^{b'}(\beta) = \int_a^{b'} v \, dt = (b' - a)v = \frac{b' - a}{b - a} \ell(\gamma),$$

and thus conclude $b' = b$, proving the claim.

The reparametrization-invariance of ℓ_a^b implies that whenever a path $\gamma \in C^\infty([a, b], M \,; p, q)$ is stationary for ℓ_a^b, all its reparametrizations are as well. Now if $\gamma_s \in C^\infty([a, b], M \,; p, q)$ is a smooth 1-parameter family of *immersed* paths for which $\gamma := \gamma_0$ happens to have constant speed $v := |\dot{\gamma}|$, we find

$$\frac{d}{ds} \ell_a^b(\gamma_s) \Big|_{s=0} = \int_a^b \frac{\partial}{\partial s} \sqrt{\langle \dot{\gamma}_s(t), \dot{\gamma}_s(t) \rangle} \Big|_{s=0} \, dt = \int_a^b \frac{1}{2\sqrt{\langle \dot{\gamma}_0(t), \dot{\gamma}_0(t) \rangle}} \frac{\partial}{\partial s} \langle \dot{\gamma}_s(t), \dot{\gamma}_s(t) \rangle \Big|_{s=0} \, dt$$
$$= \frac{1}{2v} \frac{d}{ds} E_a^b(\gamma_s) \Big|_{s=0}.$$

It follows that if γ is stationary for ℓ_a^b, then it has a reparametrization with constant speed that is stationary for E_a^b, and is therefore a geodesic. Conversely, every geodesic is stationary for ℓ_a^b, and also has constant speed. This proves:

COROLLARY 22.11. *In a Riemannian manifold (M, g), an immersed path $\gamma \in C^\infty([a, b], M \,; p, q)$ is a geodesic if and only if it is stationary for the length functional ℓ_a^b and has constant speed.* $\qquad \square$

We conclude that any path $\gamma \in C^{\infty}([a,b], M\,;\, p, q)$ which minimizes the length $\ell_a^b(\gamma)$ among all nearby paths from p to q can be parametrized by a geodesic. We will discuss a "local" converse to this in the next lecture.

REMARK 22.12. One can extend our discussion of the length functional to the indefinite case with the following modifications. The argument above that $\langle \dot\gamma, \dot\gamma \rangle$ is constant for any geodesic γ is valid for metrics of arbitrary signature, so it makes sense to distinguish between cases where this constant is positive, negative or zero. In general relativity, where the metric has signature $(1,3)$,[63] one calls a geodesic **time-like** if $\langle \dot\gamma, \dot\gamma \rangle > 0$, **space-like** if it is negative and **light-like** if it vanishes. Space-like geodesics represent paths in spacetime that would be perceived by a three-dimensional observer to move faster than the speed of light, while time-like geodesics move slower, and light-like geodesics move (unsurprisingly) at precisely the speed of light. According to the known laws of physics, all freely moving objects with positive mass traverse time-like geodesics in spacetime, and massless objects traverse light-like geodesics. Nothing can traverse a space-like geodesic; its "speed" $|\dot\gamma| := \sqrt{\langle \dot\gamma, \dot\gamma \rangle}$ as measured by the Lorentzian metric $g = \langle\ ,\ \rangle$ would be imaginary.

With this understood, the length functional makes sense for time-like paths, and Corollary 22.11 remains true on a Lorentzian manifold if one restricts attention to time-like geodesics.

23. More on geodesics

23.1. Normal coordinates. Assume M is a smooth manifold without boundary, with a connection ∇. In the previous lecture, we defined the exponential map

$$\exp : TM \supset \mathcal{O} \to M,$$

which is defined on an open subset $\mathcal{O} \subset TM$ containing the zero-section and can be characterized by the property that for each $X \in TM$, $\gamma(t) := \exp(tX)$ is the unique geodesic ($\nabla_t \dot\gamma \equiv 0$) satisfying $\dot\gamma(0) = X$, defined for t in the largest possible interval. We also observed that for each $p \in M$, the restriction

$$\exp_p := \exp|_{\mathcal{O}_p} : \mathcal{O}_p \to M \qquad \text{for } \mathcal{O}_p := \mathcal{O} \cap T_p M$$

satisfies $\exp_p(0) = p$ and has derivative equal to the identity map $T_p M \to T_p M$ at $0 \in \mathcal{O}_p$, implying that it maps a neighborhood of $0 \in T_p M$ diffeomorphically onto a neighborhood of $p \in M$. This means that $\exp_p$ can be used to define local coordinates near p, i.e. if we choose any basis $X_1, \ldots, X_n$ of $T_p M$, then

$$\varphi(t^1, \ldots, t^n) := \exp_p(t^i X_i)$$

defines a diffeomorphism from some neighborhood of $0 \in \mathbb{R}^n$ to some neighborhood $\mathcal{U} \subset M$ of p, and its inverse $x = (x^1, \ldots, x^n) : \mathcal{U} \to x(\mathcal{U}) \subset \mathbb{R}^n$ is therefore a chart sending p to 0.

Charts defined via the exponential map as described above are often referred to as **normal coordinates** about p. They have the following special property. By construction, any path through p that looks in normal coordinates like a straight line with constant velocity through the origin is a geodesic, and any path of this form is also a flow line (through p) of some vector field $Y \in \mathfrak{X}(M)$ that has constant components near p in normal coordinates. The geodesic equation thus implies that all vector fields with this property satisfy

$$\nabla_{Y(p)} Y = 0.$$

This applies in particular to the coordinate vector fields $\partial_1, \ldots, \partial_n$, as well as their linear combinations with constant coefficients, such as $\partial_i + \partial_j$. We therefore have

$$0 = \nabla_{\partial_i + \partial_j}(\partial_i + \partial_j) = \nabla_i \partial_i + \nabla_j \partial_j + \nabla_i \partial_j + \nabla_j \partial_i = \nabla_i \partial_j + \nabla_j \partial_i \quad \text{at } p,$$

---

[63]Many authors also prefer to take $(3,1)$ as the signature of a spacetime manifold, in which case the definitions of the terms "space-like" and "time-like" should be modified by a sign.

implying that the Christoffel symbols satisfy

$$\Gamma_{ij}^k + \Gamma_{ji}^k = 0 \qquad \text{at } p.$$

If the connetion is symmetric, this implies that the Christoffel symbols vanish at p, and we've proved:

PROPOSITION 23.1. *For any symmetric connection ∇ on M, the Christoffel symbols vanish in any normal coordinate system about p.* $\qquad\square$

To take this a step further, suppose (M, g) is a pseudo-Riemannian manifold with signature (k, ℓ) and ∇ is the Levi-Cività connection. In this setting we can require the basis $X_1, \dots, X_n \in T_pM$ in the construction above to be orthonormal, meaning

$$\langle X_i, X_j \rangle = \eta_{ij} := \begin{cases} 1 & \text{if } i = j \leqslant k, \\ -1 & \text{if } i = j > k, \\ 0 & \text{if } i \neq j, \end{cases}$$

and normal coordinates about p under this extra condition are called **Riemann normal coordinates**. The vectors $X_1, \dots, X_n$ match the coordinate vector fields $\partial_1, \dots, \partial_n$ at p, so the components $g_{ij} = \langle \partial_i, \partial_j \rangle$ of the metric now match η_{ij} at p, meaning that $\langle\ ,\ \rangle$ matches the "standard" inner product of signature (k, ℓ) on $\mathbb{R}^n$ at that one point. The vanishing of the Christoffel symbols at that point implies moreover that

$$\partial_k g_{ij}(p) = \partial_k \langle \partial_i, \partial_j \rangle|_p = \langle \nabla_k \partial_i, \partial_j \rangle|_p + \langle \partial_i, \nabla_k \partial_j \rangle|_p = 0$$

for all i, j, k, since the covariant derivatives of the coordinate vector fields all vanish at p. This proves:

PROPOSITION 23.2. *In any Riemann normal coordinate system about a point p in a pseudo-Riemannian manifold (M, g), the components g_{ij} of the metric satisfy*

$$g_{ij}(p) = \eta_{ij}, \qquad \text{and} \qquad \partial_k g_{ij}(p) = 0$$

for all $i, j, k \in \{1, \dots, n\}$. $\qquad\square$

Riemann normal coordinates are sometimes useful for calculations, but their existence also has theoretical importance, for the following reason. The simplest example of a pseudo-Riemannian manifold with signature (k, ℓ) is $\mathbb{R}^n$ with a metric whose components are given by the constants η_{ij} in the obvious global coordinates. In fact, the classification of quadratic forms implies (cf. §18.5) that any pseudo-Riemannian metric on $\mathbb{R}^n$ with constant components can be turned into this one by a global linear change of coordinates. When the signature is $(n, 0)$, this is what we call the *Euclidean* metric; the case of signature $(1, n - 1)$ or $(n - 1, 1)$ is called the **Minkowski metric**, and is important in special relativity. Anticipating the relevance of curvature to this discussion, we shall refer to this metric for arbitrary signatures as the **flat metric** on $\mathbb{R}^n$. The significance of Riemann normal coordinates according to Proposition 23.2 is that they make an arbitrary metric g look more like the flat one, at least at a single point—its value and first derivative at that point match the flat case. We will see when we discuss curvature that, in general, one cannot do better than this: arbitrary pseudo-Riemannian manifolds cannot be made to look like flat space on open neighborhoods of a point just by choosing the right coordinates. Attempting to do this will run into problems as soon as one tries to make the *second* derivatives of g_{ij} vanish, and this impossibility is one of the things that curvature measures.

REMARK 23.3. Another nice trick one can play with the exponential map is to obtain standardized models for arbitrary smooth submanifolds. The result, known as the **tubular neighborhood theorem**, says that if $N \subset M$ is a submanifold and M and N both have empty boundary, then

there is a diffeomorphism Φ from some neighborhood $\mathcal{U} \subset M$ of N to a neighborhood $\mathcal{O} \subset \nu N$ of the zero-section in the total space of its normal bundle (see Example 17.15), and Φ identifies N itself with the zero-section. This result is useful because vector bundles of a given rank over a given manifold are typically not so hard to classify up to bundle isomorphism, thus one obtains manageable lists of models that can describe neighborhoods of all possible embeddings of N into M. For example, one can show that all orientable vector bundles over S^1 are trivial, so one concludes that all embeddings of S^1 into an orientable n-manifold have neighborhoods that look like $S^1 \times \mathbb{D}^{n-1}$; this fact is crucial in knot theory. We refer to [**Hir94**, Chapter 4] for a general discussion of the tubular neighborhood theorem, including a version for manifolds with nonempty boundary. The case of a compact submanifold without boundary is easier, and is Exercise 23.4 below.

EXERCISE 23.4. Suppose N is a compact smooth submanifold of M, where ∂N and ∂M are empty. Choose a Riemannian metric $g = \langle \ , \ \rangle$ on M and recall from Exercise 17.16 that the subbundle $TN^\perp \subset TM|_N$ is isomorphic to the normal bundle of N. Prove:
   (a) For any $\epsilon > 0$ sufficiently small, the set
   $$\mathbb{D}_\epsilon(TN)^\perp := \left\{ X \in TN^\perp \ \middle| \ |X| < \epsilon \right\}$$
   is contained in the domain $\mathcal{O}$ of the exponential map $\exp : TM \supset \mathcal{O} \to M$.
   Hint: For every $p \in M$, the zero vector in T_pM has a neighborhood in TM that belongs to the domain of $\exp$. Use the fact that N is compact.
   (b) The derivative of the smooth map
   $$\Psi := \exp|_{\mathbb{D}_\epsilon(TN)^\perp} : \mathbb{D}_\epsilon(TN)^\perp \to M$$
   is invertible at every zero vector in $TN^\perp$. It follows via the inverse function theorem that for each $p \in N$, Ψ restricts to a diffeomorphism from some neighborhood of $0 \in T_pM$ in $\mathbb{D}_\epsilon(TN)^\perp$ to a neighborhood of p in M.
   (c) After possibly shrinking $\epsilon > 0$ further, the map Ψ in part (b) is a diffeomorphism onto an open neighborhood of N in M.

23.2. The shortest path between nearby points. [64]
Assume (M, g) is a Riemannian manifold and ∇ is the Levi-Cività connection. As you know, if (M, g) is Euclidean space, then the shortest path between any two distinct points $p, q \in M$ is a straight line, also known as a geodesic. Is this true in all Riemannian manifolds? We saw in the previous lecture for instance that any path with constant speed that is a local minimum of the length functional on paths from p to q must be a geodesic. Various subtleties can arise, however, because it is not always true that there is a unique geodesic from p to q, nor must every geodesic from p to q be shorter than all other paths; the easiest example to imagine here is the unit sphere $S^2 \subset \mathbb{R}^3$, which we'll discuss in more detail in the next lecture. More can be said however if we assume that p and q are sufficiently close to each other:

THEOREM 23.5. *For every point p in a Riemannian manifold (M, g), there is a neighborhood $\mathcal{U} \subset M$ of p such that for each $q \in \mathcal{U}$, there exists an embedded geodesic segment $\gamma : [0, 1] \to M$ from $\gamma(0) = p$ to $\gamma(1) = q$ that is strictly shorter than all paths from p to q other than the reparametrizations of γ.*

The existence of the neighborhood $\mathcal{U} \subset M$ in this theorem is the easy part: we have already seen that $\exp_p$ maps some neighborhood $\mathcal{O} \subset T_pM$ of 0 diffeomorphically to a neighborhood $\mathcal{U} \subset M$ of p. Every $q \in \mathcal{U}$ can then be written as $q = \exp_p(X)$ for a unique $X \in \mathcal{O}$, and we may as

---

[64]All details concerning indefinite metrics in this section should be considered inessential to the course and are included only out of interest—they were not covered in the lecture.

well assume $\mathcal{O} \subset T_pM$ is a **star-shaped** neighborhood, meaning that $tX \in \mathcal{O}$ for every $X \in \mathcal{O}$ and $t \in [0,1]$, so that the geodesic segment $\gamma : [0,1] \to M : t \mapsto \exp(tX)$ from p to q is also contained in $\mathcal{U}$. Note that this might not necessarily be the *only* geodesic segment connecting p to q, though it is certainly the only one that is fully contained in $\mathcal{U}$. The goal is to show that this particular geodesic segment and its reparametrizations are strictly shorter than all other paths from p to q.

The key turns out to be an observation that sounds eminently plausible in our geometric intuition, but is a bit tricky to prove: every geodesic emerging from p is *orthogonal* to the spheres of constant radius around p. By "spheres of constant radius", we mean more precisely the following: for each $r \in \mathbb{R}$, consider the set

(23.1) $$\Sigma_r := \left\{ \exp_p(X) \in M \mid X \in \mathcal{O} \text{ and } \langle X, X \rangle = r \right\} \subset \mathcal{U}.$$

The definition also makes sense when the metric is indefinite, so we have allowed r to be any real number, not just $r > 0$. The condition $\langle X, X \rangle = r$ cuts out a smooth hypersurface in T_pM for any $r \neq 0$, and this is also true at $r = 0$ with the exception of a singular point at the origin, thus Σ_r is a smooth hypersurface in M for every $r \neq 0$, and so is Σ_0 except at the isolated singular point $p \in \Sigma_0$, which we will exclude.[65]

PROPOSITION 23.6 (Gauss lemma). *Assume (M, g) is a pseudo-Riemannian manifold without boundary, $0 \in \mathcal{O} \subset T_pM$ denotes the star-shaped neighborhood described above with $\mathcal{U} = \exp_p(\mathcal{O})$, and $\Sigma_r \subset \mathcal{U}$ is the hypersurface defined in (23.1). Then for every $r \in \mathbb{R}$, any geodesic segment of the form $\gamma(t) = \exp_p(tX)$ for $X \in \mathcal{O} \backslash \{0\}$ with $\gamma(1) \in \Sigma_r$ hits Σ_r orthogonally, i.e. $\langle \dot{\gamma}(1), Y \rangle = 0$ for all $Y \in T_{\gamma(1)}\Sigma_r$.*

PROOF. Suppose $\exp_p(X) = q \in \Sigma_r$, meaning $\langle X, X \rangle = r$, and pick any $Y \in T_q\Sigma_r$. The latter can be realized as $Y = \partial_t f(1, 0)$ for a smooth map of the form

$$f : [0, 1 + \epsilon) \times (-\epsilon, \epsilon) \to M : (s, t) \mapsto \exp_p(sX(t)) \in \mathcal{U}$$

with $\epsilon > 0$ chosen sufficiently small and $X(t) \in T_pM$ a smooth path with $X(0) = X$ and $\langle X(t), X(t) \rangle = r$ for all t. The lemma will thus follow from the claim that for any map of this form,

$$\langle \partial_s f, \partial_t f \rangle \equiv 0.$$

When $s = 0$ this is immediate, because $f(0, t) = p$ for all t and thus $\partial_t f(0, t) = 0$. Using the properties of the Levi-Cività connection and the fact that $s \mapsto f(s, t) = \exp_p(sX(t))$ is a geodesic for each fixed t, we also have

(23.2) $$\partial_s \langle \partial_s f, \partial_t f \rangle = \langle \nabla_s \partial_s f, \partial_t f \rangle + \langle \partial_s f, \nabla_s \partial_t f \rangle = \langle \partial_s f, \nabla_t \partial_s f \rangle.$$

Next observe that for each t, the "speed squared"[66] $\langle \partial_s f, \partial_s f \rangle$ of the geodesic $s \mapsto f(s, t)$ is a constant independent of s, because $\nabla_s \partial_s f \equiv 0$ implies $\partial_s \langle \partial_s f, \partial_s f \rangle \equiv 0$ and thus $\langle \partial_s f(s, t), \partial_s f(s, t) \rangle = \langle \partial_s f(0, t), \partial_s f(0, t) \rangle = \langle X(t), X(t) \rangle = r$. This proves

$$0 = \partial_t \langle \partial_s f, \partial_s f \rangle = 2 \langle \nabla_t \partial_s f, \partial_s f \rangle,$$

so that (23.2) now vanishes, thus establishing that $\langle \partial_s f(s, t), \partial_t f(s, t) \rangle = \langle \partial_s f(0, t), \partial_t f(0, t) \rangle = 0$ for all (s, t). $\qquad\square$

---

[65] If the metric is positive, then Σ_0 consists only of the point p and is thus excluded from this discussion. But Σ_0 is more interesting in the indefinite case: imagine for instance the standard indefinite inner product of signature $(1, 1)$ on $\mathbb{R}^2$, so that $\langle X, X \rangle = 0$ is equivalent to the equation $x^2 - y^2 = 0$. This cuts out a smooth submanifold with an isolated singularity at the origin.

[66] The *speed* $|\dot{\gamma}(t)|$ of a geodesic γ only makes sense when the metric is positive, but "speed squared" $|\dot{\gamma}(t)|^2 := \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle$ can also be defined in the indefinite case, with the understanding that it might be negative.

REMARK 23.7. The $r = 0$ case of the Gauss lemma is vacuous when the metric is positive, and what it says in the indefinite case is slightly counterintuitive: observe that if $X \in T_pM$ is a nonzero vector with $\langle X, X \rangle = 0$, then also $\langle tX, tX \rangle = 0$ for every t and the geodesic $t \mapsto \exp(tX)$ is therefore *contained* in Σ_0, in addition to being (according to the statement of the proposition) orthogonal to it. This is not a contradiction, because while $\Sigma_0 \backslash \{p\}$ is a well-defined submanifold of M, it is not what we would call a *pseudo-Riemannian submanifold* of (M, g), i.e. the restriction of $\langle \ , \ \rangle$ to Σ_0 is degenerate and thus fails to be a pseudo-Riemannian metric. As a consequence, for $q \in \Sigma_0 \backslash \{p\}$, the "orthogonal complement" $(T_q\Sigma_0)^\perp := \left\{ Y \in T_qM \mid \langle Y, X \rangle = 0 \text{ for all } X \in T_q\Sigma_0 \right\}$ has the correct dimension but is not actually *complementary* to $T_q\Sigma_0$, but is instead contained in it. The content of Proposition 23.6 is then that for each $q \in \Sigma_0$, $(T_q\Sigma_0)^\perp$ is the 1-dimensional subspace of $T_q\Sigma_0$ spanned by the tangent vector of a geodesic connecting p to q.

PROOF OF THEOREM 23.5. Fix $\mathcal{O} \subset T_pM$ and $\mathcal{U} \subset M$ as in Proposition 23.6, assuming additionally that the metric $\langle \ , \ \rangle$ is positive and $\mathcal{O}$ has the form of a ball,

$$\mathcal{O} = \left\{ X \in T_pM \mid \langle X, X \rangle < R^2 \right\}$$

for some $R > 0$. Given $q = \exp_p(X) \in \mathcal{U} \backslash \{p\}$ with $X \in \mathcal{O} \backslash \{0\}$, the geodesic segment $\gamma_0 : [0, 1] \to \mathcal{U} : t \mapsto \exp_p(tX)$ has length $\ell_0^1(\gamma_0) = |X| =: \sqrt{r}$. For any other smooth path $\gamma : [0, 1] \to \mathcal{U}$ from $\gamma(0) = p$ and $\gamma(1) = q$, let us assume after a small perturbation that $\gamma(t) \neq p$ for all $t \neq 0$, in which case we can write

$$\gamma(t) = \exp_p(\rho(t)X(t)) \qquad \text{for all } t \in (0, 1],$$

with uniquely-determined smooth paths $\rho(t) > 0$ and $X(t) \in \Sigma_r$ satisfying $\lim_{t \to 0} \rho(t) = 0$, $\rho(1) = 1$ and $X(1) = X$. Write $f(s, t) = \exp_p(sX(t))$ as in the proof of Prop. 23.6, so the proposition implies $\langle \partial_s f, \partial_t f \rangle \equiv 0$, and since $s \mapsto f(s, t)$ is a geodesic with constant speed starting at $X(t)$ for each t, $|\partial_s f(s, t)| = |X(t)| = \sqrt{r}$ for every t. Now since $\gamma(t) = f(\rho(t), t)$, we have $\dot\gamma(t) = \partial_s f(\rho(t), t)\dot\rho(t) + \partial_t f(\rho(t), t)$, and using the Pythagorean theorem,

$$|\dot\gamma(t)|^2 = |\partial_s f(\rho(t), t)\dot\rho(t)|^2 + |\partial_t f(\rho(t), t)|^2 \geqslant r|\dot\rho(t)|^2,$$

with strict inequality unless the path $X(t)$ is constant. The latter would mean $\gamma(t) = \exp_p(\rho(t)X) = \gamma_0(\rho(t))$, so that γ traces out the same image as γ_0, with a strictly longer length unless $\rho : (0, 1] \to (0, 1]$ is a diffeomorphism, which means γ is a reparametrization of γ_0. Now if $X(t)$ is not constant, we have

$$|\dot\gamma(t)| > \sqrt{r}|\dot\rho(t)| \geqslant \sqrt{r}\dot\rho(t) \qquad \text{for all } t \in (0, 1]$$

and thus

$$\ell_0^1(\gamma) = \int_0^1 |\dot\gamma(t)|\, dt > \sqrt{r} \int_0^1 \dot\rho(t)\, dt = \sqrt{r} = \ell_0^1(\gamma_0).$$

This proves that all paths from p to q contained in $\mathcal{U}$ are strictly longer than the reparametrizations of γ_0. Any path that is *not* contained in $\mathcal{U}$ is obviously also longer, because it must cover a distance of at least $R > \sqrt{r}$ after starting at p before it can exit $\mathcal{U}$. $\qquad \square$

REMARK 23.8. It is not straightforward to formulate variants of Theorem 23.5 with indefinite metrics, but on a pseudo-Riemannian manifold with Lorentz signature $(1, n-1)$ one can say the following. Recall from Remark 22.12 that the length functional is well-defined on time-like paths γ since their velocities satisfy $\langle \dot\gamma, \dot\gamma \rangle > 0$. It is not really appropriate to call it "length" in this situation, though; physicists prefer to call it the **proper time**, because in a Lorentzian 4-manifold representing spacetime, the proper time of a time-like path is the actual amount of time elapsed on a clock that is carried along that path through spacetime. Let us therefore denote

$$\tau_a^b(\gamma) := \int_a^b \sqrt{\langle \dot\gamma(t), \dot\gamma(t) \rangle}\, dt,$$

and consider the proper time of a time-like path from p to a point $q = \exp_p(X)$ that is "nearby" in the sense that $X \in T_pM$ is close to 0. The set of points q that can be reached in this way from p is called the **light cone** of p; it is an open subset bounded by light-like paths, i.e. paths that represent objects moving at the speed of light. An interesting detail arises here that is completely unlike anything in the Riemannian case: if you look at the standard Lorentzian inner product in an orthonormal basis so that it takes the form

$$\langle X, Y \rangle = X^1 Y^1 - \sum_{j=2}^n X^j Y^j,$$

you may notice that the set of time-like vectors (satisfying $\langle X, X \rangle > 0$) has two connected components, and as a result, the light cone of p is guaranteed to have two components if the neighborhoods $\mathcal{O} \subset T_pM$ and $\mathcal{U} \subset M$ are chosen sufficiently small. This is a symptom of the fact that in the physical world, there is a distinction between time-like paths moving forward or backward in time. We can therefore label the two components of the light cone C_p^+ and C_p^-, call them the *positive* and *negative* light cone respectively, and say $q \in C_p^+$ if and only if q is in the *future* of p.

One can now ask the following: how does the proper time of the geodesic segment $\gamma_0(t) = \exp_p(tX)$ compare with that of all other future-directed time-like paths from p to q?

The following detail is important to understand first: according to Proposition 23.6, time-like paths will pass orthogonally through hypersurfaces $\Sigma_r \subset \mathcal{U}$ with $r > 0$, and the restriction of the Lorentzian metric $\langle \, , \, \rangle$ to these hypersurfaces is *negative*, i.e. it is $-h_r$ for a Riemannian metric h_r on Σ_r. One can deduce this from the fact that each geodesic $\gamma(t) = \exp(tX)$ hits Σ_r orthogonally: given that $\langle \dot\gamma(t), \dot\gamma(t) \rangle > 0$, the only way for $\langle \, , \, \rangle$ to have signature $(1, n-1)$ at the intersection point is if it is negative-definite on $T\Sigma_r$.

Now, imitating the proof of Theorem 23.5, the assumption that $q = \exp_p(X) \in C_p^+$ implies $\langle X, X \rangle = [\tau_0^1(\gamma_0)]^2 =: r$, hence $q \in \Sigma_r$, and we can consider arbitrary paths from p to q of the form $\gamma(t) = \exp_p(\rho(t)X(t)) = f(\rho(t), t)$, where $f(s, t) = \exp_p(sX(t))$, $X(t) \in \Sigma_r$, $X(1) = X$, $\rho(1) = 1$ and $\lim_{t \to 0} \rho(t) = 0$. We still have $\langle \partial_s f, \partial_t f \rangle \equiv 0$, but the big difference from Theorem 23.5 is now that $\langle \partial_t f, \partial_t f \rangle \leqslant 0$, with strict inequality unless $\partial_t X = 0$, so our previous application of the Pythagorean theorem becomes

$$\langle \dot\gamma(t), \dot\gamma(t) \rangle = r|\dot\rho(t)|^2 - |\partial_t f(\rho(t), t)|^2 \leqslant r|\dot\rho(t)|^2,$$

again with equality only if $X(t)$ is constant. Requiring γ to be time-like then imposes the condition $r|\dot\rho(t)|^2 > |\partial_t f(\rho(t), t)| \geqslant 0$, so in contrast to the Riemannian case, we can only consider paths for which $\dot\rho(t) \geqslant 0$. The end result is that either γ is a reparametrization of γ_0 or

$$\tau_0^1(\gamma) = \int_0^1 \sqrt{\langle \dot\gamma(t), \dot\gamma(t) \rangle}\, dt < \sqrt{r} \int_0^1 \dot\rho(t)\, dt = \sqrt{r} = \tau_0^1(\gamma_0),$$

thus the geodesic from p to q *maximizes* the proper time among time-like paths from p to q.

One can use this calculation to explain the famous "twins paradox" in relativity—it is not a paradox, but merely a result of the fact that the proper time is not the same for all time-like paths between two points in spacetime. The scenario is that Albert and Henry are born at the same time, but Albert stays for his whole life on Earth, while Henry becomes an astronaut and travels several light-years across the universe and back, travelling at nearly the speed of light in both directions. On return, Henry has barely aged at all, but Albert is twenty years older. The reason is that by staying on Earth, Albert followed a geodesic in spacetime, but Henry did not: his path was at best a *piecewise* smooth geodesic, because he had to accelerate abruptly in order to reverse course and return to Earth. As a result, Albert's path experienced more proper time than Henry's.

REMARK. This lecture took place online and was interrupted multiple times by internet outages, as a result of which, there was no time to cover the two sections below. The contents of §23.3 will be covered briefly in the next online lecture, and §23.4 will be the main topic in one of this week's problem sessions.

23.3. Geodesic completeness. For an arbitrary pseudo-Riemannian manifold (M, g), the domain of the exponential map is an open subset $\mathcal{O} \subset TM$, and we say that (M, g) is **geodesically complete** if $\mathcal{O} = TM$. An equivalent condition is that for every $p \in M$ and $X \in T_pM$, the unique maximal geodesic $\gamma : (a, b) \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$ is defined for all time, i.e. the interval (a, b) is $\mathbb{R}$. It is easy to find examples for which this is not true, e.g. take $\mathbb{R}^n$ with a flat metric but remove a point to define $M := \mathbb{R}^n \backslash \{p\}$; then there exist geodesics in M that are not defined for all time because they collide at some time with the missing point p. In general, this would seem to be a danger whenever M is noncompact, because any given geodesic could potentially "escape to infinity" in finite time. We will see below that there is still a danger in general even when M is compact. On the other hand, $\mathbb{R}^n$ with a flat metric is an obvious example of a noncompact geodesically complete manifold: its geodesics are precisely the straight paths $\gamma(t) = \mathbf{v} + t\mathbf{w}$, which are defined for all $t \in \mathbb{R}$.

In order to say something general about completeness, it is useful to reformulate the problem in terms of the flow of a vector field. It will not be a vector field on M, because the geodesic equation is second order, so solutions are determined by more than just their initial position; the initial velocity is also required. This suggests defining a vector field instead on the tangent bundle TM, such as

$$(23.3) \qquad \xi(X) := \mathrm{Hor}_X(X) \in T_X(TM),$$

where for $p \in M$ and $X \in T_pM$, $\mathrm{Hor}_X : T_pM \to T_X(TM)$ denotes the horizontal lift map for some connection ∇ on TM. Suppose $Y(t) \in TM$ is a flow line of $\xi \in \mathfrak{X}(TM)$, and using the bundle projection $\pi : TM \to M$, let $\gamma(t) := \pi \circ Y(t) \in M$, so γ is a path in M and Y is a vector field along γ. The condition $\partial_t Y(t) = \xi(Y(t)) = \mathrm{Hor}_{Y(t)}(Y(t))$ then implies

$$\dot{\gamma}(t) = \pi_*\left(\partial_t Y(t)\right) = \pi_* \mathrm{Hor}_{Y(t)}(Y(t)) = Y(t)$$

and, using the vertical projection $K : T(TM) \to TM$ to write the covariant derivative via (20.9),

$$\nabla_t Y(t) = K(\partial_t Y(t)) = 0.$$

In other words, Y is the velocity of γ and it is parallel along γ, hence γ is a geodesic. Conversely, if γ is any geodesic in M, then the path $Y(t) := \dot{\gamma}(t)$ in TM satisfies $\nabla_t Y(t) = 0$, implying that $\partial_t Y(t)$ is horizontal, which it means it can only be the horizontal lift of $\dot{\gamma}(t)$ and thus satisfies $\partial_t Y(t) = \xi(Y(t))$. This proves:

PROPOSITION 23.9. *Suppose ∇ is any connection on the tangent bundle $\pi : TM \to M$, and $\xi \in \mathfrak{X}(TM)$ is the vector field defined by (23.3) in terms of this connection. Then the exponential map has the same domain as the time 1 flow φ_ξ^1 of ξ, and $\exp = \pi \circ \varphi_\xi^1$.* $\qquad\square$

The flow of the vector field ξ on TM is called the **geodesic flow** for M with connection ∇. It becomes an especially useful tool if we specialize to the Levi-Cività connection of a Riemannian metric:

THEOREM 23.10. *Every compact Riemannian manifold without boundary is geodesically complete.*

PROOF. Assuming (M, g) is a compact Riemannian manifold, we define the vector field $\xi \in \mathfrak{X}(TM)$ as in (23.3) using the Levi-Cività connection and notice that it has the following useful

property: for each $r > 0$, ξ is tangent to the smooth hypersurface

$$S_r TM := \left\{ X \in TM \mid \langle X, X \rangle = r^2 \right\}.$$

This follows from the fact that geodesics have constant speed, thus all flow lines of ξ are confined to hypersurfaces of this form.

Now observe that since M is compact, $S_r TM$ is also compact for every $r > 0$: indeed, the intersection of $S_r TM$ with each fiber $T_p M$ is a compact $(n-1)$-sphere in $T_p M$, thus for any sufficiently small compact neighborhood $K \subset M$ of p, one can use a local trivialization of $TM|_K$ to show that $S_r TM \cap \pi^{-1}(K)$ is homeomorphic to the compact set $K \times S^{n-1}$. Then if $X_k \in S_r TM$ is any sequence and we write $p_k := \pi(X_k) \in M$, the compactness of M implies after restricting to a subsequence that p_k converges to some point $p \in M$, so that X_k for large k lies in a neighborhood homeomorphic to a compact set of the form $K \times S^{n-1}$ and therefore also has a convergent subsequence.

For any given $X \in TM$, one can now define $r := |X|$ and regard ξ as a vector field on the compact manifold $S_r TM$ instead of TM; since every vector field on a compact manifold has a global flow, the theorem follows. $\qquad\square$

Theorem 23.10 depends rather crucially on the assumption that ∇ is the Levi-Cività connection for a *positive* metric g. A negative metric would also be fine, but the trouble with signatures (k, ℓ) with $k, \ell > 0$ is that while the geodesic flow $\xi \in \mathfrak{X}(TM)$ is tangent to hypersurfaces of the form

$$\left\{ X \in TM \mid \langle X, X \rangle = c \right\}$$

for constants $c \in \mathbb{R}$, these hypersurfaces are not compact, even if M is. The problem is clearly visible if you look at the intersection of this hypersurface with a single fiber $T_p M$: choosing an orthonormal basis on $T_p M$ so that $\langle X, Y \rangle = \sum_{j=1}^{k} X^j Y^j - \sum_{j=k+1}^{n} X^j Y^j$, the set of vectors $X \in T_p M$ with $\langle X, X \rangle = c$ is not a sphere, it is a hyperboloid, which is definitely not compact. As such, there is no reason to expect the geodesic flow on TM to be globally defined, and in general, it is not. There are simple examples of indefinite pseudo-Riemannian manifolds that are compact but not geodesically complete.[67]

On the flip side, there are plenty of interesting Riemannian manifolds that are geodesically complete despite being noncompact; we will discuss some important examples in the next lecture. We do not have space here to prove the main result on this subject, but we plan to do so next semester, so consider the following statement a preview:

THEOREM (Hopf-Rinow theorem). *A connected Riemannian manifold (M, g) is geodesically complete if and only if it is a complete metric space with respect to the metric defined as the infimum of lengths of paths between points. Moreover, if it is complete, then for every pair of points $p, q \in M$, there exists a (not necessarily unique) geodesic segment from p to q that minimizes the length among all paths from p to q.*

23.4. Geodesics as a Hamiltonian system. The notion of the geodesic flow on TM can be placed into a wider context that connects it with symplectic geometry (cf. Lecture 14). To see this, we start with the observation that for any smooth manifold M, the cotangent bundle T^*M admits a canonical symplectic form. One defines it as follows: first let $\pi : T^*M \to M$ denote the bundle projection for the cotangent bundle, whose derivative gives a map $T\pi : T(T^*M) \to TM$ sending $T_\alpha(T^*M)$ linearly to $T_q M$ for each $q \in M$ and $\alpha \in T_q^* M$. We can thus define a 1-form $\lambda \in \Omega^1(T^*M)$ by

$$(23.4) \qquad\qquad \lambda_\alpha(\xi) := \alpha(T\pi(\xi)).$$

---

[67]See for instance the *Clifton-Pohl torus*: https://en.wikipedia.org/wiki/Clifton-Pohl_torus.

This is called the **tautological** 1-**form** on T^*M, and we will see below that

$$\omega := d\lambda \in \Omega^2(T^*M)$$

is a symplectic form. Recall from Lecture 14: this would mean that every point in T^*M has a neighborhood on which there exists a chart of the form $(p^1, q^1, \ldots, p^n, q^n)$ such that $\omega = \sum_{j=1}^n dp^j \wedge dq^j$. We claim in fact that any chart $(\mathcal{U}, (x^1, \ldots, x^n))$ on M naturally gives rise to a chart with this property on the open set $T^*M|_{\mathcal{U}} = \pi^{-1}(\mathcal{U}) \subset T^*M$. Indeed, we define n of the required $2n$ coordinates on $T^*M|_{\mathcal{U}}$ by

$$q^i := x^i \circ \pi, \qquad i = 1, \ldots, n.$$

For the remaining n coordinates, observe that the coordinates $x^1, \ldots, x^n$ give us a natural basis for each of the cotangent spaces over $\mathcal{U}$, namely the coordinate differentials, so let us define $p^1, \ldots, p^n$ on $T^*M|_{\mathcal{U}}$ by

$$(p^1, \ldots, p^n)\left(a_i\, dx^i\right) := (a_1, \ldots, a_n) \in \mathbb{R}^n.$$

Now observe: if a path $s(t) \in T^*M|_{\mathcal{U}}$ has constant coordinates $q^1, \ldots, q^n$, it means that $s(t)$ is moving within a single fiber, thus the velocity vectors $\dot{s}(t)$ belong to the vertical subbundle $V(T^*M) \subset T(T^*M)$, and in particular, this applies to the coordinate vector fields

$$\frac{\partial}{\partial p^1}, \ldots, \frac{\partial}{\partial p^n} \in V(T^*M).$$

On the other hand, for any path $s(t) \in T^*M|_{\mathcal{U}}$ whose coordinates are all constant except for one particular q^i, it follows that $\pi \circ s(t) \in \mathcal{U}$ has constant coordinates except for x^i, and thus

$$\pi_* \frac{\partial}{\partial q^i} = \frac{\partial}{\partial x^i} \qquad \text{for } i = 1, \ldots, n.$$

One sees now from the definition of $\lambda \in \Omega^1(T^*M)$ that it annihilates all vertical vectors, thus if we denote by $\alpha \in T^*M|_{\mathcal{U}}$ the point with some particular value of the coordinates $q^1, \ldots, q^n, p^1, \ldots, p^n$, then $\lambda_\alpha\left(\frac{\partial}{\partial p^i}\right) = 0$, and

$$\lambda_\alpha\left(\frac{\partial}{\partial q^i}\right) = \alpha\left(\pi_* \frac{\partial}{\partial q^i}\right) = \alpha\left(\frac{\partial}{\partial x^i}\right) = p^i.$$

The formula for λ in our chosen coordinates is therefore

$$\lambda = \sum_{i=1}^n p^i\, dq^i,$$

and it follows that $\omega = \sum_{i=1}^n dp^i \wedge dq^i$, so ω is symplectic.

REMARK 23.11. Symplectic geometers sometimes abbreviate the tautological 1-form λ and symplectic form $\omega = d\lambda$ on T^*M by "$p\, dq$" and "$dp \wedge dq$" respectively, where the symbols p and q are each meant as shorthand for n separate coordinates. It is a somewhat remarkable fact that $p\, dq$ turns out to be the same 1-form no matter how one chooses the local coordinates $q^1, \ldots, q^n$; what makes this possible is the fact that while the coordinates $q^1, \ldots, q^n$ are arbitrarily chosen on some open subset of M, the remaining n coordinates $p^1, \ldots, p^n$ are not at all arbitrary, in fact they are completely determined by $q^1, \ldots, q^n$. One cannot assume in general that any of these coordinates are globally defined, but $p\, dq$ does make sense globally, because one can also express it as in (23.4) without choosing any coordinates.

The symplectic structure of T^*M provides a natural framework for viewing second-order dynamical systems on M as Hamiltonian systems on T^*M, and the geodesic flow is the simplest

interesting example of this. In order to see it clearly, it will help to adopt the following notation: let us denote elements of T^*M as pairs

$$(q, p) \in T^*M, \qquad \text{where} \qquad q \in M \text{ and } p \in T_q^*M,$$

thus explicitly keeping track of horizontal motion via the symbol q and vertical motion via p. The easiest way to understand the tangent spaces $T_{(q,p)}(T^*M)$ then comes from choosing a connection ∇ on $\pi : T^*M \to M$, as it gives rise to a horizontal/vertical splitting

$$T_{(q,p)}(T^*M) = H_{(q,p)}(T^*M) \oplus V_{(q,p)}(T^*M)$$

such that $V_{(q,p)}(T^*M)$ is canonically isomorphic to the fiber T_q^*M and π_* gives a natural isomorphism of $H_{(q,p)}(T^*M)$ with the tangent space $T_q M$. The connection thus determines a natural isomorphism

$$T_{(q,p)}(T^*M) \cong T_q M \oplus T_q^* M,$$

and with this in mind, we shall write elements of $T_{(q,p)}(T^*M)$ as pairs of the form (Y, η) with $Y \in T_q M$ and $\eta \in T_q^* M$. For a path $\gamma(t) = (q(t), p(t)) \in T^*M$, the derivative $\dot{\gamma}(t) \in T_{\gamma(t)}(T^*M)$ is now written as a pair (Y, η) where $Y = \dot{q}(t) \in T_{q(t)}M$, and $\eta \in T_{q(t)}^* M$ is literally the projection of $\dot{\gamma}(t) \in T_{\gamma(t)}(T^*M)$ along the horizontal subspace to the vertical subspace, which means the covariant derivative, hence

$$\partial_t(q(t), p(t)) = (\dot{q}(t), \nabla_t p(t)) \in T_{q(t)} M \oplus T_{q(t)}^* M \cong T_{(q(t), p(t))}(T^*M).$$

Once one gets used to these natural isomorphisms, computations in T^*M become fairly straightforward.

Now suppose $g = \langle \ , \ \rangle$ is a pseudo-Riemannian metric on M, and the connection ∇ on $T^*M \to M$ is the connection induced on T^*M by the Levi-Cività connection of $TM \to M$. Using the musical isomorphisms to define a corresponding bundle metric on T^*M, Exercise 22.5 implies that our connection on $T^*M \to M$ is compatible with this bundle metric. The simplest Hamiltonian function that might be interesting to consider on T^*M is

$$(23.5) \qquad\qquad H : T^*M \to \mathbb{R} : (q, p) \mapsto \frac{1}{2}\langle p, p \rangle.$$

There is some physical motivation to look at this particular function: in the special case of $M = \mathbb{R}^n$ with the standard Euclidean metric, H has an interpretation as the classical kinetic energy of a moving particle with mass 1. If we assume this is also the *total* energy, meaning there is no potential energy and thus no forces acting on the particle, then the motion of the particle is along straight lines in $\mathbb{R}^n$, and these are the geodesics in Euclidean space. It is not unreasonable to hope that the same correspondence might hold on a general pseudo-Riemannian manifold, and indeed:

PROPOSITION 23.12. *The Hamiltonian vector field for the function H in (23.5) is given by* $X_H(q, p) = (p^\sharp, 0) \in T_q M \oplus T_q^* M \cong T_{(q,p)}(T^*M)$.

Before proving the proposition, let us see what the flow of X_H looks like. For a path $\gamma(t) = (q(t), p(t)) \in T^*M$, $\dot{\gamma}(t) = X_H(\gamma(t))$ now means

$$\dot{q}(t) = p(t)^\sharp \qquad \text{and} \qquad \nabla_t p(t) = 0.$$

By Exercise 22.5, $\nabla_t \dot{q} = \nabla_t(p^\sharp) = (\nabla_t p)^\sharp = 0$, so this implies that the path $q(t)$ is a geodesic and $p(t)$ is simply the image of its velocity under the musical isomorphism $\flat : T_{q(t)} M \to T_{q(t)}^* M$. This proves a "Hamiltonian version" of the main result about the geodesic flow:

PROPOSITION 23.13. *Given a pseudo-Riemannian manifold (M, g) with Levi-Cività connection ∇ and the function $H : T^*M \to \mathbb{R}$ defined in (23.5) via the metric, the exponential map on TM is related to the flow of the Hamiltonian vector field X_H on T^*M by* $\exp(Y) = \pi \circ \varphi_{X_H}^1(Y_\flat)$, *where* π *is the bundle projection $T^*M \to M$.* $\qquad\qquad\square$

Turning toward the proof of Proposition 23.12, it will be useful to have a coordinate-independent formula for ω that is more direct than calling it the exterior derivative of λ.

LEMMA 23.14. *Using the isomorphism $T_{(q,p)}(T^*M) \cong T_qM \oplus T_q^*M$, the canonical symplectic form ω on T^*M is given by*

$$\omega_{(q,p)}((Y,\eta),(Y',\eta')) = \eta(Y') - \eta'(Y).$$

PROOF. Using bilinearity and antisymmetry, it suffices to prove three more specific formulae:

(i) $\omega_{(q,p)}((0,\eta),(0,\eta')) = 0$ for all $\eta, \eta' \in T_q^*M$;
(ii) $\omega_{(q,p)}((Y,0),(Y',0)) = 0$ for all $Y, Y' \in T_qM$;
(iii) $\omega_{(q,p)}((Y,0),(0,\eta)) = -\eta(Y)$ for all $Y \in T_qM$ and $\eta \in T_q^*M$.

For all three, we will use the relation

(23.6) $$d\lambda(\partial_s f, \partial_t f) = \partial_s\left[\lambda(\partial_t f)\right] - \partial_t\left[\lambda(\partial_s f)\right],$$

which is valid for any smooth map $\mathbb{R}^2 \overset{\text{open}}{\supset} \mathcal{V} \to M : (s,t) \mapsto f(s,t)$. Indeed, this is actually just a computation of $f^*d\lambda(\partial_s, \partial_t) = d(f^*\lambda)(\partial_s, \partial_t)$, and since the coordinate vector fields ∂_s and ∂_t on $\mathcal{V} \subset \mathbb{R}^2$ commute, the relation follows from our original definition of the exterior derivative in §8.2.

With this understood, let us first prove (i). Given $\eta, \eta' \in T_q^*M$, define

$$f(s,t) = (q, p + s\eta + t\eta') \in T^*M,$$

so $f(0,0) = (q,p)$, $\partial_s f(0,0) = (0,\eta)$ and $\partial_t f(0,0) = (0,\eta')$. Since $\partial_s f$ and $\partial_t f$ are both vertical vectors for every (s,t), λ annihilates them both and both terms in (23.6) therefore vanish, thus proving $d\lambda(\partial_s f, \partial_t f) = 0$.

Next is (ii): choose f in the form

$$f(s,t) = (\gamma(s,t), \sigma(s,t)) \in T^*M$$

such that $\gamma(0,0) = q$, $\partial_s\gamma(0,0) = Y$ and $\partial_t\gamma(0,0) = Y'$, and σ is a section of T^*M along γ that satisfies $\sigma(0,0) = p$ and has vanishing covariant derivative at $(s,t) = (0,0)$. (To see that the latter is possible, one can e.g. first define $\sigma(s,0)$ by parallel transporting p along the path $s \mapsto \gamma(s,0)$, then define $\sigma(s,t)$ by parallel transporting $\sigma(s,0)$ along the path $t \mapsto \gamma(s,t)$ for each fixed s.) We now have $\partial_s f(0,0) = (Y,0)$ and $\partial_t f(0,0) = (Y',0)$, and by (23.6),

$$d\lambda(\partial_s f, \partial_t f) = \partial_s\left[\sigma(\partial_t\gamma)\right] - \partial_t\left[\sigma(\partial_s\gamma)\right] = (\nabla_s\sigma)(\partial_t\gamma) - (\nabla_t\sigma)(\partial_s\gamma) + \sigma(\nabla_s\partial_t\gamma - \nabla_t\partial_s\gamma).$$

The last term in this expression vanishes identically because the Levi-Cività connection is symmetric, and the first two terms vanish specifically at $s = t = 0$ because $\nabla\sigma = 0$ at that point, so (i) is proven.

For (iii), we choose f in the form

$$f(s,t) = (\gamma(s), \sigma(s) + t\xi(s)) \in T^*M$$

such that $\gamma(0) = q$, $\gamma'(0) = Y$, and σ and ξ are parallel sections of T^*M along γ with $\sigma(0) = p$ and $\xi(0) = \eta$, thus $\partial_s f(0,0) = (Y,0)$ and $\partial_t f(0,0) = (0,\eta)$. Since $\partial_t f(s,t) = (0,\xi(s))$ is always vertical, $\lambda(\partial_t f) \equiv 0$, and (23.6) thus gives

$$d\lambda(\partial_s f(s,t), \partial_t f(s,t)) = -\partial_t\left[\lambda(\partial_s f(s,t))\right] = -\partial_t\left[(\sigma(s) + t\xi(s))(\gamma'(s))\right] = -\xi(s)\left(\gamma'(s)\right),$$

which is $-\eta(Y)$ at $s = 0$. $\square$

PROOF OF PROPOSITION 23.12. The function $H(q,p) = \frac{1}{2}\langle p, p\rangle$ is constant in horizontal directions since parallel transport preserves the bundle metric, and in vertical directions, its differential is simply the differential at $(q,p) \in T^*M$ of the quadratic function $T_q^*M \to \mathbb{R} : p \mapsto \frac{1}{2}\langle p, p\rangle$, giving

$$dH(q,p)(Y,\eta) = \langle p, \eta\rangle.$$

Plugging $X_H(q,p) = (p^\sharp, 0)$ into the formula of Lemma 23.14 for ω gives

$$\omega(X_H(q,p),(Y,\eta)) = -\eta(p^\sharp) = -\langle \eta^\sharp, p^\sharp \rangle = -\langle p, \eta \rangle,$$

thus we've proven that X_H satisfies the defining equation $\omega(X_H, \cdot) = -dH$ of a Hamiltonian vector field. $\qquad\square$

Proposition 23.12 opens the door toward using methods from symplectic geometry in the study of geodesics, and this is a fairly large topic in modern research. As a very simple illustration, we will now give a second proof of the result that compact Riemannian manifolds are geodesically complete. Recall that a map $\varphi : X \to Y$ between topological spaces is called **proper** if the preimage of every compact set is compact.

EXERCISE 23.15. Show that $H(q,p) = \frac{1}{2}\langle p, p \rangle$ is a proper function on T^*M if and only if the bundle metric $\langle \ , \ \rangle$ is (positive or negative) definite.

The exercise combines with the following result to give another proof of Theorem 23.10.

THEOREM 23.16. *On any symplectic manifold (W, ω) with a smooth proper function $H : W \to \mathbb{R}$, the flow of the Hamiltonian vector field X_H exists globally.*

PROOF. One of the fundamental properties of Hamiltonian systems is that energy is conserved: "energy" in this case means the value of the Hamiltonian, and this value does not change along flow lines of X_H since

$$dH(X_H) = -\omega(X_H, X_H) \equiv 0.$$

It follows that every flow line of X_H stays within a level set $H^{-1}(c) \subset W$ for some $c \in \mathbb{R}$, and that set is compact if H is proper, thus the flow line can be continued for all time. $\qquad\square$

24. Euclidean and non-Euclidean geometries

In this lecture we will look at three specific examples of Riemannian manifolds whose properties lend considerable intuition to the rest of the subject. Our main goal for now will be to understand the behavior of the geodesics on these three examples, and certain qualitative differences will become apparent when we do this. We will later see that these differences are symptomatic of the distinction between positive, negative and zero curvature.

A bit of preparation is necessary before we discuss the actual examples, mainly because as a second-order nonlinear differential equation, the geodesic equation is generally not so easy to solve. We will first develop some tools that—at least in fortunate situations—make it easier.

24.1. Notation: how to write down a pseudo-Riemannian metric. In local coordinates $x^1, \ldots, x^n$, a pseudo-Riemannian metric $g = \langle \ , \ \rangle$ on a manifold M is a type $(0,2)$ tensor field, and thus takes the form

$$g = g_{ij} \, dx^i \otimes dx^j,$$

where the components satisfy the relation $g_{ij} = g_{ji}$ since $\langle \ , \ \rangle$ is symmetric. One often sees this written in the form

$$g = \sum_{i \leqslant j} g_{ij} \, dx^i \, dx^j,$$

in which the summation avoids unnecessary repetition of matching components by using the abbreviation

$$(24.1) \qquad dx^i \, dx^j := \frac{1}{2} \left(dx^i \otimes dx^j + dx^j \otimes dx^i \right).$$

So for example, the Euclidean metric on $\mathbb{R}^2$ can now be written in Cartesian coordinates (x,y) as

$$g_E = dx^2 + dy^2,$$

while the metric on the Poincaré half-plane in Exercise 22.8 becomes

$$h = \frac{1}{y^2} \left(dx^2 + dy^2 \right).$$

On $\mathbb{R}^n$ in the standard coordinates $(x^1, \ldots, x^n)$, the Euclidean metric is now

$$g_E = (dx^1)^2 + \ldots + (dx^n)^2,$$

and changing some signs gives us the standard flat pseudo-Riemannian metric of signature (k, ℓ), written as

$$(dx^1)^2 + \ldots + (dx^k)^2 - (dx^{k+1})^2 - \ldots - (dx^n)^2.$$

EXERCISE 24.1. On $\mathbb{R}^2$ with coordinates (x, y), show that the pseudo-Riemannian metric $dx\, dy$ has signature $(1, 1)$, and find a new global coordinate system (s, t) in which it takes the form $ds^2 - dt^2$.

REMARK 24.2. Algebraically, (24.1) can be regarded as a **symmetric product**, which is analogous to the wedge product but without all the minus signs. On an arbitrary vector space V, one can define a commutative product with values in $V \otimes V$ by symmetrizing the usual tensor product, thus writing $vw := \frac{1}{2} (v \otimes w + w \otimes v)$. The values of this product belong to the subspace consisting of symmetric bilinear maps $V^* \times V^* \to \mathbb{F}$, or equivalently, the kernel of the projection $\mathrm{Alt} : V \otimes V \to \Lambda^2 V$. If you don't care so much about algebra, don't worry about this.

24.2. Isometries and conformal transformations. A diffeomorphism $\varphi : M \to N$ from one pseudo-Riemannian manifold (M, g) to another (N, h) is called an **isometry** if

$$\varphi^* h = g.$$

We say in this situation that (M, g) and (N, h) are **isometric**, and indicate that φ is an isometry by writing

$$\varphi : (M, g) \to (N, h).$$

In more concrete terms, the condition means

$$h_{\varphi(p)}(\varphi_* X, \varphi_* Y) = g_p(X, Y) \qquad \text{for all } p \in M \text{ and } X, Y \in T_p M,$$

so in other words, the derivative $\varphi_* : T_p M \to T_{\varphi(p)} N$ of φ at every point $p \in M$ preserves the scalar products on these tangent spaces. In the Riemannian case (i.e. when the scalar products are positive), this has a simple geometric interpretation: one defines the lengths $|X|, |Y| \geq 0$ and angle $\theta \in [0, \pi]$ between two vectors $X, Y \in T_p M$ in this case by

$$(24.2) \qquad |X| := \sqrt{\langle X, X \rangle}, \qquad |Y| := \sqrt{\langle Y, Y \rangle}, \qquad \theta = \arccos\left(\frac{\langle X, Y \rangle}{|X| \cdot |Y|} \right),$$

and preserving inner products thus means preserving lengths of tangent vectors and angles between them. It follows that a diffeomorphism is an isometry if and only if it preserves lengths of paths and angles between intersecting paths.

Isometry is the natural notion of equivalence in the category of pseudo-Riemannian manifolds, thus it preserves all meaingful notions that are defined in terms of pseudo-Riemannian metrics. For example, it preserves geodesics, i.e. if $\varphi : (M, g) \to (N, h)$ is an isometry, then a path $\gamma : (a, b) \to M$ is a geodesic if and only if $\varphi \circ \gamma : (a, b) \to N$ is a geodesic. One easy way to see this is via the energy functional from §22.4: it is straightforward to check that $E(\varphi \circ \gamma) = E(\gamma)$ for all paths γ in M, and that γ is therefore stationary for the energy functional on $C^\infty([a, b], M ; p, q)$ if and only if $\varphi \circ \gamma$ is stationary for the energy functional on $C^\infty([a, b], N ; \varphi(p), \varphi(q))$.

The set of all isometries $(M, g) \to (M, g)$ forms a group, denoted by

$$\mathrm{Isom}(M, g) \subset \mathrm{Diff}(M).$$

This group has the useful property that it maps geodesics to geodesics. On the other hand, one should not expect this group to be nontrivial in general, as preserving distances and angles turns out to be a very stringent condition on a diffeomorphism. One can show that $\mathrm{Isom}(M, g)$ is always a Lie group (see e.g. [**Kob95**]), so in particular, it is a smooth finite-dimensional manifold. The following result imposes an absolute upper bound on its dimension: if $\dim M = n$, then $\dim \mathrm{Isom}(M, g)$ can never be larger than

$$(24.3) \qquad\qquad n + \dim \mathrm{O}(k, \ell) = n + \frac{1}{2}(n-1)n = \frac{1}{2}n(n+1).$$

This is, namely, the dimension of M plus the dimension of the space of all linear maps $T_pM \to T_qM$ for two points $p, q \in M$ that preserve the scalar product.

THEOREM 24.3. *Suppose (M, g) is a connected pseudo-Riemannian manifold, $p, q \in M$ are two points and $X_1, \ldots, X_n \in T_pM$ and $Y_1, \ldots, Y_n \in T_qM$ are orthonormal bases. Then there exists at most one isometry $\varphi \in \mathrm{Isom}(M, g)$ such that*

$$\varphi(p) = q \qquad \text{and} \qquad \varphi_*X_i = Y_i \quad \text{for all } i = 1, \ldots, n.$$

PROOF. By looking at isometries of the form $\psi^{-1} \circ \varphi$, it is equivalent to show that the only isometry $f : (M, g) \to (M, g)$ satisfying $f(p) = p$ and $T_pf = \mathbb{1} : T_pM \to T_pM$ is the identity map. Since each geodesic through p is determined by its velocity at p, and f maps geodesics to geodesics, the condition $T_pf = \mathbb{1}$ implies that f is the identity map on the open neighborhood $\mathcal{U} \subset M$ of p consisting of all points that can be reached via geodesics from p. Now suppose $q \in M$. Since M is connected, we can find a continuous path $\gamma : [0, 1] \to M$ from $\gamma(0) = p$ to $\gamma(1) = q$, and the interval $[0, 1]$ can then be partitioned into a finite union of subintervals with end points

$$0 =: t_0 < t_1 < \ldots < t_{N-1} < t_N := 1$$

such that for each $j = 1, \ldots, N$, $\gamma(t_j)$ can be reached via a geodesic starting at $\gamma(t_{j-1})$. It follows that f is also the identity on a neighborhood of $\gamma(t_1)$ and therefore $T_{\gamma(t_1)}f = \mathbb{1}$. Repeating the same argument N times then extends this conclusion to a neighborhood of $\gamma(t_N) = q$, and since the point q was arbitrary, f is therefore the identity map everywhere. $\qquad\square$

REMARK 24.4. Theorem 24.3 guarantees uniqueness, but not existence, thus (24.3) shows the dimension of $\mathrm{Isom}(M, g)$ if there exist as many isometries as the theorem allows, but in general $\dim \mathrm{Isom}(M, g)$ may be smaller. Once we have proved the basic theorems about curvature, it will begin to seem obvious that $\mathrm{Isom}(M, g)$ should be trivial for "most" pseudo-Riemannian manifolds, as the existence of nontrivial isometries will imply conditions on the curvature that are not usually satisfied.

You may recall from linear algebra that for two inner product spaces V and W, every linear map $A : V \to W$ that preserves lengths of vectors automatically preserves the inner product, and therefore also angles: this follows from the bilinearity of the inner product after expanding the relation $\langle A(v+w), A(v+w) \rangle = \langle v+w, v+w \rangle$. As a consequence, every smooth distance-preserving map between Riemannian manifolds is necessarily an isometry, and thus also an angle-preserving map. The converse however is false:

LEMMA 24.5. *For two (positive) finite-dimensional inner product spaces V and W, a linear map $A : V \to W$ preserves angles if and only if there is a constant $c > 0$ such that $\langle Av, Aw \rangle = c\langle v, w \rangle$ for all $v, w \in V$.*

PROOF. It is clear from (24.2) that the condition $\langle Av, Aw \rangle = c\langle v, w \rangle$ implies angles are preserved. Conversely, if $A : V \to W$ preserves angles, then it maps any orthonormal basis of V to a set of the form $\lambda_1e_1, \ldots, \lambda_ne_n$ where the $\lambda_1, \ldots, \lambda_n$ are positive numbers and $e_1, \ldots, e_n$ is an

orthonormal basis of W. After fixing appropriate bases, we can therefore assume without loss of generality that $V = W = \mathbb{R}^n$, both endowed with the standard Euclidean inner product, and A is represented by a diagonal matrix with positive entries $\lambda_1, \ldots, \lambda_n$. Writing $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ for the standard basis, the orthogonal vectors $\mathbf{e}_i + \mathbf{e}_j$ and $\mathbf{e}_i - \mathbf{e}_j$ for any $i \neq j$ must then be mapped by A to two orthogonal vectors, implying

$$0 = \langle A(\mathbf{e}_i + \mathbf{e}_j), A(\mathbf{e}_i - \mathbf{e}_j) \rangle = \langle \lambda_i \mathbf{e}_i + \lambda_j \mathbf{e}_j, \lambda_i \mathbf{e}_i - \lambda_j \mathbf{e}_j \rangle = \lambda_i^2 - \lambda_j^2,$$

and thus $\lambda_1 = \ldots = \lambda_n =: \lambda$. This proves $\langle Av, Aw \rangle = \lambda \langle v, w \rangle$ for all v, w. $\qquad\square$

With this lemma in mind, a diffeomorphism $\varphi : M \to N$ is called a **conformal transformation** $(M, g) \to (N, h)$ if it satisfies

$$\varphi^* h = fg \quad \text{for some smooth function } f : M \to (0, \infty),$$

where we should emphasize that the function f need not be specified in advance. This condition means that for every $p \in M$ and $X, Y \in T_p M$,

$$h_{\varphi(p)}(\varphi_* X, \varphi_* Y) = f(p) \cdot g_p(X, Y),$$

hence in the Riemannian case, one can say that the linear map $\varphi_* : T_p M \to T_{f(p)} N$ preserves angles (but not necessarily lengths), and the conformal transformations are therefore regarded as precisely those diffeomorphisms that preserve all angles between intersecting curves. The set of conformal transformations $(M, g) \to (M, g)$ also forms a group, denoted by

$$\mathrm{Conf}(M, g) \subset \mathrm{Diff}(M),$$

and it contains $\mathrm{Isom}(M, g)$ since every isometry is also a conformal transformation. The converse is false in general: for instance, for the Euclidean metric $g_E = dx^2 + dy^2$ and the Poincaré metric $h = \frac{1}{y^2}(dx^2 + dy^2)$ on the upper half-plane $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$, the identity map $(\mathbb{H}, g_E) \to (\mathbb{H}, h)$ is conformal, but is not an isometry. This example also shows that conformal transformations do not preserve geodesics in general. (For the geodesics on $(\mathbb{H}, h)$, see Exercise 22.8.)

Conformal transformations arise naturally in complex analysis, due to the following exercise.

EXERCISE 24.6. Identify $\mathbb{C}$ with $\mathbb{R}^2$ via $x + iy \leftrightarrow (x, y)$ and endow it with the standard Euclidean metric. Show that a diffeomorphism $f : \mathcal{U} \to \mathcal{V}$ between two open subsets $\mathcal{U}, \mathcal{V} \subset \mathbb{C}$ is a conformal transformation if and only if it is either holomorphic or antiholomorphic (meaning the map $\bar{f} : \mathcal{U} \to \mathbb{C}$ is holomorphic). In particular, the group of orientation-preserving conformal transformations from an open region in $\mathbb{C}$ to itself is the same as its group of holomorphic automorphisms.

24.3. Pseudo-Riemannian submanifolds. Many interesting examples of Riemannian manifolds occur as hypersurfaces in flat space, so the question arises: if Σ is a submanifold of a pseudo-Riemannian manifold (M, g) whose geodesic flow we already understand, can we compute from it the geodesics on Σ? In fortunate cases this is possible, but there are a few subtleties to be aware of. First is the metric on Σ: we would obviously like to define it as the restriction of $g = \langle \, , \, \rangle$ to $T\Sigma \subset TM$, or equivalently, the pullback $j^* g \in \Gamma(T_2^0 \Sigma)$ via the inclusion map $j : \Sigma \hookrightarrow M$. This is fine if g is positive, because the restriction will then also satisfy $\langle X, X \rangle > 0$ for all nontrivial $X \in T\Sigma$, but in the indefinite case, the nondegeneracy of g does not immediately imply the same for its restriction $j^* g$. There is a simple exercise in linear algebra to be done before we continue.

Suppose V is a finite-dimensional real vector space and $\langle \, , \, \rangle$ is a nondegenerate symmetric bilinear form on V; recall that "nondegenerate" in this situation means the map $V \to V^* : v \mapsto \langle v, \cdot \rangle$ is an isomorphism. By analogy with the case of a positive-definite inner product, we can associate to any linear subspace $W \subset V$ its orthogonal "complement"

$$W^\perp := \left\{ v \in V \mid \langle v, w \rangle = 0 \text{ for all } w \in W \right\}.$$

We put the word "complement" in parentheses here because if $\langle\ ,\ \rangle$ is not positive-definite, there is no guarantee in general that W and $W^\perp$ will actually be complementary, i.e. they might have nontrivial intersection.

LEMMA 24.7. *For any finite-dimensional real vector space V with a nondegenerate symmetric bilinear form $\langle\ ,\ \rangle$ and a subspace $W \subset V$:*

(1) $\dim W + \dim W^\perp = \dim V$,
(2) $(W^\perp)^\perp = W$,
(3) *The restriction of $\langle\ ,\ \rangle$ to W is nondegenerate if and only if $W \cap W^\perp = \{0\}$, which is true if and only if $V = W \oplus W^\perp$.*

PROOF. Degeneracy of $\langle\ ,\ \rangle|_W$ means there exists a nontrivial vector $v \in W$ such that $\langle v, \cdot\rangle|_W = 0$, which is the same thing as saying $v \in W \cap W^\perp$. To show that $\dim W + \dim W^\perp = \dim V$, it suffices to view $W^\perp \subset V$ as the kernel of the linear map

$$V \to W^* : v \mapsto \langle v, \cdot\rangle|_W,$$

and observe that this map is surjective since every linear functional $\lambda : W \to \mathbb{R}$ can be extended to a linear functional on V and then presented as $\lambda = \langle v, \cdot\rangle$ for a unique $v \in V$, due to the nondegeracy of $\langle\ ,\ \rangle$ on V. Since $W \subset (W^\perp)^\perp$ by definition, this also implies $W = (W^\perp)^\perp$, since both subspaces have the same dimension. $\qquad\square$

DEFINITION 24.8. In a pseudo-Riemannian manifold (M, g), a submanifold $\Sigma \subset M$ with inclusion map $j : \Sigma \hookrightarrow M$ is called a **pseudo-Riemannian submanifold** if j^*g is nondegenerate, so that it defines a pseudo-Riemannian metric on Σ. We call Σ a **Riemannian submanifold** if j^*g is positive.

Lemma 24.7 implies:

COROLLARY 24.9. *A submanifold Σ in (M, g) is a pseudo-Riemannian submanifold if and only if for every $p \in \Sigma$, $T_pM = T_p\Sigma \oplus (T_p\Sigma)^\perp$.* $\qquad\square$

The condition in Corollary 24.9 is satisfied for *every* submanifold $\Sigma \subset M$ if (M, g) is a Riemannian manifold, but this is not true in the indefinite case. For example, light-like paths (see Remark 22.12) in a Lorentzian manifold (M, g) trace out smooth 1-dimensional submanifolds $\Sigma \subset M$, but (Σ, j^*g) is not a pseudo-Riemannian submanifold, as j^*g in this case vanishes.

REMARK 24.10. Lemma 24.7 remains true without significant changes if $\langle\ ,\ \rangle$ is assumed antisymmetric instead of symmetric, and this observation is important in symplectic geometry. In particular, an analogue of Corollary 24.9 holds for symplectic submanifolds of a symplectic manifold.

By Corollary 24.9, every pseudo-Riemannian submanifold $\Sigma \subset (M, g)$ comes with a well-defined **orthogonal projection**

$$\pi_\Sigma : TM|_\Sigma \to T\Sigma,$$

which projects each tangent space T_pM for $p \in \Sigma$ to $T_p\Sigma \subset T_pM$ along the complementary subspace $(T_p\Sigma)^\perp \subset T_pM$.

PROPOSITION 24.11. *If ∇ is the Levi-Cività connection on (M, g) and $\Sigma \subset M$ is a pseudo-Riemannian submanifold with inclusion $j : \Sigma \hookrightarrow M$, the Levi-Cività connection on (Σ, j^*g) is uniquely determined by the relation*

$$\widehat{\nabla}_X Y = \pi_\Sigma(\nabla_X Y), \qquad \text{for } p \in \Sigma,\ X \in T_p\Sigma \text{ and } Y \in \mathfrak{X}(M) \text{ with } Y(\Sigma) \subset T\Sigma.$$

PROOF. Any vector field on Σ near p can be extended to a vector field on M using a slice chart, thus the stated relation uniquely determines a connection on Σ if we can prove that the operator $\pi_\Sigma \circ \nabla_X$ satisfies the required Leibniz rule. And it does: for $f \in C^\infty(\Sigma)$, $Y \in \mathfrak{X}(\Sigma)$ and $X \in T_p\Sigma$, we extend f and Y arbitrarily to a smooth function and vector field respectively on M, and use the Leibniz rule for ∇ to compute

$$\pi_\Sigma\left(\nabla_X(fY)\right) = \pi_\Sigma\left((\mathcal{L}_X f)Y + f\nabla_X Y\right) = (\mathcal{L}_X f)\pi_\Sigma(Y) + f\pi_\Sigma(\nabla_X Y) = (\mathcal{L}_X f)Y + f\pi_\Sigma(\nabla_X Y),$$

where we have written $\pi_\Sigma(Y) = Y$ since $Y(\Sigma) \subset T\Sigma$. This proves that $\pi_\Sigma \circ \nabla$ defines a connection on Σ, and to see that it is also compatible with the restricted metric j^*g, we take two vector fields $Y, Z \in \mathfrak{X}(\Sigma)$, extend them smoothly to vector fields on M, and then use the fact that ∇ is compatible with g, plus the fact that $Y(p)$ and $Z(p)$ are both orthogonal to $(T_p\Sigma)^\perp$:

$$\mathcal{L}_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle = \langle \pi_\Sigma(\nabla_X Y), Z \rangle + \langle Y, \pi_\Sigma(\nabla_X Z) \rangle.$$

Finally, we observe that for vector fields $Y, Z \in \mathfrak{X}(M)$ with values in $T\Sigma$ along Σ, the Lie bracket $[Y, Z] \in \mathfrak{X}(M)$ necessarily also has this property, so the torsion of $\pi_\Sigma \circ \nabla$ at $p \in \Sigma$ is

$$\widehat{T}(Y(p), Z(p)) := \pi_\Sigma(\nabla_{Y(p)} Z) - \pi_\Sigma(\nabla_{Z(p)} Y) - [Y, Z](p) = \pi_\Sigma\left(\nabla_{Y(p)} Z - \nabla_{Z(p)} Y - [Y, Z](p)\right)$$
$$= \pi_\Sigma\left(T(Y(p), Z(p))\right) = 0,$$

since ∇ is symmetric. The result now follows from the uniqueness of the Levi-Cività connection. $\square$

COROLLARY 24.12. *Assume $\Sigma \subset (M, g)$ is a pseudo-Riemannian submanifold with inclusion $j : \Sigma \hookrightarrow M$, and ∇ denotes the Levi-Cività connection on (M, g). Then a path $\gamma : (a, b) \to \Sigma$ is a geodesic on (Σ, j^*g) if and only if $\nabla_t \dot{\gamma}(t)$ is orthogonal to $T_{\gamma(t)}\Sigma$ for all t.*

PROOF. According to Proposition 24.11, the geodesic equation on (Σ, j^*g) is $\pi_\Sigma(\nabla_t \dot{\gamma}) \equiv 0$. $\square$

24.4. Three examples of Riemannian manifolds.

24.4.1. *Euclidean space.* We have already mentioned that the Christoffel symbols on $M := \mathbb{R}^n$ with the Euclidean metric

$$g = g_E := (dx^1)^2 + \ldots + (dx^n)^2$$

vanish identically, thus the geodesic equation becomes $\ddot{\gamma} = 0$ and the geodesics are straight lines. You may think there is not much more to say about this example, but that didn't stop Euclid from writing a treatise about $(\mathbb{R}^2, g_E)$ that was regarded as the basis of Western mathematics for 2000 years. Here is a modern reformulation of the first two of Euclid's five postulates, on which all of his propositions about plane geometry are based:

(E1) For every pair of distinct points $p, q \in M$, there exists a unique geodesic segment $\gamma : [0, 1] \to M$ with $\gamma(0) = p$ and $\gamma(1) = q$.

(E2) Every geodesic in (M, g) exists for all time, i.e. (M, g) is geodesically complete.

Before continuing, let us mention another property of $(\mathbb{R}^n, g_E)$ that Euclid uses constantly without mentioning it, but that is actually a quite nontrivial property for a Riemannian manifold to have. The isometry group $\mathrm{Isom}(\mathbb{R}^n, g_E)$ is as large as possible, i.e. every isometry that is permitted by Theorem 24.3 actually exists. Indeed, the isometry group of $(\mathbb{R}^n, g_E)$ contains all the translations $\mathbf{x} \mapsto \mathbf{x} + \mathbf{v}$ by vectors $\mathbf{v} \in \mathbb{R}^n$, as well as the orthogonal transformations $\mathbf{A} \in \mathrm{O}(n)$, and one can combine these to produce a transformation that takes any given point p to another given point q while effecting an arbitrary rotation or reflection on their tangent spaces. In Euclid's argumentation, this fact is used for congruence proofs, e.g. two triangles in $\mathbb{R}^2$ are seen to be "the same" because one can be overlaid upon another, which means in modern terms that there is an isometry $\mathbb{R}^2 \to \mathbb{R}^2$ mapping one to the other. One can use this to justify the notion that "all right angles are the same," which is essentially the content of Euclid's fourth postulate (E4): if α_1, α_2 are two geodesics that intersect at a right angle at $\alpha_1(s_1) = \alpha_2(s_2) =: p$ and β_1, β_2 is another pair

of geodesics with a right-angle intersection at $\beta_1(t_1) = \beta_2(t_2) =: q$, then there exists an isometry sending $p \mapsto q$, $\alpha_1(\mathbb{R}) \mapsto \beta_1(\mathbb{R})$ and $\alpha_2(\mathbb{R}) \mapsto \beta_2(\mathbb{R})$. This property is the reason why angles in the plane can be measured and meaningfully compared, even if they appear at different points. Euclid's third postulate (E3) is not so much a *property* of $(\mathbb{R}^2, g_E)$ as a "recipe" for constructing circles, which in our n-dimensional context would mean spheres: for every pair of distinct points p, q, there is a unique "$(n-1)$-sphere" centered at p containing q, which we would define as

$$\left\{ \exp_p(X) \mid X \in T_p M \text{ such that } |X| = |X_0| \text{ where } \exp_p(X_0) = q \right\}.$$

Note that the vector $X_0 \in T_p M$ in this definition is unique due to the uniqueness of the geodesic segment in (E1), and $\exp_p(X)$ is always well defined due to (E2). The point of (E3) is that it gives rise to *constructive* arguments, e.g. Euclid's proposition on bisecting triangles provides not just the *existence* of bisections but an actual *recipe* to construct them with a ruler and compass.

The most famous of Euclid's postulates is the fifth, which is better known in a reformulation that was stated by John Playfair in 1795 and shown to be equivalent to Euclid's fifth postulate whenever the first four also hold:

(E5) For any geodesic $\gamma : \mathbb{R} \to M$ and a point $p \in M$ not on the image of γ, there exists at most one (up to parametrization) geodesic through p that does not intersect γ.

This is the **parallel postulate**, and historically, it has caused a lot of trouble. We'll come back to that shortly.

24.4.2. *Spheres.* The natural Riemannian metric on the unit sphere $S^n \subset \mathbb{R}^{n+1}$ is the one that it inherits by restriction from the Euclidean metric on $\mathbb{R}^{n+1}$. For the latter, the Levi-Cività connection ∇ is the trivial one, thus according to Corollary 24.12, a path $\gamma : (a, b) \to S^n$ is a geodesic in S^n if and only if

$$\ddot{\gamma}(t) \in (T_{\gamma(t)} S^n)^\perp = \mathbb{R}\gamma(t) \qquad \text{for all } t.$$

It is easy to find paths that have this property, e.g. for any $p \in S^n$ and $\mathbf{v} \in T_p S^n = p^\perp$ with $|\mathbf{v}| = 1$, the path

$$\gamma(t) = (\cos t)p + (\sin t)\mathbf{v} \in S^n \subset \mathbb{R}^{n+1}$$

is an example since $\ddot{\gamma}(t) = -\gamma(t) \in \mathbb{R}\gamma(t)$ for all t. Geodesics of this form exist for all $t \in \mathbb{R}$, and they can be chosen so that $\gamma(0) = p$ is an arbitrary point in S^n and $\dot{\gamma}(0) = \mathbf{v}$ is an arbitrary unit vector in $T_{\gamma(0)} S^n$. It follows that *all* geodesics on S^n are either paths of this form or (depending on their speed) reparametrizations of them: their images are the intersections of S^n with arbitrary 2-dimensional subspaces (spanned by the vectors $p, \mathbf{v} \in \mathbb{R}^{n+1}$), and are known as **great circles**.

Just like Euclidean space, the sphere S^n has the largest possible isometry group: any matrix in $O(n+1)$ defines a transformation $\mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ that preserves S^n. If $\mathbf{e}_1 \in \mathbb{R}^{n+1}$ denotes the first standard basis vector, then for any other $\mathbf{v} \in S^n$, one can find $\mathbf{A} \in O(n+1)$ with $\mathbf{A}\mathbf{e}_1 = \mathbf{v}$ by defining the columns of $\mathbf{A}$ to be any orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ with $\mathbf{v}_1 = \mathbf{v}$. This construction allows considerable freedom in the choice of $\mathbf{v}_2, \ldots, \mathbf{v}_{n+1}$, and this freedom is sufficient to realize any desired orthogonal transformation on the subspace $T_{\mathbf{v}} S^n = \mathbf{v}^\perp$.

Let's see how Euclid's axioms are doing. All the geodesics mentioned above are defined for all $t \in \mathbb{R}$, so (E2) is fine. There is a problem with (E1), though: while it is certainly possible to connect any two distinct points $p, q \in S^n$ by a geodesic segment, this segment is *never* unique: every geodesic on S^n is periodic, so you can always find another segment from p to q just by traversing the circle more times. In some cases you can find a lot more: for instance, antipodal points on S^2 are connected by an infinite family of geodesics, e.g. the longitudes that connect the north and south poles on the Earth. Another consequence of this ambiguity is that a geodesic from p to q is definitely *not* always the shortest path on S^n from p to q, nor must it be a local minimum of the length functional: if you imagine for instance a path that traverses most of a great

circle in order to move from p to a nearby point q, it is easy to find non-geodesic paths nearby that are shorter. We proved in §22.4 that geodesics are stationary (i.e. critical points) for the length functional, but indeed, not every critical point must be a local minimum.

On S^2, the parallel postulate is true for a stupid reason: no two geodesics are parallel, i.e. they always must intersect! In summary, classical geometry on S^2 is an interesting subject, but it has very little to do with Euclid's postulates.

24.4.3. *Hyperbolic space.* The third example gives a reason to care about indefinite metrics even if you have no interest in physics and really just want to understand Riemannian manifolds. The idea is to do the same thing as in the previous subsection, but with the Euclidean metric on $\mathbb{R}^{n+1}$ replaced by a metric with Lorentz signature: we will call it the **Minkowski metric**, and write it in coordinates $X = (\tau, x^1, \ldots, x^n) = (\tau, \mathbf{x}) \in \mathbb{R}^{n+1}$ as

$$g_M := -d\tau^2 + (dx^1)^2 + \ldots + (dx^n)^2.$$

The sphere was obtained as a regular level set for the Euclidean metric, but using the Minkowski metric instead gives a hyperboloid:

$$\left\{ X \in \mathbb{R}^{n+1} \mid \langle X, X \rangle = -1 \right\} = \left\{ (\tau, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^n \mid \tau^2 - |\mathbf{x}|^2 = 1 \right\}.$$

This hypersurface has two connected components, distinguished by the conditions $\tau \geqslant 1$ and $\tau \leqslant -1$, so we pick one of them to define a connected n-manifold called **hyperbolic n-space**

$$H^n := \left\{ X = (\tau, \mathbf{x}) \in \mathbb{R}^{n+1} \mid \tau^2 - |\mathbf{x}|^2 = 1 \text{ and } \tau > 0 \right\}.$$

We claim that this is in fact a *Riemannian* submanifold of $(\mathbb{R}^{n+1}, g_M)$, i.e. the restriction of the Minkowski metric to H^n is positive-definite. To see this, note that as (a component of) a regular level set of the function $f(X) := \langle X, X \rangle$, the tangent space to H^n at any point $p \in H^n$ is the kernel of $Df(p) : \mathbb{R}^{n+1} \to \mathbb{R}$, where the latter is $Df(p)Y = 2\langle p, Y \rangle$, hence

$$T_p H^n = p^\perp \subset \mathbb{R}^{n+1}.$$

One needs to be careful not to use too much Euclidean intuition in reading equations like this: the symbol $\perp$ in this case is defined relative to the Minkowski metric, which is indefinite, so it is not even automatic that $p \notin p^\perp$. On the other hand, the Minkowski inner product is negative (and therefore nondegenerate) on the 1-dimensional subspace spanned by p, so it follows from Lemma 24.7 that $\mathbb{R}p \oplus p^\perp = \mathbb{R}^{n+1}$. Since $\mathbb{R}p = (T_p H^n)^\perp$, Corollary 24.12 then implies that $H^n \subset (\mathbb{R}^{n+1}, g_M)$ is a pseudo-Riemannian submanifold. Its signature can be deduced from the fact that g_M has signature $(n, 1)$ and is negative on $(TH^n)^\perp$: this is only possible if g_M restricts positively to TH^n. We therefore have a natural Riemannian metric on H^n.

REMARK 24.13. You may have wondered why we defined H^n as a component of the level set with $\langle X, X \rangle = -1$ instead of $\langle X, X \rangle = 1$, as the latter might have seemed more obviously analogous to the sphere. The reason is that we specifically wanted a *Riemannian* submanifold: the hyperboloid $\langle X, X \rangle = 1$ is also a pseudo-Riemannian submanifold, one that even has the advantage of being connected, but it has signature $(n - 1, 1)$.

What are the geodesics? Here it is useful to note that the Levi-Cività connection ∇ on Minkowski space is the same one as on Euclidean space: it is the trivial connection, as is true for every pseudo-Riemannian metric with constant coefficients. One can then write down the geodesics on H^n in almost exactly the same way as on S^n, the trick is just to replace cos and sin by their hyperbolic counterparts. Given any $p \in H^n$ and $\mathbf{v} \in T_p H^n = p^\perp \subset \mathbb{R}^{n+1}$ with $|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = 1$, the path

$$\gamma(t) := (\cosh t)p + (\sinh t)\mathbf{v} \in \mathbb{R}^{n+1}$$

satisfies

$$\langle\gamma(t),\gamma(t)\rangle = \langle(\cosh t)p + (\sinh t)\mathbf{v}, (\cosh t)p + (\sinh t)\mathbf{v}\rangle = (\cosh^2 t)\langle p, p\rangle + (\sinh^2 t)\langle\mathbf{v},\mathbf{v}\rangle$$
$$= -\cosh^2 t + \sinh^2 t = -1,$$

so it lies in H^n, and its image is the intersection of H^n with the 2-dimensional subspace of $\mathbb{R}^{n+1}$ spanned by p and $\mathbf{v}$. Moreover,

$$\ddot{\gamma}(t) = \gamma(t) \in (T_{\gamma(t)}H^n)^{\perp},$$

so Corollary 24.12 implies that γ is a geodesic. Since $\gamma(0) = p \in H^n$ and $\dot{\gamma}(0) = \mathbf{v} \in T_pH^n$ can each be chosen arbitrarily (subject to the condition $|\mathbf{v}| = 1$), every geodesic in H^n is a reparametrization of one of these.

And the isometries? The group of linear transformations on $\mathbb{R}^{n+1}$ preserving the Minkowski metric is the Lorentz group $\mathrm{O}(n,1)$, and its transformations preserve the submanifold $H^n \subset \mathbb{R}^{n+1}$. Analogously to the action of $\mathrm{O}(n+1)$ on S^n, one can show that there is a Lorentz transformation sending any point in H^n to any other one, while realizing any desired rotation or reflection on the tangent spaces. The isometry group of H^n is therefore as large as possible: in particular, for any two geodesics on H^n with the same speed, there exists an isometry identifying one with the other.

The **hyperbolic plane** H^2 made a splash when it was first discovered in the 19th century. The reason has to do with Euclid's postulates: H^2 satisfies the first four, so a large portion of Euclid's propositions on congruence, bisection of triangles etc. works just as well in hyperbolic as in Euclidean geometry. But not the fifth postulate:

EXERCISE 24.14. Find a pair of intersecting geodesics on H^2 and a third geodesic that intersects neither of them.

The parallel postulate was always perceived to be a less obviously "fundamental" statement than Euclid's first four postulates, and the belief remained popular for 2000 years after Euclid that it should be possible to deduce it logically from the other four, if only one could find the right argument. Several illustrious figures even claimed at various times to have achieved this, though their proofs invariably turned out to rely on unjustified intuitive assumptions that do not follow from the first four postulates. (For more on this history, see [**Lee13b**].) The example of the hyperbolic plane revealed finally that this effort was fruitless: the fifth postulate cannot be deduced from the other four, because there exists a geometry that satisfies those four but not the fifth.

EXERCISE 24.15. Let $B^n \subset \mathbb{R}^n$ denote the open ball of radius 1. There is a natural diffeomorphism $\varphi : B^n \to H^n$ defined via *stereographic projection*, which means the following: for $\mathbf{x} \in B^n$, define $\varphi(\mathbf{x}) \in H^n$ as the unique intersection of H^n with the line in $\mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n$ that passes through the points $(-1, 0)$ and $(0, \mathbf{x})$. The pullback $\varphi^* g_M$ thus defines a Riemannian metric on B^n making it isometric to H^n. Prove $\varphi^* g_M$ is related to the Euclidean metric $g_E = (dx^1)^2 + \ldots + (dx^n)^2$ by

$$\varphi^* g_M = \frac{4}{(1 - |\mathbf{x}|^2)^2} g_E.$$

This is called the *Poincaré disk model* of hyperbolic space.

The Poincaré disk model in Exercise 24.15 reveals that hyperbolic space is conformally flat, i.e. the metric $\varphi^* g_M$ on B^n defines the same notion of angles as the Euclidean metric. This observation becomes especially useful in the case $n = 2$, where we can use the bijection $\mathbb{R}^2 \ni (x, y) \leftrightarrow x + iy =: z \in \mathbb{C}$ to identify B^2 with

$$D := \left\{ z \in \mathbb{C} \mid |z| < 1 \right\}$$

and write $\varphi^* g_M = \frac{4}{(1 - |z|^2)^2} \left(dx^2 + dy^2 \right)$.

EXERCISE 24.16. The classical *Cayley transform* is the holomorphic map $f(z) := \frac{z-i}{z+i}$, which defines a conformal transformation from the open upper half-plane $\mathbb{H} := \left\{ x + iy \in \mathbb{C} \mid y > 0 \right\}$ to D. Prove

$$f^* \varphi^* g_M = \frac{1}{y^2} \left(dx^2 + dy^2 \right),$$

hence the Poincaré half-plane from Exercise 22.8 is another model of the hyperbolic plane.

By a standard theorem in complex analysis, the group of holomorphic automorphisms of the disk $D \subset \mathbb{C}$ consists of all maps of the form

$$z \mapsto e^{i\theta} \frac{z - a}{1 - \bar{a}z}, \qquad \text{for any } \theta \in \mathbb{R},\, a \in D.$$

Or if you prefer the Poincaré half-plane model, the holomorphic automorphisms of $\mathbb{H}$ are the fractional linear transformations

$$z \mapsto \frac{az + b}{cz + d}, \qquad \text{for any } a, b, c, d \in \mathbb{R} \text{ with } ad - bc = 1.$$

These are two alternate perspectives on the same thing, and in either case, we have a 3-dimensional group of conformal transformations, containing exactly one that maps any given point to any other given point while also realizing any desired rotation. But since both of these are isometric to the hyperbolic plane, we can say the same thing about the orientation-preserving isometries: all of the latter are of course conformal transformations, and they are therefore *all* of the conformal transformations. This proves a rather surprising fact about the hyperbolic plane:[68]

THEOREM 24.17. *On H^2, every conformal transformation is an isometry.* $\square$

This result plays a fundamental role in the theory of Riemann surfaces, due to the fact that choosing a complex structure on a surface is equivalent to choosing an orientation and a conformal structure, i.e. a conformal equivalence class of metrics. It implies that outside of a finite set of exceptions, the category of Riemann surfaces is essentially equivalent to the category of oriented surfaces with hyperbolic metrics, so that results from 2-dimensional Riemannian geometry have nontrivial consequences for complex 1-manifolds.

One of the standard theorems derivable from Euclid's five postulates is that the sum of the angles in every triangle is π. This is one of the things you lose if you remove the fifth postulate:

EXERCISE 24.18. Using whichever model you prefer, show that for any $\epsilon > 0$, H^2 contains a compact region bounded by three geodesics, each intersecting each of the others exactly once, such that the sum of the angles at the three intersections is less than ϵ.

25. Integrability and the Frobenius theorem

In this lecture we begin talking about curvature: we will consider first the setting of a general vector bundle with an arbitrary connection, and once this is understood, specialize to the tangent bundle of a pseudo-Riemannian manifold with the Levi-Cività connection. We assume as usual that

$$\pi : E \to M$$

is a smooth vector bundle, and the symbol ∇ will always mean a connection on this bundle.

---

[68]There was no time to mention Theorem 24.17 in the lecture, so it is included here only for information.

25.1. Flat sections and connections. One can motivate the topic of curvature by asking three questions whose answers in the setting of ordinary differentiation (i.e. for the trivial connection on a trivial bundle) are either obvious or are well-known results from first-year analysis. The answers turn out to be much less obvious for an arbitrary connection ∇ on E.

QUESTIONS 25.1. *Choose any point $p \in M$ in the base of the vector bundle $\pi : E \to M$.*

(1) *Given $v \in E_p$, is v the value at p of any* parallel *section $s : \mathcal{U} \to E$ defined on a neighborhood $\mathcal{U} \subset M$ of p , i.e. a section satisfying $\nabla s \equiv 0$?*
(2) *Does p have a neighborhood $\mathcal{U} \subset M$ on which for every smooth path $\gamma : [0,1] \to \mathcal{U}$ with $\gamma(0) = \gamma(1) = p$, the parallel transport map $P_\gamma^1 : E_p \to E_p$ is the identity?*
(3) *Given a coordinate chart $(x^1, \ldots, x^n)$ on a neighborhood of p, do the partial covariant derivative operators $\nabla_i := \nabla_{\frac{\partial}{\partial x^i}}$ and $\nabla_j := \nabla_{\frac{\partial}{\partial x^j}}$ for $i \neq j$ commute at p?*

The answer to all three questions is clearly yes if ∇ is the trivial connection with respect to some local trivialization of E near p. This is always the case if $\dim M = 1$, in particular, since p then has a neighborhood parametrized by a path, so parallel transport along that path can be used to define a trivialization in which the parallel sections are represented by constant functions, hence ∇ is the trivial connection. But for $\dim M \geq 2$, we will see that the answer to all three questions is no in general.

If you think of parallel sections as the generalization to vector bundles of the notion of a constant function, then it seems surprising at first that there might not exist one on any neighborhood of a point. Of course, parallel sections *along a path* do always exist; we get them from parallel transport. But if $\dim M \geq 2$ so that no neighborhood of p can be parametrized by a single path, then the effort to find a parallel section runs into trouble precisely because the answer to question 25.1(2) might be no: if a parallel section $s : \mathcal{U} \to E$ on some neighborhood $\mathcal{U} \to M$ of p exists with any given value $s(p) = v$, then paths $\gamma : [0,1] \to \mathcal{U}$ will satisfy $P_\gamma^t(v) = s(\gamma(t))$ for every t, and parallel transport along a *loop* in $\mathcal{U}$ therefore always brings us back to v. But we've already seen an example where the latter is impossible: parallel transport using the Levi-Cività connection on $TS^2 \to S^2$ along certain closed "triangular" paths in S^2 does not produce the identity map; see Figure 8 in Lecture 19. (You've learned in the mean time that the edges of the triangle in that picture are geodesic segments, and you could then deduce from the compatibility of the Levi-Cività connection with the metric that the vector field drawn along these edges really is parallel.) It follows that no parallel vector field exists on any neighborhood of that triangle.

REMARK 25.2. We intentionally phrased all three of the questions in 25.1 so that they are *local* in nature, i.e. they depend on the connection only in an arbitrarily small neighborhood of p. This is the one problem with Figure 8, since the triangle in that picture cannot be called a "small" neighborhood of anything. The reason to focus only on neighborhoods of a point is that for arbitrary paths $\gamma : [0,1] \to M$ with $\gamma(0) = \gamma(1)$ in a manifold M, it might happen for topological reasons that P_γ^1 is not the identity map even if local parallel sections always exist (see e.g. Exercise 25.6 below). One can show however (see Exercise 25.7) that if local parallel sections always exist, then P_γ^1 depends only on the homotopy class of γ. From this fact we can still conclude via Figure 8 that local parallel vector fields cannot always exist on S^2, because S^2 is simply connected, so the loop in the picture is homotopic to a constant loop (which would of course give the identity as a parallel transport map).

We now give some formal definitions. We will continue to use the word **parallel** to describe any section $s : \mathcal{U} \to E$ on an open subset $\mathcal{U} \to E$ such that $\nabla s \equiv 0$. The terms **flat**, **horizontal** and **covariantly constant** are sometimes used as synonyms for "parallel" when applied to sections.

DEFINITION 25.3. A connection ∇ on the bundle $E \to M$ is called **flat** if for every $p \in M$ and $v \in E_p$, there exists a neighborhood $\mathcal{U} \subset M$ of p and a flat section $s \in \Gamma(E|_\mathcal{U})$ with $s(p) = v$.

PROPOSITION 25.4. *A connection ∇ on $E \to M$ is flat if and only if every point $p \in M$ has a neighborhood $\mathcal{U} \subset M$ with a local trivialization $\Phi : E|_\mathcal{U} \to \mathcal{U} \times \mathbb{F}^m$ in which ∇ looks like the trivial connection (see Example 20.1).*

PROOF. In one direction this is obvious, since the trivial connection clearly admits flat sections (they look constant in the trivialization). Conversely, if ∇ is flat, then for any $p \in M$, we can choose a basis $v_1, \ldots, v_m$ of E_p and flat sections $e_1, \ldots, e_m \in \Gamma(E|_\mathcal{U})$ on some neighborhood $\mathcal{U} \subset M$ of p such that $e_i(p) = v_i$ for $i = 1, \ldots, m$; after possibly shrinking the neighborhood $\mathcal{U}$, we can assume that these also span the fiber E_q for every $q \in \mathcal{U}$, thus they form a frame for E over $\mathcal{U}$. Writing an arbitrary section $s \in \Gamma(E)$ on $\mathcal{U}$ in terms of its components as $s = s^i e_i$ with respect to the frame $e_1, \ldots, e_m$, the Leibniz rule then gives

$$\nabla_X s = ds^i(X) e_i(q) + s^i(q) \nabla_X e_i = ds^i(X) e_i(q) \qquad \text{for every } q \in \mathcal{U}, \, X \in T_q M,$$

showing that the covariant derivative is represented in this frame by the differentials of the components. This means that $e_1, \ldots, e_m$ corresponds to a local trivialization in which ∇ is the trivial connection. $\qquad\square$

It follows from Proposition 25.4 that for any flat connection, the answers to questions 25.1(2) and (3) are both affirmative.

EXERCISE 25.5. Prove that if $\dim M = 1$, then every connection on $E \to M$ is flat.

EXERCISE 25.6. Recall the nontrivial real line bundle $\ell \to S^1$ in Example 16.23. Exercise 25.5 implies that any connection ∇ on $\ell \to S^1$ is flat since $\dim S^1 = 1$. Show however that for a path $\gamma : [0, 1] \to S^1$ that winds once around the circle and ends at its starting point $\gamma(1) = \gamma(0) =: p$, $P_\gamma^1 : \ell_p \to \ell_p$ can never be the identity map.
Hint: This has to do with the fact that $\ell \to S^1$ is a non-orientable bundle.
Remark: The nontriviality of P_γ^1 in this example is detecting a topological property of $\ell \to S^1$ that has nothing to do with the connection. This is why we confine the loop in Question 25.1(2) to an arbitrarily small neighborhood of a point instead of allowing arbitrary loops.

EXERCISE 25.7. Suppose ∇ is a flat connection on $E \to M$.
  (a) Show that for any smooth map $f : N \to M$, the pullback of ∇ to a connection on $f^* E \to N$ is also flat.
  (b) Show that if $\{\gamma_s : [0, 1] \to M\}_{s \in [0,1]}$ is a smooth family of paths with fixed end points $\gamma_s(0) = p$ and $\gamma_s(1) = q$ for all $s \in [0, 1]$, then the maps $P_{\gamma_0}^1, P_{\gamma_1}^1 : E_p \to E_q$ are identical.
      Hint: Write $h(s, t) := \gamma_s(t)$ and use the fact that the pullback connection on $h^* E \to [0, 1] \times [0, 1]$ is also flat. Can you construct a global flat section of $h^* E$, and if so, how does it behave on the subsets $[0, 1] \times \{0\}$ and $[0, 1] \times \{1\}$?[69]

EXERCISE 25.8. Prove:
  (a) If ∇ a connection on $E \to M$ and $(\mathcal{U}_\alpha, \Phi_\alpha)$ and $(\mathcal{U}_\beta, \Phi_\beta)$ are two overlapping local trivializations in which ∇ looks like the trivial connection, then the transition functions relating these two trivializations are locally constant.
      Hint: Think in terms of local frames that are built out of flat sections. If $v = v^i e_i$ where $\nabla v \equiv 0$ and $\nabla e_i \equiv 0$ for every i, what can you conclude from the Leibniz rule?

---

[69]For the purposes of Exercise 25.7, you are safe in pretending that $[0, 1] \times [0, 1]$ is a smooth manifold, rather than something exotic like a "manifold with boundary and corners". If this worries you, assume that the family of paths $\gamma_s : [0, 1] \to M$ is defined for $s \in \mathbb{R}$ instead of just $s \in [0, 1]$; this does not change the situation in any significant way.

(b) Show that for any finite subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$, every G-structure on $E \to M$ naturally determines a flat connection, and conversely, if M is compact, then every flat connection on $E \to M$ arises in this way from a G-structure for some finite subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$.

Exercise 25.8 makes the existence of a flat connection seem like a rather restrictive condition: one would not expect the structure group of a vector bundle to be reducible in every case to a finite subgroup. Our main goal in this lecture is to formulate precise conditions for identifying whether a connection is flat. Along the way, we will be able to solve a related problem which is of independent interest and has nothing intrinsically to do with bundles: it leads to the theorem of Frobenius on integrable distributions.

25.2. Integrable frames. The word *integrability* has a variety of meanings in different contexts. Generally it refers to questions in which one is given some data of a linear nature, and would like to find some nonlinear data which produce the given linear data as a form of "derivative". The problem of finding antiderivatives of a smooth function f on $\mathbb{R}$ is the simplest example: it can always be solved (at least in principle) by writing down an antiderivative as a definite integral of f, and is thus not very interesting for the present discussion. A more interesting example is the generalization of this question to higher dimensions, which we examined in Lecture 13:

QUESTION 25.9. *Given a k-form ω on an n-manifold M, under what conditions is ω locally the exterior derivative of a $(k-1)$-form?*

Including the word "locally" in this question removes topological issues from the discussion: we've seen for instance that certain 1-forms λ on S^1 cannot be differentials of functions because $\int_{S^1} \lambda \ne 0$, but that is a symptom of the fact that the topological invariant $H^1_{\mathrm{dR}}(S^1)$ is nontrivial, and does not stop every 1-form $\lambda \in \Omega^1(S^1)$ from being presentable on a neighborhood $\mathcal{U} \subset S^1$ of any given point as df for some function $f : \mathcal{U} \to \mathbb{R}$. The answer to the question comes of course from the Poincaré lemma, which states that the "integrability condition"

$$d\omega = 0$$

is not only necessary but also sufficient for ω to admit local primitives.

Here is another integrability question whose answer will have some important applications.

QUESTION 25.10. *Suppose M is an n-manifold and $X_1, \ldots, X_n$ is a frame for TM over some open subset $\mathcal{U} \subset M$. Under what conditions does there exist for every point $p \in \mathcal{U}$ a chart $(\mathcal{U}', (x^1, \ldots, x^n))$ with $p \in \mathcal{U}' \subset \mathcal{U}$ such that $X_i = \frac{\partial}{\partial x^i}$ on $\mathcal{U}'$ for every $i = 1, \ldots, n$?*

In other words, every chart naturally gives rise to a local frame for TM, but we want to know when this process can be reversed: when can a frame for TM be "upgraded" locally to a chart?

Let's start with some good news: the answer in the case $n = 1$ is *always*. Indeed, the assumption in this case is that M is a 1-manifold and X_1 is a nowhere-zero vector field on some open subset $\mathcal{U} \subset M$, so a suitable chart $(\mathcal{U}', x)$ on some neighborhood $\mathcal{U}' \subset \mathcal{U}$ of any given point $p \in \mathcal{U}$ can be defined in terms of any local solution $\gamma : (-\epsilon, \epsilon) \to M$ to the initial value problem

$$\dot{\gamma}(t) = X(\gamma(t)), \qquad \gamma(0) = p,$$

namely $\mathcal{U}' := \gamma((-\epsilon, \epsilon)) \subset M$ and $x := \gamma^{-1} : \mathcal{U}' \to \mathbb{R}$. This example provides further justification for the term "integrability": solving an ordinary differential equation is sometimes referred to as *integrating* the equation, and since every ODE admits unique local solutions, every nowhere-zero vector field X_1 on a 1-manifold is integrable in this sense. More generally, it is reasonable to call a local frame $X_1, \ldots, X_n$ for TM *integrable* if it arises from a chart as described in Question 25.10.

It is easy to see on the other hand that for $n \geqslant 2$, not every local frame for TM is integrable, and the Lie bracket gives an obvious obstruction. Indeed, the coordinate vector fields induced by

a single chart always commute with each other (see Example 6.6), so $X_1, \ldots, X_n$ clearly cannot be coordinate vector fields, even in arbitrarily small neighborhoods of any given point p, unless $[X_i, X_j] \equiv 0$ for every $i, j = 1, \ldots, n$. One can easily find local frames that do not satisfy this condition, e.g. on $\mathbb{R}^2$ with coordinates (x, y), $(\partial_x, f\partial_x + g\partial_y)$ defines a frame for $T\mathbb{R}^2$ whenever $f, g : \mathbb{R}^2 \to \mathbb{R}$ are smooth functions with g never vanishing, but using Exercise 6.2, one finds

$$[\partial_x, f\partial_x + g\partial_y] = (\partial_x f)\partial_x + (\partial_x g)\partial_y,$$

which does not vanish unless $f(x, y)$ and $g(x, y)$ are both independent of x.

The really good news is that the condition on vanishing Lie brackets is not just necessary, but also sufficient:

THEOREM 25.11. *Suppose $X_1, \ldots, X_n \in \mathfrak{X}(M)$ are vector fields that all commute with each other. Then for any $p \in M$ at which $X_1(p), \ldots, X_n(p)$ form a basis of T_pM, there exists a chart $(\mathcal{U}, x)$ on M with $p \in \mathcal{U}$ such that $X_i = \frac{\partial}{\partial x^i}$ on $\mathcal{U}$ for every $i = 1, \ldots, n$.*

PROOF. For sufficiently small $\epsilon > 0$, we can use the flows of the vector fields $X_1, \ldots, X_n$ to define a smooth map

$$\psi : (-\epsilon, \epsilon)^n \to M : (t^1, \ldots, t^n) \mapsto \varphi_{X_1}^{t^1} \circ \ldots \circ \varphi_{X_n}^{t^n}(p).$$

By Theorem 6.9, the condition $[X_i, X_j] \equiv 0$ implies that the flows $\varphi_{X_i}^s$ and $\varphi_{X_j}^t$ commute with each other, thus for each $j \in \{1, \ldots, n\}$, one can reorder the composition of flows in the above definition so that $\varphi_{X_j}^{t^j}$ comes first, in which case the definition of the flow gives

$$\partial_j \psi(t^1, \ldots, t^n) = X_j(\psi(t^1, \ldots, t^n)).$$

Since $\psi(0, \ldots, 0) = p$, and the vectors $\partial_1 \psi(0, \ldots, 0), \ldots, \partial_n \psi(0, \ldots, 0)$ form a basis of T_pM, Lemma 4.2 implies that after possibly shrinking $\epsilon > 0$, ψ is the inverse of a chart on some neighborhood of p. That chart is the one we were looking for. $\square$

25.3. Integrability of distributions. We now return to the question of how to identify when a connection ∇ on the bundle $E \to M$ is flat. From a geometric perspective, a section $s : \mathcal{U} \to E$ over some open subset $\mathcal{U} \subset M$ can be characterized purely in terms of its image

$$\Sigma := s(\mathcal{U}) \subset E,$$

which is a submanifold of the total space E having exactly one intersection point with each of the fibers $E_p \subset E$ for $p \in \mathcal{U}$. The condition $\nabla s \equiv 0$ then holds if and only if this submanifold is always tangent to the horizontal subbundle $HE \subset TE$ determined by the connection, that is,

$$T_v\Sigma = H_vE \qquad \text{for all } v \in \Sigma.$$

With this picture in mind, we can now reframe the flatness question in a somewhat wider context.

DEFINITION 25.12. A smooth k-dimensional **distribution** on a manifold M is a smooth subbundle $\xi \subset TM$ of rank k. It is also sometimes called a k-**plane field**. Given such a distribution, an **integral submanifold** for ξ is a smooth k-dimensional submanifold $\Sigma \subset M$ such that

$$T_p\Sigma = \xi_p \qquad \text{for all } p \in \Sigma.$$

The distribution ξ is called **integrable** if for every point $p \in M$, ξ has an integral submanifold containing p.

Since we will not consider non-smooth distributions in this course, we will usually omit the word "smooth" and just refer to them as "distributions". Note that integral submanifolds do not need to be large in any sense, i.e. noncompact submanifolds diffeomorphic to a k-ball are fine, so

the integral submanifold through $p \in M$ may be contained in an arbitrarily small neighborhood of p, and in this sense integrability of a distribution is a purely local condition.

Thinking in terms of distributions and integral submanifolds makes possible a slight reformulation of our goal:

PROPOSITION 25.13. *A connection on a vector bundle $E \to M$ is flat if and only if its horizontal subbundle $HE \subset TE$ is an integrable distribution on the total space E.* $\square$

EXAMPLE 25.14. For any vector bundle $\pi : E \to M$, the vertical subbundle $VE \subset TE$ is also a distribution on the total space E, and it is *always* integrable. Indeed, the integral submanifolds of VE are the fibers of $\pi : E \to M$, and there is indeed one through every point.

The integrability problem for distributions bears several similarities to the frames considered in the previous section. One is that the 1-dimensional case is trivial: every 1-dimensional distribution (in a manifold of arbitrary dimension) is integrable. To see this on a neighborhood of any given point $p \in M$, one need only choose a vector field $X \in \mathfrak{X}(M)$ that is nonzero at p and takes values in ξ near p, as the flow lines of that vector field then trace out integral submanifolds of ξ, one of which passes through p. For a k-dimensional distribution $\xi \subset TM$ with $k \geqslant 2$, however, it is harder to see why integral submanifolds should exist, and in general they don't. Figure 9 for instance shows a 2-dimensional distribution on $\mathbb{R}^3$ consisting of 2-planes that "twist" in a way that would seem to prevent any surface from being tangent to them at every point. As with the frames in §25.2, there is in fact a necessary condition that can be stated easily, and it involves the Lie bracket:

LEMMA 25.15. *If $\xi \subset TM$ is an integrable distribution, then for every pair of vector fields $X, Y \in \mathfrak{X}(M)$ that take their values in ξ, the bracket $[X, Y] \in \mathfrak{X}(M)$ also takes its values in ξ.*

PROOF. Choose any point $p \in M$ and suppose $\Sigma \subset M$ is an integral submanifold containing p. Since $T_q\Sigma = \xi_q$ for all $q \in \Sigma$, vector fields $X, Y \in \mathfrak{X}(M)$ with values in ξ then define vector fields on Σ by restriction, and $[X|_\Sigma, Y|_\Sigma]$ is then (obviously) also a vector field on Σ, which necessarily matches the restriction of $[X, Y] \in \mathfrak{X}(M)$ to Σ. (You can check this by applying the operators $\mathcal{L}_X$ and $\mathcal{L}_Y$ to arbitrary smooth functions on M and their restrictions to Σ.) It follows that $[X, Y](p) \in T_p\Sigma = \xi_p$, and since p was chosen arbitrarily, $[X, Y](q)$ is therefore in ξ_q for every $q \in M$. $\square$

The best integrability theorems are those for which the obviously necessary condition is also sufficient, and that turns out to be the case here as well. The result is known as the *Frobenius integrability theorem*.

THEOREM 25.16 (Frobenius). *A distribution $\xi \subset TM$ on M is integrable if and only if for every pair of vector fields $X, Y \in \mathfrak{X}(M)$ taking values in ξ, $[X, Y] \in \mathfrak{X}(M)$ also takes values in ξ.*

The easy direction of this theorem is Lemma 25.15 above. To prove the converse, it will be more convenient at first to consider the special case where our manifold is the total space of a vector bundle $\pi : E \to M$ and the distribution is a horizontal subbundle $HE \subset TE$, meaning any subbundle of TE that is complementary to the vertical subbundle,

$$TE = VE \oplus HE.$$

We will not assume any more than this, so in particular, HE does not need to satisfy the other condition in our first definition of a connection (Definition 19.4), which was meant to guarantee that the resulting parallel transport maps are linear. As in Lemma 19.1, HE determines horizontal lift isomorphisms

$$\mathrm{Hor}_v : T_{\pi(v)}M \xrightarrow{\cong} H_vE \subset T_vE \qquad \text{for each } v \in E,$$
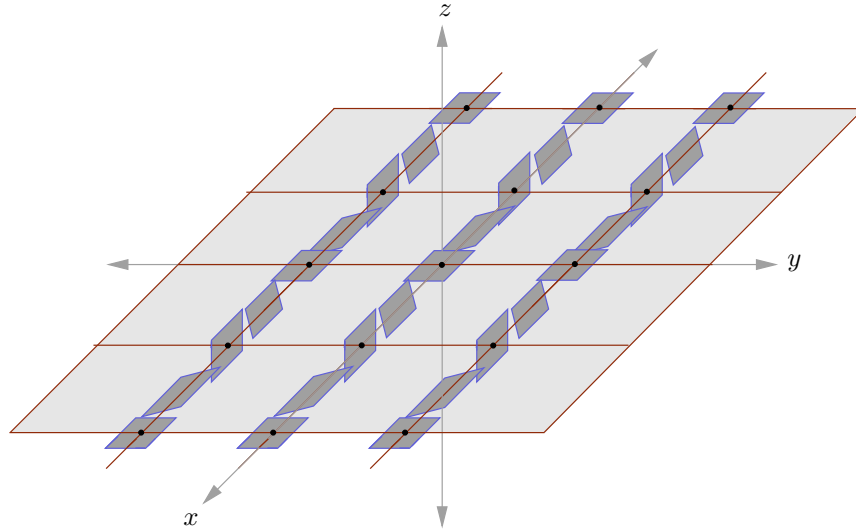
( )
( )
(∩)
(∩)
∘ ̄

′
+   ′

FIGURE 9.   A non-integrable 2-dimensional distribution on $\mathbb{R}^3$.

and every vector field $X \in \mathfrak{X}(M)$ therefore has a horizontal lift $X^h \in \Gamma(HE) \subset \mathfrak{X}(E)$, defined by

$$X^h(v) := \mathrm{Hor}_v(X(p)) \qquad \text{for } p \in M,\ v \in E_p.$$

We will denote by

$$H : TE \to HE$$

the bundle map that projects each $T_v E$ linearly to $H_v E$ along $V_v E$.

EXERCISE 25.17. Show that for any $X \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$, $\mathcal{L}_{X^h}(f \circ \pi) = (\mathcal{L}_X f) \circ \pi$.

LEMMA 25.18. *If* $\eta, \xi \in \Gamma(HE) \subset \mathfrak{X}(E)$ *satisfy* $\mathcal{L}_\eta(f \circ \pi) \equiv \mathcal{L}_\xi(f \circ \pi)$ *for every function* $f \in C^\infty(M)$, *then* $\eta \equiv \xi$.

PROOF. If $\eta(v) \neq \xi(v)$ for some $v \in E_p$ at $p \in M$, then $\pi_* \eta(v) \neq \pi_* \xi(v)$ since $\pi_* : TE \to TM$ maps $H_v E$ isomorphically to $T_p M$; we shall assume without loss of generality that $\pi_* \xi(v) \neq 0$. Then there exists a smooth function $f : M \to \mathbb{R}$ satisfying $df(\pi_* \xi(v)) \neq 0$ and $df(\pi_* \eta(v)) = 0$, which means $\mathcal{L}_\eta(f \circ \pi)(v) = 0 \neq \mathcal{L}_\xi(f \circ \pi)(v)$. □

LEMMA 25.19. *For any* $X, Y \in \mathfrak{X}(M)$, $[X,Y]^h = H \circ [X^h, Y^h]$.

PROOF. Observe first that for any $\xi \in \mathfrak{X}(E)$ and $f \in C^\infty(M)$,

$$\mathcal{L}_\xi(f \circ \pi) = \mathcal{L}_{H \circ \xi}(f \circ \pi),$$

i.e. ξ can be replaced with its horizontal part or vice versa since the difference between them is vertical, and $d(f \circ \pi)|_{VE} \equiv 0$. Then for $X, Y \in \mathfrak{X}(M)$, using Exercise 25.17,

$$\mathcal{L}_{H \circ [X^h, Y^h]}(f \circ \pi) = \mathcal{L}_{[X^h, Y^h]}(f \circ \pi) = \mathcal{L}_{X^h} \mathcal{L}_{Y^h}(f \circ \pi) - \mathcal{L}_{Y^h} \mathcal{L}_{X^h}(f \circ \pi)$$
$$= \mathcal{L}_{X^h}\left((\mathcal{L}_Y f) \circ \pi\right) - \mathcal{L}_{Y^h}\left((\mathcal{L}_X f) \circ \pi\right) = (\mathcal{L}_X \mathcal{L}_Y f) \circ \pi - (\mathcal{L}_Y \mathcal{L}_X f) \circ \pi$$
$$= (\mathcal{L}_{[X,Y]} f) \circ \pi.$$

Likewise, again applying Exercise 25.17,

$$\mathcal{L}_{[X,Y]^h}(f \circ \pi) = (\mathcal{L}_{[X,Y]} f) \circ \pi = \mathcal{L}_{H \circ [X^h, Y^h]}(f \circ \pi),$$

so the result follows from Lemma 25.18   □

We now come to the main step in the proof of the Frobenius theorem.

LEMMA 25.20. *Suppose that for every pair of vector fields* $X, Y \in \mathfrak{X}(M)$*, the vector field* $[X^h, Y^h] \in \mathfrak{X}(E)$ *takes values in* HE*. Then* $HE \subset TE$ *is an integrable distribution on* E*.*

PROOF. Since the question is purely local, we lose no generality if we replace M with a small neighborhood of some point $p \in M$ on which a chart $(x^1, \ldots, x^n)$ can be defined. Denote the resulting coordinate vector fields by $X_j := \partial_j \in \mathfrak{X}(M)$ for $j = 1, \ldots, n$. By assumption $[X_i^h, X_j^h]$ is horizontal for every i and j, thus by Lemma 25.19,

$$[X_i^h, X_j^h] = H \circ [X_i^h, X_j^h] = [X_i, X_j]^h = 0.$$

It follows that for any $v \in E_p$, we can construct an integral submanifold through v via the commuting flows of X_i^h: it is parametrized by the map

(25.1) $$\psi(t^1, \ldots, t^n) = \varphi_{X_1^h}^{t^1} \circ \ldots \circ \varphi_{X_n^h}^{t^n}(v)$$

for real numbers $t^1, \ldots, t^n$ sufficiently close to 0. $\qquad\square$

EXERCISE 25.21. Verify that the map (25.1) parametrizes an embedded integral submanifold of HE.

The last step is to observe that while the distribution we've been considering in this discussion looks like a special case, there is no actual loss of generality.

PROOF OF THEOREM 25.16. We assume $\xi \subset TM$ is a k-dimensional distribution such that $[X, Y] \in \Gamma(\xi)$ whenever $X, Y \in \Gamma(\xi)$. Since the integrability question is purely local, we can choose a chart near some point $p \in M$ so as to assume without loss of generality that M is an open subset $\mathcal{U} \subset \mathbb{R}^n$, and after a linear change of coordinates, we can also arrange that $\xi_p \subset \mathbb{R}^n$ is complementary to the subspace $\{0\} \times \mathbb{R}^{n-k} \subset \mathbb{R}^n$. After possibly shrinking the neighborhood $\mathcal{U}$, it follows that ξ_q is also complementary to this same subspace for every $q \in \mathcal{U}$. We can now view $\mathcal{U}$ as an open subset in the total space of the trivial vector bundle $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k} \to \mathbb{R}^k$: $(q, v) \mapsto q$, in which fibers take the form $\{q\} \times \mathbb{R}^{n-k}$, and ξ is therefore a horizontal subbundle. The stated condition on Lie brackets then establishes the hypothesis of Lemma 25.20, implying ξ is integrable. $\qquad\square$

EXERCISE 25.22. While integral submanifolds of a distribution $\xi \subset M$ through a given point $p \in M$ are not guaranteed to exist, show that they are unique in the following sense: if $\Sigma_1, \Sigma_2 \subset M$ are two integral submanifolds containing p, then there exist neighborhoods $\mathcal{U}_1 \subset \Sigma_1$ and $\mathcal{U}_2 \subset \Sigma_2$ of p in each such that $\mathcal{U}_1 = \mathcal{U}_2$.
Hint: It may help to think only about the special case $\xi = HE \subset TE$ *for a vector bundle* $\pi : E \to M$*, since every case locally looks like this one. Remember that a horizontal subbundle always uniquely determines parallel transport along paths.*

25.4. A tensorial characterization of flatness. In preparation for the general definition of curvature in the next lecture, we can now associate to every connection ∇ on a bundle $\pi : E \to M$ a tensor field whose vanishing is equivalent to the integrability condition in the Frobenius theorem. We continue to denote by $H : TE \to HE$ the projection along VE, and define also the complementary projection

$$V : TE \to VE,$$

which projects $T_v E$ linearly along $H_v E$ to $V_v E$ for each $v \in E$. We use these to define a bilinear map $\widehat{\Omega}_K : \mathfrak{X}(E) \times \mathfrak{X}(E) \to \Gamma(VE)$ by

$$\widehat{\Omega}_K(\eta, \xi) := -V\left([H(\eta), H(\xi)]\right).$$

The Frobenius theorem is equivalent to the statement that this map vanishes if and only if $HE \subset TE$ is an integrable distribution: indeed, every vector field on E with values in HE can be written as $H(\eta)$ for some $\eta \in \mathfrak{X}(E)$, and an arbitrary $\eta \in \mathfrak{X}(E)$ takes values in HE if and only if $V(\eta) \equiv 0$. The real reason why $\widehat{\Omega}_K$ is useful is that in addition to characterizing the flatness of a connection, it defines a tensor:

LEMMA 25.23. *The bilinear map* $(\eta, \xi) \mapsto \widehat{\Omega}_K(\eta, \xi)$ *is* C^∞*-linear in both* η *and* ξ.

PROOF. Since $\widehat{\Omega}_K$ is clearly antisymmetric, it suffices to show that it is C^∞-linear with respect to η. We use the formula $[fX, Y] = f[X, Y] - (\mathcal{L}_Y f)X$ from Exercise 6.4: for any $\eta, \xi \in \mathfrak{X}(E)$ and $f \in C^\infty(E)$,

$$
\widehat{\Omega}_K(f\eta, \xi) = -V\left([fH(\eta), H(\xi)]\right) = -V\left(f[H(\eta), H(\xi)] - \mathcal{L}_{H(\xi)}f \cdot H(\eta)\right)
$$
$$
= -fV\left([H(\eta), H(\xi)]\right) = f\widehat{\Omega}_K(\eta, \xi),
$$

where the term $V\left(\mathcal{L}_{H(\xi)}f \cdot H(\eta)\right) = \mathcal{L}_{H(\xi)}f \cdot V(H(\eta))$ disappears because $H(\eta)$ takes horizontal values and is therefore annihilated by V. $\square$

The lemma implies that $\widehat{\Omega}_K$ can be interpreted as defining a bilinear bundle map

$$
\widehat{\Omega}_K : TE \oplus TE \to VE,
$$

and since it is antisymmetric, we also think of it as a *bundle-valued* differential 2-form on E, and write

$$
\widehat{\Omega}_K \in \Omega^2(E, VE).
$$

This is one version of an object called the *curvature 2-form* determined by the connection ∇ on E; you can now regard the subscript K as either a reference to the projection $K : TE \to E$ that determines the connection (Definition 19.5), or simply as an abbreviation for the word *Krümmung*. In the next lecture we will discuss a somewhat more user-friendly variant of $\widehat{\Omega}_K$ that packages the same information. Let us first record the following consequence of the Frobenius theorem:

COROLLARY 25.24. *A connection on a vector bundle* $\pi : E \to M$ *is flat if and only if the bundle-valued 2-form* $\widehat{\Omega}_K \in \Omega^2(E, VE)$ *vanishes.* $\square$

25.5. Addendum: integrability in general. Integrability theorems are ubiquitous in differential geometry, and one should learn to recognize them. They can take different forms depending on the context in which they arise, but most fit the following paradigm: we have a manifold M whose tangent bundle TM carries some extra geometric structure defining a preferred class of local frames, which are guaranteed to exist on neighborhoods of any point. A preferred class of frames determines a preferred class of charts, namely $(x^1, \ldots, x^n)$ such that the frame formed by the coordinate vector fields $\partial_1, \ldots, \partial_n$ belongs to the preferred class. But as we saw in §25.2, not every frame comes from a chart, so it is typically harder to find a preferred chart than a preferred frame, and they don't always exist: typically some nontrivial *integrability condition* is required before the local existence of preferred charts can be guaranteed.

Theorem 25.11 fits this paradigm in a trivial way: in this case the extra geometric structure is the frame itself, and the question is whether that particular frame can arise locally from a chart.

The Frobenius theorem can also be recast in this language. Here the extra geometric structure is a distribution $\xi \subset TM$, i.e. a subbundle of the tangent bundle, and the preferred class of frames comes from Proposition 17.12: every point $p \in M$ has a neighborhood $\mathcal{U} \subset M$ on which there is a frame $X_1, \ldots, X_n$ for TM such that ξ is the span of $X_1, \ldots, X_k$ at every point $\mathcal{U}$. A frame arising naturally from a chart $x = (x^1, \ldots, x^n) : \mathcal{U} \to \mathbb{R}^n$ will have this property if and only if ξ is spanned at every point by the first k coordinate vector fields $\partial_1, \ldots, \partial_k$, in which case integral submanifolds

obviously exist through every point: they take the form $x^{-1}(\mathbb{R}^k \times \{q\})$ for constants $q \in \mathbb{R}^{n-k}$. The existence of charts of this form is in fact *equivalent* to integrability:

PROPOSITION 25.25. *A k-dimensional distribution $\xi \subset TM$ is integrable if and only if every point $p \in M$ admits a neighborhood $\mathcal{U} \subset M$ with a chart $x : \mathcal{U} \to \mathbb{R}^n$ such that the sets $x^{-1}(\mathbb{R}^k \times \{q\})$ are integral submanifolds of ξ for each $q \in \mathbb{R}^{n-k}$.*

EXERCISE 25.26. Prove Proposition 25.25.
Hint: This may seem easier if you think of ξ as a horizontal subbundle in TE for some vector bundle E.

Proposition 25.25 shows that for an integral distribution, the integral submanifolds are not just locally unique (cf. Exercise 25.22), but they also fit together into a locally-defined smooth family of smooth submanifolds. This gives rise to a decomposition of M called a **foliation** (*Blätterung*), and every connected subset $\Sigma \subset M$ formed as a union of overlapping connected integral submanifolds is called a **leaf** (*Blatt*) of the foliation. By construction, every point in M belongs to a unique leaf of the foliation, and unless $\xi = TM$, there are always uncountably many distinct leaves. It is pleasant to picture them as a smooth family of disjoint submanifolds whose union is M, though this description is not always completely accurate: the following example shows that leaves are not always submanifolds, at least not globally.

EXERCISE 25.27. Given a constant $(a, b) \in \mathbb{R}^2 \backslash \{0\}$, consider a distribution ξ on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ such that ξ at every point is the subspace spanned by the constant vector field $X = a\partial_x + b\partial_y$, where ∂_x, ∂_y are the usual coordinate vector fields of $\mathbb{R}^2$ (which are also well-defined on $\mathbb{T}^2$ since its tangent spaces are all canonically isomorphic to $\mathbb{R}^2$). This distribution is always integrable since it is 1-dimensional. Draw pictures of some representative leaves of the resulting foliation in the cases where (a, b) is $(1, 0)$, $(0, 1)$, $(1, 1)$ and $(2, 1)$. In these cases all leaves are 1-dimensional submanifolds of $\mathbb{T}^2$. Show however that if one of a or b is rational and the other is irrational, then every leaf of the foliation is dense in $\mathbb{T}^2$, and therefore cannot be a submanifold.

For your information, here are some additional examples of integrability results, most of which we will not cover in this course, though the first is an important exception.

EXAMPLE 25.28. If (M, g) is a pseudo-Riemannian manifold with signature (k, ℓ), the orthonormal frames define a preferred class of local frames for TM, equivalent to $O(k, \ell)$-compatible local trivializations. Such a frame arises from a chart $(x^1, \ldots, x^n)$ if and only if the metric g has constant components $g_{ij} \equiv \eta_{ij}$ in this chart (cf. the discussion following Proposition 23.2). We will show in the next lecture that charts of this form exist if and only if the Levi-Cività connection on (M, g) is flat, i.e. its curvature vanishes. You can already see that this is a necessary condition, since having constant components g_{ij} in some chart implies that the connection is trivial in the corresponding local trivialization.

EXAMPLE 25.29. We did not include symplectic structures among the list of "G-structures" in Lecture 18, but we could have done. The *standard symplectic structure* of $\mathbb{R}^{2m}$ is the 2-form $\omega_{\text{std}} := \sum_{j=1}^{m} dp^j \wedge dq^j$ written in global coordinates $(p^1, q^1, \ldots, p^m, q^m)$. A linear transformation $A : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$ is called **symplectic** if it preserves this structure, meaning $\omega_{\text{std}}(AX, AY) = \omega_{\text{std}}(X, Y)$ for all $X, Y \in \mathbb{R}^{2m}$, and the set of all such transformations forms the **symplectic linear group** $\text{Sp}(2m) \subset \text{GL}(2m, \mathbb{R})$. The 2-form ω_{std} is nondegenerate, meaning $\omega_{\text{std}}(X, \cdot) \neq 0 \in (\mathbb{R}^{2m})^*$ for every $X \neq 0 \in \mathbb{R}^{2m}$. Conversely, it is a simple exercise in symplectic linear algebra to show that for any real $2m$-dimensional vector space V with a nondegenerate alternating 2-form $\omega \in \Lambda^2 V^*$,

there exists a basis $(P_1, Q_1, \ldots, P_m, Q_m)$ such that

$$
\begin{aligned}
\omega(P_j, Q_j) &= 1 && \text{for all } j, \\
\text{(25.2)} \qquad \omega(P_i, Q_j) &= 0 && \text{for all } i \neq j, \\
\omega(P_i, P_j) = \omega(Q_i, Q_j) &= 0 && \text{for all } i, j,
\end{aligned}
$$

so this basis produces an isomorphism $V \cong \mathbb{R}^{2m}$ that identifies ω with ω_{std}. A procedure for finding the basis is as follows: first choose any linearly-independent P_1, Q_1 such that $\omega(P_1, Q_1) = 1$, which is possible because ω is nondegenerate and alternating. The restriction of ω to the subspace $V_1 \subset V$ spanned by P_1 and Q_1 is then nondegenerate, so by a straightforward analogue of Lemma 24.7, its *symplectic orthogonal complement*

$$
V_1^{\omega\perp} := \left\{ v \in V \mid \omega(v, \cdot)|_{V_1} = 0 \right\}
$$

satisfies $\mathbb{R}^{2m} = V_1 \oplus V_1^{\omega\perp}$, and $\omega|_{V_1^{\omega\perp}}$ is also nondegenerate. Now repeat the same argument on $V_1^{\omega\perp}$, which is 2 dimensions smaller than V, and keep repeating until there are no dimensions left. In summary, every nondegenerate alternating 2-form is equivalent to the standard symplectic form via a choice of basis.

On a real vector bundle $E \to M$ of even rank $2m$, an $\mathrm{Sp}(2m)$-structure now determines on each fiber E_p an alternating 2-form $\omega_p \in \Lambda^2 E_p^*$ that looks like ω_{std} in any $\mathrm{Sp}(2m)$-compatible local trivialization, and the map $p \mapsto \omega_p$ is then a smooth section ω of the vector bundle $\Lambda^2 E^* \to M$. The frames corresponding to $\mathrm{Sp}(2m)$-compatible trivializations consist of tuples of sections $(P_1, Q_1, \ldots, P_m, Q_m)$ that satisfy the relations in (25.2); we call them **symplectic frames**. Conversely, for any choice of section $\omega \in \Gamma(\Lambda^2 E^*)$ that is nondegenerate on every fiber, one can use the procedure outlined above to construct frames that satisfy (25.2) over sufficiently small neighborhoods of any point in M. A covering of M by neighborhoods with such frames gives rise to an $\mathrm{Sp}(2m)$-structure on $E \to M$, and we then call $E \to M$ a **symplectic vector bundle**.

If (M, ω) is a $2n$-dimensional symplectic manifold, then ω makes $TM \to M$ into a symplectic vector bundle, for which any local coordinates $(p^1, q^1, \ldots, p^n, q^n)$ in which $\omega = \sum_{j=1}^n dp^j \wedge dq^j$ give rise to a symplectic frame $\frac{\partial}{\partial p^1}, \frac{\partial}{\partial q^1}, \ldots, \frac{\partial}{\partial p^n}, \frac{\partial}{\partial q^n}$. But not every $\mathrm{Sp}(2n)$-structure on the bundle $TM \to M$ arises in this way from a symplectic form on M. According to the previous paragraph, an $\mathrm{Sp}(2n)$-structure on $TM \to M$ is equivalent to a choice of smooth 2-form $\omega \in \Omega^2(M)$ that is nondegenerate on every fiber. In this situation, local symplectic frames can always be found, but can they always also be realized as coordinate vector fields for a chart $(p^1, q^1, \ldots, p^n, q^n)$ in which $\omega = \sum_{j=1}^n dp^j \wedge dq^j$? There is an obvious necessary condition for this: ω cannot take that form in any local coordinates if it is not closed, and indeed, if $\dim M > 2$, there is no reason in general why a globally nondegenerate 2-form must also be closed. We can therefore view "$d\omega = 0$" as an integrability condition for a symplectic tangent bundle to be upgraded to a symplectic manifold. According to Darboux's theorem, this condition is also sufficient, i.e. every closed nondegenerate 2-form matches the standard symplectic form in some local coordinates. For more on both symplectic vector bundles and Darboux's theorem, see [**MS17**].

EXAMPLE 25.30. A volume form $\mu \in \Omega^n(M)$ on an n-manifold M is the same thing as an $\mathrm{SL}(n, \mathbb{R})$-structure on the bundle $TM \to M$, and the preferred class of frames consists of tuples of vector fields $X_1, \ldots, X_n$ defined on open subsets $\mathcal{U} \subset M$ such that $\mu(X_1, \ldots, X_n) \equiv 1$. The preferred class of charts $(x^1, \ldots, x^n)$ can then be characterized by the condition that μ in any such chart looks like the *standard* volume form $dx^1 \wedge \ldots \wedge dx^n$. It is very easy to turn any local frame into one that satisfies $\mu(X_1, \ldots, X_n) \equiv 1$, but less obvious in general whether every volume form can be made to look standard near every point by choosing the right coordinates. However, it is true: the necessary and sufficient integrability condition for this is $d\mu = 0$, just as with symplectic

forms, but with the important difference that it is *always* satisfied since μ is a top-dimensional form. One can prove this integrability result by a slight variation on one of the standard proofs of Darboux's theorem, using the "Moser deformation" trick.

EXAMPLE 25.31. A deep integrability theorem for almost complex structures $J \in \Gamma(\mathrm{End}(TM))$ on a $2n$-manifold M was mentioned in Exercise 8.5. An almost complex structure is equivalent to a $\mathrm{GL}(n, \mathbb{C})$-structure on TM, where $\mathrm{GL}(n, \mathbb{C})$ is identified with a subgroup of $\mathrm{GL}(2n, \mathbb{R})$ as in Example 18.10, and local frames in the preferred class take the form $(X_1, Y_1, \ldots, X_n, Y_n)$ where $Y_j = JX_j$ and $X_j = -JY_j$ for each $j = 1, \ldots, n$. A covering of M by charts that produce frames of this kind is equivalent to a covering by *complex* charts whose transition maps are holomorphic, thus making M into an n-dimensional complex manifold. An almost complex structure J is called *integrable* if M admits a covering by charts with this property, and according to the Newlander-Nirenberg theorem, the necessary and sufficient condition for J to be integrable is the vanishing of its associated Nijenhuis tensor $N \in \Gamma(T_2^1 M)$.

26. Curvature on a vector bundle

Like connections, curvature is one of those concepts that can be given several equivalent but cosmetically quite different definitions, each of which has distinct advantages in different situations. In this lecture we give two definitions[70] of the curvature of a connection on a vector bundle $\pi : E \to M$, and prove that they are equivalent. It will be immediate from one of these definitions that a connection is flat if and only if its curvature vanishes, while the other definition answers the question of when covariant partial derivatives in different directions do or do not commute.

26.1. Prelude: bundle-valued forms. We have already had a few occasions to mention bundle-valued differential forms, but have not given any formal definition of this notion so far. The time for that is now: for any vector bundle $\pi : E \to M$ and each integer $k \geq 0$, we define

$$\Omega^k(M, E)$$

as the vector space of all smooth maps

$$\omega : \underbrace{TM \oplus \ldots \oplus TM}_{k} \to E$$

such that for every $p \in M$, the restriction of ω to the fiber over p is an antisymmetric k-fold multilinear map $\omega_p : T_pM \times \ldots \times T_pM \to E_p$. As with real-valued forms, the antisymmetry condition is vacuous in the cases $k = 0, 1$, and the convention is to define $\Omega^0(M, E) := \Gamma(E)$. Another way to formulate the definition would be that $\Omega^k(M, E)$ is the space of smooth sections of the vector bundle $(\Lambda^k T^* M) \otimes E$, whose fibers can be identified canonically with the spaces of antisymmetric multilinear maps described above. Note that if $E \to M$ is a *complex* vector bundle, then it is regarded as a real bundle for the purposes of these definitions, since the fibers of TM cannot be assumed to be equipped with any complex structure.

26.2. The curvature 2-form. In §25.4 we associated to any connection ∇ on a vector bundle $\pi : E \to M$ a bundle-valued 2-form $\widehat{\Omega}_K \in \Omega^2(E, VE)$ satisfying

$$\widehat{\Omega}_K(\eta, \xi) := -V\left([H(\eta), H(\xi)]\right)$$

for all $\eta, \xi \in \mathfrak{X}(E)$, where $V : TE \to VE$ denotes the fiberwise-linear projection along HE and $H : TE \to HE$ is the complementary projection. We saw that this formula for $\widehat{\Omega}_K$ is C^∞-linear in both variables, and that by the Frobenius theorem, it vanishes if and only if the distribution $HE \to TE$ is integrable, which means the connection ∇ is flat. All of this is true for

---

[70]plus two more that will be implicit in the exercises at the end

any horizontal subbundle $HE \subset TE$, i.e. we have not yet actually used the additional requirement in Definition 19.4 that HE should be compatible with scalar multiplication,[71] i.e. the relation

$$(m_\lambda)_* (HE) = HE$$

for every $\lambda \in \mathbb{F}$, with $m_\lambda : E \to E$ denoting the smooth map $v \mapsto \lambda v$. This condition makes it possible to replace $\widehat{\Omega}_K \in \Omega^2(E, VE)$ with an object that is simpler, but equivalent. Recall from Definition 19.5 that the connection can also be characterized via a map $K : TE \to E$ that sends $T_v E$ linearly to $E_{\pi(v)}$ and vanishes on the horizontal subspaces: K is actually just the composition of the fiberwise-linear projection $TE \to VE$ with the canonical isomorphisms

$$\mathrm{Vert}_v^{-1} : V_v E \to E_p \qquad \text{for } v \in E_p,\, p \in M.$$

The condition $(m_\lambda)_*(HE) = HE$ is then equivalent to the condition

(26.1) $$K \circ Tm_\lambda = m_\lambda \circ K$$

for all $\lambda \in \mathbb{F}$. Writng $\mathrm{End}(E) := \mathrm{Hom}(E, E)$, we claim that the expression

$$\Omega_K(X, Y)v := \mathrm{Vert}_v^{-1}\left(\widehat{\Omega}_K(\mathrm{Hor}_v(X), \mathrm{Hor}_v(Y))\right) \in E_p \qquad \text{for } X, Y \in T_p M,\, v \in E_p,\, p \in M$$

defines a bundle-valued 2-form

$$\Omega_K \in \Omega^2(M, \mathrm{End}(E)).$$

We can already see that this expression is bilinear and antisymmetric in X and Y; the main thing to check is that for each fixed $X, Y \in T_p M$, the map $E_p \to E_p : v \mapsto \Omega_K(X, Y)v$ is linear. It is clearly smooth, so by Lemma 19.2, it will be sufficient to show that it is also compatible with scalar multiplication. To see this, let us associate to each vector field $X \in \mathfrak{X}(M)$ on M the "horizontal" vector field on E given by $X^h(v) := \mathrm{Hor}_v(X(p))$ for $v \in E_p$ and $p \in M$ as in §25.3. Since K is the composition of V with Vert_v^{-1}, we can rewrite Ω_K in terms of the definition of $\widehat{\Omega}_K$ as

(26.2) $$\Omega_K(X, Y)v = -K\left([X^h, Y^h](v)\right)$$

for any $X, Y \in \mathfrak{X}(M)$. Now observe that since $(m_\lambda)_*(HE) = HE$, the horizontal vector field X^h (and similarly Y^h) satisfies the relation

$$X^h(\lambda v) = Tm_\lambda(X^h(v)).$$

If $\lambda \neq 0$, so that $m_\lambda : E \to E$ is a diffeomorphism, this relation can be stated more succinctly as the condition that X^h is its own pushforward under this diffeomorphism:

$$(m_\lambda)_* X^h = X^h \in \mathfrak{X}(E).$$

By Exercise 6.5, it follows that

$$(m_\lambda)_*[X^h, Y^h] = [(m_\lambda)_* X^h, (m_\lambda)_* Y^h] = [X^h, Y^h] \in \mathfrak{X}(E),$$

thus $[X^h, Y^h] \in \mathfrak{X}(E)$ also satisfies the relation

$$[X^h, Y^h](\lambda v) = Tm_\lambda([X^h, Y^h](v)).$$

Continuing under the assumption $\lambda \neq 0$, we can now use (26.1) to conclude

$$\Omega_K(X, Y)\lambda v = -K\left([X^h, Y^h](\lambda v)\right) = -K\left(Tm_\lambda([X^h, Y^h](v))\right) = -m_\lambda \circ K\left([X^h, Y^h](v)\right)$$
$$= \lambda \Omega_K(X, Y)v.$$

If this holds for all nonzero $\lambda \in \mathbb{F}$, then by continuity it also holds for $\lambda = 0$, and the claim is thus proven: $v \mapsto \Omega_K(X, Y)v$ is linear, so Ω_K is a 2-form with values in the vector bundle $\mathrm{End}(E) \to M$.

---

[71]Without assuming the condition $(m_\lambda)_*(HE) = HE$ holds, one can still regard the horizontal subbundle $HE \subset TE$ as a connection on $\pi : E \to M$, but only after forgetting the fact that the fibers of this bundle are vector spaces and regarding them instead as smooth submanifolds, so that $\pi : E \to M$ is now an example of a *fiber bundle*.

EXERCISE 26.1. Show that $\Omega_K \in \Omega^2(M, \operatorname{End}(E))$ vanishes if and only if $\hat{\Omega}_K \in \Omega^2(E, VE)$ vanishes.

DEFINITION 26.2. We call $\Omega_K \in \Omega^2(M, \operatorname{End}(E))$ the **curvature 2-form** of the connection ∇ on $\pi : E \to M$, and say that ∇ has **vanishing curvature** if $\Omega_K \equiv 0$.

Exercise 26.1 now combines with Corollary 25.24 to prove:

COROLLARY 26.3. *A connection on a vector bundle is flat if and only if its curvature vanishes.*
□

26.3. The Riemann tensor. The definition of curvature given in the previous section does not easily lend itself to computations. In order to remedy this, let's go back to the third question in 25.1: do covariant partial derivatives in different coordinate directions commute? We've seen that the answer is yes for a flat connection, and one of the main results of the present lecture will be a converse to this: if they always commute, then the connection must be flat.

Let's first reframe the question in coordinate-invariant language. We could ask for instance whether the differential operators $\nabla_X, \nabla_Y : \Gamma(E) \to \Gamma(E)$ must commute for an arbitrary choice of two vector fields $X, Y \in \mathfrak{X}(M)$, but this is not even true in the simplest special case: for the trivial connection on the trivial real line bundle over M, $\Gamma(E)$ is identified with $C^\infty(M)$ and ∇_X and ∇_Y become the operators $\mathcal{L}_X$ and $\mathcal{L}_Y$ respectively, whose failure to commute is measured by the definition of the Lie bracket, which amounts to the formula

$$\mathcal{L}_X \mathcal{L}_Y - \mathcal{L}_Y \mathcal{L}_X = \mathcal{L}_{[X,Y]} \qquad \text{on } C^\infty(M).$$

One might extrapolate from this case and guess that the relation $\nabla_X \nabla_Y - \nabla_Y \nabla_X = \nabla_{[X,Y]}$ should hold for general connections. This turns out to be false in general, but the failure of this identity is measured by a tensor:

DEFINITION 26.4. Given a connection ∇ on a vector bundle $E \to M$, the **Riemann curvature tensor** is the unique multilinear bundle map

$$R : TM \oplus TM \oplus E \to E : (X, Y, v) \mapsto R(X, Y)v$$

such that for all $X, Y \in \mathfrak{X}(M)$ and $v \in \Gamma(E)$,

$$R(X, Y)v = \nabla_X \nabla_Y v - \nabla_Y \nabla_X v - \nabla_{[X,Y]} v.$$

The exercise below shows that this is well defined, and in particular, if $E = TM$, R is a tensor field of type $(1, 3)$ on M.

EXERCISE 26.5. Show that $R(X, Y)v$ is C^∞-linear with respect to each of its three arguments.

EXERCISE 26.6. Choosing a chart $x = (x^1, \ldots, x^n) : \mathcal{U} \to \mathbb{R}^n$ and a frame $(e_1, \ldots, e_m)$ for E over some open subset $\mathcal{U} \subset M$, define the components $R^a{}_{jkb} : \mathcal{U} \to \mathbb{F}$ of the Riemann tensor R such that

$$R(\partial_j, \partial_k)e_b = R^a{}_{jkb}e_a,$$

hence $(R(X, Y)v)^a = R^a{}_{jkb}X^j Y^k v^b$ for any $X, Y \in T_pM$ and $v \in E_p$ at $p \in \mathcal{U}$. Show that these components are given in terms of the Christoffel symbols of the connection by

$$R^a{}_{jkb} = \partial_j \Gamma^a_{kb} - \partial_k \Gamma^a_{jb} + \Gamma^a_{jc} \Gamma^c_{kb} - \Gamma^a_{kc} \Gamma^c_{jb}.$$

REMARK 26.7. Exercise 26.6 together with (22.6) shows that for the Levi-Cività connection on a pseudo-Riemannian manifold, the Riemann tensor is determined by the second derivatives of the components of the metric in any local coordinates.

EXERCISE 26.8. Suppose $\mathcal{V} \subset \mathbb{R}^2$ is an open subset with coordinates labelled (s, t), $f : \mathcal{V} \to M$ is a smooth map and $v \in \Gamma(f^*E)$ is a section of E along f. Prove the formula

$$\nabla_s \nabla_t v - \nabla_t \nabla_s v = R(\partial_s f, \partial_t f)v \qquad \text{on } \mathcal{V}.$$

Hint: On any neighborhood in $\mathcal{V}$ on which f is an embedding, you can derive this from the definition of the Riemann tensor after extending f to a diffeomorphism onto an open set in M and choosing a corresponding extension of v to a section of $E \to M$. If $\dim M \geqslant 2$, deduce the general case from this via continuity (cf. the proof of (21.2)), using the fact that any smooth map $\mathbb{R}^2 \supset \mathcal{V} \xrightarrow{f} M$ can be perturbed to become an embedding on some neighborhood of any given point. If $\dim M \leqslant 1$ then there is nothing to prove, because R vanishes (why?) and the connection ∇ is automatically flat, implying that its pullback to $f^*E \to \mathcal{V}$ is also flat (see Exercise 25.7).

It may be surprising at first sight that $R(X, Y)v$ doesn't depend on any derivatives of v: indeed, it seems to tell us less about v than about the connection itself. The main theorem in this lecture says that the Riemann tensor gives a complete characterization of the curvature of the connection—in particular, its vanishing gives yet another necessary and sufficient condition for the connection to be flat.

THEOREM 26.9. *For any vector bundle $E \to M$ with connection ∇, the Riemann tensor R and curvature 2-form Ω_K are related by*

$$R(X, Y)v = \Omega_K(X, Y)v.$$

COROLLARY 26.10. *The connection ∇ on $E \to M$ is flat if and only if for every chart $(x^1, \ldots, x^n)$, the covariant partial derivative operators ∇_i and ∇_j commute for all $i, j \in \{1, \ldots, n\}$.*

We will prove Theorem 26.9 in the next section.

If $\dim M = n$ and $\operatorname{rank}(E) = m$, then the Riemann tensor is described locally by n^2m^2 component functions $R^a{}_{jkb}$ for $j, k \in \{1, \ldots, n\}$ and $a, b \in \{1, \ldots, m\}$. This sounds like quite a lot of information, but it is often useful to notice that these components are not all independent of each other. One nontrivial relation is obvious already from the definition: since $R(X, Y)v$ is antisymmetric in X and Y, we have

$$R^a{}_{jkb} = -R^a{}_{kjb}.$$

One can say more if ∇ is compatible with a bundle metric, as is true in particular for the Levi-Cività connection on a tangent bundle:

EXERCISE 26.11. Show that if ∇ is compatible with a bundle metric $\langle\ ,\ \rangle$ on E, then the Riemann tensor satisfies

$$\langle R(X, Y)v, w \rangle + \langle v, R(X, Y)w \rangle = 0$$

for all $(X, Y, v, w) \in TM \oplus TM \oplus E \oplus E$.
Hint: Given $X, Y \in \mathfrak{X}(M)$ and $v, w \in \Gamma(E)$, apply the operator $\mathcal{L}_X \mathcal{L}_Y - \mathcal{L}_Y \mathcal{L}_X - \mathcal{L}_{[X,Y]}$ to the function $\langle v, w \rangle$.

Exercise 26.11 says that for each $X, Y \in T_pM$, the linear map $R(X, Y) : E_p \to E_p$ is antisymmetric with respect to the bundle metric on E. Let's see what this means in the case where E is the tangent bundle of an oriented Riemannian 2-manifold (Σ, g). The space of antisymmetric linear maps $T_p\Sigma \to T_p\Sigma$ in this case is 1-dimensional, and it has a canonical basis defined as follows. Let

$$J \in \Gamma(T^1_1\Sigma) = \Gamma(\operatorname{End}(T\Sigma))$$

denote the unique bundle map $T\Sigma \to T\Sigma$ such that for each $p \in \Sigma$, $J_p : T_p\Sigma \to T_p\Sigma$ is a 90-degree counterclockwise rotation; here "counterclockwise" means that (X, J_pX) is a positively-oriented

basis for each $X \neq 0 \in T_p\Sigma$. Equivalently, J_p is represented by the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ in any positively-oriented orthonormal basis of $T_p\Sigma$. Since J_p is nontrivial and antisymmetric, every antisymmetric linear map $T_p\Sigma \to T_p\Sigma$ is a scalar multiple of it. Similarly, $\dim \Lambda^2 T_p^*\Sigma = 1$, thus every alternating 2-form on $T_p\Sigma$ is a scalar multiple of the Riemannian volume form (or "area form") $d$vol at that point. These two observations, together with Exercise 26.11, imply that for the Levi-Cività connection on the tangent bundle of (Σ, g), there is a unique real-valued function

$$K : \Sigma \to \mathbb{R}$$

such that the Riemann tensor is given by the formula

(26.3) $\qquad\qquad R(X,Y)Z = -K(p)\, d\text{vol}(X,Y)\, JZ \qquad$ for $p \in \Sigma$, $X, Y, Z \in T_p\Sigma$.

This shows that despite the Riemann tensor being described on any coordinate neighborhood by a total of $2^4 = 16$ component functions, they are all determined by a single function $K : \Sigma \to \mathbb{R}$. This function is called the **Gaussian curvature** of (Σ, g), and we will have much more to say about it in the next two lectures.

REMARK 26.12. While it was convenient in the discussion above to assume Σ was oriented, the function $K : \Sigma \to \mathbb{R}$ in (26.3) is still well defined without this assumption. The reason is that reversing the chosen orientation of Σ causes sign changes in both $d$vol and J, and these two sign changes cancel each other so that (26.3) remains valid without any change in K. If Σ is not orientable, one can then define K in a small neighborhood of any point $p \in \Sigma$ by making an arbitrary choice of orientation on this neighborhood; since the result does not depend on this choice, $K : \Sigma \to \mathbb{R}$ is then well defined globally.

26.4. Covariant exterior derivatives. We will prove Theorem 26.9 by relating the bracket to an exterior derivative using a generalization of the formula

$$d\alpha(X,Y) = \mathcal{L}_X\left(\alpha(Y)\right) - \mathcal{L}_Y\left(\alpha(X)\right) - \alpha([X,Y])$$

for 1-forms $\alpha \in \Omega^1(M)$. This is possible because the definitions of Ω_K, K and R can all be expressed in terms of bundle-valued forms.

The covariant derivative gives a linear map

$$\nabla : \Gamma(E) = \Omega^0(M, E) \to \Omega^1(M, E) = \Gamma(\text{Hom}(TM, E)),$$

and by analogy with the differential $d : \Omega^0(M) \to \Omega^1(M)$, it's natural to extend this to a **covariant exterior derivative**

$$d_\nabla : \Omega^k(M, E) \to \Omega^{k+1}(M, E),$$

defined as follows. Every $\omega \in \Omega^k(M, E)$ can be expressed in local coordinates $x = (x^1, \ldots, x^n) : \mathcal{U} \to \mathbb{R}^n$ as

$$\omega = \sum_{1 \leq i_i < \ldots < i_k \leq n} \omega_{i_1 \ldots i_k}\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

for some *component sections* $\omega_{i_1 \ldots i_k} \in \Gamma(E|_\mathcal{U})$. Then $d_\nabla \omega$ is defined locally as

$$d_\nabla \omega = \sum_{1 \leq i_i < \ldots < i_k \leq n} \nabla \omega_{i_1 \ldots i_k} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

$$= \sum_{1 \leq i_i < \ldots < i_k \leq n} \nabla_j \omega_{i_1 \ldots i_k}\, dx^j \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k},$$

where in the last expression, the Einstein summation convention applies to the index j but not to $i_1, \ldots, i_k$. One can show by the same argument as for real-valued differential forms that this definition of d_∇ is independent of the choice of coordinates; see Exercise 26.13 below. Note that

wedge products $\alpha \wedge \beta$ or $\beta \wedge \alpha \in \Omega^{k+\ell}(M, E)$ can naturally be defined for $\alpha \in \Omega^k(M)$ and $\beta \in \Omega^\ell(M, E)$, but it makes no sense if *both* forms are bundle-valued.

EXERCISE 26.13. Show that $d_\nabla : \Omega^k(M, E) \to \Omega^{k+1}(M, E)$ can be defined as the unique linear operator which matches ∇ on $\Omega^0(M, E)$ and satisfies the graded Leibnitz rule

$$d_\nabla(\alpha \wedge \beta) = d_\nabla \alpha \wedge \beta + (-1)^k \alpha \wedge d\beta$$

for all $\alpha \in \Omega^k(M, E)$ and $\beta \in \Omega^\ell(M)$.

EXERCISE 26.14. Show that for $\lambda \in \Omega^1(M, E)$ written in local coordinates over $\mathcal{U} \subset M$ as $\lambda = \lambda_i \, dx^i$ with $\lambda_1, \ldots, \lambda_n \in \Gamma(E|_\mathcal{U})$, the component sections for $d_\nabla \lambda$ over $\mathcal{U}$ are given by

$$(d_\nabla \lambda)_{ij} = \nabla_i \lambda_j - \nabla_j \lambda_i.$$

Use this to prove the coordinate-free formula

$$(26.4) \qquad d_\nabla \lambda(X, Y) = \nabla_X \left(\lambda(Y)\right) - \nabla_Y \left(\lambda(X)\right) - \lambda([X, Y]).$$

Hint: For the last step, the main task is to show that the right hand side of (26.4) *gives a well-defined bundle-valued 2-form; the rest follows easily from the coordinate formula.*

PROOF OF THEOREM 26.9. We will show that both $R(X, Y)v$ and $\Omega_K(X, Y)v$ can be expressed in terms of a covariant exterior derivative of the map $K : TE \to E$. In this context, we regard K as a bundle-valued 1-form $K \in \Omega^1(E, \pi^*E)$ since it maps T_vE linearly to $E_{\pi(v)} = (\pi^*E)_v$. We use the connection ∇ on $\pi : E \to M$ to induce a natural connection on the pullback bundle $\pi^*E \to E$.

We claim first that for any $p \in M$, $v \in E_p$ and $X, Y \in T_pM$,

$$d_\nabla K(\text{Hor}_v(X), \text{Hor}_v(Y)) = \Omega_K(X, Y)v.$$

Indeed, extend X and Y to vector fields on M and use the corresponding horizontal lifts $X^h, Y^h \in \mathfrak{X}(E)$ as extensions of $\text{Hor}_v(X)$ and $\text{Hor}_v(Y) \in T_vE$ respectively. Then using (26.4), (26.2) and the fact that K vanishes on horizontal vectors,

$$d_\nabla K(X^h(v), Y^h(v)) = \nabla_{X^h(v)}\left(K(Y^h)\right) - \nabla_{Y^h(v)}\left(K(X^h)\right) - K([X^h, Y^h](v)) = \Omega_K(X, Y)v.$$

We now show that $R(X, Y)v$ can also be expressed in this way. Choose a smooth map $f(s, t) \in M$ for $(s, t) \in \mathbb{R}^2$ near $(0, 0)$ such that $\partial_s f(0, 0) = X$ and $\partial_t f(0, 0) = Y$, and extend $v \in E_p$ to a section $v(s, t) \in E_{f(s,t)}$ along f such that $v(0, 0) = v$ and $\nabla_s v(0, 0) = \nabla_t v(0, 0) = 0$. The latter can always be done e.g. by letting $v(0, 0)$ determine the values $v(s, t)$ for all $(s, t) \in \mathbb{R}^2$ near $(0, 0)$ via parallel transport along radial paths starting at the origin. (Note that this guarantees $\nabla v = 0$ at $(0, 0)$ and also that ∇v vanishes in radial directions elsewhere, but each of $\nabla_s v$ and $\nabla_t v$ might still be nonzero for $(s, t) \neq (0, 0)$; we cannot force both of these to vanish at every point unless we already know the connection is flat.) Expressing covariant derivatives via the map K (e.g. $\nabla_s v = K(\partial_s v)$) and applying (26.4) once more along with Exercise 26.8, we then find

$$R(X, Y)v = \nabla_s \nabla_t v(0, 0) - \nabla_t \nabla_s v(0, 0) = \nabla_s\left(K(\partial_t v(s, t))\right) - \nabla_t\left(K(\partial_s v(s, t))\right)\big|_{(s,t)=(0,0)}$$

$$= d_\nabla K(\partial_s v, \partial_t v)\big|_{(s,t)=(0,0)} = d_\nabla K(\text{Hor}_v(X), \text{Hor}_v(Y)),$$

where in the last step we used the assumption that $v(s, t)$ has vanishing covariant derivatives at $(0, 0)$, hence $\partial_s v(0, 0)$ and $\partial_t v(0, 0)$ are horizontal. $\qquad\square$

The exercises below exhibit two further ways that curvature can be expressed in terms of exterior derivatives.

EXERCISE 26.15. For a connection ∇ on the bundle $\pi : E \to M$, prove:

(a) For any $v \in \Gamma(E) = \Omega^0(M, E)$ and $X, Y \in T_p M$ at a point $p \in M$, $d_\nabla^2 v := d_\nabla(d_\nabla v) \in \Omega^2(M, E)$ satisfies
$$(d_\nabla^2 v)(X, Y) = R(X, Y)v.$$

(b) The connection ∇ is flat if and only if the covariant exterior derivative operators $d_\nabla : \Omega^k(M, E) \to \Omega^{k+1}(M, E)$ for all $k \geqslant 0$ satisfy $d_\nabla \circ d_\nabla = 0$.

EXERCISE 26.16. Suppose $\pi : E \to M$ has structure group $G \subset \mathrm{GL}(m, \mathbb{F})$ with Lie algebra $\mathfrak{g} \subset \mathbb{F}^{m \times m}$ and ∇ is a G-compatible connection. Recall that ∇ then associates to every G-compatible local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$ a connection 1-form $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{g})$, defined so that
$$(\nabla_X v)_\alpha = \mathcal{L}_X v_\alpha + A_\alpha(X)v_\alpha$$
for any $X \in \mathfrak{X}(\mathcal{U}_\alpha)$, where $v_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$ expresses $v|_{\mathcal{U}_\alpha} \in \Gamma(E|_{\mathcal{U}_\alpha})$ with respect to the trivialization. The corresponding **local curvature** 2-**form** $F_\alpha \in \Omega^2(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$ is defined as the local representation of $\Omega_K \in \Omega^2(M, \mathrm{End}(E))$ with respect to this trivialization, meaning that for $X, Y \in \mathfrak{X}(\mathcal{U}_\alpha)$ and $v \in \Gamma(E|_{\mathcal{U}_\alpha})$,
$$(\Omega_K(X, Y)v)_\alpha = F_\alpha(X, Y)v_\alpha.$$

(a) Prove the formula
$$F_\alpha(X, Y) = dA_\alpha(X, Y) + [A_\alpha(X), A_\alpha(Y)],$$
where the bracket on the right hand side denotes the matrix commutator $[\mathbf{A}, \mathbf{B}] := \mathbf{AB} - \mathbf{BA}$ for $\mathbf{A}, \mathbf{B} \in \mathbb{F}^{m \times m}$.
Hint: Use the Riemann tensor as a stand-in for Ω_K.

(b) If $\Phi_\beta : E|_{\mathcal{U}_\beta} \to \mathcal{U}_\beta \times \mathbb{F}^m$ is a second trivialization related to Φ_α by the transition map $g = g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to G$, show that
$$F_\beta(X, Y) = gF_\alpha(X, Y)g^{-1}.$$

(c) Show that if G is abelian, then $F_\alpha = dA_\alpha$ and it is independent of the choice of trivialization, thus defining a global 2-form $F \in \Omega^2(M, \mathfrak{g})$. (It is sometimes also called the *curvature* 2-*form* of ∇.)

Remark: By a basic result in the theory of Lie groups, the commutator $[\mathbf{A}, \mathbf{B}]$ belongs to $\mathfrak{g}$ whenever $\mathbf{A}, \mathbf{B} \in \mathfrak{g}$; this is the reason why $\mathfrak{g}$ is called the "Lie algebra" of G. It thus follows from part (a) that $F_\alpha \in \Omega^2(\mathcal{U}_\alpha, \mathfrak{g})$. In the case $G = \mathrm{O}(k, \ell)$, this is a locally trivialized analogue of Exercise 26.11, which showed that Ω_K takes values in the bundle of antisymmetric linear maps $E \to E$.

27. Curvature in pseudo-Riemannian manifolds

For the remainder of the semester, we discuss properties and applications of the curvature of the Levi-Cività connection on the tangent bundle of a Riemannian (or occasionally pseudo-Riemannian) manifold.

27.1. The covariant Riemann tensor. When the bundle under consideration is the tangent bundle of a manifold M, the Riemann tensor defines a multilinear map $TM^{\oplus 3} \to TM : (X, Y, Z) \mapsto R(X, Y)Z$ that can be regarded as a type $(1, 3)$ tensor field,
$$R \in \Gamma(T_3^1 M), \qquad R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$
Assuming ∇ is the Levi-Cività connection for a metric g, we have observed two nontrivial relations so far that R must satisfy: one is the antisymmetry
$$R(X, Y)Z = -R(Y, X)Z$$

that is obvious from its definition, and the other (from Exercise 26.11) is

$$\langle V, R(X,Y)Z \rangle + \langle R(X,Y)V, Z \rangle = 0.$$

We saw in §26.3 that when (M,g) is a 2-dimensional Riemannian manifold, these two relations imply that R is determined by a real-valued function—we'll have more to say about that below. (A version of this is also true for indefinite metrics in dimension two; see Exercise 27.1 below.) For certain purposes, it is sometimes useful to repackage the Riemann tensor as a fully covariant tensor $\mathrm{Riem} \in \Gamma(T_4^0 M)$ defined by

$$\mathrm{Riem}(V, X, Y, Z) := \langle V, R(X,Y)Z \rangle.$$

This tensor contains all the same information, and R can be recovered from it; it is essentially the result of applying a musical isomorphism $\flat : T_3^1 M \to T_4^0 M$ that associates to any $S \in \Gamma(T_3^1 M)$ the tensor $S_\flat \in \Gamma(T_4^0 M)$ defined by $S_\flat(V, X, Y, Z) := S(V_\flat, X, Y, Z)$. The two antisymmetry relations mentioned above are now equivalent to
(27.1)
$$\mathrm{Riem}(V, X, Y, Z) = -\mathrm{Riem}(V, Y, X, Z) \qquad \text{and} \qquad \mathrm{Riem}(V, X, Y, Z) = -\mathrm{Riem}(Z, X, Y, V).$$

In local coordinates, writing $R^i_{jk\ell} \partial_i = R(\partial_j, \partial_k)\partial_\ell$ for the components of R, the components of Riem are traditionally written with the same symbol but a lowered index, hence

$$R_{ijk\ell} := \mathrm{Riem}(\partial_i, \partial_j, \partial_k, \partial_\ell) = \langle \partial_i, R(\partial_j, \partial_k)\partial_\ell \rangle = \langle \partial_i, R^m_{jk\ell} \partial_m \rangle = g_{im} R^m_{jk\ell}.$$

The Riemann tensor satisfies additional symmetry relations beyond (27.1) that we will talk about next semester, but we will not yet need them at present.

EXERCISE 27.1. Assuming (M,g) is a 2-dimensional pseudo-Riemannian manifold, use the antisymmetry relations (27.1) to show that in any local coordinate system on an open set $\mathcal{U} \subset M$, the Riemann tensor is determined on $\mathcal{U}$ by the single component function $R_{1122} : \mathcal{U} \to \mathbb{R}$.

EXERCISE 27.2. The **Ricci curvature** is a tensor $\mathrm{Ric} \in \Gamma(T_2^0 M)$ derived from the Riemann tensor that plays a vital role in more advanced topics in Riemannian geometry, and also in general relativity. If (M,g) is a Riemannian manifold, it can be defined at any point $p \in M$ by choosing an orthonormal basis $e_1, \ldots, e_n \in T_p M$ and writing

$$(27.2) \qquad \mathrm{Ric}(Y,Z) := \sum_{j=1}^n \langle e_j, R(e_j, Y)Z \rangle = \sum_{j=1}^n \mathrm{Riem}(e_j, e_j, Y, Z) \in \mathbb{R}, \qquad \text{for } Y, Z \in T_p M.$$

You can convince yourself as follows that this is well defined:
- (a) Use the Einstein summation convention to give a one-line proof that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ for all pairs of square matrices $\mathbf{A}$ and $\mathbf{B}$.
- (b) Show that for linear maps $A : V \to V$ on a finite-dimensional vector space V, $\mathrm{tr}(A)$ can be defined as the trace of any matrix representing A in a basis, and it is independent of the choice of basis.
- (c) Show that $\mathrm{Ric}(Y,Z)$ according to (27.2) is the trace of the linear map $T_p M \to T_p M :$ $X \mapsto R(X,Y)Z$.
  Remark: This use of the trace demonstrates a general algebraic operation that can transform any tensor of type $(k+1, \ell+1)$ into a tensor of type (k, ℓ); it is known as a **contraction**. Notice that this also gives a definition of Ric that does not refer to the metric, and thus makes sense for an arbitrary connection on TM, including the Levi-Cività connection of an indefinite metric. (The formula (27.2) is not quite right in the indefinite case—can you see why not?)
- (d) Show that in local coordinates, the components $R_{k\ell}$ of Ric are given by $R_{k\ell} = R^i_{ik\ell}$.

A further simplification of the curvature tensor on a Riemannian manifold (M, g) can be obtained by contracting the Ricci tensor, giving rise to the **scalar curvature**

$$(27.3) \qquad \mathrm{Scal}(p) := \sum_{j=1}^{n} \mathrm{Ric}(e_j, e_j) = \sum_{j,k=1}^{n} \mathrm{Riem}(e_j, e_j, e_k, e_k) \in \mathbb{R},$$

where $e_1, \ldots, e_n \in T_pM$ again denotes an orthonormal basis. This defines a function $\mathrm{Scal} : M \to \mathbb{R}$.

(e) Show that (27.3) is independent of the choice of orthonormal basis $e_1, \ldots, e_n \in T_pM$ by reinterpreting it as the trace of the map $\mathrm{Ric}^\sharp : T_pM \to T_pM$ defined via the relation $\langle Y, \mathrm{Ric}^\sharp(Z) \rangle = \mathrm{Ric}(Y, Z)$ for $Y, Z \in T_pM$.

(f) Taking the trace in part (e) as a general definition of $\mathrm{Scal} : M \to \mathbb{R}$ for pseudo-Riemannian manifolds (M, g), rewrite (27.3) so that it is also valid when g is indefinite. (Note that unlike Ric, Scal does depend explicitly on g and not just on the connection, as the definition of $\mathrm{Ric}^\sharp$ depends on g.)

(g) Prove that if $\dim M = 2$, then the entire Riemann tensor is determined at each point $p \in M$ by the number $\mathrm{Scal}(p)$.
    Hint: Use Exercise 27.1 in coordinates chosen so that the coordinate vector fields are orthonormal at p.

(h) Show that in local coordinates, $\mathrm{Scal} = g^{k\ell} R^i{}_{ik\ell}$.

27.2. Locally flat metrics. A pseudo-Riemannian manifold (M, g) of dimension n is called **locally flat** if every point $p \in M$ admits a neighborhood $\mathcal{U} \subset M$ with a chart $(x^1, \ldots, x^n) : \mathcal{U} \to \mathbb{R}^n$ in which the components $g_{ij} = \langle \partial_i, \partial_j \rangle$ of the metric are constant functions. Recall that for metrics of signature (k, ℓ), a metric with constant components is equivalent via a linear transformation to the standard *flat metric* η of the same signature, which has components

$$\eta_{ij} := \begin{cases} 1 & \text{if } i = j \leqslant k, \\ -1 & \text{if } i = j > k, \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus (M, g) is locally flat if and only if it is *locally isometric* to the flat space $(\mathbb{R}^n, \eta)$, meaning every point has a neighborhood isometric to an open subset of $(\mathbb{R}^n, \eta)$.

We saw in §23.1 that it is always possible to find coordinates making g_{ij} match η_{ij} up to first order at a given point. Achieving $g_{ij} \equiv \eta_{ij}$ on an open neighborhood however is much more ambitious, and not usually possible. It requires an integrability condition, namely the vanishing of the curvature:

THEOREM 27.3. *A pseudo-Riemannian manifold (M, g) is locally flat if and only if its Riemann curvature tensor vanishes.*

PROOF. If $p \in M$ admits a neighborhood with a chart in which the components g_{ij} are constant, then the Christoffel symbols for the Levi-Cività connection vanish on this neighborhood, and by Exercise 26.6, so do the components $R^i{}_{jk\ell}$ of the Riemann tensor. Conversely, if $R \equiv 0$, then the Levi-Cività connection is flat, implying that any orthonormal basis $X_1, \ldots, X_n$ of T_pM can be extended to a neighborhood $\mathcal{U} \subset M$ of p as a family of parallel vector fields that form a frame for TM over $\mathcal{U}$. Since the connection is compatible with the metric, this frame is also orthonormal at every point, meaning $g(X_i, X_j) \equiv \eta_{ij}$. By the symmetry of the connection, we also have

$$[X_i, X_j] = \nabla_{X_i} X_j - \nabla_{X_j} X_i \equiv 0$$

since the vector fields $X_1, \ldots, X_n$ are all parallel. Theorem 25.11 now produces a chart near p in which $X_1, \ldots, X_n$ are the coordinate vector fields, and the components of g in this chart are precisely the constants η_{ij}. $\qquad\square$

EXERCISE 27.4. Prove that every Riemannian 1-manifold is locally flat. Give a direct proof, without mentioning the Riemann tensor. (You may have noticed that the latter vanishes for algebraic reasons whenever $\dim M = 1$.)

27.3. Gaussian curvature. The lowest dimension in which curvature is an interesting concept is 2. It was mentioned in §26.3 that for the Levi-Cività connection on a Riemannian 2-manifold (Σ, g), the Riemann curvature tensor is fully determined by a globally-defined real-valued function $K : \Sigma \to \mathbb{R}$. We would now like to clarify what geometric information this function carries, especially for surfaces embedded in $\mathbb{R}^3$.

We would also like to include the hyperbolic plane in this discussion, so in the following, we assume $\mathbb{R}^3$ with coordinates (x, y, z) is endowed with either the Euclidean or the Minkowski metric

$$g = \pm dx^2 + dy^2 + dz^2,$$

and $\Sigma \subset \mathbb{R}^3$ is a 2-dimensional Riemannian submanifold without boundary. For simplicity we also assume for now that Σ is orientable, though we will see that this assumption can be lifted. We will use the symbol

$$S_\pm^2 := \left\{ X \in \mathbb{R}^3 \mid \langle X, X \rangle = \pm 1 \right\} \subset \mathbb{R}^3$$

to denote either the unit sphere $S_+^2 := S^2$ or the two-sheeted hyperboloid $S_-^2 := \{x^2 - y^2 - z^2 = 1\}$, depending on whether $\langle \ , \ \rangle$ is the Euclidean or the Minkowski metric. An orientation of Σ now determines a unit normal vector field,

$$\nu \in \Gamma(T\Sigma^\perp) \subset \Gamma(T\mathbb{R}^3|_\Sigma), \qquad \langle \nu, \nu \rangle = \pm 1,$$

which is unique if we require that for every $p \in \Sigma$ and every positively-oriented basis (X, Y) of $T_p\Sigma$, $(\nu(p), X, Y)$ is a positively-oriented basis of $T_p\mathbb{R}^3 = \mathbb{R}^3$. Note that the sign of $\langle \nu, \nu \rangle$ is determined by the signature of $(\mathbb{R}^3, g)$: since we have assumed $\langle \ , \ \rangle$ is positive on $T\Sigma$, it must be positive on $T\Sigma^\perp$ if g is the Euclidean metric and negative for the Minkowski metric. This means that if we use the canonical isomorphisms $T_p\mathbb{R}^3 = \mathbb{R}^3$ to view ν as a map from Σ into $\mathbb{R}^3$, then it takes values in the submanifold $S_\pm^2$, giving a smooth map between surfaces

$$\nu : \Sigma \to S_\pm^2.$$

This is called the **Gauss map** of Σ. Its derivative at any point $p \in \Sigma$ has the following interesting property: $T_{\nu(p)}S_\pm^2 \subset \mathbb{R}^3$ is the orthogonal complement of $\nu(p)$, which is by definition the same subspace as $T_p\Sigma$, so the tangent map $T_p\nu$ defines a linear map of $T_p\Sigma$ to *itself*,

$$T_p\nu : T_p\Sigma \to T_p\Sigma.$$

LEMMA 27.5. *The map $T_p\nu : T_p\Sigma \to T_p\Sigma$ is self-adjoint with respect to the inner product $\langle \ , \ \rangle$.*

EXERCISE 27.6. Prove the lemma by showing that in some neighborhood $\mathcal{U} \subset \mathbb{R}^3$ of any point $p \in \Sigma$, ν can always be viewed as the restriction to Σ of the gradient of a function $f : \mathcal{U} \to \mathbb{R}$ for which $\Sigma \cap \mathcal{U} = f^{-1}(0)$. (Another proof of Lemma 27.5 will follow from more general considerations in the next section—see Remark 28.4.)

Applying the spectral theorem for self-adjoint operators, we conclude from Lemma 27.5 that $T_p\Sigma$ has an orthonormal basis X_1, X_2 consisting of eigenvectors of $T_p\nu$. The corresponding eigenspaces are called the **principal directions** of Σ at p, and their eigenvalues $\kappa_1, \kappa_2 \in \mathbb{R}$ with

$$T_p\nu(X_i) = \kappa_i X_i \qquad \text{for } i = 1, 2$$

are called the **principal curvatures** at p.

The principal curvatures at p can be interpreted in terms of the curvature of paths on Σ passing through p. In particular, fix a unit vector $X \in T_p\Sigma$ and choose a smooth path $\gamma : (-\epsilon, \epsilon) \to \Sigma$

with unit speed passing through $\gamma(0) = p$ such that $\dot{\gamma}(0) = X$. We can make some immediate observations about γ: first, since $\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle = 1$ is constant in t, differentiating it at $t = 0$ implies

$$\ddot{\gamma}(0) \in X^{\perp} \subset \mathbb{R}^3.$$

Second, $\langle \dot{\gamma}(t), \nu(\gamma(t)) \rangle = 0$ for all t since $\dot{\gamma}$ is tangent to Σ, and differentiating this at $t = 0$ then yields the relation

$$(27.4) \qquad\qquad - \langle \ddot{\gamma}(0), \nu(p) \rangle = \langle X, T_p\nu(X) \rangle =: \kappa_n(X),$$

implying that the component of $\ddot{\gamma}(0)$ pointing orthogonally to Σ depends only on the unit vector X and not on the choice of path γ. The number $\kappa_n(X) \in \mathbb{R}$ is called the **normal curvature** of Σ at p in the direction X.

REMARK 27.7. One popular interpretation of the normal curvature $\kappa_n(X)$ is expressed in terms of *plane curves*. Suppose $P \subset \mathbb{R}^3$ is a plane and $C \subset P$ is a 1-dimensional submanifold with a choice of normal vector field $\mathbf{n} \in \Gamma(TP|_C)$ along C. At any point $q \in C$, choose a smooth curve $\gamma : (-\epsilon, \epsilon) \to P$ through $\gamma(0) = q$ with unit speed $|\dot{\gamma}| \equiv 1$ that parametrizes a neighborhood of q in C. Differentiating the relation $\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle = 1$ then reveals that $\ddot{\gamma}(t)$ is always orthogonal to $\dot{\gamma}(t)$, hence

$$\ddot{\gamma}(t) = \kappa(\gamma(t))\mathbf{n}(\gamma(t))$$

for a uniquely-determined function $\kappa : C \to \mathbb{R}$. This function is independent of the choice of path γ parametrizing C; in particular, reversing the direction of γ does not change its second derivative as it passes through the same point. In this context, $\kappa : C \to \mathbb{R}$ is called the **curvature** of the curve $C \subset P$.

For the surface Σ with unit vector $X \in T_p\Sigma$, define $P \subset \mathbb{R}^3$ as the unique plane that contains p such that T_pP is spanned by X and $\nu(p)$. The intersection $P \cap \Sigma$ is then a smooth 1-dimensional submanifold near p, and the path γ in (27.4) can be chosen to be a parametrization of this submanifold, in which case $\nu(p)$ spans the orthogonal complement of $\dot{\gamma}(0) = X$ in T_pP. The normal curvature $\kappa_n(X)$ is therefore the curvature of the curve $P \cap \Sigma$ in the plane P at p.

REMARK 27.8. Yet another interpretation of $\kappa_n(X)$ comes from comparing geodesics in Σ with geodesics in the ambient space $\mathbb{R}^3$, also known as straight lines. If we choose γ in (27.4) to be the unique geodesic in Σ with initial velocity X, then Corollary 24.12 tells us $\ddot{\gamma}(0)$ is a scalar multiple of $\nu(p)$, and thus vanishes if and only if $\kappa_n(X) = 0$. From this perspective, $\kappa_n(X)$ measures the extent to which the geodesic in Σ with $\dot{\gamma}(0) = X$ deviates from being a geodesic in $\mathbb{R}^3$.

Fixing the orthonormal eigenvectors $X_1, X_2 \in T_p\Sigma$ of $T_p\nu$, every other unit vector takes the form $X = aX_1 + bX_2$ with $a^2 + b^2 = 1$, thus by (27.4)

$$\kappa_N(X) = \langle X, T_p\nu(X) \rangle = \langle aX_1 + bX_2, a\kappa_1 X_1 + b\kappa_2 X_2 \rangle = a^2\kappa_1 + b^2\kappa_2.$$

The range of values this number can take is precisely the interval in $\mathbb{R}$ bounded by the numbers κ_1 and κ_2, so this proves:

PROPOSITION 27.9. *The principal curvatures of $\Sigma \subset \mathbb{R}^3$ at $p \in \Sigma$ are the maximum and minimum values of the normal curvatures $\kappa_n(X)$ for all unit vectors $X \in T_p\Sigma$.* $\qquad\square$

Normal and principal curvatures are measurements of what is called the **extrinsic** curvature of Σ: they depend not just on the Riemannian metric of Σ but also on the way that Σ is embedded in $\mathbb{R}^3$. By contrast, the next object we will define is **intrinsic**, meaning it depends only on the metric and is thus an invariant of Riemannian 2-manifolds (Σ, g) up to isometry. This will not be obvious from the definition—proving that it is intrinsic will require a substantial effort.

DEFINITION 27.10. For a Riemannian hypersurface Σ in $\mathbb{R}^3$ with the Euclidean or Minkowski metric $g = \pm dx^2 + dy^2 + dz^2$, the **Gaussian curvature** of Σ at $p \in \Sigma$ is defined (up to a sign) as the product of its principal curvatures, that is,

$$(27.5) \qquad\qquad\qquad K_G(p) := \pm \kappa_1 \kappa_2 \in \mathbb{R},$$

where the symbol $\pm$ means $+$ if g is the Euclidean metric and $-$ for the Minkowski metric. Equivalently, $K_G(p)$ is determined from the Gauss map $\nu : \Sigma \to S^2_\pm$ as

$$K_G(p) = \pm \det \left(T_p\Sigma \xrightarrow{T_p\nu} T_p\Sigma \right).$$

REMARK 27.11. For an arbitrary n-dimensional vector space V over the field $\mathbb{F}$, one can define the determinant of a linear map $A : V \to V$ as $\det(\mathbf{A}) \in \mathbb{F}$ where $\mathbf{A} \in \mathbb{F}^{n \times n}$ is the matrix representing A in any choice of basis. The result is independent of the choice of basis since for any $\mathbf{B} \in \mathrm{GL}(n, \mathbb{F})$, $\det(\mathbf{B}\mathbf{A}\mathbf{B}^{-1}) = \det(\mathbf{A})$.

REMARK 27.12. The normal and principal curvatures all depend on the choice of normal vector field ν, but the Gaussian curvature does not, because reversing ν causes a sign change in both κ_1 and κ_2, leaving $K_G(p)$ invariant. For this reason, the Gaussian curvature can be defined even if Σ is not orientable.

For surfaces in Euclidean space, the formula $K_G(p) = \det(T_p\nu)$ implies that the Gaussian curvature is positive in any region where the Gauss map is orientation preserving, and negative wherever it is orientation reversing. It vanishes at any point where $T_p\nu$ collapses $T_p\Sigma$ to a subspace of lower dimension.

EXAMPLE 27.13. For the unit sphere $S^2 \subset \mathbb{R}^3$ in Euclidean space, the Gauss map is simply the identity, so $K_G \equiv 1$.

EXAMPLE 27.14. Consider the cylinder $Z = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 = 1\}$ in Euclidean space. The Gauss map on Z is independent of z, thus $T_p\nu$ only has rank 1 at every $p \in Z$, implying $K(p) = 0$. By Theorem 27.3 and Theorem 27.17 below, this result is equivalent to the observation that Z is locally flat: unlike a sphere, a small piece of a cylinder can easily be unfolded into a piece of a flat plane without changing lengths or angles on the surface. The same is true of the cone

$$C = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 = z^2, \ z > 0\}.$$

It is easy to check that Z and C do have nontrivial normal and principal curvatures, showing that the latter are indeed extrinsic, i.e. they depend on the specific embeddings of these surfaces in $\mathbb{R}^3$ and are not isometry invariants.

EXAMPLE 27.15. The hyperboloid $H = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 = 1\}$ in Euclidean space has everywhere negative curvature. (For a precise computation, see Exercise 27.20 below.) This is true of any surface that exhibits a "saddle" shape, for which the Gauss map is orientation reversing.

EXAMPLE 27.16. The hyperbolic plane H^2 was defined in §24.4.3 as the upper sheet of the hyperboloid S^2_- in $\mathbb{R}^3$ with the Minkowski metric. This would have positive curvature if it lived in Euclidean space, but in Minkowski space the extra sign in Definition 27.10 becomes relevant, so the curvature is negative. The situation is in fact very much analogous to the sphere in Example 27.13, because the Gauss map in this case is just the identity map on the upper sheet of S^2_-, giving $\det(T_p\nu) = 1$ at every point. We conclude $K_G \equiv -1$.

The next big result says that $K_G : \Sigma \to \mathbb{R}$ is determined by the Riemann curvature tensor, and therefore by the Riemannian metric on Σ. In fact, K_G turns out to be the same function that appeared in (26.3):

THEOREM 27.17. *Suppose Σ is an oriented Riemannian hypersurface embedded in Euclidean or Minkowski $\mathbb{R}^3$, $d\mathrm{vol} \in \Omega^2(\Sigma)$ denotes its Riemannian area form, $K_G : \Sigma \to \mathbb{R}$ is its Gaussian curvature, $R(X,Y)Z$ is its Riemann curvature tensor and $J : T\Sigma \to T\Sigma$ is the unique fiberwise linear map such that for any vector $X \in T_p\Sigma$ with $|X| = 1$, (X, JX) is a positively-oriented orthonormal basis. Then*

$$R(X,Y)Z = -K_G\, d\mathrm{vol}(X,Y)JZ.$$

We will prove this theorem in the next lecture. For arbitrary Riemannian 2-manifolds Σ, not embedded in $\mathbb{R}^3$, Theorem 27.17 can be taken as a definition of the Gaussian curvature $K_G : \Sigma \to \mathbb{R}$. Note that once again the result doesn't actually depend on an orientation (cf. Remark 27.12): locally, if the orientation of Σ is flipped, this changes the sign of both J and $d\mathrm{vol}$, leaving the function K_G unchanged.

For surfaces in Euclidean $\mathbb{R}^3$, Theorem 27.17 implies the following famous result of Gauss, which has come to be known by the Latin term for "remarkable theorem":

THEOREMA EGREGIUM. *For a surface Σ embedded in Euclidean $\mathbb{R}^3$, the Gaussian curvature $K_G : \Sigma \to \mathbb{R}$ defined in (27.5) is an invariant of the induced Riemannian metric on Σ. To be precise, if $\Sigma_1, \Sigma_2 \subset \mathbb{R}^3$ are two surfaces embedded in $\mathbb{R}^3$ with induced metrics g_1, g_2 and Gaussian curvatures K_G^1, K_G^2 respectively, and $\varphi : (\Sigma_1, g_1) \to (\Sigma_2, g_2)$ is an isometry, then*

$$K_G^1 \equiv K_G^2 \circ \varphi.$$

Example 27.14 shows that nothing similar to the Theorema Egregium is true for the normal or principal curvatures of a surface. Here are a couple of sample applications:

- There are no isometries between any open subsets of the sphere $S^2 \subset \mathbb{R}^3$ (positive curvature) and the hyperboloid of Example 27.15 or the hyperbolic plane in Example 27.16 (negative curvature).
- A Riemannian 2-manifold Σ embedded in Euclidean or Minkowski $\mathbb{R}^3$ is locally flat if and only if at least one of its principal curvatures vanishes at every point.

EXERCISE 27.18. Given a constant $r > 0$, compute K_G for:

(a) The sphere $\{x^2 + y^2 + z^2 = r^2\}$ of radius r in Euclidean $\mathbb{R}^3$;
(b) The rescaled hyperbolic plane $\{x^2 - y^2 - z^2 = r^2,\ x > 0\}$ in Minkowski $\mathbb{R}^3$.

We can deduce from Theorem 27.17 a formula for K_G in terms of the Riemann tensor. We begin by observing that the metric $\langle\ ,\ \rangle$, area form $d\mathrm{vol} \in \Omega^2(\Sigma)$ and fiberwise-linear map $J \in \Gamma(\mathrm{End}(T\Sigma))$ satisfy the relation

(27.6) $$d\mathrm{vol}(X,Y) = \langle JX, Y \rangle.$$

To see this, notice first that $(X, Y) \mapsto \langle JX, Y \rangle$ is an alternating 2-form, since J is an orthogonal transformation with $J^2 = -\mathbb{1}$, so

$$\langle JY, X \rangle = \langle J(JY), JX \rangle = \langle -Y, JX \rangle = -\langle JX, Y \rangle.$$

The 2-form $\langle J\cdot, \cdot \rangle$ is therefore a scalar multiple of $d\mathrm{vol}$ at every point, so it suffices to check that they match when evaluated on some particular basis at each point. This is true for instance for any basis of the form (X, JX) with $|X| = 1$, as this basis is positively oriented and orthonormal, so $d\mathrm{vol}(X, JX) = 1 = \langle JX, JX \rangle$, proving (27.6). Theorem 27.17 now implies

$$\langle R(X,Y)Y, X \rangle = -\langle K_G\, d\mathrm{vol}(X,Y)JY, X \rangle = -K_G\, d\mathrm{vol}(X,Y)\langle JY, X \rangle = K_G \cdot |d\mathrm{vol}(X,Y)|^2,$$

so we can write

(27.7) $$K_G(p) = \frac{\langle R(X,Y)Y, X \rangle}{|d\mathrm{vol}(X,Y)|^2} = \frac{\mathrm{Riem}(X,X,Y,Y)}{|d\mathrm{vol}(X,Y)|^2} \qquad \text{for any basis } X, Y \in T_p\Sigma.$$

We can rewrite this as follows in terms of an oriented coordinate chart (x^1, x^2) defined near p. If the components of the metric are denoted by g_{ij} and we define the symmetric matrix-valued function

$$\mathbf{g} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix},$$

we recall from Exercise 11.12 that $d$vol takes the form

$$d\text{vol} = \sqrt{\det \mathbf{g}}\ dx^1 \wedge dx^2.$$

Then applying (27.7) to the coordinate vectors $X = \partial_1$ and $Y = \partial_2$, we obtain the formula

(27.8) $$K_G = \frac{R_{1122}}{\det \mathbf{g}}.$$

If you did not already understand why the Theorema Egregium follows from Theorem 27.17, we can now prove it as follows:

PROOF OF THE THEOREMA EGREGIUM. If $\varphi : (\Sigma_1, g_1) \to (\Sigma_2, g_2)$ is an isometry and $p \in \Sigma_1$, then any chart $(\mathcal{U}, x)$ on a neighborhood $\mathcal{U} \subset \Sigma_2$ of $q := \varphi(p)$ gives rise to a chart $(\varphi^{-1}(\mathcal{U}), x \circ \varphi)$ on a neighborhood $\varphi^{-1}(\mathcal{U}) \subset \Sigma_1$ of p such that the components of the two metrics in these charts are related by $(g_1)_{ij} = (g_2)_{ij} \circ \varphi$. It follows that the components of their Riemann tensors and their Riemannian volume forms satisfy a similar relation, so by (27.8), so do their Gaussian curvatures. $\square$

EXERCISE 27.19. A Riemannian manifold (M, g) is called **homogeneous** if for every pair of points $p, q \in M$, there exists an isometry $\varphi \in \text{Isom}(M, g)$ such that $\varphi(p) = q$. Show that every homogeneous Riemannian 2-manifold has constant Gaussian curvature.
Remark: This partly explains why I claimed in §24.2 that one should not generally expect nontrivial isometries to exist. Constant curvature is a very delicate condition that is easy to destroy via small perturbations of the metric.

EXERCISE 27.20. Prove that for the hyperboloid $H \subset \mathbb{R}^3$ in Example 27.15,

$$K_G(x, y, z) = -\frac{1}{(x^2 + y^2 + z^2)^2}.$$

Hint: This can be a horrible computation, but it doesn't have to be. For instance, there are some obvious isometries that make it sufficient to consider a point of the form $(r, 0, z) \in H$ with $r^2 - z^2 = 1$, which is the intersection of the smooth curves $\alpha(t) = (\cosh t, 0, \sinh t)$ and $\beta(t) = (r \cos t, r \sin t, z)$ in H. Since H is a level set of $f(x, y, z) = x^2 + y^2 - z^2$, there is a unit normal vector field of the form $\nu = g \cdot \nabla f$ for some function $g : H \to (0, \infty)$. Try to convince yourself without any calculations that the curves α and β are tangent to the principal directions. Then consider the following: if you know $\gamma(t) \in H$ satisfies $\frac{d}{dt} \nu(\gamma(t)) = \lambda \dot{\gamma}(t)$ for some $\lambda \in \mathbb{R}$, what happens if you take the inner product of both sides with $\dot{\gamma}(t)$? Write $\nu = g \cdot \nabla f$ and use this observation to compute the two principal curvatures at $(r, 0, z)$. You will need to write down the function g for this, but you should not need to differentiate it.
Final remark: It's also possible there's an easier way to do this that I haven't thought of.

EXERCISE 27.21. Show that for any Riemannian 2-manifold (Σ, g), the scalar curvature defined in Exercise 27.2 is related to the Gaussian curvature by $\text{Scal} = 2K_G$.
Hint: Given a point $p \in \Sigma$, use coordinates for which ∂_1 and ∂_2 are orthonormal at p.

EXERCISE 27.22. Show that the Poincaré half-plane $(\mathbb{H}, h)$ from Exercise 22.8 has constant Gaussian curvature $K_G \equiv -1$.
Remark: You knew this already from Example 27.16 if you had already convinced yourself that

$(\mathbb{H}, h)$ *is isometric to the hyperbolic plane (see Exercise 24.16). But you can also compute this directly from (27.8) if you first work out the Christoffel symbols of the connection on $(\mathbb{H}, h)$ and then compute the Riemann tensor via Exercise 26.6.*

REMARK 27.23. The hyperbolic plane is a funny animal. It is the most famous and most important example of a surface with constant negative curvature—in fact it is known to be the *only* one up to isometry and scaling that is both simply connected and geodesically complete— but you may have noticed that we've never mentioned any model of it that one can look at it and say, "yes, that looks like a surface with negative curvature!". The closest thing we have is the hyperboloid model in Minkowski space, which actually looks like a *positively* curved surface, but acquires an extra minus sign in Definition 27.10, which is difficult to justify intuitively. (The justification for it is that if the minus sign were not there, Theorem 27.17 would not be true.) What I'm getting at is this: it would be nice if we could view H^2 as an embedded hypersurface in *Euclidean* $\mathbb{R}^3$ whose "saddle" shape would make the negativity of its curvature obvious. There exist *local* models of this kind, e.g. the *pseudosphere* (also called the *tractricoid*)[72] is a surface in Euclidean $\mathbb{R}^3$ that is isometric to an open subset of H^2, but not the whole thing. The reason I have not explained any global model of H^2 in Euclidean 3-space is that according to a famous theorem of Hilbert, it is impossible: there exists no embedding (nor even an immersion!) of any geodesically complete surface with constant negative curvature into Euclidean $\mathbb{R}^3$. I'd conjecture that if this theorem were not true, it would have been recognized somewhat earlier in history that Euclid's first four postulates do not imply the fifth.

28. Properties of Gaussian curvature

I owe you a proof of Theorem 27.17, specifically the formula

$$R(X, Y)Z = -K_G \, d\mathrm{vol}(X, Y)JZ,$$

which relates the Gaussian curvature $K_G : \Sigma \to \mathbb{R}$ to the Riemann tensor $R \in \Gamma(T^1_3\Sigma)$ for a Riemannian hypersurface Σ in Euclidean or Minkowski 3-space $(\mathbb{R}^3, \pm dx^2 + dy^2 + dz^2)$. We'll take care of this in §28.1 by developing a general formula to compare the Riemann tensor of any pseudo-Riemannian manifold with that of a pseudo-Riemannian submanifold embedded in it. After that, we will restrict again to dimension 2 and examine some further properties of the Gaussian curvature, in preparation for proving the Gauss-Bonnet formula.

28.1. The second fundamental form. Assume (M, g) is a pseudo-Riemannian manifold with $\dim M > n \geqslant 2$ that contains

$$\Sigma \subset M$$

as an n-dimensional pseudo-Riemannian submanifold with inclusion map $j : \Sigma \hookrightarrow M$, so the induced metric j^*g on Σ is also nondegenerate. In this situation, Corollary 24.9 produces a direct sum decomposition

$$TM|_\Sigma = T\Sigma \oplus T\Sigma^\perp,$$

so that every $X \in T_pM$ for $p \in \Sigma$ is uniquely expressible as

$$X = X^\top + X^\perp, \qquad X^\top \in T_p\Sigma \text{ and } X^\perp \in (T_p\Sigma)^\perp \subset T_pM.$$

In this notation, the Levi-Cività connections ∇ and $\widehat{\nabla}$ of (M, g) and (Σ, j^*g) are related according to Proposition 24.11 by

$$\widehat{\nabla}_X Y = (\nabla_X Y)^\top.$$

---

[72]See https://en.wikipedia.org/wiki/Pseudosphere

A vector field $X(t) \in T_{\gamma(t)}\Sigma$ on Σ along a path $\gamma(t) \in \Sigma$ is thus parallel if and only if $\nabla_t X \in T\Sigma^\perp$, and since this allows $\nabla_t X$ to be nonzero, X may fail to be parallel when regarded as a vector field on M along γ. This failure can be measured by a tensor:

LEMMA 28.1. *There exists a symmetric bilinear bundle map* $\mathrm{II} : T\Sigma \oplus T\Sigma \to T\Sigma^\perp$ *such that for any pair of vector fields* $X, Y \in \mathfrak{X}(\Sigma)$,

$$\mathrm{II}(X, Y) = (\nabla_Y X)^\perp.$$

In particular, the connections ∇ *on* M *and* $\widehat{\nabla}$ *on* Σ *are then related to each other by*

$$\nabla_Y X = \widehat{\nabla}_Y X + \mathrm{II}(X, Y).$$

PROOF. We can define $\mathrm{II} : \mathfrak{X}(\Sigma) \times \mathfrak{X}(\Sigma) \to \Gamma(T\Sigma^\perp)$ by $\mathrm{II}(X, Y) := (\nabla_Y X)^\perp$. The symmetry of II then follows easily from the symmetry of the Levi-Cività connection: extending X and Y arbitrarily to vector fields on M, we have

$$\mathrm{II}(Y, X) - \mathrm{II}(X, Y) = (\nabla_X Y - \nabla_Y X)^\perp = [X, Y]^\perp = 0,$$

since X and Y taking values in $T\Sigma$ along Σ implies that the same is true for $[X, Y]$. Now since $\mathrm{II}(X, Y)$ is manifestly C^∞-linear in Y, the symmetry implies that it is also C^∞-linear in X, and therefore gives a well-defined bundle map $T\Sigma \oplus T\Sigma \to T\Sigma^\perp$. $\qquad\square$

DEFINITION 28.2. The symmetric bundle map $\mathrm{II} : T\Sigma \oplus T\Sigma \to T\Sigma^\perp$ in Lemma 28.1 is called the **second fundamental form**[73] of the submanifold $\Sigma \subset M$.

We can now associate to any normal section $\nu \in \Gamma(T\Sigma^\perp)$ the symmetric tensor field $\mathrm{II}_\nu \in \Gamma(T_2^0\Sigma)$ defined by

$$\mathrm{II}_\nu(X, Y) := \langle \nu, \mathrm{II}(X, Y)\rangle.$$

This is especially useful in the case where $\Sigma \subset M$ is a hypersurface with orientable normal bundle, as Σ then admits a *unit* normal vector field $\nu \in \Gamma(T\Sigma^\perp)$ that is unique up to a sign. The words "second fundamental form" are also sometimes used to refer to the symmetric tensor $\mathrm{II}_\nu \in \Gamma(T_2^0\Sigma)$ in this special case.

REMARK 28.3. By now you may be wondering: what is the *first* fundamental form? This term was traditionally used for another symmetric $(0, 2)$-tensor on Σ, namely the restricted metric $\langle\,,\,\rangle|_{T\Sigma} = j^*g$. But the term has fallen somewhat out of fashion.

Since II_ν is a symmetric bilinear form on the tangent spaces of Σ for each normal section $\nu \in \Gamma(T\Sigma^\perp)$, it corresponds via the relation

$$\mathrm{II}_\nu(X, Y) = \langle X, W_\nu(Y)\rangle$$

to a unique bundle map $W_\nu : T\Sigma \to T\Sigma$ that is self-adjoint with respect to the bundle metric on $T\Sigma$. We call W_ν the **Weingarten map** associated to the normal section ν. One obtains a more revealing formula for it by differentiating the relation $\langle X, \nu\rangle \equiv 0$, which holds for any $X \in \mathfrak{X}(\Sigma)$ and $\nu \in \Gamma(T\Sigma^\perp)$: we find

$$0 = \mathcal{L}_Y \langle X, \nu\rangle = \langle \nabla_Y X, \nu\rangle + \langle X, \nabla_Y \nu\rangle = \langle \widehat{\nabla}_Y X + \mathrm{II}(X, Y), \nu\rangle + \langle X, (\nabla_Y \nu)^\top + (\nabla_Y \nu)^\perp\rangle$$

$$= \mathrm{II}_\nu(X, Y) + \langle X, (\nabla_Y \nu)^\top\rangle = \langle X, W_\nu(Y) + (\nabla_Y \nu)^\top\rangle,$$

having discarded terms that vanish due to orthogonality. The result is an interpretation of the Weingarten map as the tangential part of the covariant derivative of ν:

$$(28.1) \qquad\qquad\qquad W_\nu(X) = -(\nabla_X \nu)^\top.$$

---

[73]Do not be misled by this use of the word "form"; II is not a differential form in any sense, as it is symmetric rather than antisymmetric.

REMARK 28.4. If $\Sigma \subset M$ is a hypersurface with an orientable normal bundle, then there are two canonical choices of $\nu \in \Gamma(T\Sigma^\perp)$ determined by the normalization condition $\langle \nu, \nu \rangle \equiv \pm 1$, where the sign depends on the signatures of (M, g) and (Σ, j^*g). Differentiating $\langle \nu, \nu \rangle$ now reveals that $\langle \nabla_X \nu, \nu \rangle \equiv 0$ for all $X \in \mathfrak{X}(\Sigma)$, and $\nabla_X \nu$ is therefore tangent to Σ, thus (28.1) simplifies to

$$W_\nu(X) = -\nabla_X \nu.$$

This particular form of the Weingarten map is sometimes called the **shape operator**. In the important special case where M is $\mathbb{R}^3$ with the Euclidean or Minkowski metric, ∇ is the trivial connection, so $-W_\nu : T\Sigma \to T\Sigma$ is now the derivative of the Gauss map introduced in §27.3. The self-adjointness of W_ν thus gives a second proof of Lemma 27.5, and the Gaussian curvature of (Σ, j^*g) in this situation is $\pm \det(W_\nu)$.

Like the Gauss map and the principal curvatures in §27.3, the Weingarten map and second fundamental form belong to the *extrinsic* rather than *intrinsic* geometry of (Σ, j^*g), meaning they depend on the way that Σ is embedded as a pseudo-Riemannian submanifold of (M, g), rather than intrinsically on the metric j^*g. They are not deeply meaningful objects, but they turn out to be useful tools for deriving the Riemann tensor of (Σ, j^*g) from that of (M, g). In the following, we denote by

$$R \in \Gamma(T_3^1 M), \qquad \widehat{R} \in \Gamma(T_3^1 \Sigma)$$

the Riemann curvature tensors of (M, g) and (Σ, j^*g) respectively, along with their covariant versions

$$\text{Riem} \in \Gamma(T_4^0 M), \qquad \widehat{\text{Riem}} \in \Gamma(T_4^0 \Sigma)$$

as defined in §27.1.

PROPOSITION 28.5 (Gauss equation). *The tensors* Riem *and* $\widehat{\text{Riem}}$ *are related by*

$$\widehat{\text{Riem}}(V, X, Y, Z) = \text{Riem}(V, X, Y, Z) + \langle \text{II}(V, X), \text{II}(Y, Z) \rangle - \langle \text{II}(V, Y), \text{II}(X, Z) \rangle.$$

PROOF. We observe first that for any tuple of vector fields $V, X, Y, Z \in \mathfrak{X}(\Sigma)$, differentiating the relation $\langle V, \text{II}(Y, Z) \rangle \equiv 0$ with respect to X gives

$$(28.2) \qquad \langle V, \nabla_X (\text{II}(Y, Z)) \rangle = -\langle \nabla_X V, \text{II}(Y, Z) \rangle = -\langle \text{II}(X, V), \text{II}(Y, Z) \rangle,$$

where $\nabla_X V$ can be replaced by its normal part $\text{II}(X, V)$ in the last expression because the inner product of its tangential part with $\text{II}(Y, Z)$ necesarily vanishes. The same trick allows us in the following calculation to replace $\widehat{\nabla}$ with ∇ in several places since we are taking the inner product with V; applying also (28.2) and the relation $\nabla_X Y = \widehat{\nabla}_X Y + \text{II}(X, Y)$, we find

$$\begin{aligned}
\widehat{\text{Riem}}(V, X, Y, Z) &= \langle V, \widehat{R}(X, Y)Z \rangle = \langle V, \widehat{\nabla}_X \widehat{\nabla}_Y Z - \widehat{\nabla}_Y \widehat{\nabla}_X Z - \widehat{\nabla}_{[X,Y]} Z \rangle \\
&= \langle V, \nabla_X (\nabla_Y Z - \text{II}(Y, Z)) - \nabla_Y (\nabla_X Z - \text{II}(X, Z)) - \nabla_{[X,Y]} Z \rangle \\
&= \langle V, R(X, Y)Z \rangle + \langle \text{II}(V, X), \text{II}(Y, Z) \rangle - \langle \text{II}(V, Y), \text{II}(X, Z) \rangle.
\end{aligned}$$

$\square$

Now let's specialize this to a situation closer to that of Theorem 27.17. We assume (M, g) is a locally flat pseudo-Riemannian 3-manifold, and $\Sigma \subset M$ is a Riemannian hypersurface. In this case $T\Sigma^\perp$ is a line bundle over Σ on which the bundle metric $\langle \, , \, \rangle$ is nondegenerate, and it may be either positive or negative, depending on whether (M, g) has Riemannian signature $(3, 0)$ or Lorentz signature $(2, 1)$, which are the only two possibilites since we are assuming (Σ, g) is Riemannian. As usual it will also be convenient to assume that both Σ and its normal bundle $T\Sigma^\perp$

are orientable, though these assumptions will both be seen to be inessential in the end. Fixing an orientation of Σ determines the fiberwise-linear map

$$J : T\Sigma \to T\Sigma$$

that rotates each tangent space counterclockwise by 90 degrees. The orientability of $T\Sigma^\perp$ allows us in turn to choose a (unique up to a sign) unit normal vector field

$$\nu \in \Gamma(T\Sigma^\perp), \qquad \langle \nu, \nu \rangle \equiv \pm 1,$$

where the sign is positive if (M, g) is Riemannian and negative otherwise; in the following we will make consistent use of the symbol "$\pm$" for this sign, and write "$\mp$" whenever it gets reversed. For example, the symmetric tensors $\mathrm{II}(X, Y)$ and $\mathrm{II}_\nu(X, Y)$ are now related to each other by

$$\mathrm{II}(X, Y) = \pm \mathrm{II}_\nu(X, Y)\nu,$$

and in light of Remark 28.4, $\nabla\nu|_{T\Sigma}$ matches the shape operator $-W_\nu : T\Sigma \to T\Sigma$ and is thus related to the second fundamental form by

$$(28.3) \qquad \mathrm{II}(X, Y) = \pm \mathrm{II}_\nu(X, Y)\nu = \pm\langle X, W_\nu(Y)\rangle\nu = \mp\langle X, \nabla_Y\nu\rangle\nu.$$

Fix a point $p \in \Sigma$ and write $\nabla\nu(p) : T_p\Sigma \to T_p\Sigma$ for the restriction of $\nabla\nu$ to the tangent space at p. The symmetry of II_ν implies that $\nabla\nu(p)$ is self-adjoint with respect to the inner product $\langle\ ,\ \rangle$ on $T_p\Sigma$, thus it has an orthonormal basis of eigenvectors $X_1, X_2 \in T_p\Sigma$, and we are free to order them so that

$$X_2 = JX_1 \qquad \text{and} \qquad X_1 = -JX_2,$$

in which case they are also a positively-oriented basis and thus satisfy

$$(28.4) \qquad\qquad\qquad d\mathrm{vol}(X_1, X_2) = 1$$

for the Riemannian volume form $d\mathrm{vol} \in \Omega^2(\Sigma)$. In the case where M is the Euclidean or Minkowski $\mathbb{R}^3$, the corresponding eigenvalues

$$\kappa_1, \kappa_2 \in \mathbb{R}$$

are the principal curvatures of Σ at p. We can now use this data to turn Proposition 28.5 into a more explicit formula for the Riemann tensor of (Σ, j^*g) at p: we assumed (M, g) is flat, so $R \equiv 0$, and thus for $V, X, Y, Z \in T_p\Sigma$, using (28.3) to replace various terms in the Gauss equation gives

$$\langle V, \widehat{R}(X_1, X_2)Z \rangle = \langle \mathrm{II}(V, X_1), \mathrm{II}(X_2, Z)\rangle - \langle \mathrm{II}(V, X_2), \mathrm{II}(X_1, Z)\rangle$$
$$= \big\langle\langle V, \nabla\nu(p)X_1\rangle\nu(p), \langle Z, \nabla\nu(p)X_2\rangle\nu(p)\big\rangle - \big\langle\langle V, \nabla\nu(p)X_2\rangle\nu(p), \langle Z, \nabla\nu(p)X_1\rangle\nu(p)\big\rangle$$
$$= \pm\kappa_1\kappa_2\big(\langle V, X_1\rangle\cdot\langle Z, X_2\rangle - \langle V, X_2\rangle\cdot\langle Z, X_1\rangle\big)$$
$$= \pm\kappa_1\kappa_2\big\langle V, \langle Z, X_2\rangle X_1 - \langle Z, X_1\rangle X_2\big\rangle,$$

which implies

$$\widehat{R}(X_1, X_2)Z = \pm\kappa_1\kappa_2\left(\langle Z, X_2\rangle X_1 - \langle Z, X_1\rangle X_2\right).$$

Finally, we observe that since J maps $T_p\Sigma \to T_p\Sigma$ orthogonally and the vectors $X_2 = JX_1$ and $X_1 = -JX_2$ form an orthonormal basis,

$$\langle Z, X_2\rangle X_1 - \langle Z, X_1\rangle X_2 = \langle JZ, JX_2\rangle X_1 - \langle JZ, JX_1\rangle X_2 = -\langle X_1, JZ\rangle X_1 - \langle X_2, JZ\rangle X_2 = -JZ,$$

and combining this with (28.4) we thus obtain

$$\widehat{R}(X_1, X_2)Z = \mp\kappa_1\kappa_2 JZ = -\pm\kappa_1\kappa_2\, d\mathrm{vol}(X_1, X_2)JZ.$$

We already know there exists a unique function $K : \Sigma \to \mathbb{R}$ such that the relation $\widehat{R}(X_1, X_2)Z = -K(p)\, d\mathrm{vol}(X_1, X_2)JZ$ is satisfied, so the conclusion of this calculation is that $K(p) = \pm\kappa_1\kappa_2$, i.e. it is the Gaussian curvature K_G. This completes the proof of Theorem 27.17.

REMARK 28.6. While $\Sigma \subset M$ was allowed to be a pseudo-Riemannian submanifold of arbitrary signature in most of this section, the positivity of j^*g became essential as soon as we started talking about the orthonormal eigenvectors of the shape operator $\nabla\nu(p) : T_p\Sigma \to T_p\Sigma$. This is indeed a self-adjoint operator with respect to the bundle metric $\langle \ , \ \rangle$ in every case, but the spectral theorem does not hold in general with indefinite inner products.

28.2. Local curvature 2-forms. We haven't mentioned it in a couple of lectures, but in addition to the Riemann tensor $R \in \Gamma(T_3^1\Sigma)$, the curvature of a Riemannian 2-manifold (Σ, g) can also be characterized via a differential 2-form, the curvature 2-form $\Omega_K \in \Omega^2(\Sigma, \operatorname{End}(T\Sigma))$. You might wonder: what happens if we integrate it? This question doesn't make much sense at first glance, as Ω_K is a bundle-valued 2-form, so it's not clear what $\int_\Sigma \Omega_K$ should mean. In order to clarify this, I'd like to expand on an exercise that was stated at the end of §26.4.

Assume $\pi : E \to M$ is a vector bundle with structure group $G \subset \operatorname{GL}(m, \mathbb{F})$, denote the Lie algebra of G by $\mathfrak{g} \subset \mathbb{F}^{m \times m}$, and suppose ∇ is a G-compatible connection. We recall that for every G-compatible local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{F}^m$, ∇ can be described over $\mathcal{U}_\alpha$ via a connection 1-form $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{g})$, defined so that

$$(28.5) \qquad (\nabla_X v)_\alpha = \mathcal{L}_X v_\alpha + A_\alpha(X) v_\alpha$$

for any $X \in \mathfrak{X}(\mathcal{U}_\alpha)$. Here $v_\alpha : \mathcal{U}_\alpha \to \mathbb{F}^m$ expresses $v|_{\mathcal{U}_\alpha} \in \Gamma(E|_{\mathcal{U}_\alpha})$ with respect to the trivialization, meaning $\Phi_\alpha(v(p)) = (p, v_\alpha(p))$ for $p \in \mathcal{U}_\alpha$. The corresponding **local curvature 2-form** $F_\alpha \in \Omega^2(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$ is defined as the local representation of $\Omega_K \in \Omega^2(M, \operatorname{End}(E))$ with respect to this trivialization, meaning that for $X, Y \in \mathfrak{X}(\mathcal{U}_\alpha)$ and $v \in \Gamma(E|_{\mathcal{U}_\alpha})$,

$$(\Omega_K(X, Y)v)_\alpha = F_\alpha(X, Y)v_\alpha.$$

Let's compute $F_\alpha \in \Omega^2(\mathcal{U}_\alpha, \mathbb{F}^{m \times m})$ in terms of $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{g})$. By Theorem 26.9, we can use the Riemann tensor as a substitute for Ω_K, so plugging in the definition of $R(X, Y)v$ with a section $v \in \Gamma(E)$ and using (28.5), we find

$$\begin{aligned}
(\Omega_K(X, Y)v)_\alpha &= \left(\nabla_X\nabla_Y v - \nabla_Y\nabla_X v - \nabla_{[X,Y]}v\right)_\alpha \\
&= (\mathcal{L}_X + A_\alpha(X))(\mathcal{L}_Y + A_\alpha(Y))v_\alpha - (\mathcal{L}_Y + A_\alpha(Y))(\mathcal{L}_X + A_\alpha(X))v_\alpha \\
&\quad - \left(\mathcal{L}_{[X,Y]} + A_\alpha([X,Y])\right)v_\alpha \\
&= \left(\mathcal{L}_X\mathcal{L}_Y - \mathcal{L}_Y\mathcal{L}_X - \mathcal{L}_{[X,Y]}\right)v_\alpha \\
&\quad + A_\alpha(X)\mathcal{L}_Y v_\alpha + A_\alpha(Y)\mathcal{L}_X v_\alpha - A_\alpha(Y)\mathcal{L}_X v_\alpha - A_\alpha(X)\mathcal{L}_Y v_\alpha \\
&\quad + \left(\mathcal{L}_X(A_\alpha(Y)) - \mathcal{L}_Y(A_\alpha(X)) - A_\alpha([X,Y])\right)v_\alpha \\
&\quad + (A_\alpha(X)A_\alpha(Y) - A_\alpha(Y)A_\alpha(X))v_\alpha \\
&= (dA_\alpha(X, Y) + [A_\alpha(X), A_\alpha(Y)])v_\alpha,
\end{aligned}$$

where in the last line, we've introduced the **matrix commutator**

$$[\mathbf{A}, \mathbf{B}] := \mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A} \qquad \text{for } \mathbf{A}, \mathbf{B} \in \mathbb{F}^{m \times m}.$$

The formula for F_α is thus

$$(28.6) \qquad F_\alpha(X, Y) = dA_\alpha(X, Y) + [A_\alpha(X), A_\beta(Y)] \in \mathbb{F}^{m \times m}.$$

A basic result in the theory of Lie groups implies that $[A_\alpha(X), A_\beta(Y)]$ always lies in the Lie algebra $\mathfrak{g} \subset \mathbb{F}^{m \times m}$, hence $F_\alpha \in \Omega^2(\mathcal{U}_\alpha, \mathfrak{g})$, but this will be obvious in the case we're interested in below, so there is no need right now for a digression on Lie groups.

The local curvature 2-form depends on a choice of trivialization, so we need to pay attention to the way that it transforms when trivializations are changed. Suppose $\Phi_\beta : E|_{\mathcal{U}_\beta} \to \mathcal{U}_\beta \times \mathbb{F}^m$ is a second trivialization, related to Φ_α by the transition map $g = g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to G$. Then the

local representations of a section $v \in \Gamma(E)$ are related on the overlap $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ by $v_\beta = g v_\alpha$, thus $F_\beta(X,Y)(g v_\alpha) = F_\beta(X,Y) v_\beta = (\Omega_K(X,Y) v)_\beta = g(\Omega_K(X,Y) v)_\alpha = g F_\alpha(X,Y) v_\alpha$, implying the relation

$$(28.7) \qquad\qquad F_\beta(X,Y) = g F_\alpha(X,Y) g^{-1} \qquad \text{on } \mathcal{U}_\alpha \cap \mathcal{U}_\beta.$$

The formulas (28.6) and (28.7) have an especially interesting consequence whenever the structure group G happens to be abelian.

EXERCISE 28.7. Show that if the Lie subgroup $G \subset \mathrm{GL}(m, \mathbb{F})$ is abelian, then all matrices in G also commute with all matrices in the Lie algebra $\mathfrak{g}$, and $[\mathbf{A}, \mathbf{B}] = 0$ for all pairs $\mathbf{A}, \mathbf{B} \in \mathfrak{g}$.

In the abelian case, it now follows from (28.6) that F_α is the exterior derivative of A_α, and is thus a $\mathfrak{g}$-valued 2-form; as mentioned above, it is true in general that F_α takes values in $\mathfrak{g}$, but this is especially obvious in the abelian case. With that in mind, the values of F_α can now be seen to commute with transition functions, so (28.7) implies that $F_\alpha = F_\beta$ on the domain where they overlap, meaning there exists a *globally*-defined $\mathfrak{g}$-valued 2-form

$$F \in \Omega^2(M, \mathfrak{g})$$

that matches $F_\alpha \in \Omega^2(\mathcal{U}_\alpha, \mathfrak{g})$ for every G-compatible trivialization $(\mathcal{U}_\alpha, \Phi_\alpha)$. This 2-form is exact on $\mathcal{U}_\alpha$, and therefore closed, though it might not be globally exact since the connection 1-forms A_α are generally not globally defined.

Let's apply these observations in the special case where E is the tangent bundle of an oriented Riemannian 2-manifold (Σ, g) and ∇ is its Levi-Cività connection. The orientation and bundle metric give $T\Sigma$ the structure group $\mathrm{SO}(2)$, the group of 2-by-2 rotation matrices, which is indeed abelian. For computational purposes, it will be more convenient to replace $\mathrm{SO}(2)$ with the unitary group $\mathrm{U}(1)$, to which it is isomorphic via the transformation

$$(28.8) \qquad\qquad \mathrm{SO}(2) \to \mathrm{U}(1) : \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mapsto e^{i\theta}.$$

The Lie algebra $\mathfrak{u}(1)$ of $\mathrm{U}(1)$ is the space of purely imaginary 1-by-1 matrices, so

$$\Omega^k(\Sigma, \mathfrak{u}(1)) = \Omega^k(\Sigma, i\mathbb{R})$$

consists of imaginary-valued forms. Identifying $\mathrm{SO}(2)$ with $\mathrm{U}(1)$ in this way is equivalent to identifying $\mathbb{R}^2$ with $\mathbb{C}$ via the bijection $(x, y) \leftrightarrow x + iy$, and real local trivializations $\Phi_\alpha : T\Sigma|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{R}^2$ are thus identified with complex trivializations $T\Sigma|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$, related to each other by transition functions with values in $\mathrm{U}(1) \subset \mathbb{C}$. In this way, $T\Sigma$ can now be viewed as a complex line bundle, and according to (28.8), scalar multiplication by i on the fibers of $T\Sigma$ is represented in any $\mathrm{SO}(2)$-compatible real trivialization by the rotation matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, so in other words, it is a 90-degree counterclockwise rotation on every fiber. This is precisely the bundle map that we have previously referred to as

$$J : T\Sigma \to T\Sigma.$$

The formula relating K_G and the Riemann tensor can now be seen in a slightly new light: for any $\mathrm{U}(1)$-compatible local trivialization $\Phi_\alpha : T\Sigma|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$, we have

$$F(X,Y) Z_\alpha = (R(X,Y)Z)_\alpha = -(K_G\, d\mathrm{vol}(X,Y) J Z)_\alpha = -K_G\, d\mathrm{vol}(X,Y) i Z_\alpha,$$

implying:

PROPOSITION 28.8. *Under the identification* $\mathrm{SO}(2) \cong \mathrm{U}(1)$ *defined in* (28.8)*, the imaginary-valued 2-form* $F \in \Omega^2(\Sigma, \mathfrak{u}(1))$ *is related to the Gaussian curvature* $K_G : \Sigma \to \mathbb{R}$ *and the Riemannian volume form* $d$vol *on* (Σ, g) *by*

$$iF = K_G \, d\mathrm{vol} \in \Omega^2(\Sigma).$$

This formula strongly suggests that it might be interesting to compute integrals $\int_P K_G \, d$vol over regions $P \subset \Sigma$, especially if P is contained in the domain $\mathcal{U}_\alpha$ of a local trivialization, on which $iF = i \, dA_\alpha$, so that Stokes' theorem implies

$$\int_P K_G \, d\mathrm{vol} = i \int_P dA_\alpha = i \int_{\partial P} A_\alpha.$$

We will apply this in the next lecture to integrate K_G over disk-like regions with piecewise-smooth polygonal boundaries, e.g. triangles bounded by geodesic segments. The imaginary-valued integral $\int_{\partial P} A_\alpha$ turns out in this case to give a new perspective on one of Euclid's best-known propositions: the sum of angles in a triangle is π. As we will see, the only reason this is true on the Euclidean plane is that for that particular Riemannian 2-manifold, $K_G \equiv 0$. You can see from Figure 8 that it is not true on the positively-curved unit sphere S^2, and Exercise 24.18 shows that it is also not true on the negatively-curved hyperbolic plane.

29. The Gauss-Bonnet formula

In the previous lecture we observed that on any oriented Riemannian 2-manifold (Σ, g), the 2-form $K_G \, d$vol $\in \Omega^2(\Sigma)$ is locally (up to multiplication by i) the exterior derivative of a connection 1-form, so that $\int_P K_G \, d$vol over sufficiently simple regions $P \subset \Sigma$ should be computable via Stokes' theorem. We shall now follow this idea to its logical conclusion.

29.1. Polygons and their angles. We assume throughout this section that (Σ, g) is an oriented Riemannian 2-manifold, possibly with boundary, ∇ is its Levi-Cività connection, and its Riemannian volume form is denoted by

$$d\mathrm{vol}_\Sigma \in \Omega^2(\Sigma).$$

Our goal is to compute $\int_P K_G \, d$vol$_\Sigma$ for compact regions $P \subset \Sigma$ that have the topology of disks bounded by piecewise smooth polygons. In general, a **piecewise smooth** curve in a smooth manifold M is a continuous map $\gamma : [a, b] \to M$ for which there are finitely many points $a = t_0 < t_1 < \ldots < t_{N-1} < t_N = b$ such that the restrictions

$$\gamma|_{[t_{j-1}, t_j]} : [t_{j-1}, t_j] \to M$$

are smooth immersions for each $j = 1, \ldots, N$. The velocity $\dot{\gamma}(t)$ of such a curve is thus a smooth function of t except possibly at the finitely many points t_j for $j = 1, \ldots, N - 1$, where the two one-sided limits

$$\lim_{t \to t_j^\pm} \dot{\gamma}(t) \in T_{\gamma(t_j)} M$$

are both defined and nonzero but need not be equal, i.e. there may be jump discontinuities. The curve is called a piecewise smooth **simple closed** curve if $\gamma(b) = \gamma(a)$ and there is no other self-intersection $\gamma(t) = \gamma(t')$ for $t \neq t'$. We do not require $\dot{\gamma}(a) = \dot{\gamma}(b)$, so if we view γ as a piecewise-smooth map $S^1 \to M$ by identifying S^1 with the quotient $[a, b]/\sim$ in which $a \sim b$, the velocity of $\gamma : S^1 \to M$ may also have a jump discontinuity at the point $[a] = [b]$.

DEFINITION 29.1. A **smooth polygon in** $\mathbb{R}^2$ is the closure $P \subset \mathbb{R}^2$ of a region bounded by the image of a single piecewise-smooth simple closed curve $\gamma : [a, b] \to \mathbb{R}^2$. If we write $a = t_0 < \ldots < t_N = b$ so that $t_1, \ldots, t_{N-1}$ are the finitely-many points where γ is allowed to be nonsmooth,

then the smooth curves $\gamma([t_{j-1}, t_j])$ will be called **edges**, and their boundary points are called **vertices**. The union of all the edges will be denoted by ∂P.

REMARK 29.2. The point $\gamma(a) = \gamma(b)$ is always considered a vertex of the polygon in Definition 29.1, so there is always at least one edge and one vertex. There is also ambituity in the notion of edges and vertices since the definition requires the set $\{t_1, \ldots, t_{N-1}\}$ to contain all points where γ is not smooth, but not the converse, so there is always some freedom to add more vertices arbitrarily, even if γ is completely smooth. This is just a matter of bookkeeping, as it will never at any stage be important to *require* that $\dot\gamma$ is discontinuous at some point.

Observe that if the region P in this definition has a smooth boundary, then $\partial P \cong S^1$ inherits from the orientation of $\mathbb{R}^2$ a natural orientation as the boundary of P. This notion of orientation generalizes naturally to the piecewise-smooth case so that each edge of ∂P inherits a natural orientation, and is thus a compact oriented 1-manifold with boundary.

There are theorems in topology that give fairly strong restrictions on what a compact region bounded by a continuous simple closed curve can look like. In order to avoid too much of a digression into topology, let us single out the particular property of the curve $\gamma : [a, b] \to \mathbb{R}^2$ that we will need to know. Assuming $a < t_1 < \ldots < t_{N-1} < b$ denote the points where γ is allowed to be nonsmooth, we can define a piecewise-continuous function

$$\phi : [a, b] \backslash \{t_1, \ldots, t_{N-1}\} \to \mathbb{R}$$

that is smooth on each of the subintervals (t_{j-1}, t_j) and gives the angle between $\dot\gamma(t) \in \mathbb{R}^2$ and the first standard basis vector. There is some freedom in this definition, as any multiple of 2π can be added to ϕ on each of the subintervals (t_{j-1}, t_j), but we can reduce this freedom by restricting the jumps at $t = t_1, \ldots, t_{N-1}$ to a suitable interval, namely

(29.1) $$\Delta\phi_j := \lim_{t \to t_j^+} \phi(t) - \lim_{t \to t_j^-} \phi(t) \in [-\pi, \pi], \qquad j = 1, \ldots, N-1.$$

Here the convention is that $\Delta\phi_j > 0$ if the curve makes a sudden counterclockwise turn at t_j and $\Delta\phi_j < 0$ if it turns clockwise; these notions are well defined even in the case of a full 180-degree turn since γ is not allowed to intersect itself, and in this way we see the difference between $\Delta\phi_j = \pi$ and $\Delta\phi_j = -\pi$. With this restriction in place, the function ϕ is uniquely defined modulo a constant multiple of 2π. There is also a possible angle change at the end point $\gamma(a) = \gamma(b)$ that we will need to keep track of, so let us define this by

$$\Delta\phi_N := \phi(a) - \phi(b) + 2\pi k \in [-\pi, \pi],$$

where there is a unique choice of $k \in \mathbb{Z}$ that makes this number lie in the correct interval and satisfy the convention regarding counterclockwise/clockwise turns. The main observation we need to make now is that the total change in ϕ as t traverses the interval from a to b, including the jump discontinuities, must be exactly 2π:

(29.2) $$\int_0^1 \dot\phi(t)\, dt + \sum_{j=1}^{N} \Delta\phi_j = 2\pi.$$

This statement is obvious whenever P is e.g. a disk with smooth boundary or a convex polygon, and it will in fact be obviously true for every example we are likely to consider, thus you might as well regard it as an extra condition in Definition 29.1. It is true but not so straightforward to prove that it actually follows from the conditions already stated in that definition—if you want to know why, see §29.3 at the end of this lecture.

DEFINITION 29.3. A **smooth polygon in** Σ is a compact subset $P \subset \Sigma$ admitting an open neighborhood $\mathcal{U} \subset \Sigma$ with a chart $x : \mathcal{U} \to \mathbb{R}^2$ that identifies P with a smooth polygon P_0 in $\mathbb{R}^2$.

The points and smooth curves identified by this chart with the vertices and edges of P_0 are called the **vertices** and **edges** of P.

The orientation of Σ restricts to any smooth polygon $P \subset \Sigma$ and induces a natural orientation on its edges, whose union we again denote by ∂P. The metric also restricts to each edge $\ell \subset \partial P$ and defines a natural "volume form"

$$d\mathrm{vol}_{\partial P} \in \Omega^1(\ell).$$

Although ∂P is not generally a smooth manifold, it's easy to see that Stokes' theorem still holds:

$$\int_P d\lambda = \int_{\partial P} \lambda$$

for any $\lambda \in \Omega^1(\Sigma)$, where the integral over ∂P is defined by summing the integrals over the edges. One can prove this by an approximation argument, perturbing ∂P to a smooth loop that bounds a region P_ϵ on which $\int_{P_\epsilon} d\lambda$ is almost the same. (A similar argument was sketched in Example 12.14 for applying Stokes' theorem on the product of two manifolds with boundary, which is technically a manifold with boundary and corners.)

We can apply Stokes' theorem in particular to compute $\int_P K_G \, d\mathrm{vol}_\Sigma$ for any smooth polygon $P \subset \Sigma$. For this purpose, recall that since the bundle $T\Sigma$ is equipped with both an orientation and a positive bundle metric, it has structure group $SO(2)$, which we can identify with $U(1)$ as in §28.2, thus making $T\Sigma$ into a complex line bundle on which scalar multiplication by i is the counterclockwise 90-degree rotation map $J : T\Sigma \to T\Sigma$. From this perspective, a $U(1)$-compatible frame for $T\Sigma$ over a region $\mathcal{U} \subset \Sigma$ is simply a vector field $X \in \mathfrak{X}(\mathcal{U})$ that has unit length everywhere; indeed, one obtains a *real* orthonormal frame from this by putting X together with JX. It is now easy to see that $T\Sigma$ always admits such a frame on some neighborhood of a smooth polygon $P \subset \Sigma$: simply choose a chart $(\mathcal{U}, (x^1, x^2))$ with $P \subset \mathcal{U}$ as in Definition 29.3 and define the vector field

$$X := \frac{\partial_1}{|\partial_1|} \in \mathfrak{X}(\mathcal{U}).$$

Let us denote the corresponding local trivialization by $\Phi : T\Sigma|_{\mathcal{U}} \to \mathcal{U} \times \mathbb{C}$ and the associated connection 1-form for the Levi-Cività connection by

$$A := i\lambda \in \Omega^1(\mathcal{U}, \mathfrak{u}(1)),$$

thus defining a real-valued 1-form $\lambda \in \Omega^1(\mathcal{U})$. The discussion in §28.2 then implies $K_G \, d\mathrm{vol}_\Sigma = iF = i \, dA = -d\lambda$, hence by Stokes' theorem,

$$\int_\Sigma K_G \, d\mathrm{vol}_\Sigma = - \int_{\partial P} \lambda.$$

Our remaining task is to compute $\int_{\partial P} \lambda$.

Let us assume the boundary ∂P has $N \in \mathbb{N}$ edges, and thus N vertices at which it is not required to be smooth, and denote the angles formed between neighboring edges at these vertices by

$$\alpha_1, \ldots, \alpha_N \in [0, 2\pi].$$

Note that the definition of these angles requires the orientation: the convention is that $\alpha_j \in [0, \pi)$ if there is a counterclockwise turn and $\alpha_j \in (\pi, 2\pi]$ for a clockwise turn. The case $\alpha_j = \pi$ is allowed, and in this way we can also accommodate situations where ∂P is completely smooth.

Next, choose a parametrization of ∂P as a piecewise-smooth simple closed curve $\gamma : [0, T] \to \Sigma$, oriented so that the parametrization of each edge is orientation preserving. The length $T > 0$ of the the interval can be choosen so that $|\dot{\gamma}(t)| = 1$ for all t, except at the finitely-many parameter values

$$0 < t_1 < \ldots < t_{N-1} < T$$

where $\dot{\gamma}(t)$ may fail to exist, and we will assume $\alpha_j \in [0, 2\pi]$ is the angle formed by a vertex at time t_j for $j = 1, \ldots, N-1$, or times 0 and T for $j = N$. One can now find a piecewise-continuous function $\theta : [0, T] \to \mathbb{R}$ such that

$$\dot{\gamma}(t) = e^{i\theta(t)} X(\gamma(t)) \qquad \text{for all } t \in [0,1] \backslash \{t_1, \ldots, t_{N-1}\},$$

where θ is smooth on the open intervals (t_{j-1}, t_j) and is allowed to have jump discontinuities

$$\Delta\theta_j := \lim_{t \to t_j^+} \theta(t) - \lim_{t \to t_j^-} \theta(t) = \pi - \alpha_j \in [-\pi, \pi], \qquad j = 1, \ldots, N-1,$$

in which the orientation of Σ can be used to distinguish between $\Delta\theta_j = \pi$ and $\Delta\theta_j = -\pi$ via the same counterclockwise/clockwise convention that we used to define $\Delta\phi_j$. These conditions determine the function $\theta(t)$ uniquely modulo a constant multiple of 2π. We can also keep track of the angle α_N at $\gamma(a) = \gamma(b)$ by writing

$$\pi - \alpha_N = \Delta\theta_N := \theta(a) - \theta(b) + 2\pi k \in [-\pi, \pi],$$

for the unique choice of $k \in \mathbb{Z}$ that puts this number in the right interval and distinguishes correctly between counterclockwise and clockwise turns. With these definitions in place, the jumps $\Delta\theta_j \in [-\pi, \pi]$ are related to the angles $\alpha_j \in [0, 2\pi]$ by

$$(29.3) \qquad\qquad \alpha_j = \pi - \Delta\theta_j, \qquad j = 1, \ldots, N.$$

LEMMA 29.4. $\displaystyle \int_0^T \dot{\theta}(t)\, dt + \sum_{j=1}^N \Delta\theta_j = 2\pi.$

PROOF. It is clear from the definitions that this number is at least an integer multiple of 2π. Let $\gamma_0 := \varphi^{-1} \circ \gamma : [0, T] \to \mathbb{R}^2$, so γ_0 is a piecewise-smooth simple closed curve parametrizing ∂P_0, whose image under the embedding $\varphi : \mathcal{U}_0 \to \mathcal{U} \subset \Sigma$ is ∂P. If we equip $\mathcal{U}_0 \subset \mathbb{R}^2$ with the pullback metric $\varphi^* g$, then the way in which our frame $X \in \mathfrak{X}(\mathcal{U})$ was defined gives a new interpretation of $\theta(t)$: it is the angle of the tangent vector $\dot{\gamma}_0(t) \in \mathbb{R}^2$ relative to the standard basis vector $\mathbf{e}_1$, as measured using the metric $\varphi^* g$. If $\varphi^* g$ were the *standard* Euclidean metric on $\mathbb{R}^2$, the lemma would now just be a restatement of Equation 29.2. Unfortunately, we cannot assume $\varphi^* g$ is the standard Euclidean metric; this would be a very strong restriction, forcing (Σ, g) to be locally flat on the region $\mathcal{U}$. However, the space of all Riemannian metrics is convex, so we can define a smooth family of metrics on $\mathcal{U}_0 \subset \mathbb{R}^2$ by

$$g_s := s\varphi^* g + (1-s)g_E, \qquad s \in [0, 1],$$

where $g_E := dx^2 + dy^2$ denotes the Euclidean metric, so g_s interpolates between $g_1 = \varphi^* g$ and $g_0 = g_E$. For each $s \in [0, 1]$, we can now define a corresponding function $\theta^s(t)$ in the same manner as above, but using the metric $\varphi_* g_s$ on $\mathcal{U} \subset \Sigma$ to measure angles. The sum $\int_0^T \dot{\theta}^s(t)\, dt + \sum_{j=1}^N \Delta\theta_j^s$ depends continuously on the parameter s, and since it is always a multiple of 2π, we get the same answer for $s = 1$ and $s = 0$, so that the result in the case of the Euclidean metric is also valid in the general case. $\qquad \square$

Now let's compute $\int_{\ell_j} \lambda$ for a specific edge $\ell_j := \gamma([t_{j-1}, t_j]) \subset \partial P$. This requires computing $\lambda(\dot{\gamma}(t)) = -iA(\dot{\gamma}(t))$, which can be deduced by computing a covariant derivative in the direction of $\dot{\gamma}(t)$. In particular, $\dot{\gamma}(t)$ itself is expressed relative to our chosen frame X as the complex-valued function $e^{i\theta(t)}$, thus

$$(29.4) \qquad \begin{aligned} \nabla_t \dot{\gamma}(t) &= \left(\partial_t e^{i\theta(t)} + A(\dot{\gamma}(t)) e^{i\theta(t)} \right) X(\gamma(t)) = \left(\dot{\theta}(t) + \lambda(\dot{\gamma}(t)) \right) i e^{i\theta(t)} X(\gamma(t)) \\ &= \left(\dot{\theta}(t) + \lambda(\dot{\gamma}(t)) \right) i\dot{\gamma}(t). \end{aligned}$$

This last expression has a useful geometric interpretation.

DEFINITION 29.5. Suppose ℓ is a 1-dimensional submanifold of a Riemannian 2-manifold (Σ, g) and $\nu \in \Gamma(T\Sigma|_\ell)$ is unit normal vector field along ℓ. The (signed) **geodesic curvature** of ℓ is then defined as the unique function

$$\kappa_\ell : \ell \to \mathbb{R}$$

such that for any local parametrization $\gamma : (a, b) \to \ell$ of ℓ satisfying $|\dot\gamma| \equiv 1$,

$$\nabla_t \dot\gamma(t) = \kappa_\ell(\gamma(t))\nu(\gamma(t))$$

for all t.

This definition makes sense because if γ is parametrized with unit speed, differentiating the relation $\langle \dot\gamma(t), \dot\gamma(t) \rangle = 1$ reveals that $\nabla_t \dot\gamma(t)$ is always orthogonal to $\dot\gamma(t)$, and is therefore a real multiple of $\nu(\gamma(t))$. Moreover, one could change the local parametrization γ of ℓ, but all other parametrizations with unit speed take the form $t \mapsto \gamma(\pm t + c)$ for a constant c, so one obtains the same definition of κ_ℓ. It does depend on the choice of normal vector field: reversing ν changes κ_ℓ by a sign. It follows that κ_ℓ cannot be defined in this way if ℓ has non-orientable normal bundle, but this situation does not arise in the application that we have in mind. In the non-orientable case, one can still define an *unsigned* geodesic curvature $|\kappa_\ell| \geqslant 0$, which is actually just the norm of $\nabla_t \dot\gamma$, and the latter is given as a definition of the term "geodesic curvature" in many books. In either case, it should be emphasized that geodesic curvature is a purely *extrinsic* notion, as it depends on the embedding of the submanifold ℓ into the surface Σ. (Indeed, Exercise 27.4 shows that there is no interesting notion of intrinsic curvature for Riemannian 1-manifolds, as they are all locally flat.) The geodesic curvature is a measurement of the extent to which $\ell \subset \Sigma$ fails locally to the image of a geodesic in (Σ, g); in particular, $\kappa_\ell \equiv 0$ if and only if ℓ can be parametrized locally by geodesics.

With this definition understood, (29.4) can be reinterpreted using the observation that $i\dot\gamma(t)$ is a 90-degree counterclockwise rotation of $\dot\gamma(t)$, pointing inwards through the boundary of P. If we take this as a choice of normal vector field along ℓ_j, the relation now says:

LEMMA 29.6. *For $t \in [t_{j-1}, t_j]$, $\dot\theta(t) + \lambda(\dot\gamma(t)) = \kappa_{\ell_j}(\gamma(t))$.* $\qquad\square$

We now have enough ingredients in place to write down a revealing formula for $\int_{\partial P} \lambda$: combining Lemmas 29.6 and 29.4 with (29.3), we have:

$$\int_{\partial P} \lambda = \sum_{j=1}^N \int_{\ell_j} \lambda = \sum_{j=1}^N \int_{t_{j-1}}^{t_j} \lambda(\dot\gamma(t))\, dt = \sum_{j=1}^N \int_{t_{j-1}}^{t_j} \kappa_{\ell_j}(\gamma(t))\, dt - \int_0^T \dot\theta(t)\, dt$$

$$= \sum_{j=1}^N \int_{t_{j-1}}^{t_j} \kappa_{\ell_j}(\gamma(t))\, dt - \left(2\pi - \sum_{j=1}^N \Delta\theta_j \right) = \sum_{j=1}^N \int_{t_{j-1}}^{t_j} \kappa_{\ell_j}(\gamma(t))\, dt - 2\pi + \sum_{j=1}^N (\pi - \alpha_j)$$

$$= \sum_{j=1}^N \int_{t_{j-1}}^{t_j} \kappa_{\ell_j}(\gamma(t))\, dt + (N-2)\pi - \sum_{j=1}^N \alpha_j.$$

Since γ was parametrized to have unit speed on each edge, the integrals in this last expression are actually just the integrals of the 1-forms $\kappa_{\ell_j}\, d\mathrm{vol}_{\partial P}$ over the respective edges, and using Stokes' theorem to rewrite $\int_{\partial P} \lambda$ in terms of the Gaussian curvature, we obtain from this the first version of the Gauss-Bonnet formula:

THEOREM 29.7 (Gauss-Bonnet formula, polygon version). *Suppose (Σ, g) is an oriented Riemannian 2-manifold with Gaussian curvature $K_G : \Sigma \to \mathbb{R}$, $P \subset \Sigma$ is a smooth polygon with N*

smooth edges $\ell_1, \ldots, \ell_N \subset \partial P$ and angles $\alpha_1, \ldots, \alpha_N \in [0, 2\pi]$ at its vertices, and the signed geodesic curvature κ_{ℓ_j} of each edge ℓ_j is defined with respect to a normal vector field pointing inwards. Then

$$\sum_{j=1}^{N} \alpha_j = (N-2)\pi + \int_P K_G \, d\mathrm{vol}_\Sigma + \sum_{j=1}^{N} \kappa_{\ell_j} \, d\mathrm{vol}_{\partial P}.$$

$\square$

We have arranged this formula to look like a generalization of the fact that triangles in the Euclidean plane have angles adding up to π; that is just the case where $N = 3$, $K_G \equiv 0$ and all edges are geodesic segments. More generally, the integral of the Gaussian curvature can now be viewed as a correction term that measures the failure of this relation to hold:

COROLLARY 29.8. *If $P \subset \Sigma$ is a smooth polygon with N edges that are all geodesic segments, then the angles $\alpha_1, \ldots, \alpha_N$ at the vertices satisfy*

$$\sum_{j=1}^{N} \alpha_j = (N-2)\pi + \int_P K_G \, d\mathrm{vol}_\Sigma.$$

$\square$

REMARK 29.9. The assumption that Σ carries an orientation is not actually necessary for Theorem 29.7, because even if Σ is not globally orientable, a neighborhood of the polygon $P \subset \Sigma$ is diffeomorphic to an open subset of $\mathbb{R}^2$, so an orientation can always be chosen on this neighborhood. If one reverses the orientation, none of the terms in the Gauss-Bonnet formula actually change: for the angles α_j and the term $(N-2)\pi$ this is obvious, though it takes a bit more thought for the two integrals. We already saw in the previous lecture that K_G does not change if the orientation is switched; the geodesic curvatures also do not change since they depend on the choice of normal vector field at the boundary and this was defined independently of all orientations. Changing the orientations of Σ and ℓ_j changes the volume forms $d\mathrm{vol}_\Sigma$ and $d\mathrm{vol}_{\ell_j}$ by a sign, but a cancelling sign is caused by the fact that $\int_{-M} \omega = -\int_M \omega$ for any oriented manifold M and any top-dimensional form ω.

EXAMPLE 29.10. Now is a good moment to look again at Figure 8 in Lecture 19, which shows a geodesic triangle in the unit sphere S^2 whose angles are all $\pi/2$. This triangle occupies exactly $1/8$ of the total area of S^2, so its area is $\pi/2$, and this is also $\int_\Sigma K_G \, d\mathrm{vol}_\Sigma$ since $K_G \equiv 1$ by Example 27.13. The formula in Corollary 29.8 thus becomes $3\pi/2 = \pi + \frac{\pi}{2}$ in this case.

EXERCISE 29.11. According to Exercise 27.22, the Poincaré half-plane $(\mathbb{H}, h)$ has constant curvature $K_G \equiv -1$.

(a) Write down the Riemannian volume form on $(\mathbb{H}, h)$, and show that any region of the form $[a, b] \times [c, \infty) \subset \mathbb{H}$ for $-\infty < a < b < \infty$ and $c > 0$ has finite area, while regions of the form $[a, b] \times (0, c] \subset \mathbb{H}$ have infinite area.

(b) Show that every compact region in $(\mathbb{H}, h)$ bounded by three geodesics has area strictly less than π, though its area can be arbitrarily close to π.
    Hint: Use the result of Exercise 24.18.

29.2. Triangulation and the Euler characteristic. Next question: what happens if we integrate K_G over a region on which $T\Sigma$ is not trivializable? A nice way to approach this is by decomposing Σ into a union of polygons glued together along their edges.

DEFINITION 29.12. Let Σ be a 2-dimensional manifold, possibly with boundary. A **polygonal triangulation** of Σ is a collection of smooth polygons $\{P_\alpha \subset \Sigma\}_{\alpha \in I}$ with $\Sigma = \bigcup_{\alpha \in I} P_\alpha$, called the

faces of the triangulation, while each edge or vertex of each of these polygons is called an **edge** or **vertex** of the triangulation respectively. They are required to satisfy the following conditions:

(1) Each edge ℓ is either contained in $\partial\Sigma$ or satisfies $\ell \cap \partial\Sigma \subset \partial\ell$, and in the latter case, it is an edge of exactly two faces.
(2) Two distinct faces are either disjoint or their intersection is a union of common edges.
(3) Every vertex is a vertex of at most finitely many faces.

The sets of vertices and edges of the triangulation are sometimes denoted by $\Sigma^0, \Sigma^1 \subset \Sigma$ and also called the 0-**skeleton** and 1-**skeleton** respectively. Note that if $\partial\Sigma \neq \varnothing$, then $\partial\Sigma \subset \Sigma^1$. We say the triangulation is **finite** if it has only finitely many faces (and therefore also finite-many vertices and edges).

Polygonal triangulations are somewhere in between two similar notions that are popular in topology: they are more general than what are normally just called *triangulations* (in which all the faces are required to be actual triangles), while also being special cases of the more general notion of *CW-complexes*. It is a general fact that all smooth surfaces admit polygonal triangulations, and one can even arrange without loss of generality for them to be triangulations in the stricter sense, in which every face has three edges. A similar result (based on *simplices*, a higher-dimensional generalization of triangles) also holds for smooth manifolds of all dimensions, though not generally for *topological* manifolds above dimension 3. In practice, we will not need to have such general existence results, because for our purposes it is more interesting to look at specific examples in which explicit triangulations are not hard to construct. But just out of interest, here is the most relevant special case of the general result:

PROPOSITION 29.13. *Every compact smooth surface Σ admits a finite polygonal triangulation consisting only of triangles.*

SKETCH OF THE PROOF. Every point in Σ admits a compact neighborhood that is a smooth polygon contained in the domain of a chart, and since Σ is compact, it can be covered with finitely many such polygons. After small perturbations, we can also arrange without loss of generality that no edge of any of these polygons intersects a vertex of another one, and that whenever two edges intersect, they do so transversely (and therefore only finitely-many times). Define Σ^0 to be the union of the set of vertices of all these polygons with the finite set of intersections between their edges; we should correspondingly redefine the word "edge" to mean any potentially shorter segment of one of the original edges that is bounded by two points of Σ^0. Each connected component in the complement of the set of edges is now an open region with compact closure contained in the domain of a chart, and bounded by some disjoint union of piecewise-smooth simple closed curves. It therefore remains only to show the following: any region $P \subset \mathbb{R}^2$ bounded by piecewise-smooth simple closed curves can be decomposed into a union of smooth polygons that each have three edges and intersect each other only along matching pairs of edges. This can be achieved by adding new edges, i.e. choosing new smooth paths through the interior of P that connect previously unconnected pairs of vertices. Once you've added enough of these, every component of the complement is bounded by a triangle. $\square$

REMARK 29.14. There's a subtlety in the construction of triangulations that should be mentioned. Most authors' definition of the term "smooth triangle" is stricter than ours: we are assuming a smooth polygon in $\mathbb{R}^2$ can be any compact region bounded by a piecewise-smooth simple closed curve, and we call it a triangle if that curve has three smooth edges, but in practice, such an object does not need to look very similar to what we typically imagine as a triangle. (Try drawing an example where the edges form gratuitously complicated spirals and vertices have angles $2\pi - \epsilon$.) Most authors add the condition that a "triangle" must actually be homeomorphic to a perfectly

ordinary convex triangle with straight edges in $\mathbb{R}^2$. We are not assuming this, but it follows from our definition for somewhat nontrivial reasons, and you'll find this fact lurking in the background of Equation (29.2) if you look into the details as discussed in §29.3. Our argument in that appendix appeals to the classification of closed surfaces in order to show that every smooth polygon by our definition really is homeomorphic to a disk. One needs to be a bit careful about circularity here, because most popular proofs of the classification of surfaces are based on the fact that all surfaces can be triangulated. (There are ways to get around this, however, e.g. the proof via Morse theory in [**Hir94**] is quite illuminating and does not require triangulations.)

In practice, all useful constructions of triangulations on surfaces require some nontrivial topological input at some step to ensure that compact regions bounded by simple closed curves in $\mathbb{R}^2$ are always homeomorphic to disks. If the boundary curve is continuous but not smooth, then this fact requires a difficult classical result known as the *Schoenflies theorem* (see [**Moi77**]). In the smooth category there is a standard way to avoid this by using geodesics: the idea is to choose a Riemannian metric on Σ and carry out the proof of Proposition 29.13 so that every edge in the triangulation becomes the unique shortest geodesic between two nearby points. The final subdivision step is less obvious in this setting, but with a bit more care it can still be done, and in this way one obtains a triangulation whose edges are all geodesics. It is much easier than the Schoenflies theorem to see (e.g. by working in Riemann normal coordinates based at a vertex) that any region bounded by three short geodesics is homeomorphic to a disk.

DEFINITION 29.15. Given a finite polygonal triangulation of Σ with v vertices, e edges and f faces, the *Euler characteristic* of Σ is the integer

$$\chi(\Sigma) = v - e + f.$$

The Euler characteristic turns out to be a topological invariant of Σ, though our definition makes this far from obvious—*a priori* it appears to depend rather crucially on a choice of triangulation. It will follow from Theorem 29.17 below that this is not the case, that in fact $\chi(\Sigma)$ depends at most on the differentiable structure of Σ. Proving that it only depends on the topology of Σ requires methods from algebraic topology: the standard approach is to define $\chi(\Sigma)$ in terms of singular homology and use either cellular or simplicial homology to prove that the quantity above matches this definition for any triangulation. Details may be found in e.g. [**Hat02**,**Bre93**,**Wen18**].

EXERCISE 29.16. Taking it on faith for the moment that the Euler characteristic doesn't depend on a choice of triangulation, show that $\chi(S^2) = 2$, $\chi(\mathbb{D}^2) = 1$ and $\chi(\mathbb{T}^2) = 0$.

We shall now compute the integral of K_G over a compact surface using a finite polygonal triangulation with v vertices, e edges and f faces. Assume $e = e_0 + e_\partial$ where e_∂ is the number of edges contained in $\partial\Sigma$, and similarly $v = v_0 + v_\partial$. Observe that every vertex on $\partial\Sigma$ is a boundary point of exactly two edges on $\partial\Sigma$, and since every edge likewise has two boundary points, $e_\partial = v_\partial$.[74] By Theorem 29.7, $\int_\Sigma K_G \, d\mathrm{vol}_\Sigma$ contains a term of the form

$$-\sum_j \int_{\ell_j} \kappa_{\ell_j} \, d\mathrm{vol}_{\ell_j} + \sum_j \alpha_j - (N-2)\pi$$

for each face, assuming the face in question has N edges. Adding these up for all faces, we make the following observations:

---

[74]There is a small loop-hole in this argument: our definition of smooth polygons allows the possibility that there is only one edge, whose two boundary points then coincide to form a single vertex, but if this happens, the claim that $e_\partial = v_\partial$ remains valid.

(1) Every edge $\ell \subset \Sigma \backslash \partial\Sigma$ is an edge for two distinct faces and thus appears twice with two oppositely-oriented choices of normal vector field pointing toward different faces. The geodesic curvature terms for these edges cancel in the sum.
(2) The geodesic curvature terms for all edges $\ell \subset \partial\Sigma$ add up to

$$- \int_{\partial\Sigma} \kappa_{\partial\Sigma} \; d\mathrm{vol}_{\partial\Sigma}$$

if we define $\kappa_{\partial\Sigma}$ with respect to a normal vector field pointing inward at the boundary.
(3) The sum of all angles α_j at an interior vertex (for every face adjacent to that vertex) is 2π, and for boundary vertices the sum is π. Thus altogether these terms contribute $2\pi v_0 + \pi v_\partial = 2\pi v - \pi v_\partial$.
(4) Every interior edge is counted twice and boundary edges are counted once, so the $-(N-2)\pi$ terms add up to $-\pi(2e_0 + e_\partial - 2f) = 2\pi(f - e) + \pi e_\partial$.

Summing all these contributions, we have

$$- \int_{\partial\Sigma} \kappa_{\partial\Sigma} \; ds + 2\pi v + 2\pi(f - e) - \pi v_\partial + \pi e_\partial = - \int_{\partial\Sigma} \kappa_{\partial\Sigma} \; ds + 2\pi \chi(\Sigma).$$

This proves:

THEOREM 29.17 (Gauss-Bonnet formula, global version). *Assume (Σ, g) is a compact oriented 2-dimensional Riemannian manifold, possibly with boundary, and the signed geodesic curvature $\kappa_{\partial\Sigma} : \partial\Sigma \to \mathbb{R}$ is defined with respect to a normal vector field pointing inward at the boundary. Then*

$$\int_\Sigma K_G \; d\mathrm{vol}_\Sigma + \int_{\partial\Sigma} \kappa_{\partial\Sigma} \; d\mathrm{vol}_{\partial\Sigma} = 2\pi \chi(\Sigma).$$

REMARK 29.18. In keeping with Remark 29.9, Theorem 29.17 remains true if (Σ, g) is not oriented or orientable, though in this case the two integrals on the left hand side require some additional effort to interpret. The global volume form $d\mathrm{vol}_\Sigma \in \Omega^2(\Sigma)$ does not exist if Σ is not orientable, but recall from §11.4 that every Riemannian manifold, regardless of orientability, admits a canonical *volume element*, which is a density rather than a differential form. We can interpret both of the integrals in Theorem 29.17 as integrals of smooth real-valued functions with respect to measures defined via the canonical volume elements determined by the metric on Σ and $\partial\Sigma$. In practice, the volume element on (Σ, g) matches $|d\mathrm{vol}_\Sigma|$ on any region where an orientation can be chosen, so for instance $\int_\Sigma K_G \; d\mathrm{vol}_\Sigma$ can be computed as the sum of the terms $\int_{P_\alpha} K_G \; d\mathrm{vol}_\Sigma$ over all the faces P_α of a polygonal triangulation, where $d\mathrm{vol}_\Sigma$ and the integral are defined in each case by choosing an arbitrary orientation of $T\Sigma|_{P_\alpha}$, and Remark 29.9 shows that the result does not depend on this choice. Once this is understood, the proof of Theorem 29.17 also works in the non-oriented case.

Several wonderful things follow immediately from the global Gauss-Bonnet formula. Observe that the left hand side has nothing to do with the triangulation, while the right hand side makes no reference to the metric or curvature.

COROLLARY 29.19. *The Euler characteristic $\chi(\Sigma)$ does not depend on the choice of triangulation, and for any two diffeomorphic surfaces Σ_1 and Σ_2, $\chi(\Sigma_1) = \chi(\Sigma_2)$.* □

COROLLARY 29.20. *For a fixed compact surface Σ, the sum $\int_\Sigma K_G \; d\mathrm{vol}_\Sigma + \int_{\partial\Sigma} \kappa_{\partial\Sigma} \; d\mathrm{vol}_{\partial\Sigma}$ is an integer multiple of 2π, and is the same for any choice of Riemannian metric.* □

In particular, the latter statement imposes serious topological restrictions on the kinds of metrics that are allowed on any given surface: e.g. it is impossible to find a metric with everywhere

positive Gaussian curvature on a surface with negative Euler characteristic. To get a handle on this, it helps to have some concrete examples in mind; these are provided by the following exercises.

EXERCISE 29.21. Suppose Σ is a compact oriented surface with boundary and $\ell_1, \ell_2 \subset \partial\Sigma$ are two distinct connected components of $\partial\Sigma$. We can *glue* these two components to produce a new surface Σ' as follows: since ℓ_1 and ℓ_2 are both circles, there is an orientation reversing diffeomorphism $\varphi : \ell_1 \to \ell_2$, which we use to define

$$\Sigma' = \Sigma/\sim$$

where the equivalence identifies $p \in \ell_1$ with $\varphi(p) \in \ell_2$, thus identifying ℓ_1 and ℓ_2 to a single circle, now in the interior of Σ'. Show that $\chi(\Sigma') = \chi(\Sigma)$. Note: Σ need not be a connected surface to start with, so this trick can be used to glue together two separate surfaces along components of their boundaries.
Hint: A given triangulation of Σ may have different numbers of vertices on ℓ_1 and ℓ_2, but one can always modify the triangulation by adding more vertices and edges so that these numbers become the same. The number of edges on each boundary component will also always match the number of vertices. (Why?)

EXERCISE 29.22. Let Σ be the closed unit disk in $\mathbb{R}^2$ with two smaller disjoint open disks removed: the resulting surface is called a *pair of pants*. Show that $\chi(\Sigma) = -1$.
Similarly, a *handle* is a surface Σ diffeomorphic to the torus $\mathbb{T}^2$ with one open disk removed. Show that $\chi(\Sigma) = -1$.

EXERCISE 29.23. Suppose Σ is a compact surface with boundary. The operation of *gluing a handle* to Σ is defined as follows: choose a smoothly embedded closed disk in the interior of Σ, remove its interior, and glue the resulting surface along its new boundary component to a handle (see Exercise 29.22). Show that this operation decreases the Euler characteristic of Σ by 2.

EXERCISE 29.24. A **closed oriented surface of genus** g is any compact surface Σ without boundary that is diffeomorphic to a surface obtained from S^2 by gluing g handles. Special cases include the sphere itself $(g = 0)$ and the torus $(g = 1)$. Show that

$$\chi(\Sigma) = 2 - 2g.$$

For Σ a compact surface with boundary, we say it has *genus* g if it is diffeomorphic to a closed surface of genus g with finitely many small open disks cut out. Show that if such a surface has m boundary components, then $\chi(\Sigma) = 2 - 2g - m$.

REMARK 29.25. In case you didn't already believe this, we now have a simple proof of the fact that two closed oriented surfaces with differing genera (that is the plural of "genus") are not diffeomorphic: if they were, then their Euler characteristics would have to match. The converse is also true, but harder to prove; it follows from the topological classification of surfaces (see e.g. [Wen18, Lecture 19] or [Hir94]).

The Gauss-Bonnet theorem enables us to make some sweeping statements regarding what kinds of metrics may exist on various compact surfaces. In general, we say that a surface Σ with a Riemannian metric has *positive* (or *zero* or *negative*) curvature if its Gaussian curvature is positive (or zero or negative) at every point.

THEOREM 29.26. *Let Σ be a closed oriented surface of genus g. Then Σ admits a Riemannian metric with positive curvature if and only if $\Sigma \cong S^2$, zero curvature if and only if $\Sigma \cong \mathbb{T}^2$, and negative curvature if and only if $g \geqslant 2$.*

PROOF. We shall not provide the entire proof, but by this point the result should at any rate seem believable, and in one direction the claim is clear: the stated conditions on the genus are necessary due to the Gauss-Bonnet theorem and the formula $\chi(\Sigma) = 2 - 2g$. It's easy to see that the sphere admits a metric with positive curvature: this is true for the induced metric coming from the standard embedding of S^2 in $\mathbb{R}^3$. Things are similarly simple for the torus, though the usual embedding of $\mathbb{T}^2$ into $\mathbb{R}^3$ (as a doughnut) is the wrong picture to look at. Instead take $\mathbb{R}^2$ with its standard flat metric and define $\mathbb{T}^2$ as $\mathbb{R}^2/\mathbb{Z}^2$: the translation invariance of the Euclidean metric implies that it gives a well defined metric on the quotient, and it is indeed locally flat.

The only part that is less obvious is that every surface of genus $g \geqslant 2$ admits a metric of negative curvature—in fact, by a famous result in the theory of surfaces, one can always find a metric that has *constant* curvature -1. One approach is to take the Poincaré half-plane $(\mathbb{H}, h)$ as a model (see Exercise 27.22) and show that every such surface can be constructed by drawing a smooth polygon in $(\mathbb{H}, h)$ and identifying certain edges appropriately. We refer to [**Spi99b**, Chapter 6, Addendum 1] for details. One can also prove this using geometric PDE methods, see for instance [**Tro92**]. $\qquad\square$

REMARK 29.27. For a surface Σ of genus $g \geqslant 2$, the standard way of embedding Σ into $\mathbb{R}^3$ as a surface with g handles is misleading in some respects: as a hypersurface in $\mathbb{R}^3$, its Gaussian curvature is sometimes positive and sometimes negative. Exercise 29.28 below shows that this will always be true, for *any* embedding of Σ in Euclidean $\mathbb{R}^3$, though the Gauss-Bonnet theorem guarantees at least that the part with negative curvature is the majority. Unfortunately (from the perspective of people who like to visualize things), there is no isometric embedding of any closed surface with everywhere negative curvature into $\mathbb{R}^3$. (This is a less deep observation than Hilbert's theorem about embeddings of the hyperbolic plane, mentioned in Remark 27.23. The exercise below does not say anything about the hyperbolic plane since it is not compact.)

EXERCISE 29.28. Prove: A closed surface Σ in Euclidean $\mathbb{R}^3$ cannot have $K_G \leqslant 0$ everywhere. *Hint: For some $R > 0$, Σ must lie inside the closed ball of radius R and touch its boundary tangentially at some point.*

29.3. Addendum: Polygons are disks. You should perhaps not bother to read this section unless you felt uncomfortable calling Equation (29.2) "obvious". Here is one way I can think of to prove it, using only the assumption that $\gamma : [a, b] \to \mathbb{R}^2$ is a piecewise-smooth simple closed curve bounding a compact region P. There may also be easier ways that I haven't thought of, but the basic idea of what I have in mind is to deform γ via a so-called **regular homotopy** to a smooth loop bounding a standard disk, for which (29.2) really is obvious. Let us call $\gamma : [a, b] \to \mathbb{R}^2$ a *smoothly immersed loop* if it is smooth and satisfies $\gamma(a) = \gamma(b)$, $\dot{\gamma}(a) = \dot{\gamma}(b)$ and $\dot{\gamma}(t) \neq 0$ for all t. One can associate to any smoothly immersed loop a smooth function $\phi : [a, b] \to \mathbb{R}$, unique modulo 2π, that measures the angle of $\dot{\gamma}(t) \in \mathbb{R}^2$ relative to a standard basis vector, and $\int_a^b \dot{\phi}(t)\,dt = \phi(b) - \phi(a)$ is then $2\pi k$ for some $k \in \mathbb{Z}$, called the **twisting number** of γ. A **regular homotopy** of loops is a smooth family of smoothly immersed loops $\{\gamma_s : [a, b] \to \mathbb{R}^2\}_{s \in [0,1]}$. Given such a family, the corresponding angle functions $\phi_s(t)$ can also be chosen to depend smoothly on both s and t, so that $\int_a^b \dot{\phi}_s(t)\,dt$ depends continuously on s, and therefore so does the twisting number. Since the latter is always in integer, this implies that it is the same for γ_0 and γ_1, i.e. the twisting number is invariant under regular homotopy. Our goal in the following is thus to show that, after smoothing the angles in order to make ∂P a smooth loop, it admits a regular homotopy to the boundary of a round disk, whose twisting number is clearly 1.

Step 1: Since $\gamma : [a, b] \to \mathbb{R}^2$ has only finitely-many nonsmooth points, each one is isolated, and it is therefore easy to modify γ by a C^0-small perturbation in small neighborhoods of these points to make it a smooth embedding with $\gamma(a) = \gamma(b)$ and $\dot{\gamma}(a) = \dot{\gamma}(b)$. This is an example

of what topologists call "smoothing the corners", and the contribution to $\int_a^b \dot{\phi}(t)\,dt$ from the small neighborhoods of t_j where this modification is done then corresponds to $\Delta\phi_j$, so the left hand side of (29.2) now contains only the integral term, and computes 2π times the twisting number of γ. (Note: It is really important in this step to make sure that you're using the right convention about the distinction between $\Delta\phi_j$ being $+\pi$ or $-\pi$, i.e. counterclockwise vs. clockwise rotations!)

Step 2: The compact region $P \subset \mathbb{R}^2$ bounded by γ is now a compact oriented smooth 2-manifold with connected boundary, and we claim that it is diffeomorphic to a disk $\mathbb{D}^2$. Indeed, the classification of surfaces (see e.g. [**Wen18**, Lecture 19] or [**Hir94**]) implies that P must be diffeomorphic to the complement of an open disk in a closed orientable surface Σ_g of some genus $g \geqslant 0$, so our claim is equivalent to the assertion that $g = 0$. To see this, one can add a "point at infinity" to $\mathbb{R}^2$, making it diffeomorphic to the sphere S^2, so that the unbounded region of $\mathbb{R}^2$ lying outside of γ becomes another compact oriented smooth 2-manifold with connected boundary, embedded in S^2. Applying the classification of surfaces again, this region is diffeomorphic to the complement of an open disk in a closed orientable surface Σ_h of some genus $h \geqslant 0$. Gluing the two pieces together presents S^2 as the connected sum of Σ_g and Σ_h, which is Σ_{g+h}. But S^2 is not diffeomorphic to Σ_{g+h} unless $g + h = 0$, thus $g = h = 0$.

Step 3: We now know that P is diffeomorphic to $\mathbb{D}^2$, and by the tubular neighborhood theorem, ∂P has a neighborhood in $\mathbb{R}^2$ diffeomorphic to $(-1, 1) \times S^1$, where ∂P itself is identified with $\{0\} \times S^1$. This is enough information to construct an open neighborhood $\mathcal{U} \subset \mathbb{R}^2$ of P with a diffeomorphism $\psi : \mathcal{U} \to B_r^2$ onto the open ball $B_r^2 \subset \mathbb{R}^2$ of some radius $r > 1$ such that $\psi(P)$ is the closed unit disk $\mathbb{D}^2$. Let us equip $\mathcal{U}$ with the Riemannian metric $g := \psi^*(dx^2 + dy^2)$. The point of this definition is that we can easily understand the geodesics for this metric: they are the images under ψ^{-1} of straight lines in B_r^2. As a consequence, we can now write

$$\gamma(t) = \exp_p^g(X(t)),$$

where $p := \psi^{-1}(0)$, $[a, b] \to T_p\mathcal{U} = \mathbb{R}^2 : t \mapsto X(t)$ is a parametrization of the unit circle in $T_p\mathcal{U}$ with respect to the metric g, and the superscript g is included to emphasize that we are using this metric (rather than the Euclidean metric) to define the exponential map.

Step 4: Since $(\mathcal{U}, g)$ is isometric via ψ to a standard ball with the Euclidean metric, $\exp_p^g$ defines a diffeomorphism from a ball in $T_p\mathcal{U}$ containing the loop $X(t)$ onto $\mathcal{U}$. The family

$$\gamma_s(t) := \exp_p^g(sX(t)), \qquad s \in [\epsilon, 1]$$

therefore defines a regular homotopy between γ and some loop $\gamma_\epsilon : [a, b] \to \mathcal{U}$ that can be assumed to lie in an arbitrarily small neighborhood of p by choosing $\epsilon > 0$ small.

Step 5: Consider the smooth family of Riemannian metrics $g_s := sg + (1 - s)g_E$ on $\mathcal{U}$ for $s \in [0, 1]$, where $g_E := dx^2 + dy^2$ is the Euclidean metric. For a sufficiently small neighborhood $\mathcal{O} \subset T_p\mathcal{U}$ of 0, we can assume that the corresponding exponential maps $\exp_p^{g_s}$ are embeddings of $\mathcal{O}$ onto open neighborhoods of p in $\mathcal{U}$. Now define

$$X_s(t) \in T_p\mathcal{U} = \mathbb{R}^2$$

for each $s \in [0, 1]$ and $t \in [a, b]$ as the unique positive rescaling of $X(t) \in \mathbb{R}^2$ that makes it a unit vector with respect to the metric g_s, and define another family of smooth loops by

$$\beta_s(t) := \exp_p^{g_s}(\epsilon X_s(t)) \qquad s \in [0, 1].$$

Taking $\epsilon > 0$ small enough so that $\epsilon X_s(t) \in \mathcal{O}$ for all s, t, these loops are all embeddings, and thus define a regular homotopy between $\beta_1 = \gamma_\epsilon$ and β_0, where the latter is a parametrization of the ϵ-disk about p with respect to the Euclidean metric. We conclude that γ_ϵ and therefore also γ have the same twisting number as β_0, which is 1.

30. The first Chern class

We are not yet done extracting mileage out of the formula

$$F = dA_\alpha.$$

Recall from §28.2: this relates a local connection 1-form $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{g})$ to a globally-defined Lie algebra-valued curvature 2-form $F \in \Omega^2(M, \mathfrak{g})$ on any vector bundle $E \to M$ with abelian structure group G carrying a compatible connection. The Gauss-Bonnet formula arose from the special case where E is the tangent bundle of an oriented Riemannian 2-manifold, so that the group G was $SO(2) \cong U(1)$, but this is not the only type of vector bundle with structure group $U(1)$ one might want to consider. We will explore what else can be done with this in §30.1 and §30.2, giving a rudimentary introduction to the much larger subject of characteristic classes and Chern-Weil theory. We then apply this again to the case $E = T\Sigma$ in §30.3, and deduce yet another useful interpretation of the Euler characteristic, including some discussion of spheres with hair.

30.1. An invariant of complex line bundles. The basic object of study in this section is a smooth complex line bundle

$$\pi : E \to M$$

over a manifold M of some dimension $n \in \mathbb{N}$. We are going to construct an algebraic invariant that can detect whether two such bundles are isomorphic. There are many reasons one might want to do this. One easy one to name is that vector bundles arise naturally in the tubular neighborhood theorem (cf. Exercise 23.4), where they serve as local models for neighborhoods of submanifolds, so if one can classify the isomorphism classes of vector bundles of a given rank over a given manifold, one obtains a picture of all possible neighborhoods of embeddings of that manifold into a larger one up to diffeomorphism. Since $U(1) \cong SO(2)$, classifying complex line bundles is equivalent to classifying oriented real bundles of rank 2, which arise whenever one studies the embeddings of an oriented manifold into another oriented manifold two dimensions larger. The simplest case of the latter situation is knot theory, which studies embeddings of S^1 into 3-manifolds, and this is only one of many situations in topology and related areas where certain types of vector bundles need to be classified.

The construction of our invariant will depend on two choices of auxiliary data:

(1) A bundle metric $\langle \ , \ \rangle$, thus making $E \to M$ a *Hermitian* line bundle and reducing its structure group from $GL(1, \mathbb{C})$ to $U(1)$;
(2) A metric connection ∇, represented in any $U(1)$-compatible local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$ by an imaginary-valued connection 1-form $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{u}(1))$.

These choices are crucial for the definition of the invariant, but we will see that the invariant itself does not depend on them.

A clue about the right thing to do arises out of the observation in §28.2 that in this situation, there is a globally-defined imaginary-valued 2-form $F \in \Omega^2(M, \mathfrak{u}(1))$ that matches dA_α for every choice of local trivialization. In particular, it is obvious that F is closed, but it might not be exact since each of the individual 1-forms A_α is defined only on the domain $\mathcal{U}_\alpha$ and not necessarily on all of M. As we saw in Lecture 13, the distinction between closed and exact forms on a manifold is measured by its de Rham cohomology, so one wonders whether the cohomology class represented by F might carry interesting information. A further hint in this direction comes from the following:

LEMMA 30.1. *If $\widehat{\nabla}$ is another metric connection on $E \to M$ with curvature 2-form $\widehat{F} \in \Omega^2(M, \mathfrak{u}(1))$, then $\widehat{F} = F + i\, d\lambda$ for some $\lambda \in \Omega^1(M)$.*

PROOF. The difference between two connections is always a bundle map, i.e. there exists a smooth bilinear bundle map $B : TM \oplus E \to E$ such that $\widehat{\nabla}_X v = \nabla_X v + B(X, v)$ for all $X \in \mathfrak{X}(M)$

and $v \in \Gamma(E)$, and B can also be interpreted as a bundle-valued 1-form

$$\beta \in \Omega^1(M, \mathrm{End}(E)), \qquad \beta(X)v := B(X, v).$$

Since the fibers E_p for all $p \in M$ are 1-dimensional, all endomorphisms $E_p \to E_p$ come from scalar multiplication, giving a natural isomorphism $\mathbb{C} \to \mathrm{End}(E_p)$ so that β can be replaced with a *complex*-valued 1-form $\beta \in \Omega^1(M, \mathbb{C})$ such that $\widehat{\nabla}_X v = \nabla_X v + \beta(X)v$. Writing down this relation in the local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$ then gives the relation

$$\widehat{A}_\alpha(X) = A_\alpha(X) + \beta(X),$$

where $\widehat{A}_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{u}(1))$ is the local connection 1-form for $\widehat{\nabla}$. Since A_α and $\widehat{A}_\alpha$ are both purely imaginary-valued, the same is therefore true for β, giving $\beta = i\lambda$ for some real-valued 1-form $\lambda \in \Omega^1(M)$. Taking the exterior derivative of $\widehat{A}_\alpha = A_\alpha + i\lambda$ then gives the stated relation between F and $\widehat{F}$. $\qquad\qquad\square$

Strictly speaking, we cannot talk about the de Rham cohomology class represented by $F \in \Omega^2(M, \mathfrak{u}(1))$ without slightly altering our previous definition of de Rham cohomology, because F is not a real-valued form. But that is easily fixed, and Lemma 30.1 then tells us that the following definition is independent of the choice of metric connection:

DEFINITION 30.2. The **first Chern class** of the complex line bundle $\pi : E \to M$ with bundle metric $\langle \ , \ \rangle$ is the de Rham cohomology class

$$c_1(E) := \left[-\frac{1}{2\pi i} F \right] \in H^2_{\mathrm{dR}}(M),$$

where $F \in \Omega^2(M, \mathfrak{u}(1))$ is the curvature 2-form associated to any choice of metric connection on $E \to M$.

The reason for the factor of 2π and the minus sign in this definition will become clear when we discuss computations in §30.2.

THEOREM 30.3. *The first Chern class of complex line bundles has the following properties.*
(1) *$c_1(E)$ is independent of the choice of bundle metric $\langle \ , \ \rangle$ on $E \to M$.*
(2) *For the trivial line bundle $E^0 := M \times \mathbb{C} \to \mathbb{C}$, $c_1(E^0) = 0$.*
(3) *If $E, E' \to M$ are two complex line bundles admitting a bundle isomorphism $E \to E'$, then $c_1(E) = c_1(E')$.*
(4) *For any complex line bundle $E \to M$ and any smooth map $f : N \to M$, the pullback bundle $f^*E \to N$ has $c_1(f^*E) = f^*c_1(E) \in H^2_{\mathrm{dR}}(N)$.*

PROOF. The easiest property to prove is (2), so we start with that: on the trivial bundle we can choose ∇ to be the trivial connection, and there is an obvious global trivialization in which the resulting connection 1-form vanishes identically, implying the same for the curvature 2-form and thus $[-F/2\pi i] = 0$. Alternatively, one can reach the same conclusion without assuming anything about the connection: it suffices to observe that since a trivialization can be defined on $\mathcal{U}_\alpha := M$, there is a globally-defined connection 1-form $A_\alpha \in \Omega^1(M, \mathfrak{u}(1))$, whose exterior derivative is F, hence $-F/2\pi i$ is exact.

Moving on, we will prove a slightly stricter version of property (3) that depends on a bundle metric, and then use this to prove (1). For two line bundles $E \to M$ and $E' \to N$ equipped with bundle metrics, a smooth linear bundle map $\Psi : E \to E'$ covering a smooth map $\psi : M \to N$ will be called a **bundle isometry** if for every $p \in M$, Ψ defines an isomorphism $E_p \to E'_{\psi(p)}$ that is unitary, meaning it preserves the inner products. If $E, E' \to M$ admit a bundle isomorphism $\Psi : E \to E'$, then for any choice of bundle metric on E, there is a unique one on E' that makes Ψ a

bundle isometry. With this data in place, any metric connection ∇ on E can be "pushed forward" via Ψ to define a metric connection ∇' on E', namely by

$$\nabla'_X v := \Psi\left(\nabla_X(\Psi^{-1}v)\right).$$

It is an easy exercise to check that ∇'_X satisfies the Leibniz rule required to be a connection on E' and is also compatible with the bundle metric. We can also use Ψ to push forward local trivializations: given a trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$, we can define a trivialization $\Phi'_\alpha : E'|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$ by

$$\Phi'_\alpha := \Phi_\alpha \circ \Psi^{-1}.$$

For this choice, the section $v \in \Gamma(E)$ has the same local representation $v_\alpha : \mathcal{U}_\alpha \to \mathbb{C}$ as the section $\Psi v \in \Gamma(E')$, and the local connection 1-forms $A_\alpha, A'_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{u}(1))$ from our two connections via these two trivializations are exactly the same. It follows that the curvature 2-forms for ∇ and ∇' are identical over $\mathcal{U}_\alpha$, and since the same trick can be done on any region where E is trivializable, they are therefore identical everywhere, proving $c_1(E) = c_1(E')$.

We can now prove (1) as follows. The space of bundle metrics on $E \to M$ is convex, so any two bundle metrics $\langle\ ,\ \rangle_0$ and $\langle\ ,\ \rangle_1$ can be connected by a smooth family of bundle metrics

$$\langle\ ,\ \rangle_s := s\langle\ ,\ \rangle_1 + (1-s)\langle\ ,\ \rangle_0, \qquad s \in [0,1].$$

Let $\widehat{E} \to [0,1] \times M$ denote the pullback of $E \to M$ via the obvious projection map $[0,1] \times M \to M$, and endow $\widehat{E}$ with a bundle metric such that the inner product at the point $(s,p) \in [0,1] \times M$ is $\langle\ ,\ \rangle_s$ at the point p. Choosing a metric connection on $\widehat{E}$, parallel transport along the paths $s \mapsto (s,p)$ for each $p \in M$ now defines a bundle isometry $(E,\langle\ ,\ \rangle_0) \to (E,\langle\ ,\ \rangle_1)$, and the result of the previous paragraph thus implies that the two definitions of $c_1(E)$ via these two bundle metrics match.

Finally, we prove (4): assuming $E \to M$ is a line bundle with bundle metric $\langle\ ,\ \rangle$ and $f : N \to M$ is a smooth map, equip $f^*E \to N$ with the unique bundle metric so that the canonical bundle map $f^*E \to E$ covering f is a bundle isometry. For any metric connection ∇ on E, the pullback connection on f^*E is then also compatible with the bundle metric, and the discussion following Equation (21.3) shows that for any local trivialization $\Phi_\alpha : E|_{\mathcal{U}_\alpha} \to \mathcal{U}_\alpha \times \mathbb{C}$ with connection 1-form $A_\alpha \in \Omega^1(\mathcal{U}_\alpha, \mathfrak{u}(1))$, there is a pullback trivialization of f^*E over $f^{-1}(\mathcal{U}_\alpha) \subset N$ in which the connection 1-form for the pullback connection is f^*A_α. Taking the exterior derivative, it follows that the pullback $f^*F \in \Omega^2(N, \mathfrak{u}(1))$ of the curvature 2-form $F \in \Omega^2(M, \mathfrak{u}(1))$ for ∇ is the curvature 2-form for the pullback connection, thus $c_1(f^*E) = [-f^*F/2\pi i] = f^*[-F/2\pi i] = f^*c_1(E)$. $\square$

30.2. Computing the first Chern number. For $c_1(E) \in H^2_{\mathrm{dR}}(M)$ to be a truly useful invariant, we need a practical means of computing it. As a rule, the best way to understand a 2-dimensional cohomology class $[\omega] \in H^2_{\mathrm{dR}}(M)$ is by integrating it over closed oriented 2-dimensional submanifolds $\Sigma \subset M$: the result is independent of the 2-form $\omega \in \Omega^2(M)$ representing $[\omega]$ since, by Stokes' theorem, integrals of exact forms over closed manifolds always vanish. Integrating ω over $\Sigma \subset M$ is the same as computing $\int_\Sigma j^*\omega$ for the natural inclusion map $j : \Sigma \hookrightarrow M$. More generally, one can also consider integrals of the form $\int_\Sigma f^*\omega$ for arbitrary closed oriented surfaces Σ and smooth maps $f : \Sigma \to M$, which need not be embeddings; in this situation, if ω represents $c_1(E) \in H^2_{\mathrm{dR}}(M)$, then $f^*\omega$ represents $c_1(f^*E) \in H^2_{\mathrm{dR}}(\Sigma)$ according to Theorem 30.3. It can be deduced from de Rham's theorem and a result of Thom[75] that these integrals for all possible choices of surfaces Σ and maps $f : \Sigma \to M$ completely characterize $[\omega] \in H^2_{\mathrm{dR}}(M)$. I will not prove that here, but am mentioning it only as support for the following assertion: if you want to

---

[75]The result in question comes from the famous paper [**Tho54**], and states that every singular homology class $A \in H_k(M;\mathbb{Z})$ can be written as $A = \frac{1}{q}f_*[\Sigma]$ for some closed oriented k-manifold Σ, smooth map $f : \Sigma \to M$ and $q \in \mathbb{N}$.

compute $c_1(E)$ in general, then it suffices in principle if you know how to compute the first Chern classes of bundles over *closed oriented surfaces*, as this is what the pullback bundles $f^*E \to \Sigma$ are. Moreover, the essential information about $c_1(E)$ for a bundle $E \to \Sigma$ is contained in the integral $\int_\Sigma \omega \in \mathbb{R}$ for any choice of 2-form ω representing $c_1(E) \in H^2_{\mathrm{dR}}(\Sigma)$. This number deserves a name.

DEFINITION 30.4. For a complex line bundle E over a closed oriented surface Σ, the number

$$\int_\Sigma c_1(E) := \int_\Sigma \omega,$$

defined by choosing any representative $\omega \in \Omega^2(M)$ of the cohomology class $c_1(E) \in H^2_{\mathrm{dR}}(M)$, is called the **first Chern number** of $E \to \Sigma$.

According to the definition of $c_1(E)$, one can compute the first Chern number in principle by choosing a bundle metric and metric connection, which give rise to an imaginary-valued curvature 2-form $F \in \Omega^2(M, \mathfrak{u}(1))$, and then integrating

$$\int_\Sigma c_1(E) = -\frac{1}{2\pi i} \int_\Sigma F.$$

We already know how to do this in the case $E = T\Sigma$ with the Levi-Cività connection: the answer comes from the Gauss-Bonnet formula, which we'll come back to in §30.3 below. But without making any further assumptions about the line bundle $E \to \Sigma$ or the connection, it is possible to compute this integral in another way that has interesting applications. Let us make the following assumption, which we will see below is completely realistic: suppose Σ can be decomposed into two compact (but not necessarily connected) surfaces $\Sigma_\alpha, \Sigma_\beta \subset \Sigma$ with a common boundary consisting of a finite set of disjoint circles $C_1, \ldots, C_N \subset \Sigma$,

$$\partial \Sigma_\beta = \partial \Sigma_\alpha = \coprod_{j=1}^N C_j,$$

such that both subsets are contained in open neighborhoods $\mathcal{U}_\alpha, \mathcal{U}_\beta \subset \Sigma$ on which there exist U(1)-compatible trivializations Φ_α and Φ_β respectively. Denote the transition function relating these trivializations by

$$g := g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{U}(1),$$

and observe that it is defined in particular on each of the circles C_j. Both Σ_α and Σ_β inherit orientations from Σ such that the boundary orientations of $\partial \Sigma_\alpha$ and $\partial \Sigma_\beta$ are opposite; let us orient the individual circles C_j to match the boundary orientation of $\partial \Sigma_\beta$. Stokes' theorem then gives

$$(30.1) \quad \int_\Sigma F = \int_{\Sigma_\alpha} dA_\alpha + \int_{\Sigma_\beta} dA_\beta = \int_{\partial \Sigma_\alpha} A_\alpha + \int_{\partial \Sigma_\beta} A_\beta = \int_{\partial \Sigma_\beta} (A_\beta - A_\alpha) = \sum_{j=1}^N \int_{C_j} (A_\beta - A_\alpha).$$

A formula relating A_α and A_β to each other on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ was worked out in Exercise 20.9, namely

$$A_\alpha(X) = g(p)^{-1} A_\beta(X) g(p) + g(p)^{-1} \, dg(X), \qquad \text{for } p \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta, \, X \in T_p M,$$

and in the present case, the fact that U(1) is abelian simplifies it to

$$A_\alpha = A_\beta + g^{-1} \, dg \qquad \text{on } \mathcal{U}_\alpha \cap \mathcal{U}_\beta,$$

so that (30.1) becomes

$$\int_\Sigma F = -\sum_{j=1}^N \int_{C_j} g^{-1} \, dg.$$

This formula is further confirmation that $c_1(E)$ is an essentially topological quantity with no dependence on the choice of connection. Now observe that since g takes values in $\mathrm{U}(1)$, we can write it as $g = e^{i\theta}$ for a uniquely defined smooth function

$$\theta : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathbb{R}/2\pi\mathbb{Z}.$$

It should be emphasized that θ cannot necessarily be defined as a *real*-valued function on $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$, at least not if we want it to be continuous, though a real-valued version could indeed be defined on a sufficiently small neighborhood of any given point in $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$. Such a local function would be unique only up to the addition of constant multiples of 2π, but this means that its *differential* is uniquely defined as a perfectly ordinary closed (but not necessarily exact) real-valued 1-form

$$d\theta \in \Omega^1(\mathcal{U}_\alpha \cap \mathcal{U}_\beta).$$

We can therefore write $g^{-1}\, dg = e^{-i\theta} d(e^{i\theta}) = ie^{-i\theta}e^{i\theta}\, d\theta = i\, d\theta$, giving

$$(30.2) \qquad \int_\Sigma c_1(E) = -\frac{1}{2\pi i} \int_\Sigma F = \frac{1}{2\pi} \sum_{j=1}^{N} \int_{C_j} d\theta.$$

This last integral looks at first like it should vanish, but remember that $d\theta$ is not necessarily an exact 1-form since θ is not a real-valued function. We encountered something similar in Example 13.12, and the following definition provides a useful topological interpretation for integrals of this type.

DEFINITION 30.5. Suppose S is an oriented manifold diffeomorphic to S^1, and $f : S \to \mathbb{C}\backslash\{0\}$ is a smooth function. The **winding number**

$$\mathrm{wind}_S(f) \in \mathbb{Z}$$

of f is then the unique integer with the following property: for any smooth orientation-preserving map $\gamma : [0,1] \to S$ that satisfies $\gamma(0) = \gamma(1)$ and is an embedding on $(0,1)$, and any smooth functions $\rho, \phi : [0,1] \to \mathbb{R}$ such that $\rho > 0$ and $f(\gamma(t)) = \rho(t)e^{i\phi(t)}$ for all t,

$$\mathrm{wind}_S(f) = \frac{1}{2\pi}\left[\phi(1) - \phi(0)\right].$$

To see that $\mathrm{wind}_S(f)$ in Definition 30.5 is independent of the various choices involved, one can reinterpret it as the integral

$$(30.3) \qquad \mathrm{wind}_S(f) = \frac{1}{2\pi} \int_S d\theta,$$

where the 1-form $d\theta \in \Omega^1(S)$ is defined from the unique smooth function $\theta : S \to \mathbb{R}/2\pi\mathbb{Z}$ satisfying $f(p) = r(p)e^{i\theta(p)}$ for some positive function $r : S \to \mathbb{R}$. Indeed, suppose $\gamma : [0,1] \to S$ is an orientation-preserving parametrization of S as in the definition: then we can write $f(\gamma(t)) = r(\gamma(t))e^{i\phi(t)}$ for some smooth function $\phi : [0,1] \to \mathbb{R}$ such that $\theta(\gamma(t))$ is the image of $\phi(t)$ under the quotient projection $\mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z}$. Differentiating this relation between θ and ϕ gives $d\phi = d\theta \circ T\gamma = \gamma^* d\theta$. Now for any small $\epsilon > 0$, γ is an orientation-preserving diffeomorphism of $[\epsilon, 1 - \epsilon]$ onto its image $S_\epsilon \subset S$, so the change-of-variables formula gives $\int_{S_\epsilon} d\theta = \int_{[\epsilon,1-\epsilon]} \gamma^* d\theta$. After taking $\epsilon \to 0$, this gives

$$\int_S d\theta = \int_{[0,1]} \gamma^* d\theta = \int_{[0,1]} d\phi = \phi(1) - \phi(0),$$

thus proving (30.3). (Caution: If we had manipulated the symbols in this last equation without thinking about their meaning, we might have said $\int_{[0,1]} \gamma^* d\theta = \int_{[0,1]} d(\gamma^* \theta) = \int_{[0,1]} d(\theta \circ \gamma) = \theta(\gamma(1)) - \theta(\gamma(0)) = 0$. This is where it is crucial to remember that θ is not a *real*-valued function

on S, but instead takes values in the manifold $\mathbb{R}/2\pi\mathbb{Z}$. Thus $d(\gamma^*\theta)$ cannot be interpreted as an exact 1-form, and we cannot use Stokes' theorem to compute it.)

An easy corollary of (30.3) is that $\mathrm{wind}_S(f)$ is not only independent of the choice of parametrization $\gamma : [0,1] \to S$, but it is also *homotopy invariant* : if $f_0, f_1 : S \to \mathbb{C}\backslash\{0\}$ are two ends of a smooth family of nowhere-zero functions $\{f_s : S \to \mathbb{C}\backslash\{0\}\}_{s\in[0,1]}$, then $\mathrm{wind}_S(f_0) = \mathrm{wind}_S(f_1)$. One can see this from the fact that the integral $\int_S d\theta$ in this situation will depend continuously on the parameter s, and since it is also an integer multiple of 2π, this implies that it cannot change at all.

We can now rewrite (30.2) in terms of winding numbers:

PROPOSITION 30.6. *For the complex line bundle $E \to \Sigma = \Sigma_\alpha \cup \Sigma_\beta$ as described above with transition function $g = g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to \mathrm{U}(1)$,*

$$\int_\Sigma c_1(E) = \sum_{j=1}^N \mathrm{wind}_{C_j}(g).$$

$\square$

It should now be obvious why the factor of $1/2\pi$ was included in the definition of $c_1(E)$: it makes the value of $\int_\Sigma c_1(E)$ an integer, a fact which was far from obvious in its definition.

Proposition 30.6 is of practical use for computing first Chern numbers. It can also be used to approach the following question, which you might not expect should have a well-defined answer at all:

QUESTION 30.7. *How many zeroes should a section $s \in \Gamma(E)$ be expected to have?*

If you think of sections of a complex line bundle over a surface as something analogous to complex-valued functions on an open domain in $\mathbb{R}^2$, then this does not at first seem like a sensible question, because the number of zeroes will generally depend on the choice of function, and one could always just choose a nonzero constant function for which the answer zero. But on a line bundle over a closed surface, nowhere-zero sections might not exist—indeed, a nowhere-zero section in this situation is equivalent to a frame, so such a thing exists if and only if the bundle is globally trivial. This observation hints that the issue in Question 30.7 is fundamentally topological, at least if we have the correct interpretation of the words "how many". Let us restrict our attention to smooth sections $s : \Sigma \to E$ such that the zero set

$$s^{-1}(0) := \left\{ p \in \Sigma \mid s(p) = 0 \in E_p \right\} \subset \Sigma$$

is finite. One can show that all sections in an open and dense subset of $\Gamma(E)$ have this property (see Remark 30.12 below). One could now count the number of elements in $s^{-1}(0)$, but this notion of counting is too naive to give an answer independent of the choice of section. The right interpretation of the words "how many" turns out to be one that attaches to each individual zero an integer-valued weight, and this weight can be defined as a winding number:

DEFINITION 30.8. Suppose $p \in \Sigma$ is an isolated point in the zero set $s^{-1}(0)$ of a section $s \in \Gamma(E)$. The **index** of s at p (also sometimes called the **order** of the zero p) is defined as the integer

$$\mathrm{ind}(s;p) := \mathrm{wind}_{\partial\mathcal{D}}(s_\alpha),$$

where $\mathcal{D} \subset \Sigma$ is a small disk containing p in its interior such that $s^{-1}(0) \cap \mathcal{D} = \{p\}$, and $s_\alpha : \mathcal{D} \to \mathbb{C}$ is the local representative of s in some trivialization Φ_α of E defined on a neighborhood of $\mathcal{D}$.

REMARK 30.9. The winding number in Definition 30.8 requires $\partial\mathcal{D}$ to be oriented, so we assign it the boundary orientation, where $\mathcal{D}$ inherits an orientation from Σ. Reversing the orientation of

Σ thus changes the sign of $\text{ind}(s;p)$, and the index can only be defined up to a sign if Σ is not orientable.

EXERCISE 30.10. Use the homotopy-invariance of winding numbers to show that the index $\text{ind}(s;p)$ in Definition 30.8 does not depend on the choices of disk $\mathcal{D} \subset \Sigma$ surrounding p and local trivialization Φ_α over $\mathcal{D}$.
Hint: The crucial detail is that s does not vanish on $\mathcal{D}$ except at the point p.

EXERCISE 30.11. Recall from Exercise 19.7 that at any point $p \in s^{-1}(0)$ in the zero-set of a section $s \in \Gamma(E)$, there is a well-defined *linearization* $Ds(p) : T_p\Sigma \to E_p$. For the following statement, we can regard E_p as an oriented 2-dimensional *real* vector space by defining any basis of the form (v, iv) for $v \neq 0 \in E_p$ to be positively oriented. Convince yourself that this orientation is well defined, and then prove the following: if $Ds(p)$ is invertible, then $\text{ind}(s;p) = \pm 1$, positive if $Ds(p)$ is orientation preserving and negative if $Ds(p)$ is orientation reversing.

REMARK 30.12. We do not have time for a proper treatment of transversality theory in this course, but if you know the basic definitions, you may be able to convince yourself without much difficulty that the linearization $Ds(p) : T_p\Sigma \to E_p$ at a zero $p \in s^{-1}(0)$ is invertible if and only if the intersection at $0 \in E_p \subset E$ between the zero-section $Z := \bigcup_{p \in \Sigma} 0 \subset E$ and the submanifold $s(\Sigma) \subset E$ is transverse. General results in transversality theory (see e.g. [Hir94]) then imply that all zeroes satisfy this condition for all sections in some open and dense subset of $\Gamma(E)$. This is why we know there always exist sections whose zero-sets are finite.

EXERCISE 30.13. Suppose $s \in \Gamma(E)$ is an isolated zero $p \in s^{-1}(0)$ with $\text{ind}(s;p) \neq 0$. Show that for any neighborhood $\mathcal{U} \subset \Sigma$ of p, any sufficiently C^0-small perturbation of s must also vanish somewhere in $\mathcal{U}$. In other words, zeroes with nonvanishing index cannot be perturbed away.
Hint: Consider only perturbations of s such that the winding number along some fixed circle around p does not change.

EXERCISE 30.14. On the trivial complex line bundle $E = \mathbb{R}^2 \times \mathbb{C} \to \mathbb{R}^2$, find an example of a section $s \in \Gamma(E)$ with an isolated zero at one point $p \in \mathbb{R}^2$ with $\text{ind}(s;p) = 0$, such that s admits small perturbations with no zeroes at all.

DEFINITION 30.15. Suppose $s \in \Gamma(E)$ has a finite zero set. The **algebraic count of zeroes** of s is the integer

$$\#s^{-1}(0) := \sum_{p \in s^{-1}(0)} \text{ind}(s;p) \in \mathbb{Z}.$$

THEOREM 30.16. *Suppose Σ is a closed oriented surface and $E \to \Sigma$ is a complex line bundle. Then for any section $s \in \Gamma(E)$ with at most finitely-many zeroes,*

$$\#s^{-1}(0) = \int_\Sigma c_1(E).$$

PROOF. Assuming $s^{-1}(0) \subset \Sigma$ is finite, choose for each $p \in s^{-1}(0)$ a small closed disk $\mathcal{D}_p \subset \Sigma$ whose boundary encircles p, and assume all of these disks are small enough so that they do not intersect each other and they are contained in a neighborhood on which E is trivializable. Set

$$\Sigma_\beta := \bigcup_{p \in s^{-1}(0)} \mathcal{D}_p, \qquad \Sigma_\alpha := \overline{\Sigma \backslash \Sigma_\beta},$$

and let $v \in \Gamma(E|_{\Sigma_\beta})$ denote an arbitrary choice of section over some open neighborhood $\mathcal{U}_\beta \subset \Sigma$ of Σ_β such that $|v| \equiv 1$, hence v can be interpreted as a U(1)-compatible frame over $\mathcal{U}_\beta$ and gives rise to a corresponding trivialization Φ_β. On $\mathcal{U}_\alpha := \Sigma \backslash s^{-1}(0)$, s itself determines a U(1)-compatible

trivialization, defined by interpreting the normalized section $s/|s|$ as a U(1)-compatible frame, and we can denote the corresponding trivialization by Φ_α. This means that the local representation of s with respect to Φ_α is a positive real-valued function $s_\alpha > 0$; its representation with respect to Φ_β is related to this by $s_\beta = g s_\alpha$ for the transition map $g := g_{\beta\alpha} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to$ U(1), and is therefore just a positive rescaling of g. This proves that for each zero $p \in s^{-1}(0)$,

$$\mathrm{wind}_{\partial \mathcal{D}_p}(s_\beta) = \mathrm{wind}_{\partial \mathcal{D}_p}(g),$$

so the equality of $\#s^{-1}(0)$ and $\int_\Sigma c_1(E)$ now follows from Proposition 30.6. $\square$

EXERCISE 30.17. By counting zeroes of sections, show that for any pair of complex line bundles $E, E' \to \Sigma$ over a closed oriented surface Σ, $\int_\Sigma c_1(E \otimes E') = \int_\Sigma c_1(E) + \int_\Sigma c_1(E')$.

EXERCISE 30.18. For any vector bundle $E \to M$ over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, with dual bundle $E^* \to M$, there is a canonical bilinear bundle map from $E \otimes E^*$ to the trivial line bundle $M \times \mathbb{F} \to M$, defined at each point $P \in M$ by

$$E_p \otimes E_p^* \to \mathbb{F} : v \otimes \lambda \mapsto \lambda(v).$$

(a) Show that if $\mathrm{rank}(E) = 1$, the bundle $E \otimes E^*$ is trivial.
(b) Show that for any complex line bundle E over a compact oriented surface Σ, $\int_\Sigma c_1(E^*) = -\int_\Sigma c_1(E)$.

Exercise 30.18 reveals an interesting difference between real and complex vector bundles. For any real bundle $E \to M$, choosing a bundle metric $\langle\ ,\ \rangle$ gives rise to a bundle isomorphism

$$E \to E^* : v \mapsto \langle v,\ \rangle.$$

This trick does not work in the complex case because bundle metrics are complex linear only in one argument and complex *antilinear* in the other, so the map $E \to E^*$ above can be defined, but it is complex antilinear on each fiber and thus not a complex bundle isomorphism. Exercise 30.18 shows that this is not just a defect in our method of finding isomorphisms: the bundles E and E^* really are not generally isomorphic in the complex case, as their first Chern classes will differ whenever they are nonzero. There do exist complex line bundles with $c_1(E) \neq 0$: we will see some explicit examples in the next section, and more generally, it is not hard to construct examples over surfaces that have arbitrary interger values for $\int_\Sigma c_1(E)$. The trick is to glue simpler pieces together in clever ways, e.g. if you present the sphere S^2 with its north and south poles $p_\pm \in S^2$ as the union of the two open subsets $\mathcal{U}_\pm := S^2 \backslash \{p_\pm\}$, then you can take two trivial line bundles $E_\pm := \mathcal{U}_\pm \times \mathbb{C} \to \mathcal{U}_\pm$, and glue these together to produce a bundle $E \to S^2$ with local trivializations over $\mathcal{U}_+$ and $\mathcal{U}_-$ having any desired transition function $g : \mathcal{U}_+ \cap \mathcal{U}_- \to$ U(1).

30.3. The Poincaré-Hopf theorem on surfaces. If $E \to \Sigma$ is the tangent bundle $T\Sigma$ of a closed oriented surface with a Riemannian metric, one can choose ∇ to be the Levi-Cività connection, and by Proposition 28.8 and the Gauss-Bonnet formula, the first Chern number then becomes

$$\int_\Sigma c_1(T\Sigma) = -\frac{1}{2\pi i} \int_\Sigma F = \frac{1}{2\pi} \int_\Sigma iF = \frac{1}{2\pi} \int_\Sigma K_G\, d\mathrm{vol}_\Sigma = \chi(\Sigma).$$

This is the most famous explicit computation of a first Chern number, and is the main one that you should commit to memory if you don't have space for any others. Combining it with the results of the previous section now gives a new interpretation of the Euler characteristic:

THEOREM 30.19 (Poincaré-Hopf). *For any vector field $X \in \mathfrak{X}(\Sigma)$ with at most finitely-many zeroes on a closed oriented surface Σ, the algebraic count of zeroes is*

$$\#X^{-1}(0) = \chi(\Sigma).$$

$\square$

I recommend taking a moment to think about what this implies for the most familiar surfaces. For the torus $\mathbb{T}^2$, whose Euler characteristic according to Exercise 29.16 is 0, it is consistent with the observation that nowhere-zero vector fields on $\mathbb{T}^2$ are easy to construct. The most famous consequence of the Poincaré-Hopf theorem applies to S^2, whose Euler characteristic is 2: it is often summarized by the colorful phrase, "you cannot comb the hair on a sphere".

COROLLARY 30.20. *There does not exist a nowhere-zero vector field on S^2.* $\square$

EXERCISE 30.21. For a closed oriented surface Σ_g of genus $g \geqslant 0$, we can use the Poincaré-Hopf theorem to compute $\chi(\Sigma_g)$ without needing to choose triangulations. Recall from Exercise 29.22 the notion of a *pair of pants*.

(a) Show that a pair of pants admits a smooth vector field that is tangent to the boundary and nonzero there, and has only one zero in the interior, with index -1.
Hint: Just try to draw the flow lines. They should form the leaves of a foliation, with one singular point where two leaves intersect transversely.

(b) By gluing together pairs of pants, show that Σ_g admits a vector field with exactly $2 - 2g$ zeroes, all of index -1.

30.4. Addendum: counting zeroes in general. Having seen the definition of the first Chern class, it will surely not surprise you to learn that there is also a second, and a third and so forth: for every $k \in \mathbb{N}$ one can associate to every complex vector bundle $\pi : E \to M$ of arbitrary rank a *kth Chern class*

$$c_k(E) \in H^{2k}_{\mathrm{dR}}(M).$$

Its definition when either $k \geqslant 2$ or $\mathrm{rank}(E) > 1$ is more complicated than we have space to discuss here: this is the subject of a large sub-branch of differential topology known as *Chern-Weil theory*, which is one of the topics we might discuss near the end of next semester's followup course. One can also define analogous so-called *characteristic classes* for real vector bundles $E \to M$, such as the *Pontryagin classes*

$$p_k(E) \in H^{4k}_{\mathrm{dR}}(M)$$

for each $k \in \mathbb{N}$. In Chern-Weil theory, characteristic classes of a bundle $E \to M$ with structure group G are always constructed in terms of closed forms determined by the curvature of some chosen G-compatible connection, on which the cohomology class turns out not to depend. One can show as in §30.2 that the integrals of these classes over closed oriented submanifolds of suitable dimensions are always integers, despite this being highly nonobvious from their definition. This hints at the fact that all characteristic classes can also be constructed by completely different methods, using algebraic topology, where they live naturally in $\mathbb{Z}$-modules such as singular or Čech cohomology with integer coefficients, rather than the real vector space $H^*_{\mathrm{dR}}(M)$. (The major exceptions to this last statement are the Stiefel-Whitney classes, which can be defined for all real vector bundles and take values in cohomology with $\mathbb{Z}_2$ coefficients, thus there is no sensible way to define them in de Rham cohomology.) The fact that the Pontryagin numbers are integers played a major role e.g. in Milnor's discovery that the topological manifold S^7 admits smooth structures not diffeomorphic to its standard one. The fact that the widely differing constructions of characteristic classes via algebraic topology vs. Chern-Weil theory give equivalent results is also a deep theorem with many applications.

I'd like to add a word about one other characteristic class which places the discussion of §30.2 into a wider context. There is a certain perspective from which it is not at all surprising that the question "How many zeroes should a section $s : M \to E$ have?" might have a well-defined answer. The idea is roughly as follows: suppose $\pi : E \to M$ is an *oriented* real vector bundle of rank n over a closed oriented n-manifold M, and call a section $s \in \Gamma(E)$ *generic* if for every point $p \in s^{-1}(0)$

in its zero-set, the linearization

$$Ds(p) : T_p M \to E_p$$

is invertible. As mentioned in Remark 30.12, there is always an open and dense set of sections in $\Gamma(E)$ that satisfy this condition, and the inverse function theorem then implies that the zeroes of s are isolated; since M is assumed compact, this means there are only finitely many. Generalizing Exercise 30.11, one can now associate an index $\operatorname{ind}(s; p) = \pm 1$ to each zero by defining it to be $+1$ if $Ds(p)$ is orientation preserving and -1 if $Ds(p)$ is orientation reversing.

The key idea now is to regard the zero set $s^{-1}(0)$ as a compact oriented 0-dimensional submanifold of M, with the orientation of each point defined by the sign of $\operatorname{ind}(s; p)$. Now if $s_0, s_1 \in \Gamma(E)$ are two generic sections, we can find a smooth homotopy between them, i.e. a map

$$H : [0, 1] \times M \to E$$

such that $s_t := H(t, \cdot) \in \Gamma(E)$ for each t; such a map always exists, for instance the linear interpolation $H(t, p) := t s_1(p) + (1 - t) s_0(p)$. By a nontrivial bit of transversality theory, one can always make a small perturbation of H so as to assume without loss of generality that its image in E meets the zero-section transversely, in which case $H^{-1}(0) \subset [0, 1] \times M$ is a smooth oriented 1-dimensional submanifold with boundary. Then

$$\partial \left(H^{-1}(0) \right) = \left(\{1\} \times s_1^{-1}(0) \right) \cup \left(\{0\} \times (-s_0^{-1}(0)) \right),$$

where the minus sign on the right hand side indicates reversal of orientation. The 1-manifold $H^{-1}(0)$ will generally have multiple connected components, which come in three flavors:

(1) Circles in the interior of $[0, 1] \times M$;
(2) Arcs with one boundary point in $\{1\} \times s_1^{-1}(0)$, and the other a point in $\{0\} \times s_0^{-1}(0)$ with the same orientation;
(3) Arcs with both boundary points in either $\{1\} \times s_1^{-1}(0)$ or $\{0\} \times s_0^{-1}(0)$, having opposite orientations.

The result is that the points in the disjoint union of $s_1^{-1}(0)$ with $s_0^{-1}(0)$ come in pairs: matching pairs of zeros of s_1 and s_0, or cancelling pairs of zeros of s_1 alone or s_0 alone. Thus the count of positive points in $s_1^{-1}(0)$ minus negative points in $s_1^{-1}(0)$ is the same as the corresponding count for s_0, and we conclude that for all *generic* sections $s \in \Gamma(E)$, the algebraic count

$$\# s^{-1}(0) = \sum_{p \in s^{-1}(0)} \operatorname{ind}(s; p) \in \mathbb{Z}$$

is the same. This number is called the **Euler number** of the bundle $E \to M$, and it corresponds to an **Euler class** $e(E) \in H_{\mathrm{dR}}^n(M)$ such that $\int_M e(E)$ is the Euler number. In this context, Theorem 30.16 can be rephrased as the statement that for any complex line bundle $E \to M$, if one regards it as an oriented real vector bundle of rank 2, its Euler class matches its first Chern class. The reason however for the terminology is that when E is TM for a closed oriented manifold M, its Euler number matches the Euler characteristic:

$$\int_M e(E) = \chi(M).$$

This is the general version of the Poincaré-Hopf theorem. It can be proved in various ways, depending on whether one prefers to define $e(E)$ via Chern-Weil theory or algebraic topology. Add this to our to-do list for next semester.

If you want to read the full details on why algebraic counts like $\# s^{-1}(0)$ do not depend on the choice of generic section, and how to generalize them without always assuming $\operatorname{ind}(s; p) = \pm 1$, I highly recommend Milnor's short book [**Mil97**]. It's something every mathematics student should read sooner or later.

EXERCISE 30.22. The argument sketched above for proving $\#s_0^{-1}(0) = \#s_1^{-1}(0)$ appealed to the classification of compact 1-manifolds with boundary, i.e. their connected components are each diffeomorphic to either a circle S^1 or a compact interval $[0, 1]$. This is a basic result in topology, but one doesn't really need to use it for this purpose: the main fact we actually needed was that whenever M is a compact oriented 1-manifold with boundary, the signed count of boundary points vanishes:

$$\sum_{p \in \partial M} \varepsilon(p) = 0,$$

where $\varepsilon : \partial M \to \{1, -1\}$ is the boundary orientation (see §12.1). Prove this without assuming anything about the topology of M.

Hint: Integrate an exact 1-form over M.

31. Sectional curvature

In this last lecture of the semester, we'll introduce a generalization of the Gaussian curvature that makes sense in all dimensions, not just on surfaces. Like the Gaussian curvature on surfaces, the sectional curvature of a Riemannian n-manifold contains all the same information as the Riemann tensor, but packaged in a more useful way: for instance, sectional curvature makes it possible to define the notions of positive/negative/zero curvature in arbitrary dimensions and observe qualitative distinctions between them. We will motivate the definition by investigating a natural question about geodesics: when can we conclude that a geodesic segment connecting two (not necessarily nearby) points actually is the *only* geodesic connecting them?

The following thought experiment shows that this question clearly has something to do with curvature: suppose $\ell_1, \ell_2 \subset \Sigma$ are two distinct embedded geodesic curves connecting distinct points p and q in a Riemannian 2-manifold (Σ, g). (We know from the example of S^2 that this scenario sometimes happens.) If ℓ_1 and ℓ_2 do not intersect at any other point between p and q, and if they are also "close" to each other in some sense, then they form the edges of a smooth polygon $P \subset \Sigma$ with vertices at p and q. Their angles $\alpha, \beta \in [0, 2\pi]$ at p and q must be positive in this situation, because any two geodesics that meet tangentially must be identical. The Gauss-Bonnet formula then gives

$$0 < \alpha + \beta = \int_P K_G \, d\mathrm{vol}_\Sigma,$$

which is a contradiction if the Gaussian curvature of (Σ, g) happens to be everywhere nonpositive. This proves a uniqueness result: if curvature is never positive, then any geodesic from p to q is the *only* geodesic among nearby curves connecting those two points. Our aim is to extend this observation to arbitrary dimensions.

31.1. The second variation of the energy functional. Throughout this lecture, (M, g) is a Riemannian manifold of arbitrary dimension. A more elaborate version of the question mentioned above can be stated as follows:

QUESTION 31.1. *On a Riemannian manifold (M, g), suppose $\gamma : [a, b] \to M$ is a geodesic segment with $\gamma(a) = p$ and $\gamma(b) = q$. Is there any other geodesic segment from p to q near γ? Is γ in fact the shortest path between p and q?*

The issue here is different from what we discussed in §23.2, because we are now allowing the points p and q to be arbitrarily far apart. It is clear from examples that non-short geodesics connecting two points need not be unique: consider for instance the geodesics in $S^2 \subset \mathbb{R}^3$ from the north pole to the south pole, which come in a whole 1-parameter, all equally long. We'll find that this is only allowed because S^2 has positive curvature; it *cannot* happen on a surface with negative or zero curvature.

To better understand the global relationship between geodesics and length, we can apply an infinite-dimensional version of the "second derivative test" to the length functional. As in §22.4, it turns out to be easier for this purpose to work with the energy functional instead of the length functional, because results about the latter can be derived from results about the former, but energy is easier to compute with. Recall the following notation: for two parameter values $a < b \in \mathbb{R}$ and points $p, q \in M$, we denote by

$$\mathcal{P} := C^{\infty}([a, b], M; p, q)$$

the space of smooth paths $\gamma : [a, b] \to M$ starting at p and ending at q. We think of this space intuitively as an infinite-dimensional smooth manifold, with tangent spaces

$$T_{\gamma}\mathcal{P} := \{\eta \in \Gamma(\gamma^*TM) \mid \eta(a) = 0 \text{ and } \eta(b) = 0\}.$$

We then have two functionals $\ell, E : \mathcal{P} \to \mathbb{R}$, the length

$$\ell(\gamma) = \int_a^b |\dot{\gamma}(t)| \, dt,$$

and the energy

$$E(\gamma) = \frac{1}{2} \int_a^b |\dot{\gamma}(t)|^2 \, dt,$$

where an extra factor of $1/2$ has been inserted in front of the energy functional to make some of the expressions below look a bit nicer. For a smooth 1-parameter family of paths $\gamma_s \in \mathcal{P}$ with $\gamma_0 = \gamma$ and $\partial_s \gamma_s|_{s=0} = \eta \in T_{\gamma}\mathcal{P}$, we computed in §22.4 the *first variation* of the energy functional:

$$dE(\gamma)\eta := \frac{d}{ds} E(\gamma_s)\bigg|_{s=0} = \int_a^b \langle -\nabla_t \dot{\gamma}(t), \eta(t) \rangle \, dt.$$

We can express this more succinctly by defining an inner product on the space of sections $\Gamma(\gamma^*TM)$ for each $\gamma \in \mathcal{P}$: for two such sections ξ and η, let[76]

$$\langle \xi, \eta \rangle_{L^2} = \int_a^b \langle \xi(t), \eta(t) \rangle \, dt.$$

Informally, we can think of $\langle\ ,\ \rangle_{L^2}$ as defining a Riemannian metric on $\mathcal{P}$. Now the first variation can be expressed as

$$dE(\gamma)\eta = \langle \nabla E(\gamma), \eta \rangle_{L^2},$$

where

$$\nabla E(\gamma) := -\nabla_t \dot{\gamma} \in \Gamma(\gamma^*TM)$$

is the so-called L^2-*gradient* of the energy functional. In this notation, γ is a geodesic if and only if $\nabla E(\gamma) = 0$.

Informally again, we think of ∇E as a vector field on $\mathcal{P}$ which represents the first derivative of E, though it would be more accurate to call it a section of a vector bundle $\mathcal{E} \to \mathcal{P}$ with fibers $\mathcal{E}_{\gamma} := \Gamma(\gamma^*TM)$, as $\nabla E(\gamma)$ need not take values in the subspace $T_{\gamma}\mathcal{P} \subset \Gamma(\gamma^*TM)$. In any case, we would like to compute the derivative of this section, i.e. it's *linearization* (cf. Exercise 19.7) at a point $\gamma \in \mathcal{P}$ where $\nabla E(\gamma) = 0$, and interpret it as the Hessian of E at the critical point γ. For $\eta \in T_{\gamma}\mathcal{P}$, we choose a 1-parameter family $\gamma_s \in \mathcal{P}$ with $\gamma_0 = \gamma$ and $\partial_s \gamma_s|_{s=0} = \eta$ and define a "covariant derivative" $\nabla_{\eta}\nabla E \in \Gamma(\gamma^*TM)$ by

$$(\nabla_{\eta}\nabla E)(t) := \nabla_s(\nabla E(\gamma_s)(t))|_{s=0}.$$

---

[76]The subscript L^2 refers to the standard notation for the Hilbert space completion of $\Gamma(\gamma^*TM)$ with respect to this inner product.

A quick computation using the definition of the Riemann tensor shows that this does indeed only depend on η rather than the 1-parameter family γ_s:

$$\nabla_s(\nabla E(\gamma_s))|_{s=0} = -\nabla_s\nabla_t\partial_t\gamma_s|_{s=0} = -\nabla_t\nabla_s\partial_t\gamma_s - R(\partial_s\gamma_s, \partial_t\gamma_s)\partial_t\gamma_s|_{s=0} = -\nabla_t^2\eta - R(\eta,\dot\gamma)\dot\gamma.$$

With this calculation as motivation, define for any $\gamma \in \mathcal{P}$ a linear operator

$$(31.1) \qquad\qquad \nabla^2 E(\gamma) : \Gamma(\gamma^*TM) \to \Gamma(\gamma^*TM) : \eta \mapsto -\nabla_t^2\eta - R(\eta,\dot\gamma)\dot\gamma.$$

We can now state the *second variation formula*:

PROPOSITION 31.2. *Suppose $\gamma \in \mathcal{P}$ is a geodesic and $\gamma_{\sigma,\tau} \in \mathcal{P}$ is a smooth 2-parameter family of paths with $\gamma_{0,0} = \gamma$, with variations $\xi, \eta \in T_\gamma\mathcal{P}$ defined by*

$$\xi = \partial_\sigma\gamma_{\sigma,\tau}|_{\sigma=\tau=0} \qquad and \qquad \eta = \partial_\tau\gamma_{\sigma,\tau}|_{\sigma=\tau=0}.$$

Then

$$\frac{\partial^2}{\partial\sigma\partial\tau}E(\gamma_{\sigma\tau})\bigg|_{\sigma=\tau=0} = \langle\nabla^2 E(\gamma)\xi, \eta\rangle_{L^2}.$$

PROOF. Compute:

$$\begin{aligned}
\frac{\partial^2}{\partial\sigma\partial\tau}E(\gamma_{\sigma,\tau})\bigg|_{\sigma=\tau=0} &= \frac{\partial}{\partial\sigma}\left(\frac{\partial}{\partial\tau}E(\gamma_{\sigma,\tau})\bigg|_{\tau=0}\right)\bigg|_{\sigma=0} = \frac{\partial}{\partial\sigma}\langle\nabla E(\gamma_{\sigma,0}), \partial_\tau\gamma_{\sigma,\tau}|_{\tau=0}\rangle_{L^2}\bigg|_{\sigma=0} \\
&= \frac{\partial}{\partial\sigma}\int_a^b \langle\nabla E(\gamma_{\sigma,0})(t), \partial_\tau\gamma_{\sigma,\tau}(t)|_{\tau=0}\rangle\, dt\bigg|_{\sigma=0} \\
&= \int_a^b \langle\nabla_\sigma\nabla E(\gamma_{\sigma,0})(t)|_{\sigma=0}, \eta(t)\rangle\, dt + \int_a^b \langle\nabla E(\gamma)(t), \nabla_\sigma\partial_\tau\gamma_{\sigma,\tau}|_{\sigma=\tau=0}\rangle\, dt \\
&= \int_a^b \langle(\nabla_\xi\nabla E)(t), \eta(t)\rangle\, dt = \langle\nabla^2 E(\gamma)\xi, \eta\rangle_{L^2}.
\end{aligned}$$

Note that we used the assumption that γ is geodesic, so $\nabla E(\gamma) = 0$. □

For a 1-parameter family of paths $\{\gamma_s \in \mathcal{P}\}_{s\in(-\epsilon,\epsilon)}$, one can plug $\gamma_{\sigma,\tau} := \gamma_{\sigma+\tau}$ into the second variation formula and extract from it the first nontrivial term in the Taylor expansion of $E(\gamma_s)$ as a function of s: we have

$$\begin{aligned}
E(\gamma_s) &= E(\gamma) + \frac{1}{2}s^2 \cdot \langle\eta, \nabla^2 E(\gamma)\eta\rangle_{L^2} + O(|s|^3) \\
&= E(\gamma) - \frac{1}{2}s^2 \cdot \langle\eta, \nabla_t^2\eta + R(\eta,\dot\gamma)\dot\gamma\rangle_{L^2} + O(|s|^3)
\end{aligned}$$

If we want a criterion to guarantee that $E(\gamma_s)$ is locally minimized at $s = 0$, the first important thing to understand is whether the coefficient on the quadratic term is positive. This coefficient breaks up into two terms, and the sign of the first one can easily be ascertained after integrating by parts: since $\partial_t\langle\eta, \nabla_t\eta\rangle = |\nabla_t\eta|^2 + \langle\eta, \nabla_t^2\eta\rangle$ and η vanishes at the end points, we have

$$-\langle\eta, \nabla_t^2\eta\rangle_{L^2} = \langle\nabla_t\eta, \nabla_t\eta\rangle_{L^2} =: \|\nabla_t\eta\|_{L^2}^2 \geqslant 0,$$

with strict inequality unless η is parallel along γ, which would mean $\eta \equiv 0$ since $\eta(a) = 0$. The other term is

$$(31.2) \qquad\qquad -\langle\eta, R(\eta,\dot\gamma)\dot\gamma\rangle_{L^2} = -\int_a^b \mathrm{Riem}(\eta(t), \eta(t), \dot\gamma(t), \dot\gamma(t))\, dt,$$

and evidently the question of whether this must be positive or not depends in some way on the curvature. This question merits a more thorough discussion.

31.2. Sectional curvature. At any given point, the vectors $\dot{\gamma}(t)$ and $\eta(t)$ appearing in the integrand of (31.2) can be completely arbitrary, so the real question here is whether any meaningful condition can be formulated that would determine the signs of real numbers of the form

$$\langle X, R(X,Y)Y \rangle = \text{Riem}(X,X,Y,Y)$$

for arbitrary tangent vectors $X, Y \in T_pM$ at a point $p \in M$. We've seen products like this before: in the case $\dim M = 2$, if X and Y are taken to be a basis of T_pM, we saw in (27.7) that $\text{Riem}(X,X,Y,Y)$ determines the Gaussian curvature by the formula

$$(31.3) \qquad K_G(p) = \frac{\text{Riem}(X,X,Y,Y)}{|d\text{vol}_M(X,Y)|^2} = \frac{\text{Riem}(X,X,Y,Y)}{\langle X,X \rangle \langle Y,Y \rangle - \langle X,Y \rangle^2},$$

where in the second expression we have used Exercise 11.12 to write

$$d\text{vol}_M(X,Y) = \sqrt{\det \begin{pmatrix} \langle X,X \rangle & \langle X,Y \rangle \\ \langle Y,X \rangle & \langle Y,Y \rangle \end{pmatrix}}.$$

The fact that this formula remains equally valid for any choice of the basis $X, Y \in T_pM$ is both non-obvious and useful, and the following definition will help us generalize it to higher dimensions.

DEFINITION 31.3. Suppose (M,g) is a Riemannian manifold and $P \subset T_pM$ is a 2-dimensional subspace in the tangent space at some point $p \in M$. The **sectional curvature** $K_S(P) \in \mathbb{R}$ along P is defined as follows. Choose a sufficiently small neighborhood $0 \in \mathcal{O}_p \subset T_pM$ so that $\exp_p$ restricts to a diffeomorphism from $\mathcal{O}_p$ to a neighborhood of p in M. Then

$$\Sigma_P := \exp_p(\mathcal{O}_p \cap P) \subset M$$

is a 2-dimensional submanifold containing p, and we set

$$K_S(P) := K_G(p),$$

where $K_G : \Sigma_P \to \mathbb{R}$ is the Gaussian curvature of Σ_P with respect to the Riemannian metric induced by its embedding in (M,g).

LEMMA 31.4. *For the embedded surface $\Sigma_P \subset M$ through $p \in M$ in Definition 31.3, the second fundamental form $\text{II} : T\Sigma_P \oplus T\Sigma_P \to (T\Sigma_P)^\perp$ vanishes at p.*

PROOF. Given any $X, Y \in T_p\Sigma_P$ and constants $a, b \in \mathbb{R}$, the geodesic $\gamma(t) = \exp_p(t(aX + bY))$ lies in Σ_P for t close to 0, thus $aX + bY$ extends to a parallel vector field on Σ_P along this curve, giving

$$\text{II}(aX + bY, aX + bY) = (\nabla_{aX+bY}(aX+bY))^\perp = 0$$

at $p = \gamma(0)$. Since $a, b \in \mathbb{R}$ were arbitrary, this implies $\text{II}(X,X) = \text{II}(Y,Y) = 0$, and $\text{II}(X,Y) = 0$ then follows from $\text{II}(X+Y, X+Y) = 0$ using bilinearity and symmetry. $\square$

REMARK 31.5. If you remember the construction of Riemann normal coordinates in §23.1, you may have noticed that we used a trick from that construction in Lemma 31.4. We could alternatively have used normal coordinates to deduce the lemma: if we pick any orthonormal vectors $X_1, X_2 \in T_p\Sigma_P$ and extend them to an orthonormal basis $X_1, \ldots, X_n$ of T_pM, we obtain a normal coordinate system $(x^1, \ldots, x^n)$ near p in which Σ_P looks like the "flat" plane $\{x^3 = \ldots = x^n = 0\}$ and X_1 and X_2 match the coordinate vector fields ∂_1, ∂_2. The vanishing of II at p then follows from the fact that in normal coordinates centered at p, the Christoffel symbols (which are equivalent to the covariant derivatives $\nabla_j(\partial_k)$) vanish at p.

PROPOSITION 31.6. *On a Riemannian manifold* (M, g), *the relation*

$$\mathrm{Riem}(X, X, Y, Y) = K_S(P) \cdot \mathrm{Area}(X, Y)^2 = K_S(P) \cdot \left(\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2 \right)$$

holds for any 2-dimensional subspace $P \subset T_p M$ *and any* $X, Y \in P$, *where* $\mathrm{Area}(X, Y) \geqslant 0$ *denotes the area of the parallelogram in* $T_p M$ *spanned by* X *and* Y, *as measured with respect to the metric* $\langle \ , \ \rangle$. *In particular, whenever* X, Y *is a basis of* P, *this gives rise to the formula*

$$K_S(P) = \frac{\mathrm{Riem}(X, X, Y, Y)}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}$$

for the sectional curvature of (M, g) *along* P.

PROOF. Both sides of the relation vanish if X and Y are linearly dependent: for the left hand side this follows from the symmetries of the Riemann tensor, and on the right hand side it follows because the parallelogram spanned by X and Y has no area. We are thus free to assume X and Y form a basis of $P = T_p \Sigma_P$. Lemma 31.4 and the Gauss equation (Prop. 28.5) now imply that the Riemann tensor of the submanifold Σ_P at p is just the restriction of the Riemann tensor $R \in \Gamma(T_3^1 M)$ of (M, g) to the subspace $T_p \Sigma_P \subset T_p M$. The result thus follows from the definition of sectional curvature and the formula (31.3) for the Gaussian curvature. $\square$

Just as the Gaussian curvature determines the Riemann tensor in dimension 2, one can show that the Riemann tensor is determined in general by the sectional curvature. The proof of this requires a more thorough discussion of the symmetries of the Riemann tensor than we have given so far, thus we will save it for next semester, and merely record it here as a fact:

PROPOSITION 31.7. *The Riemann tensor* $R \in \Gamma(T_3^1 M)$ *at a point* $p \in M$ *is determined by the values of the sectional curvature* $K_S(P) \in \mathbb{R}$ *on all possible 2-dimensional subspaces* $P \subset T_p M$. *In particular, a Riemannian manifold is locally flat if and only if its sectional curvature vanishes identically.*

DEFINITION 31.8. A Riemannian manifold (M, g) is said to have **positive** (or **negative** or **zero**) **curvature** if $K_S(P)$ is positive (or negative or zero respectively) for every 2-dimensional subspace $P \subset T_p M$ at every point $p \in M$.

In §24.4, we discussed three emblematic examples of Riemannian n-manifolds that can be defined for each $n \geqslant 2$: Euclidean space $\mathbb{R}^n$, the sphere S^n and hyperbolic space H^n. The following exercises show that these are manifolds of constant sectional curvature, vanishing in the case of $\mathbb{R}^n$, positive for S^n and negative for H^n.

EXERCISE 31.9. Suppose (M, g) is a Riemannian manifold, $p, q \in M$ are points and $P \subset T_p M$ and $Q \subset T_q M$ are 2-dimensional subspaces such that there exists an isometry $\varphi \in \mathrm{Isom}(M, g)$ with $\varphi(p) = q$ and $\varphi_* P = Q$. Show that the sectional curvature of (M, g) satisfies $K_S(P) = K_S(Q)$.

EXERCISE 31.10. Show that H^n contains a point $p \in H^n$ with a 2-dimensional subspace $P \subset T_p H^n$ for which the surface $\Sigma_P \subset H^n$ appearing in the definition of sectional curvature is isometric to the hyperbolic plane H^2. Deduce from this and Exercise 31.9 that H^n has sectional curvature $K_S(P) = -1$ for all 2-dimensional subspaces $P \subset T H^n$.
Hint: What kind of submanifold is the intersection of $H^n \subset \mathbb{R}^{n+1}$ *with the subspace* $\mathbb{R}^3 \times \{0\} \subset \mathbb{R}^{n+1}$? *Recall from* §24.4 *that the geodesics in* H^n *can be written down quite explicitly.*

EXERCISE 31.11. Use a similar trick to Exercise 31.10 to prove that the sphere $S^n \subset \mathbb{R}^{n+1}$ has constant sectional curvature $+1$, and (this one is more obvious) Euclidean space $\mathbb{R}^n$ has vanishing sectionsl curvature.

An important result we will prove next semester states that up to rescaling by positive constants, these are the *only* geodesically complete and *simply connected* Riemannian n-manifolds with constant sectional curvature. This statement is no longer true if one removes the words "simply connected": there are many interesting examples of manifolds with constant sectional curvature, also called *Riemannian space forms*. But if you know a bit of covering space theory from algebraic topology, you will recognize that the specific examples S^n, $\mathbb{R}^n$ and H^n always remain relevant, because they appear as universal covers of space forms, implying that every Riemannian space form is isometric to a quotient of one of these three examples by a discrete group acting by isometries.

EXERCISE 31.12. The copy of H^2 embedded isometrically in H^n in Exercise 31.10 is an example of a totally geodesic submanifold: we say that a submanifold $N \subset M$ in a Riemannian manifold (M, g) is **totally geodesic** if all geodesics in N (with its induced metric) are also geodesics in (M, g). One nice way to recognize totally geodesic submanifolds is via isometries: suppose $N \subset M$ is a submanifold that is also the fixed-point set $N = \text{Fix}(\varphi) := \{p \in M \mid \varphi(p) = p\}$ of an isometry $\varphi \in \text{Isom}(M, g)$. Show that N is then totally geodesic.

EXERCISE 31.13. Find an example of an isometry on the hyperbolic n-space H^n whose fixed point set is a submanifold isometric to H^2.

31.3. Nonpositive curvature and geodesics. The notion of sectional curvature now provides a simply-stated condition that is sufficient to guarantee that geodesics are local minima of the length functional. We recall from §31.1 the notation $\mathcal{P} := C^\infty([a, b], M; p, q)$ and $T_\gamma \mathcal{P} := \{\eta \in \Gamma(\gamma^* TM) \mid \eta(a) = 0 \text{ and } \eta(b) = 0\}$.

LEMMA 31.14. *Suppose (M, g) is a Riemannian manifold with everywhere nonpositive sectional curvature $K_S \leqslant 0$, and $\gamma \in \mathcal{P}$ is a nonconstant geodesic. Then the second variation operator $\nabla^2 E(\gamma) : \Gamma(\gamma^* TM) \to \Gamma(\gamma^* TM)$ defined in (31.1) satisfies*

$$\langle \nabla^2 E(\gamma)\eta, \eta \rangle_{L^2} \geqslant 0$$

for all $\eta \in \Gamma(\gamma^* TM)$, and the inequality is strict for all nontrivial $\eta \in T_\gamma \mathcal{P}$.

PROOF. We have already observed that $\langle -\nabla_t^2 \eta, \eta \rangle_{L^2} = \|\nabla_t \eta\|_{L^2}^2 \geqslant 0$, with strict inequality whenever η is a nontrivial section in $T_\gamma \mathcal{P}$. Using Proposition 31.6, the other term in $\langle \nabla^2 E(\gamma)\eta, \eta \rangle_{L^2}$ is the integral from a to b of

$$-\text{Riem}(\eta(t), \eta(t), \dot{\gamma}(t), \dot{\gamma}(t)) = -K_S(P_t) \cdot \text{Area}(\eta(t), \dot{\gamma}(t))^2 \geqslant 0,$$

where $P_t \subset T_{\gamma(t)} M$ can be taken to be any 2-dimensional subspace containing $\eta(t)$ and $\dot{\gamma}(t)$. $\quad\square$

THEOREM 31.15. *Suppose (M, g) is a Riemannian manifold with nonpositive sectional curvature and $\gamma : [a, b] \to M$ is a geodesic connecting $\gamma(a) = p$ to $\gamma(b) = q$. Then for any smooth 1-parameter family of paths $\gamma_s : [a, b] \to M$ with $\gamma_s(a) = p$, $\gamma_s(b) = q$ and $\gamma_0 \equiv \gamma$ such that $\partial_s \gamma_s|_{s=0}$ is not everywhere tangent to $\dot{\gamma}$, there is a number $\epsilon > 0$ such that:*
  (1) γ is the only geodesic among the paths γ_s for $s \in (-\epsilon, \epsilon)$;
  (2) For all paths γ_s with $s \in (-\epsilon, \epsilon)$ and $s \neq 0$,

$$\ell(\gamma_s) > \ell(\gamma).$$

PROOF. The result is already clear if $p = q$ and γ is a constant path, so let us assume γ is nonconstant, in which case $\dot{\gamma}$ is everywhere nonzero since it is a geodesic.

The second statement is easily proved if we replace length with energy, because by the second variation formula and Lemma 31.14,

$$\frac{d^2}{ds^2} E(\gamma_s)\bigg|_{s=0} = \langle \nabla^2 E(\gamma)\eta, \eta \rangle_{L^2} > 0 \qquad \text{for} \qquad \eta := \partial_s \gamma_s|_{s=0} \neq 0.$$

In order to apply this result to the length functional, we can use the same trick as in the proof of Corollary 22.11 and reparametrize each of the paths $\gamma_s : [a, b] \to M$ so that they have constant speed; here we can restrict s if necessary to a neighborhood of 0 so that $\dot{\gamma}_s(t)$ can always be assumed nonzero, and after reparametrizing, its norm is a positive number indepenent of t, namely

$$|\dot{\gamma}_s(t)| =: v_s = \frac{\ell(\gamma_s)}{b - a}.$$

The assumption that $\eta = \partial_s \gamma_s|_{s=0}$ is not everywhere tangent to $\dot{\gamma}$ implies that after this reparametrization, η is still not everywhere zero. Since γ_0 is a geodesic, we know from Corollary 22.11 that $\frac{d}{ds}\ell(\gamma_s)\big|_{s=0} = 0$ and thus $\partial_s v_s|_{s=0} = 0$, so

$$\frac{d^2}{ds^2}\ell(\gamma_s)\bigg|_{s=0} = \frac{d}{ds}\int_a^b \partial_s \sqrt{|\dot{\gamma}_s(t)|^2}\, dt\bigg|_{s=0} = \frac{d}{ds}\int_a^b \frac{1}{2\sqrt{|\dot{\gamma}_s(t)|^2}}\partial_s|\dot{\gamma}_s(t)|^2\, dt\bigg|_{s=0}$$

$$= \int_a^b \frac{\partial}{\partial s}\left(\frac{1}{2v_s}\partial_s|\dot{\gamma}_s(t)|^2\right)\bigg|_{s=0} dt = \frac{1}{2v_0}\int_a^b \partial_s^2|\dot{\gamma}_s(t)|^2\big|_{s=0}\, dt$$

$$= \frac{1}{v_0}\frac{d^2}{ds^2}\frac{1}{2}\left(\int_a^b |\dot{\gamma}_s(t)|^2\, dt\right)\bigg|_{s=0} = \frac{1}{v_0}\frac{d^2}{ds^2}E(\gamma_s)\bigg|_{s=0} > 0.$$

This proves that $\ell(\gamma_s) > \ell(\gamma)$ for $s \neq 0$ close to 0.

To see that γ is the only geodesic among the family γ_s for s close to 0, we can differentiate the L^2-inner product of $\nabla E(\gamma_s) = \nabla_t \dot{\gamma}_s \in \Gamma(\gamma_s^* TM)$ with $\eta_s := \partial_s \gamma_s \in \Gamma(\gamma_s^* TM)$ at $s = 0$, using the fact that $\nabla E(\gamma) = 0$:

$$\frac{d}{ds}\langle \nabla E(\gamma_s), \eta_s\rangle_{L^2}\bigg|_{s=0} = \langle \nabla^2 E(\gamma)\eta, \eta\rangle_{L^2} + \langle \nabla E(\gamma), \nabla_s \eta_s|_{s=0}\rangle_{L^2} = \langle \nabla^2 E(\gamma)\eta, \eta\rangle_{L^2} > 0.$$

It follows that $\langle \nabla E(\gamma_s), \eta_s\rangle \neq 0$ for sufficiently small $|s| \neq 0$, implying that $\nabla E(\gamma_s)$ itself cannot be 0, so γ_s is not a geodesic. $\qquad\square$

These results give a small hint of the larger story of geodesics on manifolds with nonpositive curvature. The general pattern is that nonpositive curvature implies uniqueness phenomena that clearly do not hold in simple examples with positive curvature, such as the sphere. Here is another example, which we will prove next semester as a consequence of the Hopf-Rinow and Cartan-Hadamard theorems:

THEOREM. *Suppose (M, g) is a connected and geodesically complete Riemannian manifold with nonpositive sectional curvature, and $p, q \in M$ are two points. Then every homotopy class (with fixed end points) of paths from p to q contains exactly one geodesic segment, up to parametrization.*

The intuition here, which you will find reasonable if you compare what you know about the 2-sphere and the hyperbolic plane, is that positive curvature typically causes two geodesics emerging from the same point to come back together at a later time, whereas negative curvature tries always to force them further apart. On a simply connected manifold, the latter guarantees that they never meet again, producing an absolutely *unique* shortest path between p and q. On surfaces, a slightly weaker version of the uniqueness statement in this theorem can be derived from the Gauss-Bonnet formula:

EXERCISE 31.16. Assume (Σ, g) is a Riemannian 2-manifold with $K_G \leqslant 0$.

(a) Show that (Σ, g) does not admit any *periodic* geodesic (i.e. a geodesic $\gamma : \mathbb{R} \to \Sigma$ satisfying $\gamma(t + T) = \gamma(t)$ for all $t \in \mathbb{R}$ and some fixed $T > 0$) whose image bounds an embedded disk.

(b) Given a pair of distinct points $p, q \in \Sigma$ and a pair of geodesic segments $\gamma_0, \gamma_1 : [0, 1] \to \Sigma$ with $\gamma_0(0) = \gamma_1(0) = p$ and $\gamma_0(1) = \gamma_1(1) = q$, show that there does not exist any smooth family of paths $\{\gamma_s : [0, 1] \to \Sigma\}_{s \in [0,1]}$ from p to q, matching the given geodesics for $s = 0, 1$, such that the map $[0, 1] \times (0, 1) \to \Sigma : (s, t) \mapsto \gamma_s(t)$ is an embedding.

(c) Find an example of a periodic geodesic on a surface with nonpositive Gaussian curvature. (*Note that by part (a), it had better not form the boundary of an embedded disk.*)

(d) Show that the phenomenon ruled out by part (b) can actually happen on S^2.

31.4. Addendum: Gauss and sectional curvature with indefinite metrics. I have thus far completely excluded metrics of indefinite signature in this lecture, and also in the discussion of Gaussian curvature in §27.3. For sectional curvature on a general pseudo-Riemannian manifold (M, g), the standard procedure is to take the formula in Proposition 31.6 as a definition, namely

$$(31.4) \qquad\qquad K_S(P) = \frac{\mathrm{Riem}(X, X, Y, Y)}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2},$$

assuming X, Y to be any basis of the 2-dimensional subspace $P \subset T_p M$. There are two very large caveats here: first, the determinant in the denominator is nonzero if and only if the restriction of the metric g to the subspace $P \subset T_p M$ is nondegenerate, a condition that came for free when g was positive, but in the indefinite case it imposes a restriction on the set of subspaces $P \subset T_p M$ along which K_S can be defined. This should not be surprising however, as it is another symptom of the fact that arbitrary submanifolds are not always *pseudo-Riemannian* submanifolds, and if we want to interpret $K_S(P)$ as the Gaussian curvature at p of a submanifold $\Sigma_P \subset M$ tangent to P, we certainly want g to be nondegenerate on that submanifold. The second caveat is that it is not obvious at this stage why the right hand side of (31.4) should be independent of the choice of basis $X, Y \in P$. If $g|_P$ is positive, then $\Sigma_P \subset M$ is a Riemannian submanifold and we can again recognize (31.4) as a formula for the Gaussian curvature of Σ_P, which proves independence of the choices. If $g|_P$ is nondegenerate but not positive, then a similar argument will work, but we must first discuss how to define K_G on a surface with an indefinite metric.

For a pseudo-Riemannian surface (Σ, g), g is either positive or negative or has signature $(1, 1)$. The negative case is essentially the same as the positive case: it means simply that $(\Sigma, -g)$ is a Riemannian manifold, and since the Levi-Cività connections and volume forms with respect to g and $-g$ are the same, their Riemann tensors and Gaussian curvatures are also the same.

The more interesting case is where (Σ, g) is a *Lorentzian* 2-manifold, with signature $(1, 1)$. The formula $R(X, Y)Z = -K_G \, d\mathrm{vol}(X, Y)JZ$ in the Riemannian case was based on the fact that at each point $p \in \Sigma$, the space of antisymmetric bilinear forms on $T_p \Sigma$ is 1-dimensional, and so is the space of antisymmetric linear maps $T_p \Sigma \to T_p \Sigma$, for which we chose the 90-degree counterclockwise fiberwise rotation $J : T\Sigma \to T\Sigma$ as a canonical basis at each point. If we instead have an indefinite metric $\langle \ , \ \rangle$ on $T\Sigma$, then antisymmetry of a map $T_p \Sigma \to T_p \Sigma$ with respect to this metric means something qualitatively different, and we will have to choose a new generator $J : T\Sigma \to T\Sigma$. In the Riemannian case, our generator was characterized by the property that for any unit vector $X \in T_p \Sigma$, (X, JX) forms a positively-oriented orthonormal basis. This turns out to be a reasonable condition to generalize.

LEMMA 31.17. *Assume V is an oriented 2-dimensional vector space with a nondegenerate symmetric bilinear form $\langle \ , \ \rangle$ of signature $(1, 1)$. Then there exists a unique linear map $J : V \to V$ with the property that for every positively-oriented basis (v, w) of V that is orthonormal with respect to $\langle \ , \ \rangle$,*[77]

$$Jv = w \qquad and \qquad Jw = v.$$

---

[77] Note that when we talk about an orthonormal basis (v, w) of V with respect to an indefinite inner product of signature $(1, 1)$, the order matters: our convention is that $\langle v, v \rangle = 1$ and $\langle w, w \rangle = -1$.

Moreover, J is antisymmetric with respect to $\langle \ , \ \rangle$, and also satisfies $J^2 = \mathbb{1}$ and $\langle Jv, Jw \rangle = -\langle v, w \rangle$ for all $v, w \in V$.

PROOF. Choosing a positively-oriented orthonormal basis of V allows us to assume without loss of generality that $V = \mathbb{R}^2$ with the standard Minkowski inner product $\langle v, w \rangle = v^T \eta w$ determined by the matrix $\eta := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. The standard basis is then positively oriented and orthonormal, so if a transformation $J : \mathbb{R}^2 \to \mathbb{R}^2$ with the stated properties exists, then it is clearly unique: it is given by the matrix

$$(31.5) \qquad\qquad \mathbf{J}_0 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which reflects $\mathbb{R}^2$ orthogonally across the diagonal subspace $\{(x, x)\} \subset \mathbb{R}^2$. To see that this does the trick, observe that every positively-oriented orthonormal basis of $(\mathbb{R}^2, \langle \ , \ \rangle)$ is of the form (v, w) with

$$v = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad w = \begin{pmatrix} y \\ x \end{pmatrix} \quad \text{such that} \quad x^2 - y^2 = 1,$$

so $J := \mathbf{J}_0$ does indeed send $v \mapsto w$ and $w \mapsto v$. This proves that $J : V \to V$ is defined independently of the choice of positively-oriented orthonormal basis on V. The relations $J^2 = \mathbb{1}$ and $\langle Jv, Jw \rangle = -\langle v, w \rangle$ are now quick computations, and antisymmetry follows: $\langle v, Jw \rangle = -\langle Jv, J^2 w \rangle = -\langle Jv, w \rangle$. $\qquad\square$

REMARK 31.18. The fact that the reflection $J : V \to V$ in Lemma 31.17 is represented by the same matrix in any positive orthonormal basis reveals that the subspaces $\ell_\pm \subset V$ spanned by the vectors represented by $(1, \pm 1)$ in coordinates do not actually depend on the chosen coordinates. Their invariant description is as follows: $\ell_+ \cup \ell_-$ is the set of all *light-like* vectors $v \in V$, i.e. those which satisfy $\langle v, v \rangle = 0$. One can use the orientation on V to distinguish between ℓ_+ and ℓ_-, and then define J as the reflection along ℓ_- about ℓ_+.

REMARK 31.19. On $\mathbb{R}^2$ with the Minkowski inner product, the group of orientation-preserving isometries is the 1-dimensional abelian group

$$\text{SO}(1, 1) := \text{O}(1, 1) \cap \text{SL}(2, \mathbb{R}) = \left\{ \pm e^{\theta \mathbf{J}_0} = \pm \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} \;\middle|\; \theta \in \mathbb{R} \right\} \subset \text{GL}(2, \mathbb{R}),$$

where $\mathbf{J}_0$ is the matrix defined in (31.5). From this perspective, $\mathbf{J}_0$ can be regarded as a canonical generator of the Lie algebra $\mathfrak{so}(1, 1)$, and the fact that $\text{SO}(1, 1)$ is abelian is what makes it possible to define $J : V \to V$ in Lemma 31.17 via this matrix without it depending on the choice of positively-oriented orthonormal basis.

By Lemma 31.17, we can define on our oriented Lorentzian surface (Σ, g) a canonical antisymmetric bundle map

$$J : T\Sigma \to T\Sigma$$

such that whenever (X_1, X_2) is a positively-oriented orthonormal basis of some tangent space $T_p\Sigma$,

$$X_2 = JX_1 \qquad \text{and} \qquad X_1 = JX_2.$$

Every antisymmetric bundle map $T\Sigma \to T\Sigma$ is then of the form fJ for some function $f : \Sigma \to \mathbb{R}$, so the symmetries of the Riemann tensor imply

$$(31.6) \qquad\qquad R(X, Y)Z = -K_G \, d\text{vol}(X, Y)JZ$$

for a uniquely determined function $K_G : \Sigma \to \mathbb{R}$, which will be known henceforth as the Gaussian curvature of (Σ, g). We can compute K_G from the Riemann tensor in much the same way as in

the Riemannian case: since J is antisymmetric, one checks by plugging in a positively-oriented orthonormal basis that

$$d\mathrm{vol}(X, Y) = \langle X, JY \rangle$$

for all $X, Y \in T_p\Sigma$, thus

$$\mathrm{Riem}(X, X, Y, Y) = \langle X, R(X, Y)Y \rangle = -K_G(p) \cdot \left| d\mathrm{vol}(X, Y)^2 \right|^2$$

and if X and Y are chosen to be linearly independent, we can write

$$(31.7) \qquad K_G(p) = -\frac{\mathrm{Riem}(X, X, Y, Y)}{|d\mathrm{vol}(X, Y)|^2}.$$

If you look at (27.7), you'll notice that this formula has an extra minus sign compared with the Riemannian case, but this is more sensible than you might think. Recall from Exercise 18.30 how the canonical volume form is computed in the indefinite case: when $\langle \, , \, \rangle$ has signature $(1, 1)$, the symmetric matrix

$$\mathbf{g} := \begin{pmatrix} \langle X, X \rangle & \langle X, Y \rangle \\ \langle Y, X \rangle & \langle Y, Y \rangle \end{pmatrix}$$

has one positive and one negative eigenvalue, so its determinant is *negative*, implying

$$d\mathrm{vol}(X, Y) = \sqrt{-\det \mathbf{g}} = \sqrt{-\langle X, X \rangle \langle Y, Y \rangle + \langle X, Y \rangle^2}.$$

The extra minus sign turns (31.7) into

$$(31.8) \qquad K_G(p) = \frac{\mathrm{Riem}(X, X, Y, Y)}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2},$$

a formula that is equally valid in the Riemannian and Lorentzian cases. In local coordinates, (27.8) now generalizes in the form

$$K_G = \frac{R_{1122}}{g_{11}g_{22} - (g_{12})^2}.$$

With this definition of K_G in place, our original definition of the sectional curvature $K_S(P)$ also makes sense in pseudo-Riemannian manifolds: it is the Gaussian curvature at p of a particular 2-dimensional pseudo-Riemannian submanifold $\Sigma_P \subset M$ tangent to P at p, and (31.4) gives a correct formula for computing it, due to (31.8).

In order for K_G to be truly useful, we'd like to be able to generalize the results of §27.3 and have a convenient way of computing it for surfaces embedded as pseudo-Riemannian submanifolds in $\mathbb{R}^3$. In order for $\Sigma \subset \mathbb{R}^3$ to have signature $(1, 1)$, we need to assume $(\mathbb{R}^3, g)$ has signature either $(2, 1)$ or $(1, 2)$, so let's take the metric to be plus or minus the Minkowski metric,

$$g := \pm g_M, \qquad g_M = -dx^2 + dy^2 + dz^2.$$

If we take the plus sign in this definition, then $(\mathbb{R}^3, g)$ has signature $(2, 1)$, and g is thus positive on the normal bundle $T\Sigma^\perp \subset T\mathbb{R}^3|_\Sigma$ of any pseudo-Riemannian surface $\Sigma \subset \mathbb{R}^3$ with signature $(1, 1)$. Taking the minus sign makes the signature of $(\mathbb{R}^3, g)$ into $(1, 2)$, so g is then negative on $T\Sigma^\perp$. A unit normal vector field $\nu \in \Gamma(T\Sigma^\perp)$ can thus be said to satisfy

$$\langle \nu, \nu \rangle = \pm 1,$$

so that it takes values in the connected hyperboloid

$$H_L^2 := \left\{ X \in \mathbb{R}^3 \mid \langle X, X \rangle = \pm 1 \right\} = \{ g_M(X, X) = 1 \} = \left\{ y^2 + z^2 - x^2 = 1 \right\} \subset \mathbb{R}^3,$$

defining a Gauss map

$$\nu : \Sigma \to H_L^2,$$

whose derivative at any point $p \in \Sigma$ again defines a linear self-map of $T_p\Sigma$:

$$T_p\nu : T_p\Sigma \to T_{\nu(p)}H_L^2 = \nu(p)^\perp = T_p\Sigma.$$

PROPOSITION 31.20. *For a 2-dimensional pseudo-Riemannian submanifold Σ in $(\mathbb{R}^3, \pm g_M)$, the Gaussian curvature $K_G : \Sigma \to \mathbb{R}$ is related to the Gauss map $\nu : \Sigma \to \mathbb{R}$ by*

$$K_G(p) = \pm \det(T_p\nu).$$

PROOF. The arguments of §28.1 apply equally well to this situation up to and including Equation 28.3. The rest of the proof requires modification, because while $\nabla\nu(p) : T_p\Sigma \to T_p\Sigma$ is self-adjoint with respect to the bundle metric $\langle \, , \, \rangle$ on $T\Sigma$, that does not generally imply that it is diagonalizable if the metric is indefinite. The main danger here is that there could be a light-like eigenspace, which is its own orthogonal complement and thus does not imply the existence of any complementary eigenspace. (If you try writing down simple examples of symmetric operators on Minkowski $\mathbb{R}^2$, you will find that this really can happen.)

As luck would have it, the orthonormal basis of eigenvectors we used for proving Theorem 27.17 in §28.1 was convenient, but not truly necessary. We could have argued more generally as follows. Choose any positively-oriented orthonormal basis (X_1, X_2) of $T_p\Sigma$; recall that in the Lorentzian case, this means

(31.9) $\langle X_1, X_1 \rangle = 1, \quad \langle X_2, X_2 \rangle = -1, \quad \langle X_1, X_2 \rangle = 0 \qquad$ and $\qquad d\text{vol}(X_1, X_2) = 1,$

and the defining property of the reflection $J : T\Sigma \to T\Sigma$ then implies

(31.10) $X_2 = JX_1, \qquad$ and $\qquad X_1 = JX_2.$

In this basis, the transformation $\nabla\nu(p) : T_p\Sigma \to T_p\Sigma$ is represented by a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so applying the Gauss equation as in §28.1 gives

$$\langle V, \widehat{R}(X_1, X_2)Z \rangle = \langle \text{II}(V, X_1), \text{II}(X_2, Z) \rangle - \langle \text{II}(V, X_2), \text{II}(X_1, Z) \rangle$$
$$= \big\langle \langle V, \nabla\nu(p)X_1 \rangle \nu(p), \langle Z, \nabla\nu(p)X_2 \rangle \nu(p) \big\rangle - \big\langle \langle V, \nabla\nu(p)X_2 \rangle \nu(p), \langle Z, \nabla\nu(p)X_1 \rangle \nu(p) \big\rangle$$
$$= \pm \left(\langle V, \nabla\nu(p)X_1 \rangle \langle Z, \nabla\nu(p)X_2 \rangle - \langle V, \nabla\nu(p)X_2 \rangle \langle Z, \nabla\nu(p)X_1 \rangle \right)$$
$$= \pm \left(\langle V, aX_1 + cX_2 \rangle \langle Z, bX_1 + dX_2 \rangle - \langle V, bX_1 + dX_2 \rangle \langle Z, aX_1 + cX_2 \rangle \right)$$
$$= \langle V, \pm \left(\langle Z, bX_1 + dX_2 \rangle (aX_1 + cX_2) - \langle Z, aX_1 + cX_2 \rangle (bX_1 + dX_2) \right) \rangle,$$

implying

$$\widehat{R}(X_1, X_2)Z = \pm \langle Z, bX_1 + dX_2 \rangle (aX_1 + cX_2) \mp \langle Z, aX_1 + cX_2 \rangle (bX_1 + dX_2)$$
$$= \pm (ad - bc) \left(\langle Z, X_2 \rangle X_1 - \langle Z, X_1 \rangle X_2 \right).$$

To simplify the last expression in parentheses, recall that J preserves $\langle \, , \, \rangle$ with an extra sign, so using (31.9) and (31.10),

$$\langle Z, X_2 \rangle X_1 - \langle Z, X_1 \rangle X_2 = -\langle JZ, X_1 \rangle X_1 + \langle JZ, X_2 \rangle X_2 = -JZ,$$

and our computation thus becomes

$$\widehat{R}(X_1, X_2)Z = \mp \det(\nabla\nu(p))JZ = \mp \det(\nabla\nu(p)) \, d\text{vol}(X_1, X_2)JZ = -K_G \, d\text{vol}(X_1, X_2)JZ.$$

$\square$

EXAMPLE 31.21. The hyperboloid $H_L^2 = \{y^2 + z^2 - x^2 = 1\}$ has the identity map $H_L^2 \to H_L^2$ as its Gauss map, so it has $K_G \equiv 1$ with the metric g_M, and $K_G \equiv -1$ if the metric is $-g_M$.

EXERCISE 31.22. In $(n+1)$-dimensional Minkowski space $(\mathbb{R}^{n+1}, g_M)$ with coordinates $(\tau, x^1, \ldots, x^n)$ and the metric $g_M := -d\tau^2 + (dx^1)^2 + \ldots + (dx^n)^2$, the connected n-dimensional hyperboloid

$$H_L^n := \left\{ X \in \mathbb{R}^{n+1} \mid \langle X, X \rangle = 1 \right\} = \left\{ (\tau, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^n \mid |\mathbf{x}|^2 - \tau^2 = 1 \right\}$$

is a Lorentzian submanifold, i.e. its signature is $(n-1, 1)$.

(a) Show that for $p \in H_L^n$ and $\mathbf{v} \in T_p H_L^n = p^\perp \subset \mathbb{R}^{n+1}$, the following curves $\gamma : \mathbb{R} \to \mathbb{R}^{n+1}$ are all examples of geodesics in H_L^N:

$$\gamma(t) = \begin{cases} (\cosh t)p + (\sinh t)\mathbf{v} & \text{if } \langle \mathbf{v}, \mathbf{v} \rangle = -1, \\ p + t\mathbf{v} & \text{if } \langle \mathbf{v}, \mathbf{v} \rangle = 0, \\ (\cos t)p + (\sin t)\mathbf{v} & \text{if } \langle \mathbf{v}, \mathbf{v} \rangle = 1. \end{cases}$$

Show moreover that every geodesic in H_L^n is a parametrization of one of these curves.

(b) Show that the isometry group of H_L^n is as large as possible (in the sense of Theorem 24.3). *Hint: Use the Lorentz group* $\mathrm{O}(n, 1)$.

(c) Find an isometrically embedded copy of H_L^2 in H_L^n, and deduce via Example 31.21 and the abundance of isometries that H_L^n has constant sectional curvature 1.

REMARK 31.23. If we instead regard the hyperboloid H_L^n in Exercise 31.22 as a submanifold of signature $(1, n-1)$ in $(\mathbb{R}^{n+1}, -g_M)$, then its constant sectional curvature becomes -1.

We can now adapt Theorem 31.15 to the study of time-like geodesics in Lorentzian manifolds. Here it is necessary to make a choice as to whether "Lorentzian" means the signature is $(n-1, 1)$ or $(1, n-1)$; I have used the latter convention in Remarks 22.12 and 23.8, and will thus stick with it here, even though the other convention seems to be slightly more popular in the literature on mathematical relativity. In a manifold (M, g) of signature $(1, n-1)$, we say a tangent vector $X \in TM$ is *time-like* if $\langle X, X \rangle$ is positive and *space-like* if it is negative. If we were instead using signature $(n-1, 1)$, we would have to define the terms "time-like" and "space-like" the other way around; the rule of thumb in order for these terms to be meaningful is that there should always be exactly one time dimension, though we can consider arbitrarily many spatial dimensions.

The following detail requires slightly more care than in the Riemannian case: if $\{\gamma_s \in \mathcal{P}\}_{s \in (-\epsilon, \epsilon)}$ is an arbitrary smooth family of time-like paths with $\gamma := \gamma_0$ a geodesic, then for the vector field $\eta := \partial_s \gamma_s|_{s=0} \in \Gamma(\gamma^* TM)$ along γ, the integral

$$\int_a^b \langle \eta(t), \eta(t) \rangle \, dt$$

can no longer be viewed as the square of an L^2-norm, and its integrand might be sometimes positive and sometimes negative. This is where it becomes important to assume there is only one time dimension, because in this case, reparametrizing to achieve constant speed removes the uncertainty:

LEMMA 31.24. *On a pseudo-Riemannian manifold* (M, g), *suppose* $\{\gamma_s \in \mathcal{P}\}_{s \in (-\epsilon, \epsilon)}$ *is a smooth family of paths with fixed end points* $\gamma_s(a) = p$, $\gamma_s(b) = q$ *such that* $\gamma := \gamma_0$ *is a geodesic segment and for each* s, *the "speed squared"* $\langle \dot{\gamma}_s(t), \dot{\gamma}_s(t) \rangle \in \mathbb{R}$ *is nonzero and independent of* t. *Then the vector field* $\eta := \partial_s \gamma_s|_{s=0}$ *along* γ *and its covariant derivatives* $\nabla_t^k \eta \in \Gamma(\gamma^* TM)$ *of arbitrary orders* $k \geqslant 0$ *are everywhere orthogonal to* $\dot{\gamma}$.

PROOF. By assumption $\partial_t \langle \dot{\gamma}_s(t), \dot{\gamma}_s(t) \rangle$ vanishes for all s and t, thus

$$0 = \partial_s \partial_t \langle \dot{\gamma}_s(t), \dot{\gamma}_s(t) \rangle|_{s=0} = \partial_t \partial_s \langle \dot{\gamma}_s(t), \dot{\gamma}_s(t) \rangle|_{s=0} = 2\partial_t \langle \nabla_s \partial_t \gamma_s(t)|_{s=0}, \dot{\gamma}(t) \rangle$$
$$= 2\partial_t \langle \nabla_t \eta(t), \dot{\gamma}(t) \rangle = 2 \langle \nabla_t^2 \eta(t), \dot{\gamma}(t) \rangle,$$

where in the last line the term involving $\nabla_t \dot{\gamma}(t)$ does not appear since γ is a geodesic. This proves that $\nabla_t^2 \eta \in \Gamma(\gamma^*TM)$ is everywhere orthogonal to $\dot{\gamma}$. Since $\langle \dot{\gamma}, \dot{\gamma} \rangle \neq 0$ by assumption, the image of γ is a pseudo-Riemannian submanifold of (M, g), and it follows that $T_{\gamma(t)}M = \mathbb{R}\dot{\gamma}(t) \oplus \dot{\gamma}(t)^\perp$ for each t, so that $\eta \in \Gamma(\gamma^*TM)$ splits uniquely into a sum

$$\eta(t) = f(t)\dot{\gamma}(t) + \eta^\perp(t)$$

with a function $f : [a, b] \to \mathbb{R}$ and a section $\eta^\perp \in \Gamma(\gamma^*TM)$ that is everywhere orthogonal to $\dot{\gamma}$. Since $\dot{\gamma}$ is parallel along γ, differentiating the relation $\langle \dot{\gamma}, \eta^\perp \rangle \equiv 0$ gives $\langle \dot{\gamma}, \nabla_t \eta^\perp \rangle \equiv 0$, and applying the Leibniz rule to $f\dot{\gamma}$, we obtain

$$\nabla_t \eta(t) = \dot{f}(t)\dot{\gamma}(t) + \nabla_t \eta^\perp(t)$$

with $\nabla_t \eta^\perp$ also everywhere orthogonal to $\dot{\gamma}$. Repeating this same argument one more time gives

$$\nabla_t^2 \eta(t) = \ddot{f}(t)\dot{\gamma}(t) + \nabla_t^2 \eta^\perp(t)$$

with $\nabla_t^2 \eta^\perp$ also everywhere orthogonal to $\dot{\gamma}$, and since we already know the same holds for $\nabla_t^2 \eta$, it follows that $\ddot{f} \equiv 0$. The function $f : [a, b] \to \mathbb{R}$ is therefore affine, and since $\eta(a) = 0$ and $\eta(b) = 0$, the condition $f(a) = f(b) = 0$ then implies $f \equiv 0$, so η is everywhere orthogonal to $\dot{\gamma}$. The same now follows for all derivatives $\nabla_t^k \eta$ by repeatedly differentiating the relation $\langle \eta, \dot{\gamma} \rangle \equiv 0$. $\square$

Let us now assume we are in the setting of Theorem 31.15, but with the following modifications: (M, g) has signature $(1, n-1)$ and the geodesic $\gamma : [a, b] \to M$ is time-like. The paths in the family $\gamma_s : [a, b] \to M$ with $\gamma_0 = \gamma$ and $\eta := \partial_s \gamma_s|_{s=0}$ can then be assumed to be all time-like after restricting s to a small enough neighborhood of 0, and we are also free to reparametrize them so that they all have constant speed. Lemma 31.24 then implies $\langle \eta(t), \dot{\gamma}(t) \rangle = 0$ for all t, making η a section of the orthogonal complement bundle along the image of γ. Since γ satisfies $\langle \dot{\gamma}, \dot{\gamma} \rangle > 0$ and the signature of (M, g) is $(1, n-1)$, the restriction of g to this complementary subbundle must be strictly *negative*, implying

$$-\langle \nabla_t^2 \eta, \eta \rangle_{L^2} = \langle \nabla_t \eta, \nabla_t \eta \rangle_{L^2} \leqslant 0,$$

with strict inequality unless η vanishes. This is a significant change compared with the proof of Lemma 31.14, but the second term in $\langle \nabla^2 E(\gamma)\eta, \eta \rangle_{L^2}$ undergoes a similar change: whenever $\eta(t)$ is nonzero, the restriction of the metric to the space spanned by $X := \eta(t)$ and $Y := \dot{\gamma}(t)$ has signature $(1, 1)$, making the determinant

$$\det \begin{pmatrix} \langle X, X \rangle & \langle X, Y \rangle \\ \langle Y, X \rangle & \langle Y, Y \rangle \end{pmatrix}$$

negative, so that this determinant is not $\text{Area}(X, Y)^2$, but instead $-\text{Area}(X, Y)^2 \leqslant 0$. Using (31.4), we thus have

$$-\text{Riem}(\eta(t), \eta(t), \dot{\gamma}(t), \dot{\gamma}(t)) = K_S(P_t) \cdot \text{Area}(\eta(t), \dot{\gamma}(t))^2 \leqslant 0,$$

assuming $K_S \leqslant 0$. The result is that in contrast to Lemma 31.14, the operator $\nabla^2 E(\gamma)$ is now *negative*-definite, and Theorem 31.15 becomes the following statement about the *proper time* (cf. Remark 23.8) of a time-like geodesic:

THEOREM 31.25. *Suppose (M, g) is a pseudo-Riemannian manifold of signature $(1, n-1)$ with nonpositive sectional curvature and $\gamma : [a, b] \to M$ is a time-like geodesic connecting $\gamma(a) = p$ to $\gamma(b) = q$. Then for any smooth 1-parameter family of paths $\gamma_s : [a, b] \to M$ with $\gamma_s(a) = p$, $\gamma_s(b) = q$ and $\gamma_0 \equiv \gamma$ such that $\partial_s \gamma_s|_{s=0}$ is not everywhere tangent to $\dot{\gamma}$, there is a number $\epsilon > 0$ such that:*

(1) γ is the only geodesic among the paths γ_s for $s \in (-\epsilon, \epsilon)$;

(2) *For all paths γ_s with $s \in (-\epsilon, \epsilon)$ and $s \neq 0$, the proper time $\tau(\gamma_s)$ satisfies*

$$\tau(\gamma_s) < \tau(\gamma).$$

<div align="right">□</div>

REMARK 31.26. In the literature, results similar to Theorem 31.25 are often stated with signature $(n-1, 1)$ instead of $(1, n-1)$, in which case the definitions of the terms "time-like" and "space-like" must be interchanged. The argument via Lemma 31.24 above then leads to the conclusion that $\langle \nabla_t \eta, \nabla_t \eta \rangle_{L^2} \geqslant 0$, so that the hypothesis necessary for the proof to work becomes

$$K_S \geqslant 0$$

instead of $K_S \leqslant 0$. To get an idea of why this makes sense, take another look at the 2-dimensional case of Exercise 31.22. The hyperboloid H_L^2 has constant positive curvature in $(\mathbb{R}^3, g_M)$, and since signature $(1,1)$ can be interpreted as either $(n-1, 1)$ or $(1, n-1)$, we have some freedom here as to which curves we choose to call time-like vs. space-like. If we view the signature as $(n-1, 1)$, then a tangent vector X is considered time-like if $\langle X, X \rangle < 0$, so the time-like geodesics have the form $\gamma(t) = (\cosh t)p + (\sinh t)\mathbf{v}$, and there is indeed only one connecting any given pair of points, consistent with the conclusions of Theorem 31.25. But if the signature is instead viewed as $(1, n-1)$, then time-like means $\langle X, X \rangle > 0$, and the time-like geodesics are all *periodic* curves of the form $(\cos t)p + (\sin t)\mathbf{v}$. Note that we have not changed the metric, so the curvature is still $+1$, making this a situation in which Theorem 31.25 simply does not apply, and we can see this explicitly since the geodesic segments from a point p to itself are clearly not isolated.[78] In order to apply the literal statement of Theorem 31.25 to this example, one should follow the advice of Remark 31.23 and view it as a submanifold of $(\mathbb{R}^3, -g_M)$ with signature $(1, n-1)$, so that its curvature becomes -1, and the time-like geodesics are again the paths $\gamma(t) = (\cosh t)p + (\sinh t)\mathbf{v}$.

I have one final remark about sectional curvature in the indefinite case. On any manifold M, the set of all 2-dimensional subspaces $P \subset T_p M$ at points $p \in M$ can be given the structure of a smooth manifold,

$$\mathrm{Gr}_2(TM) := \left\{ P \subset T_p M \mid p \in M, \ \dim P = 2 \right\},$$

known as a *Grassmannian*. This manifold is compact if M is, so if (M, g) is a compact Riemannian manifold, its sectional curvature can be viewed as a smooth function on a compact manifold,

$$K_S : \mathrm{Gr}_2(TM) \to \mathbb{R},$$

and is therefore necessarily bounded. In the indefinite case, this is no longer true, because K_S is not defined on all of $\mathrm{Gr}_2(TM)$, but only on the open subset

$$\mathrm{Gr}_2^*(TM, g) := \left\{ P \in \mathrm{Gr}_2(TM) \mid g|_P \text{ is nondegenerate} \right\},$$

which is always noncompact when $\dim M \geqslant 3$ unless g is positive or negative. It's not just that K_S *can* be unbounded in this case: by a theorem of Kulkarni and Nomizu, it *must* be unbounded, both above and below, outside of exceptional cases like the hyperboloid H_L^n in Exercise 31.22 for which K_S is constant. More details on this theorem can be found e.g. in [**Bau06**].

---

[78]Let it be said that this would in any case be an extremely strange spacetime manifold to consider: all time-like geodesics being periodic means a universe in which time travel is not just possible, but mandatory!

Bibliography

[AF03]   R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003.

[Bau06] H. Baum, *Eine Einführung in die Differentialgeometrie* (2006). Notes from a course at HU Berlin, available at https://www.mathematik.hu-berlin.de/~baum/Skript/diffgeo1.pdf.

[Bre93] G. E. Bredon, *Topology and geometry*, Springer-Verlag, New York, 1993.

[CR53]   E. Calabi and M. Rosenlicht, *Complex analytic manifolds without countable base*, Proc. Amer. Math. Soc. **4** (1953), 335–340.

[DK90]   S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1990. Oxford Science Publications.

[Gro85] M. Gromov, *Pseudoholomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), no. 2, 307–347.

[Hat02] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002.

[Hir94] M. W. Hirsch, *Differential topology*, Springer-Verlag, New York, 1994.

[Kob95] S. Kobayashi, *Transformation groups in differential geometry*, Classics in Mathematics, Springer-Verlag, Berlin, 1995. Reprint of the 1972 edition.

[Lee11] J. M. Lee, *Introduction to topological manifolds*, 2nd ed., Graduate Texts in Mathematics, vol. 202, Springer, New York, 2011.

[Lee13a] _____, *Introduction to smooth manifolds*, 2nd ed., Graduate Texts in Mathematics, vol. 218, Springer, New York, 2013.

[Lee13b] _____, *Axiomatic geometry*, Pure and Applied Undergraduate Texts, vol. 21, American Mathematical Society, Providence, RI, 2013.

[MS17]   D. McDuff and D. Salamon, *Introduction to symplectic topology*, 3rd ed., Oxford University Press, 2017.

[Mil97] J. W. Milnor, *Topology from the differentiable viewpoint*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Based on notes by David W. Weaver; Revised reprint of the 1965 original.

[Moi77] E. E. Moise, *Geometric topology in dimensions 2 and 3*, Springer-Verlag, New York-Heidelberg, 1977. Graduate Texts in Mathematics, Vol. 47.

[nLa]   nLab, *Shrinking lemma*. exposition based on a blog post by Matt Rosenzweig, available at https://ncatlab.org/nlab/show/shrinking+lemma.

[Rud69] M. E. Rudin, *A new proof that metric spaces are paracompact*, Proc. Amer. Math. Soc. **20** (1969), 603.

[Sal16] D. A. Salamon, *Measure and integration*, EMS Textbooks in Mathematics, European Mathematical Society (EMS), Zürich, 2016. MR3469972

[Spi99a] M. Spivak, *A comprehensive introduction to differential geometry*, 3rd ed., Vol. 1, Publish or Perish Inc., Houston, TX, 1999.

[Spi99b] _____, *A comprehensive introduction to differential geometry*, 3rd ed., Vol. 3, Publish or Perish Inc., Houston, TX, 1999.

[Tho54] R. Thom, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954), 17–86 (French).

[Tro92] A. J. Tromba, *Teichmüller theory in Riemannian geometry*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1992.

[Wen18] C. Wendl, *Topology I and II* (2018). Notes from the course at HU Berlin, available at https://www.mathematik.hu-berlin.de/~wendl/Winter2018/Topologie2/lecturenotes.pdf.

[Wen19] _____, *Integration auf Untermannigfaltigkeiten* (2019). Skript zur Vorlesung Analysis III an der HU Berlin, verfügbar unter https://www.mathematik.hu-berlin.de/~wendl/Winter2019/Analysis3/Skript_DifferentialFormen.pd