These course-notes are a minor revision (from 2003) of a first draft that was prepared for a course in spring 2000 at ASU. Some type-ohs continue to be corrected at irregular times.

They are extremely close to (but nowhere as accurate) as Spivak’s books – and they can be justified only by Spivak being out of print when the 2000 spring semester began. Since then Spivak’s volumes have been republished (now typeset, no longer type-written), and every user of these notes is expected to eventually buy the originals by Spivak – he deserves his royalties!

These notes are based on class-notes taken from a class taught by Al Lundell at the University of Colorado in 1983/84, and on two classes taught in 1989 and 1991 at Arizona State University. Originally, these notes may have been quite independent, but upon efforts to make them more comprehensive, more precise, they have again converged very much to Spivak’s treatment.

However, in some places the presentation departs from Spivak and material follows more closely e.g. Sternberg (e.g. notation and terminology involving tensors), Boothby (e.g. Riemannian basics), Marsden and others. The main practical value of these notes is that they use the same notation (even if it is just $u^i$ in place of $x^i$) that the instructor has become too accustomed to, and will use in class... and they integrate questions / exercises for our class.

The current version will include more diagrams – which are essential for readability. Moreover, sections such as the reviews of basic topology and differentiability, which really belong into an appendix, are included in the order that the class actually covered them.

The affiliated explorations that use computer algebra system have not yet been integrated into these notes.


1 Curves in the plane and 3-space

This first section addresses mostly prerequisite material and is not completely self-contained. It provides some basic definitions and discusses some fundamental theorems. Central objectives are to raise some questions that will have to be addressed when working in more general settings, and to set the stage for the questions about geometric properties.

1.1 Basic definition of a curve

In many settings it may be appropriate to think of a curve as a set of points in the plane or in 3-space. However, in differential geometry and other advanced settings, it is generally more convenient to work with a different notion – basically calling what previously was named a parameterization the "curve".

Definition 1.1 A curve is a continuous function defined on an interval $I \subseteq \mathbb{R}$, taking values in a (topological) space $M$ (in this section $M$ is assumed to be $\mathbb{R}^n$). (The interval in this definition may be open or closed, finite, semi-infinite or the entire real line. At this time we only assume enough structure on the space $M$ so that we can talk about continuity.)

The key difference is that with this definition a curve is a function. Consequently it has a richer structure than just a set of points – a structure that facilitates technical analysis. Moreover, this definition easily carries over to much more general settings – e.g. we may think of a vibrating membrane as a curve in an appropriate space $M$ of functions of two variables. What matters is that the space has enough structure (at least a topology) so that we may talk about continuity. (Later we will require additional structures on the space $M$ so that we can differentiate curves.)

Several properties of curves deserve their own names. A curve $\gamma: I \mapsto M$ is called closed if $I = [a,b]$ is a (finite) closed interval and $\gamma(a) = \gamma(b)$. If the restriction of a closed curve $\gamma$ to $[a,b]$ is one-to-one, then $\gamma$ is called a simple closed curve.

Example 1.1 The circle $S^1 = \{(x,y) \in \mathbb{R}^2: x^2 + y^2 = 1\}$ is the image of the simple closed curve $\gamma: [0,2\pi] \mapsto \mathbb{R}^2$ defined by $\gamma(t) = (\cos t, \sin t)$. (Note that the circle $S^1$ is not a curve!)

Definition 1.2 If $\gamma: I \mapsto M$ is a curve defined on a finite interval $I$ and $\phi: J \mapsto I$ is a continuous function that maps a finite interval $J$ onto $I$ (mapping endpoints to endpoints), then the curve $\sigma = \gamma \circ \phi: J \mapsto M$ is called a reparameterization of $\gamma$.

(Commonly one requires in addition that $\phi$ is also one-to-one.)

The notion of reparameterization may be extended to infinite intervals provided one suitably modifies the notion of endpoint to endpoint, e.g. requiring the existence of the limits $\lim_{t \to \pm \infty}$ and that these equal $-\infty$ and $\infty$. Among one-to-one reparameterizations one distinguishes orientation-preserving and orientation-reversing reparameterizations according to whether the map $\phi$ maps the left endpoint of $J$ to the left or right endpoint of $I$. 


1.2 Differentiable curves and arc-length

An intuitive notion of the length of a curve in $\mathbb{R}^n$ may be built on successive approximations by polygonal approximations. More specifically, suppose that $\gamma: [a, b] \rightarrow \mathbb{R}^n$. Let $\| \cdot \|$ denote the Euclidean norm $\|(x_1, \ldots, x_n)\| = \sqrt{x_1^2 + \ldots + x_n^2}$ in $\mathbb{R}^n$. Define the length $L(\gamma)$ of the curve as the supremum (possibly infinite)

$$L(\gamma) = \sup \sum_{i=0}^{n(\mathcal{P})} \| \gamma(t_{i+1}) - \gamma(t_i) \|$$

(1)

where $\mathcal{P}$ ranges over all partitions $\mathcal{P} = \{ t_i : 0 \leq i \leq n(\mathcal{P}) \}$ such that $a = t_0 < t_1 < \ldots < t_n(\mathcal{P}) = b$. It is very important to note that this definition of length cannot directly generalize to spaces for which one does not have an a-priori notion of distance – i.e. where $\| \cdot \|$ has no meaning (yet).

The key idea is that for differentiable curves there is a natural alternative – the length is the integral of the speed, and this notion will generalize, even give rise to the concept of Riemannian manifolds. Loosely speaking, the main idea is to rewrite

$$\sum_{j=0}^{n} \| \gamma(t_{i+1}) - \gamma(t_i) \| = \sum_{j=0}^{n} \| \gamma(t_{i+1}) - \gamma(t_i) \| \cdot (t_{i+1} - t_i) \rightarrow \int_{a}^{b} \| \gamma'(t) \| dt =: L(\gamma)$$

(2)

By fairly straightforward (advanced) calculus arguments one may make this idea rigorous, i.e. show that for any continuously differentiable curve defined on a finite closed interval there exists a unique limit which defines the length of the curve.

The key to most of our later work will be to develop a natural notion of a tangent vector to a curve (taking values in an abstract manifold) that does not require any prior notion of a difference $\gamma(t_{i+1}) - \gamma(t_i)$ of two points in that space. (However, on a more advanced level, a key idea is to make sense of such differences by interpreting objects such as points as linear functionals on the space of smooth functions on the manifold – material for the next class!) Once we have such a generalized notion of a tangent vector, much of the following fundamental notions, structures, calculations and arguments will carry over to the general case of abstract manifolds.

**Definition 1.3** A curve $\gamma: (a, b) \rightarrow \mathbb{R}^n$ is called differentiable if for every $t \in (a, b)$ the limit $\lim_{h \to 0} \frac{1}{h} (\gamma(t + h) - \gamma(t))$ exists. If the limit exists, it is denoted $\gamma'(t)$ and called the velocity at $\gamma(t)$ (or at $t$).

The second derivative $\gamma''(t)$ is defined analogously, and is called the acceleration at $\gamma(t)$ or at $t$.

The magnitude $\| \gamma'(t) \|$ of the velocity is called the speed.

In the case of plane and space curves one routinely identifies the point $\gamma(t) \in \mathbb{R}^n$ with the arrow (vector) from the origin to this point. On the other hand the velocity and acceleration are commonly visualized as arrows (vectors) rooted at the point $\gamma(t)$, or even at $\gamma(t) + \gamma'(t)$. There appears to be a certain arbitrariness about this representation – but it seems to make sense after a little thought. The upcoming construction of the tangent bundle will illuminate the situation and provide clarifying distinctions. A helpful preparation at this time is to think about possible alternative representations, and to find good arguments why the usual placements of the arrows are a good choice without any compelling alternative. Also think of how these arrows are affected by changes of units, e.g. going from inches to centimeters, or from minutes to seconds.

**Exercise 1.1** Show that if the velocity $\gamma'$ is constant then the (image of the) curve $\gamma$ is a straight line, but the converse is not true.
Exercise 1.2 Verify that the curve \( \gamma: \mathbb{R} \rightarrow \mathbb{R}^2 \) defined by \( \gamma(t) = (\cos(t), \sin(t)) \) has constant speed, but nonzero acceleration.

Exercise 1.3 Prove that the acceleration \( \gamma'' \) is orthogonal to the velocity \( \gamma' \) (for all \( t \)), i.e. \( \langle \gamma'(t), \gamma''(t) \rangle \equiv 0 \) if and only if the speed is constant.
(Hint: Differentiate \( c \equiv \|\gamma'(t)\|^2 = \langle \gamma'(t), \gamma'(t) \rangle \). Read the identities both directions.)

Exercise 1.4 Verify that the curve \( \gamma: \mathbb{R} \rightarrow \mathbb{R}^2 \) defined by \( \gamma(t) = (t^2, 0) \) if \( t \geq 0 \), and \( \gamma(t) = (0, t^2) \) if \( t < 0 \) is continuously differentiable, yet its image in the plane has a corner.

Blanket assumption:
For most of the following we shall assume that all curves under consideration are twice continuously differentiable and that \( \|\gamma'(t)\| \neq 0 \) for all \( t \).

In many cases we will for convenience even assume that the curve is smooth, i.e. that it has continuous derivatives of all orders. Note that the assumption that the speed is never zero eliminates such nuisances as the corners exhibited by the continuously differentiable curve of exercise 1.4. Moreover, it also prohibits such nuisances as exhibited by the curve \( \gamma(t) = (\cos(t - t^3), \sin(t - t^3)) \) which “go back and forth” along the image of the curve. If there is a need to allow for such behaviours it is usually easy to either consider the curve in pieces, or relax the requirements in specific cases and then adapt the desired theorems as needed.

Obviously many different curves may have the same image – and there seems an arbitrariness about picking a specific parameterization. For theoretical purposes it is often convenient to work with a canonical reparameterization of a differentiable curve \( \gamma \). A natural choice is such that the parameter \( t \) (often considered as time) agrees with the distance traveled along the curve.

**Definition 1.4** The arc-length of a continuously differentiable curve \( \gamma: [a, b] \rightarrow \mathbb{R}^n \) is defined as the function \( s: [a, b] \rightarrow \mathbb{R} \),

\[
s = \psi(t) = \int_a^t \langle \gamma'(\tau), \gamma'(\tau) \rangle^{1/2} \, d\tau \tag{3}
\]

The curve \( \gamma \) is called parameterized by arc-length if \( \psi(t) \equiv t \), or, equivalently, \( \langle \gamma', \gamma' \rangle \equiv 1 \).

This definition relies on the standard inner product \( \langle \cdot, \cdot \rangle \) which in \( \mathbb{R}^n \) is almost synonymous with the notion of (Euclidean) distance. The key idea underlying Riemannian geometry – see chapter 4 – is that once one has a suitable generalization of this inner product, then most of the notions and properties naturally carry over to abstract settings.

To explicitly reparameterize a given curve \( \gamma: [a, b] \rightarrow \mathbb{R}^n \) by arc-length generally requires one not only to evaluate the integral in equation (3) in closed form, but in addition, to solve the equation \( s = \psi(t) \) for \( t \) in terms of \( s \) – generally a hopeless task if looking for closed form expressions in terms of the traditional elementary functions. Thus one usually has to choose between either formal expressions (for theoretical purposes), and numerical techniques (in practical calculations). We will see that curves that are parameterized by arc-length allow for particular elegant descriptions of their geometry.

**Exercise 1.5** Reparameterize the curve \( \gamma: [0, 1] \rightarrow \mathbb{R}^2 \) defined by \( \gamma(t) = (\sqrt{1-t^2}, t) \) by arc-length by explicitly integrating equation (3) and solving for \( t \) in terms of \( s \).
Exercise 1.6 Calculate and sketch the graphs of the arc-lengths of the curves $\gamma_1(t) = (t, t)$, $\gamma_2(t) = (\cos 2\pi t, \sin 2\pi t)$, $\gamma_3(t) = (t, t^2)$, and $\gamma_4(t) = (\cos 2\pi t, \sin 2\pi t, ct)$, all defined for $t \geq 0$. If feasible, reparameterize each curve by arc-length.

Exercise 1.7 Verify by direct calculation that arc-length of plane curves is invariant under orthogonal linear transformations: More specifically, let $\gamma: [a, b] \mapsto \mathbb{R}^2$ and $\sigma: [a, b] \mapsto \mathbb{R}^2$ be two differentiable curves that are related by $\sigma = A \cdot \gamma$ where $A$ is a $2 \times 2$ rotation matrix with $a_{11} = \pm a_{22} = \cos \theta$ and $\mp a_{12} = a_{21} = \sin \theta$ for some value of $\theta \in \mathbb{R}$. Calculate and compare the arc-length functions associated to $\sigma$ and $\gamma$.

Repeat for reflections with $a_{11} = -a_{22} = \cos \theta$ and $a_{12} = a_{21} = \sin \theta$.

Explain in geometric terms – e.g. using the earlier definition in terms of polygonal approximations – why this is expected. Try to make this into a rigorous argument that applies to $\mathbb{R}^n$ for any dimension $n \geq 1$.

1.3 Curvature of plane and space curves

In a very general sense, differentiability makes precise the intuitive idea of being approximable by a linear object, think of tangent lines and planes, or more generally by linear functions and maps. Curvature then may loosely be thought of as a quantification how far the object is from locally being linear. Curvature is the central concept of differential geometry.

In the case of graphs of functions $y = f(x)$ all calculus students learn that the second derivative is somehow related to how much the graph curves – but it is important to fully understand that, and why, the second derivatives does not represent curvature.

Exercise 1.8 Consider the graph of $f_0(x) = x^2$ for $0 \leq x \leq b$. For small angles $\theta$ and small values of $b > 0$ the image of the graph under a rotation by an angle $\theta$ about the origin is again the graph of a function $f_\theta(x)$. Find an explicit formula for $f_\theta$ and show that its second derivative is not constant equal to $f_0'' \equiv 2$.

Suggestion: Use $(x, y)$ to denote points on the original curve $y = x^2$ and let $(\xi, \eta)$ denote points on the rotated curve. Express $x$ and $y$ in terms of $\xi$ and $\eta$ (compare exercise 1.7), substitute into $y = x^2$ and solve for $\eta$ in terms of $\xi$. Finally calculate $\frac{d^2\eta}{d\xi^2}$.

Compare the associated MAPLE worksheet.

There are two aspects of the second derivative that do not make it suitable for immediate use to denote a notion of curvature: First the derivative in the exercise is taken with respect to the first coordinate $x$, as opposed to the intrinsic arc-length parameter. Secondly, the slopes are not the same as the direction of the curve – the tan in $y' = \tan \alpha$ distorts the description.

In the following $T, N, \sigma, \kappa, \ldots$ are correct function names. However, for emphasis only, we often will write $T(s), N(s), \kappa(s), \ldots$ etc. This will also help distinguish from these from the compositions $T \circ \psi, \kappa \circ \psi, \ldots$ which with common abuse of notation, often are written as $T(t), \kappa(t), \ldots$. One might even want to instead consider the functions (in our notation) $T \circ \sigma^{-1}, \kappa \circ \sigma^{-1}, \ldots$ whose domain are points in the image of the curve. However, for the purposes of differentiation etc., is is much easier to consider $T, N, \kappa, \ldots$ as functions defined on the parameter interval $J$.

Thus we first consider smooth curves $\sigma: I \mapsto \mathbb{R}^2$ and $\sigma: I \mapsto \mathbb{R}^3$ that are parameterized by arc-length. This implies that $\|\sigma'(s)\| \equiv 1$, i.e. the velocity is a unit tangent vector to the curve at
any time \( s \in I \), suggesting the notation \( T(s) \) for \( \sigma'(s) \). In the case of plane curves it is convenient to complete \( \{T(s)\} \) to a positively oriented orthonormal basis (frame”) \( \{T(s), N(s)\} \).

To describe the rate of change of this direction differentiate again. We define the curvature to be the (signed) magnitude \( \kappa = \pm\|\sigma''\| \) of this derivative. More specifically, in the case of plane curves it is convenient to define \( \kappa(s) = \langle T'(s), N(s) \rangle \) (allowing both positive and negative values). In the case of space curves, which will be considered from here on, the natural way of choosing a direction for a normal \( N \) is to require that the curvature is nonnegative and use \( \sigma'' = \kappa N \) as the defining equation for both \( \kappa \) and \( N \) – of course, in the case that for some \( s_0 \in J \), \( \|\sigma''(s_0)\| = 0 \) this only defines \( \kappa(s_0) \), but does not determine a direction \( N(s_0) \). Recall from exercise 1.3 that \( \langle \sigma', \sigma' \rangle \equiv 1 \) implies that \( \sigma'' \perp \sigma' \), and hence for each \( s \in I \) if \( \kappa(s) \neq 0 \) then \( N(s) \perp T(s) \).

Before differentiating further, define for each \( s \in I \) where \( \kappa(s) \neq 0 \) a third unit vector \( B(s) = T(s) \times N(s) \) (using the standard cross-product in 3-space). (Observe the analogy to predetermining \( N \) in the planar case as the last vector to complete an orthonormal frame – and allowing the coefficient to have both positive and negative values. This also suggests an obvious generalization to higher dimensions \( n > 3 \).) The triple \( (T(s), N(s), B(s)) \) is called the Frenet frame along the curve \( \sigma \) – it is only defined at points where \( \kappa(s) \neq 0 \).

To continue the investigations differentiate \( N \) (with respect to \( s \)). By an argument analogous to the one employed earlier, \( N' = \frac{d}{ds}N \) is orthogonal to \( N \), and hence may be written as a linear combination \( \frac{d}{ds}N(s) = a_{21}(s)T(s) + \tau(s)B(s) \). The function \( \tau \) is called the torsion of the curve.

Intuitively, the torsion quantifies the rate at which the curve twists out of a plane – compare the exercise 1.13.

To identify the parameter \( a_{21} \), differentiate the identity \( 0 \equiv \langle T, N \rangle \) and find that

\[
0 \equiv \kappa \langle N, N \rangle + a_{21} \langle T, T \rangle + \tau \langle T, B \rangle.
\]

Since the third term vanishes, it is clear that \( a_{21} = -\kappa \). To complete the analysis differentiate the identities \( 0 \equiv \langle T, B \rangle \), \( 0 \equiv \langle N, B \rangle \), and \( 1 \equiv \langle B, B \rangle \) to obtain \( \frac{d}{ds}B = -\tau N \).

Taken together, these equations form the famous Frenet-Serret formulas:

\[
\begin{align*}
T' &= \kappa N \\
N' &= -\kappa T + \tau B \\
B' &= -\tau N
\end{align*}
\] (4)

To emphasize the characteristic structure of this set of equations, we rewrite it by forming a matrices \( R = (T|N|B) \) and \( R' = (T'|N'|B') \) whose columns are the representations of these vectors with respect to the standard coordinates in \( \mathbb{R}^3 \). Note that \( R \) is an orthogonal matrix, i.e. \( R^T R = R R^T = I_{3 \times 3} \). With this notation, the Frenet-Serret formulas become

\[
R' = RA \quad \text{where} \quad A = \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix}
\] (5)

Preview: We will later consider \( R \) as a point (!) in the three dimensional manifold (actually a Lie group) \( SO(3) \) of \( 3 \times 3 \) orthogonal matrices. The product \( RA \), which again is a \( 3 \times 3 \) orthogonal matrix (though generally not orthogonal) will be considered a tangent vector at \( R \), and it will be clear from general considerations (in some sense generalizing the arguments made above about the derivatives of the inner products between the columns of the matrix \( R \)) why \( A \) has to be skew symmetric, i.e. \( A^T = -A \).
When given an initial reference frame $R(0)$ consisting of three orthogonal unit vectors $T(0)$, $N(0)$ and $B(0)$ together with two sufficiently regular functions $κ(s)$ and $τ(s)$ this system of differential equations uniquely determines $R(s)$ for all times $s$, considering $R$ as a curve in $SO(3)$. Continuing further, if in addition an initial point $σ(0) ∈ \mathbb{R}^3$ has been specified, then the Frenet Serret formulas together with the differential equation $σ’ = T(s)$ uniquely determine a curve in $\mathbb{R}^3$. It is straightforward to verify that the curvature and torsion of this curve agree with the data provided to the differential equation. A most important corollary of this study is that the curvature and torsion completely determine a smooth curve up to translation (as determined by $σ(0)$) and rotation (as determined by $R(0)$).

**Exercise 1.9** Explain how this gives as a corollary that curvature and torsion are invariant under translation and rotation (compare also exercise 1.7).

**Exercise 1.10** Explore how complicated a brute force linear algebra calculation is (similar to exercise 1.7) that directly shows that curvature and torsion are invariant under rotations and translation. (It may be appropriate to use a computer algebra system for part of this work.)

**Exercise 1.11** Show that if the curvature $κ ≡ 0$ of a plane curve vanishes identically, then the curve is a straight line.

Is the same true for a curve in 3-space? Explain!

**Exercise 1.12** Show that if the curvature $κ ≡ c ≠ 0$ of a plane curve is constant, then the curve is a circle with radius $1/c$.

Is the same true for a curve in 3-space? Explain! (Remark: Feel free to consult the literature for elegant arguments – a direct brute-force approach quickly can get very messy!)

**Exercise 1.13** Show that if the torsion $τ ≡ 0$ of a space curve vanishes identically, then the curve lies in a plane.

The Frenet-Serret formulas provide a most beautiful and comprehensive geometric description of the curves in 3-space. They appear to intrinsically rely on working with parameterizations by arc-length, yet for most curves explicit closed-form formulas for parameterizations by arc-length are beyond reach. However, note that all these formulas only involve derivatives of the curve $σ$. Consequently, there is no need to ever explicitly calculate the arc-length. All that is needed is the integrand of the formula (3) – the chain-rule does the rest.

Consider a smooth curve $γ: I = [a, b] \mapsto \mathbb{R}^3$. Define $ψ(t) = \int_a^t √⟨γ’(τ), γ’(τ)⟩ dτ$. As usual, assume that $∥γ’(t)∥ > 0$ for all $t ∈ I$. Then the curve $σ: J = [0, L(γ)] \mapsto \mathbb{R}^3$ defined by $σ = γ ∘ ψ^{-1}$ is the reparameterization of $γ$ by arc-length. Differentiating $γ = σ ∘ ψ$, the chain rule relates the velocities $γ’ = (σ’ ∘ ψ) ψ’ = ∥γ∥ T ∘ ψ$ – i.e. for each $t ∈ I$, the velocity vector $v(t) = γ’(t)$ points in the same direction as $T(ψ(t))$ but it generally has non-unit magnitude (or “speed”) $∥γ’(t)∥$. In practical calculations one typically first obtains $γ’$, then $∥γ’∥$ and $T$. The key to avoiding excessively unpleasant calculations is to never differentiate normalized expressions such as $T$ or $N$, but rather first take suitable cross- and dot-product of derivatives of $γ$. The next step is to note that $γ'' = (σ'' ∘ ψ) (ψ')^2 + (σ’ ∘ ψ) ψ''$ implies that for every $t ∈ I$ the vector $γ''(t)$ lies in the plane spanned by $T(ψ(t))$ and $N(ψ(t))$. (This plane is called the osculating plane.) In practical calculations in 3-space one calculates $γ''$, then calculates $B$ by normalizing the cross-product $γ’ × γ''$. Only afterwards(!) one calculates
$N = B \times T$. Returning to the acceleration $\gamma''$, the magnitudes of its tangential and normal components are easily calculated as

$$a_\parallel = (T \circ \psi) \cdot \gamma'' = \frac{\gamma' \cdot \gamma''}{\|\gamma'\|} \quad \text{and} \quad a_\perp = (N \circ \psi) \cdot \gamma'' = \pm \sqrt{\|\gamma''\|^2 - a_\parallel^2} \quad (6)$$

The curvature $\kappa$ (and radius of curvature $\varrho = \frac{1}{\kappa}$) and the torsion may be obtained in various ways. Typical formulas suitable for practical calculations (for space curves) are

$$\kappa \circ \psi = \frac{|a_\perp|}{\|\gamma'^2\|} = \frac{\|\gamma' \times \gamma''\|}{\|\gamma'\|^3} \quad \text{and} \quad \tau \circ \psi = \frac{(\gamma' \times \gamma'') \cdot \gamma''}{\|\gamma' \times \gamma''\|^2} \quad (7)$$

**Exercise 1.14** Derive the formulas presented above for $\kappa \circ \psi$ and for $\tau \circ \psi$ from the definitions of $\kappa$ and $\tau$ in terms of the Frenet formulas.

**Exercise 1.15** Consider a smooth planar curve $\gamma: I \rightarrow \mathbb{R}^2$, not necessarily parameterized by arclength. Devise a practical strategy to calculate $T, N, \kappa$ with minimal effort. (Note, that from $T$ one easily obtains $N$ by interchanging the components and changing the sign of one of the components. Which one? Why?)

**Exercise 1.16** For curves in the plane given as graphs of functions, i.e. $\gamma(t) = (t, f(t))$ or casually $y = f(x)$, derive the usual formula $\kappa = y''/(1 + (y')^2)^{3/2}$ for the curvature. Note that technically the above stands for $\kappa \circ \gamma_1 \circ \psi = y''/(1 + (\gamma_2')^2)^{3/2}$.

**Exercise 1.17** Verify that the curve $\gamma: \mathbb{R} \rightarrow \mathbb{R}^2$ defined by $\gamma(t) = (\exp(-1/t^2), 0)$ if $t > 0$, $\gamma(0) = (0, 0)$ and $\gamma(t) = (0, \exp(-1/t^2), 0)$ if $t < 0$ is infinitely many times continuously differentiable on the whole real line. Describe the $(T(s), N(s))$-frame for this curve (this is very easy), with special attention to what happens when $t = 0$.

**Project 1.18** Write a MAPLE procedure that takes as input a space curve, i.e. algebraic expressions for $x(t), y(t)$ and $z(t)$ (and possibly other parameters such as the domain $I$), and which gives as output an animation of the Frenet frame along the curve.

As test curves consider a helix $\gamma(t) = (\cos t, \sin t, t)$ and the $(2,3)$-torus-knot $\gamma = u^{-1} \circ \ell$ where $\ell(t) = (2t, 3t)$ for $t \in [0, 2\pi]$ and $u^{-1}(\theta, \phi) = ((R + r \sin \phi) \cos \theta, (R + r \sin \phi) \sin \theta, r \cos \phi)$, e.g. for $R = 5$ and $r = 2$.

**Project 1.19** Consider a family of closed curves $\gamma(s, t)$ that are parameterized by arc-length $s$ and that evolve with time $t$ according to e.g. the heat-equation $\frac{\partial^2}{\partial s^2} \kappa(s, t) - \frac{\partial}{\partial t} \kappa(s, t) = 0$. Intuitively, this is the easiest model that describes how loops may try to straighten out under the influence of tension.

For simplicity, start with initial curvature functions $\kappa(s, 0)$ that are expressed as (finite) Fourier polynomials $\kappa(s, 0) = a_0 + \sum_{j=1}^N (a_j \cos jt + b_j \sin jt)$. This allows one to explicitly write out the solutions $\kappa(s, t) = a_0 + \sum_{j=1}^N e^{-jt}(a_j \cos jt + b_j \sin jt)$ of the heat equation.

First find conditions on the Fourier coefficients that assure that the associated curve is closed. Then integrate the two-dimensional analogue of the Frenet-Serret formulas to obtain the associated curve $\sigma(t, s)$. Animate the images of the curves.

An exploratory worksheet that addresses this project is available from the WWW-site http://math.asu.edu/kawski/MAPLE/MAPLE.html. However, it has at least a cosmetic flaw as it arbitrarily fixes $\sigma(0, t)$ and $\frac{\partial}{\partial s} \sigma(0, t)$ — a nicer solution would include a more physical solution,
e.g. fixing the center of mass of the curve by translating the curve as needed. Even better, it would be nice to add, if necessary, a rotation so that the total angular momentum as determined from $\frac{d}{dt}\sigma(t,0)$ is constant equal to zero. Any improvements of this worksheet are highly welcome. They likely will lead to further, even more interesting applications – starting with an analogous exploration of loops in 3-space!

\begin{quote}
In a future version add a little classical stuff involving evolutes and involutes – much of this can be done in exercises. Alternatively, do this in some MAPLE worksheets. Till then refer to Opreah’s book as a nice reference.
Lots of quick insight may also be gained from the superb Famous curves applets from St. Andrews University, and available free on-line at http://www-history.mcs.st-andrews.ac.uk/Java/.
\end{quote}
2 Manifolds

2.1 Introduction

We want to think of manifolds as abstractions and generalizations of the intuitive notions of curves and surfaces. This subsection reviews a few key ideas, purposes and examples. The next subsection provides a few fundamental topological notions to prepare for a precise definition of manifolds, first in the topological category, and then in the differential category.

The upcoming definition will characterize a manifold as a space which is such that every point in it has a neighbourhood that is homeomorphic to an open subset of a Euclidean space $\mathbb{R}^n$. In particular, we shall not allow for edges and boundaries to avoid the associated technical complications. Next, we will equip manifolds with differentiable structures that allow for notions such as dynamical systems evolving on the manifolds, and for generalized notions of curvature. Typical objectives are to analyze the effects of curvature on the global topological structure or on the behaviour of dynamical systems. A need to integrate over (subsets of) manifolds arises naturally. A major role of local coordinate charts is to transfer these differential (and integral) concepts back into familiar Euclidean space where standard techniques may be employed for calculations.

Throughout we will emphasize geometric points of view – as a simple example what we don’t want think of the two dimensional sphere $S^2$ as (the union of) the graph(s) of two functions $z = \pm \sqrt{x^2 + y^2}$. This rather arbitrary preferential treatment of $z$ versus $x$ and $y$ begins to hide the full symmetry of the sphere under a group of rotations and reflections.

Before proceeding to technical descriptions let us take a brief look at some typical examples that should be included in our notion of manifold.

Curves and surfaces, especially the Euclidean spaces $\mathbb{R}^n$, and (open) subsets of Euclidean spaces should be manifolds. However, we may impose conditions so as to avoid e.g. self-intersections, boundaries, and, in the category of differentiable manifolds, cusps, corners and the like.

The characterization of the two dimensional sphere $S^2 \subseteq \mathbb{R}^3$ as the set of all $(x, y, z) \in \mathbb{R}^3$ that satisfy $x^2 + y^2 + z^2 = 1$, invites a natural generalization to higher dimensional analogues of surfaces as subsets of $\mathbb{R}^n$ that may be characterized by (sets of) equations $F_k(x_1, x_2, \ldots x_n) = 0$ ($k = 1, \ldots p$). To avoid cusps and corners one usually imposes a condition that the gradient (or a higher dimensional analogue) does not vanish.

As a special case, this description immediately opens the door to objects such as the group of special orthogonal $n \times n$ matrices $SO(n)$. The defining equation $A^T A = I_{n \times n}$ is of the same form as the equation of the sphere given above. What makes these matrix manifolds particularly interesting is their natural group structure – there is a natural notion of multiplying points on the generalized surface – this is the starting point for Lie groups.

A different way that many manifolds of interest are obtained is by taking quotients. In the most simple case the circle $S^1$ arises as a quotient of $\mathbb{R}$ by $\mathbb{Z}$. Intuitively, for any periodic function $f$ with period $p > 0$, i.e. $f(x + p) = f(x)$ for all $x \in \mathbb{R}$, one may consider as its natural domain any interval $[a, a + p]$ with endpoints identified. More abstractly, consider the equivalence relation $\sim$ defined on $\mathbb{R}$ by $x \sim y \iff (x - y)/p \in \mathbb{Z}$. Then each point on the circle $\Theta$ represents an equivalence class $[\Theta] = \{\Theta + kp : k \in \mathbb{Z}\}$.

In an analogous way, the torus arises naturally (e.g. very commonly in dynamical systems) as the quotient of the plane $\mathbb{R}^2$ by $\mathbb{Z}^2$. One commonly visualizes the torus as the unit square $[0, 1] \times [0, 1]$ with opposing edges identified.
If one starts with the same square, but identifies one (or two) sets of opposing edges with orientation reversed one arrives at the Klein bottle and at the projective plane. Neither one of these can be visualized in the usual way as a *surface* in $\mathbb{R}^3$, but apparently each shares many properties with the torus due to their analogous construction.

More abstractly, projective spaces arise when considering the *spaces* of all (straight) lines in $\mathbb{R}^n$ that pass through the origin. Before looking at this more closely, recall the simple case of considering the space of all (semi-infinite, open) rays emanating from the origin. Each of these rays may be naturally identified with the point on the unit sphere (unit circle) through which it passes. Thus we may think of the spheres $S^{n-1}$ as arising from $\mathbb{R}^n \setminus \{0\}$ as quotients under the equivalence relation $x \sim y \iff$ if there exists $\lambda \in \mathbb{R}$, $\lambda > 0$ such that $x = \lambda y$. In a practical sense this is closely related to considering only the angle(a) $\theta$ (or $(\theta, \phi)$) when working with polar (or spherical coordinates).

In analogy, if one discards the requirement $\lambda > 0$ in the preceding definition of equivalence, then the equivalence classes are the *lines* through the origin. (More precisely, since we started with $\mathbb{R}^n \setminus \{0\}$, the origin is removed from each line, or the line really consists of two rays.) One may visualize the resulting quotient space as the space of *pairs of opposite points* on the sphere $S^{n-1}$, or as a semi-sphere with two halves of the *equator* (which is a sphere $S^{n-2}$ by itself) identified, or *glued together* with careful attention to the orientation of each piece.

From these projective spaces it is only a small step to Grassmannian manifolds which may be thought of as spaces of $m$-dimensional (hyper-)planes in $n$-dimensional Euclidean space.

A typical application where these appear naturally is in the classification of linear control systems $\dot{x} = Ax + Bu$, with *state* $x \in \mathbb{R}^n$ and *control* $u \in \mathbb{R}^m$. Here one considers two systems equivalent if one may be transformed into the other by coordinates changes, in state $\tilde{x} = Rx$ and control space $\tilde{u} = Su$, and/or under *feedback transformations* $\tilde{u} = u + Kx$, $\ldots$

All the above examples clearly have (preserve) some additional structure beyond just being sets of points. In order to be able to work with concepts such as continuity and notions of derivatives one intuitively needs some notion of distance. Indeed, while one can start with even more general topological spaces, in the finite dimensional setting very little is lost if one requires that the set is equipped with at least some a-priori notion of distance. However, this basic notion of distance will primarily be used only as a foundation for e.g. continuity, and should not be confused with the Riemannian metrics that we will study later, and which have a deeply connected with curvature.
2.2 Some basic topological notions

This subsection reviews some basic definitions and properties of metric spaces and objects in topology.

Definition 2.1 A metric on a set \( X \) is a function \( d: X \times X \mapsto \mathbb{R} \) that satisfies

(i) For all \( x, y \in X \), \( d(x, y) \geq 0 \), and \( d(x, y) = 0 \) if and only if \( x = y \) (positive definiteness),

(ii) For all \( x, y \in X \), \( d(x, y) = d(y, x) \) (symmetry), and

(iii) For all \( x, y, z \in X \), \( d(x, z) \leq d(x, y) + d(y, z) \) (triangle inequality).

A metric space is a pair \((X, d)\) where \( X \) is a set and \( d \) is a metric of \( X \).

A set \( X \) may be equipped with different metrics, which may result in different, or the same, notions of continuity. For example, common metrics on \( \mathbb{R}^n \) are

- the usual Euclidean distance defined by \( d_2(x, y) = \|y - x\| = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2} \)
- the taxi-cab metric defined by \( d_1(x, y) = \sum_i |y_i - x_i| \), and
- the sup-norm \( d_{\infty}(x, y) = \max_i |y_i - x_i| \).

Two metrics \( d \) and \( d' \) on a space \( X \) are called equivalent if there exists constants \( c, C > 0 \) such that for all \( x, y \in X \), \( cd(x, y) \leq d(x, y) \leq Cd(x, y) \).

Definition 2.2 Suppose \((X, d)\) is a metric space. A set \( O \subseteq X \) is called open if for every \( p \in O \) there is an \( \varepsilon > 0 \) such that the (open) \( \varepsilon \)-ball \( B_p(\varepsilon) = \{ x \in X : d(x, p) < \varepsilon \} \) is contained in \( O \), i.e. \( B_p(\varepsilon) \subseteq O \). A set \( F \subseteq X \) is called closed if \( X \setminus F \) is open.

Exercise 2.1 Show that the three metrics on \( \mathbb{R}^n \) discussed above are equivalent. Show pictorially the meaning of above inequalities in terms of nested \( \varepsilon \)-balls with respect to the different metrics. Conclude that the open sets in \( \mathbb{R}^n \) are the same, independent of the metric employed to define open balls.

On any set \( X \) the discrete metric may be defined by \( d(x, y) = 1 \) is \( x \neq y \) and \( d(x, y) = 0 \) if \( x = y \).

Given a metric \( d \) on a space \( X \) one may construct from it a bounded metric \( \tilde{d} \) by setting \( \tilde{d}(x, y) = d(x, y) \) if \( d(x, y) \leq 1 \) and \( \tilde{d}(x, y) = 1 \) else. Another useful bounded metric may be obtained by defining \( \tilde{d} = d/(1 + d) \).

Exercise 2.2 Verify that the discrete and bounded metric described above are indeed metrics.

Exercise 2.3 Let \( X \) be the set of all lines in the plane that pass through the origin. For lines \( \ell_1 \) and \( \ell_2 \) let \( d(\ell_1, \ell_2) \) to be the (smaller) angle between them. Verify that \( d \) is a metric on \( X \).

Exercise 2.4 Show that if \( d \) is the discrete metric on a set \( X \) then every subset \( S \subseteq X \) is both open and closed.

Exercise 2.5 Consider a metric space \((X, d)\) and the bounded metric \( \tilde{d} \) (or \( \bar{d} \)) constructed from \( d \) as above. Show that a subset \( S \subseteq X \) is open in \((X, \bar{d})\) if and only if it is open in \((X, d)\).
The notions of open and closed do not require an underlying metric structure. The following axioms allow for a generalization to spaces without a metric:

**Definition 2.3** A topology on a set $X$ is a collection $\mathcal{T}$ of subsets of $X$ that satisfies

(i) $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$,

(ii) $\mathcal{T}$ is closed under (arbitrary) unions, i.e., if $\{O_\alpha : \alpha \in A\} \subseteq \mathcal{T}$ then $\bigcup_{\alpha \in A} O_\alpha \in \mathcal{T}$, and

(iii) $\mathcal{T}$ is closed under finite intersections, i.e., if $O_k \in \mathcal{T}$, $k = 1, 2, \ldots, n$, then $\bigcap_{k=1}^n O_k \in \mathcal{T}$.

A subset $O \subset X$ is called open if $O \in \mathcal{T}$. A subset $F \subset X$ is called closed if $X \setminus F \in \mathcal{T}$.

A topological space is a pair $(X, \mathcal{T})$ where $X$ is a set and $\mathcal{T}$ is a topology on $X$.

Note that an infinite intersection of open sets is not required to be open. The standard example is the real line with the usual topology and $O_k = \left(-\frac{1}{k}, \frac{1}{k}\right)$. Clearly $\bigcap_{k=1}^\infty O_k = \{0\}$ which is not open (in the standard topology).

One commonly uses the term (open) neighborhood of $p \in X$ for an open set which contains $p$.

While technically a topological space is a pair $(X, \mathcal{T})$, one often refers to $X$ alone as a topological space. In such cases it is usually understood from the context which topology $\mathcal{T}$ on $X$ is meant.

Commonly one specifies a topology by describing a smaller set of basic open sets.

**Definition 2.4** If $X$ is a set, a collection $\mathcal{B}$ of subsets of $X$ is a basis for a topology on $X$ if it satisfies

(i) For every $x \in X$ there exists $B \in \mathcal{B}$ such that $x \in B$.

(ii) For all $B_1, B_2 \in \mathcal{B}$ and all $x \in B_1 \cap B_2$ there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

The topology generated by $\mathcal{B}$ consists of all subsets $O \subseteq X$ such that for every $x \in O$, there exists a set $B \in \mathcal{B}$ such that $x \in B \subseteq O$.

**Exercise 2.6** Show that if $Y \subseteq X$ is a subset of a topological space $(X, \mathcal{T})$ then the collection $\mathcal{T}' = \{O \cap Y : O \in \mathcal{T}\}$ defines a topology on $Y$. This topology is called the subspace topology.

**Exercise 2.7** Suppose that $(X, d)$ is a metric space. Show that the collection of open sets (in the sense of open in a metric space) defines a topology on $X$. (This topology is called the metric topology on $(X, d)$.)

**Definition 2.5** A topological space $(X, \mathcal{T})$ is first countable if at every $x \in X$ it has a countable basis, that is, for every $x \in X$ there exists a countable collection $\{B_k^x \in \mathcal{T} : k \in \mathbb{Z}^+\}$ such that for every $O \in \mathcal{T}$, if $x \in O$ then there exists $k \in \mathbb{Z}$ such $x \in B_k^x \subseteq O$.

A topological space $(X, \mathcal{T})$ is second countable if it has a countable basis.

A topological space $(X, \mathcal{T})$ is separable if there exists a countable dense subset $Y \subseteq X$, i.e. a subset $Y \subseteq X$ such that for every $x \in X$ and every $O \in \mathcal{T}$, if $x \in O$ then $O \cap Y \neq \emptyset$.

**Example 2.1** Every metric space is first countable.

Every separable metric space is second countable.

Every Euclidean space $\mathbb{R}^n$ is second countable.

**Exercise 2.8** Prove the assertions made in example 2.1.
Definition 2.6 Suppose $X$ and $Y$ (or, more precisely $(X,T_X)$ and $(Y,T_Y)$) are topological spaces. A map $f: X \mapsto Y$ is called continuous if for every open set $O \subseteq Y$ the preimage $f^{-1}(O) \subseteq X$ is open (i.e. $O \in T_Y \implies f^{-1}(O) \in T_X$). (This is equivalent to $f^{-1}(F) \subseteq X$ closed for every $F \subseteq Y$ closed.) A map $f: X \mapsto Y$ is called open if for every open set $O \subseteq X$ the image $f(O) \subseteq Y$ is open. A map $f: X \mapsto Y$ is called closed if for every closed set $F \subseteq X$ the image $f(F) \subseteq Y$ is closed.

In the case that $T_X$ and $T_Y$ are the metric topologies associated with metrics $d_X$ and $d_Y$ on $X$ and $Y$, respectively, this notion of continuity agrees with the standard $\varepsilon$-$\delta$ characterization of continuity. A function $f: X \mapsto Y$ is continuous (as defined above) if and only if for every $p \in X$ and for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $q \in X$ if $d_X(p,q) < \delta$ then $d_Y(f(p), f(q)) < \delta$. Practically the notion of continuity captures the concept that small changes in the input of a function cause only small changes in the output.

Exercise 2.9 Consider the set $\mathbb{R}$ of real numbers with the usual topology $T_2$, with the indiscrete topology $T_1 = \{\emptyset, \mathbb{R}\}$, and with the discrete topology $T_3$ in which every subset of $\mathbb{R}$ is open. For each pair $(i,j)$ with $i,j = 1,2,3$ describe the set of continuous functions from $(\mathbb{R},T_i)$ to $(\mathbb{R},T_j)$. (Make a $3 \times 3$ table.) In particular, for which pairs is the identity function $id: x \mapsto x$ continuous? For which pair(s) are (only) the constant functions continuous, and for which pair(s) are all functions continuous?

Exercise 2.10 Verify the assertion that in metric spaces the standard $\varepsilon$-$\delta$ characterization of continuity agrees with the definition given above.
Definition 2.7
A map \( f : X \mapsto Y \) between topological spaces \( X \) and \( Y \) is called a homeomorphism if

(i) \( f \) is a bijection, i.e. one-to-one and onto,

(ii) \( f \) is continuous, and

(iii) \( f^{-1} \) is continuous (i.e. \( f \) is open).

Two spaces topological spaces \( X \) and \( Y \) are called homeomorphic if there exists a homeomorphism \( f \) that maps \( X \) onto \( Y \).

Do not confuse the term homeomorphism discussed here with homomorphism which refers to maps that preserve algebraic relationships as in \( f(p \cdot q) = f(p) \cdot f(q) \).

From a topological point of view homeomorphic spaces are basically considered as identical.

Exercise 2.11 Verify that the map \( f : (0,1) \mapsto \mathbb{R} \) defined by \( f : x \mapsto (1 - 2x)/(x(x - 1)) \) is a homeomorphism. (Calculate \( f' \) and consider \( \lim_{x \to 0^+} f(x) \) and \( \lim_{x \to 1^-} f(x) \).

Exercise 2.12 Verify that the map \( f : \mathbb{R}^2 \mapsto B_0^1(1) = \{x \in \mathbb{R}^2 : \|x\| < 1\} \) defined by \( f : x \mapsto x/(1 + \|x\|) \) is a homeomorphism.

While it appears intuitively clear that \( \mathbb{R}^n \) and \( \mathbb{R}^m \) are not homeomorphic for \( n \neq m \), the proof (for general \( m \) and \( n \)) is surprisingly hard – usually utilizing tools from algebraic topology. Note that there exists for example a function \( f : [0,1] \mapsto [0,1] \times [0,1] \) that is both continuous and onto. However, \( f \) cannot be a homeomorphism, and in particular it cannot also be one-to-one. For more details on the construction of such space-filling curves see e.g. Munkres, Topology, p. 271.

Definition 2.8 A subset \( A \subseteq X \) of a topological space \( X \) is called connected if whenever \( B, C \subseteq X \) are disjoint open sets such that \( B \cup C = A \) then \( A \subseteq B \) or \( A \subseteq C \) (i.e. \( A \cap B = \emptyset \) or \( A \cap C = \emptyset \)). If \( A \) is not connected then it is called disconnected.

Exercise 2.13 Suppose that \( f : X \mapsto Y \) is a continuous map. Show that if \( f \) is onto and \( X \) is connected then \( Y \) is connected.

Exercise 2.14 On a topological space \( X \) define the relation \( \sim \subseteq X \times X \) by \( x \sim y \) if there exists a connected subset \( C \subseteq X \) such that \( x \in C \) and \( y \in C \). Show that \( \sim \) is an equivalence relation. (The equivalence classes of this relation are called the (connected) components of the space \( X \).)

In general topological spaces one works with a number of different notions of connectedness. Here we only mention the following other notion, which is stronger than connectedness:

Definition 2.9 A subset \( A \subseteq X \) of a topological space \( X \) is called path-connected if whenever \( p, q \in A \) then there exist a continuous map \( \gamma : [0,1] \mapsto A \) such that \( \gamma(0) = p \) and \( \gamma(1) = q \).

Exercise 2.15 Show that every path-connected set is connected.

Exercise 2.16 Show that the closure of the topologist’s sine curve, i.e. the set \( \{(x, \sin(\frac{x}{x})) \in \mathbb{R}^2 : x \neq 0\} \cup \{0\} \times [-1,1] \) is connected but not path-connected.
One of the most common uses of connectedness is the argument that if a function \( f: X \mapsto \mathbb{R} \) is continuous and locally constant then it is constant provided the domain \( X \) is connected. Here, \textit{locally constant} means that every \( p \in X \) has an open neighbourhood \( U \) (containing \( p \)) such that the restriction of \( f \) to \( U \) is constant.

To clarify this argument, consider the function \( f: \mathbb{R} \setminus \{0\} \mapsto \mathbb{R} \) defined by \( f: x \mapsto 0 \) if \( x < 0 \) and \( f: x \mapsto 1 \) if \( x > 0 \). Clearly the derivative \( f' \equiv 0 \) vanishes identically, but \( f \) is not constant. Of course, the key is that the domain is not connected. Consequently, the vanishing of the derivative only assures that \( f \) is locally constant. It does not assure that \( f \) is constant.

Arguably the most important topological concept for us is compactness. One may think of it as an outgrowth of the desire to generalize, or to get to the underlying foundation of the important theorem that every continuous function \( f: [a, b] \mapsto \mathbb{R} \) defined on a closed bounded interval attains its minimum and its maximum, i.e. there exist points \( x_1, x_2 \in [a, b] \) such that for all \( x \in [a, b] \), \( f(x_1) \leq f(x) \leq f(x_2) \). It is well-known from calculus that this assertion may fail if either of the closedness or boundedness hypotheses is omitted. The closedness requirement naturally generalizes to general topological spaces, but the boundedness does not. For example, if \( \mathbb{R}^n \) is equipped with the bounded metric \( d = d_2/(1 + d_2) \) (where \( d_2 \) is the standard Euclidean metric), then \( (\mathbb{R}^n, d) \) still has the same topology as \( (\mathbb{R}^n, d_2) \) yet while \( K = \mathbb{R}^n \) is closed and bounded in \( (\mathbb{R}^n, d) \), it is not in \( (\mathbb{R}^n, d_2) \). Many different generalizations have been proposed to generalize the basic idea of “closed and bounded” which is so useful in \( \mathbb{R}^n \) (with its usual metric).

Any introductory course in point-set topology will discuss such different notions of compactness. It was not until quite late into the 20th century that the following notion finally crystallized, and it became clear that it captures the fundamental features of the desired properties.

**Definition 2.10** A subset \( K \subseteq X \) of a topological space \( X \) is called compact if every open cover of \( K \) has a finite subcover, i.e. if \( \{O_\alpha \subseteq X : \alpha \in A\} \) is a collection of open sets such that \( K \subseteq \bigcup_{\alpha \in A} O_\alpha \) then there exists a finite subcollection \( \{O_{\alpha_j} : j = 1, 2, \ldots, n\} \) such that \( K \subseteq \bigcup_{j=1}^n O_{\alpha_j} \).

The Heine-Borel theorem asserts that every bounded closed interval in \( \mathbb{R} \) is compact. Its proof may be found in any advanced calculus text.

**Exercise 2.17** Prove that if \( f: X \mapsto Y \) is continuous and \( K \subseteq X \) is compact then the image \( f(K) \subseteq Y \) is compact.

In the case of \( Y = \mathbb{R} \) this implies that there exist points \( p, q \in X \) at which \( f \) attains its global minimum and global maximum, i.e. such that \( f(p) \leq f(x) \leq f(q) \) for all \( x \in X \).

**Definition 2.11** A sequence \( \{a_k\}_{k \in \mathbb{N}} \subseteq X \) in a topological space \( X \) is said to converge if there exists \( \bar{x} \in X \) such that for every open set \( O \subseteq K \) containing \( \bar{x} \) there exist a finite natural number \( N \) such that \( a_n \in O \) for all \( n > N \).

**Exercise 2.18** Suppose \( \{x_k\}_{k \in \mathbb{N}} \) is an infinite sequence with values in a compact topological space \( K \). Show that \( \{x_k\}_{k \in \mathbb{N}} \) has an accumulation point \( \bar{x} \in K \). If, in addition, \( K \) is first countable then there exists a converging subsequence \( \{x_{k_j}\}_{j \in \mathbb{N}} \).
Finally we mention a few *separation axioms* which on occasion are used as essential hypotheses in differential geometry.

**Definition 2.12**

- A topological space $X$ is called a Hausdorff space if for every pair of distinct points $p, q \in X$ there exist disjoint open sets $U$ and $V$ such that $p \in U$ and $q \in V$.

- A Hausdorff space $X$ is called completely regular if one-point sets are closed in $X$ and if for every point $p \in X$ and every closed set $F \subseteq X$ not containing $p$ there exists a continuous function $f: X \mapsto \mathbb{R}$ such that $f(p) = 0$ and $f(x) = 1$ for every $x \in F$.

- A Hausdorff space $X$ is called normal if one-point sets are closed in $X$ and if for every pair of disjoint closed sets $F_1, F_2 \subseteq X$ there exists disjoint open sets $U$ and $V$ such that $F_1 \subseteq U$ and $F_2 \subseteq V$.

These notions become useful when patching together local results, e.g. obtained in one coordinate chart at a time. This will be made precise when discussing partitions of unity in a subsequent section.

For the sake of completeness we here also give the following two technical definitions:

**Definition 2.13** A map $f: X \mapsto Y$ is called proper if for every compact set $K \subseteq Y$ the preimage $f^{-1}(K) \subseteq X$ is compact.

**Definition 2.14** A subset $S \subseteq X$ of a Hausdorff space $X$ is called paracompact if every open cover of $S$ has a locally finite open refinement. This means that if $U_\alpha, \alpha \in A$ are open sets such that $S \subseteq \bigcup_{\alpha \in A} U_\alpha$ then there exist a collection of open sets $V_\beta, \beta \in B$ such that

- For every $\beta \in B$ there exists an $\alpha \in A$ such that $V_\beta \subseteq U_\alpha$,

- $S \subseteq \bigcup_{\beta \in B} V_\beta$, and

- every $p \in S$ has an open neighbourhood $W$ which intersects only a finite number of the sets $V_\beta, \beta \in B$.

Paracompactness is very close to metrizability, (indeed, metrizability is equivalent to paracompactness and local metrizability). Thus many authors use paracompactness as a basic requirement when defining manifolds.
2.3 Local coordinate charts

This section defines the concept of a topological manifold which is to serve as a spring board for the subsequent definition of a differentiable manifold. The main focus is on the concept of local coordinate charts.

**Definition 2.15** A topological manifold $M$ is a metric space $(M, d)$ such that for every $p \in M$ there exist $n \in \mathbb{N}$, an open set $U \subseteq M$ containing $p$ and a homeomorphism $u: U \rightarrow \mathbb{R}^n$.

A few remarks:

- The metric $d$ plays little role in the future – what is needed is really only a reasonably nice topological space. Metric, or more accurately, metrizable spaces just happen to have about the right set of properties needed throughout the standard development.

- The statement that “$U$ is homeomorphic to $\mathbb{R}^n$” may be replaced by “$U$ is homeomorphic to an open subset of $\mathbb{R}^n$.”

- If $M$ is connected, then $n$ is constant and is called the dimension of the manifold $M$.

- The functions $u^i = x^i \circ u: U \rightarrow \mathbb{R}$ are called local coordinates. We use $x^i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \ldots, n$ to denote the standard coordinate functions, defined by $x^i: (a_1, \ldots, a_n) \mapsto a_i$ for every $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. The pair $(u, U)$ is called a local coordinate chart about $p$. Note that if $(u, U)$ is a local coordinate chart about $p$ then $(\tilde{u}, U)$ defined by $\tilde{u}: q \mapsto u(q) - u(p)$ for $q \in U$ is a local coordinate chart about $p$ such that $\tilde{u}(p) = 0$. This is usually written as $\tilde{u}: (U, p) \mapsto (\mathbb{R}^n, 0)$.

If $(u, U)$ and $v, V$ are coordinate charts about $p \in M$ (i.e. in particular $p \in U \cap V$, then

$$v \circ u^{-1}: u(U \cap V) \rightarrow v(U \cap V)$$

is a homeomorphism between open subsets of $\mathbb{R}^n$.

We continue with a short list of examples of manifolds and coordinate charts.

- For any $n \geq 0$ the $n$-dimensional Euclidean space $\mathbb{R}^n$ is an $n$-dimensional manifold with the single coordinate chart $(\text{id}, \mathbb{R}^n)$. 

---

![Diagram of a manifold with local coordinate charts](image-url)

If $(u, U)$ and $v, V$ are coordinate charts about $p \in M$ (i.e. in particular $p \in U \cap V$, then

$$v \circ u^{-1}: u(U \cap V) \rightarrow v(U \cap V)$$

is a homeomorphism between open subsets of $\mathbb{R}^n$.

We continue with a short list of examples of manifolds and coordinate charts.

- For any $n \geq 0$ the $n$-dimensional Euclidean space $\mathbb{R}^n$ is an $n$-dimensional manifold with the single coordinate chart $(\text{id}, \mathbb{R}^n)$. 

---

![Diagram of a manifold with local coordinate charts](image-url)

If $(u, U)$ and $v, V$ are coordinate charts about $p \in M$ (i.e. in particular $p \in U \cap V$, then

$$v \circ u^{-1}: u(U \cap V) \rightarrow v(U \cap V)$$

is a homeomorphism between open subsets of $\mathbb{R}^n$.

We continue with a short list of examples of manifolds and coordinate charts.

- For any $n \geq 0$ the $n$-dimensional Euclidean space $\mathbb{R}^n$ is an $n$-dimensional manifold with the single coordinate chart $(\text{id}, \mathbb{R}^n)$.
• The open ball $B_p(r) = \{ x \in \mathbb{R}^n : \|x-p\| < r \}$ of radius $r$ about $p \in \mathbb{R}^n$ is an $n$-dimensional manifold with a single chart given by $U = B_p(r)$ and $u(x) = (x - p)/(r - \|x - p\|)$.

**Exercise 2.19** Verify that the inverse is given by $u^{-1}(y) = p + ry/(1 + \|y\|)$.

• Every open subset of an $n$-dimensional manifold is itself an $n$-dimensional manifold.

• Identify the space $M_{m,n}(\mathbb{R})$ of $m$-by-$n$ matrices with real entries with the space $\mathbb{R}^{mn}$. E.g. in the $2 \times 2$-case simply identify

$$u: \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

Note that with this identification $M_{m,n}(\mathbb{R})$ inherits a natural metric structure. Clearly this shows that the entire space $M_{m,n}(\mathbb{R})$ is an $nm$-dimensional manifold.

Of more interest are various subspaces of $M_{m,n}(\mathbb{R})$. Typical examples are the general linear group $GL(n, \mathbb{R}) = \{ A \in M_{m,n}(\mathbb{R}) : \det A \neq 0 \}$ and its subset of orientation preserving nonsingular matrices $GL^+(n, \mathbb{R}) = \{ A \in M_{m,n}(\mathbb{R}) : \det A > 0 \}$. Both are $n^2$-dimensional manifolds. The argument uses that det is a polynomial function in the entries of the matrix, and hence it is continuous. Consequently, the preimages $\det^{-1}(\mathbb{R} \setminus \{0\})$ and $\det^{-1}(0, \infty)$ of open sets are open subsets of $M_{m,n}(\mathbb{R})$.

Further examples are the special linear groups $SL(n, \mathbb{R}) = \{ A \in M_{m,n}(\mathbb{R}) : \det A = 1 \}$, the orthogonal groups $O(n) = \{ A \in M_{m,n}(\mathbb{R}) : A^T A = I \}$, and the special orthogonal groups $SO(n) = O(n) \cap SL(n, \mathbb{R})$. Unlike the previous examples – which are open subsets – these manifolds are defined by closed conditions (i.e. “=” as opposed to “≠”, “<” or “>”). We return to these examples later when tools from differential calculus on manifolds will make it easy to establish when such subsets defined by closed conditions give rise to manifolds.

• The $m$-sphere $S^m(r) = \{ x \in \mathbb{R}^{m+1} : \|x\| = r \}$ is a $m$-manifold. The case $m = 0$ is special with $S^0(r) = \{ -r, r \} \subseteq \mathbb{R}$ consisting of only two points, and thus being disconnected. The 1-sphere is a circle, and one needs at least two charts, e.g. $U = S^1(r) \setminus \{-r, 0\}$ and $V = S^1(r) \setminus \{r, 0\}$. Use $u(x, y) = \text{atan}2(x, y)$ taking values in $(-\pi, \pi)$ and $r(x, y) = \text{atan}2(x, y)$ taking values in $(0, 2\pi)$. Note that the local coordinates agree essentially with the angle of polar coordinates.

**Exercise 2.20** Construct a collection of charts for the 2-sphere $S^2(r)$ by explicitly adapting the spherical coordinates $u(x, y, z) = (\theta(x, y, z), \phi(x, y, z)) \in (0, 2\pi) \times (0, \pi)$ to different subsets resulting from different “cuts”. What are the minimal number of cuts, and the minimal number of charts needed to cover $S^2$?

A different set of coordinates, that is particularly useful for higher dimensional spheres is based on stereographic projections: Consider the two subsets $U_\pm = S^m(r) \setminus \{0, 0, \ldots, \pm r\}$ and define the maps $u_\pm : U_\pm \mapsto \mathbb{R}^m$ by

$$u_\pm(x) = \frac{2r}{r \mp x_{m+1}}(x_1, \ldots x_m)$$  \hspace{1cm} (9)

Graphically $(u_\pm(x), \mp r) \in \mathbb{R}^{m+1}$ is the point where the hyperplane $x_{m+1} = \mp r$ intersects the line that passes through the point $x \in S^m(r)$ and through $(0, 0, \ldots, \pm r)$.
Exercise 2.21  For the inverse maps $u^{-1}_\pm: \mathbb{R}^m \hookrightarrow U_\pm \subset S^m(r)$, derive the formulae

$$x = u^{-1}_\pm(y) = \left(\frac{2r}{1 + \|y/2r\|^2} y_1 \frac{2r}{1 + \|y/2r\|^2} \cdots \frac{2r}{1 + \|y/2r\|^2} y_m, \mp r \cdot \frac{1 - \|y/2r\|^2}{1 + \|y/2r\|^2}\right) \quad (10)$$

Use this to obtain explicit formulae for the “transition maps” $u_\pm \circ u^{-1}_\pm: \mathbb{R}^m \hookrightarrow \mathbb{R}^m$.

What do these maps do graphically – e.g. which sets do they leave fixed?

What are the images of (special) lines and circles?

- If $M^m$ and $N^n$ are $m$- and $n$-dimensional manifolds, respectively, then the Cartesian product $M \times N$ is an $(m + n)$-dimensional manifold: Suppose $(p, q) \in M \times N$ and $(u, U)$ and $(v, V)$ are coordinate charts about $p$ and $q$, then $(u \times v, U \times V)$ is a coordinate chart about $(p, q)$ where $(u, v)(a, b) = (u(a), v(b)) \in \mathbb{R}^m \times \mathbb{R}^n$.

A typical example uses that the circle $S^1 = \{x \in \mathbb{R}^2: \|x\| = 1\}$ is a manifold to establish that the torus $T^2 = S^1 \times S^1$ is a 2-dimensional manifold.

- We briefly return to the real projective spaces, now illustrating coordinate charts. On $\mathbb{R}^{m+1} \setminus \{0\}$ define the equivalence relation $x \sim y$ if $x = \lambda y$ for some $\lambda \in \mathbb{R}$. The $m$-dimensional real projective space is defined as the quotient $\mathbb{P}^m = (\mathbb{R}^{m+1} \setminus \{0\}) / \sim$, i.e. a point $[x] \in \mathbb{P}^m$ is the equivalence class $[x] = \{y \in \mathbb{R}^{m+1} \setminus \{0\}: y \sim x\}$. Intuitively think of $\mathbb{P}^m$ as the space of all lines in $\mathbb{R}^{m+1}$ that pass through the origin, or as the $m$-sphere $S^m$ with antipodal points $x$ and $-x$ identified. More graphically, one may obtain $\mathbb{P}^2$ by sewing a disk to the (only one!) edge of a Möbius strip. For $j = 1, \ldots, m$ consider the sets $U_j = \{[x] \in \mathbb{P}^m: x_j \neq 0\}$ and coordinates maps (homogeneous coordinates)

$$u_j([x]) = \left(\frac{x_1}{x_j}, \ldots, \frac{x_{j-1}}{x_j}, \frac{x_{j+1}}{x_j}, \ldots, \frac{x_{m+1}}{x_j}\right) \quad (11)$$

It is straightforward to verify that the value of $u_j([x])$ does not depend on the choice of the representative $x \in [x]$. The inverse is given by $u^{-1}_j(y_1, \ldots, y_m) = [y_1, \ldots, y_{i-1}, 1, y_{i+1}, \ldots, y_m]$.

A complete discussion of these coordinate maps (they are supposed to be homeomorphisms) is straightforward but technical, in terms of the quotient topology. For an introductory discussion of quotient maps, and quotient topologies see e.g. Munkres (Topology, a first course", p.134). A main issue is to assure that the quotient topology is not pathological. Just as a reference, a surjective map $f: X \to Y$ is called a quotient map if $O \subset Y$ is open if and only if $f^{-1}(O) \subset X$ is open. For any map $f: X \to A$ from a topological space $X$ to a set $A$ there is exactly one topology on $A$, called the quotient topology, such that $f$ is a quotient map. In the case that $A$ is a set of equivalence classes on $X$, $A$ with this topology is called a quotient space of $X$.

- Two-dimensional surfaces in $\mathbb{R}^3$, or more generally $m$-dimensional hypersurfaces in $\mathbb{R}^{m+1}$ are some of familiar manifolds. Clearly every graph $\{(x, f(x)): x \in \mathbb{R}^m\} \subset \mathbb{R}^{m+1}$ of any continuous function $f: \mathbb{R}^m \to \mathbb{R}$ is a manifold with a single chart $u: (x, f(x)) \to x \in \mathbb{R}^m$.

More interesting are hypersurfaces that arise as preimages ("zero-sets") $M = F^{-1}(\{0\}) = \{x \in \mathbb{R}^{m+1}: F(x) = 0\}$ of functions $F: \mathbb{R}^{m+1} \to \mathbb{R}$, or that are given by parameterizations $F: \mathbb{R}^m \to \mathbb{R}^{m+1}$ and $M = F(\mathbb{R}^m)$.
More generally, if $M$ and $N$ are manifolds and $\Phi: M \hookrightarrow N$ then $\Phi(M) \subseteq N$ may be a manifold. Similarly, if $P \subseteq N$ is a submanifold, then $\Phi^{-1}(P) \subseteq M$ might be a submanifold of $M$. To avoid unnecessary duplication and difficulties we shall discuss these constructions only in the setting of differentiable manifolds, in a subsequent section. Here we only briefly mention two examples: Let $f, g: \mathbb{R}^2 \hookrightarrow \mathbb{R}$ be defined by $f: (x, y) \mapsto x^2 + y^2 - 1$ and $g: (x, y) \mapsto xy$. Then $f^{-1}(0)$ is the 1-sphere, while $g^{-1}(0)$ is not a manifold. The standard criterion that distinguishes these examples relies on derivatives: While $(Df)$ never vanishes where $f$ vanishes, $(Dg)$ and $g$ have common zeros – which are potentially troublesome points.

- There is a very rich world of complex manifolds – but we will not have the opportunity to explore this in any depth in this course. Within the frame of this section – coordinate charts for topological manifolds – complex manifolds do not offer any new features. But in the framework of differentiable manifolds, the much richer structure of complex differentiability opens completely new worlds, far beyond our course . . .
2.4 Differentiation: Notation and review

This section fixes some notation for partial derivatives of maps between Euclidean spaces, and contrasts this with a coordinate-free description of differentiation.

Let \( e_i = (0, 0, \ldots, 0, 1, 0 \ldots 0)^T \in \mathbb{R}^n \) denote the standard \( i \)-th basis vector. For a function \( f: U \rightarrow \mathbb{R} \) defined on an open subset \( U \subseteq \mathbb{R}^n \) and \( a \in U \) the \( i \)-the partial derivative of \( f \) at \( a \) is defined (and denoted) by

\[
(D_i f)(a) = \lim_{h \to 0} \frac{1}{h} (f(a + he_i) - f(a))
\]

provided the limit exists. To denote different degrees of regularity we use the following notation:

\[
\begin{align*}
\text{If} & \quad f \in C^0(U), \quad \text{f is continuous on } U \quad (\text{i.e., } f \text{ is continuous at every } p \in U) \\
\text{If} & \quad f \in C^r(U), \quad \text{all partial derivatives of up to order } r \text{ of } f \text{ exist and are continuous on } U, \\
\text{If} & \quad f \in C^\infty(U), \quad \text{all partial derivatives of all orders of } f \text{ exist and are continuous on } U, \\
\text{If} & \quad f \in C^\omega(U), \quad \text{f is real analytic on } U \quad (f \in C^\infty(U) \text{ and } f \text{ agrees (locally) with its Taylor series}).
\end{align*}
\]

For a function \( f: A \rightarrow \mathbb{R} \) defined on a set \( A \subseteq \mathbb{R}^m \) we say \( f \in C^\alpha(A) \) if there exist an open set \( O \subseteq \mathbb{R}^m \) such that \( A \subseteq O \) and an extension \( \bar{f} \) of \( f \) to \( O \) (i.e., \( \bar{f} |_A = f \)), and \( \bar{f} \in C^\alpha(O) \).

If \( f: A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m \), then \( f \in C^\alpha(A, \mathbb{R}^m) \) if each coordinate function \( f^k = x^k \circ f \in C^\alpha(A) \).

If \( f \in C^1(\mathbb{R}^n, \mathbb{R}^m) \) then the Jacobian matrix is

\[
(Df) = (D_j f^i)_{i=1, \ldots, m, \ j=1, \ldots, n} = \begin{pmatrix} D_1 f^1 & \ldots & D_n f^1 \\
\vdots & \ddots & \vdots \\
D_1 f^m & \ldots & D_n f^m \end{pmatrix}
\]

For convenience we identify the space \( M_{m,n}(\mathbb{R}) \) of real \( m \times n \) matrices with \( \mathbb{R}^{mn} \). Thus, if \( f \in C^r(\mathbb{R}^n, \mathbb{R}^m) \), then \( (Df) \in C^{r-1}(\mathbb{R}^n, \mathbb{R}^{mn}) \), and \( (D^k f) \in C^{r-k}(\mathbb{R}^n, \mathbb{R}^{mn^k}) \).

The chain rule asserts that if \( U \subseteq \mathbb{R}^m \) and \( V \subseteq \mathbb{R}^n \) are open sets, \( g \in C^r(U, \mathbb{R}^n) \) and \( f \in C^r(V, \mathbb{R}^p) \) with \( r \geq 1 \), \( g(U) \subseteq V \) and \( a \in U \), then \( f \circ g \) is differentiable at \( a \) and \( D(f \circ g)(a) = (Df)(g(a)) \cdot (Dg)(a) \) (matrix-multiplication).

**Theorem 2.1 (Implicit function theorem)** Suppose \( U \subseteq \mathbb{R}^{m+n} \) is open, \( (a, b) \in U \) and \( f \in C^r(U, \mathbb{R}^m) \) with \( r \geq 1 \), and \( f(a, b) = 0 \). If the matrix of partial derivatives \( (D_{n+j} f^i)_{i,j=1, \ldots, m} \) is nonsingular then there exist open sets \( V \subseteq \mathbb{R}^n \) and \( W \subseteq \mathbb{R}^m \) with \( a \in V \) and \( b \in W \) and a unique function \( g: V \rightarrow W \) such that \( f(x, g(x)) = 0 \) for all \( x \in V \). Moreover, \( g \in C^r(V, W) \).

Differentiability (as opposed to mere existence of partial derivatives) may be described in a coordinate-free way. Consider finite dimensional normed linear spaces \( V, W \) and let \( U \subseteq V \) be open. Recall, a norm is a map \( \| \cdot \|: V \rightarrow \mathbb{R} \) such that \( \|v\| \geq 0 \) for all \( v \in V \), \( \|v\| = 0 \) if and only of \( v = 0 \), \( \|\lambda v\| = |\lambda| \|v\| \) for all \( v \in V \) and \( \lambda \in \mathbb{R} \), and \( \|v + w\| \leq \|v\| + \|w\| \) for all \( v, w \in V \). As finite dimensional linear spaces \( V \) and \( W \) are isomorphic to some Euclidean spaces \( \mathbb{R}^n \) and \( \mathbb{R}^m \), but here the objective is to not fix any bases.

**Definition 2.16** A map \( f: U \rightarrow W \) is differentiable at \( a \in U \) if there exists a linear map \( L = L_{a,f} \in \text{Hom}(V, W) \) such that \( f(a + h) = f(a) + L(h) + o(\|h\|). \) (This means that there exists a map \( \eta: U \rightarrow W \) (depending on \( a \) and \( f \)) such that \( f(a + h) = f(a) + L(h) + \|h\| \eta(h) \) and \( \|\eta(h)\| \rightarrow 0 \) at \( \|h\| \rightarrow 0 \).)
Exercise 2.22  Show that if \( f \) is differentiable at \( a \) then the linear map of the preceding definition is uniquely determined. (Suppose there are two such linear maps. Show that their difference satisfies \( L(h) - M(h) = o(\|h\|) \). Therefore it is justified to talk about the derivative of \( f \) at \( a \) and we will use the notation \( f'(a) \).

A map \( f \) is called differentiable on an open set \( U \) if \( f \) is differentiable at all \( a \in U \).

Note that if \( f: U \subseteq V \rightarrow W \) then \( f': U \mapsto \text{Hom}(V, W) \). But the space \( \text{Hom}(V, W) \) of linear maps from \( V \) to \( W \) is itself a linear space (finite dimensional, normed – with the induced norm \( \|L\| = \max \{\|Lv\|: \|v\| = 1\} \)). Hence one may naturally define higher order derivatives as linear maps \( f''': U \mapsto \text{Hom}(V, \text{Hom}(V, W)) \approx \text{Hom}(V \otimes V, W) \) and inductively \( f^{(k)}: U \mapsto \text{Hom}(\bigotimes_{i=1}^{k} V, W) \).

For comparison, if \( f: \mathbb{R}^n \mapsto \mathbb{R}^m \) then \( (Df) \) is an \( (m \times n) \) matrix, and, naively, \( (D^2 f) \) is some sort of \( (m \times n \times n) \) object which takes three inputs, a point where it is evaluated and two vectors, and its output is an \( m \)-vector.

We summarize a few basic properties

(i) If \( f \) is differentiable then \( f \) is continuous.

(ii) If \( f \) and \( g \) are differentiable and \( \lambda \in \mathbb{R} \) then \( (f + g)' = f' + g' \) and \( (\lambda f)' = \lambda f' \).

(iii) If \( f \) is constant then \( f' \equiv 0 \).

(iv) If \( L \in \text{Hom}(V, W), \ b \in W \) and \( f = L + b \) (i.e. \( f: v \mapsto L(v) + b \)) then \( f' = L \).

(v) (Chain-rule). Let \( V_1, V_2, V_3 \) be normed linear spaces. Suppose \( U_1 \subseteq V_1 \) and \( U_2 \subseteq V_2 \) are open, \( g: U_1 \mapsto V_2 \), \( f: U_2 \mapsto V_3 \), and \( g(U_1) \subseteq U_2 \). If \( f \) and \( g \) are differentiable, then so is \( f \circ g \) and \( (f \circ g)' = (f' \circ g) \cdot g' \).

This notation is to be interpreted as follows: For \( p \in U_1 \) and \( v \in V_1 \), \( g'(p)(v) \in \text{Hom}(V_1, V_2) \) and hence \( g'(p)(v) \in V_2 \). Similarly, \( f'(g(p)) \in \text{Hom}(V_2, V_3) \) and hence \( f'(g(p))(g'(v)) \in V_3 \).

This matches with \( (f \circ g)'(p) \in \text{Hom}(V_1, V_3) \) and hence \( (f \circ g)'(p)(v) \in V_3 \).

2.5 Differentiable structures

In general, manifolds do not have a linear, not even an additive structure. Thus expressions reminiscent of \( f'(x) = \lim_{h \to 0} \frac{1}{h} (f(x + h) - f(x)) \) are meaningless for maps \( f: M \mapsto N \) between manifolds (unless one considers points in \( M \) as distributions on the algebra of smooth functions).

A natural way to define a notion of differentiability on manifolds is to utilize coordinate charts \( (u, U) \) on \( M \) and \( (v, V) \) on \( N \) to relate \( f: U \mapsto V \) to the map \( v \circ f \circ u^{-1} \) between Euclidean spaces. The main concern is to ensure that any so-defined notion of differentiation on a manifold does not depend on the particular choice of coordinates. This leads naturally to the concept of differentiable structures.

Definition 2.17 Two charts \((u_1, U_1)\) and \((u_2, U_2)\) on a manifold \( M \) are \( C^r \)-related \((r = 1, 2, \ldots, \infty, \omega)\) if the maps \( u_2 \circ u_1^{-1} \) and \( u_1 \circ u_2^{-1} \) are \( C^r \)-maps as maps between Euclidean spaces) on their respective domains \( u_1(U_1 \cap U_2) \) and \( u_2(U_1 \cap U_2) \).

Definition 2.18 A \( C^r \)-differentiable structure on a manifold \( M \) is a maximal atlas \( \mathcal{D} \), that is, a maximal collection of coordinate charts that covers \( M \) and which is such that any two charts \((u_1, U_1), (u_2, U_2) \in \mathcal{D} \) are \( C^r \)-related. A manifold \( M \) together with a \( C^r \)-differentiable structure \( \mathcal{D} \) is called a \( C^r \)-manifold, or simply, a differentiable manifold (if \( r \) is understood, usually \( r = \infty \)).
**Definition 2.19** Two $C^r$-manifolds $(M, \mathcal{D})$ and $(N, \mathcal{D}')$ are called diffeomorphic if there exists a bijection $\Phi: M \mapsto N$ such that $(v, V) \in \mathcal{D}'$ if and only if $(v \circ \Phi, \Phi^{-1}(V)) \in \mathcal{D}$.

The map $\Phi$ is called a diffeomorphism.

It is easy to see that every diffeomorphism must be continuous. Since the inverse $\Phi^{-1}$ is automatically a diffeomorphism, $\Phi$ is automatically a homeomorphism. However, manifolds may be homeomorphic without being diffeomorphic, see below.

**Proposition 2.2** Every $C^r$-atlas is contained in a unique $C^r$-differentiable structure.

**Proof.** Let $\mathcal{U} = \{(u_\alpha, U_\alpha): \alpha \in A\}$ be a $C^r$-atlas for a manifold $M$ – i.e. any two charts $(u_\alpha, U_\alpha), (u_\beta, U_\beta) \in \mathcal{U}$ are $C^r$-related and $M \subseteq \bigcup_{\alpha \in A} U_\alpha$. Define $\mathcal{D}$ to be the collection of all coordinate charts $(v_\alpha, V_\alpha)$ on $M$ which (each) are $C^r$-related to every $(u_\alpha, U_\alpha) \in \mathcal{U}$.

Maximality of $\mathcal{D}$ is clear. To verify that any two charts $(v_\alpha, V_\alpha), (v_\beta, V_\beta) \in \mathcal{D}$ are $C^r$-related it suffices to show that $v_\beta \circ v_\alpha^{-1} : v_\alpha(V_\alpha \cap V_\beta) \mapsto \mathbb{R}^m$ is locally $C^r$. Thus suppose that $x \in v_\alpha(V_\alpha \cap V_\beta) \subseteq \mathbb{R}^m$ and let $p = v_\alpha^{-1}(x) \in M$. Since $\mathcal{U}$ is an atlas of $M$, it contains a chart $(u, U) \in \mathcal{U}$ about $p$. On the set $v_\alpha(U \cap V_\alpha \cap V_\beta)$ we may write

$$v_\beta \circ v_\alpha^{-1} = (v_\beta \circ u^{-1}) \circ (u \circ v_\alpha^{-1})$$

as a composition of $C^r$-maps – and hence $v_\beta \circ v_\alpha^{-1}$ is locally $C^r$ on $v_\alpha(U \cap V_\alpha \cap V_\beta)$. Since the latter set is open in $\mathbb{R}^m$, $v_\beta \circ v_\alpha^{-1}$ is $C^r$.

Regarding uniqueness, suppose $\mathcal{D}'$ is any $C^r$ differentiable structure containing $\mathcal{U}$. Then by definition every $(v, V) \in \mathcal{D}'$ is $C^r$-related to every $(u, U) \in \mathcal{U}$, and consequently $(v, V) \in \mathcal{D}$, i.e. $\mathcal{D}' \subseteq \mathcal{D}$. Since a differentiable structure is maximal by definition, also $\mathcal{D} \subseteq \mathcal{D}'$, i.e. $\mathcal{D} = \mathcal{D}'$. ■

**Definition 2.20** Let $(M, \mathcal{D})$ and $(N, \mathcal{D}')$ be differentiable manifolds of class $C^r$ and $C^s$, respectively, and $k \leq \min(r, s)$. A map $\Phi: M \mapsto N$ is called differentiable of class $C^k$ if for any charts $(u, U) \in \mathcal{D}$ and $(v, V) \in \mathcal{D}'$ the map $v \circ \Phi \circ u^{-1}$ is of class $C^k$ on its domain.

- If $M = \mathbb{R}^m$ and $N = \mathbb{R}^n$ (with the usual differentiable structures) then $\Phi: M \mapsto N$ is differentiable in the usual sense.
- A map $\Phi: M \mapsto \mathbb{R}^n$ is differentiable if and only if each coordinate function $\Phi^k = x^k \circ \Phi$ is differentiable.
- Coordinate maps $u$ are diffeomorphisms from $U$ to $u(U)$.
- A map $\Phi: M \mapsto \mathbb{R}^n$ is a diffeomorphism if and only if it is bijective, differentiable, and its inverse is differentiable.

**Proposition 2.3** Every differentiable map $\Phi: M \mapsto N$ between manifolds is continuous.

**Proof.** Let $W \subseteq N$ be open. We will show that $\Phi^{-1}(W) \subseteq M$ is open. Let $\mathcal{U} = \{(u_\alpha, U_\alpha): \alpha \in A\}$ and $\mathcal{V} = \{(v_\beta, V_\beta): \beta \in B\}$ be atlases for $M$ and $N$, respectively. Since $W = \bigcup_{\beta \in B} (V_\beta \cap W)$ we have $\Phi^{-1}(W) = \bigcup_{\beta \in B} \Phi^{-1}(V_\beta \cap W)$. Using that for each $\alpha \in A$, $u_\alpha$ is one-to-one, image under $u_\alpha$ distributes over intersections, and

$$u_\alpha(U_\alpha \cap \Phi^{-1}(W)) = u_\alpha(U_\alpha \cap \bigcup_{\beta \in B} \Phi^{-1}(V_\beta \cap W)) = \bigcup_{\beta \in B} u_\alpha(U_\alpha) \cap (u_\alpha \circ \Phi^{-1} \circ v_\beta^{-1})(v_\beta(V_\beta \cap W))$$

(14)
Since \( v_\beta(V_\beta \cap W) \subseteq \mathbb{R}^n \) is open for each \( \beta \), and as a differentiable map between Euclidean spaces \((v_\beta \circ \Phi \circ u_\alpha^{-1})\) is continuous for each \( \alpha, \beta \), each of the sets \((u_\alpha \circ \Phi^{-1} \circ v_\beta^{-1})(V_\beta \cap W)\) is open. Consequently, each \( u_\alpha(U_\alpha \cap \Phi^{-1}(W)) \subseteq \mathbb{R}^m \) is open, and thus each \( U_\alpha \cap \Phi^{-1}(W) \subseteq M \) is open. Finally, \( \Phi^{-1}(W) = (\bigcup_\alpha U_\alpha) \cap \Phi^{-1}(W) = \bigcup_\alpha (U_\alpha \cap \Phi^{-1}(W)) \) is a union of open sets, and thus is open.  

**Exercise 2.23** Show that the identity map \( \text{id}_M : M \mapsto M \) on a \( C^r \)-manifold \( M \) is a \( C^r \)-map. Suppose that \( M \), \( N \), and \( P \) are \( C^r \)-manifolds with \( A \subseteq M \) and \( B \subseteq N \). Show that if \( g \in C^r(A, N) \) and \( f \in C^r(B, P) \) then \( f \circ g \in C^r(A \cap g^{-1}(B), P) \).

**Exercise 2.24** Consider \( M = \mathbb{R} \) with charts \((u_k, U)\), \( k = 1, 2, 3 \) where \( U = \mathbb{R} \) and \( u_1(x) = x^3 \), \( u_2(x) = x \), and \( u_3(x) = x^{1/3} \). In each case provide an example of another chart \((v_k, V_k)\) that is contained in the unique \( C^1 \)-differentiable structure \( \mathcal{D}_k \) on \( M \) that contains \((u_k, U_k)\).

Explain why \( \mathcal{D}_k \) are (pairwise) different, give examples of charts contained in their intersections (or explain why the intersections are empty). Demonstrate that \((M, \mathcal{D}_k)\) are diffeomorphic.

**Exercise 2.25** (Continuation of exercise 2.24). Consider the maps \( \Phi_{i,j,k} : (M, \mathcal{D}_i) \mapsto (M, \mathcal{D}_j) \) defined by \( \Phi_{i,j,k}(x) = x^k \) for \( k = \frac{1}{9}, \frac{1}{3}, 1, 3, 9 \). Which of these maps are differentiable? Which maps are diffeomorphisms?

Note that every differentiable structure \( \mathcal{D} \) of class \( C^r \), \( r \geq 1 \) may be regarded as an atlas of class \( C^{r-1} \), and hence is contained in a unique differentiable structure \( \mathcal{D}' \) of class \( C^{r-1} \). Consequently the class of a manifold can be lowered at will, by adding new charts to an atlas.

More important is that every \( C^r \)-differentiable structure with \( r \geq 1 \) contains a \( C^\infty \) differentiable structure (see also below). Consequently one routinely restricts one’s attention to \( C^\infty \)-manifolds.

**Blanket hypothesis.** Unless otherwise stated, all manifolds and maps considered henceforth are assumed to be \( C^\infty \) manifolds and \( C^\infty \)-maps, respectively.

The following remarks are taken from my class-notes from UC Boulder in 1983 – they have not independently been verified (nor updated) ...

A (very hard) theorem by Whitney asserts that every \( C^1 \) differentiable structure contains a \( C^\infty \) structure. On the other hand, Kervaire has given an example of a 10 dimensional \( C^0 \) manifold which admits no differentiable structure (hard!).

The spheres \( S^n \) have unique differentiable structures for \( n \leq 6 \), but Milnor showed in 1958 that \( S^7 \) admits 28 non-diffeomorphic differentiable structures! However, for any \( n \) there is only a finite number of “diffeo-classes” on \( S^n \).

For \( n \neq 4 \) the Euclidean space \( \mathbb{R}^n \) has a unique differentiable structure, but \( \mathbb{R}^4 \) has at least 3 non-diffeomorphic differentiable structures (1982).
2.6 Partitions of unity

It is very common that one can easily construct objects locally, e.g. working on coordinate charts. Partitions of unity are a versatile tool to patch together such objects into a globally defined one. From a different point of view, partitions of unity demonstrate that there are plenty of $C^\infty$ functions on a differentiable manifold (as opposed to comparatively few $C^\omega$-functions).

We begin with some fundamental constructions in Euclidean spaces.

**Lemma 2.4** Let $m \geq 1$, $0 \leq a < b$ and $p \in \mathbb{R}^m$. Then there exists a map $k \in C^\infty(\mathbb{R}^m, \mathbb{R})$ such that $k(x) = 0$ for $\|x-p\| \geq b$, $k(x) = 1$ for $\|x-p\| \leq a$, and $0 < k(x) \leq 1$ for $\|x-p\| \leq b$.

**Proof.** Let $f(t) = \exp(-1/t)$ for $t > 0$ and $f(t) = 0$ else. Then $f \geq 0$ and $f \in C^\infty(\mathbb{R})$. Next define

$$g(t) = \frac{f(t)}{f(t) + f(b-t)} \text{ if } t > 0$$

and $g(t) = 0$ for $t \leq 0$. Then $g(t) = 1$ for all $t \geq b$ and $g'(t) > 0$ for $0 < t < b$. Define

$$h(t) = g\left(\frac{b(t+b)}{b-a}\right) \cdot g\left(\frac{b(t-b)}{b-a}\right).$$

and finally set $k(x) = h(\|x-p\|)$.

One may replace the balls $B_p(r)$ in the lemma by cubes $C^m_p(r) = \{x \in \mathbb{R}^m : |x_i - p_i| \leq r\}$ by taking $k(x) = h(x_1 - p_1) \ldots h(x_m - p_m)$.

**Exercise 2.26** Prove that $f \in C^\infty(\mathbb{R})$ as claimed in the preceding proof. (Use induction.)

**Proposition 2.5** Let $M^m$ be a $C^\infty$-manifold, $V \subseteq M$ open, and $K \subseteq V$ compact. Then there exists a function $\phi \in C^\infty(M, [0, 1])$, $\phi|_K \equiv 1$ and $\phi|_W \equiv 0$ for some open set $W \supseteq M - V$.

**Proof.** Use the compactness of $K$ to select charts $(u_i, U_i)$, $i = 1, \ldots N_1$ such that $K \subseteq \bigcup_{i=1}^{N_1} U_i$. Since $u_i(U_i \cap V) \subseteq \mathbb{R}^m$ is open (and w.l.o.g. nonempty), for every $y \in u_i(U_i \cap V)$ there exist $r_{y,i} > 0$ such $C_{y}(r_{y,i}) \subseteq u_i(U_i \cap V)$. The collection $\{u_i^{-1}(C_y(r_{y,i})) : i \leq N_1, u_i^{-1}(y) \in K\}$ is an open cover of $K$.

Choose a finite subcover $\{u_i^{-1}(C_{y_{i,j}}(r_j)) : i \leq N_1, j \leq N_2(i)\}$, (writing $r_{i,j}$ for $r_{y_{i,j},i}$). By the preceding proposition there exist functions $h_{i,j} : \mathbb{R}^m \mapsto [0, 1]$ such that $h_{i,j}(x) = 1$ for all $x \in C_{y_{i,j}}(\frac{3}{4}r_j)$ and $h_{i,j}(x) = 0$ for all $x \notin C_{y_{i,j}}(\frac{3}{4}r_j)$. Define $\phi_{i,j} = h_{i,j} \circ u_i$ on $U_i$ and extend to $M$ by setting $\phi_{i,j}(q) = 0$ for $q \notin U_i$. Combine these functions into

$$\phi(x) = 1 - \prod_{i,j} (1 - \phi_{i,j}(x))$$

and set

$$W = M \setminus \bigcup_{i,j} u_i^{-1}\left(C_{y_{i,j}}(\frac{3}{4}r_j)\right) \supseteq M \setminus V.$$

Note that $\phi|_K \equiv 1$ since $K \subseteq \bigcup_{i,j} u_i^{-1}(C_{y_{i,j}}(\frac{1}{4}r_j))$. Clearly $\phi|_W \equiv 0$.

The objective is to use these bump-functions to patch together local results. It is a natural to require that at any fixed point only a finite number of the local results are needed, or may be
selected. This is a place where our assumptions come into play that a manifold’s topology be reasonably nice. For example, every metric (metrizable) space is paracompact (Stone’s theorem), and hence normal. The following shrinking lemma is a direct consequence of these properties. The construction in its proof is a good exercise to practice working with paracompactness and normality, and a good check for understanding. The lemma itself plays a fundamental role in the desired partitions of unity which are used to patch together the local results.

Before proceeding with the shrinking lemma, we provide a few optional side-remarks.

Compactness may be characterized in the following way, which may seem unusual, but which lends itself a natural generalization: “A space K is compact if every open cover of X has a finite open refinement that covers X.” From here it is only a small step to paracompactness, which weakens “finite” to “locally finite” (and traditionally explicitly requires that the space is Hausdorff). According to Munkres (Topology, a first course): “The concept of paracompactness is one of the most useful generalizations of compactness that has been discovered in recent years. Particularly is it useful for applications in algebraic topology, differential geometry, . . . ”. In point set topology its close connection with metrizability is utilized.

The following simple example illustrates how the definition works. Consider the real line, $\mathbb{R}$, of which lends itself a natural generalization: “A space $K$ is compact if every open cover of $X$ has a finite open refinement that covers $X$.” From here it is only a small step to paracompactness, which weakens “finite” to “locally finite” (and traditionally explicitly requires that the space is Hausdorff). According to Munkres (Topology, a first course): “The concept of paracompactness is one of the most useful generalizations of compactness that has been discovered in recent years. Particularly is it useful for applications in algebraic topology, differential geometry, . . . ”. In point set topology its close connection with metrizability is utilized.

The following simple example illustrates how the definition works. Consider the real line, $\mathbb{R}$, (For example think of $U = \{(-\alpha, \alpha) : \alpha > 0\}$.) Using the compactness of the finite closed intervals $[n, n+1]$, $n \in \mathbb{Z}$, there exist for each $n$ a finite number of indices $\alpha_1^{(n)}, \ldots, \alpha_k^{(n)} \in A$ such that $[n, n+1] \subseteq \bigcup_{j=1}^{k(n)} U_{\alpha_j^{(n)}}$. Define $V_j^{(n)} = U_{\alpha_j^{(n)}} \cap (n-1, n+2)$. Then $\{V_j^{(n)} : n \in \mathbb{Z}, j \leq k(i)\}$ is the desired locally finite refinement that covers $\mathbb{R}$.

The following make implicitly use of some technical properties of manifolds: As an immediate consequence of coordinate charts being homeomorphisms onto $\mathbb{R}^n$, every manifold $M$ is locally compact, meaning that every point $p \in M$ has an open neighbourhood with compact closure. Moreover, as a metric (metrizable) space, every manifold is also paracompact and hence normal. Indeed, even stronger than paracompactness, every open cover $U$ of a manifold $M$ has a locally finite open refinement $V$ such that every $V \in V$ is diffeomorphic to $\mathbb{R}^n$. In addition, every connected manifold is $\sigma$-compact, meaning that it is a countable union of compact subsets (Compare Spival vol.I Ch. I, theorem 2 and App. A thm 1). Thus, every open cover of a $\sigma$-compact space contains a countable subcover.

**Proposition 2.6 (Shrinking lemma)** Let $U = \{U_\alpha : \alpha \in A\}$ be a locally finite open cover of a normal space $X$. Then there exists an open cover $V = \{V_\alpha : \alpha \in A\}$ such that $V_\alpha \subseteq \overline{V}_\alpha \subseteq U_\alpha$ for every $\alpha \in A$.

**Proof.** We consider ourselves here to the case of $X = M$ being a connected manifold, and hence $\sigma$-compact. (else apply the following argument to each connected component of $M$). Hence assume that the open cover is countable, i.e. $U = \{U_i : i \in \mathbb{Z}^+\}$.

Define $F_1 = M \setminus \bigcup_{i \geq 1} U_i$. Clearly $F_1 \subseteq M$ is closed and $F_1 \subseteq U_1$. Using that $M$ is normal, there is an open set $V_1 \subseteq M$ such that $F_1 \subseteq V_1 \subseteq \overline{V}_1 \subseteq U_1$. Next define $F_2 = U_2 \setminus (V_1 \cup \bigcup_{i \geq 2} U_i)$. Again, $F_2 \subseteq M$ is closed, $\{V_1\} \cup \{U_i : i \geq 2\}$ is still an open cover of $M$. 


Continue by inductively selecting open sets \( V_k \subseteq M \) such that \( F_k \subseteq V_k \subseteq \bar{V}_k \subseteq U_k \) and defining \( F_{k+1} = U_{k+1} \setminus \left( \bigcup_{i \leq k} V_i \cup \bigcup_{i > k+1} U_i \right) \). At each stage \( \{ V_i : i \leq k \} \cup \{ U_i : i \geq k+1 \} \) is an open cover of \( M \). Therefore each \( F_{k+1} \subseteq M \) is closed.

We verify that the collection \( \{ V_i : i \geq 1 \} \) still covers \( M \): Suppose \( p \in M \). Since \( U \) is locally finite, there exists for a finite number \( n(p) \) such that \( p \notin U_i \) for all \( i > n(p) \). Consequently, \( p \notin V_i \) for all \( i > n(p) \). In other words, \( p \in \bigcup_{i=1}^{n(p)} V_i \). ■

**Theorem 2.7 (Partition of unity)** Let \( U = \{ U_\alpha \}_{\alpha \in A} \) be an open cover of a manifold \( M \). Then there exist \( C^\infty \)-functions \( \phi_\alpha : M \mapsto [0,1] \) such that the collection of closed sets \( W_\alpha = \{ p \in M : \phi_\alpha(p) \neq 0 \} \), \( \alpha \in A \) is locally finite, \( W_\alpha \subseteq U_\alpha \) for each \( \alpha \in A \), and \( \sum_{\alpha \in A} \phi_\alpha \equiv 1 \).

The collection \( \{ \phi_\alpha \}_{\alpha \in A} \) is called a partition of unity subordinate to \( U \).

**Proof.** Using the paracompactness of \( M \) we may assume that \( U \) is locally finite.

Else, find a locally finite open refinement \( V \) that covers \( M \), and relabel \( V \) to be \( U \).

First consider the special case that \( \bar{U}_\alpha \subseteq M \) is compact for each \( \alpha \in A \). Applying the shrinking lemma, obtain an open cover \( V = \{ V_\alpha \}_{\alpha \in A} \) such that \( V_\alpha \subseteq U_\alpha \) for each \( \alpha \in A \).

As a closed subset of the compact set \( \bar{U}_\alpha \) the set \( V_\alpha \) is again compact, and proposition 2.5 applies. Thus there are maps \( \psi_\alpha \in C^\infty(M) \) such that \( \psi_\alpha|_{\bar{V}_\alpha} \equiv 1 \). Moreover, if \( Z_\alpha = \{ x \in M : \psi_\alpha(x) \neq 0 \} \) then \( \bar{Z}_\alpha \subseteq U_\alpha \). ((Using the notation of the proposition 2.5, \( Z_\alpha \subseteq W_\alpha \), and since each \( W_\alpha \) is closed, it follows that \( \bar{Z}_\alpha \subseteq W_\alpha \), and hence \( Z_\alpha \subseteq U_\alpha \).)

Every point \( p \in M \) has an open neighborhood \( O \subseteq M \) that meets only finitely many \( U_\alpha \). Consequently, all but a finite number of the functions \( \psi_\alpha \) vanish identically on \( O \), and the sum \( \sum_{\alpha \in A} \psi_\alpha \) is well-defined on \( O \), and hence on all of \( M \). Moreover, since \( V \) is a cover for \( M \), for every point \( p \in M \) there exists some \( \alpha \in A \) such that \( p \in\bar{V}_\alpha \) and thus \( \psi_\alpha(p) > 0 \). Define

\[
\phi_\alpha = \frac{\psi_\alpha}{\sum_{\beta \in A} \psi_\beta}.
\]

Clearly \( 0 \leq \phi_\alpha \leq 1 \) for all \( \alpha \in A \) and \( \sum_{\alpha \in A} \phi_\alpha \equiv 1 \). Moreover the support \( \text{supp}(\phi_\alpha) = \{ x \in M : \phi(x) \neq 0 \} \) is contained in \( U_\alpha \) since \( \text{supp}(\phi_\alpha) \subseteq \bar{Z}_\alpha \subseteq U_\alpha \).

We now use this special case to prove a strengthened version of proposition 2.5, and then use that strengthened version to prove the existence of a partition of the unity in the general case.

Suppose \( F \subseteq M \) is closed (not necessarily compact), \( O \subseteq M \) open and \( F \subseteq O \). For each \( x \in F \) choose an open neighbourhood \( V(x) \subseteq O \) such that \( \bar{V(x)} \) is compact. For each \( x \notin F \) choose an open neighbourhood \( V(x) \) such that \( F \cap \bar{V(x)} = \emptyset \) and such that \( \bar{V(x)} \) is compact. (This uses the normality of \( M \).) The open cover \( \{ V(x) \}_{x \in M} \) has a locally open refinement \( \{ Z(x) \}_{x \in M} \) that covers \( M \). (Note that \( Z(x) = \emptyset \) may happen for many \( x \in M \).)

Since the sets \( \bar{Z(x)} \) are compact, the special case of the partition of the unity theorem applies. This means that there are functions \( \phi_x \in C^\infty(M, [0,1]) \) such that the collection \( \{ y \in M : \phi_x(y) > 0 \} : x \in M \) is locally finite, \( \{ y \in M : \phi_x(y) > 0 \} \subseteq Z(x) \) and \( \sum_{x \in M} \phi_x \equiv 1 \).

Set \( f = \sum_{x \in F} \phi_x \). Clearly \( f \in C^\infty(F, [0,1]) \). If \( x \in F \), then \( \text{supp}(\phi_x) \subseteq Z(x) \subseteq \bar{V(x)} \subseteq M \setminus F \). Consequently, \( f|_F \equiv 1 - \sum_{x \notin F} \phi_x|_F \equiv 1 - 0 = 1 \). ■

In the next subsection we will use the partition of unity theorem to prove that every compact \( C^\infty \) manifold may be embedded in some Euclidean space. (This is even true without the compactness assumption, but considerably harder to prove.)
2.7 Differentiating differentiable maps

We start this section by defining (partial) derivatives of maps on coordinate charts. This notion will enable us to define submersions, immersions, and embeddings. These in turn are useful tools to construct new differentiable manifolds from known ones. As examples we revisit hypersurfaces and some matrix submanifolds, and we show that every compact manifold can be imbedded into a Euclidean space.

Recall the definition of the $i$-th partial derivative of a function $f: \mathbb{R}^n \mapsto \mathbb{R}$ at a point $a \in \mathbb{R}^n$:

$$D_i f(a) = \lim_{h \to 0} \frac{1}{h} (f(a_1, \ldots, a_{i-1}, a_i + h, a_{i-1}, \ldots, a_n) - f(a)).$$

**Definition 2.21** Suppose $M$ is a differentiable manifold and $f: M \mapsto \mathbb{R}$. The $i$-th partial derivatives of $f$ in a local coordinate chart $(u, U)$ at a point $p \in U$ is defined as

$$\frac{\partial f}{\partial u^i} \bigg|_p = D_i(f \circ u^{-1}) \bigg|_{u(p)}$$

(20)

Note, if $\gamma: (-\varepsilon, \varepsilon) \mapsto U \subseteq M$ is a curve such that $\gamma(0) = p$, $u^i(\gamma(t)) = t$ and $u^j(\gamma(t)) = u^j(p)$ for $j \neq i$ then for $f: U \mapsto \mathbb{R}$

$$\frac{\partial f}{\partial u^i} \bigg|_p = D_i(f \circ u^{-1}) \bigg|_{u(\gamma(0))} = \lim_{h \to 0} \frac{1}{h} (f(\gamma(h)) - f(p))$$

(21)

This curve $\gamma$ is adapted in a very special way to the specific local coordinate chart. The next section will generalize this setting along its way to construct the tangent bundle.

- If $f = u^j$ then $\frac{\partial u^j}{\partial u^i} = \delta_{i,j}$ (Kronecker delta, i.e., $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ else).
- If $M = \mathbb{R}^n$ with the identity chart then $\frac{\partial f}{\partial u^i} = D_i f$.

**Proposition 2.8 (Chain-rule)** Let $(u, U)$ and $(v, V)$ be local coordinate charts, $p \in U \cap V$ and $f: M \mapsto \mathbb{R}$. Then

$$\frac{\partial f}{\partial v^j} \bigg|_p = \sum_{i=1}^{n} \frac{\partial f}{\partial u^i} \bigg|_p \frac{\partial u^i}{\partial v^j} \bigg|_p$$

(22)

In this case, the relation uses only two indices and may be nicely be written in matrix from

$$\left( \frac{\partial f}{\partial v^1}, \ldots, \frac{\partial f}{\partial v^n} \right) = \left( \frac{\partial f}{\partial u^1}, \ldots, \frac{\partial f}{\partial u^n} \right) \left( \begin{array}{cccc} \frac{\partial u^1}{\partial v^1} & \cdots & \frac{\partial u^1}{\partial v^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial u^n}{\partial v^1} & \cdots & \frac{\partial u^n}{\partial v^n} \end{array} \right).$$

(23)

However, many subsequent formulas will involve more indices and hence generally will not be amenable to simple matrix thinking. Instead, most helpful is a careful and consistent choice of indices written as subscripts and superscripts. This goes as far as the “Einstein summation convention” (not used in these notes), which omits the summation signs and instead tacitly implies that if an index occurs once as a lower and once as an upper index, then it is understood that this index is to be summed over. yet even without using that summation convention, we will find it helpful to carefully check balanced occurrences as upper and lower indices.
Proof. Use the coordinate maps and the chain-rule in Euclidean spaces
\[
\left.\frac{\partial f}{\partial \omega} \right|_p = D_j(f \circ v^{-1})\bigg|_{v(p)} = D_j((f \circ u^{-1}) \circ (u \circ v^{-1}))\bigg|_{v(p)} = \sum_{i=1}^{n} D_i((f \circ u^{-1})|_{(u \circ v^{-1})(v(p))}) \cdot D_k(u^i \circ v^{-1})\bigg|_{v(p)} = \sum_{i=1}^{n} \left.\frac{\partial f}{\partial \omega_i}\right|_p \left.\frac{\partial u_i}{\partial \omega_j}\right|_p.
\]
(24)

Proposition 2.9 (Product-rule) Suppose \((u, U)\) is a chart, \(p \in U\) and \(f, g : M \rightarrow \mathbb{R}\). Then
\[
\left.\frac{\partial (fg)}{\partial v^j}\right|_p = \left.\frac{\partial f}{\partial v^j}\right|_p \cdot g(p) + f(p) \cdot \left.\frac{\partial g}{\partial v^j}\right|_p.
\]
(25)

Exercise 2.27 Prove the product rule for functions on a manifold as stated above.

Proposition 2.10 Let \((u, U)\) and \((v, V)\) be local coordinate charts, about \(p \in U \cap V\). Then the Jacobian matrix
\[
\left(\left.\frac{\partial u^i}{\partial v^j}\right|_p\right)_{i=1, \ldots, n, \ j=1, \ldots, n}
\]
(26)
is invertible.

Proof. The chain-rule gives \(\delta_{i,k} = \left.\frac{\partial u^i}{\partial v^k}\right|_p = \sum_{j=1}^{n} \left.\frac{\partial u^i}{\partial v^j}\right|_p \cdot \left.\frac{\partial v^j}{\partial v^k}\right|_p\). Hence the Jacobian matrices
\[
A = \left(\left.\frac{\partial u^i}{\partial v^j}\right|_p\right)_{i=1, \ldots, n, \ j=1, \ldots, n}
\]
(27) and \(B = \left(\left.\frac{\partial u^j}{\partial v^k}\right|_p\right)_{i=1, \ldots, n, \ j=1, \ldots, n}\) satisfy \(A \cdot B = I\), and similarly \(B \cdot A = I\), i.e. \(A\) is invertible. □

Proposition 2.11 Let \(\Phi : M^m \rightarrow N^n\) be a differentiable map between manifolds. If \((u, U)\) and \((\bar{u}, \bar{U})\) are local coordinate charts about \(p \in M\), and \((v, V)\) and \((\bar{v}, \bar{V})\) are local coordinate charts about \(\Phi(p) \in N\), then the Jacobian matrices (at \(p \in U \cap \bar{U}\))
\[
\left(\left.\frac{\partial (v^i \circ \Phi)}{\partial v^j}\right|_p\right)_{i=1, \ldots, n, \ j=1, \ldots, n}
\]
and \(\left(\left.\frac{\partial (\bar{v}^i \circ \Phi)}{\partial \bar{v}^j}\right|_p\right)_{i=1, \ldots, n, \ j=1, \ldots, n}\) have the same rank
(28)

Proof. Applying the chain-rule twice yields
\[
\left.\frac{\partial (v^i \circ \Phi)}{\partial \bar{v}^j}\right|_p = D_j(\bar{v}^i \circ \Phi \circ \bar{u}^{-1})\bigg|_{\bar{u}(p)} = D_j((\bar{v}^i \circ v^{-1}) \circ (v \circ \Phi \circ u^{-1}) \circ (u \circ \bar{u}^{-1}))\bigg|_{\bar{u}(p)} = \sum_{k=1}^{m} \sum_{\ell=1}^{m} D_k(\bar{v}^i \circ v^{-1})\bigg|_{(v \circ \Phi(p))}\cdot D_{\ell}(v^k \circ \Phi \circ u^{-1})\bigg|_{u(p)} \cdot D_j(u^\ell \circ \bar{u}^{-1})\bigg|_{\bar{u}(p)} = \sum_{k=1}^{m} \sum_{\ell=1}^{m} D_k(\bar{v}^i \circ v^{-1})\bigg|_{\Phi(p)} \cdot \left.\frac{\partial (v^k \circ \Phi)}{\partial v^\ell}\right|_p \frac{\partial u^\ell}{\partial \bar{v}^j}\bigg|_p.
\]
(29)
The proposition follows from the invertibility of the Jacobian matrices
\[
\left(\left.\frac{\partial \bar{v}^i}{\partial v^j}\right|_{\Phi(p)}\right)_{i=1, \ldots, n, \ j=1, \ldots, n}
\]
(30) and
\[
\left(\left.\frac{\partial u^\ell}{\partial \bar{v}^j}\right|_p\right)_{j=1, \ldots, m, \ \ell=1, \ldots, m}.
\]
Definition 2.22 Let $\Phi: M^m \to N^n$ be a differentiable map between manifolds. The rank of $\Phi$ at $p \in M$ is defined as the rank of the Jacobian matrix
\[
\left( \frac{\partial (v^i \circ \Phi)}{\partial u^j} \right)_{p} \quad i = 1, \ldots, n \quad j = 1, \ldots, m \quad (31)
\]
where $(u,U)$ and $(v,V)$ are any local coordinate charts about $p \in M$ and $\Phi(p) \in N$, respectively.

This definition is justified – the rank is well-defined – by virtue of the preceding proposition: The rank of $\Phi$ at $p$ (written $\text{rank}_p \Phi$) is independent of the choice of local coordinates employed.

Definition 2.23 Suppose $\Phi: M \to N$ is a differentiable map between manifolds.

- $p \in M$ is called a regular point of $\Phi$ if $\text{rank}_p \Phi = \dim N$.
- $p \in M$ is called a critical point of $\Phi$ if $\text{rank}_p \Phi < \dim N$.
- $q \in N$ is called a regular value of $\Phi$ if every point $p \in \Phi^{-1}(q) \subseteq M$ is a regular point of $\Phi$.
- $q \in N$ is called a critical value of $\Phi$ if $q = \Phi(p)$ for some critical point $p$ of $\Phi$.

Note that every point $q \in N \setminus \Phi(M)$ is automatically a regular value of $\Phi$. Intuitively one expects that the set of critical values is a small subset of $N$. To make this precise it suffices to have a notion of sets of measure zero on the manifold. (A notion of measure zero does not require a notion of a measure on the manifold!) Recall that a subset $A \subseteq \mathbb{R}^n$ has measure zero, if for every $\varepsilon > 0$ there exist a countable collection of open cubes
\[
C_i(p_i, r_i) = \{ q \in \mathbb{R}^n : \| q - p \|_\infty < r_i \} \quad \text{such that} \quad A \subseteq \bigcup_{i=1}^{\infty} C_i(p_i, r_i) \quad \text{and} \quad \sum_{i=1}^{\infty} r_i^n < \varepsilon. \quad (32)
\]

Definition 2.24 A subset $S \subseteq M^n$ of a manifold $M$ has measure zero if there exist a sequence of charts $(u_i, U_i)$, $i \in \mathbb{Z}^+$ such that $S \subseteq \bigcup_{i=1}^{\infty} U_i$ and each $u_i(U_i \cap S) \subseteq \mathbb{R}^n$ has measure zero.

Exercise 2.28 Suppose $\Phi \in C^1(M,N)$ is a differentiable map between manifolds and $S \subseteq M$ has measure zero. Show that $\Phi(S) \subseteq N$ has measure zero.

Theorem 2.12 (Sard’s theorem, simple version) Suppose $M$ and $N$ are differentiable manifolds with $\dim M = \dim N$ and $M$ has at most countably many connected components. If $\Phi \in C^1(M,N)$ then the set of critical values of $\Phi$ has measure zero in $N$.

Note that with our definition of sets of measure zero this theorem is a direct consequence of the analogous statement for maps between Euclidean spaces. A proof of Sard’s theorem for such maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ may be found in e.g. Spivak, Calculus on manifolds (p.72).
Definition 2.25 Suppose a differentiable map \( \Phi: M^m \mapsto N^n \) has constant rank \( k \) at all \( p \in M \).

- If \( k = \dim N \) then \( \Phi \) is called a submersion.
- If \( k = \dim M \) then \( \Phi \) is called an immersion and \( \Phi(M) \subseteq N \) an immersed submanifold.
- If \( \Phi \) is an immersion and \( \Phi \) is also a homeomorphism onto its image \( \Phi(M) \) (in the subspace topology inherited from \( N \)), then \( \Phi \) is called an embedding, and \( \Phi(M) \subseteq N \) is called an embedded submanifold, or simply a submanifold.

These definitions apply in particular to the case when the manifold \( M \) is also a subset of \( N \). In this case, the map \( \Phi: M \mapsto N \) is naturally taken as the inclusion map \( \iota: M \hookrightarrow N \) defined by \( \iota(p) = p \). Hence, the manifold \( M \) is an immersed submanifold of \( N \) if \( \iota \) is an immersion, and \( M \) is a submanifold of \( N \) if \( \iota \) is an imbedding. In the case that \( M \) is a submanifold and is also a closed subset of \( N \) one also calls \( M \) a closed submanifold.

Note that an immersion is not required to be one-to-one. However, since it has maximal rank, the map is locally one-to-one. This means that every point \( p \in M \) has a neighbourhood on which the map is one-to-one (i.e. the map is a topological immersion).

Examples (immersions and imbeddings):

- The map \( \Phi: \mathbb{R} \mapsto \mathbb{R} \), \( \Phi(x) = x^3 \) is \( C^\infty \) and is a homeomorphism, but not an immersion.
- The map \( g: \mathbb{R} \mapsto \mathbb{R}^2, g(t) = \left(2 \cos(t - \frac{\pi}{2}), \sin(2t - \pi)\right) \) is an immersion but not one-to-one.
- Define \( h: \mathbb{R} \mapsto \mathbb{R} \) by \( h(t) = 2 \tan^{-1} t \). Then \( (g \circ h): \mathbb{R} \mapsto \mathbb{R}^2 \) is a one-to-one immersion, but not an imbedding: For every neighbourhood \( U \) of \( (g \circ h)(0) = (0, 0) \) in \( N = \mathbb{R}^2 \) there exists a \( a > 0 \) such that \( (-\infty, -a) \cup (a, \infty) \subseteq (g \circ h)^{-1}(U) \).
- A skew line on the torus is the image of a curve \( \gamma: \mathbb{R} \mapsto \mathbb{R}^2/\mathbb{Z}^2 \gamma: t \mapsto (t, qt) \mod \mathbb{Z}^2 \) for an irrational number \( q \in \mathbb{R} \setminus \mathbb{Q} \). The irrationality assures that \( \gamma \) is one-to-one. But the image of \( \gamma \) is a dense subset on the torus, and hence it is not an embedding.
Theorem 2.13 Suppose $\Phi \in C^\infty(M^m,N^n)$ ($r > 0$) is a smooth map between manifolds and $p \in M$. If $\Phi$ has constant rank $k$ on a neighbourhood of $p$ then there exist local coordinate charts $(u,U)$ and $(v,V)$ about $p$ and $\Phi(p)$, respectively, such that for all $x \in u(U) \subseteq \mathbb{R}^m$

$$v \left((\Phi \circ u^{-1})(x^1,\ldots,x^m)\right) = (x^1,\ldots,x^k,0,\ldots,0)$$

Proof. Start with any charts $(\bar{u},\bar{U})$ and $(\bar{v},\bar{V})$ about $p$ and $\Phi(p)$, respectively. Without loss of generality assume that the first $(k \times k)$-minor of the Jacobian has full rank, i.e.

$$\det \left( \frac{\partial (\bar{v} \circ \Phi)}{\partial \bar{w}^j} \right)_{i=1,\ldots,k, \ j=1,\ldots,k} \neq 0$$

(else permute and relabel the components of $\bar{u}$ and/or $\bar{v}$). This rank conditions allows one to effectively use $(\bar{v} \circ \Phi)$, $i = 1,\ldots,k$ as coordinate functions in place of $u^i$, $i = 1,\ldots,k$. More specifically define a new map $\bar{u}: \bar{U} \mapsto \mathbb{R}^m$ by $\bar{u}^j = \bar{v}^j \circ \Phi$ if $j \leq k$ and $\bar{u}^j = \bar{u}^j$ else. To verify that this is indeed a legitimate local coordinate change calculate the Jacobian matrix of partial derivatives $\frac{\partial \bar{u}^i}{\partial \bar{w}^j}$. This matrix has an upper block triangular structure. The first $(k \times k)$ minor agrees by construction with the one in (33), while the bottom right block is the $(m-k) \times (m-k)$ identity matrix. Consequently, the Jacobian has full rank at $p$, and hence in some open neighbourhood of $p$. By virtue of the inverse function theorem, there exists a neighbourhood $U \subseteq \bar{U}$ of $p$ such that the restriction $u$ of $\bar{u}$ to $U$ is a diffeomorphism (onto its image), and hence $(u,U)$ is a chart about $p$. Thus

$$\bar{v} \left((\Phi \circ u^{-1})(x^1,\ldots,x^m)\right) = (x^1,\ldots,x^k,\bar{v}^{k+1}(x),\ldots,\bar{v}^n(x))$$

(34) for suitable functions $\bar{v}^{k+1},\ldots,\bar{v}^n$: $u(U) \mapsto \mathbb{R}$. The Jacobian matrix of partial derivatives $D_j(\bar{v}^i \circ \Phi \circ u^{-1})(x)$, with $i = 1,\ldots,n$, and $j = 1,\ldots,m$ has a lower block triangular structure. The top left $(k \times k)$ block is the identity, while the bottom right $(m-k) \times (n-k)$-block consist of the partial derivatives $D_j \bar{v}^i(x)$, $i = k+1,\ldots,n$ and $j = k+1,\ldots,m$. Since the matrix is assumed to have constant rank $k$ in a neighbourhood of $u(p) \in \mathbb{R}^m$, one concludes that the bottom right $(m-k) \times (n-k)$-block is identically equal to zero. This means that the functions $\bar{v}^i$ for $i = k+1,\ldots,n$ do not depend on $x^j$, $j = k+1,\ldots,m$. Consequently there are functions $\psi^i$, $i = k+1,\ldots,n$, defined on a suitable subset of $\mathbb{R}^k$ such that

$$\bar{v}^i(x^1,\ldots,x^k,x^{k+1},\ldots,x^m) = \bar{v}^i(x^1,\ldots,x^k) \quad \text{for all} \quad x = (x^1,\ldots,x^m) \in u(U)$$

(35)

Restrict the maps $\bar{v}$ and $\bar{v}$ to $V = \bar{V} \cap \Phi(U)$, denoted $\bar{v} = \bar{v}|_V$ and $\psi = \bar{v}|_V$. Finally define $v: V \mapsto \mathbb{R}^n$ by $v^i = \bar{v}^i$ if $i \leq k$ and $v^i = \bar{v}^i - \psi^i \circ (\bar{v}^1,\ldots,\bar{v}^k)$ if $i > k$; i.e. for $y \in v(V \cap \Phi(U)) \subseteq \mathbb{R}^n$

$$(v^i \circ \bar{v}^{-1})(y^1,\ldots,y^n) = (y^1,\ldots,y^k,y^{k+1} - \psi^{k+1}(y^1,\ldots,y^k),\ldots,y^n - \psi^n(y^1,\ldots,y^k))$$

(36)

This assures that, as desired, for $x \in u(U \cap \Phi^{-1}(V))$

$$(v \circ \Phi \circ u^{-1})(x^1,\ldots,x^m) = (v \circ \bar{v}^{-1}) \circ (\bar{v} \circ \Phi \circ u^{-1})(x^1,\ldots,x^m)$$

$$= (v \circ \bar{v}^{-1})(x^1,\ldots,x^k,\psi^{k+1}(x^1,\ldots,x^k),\ldots,\psi^n(x^1,\ldots,x^k))$$

$$= (x^1,\ldots,x^k,\psi^{k+1}(x^1,\ldots,x^k) - \psi^{k+1}(x^1,\ldots,x^k),\ldots,\psi^n(x^1,\ldots,x^k) - \psi^n(x^1,\ldots,x^k))$$

$$= (x^1,\ldots,x^k,0,\ldots,0).$$
Note that in general the construction in the proof works under the assumption that $\Phi \in C^r$ as long as $r > 0$. The strong version of the inverse function theorem yields local coordinates $u$ and $v$ of the same degree of smoothness as $\Phi$.

The special case of $k = \dim M$ is sometimes referred to as the "local immersion theorem", while the special case of $k = \dim N$ is referred to as the "local submersion theorem". In the first case the proof is a little shorter as it is immediately clear that $(v \circ \Phi)$ defines local coordinates about $p$. In the second case, the construction becomes shorter as there is no need for the functions $\bar{\psi}$.

**Corollary 2.14** Suppose $\Phi \in C^\infty(M^m, N^n)$ is a smooth map between manifolds and $q \in N$. If $\Phi$ has constant rank $k$ on a neighbourhood of $\Phi^{-1}(q)$ then $\Phi^{-1}(q) \subseteq M$ is a closed submanifold of dimension $(m - k)$ (or it is empty). In particular, if $q \in N$ is a regular value of $\Phi$ then $\Phi^{-1}(q) \subseteq M$ is an $(m - n)$-dimensional submanifold on $M$.

**Proof.** Suppose $p \in \Phi^{-1}(q)$. Then there exist charts $(u, U)$ and $(v, V)$ about $p$ and $q$, respectively, such that w.l.o.g. $u(p) = 0 \in \mathbb{R}^m$, $v(q) = 0 \in \mathbb{R}^n$, and for $x \in u(U \cap \Phi^{-1}(V))$

$$ (v \circ \Phi \circ u^{-1})(x^1, \ldots, x^k, x^{k+1}, \ldots, x^m) = (x^1, \ldots, x^k) \quad (38) $$

Let $W = U \cap \Phi^{-1}(q)$ and define $w: W \mapsto \mathbb{R}^{m-k}$ by $w^j = u^{k+j}$ for $j = 1, \ldots, (m - k)$. Then $(w, W)$ is a chart for $\Phi^{-1}(M)$ about $p$. ■

This theorem is a special case of a more general result that applies to the inverse image $\Phi^{-1}(P) \subseteq M$ of a submanifold $P \subseteq N$ under a smooth map $\Phi: M \mapsto N$. To illustrate that some additional hypotheses are needed, consider the example of $M = \mathbb{R}$, $N = \mathbb{R}^2$, $\Phi(x) = (x, 0)$ and $P = \{(x, f(x): x \in \mathbb{R}\}$ is the graph of the $C^\infty$ function $f: \mathbb{R} \mapsto \mathbb{R}$, defined by $f(x) = \exp(-\frac{1}{x})$ if $x > 0$ and $f(x) = 0$ else. Then $P \subseteq N$ is an (imbedded) 1-dimensional submanifold, $\Phi \in C^\infty(M, N)$ is a submersion (rank$_x\Phi = 1$ for all $x \in M$). However, $\Phi^{-1}(P) = (-\infty, 0] \subseteq M$ is not a submanifold (because no neighbourhood of $0 \in \Phi^{-1}(P) \subseteq M$ is homeomorphic to $\mathbb{R}$).

**Exercise 2.29** Prove the submersion theorem: If $\Phi \in C^\infty(M^m, N^n)$ is a submersion and $P \subseteq N$ is a submanifold of dimension $p$ then $\Phi^{-1}(P) \subseteq M$ is a submanifold of dimension $m - (n - p)$.

**Exercise 2.30** Explore where an attempt (using constructions similar to the ones used in the proofs of the preceding theorem and corollary) to prove the statement of the preceding exercise without the assumption that rank$\Phi = n$ breaks down (unless other assumptions are added).
Examples (Submanifolds from submersions):

- If \( f: \mathbb{R}^2 \rightarrow \mathbb{R} \), \( \Phi(x, y) = x^2 + y^2 \), then \( f^{-1}(1) = S^1 \subseteq \mathbb{R}^2 \) is an (imbedded) circle.
- If \( f: \mathbb{R}^{m+1} \rightarrow \mathbb{R} \), \( \Phi(x) = \|x\|_2^2 \), then \( f^{-1}(1) = S^m \subseteq \mathbb{R}^{m+1} \) is an (imbedded) \( m \)-sphere.

Exercise 2.31 (Hopf map) Let \( \tilde{\Phi}: \mathbb{R}^4 \rightarrow \mathbb{R}^3 \) be defined by \( \tilde{\Phi}(a) = (2a_1a_3 + 2a_2a_4, 2a_2a_3 - 2a_1a_4, a_1^2 + a_2^2 - a_3^2 - a_4^2) \).

- Calculate the rank of \( \tilde{\Phi} \) at any \( a \in \mathbb{R}^4 \).
- Verify that \( \tilde{\Phi}(S^3) \subseteq S^2 \).
- Let \( \Phi \) be the restriction of \( \tilde{\Phi} \) to \( S^3 \subseteq \mathbb{R}^4 \), and considered as a map into \( S^2 \subseteq \mathbb{R}^3 \).
- Calculate the rank of \( \tilde{\Phi} \) at \( p \in S^3 \) using the definition in terms of coordinate charts. (Use stereographic projections to avoid square-roots!)
- For \( q \in S^2 \) describe the preimage \( \tilde{\Phi}^{-1}(q) \subseteq S^3 \). Is it a submanifold? If so, what is its dimension? Which manifold is it?

Exercise 2.32 Define \( \Phi: M_{n \times n}(\mathbb{R}) \rightarrow M_{n \times n}(\mathbb{R}) \) by \( \Phi(A) = A^T A \). Show that the set \( \Phi^{-1}(I_{n \times n}) \) of orthogonal matrices is a closed submanifold of \( M_{n \times n}(\mathbb{R}) \). What is its dimension?

Calculations for the example of the special linear group. If \( \Phi = \det: M_{n \times n}(\mathbb{R}) \rightarrow \mathbb{R} \) is the determinant function, then the special linear group \( SL(n, \mathbb{R}) \) is the preimage \( \det^{-1}(1) \). The following calculations show that \( SL(n, \mathbb{R}) \) is a closed \( (n^2 - 1) \)-dimensional submanifold of \( M_{n \times n}(\mathbb{R}) \). It suffices to show that \( \det \) has rank one at every \( A \in \det^{-1}(1) \subseteq M_{n \times n}(\mathbb{R}) \). The following calculations actually provide a little more, including a formula for the derivative of \( \det \) which is of interest in the next chapter. The calculation of the derivative in \( M_{n \times n}(\mathbb{R}) \) is basically the same as in \( \mathbb{R}^{n^2} \). Since we need inverses, we consider the restriction \( f \) of \( \det \) to the the general linear group \( GL(n, \mathbb{R}) = \{ A \in M_{n \times n}(\mathbb{R}) : \det A \neq 0 \} \) of invertible matrices. To make the connection with the next chapter, recall that \( GL(n, \mathbb{R}) \) is an open submanifold of \( M_{n \times n}(\mathbb{R}) \) – and thus it is appropriate to consider directional derivatives of at points \( A \in GL(n, \mathbb{R}) \) in all possible directions \( B \in M_{n \times n}(\mathbb{R}) \):

\[
f'(A)(B) = \lim_{h \to 0} \frac{1}{h} \left( f(A + hB) - f(A) \right)
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left( \det(A + hB) - \det A \right)
\]

\[
= \det A \cdot \lim_{h \to 0} \frac{1}{h} \cdot \left( h^n \cdot \det(\frac{1}{h} I - (-A^{-1}B)) - 1 \right)
\]

\[
= \det A \cdot \lim_{h \to 0} \frac{1}{h} \cdot \left( h^n \cdot \left( \frac{1}{h^n} - \frac{1}{h^{n-1}} \cdot \text{tr}(-A^{-1}B) + \ldots + \det(-A^{-1}B) \right) - 1 \right)
\]

Choosing \( B = A \) shows \( f'(A)(A) = n \det A \neq 0 \) and hence \( f'(A) \) has full rank equal to one, the dimension of the range. Thus \( SL(n, \mathbb{R}) \) is a an \( (n^2 - 1) \)-dimensional closed submanifold.

As a final note in this chapter we combine imbeddings with a variation of partitions of the unity.

Theorem 2.15 Any compact smooth manifold \( M^n \) can be imbedded in some Euclidean space \( \mathbb{R}^N \).
This is a very weak version – with a very simple proof – of much stronger results. Indeed, it is not too hard to actually show that it is always possible to take $N = 2n + 1$ (Spivak I, ex. 3.33). Much harder is Whitney’s theorem which asserts that actually $N = 2n$ is possible, even without the compactness assumption. A similar result for isometric (i.e. distance-preserving) imbeddings of Riemannian manifolds is known as Nash’s theorem.

Proof. Since $M^m$ is compact there exists a finite atlas $U = \{(u_i, U_i)\colon i = 1 \ldots n\}$ for $M$. Use the shrinking lemma to obtain an open cover $V$ of $M$ by open sets $V_i \subseteq M$ whose closures are contained in $U_i$, i.e. $\overline{V}_i \subseteq U_i$, $i = 1 \ldots n$. Using normality, there are smooth functions $\phi_i\colon M \mapsto [0,1]$, $i = 1, \ldots n$ such that $\phi_i|_{\overline{V}_i} \equiv 1$ and such that each has support $\text{supp}(\phi_i) = \{p \in M\colon \phi_i(p) \neq 0\} \subseteq U_i$. Extend each map $u_i$ to all of $M$ by setting $u_i(p) = 0$ if $p \notin U_i$. Let $N = nm + n$ and define $\Phi\colon M \mapsto \mathbb{R}^N$ by

$$\Phi = (\phi_1 \cdot u_1, \ldots \phi_n \cdot u_n, \phi_1, \ldots \phi_n)$$

(39)

To see that $\Phi$ is one-to-one, suppose $p, q \in M$ are such that $\Phi(p) = \Phi(q)$. Since $V$ covers $M$ there exists $i_0$ such that $\phi_{i_0}(p) \neq 0$. Since $\Phi(p) = \Phi(q)$ this implies that $\phi_{i_0}(q) = \phi_{i_0}(p) \neq 0$, too. Consequently both $p \in U_{i_0}$ and $q \in U_{i_0}$. Again use $\Phi(p) = \Phi(q)$ to conclude that $\phi_{i_0}(q) u_{i_0}(q) = \phi_{i_0}(p) u_{i_0}(p)$, hence $u_{i_0}(q) = u_{i_0}(p)$, i.e. $p = q$.

Finally verify that $\Phi$ is an immersion: Let $p \in M$ and choose $i_0$, as above, that $p \in V_{i_0}$, i.e. $\phi_{i_0} \equiv 1$ in some neighbourhood of $p$. Consequently $\Phi^{m(i_0-1)+j} = u_{i_0}^j$, $j = 1, \ldots m$ in a neighbourhood of $p$. This guarantees that the $(N \times m)$-Jacobian $D(\Phi \circ u_{i_0}^{-1})$ contains an $(m \times m)$-identity block, i.e. $\Phi$ has full rank in a neighbourhood of $p$. □
3 The tangent bundle

3.1 Introduction

This is a good time to reflect why we want a notion of tangent spaces and tangent maps in the first place. Said differently, what do we expect this notion to deliver? What properties should tangent vectors and tangent spaces have? What are the tangent spaces to the line $\mathbb{R}$ and the plane $\mathbb{R}^2$—two of the most familiar manifolds?

We want to use our experience with tangent lines to curves and tangent planes to surfaces in two- and three dimensional Euclidean spaces as guidance. However, in general we do not want our notion of tangent objects to depend on, or be constrained by imbeddings of the manifold into some Euclidean space. Thus without any surrounding space available, the pictorial arrows become untenable. Before reading on, you should close the notes and brainstorm some ideas . . .

Some ideas which come to mind are:

- Tangent vectors should be vectors, i.e. be members of a linear space that provides for addition and scalar multiplication.
- The dimension of the linear tangent space(s) should equal the dimension of the manifold.
- Tangent objects should provide for notions of linear approximations, of objects on manifolds, and of maps between manifolds. Recall that we already have notions of derivatives of maps $\Phi: M \rightarrow N$, but only with respect to coordinate charts $(u, U)$ and $(v, V)$, in terms of the maps $(v \circ \Phi \circ u^{-1}): \mathbb{R}^m \rightarrow \mathbb{R}^n$. Clearly a coordinate-free notion is desirable.
- Vector fields are intimately connected to differential equations / dynamical systems. Thus tangent vectors should provide a means to describe dynamical systems on manifolds.
- Some vector fields are gradient vector fields, i.e. are the derivatives of some potential function. Said differently, vector fields should generalize partial differential equations for unknown functions. [[Aside: This will lead to a notion of co-tangent vector fields.]]
- We defined arc-length as an integral of the speed. In general tangent vectors may provide a means on which to base a generalized notion of distance. [[This will lead to Riemannian metrics in the second half of this course.]]
- We defined curvature for curves in terms of the rate of change of the tangent vectors. Thus we expect that a general notion of (comparing) tangent spaces (at different points) should provide for a notion of curvature. [[This raises the question of which of the two comes first, the notion of curvature, or the means to compare tangent spaces at different points.]]

Should our definition allow tangent spaces at different points of a manifold to have nonempty intersection? E.g. consider the unit-circle $S^1$ imbedded in the plane $\mathbb{R}^2$. The tangent lines to $S^1$ at $p = (1, 0)$ and $q = (0, 1)$ intersect nontrivially at $(1, 1)$. On the other hand, if we think of the tangent vectors $v_p = (0, 1)$ and $v_q = (1, 0)$ as arrows based at $p$ and $q$ respectively, then we certainly think of them as different.

This brings up a larger issue of distinguishing vectors (arrows) that may be moved around and vectors that are rooted at a fixed point. There are many applications where it is advantageous to consider equivalence classes of directed line-segments, equivalence meaning that they may be transformed into each other by parallel translation. On the other hand, there are many places where it is appropriate to consider vectors that are rooted, or fixed at their base points (e.g. velocity vectors to a curve).

We shall use the next section as an opportunity to bring clarity to these issues and make very precise definitions (which may always be relaxed where this causes no trouble).
3.2 Tangent spaces

There are many different ways in which one may motivate an eventual construction of tangent spaces to a general manifold. One typically starts from surfaces in Euclidean spaces, then considers more abstractly immersed manifolds in higher dimensional Euclidean spaces, and eventually tries to develop a notion that works in abstract settings, yet reduces to the familiar ones in Euclidean settings. For a lengthy such discussion see Spivak Vol. I ch. 3.

An intuitive (and very useful) way to define tangent vectors to a manifold \( M \) at a point \( p \) is as equivalence classes of curves. Roughly, two curves are equivalent if they have the same velocity vector at \( p \) – but this would be circular as we don’t have notions of velocity vectors for general curves on manifolds. So the next best thing is to declare any two smooth curves \( \sigma, \gamma : (-\varepsilon, \varepsilon) \to M \) (with \( \sigma(0) = \gamma(0) = p \)) equivalent if for every smooth function \( f \in C_\infty(p) \) (that is, smooth function defined on some neighborhood of \( p \)) \( \frac{d}{dt}|_{t=0}(f \circ \sigma) = \frac{d}{dt}|_{t=0}(f \circ \gamma) \).

**Exercise 3.1** Further explore how this leads to a notion of tangent spaces that is basically the same as the one we define below. In particular, equip the collection of equivalence classes with an addition and scalar multiplication (make sure that these are well-defined). Check that the space of tangent vectors at a point is indeed an \( m \)-dimensional vector space. In a coordinate chart find a basis for the tangent space (e.g. provide representatives (curves) for \( m \) equivalence classes that form a basis). Show how to write any tangent vector as a linear combination of this basis. Analyze how the coordinates of a tangent vector transform under local coordinate changes on the manifold. Match this notion of tangent spaces to the one provided below.

The exercise already hinted at a useful connection between tangent vectors and derivatives. Indeed, going back to Euclidean spaces, say e.g. \( M = \mathbb{R}^2 \) consider the relation between a vector \( \vec{v} = (v^1, v^2) = v^1 \vec{i} + v^2 \vec{j} \) and the directional derivative (operator) \( D_{\vec{v}}|_p \), defined by \( D_{\vec{v}}|_p f = v^1 \cdot (D_1 f)(p) + v^2 \cdot (D_2 f)(p) \), commonly also written as \( D_{\vec{v}}f(p) = \langle \vec{v} \rangle \nabla f(p) \).

Clearly any vector \( \vec{v} \) uniquely determines a (directional) derivative operator \( D_{\vec{v}}|_p \). Conversely, one can easily recover the vector \( \vec{v} \) from the directional derivative \( D_{\vec{v}} \) by simply evaluating the latter on suitable functions: For example, evaluating \( D_{\vec{v}}|_p \) on the coordinate functions \( f_1(x^1, x^2) = x^1 \) and \( f_2(x^1, x^2) = x^2 \) immediately recovers the coordinates of \( \vec{v} \) as \( v^1 = (D_{\vec{v}}x^1)(p) \) and \( v^2 = (D_{\vec{v}}x^2)(p) \). Thus there appears to be no harm in identifying the vector \( \vec{v} \) with the directional derivative operator \( D_{\vec{v}}(\cdot)(p) \). As we shall soon see, this idea will prove most beneficial since operators on spaces of functions are automatically endowed with a rich algebraic structure which is ready for us to use! Following Chevalley, we define

**Definition 3.1** A tangent vector to a manifold \( M \) at a point \( p \in M \) is a function \( X_p : C_\infty(p) \to \mathbb{R} \) which is linear over \( \mathbb{R} \) and is a derivation on \( C_\infty(p) \), i.e. which satisfies for \( \lambda \in \mathbb{R} \) and for all \( f, g \in C_\infty(p) \) on their common domain,

\[
X_p(\lambda f + g) = \lambda X_p(f) + X_p(g).
\]

\[
X_p(f \cdot g) = X_p(f) \cdot g(p) + f(p) \cdot X_p(g).
\]

The set of all tangent vectors to \( M \) at \( p \) is called the tangent space to \( M \) at \( p \), denoted \( T_pM \).

Here \( f \in C_\infty(p) \) means that there exists an open neighborhood \( U \) of \( p \) (depending on \( f \)) such that \( f \in C_\infty(U, \mathbb{R}) \). (This is a special case of the earlier definition of \( C_\infty(A) \) for subsets \( A \subseteq \mathbb{R}^n \) that are not necessarily open.) The following lemma only uses the abstract notion of a derivation, no coordinates are required:
**Lemma 3.1** If $c: M \rightarrow \mathbb{R}$ is a constant function and $X_p \in T_p M$ then $(X_p c) = 0$

**Proof.** Use the linearity and Leibniz rule for differentiating products to establish
\[ c \cdot X_p(1) = X_p(c \cdot 1) = (X_p c) \cdot 1 + c \cdot (X_p1) \quad \text{and hence} \quad (X_p c) = 0. \]

**Observations:**
- $T_p M$ is a vector space: In particular, if $X_p, Y_p \in T_p M$ and $\lambda \in \mathbb{R}$ then $(X_p + \lambda Y_p) \in T_p M$. The addition and scalar multiplication are inherited from pointwise evaluations, i.e. $(X_p + \lambda Y_p)f = (X_p f) + \lambda(Y_p f)$.
- Tangent vectors are local operators: If $f, g \in C^\infty(M)$ agree on some neighborhood $U$ of $p$, i.e. $f|_U \equiv g|_U$ then for all $X_p \in T_p M$, $X_p f = X_p g$.
- Technically, the “natural domain” of tangent vectors $X_p \in T_p M$ are germs of functions at $p$: The germ of a function $f \in C^\infty(p)$ is defined as the set of all $g \in C^\infty(p)$ for which there exists an open neighborhood $U$ of $p$ so that $f|_U \equiv g|_U$.
- If $M = \mathbb{R}^m$ then the tangent vectors to $M$ at a point $p$ are precisely the directional derivatives evaluated at $p$. (One direction is obvious. For the other see the calculation below for a general manifold.)
- If $(u, U)$ is a chart about $p \in M$, then for $j = 1, \ldots, m$, $i \frac{\partial}{\partial u^i}|_p \in T_p M$. Thus also
\[
\sum_{j=1}^m a_j \frac{\partial}{\partial u^j}|_p \in T_p M \text{ for all } a^j \in \mathbb{R}.
\]

To verify this assertion, recall the definition
\[
\frac{\partial f}{\partial u^i}|_p = D_i(f \circ u^{-1})|_{u(p)}
\]

and use the familiar properties of partial derivatives in $\mathbb{R}^n$ to manipulate e.g.
\[
\frac{\partial (f \circ g)}{\partial u^i}|_p = D_i((f \circ g) \circ u^{-1})|_{u(p)}
\]
\[
= D_i((f \circ u^{-1}) \cdot (g \circ u^{-1}))|_{u(p)}
\]
\[
= D_i(f \circ u^{-1})|_{u(p)} \cdot (g \circ u^{-1})(u(p)) + (f \circ u^{-1})(u(p)) \cdot D_i(g \circ u^{-1})|_{u(p)}
\]
\[
= \frac{\partial f}{\partial u^i}|_p \cdot g(p) + f(p) \cdot \frac{\partial g}{\partial u^i}|_p
\]

More interesting is the converse, i.e. that in any chart $(u, U)$ about $p \in M$ every tangent vector $X_p \in T_p M$ may be expressed as a linear combination of the partial derivatives $\frac{\partial}{\partial u^i}$, $i = 1, \ldots, m$:

**Theorem 3.2** If $(u, U)$ is a chart about $p \in M^m$ then $\{ \frac{\partial}{\partial u^1}|_p, \ldots, \frac{\partial}{\partial u^m}|_p \}$ is a basis for $T_p M$.

**Corollary 3.3** If $(u, U)$ is a chart about $p \in M^m$ and $X_p \in T_p M$ then $X_p = \sum_{i=1}^m (X_p u^i) \cdot \frac{\partial}{\partial u^i}|_p$.

**Corollary 3.4** If $(u, U)$ and $(v, V)$ are charts about $p \in M^m$ then $\frac{\partial}{\partial v^i}|_p = \sum_{i=j}^m \frac{\partial u^i}{\partial v^j}|_p \cdot \frac{\partial}{\partial u^i}|_p$.

Consider this last corollary as a statement about (linear) bases changes in the tangent space associated to (nonlinear) changes of local coordinates on the manifold.

Before proving this theorem we establish a lemma using an elegant construction.
Lemma 3.5 Let \((u,U)\) be a chart about \(p \in M^m\) with \(u(p) = x_0\) and \(f \in C^\infty(p)\). Then there exist \(f_i \in C^\infty(p)\) such that \(f_i(p) = \frac{\partial f}{\partial u^i}\bigg|_p\) for \(i = 1, \ldots, m\), and \(f(q) = f(p) + \sum_{i=1}^m (u^i(q) - u^i(p)) f_i(q)\).

Compare this assertion to a first-order Taylor approximation. Here we have equality (as opposed to an approximation) – but the functions \(f_i\) are evaluated at the variable point \(q\) as opposed to fixed derivatives evaluated at the fixed point \(p\) in the Taylor approximation.

Proof (of the lemma). Using the local coordinates we reduce the proof to Euclidean spaces: Write \(x = u(q)\) and \(x_0 = u(p)\), rewrite the second statement of the lemma as

\[
f(u^{-1}(x)) = f(u^{-1}(x_0)) + \sum_{i=1}^m (u^i(u^{-1}(x)) - u^i(u^{-1}(x_0))) f_i(u^{-1}(x))
\]

Write \(g\) for \(f \circ u^{-1}: U \subseteq \mathbb{R}^m \mapsto \mathbb{R}\). The desired functions \(g_i: U \subseteq \mathbb{R}^m \mapsto \mathbb{R}\) are such that

\[
g(x) = g(x_0) + \sum_{i=1}^m (x^i - x_0^i) g_i(x).
\]

After shrinking the neighborhood \(U\), if necessary, we may assume that \(u(U) \subseteq \mathbb{R}^m\) is star-shaped with respect to \(x_0\), i.e. for every \(x \in u(U)\) the line segment \(\{x_0 + t \cdot (x - x_0) : t \in [0, 1]\}\) is contained in \(u(U)\). For any fixed \(x \in u(U)\) consider the curve \(\sigma_x: [0, 1] \mapsto u(U)\) defined by \(\sigma(t) = x_0 + t \cdot (x - x_0)\). Via the fundamental theorem of calculus and the chain rule

\[
g(x) = g(\sigma_x(1)) = g(\sigma_x(0)) + \int_0^1 \frac{d}{dt} g(\sigma_x(t)) \, dt
\]

\[
= g(\sigma_x(0)) + \int_0^1 \sum_{j=1}^m (D_j g)(\sigma_x(t)) \cdot \frac{d\sigma_x^j}{dt}(t) \, dt
\]

\[
= g(\sigma_x(0)) + \sum_{j=1}^m (x^j - x_0^j) \int_0^1 (D_j g)(\sigma_x(t)) \, dt
\]

Define as \(g_j(x)\)

Note the constant derivative \(\sigma'_x = (x - x_0)\) for the curve in \(\mathbb{R}^m\) – this makes no sense on a general manifold, but is coordinate dependent. One immediately verifies that with this definition \(g_j(x_0) = (D_j g)(x_0)\) (in this case \(\sigma_x\) is a constant curve). Consequently \(f_j(p) = g_j(u(p)) = (D_j g)(u(p)) = (D_j (f \circ u^{-1}))(u(p)) = \frac{\partial f}{\partial u^i}\bigg|_{u(p)}\). Since \(g \in C^\infty\), also \(g_j \in C^\infty\) and \(f_j \in C^\infty\). ■

Proof (of the theorem). Suppose \((u,U)\) is a chart about \(p \in M^m\), \(X_p \in T_p M\) and \(f \in C^\infty(M)\). Using lemma 3.5 there exist (on some open neighborhood of \(p\)) suitable functions \(f_i\) such that we may rewrite \(X_pf\) as \(X_pf = X_p \left( f(p) + \sum_{j=1}^m (u^j - u^j(p)) f_j \right)\). Using the linearity of \(X_p\), the Leibniz rule, and that \(X_p c = 0\) for any constant function this yields:

\[
X_p f = 0 + \sum_{j=1}^m \left( (X_p(u^j) - 0) f_j(p) + (u^j(p) - u^j(p)) \cdot (X_p f) \right)
\]

i.e. \((X_p f) = \sum_{j=1}^m (X_p u^j) \cdot f_j(p)\). By lemma 3.5 this is equal to \((X_p f) = \sum_{j=1}^m (X_p u^j) \frac{\partial f}{\partial u^i}\bigg|_p\).

Since this holds for all \(f \in C^\infty(p)\), we conclude \(X_p = \sum_{j=1}^m (X_p u^j) \frac{\partial}{\partial u^i}\bigg|_p\). ■
3.3 Tangent maps, part I

For every smooth map \( F: \mathbb{R}^n \to \mathbb{R}^m \) between Euclidean spaces and any point \( x \in \mathbb{R}^n \) the derivative of \( F \) at \( x \) is a linear map \( (DF)(x)(\cdot): \mathbb{R}^n \to \mathbb{R}^m \). Now that we have tangent spaces to manifolds, we are ready to associate analogous (linear) tangent maps (between tangent spaces) to smooth maps (between manifolds).

**Definition 3.2** Suppose \( \Phi: M \to N \) is a smooth map between manifolds and \( p \in M \). The tangent map \( \Phi_*: T_pM \to T_{\Phi(p)}N \) (of \( \Phi \) at \( p \)) is defined for \( X_p \in T_pM \) and \( f \in C^\infty(\Phi(p)) \) by

\[
(\Phi_*X_p)f = X_p(f \circ \Phi)
\]

(40)

How else could \( \Phi_* \) be defined? It is immediate that if \( \Phi = \text{id}_M \) then \( \Phi_* = \text{id}_{T_pM} \). Also, it follows immediately from the definition that

**Proposition 3.6** Suppose \( \Phi: M^m \to N^n \) and \( \Psi: N \to P \) are smooth maps between manifolds, and \( p \in M \). Then (note the preservation of the order of \( \Phi \) and \( \Psi \))

\[
(\Psi \circ \Phi)_* = \Psi_*\Phi_* \circ \Phi_*
\]

(41)

**Exercise 3.2** Prove proposition 3.6.

In local coordinates the tangent map is given by matrix multiplication. More specifically, suppose \( (u,U) \) and \( (v,V) \) are charts about \( p \in M^m \) and \( \Phi(p) \in N^n \), respectively, and \( f \in C^\infty(\Phi(p)) \). From the definitions calculate (using the chain-rule)

\[
\left( \Phi_* \frac{\partial}{\partial u^i} \bigg|_p \right) f = \frac{\partial}{\partial u^i} \bigg|_p (f \circ \Phi)
\]

\[
= D_j(f \circ \Phi \circ u^{-1})|_{u(p)}
\]

\[
= D_j((f \circ v^{-1}) \circ (v \circ \Phi \circ u^{-1}))|_{u(p)}
\]

\[
= \sum_{i=1}^n D_i(f \circ v^{-1})|_{v(p)} \cdot D_j(v^i \circ \Phi \circ u^{-1})|_{u(p)}
\]

\[
= \left( \sum_{i=1}^n \frac{\partial(v^i \circ \Phi)}{\partial u^j} \bigg|_p \cdot \frac{\partial}{\partial u^i} \bigg|_{\Phi(p)} \right) f
\]

In concrete examples, using local coordinates, it is convenient to express tangent vectors as column vectors. E.g. suppose, as before, \( \Phi \in C^\infty(M^m, N^n) \), and \( (u,U) \) and \( (v,V) \) are charts about \( p \in M^m \) and \( \Phi(p) \in N^n \). If \( X_p \in T_pM \) is a tangent vector at \( p \), let \( a = (a^1, \ldots, a^m)^T \) be the column vector with components \( a^i = X_p u^i \), representing \( X_p = \sum_{i=1}^m a^i \frac{\partial}{\partial u^i}\bigg|_p \). Similarly, let \( b = (b^1, \ldots, b^n)^T \) be the column vector with components \( b^j = (\Phi_* X_p) v^j \), representing the image \( \Phi_* X_p \in T_{\Phi(p)}N \). These column vectors \( a \) and \( b \) are related by matrix multiplication \( b = Ca \) where \( C \) is the \((n \times m)\) matrix with components \( C_{ij} = \frac{\partial(v^j \circ \Phi)}{\partial u^i} \bigg|_p \).

Formally, one may go further, and write \( \alpha \) for the row-vector with components \( \alpha_i = \Phi_* \frac{\partial}{\partial u^i} \bigg|_p \) and \( \beta \) for the row-vector with components \( \alpha_i = \frac{\partial}{\partial u^i} \bigg|_{\Phi(p)} \). Then formally, the images of the basis vectors \( \frac{\partial}{\partial u^i}\bigg|_p \) are obtained by right matrix multiplication, i.e. \( \alpha = \beta \cdot C \). This matches with the observation that formally \( \Phi_* (X_p) = aa = (\beta C)a = \beta(Ca) = \beta b \). While this is all simple (formal) matrix algebra, it is worthwhile to remember that when transforming formal vectors of basis elements these are multiplied by the transformation matrix in a way opposite to the multiplication familiar for transforming specific vectors.
Exercise 3.3 Suppose $\Phi \in C^\infty(M^m, N^n)$ and $\Psi \in C^\infty(N^n, Pr)$ are smooth maps between manifolds, $p \in M$ and $f \in C^\infty(P)$. Furthermore, suppose $(u, U)$, $(v, V)$ and $(w, W)$ are local coordinate charts about $p \in M$, $\Phi(p) \in N$ and $(\Psi \circ \Phi)(p) \in P$, respectively. Verify that the matrix representing $(\Psi \circ \Phi)_p$ with respect to $(u, U)$ and $(w, W)$ is the (matrix-)product of the matrices representing $\Phi_p$ (with respect to $(u, U)$ and $(v, V)$) and $\Psi_\Phi(p)$ (with respect to $(v, V)$ and $(w, W)$).

We digress a little to consider tangent spaces of immersed manifolds in $\mathbb{R}^n$ which justify the familiar pictures of tangent planes. Suppose that $\Phi \in C^\infty(M^m, \mathbb{R}^n)$ is an immersion at $p \in M$, i.e. $\text{rank}_p \Phi = m$. Using local coordinates $(u, U)$ about $p \in M^m$ and the standard coordinates $(x, \mathbb{R}^n)$ in the range, the rank condition says that the $(n \times m)$-matrix with components $\frac{\partial x^i \circ \Phi}{\partial u^j}$ has rank $m$, and $\Phi_p$ is a monomorphism (a linear one-to-one map) from $T_p M$ to $T_{\Phi(p)} \mathbb{R}^n$.

The tangent vectors $\Phi_p \left( \frac{\partial}{\partial u^i} \bigg|_p \right) \in \Phi_p (T_p M) \subseteq T_{\Phi(p)} \mathbb{R}^n$ are linearly independent, and span an $m$-dimensional subspace of $T_{\Phi(p)} \mathbb{R}^n$ which is usually pictured as a tangent line/plane, . . . .

The image of any tangent vector $X_p \in T_p M$ in the standard coordinates may be written in the form $\Phi_p X_p = \sum_{i=1}^n b_i D_i |_{\Phi(p)}$. Now, if $f \in C^\infty(\mathbb{R}^n)$ is a function such that $f \circ \Phi \equiv 0$ in a neighborhood of $p \in M$, then

$$0 \equiv X_p (f \circ \Phi) = (\Phi_p X_p) f = \sum_{i=1}^n b_i (D_i f) \bigg|_{\Phi(p)}$$

which in calculus notation might be written as $\Phi_p X_p \perp (\text{grad} f) (\Phi(p))$, or $0 = \langle b, \nabla f (\Phi(p)) \rangle$.

Recall if $(\text{grad} f)(\Phi(p)) \neq 0$, then $f^{-1}(0)$ is (locally) a smooth hypersurface in $\mathbb{R}^n$ (near $\Phi(p)$) and thus $(\Phi_p X_p)$ may be pictured as lying in the tangent hyperplane to the hypersurface $f^{-1}(0)$ at $\Phi(p)$.

Exercise 3.4 Generalize this discussion to the case when there are functions $f^1, \ldots, f^{n-m} \in C^\infty(\Phi(p))$ with $f^i \cdot \Phi \equiv 0$ and linearly independent gradients $(\text{grad} f^i)(\Phi(p))$.

Example 3.1 As a hands-on example consider an immersion of the Moebius-strip into $\mathbb{R}^3$. One way to represent the Moebius-strip is as the quotient $M = R / \sim$ of the rectangle $R = [0, 2\pi] \times (-1, 1)$ two of whose edges have been identified by $(0, t) \sim (2\pi, -t)$. Define a map $\Phi: M \mapsto \mathbb{R}^3$ by $\Phi(\theta, t) = \left( (2 + t \cos(\frac{\theta}{2})) \cos \theta, \ 2 + t \cos(\frac{\theta}{2}) \sin \theta, \ t \sin(\frac{\theta}{2}) \right)$.

Exercise 3.5 Explicitly calculate the images $\Phi_p \left( \frac{\partial}{\partial \theta} \bigg|_p \right)$ and $\Phi_p \left( \frac{\partial}{\partial t} \bigg|_p \right)$ at any point $p = (\theta, t) \in U = (0, 2\pi) \times (-1, 1) \subseteq M$. Verify that $\Phi$ is indeed an immersion.

The image $\Phi(M) \subseteq \mathbb{R}^3$ is a surface in the usual sense. The images of the tangent vectors to $M$ calculated in the exercise may be visualized as the usual arrows that are tangent to the surface. It is easily seen that the map $\Phi$ is indeed well-defined on $M$ (as opposed to only on the rectangle $R$) because $\Phi(0, t) = \Phi(2\pi, -t)$. For $p \in U$ the map $\Phi_p$ is well-defined, but problems arise when trying to extend $\Phi$ and $\Phi_q$ continuously to all $q \in M$. Indeed, $((\theta, t), U)$ is a local coordinate chart of $M$, but it does not cover all of $M$. In the language of the next sections $\Phi_*$ maps the coordinate vector fields $\frac{\partial}{\partial \theta^i} = \frac{\partial}{\partial \theta^i} \Phi$ and $\frac{\partial}{\partial \theta^i} = \frac{\partial}{\partial \theta^i} \Phi$ (which are only defined on $U$) to vector fields on $\Phi(U)$, but the vector field $\frac{\partial}{\partial \theta^i}$ cannot be continuously extended to a vector field on all of $M$. 
We conclude this first section on tangent maps with a generalization of our earlier local submersion theorem.

**Theorem 3.7** Suppose that $\Phi \in C^\infty(M^m,N^n)$ is a smooth map between manifolds and $P^r \subseteq N$ is a smooth submanifold. If

$$\text{for every } p \in \Phi^{-1}(P), \quad \Phi_*(T_pM) + T_{\Phi(p)}P = T_{\Phi(p)}$$

then $\Phi^{-1}(P) \subseteq M$ is a submanifold of $M$ of dimension $(m - (n - r))$.

In general one calls a smooth map $\Phi \in C^\infty(M,N)$ between manifolds transverse to (a submanifold) $P \subseteq N$ along (a submanifold) $L \subseteq M$ if $\Phi_*T_pM + T_{\Phi(p)}P = T_{\Phi(p)}N$ for all $p \in L \cap \Phi^{-1}(P)$.

The theorem motivates the notion of codimensions as opposed to dimensions of submanifolds. More specifically, for an $r$-dimensional submanifold $P^r \subseteq N^n$ of an $n$-dimensional manifold $N$ the codimension of $P$ (in $N$) is defined as $(n - r)$. The theorem then simply states that if $P \subseteq N$ is a submanifold of codimension $k$ and $\Phi$ is transversal to $P$ (along $M$) then $\Phi^{-1}(P) \subseteq M$ is a submanifold of the same codimension $k$.

**Proof.** Suppose that $\Phi \in C^\infty(M^m,N^n)$ and $p \in \Phi^{-1}(P) \subseteq M$ is in the preimage of a smooth submanifold $P^r \subseteq N$. Using theorem 2.13 choose an adapted chart $(v,V)$ about $\Phi(p) \in N$ such that $v(\Phi(p)) = 0$ and such that the restriction of $w = (v^1, \ldots, v^r)$ to the set $W = \{q \in N: v^{r+1}(q) = \ldots = v^n(q) = 0\}$ is a chart of $P$ about $\Phi(p)$.

Define $\Psi: V \mapsto \mathbb{R}^{n-p}$ by $\Psi = (v^{r+1}, \ldots, v^n)$. Let $U = \Phi^{-1}(V) \subseteq M$.

Then $p \in \Phi^{-1} \circ \Psi^{-1}(0) = (\Psi \circ \Phi)^{-1}(0)$ and $(\Psi \circ \Phi)_p = \Psi_{\Phi(p)} \circ \Phi_p: T_pM \mapsto T_0\mathbb{R}^{n-r}$.

Use that $D(\Psi \circ v^{-1}) = (0 \quad I_{n-r})$ and that the kernel of $\Psi_{\Phi(p)}$ is precisely $T_{\Phi(p)}$. Together with $\Phi_*T_pM + T_{\Phi(p)}P = T_{\Phi(p)}N$ this establishes that the restriction of $\Psi_{\Phi(p)}$ to the image of $\Phi_p$ (i.e. to $\Phi_p(T_pM)$) has full rank, and hence rank($\psi \circ \Phi_p$) = $n - p$ (using corollary 2.14). \[\square\]

**Exercise 3.6** Revisit the Hopf map $\Phi: S^3 \mapsto S^2$ (of exercise 2.31), i.e., the restriction (to $S^3$) of $\Phi: \mathbb{R}^4 \mapsto \mathbb{R}^3$ defined by $\Phi(a) = (2a_1a_3 + 2a_2a_4, 2a_2a_3 - 2a_1a_4, a_1^2 + a_2^2 - a_3^2 - a_4^2)$.

Considering the usual imbeddings of the spheres into $\mathbb{R}^4$ and $\mathbb{R}^3$, respectively, describe the preimages $\Phi^{-1}(P_c)$ of the meridians $P_c = S^2 \cap \{(x_1, x_2, x_3) \in \mathbb{R}^3: x_3 = c\}$ for $-1 \leq c \leq 1$. 


3.4 The tangent bundle

It is natural to assemble all tangent spaces of a manifold together into a new object – and conceivably, this set should again have a natural manifold structure. We will omit some of the more technical details of this construction and refer to e.g. Spivak vol.I ch.3, especially exercise 1. As discussed earlier, it is desirable to distinguish between tangent vectors at different points, and we define:

**Definition 3.3** The tangent bundle $TM$ (as a set) of a manifold $M$ is the (disjoint) union of all tangent spaces to $M$ at all points $p \in M$.

$$TM = \left\{ (p, X_p) \in M \times \bigcup_{p \in M} T_pM : X_p \in T_pM \right\}$$

(44)

The bundle projection $\pi : TM \mapsto M$ is defined by $\pi(p, X_p) = p$. The fiber over $p \in M$ is the preimage $\pi^{-1}(p) = \{p\} \times T_pM$.

A section of $TM$, or tangent vector field, is a map $X : M \mapsto TM$ that satisfies $\pi \circ X = \text{id}_M$.

Basically, vector fields are functions that assign to each $p \in M$ a tangent vector in $T_pM$.

Often we conveniently identify the fiber $\pi^{-1}(p)$ with $T_pM$ and the pair $(p, X_p) \in TM$ with tangent vector $X(p) = X_p \in T_pM$. Technically this involves a tacit projection onto the second factor or a tacit use of the inclusion map $\iota_p : T_pM \mapsto \{p\} \times T_pM \subseteq TM$. However, in some instances, e.g. when working with $TTM$, more precision is indicated.

To illustrate that tangent bundles of manifolds are candidates to be considered manifolds themselves, consider the example of $M = S^1$. The naive collection of all tangent lines to the imbedded circle $S^1 \subseteq \mathbb{R}^2$ is full of intersections. More suitable for our purposes is to imbed the circle in $\mathbb{R}^3$ as $\bar{S}^1 = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 = 1, \ x_3 = 0\}$ and attach at every point $p \in \bar{S}^1$ a vertical (!) line, yielding a cylinder. As a set, this cylinder is in bijection with the (disjoint) collection of all tangent lines to the circle imbedded in the plane. Intuitively one can consistently choose an orientation of the lines, and even more a consistent scaling. For example, identify the naive tangent vector $\left((\cos \Theta, \sin \Theta), (-L \sin \Theta, L \cos \Theta)\right)$ with the point $(\cos \Theta, \sin \Theta, L) \in \mathbb{R}^3$.

Within this picture, a vector field on the circle may be visualized as the graph of a function $\Theta \mapsto (\cos \Theta, \sin \Theta) \mapsto L(\Theta)$. If the vector field is continuous and nonvanishing, then the graph lies entirely above, or entirely below the plane $x_3 = 0$.

In analogy, we may intuitively think of the tangent bundle $T\mathbb{R}$ of the real line $\mathbb{R}$ as the plane $\mathbb{R}^2$. However, due to dimensional reasons it is clear that these two examples are the only tangent bundles amenable to such immediate visualization. How quickly things get complicated becomes clear if one tries to think of $TS^2$ as a sphere with a plane attached to each of its points. A vector field on the sphere simply selects one point on each plane. However, from algebraic topology it is known that there does not exist any continuous (yet to be defined!) vector field on the sphere that vanishes nowhere. In our picture this means that it is impossible to continuously select one point on each tangent plane avoiding the origin (zero-vector) in each $T_pS^2$. Intuitively $TS^2$ must be a nontrivially twisted, (when compared to e.g. $TS^1$ which is the very tame cylinder), i.e. it must be very different from the trivial Cartesian product $S^2 \times \mathbb{R}^2$. 
We proceed more abstractly to endow the tangent bundle $TM$ of a smooth manifold $M$ with a manifold structure. The key idea is that locally, above a chart $(u,U)$ (which itself is homeomorphic to $\mathbb{R}^m$) the tangent bundle basically looks like $\mathbb{R}^m \times \mathbb{R}^m \cong \mathbb{R}^{2m}$. This observation is captured in the concept of local triviality, i.e., if $(u,U)$ is a chart of an $m$-dimensional manifold $M$, then $TU \cong U \times \mathbb{R}^m$. Thus the topology and geometry of $M$ are captured in the global structure of the tangent bundle, by how the trivial bundles are pieced together in some twisted way.

Starting with an atlas of local coordinate charts for $M$ we shall find it easy to obtain a candidate atlas of charts for $TM$. There is a natural candidate for a topology on $TM$ that makes all candidate charts for $TM$ into homeomorphisms, and automatically guarantees that their transition maps are smooth. However, it takes a little more advanced arguments to show that this topology is indeed sufficiently nice (metrizable or paracompact) so that $TM$ qualifies as a manifold.

Suppose $(u,U)$ is a chart of $M$ about $p$. Consider the subset $\bar{U} = \pi^{-1}(U) \subseteq TM$. Every point $Q \in \bar{U}$ is a pair $Q = (q, X_q)$ with $X_q \in T_qM$. Since $\{ \frac{\partial}{\partial u^j}|_q : 1 \leq j \leq m \}$ is a basis for $T_qM$ there exists functions $w^j: \bar{U} \mapsto \mathbb{R}$ (indeed, $w^j(Q) = X_q(u^j)$) such that

$$Q = \left( \pi(Q), \sum_{j=1}^{m} w^j(Q) \frac{\partial}{\partial u^j}|_{\pi(Q)} \right).$$

It is natural to define a map $\bar{u}: \bar{U} \mapsto \mathbb{R}^2m$ by

$$\bar{u} = \left( u^1 \circ \pi, \ldots, u^m \circ \pi, w^1, \ldots, w^m \right)$$

It is clear that $\bar{u}$ is a bijection from $\bar{U}$ to $\mathbb{R}^{2m}$ (This assumes that $u$ is a bijection from $U$ to $\mathbb{R}^m$, as originally mandated. Alternatively, $\bar{u}$ is a bijection from $\bar{U}$ onto $u(U) \times \mathbb{R}^m \subseteq \mathbb{R}^{2m}$.)

For each chart $(u,U)$ in the $C^\infty$-differentiable structure of $M$, we want the associated map $\bar{u}$ to be a homeomorphism. This is achieved by endowing $TM$ with the weakest (i.e., coarsest) topology in which all maps $\bar{u}$ are continuous. More constructively, we consider the collection $T$ of all subsets $O \subseteq TM$ which are such that for every point $Q \in TM$ and every chart $(u,U)$ of $M$ for which $\pi(Q) \in U$, there exists an open set $W \in \mathbb{R}^{2m}$ containing $\bar{u}(Q)$ such that $\bar{u}^{-1}(W) \subseteq O$.

**Exercise 3.7** Show that the collection $T$ of subsets of $TM$ is a topology on $TM$, i.e., show that $\emptyset, TM \in T$, and that $T$ is closed under finite intersections and arbitrary unions.

**Exercise 3.8** Check that when $TM$ is endowed with this topology $T$ then for every chart $(u,U)$ the associated map $\bar{u}$ is a homeomorphism, i.e. is continuous with continuous inverse.

We next calculate the transition maps $(\bar{v} \circ \bar{u}^{-1}): \bar{u}(\bar{U} \cap \bar{V}) \subseteq \mathbb{R}^{2m} \mapsto \bar{v}(\bar{U} \cap \bar{V}) \subseteq \mathbb{R}^{2m}$ to verify that the charts are indeed $C^\infty$-related.

Thus assume that $Q \in \bar{U} \cap \bar{V} \subseteq TM$ and that $\bar{u}(Q) = (x,y) \in \mathbb{R}^m \times \mathbb{R}^m$. Calculate

$$\bar{v}(Q) = (\bar{v} \circ \bar{u}^{-1})(x,y) = \bar{v}(u^{-1}(x), \sum_{j=1}^{m} y^j \frac{\partial}{\partial u^j}|_{u^{-1}(x)})$$

(47)
To obtain the second $m$-components of $\bar{v}(Q)$ change the basis in $T_{\pi(Q)}M$ from the \( \left\{ \frac{\partial}{\partial u^i}\big|\pi(Q)\right\}_{j=1}^m \) to the \( \left\{ \frac{\partial}{\partial v^i}\big|\pi(Q)\right\}_{j=1}^m \), yielding

\[
(\bar{v} \circ \bar{u}^{-1})(x,y) = \bar{v}\left((u^{-1}(x), \sum_{j=1}^m y^j \sum_{i=1}^m \frac{\partial v^i}{\partial u^j}|_{u^{-1}(x)} \frac{\partial}{\partial u^i}|_{u^{-1}(x)}\right) \tag{48}
\]

Interchange the order of summation and regroup to read off the components of $\bar{v}(Q)$

\[
(\bar{v} \circ \bar{u}^{-1})(x,y) = ((v \circ u^{-1})(x), \sum_{j=1}^m y^j \frac{\partial v^i}{\partial u^j}|_{u^{-1}(x)}, \ldots, \sum_{j=1}^m y^j \frac{\partial v^m}{\partial u^j}|_{u^{-1}(x)}) \tag{49}
\]

It is easily seen that the map $(\bar{v} \circ \bar{u}^{-1}): \bar{u}(\bar{U} \cap \bar{V}) \subset \mathbb{R}^{2m} \mapsto \bar{v}(\bar{U} \cap \bar{V}) \subset \mathbb{R}^{2m}$ is a smooth map: By hypothesis the first $m$-components are smooth maps. The second $m$ components are linear in $y$ and smooth functions of $x$, and hence the combined map is smooth. If working in the class of $C^\infty$-manifolds, this calculation shows that when one starts with a $C^r$ atlas for $M$, then one obtains, as might be expected, a $C^{r-1}$ atlas for $TM$.

In order for $TM$ to qualify as a smooth manifold we still need that the topology is reasonably nice (metrizable, or equivalently that $(TM,T)$ is paracompact). For the technical details we refer to Spivak vol.I ch.3, especially exercise 1. Here we sketch only some basic ideas. Since a manifold $M$ is locally homeomorphic to a Euclidean space and it is assumed to be metrizable (or equivalently paracompact), each connected component of $M$ is second countable. [[This means that there is a countable basis for the topology on each connected component of $M$. A basis for a topology is a collection of open sets that covers the space, and such that whenever a point $x$ is contained in basic open sets $B_1$ and $B_2$, then there exists a basic open set $B_3$ such that $x \in B_3 \subseteq B_1 \cap B_2$.]] Following Spivak vol.I ch.3 ex. 1, construct a sequence of functions that separates points and closed sets, use these to produce a sequence of bounded metrics $d_i$ and finally piece these together e.g. via $d = \sum_{i=1}^{\infty} 2^{-i}d_i$.

This construction of the tangent bundle shall serve as a model for similar constructions of more general vector bundles in which the tangent spaces $T_pM$ are replaced by other suitable vector spaces. Formally, a vector bundle is a triple $(E,B,\pi)$ (or actually, a five-tuple $(E,B,\pi,\oplus,\odot)$) consisting of a total space $E$, a base space $B$ and a bundle projection $\pi: E \mapsto B$ which is a continuous surjective map. The linear operations $\oplus$ and $\odot$ are defined on the fibres $\pi^{-1}(p)$ for $p \in B$, making each fibre a vector space. A distinguishing condition is that a vector bundle is locally trivial, i.e. every $p \in B$ has an open neighborhood $U$ together with a homeomorphism $\beta: \pi^{-1}(U) \mapsto U \times \mathbb{R}^m$ such that for each $q \in U$ the restriction $\beta|_{\pi^{-1}(q)}$ is a vector space isomorphism from the fibre $\pi^{-1}(q)$ to $\{q\} \times \mathbb{R}^m$.

The Moebius strip is an example of a nontrivial line-bundle over the circle $S^1$. An upcoming section will introduce the cotangent bundle in which the fibre are the spaces of linear functionals on the corresponding tangent spaces.

In some places it is convenient to work with functions that assign to each point $p \in M$ a pair, triple, or $m$-tuple of (co)-tangent vectors. These may be thought of as sections of bundles in which each fibre is a product of two, three, or $n$ copies of the (co)tangent space.

Beyond vector bundles are fibre bundles in which the fibres need not necessarily be vector spaces. Arguably the most important such is a/the principal bundle in which each fibre is a copy of the general linear group $GL(m,\mathbb{R})$ (the space of all invertible linear maps from $\mathbb{R}^m$ to...
A manifold $\mathbb{R}^m$). Its distinguishing feature is that each section $L: M \mapsto P$ acts on e.g. local coordinates via composition: If $(u, U)$ is a chart then $(L \circ u, U)$ is another chart (to be read as $(L \circ u: q \mapsto L_q \circ u(q), U)$ where $L_q: \mathbb{R}^m \mapsto \mathbb{R}^m$ is a linear map.)

$$
\text{To be added:}\quad \text{Use tangent bundles for a geometric definition of orientability for a manifold } M, \text{ or of vector bundle - as opposed to the purely algebraic condition in terms of charts, whether there exists atlas } \mathcal{A} \text{ such that } \det(D(v \circ u^{-1})) > 0 \text{ for all } (u, U), (v, V) \in \mathcal{A}.
$$

We digress with a brief discussion of the (lack of) triviality of the tangent bundles of spheres and its consequences. Consider the usual imbeddings of the spheres $S^m \hookrightarrow \mathbb{R}^{m+1}$ and use the standard coordinates in $\mathbb{R}^{m+1}$. Note that the tangent bundles $TS^m$ are diffeomorphic to the subsets $\{(a, b) \in S^m \times \mathbb{R}^{m+1} : \langle a, b \rangle = 0\} \subseteq \mathbb{R}^{m+1} \times \mathbb{R}^{m+1}$, using the standard inner product in $\mathbb{R}^{m+1}$. [[This is completely different from asserting that $TS^m$ were trivial, or diffeomorphic to $S^m \times \mathbb{R}^m$.]]

When $m = (2k - 1)$ is odd, then $X = x_2D_1 - x_1D_2 + x_4D_3 - x_3D_4 + \ldots x_{2k}D_{2k-1} - x_{2k-1}D_{2k}$ (all $D_j$ evaluated at $x$) is a global nonvanishing (tangent) vector field on $S^m \subseteq \mathbb{R}^{m+1}$. Conversely, sing tools from algebraic topology one may show that if $m = 2k$ is even, then not even a single globally defined continuous nonvanishing vector field exists on the sphere $S^{2k}$. Very special is $S^3$ which admits three smooth vector fields that are everywhere linearly independent:

$$
\begin{align*}
X &= -x_2D_1 + x_1D_2 - x_4D_3 + x_3D_4 \\
Y &= -x_3D_1 + x_4D_2 + x_1D_3 - x_2D_4 \\
Z &= -x_4D_1 - x_3D_2 + x_2D_3 + x_1D_4
\end{align*}
$$

Mimicking (and repeating this construction, similar to the example for $S^{2k-1}$ above) one may construct from these three vector fields on $S^3$ three everywhere linearly independent vector fields on any sphere $S^{4k-1}$. However, it can be shown that on $S^{4k+1}$ any two smooth vector fields are linearly dependent at some point. This example of the frame of the three everywhere linearly independent vector fields on $S^3$ motivates the notion of a parallelizable manifolds [[Abraham-Marsden p.218; Boothby p.219; not in Spivak]]:

**Definition 3.4** A manifold $M^m$ is called parallelizable if it admits a frame of $m$ everywhere linearly independent vector fields. 

It is straightforward to see that a [[finite-dimensional, c.f. Abraham-Marsden]] manifold is parallelizable if and only if its tangent bundle is trivial. So far we have seen that Euclidean spaces $\mathbb{R}^m$, hence all coordinate charts $(u, U)$ are parallelizable. Also, every Lie group is parallelizable.

$$
\text{Should have been done much earlier, in chapter 2: A Lie group is a differentiable manifold } G \text{ with a group structure such that both the multiplication } : G \times G \mapsto G \text{ defined by } (p, q) \mapsto pq \text{ and the inverse } G \mapsto G, \text{ defined by } p \mapsto p^{-1} \text{ are } C^\infty \text{ maps.}
$$

We already encountered several examples of Lie groups: the general linear groups $\text{GL}(n, \mathbb{R})$ of invertible linear maps on $\mathbb{R}^n$, the special linear groups $\text{SL}(n, \mathbb{R})$ (linear maps with determinant one), and the orthogonal groups $\text{O}(n)$ and special orthogonal groups $\text{SO}(n)$. Note that as a manifold $S^1$ is diffeomorphic to $\text{SO}(2)$. Similarly, $S^3$ is a double-cover of the projective space $P^3$ which is diffeomorphic to $\text{SO}(3)$ – thus shedding some light on this most versatile example.
Exercise 3.9 Suppose \( f \in C^\infty(\mathbb{R}^m, \mathbb{R}) \). Show that the graph \( \{ (x, f(x)) : x \in \mathbb{R}^m \} \) is a parallelizable submanifold of \( \mathbb{R}^{m+1} \). Is the same necessarily true for functions \( f : \mathbb{R}^m \mapsto \mathbb{R}^n \)?

Returning to the tangent bundles of the spheres: It is known that the only parallelizable spheres are \( S^1, S^3 \), and \( S^7 \) [Spivak vol.I, ch.3, ex. 19]. It is no coincidence that these are the only dimensions in which one may endow the Euclidean spaces with some a multiplicative structure: In \( \mathbb{R}^2 \) this is the field of complex numbers, in \( \mathbb{R}^4 \) this yields the noncommutative, but still associative quaternions (or Hamilton numbers), and in \( \mathbb{R}^8 \) these are the Cayley numbers whose multiplication is not even associative.

3.5 Smooth vector fields and Lie products

We defined a vector field on a manifold to be a section of the tangent bundle, that is, a function \( X : M \mapsto TM \) such that its composition \( \pi \circ X \) with the bundle projection is the identity on \( M \). Rather than considering arbitrary such functions, our interest is primarily in those that vary smoothly (in topological considerations continuity may suffice). Since a vector field is defined as a map between manifolds \( M \) and \( TM \) we already have a notion of smoothness: A vector field \( X : M \mapsto TM \) is \( C^r \) if for every point \( p \in M \) and coordinate charts \((u, U)\) and \((v, V)\) about \( p \) and \( X(p) \), respectively, the map \( v \circ X \circ u^{-1} : \mathbb{R}^m \mapsto \mathbb{R}^{2m} \) is a \( C^r \)-map between Euclidean spaces. We write \( \Gamma^\infty(M) \) for the set of all smooth vector fields on \( M \).

On the other hand recall that every tangent vector \( X_p \in T_pM \) maps \( C^\infty(p) \mapsto \mathbb{R} \). Consequently, we may view a vector field \( X \) as a mapping of the algebra \( C^\infty(M) \) of smooth functions to itself. We expect that if \( X \) is a smooth vector field and \( f \in C^\infty(M) \) then \( (Xf) \in C^\infty(M) \).

In particular, any smooth vector field is a derivation on the algebra \( C^\infty(M) \) (i.e. it satisfies \( X(fg) = (Xf)g + f(Xg) \) for all \( f, g \in C^\infty(M) \)).

Proposition 3.8 A vector field \( X : M \mapsto TM \) is a \( C^\infty \) vector field, written \( X \in \Gamma^\infty(M) \), if and only if for every open set \( U \subseteq M \) and every function \( f \in C^\infty(U) \) the function \( (Xf) : p \mapsto X(p)f \) is again in \( C^\infty(U) \).

This proposition follows easily from the following exercise upon expanding \( (Xf)(p) \) on a chart \((u, U)\) about \( p \) in terms of local coordinates \( (Xf)(q) = \sum_{j=1}^m (Xu^j)(q) \frac{\partial f}{\partial u^j} \bigg|_q \) and writing out the components of the map \( \bar{v} \circ X \circ u^{-1} : \mathbb{R}^m \mapsto \mathbb{R}^{2m} \).

Exercise 3.10 Verify directly that a vector field \( X : M \mapsto TM \) is \( C^\infty \) if and only if for every coordinate chart \((u, U)\) of \( M \) the functions \( Xu^j : U \mapsto \mathbb{R} \) are smooth.

Since every (smooth) vector field \( X \in \Gamma^\infty(M) \) maps \( C^\infty(M) \) back into itself, it is natural to consider compositions of two vector fields \( X, Y \in \Gamma^\infty(M) \). Clearly \( X \circ Y \), also written \( XY \), is again a map from \( C^\infty(M) \) into itself. However, for two functions \( f, g \in C^\infty(M) \) we calculate

\[
XY(fg) = X\left((Yf)g + f(Yg)\right) = (XYf)g + (Yf)(Xg) + (Xf)(Yg) + f(XYg)
\]  

(51)

In general there is no reason for the terms \((Yf)(Xg)\) and \((Xf)(Yg)\) to cancel each other, hence in general \( XY \) is not a derivation, and thus is not a vector field! However, the commutator \( XY - YX \) clearly will be a derivation. Thus we have a product structure on \( \Gamma^\infty(M) - \) which equips this space of all smooth vector field with an important algebraic structure that invites deeper study:
Definition 3.5 The Lie bracket or Lie product of vector fields is the map $[\cdot, \cdot]: \Gamma^\infty(M) \times \Gamma^\infty(M) \mapsto \Gamma^\infty(M)$, defined for $f \in C^\infty(M)$ by $[X,Y]f = X(Yf) - Y(Xf)$.

Exercise 3.11 Let $\xi = (\xi^1, \ldots, \xi^n)^T$ and $\eta = (\eta^1, \ldots, \eta^n)^T$ be column vector fields representing two vector fields $X, Y \in \Gamma^\infty(M)$ in a coordinate chart $(u, U)$, i.e. $\xi^i = (Xu^i)$ and $\eta^i = (Yu^i)$. Verify that in these coordinates the Lie product $[X, Y]$ is represented by the column vector $(D\eta)\xi - (D\xi)\eta$ where $D$ denotes the Jacobian matrix of partial derivatives.

Definition 3.6 A linear vector space $L$ equipped with a bilinear mapping $[\cdot, \cdot]: L \times L \mapsto L$ is a Lie algebra if this map is anti-commutative and satisfies the Jacobi identity:

$$
\begin{align*}
\text{for all } x, y \in L, \quad & 0 = [x, y] + [y, x] \\
\text{for all } x, y, z \in L, \quad & 0 = [x, [y, z]] + [y, [z, x]] + [z, [x, y]]
\end{align*}
$$ (52)

Exercise 3.12 Verify that $\mathbb{R}^3$ equipped with the standard cross product is a Lie algebra.

Exercise 3.13 Verify that the space $\mathfrak{so}(3)$ of skew symmetric $3 \times 3$ matrices with the product $[A, B] = AB - BA$ (matrix product) is a three dimensional Lie algebra. Find a basis for $\mathfrak{so}(3)$ and establish a Lie algebra isomorphism from $\mathfrak{so}(3)$ to $\mathbb{R}^3$ with the cross-product – i.e. explicitly give a bijective linear map (between vector spaces) that is also a Lie algebra homomorphism, meaning in this case $\Phi([A, B]) = \Phi(A) \times \Phi(B)$ for all $A, B \in \mathfrak{so}(3)$.

Exercise 3.14 Verify by direct calculation that the Lie product of vector fields as defined above equips $\Gamma^\infty(M)$ with a Lie algebra structure. Note that this means verifying that $[\cdot, \cdot]$ is linear over $\mathbb{R}$, i.e. $[aX + Y, Z] = a[X, Z] + [Y, Z]$, that it is anti-commutative (obvious) and that it satisfies the Jacobi identity – simply expand $[X, [Y, Z]]f + [Y, [Z, X]]f + [Z, [X, Y]]f$.

Exercise 3.15 Show that any associative algebra $(A, \cdot)$ is a Lie algebra under the commutator product defined by $[x, y] = x \cdot y - y \cdot x$. In particular, the set $D(A)$ of derivations on an associative algebra, that is of linear maps $\ell: A \mapsto A$ satisfying $\ell(xy) = (\ell(x))y + x(\ell(y))$ for all $x, y \in A$ is an associative algebra under composition and thus a Lie algebra under the commutator as above.

Exercise 3.16 (This is a preview of an example which geometrically is situated in the cotangent bundle and symplectic geometry). On the set of all smooth functions $C^\infty(\mathbb{R}^{2m})$ define a product, the Poisson bracket, using coordinates $(q_1, \ldots, q_m, p_1, \ldots, p_m)$ on $\mathbb{R}^{2m}$ by

$$
\{f, g\} = \sum_{i=1}^m \frac{\partial f}{\partial p^i} \frac{\partial g}{\partial q^i} - \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p^i} .
$$ (53)

Verify that this product equips $C^\infty(\mathbb{R}^{2m})$ with the structure of a Lie algebra.

Recall that if $X, Y \in \Gamma^\infty(M)$ and $f, g \in C^\infty(M)$ then $fX \in \Gamma^\infty(M)$, and the usual distributive and mixed associative properties hold, e.g. $(fg)X = f(gX)$, $f(X + Y) = fX + fY$, $(f + g)X = fX + gX$, $1 \cdot X = X$, \ldots. This means that $\Gamma^\infty(M)$ is not only a vector space over $\mathbb{R}$, but also a (left) $C^\infty(M)$-module. (It is not a vector space over $C^\infty(M)$ since the ring of smooth functions
is not a field.) Given this $C^\infty(M)$-module structure it is natural to ask how the Lie bracket on $\Gamma^\infty(M)$ relates to it. For $X,Y \in \Gamma^\infty(M)$ and $f,g \in C^\infty(M)$ we calculate

\[
[X,Y] g = (fX)(Yg) - Y(fX(g)) = f\left( X(Yg) - Y(Xg) \right) - (Yf) \cdot (Xg) = \left( f[X,Y] - (Yf)X \right) g
\]

and conclude that the Lie bracket $[\cdot,\cdot]$ is not linear over $C^\infty(M)$.

In a chart $(u,U)$ (compare also exercise 3.20) we calculate

\[
\left[ \frac{\partial}{\partial u^i} , \frac{\partial}{\partial u^j} \right] f = \frac{\partial}{\partial u^i} (D_j (f \circ u^{-1}) \circ u) - \frac{\partial}{\partial u^j} (D_i (f \circ u^{-1}) \circ u) = D_i (D_j (f \circ u^{-1}) \circ u \circ u^{-1}) \circ u - D_j (D_i (f \circ u^{-1}) \circ u \circ u^{-1}) \circ u = \left( D_i D_j (f \circ u^{-1}) - D_j D_i (f \circ u^{-1}) \right) \circ u = 0
\]

since the mixed partial derivatives on $C^\infty \mathbb{R}^m$ are equal. As an important corollary we obtain:

**Proposition 3.9** If $X,Y \in \Gamma^\infty(M)$ and $U \subseteq M$ is an open set with $[X,Y] \big|_U \neq 0$ then there does not exist a map $u: U \mapsto \mathbb{R}^m$ such that $(u,U)$ is a chart on $M$ with $X\big|_U = \frac{\partial}{\partial u^i}$ and $Y\big|_U = \frac{\partial}{\partial u^j}$.

Indeed, in subsequent sections we will see that in the neighborhood of any point $p$ at which a smooth vector field $X$ does not vanish, there are always coordinates $(u,U)$ such that $X = \frac{\partial}{\partial u^i}$. On the other hand, generalizing the above criterion to sets of vector fields will lead to important Frobenius integrability theorem.

**Exercise 3.17** [[This exercise is somewhat frivolous – but it is a good practice for hands-on calculations, and it hits hard at common misperceptions.]] Consider the upper half plane $M = \{(x_1,x_2) \in \mathbb{R}^2 : x_2 > 0\}$ with standard rectangular coordinates $(x,y)$, with polar coordinates $(r,\theta)$ and with the mixed coordinates $(\rho,\xi)$ defined by $\rho = r$ and $\xi = x$.

- Explicitly express the coordinate vector fields $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \theta}$ as linear combinations of $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$. (In particular express the coefficients in terms of $x$ and $y$).
- Use these expressions to verify by direct calculation that $[\frac{\partial}{\partial r} , \frac{\partial}{\partial \theta}] \equiv 0$.
- Verify by direct calculation that $[\frac{\partial}{\partial \rho} , \frac{\partial}{\partial \xi}] \neq 0$.
- Explain why this does not contradict that $(\xi,\rho) = (x,r)$ are admissible local coordinates. Calculate the $(\xi,\rho)$ coordinates of the points $(1,0.1)$, $(1,1)$, and $(0,1)$.
- Calculate $\frac{\partial}{\partial \xi}$ and $\frac{\partial}{\partial r}$, e.g. write these as linear combinations of $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$, and sketch these coordinate vector fields as arrows in the half-plane. Describe in words in which directions these arrows point.
- Explain the basis for possible misconceptions. In some sense, for partial derivatives it is less important what varies than what held fixed ... Revisit this discussion later using differential forms $dx, d\xi, d\rho, d\eta$. 


3.6 The tangent map and vector fields

Having assembled the tangent spaces $T_pM$ at all points $p \in M$ into the tangent bundle $TM$ as a manifold, it is natural to combine the tangent maps $\Phi_{*p}$ associated to a map $\Phi \in C^\infty(M,N)$ between manifolds into a map $\Phi_* : TM \mapsto TN$. This is a straightforward definition with no or few ensuing surprises. However, in general, tangent maps need not map vector fields to vector fields.

**Definition 3.7** For any map $\Phi \in C^\infty(M,N)$ define an associated tangent map $\Phi_* : TM \mapsto TN$ for $q \in M$, $(q, X_q) \in \pi^{-1}(q)$ by

$$\Phi_*(q, X_q) = (\Phi(q), \Phi_{*q}(X_q))$$ (56)

Note that the tangent map $\Phi_*$ has the map $\Phi$ built in. It is straightforward to verify the following:

**Proposition 3.10** If $\Phi \in C^\infty(M,N)$ and $\Psi \in C^\infty(N,P)$ then (note preservation of order)

$$(\Psi \circ \Phi)_* = \Psi_* \circ \Phi_*$$ (57)

**Exercise 3.18** Show that if $\Phi \in C^\infty(M,N)$ then $\Phi_* \in C^\infty(TM,TN)$. (Use the definition of differentiability of a map between manifolds in terms of charts $(\bar{u}, \bar{U})$ and $(\bar{v}, \bar{V})$ for $M$ and $N$, respectively.)

It is important to understand that in general there is no hope that a tangent map associated to a smooth $\Phi : M \mapsto N$ between manifold will map a vector field $X$ on $M$ to a vector field on $N$. This is immediately clear if we recall that a vector field on $N$ is a function from $N$ to $TN$. Thus if $p_1 \neq p_2 \in M$ but $\Phi(p_1) = \Phi(p_2) \in N$ then problems arise unless $\Phi_{*p_1} X_{p_1} = \Phi_{*p_2} X_{p_2}$. Similarly, if $\Phi$ is not onto, then $\Phi_*$ can at best yield a partially defined vector field on $N$.

If $\Phi : M \mapsto N$ is a diffeomorphism and $X \in \Gamma^\infty(M)$ then we define $\Phi_* X : N \mapsto TN$ by

$$(\Phi_* X)_q = \Phi_{*\Phi^{-1}(q)} X_{\Phi^{-1}(q)} \quad \text{for} \quad q \in N.$$ (58)

Sometimes this is written suggestively as $(\Phi_* X)_q = (\Phi_* \circ X \circ \Phi^{-1})(q)$. It is clear that $\pi \circ (\Phi_* X) = \text{id}_N$ and that $\Phi_*$ is a smooth map, and hence $(\Phi_* X) \in \Gamma^\infty(N)$.

The most important application is when $\Phi = u : U \mapsto \mathbb{R}^m$ is a coordinate map. Indeed, we have routinely used the map $u_*$ which maps e.g. coordinate vector fields $\frac{\partial}{\partial x_j}$ to the fields $D_j$ on $\mathbb{R}^m$. Note that it is not required for $\Phi$ to be a diffeomorphism in order for $(\Phi_* X)$ to make sense as a vector field on $N$ – as long as $\Phi$ is a smooth map such that $p_1 = p_2 \in M$ implies $\Phi_{*p_1} X_{p_1} = \Phi_{*p_2} X_{p_2}$ the definition (58) still makes sense. The following example gives a preview on how this may be used in the case that the vector field has some infinitesimal symmetries as they will be defined in the section on Lie derivatives.

**Exercise 3.19** Consider $M = \mathbb{R}^2 \setminus \{0\}$ and $N = \mathbb{P}^1$. Let $X(x) = (ax^1 + bx^2) \, D_1|_x + (cx^1 + dx^2) \, D_2|_x$ be a linear vector field on $M$. Define a relation $\sim$ on $M$ by $x \sim y$ if there exists $\lambda \in \mathbb{R}$ such that $x = \lambda y$. Verify that $\sim$ is an equivalence relation on $M$.

Let $\Phi : M \mapsto N = \mathbb{P}^1 = \overline{\mathbb{R}} / \sim$ be the canonical projection map which maps each $x \in M$ to its equivalence class $[x] = \{y \in M : y \sim x\}$. Verify that if $x \sim y$ then $\Phi_{*x}(X_x) = \Phi_{*y}(X_y)$ and hence we may $\Phi_*(X)$ does define a smooth vector field on $\mathbb{P}^1$.

Consider the local coordinate chart $(m, U)$ on $\mathbb{P}^1$ where $U = \{(x_1, x_2) : x_1 \neq 0\}$ and $m(\{(x_1, x_2)\})$
is the slope of the line through the points $[(x_1, x_2)]$, i.e. $m([(x_1, x_2)]) = \frac{x_2}{x_1}$. Find an explicit expression for $(\Phi_X)_m = f(m) \frac{\partial}{\partial m} m$. Interpret $(\Phi_X)$ as defining a dynamical system (via $\dot{m} = f(m)$) on the space of lines through the origin. In detail discuss the special cases when $a = d$ and either $b = c = 0$ or $b = -c = -1$. In general relate the stationary points of $\Phi_X$ (i.e. the zeros of $f(\theta)$) to the eigenspaces of the $2 \times 2$-matrix with entries $a, b, c$ and $d$.

Exercise 3.20 Extend the previous exercise 3.19 to a higher dimensional case. Let $X(x) = \sum_{i,j=1}^m a_{ij} x^i D_j |_x$ be a linear vector field on $\mathbb{R}^m$. Use the coordinates $y = (y^1, \ldots, y^{m-1})$ on the subset $U \subseteq \mathbb{R}^{m-1} = (\mathbb{R}^m \setminus \{0\})/\sim$ defined by $y^j([x]) = \frac{x^j}{x^m}$ and verify that in these coordinates $(\Phi_* X)$ is a quadratic vector field (representing a Riccati differential equation).

As an illustration explicitly write out the formula for $(\Phi_* X)$ in the case of $m = 1$. For fun explore the case where the matrix $(a_{ij})$ has a triple eigenvalue with a single Jordan block, e.g. $a_{ii} = \lambda \neq 0$, $a_{12} = a_{23} = 1$ and $a_{ij} = 0$ else. In particular, sketch the phase portrait for $(\Phi_* X)$ near $y = 0$ and relate it to the integral curves on $\mathbb{P}^2$ (or on $S^2$ which may be easier to visualize).

3.7 The cotangent bundle and differential one forms

Associated to each tangent space $T_p M$ of a manifold $M$ at a point $p$ is a well-defined dual space whose elements are the linear functionals on $T_p M$. Assembling all these dual spaces one obtains the cotangent bundle. Its sections, the analogues to (tangent) vector fields, are differential forms. While such dual objects appear to be considerably less tangible to the novice, they do have better algebraic properties than tangent vector fields. This makes them the preferred choice in the many settings where one may choose between describing objects and properties using tangent fields or cotangent fields. We begin with a brief linear algebra review.

Let $V$ be a finite dimensional vector space (over a field, here always taken to be $\mathbb{R}$). A linear functional on $V$ is a linear map $\lambda: V \mapsto \mathbb{R}$ (i.e. $\lambda(cv + w) = c\lambda(v) + \lambda(w)$ for all $v, w \in V$ and all $c \in \mathbb{R}$). The set $V^*$ of all linear functionals on $V$ inherits a scalar multiplication and addition from the range $\mathbb{R}$, i.e. for linear functionals $\lambda_1, \lambda_2$ on $V$, $c \in \mathbb{R}$, and $v \in V$ define $(c\lambda_1 + \lambda_2)(v) = c\lambda_1(v) + \lambda_2(v)$. It is a straightforward to check that with these operations the set $V^*$ is a vector space over $\mathbb{R}$.

Exercise 3.21 Suppose $\beta = \{v_1, \ldots, v_m\}$ is a basis for a vector space $V$. Consider the maps $\lambda^i: V \mapsto \mathbb{R}$ defined by

$$\lambda^i \left( \sum_{j=1}^m c^j v_j \right) = c^i \quad \text{where} \quad c^k \in \mathbb{R}. \quad (59)$$

- Verify that $\lambda^i \in V^*$.
- Show that $\gamma = \{\lambda^1, \ldots, \lambda^m\}$ are linearly independent.
- Show that every linear functional $\lambda \in V^*$ is a linear combination of $\gamma$.

The exercise establishes, in particular, that $V^*$ is of the same dimension as $V$. The basis $\gamma$ for $V^*$, described in this exercise, is called the dual basis to $\beta$.

Novices to linear algebra often seem troubled that unlike the elements of the given vector space $V$ the elements of $V^*$ seem to be less tangible, and that they can be represented by an somewhat arbitrary collection of different objects. However, this drawback is easily compensated for by their superior algebraic properties... The following exercise may help a little pinning down what the linear functionals are (and what they are not).
Exercise 3.22 [[This is not meant to be deep, but should be fun and provide a hands-on different point of view.]] Consider the vector space \( V \) of all quadratic polynomial functions on the real line. In the usual shorthand notation \( V = \{ a + bx + cx^2 : a, b, c \in \mathbb{R} \} \).

- Verify that \( \lambda^1 : p \mapsto p(1), \lambda^2 : p \mapsto p''(23), \lambda^3 : p \mapsto \int_0^1 p(t) \, dt, \) and \( \lambda^4 : p \mapsto \int_{-\infty}^\infty e^{-t^2} \, p(t) \, dt \), are linear functionals on \( V \).
- Show that \( \{ \lambda^1, \lambda^2, \lambda^3 \} \) is a basis for \( V^* \).
- Write \( \lambda^4 \) as a linear combinations of \( \lambda^1, \lambda^2 \) and \( \lambda^3 \).
- Find a basis for \( V^* \) that is dual to the basis \( \{ 1, x, x^2 \} \) for \( V \).
- Find a basis for \( V \) that is dual to the point evaluations \( e^j : p \mapsto p(j) \) for \( j = 1, 2, 3 \).
- Explain why for every fixed integer \( N > 0 \) and every fixed interval \( [a, b] \) there exist fixed numbers \( \alpha_j, \xi_j \in \mathbb{R} \) (not necessarily in \( [a, b] \)) such that for every polynomial function \( p \) of degree at most \( (N - 1) \), \( \int_a^b p(t) \, dt = \sum_{j=1}^N \alpha_j p(\xi_j) \). (E.g. use that Vandermonde matrices are nonsingular.)

This example will be revisited in the next chapter in the context of inner product spaces.

Returning to differential geometry, define

Definition 3.8 Suppose \( M \) is a smooth manifold and \( p \in M \). The cotangent space to \( M \) at \( p \), denoted \( T^*_pM \), is the space of all linear functionals on \( T_pM \), i.e. \( T^*_pM = (T_pM)^* \).

Recall that we defined tangent vectors \( X_p \in T_pM \) to be linear mappings from \( C^\infty(p) \) to \( \mathbb{R} \). Turning this around we define:

Definition 3.9 For \( p \in M \) and \( f \in C^\infty(p) \) define a map \( (df)_p : T_pM \mapsto \mathbb{R} \), called the differential of \( f \) at \( p \), by

\[
(df)_p(X_p) = (X_pf)
\]

Exercise 3.23 Verify that for each \( p \in M \) and each \( f \in C^\infty(p) \) the differential \( (df)_p \) is a linear functional on \( T_pM \), i.e. \( (df)_p \in T^*_pM \).

Proposition 3.11 Suppose that \( (u, U) \) is a chart about \( p \in M^m \). Then the set \( \{(du^1)_p, \ldots, (du^m)_p\} \) of differentials at \( p \) is a basis for \( T^*_pM \), dual to the basis \( \{ \frac{\partial}{\partial u^1}|_p, \ldots, \frac{\partial}{\partial u^m}|_p \} \) of \( T_pM \).

Proof. From the definition it is clear that

\[
(du^i)_p(\frac{\partial}{\partial u^j}|_p) = D_j(u^i \circ u^{-1})|_u(p) = \delta^i_j
\]

which shows the linear independence of \( \gamma = \{(du^1)_p, \ldots, (du^m)_p\} \). Since the cardinality of \( \gamma \) matches the dimension on \( T_pM \), this also establishes that \( \gamma \) is a basis for \( T^*_pM \). \( \blacksquare \)

Note that in a chart \( (u, U) \) the coordinates \( \omega_j \) of any element \( \omega = \sum_{j=1}^m \omega_j (du^j)_p \in T^*_pM \) are immediately obtained by evaluating \( \omega_j = \omega(\frac{\partial}{\partial u^j}|_p) \). In particular, if \( f, g \in C^\infty(p) \) are such that for all \( j = 1, \ldots, m \), \( \frac{\partial f}{\partial u^j}|_p = \frac{\partial g}{\partial u^j}|_p \) then \( (df)_p = (dg)_p \) as elements of \( T^*_p(M) \).

It is useful to compare the notion of differential forms developed here to the common usage in calculus. For illustration consider the function \( z = x^2 + y^2 \) (i.e. \( z : \mathbb{R}^2 \mapsto \mathbb{R} \)), whose differential is \( dz = 2x \, dx + 2y \, dy \). Commonly \( dz \) is considered as a function of the four variables \( x, y, z \),...
often one finds some ambiguous language that characterizes the differentials \( dx \), \( dy \), and \( dz \) as infinitesimal objects, yet allows the function \( dz \) to be evaluated at a point like \((x, y, dx, dy) = (2, 3, 0.2, -0.1)\). Thus \( dz \) is now considered as a function \( dz: \mathbb{R}^4 \to \mathbb{R} \). It is apparent that \((2, 3, 0.2, -0.1)\) denotes (are the coordinates of) the (infinitesimal?) tangent vector \((0.2, -0.1)\) at \((2, 3)\). In our notation this tangent vector is written as \(0.2 \frac{\partial}{\partial x}(2,3) - 0.1 \frac{\partial}{\partial y}(2,3)\). Note that it is quite consistent with our language to use \(dx\) and \(dy\) as coordinates in the tangent plane – we merely may regard them as linear functions, here on \(\mathbb{R}^2\).

After this brief discussion of smoothness and local representations we take a look at the algebraic structure and properties of \(\Omega^1\). In particular, \(dx = 0.2\) is simply a shorthand for \((dx)(2,3)(0.2 \frac{\partial}{\partial x}(2,3) - 0.1 \frac{\partial}{\partial y}(2,3)) = 0.2\). Indeed, with differential forms we now alternatively may express a tangent vector \(X_p \in T_pM\) in a chart \((u, U)\) about \(p\) as

\[
X_p = \sum_{j=1}^{m} (X_p u^j) \frac{\partial}{\partial u^j} \bigg|_p \quad \text{or} \quad X_p = \sum_{j=1}^{m} (du^j)_p (X_p) \frac{\partial}{\partial u^j} \bigg|_p
\]

and \((du^j)_p, j = 1, \ldots, m\) are legitimate coordinate functions, or simply “coordinates” (?) of tangent vectors.

In complete analogy to the tangent bundle we assemble all cotangent spaces \(T^*_p\) into the cotangent bundle, denoted \(T^*M\). It is a vector bundle over \(M\) with bundle projection again denoted by \(\pi\). For any chart \((u, U)\) of \(M\) define \(\tilde{U} = \pi^{-1}(U)\), and, \(\tilde{u}: \tilde{U} \to \mathbb{R}^{2m}\) by

\[
\tilde{u}(p, \omega_p) = (u^1(p), \ldots, u^m(p), \omega_p(\frac{\partial}{\partial u^1}) |_p, \ldots, \omega_p(\frac{\partial}{\partial u^m}) |_p).
\]

Using proposition 3.11 it is clear that \(\tilde{u}\) is a bijection onto its image. As in the case of \(TM\), it is possible to equip \(T^*M\) with a topology such that the maps \(\tilde{u}\) are homeomorphisms (onto their respective images). In more technical work one may show that the topology is metrizable, and via the next exercise, \(T^*M\) is a smooth manifold.

**Exercise 3.24** Suppose \((u, U)\) and \((v, V)\) are charts on \(M\), and \((\tilde{u}, \tilde{U}), (\tilde{v}, \tilde{V})\) are defined as above. Verify that the transition maps \(\tilde{v} \circ \tilde{u}^{-1}: \tilde{u} (\tilde{U} \cap \tilde{V}) \to \tilde{v} (\tilde{U} \cap \tilde{V}) \subseteq \mathbb{R}^{2m}\) are smooth bijections between subsets of Euclidean spaces.

**Definition 3.10** The (smooth) sections of the cotangent bundle, that is, the (smooth) functions \(\omega: M \to T^*M\) satisfying \(\pi \circ \omega = \text{id}_M\) are called (smooth) differential one forms. The space of all smooth differential one-forms on \(M\) is denoted by \(\Omega^1(M)\).

As a map between smooth manifolds, a section \(\omega: M \to T^*M\) is smooth if for all coordinate charts \((u, U)\) of \(M\) and \((\tilde{v}, \tilde{V})\) of \(T^*M\) the maps \(\tilde{v} \circ \omega \circ u^{-1}: u(U \cap V) \to \tilde{v}(\tilde{U} \cap \tilde{V}) \subseteq \mathbb{R}^{2m}\) are smooth as maps from subsets of \(\mathbb{R}^m\) to subsets of \(\mathbb{R}^{2m}\). In particular, in a chart \((u, U)\) a differential one-form \(\omega \in \Omega^1(M)\) may be written as a linear combination \(\omega = \sum_{j=1}^{m} \omega_j \, du^j\) with smooth functions \(\omega_j \in C^\infty(M)\) defined by \(\omega_j = \omega(\frac{\partial}{\partial u^j})\).

For every function \(f \in C^\infty(U)\) defined on an open subset \(U \subseteq M\) define a smooth differential one-form \(df \in \Omega^1(U)\) by \((df)(p) = (df)_p\) for \(p \in U\). (More pedantic writers may prefer \((df)(p) = (p, (df)_p))\.) In a coordinate chart \((u, U)\) this becomes \(df = \sum_{j=1}^{m} \frac{\partial f}{\partial u^j} \, du^j\).

After this brief discussion of smoothness and local representations we take a look at the algebraic structure and properties of \(\Omega^1(M)\). We already have routinely combined smooth functions \(f, g \in \)
$C^\infty(M)$ and differential forms $\omega, \eta \in \Omega^1(M)$ to write e.g. $(f \omega + g \eta)$. A brief reflection shows, that this is permissible, and indeed yields new differential forms in $\Omega^1(M)$: Indeed any $T^*_pM$ is a $\mathbb{R}$-vector space and for any $X_p \in T_pM$ we interpret $(f \omega + g \eta)_p(X_p)$ as $f(p)\omega_p(X_p) + g(p)\eta_p(X_p)$. The next exercise addresses the smoothness.

**Exercise 3.25** Suppose $f, g \in C^\infty(M)$ and $\omega, \eta \in \Omega^1(M)$. Argue (from the definition of smoothness of maps between manifolds) why $f \omega + g \eta$ is indeed a smooth differential form on $M$.

It is a straightforward to verify that the usual mixed associative and mixed distributive laws hold, and thus $\omega \in \Omega^1(M)$ has the structure of a $C^\infty(M)$ module.

On the other hand, every differential one-form $\omega \in \Omega^1(M)$ is naturally also a functional that maps $\omega: \Gamma^\infty(M) \mapsto C^\infty(M)$, defined pointwise by $\omega(X)(p) = \omega_p(X_p)$. To verify that $\omega(X)$ is indeed a smooth map locally expand $\omega(X)$ in a coordinate chart $(u,U)$

$$
\omega(X) = \left(\sum_{i=1}^{m} \omega_i \frac{\partial}{\partial u^i}\right) \left(\sum_{j=1}^{m} (Xu^j) \frac{\partial}{\partial u^j}\right) = \sum_{i=1}^{m} \omega_i \cdot (Xu^i) 
$$

and use that $\omega_i$ and $(Xu^i)$ are smooth functions since $\omega$ and $X$ are a smooth differential form and a smooth vector field, respectively.

Moreover, one readily observes that if $f \in C^\infty(M)$ (in a chart $(u,U)$

$$
\omega(fX) = \sum_{i=1}^{m} \omega_i \frac{\partial}{\partial u^i} \left(\sum_{j=1}^{m} (fXu^j) \frac{\partial}{\partial u^j}\right) = \sum_{i=1}^{m} f \cdot \omega_i \cdot Xu^i = f \sum_{i=1}^{m} \omega_i \frac{\partial}{\partial u^i} \left(\sum_{j=1}^{m} (Xu^j) \frac{\partial}{\partial u^j}\right) = f \omega(X)
$$

establishing that any $\omega \in \Omega^1(M)$ is not only an $\mathbb{R}$-linear map, but indeed a $C^\infty(M)$-linear map from $\omega \in \Gamma^\infty(M)$ to $C^\infty(M)$, written

$$
\Omega^1(M) \subseteq \text{Hom}_{C^\infty(M)}(\Gamma^\infty(M), C^\infty(M))
$$

### 3.8 Cotangent maps and pullbacks of differential forms

The next step is to analyze the analogues of the tangent maps associated to a smooth function between manifolds. Recall from linear algebra that every linear map $\phi: V \mapsto W$ between vector spaces induces a dual map $\phi^*: W^* \mapsto V^*$, defined by $(\phi^*\lambda)(v) = (\lambda \circ \phi)(v)$ for $v \in V$ and $\lambda \in W^*$.

**Definition 3.11** Suppose $\Phi \in C^\infty(M,N)$ is a smooth map and $p \in M$. Define the cotangent map $\Phi^*_p: T^*_p \mapsto T^*_p$, as the dual of the tangent map $\Phi_p$, i.e. $\Phi^*_p = (\Phi_p)^*$.

Note that this means if $\omega_{\Phi(p)} \in T^*_p N$ and $X_p \in T_p M$ then

$$
(\Phi^*_p \omega_{\Phi(p)})(X_p) = \omega_{\Phi(p)}(\Phi_p X_p)
$$

**Exercise 3.26** Let $\Phi \in C^\infty(M,N)$, $\Psi \in C^\infty(N,P)$, and $p \in M$. Verify $(\Psi \circ \Phi)^*_p = \Phi^*_p \circ \Psi^*_p$.

Very unlike the situation of the tangent bundle it is in general not possible to combine all maps $\Phi^*_p$, $p \in M$ together to get a well-defined map from $T^*N$ to $T^*M$. Indeed, the first hint at problems is that the maps $\Phi^*_p$ are naturally indexed not by their domains but by their codomains! Indeed, if $p, q \in M$ are such that $z = \Phi(p) = \Phi(q) \in N$ then there are well-defined maps $\Phi^*_p: TN \mapsto T^*_p M$, $\Phi^*_q: TN \mapsto T^*_q M$ with the same domain, but different ranges (unless
Suppose that \(\Phi: M \rightarrow N\) is a diffeomorphism (i.e. \(\Phi\) is a homeomorphism) and \(\omega \in \Omega^1(N)\) define the pullback \(\Phi^*\omega: M \rightarrow T^*M\) of \(\omega\) by \(\Phi\) at \(p \in M\) to \(\Phi(p) \in N\) by
\[
(\Phi^*\omega)(X_p) = \omega_{\Phi(p)}(\Phi_*X_p)
\]

**Exercise 3.27** Suppose \(\Phi \in C^\infty(M, N)\) and \(\omega \in \Omega^1(N)\). Verify directly that \((\Phi^*\omega) \in \Omega^1(M)\), i.e. that \(\Phi^*\omega\) is smooth.

This is a good place to comment about some unfortunate terminology. Associated to a map \(\Phi: M \rightarrow N\) are two maps, \(\Phi_*: TM \rightarrow TN\), going in the same direction, and \(\Phi^*: T^*N \rightarrow T^*M\), going in the opposite direction. Modern language would use the attribute *covariant* for the first, and the attribute *contravariant* for the latter. Unfortunately, classical language used the same words for co-tangent and tangent vector fields. Quoting from Spivak vol.I, p.156 “...and no one had the gall or authority to reverse terminology so sanctioned by years of usage. So it’s very easy to remember which kind of vector field is covariant, and which is contravariant – it’s just the opposite of what it logically ought to be. (i.e. sections \(X: M \rightarrow TM\) are called contravariant vector fields, and sections \(\omega: M \rightarrow T^*M\) are called covariant vector fields ...)

Pullbacks of cotangent vector fields are especially useful when working with imbedded submanifolds. More specifically, suppose that \(M \subseteq N\) is a submanifold and consider the inclusion map \(\iota: M \rightarrow N\). Then every differential form \(\omega \in \Omega^1(N)\) immediately gives rise to a differential form \(\iota^*(\omega) \in \Omega^1(M)\). Indeed, this is used so often that one routinely even uses the same symbol \(\omega\) for \(\iota^*(\omega)\). On the side note that there is no equivalent to this for tangent vector fields: Indeed, for any vector field \(X \in \Gamma^\infty(M)\) there are in general many extensions to a vector field on \(N\). Conversely if \(N \subseteq M\) is a submanifold of positive codimension and \(\Phi \in C^\infty(M, N)\) then \(\Phi\) is necessarily many-to-one and unless something special happens there is little hope that the collection of tangent vectors \(\Phi_*X_p\) (with \(p \in M\)) are the image of a vector field on \(N\).

In practical examples one routinely needs to calculate the pullbacks of differential forms in terms of local coordinates. Thus consider a smooth map \(\Phi \in C^\infty(M, N)\), local and coordinate charts \((u, U)\) about a point \(p \in M\) and \((v, V)\) about \(\Phi(p) \in N\). Due to the linearity of \(\Phi_*\) it suffices to consider the pullbacks \(\Phi^*(dv^i)\). As an immediate consequence of the earlier calculations (3.3) of the tangent map in coordinates find
\[
(\Phi^*(dv^i)) \left( \frac{\partial}{\partial u^j} \right)_p = dv^i \left( \Phi_* \left( \frac{\partial}{\partial v^j} \right) \Phi \right) = dv^i \left( \sum_{\ell=1}^n \frac{\partial (v^\ell \circ \Phi)}{\partial u^j} \right)_p \cdot \left( \frac{\partial}{\partial v^j} \Phi \right)_p = \frac{\partial (v^\ell \circ \Phi)}{\partial u^j} \right)_p
\]

and consequently for \(\omega_i \in C^\infty(N)\)
\[
\Phi^* \left( \sum_{i=1}^n \omega_i \, dv^i \right) = \sum_{j=1}^m \left( \sum_{i=1}^n \omega_i \frac{\partial (v^\ell \circ \Phi)}{\partial u^j} \right) \cdot du^j
\]
As expected this means that the coordinates transform by matrix-multiplication. One may look at this in different ways: If assembling the coordinates \( \omega_i \) into column vectors then the coordinates of the image are obtained by left multiplication by the transpose of the usual Jacobian matrix with components \( \frac{\partial(v^i \circ \Phi)}{\partial u^j} \). A more elegant way to interpret the sum in equation (70) is in terms of right multiplication of row vectors by the standard Jacobian matrix – no transpose. Thus if we write \( a = (\omega_1, \ldots, \omega_n) \) and \( b = (du^1(\Phi^* \omega), \ldots, du^n(\Phi^* \omega)) \) then \( b = aC \) where \( C \) is the matrix with components \( C_{ij} = \frac{\partial(v^i \circ \Phi)}{\partial u^j} \).

Consistently using this convention of representing (in local coordinates) tangent vector fields by column vectors and differential forms by row vectors facilitates many calculations. In particular, the evaluation of a differential form on a tangent vector field becomes in coordinates simply the matrix product. Furthermore, suppose \( X = \omega(\Phi_s X_p) \) is simply interpreted as associativity of matrix multiplication:

\[
(\Phi^* \omega)X_p = \omega(\Phi_s X_p) \quad \longleftrightarrow \quad (aC)\xi = a(C\xi)
\]

Formally, it is at times convenient to assemble the basis vectors into formal row and column vectors. To be consistent introduce the formal column vectors \( \{\partial u^1, \ldots, \partial u^n\} \). Then \( a = C\beta \) from (69). Together with the notation of the previous paragraph, this provides for such nice shorthand notation as

\[
\Phi^* \omega = a a = a(C\beta) = b\beta
\]

**Exercise 3.28** Suppose \( \Phi \in C^\infty(M^m, N^n) \) and \( \Psi \in C^\infty(N^n, P^p) \) are smooth maps between manifolds, \( p \in M \) and \( X_p \in T_pM \). Furthermore, suppose \( (u,U), (v,V) \) and \( (w,W) \) are local coordinate charts about \( p \in M \), \( \Phi(p) \in N \) and \( \Psi \circ \Phi(p) \in P \), respectively. Verify that the matrix representing \( (\Psi \circ \Phi)^* \) with respect to \( (u,U) \) and \( (w,W) \) is the product of the matrices representing \( \Phi^* \) (with respect to \( (u,U) \) and \( (v,V) \)) and \( \Psi^* \) (with respect to \( (v,V) \) and \( (w,W) \)).

**Exercise 3.29** Revisit the exercise 3.17 with \( M = \{x \in \mathbb{R}^2 : x^2 > 0\} \) equipped with rectangular coordinates \( (x,y) \), polar coordinates \( (r,\theta) \), and the mix \( (\xi,\rho) \) defined by \( \xi = x \) and \( \rho = r \).

For each pair of coordinates calculate the Jacobian matrix \( C \) with components \( C_{ij} = \frac{\partial(v^i \circ \Phi)}{\partial u^j} \), when \( \Phi = \text{id}_M \) is the identity map, and use this to write each of the basic differential forms \( \{du^1, du^2\} \) as a linear combination of each other set \( \{dv^1, dv^2\} \). In particular sketch these basic co-tangent vector fields using arrows .... Compare to the pictures for the basic tangent vector fields ....

**Exercise 3.30** Consider the imbedded sphere \( S^2 \subseteq \mathbb{R}^3 \) (i.e. \( M = S^2 \), \( N = \mathbb{R}^3 \)) and the inclusion map \( \Phi = \iota : S^2 \hookrightarrow \mathbb{R}^3 \) and the standard spherical coordinates \( (u,U) = ((\theta,\phi),U) \), e.g. with \( U = (\theta,\phi)^{-1}(-\pi,\pi) \times (0,\pi) \) on \( M \) and the Cartesian coordinates \( (v,V) = ((x^1, x^2, x^3),\mathbb{R}^3) \) on \( N \).

Explicitly calculate the pullbacks \( \iota^* du^i \) for \( i = 1, 2, 3 \) in terms of \( du^1 = d\theta \) and \( du^2 = d\phi \).

Locate all points \( p \in M \) where any of these cotangent vector fields vanish. Describe the vector fields pictorially, both as arrows on the sphere, and as arrows on \((-\pi,\pi) \times (0,\pi) \) (technically, this means sketching the vector fields \( (u^{-1})^* \circ \iota^* dv^3 \)).
In a subsequent section we will return to differential forms to investigate when a differential form $\omega$ is the differential of a smooth function. This generalization of the notion of gradient fields will lead to powerful integrability theorems. The reader is encouraged to continue comparing and contrasting the algebraic ease of working with differential form and the more tangible, visual aspects of tangent vector fields.
3.9 Differential equations and local flows

Vector fields are intimately related to differential equations. But we argue that as functions with clearly defined domain and codomain, they are mathematically both more precise and more versatile. E.g. are the two equations $\dot{y} = y$ and $\dot{y} - y = 0$, or $\dot{\frac{y}{y}} = 1$ “the same”?

To translate between these languages, consider in classical terminology a system of ordinary differential equations on $\mathbb{R}^n$

$$
\begin{align*}
\dot{x}_1 &= f_1(x) \\
\dot{x}_2 &= f_2(x) \\
&\vdots \\
\dot{x}_n &= f_n(x)
\end{align*}
$$

(73)

One may say that “a solution of this system of differential equations is a function $x(t)$ that satisfies the system, i.e. makes it true when $x(t)$ is substituted for $x$ . . .” Clearly there is nothing lost when going from writing out equations with symbols such as $\dot{x}_i$ to focusing on the vector field $X = \sum_{i=1}^n f^i D_i$ (we change the indices to traditional superscripts of differential geometry). Then the notion of a solution of a differential equation is made precise:

**Definition 3.13** Suppose $X \in \Gamma^\infty(M)$ is a smooth vector field on a smooth manifold $M$. An integral curve of $X$ is a curve $\sigma: I \hookrightarrow M$ (for some interval $I \subseteq \mathbb{R}$) such that for all $t \in I$

$$
\sigma_*(\frac{d}{dt}|_t) = (X \circ \sigma)(t), \quad \text{or in shorhand notation:} \quad \dot{\sigma} = X \circ \sigma
$$

(74)

To connect this with the notion of tangent vectors as first order partial differential operators note that this definition is equivalent to saying that $\sigma$ is an integral curve if for every point $p \in M$ and every smooth function $\phi \in C^\infty(p)$ defined in some neighborhood of $p$

$$
\frac{d}{dt} (\phi \circ \sigma) = (X\phi) \circ \sigma
$$

(75)

In particular, in a chart $(u,U)$ taking $\phi = u^k$ yields $\frac{d}{dt}(u^k \circ \sigma) = (Xu^k) \circ \sigma$.

For later reference we formally distinguish a vector field along a curve from the restriction of a vector field on the manifold to the image of a curve. In the case that the curve is an imbedding, one easily can go back and forth. However, if the curve is not one-to-one, a vector field along a curve need not arise from a vector field on the manifold.

**Definition 3.14** Suppose $\sigma: I \hookrightarrow M$ is a smooth curve. A smooth function $Y: I \hookrightarrow TM$ such that $\pi \circ Y = \sigma$ is called a vector field along the curve $\sigma$.

Note that if $X: M \hookrightarrow TM$ is a smooth vector field on manifold $M$, and $\sigma: I \hookrightarrow M$ is a smooth curve, then the composition $X \circ \sigma$ is a vector field along the curve $\sigma$. However, if $Y: I \hookrightarrow TM$ is a vector field along the curve $\sigma: I \hookrightarrow M$, then there need not exist any smooth vector field $X$ on $M$ such that $Y$ arises from $X$ in the afore described way.

**Exercise 3.31** Discuss vector fields along curves as opposed to vector fields on manifolds in the special cases where the curve is not one-to-one, or is an immersion, but not an imbedding. Do you see similar, related issues with vector fields along curves in nonorientable manifolds such as the Moebius strip?
More interesting than just isolated integral curves is the collection of all integral curves of a vector field. However, in general some technical difficulties arise due to the common lack of common domains. We start with a preliminary notion of solutions as parameterized families of integral curves, parameterized by the initial conditions.

**Definition 3.15 (Preliminary definition)** The/a flow of a vector field $X: M \mapsto TM$ is a function $\Phi$ defined on a suitable (maximal?) subset of $\mathbb{R} \times M$ with values in $M$ that satisfies

$$
\begin{align*}
\Phi(0,p) &= p \quad \text{for all } p \in M \\
\frac{d}{dt}(f \circ \Phi) &= (Xf) \circ \Phi \quad \text{for all } f \in C^\infty(M)
\end{align*}
$$

When holding the initial condition $\Phi(0,p) = p \in M$ fixed the map $\Phi(\cdot,p): t \mapsto \Phi(t,p)$ is an integral curve of $X$. On the other hand, when holding the time $t \in \mathbb{R}$ fixed the map $\Phi(t,\cdot): M \mapsto M$ is a map between manifolds. In different places either point of view may be more suitable. But in the sequel, often the critical step is to mix the roles of the time and space variables, e.g. consider the flow-box theorem 3.17 and the geodesic (normal) coordinates (or the exp-map) (theorem ??). The basic technical questions are about the existence and uniqueness of a maximal flow (i.e. with maximal domain), and about its smoothness. A few examples will clearly delineate the limitations – most importantly, for generic vector field, one should not expect a global flow to exist that has all desirable properties. On the other hand, locally there is hardly any difference between flows on Euclidean spaces and on manifolds. Thus the remainder of this section provides a brief review of basic properties and background material of ordinary differential equations in $\mathbb{R}^n$.

**Exercise 3.32** For each $p \in \mathbb{R}$ solve the initial value problem $\dot{y} = y^2$, and $\dot{y} = 1 + y^2$, $y(0) = p$. In each case, precisely describe the maximal domain of the flow.

The phenomenon exemplified by the solution curves in the preceding exercise is termed finite escape time.

**Exercise 3.33** Verify that for each pair of times $t_1, t_2 \in \mathbb{R}$ with $t_1 \leq t_2$ the function

$$
y(t) = \begin{cases} 
-(t_1 - t)^{3/2} & \text{if } t \leq t_1 \\
0 & \text{if } t_1 \leq t \leq t_2 \\
(t - t_2)^{3/2} & \text{if } t \geq t_2
\end{cases}
$$

is a solution of the differential equation $\dot{y} = \frac{2}{3}y^{1/3}$

Precisely describe the maximal domain of the flow – pay special attention to the solutions which are not of the form above, i.e. with one or both of $t_1 = -\infty$ and $t_2 = \infty$.

**Exercise 3.34** Let $f: \mathbb{R} \mapsto \mathbb{R}$ be defined by $f(x) = 1$ of $x \geq 0$ and $f(x) = -1$ else. Discuss and suggest notions of solutions for the differential equations $\dot{x} = f(x)$ and $\dot{x} = -f(x)$. Focus on the maximal domains on which solutions make sense, and on initial conditions $x(t_0) = 0$.

**Exercise 3.35 (Challenge problem)** Prove that for every initial condition $y(0) = p \in \mathbb{R}$ the differential equation $\dot{y} = 1 + y^{1/3}$ has a unique solution. [[There are lots of less-known conditions in the classical literature for the uniqueness of solutions to initial value problems which are much weaker than the traditionally taught Lipschitz-conditions. In this case, Hölder continuity combined with transversality do the job – contact M. Kawski for details and a recent, cute application. (“Stabilization of nonlinear systems in the plane”, Systems and Control Letters, vol. 12 no. 3 (1989) pp. 169-175.)]] Going even further: What about $\dot{y} = y + y^{1/3}$?.
The examples in these exercises illustrate that if the vector field (alas differential equation) is not sufficiently smooth then even locally neither existence and nor uniqueness of solutions are guaranteed. Moreover, without stringent growth conditions there is no hope for global existence. The standard existence and uniqueness result for solutions of ordinary differential equation relies on Gronwall inequality and Picard iteration.

**Lemma 3.12 (Gronwall inequality)** Suppose $f : [0, T] \mapsto [0, \infty)$ is continuous and $C, K \geq 0$ are such that for all $t \in [0, T]$, $f(t) \leq C + \int_0^t K f(s) \, ds$ Then for all $t \in [0, T]$, $f(t) \leq C e^{Kt}$.

**Proof.** Suppose $f : [0, T] \mapsto [0, \infty)$ is continuous and $C, K \geq 0$ are such that for all $t \in [0, T]$, $f(t) \leq C + \int_0^t K f(s) \, ds$. Define $u : [0, T] \mapsto [0, \infty)$ by $u(t) = C + \int_0^t K f(s) \, ds$. Then $u$ is continuously differentiable, and $0 \leq u' = K f \leq u$ and hence for all $t \in [0, T]$, $\frac{d}{dt}(u(t)e^{-Kt}) \leq 0$, and, $0 \leq u(t) \leq u(0)e^{Kt}$ from which the desired inequality follows readily. ■

**Definition 3.16** Suppose $U \subseteq \mathbb{R}^m$ is open. A function $f : U \mapsto \mathbb{R}^n$ is called locally Lipschitz continuous on $U$ of for every $p \in U$ there exists an open neighborhood $W \subseteq U$ of $p$ and a constant $L$, called the Lipschitz constant, such that for all $q, q' \in W$

$$\|f(q) - f(q')\|_{\mathbb{R}^n} \leq L \cdot \|q - q'\|_{\mathbb{R}^n} \tag{78}$$

**Theorem 3.13 (Picard Lindelöf)** Let $U \subseteq \mathbb{R}^m$ be a connected open set and suppose $f : U \mapsto \mathbb{R}^n$ is locally Lipschitz continuous. Then there exists an open set $W \subseteq \mathbb{R} \times U$ such that for each fixed $y \in U$ the set $\{t \in \mathbb{R} : (t, y) \in W\}$ is an interval $(a(y), b(y))$, and there exists a unique continuous map $\Psi : W \mapsto U$ with maximal domain $W$ such that

$$\frac{d}{dt} \Psi(t, y) = f(\Psi(t, y)) \quad \text{for all } (t, y) \in W$$

$$\Psi(0, y) = y \quad \text{for all } y \in U \tag{79}$$

**Definition 3.17** The map $\Psi : W \mapsto U$ of the preceding theorem is called the flow of the ordinary differential equation $\dot{y} = f(y)$.

The basic idea behind the proof is to use Picard-iteration, that is to transform the differential equation into an integral equation that is solved by successive iteration:

$$y_0(t) = y_0$$

$$y_{k+1}(t) = y_0 + \int_0^t f(s, y_k(s)) \, ds \tag{80}$$

The standard argument uses that for sufficiently small $\varepsilon > 0$ the map $y_n \mapsto y_{n+1}$ on the metric space $C^0([t_0, t_0 + \varepsilon])$ is a contraction, and hence must have a fixed point. With some additional work one may establish some additional regularity. The basic idea is that due to the integration in the Picard-iteration scheme the function $y_{k+1}$ is at least one degree smoother than the lesser of the degrees of smoothness of $y_n$ and $f$, leading to a boots-trapping argument . . . But with all technical details this is a fairly hard theorem . . .

**Theorem 3.14 (Smoothness of the flow)** Suppose $f, U, W, \Psi$ are as in the preceding theorem. Then if $f \in C^r(U, \mathbb{R}^n)$ for $1 \leq r \leq \infty$ then $\Psi \in C^r(W, U)$ (i.e. $\Psi$ is $C^r$ in both arguments).
Exercise 3.36  Carry out the Picard iteration scheme for the initial value problem $\dot{y} = y, \ y(0) = c \in \mathbb{R},$ e.g. explicitly calculate the first three iterates, and by induction find a formula for the $n$-th iterate. Verify that this sequence of functions converges globally (i.e. for all $t \in \mathbb{R}$ to an analytic function which is the solution of the initial value problem.

Exercise 3.37  Carry out the Picard iteration scheme for the initial value problem $\dot{y}_1 = -y_2, \ \dot{y}_2 = y_1, \ y(0) = (c_1, c_2) \in \mathbb{R}^2,$ e.g. explicitly calculate the first five iterates, and by induction find a formula for the $n$-th iterate. Verify that this sequence of functions converges globally (i.e. for all $t \in \mathbb{R}$) to an analytic function which is the solution of the initial value problem.

Suitably adapted to our settings of smooth vector fields on manifolds one obtains:

Theorem 3.15 (Existence of smooth, local flows) Suppose $X \in \Gamma^\infty(M)$ is a smooth vector field and $p \in M.$ Then there exists an open neighborhood $O \subseteq M$ of $p,$ $\varepsilon > 0$ and a one-parameter family of diffeomorphisms $\Phi_t: O \mapsto \Phi_t(A) \subseteq M$ onto their respective images, parameterized by $t \in (-\varepsilon, \varepsilon)$ such that

(i) $\Phi: (-\varepsilon, \varepsilon) \times O \mapsto M$ is $C^\infty.$

(ii) If $|s|, |t|, |s + t| < \varepsilon,$ and $q, \Phi_t(q) \in O$ then $\Phi_s(\Phi_t(q)) = \Phi_{s+t}(q)$ and $\Phi_0(q) = q.$

(iii) For all $q \in O,$ $X_q$ is tangent to the curve $\gamma: (-\varepsilon, \varepsilon) \mapsto O$ where $\gamma(t) = \Phi_t(q)$ at $t = 0.$

The proof is a rather straightforward construction, using at some place compactness in order to achieve a constant time interval, i.e. that is the same for all points in some neighborhood.

Corollary 3.16  If $K \subseteq M$ is compact and $X \in \Gamma^\infty M$ vanishes on $M \setminus K$ then $X$ generates a unique one parameter family of diffeomorphisms (i.e. the flow is global, defined for all $t \in \mathbb{R}$).
Most useful for theoretical purposes is the following result about the possibility to stratify any vector field at any points where it does not vanish.

**Theorem 3.17 (Flow box theorem)** Suppose $X \in \Gamma^\infty(M)$ and $p \in M$ is such that $X_p \neq 0$. Then there exists a chart $(u, U)$ about $p$, w.l.o.g. $u(p) = 0$, such that
\[
X|_u = \frac{\partial}{\partial u^j} \quad \text{i.e.} \quad X_q = \frac{\partial}{\partial u^j}|_q \quad \text{for all} \quad q \in U
\] (81)

Note that this effectively says that modulo a local coordinate change every (nonlinear!) system of differential equations is in the neighborhood of any regular point locally equivalent to the system
\[
\begin{align*}
\dot{x}_1 &= 1 \\
\dot{x}_2 &= 0 \\
&\vdots \\
\dot{x}_n &= 0
\end{align*}
\] (82)

Of course, in general there is no hope of finding explicit, so called “closed-form” solutions for nonlinear differential equations – which is really synonymous with finding a formula for the coordinates $u$, or more practically a change of coordinates to $(u, U)$.

**Proof.** Start with any coordinates $(w, W)$ about $p$ and suppose $\dot{w}(X_p) = (a^1, \ldots, a^m) = a \in \mathbb{R}^m$, and $w(p) = 0$. Without loss of generality we may assume that $a^1 \neq 0$ (else permute the coordinates in $w$). Define a new coordinate chart $(v, V = W)$ by setting for $q \in W$
\[
v^1(q) = \frac{1}{a^1}w^1(q), \quad \text{and for} \quad j \geq 2, \quad v^j(q) = w^j(q) - \frac{a^j}{a^1}w^1(q).
\] (83)

Verify that this constant linear coordinate change aligns the new coordinates with the vector field at the point $p$, i.e., $X_p = \frac{\partial}{\partial v^1}|_p$.
\[
\frac{\partial}{\partial v^1}|_p = \sum_{j=1}^m \frac{\partial w^j}{\partial v^1}|_p = \frac{\partial}{\partial w^j}|_p = \sum_{j=1}^m a^j \frac{\partial}{\partial w^j}|_p = X_p
\] (84)

In a second step, use the local flow $\Phi: (-\varepsilon, \varepsilon) \times \tilde{U} \mapsto M$ of $X$ defined on some neighborhood $\tilde{U} \subseteq V$ of $p$, to define a map $u^{-1}: v(\tilde{U}) \mapsto M$ by
\[
u^{-1}(a^1, a^2, \ldots, a^m) = \Phi(a^1, v^{-1}(0, a^2, a^3, \ldots a^m))
\] (85)

This map $u^{-1}$ is well-defined on some open set $v(\tilde{U})$ about $(0, v(p) \subseteq \mathbb{R}^{m+1})$, and it is clearly $C^\infty$, but we need to verify that it is invertible (possibly after further restricting its domain), i.e. verify that there exists a map $u$ such that $(u^{-1})^{-1} = u$ indeed defines coordinates on some neighborhood of $p$. First calculate for $a \in v(\tilde{U}) \subseteq \mathbb{R}^m$ and $f \in C^\infty(v(\tilde{U}))$
\[
((u^{-1})_*(D_1|_a)}(f) = \lim_{h \to 0} \frac{1}{h} \left( f \circ u^{-1}(a + h, a^2, \ldots a^m) - f \circ u^{-1}(a) \right) = \lim_{h \to 0} \frac{1}{h} \left( f \circ \Phi_h \circ v^{-1}(0, a^2, \ldots a^m) - f \circ u^{-1}(a) \right) = \lim_{h \to 0} \frac{1}{h} \left( (f \circ \Phi_h)(u^{-1}(a)) - f(u^{-1}(a)) \right) = X_{u^{-1}(a)} f
\] (86)
which shows that for all \( a \in v(\tilde{U}) \), \(((u^{-1})_* D_1|_a = X_{u^{-1}(a)}\). Similarly, for \( k > 1 \) calculate

\[
((u^{-1})_* D_k|_0)(f) = \lim_{h \to 0} \frac{1}{h} (f \circ u^{-1}(0, \ldots, 0, h, 0, \ldots) - f \circ u^{-1}(0)) = \lim_{h \to 0} \frac{1}{h} (f \circ v^{-1}(0, \ldots, 0, h, 0, \ldots) - f \circ v^{-1}(0)) = \lim_{h \to 0} \frac{1}{h} (f \circ u^{-1}(0, \ldots, 0, h, 0, \ldots) - f \circ v^{-1}(0)) = D_k|_0 (f \circ v^{-1}) = \frac{\partial f}{\partial v^k}\bigg|_p.
\]

Therefore for all \( k \geq 1 \), \(((u^{-1})_* D_k|_0 = \frac{\partial}{\partial v^k}\bigg|_p\) showing that \((u^{-1})_* : T_0\mathbb{R}^m \mapsto T_pM\) has full

rank, indeed, \(v_p \circ (u^{-1})_*\) is the identity on \(T_pM\). By the inverse function theorem, there exist

open neighborhoods \( U \subseteq \tilde{U} \) of \( p \) and \( Z \subseteq v(U) \) of 0 such that

\(u^{-1}|_Z : Z \mapsto U\) is a bijection and

\[
u \equiv (u^{-1}|_Z)^{-1} : (U, p) \mapsto (\mathbb{R}^m, 0)
\]

defines a local coordinate chart about \( p \) in which \( X \equiv \frac{\partial}{\partial u^1}\).

**Example 3.2** Consider \( M = \mathbb{R}^2\) with standard coordinates \(((x_1, x_2), \mathbb{R}^2)\), the vector field \( X = \frac{\partial}{\partial x_1} + 2x_1 \frac{\partial}{\partial x_1}\) and the point \( p = (0, 0)\). At \( p \), \( X_p = \frac{\partial}{\partial x_1}\bigg|_p\), i.e. no preliminary constant linear coordinate change is needed. To calculate the flow of \( X \) solve the initial value problems

\[
\begin{align*}
\dot{x}_1 &= 1 & x_1(0) &= a \\
\dot{x}_2 &= 2x_1 & x_1(0) &= b
\end{align*}
\]

yielding the formula for the flow

\[
\Phi_1(a, b) = (a + t, b + (a + t)^2 - a)
\]

To define the new coordinates first construct the inverse

\[
x = u^{-1}(y_1, y_2) \overset{\text{def}}{=} \Phi_{y_1}(0, y_2) = (y_1, y_2 + y_1^2)
\]

Verify that in the new coordinates \( u = (y_1, y_2) = (x_1, x_2 - x_1^2)\) the vector field is straightened out, indeed, \( X = \frac{\partial}{\partial y_1}\) and \( \frac{\partial}{\partial y_2} = \frac{\partial}{\partial x_2}\).
3.10 Lie derivatives

The (local) flows obtained in section 3.9 provide some minimal analogue of an additive structure on a manifold. An important application is to use these for calculus-like, coordinate free characterization and definitions of Lie derivatives of functions, of vector fields, and in more generality, later of any tensors. Instead of starting with a vector field, it is often convenient to start with a family of diffeomorphisms. Formally define:

**Definition 3.18** A one-parameter family of diffeomorphisms on a manifold \( M \) is a smooth map \( \Phi: \mathbb{R} \times M \rightarrow M \) satisfying

- For every fixed \( t \in \mathbb{R} \) the map \( \Phi_t: M \rightarrow M \), defined by \( \Phi_t(p) = \Phi(t, p) \) for \( p \in M \), is a diffeomorphism of \( M \).
- \( \Phi_s \circ \Phi_t = \Phi_{s+t} \) for all \( s, t \in \mathbb{R} \).

Similarly, a local one-parameter family of diffeomorphisms is a smooth map \( \Phi: (\varepsilon, \varepsilon) \times U \rightarrow M \), defined for some \( \varepsilon > 0 \) and some open set \( U \subseteq M \), which satisfies

- For every fixed \( |t| < \varepsilon \) the map \( \Phi_t: U \rightarrow \Phi_t(U) \subseteq M \), is a diffeomorphism of \( U \) onto its image.
- \( \Phi_s \circ \Phi_t = \Phi_{s+t} \) for all \( s, t \) such that \( |s|, |t|, |s + t| < \varepsilon \).

It is easily recognized that such (local) one-parameter families of diffeomorphism arise is (local) flows of smooth vector fields \( X \). In turn, given such a family, one can recover the vector field that locally generates this family as its flow, and one calls the infinitesimal generator of the family \( \Phi \) of (local) diffeomorphism. This is based on the observation that one may characterize the derivative \( Xf \) of a function \( f \in C^\infty(M) \) in the direction of a vector field \( X \) entirely in terms of the flow of \( X \):

**Definition 3.19** The Lie derivative \( L_Xf \) of a smooth function \( f \in C^\infty(M) \) with respect to a vector field \( X \in \Gamma^\infty(M) \) with flow \( \Phi \) is defined as

\[
(L_Xf)(p) = \lim_{h \to 0} \frac{1}{h} \left( (f \circ \Phi_h)(p) - f(p) \right) \quad \text{for } p \in M
\]

One readily sees that this Lie derivative \( L_Xf \) is the same as \( Xf \) – but the notation \( L_Xf \) is still very suggestive and nicely carries over to Lie derivatives of other objects:

**Definition 3.20** The Lie derivatives \( L_XY \) and \( L_X\omega \) of a smooth (tangent) vector field \( Y \in \Gamma^\infty(M) \) and of a smooth co-tangent vector field \( \omega \in \Gamma^{\infty\ast}(M) \) with respect to \( X \in \Gamma^\infty(M) \) with flow \( \Phi \) are defined (at any point \( p \in M \)):

\[
(L_XY)_p = \lim_{h \to 0} \frac{1}{h} \left( Y_p - (\Phi_h)_pY_p \right),
\]

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

**Proposition 3.18** Suppose \( f \in C^\infty(M) \), \( X, Y, Z \in \Gamma^\infty(M) \), and \( \omega, \eta \in \Omega^1(M) \). Then

\[
(i) \quad L_X(Y + Z) = L_XY + L_XZ
\]

\[
(ii) \quad L_X(\omega + \eta) = L_X\omega + L_X\eta
\]

\[
(iii) \quad L_X(fY) = (L_Xf) \cdot Y + f \cdot L_XY
\]

\[
(iv) \quad L_X(f\omega) = (L_Xf)\omega + f \cdot (L_X\omega)
\]

\[
(v) \quad L_X(\omega(Y)) = (L_X\omega)(Y) + \omega(L_XY)
\]
The first two parts are immediate consequences of the \( \mathbb{R} \)-linearity of the maps \( \Phi_{h*} \) and \( \Phi^*_h \). We will prove the product rules (iii) and (v) and leave (iv) as an exercise.

**Proof** (of part (iii)). Suppose \( f \in C^\infty(M) \), \( X, Y \in \Gamma^\infty(M) \), \( p \in M \) and let \( \Phi \) be the local flow of \( X \) defined in a neighborhood of \( p \). Using basic properties of the tangent maps \( \Phi_{h*} \), especially the linearity, and, as expected for product rules, add and subtract suitable terms:

\[
(L_X(fY))_p = \lim_{h \to 0} \frac{1}{h} \left( (fY)_p - (\Phi_{h*}(fY))_p \right)
= \lim_{h \to 0} \frac{1}{h} \left( f(p) \cdot Y_p - \Phi_{h*} \Phi_{-h}(p) \left((f \circ \Phi_{-h})(p) \cdot Y_{-h}(p)\right)\right)
= \lim_{h \to 0} \frac{1}{h} \left( f(p)Y_p - (f \circ \Phi_{-h})(p) \cdot \Phi_{h*} \Phi_{-h}(p)Y_{-h}(p)\right)
= \lim_{h \to 0} \frac{1}{h} \left( f(p) - (f \circ \Phi_{-h})(p)\right) \cdot Y_p
+ \lim_{h \to 0} (f \circ \Phi_{-h})(p) \cdot \lim_{h \to 0} \frac{1}{h} (Y_p - (\Phi_{h*}Y)_p)
= (L_Xf)(p) \cdot Y_p + f(p) \cdot (L_XY)_p
\]

As long as all the limits exist there is no problem with breaking them up. Moreover, from continuity of \( f \) at \( p \) it follows that \( \lim_{h \to 0} (f \circ \Phi_{-h})(p) = f(p) \). 

Instrumental for the proofs of part (iv) and (v) of proposition 3.18 is the observation:

**Lemma 3.19** Suppose \( \omega \in \Omega^1(M) \), \( p \in M \) and \( \Phi \) is the local flow of \( X \in \Gamma^\infty(M) \) near \( p \). Then

\[
\lim_{h \to 0} (\Phi^*_h \omega)_p = \omega_p
\]

**Proof.** Suppose that \( (u,U) \) is a chart about \( p \). Then for \( |h| \) sufficiently small \( \Phi_h(p) \in U \) and (70) (with \( v = u \)) yields

\[
(\Phi^*_h \omega)_p = \sum_{i=1}^n \omega_i(\Phi_h(p)) \Phi^*_h (du^i_{\Phi_h(p)}) = \sum_{j=1}^m \left( \sum_{i=1}^n \omega_i(\Phi_h(p)) \frac{\partial (u^i \circ \Phi_h)}{\partial u^j} \right)_p \cdot du^j
\]

Since \( \omega \) is smooth, clearly \( \lim_{h \to 0} \omega_i(\Phi_h(p)) = \omega_i(p) \). Regarding the second factor, since \( \Phi \) is smooth in both variables, and \( \Phi_0 \) is the identity map on some neighborhood of \( p \), it follows that \( \lim_{h \to 0} \frac{\partial (u^i \circ \Phi_h)}{\partial u^j} \big|_{\Phi_h(p)} = \delta^j_i \). [[Think carefully about this limit!]]

**Exercise 3.38** Prove the product rule (iv) of proposition 3.18 (use lemma 3.19).
Proof (of part (v) of proposition 3.18). Let \( X, Y \in \Gamma^\infty(M) \), \( \omega \in \Gamma^\infty(M) \) \( p \in M \) and let \( \Phi \) be the local flow of \( X \) defined in a neighborhood of \( p \).

\[
L_X(\omega(Y))(p) = \lim_{h \to 0} \frac{1}{h} \left( (\omega(Y)) \circ \Phi_h(p) - (\omega(Y))(p) \right)
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left( \omega_{\Phi_h(p)}(Y_{\Phi_h(p)}) - \omega_p(Y_p) \right)
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left( \omega_{\Phi_h(p)} \left( (\Phi_{h*}^\omega)(\Phi_{-h*}^\omega)(p) Y_{\Phi_h(p)} \right) - \omega_p(Y_p) \right)
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left( (\Phi_{h*}^\omega)_p (\Phi_{-h*}^\omega)_p - \omega_p(Y_p) \right)
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left( (\Phi_{h*}^\omega)_p (\Phi_{-h*}^\omega)_p(Y_p) + (\Phi_{h*}^\omega)_p \left( \left( \frac{1}{h} \right) \left( Y_p - (\Phi_{-h*}^\omega)_p \right) \right) \right)
\]

\[
= \left( \lim_{h \to 0} \frac{1}{h} \left( (\Phi_{h*}^\omega)_p - \omega_p \right) \right)(Y_p)
\]

\[
+ \left( \lim_{h \to 0} \frac{1}{h} \left( (\Phi_{h*}^\omega)_p(Y_p) \right) \right)(Y_p - (\Phi_{-h*}^\omega)_p))
\]

\[
= (L_X\omega)_p(Y_p) + \omega_p(L_XY)_p . \tag{97}
\]

The next major goal is to prove that the Lie derivative of a vector field is the same as the Lie bracket defined earlier: \( L_X Y = [X, Y] \). Given the obvious anti-symmetry \([Y, X] = -[X, Y]\) of the Lie bracket this sheds some important light on the Lie derivative which at first view seems to assign very different roles to the vector fields \( X \) and \( Y \), evaluating \( Y \) along the integral curves of \( X \) ... We will eventually provide an alternative coordinate-free proof, but first will use this as an opportunity to figure out how to calculate Lie derivatives in coordinates.

Lemma 3.20 Suppose \((u, U)\) is a coordinate chart and \( X \in \Gamma^\infty(M) \). Then

\[
L_X(du^i) = \sum_{j=1}^{m} \frac{\partial(Xu^i)}{\partial u^j} du^j
\]

Proof. Suppose \((u, U)\) is a coordinate chart about \( p \in M \) and \( \Phi \) is the local flow of \( X \in \Gamma^\infty(M) \) near \( p \). For sufficiently small \( |h| \) we may assume that \( \Phi_h(p) \in U \). The key step is the interchange of the order of differentiation in time and space directions, and thus requires the use of the definition of \( X \) as the generator of the flow \( \Phi \ldots \ldots \)

\[
(L_X(du^i))_p = \lim_{h \to 0} \frac{1}{h} \left( \left( \Phi_h^*(du^i)_p - (du^i)_p \right) \right)
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left( \sum_{j=1}^{m} \left( \frac{\partial(u^i \circ \Phi_h)}{\partial u^j} \right)_p (du^j)_p - \delta^i_j (du^j)_p \right)
\]

\[
= \sum_{j=1}^{m} \lim_{h \to 0} \frac{1}{h} \left( \left( \frac{\partial(u^i \circ \Phi_h)}{\partial u^j} \right)_p \right) (du^j)_p
\]

\[
= \sum_{j=1}^{m} \lim_{h \to 0} \frac{1}{h} \left( \left( \frac{\partial}{\partial u^j} (u^i \circ \Phi) \right) \bigg|_{(h, p)} - \frac{\partial}{\partial u^j} (u^i \circ \Phi) \bigg|_{(0, p)} \right) (du^j)_p \tag{99}
\]

\[
= \sum_{j=1}^{m} \frac{\partial}{\partial u^j} (u^i \circ \Phi) \bigg|_{(0, p)} (du^j)_p
\]

\[
= \sum_{j=1}^{m} \frac{\partial}{\partial u^j} (u^i \circ \Phi) \bigg|_{(0, p)} (du^j)_p
\]

\[
= \sum_{j=1}^{m} \frac{\partial(Xu^i)}{\partial u^j} (du^j)_p . \tag{100}
\]
Rather than performing a similar direct calculation of the Lie derivative $L_X \left( \frac{\partial}{\partial u^i} \right)$, we use the product rule of part (v) of proposition 3.18, and observe

$$0 = L_X \delta^i_j = L_X \left( (du)^i \left( \frac{\partial}{\partial u^j} \right) \right) = \left( L_X (du)^i \right) \left( \frac{\partial}{\partial u^j} \right) + (du)^i \left( L_X \frac{\partial}{\partial u^j} \right) \quad \text{(100)}$$

Using the lemma 3.20 this establishes:

**Corollary 3.21** Suppose $(u, U)$ is a coordinate chart and $X \in \Gamma^\infty(M)$. Then

$$L_X \frac{\partial}{\partial u^j} = \sum_{j=1}^m - \frac{\partial (Xu^i)}{\partial u^j} \frac{\partial}{\partial u^i} \quad \text{(101)}$$

We are now ready to state and give a direct proof of the theorem:

**Theorem 3.22** If $X, Y \in \Gamma^\infty(M)$ then $L_X Y = [X, Y]$.

**Corollary 3.23** If $X, Y \in \Gamma^\infty(M)$ and $f \in C^\infty(M)$ then $L_{fX} Y = fL_X Y - (LYf) X$.

**Proof.** Work locally in a coordinate chart $(u, U)$ and use the corollary 3.23.

\[
L_X Y = L_X \left( \sum_{j=1}^m (Y u^j) \frac{\partial}{\partial u^j} \right) \\
= \sum_{j=1}^m \left( L_X (Y u^j) \frac{\partial}{\partial u^j} + (Y u^j) L_X \left( \frac{\partial}{\partial u^j} \right) \right) \\\n= \sum_{j=1}^m \left( (XY u^j) \frac{\partial}{\partial u^j} + (Y u^j) \sum_{i=1}^m \left( - \frac{\partial (Xu^i)}{\partial u^j} \frac{\partial}{\partial u^i} \right) \right) \\\n= \sum_{j=1}^m \left( (XY - YX) u^j \frac{\partial}{\partial u^j} \right) = [X, Y]. \quad \blacksquare
\]

**Proof (coordinate-free version, following Spivak vol.1, p.213).** Suppose $p \in M$, $f \in C^\infty(p)$, and $\Phi$ is a local flow of $X$ defined in a neighborhood of $p$. Begin with some preliminary constructions. For $|\tau|$ sufficiently small and $q$ sufficiently near $p$ define

$$g(\tau, q) = \int_0^1 \frac{\partial (f \circ \Phi)}{\partial t}(s \tau, q) \, ds \quad \text{(103)}$$

Integration of $\tau g(\tau, q)$ yields

$$\tau g(\tau, q) = \int_0^1 \frac{\partial (f \circ \Phi)}{\partial t}(s \tau, q) \, \tau \, ds \quad \text{(104)}$$

Consequently $(f \circ \Phi_t)(q) = f(q) + tg_t(q)$, and, moreover,

$$g_0(q) = \int_0^1 \frac{\partial (f \circ \Phi)}{\partial t}(0 \tau, q) \, d\tau = \frac{\partial (f \circ \Phi)}{\partial t} |_{(0, q)} = (X_q f)(q) = (Xf)q \quad \text{(105)}$$

This allows us to write

$$(\Phi_h Y) \, f = Y_{\Phi_{-h}(p)} (f \circ \Phi_h) = Y_{\Phi_{-h}(p)} \, f + h \, Y_{\Phi_{-h}(p)} g_h \quad \text{(106)}$$
Finally we are ready to use these ingredients to prove the theorem

\[
(L_XY)_p f = \lim_{h \to 0} \frac{1}{h} \left( Y_p - (\Phi_{h*}Y)_p \right) f \\
= \lim_{h \to 0} \frac{1}{h} \left( (Y_p f) - Y_{\Phi_{-h}(p)} f - h Y_{\Phi_{-h}(p)} g_h \right) \\
= \lim_{h \to 0} \left( \frac{1}{h} \right) \left( (Yf) \circ \Phi_{-h}(p) - (Yf)(p) \right) - \lim_{h \to 0} \left( (Yg_h) \circ \Phi_{-h}(p) \right) \\
= (L_X(Yf) - Y(Xf))(p) = [X,Y]_p f.
\] (107)

**Corollary 3.24** If \( X \in \Gamma_{\infty}(M) \) then \( L_X X \equiv 0 \).

**Proof.** This is an immediate consequence of \( L_X X = [X,X] = -[X,X] \), using the anticommutativity of the Lie bracket. However, as a useful exercise in working with the definitions we provide a direct proof in terms of flows. Thus let \( p \in M \), \( f \in C^\infty(p) \) and let \( \Phi \) be the local flow of \( X \) defined near \( p \). Then

\[
(L_X X)_p f = \lim_{h \to 0} \frac{1}{h} \left( (X_p f - (\Phi_{h*}X)_p f) \right)
\]

We will show directly that the difference inside the limit is identically zero.

\[
(\Phi_{h*}X)_p f = (\Phi_{h*}\Phi_{-h}(p))X_{\Phi_{-h}(p)}) f
= X_{\Phi_{-h}(p)}(f \circ \Phi_h)
= \lim_{s \to 0} \frac{1}{h} \left( ((f \circ \Phi_h) \circ \Phi_s)(\Phi_{-h}(p)) - (f \circ \Phi_h)(\Phi_{-h}(p)) \right) \\
= \lim_{s \to 0} \frac{1}{h} \left( (f \circ \Phi_s)(p) - f(p) \right) \\
= X_p f.
\] (108)

**Exercise 3.39** Suppose that \( X,Y \in \Gamma^n(M) \) are analytic vector fields, \( p \in M \), and \( \Phi \) is the local flow of \( X \) defined near \( p \). For \( \varepsilon > 0 \) sufficiently small consider the curve \( \sigma: (-\varepsilon,\varepsilon) \to T_p M \) defined by \( \sigma(t) = (\Phi_{ts}Y)_p \). Rewrite the definition 3.20 of the Lie derivative \( L_X Y \) in the form \( (L_XY)_p f = \frac{d}{dt} \mid_{t=0} (\Phi_{ts}Y)_p(p) \) to establish (via repeated differentiation) the Taylor series expansion \( \sigma(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} (L_XY)_p \). This is often suggestively written as

\[
\Phi_{ts} \circ Y \circ \Phi_{-t} = \sum_{k=0}^{\infty} \frac{t^k}{k!} (\text{ad}^k_X Y)_p \quad \text{where} \quad (\text{ad}^{k+1}_X Y) = [X,(\text{ad}^k_X Y)],
\] (109)

Compare also proposition 3.28 for the flow of the vector field \( \Phi_{ts} \).
3.11 Lie derivatives and integrability

The main goal of this section is to prove a first integrability theorem which asserts that the vanishing of the Lie brackets \([X_i, X_j]\) is also sufficient for the (local) existence of a submanifold such that \(X_i\) are tangent vector fields for this submanifold. This complements proposition 3.9 which asserted that the vanishing of the Lie bracket is necessary for the existence of local coordinates such that given vector fields are coordinate vector fields.

**Theorem 3.25** Suppose \(p \in M^m\) and \(X_i \in \Gamma^\infty(M)\), \(i = 1, \ldots k\) are smooth vector fields such that \(\{X_1, \ldots, X_k\} \subseteq T_p M\) are linearly independent. If the vector fields commute pairwise, i.e. \([X_i, X_j] \equiv 0\), then there exists a chart \((u, U)\) about \(p\) such that \(X_i|_p = \frac{\partial}{\partial u^i}, \ i = 1, \ldots k\).

Towards the end of this section we shall provide a constructive (is it?) proof. The basic idea is to use the flows \(\Phi_j\) of the vector fields \(X_j\) and define a set of local coordinates \(u^j\) basically as the inverse of the map \((x^1, \ldots, x^k) \mapsto \Phi_{x^k}^k \circ \ldots \circ \Phi_{x_2}^2 \circ \Phi_{x_1}^1(p)\). (It is rather straightforward to accommodate the case that \(k < m\).) The most interesting facet of this construction is how the *times* spent flowing along each integral curve become *spatial* coordinates. The bulk of this section shall be devoted to prove some basic facts about Lie derivatives and their flows, which are both illuminating, and helpful in the eventual proof of the theorem.

We begin with some geometrical explorations: We defined the Lie bracket \([X, Y]\) of two vector fields \(X, Y \in \Gamma^\infty(M)\) in terms of the action on smooth functions \(f \in C^\infty(M)\), i.e. \([X, Y]f = X(Yf) - Y(Xf)\). The following relates the bracket to the commutative properties (or the lack thereof) of the associated flows. More specifically, suppose \(p \in M\) and \(\Phi\) and \(\Psi\) are the local flows of \(X\) and \(Y\) defined near \(p\). It is natural to ask how the Lie bracket \([X, Y]_p\) relates e.g. to how \(\Psi_s \circ \Phi_t(p)\) and \(\Phi_t \circ \Psi_s(p)\) compare. It turns out to be more convenient to instead compare \(\Psi_{-s} \circ \Phi_{-t} \circ \Psi_s \circ \Phi_t(p)\) with \(p\). In the following, let \(s = t\) and consider the terminal points of this concatenation of flows for small values of \(t\). Since everything is smooth, the endpoints as a function of \(t\) define a smooth curve, which passes through \(p\) at \(t = 0\). The basic observation is that the *tangent direction* of this curve at \(p\) is basically the Lie bracket of \([X, Y]_p\).

**Proposition 3.26** Suppose \(p \in M\) and \(X, Y \in \Gamma^\infty(M)\) generate the local flows \(\Phi\) and \(\Psi\) near \(p\). For \(\varepsilon > 0\) sufficiently small define the curve \(\sigma\) : \((-\varepsilon, \varepsilon) \mapsto M\) by \(\sigma(t) = \Psi_{-t} \circ \Phi_{-t} \circ \Psi_t \circ \Phi_t(p)\). Then \(\sigma'(0) = 0\) and \(\sigma''(0) = 2[X, Y]_p\).

Note that in general second derivatives do not define tangent vectors – it is only because the first derivative vanishes that the second derivative is a tangent vector! Before proving the proposition, establish the following intermediate step from the Lie bracket as a derivations on functions and to a more symmetric description of the Lie derivative in terms of flows.

**Lemma 3.27** Suppose \(p \in M\) and \(X, Y \in \Gamma^\infty(M)\) generate the local flows \(\Phi\) and \(\Psi\) near \(p\) and \(f \in C^\infty(p)\). Then

\[
[X, Y]_p f = \lim_{s, t \to 0} \frac{1}{st} \left( (f \circ \Psi_s \circ \Phi_t)(p) - (f \circ \Phi_t \circ \Psi_s)(p) \right)
\]
Proof (of the lemma). Let $p, X, Y, \Phi, \Psi$ and $f$ be as in the lemma. Consider the first term in $[X, Y]_p f = X_p(Y f) - Y_p(X f)$, using the definition for the Lie derivative of a function twice:

$$X_p(Y f) = \lim_{t \to 0} \frac{1}{2} ( (Y f)(\Phi_t(p)) - (Y f)(p) )$$

Switched order??

$$= \lim_{t \to 0} \frac{1}{2} \left( \lim_{s \to 0} \frac{1}{2} \left( (f \circ \Phi_s \circ \Phi_t)(p) - (f \circ \Phi_t)(p) \right) \right)$$

$$- \lim_{s \to 0} \frac{1}{2} \left( (f \circ \Psi_s \circ \Phi_t)(p) - (f \circ \Phi_t)(p) \right)$$

Combine this with the analogous expression for $Y_p(X f)$, namely

$$Y_p(X f) = \lim_{s, t \to 0} \frac{1}{2} \left( (f \circ \Phi_t \circ \Psi_s)(p) - (f \circ \Psi_s)(p) - (f \circ \Phi_t)(p) + f(p) \right)$$

to obtain the desired result.

Exercise 3.40 Suppose $f : \mathbb{R} \mapsto \mathbb{R}$ is twice continuously differentiable at $a \in \mathbb{R}$. Show that in this case the usual difference quotient $f''(a) = \lim_{h \to 0} \frac{1}{h^2} (f(a+h) - 2f(a) + f(a-h))$ simplifies to $f''(a) = \lim_{h \to 0} \frac{2}{h^2} (f(a + h) - f(a))$

Proof (of the proposition). Let $p, X, Y, \Phi, \Psi$ and $f$ be as in the proposition, and suppose $f \in C^\infty(p)$. Calculate

$$\sigma'_0 f = (f \circ \sigma)'(0) = \frac{d}{dt} \bigg|_0 (f \circ \Psi_{-t} \circ \Phi_{-t} \circ \Psi_t \circ \Phi_t)(p)$$

$$= \lim_{t \to 0} \frac{1}{2} \left( (f \circ \Psi_{-t} \circ \Phi_{-t} \circ \Psi_t \circ \Phi_t)(p) - f(p) \right)$$

Apply the chain-rule repeatedly, or directly, add and subtract suitable terms and take the limits of each difference separately

$$\sigma'_0 f = \lim_{t \to 0} \frac{1}{2} \left( (f \circ \Psi_{-t})(\Phi_{-t} \circ \Psi_t \circ \Phi_t)(p) - f(\Phi_{-t} \circ \Psi_t \circ \Phi_t)(p) \right)$$

$$+ \lim_{t \to 0} \frac{1}{2} \left( (f \circ \Phi_{-t})(\Psi_t \circ \Phi_t)(p) - f(\Psi_t \circ \Phi_t)(p) \right)$$

$$+ \lim_{t \to 0} \frac{1}{2} \left( (f \circ \Psi_t)(\Phi_t)(p) - f(\Phi_t)(p) \right)$$

$$+ \lim_{t \to 0} \frac{1}{2} \left( (f \circ \Phi_t)(p) - f(p) \right)$$

$$= (-Y f - X f + Y f + X f)(p) = 0$$

This last step uses the continuity of $X f$ and $Y f$, e.g. that $\lim_{t \to 0} (X f)(\Psi_t \circ \Phi_t)(p)) = (X f)(p)$, compare exercise 3.41.

Regarding the second part of the proposition, recognize that it suffices to consider the difference quotient $\frac{d^2}{dt^2} (f(\sigma(t)) - f(p))$ because the first derivative vanishes at $t = 0$, compare exercise 3.40. To directly employ lemma 3.27 insert the identity written as $\Phi_{-t} \circ (\Psi_{-t} \circ \Psi_t) \circ \Phi_t$, and use that $[X, Y]$ is continuous:

$$\frac{d^2}{dt^2} \bigg|_0 (f \circ \sigma)(t) = \lim_{t \to 0} \frac{2}{2} \left( f \circ \Psi_{-t} \circ \Phi_{-t} \circ \Psi_t \circ \Phi_t(p) - f(p) \right)$$

$$= \lim_{t \to 0} \frac{2}{2} \left( f(\Psi_{-t} \circ \Phi_{-t}(\Psi_t \circ \Phi_t)(p)) - (f \circ \Phi_{-t} \circ \Psi_{-t})(\Psi_t \circ \Phi_t(p)) \right)$$

$$= 2 [X, Y]_p f .$$
Exercise 3.41 Suppose $f, g: \mathbb{R} \mapsto \mathbb{R}$ with $g(0) = p \in \mathbb{R}$, $g \in C^0 (0)$, and $f \in C^1 (p)$. Rigorously prove that $f'(p) = \lim_{h \to 0} (f(g(h) + h) - f(g(h))).$

It may be convenient to employ the function $F: (s, t) \mapsto F(g(s) + t)$, or alternatively estimate the integral $\frac{1}{h} \int_{g(h)}^{g(h)+h} (f'(s) - f(p)) \, ds.$

Explain where your argument breaks down in the case of $p = 0$ and $f$ defined by $f(0) = 0$ and $f(x) = x^2 \cdot \sin \frac{1}{x}$ else (which is differentiable, but not continuously differentiable at zero).

Use this to rigorously justify the limits in the calculation of $\sigma'_0 f$ in the proof of the proposition. E.g. first consider the simple case of functions

Exercise 3.42 Under the hypotheses of lemma 3.27 show that

$$[X, Y]_p f = \lim_{t \to 0} \frac{1}{t} \left( (f \circ \Psi_{\sqrt{t}} \circ \Phi_{\sqrt{t}} \circ \Psi_{\sqrt{t}} \circ \Phi_{\sqrt{t}})(p) - f(p) \right)$$

(111)

Proof (of theorem 3.25).

Suppose $p \in M^n$, $X_1 \in \Gamma^\infty (M)$, $i = 1, \ldots k$ are such that $\{X_{1p}, \ldots X_{kp}\}$ are linearly independent, and that $[X_i, X_j] \equiv 0$ on an open neighborhood $O \subseteq M$ of $p$.

Since the tangent vectors $\{X_{1p}, \ldots X_{kp}\}$ are linearly independent at $p$ they necessarily are independent in some neighborhood $O' \subseteq O$ of $p$. Now suppose $(v, V)$ is a chart near $p$ such that $V \subseteq O'$ and w.l.o.g $\frac{\partial}{\partial v^j} |_p = X_{ip}$ for $i = 1, \ldots k$.

Exercise 3.43 Verify that one always can obtain such an adapted coordinate chart after performing, if necessary, a constant linear change of coordinates.

Let $\Phi^i$, $i = 1, \ldots k$ be the local flows of the vector fields $X_i$ defined near $p$. Without loss of generality, after restricting the domains as necessary, we may assume that each flow $\Phi^i$ is defined for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$ and for all $q$ in some open neighborhood $V' \subseteq V$ of $p$. Let $W = \{x \in \mathbb{R}^m : |x^i| < \varepsilon \text{ for } i = 1, \ldots m\}$ and define $V'' = V' \cap v^{-1}(W)$, which is still an open neighborhood of $p$.

Define the map $\Psi: v(V') \cap W \mapsto M$ by

$$\Psi(x) = \Phi^{k}_x \circ \cdots \circ \Phi^{2}_x \circ \Phi^{1}_x \circ v^{-1}(0, \ldots, 0, x^{k+1}, \ldots x^m)$$

(112)

Clearly at $a = 0$ the tangent vectors $\Psi_{a0}(D|_a 0) = \frac{\partial}{\partial v^j}|_a$ for $i = 1, \ldots m$ are aligned with the coordinates. Hence $\Psi_{a0}$ has full rank, and thus $\Psi_{x}x \in \Gamma^\infty (M)$ has full rank for $x$ near $0 \in \mathbb{R}^m$. Consequently the restriction of $\Psi$ to some open neighborhood $W' \subseteq W$ of $0 \in \mathbb{R}^m$ is a diffeomorphism (onto its image). Let $U = \Psi(W') \subseteq M$ and define $u: U \mapsto \mathbb{R}^m$ by $u = (\Psi|_{W'})^{-1}$. Clearly for all $k \leq n$, $\frac{\partial}{\partial w^k} \equiv X_k|_{U}$. Using that the vector fields $X_i$ commute on $U$, write

$$\Psi(x) = \Phi^i_x \circ (\Phi^k_{x^i} \circ \cdots \Phi^{i+1}_{x^{i+1}}) \circ (\Phi^{i-1}_{x^{i-1}} \circ \cdots \circ \Phi^2_{x^2} \circ \Phi^1_x) \circ v^{-1}(0, \ldots, 0, x^{k+1}, \ldots x^m)$$

(113)

and one readily sees that also $\frac{\partial}{\partial w^i} \equiv X_i|_{U}$ for all $i = 1, \ldots k$. ■
3.12 Distributions and integrability theorems

This section generalizes the first integrability theorem proven in the preceding section. It relaxes the condition that the vector fields pairwise commute to the weaker condition that the Lie brackets are linear combinations of the set of given vector fields. This main theorem is known as Frobenius integrability theorem. The natural language to address this situation is in terms of the linear spans of the vector fields, termed distributions. The section introduces some elegant language and terminology, but the main ideas – exploiting Lie brackets and commutativity – come from the preceding section.

**Definition 3.21** A distribution on a manifold $M$ is a function $\Delta$ that assigns to every $p \in M$ a subspace $\Delta_p(M) \subseteq T_pM$. If $\dim T_pM = k$ for all $p \in M$ then $\Delta$ is called a $k$-dimensional distribution. A distribution $\Delta$ is called smooth if for every $p \in M$ there exist an open neighborhood $U$ of $p$ and $k$ smooth vector fields $X_k \in \Gamma^\infty(U)$ such that for every $q \in U$, $\Delta_q$ is the linear span of $X_k(q), \ldots X_k(q)$. A vector field $X \in \Gamma^\infty(m)$ is said to belong to a distribution $\Delta$ if for all $p \in M$, $X_p \in \Delta(p)$. A $k$-dimensional submanifold $N \subseteq M$ is called an integral manifold of $\Delta$ if for all $q \in M$, $\iota_q(T_qN) = \Delta_q$ (where $\iota : N \hookrightarrow M$ is the inclusion map.)

A few comments are in order:

- Integral manifolds of distributions are similar to integral curves of vector fields, but without the time-parameterization.
- A collection of vector fields always determines a distribution. But in general this distribution need not have a well-defined dimension as is illustrated by a (single) vector field that vanishes at some (but not all) points.
- Not every $k$-dimensional distribution is determined by a collection of $k$-vector fields: Consider the Moebius-strip $M$ as the quotient of $\mathbb{R}^2$ under the equivalence relation $(x, y) \sim (x', y')$ if and only if $x' - x = 2\pi k$ and $y' = (-)^k y$ for some integer $k$ and let $\pi : \mathbb{R}^2 \mapsto M$ be the associated quotient map. Note that the vector field $D_2$ on $\mathbb{R}^2$ does not map to a vector field on $M$. However, the distribution $\Delta$ generated by $D_2$ maps to a distribution $\pi_* (\Delta)$ on $M$ which is not generated by any single vector field on $M$.
- While for every smooth vector field there exist unique integral curves through every point, in general a smooth $k$-dimensional distribution will not have any integral manifold if its dimension is larger than one. A most simple example is the smooth distribution on $\mathbb{R}^3$ spanned by the two vector fields $X = D_1 + yD_3$ and $Y = D_2 - xD_3$. As will become clear in the following this is an immediate consequence of the Lie bracket $[X, Y]$ being linearly independent of $\{X, Y\}$ (over $C^\infty(M)$).

**Exercise 3.44** Use a computer algebra system (or other drawing software) to generate images of the distribution $\Delta$ on $\mathbb{R}^3$ that is spanned by the vector fields $X = D_1 + xD_3$ and $Y = D_2 - xD_3$ (or $Y = D_2$ is an even simpler example). Explain graphically why $\Delta$ does not have any integral manifold.

We next provide a few useful notions before proceeding to Frobenius’ theorem.

**Definition 3.22** Suppose $\Phi \in C^\infty(M^m, N^n)$. Two vector fields $X \in \Gamma^\infty(M)$ and $Y \in \Gamma^\infty(N)$ are called $\Phi$-related if $\Phi_pX_p = Y_{\Phi(p)}$ for all $p \in M$. 

**Proposition 3.28** Suppose $\Phi: M \rightarrow N$ is a diffeomorphism and $Y \in \Gamma^\infty(M)$ generates the local flow $\Psi_t$. The $(\Phi_*Y) \in \Gamma^\infty(N)$ generates the local flow $\Phi \circ \Psi_t \circ \Phi^{-1}$.

A typical application is when $M = N$ and $\Phi$ is the local flow of a vector fields $X$ (evaluated at a fixed time $t$). Compare also exercise 3.39 for a Taylor expansion of the infinitesimal generator $(\Phi_*Y)$.

**Proof.** Suppose $p \in M$, $q = \Phi(p) \in N$, $f \in C^\infty(q)$, and calculate

\[
(\Phi_*Y)_q(f) = (\Phi_*\Phi^{-1}(q)Y_{\Phi^{-1}(q)})(f) = Y_{\Phi^{-1}(q)}(f \circ \Phi) = \lim_{h \to 0} \frac{1}{h} \left( ((f \circ \Phi) \circ \Psi_h)(\Phi^{-1}(q)) - (f \circ \Phi)(\Phi^{-1}(q)) \right) = \lim_{h \to 0} \frac{1}{h} \left( (f \circ (\Phi \circ \Psi_h \circ \Phi^{-1}))(q) - f(q) \right).
\]

**Proposition 3.29** Suppose $\Phi \in C^\infty(M^m, N^n)$, $X_1 \in \Gamma^\infty(M)$ is $\Phi$-related to $Y_1 \in \Gamma^\infty(N)$, and $X_2 \in \Gamma^\infty(M)$ is $\Phi$-related to $Y_2 \in \Gamma^\infty(N)$. Then $[X_1, X_2] \in \Gamma^\infty(M)$ is $\Phi$-related to $[Y_1, Y_2] \in \Gamma^\infty(N)$.

**Proof.** Suppose $\Phi \in C^\infty(M^m, N^n)$, $p \in M$, $f \in C^\infty(\Phi(p))$ and $X_i, Y_i$ are $\Phi$-related as in the proposition. Thus, by hypothesis, $(\Phi_*X_i)_p = Y_i_{\Phi(p)}$, i.e. $X_ip(f \circ \Phi) = Y_i_{\Phi(p)}f = (Yf)(\Phi(p))$ for $i = 1, 2$. Calculate:

\[
[X_1, X_2]_p(f \circ \Phi) = X_1(X_2(f \circ \Phi) - X_2(X_1(f \circ \Phi)))(p) = \left( X_1(Y_2f) \circ \Phi - X_2(Y_1f) \circ \Phi \right)(p) = \left( Y_1(Y_2f) - Y_2(Y_1f) \right)(\Phi(p)) = [Y_1, Y_2]_{\Phi(p)}f.
\]

**Definition 3.23** A smooth distribution $\Delta$ on a manifold $M$ is called involutive if whenever $X, Y \in \Gamma^\infty(M)$ belong to $\Delta$, then $[X, Y]$ belongs to $\Delta$.

**Proposition 3.30** Suppose $\Delta$ is a $k$-dimensional smooth distribution on $M$. Then $\Delta$ is involutive iff for every $p \in M$ there exists a neighborhood $U \subseteq M$ of $p$ and vector fields $X_1, \ldots, X_k \in \Gamma^\infty(M)$ which span $\Delta$ on $U$, and there exist functions $c^i_{ij} \in C^\infty(U)$ such that

\[
[X_i, X_j] = \sum_{\ell=1}^k c^\ell_{ij}X_\ell \quad \text{for all } i, j = 1, \ldots, k
\]

**Proof.** The only if part is clear from the definition. Conversely suppose $X, Y \in \Gamma^\infty(U)$ belong to $\Delta$. This means that there are smooth functions $f^i, g^j \in C^\infty(U)$ (compare the following exercise) such that $X = \sum_{i=1}^k f^i X_i$ and $Y = \sum_{j=1}^k g^j X_j$. Consequently,

\[
[X, Y] = \sum_{i, j=1}^k \left( f^i g^j [X_i, X_j] + f^i (X_i g^j) X_j - g^j (X_j f^i) X_i \right) = \sum_{j=1}^k \left( \sum_{i, \ell=1}^k f^\ell g^j c^\ell_{ij} + \sum_{i=1}^k (f^i (X_i g^j) - g^j (X_j f^i)) \right) X_j
\]

which clearly belongs to $\Delta$.

**Exercise 3.45** Elaborate why the functions $f^i$ and $g^j$ in the preceding proof are smooth.
Theorem 3.31 (Frobenius) Suppose $\Delta$ is an involutive $k$-dimensional smooth distribution on $M$ and $p \in M$. Then there exists a chart $(u, U)$ about $p$ with $u(p) = 0$ and $u'(p) = (-\varepsilon, \varepsilon)^m \subseteq \mathbb{R}^m$ such that for every $c = (c^{k+1}, \ldots c^m) \in (-\varepsilon, \varepsilon)^{m-k}$ the set $N_c = \{ q \in U : u^j(q) = c^j, \; j = k+1, \ldots m \}$ is an integral manifold of $\Delta$. Moreover, every connected integral manifold of the restriction of $\Delta$ to $U$ is contained in some $N_c$ as above.

Proof. Suppose $\Delta$ is an involutive $k$-dimensional smooth distribution on $M$ and $p \in M$. Start with any chart $(v, V)$ about $p$ such that $\Delta(p)$ is spanned by $\{ \frac{\partial}{\partial v^1}|_p, \ldots \frac{\partial}{\partial v^m}|_p \}$. (As in the previous section, this may always be achieved via a linear change of coordinates).

Let $W = \{ q \in V : v^{k+1}(q) = \ldots = v^m(q) = 0 \}$ and consider the projection $\pi : V \rightarrow W$ defined by $\pi(q) = v^{-1}(v^1(q), \ldots v^j(q), 0, \ldots 0)$. Then $\pi_*|_{\Delta(q)} : \Delta(p) \mapsto T_qW \subseteq T_pM$ is the identity map. Hence, by continuity, $\pi_*|_{\Delta(q)} : \Delta(q) \mapsto T_qW \subseteq T_qM$ is one-to-one for $q$ sufficiently close to $p$.

This allows us to define vector fields $X_j$ by setting $X_j(q)$ to be the unique tangent vector

$$X_j(q) \in \pi_*^{-1} \left( \frac{\partial}{\partial v^j}|_q \right) \cap \Delta(q)$$

(117)

Note that this defines smooth vector fields $X_j$ near $p$. [An alternative approach starts with vector fields $Y_j$ that locally span $\Delta$ and then find $\alpha_j^i \in C^\infty(V)$ such that $\sum_{j=1}^k \alpha_j^i \pi_* Y_j = \frac{\partial}{\partial v^i}|_q$]

The crux is that the vector fields $X_j$ and $\frac{\partial}{\partial v^j}$ are $\pi$-related (for each $j = 1, \ldots k$). Since

$$\pi_*[X_i, X_j]q = [\frac{\partial}{\partial v^i}, \frac{\partial}{\partial v^j}]|_{\pi(q)} = 0$$

(118)

and $\pi_*|_{\Delta(q)}$ is one-to-one this implies that $[X_i, X_j]q = 0$. Since $\{X_{1q}, \ldots X_{kq}\}$ also form a basis for $\Delta(q)$ for $q$ sufficiently close to $p$, the integrability theorem 3.25 of the previous section applies and yields a coordinate chart $(u, U)$ about $p$ with $U \subseteq V$ such that $X_j|_{u(U)} \equiv \frac{\partial}{\partial u^j}$ for $j = 1, \ldots k$.

The manifolds $N_c = \{ q \in U : u^{k+1} = c^{k+1}, \ldots u^m(q) = c^m \}$ are integral manifolds for $\Delta|_U$.

Conversely, suppose $N \subseteq U$ is any connected integral manifold of $\Delta|_U$, and $\iota : N \rightarrow U$ is the inclusion map. If $q \in N$ and $Y_q \in T_qN$ then $Y_q(w^j \circ \iota) = (\iota_* Y_q) w^j = 0$ for $j = k+1, \ldots m$ because $\iota_* (Y_q) \in \Delta(q)$ by construction. This shows that $w^j, j = k+1, \ldots m$ are constant on $N$ and hence $N \subseteq N_c$ for some $c$ as above. ■

Exercise 3.46 Consider $M = \mathbb{R}^3 \setminus \{(0, 0, 0)\}$ and consider the analytic distribution $\Delta$ spanned by the vector fields $X_1 = -x^2 D_1 + x^1 D_2$, $X_2 = -x^3 D_2 + x^2 D_3$, and $X_3 = -x^1 D_3 + x^3 D_1$. Verify that $\Delta_p \subseteq T_p M$ is a 2-dimensional subspace for every $p \in M$. Explicitly calculate the associated flows $\Phi^1$, $\Phi^2$, and $\Phi^3$. Verify that these exist globally on $\mathbb{R} \times M$.

Calculate a multiplication table for the Lie brackets $[X_i, X_j]$ to verify that $\Delta$ is involutive, and conclude that $M$ is “foliated” by 2-dimensional integral manifolds of $\Delta$ (identify these!).

Exercise 3.47 (raw and untested, continuation of exercise 3.46).

Use the notation of the preceding exercise and consider the set $G$ of all compositions $G = \{ \Phi^2_{x^3} \circ \Phi^2_{x^2} \circ \Phi^1_{x^1} : x^1, x^2, x^3 \in \mathbb{R}^3 \} \subseteq \text{Diff}(M)$ as a subset of all diffeomorphisms of $M$. Verify that $G$ is a Lie group, i.e. it combines the structures of a smooth manifold and a group in such a way that multiplication and inverses are smooth operations (with respect to the $C^\infty$-topology on $G$ [[xxx reference xxx]])

Find a Lie group isomorphism (i.e. a diffeomorphism that preserves the group operations) from $G$ to the set $\text{SO}(3)$ of special orthogonal $3 \times 3$ matrices, i.e. $3 \times 3$ matrices $A$ satisfying $A^T A = \text{Id}_{3 \times 3}$. Exhibit a local coordinate chart $(u, U)$ of $\text{SO}(3)$ about $p = \text{Id}_{3 \times 3}$ and explicitly find formulae for the images of the vector fields $X_i$ under the bijection from above.
the word: foliation infinitesimal symmetries revisit exercise 3.19
3.13 Tensors and alternating forms

The goal of this section is to prepare for an alternative way of characterizing integrability, which relies on differential forms (as opposed on tangent vector fields). The algebraic structure of differential k-forms, i.e. their products and exterior derivatives require some preliminary constructions involving tensors. This section can only give a very brief survey of some basic concepts, terminology and select fundamental properties. The language of tensors developed here will also be beneficial later for precise descriptions of objects in Riemannian geometry.

To motivate tensors (and subsequently tensor fields) we explore two basic examples – linear maps between two vector spaces and quadratic forms on a vector space. The key points will be to contrast multi-linear with linear maps, and contrast Cartesian products and direct sums with the spaces of linear maps and tensor products. The examples motivate an algebraic characterization of tensor products as objects that naturally allow one to factor all multi-linear (bilinear) maps. Here we restrict the attention to finite dimensional vector spaces over the field \( \mathbb{R} \) of real numbers (similar notions may be developed for infinite dimensions and for modules over commutative rings). Note that there are many settings in differential geometry where a vector space like the space \( \Gamma^0(M) \) of smooth vector fields may not only be considered as a vector space over the field \( \mathbb{R} \), but also its structure as a module over the ring \( C^\infty(M) \) is utilized often, for example one may consider \( \Gamma^0(M) \) and \( \Omega^1(M) \) not only as vector spaces over \( \mathbb{R} \), but it also has the structure of a module over the ring \( C^\infty(M) \) of smooth functions – and consequently one needs to meticulously distinguish different tensor products such as

\[
\Gamma^\infty(M) \otimes_{\mathbb{R}} \Omega^1(M) \quad \text{and} \quad \Gamma^\infty(M) \otimes_{C^\infty(M)} \Omega^1(M)
\]

However, in this section we will mostly consider tensor products at individual points \( p \in M \), which only involve vector spaces like \( T_pM \) and \( T_p^*M \) over the field \( \mathbb{R} \).

Let \( V \) and \( W \) be finite dimensional vector spaces. Write \( \text{Hom}(V, W) \) for the vector space of linear maps from \( V \) to \( W \). To exhibit the similarity between the Cartesian product \( V^* \times W \) and the space of linear maps \( \text{Hom}(V, W) \), define a bilinear map \( \Phi: V^* \times W \rightarrow \text{Hom}(V, W) \). To characterize \( \Phi(\lambda, w) \in \text{Hom}(V, W) \) (for every \( \lambda \in V^* \) and every \( w \in W \)) we need to specify \( \Phi(\lambda, w)(v) \in W \) for every \( v \in V \). The (only, nontrivial, natural) way to do this, is to define

\[
\Phi : V^* \times W \rightarrow \text{Hom}(V, W) \quad \text{for} \quad \lambda \in V^*, \ v \in V \quad \text{and} \quad w \in W \quad \text{by} \quad \Phi(\lambda, w)(v) = \lambda(v)w \quad (120)
\]

Bilinearity means that for any \( \lambda_1, \lambda_2 \in V^*, \ w_1, w_2 \in W \) and \( c \in \mathbb{R} \)

\[
\Phi(c\lambda_1 + \lambda_2, w) = c \cdot \Phi(\lambda_1, w) + \Phi(\lambda_2, w) \\
\Phi(\lambda, cw_1 + w_2) = c \cdot \Phi(\lambda, w_1) + \Phi(\lambda, w_2) \quad (121)
\]

Note that the map \( \Phi \) is not one-to-one – indeed for any \( c \in \mathbb{R} \setminus \{0\}, \lambda \in V^* \), and \( w \in W \), the linear maps \( \Phi(\lambda, w) = \Phi(c\lambda, \frac{1}{c}w) \) are identical. Moreover, the additive structure of the direct sum \( V^* \oplus W \) (which as a set is the same because \( V^* \times W \) i.e. \( (\lambda_1, w_1) + (\lambda_2, w_2) = (\lambda_1 + \lambda_2, w_1 + w_2) \) is also mismatched by the bilinear (as opposed to linear) map \( \Phi \) as

\[
\Phi(\lambda_1 + \lambda_2, w_1 + w_2)(v) = (\lambda_1 + \lambda_2)(v) \cdot (w_1 + w_2) \\
= \lambda_1(v)w_1 + \lambda_2(v)w_2 + \lambda_1(v)w_1 + \lambda_2(v)w_2 \\
\neq \lambda_1(v)w_1 + \lambda_2(v)w_2 \\
= \Phi(\lambda_1, w_1)(v) + \Phi(\lambda_2, w_2)(v) \quad (122)
\]
Exercise 3.48 Suppose that \( \{\lambda^1, \ldots, \lambda^m\} \) and \( \{w_1, \ldots, w_n\} \) are bases for vector spaces \( V^* \) and \( W \), respectively. With \( \Phi \) as defined above, show that \( \{\Phi(\lambda^i, w_j) : i \leq m, j \leq n\} \) is a basis for \( \text{Hom}(V, W) \).

As a consequence of the exercise we conclude that for every vector space \( Z \) every bilinear map \( \beta : V^* \times W \to Z \) factors uniquely into a composition \( \beta = \Psi_\beta \circ \Phi \) with \( \Phi \) as above and \( \Psi_\beta \) a uniquely determined linear map \( \Psi_\beta : \text{Hom}(V, W) \to Z \).

\[
\begin{array}{ccc}
V^* \times W & \overset{\Phi}{\longrightarrow} & \text{Hom}(V, W) \\
\beta \downarrow & & \downarrow \Psi \\
Z & & 
\end{array}
\]

As a second example consider quadratic (bilinear) forms \( Q : V \times V \to k = \mathbb{R} \) on a finite dimensional vector space \( V \). Bilinearity means that for \( v, v_1, v_2 \in V \) and \( c \in \mathbb{R} \)
\[
\begin{align*}
Q(cv_1 + v_2, v) &= cQ(v_1, v) + Q(v_2, v) \\
Q(v, cv_1 + v_2) &= cQ(v, v_1) + Q(v, v_2)
\end{align*}
\] (123)

We will factor \( Q \) as a linear map composed with a universal bilinear map . . .

Exercise 3.49 Suppose \( V, Q, \) and \( \{\lambda^1, \ldots, \lambda^n\} \) are as above. Show that there exist uniquely determined \( a_{ij} \in \mathbb{R} \) such that
\[
Q(v, v') = \sum_{i,j=1}^n a_{ij} \lambda^i(v) \cdot \lambda^j(v') \quad \text{for all } v, v' \in V.
\] (124)

In the following we shall characterize the space in which \( \sum_{i,j=1}^n a_{ij} \lambda^i(\cdot) \lambda^j(\cdot) \) naturally lives in. Note, if \( c \neq 0,1 \) then \( (\lambda, \lambda) \) and \( (c\lambda, \frac{1}{c}\lambda) \) are different as elements of \( V^* \times V^* \). However, they determine the same quadratic form \( Q \) via the map \( \Phi \) from \( V^* \times V^* \) to the vector space of quadratic bilinear forms on \( V \), defined by
\[
\Phi(\lambda, \lambda')(v, w) = \lambda(v) \cdot \lambda'(w)
\] (125)

Both examples suggest to build from the Cartesian product of vector spaces a new vector space with a linear structure analogous to that of spaces of linear maps. More specifically we want to identify (on \( V \times W \)) e.g. \( (cv, w) \) with \( (v, cw) \) and have an additive structure such that \( (v + v', w + w') = (v, w) + (v', w) + (v', w') \).

Abtractly, consider finite dimensional vector spaces \( V_1, \ldots, V_k \). We would like to find a vector space \( T \) together with a multi-linear map \( \Phi : V_1 \times \ldots \times V_k \to T \) such that for every vector space \( W \) and every multi-linear map \( \mu : V_1 \times \ldots \times V_k \to W \) there exists a unique linear map \( \Psi_\mu : T \to W \) such that \( \mu = \Psi_\mu \circ \Phi \).

\[
\begin{array}{ccc}
V \times \ldots \times V_k & \overset{\Phi}{\longrightarrow} & T \\
\mu \downarrow & & \downarrow \Psi \\
W & & 
\end{array}
\]
Here multi-linearity means that for \( v_1 \in V_1, \ldots, v_i, v_i' \in V_i, \ldots, v_k \in V_k \) and \( c \in \mathbb{R} \)
\[
\mu(v_1, \ldots, v_{i-1}, cv_i + v_i', v_{i+1} \ldots v_k) = c \cdot \mu(v_1, \ldots, v_i, \ldots v_k) + \mu(v_1, \ldots, v_i', \ldots v_k)
\]
Using standard algebraic arguments one easily shows that such a universal space \( T \) and associated map \( \Phi \) always exist, and they they are unique, up to isomorphisms. [[Following Sternberg]] the construction starts with the Cartesian product \( V = V_1 \times \ldots \times V_k \), then considers the free vector space \( \overline{V} \) generated by \( V \) (i.e. the set of all formal linear combinations of elements in \( V \)). Let \( \overline{R} \subseteq \overline{V} \) be the subspace generated by all elements of the from
\[
c(v_1, \ldots v_k) - (v_1, \ldots, v_{i-1}, cv_i, v_{i+1} \ldots v_k) \text{ and all elements of the form }
(v_1, \ldots, v_{i-1}, v_i + v_i', v_{i+1} \ldots v_k) - (v_1, \ldots, v_i, \ldots v_k) - (v_1, \ldots, v_i', \ldots v_k)
\]
where \( c \in \mathbb{R}, v_j \in V_j \) for \( 1 \leq j \leq k \) and \( v_i' \in V_i \) for some \( i \leq k \).

**Definition 3.24** For \( V_1, \ldots, V_k, \overline{V} \) and \( \overline{R} \) as above define the tensor product of the spaces \( V_j \) as the quotient
\[
V_1 \otimes \ldots \otimes V_k = \overline{V} / \overline{R}.
\]

Define the multi-linear map \( \Phi: V_1 \times \ldots \times V_k \mapsto V_1 \otimes \ldots \otimes V_k \) which maps \((v_1, \ldots, v_k)\) to its coset mod \( \overline{R} \) (viewing \((v_1, \ldots, v_k)\) as an element of \( \overline{V} \)). Write \( v_1 \otimes \ldots v_k \) for \( \Phi(v_1, \ldots, v_k) \).

For finite dimensional vector spaces \( V_1, \ldots, V_k \) over \( \mathbb{R} \) the tensor product \( T = V_1 \otimes \ldots \otimes V_k \) (together with the canonical map \( \Phi \)) is the desired universal space:

**Proposition 3.32** For every vector space \( W \) and every multi-linear map \( \mu: V_1 \times \ldots \times V_k \mapsto W \) there exists a unique linear map \( \Psi_\mu: V_1 \otimes \ldots \otimes V_k \mapsto W \) such that \( \mu = \Psi_\mu \circ \Phi \) (with \( \Phi \) as above).

**Proof.** Suppose \( \mu: V_1 \times \ldots \times V_k \mapsto W \) is a multi-linear map. Define a map \( \overline{\Psi}_\mu: \overline{V} \mapsto W \) on generators by \( \overline{\Psi}_\mu(v_1, \ldots, v_k) = \mu(v_1, \ldots, v_k) \) and extend linearly. Since \( \mu \) is multi-linear, the restriction of \( \overline{\Psi}_\mu \) to \( \overline{R} \) is identically zero, and hence this defines a map \( \Psi_\mu \) on the quotient. Uniqueness is clear as the image of \( \Phi \) contains a set of generators for \( V_1 \otimes \ldots \otimes V_k \) and any other such map \( \Psi_\mu \) necessarily must agree with \( \Psi_\mu \) on this set, hence agree with \( \Psi_\mu \) everywhere. \( \blacksquare \)

Note that, as defined above, the tensor product is naturally associative, it is not commutative as clearly \( v \otimes w \) and \( w \otimes v \) denote different objects in different spaces – but there is a natural isomorphism between e.g. \( V \otimes W \) and \( W \otimes V \).

It is instructive to connect this notion of tensor products to familiar objects like matrices. Going back to the first example, suppose \( \beta = \{v_1, \ldots, v_m\} \) and \( \gamma = \{w_1, \ldots, w_n\} \) are bases for vector spaces \( V \) and \( W \) and \( \beta^* = \{\lambda^1, \ldots, \lambda^m\} \) is the associated dual basis for \( V^* \). Moreover, suppose \( L \in \text{Hom}(V,W) \) is a linear map from \( V \) to \( W \). Then there exist unique \( a_{ij} \in \mathbb{R} \), such that for all \( v \in V \), \( L(v) = \sum_{j=1}^n \sum_{j=1}^m a_{ij} \lambda^j(v) w_i \). Pictorially identify the product \( \lambda^j \otimes w_i \) with the matrix
\[
\lambda^j \otimes w_i \leftrightarrow \begin{pmatrix}
0 & \ldots & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots \\
0 & \ldots & 0 & \ldots & 0
\end{pmatrix} 
\]
where \( j \)-th column \( \downarrow \) \( i \)-th row \( \leftarrow \) (129)
[Note that if, as usual, $\lambda^j$ is written as a row vector and $w_i$ is written as a column vector then their tensor product $\lambda^j \otimes w_i$ corresponds to the matrix product "$w_i \cdot \lambda^j$" in reversed order . . .] The linear combination of such matrices with coefficients $a_{ij}^k$ yields a matrix $A$ which is the matrix representation of the linear map $L$ with respect to the bases $\beta$ and $\gamma$. In the case that $\beta$ and $\gamma$ are the standard bases for $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$ one usually writes $v_i = w_i = e_i$ and $\lambda^j = e^j$ and the matrix displayed above corresponds to $e^j \otimes e_i$. [[As above, some may prefer to change the order of the factors and write $e_i \otimes e^j$ to have a closer match to matrix-products]].

Exercise 3.50 Consider the example of $V = W = \mathbb{R}^2$ with standard basis $\{e_1, e_2\}$, and the bilinear map $\Phi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2 \otimes \mathbb{R}^2$ defined on generators by $\Phi(e_i, e_j) = e_i \otimes e_j$. Identify $e_i \otimes e_j$ with the matrix whose entry in the $i$-th row and $j$-th column is 1 and whose other entries are all zero. (i) Show that there does not exist $(v, w) \in V \times W$ such that $\Phi(v, w) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

(ii) Describe the set of all matrices that lie in the image of $\Phi$. [[Note that not every matrix is in the image, but every $2 \times 2$ matrix is a linear combination of $\Phi(e_i, e_j)$ with $i, j = 1, 2$ (all combinations).]] (iii) Calculate the derivative of $\Phi$, and its rank.

In the classical literature one often finds a characterization of tensors as objects that obey certain rules for transformations. In the context of the preceding discussion, these refer to changes of bases in the factors $V_i$. In the general context of differential manifolds these changes of bases in the tangent and cotangent bases arise from changes of local coordinates on the manifold. For illustration of the kinds of expression involved consider a single vector space $V$ with bases $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_n\}$. Let $V^*$ be the dual space with associated dual bases $\beta^* = \{\lambda^1, \ldots, \lambda^n\}$ and $\gamma^* = \{\mu^1, \ldots, \mu^n\}$. For illustration consider the tensor product $T = V \otimes V \otimes V^*$. Bases for this space is given by $\{v_i \otimes v_j \otimes \lambda^k : i, j, k \leq n\}$ and $\{w_i \otimes w_j \otimes \mu^k : i, j, k \leq n\}$. Thus any element $z \in T$ can be written uniquely as linear combination

$$z = \sum_{i,j,k} a^{ij}_k \ v_i \otimes v_j \otimes \lambda^k = \sum_{i,j,k} b^{ij}_k \ w_i \otimes w_j \otimes \mu^k \quad (130)$$

The transformation rules refer to the identities that relate the coordinates $a^{ij}_k$ to the coordinates $b^{ij}_k$. Specifically, there exist $c^j_i, d^i_j \in \mathbb{R}$ such that

$$w_i = \sum_j c^j_i \ v_j \quad \text{and} \quad \mu^i = \sum_j d^i_j \ \lambda^j \quad \text{and} \quad (131)$$

The duality of the bases means that $\lambda^i(v_j) = \mu^i(w_j) = \delta_{i,j}$. The following small calculation reaffirms that the matrices $c^j_i$ and $d^i_j$ are inverses of each other:

$$\delta_{i,j} = \mu^i(w_j) = \left( \sum_k d^i_k \lambda^k \right) \left( \sum_\ell c^\ell_j v_\ell \right) = \sum_{k,\ell} d^i_k c^\ell_j \lambda^k(v_\ell) = \sum_{k,\ell} \delta_{k,\ell} c^\ell_j \lambda^k(v_\ell) = \sum_k d^i_k c^\ell_j \lambda^k(v_\ell) = \sum_k d^i_k \delta_{k,\ell} (132)$$

To obtain the transformation rules from $a^{ij}_k$ to $b^{ij}_k$ calculate

$$z = \sum_{i,j,k} a^{ij}_k \ v_i \otimes v_j \otimes \lambda^k = \sum_{r,s,t} \left( \sum_{i,j,k} a^{ij}_k \ c^j_i \ d^i_j \delta_{k,t} \right) w_r \otimes w_s \otimes \mu^t \quad (133)$$
In addition to the standard identification of $\text{Hom}(V, W)$ with $V^* \otimes W$ there are many similar, useful relations.

- A natural isomorphism between $\text{Hom}(U \otimes V, W)$ and $\text{Hom}(U, \text{Hom}(V, W))$ is provided by
  \[
  (\Phi(F)(u))(v) = F(u \otimes v) \quad (134)
  \]
- A natural isomorphism between $\text{Hom}(U, V) \otimes W$ and $\text{Hom}(U, V \otimes W)$ is provided by
  \[
  \Phi(F \otimes w)(u) = F(u) \otimes w \quad (135)
  \]
  In the special case that $V = \mathbb{R}$ this yields the isomorphism between $U^* \otimes W$ and $\text{Hom}(U, W)$.
- A natural isomorphism between $U \otimes \text{Hom}(V, W)$ and $\text{Hom}(\text{Hom}(U, V), W)$ is provided by
  \[
  \Phi(u \otimes G)(F) = F(G(u)) \quad (136)
  \]
  The special case that $V = W = \mathbb{R}$ yields the isomorphism between $U^{**}$ and $U$.
- A natural isomorphism between $\text{Hom}(U, V) \otimes \text{Hom}(W, Z)$ and $\text{Hom}(U \otimes V, W \otimes W)$ is provided by
  \[
  \Phi(F \otimes G)(u \otimes v)) = F(u) \otimes G(v) \quad (137)
  \]
  The special case that $V = Z = \mathbb{R}$ yields the isomorphism between $U^* \otimes V^*$ and $(U \otimes V)^*$.

Our primary interest is in tensor products of tangent spaces and co-tangent spaces, i.e. in tensor products where each factor is either the same vector space $V$ or its dual $V^*$. Due to the lack of commutativity there are many different higher order products that all arise from a single vector space and its dual. Taking formal linear combinations one conveniently may combine all possible tensor products into an algebra of tensor products:

**Definition 3.25** [[Sternberg defines]] the tensor algebra [[?]] of a vector space $V$ is the direct sum of all tensor products of $V$ and $V^*$, i.e.

\[
\mathcal{T}(V) = \mathbb{R} \oplus V \oplus V^* \oplus V \otimes V \oplus V \otimes V^* \oplus V^* \otimes V \oplus V^* \otimes V^* \oplus V \otimes V \otimes V \otimes V \oplus \ldots \quad (138)
\]

We shall usually call the subalgebra $\mathcal{T}(V)$ of contravariant tensors the tensor algebra of $V$:

\[
\mathcal{T}(V) = \mathbb{R} \oplus V \oplus V \otimes V \oplus V \otimes V \otimes V \oplus V \otimes V \otimes V \otimes V \oplus \ldots \quad (139)
\]

An element of a tensor-product that contains $r$ factors of $V$ and $s$ factors of $V^*$ is also called a tensor of contravariant degree $r$ and covariant degree $s$, or briefly a tensor of type $(r, s)$. The space of all tensors of type $(r, s)$ is denoted $\mathcal{T}_r^s(V)$. [[Note that unless $r = 0$ or $s = 0$ this does not identify the space as e.g. $V \otimes V^* \neq V^* \otimes V$ are both of type $(1, 1)$.]]

In addition to linear combinations and tensor products, the tensor algebra is equipped with the following operation which generalizes the *trace* of a linear map (or of a matrix):
Definition 3.26 Consider a tensor product $V_1 \otimes \ldots \otimes V_{r+s}$ of type $(r,s)$, i.e. where $r$ of the $V_k$ are equal to $V$ and $s$ of the $V_k$ are equal to $V^*$ and in particular $V_1 = V$ and $V_2 = V^*$. The contraction (of the $i$-th contravariant and $j$-th covariant index) of tensors in this space is defined by
\[
(v_1 \otimes \ldots \otimes v_{r+s}) \mapsto \begin{cases} 
 v_j(v_i) \cdot (v_1 \otimes \ldots \otimes v_{i-1} \otimes v_{i+1} \ldots v_{j-1} \otimes v_{j+1} \otimes \ldots \otimes v_{r+s}) & \text{if } i < j \\
 v_j(v_i) \cdot (v_1 \otimes \ldots \otimes v_{j-1} \otimes v_{j+1} \ldots v_{i-1} \otimes v_{i+1} \otimes \ldots \otimes v_{r+s}) & \text{else}.
\end{cases}
\] (140)

Proposition 3.33 Contractions are multi-linear maps. A contraction of a tensor of type $(r,s)$ is again a tensor, and it is a tensor of type $(r-1,s-1)$.

A very useful result is that every linear map between vector spaces uniquely extends to tensor products. This is an immediate corollary of a fundamental algebraic property:

Proposition 3.34 If $A$ is an associative algebra with 1 over $\mathbb{R}$ and $\phi: V \mapsto \mathbb{R}$ is linear then there exists a unique extension $\Phi: T(V) \mapsto A$ which is linear, which preserves products, and which is such that $\Phi(1) = 1$.

Corollary 3.35 Any $\phi \in \text{Hom}(V,W)$ extends to a unique homomorphism $\Phi: T(V) \mapsto T(W)$. The universal example is the map that maps a basis of an $n$-dimensional vector space $V$ to a set $x^1, \ldots, x^n$ of noncommuting indeterminates. Its unique extension maps the tensor algebra $T(V)$ to the algebra of noncommuting polynomials in $x_1, \ldots, x_n$.

Arguably the most useful subspace of the tensor algebra is the space of alternating or antisymmetric tensors – it provides a very concise and elegant description of higher order differential forms. Along the way it is convenient to also define symmetric tensors which have less dramatic properties.

In the following we shall consider a fixed (finite dimensional) vector space $V$ (which in differential geometry typically stands for either $T_p M$ or $T^*_p M$). Write $\Sigma_k$ for the symmetric group, or group of permutations of $k$ letters. The group $\Sigma_k$ naturally acts on $k$-fold tensor products of $V$. For $\sigma \in \Sigma_k$ define a map $\tilde{\sigma}: T_k(V) \mapsto T_k(V)$ on generators by \[[\text{compare polynomials/monomials }]
\[
\tilde{\sigma}(v_1 \otimes \ldots \otimes v_n) = v_{\sigma(1)} \otimes \ldots \otimes v_{\sigma(n)} \quad \text{for } v_j \in V
\] (141)
and extend linearly to $T_k(V)$.

Definition 3.27 A tensor $x \in T_k(V)$ is called symmetric if $\tilde{\sigma}(x) = x$ for all $\sigma \in \Sigma_k$ and it is called alternating (or skew symmetric or antisymmetric) if $\tilde{\sigma}(x) = \text{sgn}(\sigma) x$ for all $\sigma \in \Sigma_k$. Write $S_k(V)$ and $A_k(V)$ for the subsets of all symmetric and alternating tensors in $T_k(V)$, respectively. It is convenient to also define the set of alternating $k$-forms $\Lambda^k(V) = A_k(V^*) \subset T^k(V)$.

Definition 3.28 A linear map $L: T_k(V) \mapsto W$ is called symmetric (resp. alternating) if $L \circ \tilde{\sigma} = L$ (resp. $L \circ \tilde{\sigma} = \text{sgn}(\sigma) L$) for all $\sigma \in \Sigma_k$

(142)

Proposition 3.36 The maps $\text{Sym}_k, \text{Alt}_k : T_k(V) \mapsto T_k(V)$ defined by
\[
\text{Sym}_k(x) = \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \tilde{\sigma}(x) \quad \text{and} \quad \text{Alt}_k(x) = \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \text{sgn}(\sigma) \cdot \tilde{\sigma}(x)
\] (143)
are projections onto the symmetric and alternating tensors $S_k(V)$ and $A_k(V)$, respectively.
Exercise 3.51 Verify that \( S_k(V) \) and \( A_k(V) \) are subspaces of \( T_k(V) \), and prove proposition 3.36.

Proposition 3.37 Suppose \( \Phi : V \to W \) is a linear map between vector spaces. Then the map \( \Phi^* : T(W) \leftrightarrow T(V) \) defined on generators by \( \Phi^*(w^1 \otimes \ldots \otimes w^k) = ((\Phi^*w^1) \otimes \ldots \otimes \Phi^*(w^k)) \) restricts for any \( k \geq 0 \) to a map \( \Phi^* \) mapping \( \Lambda^k(W) \) to \( \Lambda^k(V) \).

It is convenient to define \( \Lambda^0(V) = \mathbb{R} \). Note that \( \Lambda^1(V) = V^* \) (there is only one permutation \( \sigma \) in \( \Sigma_1 \), and clearly \( \tilde{\sigma}(x) = 1 \cdot x = \text{sgn}(\sigma) \cdot x \). Also it is not very hard to see that the dimension of \( \Lambda^n(V) \) is one (if \( \dim(V) = n \) – and a basis for it is the determinant function suitably interpreted – the usual proof in linear algebra [[see e.g. Strang]] first show that for any pair of alternating functions on \( \mathbb{R}^n \) one is a multiple of the other …

Definition 3.29 Define \( \Lambda(V) = \Lambda^0(V) \oplus \Lambda^1(V) \oplus \Lambda^2(V) \oplus \ldots \) to be the direct sum of all alternating forms of all orders \( k \), i.e. the space of all formal linear combinations of alternating forms. Define the wedge product \( \wedge : \Lambda(V) \leftrightarrow \Lambda(V) \) on homogeneous elements \( \omega \in \Lambda^k(V) \) and \( \eta \in \Lambda^\ell(V) \) by

\[
\omega \wedge \eta = \frac{(k+\ell)!}{k!\ell!} \text{Alt}(\omega \otimes \eta) \quad \text{and extend linearly to} \quad \Lambda(V)
\]

(144)

From the construction it is clear that \( \wedge \) is bilinear and that \( \eta \wedge \omega = (\) \(-1)^{k\ell} \omega \wedge \eta \). In particular, if \( k \) is odd, then \( \omega \wedge \omega = 0 \). It takes a little more work to show associativity. Intermediate steps are to first establish the following:

Exercise 3.52 Suppose \( x, y \in T(V) \) and \( \text{Alt}(x) = 0 \). Then \( \text{Alt}(x \otimes y) = \text{Alt}(y \otimes x) = 0 \).

Suppose \( x, y \in T(V) \). Then \( \text{Alt}(\text{Alt}(x \otimes y) \otimes z) = \text{Alt}(x \otimes y \otimes z) = \text{Alt}(x \otimes \text{Alt}(y \otimes z)) \).

(Show this for homogeneous tensors \( x \in T^k(V), y \in T^\ell(V), z \in T^m(V) \) first.)

Proposition 3.38 The vector space \( \Lambda \) with the wedge product \( \wedge \) is an associative algebra.

Theorem 3.39 Suppose \( \{\lambda^1, \ldots, \lambda^n\} \) is a basis for \( V^* \). Then

\[
\{\lambda^{i_1} \wedge \ldots \wedge \lambda^{i_k} : 1 \leq i_1 \leq \ldots \leq i_k \leq n\}
\]

is a basis for \( \Lambda^k(V) \) and consequently \( \dim \Lambda^k(V) = \binom{n}{k} \).

Proof. Suppose that \( \{v_1, \ldots, v_n\} \) is a basis for the vector space \( V \) and \( \{\lambda^1, \ldots, \lambda^n\} \) is the dual basis for \( V^* \). Suppose \( \lambda \in \Lambda^k \). Then there exist \( c_{i_1 \ldots i_k} \in \mathbb{R} \) such that

\[
\lambda = \text{Alt}(\lambda) = \text{Alt} \left( \sum_{i_1 \ldots i_k \leq n} c_{i_1 \ldots i_k} \lambda^{i_1} \otimes \ldots \otimes \lambda^{i_k} \right) = \sum_{i_1 \ldots i_k \leq n} c_{i_1 \ldots i_k} \text{Alt} \left( \lambda^{i_1} \otimes \ldots \otimes \lambda^{i_k} \right)
\]

(146)

Each \( \text{Alt} \left( \lambda^{i_1} \otimes \ldots \otimes \lambda^{i_k} \right) \) is either zero or of the form \( \pm \frac{1}{k!} \lambda^{j_1} \wedge \ldots \wedge \lambda^{j_k} \) for some \( 1 \leq j_1 \leq \ldots \leq j_k \leq n \) and hence the elements of this form span \( \Lambda^k(V) \).

On the other hand suppose a linear combination \( \sum_j c_{i_1 \ldots i_k} \lambda^{i_1} \wedge \ldots \wedge \lambda^{i_k} \) is zero, where the sum
extends over all increasing $k$-tupels $I$, i.e. all $(i_1, \ldots, i_k)$ such that $1 \leq i_1 \leq \ldots i_k \leq n$. Then evaluate this linear functional on $v_{j_1} \otimes \ldots \otimes v_{j_k}$ for any $1 \leq j_1, \ldots, j_k \leq n$ to obtain

$$0 = \left( \sum_I c_{i_1, \ldots, i_k} \lambda^{i_1} \wedge \ldots \wedge \lambda^{i_k} \right) \left( v_{j_1} \otimes \ldots \otimes v_{j_k} \right)$$

$$= \left( \sum_I c_{i_1, \ldots, i_k} \frac{(1+\ldots+1)}{1! \ldots 1!} \text{Alt}(\lambda^{i_1} \otimes \ldots \otimes \lambda^{i_k}) \right) \left( v_{j_1} \otimes \ldots \otimes v_{j_k} \right)$$

$$= \left( \sum_I \sum_{\sigma \in S_k} c_{i_1, \ldots, i_k} \text{sgn}(\sigma) \cdot \lambda^{i_1} \otimes \ldots \otimes \lambda^{i_k} \circ \sigma \right) \left( v_{j_1} \otimes \ldots \otimes v_{j_k} \right)$$

$$= c_{j_1, \ldots, j_k}, \quad \text{thus proving linear independence.} \Box$$

Corollary 3.40 $\lambda^1, \ldots, \lambda^k \in \Lambda^1(V)$ are linearly independent if and only if $\lambda^1 \wedge \ldots \wedge \lambda^k \neq 0$

Proof. Suppose that $\lambda^1, \ldots, \lambda^k \in \Lambda^1(V)$ are linearly independent. Then this set may be extended to a basis $\{\lambda^1, \ldots, \lambda^n\}$ of $\Lambda^1(V)$ which is dual to some basis $\{v_1, \ldots, v_n\}$ of $V$. Hence

$$\left( \lambda^1 \wedge \ldots \wedge \lambda^k \right) \left( v_1 \otimes \ldots \otimes v_k \right) = \lambda^1(v_1) \cdot \ldots \lambda^k(v_k) = 1 \neq 0, \quad (148)$$

i.e. $\lambda^1 \wedge \ldots \wedge \lambda^k$ is a nonzero element of $\Lambda^k(V)$. Conversely, suppose that w.l.o.g. $\lambda^1$ is a linear combination of $\lambda^2, \ldots, \lambda^k$, i.e. there are $c_j$ such that $\lambda^1 = \sum_{j=1}^k c_j \lambda^j$. But then

$$\lambda^1 \wedge \ldots \wedge \lambda^k = \left( \sum_{j=1}^k c_j \lambda^j \right) \wedge \lambda^2 \wedge \ldots \wedge \lambda^k = \sum_{j=1}^k c_j \lambda^j \wedge \left( \lambda^2 \wedge \ldots \wedge \lambda^k \right) = 0. \quad \Box \quad (149)$$
3.14 Exterior derivatives and integrability

This section provides an introduction to the algebra differential $k$-forms on a smooth manifold, and to the exterior derivative $d$. The cap-stone of this section is an alternative way of characterizing integrability, in terms of differential forms as opposed to tangent vector fields).

The first step is to take the notions of tensor and exterior products, developed in the previous section for a single vector space $V$, to the setting of smooth manifolds. The construction of the associated vector bundles is completely analogous to that of the cotangent bundle: But this time replace the fibres $T_pM$ of the tangent bundle by tensor products $T_p^kM$, or by symmetrized tensor products, or alternating products. The following chapter utilizes symmetric tensor fields of type $(0, 2)$, this section is concerned with alternating tensors. We shall slightly abuse notation and write $Λ^k(M)$ for the vector bundle whose fibres over $p \in M$ are the spaces $Λ^k(T_pM)$, $k \geq 0$ (generally it will be clear from the context what $Λ^k$ stands for – with possibly minor confusion arising when the manifold $M$ is a vector space itself ...). Moreover it will be convenient to also introduce the vector bundle $Λ(M)$ whose fibres are the direct sums

$$Λ^0(T_pM) \oplus Λ^1(T_pM) \oplus \ldots \oplus Λ^m(T_pM)$$

(150)

There are no problems with these bundle constructions, and we proceed to focus our interest on functions that take values in these bundles:

**Definition 3.30** Suppose that $M^n$ is a smooth manifold. A differential $k$-form is a smooth section $ω: M \mapsto Λ^k(M)$. The space of all differential $k$-forms on $M$ is denoted by $Ω^k(M)$.

This means that $π \circ ω = id_M$ and for all $X_1, \ldots, X_k \in Γ^∞(M)$ the map $p \mapsto ω_p(X_1_p, \ldots X_k_p)$ is a smooth function on $M$. [[Be warned that the notation and terminology vary substantially among major texts. For example, Spivak uses $Ω^k$ to denote the space that are $Λ^k$ in our notation but does not introduce any special symbol for the space of to differential $k$-forms.]]

Again it is convenient to allow formal sums of differential forms of different degrees and we write

$$Ω = Ω^0(M) \oplus Ω^1(M) \oplus \ldots \oplus Ω^m(M)$$

(151)

for the space of all such formal sums. One naturally defines sums and exterior products of differential forms pointwise, e.g. $(ω + η)(p) = ω_p + η_p$ and $(ω \wedge η)(p) = ω_p \wedge η_p$.

**Exercise 3.53** Suppose $Φ ∈ C^∞(M, N)$ and $η ∈ Ω^k(N)$. Verify that $Φ^∗η$ defined pointwise by $(Φ^∗η)(p) = Φ^*_p(η_{Φ(p)})$ is a (smooth) differential $k$-form. Compare proposition 3.37 for the definition of $Φ^*_p(η_{Φ(p)})$, i.e., use $(Φ^*_p(η_{Φ(p)}))(X_1_p, \ldots, X_k_p) = η(Φ_pX_1_p, \ldots, Φ_pX_k_p))$.

With these definitions one easily verifies the following:

**Proposition 3.41** Suppose $ω, ω' \in Ω^k(N), η \in Ω^ℓ(N)$, $f \in C^∞(N)$, and $Φ \in C^∞(M, N)$. Then

\[\begin{align*}
(ω + ω') \wedge η &= ω \wedge η + ω' \wedge η \\
ω \wedge η &= (-1)^{kℓ} η \wedge ω \\
(fω) \wedge η &= f(ω \wedge η) = ω \wedge (fη) \\
Φ^∗(ω \wedge η) &= (Φ^∗ω) \wedge (Φ^∗η)
\end{align*}\]

(152)
Recall from 3.10 that locally, at every point \( p \) in a chart \((u, U)\), a basis for the cotangent space is given by \( \{du_1^q, \ldots, du^n_q\} \). Consequently the set
\[
\{du_1^q \wedge du_2^q \wedge \ldots \wedge du^n_q : 1 \leq i_1 < i_2 < \ldots < i_k \leq n \}
\] (153)
spans the vector space \( \Lambda^k(T_q U) \) for each \( q \in U \). It is convenient to introduce the multi-index \( I = (i_1, i_2, \ldots, i_k) \) and write \( du^I \) for the differential \( k \)-form \( du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k} \) which is defined for \( q \in U \) by \( du^I(q) = du_1^{i_1} \wedge du_2^{i_2} \wedge \ldots \wedge du^n_q \).

Consequently, for every \( \omega \in \Omega^k(M) \) there exist (in a chart \((u, U)\)) smooth functions \( \omega_1 \in C^\infty(U) \) such that \( \omega = \sum_I \omega_1 du^I \) where the sum is again taken over all multi-indices \( I = (i_1, i_2, \ldots, i_k) \) with \( 1 \leq i_1 < i_2 < \ldots < i_k \leq n \).

**Exercise 3.54** Suppose that \( \omega \in \Omega^k(M) \) and \((u, U), (v, V)\) are charts about \( p \). Let \( \omega_1, \omega'_1 \) be smooth functions on \( U \cap V \) such that \( \omega = \sum_I \omega_1 du^I = \sum_I \omega'_1 dv^I \). Calculate a formula for \( \omega'_1 \) in terms of \( \omega_1 \) and the partial derivatives \( \frac{\partial}{\partial u^i} \).

An important special case is when \( k = m \) which determines the **volume form** on the manifold:

**Proposition 3.42** Suppose \((u, U)\) and \((v, V)\) are charts for \( M^m \). Then
\[
dv^1 \wedge dv^2 \wedge \ldots \wedge dv^m = \det \left( \frac{\partial}{\partial u^j} \right)_{j=1}^m \ dv^1 \wedge dv^2 \wedge \ldots \wedge dv^m \quad (154)
\]

**Proof.** Suppose \((u, U)\) and \((v, V)\) are charts for \( M^m \) and \( p \in U \cap V \neq \emptyset \). Use the duality \( du^i(\frac{\partial}{\partial u^j}) = \delta^i_j \) and \( dv^i(\frac{\partial}{\partial v^j}) = \delta^i_j \), and the transformation formula \( \frac{\partial}{\partial v^j} = \sum_j \frac{\partial u^i}{\partial v^j} \frac{\partial}{\partial u^i} \) to calculate
\[
dv^1 \wedge dv^2 \wedge \ldots \wedge dv^m \left( \frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2}, \ldots, \frac{\partial}{\partial u^m} \right) = \\
= dv^1 \wedge dv^2 \wedge \ldots \wedge dv^m \left( \sum_{j=1}^m \frac{\partial u^1}{\partial v^j}, \sum_{j=1}^m \frac{\partial u^2}{\partial v^j}, \ldots, \sum_{j=1}^m \frac{\partial u^m}{\partial v^j} \right) \\
= \sum_{i_1, \ldots, i_m=1}^m \frac{\partial u^{i_1}}{\partial v^1} \frac{\partial u^{i_2}}{\partial v^2} \ldots \frac{\partial u^{i_m}}{\partial v^m} dv^1 \wedge dv^2 \wedge \ldots \wedge dv^m \left( \frac{\partial}{\partial v^{i_1}}, \frac{\partial}{\partial v^{i_2}}, \ldots, \frac{\partial}{\partial v^{i_m}} \right) \\
= \sum_{\sigma \in \Sigma_m} \text{sgn}(\sigma) \frac{\partial u^{\sigma(1)}}{\partial v^1} \frac{\partial u^{\sigma(2)}}{\partial v^2} \ldots \frac{\partial u^{\sigma(m)}}{\partial v^m}
\]
which is equal to the determinant of the Jacobian matrix of partial derivatives corresponding to the coordinate change. Note that in the key step all summands over multi-indices \( I = (i_1, i_2, \ldots, i_m) \) that are not permutations vanish due to the alternating nature of the exterior product. ■

**Exercise 3.55** Consider \( M = \mathbb{R}^2 \setminus \{(0, 0)\} \) with the standard Cartesian coordinates \((x, y)\) and the polar coordinates \((r, \theta)\) defined on a suitable subset. Express \( dx \wedge dy = \frac{1}{2}(dx \otimes dy - dy \otimes dx) \) via direct calculation in terms of \( dr \wedge d\theta \). [The result is known from calculus – but it is instructive to carry out all steps of the calculation starting with \( dx = \cos \theta \, dr - r \sin \theta \, d\theta \) etc.]

Just as in the case of differential one forms – sections of the cotangent bundle – differential \( k \)-forms are predestined to be **pulled back** along a smooth map between manifolds, compare exercise 3.53 and proposition 3.53.
Exercise 3.56 Suppose $\Phi \in C^\infty(M^m,N^n)$, $p \in M$ and $(u,U)$ and $(v,V)$ are charts about $p \in M$ and $\Phi(p) \in N$, respectively. Use the familiar formulas $\Phi_*(\frac{\partial}{\partial u^i}|_p) = \sum_{j=1}^m \frac{\partial(\omega^j \circ \Phi)}{\partial u^i}|_p \cdot \frac{\partial}{\partial v^j}|_{\Phi(p)}$ and thus $(\Phi^*dv^i)_p = \sum_{j=1}^m \frac{\partial(\omega^j \circ \Phi)}{\partial u^i}|_p \cdot du^i_p$ to find formulas for the pullbacks

$$\Phi^*(dv^i \otimes dv^{i_2} \otimes \ldots \otimes dv^{i_k}) \quad \text{and} \quad \Phi^*(dv^{i_1} \wedge dv^{i_2} \wedge \ldots \wedge dv^{i_k}) \quad (155)$$

in terms of $du^{i_1} \otimes du^{i_2} \otimes \ldots \otimes du^{i_k}$, and $du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k}$, respectively.

One of the most often used versions of pullbacks are those induced by inclusion maps or imbeddings. Specifically, if $M \subseteq N$ is a submanifold and $\iota: M \rightarrow N$ is the inclusion map, then every differential form $\omega \in \Omega(M)$ immediately gives rise to a differential form $\iota^*\omega \in \Omega(M)$. This is so commonly used, and so immediate that in many places one simply uses the same symbol $\omega$ for the pull-back $\iota^*\omega$. For example if $X \in \Gamma^\infty(M)$ is a vector field on $M$ and $\omega \in \Omega(N)$ then one often sees $\omega(X)$ for what should have been written more precisely as $(\iota^*\omega)(X).

Exercise 3.57 Consider the inclusion map $\iota: S^2 \hookrightarrow \mathbb{R}^3$ of the (imbedded) 2-sphere into 3 dimensional Euclidean space, equipped with the spherical coordinates $(\theta, \phi)$ and the Cartesian coordinates $(x^1, x^2, x^3)$. Calculate the pullbacks $\iota^*(dx^j) = d(x^j \circ \iota)$ for $j = 1, 2, 3$. [[Note, one often sloppily writes $dx^j$ when really referring to $\iota^*(dx^j).$]] Express $d\theta \wedge d\phi$ in terms of $\iota^*(dx^j \wedge dx^\ell)$ with $1 \leq j < \ell \leq 3$.

Arguably the most important map associated to differential forms is the exterior derivative which maps differential $k$-forms to differential $(k+1)$-forms. There are many intriguing relations between exterior derivatives of differential forms and Lie brackets of (tangent) vector fields – but the superior algebraic properties of the exterior derivative are hard to overestimate. Loosely following Spivak, we discuss three different ways to introduce exterior derivatives.

The exterior derivative of a differential zero-form, i.e. of a smooth function should just be the usual differential. In particular, in a chart $(u,U)$ we want

$$df = \sum_{j=1}^m \frac{\partial f}{\partial u^j} \cdot du^j \quad (156)$$

From here it is an obvious road to define for any differential $k$-form $\omega$, in a chart $(u,U)$

$$\text{If} \quad \omega = \sum_I \omega_I \cdot du^I \quad \text{then} \quad d\omega = \sum_I \sum_{j=1}^m \frac{\partial \omega_I}{\partial u^j} \cdot du^j \wedge du^I \quad (157)$$

(where $I = (i_1, i_2, \ldots, i_k)$ is any multi-index with $1 \leq i_j \leq n$. Such a definition has the obvious disadvantage that it requires an uninviting calculation to check that the definition is independent of the choice of local coordinates.

Exercise 3.58 Suppose $\omega$ is a differential form and $(u,U), (v,V)$ are local coordinate charts. Suppose that $\omega = \sum_I \omega_I \cdot du^I = \sum_J \omega'_J \cdot dv^J$. Verify directly that

$$\sum_I \sum_{j=1}^m \frac{\partial \omega_I}{\partial u^j} \cdot du^j \wedge du^I = \sum_J \sum_{j=1}^m \frac{\partial \omega'_J}{\partial v^j} \cdot dv^j \wedge dv^J \quad (158)$$
More elegant is an axiomatic characterization of the exterior derivative. The most interesting question is whether there is a compelling very short list of properties that uniquely characterizes the desired exterior derivative. Following Spivak:

**Proposition 3.43** Suppose $M$ is a smooth manifold. Then there is a unique map $d: \Omega^k(M) \rightarrow \Omega^{k+1}(M)$ that has the following properties:

(i) If $\omega, \eta \in \Omega(M)$ then $d(\omega + \eta) = d\omega + d\eta$.
(ii) If $\omega \in \Omega^k(M)$, $\eta \in \Omega(M)$ then $d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-)^k \omega \wedge (d\eta)$.
(iii) $d^2 = 0$.
(iv) If $(u, U)$ is a chart and $f \in C^\infty(U) = \Omega^0(U)$ then $df = \sum_{j=1}^m \frac{\partial f}{\partial u^j} \cdot du^j$.

**Exercise 3.59** Prove proposition 3.43. [Suppose there are maps $d, d'$ which both satisfy all properties as above. By the first property (additivity) it suffices to consider terms of the form $\omega_i dx^i$. Then compare $d\omega = d'\omega$, using the second and third properties. Finally show by induction that $d$ and $d'$ agree on all differential $k$-forms.]

The third alternative is to provide an invariant definition.

**Proposition 3.44** Suppose $\omega \in \Omega^k(M)$. There exists a unique differential $(k+1)$-form denoted $d\omega$ that satisfies for all sequences $X_1, X_2, \ldots X_{k+1} \in \Gamma^\infty(M)$ of smooth vector fields.

$$d\omega(X_1, X_2, \ldots X_{k+1}) = \sum_{i=1}^{k+1} X_i (\omega(X_1, \ldots X_{i-1}, \hat{X}_i \ldots X_{k+1})) +$$

$$+ \sum_{1 \leq i < j \leq k+1} (-)^{i+j} (\omega([X_i, X_j], X_1, \ldots X_{i-1}, X_{i+1} \ldots X_{j-1}, X_j \ldots X_{k+1})). \quad (159)$$

**Exercise 3.60** Prove proposition 3.44 (existence and uniqueness). [The key if to observe that if $\omega \in \Omega^k(M)$, $X_i \in \Gamma^\infty(M)$, and $f \in C^\infty(M)$, then]

$$d\omega(X_1, \ldots X_{i-1}, fX_i, X_{i+1} \ldots X_k) = f \cdot d\omega(X_1, \ldots X_{i-1}, X_i, X_{i+1} \ldots X_k) \quad \text{[].} \quad (160)$$

**Definition 3.31** Define $d: \omega(M) \rightarrow \Omega(M)$ to be the unique linear map that is characterized in proposition 3.44.

**Exercise 3.61** Suppose $d$ is defined as above. Verify that it satisfies the characterization (157).

**Exercise 3.62** Suppose $d$ is defined as above. Verify that it has the properties itemized in 3.43.

In the special case – the most important one – of $k = 1$ one has:

**Corollary 3.45** Suppose $X, Y \in \Gamma^\infty(M)$ and $\omega \in \Gamma^\infty(M)$. Then

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]) \quad (161)$$

**Proposition 3.46** Suppose $\Phi \in C^\infty(M, N)$ and $\omega \in \Omega(N)$. Then $\Phi^*(d\omega) = d(\Phi^*\omega) \in \Omega(M)$.

**Exercise 3.63** Prove proposition 3.46. [[Suggestion: One approach is to work in a chart and consider forms of the form $f du^i$ and proceed by induction on the length of $I$. Alternatively, use the invariant definition 3.31.]]
Finally we return to Frobenius integrability theorem. The invariant definition 3.31 of the exterior derivative $d$, and even better corollary 3.45 demonstrate the relationship between the Lie bracket of vector fields and the exterior derivative of differential forms. To utilize this relation it remains to characterize distributions in terms of differential forms: At every point $p \in M$ a smooth distribution determines a subspace $\Delta(p) \subseteq T_p M$. Instead of characterizing this subspace by a set of tangent vectors that span it, one may describe it by the linear functionals $\lambda : T_p M \mapsto \mathbb{R}$ such that the restriction $\lambda|_{\Delta(p)}$ is identically equal to zero.

On the side note that one may associate to a smooth distribution $\Delta$ a co-distribution $\Delta^\perp$ which assigns to each $p$ the subspace (!) of all $\lambda \in T^*_p M$ that annihilate $\Delta$. One may again define smooth co-distributions as those co-distributions that are spanned by smooth differential one-forms. Note that while the rank of a smooth distribution (i.e. $\dim \Delta(p)$) may decrease at points (or more generally on closed subsets), the rank of a distribution that is defined as

$$\Delta(p) = \{ X_p \in T_p M : \omega_p(X_p) = 0 \text{ for all } \omega(p) \in \Delta^*(p) \}$$

(can only increase at points if $\Delta^*$ is a smooth co-distribution.

In addition to this linear structure of the set of annihilating differential one-forms, one may also consider the algebra structure of the annihilating differential $k$-forms. The exercise motivates the subsequent definition that associates with $\Delta$ an ideal which provides for particularly elegant statements of integrability conditions – analogous to the concept of involutivity.

**Exercise 3.64** Suppose $\Delta$ is a smooth distribution on $M$, $f \in C^\infty(M)$, and $\omega, \eta \in \Omega^1$ annihilate $\Delta$, i.e. $\omega_p(X_p) = \eta_p(X_p) = 0$ for all $X_p \in \Delta(p)$. Verify that

- $(f \omega + \eta)_p(X_p) = 0$ for all $X_p \in \Delta(p)$, and
- $(\omega \wedge \theta)_p(Y_p, X_p) = 0$, for all $X_p, Y_p \in \Delta(p)$ and all $\theta \in \Omega^1(M)$.

**Definition 3.32** Suppose that $\Delta$ is a distribution on $M$. Let $\mathcal{J}(\Delta) \subseteq \Omega(M)$ be the ideal that is generated by all differential forms $\omega \in \Omega(M)$ such that

$$\omega(X_1, \ldots, X_k) = 0 \text{ whenever } X_1, \ldots, X_k \text{ belong to } \Delta$$

**Theorem 3.47** (Frobenius) Suppose $\Delta$ is a smooth distribution and $\mathcal{J}(\Delta)$ is the subring (actually, by exercise 3.64 $\mathcal{J}(\Delta)$ is an ideal) of $\Omega(M)$ generated by all smooth differential forms $\omega$ such that $\omega(X_1, \ldots, X_k) = 0$ for all smooth vector fields $X_j \in \Gamma^\infty(M)$ that belong to $\Delta$.

Then $\Delta$ is integrable if and only if $d(\mathcal{J}(\Delta)) \subseteq \mathcal{J}(\Delta)$.

**Corollary 3.48** Suppose $\Delta$ is a smooth distribution on $M$, $p \in M$, and $\omega^{k+1}, \ldots, \omega^m \in \Omega^1$ generate $\mathcal{J}(\Delta)$ on a neighborhood of $p$. Then $\Delta$ is locally integrable around $p$ if and only if there exist $\theta^i_k \in \Omega^1$ such that

$$d\omega^i = \sum_j \theta^i_j \wedge \omega^j \text{ for all } i = k + 1, \ldots, m.$$  

**Proof.** (of Frobenius’ theorem). Clearly $\mathcal{J}(\Delta)$ is locally generated by smooth one forms. Thus suppose $p \in M$ and $\omega^{k+1}, \ldots, \omega^m \in \Omega^1$ generate $\mathcal{J}(\Delta)$ on some neighborhood $U$ of $p$. Extend to a co-frame $\omega^1, \ldots, \omega^k, \omega^{k+1}, \ldots, \omega^m \in \Omega^1$, i.e. such that $\{\omega^1_q, \ldots, \omega^m_q\}$ is a basis for each $q \in U$. There exist smooth vector fields $X_1, \ldots, X_m \in \Gamma^\infty(M)$ such that $\omega^i(X_j) = \delta^i_j$ on $U$ for all $i, j \leq m$. Clearly the vector fields $X_1, \ldots, X_k$ span $\Delta$ on $U$. 

Recall the earlier version of Frobenius theorem which said that $\Delta$ is integrable if and only if there are functions $c^r_{ij} \in C^\infty(M)$ such that $[X_i, X_j] = \sum_{\ell=1}^m c^r_{ij} X_\ell$. Calculate

$$d\omega^r(X_i, X_j) = X_i(\omega^r(X_j)) - X_j(\omega^r(X_i)) - \omega^r([X_i, X_j])$$

(165)

Thus if $i, j \leq k$ and $r > k$ then the first two terms on the right hand side vanish. If $\Delta$ is integrable then the last term may be rewritten as

$$-\sum_{\ell=1}^k c^r_{ij} \omega^r X_\ell = 0$$

(166)

and it is zero since $\omega^r$ annihilates $\Delta$.

Conversely, if $d\omega^r(X_i, X_j) = 0$ for all $r > k$ then the third term on the right hand side must be zero and one concludes that $[X_i, X_j]$ belongs to $\Delta$, i.e. $[X_i, X_j]$ may be rewritten as a (smooth) linear combination of $X_1, \ldots, X_\ell$ and by Frobenius theorem, first version, $\Delta$ is integrable. ■

Frobenius integrability theorem clearly relies in an essential way on the key property (iii) of proposition 3.43 that $d^2 = 0$ - which may be considered an elegant invariant way to restate that mixed partial derivatives (of smooth functions) are equal. Since $d\omega = 0$ if $\omega = d\eta$ for some $\eta \in \Omega(M)$, one naturally may ask whether $d\omega = 0$ implies that there exists $\eta \in \Omega(M)$ such that $\omega = d\eta$.

**Definition 3.33** A differential form $\omega \in \Omega(M)$ is called closed if $d\omega = 0$ and is called exact if there exists $\eta \in \Omega$ such that $\omega = d\eta$.

In this language the above may be restated as every exact form is closed and the question is whether a closed form is necessarily exact. In general the answer is negative:

**Exercise 3.65** Consider $\omega = \frac{1}{x^2+y^2} (-y \, dx + x \, dy) \in \Omega^1(M)$ where $M = \mathbb{R}^2 \setminus \{(0,0)\}$. Verify that $d\omega = 0$, and that on any set $U_\alpha = \{(r \cos \theta, r \sin \theta): r > 0, \alpha - \pi < \theta \leq \alpha + \pi\}$ there is a function $\Theta \in \Omega^0(U_\alpha)$ such that $\omega = d\Theta$. Show that, on the other hand, there does not exist any (continuous !) $f \in \Omega^0(M)$ such that $\omega = df$.

As this exercise suggests the question which closed forms are also exact is intimately linked to the topology of the manifold. In the example in the exercise the nonexactness is traced back to the manifold not being simply connected (or not being smoothly contractible to a point. [The further study of this question leads one to generalize line integrals, Stokes’ theorem etc. to differentiable manifolds. Eventually one is led to define an equivalence relation by $\omega \sim \eta$ if $\eta - \omega$ is closed. The deRham cohomology formalizes the algebraic properties of the associated equivalence classes, and connects them to global topological properties such as “the number of holes” in the manifold (using colloquial language.)]

**Exercise 3.66** Explore the statement that curl (grad $f$) = 0 and div (curl $F$) = 0 are just special cases of the general property $d^2 = 0$. [For a function $f \in C^\infty(\mathbb{R}^3)$ it is easy to relate $df$ and the gradient of $f$ (are they equal, or is the gradient a tangent vector field?). In the case of a vector field $F(x) = (F_1(x), F_2(x), F_3(x))$ discuss different possible points of view, e.g. considering $\tilde{F} = F_1 dx^1 + F_2 dx^2 + F_3 dx^3$ and $\tilde{F} = F_1 dx^2 \wedge dx^3 + F_2 dx^3 \wedge dx^1 + F_3 dx^1 \wedge dx^2$. Calculate and discuss $d\tilde{F}$ and $\tilde{dF}$. In this context discuss the domain and range of such (questionable?) compositions as curl (curl $F$) and $\Delta f = \text{div} \, (\text{grad} \, f)$. – Note that a complete discussion of these relationships needs to take into account the Riemannian structure of $\mathbb{R}^3$ and thus is a natural topic of the next chapter.]
in progress

Application to / interpretation in terms of partial differential equations

volume form, orientation, and integration

in progress
4 Riemannian geometry

4.1 Introduction

A natural first step towards a general concept of curvature is to develop a general notion of distance on a manifold. (Recall, originally we required a manifold to be a metric space – but this metric topology really only served to rule out pathological topologies. What was used was \textit{metrizability} or \textit{paracompactness} - the metric itself was rather unimportant.) The concept of a Riemannian metric has proved to be very successful and led to a very rich body of structures and results, with many applications, most notably Hamiltonian mechanics. \[[ We note on the side that on a more advanced level one may work with \textit{connections} without first developing any Riemannian metric and develop notions of curvature and even of geodesics exclusively in terms of connections. On the other hand, Riemannian metrics are just a special case of more general metrics. The first generalization is to replace the \textit{quadratic} formulas by more general expressions, leading to Finsler metrics, and one may go well beyond. At this time it is only important to not consider Riemannian metrics as the ultimate notion, but rather a natural first step. \]]

The key idea is to employ the local (or rather infinitesimal) resemblance of a manifold to a Euclidean space to endow the manifold with an analogue of the notion of the Euclidean distance. A basic observation is that the notion of the usual distance in Euclidean spaces is inextricably intertwined with that of an inner product (see the next section for technical details)

\[
d_2(p,q) = (q - p, q - p)^{1/2} = \left( \sum_{k=1}^{n} |q^k - p^k|^2 \right)^{1/2}
\]

(167)

From here it is just a small step to recognize that the tangent spaces \(T_pM\), as linear spaces, are the natural place to put inner products. The physical interpretation is that such inner products associate a notion of speed (i.e. a magnitude) to every velocity vector. Given such notion of speed, define (!) the (arc-)length of a (smooth) curve as the usual integral of the speed. Note, that this is fundamentally different from the notion of arc-length of curves \(\gamma : [a,b] \mapsto \mathbb{R}^n\) in Euclidean spaces which is defined as the supremum

\[
L(\gamma) = \sup_{\mathcal{P}} \|\gamma(t_{i+1}) - \gamma(t_i)\|
\]

(168)

over all partitions \(\mathcal{P} = \{a = t_0 < t_1 \leq \ldots t_n = b\}\). In that setting the calculus formula for the length of sufficiently smooth curves

\[
L(\gamma) = \int_{a}^{b} \|\gamma'(t)\| \, dt
\]

(169)

is simply a formula to calculate the length defined before. In the setting of differentiable manifolds the definition (168) makes no sense, but one may now take the obvious analogue of (169) as the \textit{definition} of the length – see the next sections for the technical details.

Inner products provide a notion of angle between any two vectors \(v, w \in V\), defined as

\[
\angle(v, w) = \cos^{-1} \left( \frac{\langle v, w \rangle}{\langle v, v \rangle^{1/2} \langle w, w \rangle^{1/2}} \right)
\]

(170)
From here it is a stroll to explore different avenues to develop notions of curvature of manifolds. Clearly, one of the most obvious routes considers the sum of the angles in a triangle with the analogue of straight sides. The excess (or defect) of the angle sum (when compared to the constant sum $\pi$ for the flat Euclidean setting) provides a natural measure of being curved (after suitable rescaling by the size of the geodesic triangle). Instrumental for this approach is the notion of geodesics, which, roughly speaking, as the “straightest” curves, and also, in some sense, the shortest curves between pairs of points on a manifold provide natural analogues to straight lines. There are many technical details that need to be resolved, but many address properties that are intuitive and familiar from playing with e.g. the embedded 2-sphere and other embedded 2-surfaces in $\mathbb{R}^3$.

An central question is which inner product to put on each tangent space – after all there is a continuum of inner products on every vector space. It is clear that we want to have inner products that are in a natural sense smoothly varying from point to point – here the language of sections of vector bundles is a particularly elegant way to make this precise. The most exciting innovations – completely trivial notions in Euclidean spaces – concern the problem of comparing (the inner products on) different tangent spaces with each other. This will lead to new notions of derivatives (connections) and associated notions of parallel transport. There will be some similarity to Lie derivatives with respect to a vector field, and the mapping of tangent spaces induced by the flow of the vector field. But in the end these derivatives will turn out to be fundamentally different from Lie derivatives. In particular, they yield a crisp formalization of the naive concept of curvature. One question is: Which comes first, a notion of derivative or a notion of parallel transport.

This still leaves too much choice. Consider a coordinate chart $(u,U)$ about a point $p \in M$. It is rather straightforward to pull back the standard inner product $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_u(q)$ from $u(U) \subseteq \mathbb{R}^m$ to provide an inner product on $T_qM$ for any $q \in U$ – but there is no natural way to select one preferred coordinate chart about $p$. Indeed, recall that for any two inner products on a Euclidean space $\mathbb{R}^n$ each may be obtained from the other via a linear transformation or a basis change (Declaring a basis to be orthonormal effectively defines an inner product.)

This simple observation makes it clear that the Riemannian metric must come from the outside – a differentiable manifold has to be given a metric, or, generally, a differentiable manifold does not have its own natural Riemannian metric that only needs to be uncovered. While developing the tools and intuition one rightfully looks at embedded surfaces in Euclidean space for guidance – technically this means that the manifold inherits an induced metric as the pullback of the Euclidean metric via the inclusion map (imbedding).

Among practical applications one of the most tangible ways that Riemannian metrics arise is as inertia tensors in (Hamiltonian and Lagrangian) mechanics. Recall that the kinetic energy of a collection of point masses and/or rigid bodies consists of terms that are quadratic in the (angular) velocities and multiplied by terms that depend on the mass and shape (or mass distribution).

For example consider a double pendulum with masses $m_1$ and $m_2$ and lengths $\ell_1$ and $\ell_2$. Let $\theta$ and $\phi$ denote the angles with the vertical. Then the total kinetic energy is

$$E_{\text{kin}} = \frac{1}{2} \begin{pmatrix} \dot{\theta} \\ \dot{\phi} \end{pmatrix}^T \begin{pmatrix} m_1 \ell_1^2 + m_2 \ell_2^2 & m_2 \ell_1 \ell_2 \cos(\phi - \theta) \\ m_2 \ell_1 \ell_2 \cos(\phi - \theta) & m_2 \ell_2^2 \end{pmatrix} \begin{pmatrix} \dot{\theta} \\ \dot{\phi} \end{pmatrix} \tag{171}$$

If we allow full rotations (over the top . . . ), then the natural state space is the torus $(\theta, \phi) \in T^2 = S^1 \times S^1$ and the angular velocities $(\dot{\theta}, \dot{\phi}) \in T_{(\theta, \phi)}T^2$ are tangent vectors. For our purposes the general form of this expression is important: quadratic in the (angular) velocities and the
inner product depends smoothly on the state \((\theta, \phi) \in T^2\) of the system which is a point on the manifold \(T^2\).

A similar example describes the motion of a rigid body in the absence of external forces and torques. Its center of mass travels at constant speed along a straight line while its orientation (relative to a fixed frame) may at any time be completely described by a rotation matrix, i.e. a point \(R\) on the manifold \(SO(3)\). The angular velocity \(\omega\) is an element of the tangent space \(T_R SO(3)\) and the rotational kinetic energy takes (with slight abuse of notation (identifying the tangent spaces with \(\mathbb{R}^3\)) the form

\[
E_{\text{rot}} = \frac{1}{2} \omega^T \cdot J(R) \cdot \omega
\]  

(172)

where \(J(R)\) is a \(3 \times 3\) matrix whose entries are the moments of inertia about the chosen system of coordinate axes and coordinate planes. Again the kinetic energy is a quadratic form on the (angular) velocities with a matrix that is determined by the specific data of the system – here the distribution of the mass.

The interplay between Riemannian differential geometry and (Hamiltonian and Lagrangian) mechanics extends a lot further: Not only does the physics provide the metrics, but the geodesics (or shortest curves) with respect to this metric may also be interpreted as the curves that describe how the physical system evolves dynamically. In its most simplistic form – in the absence of any potential and external forces – the Lagrangian/Hamiltonian model says that the mechanical system will evolve in a way that minimizes the action integral, an integral defined in terms of the energy(ies) – this can be translated into minimizing distance in the setting of differential geometry. However, at this stage in this course, it would take us too long to work out all the details of systems that involve a nontrivial potential energy. We refer the reader to the original literature, with (Abraham and) Marsden’s text a good starting point, or see the recent (2003) text by Bloch et.al. In the modern treatment, further nonholonomic constraints (i.e. constraints on the allowable velocities only, but not on the state) are naturally addressed in the language of differential geometric control, with the Pontryagin maximum principle taking the place of classical calculus of variations.

What is important for our course is that the applications provide specific metrics, and they have use and interpretations of such quantities as geodesics, curvature and parallel transport. Moreover, from considering only the simple double pendulum it is clear that almost any positive definite (symmetric) matrix as above may arise from a real physical problem. This suggests that we should study all possible inner products, and create any examples as we please, not only those that arise from a-priori imbedded surfaces in Euclidean spaces. . . .
4.2 Review of inner product spaces

We begin with a brief review on the relation between metrics, norms and inner products. Recall that a metric on a set $Q$ is a function $d: Q \times Q \rightarrow \mathbb{R}$ such that for all $p, q, z \in Q$

- $d(p, q) \geq 0$
- $d(p, q) = 0 \iff p = q$
- $d(p, q) + d(q, z) \geq d(p, z)$

On a vector space $V$ one naturally has the stronger notion of a norm. This is a function $\|\cdot\|: V \times V \rightarrow \mathbb{R}$ that for all $v, w \in V$ and for all $c \in \mathbb{R}$ satisfies

- $\|v\| \geq 0$ and $\|v\| = 0$ only if $v = 0$
- $\|cv\| = |c| \cdot \|v\|$, and
- $\|v + w\| \leq \|v\| + \|w\|$.

Exercise 4.1

Verify that every norm on a vector space $V$ gives rise to a metric by $d(v, w) = \|w - v\|$. Give an example of a metric on a vector space that does not arise in this way from any norm.

From now on, suppose $V$ is a finite dimensional real vector space (later typically either the tangent or the cotangent space at a point on a manifold). An inner product on $V$ is a function $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ that satisfies all $v, v', w \in V$ and all $\lambda \in \mathbb{R}$

- $\langle v, v \rangle \geq 0$, and $\langle v, v \rangle = 0$ iff $v = 0$ (positive definiteness)
- $\langle v, w \rangle = \langle w, v \rangle$ (symmetry)
- $\langle \lambda v + v', w \rangle = \lambda \langle v, w \rangle + \langle v', w \rangle$ (together with symmetry this implies bilinearity)

In complex vector spaces, symmetry is replaced by the requirement of conjugate symmetry, i.e., for all $v, w \in V$, $\langle w, v \rangle = \overline{\langle v, w \rangle}$, and consequently, for all $v, w \in V$, $\lambda \in \mathbb{C}$, $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle = \langle v, \overline{\lambda} w \rangle$. This quadratic form is no longer bilinear. (Mathematicians generally prefer to require linearity in the first argument, while physicists prefer linearity in the second argument.) For applications such as relativity one replaces the requirement of positive definiteness by the condition

- For all $v \in V$, if for all $w \in V$, $\langle v, w \rangle = 0$ then $v = 0$ (nondegeneracy).

In the sequel we shall always work with real inner products and, in particular, assume positive definiteness – although many of the following constructions can be made also for indefinite (nondegenerate) bilinear forms.

As verified in the exercise, every positive definite inner product gives rise to a norm.

Exercise 4.2 Suppose that $\langle \cdot, \cdot \rangle$ is an inner product on a vector space $V$. Show that the function $\|\cdot\|: V \times V \rightarrow \mathbb{R}$ defined by $\|v\| = (\langle v, v \rangle)^{1/2}$ is a norm on $V$, and prove the Cauchy-Schwarz inequality, i.e., for all $v, w \in V$, $\|\langle v, w \rangle\| \leq \|v\| \cdot \|w\|$.

On the other hand there are norms that do not come from any inner product in the way described in the exercise. The parallelogram law provides a necessary and sufficient condition:

Proposition 4.1 If $\langle \cdot, \cdot \rangle$ is an inner product on a vector space $V$, then for all $v, w \in V$

$$\langle v + w, v + w \rangle + \langle v - w, v - w \rangle = 2 \langle v, v \rangle + 2 \langle w, w \rangle$$ (173)
Proposition 4.2 Suppose that \( d \) is a norm on a finite dimensional vector space \( V \) that satisfies the parallelogram law, i.e. for all \( v, w \in V \)
\[
\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2.
\]
Then there exists an inner product \( \langle \cdot, \cdot \rangle \) on \( V \) such that, for all \( v \in V \), \( \|v\| = \langle v, v \rangle^{1/2} \).

Proof. Use the polarization identity to define a function \( Q: V \times V \to \mathbb{R} \) by
\[
Q(v, w) = \frac{1}{4} (\|v + w\|^2 - \|v - w\|^2)
\]
and verify that this function is indeed an inner product.

Exercise 4.3 Verify that neither the taxi-cab norm \( \|v\|_1 = \sum_{j=1}^{n} |v_j| \) nor the sup-norm \( \|v\|_\infty = \max_{j=1}^{n} |v_j| \) is the norm associated to any inner product on \( \mathbb{R}^n \).

In addition to length (or norm), every inner product also provides a notion of angle between any \( v, w \in V \). The Cauchy-Schwarz inequality, exercise 4.2, guarantees that the following is real
\[
\angle(v, w) = \cos^{-1} \left( \frac{\langle v, w \rangle}{\|v\| \cdot \|w\|} \right)
\]
A set of vectors \( \{v_1, \ldots, v_k\} \subseteq V \) in an inner product space \( V \) is called orthogonal if \( \langle v_i, v_j \rangle = 0 \) whenever \( i \neq j \). If in addition, for all \( i \leq k \), \( \langle v_i, v_i \rangle = 1 \), then the set is called orthonormal.

Using the language of tensors identify an inner product on a vector space with a symmetric tensor of type \((0,2)\), i.e. \( \langle \cdot, \cdot \rangle \in T_0^2(V) \). Consequently, if \( \Phi: V \to W \) is a linear map between vector spaces, and \( \langle \cdot, \cdot \rangle_W \) is an inner product on \( W \), then \( \Phi^*(\langle \cdot, \cdot \rangle) \) defines a symmetric bilinear form on \( V \). However, in general this form may fail to be nondegenerate. As an example consider \( \mathbb{R}^2 \) with the nondegenerate bilinear form \( \langle v, w \rangle = v_1w_1 - v_2w_2 \) and the map \( \Phi: \mathbb{R} \to \mathbb{R}^2 \) defined by \( \Phi(v) = (v, v) \).

Exercise 4.4 Suppose that \( \Phi: V \to W \) is a bijective linear map between vector spaces and \( \langle \cdot, \cdot \rangle_W \) is an inner product on \( W \). Verify that \( \Phi^*(\langle \cdot, \cdot \rangle) \) is an inner product on \( V \).

Exercise 4.5 Suppose that \( \Phi: V \to W \) is a linear map between vector spaces and \( \langle \cdot, \cdot \rangle_W \) is an inner product on \( W \). Show that \( \Phi^*(\langle \cdot, \cdot \rangle) \) is an inner product on \( V \) if and only if \( \Phi \) is one-to-one.

An inner product on a finite dimensional vector space \( V \) gives rise to a natural isomorphism
\[
\alpha: V \to V^* \quad \text{defined by} \quad \alpha(v) = \langle v, \cdot \rangle
\]
Exercise 4.6 Show that the map \( \alpha \) defined above is bijective.

Using the natural isomorphism between \( V \) and \( V^{**} \) for finite dimensional spaces, one obtains:

Proposition 4.3 The inverse \( \alpha^{-1}: V^* \to V \) defines an inner product \( \langle \cdot, \cdot \rangle^* \) on \( V^* \) by
\[
\text{for } \lambda, \lambda' \in V^*, \quad \langle \lambda, \lambda' \rangle^* = (\lambda' \circ \alpha^{-1})(\lambda)
\]
(178)
Proof. The symmetry of $\langle \cdot , \cdot \rangle^*$ is an immediate consequence of the symmetry of $\langle \cdot , \cdot \rangle$.

From the definitions, for $\lambda, \lambda' \in V^*$
\[
\langle \lambda, \lambda' \rangle^* = \lambda'(\alpha^{-1}(\lambda)) = (\alpha(\alpha^{-1}(\lambda'))) (\alpha^{-1}(\lambda)) = \langle \alpha^{-1}(\lambda'), \alpha^{-1}(\lambda) \rangle
\] (179)

Bilinearity of $\langle \cdot , \cdot \rangle^*$ is clear from the symmetry and the fact that $\alpha^{-1}$ is the inverse of a bijective linear map. Nondegeneracy, and, in analogy, positive definiteness, follow from the respective properties of $\langle \cdot , \cdot \rangle$ since $\alpha^{-1}$ is one-to-one, again using (179).

Finally suppose that $\{v_1, \ldots, v_n\}$ is a basis for a vector space $V$ and $\langle \cdot , \cdot \rangle$ is an inner product on $V$. Let $\{\lambda^1, \ldots, \lambda^n\}$ denote the dual basis for $V$. Then there exist $g_{ij} \in \mathbb{R}$ such that
\[
\langle \cdot , \cdot \rangle = \sum_{i,j=1}^n g_{ij} \lambda^i \otimes \lambda^j.
\] (180)

These coefficients are determined as
\[
g_{ij} = \sum_k \sum_{\ell} g_{k\ell} \delta_{i}^{k} \delta_{j}^{\ell} = \sum_k g_{k\ell} \lambda^k \otimes \lambda^\ell(v_i, v_j) = \langle v_i, v_j \rangle.
\]

From a different point of view, the matrix $G$ with components $g_{ij}$ is also the matrix associated with the linear map $\alpha$ with respect to the bases $\{v_1, \ldots, v_n\}$ and $\{\lambda^1, \ldots, \lambda^n\}$, respectively:
\[
\alpha(v_i) = \langle v_i, \cdot \rangle = \sum_{j=1}^n g_{ij} \lambda^j \quad \text{(because } \langle v_i, v_k \rangle = g_{ik} = \sum_{j=1}^n g_{ij} \lambda^j(v_k) \text{).}
\] (181)

**Corollary 4.4** Suppose that $\langle \cdot , \cdot \rangle$ and $\langle \cdot , \cdot \rangle^*$ are inner products on $V$ and $V^*$ as above, and that $\{v_1, \ldots, v_n\}$ and $\{\lambda^1, \ldots, \lambda^n\}$ are dual bases for $V$ and $V^*$, respectively. If $g_{ij} = \langle v_i, v_j \rangle$ and $g_{ij}^* = \langle \lambda^i, \lambda^j \rangle^*$ then the matrices with components $(g_{ij})$ and $(g_{ij}^*)$ are inverses of each other, i.e.
\[
\sum_{k=1}^n g_{ik}^* g_{kj} = \delta_{ij}.
\] (182)

**Proof.** The matrix with components $g_{ij}$ is invertible since $\langle \cdot , \cdot \rangle$ is nondegenerate. Recall
\[
\alpha(v_i) = \langle v_i, \cdot \rangle = \sum_{j=1}^k g_{ij} \lambda^j \quad \text{and } \quad \langle \lambda^j, \cdot \rangle^* = \alpha^{-1}(\lambda^j)
\] (183)

Combine these definitions and use bilinearity to obtain
\[
\sum_{j=1}^n g_{ij} g_{jk}^* = \sum_{j=1}^n g_{ij} \lambda^j \lambda^k = \lambda^k \circ \alpha^{-1} \left( \sum_{j=1}^n g_{ijj} \lambda^j \right) = \lambda^k(v_i) = \delta_{ik}^*.
\] (184)

To complete this linear algebra review recall that symmetric matrices are orthogonally diagonalizable: If $G$ is an $n \times n$ matrix with $G^T = G$ then there exists a diagonal matrix $D$ and an orthogonal matrix $V$ such that $G = VDV^T$.

In the following sections it will be of importance that a set of vector fields may locally be *smoothly* orthonormalized. This will be obvious from the nature of the steps in the Gram-Schmidt orthonormalization algorithm which we here review for a single (finite dimensional) vector space.
Suppose that \( \{v_1, \ldots, v_n\} \) is a basis for a vector space \( V \) with an inner product \( \langle \cdot, \cdot \rangle \). Inductively construct a new basis by first setting \( u_1 = \frac{v_1}{\|v_1\|} \). For the induction step assume that an orthonormal set \( \{u_1, \ldots u_k\} \) has been constructed that spans the same subspace as \( \{v_1, \ldots v_k\} \), and define

\[
    w_{k+1} = v_{k+1} - \sum_{j=1}^{k} \langle v_{k+1}, u_j \rangle u_j, \quad \text{and} \quad u_{k+1} = \frac{w_{k+1}}{\|w_{k+1}\|}
\]

(185)

**Exercise 4.7** Verify that the set \( \{u_1, \ldots u_n\} \) is indeed an orthonormal basis for \( V \).

---

volume form and Hodge star here in "lin alg" review, or later on manifolds for diff forms?

---

in progress
4.3 Riemannian manifolds

As suggested in the introduction, a Riemannian manifold is simply a manifold with a smoothly varying inner product on each tangent space. We begin with some formal definitions:

**Definition 4.1** A Riemannian manifold is a differentiable (smooth) manifold together with a function $\langle \cdot, \cdot \rangle : M \mapsto T_0^2(M)$ called a Riemannian metric, that satisfies

- For every $p \in M$ the function $\langle \cdot, \cdot \rangle_p$ is a (positive definite) inner product on $T_pM$.
- If $X, Y \in \Gamma^\infty(M)$ are smooth vector fields, then the function $p \mapsto \langle X_p, Y_p \rangle_p$ is smooth.

The following exercise shows that every manifold can be endowed with some Riemannian metric - thus settling existence. In many applications the problem provides the metric (compare the discussion of problems from mechanics). Another important means of obtaining a Riemannian metric on a manifold is as the pullback from another Riemannian manifold. More specifically:

**Proposition 4.5** If $\Phi: M \mapsto N$ is an immersion (i.e. for all $p \in M^m$, $\Phi_\ast p$ has rank $m$), and $(N, \langle \cdot, \cdot \rangle^N)$ is a Riemannian manifold, then $\Phi_\ast(\langle \cdot, \cdot \rangle^N)$ is a Riemannian metric on $M$.

One of the most common applications of the proposition occurs when $M$ is an imbedded submanifold of a Riemannian manifold and $\Phi = i : M \mapsto N$ is the inclusion mapping. In this case one calls $i_\ast(\langle \cdot, \cdot \rangle^N)$ the induced, or inherited, (Riemannian) metric on $M$. Subsequent examples will explore these in detail for the case of surfaces in Euclidian 3-space.

**Proof.** It is clear that at every point $p \in M$ the pullback $\Phi_\ast(\langle \cdot, \cdot \rangle^N)_p$ yields a bilinear map $(\Phi_\ast(\langle \cdot, \cdot \rangle^N))_p : T_pM \times T_pM \mapsto \mathbb{R}$. Its positive definiteness is a consequence from $\text{rank} \Phi_\ast p = m$ (compare exercise 4.5 of the linear algebra review).

Regarding smoothness, suppose $X, Y \in \Gamma^\infty(M)$ are smooth vector fields on $M$. Since $\Phi$ is an immersion, one can find for any $p \in M$ neighborhoods $U$ of $p \in M$ and $V$ of $\Phi(p) \in N$ together with smooth vector fields $\tilde{X}$ and $\tilde{Y}$ defined on $V$ such that for every $q \in U$, $\Phi_\ast q X_q = \tilde{X}_{\Phi(q)}$ and $\Phi_\ast q Y_q = \tilde{Y}_{\Phi(q)}$. Denoting $(\Phi_\ast(\langle \cdot, \cdot \rangle^N))_p$ by $\langle \cdot, \cdot \rangle^M_p$ this yields that

$$p \mapsto \langle X_p, Y_p \rangle^M_p = \langle \Phi_\ast p X_p, \Phi_\ast p Y_p \rangle^N_{\Phi(p)} = \langle \tilde{X}_{\Phi(p)}, \tilde{Y}_{\Phi(p)} \rangle^N_{\Phi(p)}$$

is a smooth function. ■

**Exercise 4.8** Show that every manifold $M$ can be endowed with a Riemannian metric. Suggestion: Find a locally finite collection of charts $U = (u_\alpha, U_\alpha)$ that covers $M$. Then use a smooth partition of the unity subordinate to $U$ to paste together the locally defined pullbacks $u_\alpha_\ast(\langle \cdot, \cdot \rangle)$ of any positive definite metrics on Euclidean spaces. Verify that this construction indeed yields Riemannian metric!

**Definition 4.2** A diffeomorphism $\Phi : (M, \langle \cdot, \cdot \rangle^M) \mapsto (N, \langle \cdot, \cdot \rangle^N)$ between Riemannian manifolds is called an isometry if $\Phi_\ast(\langle \cdot, \cdot \rangle^N) = \langle \cdot, \cdot \rangle^M$.

If such an isometry exists, then the Riemannian manifolds $M$ and $N$ are called isometric.

In other words, the diffeomorphism $\Phi$ is an isometry if for all $p \in M$, and for all $X_p, Y_p \in T_pM$

$$\langle X_p, Y_p \rangle^M_p = \langle \Phi_\ast p X_p, \Phi_\ast p Y_p \rangle^N_{\Phi(p)}.$$  

**Definition 4.3** An submanifold $M$ of a Riemannian manifold $N$ is called a Riemannian submanifold if the inclusion map $i : M \mapsto N$ is an isometry.
Definition 4.4 If a Riemannian manifold $M$ can be covered with charts $(u, U)$ such that each $u: U \to \mathbb{R}^n$ is an isometry, then the manifold is called flat.

An important example is the flat torus, which is defined as the quotient of $\mathbb{R}^2$ under the equivalence relation $(x, y) \sim (x', y')$ if and only if $(x - x', y - y') \in \mathbb{Z}^2$, usually pictured as the unit square in the plane with edges appropriately identified. It is a straightforward exercise to define a suitable atlas and impose that the maps $u$ are isometries. On the other hand, as we will soon see (basically Riemann’s main theme), it is impossible to cover e.g. the usual embedded 2-sphere with charts such that each $u$ is an isometry – i.e. while there exists a flat torus, there does not exist a flat 2-sphere.

A question to ponder: Can you prove, with the tools developed so far that there does not exist an isometry between two imbedded spheres $S^n(r)$, $S^n(r') \subseteq \mathbb{R}^{n+1}$ with different radii $r' \neq r$?

In a coordinate chart $(u, U)$ the Riemannian metric determines smooth functions

$$g_{ij} = \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \rangle$$

which measure, in some sense, how far $u$ is from being an isometry between $U$ and $u(U) \subseteq \mathbb{R}^m$. Clearly, $u$ is an isometry if and only if $g_{ij} = \delta_{ij}$. In general, coordinate vector fields $\{\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^m}\}$ do not form an orthonormal $m$-frame, i.e., do not at each point $p \in U$ form an orthonormal basis for $T_pM$. On the other hand, locally one always can find orthonormal frames, that is, collections $\{X_1, \ldots, X_m\}$ of $m$ smooth vector fields such that $\{X_{ip}, \ldots X_{mp}\}$ forms an orthonormal basis for $T_pM$ for each $p$ in some open set. The price one pays is that the vector fields of such an orthonormal frame generally do not commute, and hence are not coordinate vector fields.

Exercise 4.9 Starting with a set of coordinate vector fields $\{\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^m}\}$, and verify that the Gram-Schmidt algorithm yields a smooth orthonormal frame on $U$.

In general there is little hope that there exist global orthonormal frames: For example, on the 2-sphere there does not even exist a single nonvanishing vector field, and thus not even one orthonormal vector field. On the other hand, both the flat and the embedded 2-torus, and the embedded 3-sphere $S^3 \hookrightarrow \mathbb{R}^4$ have (global) orthonormal 2 and 3-frames, respectively.

Exercise 4.10 Explicitly construct an orthonormal frame on the imbedded torus $T(R, r) = \{(x, y, z) \in \mathbb{R}^3: (\sqrt{x^2 + y^2} - R)^2 + z^2 = r^2\}$. Is it possible to obtain a frame of coordinate vector fields? Why, or why not? Suggestion: Start with the coordinate vector fields $\frac{\partial}{\partial \theta}$ and $\frac{\partial}{\partial \phi}$ defined by the parametrization $F(\theta, \phi) = (R + r \cos \phi) \cos \theta, (R + r \cos \phi) \sin \theta, r \sin \phi)$. Analyze the Lie bracket of the normalized/orthogonalized fields.

Familiar examples – which motivated the development of these structures in the first place – are hypersurfaces in Euclidean spaces, i.e. $m$-manifolds imbedded in $\mathbb{R}^{m+1}$. For clarity of exposition consider the special case of $m = 2$ and distinguish between graphs of functions and more general parameterized surfaces. The objective here is to relate the classical notation and terminology to the general language developed for Riemannian manifolds.

If $f: \mathbb{R}^2 \to \mathbb{R}$ is a smooth function it is natural to consider the atlas for the graph of $f$, i.e., $M = \{(x^1, x^2, f(x^1, x^2)) : (x^1, x^2) \in \mathbb{R}^2\}$ that consists of the single chart $(u, U)$ with $U = M$ and $u^{-1}(x^1, x^2) = (x^1, x^2, f(x^1, x^2))$. 


99
Let \( \iota: M \hookrightarrow \mathbb{R}^3 \) denote the inclusion map and \( \langle \cdot, \cdot \rangle \) be the Euclidean metric on \( \mathbb{R}^3 \).

With slight abuse of notation, calculate the induced metric

\[
 g_{ij} = \iota^*\left( \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right)
 = \langle \iota_* \frac{\partial}{\partial x^i}, \iota_* \frac{\partial}{\partial x^j} \rangle
 = \langle (D_i + (D_i f)D_3, D_j + (D_j f)D_3) \rangle
 = \delta_{ij} + (D_i f)(D_j f)
\]

It is a good exercise to check in detail where each of these terms naturally lives, where each term is evaluated. E.g. in the product \( (D_i f)D_3 \) the \( D_i \) is a derivative in \( \mathbb{R}^2 \), the domain of \( f \), whereas the \( D_3 \) is a derivative in the ambient space \( \mathbb{R}^3 \). Keeping track of all details the same calculation may appear as follows, for \( p = (x^1, x^2, f(x^1, x^2)) \in M \),

\[
 g_{ij}(p) = \iota^*\left( \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^j} \right)
 = \langle \iota_* \frac{\partial}{\partial x^1}, \iota_* \frac{\partial}{\partial x^j} \rangle
 = \langle (D_i|_p + (D_i f)|_p) \cdot D_3|_p, (D_j|_p + (D_j f)|_p) \cdot D_3|_p \rangle
 = \delta_{ij} + (D_i f)(D_j f)(u(p))
\]

The calculation of \( \iota_* \left( \frac{\partial}{\partial x^i} \right) \) uses the definition – if \( h \in C^\infty(p) \) is a smooth function defined in some neighborhood of \( p = \iota(p) \) in \( \mathbb{R}^3 \) – not just defined on \( M \) – then, the chain rule applies, and

\[
 \left( \iota_* \left( \frac{\partial}{\partial x^i} \right) \right) h = \frac{\partial}{\partial \nu^i}|_p(h \circ \iota) = D_i(h \circ \iota \circ u^{-1})|_{u(p)}
 = (D_i h)(\iota(p)) + (D_3 h)(\iota(p)) \cdot (D_i f)(u(p))
\]

using that \( (h \circ \iota \circ u^{-1})(x^1, x^2) = h(x^1, x^2, f(x^1, x^2)) \). For the record let us write out the matrix whose components are the functions \( g_{ij} \) in traditional notation using subscripts to denote partial derivatives – the induced metric on the graph of a function \( f(x^1, x^2) \) using standard coordinates

\[
 (g_{ij})_{i,j=1..m} = \left( \begin{array}{cc}
 1 + f_x^2 & f_x f_y \\
 f_x f_y & 1 + f_y^2
 \end{array} \right) \circ u
\]

Other (hyper-)surfaces are described, in elementary language, as parameterized surfaces. For example the embedded sphere and torus in \( \mathbb{R}^3 \) may be described as the images of the parameterizations (with standard identification of the edges, and for constants \( R > r > 0 \)

\[
 F: [-\pi, \pi] \times [0, \pi] \hookrightarrow \mathbb{R}^3 \quad F(\theta, \phi) = (r \cos \theta \ r \sin \phi, \sin \theta \ r \sin \phi, \cos \phi)
 F: [-\pi, \pi] \times [-\pi, \pi] \hookrightarrow \mathbb{R}^3 \quad F(\theta, \phi) = ((R + r \cos \phi) \cos \theta, (R + r \cos \phi) \sin \theta, r \sin \phi)
\]

In either case does there exist an atlas consisting of a single local coordinate chart – but it is straightforward to introduce suitable cuts and modify the maps as necessary. Thus here we only consider a typical chart. In the specific examples we may take the images of the open rectangles, i.e. \( U = F((-\pi, \pi) \times (0, \pi)) \) and \( U = F((-\pi, \pi) \times (-\pi, \pi)) \), respectively, and define \( u = F^{-1} \) as the inverse of the parameterization. In particular, we consider \( \theta = u^1 \) and \( \phi = u^2 \) as coordinate functions defined on the manifolds (surfaces). As before, we may use the inclusion map \( \iota: M \hookrightarrow \mathbb{R}^3 \) to pull back the Euclidean metric to the surfaces. With the usual abuse of
notation we calculate the induced metric with respect to the coordinates $u = (\theta, \phi)$ as

$$g_{ij} = \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \rangle$$

$$= \langle (D_i F^1) D_1 + (D_i F^2) D_2 + (D_i F^3) D_3, (D_j F^1) D_1 + (D_j F^2) D_2 + (D_j F^3) D_3 \rangle$$

$$= (D_i F^1) (D_j F^1) + (D_i F^2) (D_j F^2) + (D_i F^3) (D_j F^3).$$

Using the specific coordinates $u = (\theta, \phi)$ for $M^2 \subseteq \mathbb{R}^3$ we obtain

$$(g_{ij})_{i,j=1,2} = \begin{pmatrix}
\frac{\partial F^1}{\partial \theta} \frac{\partial F^1}{\partial \phi} + \frac{\partial F^2}{\partial \theta} \frac{\partial F^2}{\partial \phi} + \frac{\partial F^3}{\partial \theta} \frac{\partial F^3}{\partial \phi} & \frac{\partial F^1}{\partial \theta} \frac{\partial F^2}{\partial \phi} + \frac{\partial F^2}{\partial \theta} \frac{\partial F^3}{\partial \phi} + \frac{\partial F^3}{\partial \theta} \frac{\partial F^1}{\partial \phi} \\
\frac{\partial F^1}{\partial \theta} \frac{\partial F^2}{\partial \phi} + \frac{\partial F^2}{\partial \theta} \frac{\partial F^3}{\partial \phi} + \frac{\partial F^3}{\partial \theta} \frac{\partial F^1}{\partial \phi} & \frac{\partial F^2}{\partial \theta} \frac{\partial F^2}{\partial \phi} + \frac{\partial F^3}{\partial \theta} \frac{\partial F^3}{\partial \phi}
\end{pmatrix}$$

which in the special cases of the sphere and the torus evaluates to the explicit formulae

$$(g_{ij})_{i,j=1,2}^{\text{sphere}} = \begin{pmatrix}
r^2 \sin^2 \phi & 0 \\
0 & r^2
\end{pmatrix} \quad \text{and} \quad (g_{ij})_{i,j=1,2}^{\text{torus}} = \begin{pmatrix}
(R + r \cos \phi)^2 & 0 \\
0 & r^2
\end{pmatrix}$$

**Exercise 4.11** Verify the formulas (196). Calculate the metric in coordinates for some other parameterized surface of your choice – make sure to explicitly define your choice of coordinates.

**Exercise 4.12** Note that both matrices in (196) have columns that are mutually orthogonal. Explain the practical meaning in terms of the level curves $u^i = \text{const}$ (as curves in $\mathbb{R}^3$).

This is an opportune point to revisit differential $m$-forms: Recall that $\dim \Lambda^m(M) = 1$ if $\dim M = m$. Thus every nonvanishing $m$-form is a generator for this one-dimensional vector space, but until now there was no natural way of singling out a preferred $m$-form. Now the Riemannian structure suggests a distinguished element:

**Proposition 4.6** Suppose $(M^m, \langle \cdot, \cdot \rangle)$ is an $m$-dimensional orientable manifold. Then there is a unique $m$-form in $\Omega^m(M)$, denoted $dV$ and called the volume form on $M$ which is such that for every positively oriented, orthonormal sets of vectors $\{X_{1p}, X_{2p}, \ldots, X_{mp}\} \subseteq T_p M$,

$$dV_p(X_{1p}, X_{2p}, \ldots, X_{mp}) = 1.$$  \hspace{1cm} (197)

**Proof.** Suppose that $\{X_{1p}, X_{2p}, \ldots, X_{mp}\} \subseteq T_p M$ is a positively oriented, orthonormal sets of vectors. There exists a unique element $dV_p \in \Lambda^m(T_p M)$ such that $dV_p(X_{1p}, X_{2p}, \ldots, X_{mp}) = 1$. If $\{Y_{1p}, Y_{2p}, \ldots, Y_{mp}\} \subseteq T_p M$ is another, positively oriented, orthonormal sets of vectors then the formula (154) for the change of basis implies that $dV_p(Y_{1p}, Y_{2p}, \ldots, Y_{mp}) = 1$ since the transformation matrix between two orthonormal bases has determinant $\pm 1$, and $+1$ if both have the same orientation. Suppose that $(u, U)$ is a chart and $\{X_1, \ldots, X_m\}$ is a positively oriented orthonormal $m$-frame on $U$. Then

$$f = du^1 \wedge du^2 \wedge \ldots \wedge du^m(X_{1p}, X_{2p}, \ldots, X_{mp})$$

is a smooth, nowhere vanishing function on $U$ and $dV = \frac{1}{f} \cdot du^1 \wedge du^2 \wedge \ldots \wedge du^m$ is the unique differential $m$-form on $U$ with the desired properties. Since $M$ is orientable this determines a unique, globally defined $m$-form $dV \in \Omega^m(M)$. $\blacksquare$
Exercise 4.13 Suppose that \((u, U)\) is a chart on an orientable \(m\)-manifold \(M\) with metric on \(U\) defined by the matrix \(G\) with components \(g_{ij} = \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\rangle\).
Show that \(dV = (\det G) \cdot du^1 \wedge du^2 \wedge \ldots \wedge du^m\).

Note, even in the case that \(M\) is not orientable, \(dV = (\det G) \cdot |du^1 \wedge du^2 \wedge \ldots \wedge du^m|\) globally defines a volume element which may be used for integration on Riemannian manifolds.

Exercise 4.14 Calculate the volume form \(dV\) on the embedded sphere with radius \(r\) in \(\mathbb{R}^3\).

Exercise 4.15 Calculate the volume form \(dV\) on the embedded torus with inner and outer radii \(r\) and \(R\), respectively. Compare (193).

Exercise 4.16 Returning to the examples (196), explain in practical, graphical the dependence of \(g_{11}\) on \(\phi\) in terms of the areas of rectangular regions \((\theta, \phi)^{-1}([\theta_1, \theta_2] \times [\phi_1, \phi_2])\).

Recall that for every \(n\)-dimensional vector space \(V\), the space of alternating \(k\)-forms \(\Lambda^k(V) = \mathcal{A}_k(V^*) \subseteq T^k(V)\) has dimension \(\binom{n}{k} = \frac{n!}{k!(n-k)!}\). In particular, \(\Lambda^k(V)\) and \(\Lambda^{n-k}(V)\) have the same dimensions, suggesting a possible duality. However, without any additional structure on \(V^*\), there is no natural way to single out any one map to implement this duality.

On the other hand, an inner product on \(V\) immediately gives rise to a well-defined volume form as above, as well as distinguished orthonormal bases. Using these, there is a natural way to match elements in \(\Lambda^k(V)\) and \(\Lambda^{n-k}(V)\).

We proceed with an invariant definition of the Hodge star operator in the framework of Riemannian manifolds.

Definition 4.5 Suppose \(M\) is an oriented \(n\)-dimensional Riemannian manifold with volume form \(dV\), and \(0 \leq k \leq n\). Then there exists a unique linear map, called the Hodge star operator, \(*: \Omega^k(M) \mapsto \Omega^{(n-k)}(M)\) such that for every \(k\)-form \(\omega \in \Omega^k(M)\),
\[
\omega \wedge *\omega = dV. \tag{199}
\]

Briefly return to the framework of general inner product space \(V\). Suppose \(\{v_1, \ldots, v_n\}\) is any oriented orthonormal basis for \(V\), and \(\omega \in \Lambda^k(V)\) is an alternating \(k\)-form. To define \(*\omega \in \Lambda^{n-k}(V)\), it suffices to give a formula of what \(*\omega\) does to any \((n-k)\)-tuple of elements of the elements of the orthonormal basis. Observe that for every \(1 \leq i_1 < i_2 < \ldots < i_{n-k} \leq n\), there exist uniquely defined \(1 \leq j_1 < j_2 < \ldots < j_k \leq n\) such that (as unordered sets) \(\{j_1, \ldots, j_k, i_1, \ldots, i_{n-k}\} = \{1, \ldots, n\}\) and consequently \(dV(v_{j_1}, \ldots, v_{j_k}, v_{i_1}, \ldots, v_{i_{n-k}}) = \varepsilon_I \in \{1, -1\}\). Therefore \((*\omega)(v_{i_1}, \ldots, v_{i_{n-k}}) = \varepsilon_I \omega(v_{j_1}, \ldots, v_{j_k})\).

Exercise 4.17 Verify that the above characterization of \(\omega\) in terms of bases indeed gives the same result as the preceding invariant definition, and that it is well-defined, i.e. yielding the same result for every orthonormal basis.

Exercise 4.18 For a local coordinate chart \((u, U)\) on an oriented Riemannian manifold \(M^n\), and a \(k\)-form \(\omega = \sum_I \omega_I du^I\) provide an explicit formula for \(*\omega\).
The Hodge star operator has many interesting properties, and serves as a powerful tool to relate properties of objects of complementary dimensions. We briefly note that up to a possible change of sign, the Hodge star is almost its own inverse \( **\omega = (-1)^k(n-k)\omega \). In particular, when \( \dim V = 3 \) or \( \dim M = 3 \) then for every form \( \omega \) (i.e., zero, one, two, or three-form), \( **\omega = \omega \), making the application of the Hodge star even less visible in the familiar setting of \( \mathbb{R}^3 \). Indeed, together with the map \( \alpha \) of equation (181) that raises indices, the Hodge star provides many different ways to match objects in e.g. \( \mathcal{T}\mathbb{R}^3, \mathcal{T}^*\mathbb{R}^3, \Lambda^2\mathbb{R}^3 \): 

**Exercise 4.19** Using \( (x, y, z) = (x^1, x^2, x^3) \) as the standard coordinates on \( \mathbb{R}^3 \), elaborate how \( \alpha \), its inverse \( \alpha^{-1} \), and the Hodge star \( * \) map the basic sets \( \{ \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \} \), \( \{ dx, dy, dz \} \) and \( \{ dx \wedge dy, dy \wedge dz, dz \wedge dx \} \) to each other, and write out the images of any tangent vector, one-from, and two form. Moreover, how to represent (infinitesimal) parallelograms? Combine these maps with the exterior derivative, and use these maps to rewrite the familiar operations of gradient, curl, and divergence in terms of \( \alpha, \alpha^{-1}, *, \) and \( d \). Note, that the gradient of a smooth function, i.e. of a smooth zero-form \( f \in \Omega^0(\mathbb{R}^3) \) is a tangent vector field. Finally, observe that the Laplacian \( \triangle : \Omega^0(\mathbb{R}^3) \mapsto \Omega^3(\mathbb{R}^3) \equiv \Omega^0(\mathbb{R}^3) \) may be nicely expressed \( \triangle = d \ast d \).

**Exercise 4.20** Repeat the preceding exercise for a common orthonormal, but curvilinear coordinate system such as standard spherical coordinates. Explore how the formulas change for an arbitrary local coordinate system \( (u, U) \) on \( \mathbb{R}^3 \).

This is just the very beginning – there is a lot more. For example the composition \( \ast d \ast \) yields an interesting anti derivation, mapping \( k \)-forms to \( (k-1) \) forms. For an in-depth treatment, leading up to the Hodge decomposition theorem and Bochner’s theorem, see Spivak 2nd edition, vo. 4, Ch. 7, addendum 2, pp. 202–218. In particular, there is more than one way of extending the Laplacian to these settings, e.g. contrast the Laplacian studied on these pages pp. 202–218 with the Laplacian of addendum 1, pp. 187–201.

__________________________end of “in progress”, move into linear algebra section ?? ___________________

**Definition 4.6** Suppose that \( \sigma : [a, b] \mapsto M \) is a piecewise smooth curve, i.e. \( \sigma \) is continuous and there exist \( a = t_0 < t_1 < \ldots t_r = b \) such that the restrictions of \( \sigma \) to any open interval \( (t_i, t_{i+1}) \) are smooth. Define the length of \( \sigma \) as

\[
L(\sigma) = L^b_a(\sigma) = \int_a^b \langle \sigma'(t), \sigma'(t) \rangle \, dt.
\]

(200)

One easily checks that this is well-defined as one always may consider a finer partition obtained by taking the union of partitions \( \{ t_0, t_1, \ldots t_r \} \) and \( \{ t'_0, t'_1, \ldots t'_s \} \).

**Definition 4.7** For \( p, q \in M \) define the Riemannian distance from \( p \) to \( q \) as

\[
d_R(p, q) = \inf_{\sigma} L^b_{a_\sigma}(\sigma)
\]

(201)

where the infimum is taken over all piecewise smooth curves \( \sigma : [a_\sigma, b_\sigma] \mapsto M \) with \( \sigma(a_\sigma) = p \) and \( \sigma(b_\sigma) = q \).

In the case that \( p \) and \( q \) lie in different components of \( M \), i.e., when there is no piecewise smooth curve connecting them, then \( d_R(p, q) = \infty \). In all other cases clearly \( d_R(p, q) < \infty \). Nonetheless,
in general there need not exist any piecewise smooth curve that attains the infimum. The most basic example is that of \( M = \mathbb{R}^2 \setminus \{(0,0)\} \) with the standard Euclidean metric and \( p = (-1,0) \) and \( q = (1,0) \).

**Proposition 4.7** Suppose that \( M \) is a connected Riemannian manifold. Then the function \( d_R: M \times M \mapsto \mathbb{R} \) is a metric on \( M \).

**Proposition 4.8** The metric \( d_R: M \times M \mapsto \mathbb{R} \) generates the topology of \( M \).

**Proof.** (For both propositions combined.) It is straightforward to see that \( d_R \) is symmetric as the length of \( \sigma: [a, b] \mapsto M \) and \( \sigma^{-1}: [-b, -a] \mapsto M \), defined by \( \sigma^{-1}(t) = \sigma(-t) \) are the same.

To verify that the triangle inequality is satisfied suppose \( \sigma_1: [a, b] \mapsto M \) and \( \sigma_2: [c, d] \mapsto M \) are piecewise smooth curves with \( \sigma_1(a) = p \), \( \sigma_1(b) = \sigma_2(c) = z \), and \( \sigma_2(d) = q \). Define a curve \( \sigma_1 \sharp \sigma_2: [a, b + d - c] \mapsto M \) by \( \sigma_1 \sharp \sigma_2(t) = \sigma_1(t) \) if \( t < b \) and \( \sigma_1 \sharp \sigma_2(t) = \sigma_2(t - c + b) \) else. Then \( \sigma_1 \sharp \sigma_2 \) is a piecewise smooth curve connecting \( p \) to \( q \) and \( L(\sigma_1 \sharp \sigma_2) = L(\sigma_1) + L(\sigma_2) \). Consequently the infimum over all piecewise smooth curves connecting \( p \) to \( q \) is no larger than the infimum over those curves of the form \( \sigma_1 \sharp \sigma_2 \), thus showing that \( d_R(p, q) \leq d_R(p, z) + d_R(z, q) \).

It remains to be shown that if \( p \neq q \) then \( d_R(p, q) > 0 \). While intuitive – after all a Riemannian metric should on a small scale look very much like Euclidean space, it remains to be shown that it is impossible to have a sequence of piecewise smooth curves \( \sigma_k: [a_k, b_k] \mapsto M \) with \( \sigma_k(a_k) = p \), \( \sigma_k(b_k) = q \), for all \( k \in \mathbb{Z} \), such that \( \inf_k L(\sigma_k) = 0 \) yet \( p \neq q \).

\[
\text{in progress}
\]

\begin{itemize}
  \item Coordinate maps \( u \) are by definition diffeomorphisms. Thus instead of worrying about open sets on \( M \) w.r.t. its native (metric/metrizable) topology only consider balls \( B_0(\varepsilon) \subseteq \mathbb{R}^m \) (w.r.t. the standard Euclidean metric).
  \item Use (the inverse of) the coordinate map \( u \) to pull back the Riemannian metric \( \langle \cdot, \cdot \rangle \) from \( U \subseteq M \) to a Riemannian metric \( (u^{-1})^* \langle \cdot, \cdot \rangle \) on \( u(U) \subseteq \mathbb{R}^m \). With this metric \( u(U) \) is clearly isometric to \( U \).
  \item basically use the continuity of this metric and its positive definiteness to bound it away from zero – in the end it is about eigenvalues of positive definite symmetric matrices. These vary continuously! Thus need only get some compact set. Start with some compact set \( K_0 \) containing an open neighborhood of the point \( u(p) \), then consider the set \( K = \{(q, X_q) \in TK_0 \subseteq T\mathbb{R}^n : q \in K, \|X_q\| = 1\} \). (202)
\end{itemize}

On this set find the minimum and maximum of \( \langle q, X_q \rangle \mapsto (u^{-1})^* \langle X_p, X_p \rangle \) which are strictly positive and finite. The rest is straightforward lower and upper estimates of the length of any curve in \( K \) in terms of these bounds, using the homogeneity of the metric, i.e. \( \langle \sigma', \sigma' \rangle = \|\sigma'\|\langle X_q, X_q \rangle \) if \( X^q \) is a unit vector such that \( \sigma' = \|\sigma'\| \cdot X^q \).

\[
\text{in progress}
\]

**Exercise 4.21** WHERE SHOULD THIS GO? Let \( S^n(r) \subseteq \mathbb{R}^{n+1} \) denote the imbedded \( n \)-sphere with radius \( r \), and equipped with the induced metric. Show that \( S^n(r) \) and \( S^n(r') \) are not isometric if \( r \neq r' \).
4.4 Geodesics and the Christoffel symbols

There are many ways in which to approach geodesics: closest to the original sense of this word is the characterization of geodesics as curves that minimize distance in a sense yet to be made precise. Closely related is their characterization as the analogues to the straight lines of Euclidean spaces. Geodesics also govern the evolution of mechanical systems, as the problems of minimizing length, minimizing curvature, and minimizing action (the time integral of energy) are intimately related. We start by studying the problem of minimizing distance, and then change focus to the intricate interplay of geodesics and notions of curvature: One might take the excess (or deficit) of the sum of the interior angles of a triangle – made up of “straight” sides as an intuitive measure of the total curvature (of a triangular region).

The classical starting point for the calculus of variations are minimization problems of the following form: On some subset \(U \subseteq \mathbb{R}^{2n+1}\), given a running cost \(L \in C^2(U)\), find among all piecewise \(C^2\)-curves \(\gamma: [a, b] \mapsto \mathbb{R}^n\) that satisfy \(\gamma(a) = p\) and \(\gamma(b) = q\) the one(s) that minimize

\[
J(\gamma) = \int_a^b L(t, \gamma(t), \gamma'(t)) \, dt \tag{203}
\]

Simple examples in the case of \(n = 1\) are \(L(t, x, y) = \sqrt{1 + y^2}\) in which case \(J(\gamma)\) is the arc-length of the graph of \(\gamma\), and \(L(t, x, y) = 2\pi \cdot \sqrt{1 + y^2}\) in which case \(J(\gamma)\) is the area of the surface of revolution obtained by revolving the graph of \(\gamma(t)\) about the \(t\)-axis. For general \(n \geq 1\) the
\[
L(t, x, y) = \left( \sum_{k=1}^n \|y^k\|_2 \right)^{1/2}
\]

yields the length of (the image of the curves \(\gamma: [a, b] \mapsto \mathbb{R}^n\).

Our interest lies in curves that take values in a Riemannian manifold and the integrand is a function \(L: TM \mapsto \mathbb{R}\) defined by \(L(p, X_p) = \langle X_p, X_p \rangle^{1/2}\).

The basic idea is that if the functional \(J\) is differentiable at a curve \(\hat{\gamma}\) and \(J'(\hat{\gamma}) \neq 0\) then there exist curves \(\gamma\) near \(\hat{\gamma}\) such that \(J(\gamma) < J(\hat{\gamma})\), i.e. \(\hat{\gamma}\) cannot possible be a minimizer. For the purposes considered here it is not necessary to develop a full notion of differentiability of functionals \(J\) with respect to curves \(\gamma\), but it suffices to consider directional derivatives obtained from one-parameter families of variations.

In the basic case of curves in Euclidean space \(\mathbb{R}^n\) one may restrict to additive variations. More specifically, for two fixed \(C^2\)-curves \(\hat{\gamma}, \eta: [a, b] \mapsto \mathbb{R}^n\) with \(\hat{\gamma}(a) = p\) and \(\hat{\gamma}(b) = q\), while \(\eta(a) = 0\) and \(\eta(b) = 0\), consider for \(\varepsilon \in \mathbb{R}\) sufficiently small (e.g. so that for all \(t \in [a, b]\), \(\gamma_{\varepsilon}(t) \in U\) the family of curves \(\gamma_{\varepsilon}: [a, b] \mapsto \mathbb{R}^n\) defined by \(\gamma_{\varepsilon} = \hat{\gamma} + \varepsilon \eta\). Note that for all \(\varepsilon \in \mathbb{R}\), \(\gamma_{\varepsilon}(a) = p\) and \(\gamma_{\varepsilon}(b) = q\) If \(\hat{\gamma}\) minimizes \(J\) then necessarily for all \(\varepsilon\), \(J(\hat{\gamma}) \leq J(\gamma_{\varepsilon})\). Assuming sufficient
Suppose \( f \) yields a contradiction, thus establishes that the integrands must be identically equal to zero for each \( k \) curves \( \eta \). 

\[ \frac{d}{d\varepsilon} \bigg|_{\varepsilon=0} J(\gamma_\varepsilon) = \int_a^b L(t, \gamma_\varepsilon(t), \gamma'_\varepsilon(t)) \, dt \]

\[ = \int_a^b \frac{d}{d\varepsilon} \bigg|_{\varepsilon=0} L(t, \gamma(t) + \varepsilon\eta(t), \gamma'(t) + \varepsilon\eta'(t)) \, dt \]

\[ = \int_a^b \sum_{k=1}^n \left( L_{xk}(t, \gamma(t), \gamma'(t)) \cdot \eta^k(t) + L_{yk}(t, \gamma(t), \gamma'(t)) \cdot \eta'^k(t) \right) \, dt \]

\[ = \sum_{k=1}^n \left( L_{xk}(t, \gamma(t), \gamma'(t)) \cdot \eta^k(t) \right)_{t=a}^{t=b} + \sum_{k=1}^n \int_a^b \left( L_{xk}(t, \gamma(t), \gamma'(t)) - \frac{d}{dt} L_{yk}(t, \gamma(t), \gamma'(t)) \right) \cdot \eta^k(t) \, dt \]

Here \( L_{xk} \) and \( L_{yk} \) stand for the partial derivatives \( D_{k+1}L \) and \( D_{n+k+1}L \). The key is the integration by parts in the last step together with the observation that the boundary terms vanish since \( \eta(a) = \eta(b) = 0 \), i.e., since the variations hold the endpoints fixed.

Now one argues that if \( \gamma \) minimizes \( J \) then this derivative must be zero for all \( C^1 \)-curves \( \eta: [a,b] \mapsto \mathbb{R}^n \) with \( \eta(a) = \eta(b) = 0 \). Using the subsequent lemma, and considering special curves \( \eta \) such that all coordinate functions \( \eta^k \) except one are identically equal to zero one establishes that the integrands must be identically equal to zero for each \( k \leq n \):

**Proposition 4.9** Suppose \( L \in C^2(\mathbb{R}^{2n+1}), \gamma \in C^2([a,b], \mathbb{R}^n), \) and \( J: C^2([a,b], \mathbb{R}^n) \mapsto \mathbb{R} \) is defined by \( J(\gamma) = \int_a^b L(t, \gamma(t), \gamma'(t)) \, dt \). If \( J(\hat{\gamma}) \leq J(\hat{\gamma} + \varepsilon\eta) \) for all \( \eta \in C^1([a,b], \mathbb{R}^n) \) with \( \eta(a) = \eta(b) = 0 \), then \( \hat{\gamma} \) must satisfy the Euler-Lagrange equations

\[ L_{xk}(t, \hat{\gamma}(t), \hat{\gamma}'(t)) - \frac{d}{dt} L_{yk}(t, \hat{\gamma}(t), \hat{\gamma}'(t)) = 0 \quad \text{for all} \quad t \in [a,b], \ k = 1, \ldots, n \]

**Lemma 4.10** Suppose \( f \in C^0([a,b]) \). If for all \( g \in C^\infty([a,b]) \) with \( g(0) = g(b) = 0 \), \( \int_a^b f(t)g(t) \, dt \) then \( f \equiv 0 \).

The lemma is stated here for the case of the integral vanishing for all smooth \( g \). It is an area of continued research to find ever stronger, similar conditions for various regularity assumptions. Typically, e.g. in control theory, one has much weaker regularity assumptions on \( L \) and \( \gamma \), while considering more innovative families of variations.

**Proof.** Suppose \( t_0 \in (a, b) \) is such that \( f(t_0) \neq 0 \). W.l.o.g. assume that \( f(t_0) > 0 \). Since \( f \) is continuous, there exists \( \delta > 0 \) such that \( a < t_0 - 2\delta < t_0 + 2\delta < b \), and such that \( f(t) \geq \frac{1}{2} f(t_0) \) for all \( t \in [t_0 - 2\delta, t_0 + 2\delta] \). There exists a smooth function \( g \in C^\infty([a, b]) \) \( g \geq 0 \), such that \( g(t) = 1 \) if \( t \in [t_0 - \delta, t_0 + \delta] \) and \( g(t) = 0 \) if \( t \not\in [t_0 - 2\delta, t_0 + 2\delta] \). Then

\[ \int_a^b f(t)g(t) \, dt = \int_{t_0 - 2\delta}^{t_0 + 2\delta} f(t)g(t) \, dt > \int_{t_0 - \delta}^{t_0 + \delta} f(t)g(t) \, dt > 0 \]

yields a contradiction, thus \( f(t) = 0 \) for all \( t \in (a, b) \), and by continuity also for all \( t \in [a, b] \). ■
To provide a feeling for the Euler-Lagrange equations it is encouraged to work a few classical examples from calculus of variations.

Exercise 4.22 Evaluate the Euler-Lagrange equations for the problem of minimizing the arc-length of the graph among all $C^2$-functions $\gamma: [a, b] \mapsto \mathbb{R}^n$ that satisfy $f(a) = p$, and $f(b) = q$ (for fixed endpoints $p, q \in \mathbb{R}^n$). (Hint: Use $L(t, x, y) = \sqrt{1 + \|y\|^2}$.) Carefully argue why a minimizer must exist, why it is unique, and then identify the minimizing function.

Exercise 4.23 Evaluate the Euler-Lagrange equations for the problem of minimizing the arc-length among all $C^2$-curves $\gamma: [a, b] \mapsto \mathbb{R}^n$ that satisfy $f(a) = p$, and $f(b) = q$ (for fixed endpoints $p, q \in \mathbb{R}^n$. (Hint: Use $L(t, x, y) = \|y\|$.) Carefully argue why a minimizer must exist, why it is unique, and then identify the minimizing function.

Exercise 4.24 Evaluate the Euler-Lagrange equations for the problem of minimizing the area of the surface of revolution obtained by revolving the graph of a $C^2$-function $\gamma: [a, b] \mapsto \mathbb{R}$ that satisfy $f(a) = p$, and $f(b) = q$ about the $t$-axis. (Hint: Use $L(t, x, y) = 2\pi x \cdot \sqrt{1 + \|y\|^2}$.) Carefully argue why a minimizer must exist, why it is unique, and then identify the minimizing function by integrating the differential equations. The solution is commonly expressed in terms of the hyperbolic cosine.

Similar classical examples involve the shape of a physical chain hanging under its own weight between two points (the solution is also a catenary), and the curve of fastest descent under gravity (brachistochrone).

These exercises suggest that it may be worth to spend some extra effort to circum-navigate the unpleasant radical in the integrand of above problems. Indeed, as the subsequent analysis shows, this is indeed a preferable approach. The next exercise will give some sense for what is going on.

Exercise 4.25 Evaluate the Euler-Lagrange equations for the problem of minimizing the integral of the energy among all $C^2$-curves $\gamma: [a, b] \mapsto \mathbb{R}^n$ that satisfy $f(a) = p$, and $f(b) = q$ (for fixed endpoints $p, q \in \mathbb{R}^n$, i.e. use the cost $L(t, x, y) = \|y\|^2$.) Compare the resulting differential equations, and their solutions with those of minimizing arc-length as in exercise 4.23.

Exercise 4.26 Euler-Lagrange equations when integrand does not depend on $x$

Exercise 4.27 Euler-Lagrange equations when integrand ....
On Riemannian manifolds, the notion of *additive* variations \( \gamma_\varepsilon = \hat{\gamma} + \varepsilon \eta \) generally does not make sense. Hence, for a fixed curve \( \hat{\gamma} : [a, b] \mapsto M \) one may consider families of variations of the form

\[
\gamma : (-\varepsilon_0, \varepsilon_0) \times [a, b] \mapsto M
\]

such that for all \( \varepsilon \in (-\varepsilon_0, \varepsilon_0) \), \( \gamma(\varepsilon, a) = \hat{\gamma}(a) \), \( \gamma(\varepsilon, b) = \hat{\gamma}(b) \), and such that for all \( t \in [a, b] \), \( \gamma(0, t) = \hat{\gamma} \).

At this point the road forks: A clean geometric treatment would develop the analogous theory for cost functions \( L \in C^2([a, b] \times TM) \) leading to differential equations on \( TT^n \). This approach involves a little more abstraction, and quite a bit extra notation that reflects the inherent symmetries of the \( 4n \)-dimensional manifold \( TT^n \), e.g. each \( T_XTM \) contains two copies of \( T_qM \). We here follow a more pedestrian approach by working in coordinate charts which is sufficient for the local results that we desire, which is considerably faster but less elegant. In particular, this route demands that one pay attention to curves leaving charts and entering neighboring charts. It is left as an exercise to work out the details for deriving the Euler-Lagrange equations for general, not necessarily additive variations, in the setting of curves in Euclidean space – which in some sense is a first step towards the geometric alternative road.

**Exercise 4.28** Work out an analogous calculation to (204) that led to proposition 4.9, for the more general case which uses not necessarily additive variations \( \gamma : (-\varepsilon_0, \varepsilon_0) \times [a, b] \mapsto \mathbb{R}^n \) as discussed above. Hint: Now \( \eta \) is replaced by \( \frac{\partial}{\partial \varepsilon} \big|_{\varepsilon=0} \gamma \) and a key step uses that \( \frac{\partial}{\partial t} \frac{\partial}{\partial \varepsilon} \gamma = \frac{\partial}{\partial \varepsilon} \frac{\partial}{\partial t} \gamma \).

---

Change to a modern presentation, along the lines of

Back to Riemannian manifolds, apply the calculus of variations argument in one each chart at a time. It will turn out helpful later to allow the reference curves to only be piecewise $C^2$, and also allow more flexibility with regard to the endpoints. Specifically, suppose $M$ is a Riemannian manifold and $p \neq q \in M$ are two points joined by a piecewise $C^2$-curve $\tilde{\sigma} : [a, b] \mapsto M$, i.e. $\tilde{\sigma}(a) = p$ and $\tilde{\sigma}(b) = q$. Suppose that $a = t_0 < t_1 < \ldots t_N = b$ are times such that the restriction of $\tilde{\sigma}$ to each of the intervals $(t_{r-1}, t_r)$ is $C^2$. W.l.o.g. one may assume that for each subinterval $[t_{r-1}, t_r]$ there is a chart $(u_r, U_r)$ such that $\tilde{\sigma}([t_{r-1}, t_r]) \subset U_r$. To avoid excessive notation we do not introduce additional symbols to denote the sequence $(u_r, U_r)$ of charts, and simply write $(u, U)$ - but care must be taken later as e.g. there will be functions $g_{ij}$ that are different from chart to chart.

As we shall justify later, instead of minimizing length, we shall investigate the problem of minimizing the integral of the energy $\int (\sigma', \sigma')dt$. Writing $\gamma = u \circ \sigma : [t_{r-1}, t_r] \mapsto u(U) \subseteq \mathbb{R}^n$, the problem is to minimize the sum of the integrals (sum over all subintervals)

$$E(\sigma |_{[t_{r-1}, t_r]}) = \int_{t_{r-1}}^{t_r} \frac{1}{2} (\sigma'(t), \sigma'(t))_{\sigma(t)} dt$$

$$= \int_{t_{r-1}}^{t_r} \frac{1}{2} (u^{-1})^*(\gamma'(t), \gamma'(t))_{\gamma(t)} dt = \int_{t_{r-1}}^{t_r} \frac{1}{2} \sum_{i,j=1}^n g_{ij}(\sigma(t)) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} dt$$  \hspace{1cm} (207)

In general, consider variations that preserve the piecewise $C^2$-nature of the class in which we are looking for extremals,

$$\sigma : [a, b] \times (-\varepsilon_0, \varepsilon_0) \mapsto M, \text{ such that for all } \varepsilon \in (-\varepsilon_0, \varepsilon_0), \sigma(a, \varepsilon) = p \text{ and } \sigma(b, \varepsilon) = q. \hspace{1cm} (208)$$

and for each $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ fixed, and each $1 \leq k \neq N$, the restriction of the curve $t \mapsto \sigma(t, \varepsilon) \in M$ to the interval $(t_{r-1}, t_r)$ is $C^2$. (More generally, one could have that the subintervals themselves depend on the parameter, i.e. $(t_{r-1}(\varepsilon), t_r(\varepsilon))$, which in similar applications may lead to stronger necessary conditions for optimality. But there is no need for this here.) On each of the subintervals, argue as before: integrate by parts and obtain the analogue of the Euler Lagrange equations for this energy functional. However, in general the boundary terms need no longer vanish at the endpoints of the subintervals. Thus one expects that the final necessary condition contains also various terms evaluated at these times $t_k$. However, to simplify matters, one may in a first step only consider variations that are supported on a single, fixed subinterval, i.e., for all $t \in [a, b] \setminus (t_{r-1}, t_r)$, and all $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$, $\sigma(t, \varepsilon) = \sigma(t, 0) = \tilde{\sigma}(t)$. Note that by only considering such special kinds of variations one potentially obtains a weaker necessary condition for extremals – i.e., if some derivative has to be zero for all variations, then it must, a fortiori, be zero for the special class.

Two last notational items before proceeding to the calculation: Assuming that $(u, U)$ is a chart such that $\sigma([t_{r-1}, t_r] \times (-\varepsilon_0, \varepsilon_0) \subseteq U$, write $\gamma = u \circ \sigma$, and for $1 \leq i, j \leq n$ write $g_{ij} = \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \rangle$. Note that fully written out in terms of the variables $t$ and $\varepsilon$, the integrand in the energy integral is

$$\langle \sigma'(t, \varepsilon), \sigma'(t, \varepsilon) \rangle_{\sigma(t, \varepsilon)} = \sum_{i,j=1}^n g_{ij}(\sigma(t, \varepsilon)) \frac{d\gamma^i}{dt}(t, \varepsilon) \frac{d\gamma^j}{dt}(t, \varepsilon). \hspace{1cm} (209)$$

using that for every $(t, \varepsilon)$, $\sigma'(t, \varepsilon) = \sum_{i=1}^n \frac{d}{dt}(u^i \circ \sigma)(t, \varepsilon) \frac{\partial}{\partial u^i}|_{\sigma(t, \varepsilon)}$. In particular, the $g_{ij}$ are evaluated along (the image of) $\sigma$, whereas the first and second derivatives are of the coordinate functions $\gamma^i = u^i \circ \sigma$. 

As before, for the Euler Lagrange equations, as before integrate by parts

\[
\frac{d}{d\epsilon} \bigg|_{\epsilon=0} E(\sigma_\epsilon |_{t_{r-1}, t_r}) = \int_{t_{r-1}}^{t(1)} \frac{1}{2} \sum_{i,j=1}^{n} \frac{d}{d\epsilon} \bigg|_{\epsilon=0} \left( (g_{ij} \circ \sigma) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} \right) dt
\]

\[
= \int_{t_{r-1}}^{t_r} \sum_{i,j=1}^{n} \left( \sum_{i,j=1}^{n} \frac{1}{2} \left( \frac{\partial g_{ij}}{\partial u^k} \circ \sigma \right) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} - \sum_{j=1}^{n} \frac{d}{dt} \left( (g_{ik} \circ \sigma) \frac{d\gamma^i}{dt} \right) \right) \left( \frac{\partial}{\partial \epsilon} \bigg|_{\epsilon=0} \gamma^k \right) dt \quad (210)
\]

For later use, record that in general in this calculation one would have for each subinterval \((t_{r-1}, t_r)\) the additional boundary terms

\[
\sum_{j,\ell=1}^{n} (g_{ik} \circ \sigma) \frac{d\gamma^i}{dt} \left( \frac{\partial}{\partial \epsilon} \bigg|_{\epsilon=0} \gamma^k \right) \bigg|_{t_r} - \bigg|_{t_{r-1}} \quad (211)
\]

Again use lemma 4.10 to obtain for \(1 \leq k \leq n\) (now considering only variations whose support is contained in \([t_{r-1}, t_r]\) the following equations. For notational simplicity one may from now on again write \(\sigma(\cdot)\) and \(\gamma(\cdot)\) for \(\sigma(\cdot, 0)\) and \(\gamma(\cdot, 0)\).

\[
0 = \frac{1}{2} \sum_{i,j=1}^{n} \left( \frac{\partial g_{ij}}{\partial u^k} \circ \sigma \right) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} - \left( \sum_{i,\ell=1}^{n} \left( \frac{\partial g_{ik}}{\partial u^\ell} \circ \sigma \right) \frac{d\gamma^i}{dt} \frac{d\gamma^\ell}{dt} + \sum_{i=1}^{n} (g_{ik} \circ \sigma) \frac{d^2\gamma^i}{dt^2} \right) \quad (212)
\]

This is customarily rewritten in the more symmetric form

\[
\sum_{i=1}^{n} (g_{ik} \circ \sigma) \frac{d^2\gamma^i}{dt^2} = \frac{1}{2} \sum_{i,j=1}^{n} \left( \frac{\partial g_{ij}}{\partial u^k} - \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i} \circ \sigma \right) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} \quad (213)
\]

Interpret this “first derivative is equal to zero” condition as a system of \(n\) linear equations in the \(n\) “unknowns” \(\frac{d^2\gamma^i}{dt^2}\), \(1 \leq k \leq n\). To solve for the second derivatives, “multiply” by the inverse of the matrix \(G = (g_{ij} \circ \sigma))_{i,j=1}^{n}\), i.e.

\[
\frac{d^2\gamma^k}{dt^2} = \sum_{\ell=1}^{n} \delta^k_{\ell} \frac{d^2\gamma^i}{dt^2} = \sum_{i,\ell=1}^{n} (g^{k\ell} \circ \sigma)(g_{i\ell} \circ \sigma) \frac{d^2\gamma^i}{dt^2} \quad (214)
\]

to obtain the geodesic equations, for \(1 \leq k \leq n\),

\[
0 = \frac{d^2\gamma^k}{dt^2} + \frac{1}{2} \sum_{i,j,\ell=1}^{n} (g^{k\ell} \circ \sigma) \left( \left( \frac{\partial g_{i\ell}}{\partial u^j} + \frac{\partial g_{j\ell}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^\ell} \circ \sigma \right) \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt} \right) \quad (215)
\]

These are usually abbreviated by introducing the Christoffel symbols of the first kind

\[
[ij, \ell] = \frac{1}{2} \left( \frac{\partial g_{i\ell}}{\partial u^j} + \frac{\partial g_{j\ell}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^\ell} \circ \sigma \right) \quad (216)
\]

and the Christoffel symbols of the second kind

\[
\Gamma^k_{ij} = \sum_{\ell=1}^{n} g^{k\ell} [ij, \ell] = \sum_{\ell=1}^{n} g^{k\ell} \frac{1}{2} \left( \frac{\partial g_{i\ell}}{\partial u^j} + \frac{\partial g_{j\ell}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^\ell} \circ \sigma \right). \quad (217)
\]
Using these abbreviations the geodesic equations become

\[ 0 = \frac{d^2 \gamma^k}{dt^2} + \sum_{i,j=1}^{n} \left( \Gamma^k_{ij} \circ \sigma \right) \frac{d \gamma^i}{dt} \frac{d \gamma^j}{dt}. \tag{218} \]

revisit the piecewise smooth problem, track down the corner conditions

somewhere: write out \( \Gamma^k_{ij} \) for some examples - surfaces in \( \mathbb{R}^3 \). Refer to MAPLE worksheets

**Proposition 4.11** Suppose \( \sigma : [a, b] \mapsto M \) is piecewise smooth and a critical point of the energy functional \( E_a^b \). Then \( \sigma \) is smooth.

**Proposition 4.12** If \( \sigma : [a, b] \mapsto M \) is a critical point for the energy \( E \), i.e. it satisfies (218) then it is parameterized proportionally to arc-length, i.e. \( \langle \sigma' \, , \sigma' \rangle : (a, b) \mapsto \mathbb{R} \) is constant.

**Proof.** This is a straightforward calculation - mainly requiring attention to the several summation indices. Suppose that \( \sigma : [a, b] \mapsto M \) is a critical point for the energy functional \( E_a^b \), i.e. it satisfies (218), and let \((u, U)\) be a chart about \( t_0 \in (a, b) \). For \( t \) sufficiently close to \( t_0 \) let \( \gamma = u \circ \sigma \). For notational convenience omit the argument of \( g_{ij} \) (which technically is to evaluated at \( \sigma(t) \)), and calculate:

Continuing the common notational abuse or simplification, in the following calculation write \( g_{ij} \) where correctly one should use \( g_{ij} \circ \sigma \).

\[
\frac{d}{dt} \langle \sigma' \, , \sigma' \rangle = \frac{d}{dt} \sum_{i,j=1}^{m} g_{ij} \cdot \frac{d \gamma^i}{dt} \cdot \frac{d \gamma^j}{dt} \\
= \sum_{i,j=1}^{m} \left( \sum_{\ell=1}^{m} \left( \frac{\partial g_{ij}}{\partial u^\ell} \cdot \frac{d \gamma^\ell}{dt} \cdot \frac{d \gamma^i}{dt} \cdot \frac{d \gamma^j}{dt} \right) + 2 \cdot g_{ij} \cdot \frac{d^2 \gamma^i}{dt^2} \cdot \frac{d \gamma^j}{dt} \right) \\
= \sum_{j=1}^{m} \frac{d \gamma^j}{dt} \left( \sum_{i=1}^{m} \left( \frac{\partial g_{ij}}{\partial u^\ell} \cdot \frac{d \gamma^\ell}{dt} \cdot \frac{d \gamma^i}{dt} \right) - 2 g_{ij} \sum_{r,s=1}^{m} \Gamma^i_{rs} \cdot \frac{d \gamma^r}{dt} \cdot \frac{d \gamma^s}{dt} \right) \\
= \sum_{j=1}^{m} \frac{d \gamma^j}{dt} \left( \sum_{i=1}^{m} \left( \frac{\partial g_{ij}}{\partial u^\ell} \cdot \frac{d \gamma^\ell}{dt} \cdot \frac{d \gamma^i}{dt} \right) - 2 g_{ij} \sum_{k,r,s=1}^{m} g^{ik} \left( \frac{\partial g_{rj}}{\partial u^k} + \frac{\partial g_{kj}}{\partial u^r} - \frac{\partial g_{rs}}{\partial u^k} \right) \cdot \frac{d \gamma^r}{dt} \cdot \frac{d \gamma^s}{dt} \right) \\
= \sum_{j=1}^{m} \frac{d \gamma^j}{dt} \left( \sum_{i=1}^{m} \left( \frac{\partial g_{ij}}{\partial u^\ell} \cdot \frac{d \gamma^\ell}{dt} \cdot \frac{d \gamma^i}{dt} \right) - 2 \sum_{r,s=1}^{m} \left( \frac{\partial g_{rj}}{\partial u^s} + \frac{\partial g_{sj}}{\partial u^r} - \frac{\partial g_{rs}}{\partial u^s} \right) \cdot \frac{d \gamma^r}{dt} \cdot \frac{d \gamma^s}{dt} \right) \\
= \sum_{i,j=1}^{m} \left( \frac{\partial g_{ij}}{\partial u^\ell} - \frac{\partial g_{ij}}{\partial u^\ell} - \frac{\partial g_{ij}}{\partial u^\ell} + \frac{\partial g_{ij}}{\partial u^\ell} \right) \cdot \frac{d \gamma^i}{dt} \cdot \frac{d \gamma^j}{dt} \cdot \frac{d \gamma^\ell}{dt} = 0. \quad \blacksquare
\]
Now return to the original problem of minimizing the length of the curve joining two points $p$ and $q$, i.e. consider the functional

$$L_a^b: \sigma \mapsto \int_a^b \sqrt{\langle \sigma'(t) \cdot \sigma'(t) \rangle} \, dt$$

(219)

The calculus of variations is essentially as before, but in order to satisfy the differentiability hypotheses in the derivation of the Euler-Lagrange equations first only consider curves that are immersions, i.e $\sigma'(t) \neq 0$ for all $t \in [a, b]$.

To outline the basic steps, informally write $\sqrt{\ldots} = \sqrt{F(x(t), \dot{x}(t))}$ for the integrand in the minimization problem. The Euler-Lagrange equations evaluate to

$$0 = \frac{1}{\sqrt{\ldots}} \cdot F_x \frac{d}{dt} \left( F_y \cdot \frac{1}{\sqrt{\ldots}} \right) = \frac{1}{\sqrt{\ldots}} \left( F_x \frac{d}{dt} F_y \right) + \frac{\frac{d}{dt} F}{\sqrt{\ldots}^3} = \frac{1}{\sqrt{\ldots}} \left( F_x \frac{d}{dt} F_y \right) + \frac{\frac{d}{dt} F}{\sqrt{\ldots}^3}$$

(220)

Exercise 4.29 Write out the details to derive the (real) geodesic equation, i.e. the condition for a curve $\sigma$ to be a critical point of the length functional $L_a^b$. The square root in the denominator plays no role. The next two terms are exactly the same as in the minimization of the energy functional. The last term is (in the problem of minimizing length) recognized as

$$\frac{d}{dt} \langle \sigma'(t) \cdot \sigma'(t) \rangle$$

(221)

In the case that the speed of $\sigma$ is constant, i.e. when $\sigma$ is parameterized proportionally to arc-length then this term vanishes identically, and thus:

**Proposition 4.13** If $\sigma$ is a critical point of the energy functional $E_a^b$, then it is also a critical point of the length functional $L_a^b$.

Conversely, suppose $\sigma$ is a critical point of $L_a^b$, and, in particular, $\sigma'(t) \neq 0$ for all $t \in [a, b]$. Let $\tilde{\sigma}$ be the reparameterization of $\sigma$ by arc-length, i.e.

$$\tilde{\sigma} = \sigma \circ s^{-1} \quad \text{where} \quad s: t \mapsto L_a^t(\sigma|_{[a,t]}) = \int_a^t \langle \sigma'(\tau) \cdot \sigma'(\tau) \rangle^{\frac{1}{2}} d\tau.$$

(222)

The total length $T = L_a^b(\sigma)$ of $\sigma$ clearly is the same as the total length $L_a^T(\tilde{\sigma})$. Thus $\tilde{\sigma}$ is also a critical point of the length functional (but generally for the different interval $[a, t]$). Hence it also satisfies the Euler-Lagrange equations for the length functional – but due to the constant speed the last term (221) vanishes identically. Consequently:

**Proposition 4.14** If $\sigma$ is a critical point of the length functional (and in particular $\sigma'(t) \neq 0$ for all $t$), then $\sigma$ is also a critical point of the energy functional.

These two proposition justify that one generally simply works with the energy functional as opposed to the messier length functional – except for one technicality: A length minimizing curve between two points $p$ and $q$ might conceivably be smooth, but have a corner (at which the speed must take the value zero):

**Proposition 4.15** Suppose $\sigma: [a, b] \mapsto M$ is a piecewise smooth curve with $\sigma(a) = p$ and $\sigma(b) = q$, and $\sigma$ has a corner at $t_0 \in (a, b)$, i.e. the left and right hand limits $\sigma'(t_0-0) \neq \sigma'(t_0+0)$ do not agree. Then for every $\varepsilon > 0$ and all $\delta > 0$ sufficiently small, there exists a curve $\gamma: [a, b] \mapsto M$ which agrees with $\sigma$ on $[a, b] \setminus (t_0 - \delta, t_0 + \delta)$, which is smooth on $[t_0 - \delta, t_0 + \delta]$ and whose length $L_a^b(\gamma)$ is strictly less than $L_a^b(\sigma)$.
Exercise 4.30 Prove the proposition. [[The assertion is rather intuitive – and a direct construction guided by intuition does go through – however there are lots of technical details which provide for a rather long argument. Since the argument is of a local nature it makes sense to work in coordinates, and use the metric \((u^{-1})^*\langle \cdot, \cdot \rangle\) on \(u(U) \subseteq \mathbb{R}^m\). The obvious construction of shortcuts avoiding the corner is straightforward, but the estimates of the length at the ends where the short-cut smoothly joins \(\sigma\) require careful attention to detail.]].

So far it is clear that for every initial point \(p \in M\) and every initial velocity \(X_p\) there exists a unique geodesic \(\sigma\) with \(\sigma(0) = p\) and \(\sigma'(0) = X_p\), which is defined on some interval \((-\varepsilon, \varepsilon)\) with \(\varepsilon > 0\). In order to develop suitable tools to prove e.g. minimality, and to facilitate later characterizations of curvature, it is desirable to shift the variability and obtain descriptions that show more uniformity. The key steps will be to consider open sets of initial points, and to obtain fixed time intervals on which the geodesics are defined (at the cost of limiting the initial velocity). In the end, it is most elegant to only consider the evaluation of geodesics at time \(t = 1\), while letting the initial point and velocity range over some subsets of the tangent bundle \(TM\). We proceed in several small steps.

Theorem 4.16 For every \(p \in M\) there exist a neighborhood \(U\) of \(p\) and \(r_0 > 0\) such that for every \(q \in U\) and every \(X_q \in T_qM\) with \(|X_q| < r_0\) there exists a unique geodesic \(\sigma: (-2, 2) \mapsto M\) such that \(\sigma(0) = q\) and \(\sigma'(0) = X_q\).

Proof. By virtue of the existence and uniqueness theorem for initial value problems there exists for every \(p \in M\) some neighborhood \(U\) of \(p\) and some \(r_0, t_0 > 0\) such that for every \(q \in U\) and every \(X_q \in T_qU\) with \(|X_q| < r_0\) there exists a unique geodesic \(\gamma(X_q, \cdot)\) defined on the interval \((-t_0, t_0)\) such that \(\gamma(X_q, 0) = q\) and \(\gamma'(X_q, 0) = X_q\). Fix any \(\varepsilon < 2t_0r_0\). If \(q \in U\), \(X_q \in T_qM\) with \(|X_q| < \varepsilon\) and \(|t| < 2\), then \(Y_q = \frac{1}{2t_0}X_q\) satisfies \(|Y_q| < \frac{1}{2t_0} \varepsilon < r_0\) and \(|\frac{1}{2t_0} \cdot t| < t_0\). Hence one may define for \(|X_q| < \varepsilon\) and \(|t| < 2\),

\[
\sigma(X_q, t) = \gamma(Y_q, \frac{1}{2t_0} \cdot t). \quad (223)
\]

Definition 4.8 Let \(V \subseteq TM\) be the set of all initial points/velocities \((q, X_q)\) such that there exists a geodesic \(\sigma = \sigma_{(q, X_q)}: [0, 1] \mapsto M\) with \(\sigma(0) = q\) and \(\sigma'(0) = X_q\). Define the map \(\exp: V \mapsto M\), called exponential by \(\exp(X_q) = \exp_q(X_q) = \sigma_{(q, X_q)}(1)\).

Since the geodesic flow as a solution of a smooth initial value problem is smooth, it is clear that \(\exp\) is a \(C^\infty\) map. Since the geodesic equation is a second order differential equation, there is little surprise that the tangent map of \(\exp\) comes into play. It maps a suitable subset of \(T(TM)\) to \(TM\). It is common practice to use abbreviated notation as the \(4m\)-dimensional manifold \(T(TM)\) contains some duplication that is rarely of importance. The following makes an effort to be precise, without being overly pedantic – but it may help to just recognize that difficulties may simply due to notational problems. For any fixed point \(p \in M\) the map \(\exp_p\) maps some subset of \(T_pM\) to \(M\). At any \(X_p \in T_pM\) the tangent map \((\exp_p)'(X_p)\) maps \(T_{T_pM}(T_pM)\) to \(T_{(\exp_p)(X_p)}M\).

In the special case that \(X_p = 0 = 0_p\), clearly \(\exp_p(0_p) = p\) and the tangent map \((\exp_p)'(0_p)\) maps \(T_0_p(T_pM)\) to \(T_pM\).

Lemma 4.17 The map \((\exp_p)'(0_p): T_{0_p}(T_pM) \mapsto T_pM\) is the identity map under the natural identification of the vector spaces \(T_{0_p}(T_pM)\) with \(T_pM\).
Theorem 4.18 \textit{[almost verbatim from Spivak 2nd ed., p.454]}  
For every \( p \in M \) there exist a neighborhood \( W \) and \( \varepsilon_0 > 0 \) such that

- Any two points \( q \) and \( q' \) in \( W \) are joined by a unique geodesic of length less than \( \varepsilon \).
- Let \( \gamma(q, q') \in T_q M \) be the unique tangent vector with \( \gamma(q, q')\| < \varepsilon \) such that \( \exp_q(\gamma(q, q')) = q' \). Then \( \gamma \in C^\infty(\exp W \times W, TM) \).
- For each \( q \in W \) the map \( \exp_q \) maps the open ball \( \{X_q \in T_q M : \|X_q\| < \varepsilon\} \) diffeomorphically onto an open set \( U_q \supseteq W \).

Proof. – not really hard, main difficulty is notational involving objects in \( TTM \). –  
The main observation is that the map \( \Phi \) mapping a neighborhood of \( (p, 0_p) \) in \( TM \) into \( M \times M \) must be one-to-one on some sufficiently small neighborhood. This is immediate from the observation that the tangent map \( F_{(p, 0_p)} \) is the identity from \( T_{(p, 0_p)} TM \) to \( T_{(p, p)} (M \times M) \equiv T_p M \times T_p M \).  
\ldots to be completed \ldots \]

Definition 4.9 Suppose \( p \in M \) and \( r_0 > 0 \) is such that \( \exp_p \) is defined for all \( X_p \in T_p M \) with \( \|X_p\| < r_0 \). Then for \( r < r_0 \) the image of the set \( \{X_p \in T_p M : \|X_p\| = r\} \) under the map \( \exp_p \) is called the geodesic sphere about \( p \) of radius \( r \).

Exercise 4.31 Use a computer algebra system, possibly combined with a numerical package to explore the shape of geodesic spheres for basic manifolds, such as the graphs of \( z = x^2 + y^2 \) and \( z = 1/(1 + x^2 + y^2) \). Pay special attention for which magnitudes of the initial velocities the spheres are imbedded spheres, and for which values they develop foldovers.  
[[Compare the existing MAPLE worksheets and extend the animations provided therein.]]

Proposition 4.19 (Gauss lemma) Suppose \( p \in M \). Then for \( t > 0 \) sufficiently small the geodesics through \( p \) are orthogonal to the submanifolds \( \{\exp_p(X_p) : X_p \in T_p M, \|X_p\| = t\} \).

Proof. Let \( r_0 > 0 \) be so that \( \exp_p \) is defined and one-to-one on the ball \( \{X_p \in T_p M : \|X_p\| \leq r_0\} \). Fix any \( r \in (0, r_0) \) and consider any curve \( \gamma : (-\varepsilon_0, \varepsilon_0) \mapsto T_p M \) (for some \( \varepsilon_0 > 0 \)) with \( \|\gamma(\varepsilon)\| = r \) for all \( \varepsilon \in (-\varepsilon_0, \varepsilon_0) \). Define the map

\[
\Phi : (-\varepsilon_0, \varepsilon_0) \times [0, 1] \mapsto M \quad \text{by} \quad \Phi(\varepsilon, t) = \exp_p(t \cdot \gamma(\varepsilon)).
\] (225)

For each fixed value of \( \varepsilon \) the curve \( t \mapsto \Phi(\varepsilon, t) \) has energy

\[
E_0^1(\Phi(\varepsilon, \cdot)) = \int_0^1 \|D_2 \Phi(\varepsilon, t)\|^2 dt = \int_0^1 r^2 dt = r^2
\] (226)

Considering \( \Phi(\varepsilon, \cdot) \) as a one-parameter family of smooth variations of the geodesic \( \sigma = \Phi(0, \cdot) \), the first variation formula yields

\[
0 = \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} E_0^1(\Phi(\varepsilon, \cdot)) = -\langle D_1 \Phi(0, 1), D_2 \Phi(0, 1) \rangle - \langle D_1 \Phi(0, 0), D_2 \Phi(0, 0) \rangle.
\] (227)
Suppose $\sigma'(r) = D_2 \Phi(0,1) \perp D_1 \Phi(0,1)$. The result follows from the fact that any tangent vector in the tangent space to the submanifold $\{\exp_p(X_p) : X_p \in T_p M, \|X_p\| = t\}$ is of the form $(D_1 \Phi)(0,1)$ for some curve $\gamma$ as above.

**Corollary 4.20** Suppose $p \in M$ and $r_0 > 0$ is such that the $\exp_p$ maps the ball $B_0(r_0)$ diffeomorphically onto some open neighborhood $U$ of $p$. Suppose $\sigma : [a,b] \mapsto U \setminus \{p\}$ is a piecewise smooth curve. Let $r : [a,b] \mapsto [0,r_0)$ and $\theta : [a,b] \mapsto \{X_p \in T_p M : \|X_p\| = 1\}$ be the unique curves such that $\exp(r \theta) = \sigma$. Then $L_a^b \sigma \geq |r(b) - r(a)|$ with equality only if $\theta$ is constant and $r$ is monotonically increasing or monotonically decreasing.

Pictorially, equality means that $\sigma$ agrees (up to reparameterization) with a radial geodesic that connects two geodesic spheres.

**Proof.** — fairly straightforward —

**Corollary 4.21** Suppose $p \in M$ and $r_0 > 0$ is such that the $\exp_p$ maps the ball $B_0(r_0)$ diffeomorphically onto some open neighborhood $U$ of $p$. Suppose $\sigma : [a,b] \mapsto U \setminus \{p\}$ is a piecewise smooth curve. Let $r : [a,b] \mapsto [0,r_0)$ and $\theta : [a,b] \mapsto \{X_p \in T_p M : \|X_p\| = 1\}$ be the unique curves such that $\exp(r \theta) = \sigma$. Then $L_a^b \sigma \geq |r(b) - r(a)|$ with equality only if $\theta$ is constant and $r$ is monotonically increasing or monotonically decreasing.

**Exercise 4.32** — guided explorations with MAPLE (first provide intuitive and technical definitions of conjugate point and envelopes from calculus of variations — then explore them in action. see also maple worksheets —

**Exercise 4.33** Use a combination of a computer algebra system (to generate the code for the geodesic equations) and a numerical differential equations solver (e.g. MATLAB) to animate the evolution of the geodesic spheres, i.e the images of the spheres $\{X_p \in T_p M : \|X_p\| = c\}$ for increasing values of the radius $c$ under the exp map. A good starting point might be the graph of the function $f : (x,y) \mapsto \varepsilon \sin(x) \cdot \sin(y)$ for a suitable parameter $\varepsilon$.

[[First experiments suggest a very rapid development of many foldovers. Choosing smaller values for $\varepsilon$ will produce calmer dynamics. In the end, any such direct approach may soon become impractical and much more advanced methods, e.g. involving viscosity solutions for solving associated partial differential equations might be more appropriate . . . speculations . . . enjoy!]]

**Definition 4.10** A Riemannian manifold $M$ is called geodesically complete if every geodesic can be extended to a geodesic defined on the whole real line $(-\infty, \infty)$.

**Theorem 4.22** (Hopf-Rinow) A Riemannian manifold is geodesically complete if and only if it is complete (w.r.t. its native metric). Any two points $p$ and $q$ in a geodesically complete manifold can be joined by a geodesic of minimal length.

**Proof.** Use exp map and derive contradiction from assuming that there is a finite time beyond which geodesic cannot be extended. Obvious Cauchy sequences in other direction. [[Spivak pp.462–464, Boothby pp.347–350.]]

Work out geodesics on sphere and hyperbolic space (Poincare half-plane — emphasize: constant curvature in calculations)
• Work out geodesics on sphere and hyperbolic space
  (Poincare half-plane – emphasize: constant curvature in calculations)