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 \( \gamma: [a, b] \mapsto \mathbb{R}^n \). 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)

\[
L(\gamma) = \sup_{\mathcal{P}} \sum_{i=0}^{n(\mathcal{P})} \| \gamma(t_{i+1}) - \gamma(t_i) \| \tag{1}
\]

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

\[
\sum_{j=0}^{n} \| \gamma(t_{i+1}) - \gamma(t_i) \| = \sum_{j=0}^{n} \frac{\| \gamma(t_{i+1}) - \gamma(t_i) \|}{(t_{i+1} - t_i)} \cdot (t_{i+1} - t_i) \rightarrow \int_{a}^{b} \| \gamma'(t) \| \, dt =: L(\gamma) \tag{2}
\]

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 \( \gamma: (a, b) \mapsto \mathbb{R}^n \) is called differentiable if for every \( t \in (a, b) \) the limit

\[
\lim_{h \to 0} \frac{1}{h} (\gamma(t + h) - \gamma(t)) \text{ exists. If the limit exists, it is denoted } \gamma'(t) \text{ and called the velocity at } \gamma(t) \text{ (or at } t). \]

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 $\gamma: \mathbb{R} \to \mathbb{R}^2$ 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 $\gamma: \mathbb{R} \to \mathbb{R}^2$ 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 fourth” 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 $\gamma: [a, b] \to \mathbb{R}^n$ is defined as the function $s: [a, b] \to \mathbb{R}$,

$$ s = \psi(t) = \int_a^t \langle \gamma'(\tau), \gamma'(\tau) \rangle^{1/2} d\tau $$(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 $\gamma: [a, b] \to \mathbb{R}^n$ 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 $\gamma: [0, 1) \to \mathbb{R}^2$ 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 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 \( f_0(x) \). Find an explicit formula for \( f_0 \) 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 predetermination of \( N \) 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:

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

To emphasize the characteristic structure of this set of equations, we rewrite it by forming a matrices \( R = (T \mid N \mid B) \) and \( R' = (T' \mid N' \mid B') \) 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

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

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 $\kappa(s)$ and $\tau(s)$ 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 $\sigma(0) \in \mathbb{R}^3$ has been specified, then the Frenet-Serret formulas together with the differential equation $\sigma' = T(s)$ 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 $\sigma(0)$) 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 $\kappa \equiv 0$ 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 $\kappa \equiv c \neq 0$ 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 $\tau \equiv 0$ 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 $\sigma$. 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 $\gamma : I = [a, b] \mapsto \mathbb{R}^n$. Define $\psi(t) = \int_0^t \sqrt{\langle \gamma'(\tau), \gamma'(\tau) \rangle} \, d\tau$. As usual, assume that $\|\gamma'(t)\| > 0$ for all $t \in I$. Then the curve $\sigma : J = [0, L(\gamma)] \mapsto \mathbb{R}^n$ defined by $\sigma = \gamma \circ \psi^{-1}$ is the reparameterization of $\gamma$ by arc-length.

Differentiating $\gamma = \sigma \circ \psi$, the chain rule relates the velocities $\gamma' = (\sigma' \circ \psi) \psi' = \|\gamma\| T \circ \psi$ – i.e. for each $t \in I$, the velocity vector $v(t) = \gamma'(t)$ points in the same direction as $T(\psi(t))$ but it generally has non-unit magnitude (or “speed”) $\|\gamma'(t)\|$. In practical calculations one typically first obtains $\gamma'$, then $\|\gamma'\|$ 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 $\gamma$. The next step is to note that $\gamma'' = (\sigma'' \circ \psi)(\psi')^2 + (\sigma' \circ \psi) \psi''$ implies that for every $t \in I$ the vector $\gamma''(t)$ lies in the plane spanned by $T(\psi(t))$ and $N(\psi(t))$. (This plane is called the osculating plane.) In practical calculations in 3-space one calculates $\gamma''$, then calculates $B$ by normalizing the cross-product $\gamma' \times \gamma''$. 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

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

(6)

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

\[
\kappa \circ \psi = \frac{|a_\parallel|}{\|\gamma'\|^2} = \frac{\|\gamma \times \gamma''\|}{\|\gamma'\|^3} \quad \text{and} \quad \tau \circ \psi = \frac{(\gamma' \times \gamma'') \cdot \gamma'''}{\|\gamma' \times \gamma''\|^2}
\]  

(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 \( \gamma : I \mapsto \mathbb{R}^2 \), 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 \( \kappa \circ \gamma_1 \circ \psi = \gamma_2''/(1 + (\gamma_2')^2)^{3/2} \).

Exercise 1.17 Verify that the curve \( \gamma: \mathbb{R} \mapsto \mathbb{R}^2 \) 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 \( \kappa(s, t) = a_0 + \sum_{j=1}^{N} e^{-j^2t}(a_j \cos jt + b_j \sin jt) \) 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!

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/.
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 \textit{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 \textit{integrate} over (subsets of) manifolds arises naturally. A major role of \textit{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 \textit{geometric} points of view – as a simple example what we don’t want think of the two dimensional sphere \( S^2 \subseteq \mathbb{R}^3 \) 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 \textit{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 \textit{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 \([0, 1] \times [0, 1]\) \textit{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 $\tilde{x} = R x$ and control space $\tilde{u} = S u$, and/or under feedback transformations $\tilde{u} = u + K x$, $\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 $d_2(x, y) = \|y - x\| = \sqrt{(y_1 - x_1)^2 + \ldots + (y_n - x_n)^2}$
- 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 $d(x, y) = 1$ if $x \neq y$ 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 $\bar{d} = d/(1 + d)$.

**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 $(X, \tilde{d})$ if and only if it is open in $(X, \bar{d})$. 
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 $T$ of subsets of $X$ that satisfies

(i) $\emptyset \in T$ and $X \in T$,

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

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

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

A topological space is a pair $(X, T)$ where $X$ is a set and $T$ 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 $O_k = (-\frac{1}{k}, \frac{1}{k})$. 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 $(X, T)$, one often refers to $X$ alone as a topological space. In such cases it is usually understood from the context which topology $T$ 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 $B$ of subsets of $X$ is a basis for a topology on $X$ if it satisfies

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

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

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

**Exercise 2.6** Show that if $Y \subseteq X$ is a subset of a topological space $(X, T)$ then the collection $T' = \{O \cap Y : O \in T\}$ 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 $(X, T)$ 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 $\{B^x_k : k \in \mathbb{Z}^+\}$ such that for every $O \in T$, if $x \in O$ then there exists $k \in \mathbb{Z}$ such $x \in B^x_k \subseteq O$.

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

A topological space $(X, T)$ 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 $O \in T$, if $x \in O$ then $O \cap Y \neq \emptyset$.

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

Every separable first countable 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 \( f : \mathbb{R}^2 \mapsto B^2_0(1) = \{ x \in \mathbb{R}^2 : \|x\| < 1 \} \) 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 \( \{(x, \sin(\frac{1}{x})) \in \mathbb{R}^2 : x \neq 0\} \cup \{0\} \times [-1, 1] \) 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 20\textsuperscript{th} 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 $f : X \to \mathbb{R}$ 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 $f : X \to Y$ 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 $V_\beta \subseteq U_\alpha$,

- $S \subseteq \bigcup_{\beta \in A} V_\beta$, 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 $u: U \mapsto \mathbb{R}^n$.

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 $u^i = x^i \circ u: U \mapsto \mathbb{R}$ are called local coordinates. We use $x^i: \mathbb{R}^n \mapsto \mathbb{R}$, $i = 1, \ldots, n$ 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

$$v \circ u^{-1}: u(U \cap V) \mapsto v(U \cap V) \tag{8}$$

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 $M_{n,n}(\mathbb{R})$.

Further examples are the special linear groups $SL(n, \mathbb{R}) = \{ A \in M_{n,n}(\mathbb{R}) : \det A = 1 \}$, the orthogonal groups $O(n) = \{ A \in M_{n,n}(\mathbb{R}) : A^T A = I \}$, 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 $u(x, y) = \text{atan2}(x, y)$ taking values in $(-\pi, \pi)$ and $r(x, y) = \text{atan2}(x, y)$ 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

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

(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 $u_{\pm}^{-1} : \mathbb{R}^m \mapsto U_{\pm} \subseteq S^m(r)$, derive the formulae

$$x = u_{\pm}^{-1}(y) = \left( \frac{2r}{1 + \| \tfrac{y}{2r} \|^2} \left( \frac{y_1}{2r} \right), \ldots, \frac{2r}{1 + \| \tfrac{y}{2r} \|^2} \left( \frac{y_m}{2r} \right), \mp r \cdot \frac{1 - \| \tfrac{y}{2r} \|^2}{1 + \| \tfrac{y}{2r} \|^2} \right)$$

(10)

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

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 $\mathbb{P}^m = (\mathbb{R}^{m+1} \setminus \{ 0 \}) \slash \sim$, 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)

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

(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 $u_j^{-1}(y_1, \ldots, y_m) = \{ y_1, \ldots, y_{i-1}, 1, y_{i+1}, \ldots, y_m \}$.

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 $f : X \mapsto Y$ is called a quotient map if $O \subseteq Y$ is open if and only if $f^{-1}(O) \subseteq X$ is open. For any map $f : X \mapsto A$ 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 $\{ (x, f(x)) : x \in \mathbb{R}^m \} \subseteq \mathbb{R}^{m+1}$ of any continuous function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is a manifold with a single chart $u : (x, f(x)) \mapsto x \in \mathbb{R}^m$.

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 $F : \mathbb{R}^{m+1} \mapsto \mathbb{R}$, or that are given by parameterizations $F : \mathbb{R}^m \mapsto \mathbb{R}^{m+1}$ and $M = F(\mathbb{R}^m)$. 
More generally, if $M$ and $N$ are manifolds and $\Phi: M \to N$ 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 $f, g: \mathbb{R}^2 \to \mathbb{R}$ 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 $e_i = (0, 0, \ldots, 0, 1, 0, \ldots) \in \mathbb{R}^n$ 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

$$(D_i f)(a) = \lim_{h \rightarrow 0} \frac{1}{h} (f(a + he_i) - f(a)) \quad (12)$$

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

- $f \in C^0(U)$ if $f$ is continuous on $U$ (i.e., $f$ is continuous at every $p \in U$)
- $f \in C^r(U)$ if all partial derivatives of up to order $r$ of $f$ exist and are continuous on $U$,
- $f \in C^\infty(U)$ if all partial derivatives of all orders of $f$ exist and are continuous on $U$,
- $f \in C^\omega(U)$ if $f$ is real analytic on $U$ ($f \in C^\infty(U)$ and $f$ agrees (locally) with its Taylor series).

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 $O \subseteq \mathbb{R}^n$ such that $A \subseteq O$ and an extension $\tilde{f}$ of $f$ to $O$ (i.e., $\tilde{f}|_A = f$), and $\tilde{f} \in C^\alpha(O)$.

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

$$(Df) = (D_j f^i)_{i=1, \ldots, m \atop j=1, \ldots, n} = \left( \begin{array}{ccc} D_1 f^1 & \cdots & D_n f^1 \\ \vdots & \ddots & \vdots \\ D_1 f^m & \cdots & D_n f^m \end{array} \right) \quad (13)$$

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 $(D^k f) \in C^{r-k}(\mathbb{R}^n, \mathbb{R}^{mn})$.

The chain rule asserts that if $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^o$ are open sets, $g \in C^r(U, \mathbb{R}^o)$ 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 $(D_{n+j} f^i(a, b))_{i,j=1,\ldots,m}$ 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 $f: U \subseteq V \mapsto W$ then $f': U \mapsto \text{Hom}(V,W)$. But the space Hom$(V,W)$ of linear maps from $V$ to $W$ is itself a linear space. Hence one may naturally define higher order derivatives as linear maps $f''(U) \mapsto \text{Hom}(V,	ext{Hom}(V,W)) \cong \text{Hom}(V \otimes V,W)$ 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 $v \in V_{1}$, $g'(p) \in \text{Hom}(V_{1},V_{2})$ 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. $\mathcal{D} = \mathcal{D}'$. $\blacksquare$

**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 $u_\alpha$ is one-to-one for each $\alpha \in A$

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

(14)
Since \( v_\beta(V_\beta \cap W) \subseteq \mathbb{R}^n \) is open for each \( \beta \), and \( (v_\beta \circ \Phi \circ u_\alpha^{-1}) \) is continuous for each \( \alpha, \beta \), each of the sets \( (u_\alpha \circ \Phi^{-1} \circ v_\beta^{-1}) \circ v_\beta(V_\beta \cap W) \) 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 \( D_k \) on \( M \) that contains \((u_k, U_k)\). Explain why \( D_k \) are (pairwise) different, give examples of charts contained in their intersections (or explain why the intersections are empty). Demonstrate that \((M, D_k)\) are diffeomorphic.

**Exercise 2.25** (Continuation of exercise 2.24). Consider the maps \( \Phi_{i,j,k} : (M, D_i) \mapsto (M, D_j) \) 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 \( D \) 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 \( D' \) 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

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

(15)

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

\[ h(p) = g\left(\frac{b(b + t)}{(b - a)}\right) \cdot g\left(\frac{b(b - t)}{(b - a)}\right) \cdot \]

(16)

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

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

**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), there exist $r_{y,i} > 0$ for every $y \in u_i(U_i \cap V)$ 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(r_{y,i})): i \leq N_1, j \leq N_2(i)\}, (writing $r_{y,i,j}$ 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

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

(17)

and set

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

(18)

Note that $\phi|_K \equiv 1$ since $K \subseteq \bigcup_{i,j} u_i^{-1}(C_{y_{i,j}}(\frac{3}{4}r_{j}))$. 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, $\mathbb{R}$, which is paracompact, but not compact. Suppose $U = \{(-\alpha, \alpha) : \alpha > 0\}$. 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 $\alpha^{(n)}_1, \ldots, \alpha^{(n)}_{k(n)} \in A$ such that $[n, n+1] \subseteq \bigcup_{j=1}^{k(n)} U_{\alpha^{(n)}_j}$. Define $V^{(n)}_j = U_{\alpha^{(n)}_j} \cap (n-1, n+2)$. Then $\{V^{(n)}_j : n \in \mathbb{Z}, j \leq k(i)\}$ is the desired locally finite refinement that covers $\mathbb{R}$.

The following make implicitly use of some technical properties of manifolds: Specifically, every connected manifold is $\sigma$-compact, meaning that it is a countable union of compact subsets. Thus, every open cover of a $\sigma$-compact space contains a countable subcover.

By definition, every manifold $M$ is also locally compact, meaning that every point $p \in M$ has an open neighbourhood with compact closure. (This is an immediate consequence of coordinate charts being homeomorphisms onto $\mathbb{R}^n$.)

Moreover, as a metric (metrizable) space, every manifold is also paracompact and hence normal. Indeed, even stronger than paracompactness, every open cover $\mathcal{U}$ of a manifold $M$ has a locally finite open refinement $\mathcal{V}$ such that every $V \in \mathcal{V}$ is diffeomorphic to $\mathbb{R}^n$.

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

**Proof.** Without loss of generality assume that $M$ is connected (else the following argument applies to each connected component). Hence we may also assume that the open cover is countable, i.e. $\mathcal{U} = \{U_i : i \in \mathbb{Z}^+\}$

Define $F_1 = M \setminus \left( \bigcup_{i \geq 1} U_i \right)$. 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 $F_1 \subseteq V_1 \subseteq \bar{V}_1 \subseteq U_1$.

Next define $F_2 = U_2 \setminus \left( V_1 \cup \bigcup_{i \geq 2} U_i \right)$. Again, $F_2 \subseteq M$ is closed, $\{V_1\} \cup \{U_i : i \geq 2\}$ is still an open cover of $M$
Let $V$ be an open neighbourhood.

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

We verify that the collection 
\[ \{ \psi_i : i > n(p) \} \] 
for all $i > n(p)$. Consequently, $p \not\in V_i$ for all $i > n(p)$. In other words, $p \in \bigcup_{i=1}^{n(p)} V_i$. 

\[ \text{Theorem 2.7 (Partition of unity)} \] 

Let $U = \{ U_{\alpha} \}_{\alpha \in A}$ be an open cover of a manifold $M$. Then there exist $C^\infty$-functions $\phi_{\alpha} : M \to [0,1]$ such that the collection of closed sets $W_{\alpha} = \{ p \in M : \phi(p) \neq 0 \}$, $\alpha \in A$, 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 $\tilde{U}$ that covers $M$, and relabel $\tilde{U}$ to be $U$.

First consider the special case that $\tilde{U}_\alpha \subseteq M$ 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 $\tilde{U}_\alpha$ the set $\tilde{V}_\alpha$ is again compact, and proposition 2.6 applies. Thus there are maps $\psi_{\alpha} \in C^\infty(M)$ such that $\psi_{\alpha}|_{\tilde{V}_\alpha} \equiv 1$. Moreover, if $Z_{\alpha} = \{ x \in M : \psi_{\alpha}(x) \neq 0 \}$

then $Z_{\alpha} \subseteq U_{\alpha}$. ((Using the notation of the proposition 2.6, $Z_{\alpha} \subseteq W_{\alpha}$, and since each $W_{\alpha}$ is closed, it follows that $\overline{Z}_{\alpha} \subseteq W_{\alpha}$, and hence $\overline{Z}_{\alpha} \subseteq U_{\alpha}$.)

Every point $p \in M$ has an open neighbourhood $O \subseteq M$ 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 $p \in V_{\alpha}$ and thus $\psi_{\alpha}(p) > 0$. Define

\[ \phi_{\alpha} = \frac{\psi_{\alpha}}{\sum_{\beta \in A} \psi_{\beta}}. \] (19)

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 $\text{supp}(\phi_{\alpha}) \subseteq Z_{\alpha} \subseteq U_{\alpha}$.

We now use this special case to prove a strengthened version of proposition, 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 $V(x)$ is compact. For each $x \not\in F$ choose an open neighbourhood $V(x)$ such that $F \cap V(x) = \emptyset$ and such that $\overline{V(x)}$ 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 $\overline{Z(x)}$ 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 $\text{supp}(\phi_x) \subseteq \overline{Z(x)} \subseteq \overline{V(x)} \subseteq M \setminus F$. Consequently, $f|_F = 1 - \sum_{x \not\in F} \phi_x|_F = 1 - 0 = 1$.

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.)