Lecture Notes for MA455 Manifolds

David Mond

March 7, 2008

Contents

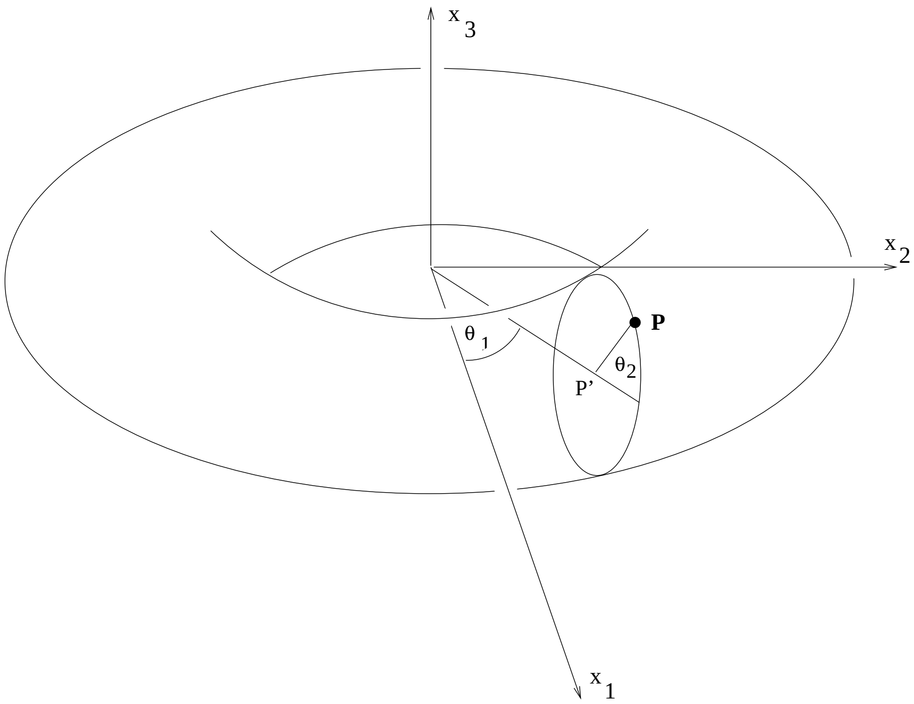
1 Foundations

Check the **Appendices** *at the end of these notes for revision of basic notions of linear algebra and of the derivative of a smooth map.*

1.1 First definitions

Manifolds, as we will first encounter them, are certain subsets of Euclidean space $\mathbb{R}^n$. Later we will introduce a more sophisticated notion of Manifold, which does not require an ambient space to live in.

Example 1.1. Choose real numbers $0 < b < a$. The *torus* $\mathbb{T} = \mathbb{T}_{ab} \subset \mathbb{R}^3$ is shown in the following diagram. It is the set of points a distance b from the circle C_a (not shown) in the x_1x_2 plane with centre 0 and radius a.



Each point $P \in \mathbb{T}$ is determined by the two angles θ_1, θ_2 shown; in terms of these angles its co-ordinates in the ambient $\mathbb{R}^3$ are

$$((a + b\cos\theta_2)\cos\theta_1, (a + b\cos\theta_2)\sin\theta_1, b\sin\theta_2).$$

This formula specifies a bi-periodic map Φ from the $\theta_1\theta_2$-plane to the torus. In fact the torus is the image of the square $[0, 2\pi]\times[0, 2\pi]$ under Φ, or indeed of any square $[c, c+2\pi]\times[d, d+2\pi]$.

A smooth map taking an open set in the plane onto an open set in a surface is called a (smooth) *parametrisation* of that part of the surface.

We can also define $\mathbb{T}$ by an equation: for any point $P \in \mathbb{R}^3 \setminus \{x_3 - \text{axis}\}$ there is a unique closest point P' on C_a. Then $P \in \mathbb{T}$ if and only if $\|P - P'\| = b$. If P has co-ordinates (x_1, x_2, x_3) then P' has co-ordinates

$$\left(\frac{ax_1}{\sqrt{x_1^2 + x_2^2}}, \frac{ax_2}{\sqrt{x_1^2 + x_2^2}}, 0 \right)$$

so $\mathbb{T}$ is the set of points (x_1, x_2, x_3) satisfying

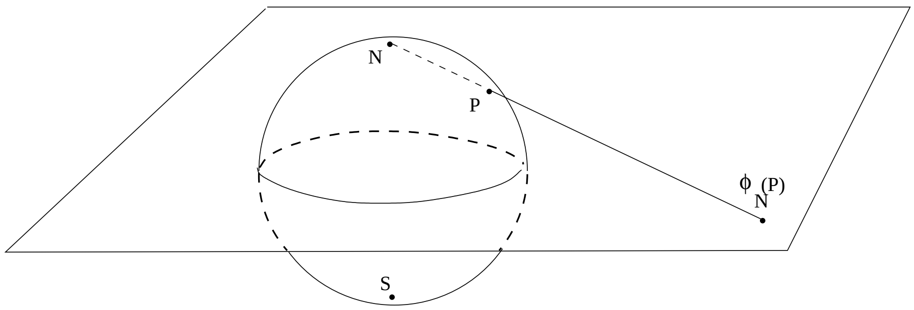
$$\left(x_1 - \frac{ax_1}{\sqrt{x_1^2 + x_2^2}} \right)^2 + \left(x_2 - \frac{ax_2}{\sqrt{x_1^2 + x_2^2}} \right)^2 + x_3^2 = b^2.$$

Notice that this equation only makes sense in $\mathbb{R}^3 \setminus \{x_3 - \text{axis}\}$. □

Example 1.2. The sphere $S^2 \subset \mathbb{R}^3$ is the set of points defined by the equation

$$x_1^2 + x_2^2 + x_3^2 = 1.$$

The north and south poles N and S are the points $(0, 0, 1)$ and $(0, 0, -1)$. *Stereographic projection* $\phi_N : S^2 \setminus \{N\} \to \mathbb{R}^2$ is defined by the following diagram. That is, $\phi_N(P)$ is the point where the line joining N to P meets the equatorial plane.



Evidently $\phi_N(P) = P + \lambda NP$ for some constant λ. The requirement that $\phi_N(P)$ lie in the equatorial plane determines the value of λ: it is $x_3/(1 - x_3)$. Hence

$$\phi_N(P) = (x_1, x_2, x_3) + \frac{x_3}{1 - x_3}(x_1, x_2, x_3 - 1) = \left(\frac{x_1}{1 - x_3}, \frac{x_2}{1 - x_3} \right)$$

(we ignore the last coordinate, 0, of $\phi_N(P)$ in order to regard $s_N(P)$ as a point in $\mathbb{R}^2$). Stereographic projection $\phi_S : S^2 \setminus \{S\} \to \mathbb{R}^2$ is defined by a similar procedure, resulting in the formula

$$\phi_S(x_1, x_2, x_3) = \left(\frac{x_1}{1 + x_3}, \frac{x_2}{1 + x_3} \right)$$

Each is a homeomorphism from an open set in S^2 to the plane. Such maps are called *charts*, in line with the old nautical term: they are planar representations of parts of the surface (in this case the sphere). $\square$

Charts and parametrisations are mutually inverse procedures:

Example 1.1. (again).Given $(x_1, x_2, x_3) \in \mathbb{T}$, we can find (θ_1, θ_2) such that $(x_1, x_2, x_3) = \Phi(\theta_1, \theta_2)$ by means of the formula

$$\theta_1 = \begin{cases} \arctan x_2/x_1 & \text{if } x_1 > 0 \quad , \quad \arctan x_2/x_1 + \pi \quad \text{if } x_1 < 0 \\ \pi/2 - \arctan x_1/x_2 & \text{if } x_2 > 0 \quad , \quad -\pi/2 - \arctan x_1/x_2 \quad \text{if } x_2 < 0 \end{cases}$$

and a slightly more complicated formulae for θ_2. Each point on the torus is in the interior of the domain of a map which gives (θ_1, θ_2) in terms of smooth functions of the coordinates (x_1, x_2, x_3). Each is a homeomorphism from some open set in $\mathbb{T}$ to an open set in the plane. But these charts do not piece together to give a smooth map from all of the torus to some region in $\mathbb{R}^2$.

Example 1.2. (again) **Exercise:** Find a formula for the inverse of $\phi_N : S^2 \setminus \{N\} \to \mathbb{R}^2$. $\square$

Remarks on these examples

1. The charts we've seen in the examples are defined in terms of ambient coordinates, even though we're only interested in their behaviour on the surfaces in question. As functions in the ambient co-ordinates, they are generally defined on open sets in $\mathbb{R}^3$, not on all of $\mathbb{R}^3$.

2. In general we need several charts to cover the whole surface. Their domains will overlap, so we will have several different planar representations of the same portion of surface. In order to extract unambiguous information from these charts we need to understand how the picture in one chart is transformed when we view the same portion of surface in another.

Example 1.2. (yet again): the composite $\phi_S \circ \phi_N^{-1}$ (which maps $\phi_N(\text{domain}(\phi_N) \cap \text{domain}(\phi_S)) = \mathbb{R}^2 \setminus \{0\}$ to $\phi_S(\text{domain}(\phi_N) \cap \text{domain}(\phi_S)) = \mathbb{R}^2 \setminus \{0\}$) is easily seen to be the map

$$(y_1, y_2) \to \frac{1}{y_1^2 + y_2^2}(y_1, y_2).$$

This is "inversion in the unit circle" in the language of classical geometry. It has the interesting property that it preserves angles, in the sense that if the smooth curves C_1 and C_2 in $\mathbb{R}^2 \setminus \{0\}$ meet at the point Q, then their images under $\phi_S \circ \phi_N^{-1}$ meet *with the same angle* at $\phi_S \circ \phi_N^{-1}(Q)$. A map from the plane to the plane with this property is called *conformal.* You can easily check this property here by differentiating $\phi_S \circ \phi_N^{-1}$, and checking that for each point Q and any two vectors u, v, there is a constant $C(Q)$ (depending on Q) such that

$$d_Q(\phi_S \circ \phi_N^{-1})(u) \cdot d_Q(\phi_S \circ \phi_N^{-1})(v) = C(Q)u \cdot v,$$

from which the property of preserving angles follows easily.

We all know how to measure the angle between smooth curves on the sphere, of course. But now imagine that we do not (e.g. the sphere is to big for us to climb onto). A sensible, consistent way of assigning an "angle" to the intersection of two curves C_1, C_2 on the sphere would be to measure the angle between their images in either of the charts ϕ_S, ϕ_N. The crucial point is that it would not matter which we use, because since $\phi_S \circ \phi_N^{-1}$ is conformal, we would get the same answer either way. $\square$

Definition 1.3. (i) Let $U \subset \mathbb{R}^n$ be open. A map $f : U \to \mathbb{R}^p$ is smooth if its partial derivatives of all orders exist and are continuous.
(ii) If $U, V \subset \mathbb{R}^n$ are open then $f : U \to V$ is a *diffeomorphism* if it is smooth and has a smooth inverse.

We need openness of U in (i) to be able to speak of partial derivatives: since

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}$$

(where e_i is the i'th standard basis vector), we need $a + he_i$ to be in U, for all a and small enough h, in order to define the limit, and this is exactly what openness of U guarantees.

Definition 1.3. (continued) (iii) Let $X \subset \mathbb{R}^n$. A map $f : X \to \mathbb{R}^p$ is **smooth** (new) if for each $x \in X$ there is a neighbourhood U of x in $\mathbb{R}^n$ and a smooth map (in the old sense) $F : U \to \mathbb{R}^p$ such that on $U \cap X$, f and F coincide.
(iv) If $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^p$ then a map $f : X \to Y$ is a **diffeomorphism** (new) if it is smooth (in the sense of (iii)) and has a smooth inverse (also in the sense of (iii)).

We will refer to maps which are smooth in the sense of (i) and (iii) as smooth (old) and smooth (new), until it has become clear that our new definition does not clash with the old.

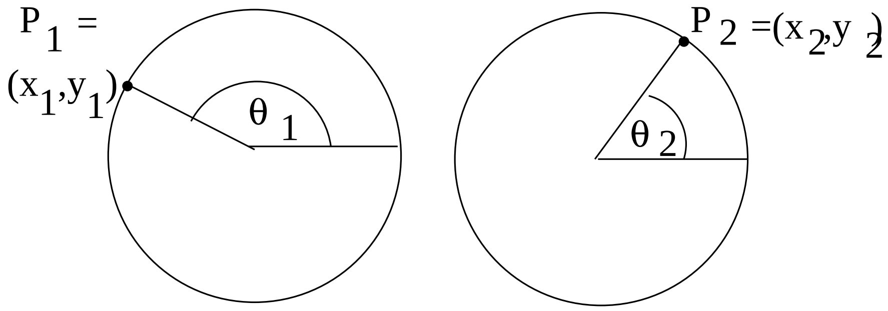
Proposition 1.4. *(i) The composite of two smooth (new) maps is smooth (new).*
(ii) If the domain of a smooth (new) map is in fact open in a Euclidean space, then the map is smooth (old).

Proof If $X \xrightarrow{f} Y \xrightarrow{g} \mathbb{R}^q$ are smooth (new), with X, Y in $\mathbb{R}^n$ and $\mathbb{R}^p$ respectively, suppose $x \in X$. By definition of smoothness (new) of f, there is a neighbourhood U of x in $\mathbb{R}^n$ and a local smooth extension $F : U \to \mathbb{R}^p$ of f — that is, a smooth (old) map such that on $U \cap X$, f and F coincide. Similarly, by the smoothness (new) of g, there is a neighbourhood V of $f(x)$ in $\mathbb{R}^p$ on which g has a local smooth extension G. Then $G \circ F$ is a local smooth extension of $g \circ f$ on U, so $g \circ f$ is smooth (new).

The second statement is obvious from the definition — take $F = f$. $\square$

The purpose of part (ii) here is to make clear that the new definition never clashes with the old.

Example 1.5. The torus $\mathbb{T} \subset \mathbb{R}^3$ is diffeomorphic to $S^1 \times S^1 \subset \mathbb{R}^4$.
Proof



The map $\mathbb{R}^2 \to \mathbb{T}$ introduced in Example 1.1 used the trigonometric functions of the two angles θ_1, θ_2. These can be expressed in terms of the cartesian coordinates of the point $(P_1, P_2) \in S^1 \times S^1$: $\cos\theta_1 = x_1, \sin\theta_1 = y_1, \cos\theta_2 = x_2, \sin\theta_2 = y_2$. Substituting these in the formula for the parametrisation, we get a map $S^1 \times S^1 \to \mathbb{T}$ defined by

$$(x_1, y_1, x_2, y_2) \mapsto ((a + bx_2)x_1, (a + bx_2)y_1, by_2).$$

Exercise Find its inverse and check that it is smooth. $\square$

Definition 1.6. (i) $M \subset \mathbb{R}^n$ is a *smooth manifold of dimension* m if for all $x \in M$ there is a neighbourhood U of x in M, an open set $V \subset \mathbb{R}^m$, and a diffeomorphism $\phi : U \to V$. Such a map ϕ is called a *chart* on M around x.
(ii) A collection of charts whose domains together cover all of M is called an *atlas* on M

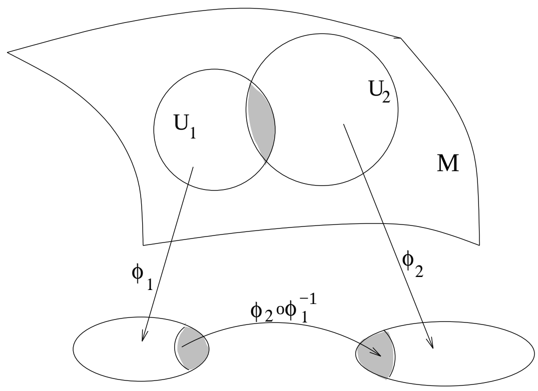
Example 1.7. (i) $\mathbb{T} \subset \mathbb{R}^3$ is a smooth manifold of dimension 2. The locally defined inverses to the periodic parametrisation give charts.
(ii) $S^2 \subset \mathbb{R}^3$ is a manifold of dimension 2 - use the two stereographic projections as charts.
(iii) If $M \subset \mathbb{R}^M$ and $N \subset \mathbb{R}^N$ are manifolds of dimension m and n then $M \times N \subset \mathbb{R}^{M+N}$ is a manifold of dimension $m + n$.
(iv) If $U \subset \mathbb{R}^n$ is open, and $f : U \to \mathbb{R}^p$ is smooth (old) then the graph of f, $\{(x, y) \in U \times \mathbb{R}^p : y = f(x)\}$ is a manifold of dimension n. For the restriction to $\mathrm{gr}(f)$ of the projection $V \times \mathbb{R}^p \to V$ is smooth with smooth inverse the map $x \mapsto (x, f(x))$, and is thus a diffeomorphism. In this case we need only one chart to get an atlas.

In fact, we will see later that every n-dimensional manifold $M \subset \mathbb{R}^N$ is locally the graph of a smooth map mapping an open set in some co-ordinate n-plane to its complementary $N - n$-plane.
(v) If, in (iv), f is not smooth, then the graph of f is not a smooth manifold (Exercise).
(vi) Although this might seem to be stretching a point, a 0-dimensional manifold is a collection of discrete points. For we (perhaps artificially) declare that $\mathbb{R}^0 = \{0\}$ is a single point,

and it follows that the only non-empty open set in $\mathbb{R}^0$ is $\mathbb{R}^0$ itself. Thus, a set of points $X \subset \mathbb{R}^n$ is a 0-dimensional manifold if each point $x \in X$ has a neighbourhood in X which is diffeomorphic to $\mathbb{R}^0$. This means precisely that each point has a neighbourhood in X of which it is the only member, and thus that X is a collection of discrete points.

□

Remark 1.8. Suppose that $\phi_1 : U_1 \to V_1$ and $\phi_2 : U_2 \to V_2$ are charts on the manifold M, and that their domains, U_1 and U_2 have non-empty intersection. Then there is a commutative diagram



in which the map $\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$ is known a the *crossover map* from the chart ϕ_1 to the chart ϕ_2. It is a consequence of 1.4(ii) that all such crossover maps are automatically smooth (old). In the more abstract version of the definition of manifold, which we will meet in Chapter 5, and which you may already have come across elsewehere, we do not insist that our manifolds be contained in some Euclidean space $\mathbb{R}^n$, and we then have to *insist*, as an explicit hypothesis, that in any atlas, all the crossover maps *must* be smooth.

Almost everything one can prove about manifolds at the outset makes use of the Inverse Function Theorem:

Theorem 1.9. *Let U be an open set in $\mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ a smooth map. If the derivative $d_x f : \mathbb{R}^n \to \mathbb{R}^n$ of f is an isomorphism then there are neighbourhoods U_1 of x in $\mathbb{R}^n$ and V_1 of $f(x)$ in $\mathbb{R}^n$ such that $f : U_1 \to V_1$ is a diffeomorphism.* □

Recall that the derivative of f at x, $d_x f$, (when it exists) is a linear map such that $f(x) + d_x f(v)$ is a "reasonable" approximation to $f(x + v)$. More precisely, it is the unique linear map with the property that

$$f(x + v) = f(x) + d_x f(v) + \|v\| E(x, v)$$

for some error term E such that $E \to 0$ as $v \to 0$. A sufficient condition for the existence of the derivative is the existence and continuity of the first order partial derivatives of each of the component functions of f. Thus every smooth map on an open set in $\mathbb{R}^n$ has a derivative at every point.

Recall also the Chain Rule for derivatives:

Theorem 1.10. $d_x(g \circ f) = d_{f(x)}g \circ d_x f$. □

And remember that in the case of a function f of 1 variable, the notion of derivative, in this sense, is linked to the old notation $f'(t)$ as follows:

$$f'(t) = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h} = \lim_{h \to 0} \frac{f(t) + d_t f(h) + |h|E(t,h) \; - \; f(t)}{h} =$$

$$= \lim_{h \to 0} \frac{h d_t f(1) + |h|E(t,h)}{h} = d_t f(1).$$

A similar argument shows that for a function f of n variables,

$$\frac{\partial f}{\partial x_j}(x) = d_x f(e_j),$$

where e_j is the j'th basis vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ (with 1 in the j'th place.) Because of this, we have the following useful formula: suppose that $\gamma : \mathbb{R} \to U \subset \mathbb{R}^n$ is a smooth parametrised curve and $f : U \to \mathbb{R}^p$ is a smooth (old) map. Then if $\gamma(0) = x$, and we let $v = \gamma'(0)$, then

$$d_x f(v) = d_x f(\gamma'(0)) = d_x f(d_0\gamma(1)) = d_0(f \circ \gamma)(1) = (f \circ \gamma)'(0)$$

which we will use a lot in what follows.

In order to use the inverse function theorem we need first a notion of derivative for maps between manifolds.

Definition/Proposition 1.11. Let $M^m \subset \mathbb{R}^n$ be a manifold, and $x \in M$. The *tangent space to M at x*, denoted T_xM, is the space defined by the following two equivalent formulae:
(i) Let $\phi : U \to V \subset \mathbb{R}^m$ be a chart on M around x, and let $a = \phi(x)$. Then

$$T_xM = d_a(\phi^{-1})(\mathbb{R}^m).$$

Note that ϕ^{-1} is smooth (old) and so has an already-defined derivative.
(ii)

$$T_xM = \{\gamma'(0) : \gamma : \mathbb{R}, 0 \to M, x \;\text{ is smooth curve}\}.$$

Observe that neither formulation is completely satisfactory: the first apparently makes use of a choice, and might conceivably depend on that choice, while the second seems impossible to compute, and, unlike the first, has no obvious vector-space structure. On the other hand, each formulation makes up for the other's deficiencies, and
Claim The two formulations are equivalent.
Proof If $\gamma : \mathbb{R}, 0 \to M, x$ is a smooth curve then $\gamma = \phi^{-1} \circ \phi \circ \gamma$. Write $\phi \circ \gamma =: \sigma$. It is a curve in V. By the chain rule we have

$$\gamma'(0) = (\phi^{-1} \circ \sigma)'(0) = d_a(\phi^{-1})(\sigma'(0)).$$

Thus the set defined by the second formula is contained in the set defined by the first.

Conversely, if $v \in \mathbb{R}^m$, we can choose a curve $\sigma : \mathbb{R}, 0 \to V, a$ such that $\sigma'(0) = v$: e.g. $\sigma(t) = a + tv$. If we denote by γ the composite $\phi^{-1} \circ \sigma$, then we have

$$d_a(\phi^{-1})(v) = d_a(\phi^{-1})(\sigma'(0)) = (\phi^{-1} \circ \sigma)'(0) = \gamma'(0).$$

It follows that the set defined by the first formula is contained in the set defined by the second. □

Corollary 1.12. *T_xM is an m-dimensional vector space.*

Proof That it is a vector space follows from the first formulation of the definition. Here is why it is m-dimensional: choose a local smooth extension Φ to ϕ around x. That is, Φ is defined and smooth (old) on some neighbourhood of x in $\mathbb{R}^n$, and on $U \cap M$ Φ and ϕ coincide. Then

$$\Phi \circ \phi^{-1} = \phi \circ \phi^{-1} = 1_V,$$

so

$$d_x\Phi \circ d_a(\phi^{-1}) = d_a 1_V = 1_{\mathbb{R}^m}.$$

Hence $d_a(\phi^{-1})$ is injective, and its image is an m-dimensional vector space. □

Example 1.13. (i) If $U \subset \mathbb{R}^k$ is open then U is a smooth manifold. It is easy to see, by either of the two versions of the definition of the tangent space, that for each $x \in U$, $T_xU = \mathbb{R}^k$. To apply the first version of the definition, just take, as chart on U, the identity map. To apply the second version, let v be any vector in $\mathbb{R}^k$ and differentiate the curve $\gamma(t) = x + tv$.

(ii) If $V \subset \mathbb{R}^n$ is a real vector space then it is linearly isomorphic to some $\mathbb{R}^k$, and is thus a k-dimensional manifold. The last argument in (i) shows that for all $x \in V$, $T_xV = V$.

(iii) Suppose that the surface $S \subset \mathbb{R}^3$ is defined by an equation $h = 0$ (e.g. in the case of the sphere S^2, $h(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 1$). Suppose also that at $x \in S$, $d_xh \neq 0$. Then

$$T_xS = \ker d_xh.$$

For since $d_xh \neq 0$, both sides in this equation are 2-dimensional vector spaces; and the left hand side is contained in the right, for the obvious reason that if $\gamma : \mathbb{R}, 0 \to S, x$ is any curve then $h \circ \gamma$ is constant, so that $0 = (h \circ \gamma)'(0) = d_xh(\gamma'(0))$. □

Definition/Proposition 1.14. Let $f : M \to N$ be a smooth map between manifolds. If $x \in M$, the *derivative of f at x* is the linear map

$$d_xf : T_xM \to T_{f(x)}N$$

defined by any one of the following equivalent formulations:

(i) Let $v \in T_xM$. Choose a curve γ in M such that $\gamma'(0) = v$. Then $d_xf(v) = d_xf(\gamma'(0)) = (f \circ \gamma)'(0)$.

(ii) Choose a smooth (old) extension F of f around x. Then

$$d_xf = d_xF_{|T_xM}.$$

Again, each definition on its own is deficient, in this case because each involves an element of choice. However,
Claim The two versions of the definition coincide.
Proof $d_xF(\gamma'(0)) = (F \circ \gamma)'(0) = (f \circ \gamma)'(0)$ (the last equality because F coincides with f on all of M, and the image of γ lies in M). $\square$

Because of this equality, (i) is independent of the choice of γ such that $\gamma'(0) = v$, and (ii) is independent of the choice of smooth local extension of f, so both are unambiguous definitions. Note also that version (ii) makes clear that if the manifold M is actually an open set in a Euclidean space $\mathbb{R}^n$, (so that for a map with domain M, smooth (new) is the same as smooth (old)), then our new definition of derivative d_xf is the same as the old definition.

Example 1.15. Consider the map $f : S^1 \to S^1$ defined, using a complex co-ordinate z, by $f(z) = z^2$. In real co-ordinates x, y, this is

$$f(x, y) = (x^2 - y^2, 2xy).$$

This formula uses ambient coordinates, and therefore defines a smooth extension of f (in fact to all of $\mathbb{R}^2$). Thus $d_{(x,y)}f$ is the restriction to $T_{(x,y)}S^1$ of the linear map with matrix (with respect to the standard basis of the ambient space)

$$\begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}.$$

Consider the curve γ in S^1 (in fact, parametrising all of S^1, of course!) $\gamma(t) = (\cos t, \sin t)$. The vector $\gamma'(t)$ is a unit-length generator of $T_{\gamma(t)}S^1$. We have

$$d_{\gamma(t)}f(\gamma'(t)) = \begin{pmatrix} 2\cos t & -2\sin t \\ 2\sin t & 2\cos t \end{pmatrix} \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} = \begin{pmatrix} -4\cos t \sin t \\ 2\cos^2 t - 2\sin^2 t \end{pmatrix} = \begin{pmatrix} -2\sin 2t \\ 2\cos 2t \end{pmatrix} = 2\gamma'(2t).$$

That is, the image under $d_{(x,y)}f$ of the unit generator of $T_{(x,y)}S^1$ is twice the unit generator of $T_{f(x,y)}S^1$. Of course, this is not surprising — we know that z^2 goes round the circle twice as fast as z does. $\square$

Proposition 1.16. Chain rule for smooth maps between manifolds
If $M \xrightarrow{f} N \xrightarrow{g} P$ *are smooth maps, then*

$$d_x(g \circ f) = d_{f(x)}g \circ d_xf.$$

Proof Obvious, using definition (i) of the derivative. □

Example 1.17. A *Lie group* is a manifold G which is also a group, for which the operations of

$$\text{multiplication:} \quad \begin{cases} p : G \times G \to G \\ (g_1, g_2) \mapsto g_1 g_2 \end{cases} \qquad \text{and inversion:} \quad \begin{cases} i : G \to G \\ g \mapsto g^{-1} \end{cases}$$

are smooth maps. We can give only rather few examples at this stage: they are

1. $Gl(n, \mathbb{R})$, the set of invertible $n \times n$ real matrices, under matrix multiplication.

2. $\mathbb{R}^2 \setminus \{0\}$, under complex multiplication, and its subgroup S^1, the unit circle.

3. $\mathbb{R}^4 \setminus \{0\}$ under quaternionic multiplication, which is defined as follows: write $(x_0, x_1, x_2, x_3) \in \mathbb{R}^4$ as $x_0\mathbf{1} + x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$ and then define

$$\mathbf{1i} = \mathbf{i},\ \ \mathbf{1j} = \mathbf{j}\ \mathbf{1k} = \mathbf{k}$$

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$$

$$\mathbf{ij} = \mathbf{k},\ \ \mathbf{jk} = \mathbf{i},\ \ \mathbf{ki} = \mathbf{j}$$

with each of the last three products changing sign if the order of the factors is reversed. Extend 'by linearity' to all of $\mathbb{R}^4 \setminus \{0\}$, so, back in the usual coordinate notation, we get

$$(x_0, x_1, x_2, x_3) \cdot (y_0, y_1, y_2, y_3) =$$

$$(x_0y_0 - x_1y_1 - x_2y_2 - x_3y_3, x_0y_1 + x_1y_0 + x_2y_3 - x_3y_2, x_0y_2 + x_2y_0 + x_3y_1 - x_1y_3, x_0y_3 + x_3y_1 + x_1y_2 - x_2y_1).$$

It is straightforward to check that if $x, y \in Q$ then $\|x \cdot y\| = \|x\| \times \|y\|$; as a consequence, the set of *unit quaternions*, $\{x \in Q : \|x\| = 1\}$, is a Lie subgroup. As a manifold, this set is just the sphere S^3.

The following proposition provides a good example of the flexibility of the definition of the derivative.

Proposition 1.18. *Let e denote the neutral element of the Lie group G. Then*
(a) the derivative $d_e p : T_eG \times T_eG \to T_eG$ of the group multiplication is just addition:

$$d_e p(\hat{x}_1, \hat{x}_2) = \hat{x}_1 + \hat{x}_2.$$

(b) The derivative $d_e i : T_eG \to T_eG$ of the inversion map i is just multiplication by -1:

$$d_e i(\hat{x}) = -\hat{x}.$$

Proof (i) We will use part (i) of the definition of the derivative, but we have to apply it with a little ingenuity. To calculate $d_e p(\hat{x}_1, \hat{x}_2)$, we might look for a curve $\gamma = (\gamma_1, \gamma_2) : \mathbb{R} \to G \times G$ such that $\gamma(0) = (e, e)$ and $(\gamma_1'(0), \gamma_2'(0)) = (\hat{x}_1, \hat{x}_2)$. Then by definition of the derivative,

$$d_e p(\hat{x}_1, \hat{x}_2) = (p \circ \gamma)'(0).$$

Now $p \circ \gamma$ is just the product of the two components γ_1 and γ_2 of γ. If we had some information about the value of this product, we might be able to use it to find $(p \circ \gamma)'(0)$. Unfortunately, without detailed information on the particular Lie group G, we cannot in general say anything useful about $\gamma_1 \gamma_2$. But there are cases where the fact that G is a group is enough. If, for example, $\hat{x}_1 = 0$ then we can choose $\gamma_1(t) = e$ for all t. Then $p(\gamma_1(t), \gamma_2(t)) = \gamma_2(t)$ for all t, so $d_e p(\gamma_1'(0), \gamma_2'(0)) = \gamma_2'(t)$. In other words,

$$d_e p(0, \hat{x}_2) = \hat{x}_2.$$

Similarly,

$$d_e p(\hat{x}_1, 0) = \hat{x}_1.$$

Now it is clear how to calculate $d_e p(\hat{x}_1, \hat{x}_2)$:
since the derivative is linear, we know that

$$d_e p(\hat{x}_1, \hat{x}_2) = d_e p(\hat{x}_1, 0) + d_e p(0, \hat{x}_2),$$

and by what has just been said,

$$d_e p(\hat{x}_1, 0) + d_e p(0, \hat{x}_2) = \hat{x}_1 + \hat{x}_2.$$

(ii) To calculate $d_e i(\hat{x}))$ we use the result we have just proved. Suppose that γ is a curve in G with $\gamma(0) = e$ and $\gamma'(0) = \hat{x}$. By definition of p and i,

$$p\big(i(\gamma(t)), \gamma(t)\big) = e$$

for all t. By the Chain rule,

$$0 = \frac{d}{dt} p\big(i(\gamma(t)), \gamma(t)\big)_{|t=0} = d_e p\big((i \circ \gamma)'(0), \gamma'(0)\big)$$

and by (i) this is equal to

$$(i \circ \gamma)'(0) + \gamma'(0),$$

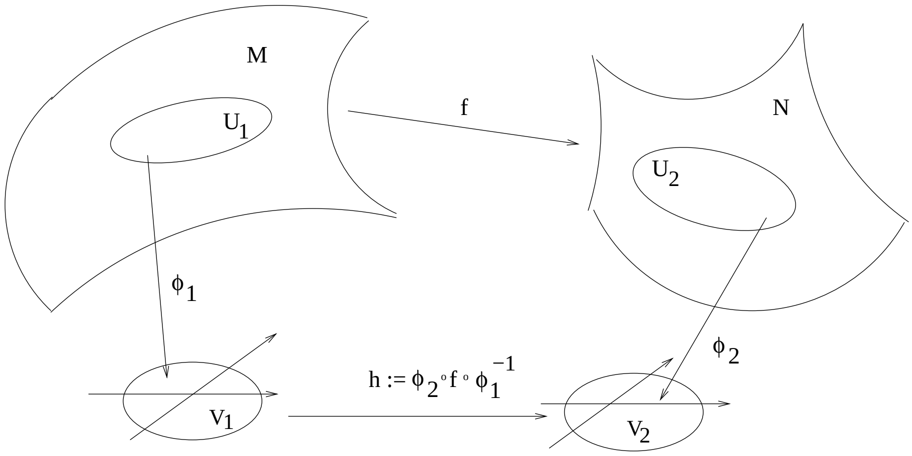
i.e. to

$$d_e i(\hat{x}) + \hat{x}.$$

It follows that $d_e i(\hat{x}) = -\hat{x}$. $\square$

Theorem 1.19. (Inverse Function Theorem for Manifolds). *Suppose that $f : M^m \to N^m$ is a smooth map between manifolds, and suppose also that $d_x f : T_x M \to T_{f(x)} N$ is an isomorphism. Then there are neighbourhoods W_1 of x in M and W_2 of $f(x)$ in N such that $f(W_1) = W_2$ and $f : W_1 \to W_2$ is a diffeomorphism.*

Proof We draw the standard diagram

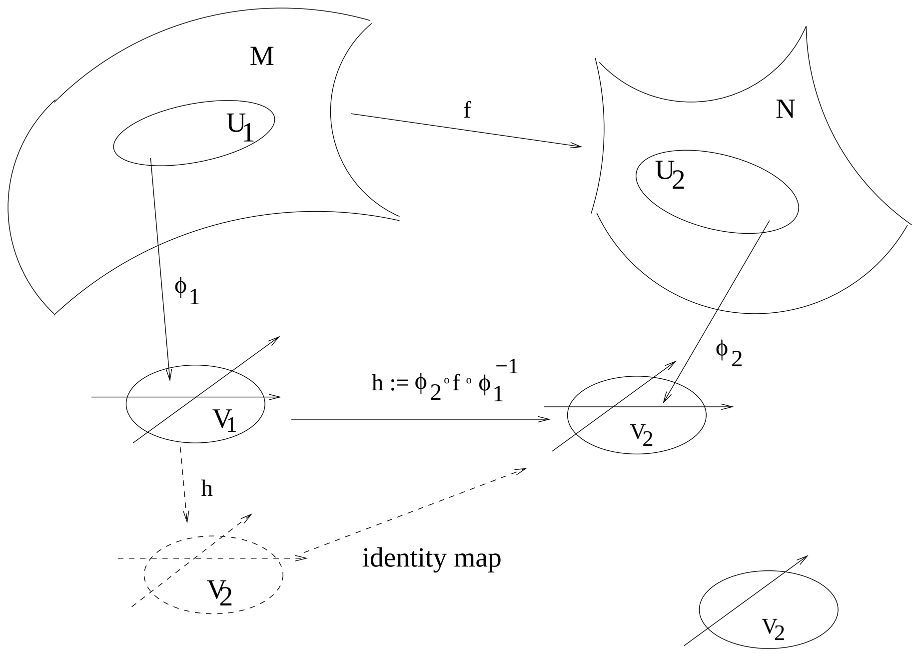


in which ϕ_1 and ϕ_2 are charts on M and N around x and $f(x)$ respectively. Write $a = \phi_1(x)$.
Typical Step 1: use charts to turn the hypothesis on the smooth (new) map into a hypothesis on a smooth (old) map
Commutativity of this diagram implies commutativity of the diagram

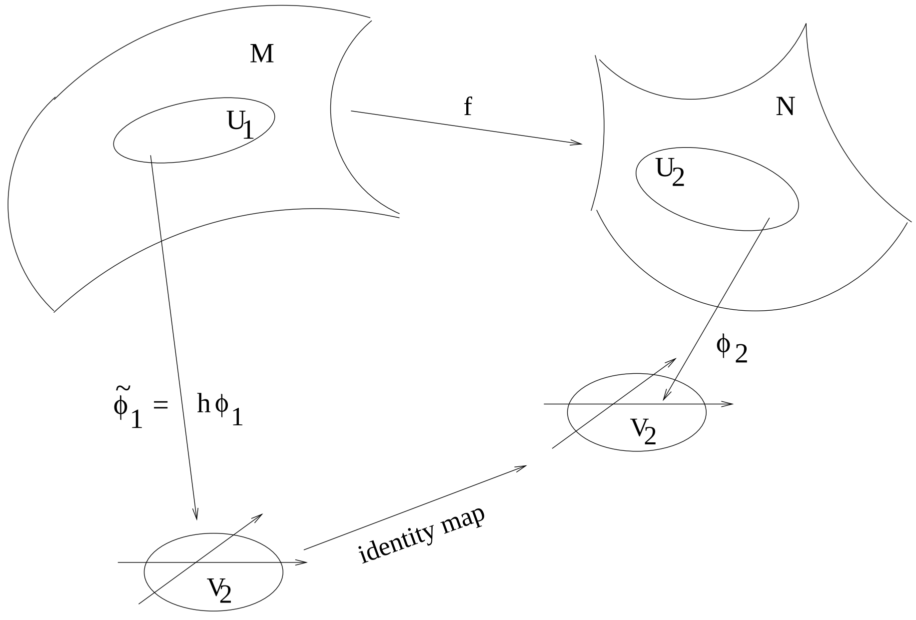
$$\begin{array}{ccc} T_xM & \xrightarrow{d_xf} & T_{f(x)}N \\ d_x\phi_1 \downarrow & & \downarrow d_{f(x)}\phi_2 \\ \mathbb{R}^m & \xrightarrow{d_ah} & \mathbb{R}^m \end{array}$$

by the chain rule 1.16. The vertical maps here are isomorphisms, and therefore because d_xf is an isomorphism, so is d_ah.
Typical Step 2: Apply an old theorem
By the inverse function theorem (old), there are neighbourhoods Z_1 of a in $\mathbb{R}^m$ and Z_2 of $h(a)$ in $\mathbb{R}^m$, such that $h : Z_1 \to Z_2$ is a diffeomorphism. Call its inverse g.
Typical Step 3: Use charts to turn the conclusion about the smooth (old) map into a conclusion about the smooth (new) map.
Let $W_1 = \phi_1^{-1}(Z_1), W_2 = \phi_2^{-1}(Z_2)$. Then $f(W_1) = W_2$ and $f : W_1 \to W_2$ has inverse $\phi_1^{-1} \circ g \circ \phi_2$, and so is a diffeomorphism. $\square$

Remark 1.20. The inverse function theorem can be viewed as saying that if d_xf is an isomorphism then *with respect to suitable coordinates on M around x and on N around $y = f(x)$, f is nothing but the identity map.* To see why, consider the diagram in the proof of 1.19. If we add to it the extra maps indicated with dashed lines,



and then rub out some of the maps and spaces, we are left with



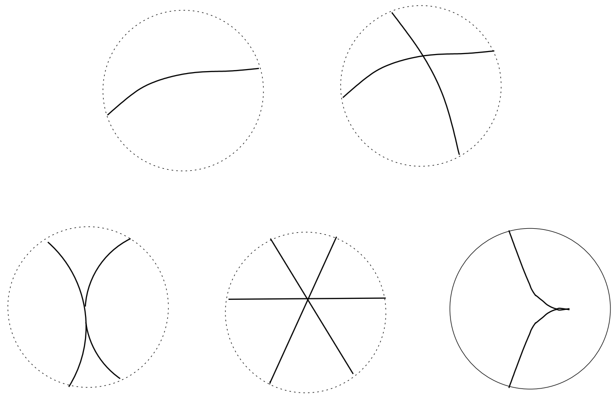
The expression of f with respect to the new coordinates on M provided by the chart $\widetilde{\phi}_1 = h \circ \phi_1$, and the old coordinates on N, is simply the identity.

1.2 Submersions and Immersions

We begin with a geometrical example.

Example 1.21. Consider a curved and twisted piece of wire. Suppose it does not have any kinks, and does not touch itself. If we neglect its thickness, we can think of it as a one-dimensional manifold contained in $\mathbb{R}^3$.

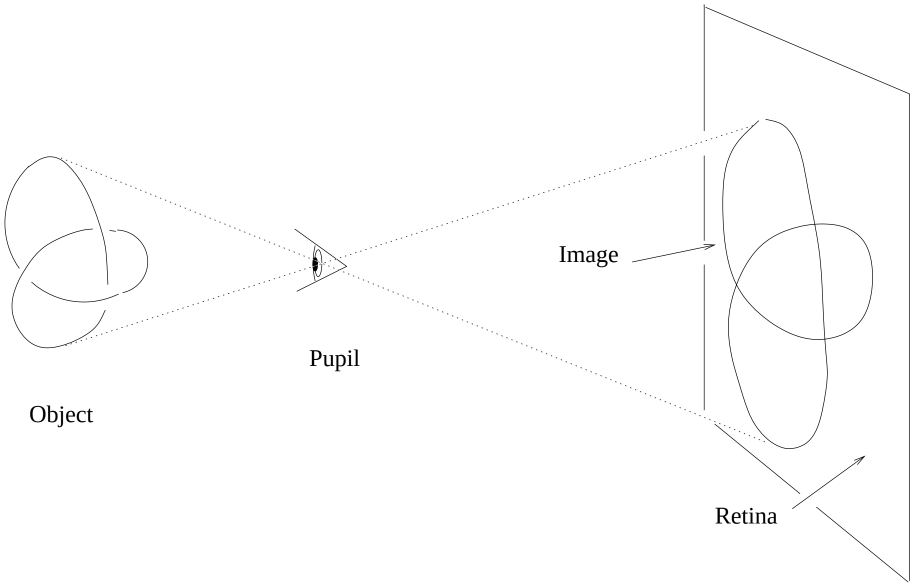
By looking at the wire from different positions, we see a number of qualitatively different "local pictures", some of which are shown below. The local pictures on the first row can be seen from all points in space, or at least from the points of an open set; those on the second row can only be seen from the points of certain surfaces. Other more complicated local pictures may be seen from certain curves, or even from isolated points.



All of the local pictures shown are easy to understand except possibly the last, the cusp. How can a smooth curve appear to acquire a sharp point when viewed from certain positions? And from which positions do we see such a view?

If you have a piece of wire to hand, even a loosely coiled spring, it is a good idea to experiment at this point, turning it over and examining it from different points of view.

It is not hard to find, by experiment, that we see a cusp by looking at the curve along one of its tangent lines. The set of all the tangent lines together make up the *tangent developable surface* of the curve, and this is the set of viewpoints from which the cusp may be seen. To explain this, we simplify the geometry of vision as shown in the following diagram.



If the eye is situated at the origin of coordinates and the retina is the plane $\{y = -h\}$, then the point on the object with coordinates (x, y, z) has image

$$-\frac{h}{y}(x, y, z).$$

Taking coordinates (x, z) on the retinal plane, our simplified version of vision is thus the map

$$V : (x, y, z) \mapsto -h(x/y, z/y),$$

which is well defined, and smooth, on $\mathbb{R}^3 \setminus \{y = 0\}$. Its derivative $d_{(x,y,z)}V$ has matrix

$$-h\begin{pmatrix} 1/y & -x/y^2 & 0 \\ 0 & -z/y^2 & 1/y \end{pmatrix}$$

Now suppose that $\gamma : (a, b) \to \mathbb{R}^3$ parametrises the curve C. The image of C is parametrised by $V \circ \gamma$. Its derivative

$$(V \circ \gamma)'(t) = d_{\gamma(t)}V(\gamma'(t))$$
$$= -h\Big(\frac{\gamma_1'(t)}{\gamma_2(t)} - \frac{\gamma_1(t)\gamma_2'(t)}{\gamma_2(t)^2}, \frac{\gamma_3'(t)}{\gamma_2(t)} - \frac{\gamma_3(t)\gamma_2'(t)}{\gamma_2(t)^2}\Big)$$

can be expressed as

$$\frac{-h}{\gamma_2(t)^2}\Big(\begin{vmatrix} \gamma_1'(t) & \gamma_2'(t) \\ \gamma_1(t) & \gamma_2(t) \end{vmatrix}, \begin{vmatrix} \gamma_3'(t) & \gamma_2'(t) \\ \gamma_3(t) & \gamma_2(t) \end{vmatrix}\Big).$$

Since we are assuming that $\gamma_2(t) \neq 0$, the vanishing of the two determinants in this expression is equivalent to $\gamma(t)$ and $\gamma'(t)$ being parallel. In other words, *the composed parametrisation* $V \circ \gamma$ *has zero derivative precisely when we look at the curve* C *along its tangent vector.*

All the other local views, in which the image of C is made up of (possibly superimposed) pieces of smooth curve, are seen from other viewpoints. □

Now put this calculation together with the experimental evidence that we see a cusp at P when we look along the tangent line to the curve at P, and that if we don't look along the tangent line then the curve appears smooth. It suggests that $V(C)$ is a smooth manifold in the neighbourhood of $V(P)$ provided that $d_P V$ is 1-1 on $T_P C$. The following definition and theorems show that this is the case. We introduce not only the definition of *immersion*, a map whose derivative at each point is injective, but also the dual notion of *submersion*, where the derivative at each point is surjective.

Exercise 1.22. Suppose that V and W are vector spaces of finite dimensions m and n respectively, and that $A : V \to W$ is a linear map.
(i) Show that if A is injective, then with respect to suitable bases E of V and F of W,

$$[A]_F^E = \begin{pmatrix} I_m \\ 0 \end{pmatrix}.$$

(ii) Show that if A is surjective, then with respect to suitable bases E of V and F of W,

$$[A]_F^E = \begin{pmatrix} I_n & 0 \end{pmatrix}.$$

Definition 1.23. *Let $f : M \to N$ be a smooth map. It is an* **immersion** *at x if $d_x f$ is injective, and a* **submersion** *at x if $d_x f$ is surjective.*

In terms of matrices, the two conditions greatly resemble each other: if $m \leq n$ then the linear map $\mathbb{R}^m \to \mathbb{R}^n$ defined by a matrix A is injective if it has rank m; that is, if its m columns are linearly independent. If $m \geq n$ then A is surjective if it has rank n; that is, if, among its m columns, some n make up a basis for $\mathbb{R}^n$. In both cases, the condition holds *if an only if the matrix A has maximal rank — the biggest rank it can have, given its dimensions.*

Example 1.24. The **standard immersion** $\mathbb{R}^m \to \mathbb{R}^{m+k}$ is the map

$$(x_1, \ldots, x_m) \mapsto (x_1, \ldots, x_m, 0, \ldots, 0).$$

The **standard submersion** $\mathbb{R}^{m+k} \to \mathbb{R}^m$ is the map

$$(x_1, \ldots, x_{m+k}) \mapsto (x_1, \ldots, x_m).$$

$\square$

These are the maps defined by the matrices of Exercise 1.22.

Example 1.25. Consider the torus $\mathbb{T}$ situated in $\mathbb{R}^3$ as shown, and let $f : \mathbb{T} \to \mathbb{R}$ be the third coordinate function — i.e. f is the restriction to $\mathbb{T}$ of the orthogonal projection to the x_3 axis. Let us denote the orthogonal projection $\mathbb{R}^3 \to \mathbb{R}$ by F. To calculate $d_x f : T_x\mathbb{T} \to T_{f(x)}\mathbb{R} = \mathbb{R}$, we use two key facts:

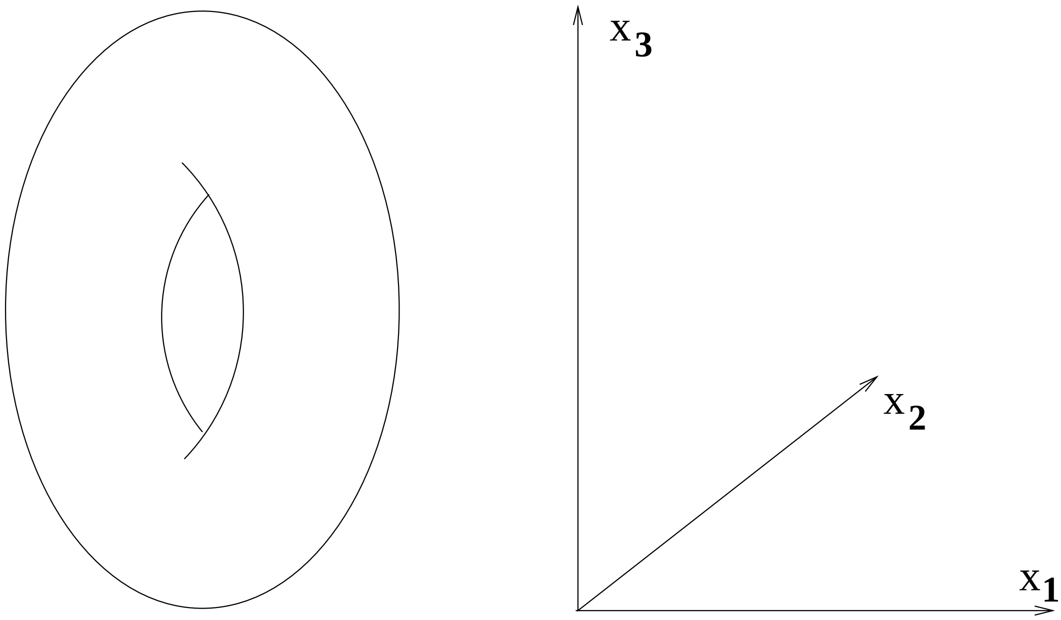
- By version (ii) of the definition of the derivative, $d_x f = d_x F$,

- As F is a linear map, $d_xF = F$.

It follows that

$$d_xf : T_x\mathbb{T} \to \mathbb{R} \text{ is just the orthogonal projection from } T_x\mathbb{T} \text{ to the } x_3\text{-axis.}$$

Thus f is a submersion at each point $x \in \mathbb{T}$ where the tangent plane is not horizontal. There are just four points on $\mathbb{T}$ where this fails.



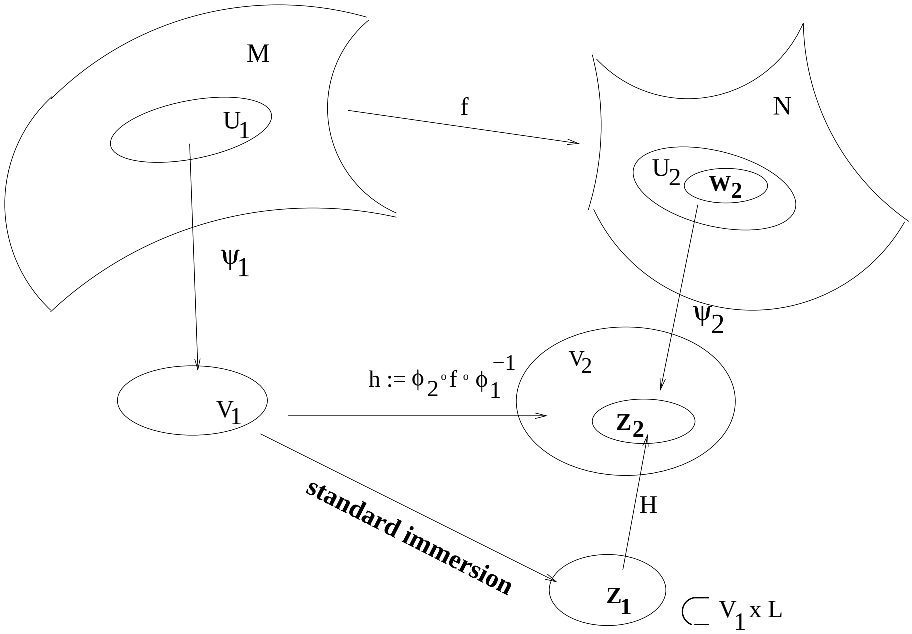
Theorem 1.26. Local normal form for immersions *If $f : M \to N$ is an immersion at x then there are charts $\phi_1 : U_1 \to V_1$ on M around x and $\phi_2 : U_2 \to V_2$ on N around $y := f(x)$ such that $\phi_2 \circ f \circ \phi_1^{-1}$ is the standard immersion.*

Proof Begin with any two charts $\psi_1 : U_1 \to V_1$ and $\psi_2 : U_2 \to V_2$ on M and N around x and y respectively. Write $a = \psi_1(x), b = \psi_2(y)$. Because f is an immersion at x, so is $h := \psi_2 \circ f \circ \psi_1^{-1}$ (by Typical Step 1).

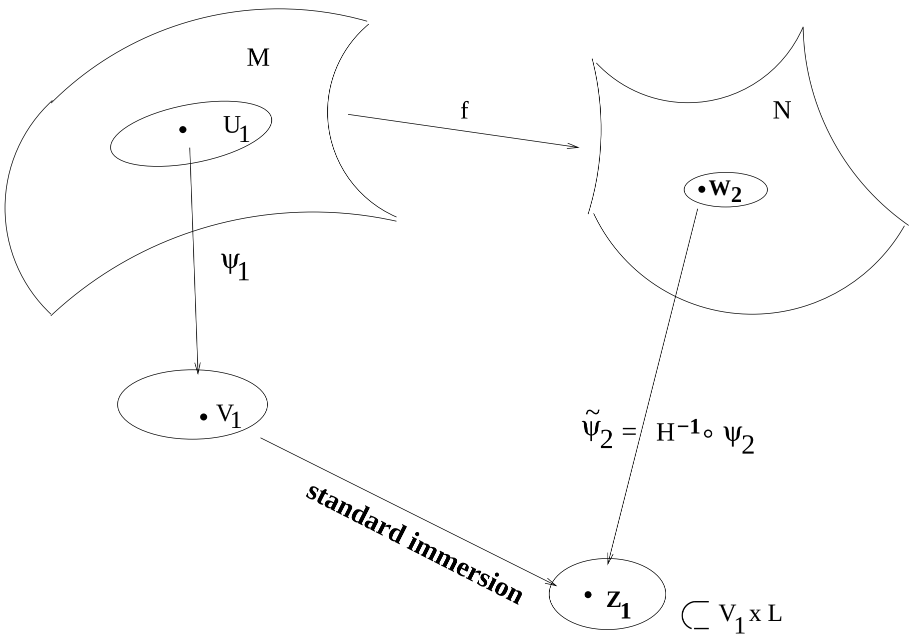
Let $L \subset \mathbb{R}^n$ be an $n - m$-dimensional vector subspace which meets $d_ah(\mathbb{R}^m)$ only at 0. Then $d_ah(\mathbb{R}^m) + L = \mathbb{R}^n$. Define $H : V_1 \times L \to \mathbb{R}^n$ by

$$H(x, v) = h(x) + v.$$

Then $d_{(a,0)}H$ is surjective, since its image contains $d_ah(\mathbb{R}^m)$ and L. Hence it is an isomorphism. By the inverse function theorem there are neighbourhoods Z_1 of $(a, 0)$ in $V_1 \times L$ and Z_2 of $b = H(a, 0)$ in V_2 such that $H : Z_1 \to Z_2$ is a diffeomorphism.



Since $h = H \circ$ standard immersion, we let $W_2 = \psi_2^{-1}(Z_2)$ in N and take as new chart on N the map $\tilde{\psi}_2 = H^{-1} \circ \psi_2 : W_2 \to Z_1$. It follows that $\tilde{\psi}_2 \circ f \circ \psi_1^{-1} =$ standard immersion, as required.



$\square$

An analogous result is true for submersions.

Theorem 1.27. Local normal form for a submersion. *If $f : M \to N$ is a submersion at $x \in M$ then there are charts ϕ_1 on M around x and ϕ_2 on N around $y = f(x)$ such that $\phi_2 \circ f \circ \phi_1^{-1}$ is the standard submersion.*

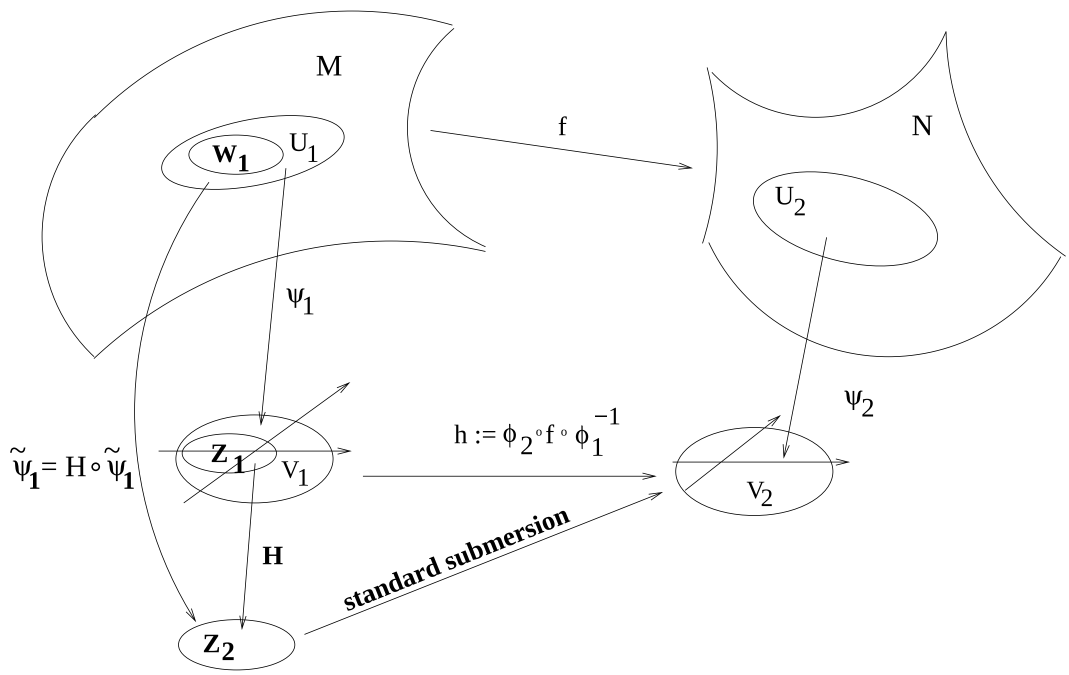
Proof Begin with any two charts $\psi_1 : U_1 \to V_1$ and $\psi_2 : U_2 \to V_2$ on M and N around x and y respectively. Write $a = \psi_1(x), b = \psi_2(y)$. Because f is a submersion at x, so is $h := \psi_2 \circ f \circ \psi_1^{-1}$ (this is Typical Step 1).

Typical Step 2: Let $K \subset \mathbb{R}^m$ be the kernel of $d_a h$, and let $\pi : V_1 \to K$ be a linear projection. As $d_a h$ is surjective, $\dim K = m - n$. Define $H : V_1 \to V_2 \times L$ by

$$H(x) = (h(x), \pi(x)).$$

Then $d_a H$ is injective, and hence an isomorphism: for if $d_a H(v) = 0$ then $d_a h(v) = 0$ and $d_a \pi(v) = 0$, so in particular $v \in K$; but this means $d_a \pi(v) = v$, since π is a linear projection. Hence $v = 0$. It follows from the inverse function theorem (old) that H is a diffeomorphism from some neighbourhood Z_1 of a in V_1 to some neighbourhood Z_2 of $H(a)$ in $V_2 \times K$.

We note that $h =$ standard submersion $\circ\, H$. So take as new chart on M around x the map $\tilde{\psi}_1 := H \circ \psi_1 : \psi_1^{-1}(Z_1) \to Z_2$.



Thus,

$$\psi_2 \circ f \circ \tilde{\psi}_1^{-1} = \text{standard submersion}$$

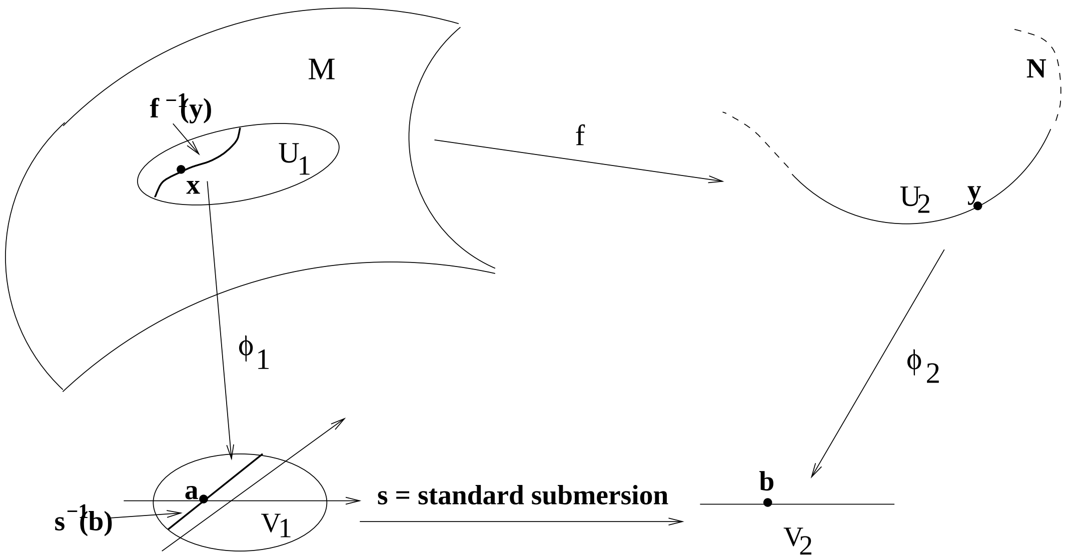
by the commutativity of all these diagrams of mappings. □

We will explore the consequences of 1.26 and 1.27 in the rest of this chapter.

Definition 1.28. Let $f : M \to N$ be a smooth map of manifolds. The point $x \in M$ is a *regular point* of f if $d_x f$ is surjective, and a *critical point* otherwise. The point $y \in N$ is a *regular value* if every preimage $x \in f^{-1}(y)$ is a regular point, and a *critical value* otherwise. Thus, y is a critical value if and only if it is the image of a critical point. In particular, if $f^{-1}(y) = \emptyset$ then y is a regular value.

Corollary 1.29. (of 1.27) *If $f : M^m \to N^n$ is a smooth map of manifolds and $y \in N$ is a regular value with $f^{-1}(y) \neq \emptyset$ then $f^{-1}(y)$ is a manifold of dimension $m - n$, and for any $x \in f^{-1}(y)$, $T_x f^{-1}(y) = \ker d_x f$.*

Proof Let $x \in f^{-1}(y)$. Then f is a submersion at x. Coose charts as in 1.27. Denote the standard submersion $\mathbb{R}^m \to \mathbb{R}^n$ by s. Evidently $s^{-1}(b) \cap V_1 = (\{b\} \times \mathbb{R}^{m-n}) \cap V_1$, and this is diffeomorphic to an open set in $\mathbb{R}^{m-n}$ (just project to lose the b). Moreover ϕ_1 defines a diffeomorphism $f^{-1}(y) \cap U_1 \simeq s^{-1}(b) \cap V_1$. It follows that x has a neighbourhood $f^{-1}(y) \cap U_1$ in $f^{-1}(y)$ which is diffeomorphic to an open set in $\mathbb{R}^{m-n}$. Thus $f^{-1}(y)$ is a manifold of dimension $m - n$.

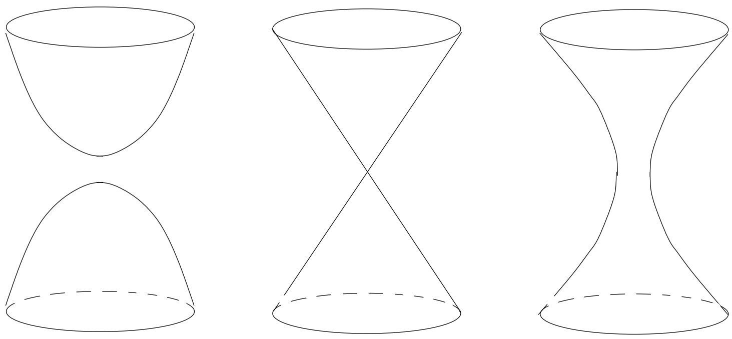


For the statement about the tangent space, let γ be a smooth curve in $f^{-1}(y)$ with $\gamma(0) = x$. Then $f \circ \gamma$ is constant, so $d_x f(\gamma'(0)) = (f \circ \gamma)'(0) = 0$, and $\gamma'(0) \in \ker d_x f$. Thus

$$T_x f^{-1}(y) \subset \ker d_x f.$$

Both ker $d_x f$ and $T_x f^{-1}(y)$ have dimension $m-n$, so they must be equal. $\square$

Example 1.30. (i) Consider the function $f:\mathbb{R}^3 \to \mathbb{R}$, $f(x_1,x_2,x_3)=x_1^2+x_2^2-x_3^2$. The only critical point of f is $(0,0,0)$, so the only critical value is 0. If $t<0$ or $t>0$ then $f^{-1}(t)$ is a smooth manifold of dimension 2, by Theorem 1.29. But $f^{-1}(0)$ is not a smooth manifold at the critical point of f.



t<0 t=0 t>0

If we think of t as time and view this sequence of pictures with time in reverse, it resembles a drip of water in the moments before and after it separates from the water on the edge of the tap and falls.

(ii) The matrix group $O(n)$ is the set of $n\times n$ real matrices A satisfying $A^tA=I_n$. Thus, $O(n)$ is the preimage of the point I_n under the map $\mathrm{Mat}_{n\times n}(\mathbb{R})\to \mathrm{Mat}_{n\times n}(\mathbb{R})$, where $\mathrm{Mat}_{n\times n}(\mathbb{R})$ is the space of real $n\times n$ matrices. Thinking of $\mathrm{Mat}_{n\times n}(\mathbb{R})$ as the same as $\mathbb{R}^{n^2}$, and thus a manifold, we can ask whether I_n is a regular value of f. The answer is clearly no, since if it were a regular value then its preimage $O(n)$ would have dimension equal to dimension(source)-dimension(target)= 0. As $O(n)$ plainly is not just a collection of isolated points, this cannot be the case. In fact, however, for every A the matrix A^tA is *symmetric*, and so we can redesignate f as a map $\mathrm{Mat}_{n\times n}(\mathbb{R})\to \mathrm{Sym}_n(\mathbb{R})$, where $\mathrm{Sym}_n(\mathbb{R})$ is the space of $n\times n$ symmetric real matrices. Now I_n *is* a regular value. To see this we first compute the derivative d_Af, or rather, its value on a matrix $B\in T_A\mathrm{Mat}_{n\times n}(\mathbb{R})$ (note that since $\mathrm{Mat}_{n\times n}(\mathbb{R})$ is a vector space, its tangent space at any point is canonically identified with $\mathrm{Mat}_{n\times n}(\mathbb{R})$ itself). We have

$$d_Af(B)=\lim_{h\to 0}\frac{f(A+hB)-f(A)}{h}=\lim_{h\to 0}\frac{(A+hB)^t(A+hB)-A^tA}{h}$$

$$=\lim_{h\to 0}\frac{h(A^tB+B^tA)+h^2B^tB}{h}=A^tB+B^tA.$$

We have to show that for each $A\in O(n)$ and each matrix $S\in T_{I_n}\mathrm{Sym}_n(\mathbb{R})$, there exists a matrix B such that $d_Af(B)=S$. $\mathrm{Sym}_n(\mathbb{R})$ is once again a vector space, so S here is any symmetric matrix. Solutions to this equation will not be unique - in fact we expect a $1/2n(n-1)$ dimensional affine space of solutions. But here is one: let $B=1/2AS$. Then $A^tB=1/2A^tAS=1/2S$, and $B^tA=1/2S^tA^tA=1/2S^t=1/2S$. Hence $A^tB+B^tA=S$

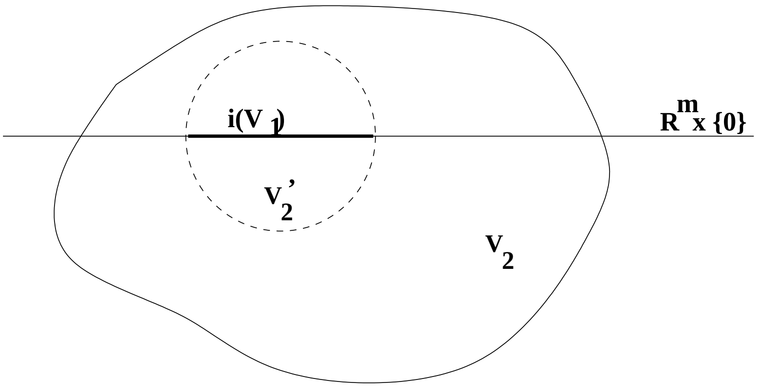
as required. This shows that I_n is a regular value of f, and thus that $O(n)$ is a manifold of dimension equal to $n^2 - \dim \mathrm{Sym}_n(\mathbb{R}) = n(n-1)/2$.

The group $O(n)$ is contained in the group $Gl(n, \mathbb{R})$ of invertible $n \times n$ matrices, which is an open set in the Euclidean space $\mathrm{Mat}_{n\times n}(\mathbb{R})$. The operations of matrix multiplication and matrix inversion are smooth maps $Gl(n, \mathbb{R}) \times Gl(n, \mathbb{R}) \to Gl(n, \mathbb{R})$ and $Gl(n, \mathbb{R}) \to Gl(n, \mathbb{R})$. Therefore they are smooth (new) on $O(n)$, and $O(n)$ is a Lie group.

$\square$

Corollary 1.31. of 1.26 *If $M^m \subset N^n$ are manifolds then for each point $x \in M$ there is a neighbourhood U of x in N and a smooth map $g : U \to \mathbb{R}^{n-m}$ such that* 0 *is a regular value of g and $M \cap U = g^{-1}(0)$.*

Proof Apply 1.26 to the inclusion $f : M \hookrightarrow N$ to get charts $\phi_1 : U_1 \to V_1 \subset \mathbb{R}^m$ on M around x and $\phi_2 : U_2 \to V_2 \subset \mathbb{R}^n$ on N around $f(x) = x$, so that $\phi_2 \circ f \circ \phi_1^{-1}$ is the standard immersion i. Now $i(V_1) \subset V_2 \cap (\mathbb{R}^m \times \{0\})$, but this inclusion is not necessarily an equality. On the other hand, because V_1 is open in $\mathbb{R}^m$, $i(V_1)$ is open in $\mathbb{R}^m \times \{0\}$, and so there exists an open set V_2' in $\mathbb{R}^n$, contained in V_2, such that $i(V_1) = V_2' \cap (\mathbb{R}^{n-m} \times \{0\})$. [1]



In V_2', $i(V_1)$ is the set of point where $x_{m+1}, \ldots, x_n$ all vanish. So take $U = \phi_2^{-1}(V_2')$, and take, as $g : U \to \mathbb{R}^{n-m}$, the map $(x_{m+1} \circ \phi_2, \ldots, x_n \circ \phi_2)$. The maps

$$(x_1, \ldots, x_n) \mapsto (x_{m+1}, \ldots, x_n)$$

and

$$y \in U \mapsto \phi_2(y)$$

are both submersions, and hence so is their composite, g. $\square$

[1]Editorial comment: Actually this is the "trivial" statement that the subspace topology on $\mathbb{R}^m \times \{0\} \subset \mathbb{R}^n$ coincides with the topology that $\mathbb{R}^m \times \{0\}$ inherits from its identification with $\mathbb{R}^m$.

The components of a map $g : U \to \mathbb{R}^{n-m}$ such that $M \cap U = g^{-1}(0)$ are equations for M in U; if in addition 0 is a regular value of g, then g is a set of *regular local equations for M in U*. So 1.31 establishes that every submanifold $M \subset N$ has regular local equations everywhere.

Remark 1.32. Let $M^m \subset N^n$. The number of *regular* equations needed to define M locally in N is $n - m$. If $n - m > 1$, it is possible to define M with fewer equations - for example, if $g_1, \ldots, g_{n-m}$ are equations defining M in the open set $U \subset N$, then the single equation $G := g_1^2 + \cdots + g_{n-m}^2$ also defines M in U. But *is not regular*: 0 is not a regular value of G. This is an immediate consequence of 1.29: if 0 were a regular value then $G^{-1}(0)$ would have dimension $n - 1$, instead of m. It is instructive to see, by applying the chain rule to $d_x G$ for $x \in M \cap U$, why no such point x can be a regular point of G.

1.3 Finding Regular Equations

How does one find regular equation? It can be very hard. Here I give some relatively easy examples and exercises. Mostly, finding equations is a question of paying careful attention to the description of the submanifold M for which one wants to find equations.

Example 1.33. (1) Let $(0, 0) \neq (a, b) \in \mathbb{R}^2$, and let

$$M = \{(u, v) \in \mathbb{R}^2 : (x, y) \text{ is orthogonal to } (a, b)\}.$$

It is easy to find an equation for M:

$$(x, y) \in M \quad \Leftrightarrow \quad (a, b) \cdot (x, y) = 0 \quad \Leftrightarrow \quad ax + by = 0.$$

The map $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g(x, y) = ax + by$ is a regular equation for M in all of $\mathbb{R}^2$, because the matrix $[d_{(x,y)}g]$ is equal to $[a\ b]$ (as g is linear, it is equal to its own derivative at every point), and $[a\ b]$ defines a surjective linear map (remember that we are assuming $(a, b) \neq (0, 0)$.)

(2) Now let $M = \{(u, v) \in \mathbb{R}^2 : (u, v) \text{ is parallel to } (x, y)\}$. How to find equation(s) for M?

First approach The vector $(-b, a)$ is orthogonal to (a, b), so (x, y) is parallel to (a, b) if it is orthogonal to $(-b, a)$. Thus, $M = \{(x, y) : -bx + ay = 0\}$ and we can take $g(x, y) = -bx + ay$. Similar reasoning to (i) shows g is a regular equation for M in all of $\mathbb{R}^2$.

Second approach Two vectors in $\mathbb{R}^2$ are parallel if they are linearly dependent. Linear dependence of two vectors in $\mathbb{R}^2$ (and in general of n vectors in $\mathbb{R}^n$) is detected by the vanishing of a determinant:

$$M = \{(x, y) : \det \begin{pmatrix} a & b \\ x & y \end{pmatrix} = 0\}.$$

As the determinant here is equal to $-bx + ay$, we have arrived at the same conclusion as with the first approach. Perhaps this approach is better. It's easier to see how to generalise it to higher dimensions.

(3) Let

$$M = \{(a, b, x, y) \in \mathbb{R}^4 : (a, b) \neq (0, 0),\ (x, y) \ \text{ is parallel to } \ (a, b)\}.$$

Is M a manifold? The condition for membership is the same as before: that the determinant should be zero. But now we are treating a and b as variables as well. So if we take

$$g(a, b, x, y) = \det \begin{pmatrix} a & b \\ x & y \end{pmatrix}$$

as before, is g a regular equation for M? Notice that in the definition of M we stipulate that $(a, b) \neq (0, 0)$, so we only need to think of the behaviour of g in $U := (\mathbb{R}^2 \smallsetminus \{(, 0)\}) \times \mathbb{R}^2$. We have

$$[d_{(a,b,x,y)}g] = [y, -x, -b, a]$$

and this is a surjective linear map for every $(a, b, x, y) \in U$. In fact we see that it is surjective if $(a, b, x, y) \neq (0, 0, 0, 0)$, so g defines a manifold in $\mathbb{R}^4 \smallsetminus \{0\}$.

(4) In all of the examples we've looked at so far, one equation was enough. If M has codimension greater than 1 then more equations will be needed. Let $0 \neq (a, b, c) \in \mathbb{R}^3$, and let

$$M = \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \ \text{ is parallel to } \ (a, b, c)\}.$$

Clearly M is a line in 3-space, so will need two equations in the neighbourhood of every point. But which two equations? The condition that the two vectors be parallel is equivalent to the the matrix

$$\begin{pmatrix} a & b & c \\ x & y & z \end{pmatrix}$$

having rank 1. This means that all of its 2×2 minors must vanish. But it has three 2×2 minors. Which two should we choose? The answer is that it depends on the vector (a, b, c). If $(a, b, c) = (1, 2, 3)$ then the three minors, in the order $\det(\text{col } 2, \text{col } 3), \det(\text{col } 1, \text{col } 3), \det(\text{col } 1, \text{col } 2)$, are

$$2z - 3y, z - 3x, y - 2x$$

and in fact *any* two of these will do, as if both are zero then so is the third. For example

$$y - 2x = -\frac{1}{3}(2z - 3y) + \frac{2}{3}(z - 3x).$$

Exercise 1.34. (1) Assume $(a, b, c) \neq 0$ and let

$$M = \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \ \text{ is parallel to } \ (a, b, c)\}.$$

Show that if $a \neq 0$ then $\det(\text{col } 2, \text{col } 3)$ can be written as a linear combination of $\det(\text{col } 1, \text{col } 2)$ and $\det(\text{col } 1, \text{col } 3)$, and deduce that M has, as regular equations, $\det(\text{col } 1, \text{col } 3)$ and $\det(\text{col } 1, \text{col } 2)$. Show that if $a = 0$ then these are no longer regular equations for M. Find regular equations if $b \neq 0$, and if $c \neq 0$.

(2) Let

$$M = \{(a,b,c,x,y,z) \in \mathbb{R}^6 : \text{rank}\begin{pmatrix} a & b & c \\ x & y & z \end{pmatrix} = 1\}.$$

Find regular equations for M. Note that

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq M,$$

so we only need be concerned with equations in $\mathbb{R}^6 \smallsetminus \{0\}$. Note also that by (i), there is no set of regular equations which is good for all of $\mathbb{R}^6 \smallsetminus \{0\}$. Instead, you will have to find different sets of regular equations for different open subsets of $\mathbb{R}^6 \smallsetminus \{0\}$.

Example 1.35. (1) Let $\gamma(t) = (\cos t, \sin t, t)$. The image of γ is a helix. It's easy to find regular equations for it: since $\gamma_3(t) = t$, we can take the equations

$$x_1 - \cos x_3, \quad x_2 - \sin x_3. \tag{1.1}$$

(2) Now let $\sigma(t) = (\cos t, \sin t, t^3)$ The image of σ (sketch it!) is once again a manifold; in the next subsection we will give conditions on immersions to guarantee that their images are manifolds. But now finding regular equations is not so easy. Near any point on the curve where $x_3 \neq 0$, we can recover the parameter t as $(x_3)^{\frac{1}{3}}$, and then take, as equations

$$x_1 - \cos(x_3^{\frac{1}{3}}), \quad x_2 - \sin(x_3^{\frac{1}{3}}). \tag{1.2}$$

These are smooth functions on $\mathbb{R}^3 \smallsetminus \{x_3 = 0\}$, and give regular equations there (Exercise 0.4(1). But what about points where $x_3 = 0$? The function $(x_3)^{\frac{1}{3}}$ is not smooth at $x_3 = 0$. There is such a point on the curve, namely $\sigma(0) = (1,0,0)$. Fortunately when t is near 0 we can recover it from $\sin t$, using the arcsin function, which defines a diffeomorphism $(-1,1) \to (-\pi/2, \pi/2)$. Thus in the neighbourhood of $(1,0,0)$ we can use the equations

$$x_1 - \cos(\arcsin x_2), \quad x_3 - (\arcsin x_2)^3. \tag{1.3}$$

Exercise 1.36. (1) Show that the equations (1.1), (1.2) and (1.3) are all regular equations (for the manifolds in question) where they claim to be.

(2) Is the image of the parametrisation $\rho(t) = (\cos t, \sin t, t^2)$ a manifold? If it is, find regular equations for it.

(3) Find regular equations for the image of the map $\mathbb{R} \to \mathbb{R}^3$ given by $f(t) = (\cos t, \sin t, \cos t \sin t)$.

(4) Find equations for the image of the map $\mathbb{R} \to \mathbb{R}^3$ given by $g(t) = (\cosh t, \sinh t, e^t)$.

(5) Find regular equations for the set

$$\{(a,b,c,d,w,x,y,z) \in \mathbb{R}^8 \smallsetminus \{0\} : \operatorname{rank}\begin{pmatrix} a & b & c & d \\ w & x & y & z \end{pmatrix} = 1\}.$$

(6) For each fixed $k \leq p \leq n$, the set

$$\{A \in \mathrm{Mat}_{p\times n}(\mathbb{R}) : \operatorname{rank} A = k\}$$

is a manifold of codimension $(p-k)(n-k)$. Hints for a proof can be found in Exercises I number 18. Exercises 1.34(2) and 1.36(4) above give alternative proofs for the case of matrices of rank 1 in the space of 2×3 matrices, and matrices of rank 1 in the space of 2×4 matrices.

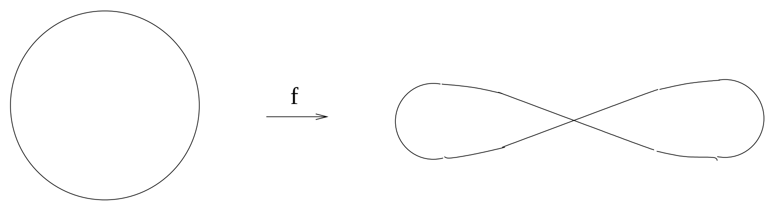
To do: Find regular equations for this manifold when $p = 3, n = 3, k = 1$.

1.4 Images of Immersions

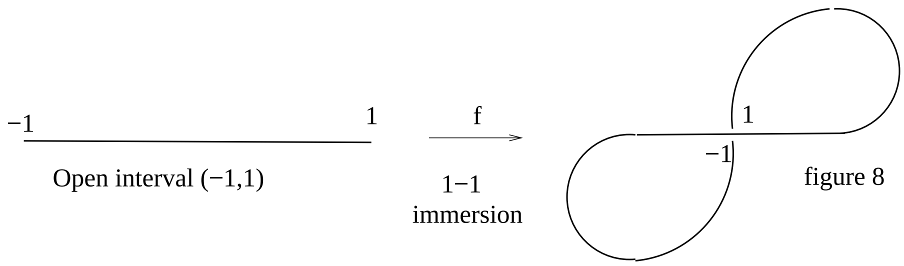
The proof of theorem on the preimage of a regular value (1.29) amounts to little more than the following two facts: first, the obvious fact that the preimage of any point under the standard submersion $\mathbb{R}^{n+k} \to \mathbb{R}^n$ is a manifold of dimension k (in fact an affine subspace of $\mathbb{R}^{n+k}$), and second, the local normal form for submersions 1.27, that that a submersion locally looks just like the standard submersion.

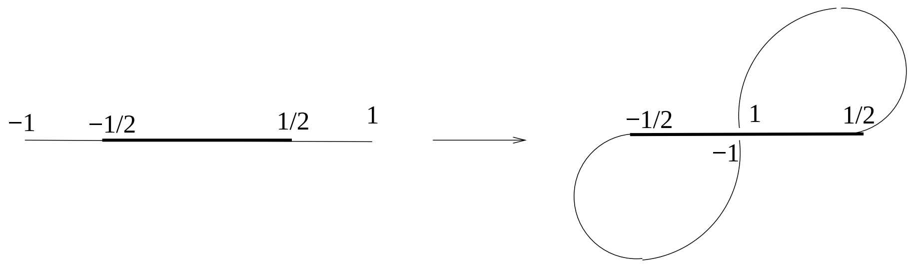
Since dual versions of these two statements hold for immersions — that the *image* of the standard immersion $\mathbb{R}^m \to \mathbb{R}^{m+k}$ is a manifold, and the local normal form for an immersion 1.26 — one might imagine that a dual version of 1.29 would hold, and that the image of an immersion is a manifold. But life is more interesting than that:

Example 1.37. There is an immersion twisting a circle into a figure 8 as shown (which one can easily describe in co-ordinates — **Exercise**).



So the image of an immersion is not necessarily a manifold. Perhaps we have to ask that the immersion be 1-1 in order to guarantee that the image is a manifold? In fact this is neither sufficient, as the next diagram shows, nor necessary, as we shall see later.





Here the injective immersion f twists the open interval $(-1,1)$ into a figure 8. If the interval were the *closed* interval $[-1,1]$ then the points $-1,0$ and 1 would all have the same image. But -1 and 1 are not there, and so f is injective. $\square$

It turns out that the problem with the map in this example is that it is not *open onto its image*, as you can see from the lower drawing – the image of $(-1/2,1/2)$ is not open in $f(-1,1)$.

Definition 1.38. Let $f : X \to Y$ be a map of topological spaces.
(i) f is *open* (or an *open map*) if for every subset U open in X, $f(U)$ is open in Y.
(ii) f is *open onto its image* if for every subset U open in X, $f(U)$ is open in $f(X)$.

Proposition 1.39. *(i) If $f : M \to N$ is a submersion (with M, N manifolds), then it is an open map.*
(ii) If $M^m \subset N^m$ is an inclusion of manifolds of the same dimension, then M is open in N.

Proof Exercise. For (i), use the local normal form for a submersion — the standard submersion is certainly open. For (ii), apply (i) to the inclusion $M \hookrightarrow N$. $\square$

Theorem 1.40. *Suppose that $f : M^m \to N^n$ is an immersion which is open onto its image $f(M)$. Then $f(M)$ is a manifold of dimension m, and for each $x \in M$, $d_xf : T_xM \to T_{f(x)}f(M)$ is an isomorphism.*

Proof Let $y \in f(M)$. We have to find a neighbourhood W of y in $f(M)$ which is diffeomorphic to some open set in $\mathbb{R}^m$. Choose $x \in f^{-1}(y)$, and choose charts $\phi_1 : U_1 \to V_1$ on M around x, and $\phi_2 : U_2 \to V_2$ on N around y, as in theorem on the local normal form for an immersion, 1.26. Then $\phi_2 \circ f \circ \phi_1^{-1}$ is the standard immersion i. The image of i in V_2 is certainly a manifold: it is diffeomorphic to the open set V_1 in $\mathbb{R}^m$, since $i : V \to i(V)$ has a smooth inverse, namely the projection which forgets the last $n - m$ co-ordinates. And ϕ_2 restricts to a diffeomorphism $f(U_1) \to i(V_1)$. So what more do we need? We need $f(U_1)$ to be a neighbourhood of y in $f(M)$. But this is guaranteed by the hypothesis that f be open onto its image.

The last statement holds because $d_x f : T_x M \to T_{f(x)} f(M)$ is a linear injection, and its source and target have the same dimension. □

Example 1.41. (i) Consider the map $S^1 \to \mathbb{R}^2$ given, with respect to complex co-ordinate z, by $z \mapsto z^n$ ($n \in \mathbb{Z} \setminus \{0\}$. It's an easy matter to check that f is an immersion. In Example 1.15 we do this rather painstakingly for the case $n = 2$; using the holomorphic derivative $f'(z) = nz^{n-1}$ gives a much quicker way. Since the image of f is S^1, that f is open onto its image follows from Proposition 1.39 above.
(ii) More interesting example: consider the map $S^2 \to \mathbb{R}^6$ given by all the quadratic monomials:

$$f(x_1, x_2, x_3) = (x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3).$$

Since this map identifies antipodal points (i.e. $f(x) = f(-x)$), it passes to the quotient to define a continuous map

$$\tilde{f} : S^2/\sim \;\to f(S^2).$$

where $S^2/\sim$ is the quotient of S^2 by the equivalence relation $x \sim -x$, with the standard quotient topology. Since f *only* identifies antipodal points, $\tilde{f}$ is a 1-1 continuous map onto $f(S^2)$. As $S^2/\sim$ is the continuous image of a compact space, it is compact. It follows that $\tilde{f}$ is a homeomorphism. In fact our theorem (1.40) tells us that $f(S^2)$ is a manifold, as long as we know that f is open onto its image. But this follows from the fact that $S^2 \to S^2/\sim$ is an open map (which you should check).

In fact $S^2/\sim$ is the *real projective plane*, which can also be viewed as the space of lines in $\mathbb{R}^3$ through 0. The latter is the quotient of $\mathbb{R}^3 \setminus \{0\}$ by the equivalence relation

$$(x_1, x_2, x_3) \sim_1 (y_1, y_2, y_3) \quad \text{if there exists } \lambda \neq 0 \text{ such that } \lambda(x_1, x_2, x_3) = (y_1, y_2, y_3)$$

(whose equivalence classes are indeed lines through 0 in $\mathbb{R}^3$). As an exercise in quotient topologies, you might prove

Proposition 1.42. *There is a natural homeomorphism*

$$S^2/\sim \quad \simeq \quad \mathbb{R}^3/\sim_1 .$$

□

Exercise 1.43. Find equations for the image $f(S^2)$ in Example 1.41. If you have found the right collection of equations, then in a neighbourhood of each point, some four of them will be regular equations for the image. However, there is no guarantee that the same four will be regular equations at all $x \in f(S^2)$. $\square$

Exercise 1.44. Consider the (bijective) map $[0, 2\pi) \to S^1$ sending θ to $(\cos\,\theta, \sin\,\theta)$. Find an open set in $[0, 2\pi)$ whose image in S^1 is not open. $\square$

So far, the principal tool we have for proving that a map is open onto its image is the normal form for submersions. But this can only be used if the image is already known to be a manifold, so is no use if, for example, we want to apply 1.40 to prove that the image is a manifold. A very useful condition under these circumstances is the following notion.

Definition 1.45. A map of topological spaces $f : X \to Y$ is *proper* if for every compact set $K \subset Y$, $f^{-1}(K)$ is also compact.

Example 1.46. (i) If X is compact, f is continuous and Y is Hausdorff (e.g. $Y \subset \mathbb{R}^n$), then f is proper. For

$$K \subset Y \quad \text{compact} \Rightarrow \ K \ \text{ closed in } Y \Rightarrow \ f^{-1}(K) \text{ closed in } X \ \Rightarrow \ f^{-1}(K) \ \text{ compact.}$$

Here the three hypotheses are used in reverse order to justify the three implications.
(ii) The inclusion $S^2 \setminus \{\text{point}\} \hookrightarrow \mathbb{R}^3$ is not proper. For example, the preimage of the compact set S^2 is not compact. So in general the property of properness is not inherited by the restriction of a proper map. On the other hand, if $f : X \to Y$ is proper, then $f : X \to f(X)$ is also proper.
(iii) The map $g : \mathbb{R} \to \mathbb{R}$ defined by $g(t) = t^3$ is proper. So is the map $f : \mathbb{R} \to \mathbb{R}^3$ of Example 1.35(2), $f(t) = (\cos\,t, \sin\,t, t^3)$, is proper (Exercise). $\square$

Proposition 1.47. *Suppose that $f : X \to Y$ is a 1-1, continuous and proper map, with X and Y second countable* [2] *and Y Hausdorff (for example, $X, Y \subset \mathbb{R}^n$). Then f is open onto its image, and is thus a homeomorphism onto its image.*

Proof f is a bijection onto its image, so "open onto its image" is equivalent to "closed onto its image". We show that f is closed onto its image. Second countability is inherited by subspaces. Suppose that $C \subset X$ is closed. We have to show $f(C)$ closed in $f(X)$. Let (z_n) be a sequence in $f(C)$, converging to $z \in f(X)$. The set $K := \{z_n : n \in \mathbb{N}\} \cup \{z\}$ is compact, so $f^{-1}(K)$ is compact, as f is proper. Choose $x_n \in C$ such that $f(x_n) = z_n$. As (x_n) is a sequence in the compact set $f^{-1}(K)$, it has a convergent subsequence. Replacing (x_n) by this convergent subsequence, we may suppose that $(x_n) \to x \in C$ (recall that C is closed in X). By the continuity of f, $z_n = (f(x_n)) \to f(x)$. By uniqueness of limits in Hausdorff spaces, $f(x) = z$. So $z \in f(C)$, and $f(C)$ is closed in $f(X)$. $\square$

[2] i.e. whose topology has a countable basis. This is equivalent to the fact that we use here, namely that a subset Z is closed if and only if every convergent sequence (z_n) of points in Z has its limit in Z

Theorem 1.48. *If $f : M^m \to N^n$ is a proper, 1-1 immersion then $f : M \to f(M)$ is a diffeomorphism.*

Proof By 1.47, f is open onto its image. By 1.40 $f(M)$ is a manifold, and for each $x \in M$, $d_x f : T_x M \to T_{f(x)} f(M)$ is an isomorphism. Hence f, thought of as a map $M \to f(M)$, is a local diffeomorphism, with a smooth locally-defined inverse around each point $y \in f(M)$. But f is also a homeomorphism onto its image, so f has a globally defined inverse. This must coincide with the local inverses wherever both are defined; hence f^{-1} is actually smooth, and $f : M \to f(M)$ is a diffeomorphism. $\square$

Terminology: $f : M \to N$ is an *embedding of M in N* if f is a diffeomorphism onto its image.

Example 1.49. Let $f : M \to N$ be a smooth map. Define a map $\Gamma_f : M \to M \times N$ by $\Gamma_f(x) = (x, f(x))$. Its image is the graph of f. Γ_f has smooth inverse $(x, y) \mapsto x$, and so is an embedding. Its derivative at x is the map

$$\hat{x} \mapsto (\hat{x}, d_x f(\hat{x})),$$

so by Theorem 1.40,

$$T_{(x,f(x))}\mathrm{graph}(f) = \{(\hat{x}, d_x f(\hat{x})) : \hat{x} \in T_x M\}.$$

$\square$

Our last result in this direction is an odd-looking technical result whose purpose is not readily apparent yet:

Theorem 1.50. *Suppose $f : M^m \to N^m$ is a smooth map of manifolds, and that $X \subset M$ is a submanifold (possibly of lower dimension than M). If*

1. *f is 1-1 on X, and*
2. *$d_x f : T_x M \to T_{f(x)} N$ is an isomorphism for all $x \in X$,*

then there is a neighbourhood U of X in M such that $f : U \to f(U)$ is a diffeomorphism.

Proof I give the proof only for X compact, where it is substantially easier than in the general case.
By hypothesis 2. and the inverse function theorem, f maps some neighbourhood U_x of each point $x \in X$ diffeomorphically to a neighbourhood $V_{f(x)}$ of $f(x)$ in N. The union U_1 of all of the U_x is a neighbourhood of X, and at each point $x \in U_1$, $d_x f$ is an isomorphism.

I claim that there is another neighbourhood U_2 of X in M, on which f is 1-1. For suppose that there is not. Then in every neighbourhood of X there exist points $a \neq b$ such that $f(a) = f(b)$. In particular, in $U^{(n)} := \cup_{x \in X} B(x, 1/n)$ there are distinct points a_n, b_n

such that $f(a_n) = f(b_n)$. Now the set U^1 has compact closure (in the ambient Euclidean space of M), simply because X, being compact, is bounded. Hence so does $U^{(1)} \times U^{(1)}$, and so the sequence (a_n, b_n), which is contained in $U^{(1)} \times U^{(1)}$, has a subsequence converging to some point (a, b) in the closure of $U^{(1)} \times U^{(1)}$. By replacing (a_n, b_n) by this subsequence we can suppose that $a_n \to a, b_n \to b$ as $n \to \infty$.

So far we are only guaranteed that a and b lie in the closure of $U^{(1)}$. But in fact they must lie in X; for suppose, for example, that $a \notin X$. As X is compact, there exists $\varepsilon > 0$ such that $B(a, \varepsilon) \cap X = \emptyset$. Then for $n > 2/\varepsilon$, $\|a_n - a\| > \varepsilon/2$, making it impossible for (a_n) to converge to a.

Now $f(a) = f(b)$, by continuity. So as f is 1-1 on X, $a = b$. But because f is a diffeomorphism on some neighbourhood U_a of $a = b$ in M, f is 1-1 on U_a. This contradicts the assumption that $a_n \neq b_n$ and $f(a_n) = f(b_n)$ for all n, since all but finitely many of the a_n and b_n are in U_a.

We conclude that f must indeed be 1-1 on some neighbourhood U_2 of X. Now let $U = U_1 \cap U_2$. Then $f : U \to N$ is a 1-1, open immersion (for openness use 1.39(i)). Hence $f : U \to f(U)$ is a diffeomorphism. □

1.5 Transversality

Given $f : M \to N$ and $Z \subset N$, when is $f^{-1}(Z)$ a manifold? If Z is just a point z_0, the answer is given by 1.29: when z_0 is a regular value of f. Building on this result, we now give a condition for $f^{-1}(Z)$ to be a manifold.

Definition 1.51. (i) If M and Z are both submanifolds of N, we say that M is transverse to Z at $x \in M \cap Z$, written $M \pitchfork Z$ at x, if

$$T_xM + T_xZ = T_xN,$$

and M is transverse to Z if this holds for all $x \in M \cap Z$.

(ii) If $Z \subset N$ and now f is a map from $M \to N$, we say f is *transverse to Z at $x \in M$* (written "$f \pitchfork Z$ at x") if

$$d_xf\,(T_xM) + T_{f(x)}Z = T_{f(x)}N. \tag{1.4}$$

f is *transverse to* Z if it is transverse to Z at every $x \in f^{-1}(Z)$. (So in particular, if $f^{-1}(Z) = \emptyset$, f is transverse to Z.)

In fact (i) is the special case of (ii) in which $f : M \to N$ is the inclusion.

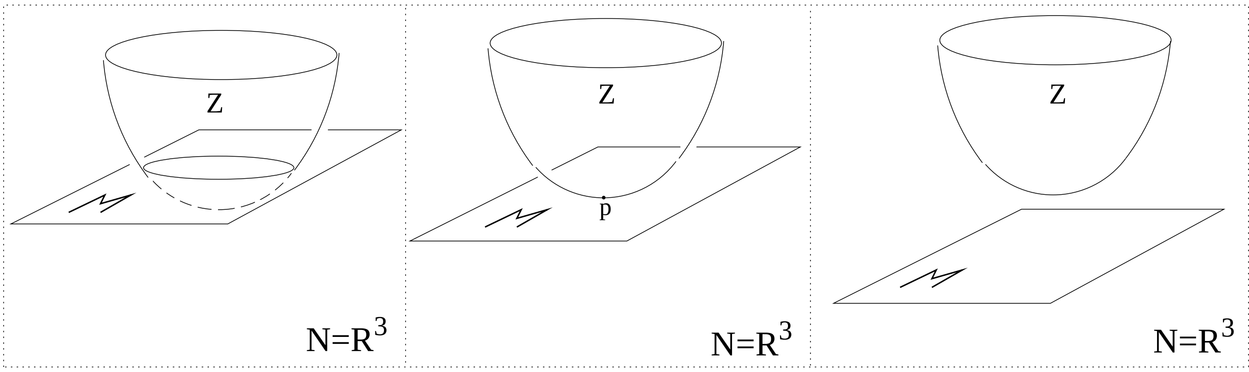
It is important to note that the relation of transversality of two submanifolds of N involves *three* manifolds: the two which are transverse to one another, or not — M and Z in the definition just given — and the ambient manifold N. For example, if M is the x-axis and Z

is the y-axis then M and N meet transversely in $N = \mathbb{R}^2$, for at their point O of intersection, we have

$$T_O M + T_O Z = \mathrm{Sp}\{(1,0)\} + \mathrm{Sp}\{(0,1)\} = \mathbb{R}^2 = T_O \mathbb{R}^2.$$

However, if we now think of $\mathbb{R}^2$, and therefore also M and Z, as contained in $\mathbb{R}^3$, then M and Z are no longer transverse. Indeed, no two 1-manifolds meeting at a point x in $\mathbb{R}^3$ can be transverse in $\mathbb{R}^3$ — $T_xM + T_xZ$ has dimension at most 2, and thus cannot be equal to $T_x\mathbb{R}^3 = \mathbb{R}^3$. For this reason, the notation "$M \pitchfork Z$" should ideally be replaced by something like "$(M \pitchfork Z)_N$" to emphasize the role of the ambient space.

On the other hand, part (ii) of the definition of transversality does not have this ambiguity, since the manifold N is now present in the data, as the target of the map f. Nevertheless, this is an issue in which one must be careful.



In the picture on the left, $M \pitchfork Z$; for at each point where they meet, their tangent planes are distinct, and any two distinct 2-dimensional subspaces of $\mathbb{R}^3$ together generate all of $\mathbb{R}^3$. In the middle picture, M and Z are not transverse, since at the point p where they meet, $T_pM = T_pZ$. In the right-hand picture again $M \pitchfork Z$, this time because their intersection is empty.

Theorem 1.52. *If $f : M \to N$ is transverse to Z then $f^{-1}(Z)$, if not empty, is a manifold, and*

$$\dim M \; - \; \dim f^{-1}(Z) = \dim N \; - \; \dim Z.$$

Moreover,

$$T_x f^{-1}(Z) = (d_x f)^{-1}(T_{f(x)} Z).$$

Proof Let $x \in f^{-1}(Z)$. By 1.31 there is a neighbourhood U of z in N on which Z has regular equations g — that is, a smooth map $g : U \to \mathbb{R}^c$ such that $Z \cap U = g^{-1}(0)$ and 0 is a regular value of g. Here $c = \dim N - \dim Z$. Then $f^{-1}(U)$ is a neighbourhood of x in M, and

$$f^{-1}(Z) \cap f^{-1}(U) = f^{-1}(Z \cap U) = f^{-1}(g^{-1}(0)) = (g \circ f)^{-1}(0).$$

Now we show that 0 is a regular value of $g \circ f$. By applying $d_{f(x)}g$ to the equality (1.4) we obtain

$$d_x(g \circ f)(T_xM) + d_{f(x)}g(T_{f(x)}Z) = d_{f(x)}g(T_{f(x)}N) \tag{1.5}$$

Since $T_{f(x)}Z = \ker d_{f(x)}g$, the second term on the left hand side is 0. And since 0 is a regular value of g, $d_{f(x)}g$ is an epimorphism and the right hand side is equal to $\mathbb{R}^c$. We conclude that 0 is a regular value of $g \circ f$. Since $f^{-1}(Z) \cap f^{-1}(U) = (g \circ f)^{-1}(0)$, $f^{-1}(Z) \cap f^{-1}(U)$ is a manifold. Being a manifold is a local property, so $f^{-1}(Z)$ is a manifold.

The second statement follows from the fact that $f^{-1}(Z)$ is locally the preimage of a regular value of a map into $\mathbb{R}^c$.

For the third statement, we have

$$T_x f^{-1}(Z) = \ker d_x(g \circ f) = \ker d_{f(x)}g \circ d_x f = (d_x f)^{-1}(\ker d_{f(x)}g) = (d_x f)^{-1}(T_{f(x)}Z),$$

1.29. □

Let $M \subset N$ be manifolds. The *codimension* of M in N is defined to be $\dim N - \dim M$. Thus, part (ii) of the previous theorem says that if $f \pitchfork Z$ and $f^{-1}(Z) \neq \emptyset$, then the codimension in M of $f^{-1}(Z)$ is equal to the codimension in N of Z.

Remark 1.53. (i) The above proof makes clear that 0 is a regular value of $g \circ f$ if and *only if* equation (1.4) holds for all x in $f^{-1}(Z \cap U)$.

(ii) The equality in (1.4) can be rewritten as surjectivity of the composite

$$T_x M \xrightarrow{d_x f} T_{f(x)}N \to \frac{T_{f(x)}N}{T_{f(x)}Z}.$$

The kernel of the composite is $(d_x f)^{-1}(T_{f(x)}Z)$. The rank-nullity theorem of linear algebra tells us that a linear map $\ell : V_1 \to V_2$ is surjective if and only if the dimension of $\ker \ell$ is $\dim V_1 - \dim V_2$. So the composite map $T_x M \to T_{f(x)}N/T_{f(x)}Z$ is surjective if and only if $(d_x f)^{-1}(T_{f(x)}Z)$ is as *small* as possible, with dimension $\dim M - (\dim N - \dim Z)$.

(iii) Suppose y is a regular value of the map $g : N \to P$ and $Z := g^{-1}(y) \neq \emptyset$. Since $T_z Z = \ker d_z g$ for each $z \in Z$, $d_z g$ gives rise to an isomorphism

$$\frac{T_z N}{T_z Z} \simeq T_{g(z)}P.$$

Example 1.54. We give three examples which illustrate the definition of transversality in action. The first uses it directly; the second uses it via the interpretation given in the proof of Theorem 1.52, that $f \pitchfork Z$ if and only if regular equations for Z, when composed with f, become regular equations for $f^{-1}(Z)$.

(i) Suppose $S \subset \mathbb{R}^n \times \mathbb{R}^p$ is a manifold, and let $\pi : S \to \mathbb{R}^p$ be projection. Then for each point $y \in \mathbb{R}^p$,

$$\big(\mathbb{R}^n \times \{y\}\big) \pitchfork S \text{ at } (x, y) \quad \Leftrightarrow \quad d_{(x,y)}\pi : T_{(x,y)}S \to T_y\mathbb{R}^p = \mathbb{R}^p \text{ is surjective.}$$

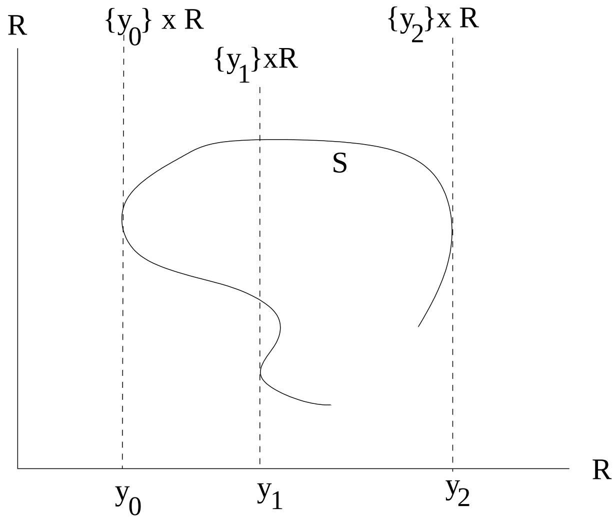
For, by the definition of transversality, the LHS means that for all $(\hat{x}, \hat{y}) \in T_{(x,y)}\mathbb{R}^n \times \mathbb{R}^p$, there exist $(\hat{x}_1, \hat{y}_1) \in T_{(x,y)}\big(\mathbb{R}^p \times \{y\}\big)$ and $(\hat{x}_2, \hat{y}_2) \in T_{(x,y)}S$ such that

$$(\hat{x}_1, \hat{y}_1) + (\hat{x}_2, \hat{y}_2) = (\hat{x}, \hat{y}).$$

As $T_{(x,y)}\mathbb{R}^n \times \{y\} = \mathbb{R}^p \times 0$, $\hat{y}_1 = 0$, and so we must have $\hat{y}_2 = \hat{y}$. Since $d_{(x,y)}\pi(\hat{x}_2, \hat{y}) = \hat{y}$, and $\hat{y}$ was arbitrary, this shows that $d_{(x,y)}\pi : T_{(x,y)}S \to T_y\mathbb{R}^p$ is surjective.

The argument in the opposite direction is equally simple. Suppose $d_{(x,y)}\pi$ is surjective. Then given $(\hat{x}, \hat{y}) \in T_{(x,y)}\mathbb{R}^n \times \mathbb{R}^p$, there exists a vector of the form $(\hat{x}_1, \hat{y}) \in T_{(x,y)}S$. The vector $(\hat{x} - \hat{x}_1, 0)$ is in $T_{(x,y)}\mathbb{R}^n \times \{y\}$. Thus $(\hat{x}, \hat{y})$ is the sum of a vector in $T_{(x,y)}S$ and a vector in $T_{(x,y)}\mathbb{R}^n \times \{y\}$.

The geometry of the situation can be appreciated in the following picture, in which $n = p = \dim S = 1$. The points where $\{y\} \times \mathbb{R}$ fails to be transverse to S are precisely the critical points of $\pi : S \to \mathbb{R}$.



(ii) Now suppose that $F : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n$ is a smooth map, $Z \subset \mathbb{R}^n$ is a submanifold, and $F \pitchfork Z$. For each $y \in \mathbb{R}^p$, denote by $f_y : \mathbb{R}^m \to \mathbb{R}^n$ the map $f_y(x) = F(x, y)$. Then for each $(x, y) \in F^{-1}(Z) \subset \mathbb{R}^m \times \mathbb{R}^p$,

$$\big(\mathbb{R}^m \times \{y\}\big) \pitchfork F^{-1}(Z) \text{ at } (x, y) \quad \Leftrightarrow \quad f_y \pitchfork Z \text{ at } x.$$

To see this, suppose that $\operatorname{codim} Z = c$ and $g_1, \ldots, g_c$ are regular defining equations for Z in some neighbourhood U of $F(x, y)$ in $\mathbb{R}^n$, and let $g : \mathbb{R}^n \to \mathbb{R}^c$ be the map with component functions $g_1, \ldots, g_c$. Then by what we saw in the proof of Theorem 1.52, the transversality of F to Z means that 0 is a regular value of $g \circ F$, i.e. $g_1 \circ F, \ldots, g_c \circ F$ are regular defining equations for $F^{-1}(Z)$ in $F^{-1}(U)$. Moreover, by the same argument,

$$f_y \pitchfork Z \text{ at } x \quad \Leftrightarrow \quad d_x(g \circ f_y) \text{ is surjective.} \tag{1.6}$$

Denote by $i_y : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^p$ sending x to (x, y). We have

$$\big(\mathbb{R}^m \times \{y\}\big) \pitchfork F^{-1}(Z) \text{ at } (x, y) \quad \Leftrightarrow \quad i_y \pitchfork F^{-1}(Z) \text{ at } x.$$

Since the $g_i \circ F$ are regular equations for $F^{-1}(Z)$, the argument of 1.52 gives

$$i_y \pitchfork F^{-1}(Z) \text{ at } x \quad \Leftrightarrow \quad d_x(g \circ F \circ i_y) \text{ is surjective.} \tag{1.7}$$

But $f_y = F \circ i_y$, so the right hand sides of (1.6) and (1.7) are the same.

Exercise 1.55. (i) Suppose M and P are manifolds, and $S \subset M \times P$ is a manifold. Let $\pi : S \to P$ be the restriction of the standard projection $M \times P \to P$. Show that if $y \in P$ then

$$\big(M \times \{y\}\big) \pitchfork S \quad \Leftrightarrow \quad y \text{ is a regular value of } \pi.$$

(ii) Suppose that $F : M \times P \to N$ is a smooth map, $Z \subset N$ is a manifold, and $F \pitchfork Z$. For each $y \in P$, denote by $f_y : M \to N$ the map $f_y(x) = F(x, y)$. Show that

$$f_y \pitchfork Z \quad \Leftrightarrow \quad M \times \{y\} \pitchfork F^{-1}(Z).$$

(iii) Show, in the situation of (ii), that

$$f_y \pitchfork Z \quad \Leftrightarrow \quad y \text{ is a regular value of } \pi : F^{-1}(Z) \to P.$$

2 Sard's theorem and the density of transversality

2.1 Tangent and normal bundles

Let $M \subset \mathbb{R}^n$ be a manifold. We define the *tangent bundle* TM to be the set

$$TM = \{(x, v) \in M \times \mathbb{R}^n : v \in T_x M\},$$

and denote the projection $TM \to M$ sending (x, v) to x by p.

Proposition 2.1. *Let $M^m \subset \mathbb{R}^n$ be a manifold. Then TM is a manifold of dimension $2m$, and the projection $p : TM \to M$ is a submersion.*

Proof Suppose that $\varphi : U \to V \subset \mathbb{R}^m$ is a chart on M. Then $p^{-1}(U)$ is open in TM. We show that there is a diffeomorphism $p^{-1}(U) \simeq V \times \mathbb{R}^m$. As TM is covered by open sets of the form $p^{-1}(U)$, these diffeomorphisms provide an atlas for TM. For each $x \in U$, $T_x M = d_{\varphi(x)}(\varphi^{-1})(\mathbb{R}^m)$. It follows that the map $V \times \mathbb{R}^m \to p^{-1}(U)$ sending $(y, v) \mapsto (\varphi^{-1}(y), d_y\varphi^{-1}(v))$ is a smooth surjection. It has the smooth inverse $T\varphi$ defined by $T\varphi(x, v) = (\varphi(x), d_x\varphi(v))$, and is thus a diffeomorphism.

The map $p : TM \to M$ is a submersion, because it is a composite of submersions; the diagram

$$\begin{tikzcd} p^{-1}(U) \arrow[r, "T\varphi"] \arrow[d, "p"'] & V \times \mathbb{R}^m \arrow[d, "\mathrm{pr}"] \\ U \arrow[r, "\varphi"'] & V \end{tikzcd}$$

commutes, so $p = \varphi^{-1} \circ \mathrm{pr} \circ T\varphi$. $\square$

Remark 2.2. By juxtaposing the diagram in this proof with the commutative diagram

$$\begin{tikzcd} V \times \mathbb{R}^m \arrow[r, "\varphi^{-1}\times \mathrm{id}"] \arrow[d] & U \times \mathbb{R}^m \arrow[d, "\mathrm{pr}"] \\ V \arrow[r, "\varphi^{-1}"'] & U \end{tikzcd}$$

we obtain the cummutative diagram

$$\begin{tikzcd} p^{-1}(U) \arrow[r] \arrow[d, "p"'] & U \times \mathbb{R}^m \arrow[d, "\mathrm{pr}"] \\ U \arrow[r, "\mathrm{id}"'] & U \end{tikzcd}$$

more succinctly displayed as

$$\begin{tikzcd} p^{-1}(U) \arrow[rr] \arrow[dr, "p"'] & & U \times \mathbb{R}^m \arrow[dl, "\mathrm{pr}"] \\ & U & \end{tikzcd} \tag{2.1}$$

in which (check this!) restriction of the horizontal arrow to each fibre $p^{-1}(x) = T_xM$ is a linear isomorphism $T_xM \to \{x\} \times \mathbb{R}^m$.

Let $M^m \subset \mathbb{R}^n$ be a manifold. The *normal space* to M at x, $\nu_x(M, \mathbb{R}^n)$, is the orthogonal complement $(T_xM)^\perp$ to T_xM in $\mathbb{R}^n$. We define the *normal bundle of M in $\mathbb{R}^n$*, denoted by $\nu(M, \mathbb{R}^n)$, by

$$\nu(M, \mathbb{R}^n) = \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n : x \in M, v \in (T_xM)^\perp\},$$

More generally, let $M \subset N \subset \mathbb{R}^n$ be manifolds. The *normal bundle to M in N*, which we denote by $\nu(M, N)$, is the set

$$\{(x, v) \in M \times T_xN : v \in (T_xM)^\perp\}.$$

We denote the projections $\nu(M, \mathbb{R}^n) \to M$ by and $\nu(M, N) \to M$ by p.

Proposition 2.3. *Let $M \subset \mathbb{R}^n$ be a manifold of dimension m, and write $c = n - m$.*
(i) If in the open set $U \subset \mathbb{R}^n$ M has regular equations $g : U \to \mathbb{R}^c$ then for each $x \in U$, $\nabla g_1(x), \ldots, \nabla g_c(x)$ form a basis for $\nu_x(M, \mathbb{R}^n)$.
(ii) $\nu(M, \mathbb{R}^n)$ is a manifold of dimension n.

Proof (i) Recall that $\nabla g_i(x)$ is the (unique) vector in $\mathbb{R}^n$ such that for every vector v,

$$\nabla g_i(x) \cdot v = d_x g_i(v)$$

(in this connection, see Exercise 2.6(6) below). Since $T_xM = \ker d_xg$, we have $T_xM = \cap_i \ker d_xg_i = \cap_i(\nabla g_i(x))^\perp = \mathrm{Sp}\{\nabla g_1(x), \ldots, \nabla g_c(x)\}^\perp$. Therefore

$$(T_xM)^\perp = \mathrm{Sp}\{\nabla g_1(x), \ldots, \nabla g_c(x)\}^{\perp\perp} = \mathrm{Sp}\{\nabla g_1(x), \ldots, \nabla g_c(x)\}.$$

For each $x \in M$, the vectors $\nabla g_1(x), \ldots, \nabla g_c(x)$ are linearly independent, as they are the rows of the matrix of $d_x g$, which has rank c. Hence they form a basis for $\nu_x(M, \mathbb{R}^n)$.

(ii) Let $x \in M$ and let U be a neighbourhood of x in $\mathbb{R}^n$ on which M has regular equations $g : U \to \mathbb{R}^c$. Write $U \cap M = U_0$. Then $p^{-1}(U_0) = \nu(M, \mathbb{R}^n) \cap (U \times \mathbb{R}^n)$ is open in $\nu(M, \mathbb{R}^n)$, and by (i)

$$\nu(M, \mathbb{R}^n) \cap (U \times \mathbb{R}^n) = \{(x, v) : x \in M, v \in \mathrm{Sp}\{\nabla g_1(x), \ldots, \nabla g_c(x)\}\}.$$

There is a smooth surjection $\psi_{U_0} : U_0 \times \mathbb{R}^c \to p^{-1}(U_0)$ given by

$$(x, t_1, \ldots, t_c) \mapsto (x, t_1 \nabla g_1(x) + \cdots + t_c \nabla g_c(x)).$$

Claim: ψ_{U_0} is a diffeomorphism. To show this I will exhibit a smooth inverse. The inverse is the following:

$$(x, v) \mapsto (x, \text{coefficients of } v \text{ with respect to the basis } \nabla g_1(x), \ldots, \nabla g_c(x) \text{ of } \nu_x(M, \mathbb{R}^n)).$$

The expression of a vector in terms of the vectors of a basis depends smoothly on the vector (this is essentially Cramer's rule). Hence $\psi_{U_0}^{-1}$ is smooth. Because $U_0 \times \mathbb{R}^c$ is a manifold, so is $p^{-1}(U_0)$. Clearly, every point in $\nu(M, \mathbb{R}^n)$ has a neighbourhood of this form. Therefore $\nu(M, \mathbb{R}^n)$ is a manifold. □

Exercise 2.4. Show that there is a commutative diagram

$$\begin{tikzcd} p^{-1}(U_0) \arrow[rr] \arrow[dr, "p"'] & & U_0 \times \mathbb{R}^c \arrow[dl, "\mathrm{pr}"] \\ & U_0 & \end{tikzcd} \tag{2.2}$$

in which the restriction of the horizontal map to each fibre $p^{-1}(x) = \nu_x(M, \mathbb{R}^n)$ is a linear isomorphism $\nu_x(M, \mathbb{R}^n) \to \{x\} \times \mathbb{R}^c$.

Both TM and $\nu(M, \mathbb{R}^n)$, or rather the maps $TM \xrightarrow{p} M$ and $\nu(M, \mathbb{R}^n) \xrightarrow{p} M$, are examples of *vector bundles* over M — maps $X \xrightarrow{p} M$ characterised by the existence of diagrams like (2.1) and (2.2).

Proposition 2.5. *Suppose that $M^m \subset N^n \subset \mathbb{R}^p$ are manifolds. Then $\nu(M, N)$ is a manifold of dimension n.*

Proof $\nu(M, N)$ is the intersection (in $N \times \mathbb{R}^p$) of $\nu(M, \mathbb{R}^p)$ and TN. I leave as an exercise (2.6(3) below) the proof that this intersection is transverse. The corollary follows. □

Exercise 2.6. 1. The normal bundle of the circle S^1 in $\mathbb{R}^2$, $\nu(S^1, \mathbb{R}^2)$ is the set

$$\{(x, v) \in S^1 \times \mathbb{R}^2 : v \in (T_x S^1)^\perp\} = \{(x, v) \in S^1 \times \mathbb{R}^2 : v = tx \text{ for some } t \in \mathbb{R}\}.$$

(i) Find regular equations for $\nu(S^1, \mathbb{R}^2)$ in $\mathbb{R}^2 \times \mathbb{R}^2$.
(ii) Show (by an explicit formula) that there is a commutative diagram

$$\begin{tikzcd} \nu(S^1, \mathbb{R}^2) \arrow[rr] \arrow[dr, "p"'] & & S^1 \times \mathbb{R} \arrow[dl, "\mathrm{pr}"] \\ & S^1 & \end{tikzcd}$$

in which the restriction of the horizontal map to each fibre $p^{-1}(x) = \nu_x(S^1, \mathbb{R}^2)$ is a linear isomorphism $\nu_x(S^1, \mathbb{R}^2) \to \{x\} \times \mathbb{R}$. Find also an explicit formula for the inverse to the horizontal map.
(iii) Generalise (ii) to $S^n \subset \mathbb{R}^{n+1}$.

(2) The tangent bundle of S^n is the set

$$\{(x, v) \in S^n \times \mathbb{R}^{n+1} : v \in T_x S^n\} = \{(x, v) \in S^n \times \mathbb{R}^{n+1} : \ v \cdot x = 0\}.$$

(i) Find regular equations for TS^n in $\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}$
(ii) Show (by an explicit formula) that there is a commutative diagram

$$\begin{tikzcd} TS^1 \arrow[rr] \arrow[dr, "p"'] & & S^1 \times \mathbb{R} \arrow[dl, "\mathrm{pr}"] \\ & S^1 & \end{tikzcd}$$

in which the restriction of the horizontal map to each fibre $p^{-1}(x) = T_x S^1$ is a linear isomorphism $T_x S^1 \to \{x\} \times \mathbb{R}$. Find also an explicit formulae for the inverse to the horizontal diffeomorphism. Hint: the inverse may be easier; at each point $(x, y) \in S^1$, $T_{(x,y)}S^1$ is generated by the vector $(-y, x)$.
Part (ii) of (2) does *not* generalise to S^n for all n; in particular, it fails for S^2, though it is true for S^3. This is a consequence of a rather dramatic theorem called the Hairy Ball Theorem which we will prove in Section 3.

(3) Show, by making a calculation for a suitable example (e.g. $S^1 \subset \mathbb{R}^2$), that in general for a manifold M,

$$T_{(x,v)}TM \neq T_x M \times T_v(T_x M).$$

(4) Complete the proof of 2.5 by showing that TN and $\nu(M, \mathbb{R}^p)$ are transverse *in* $N \times \mathbb{R}^p$. It is worth noting that since both TN and $\nu(M, \mathbb{R}^p)$ are contained in $N \times \mathbb{R}^p$, so are their tangent spaces at each point contained in the tangent space to $N \times \mathbb{R}^p$. It follows that *the sum of these two spaces is at best equal to the tangent space to $N \times \mathbb{R}^p$.* Thus, for example, it would be a mistake here to try to prove that TN and $\nu(M, \mathbb{R}^p)$ are transverse in the bigger space $\mathbb{R}^p \times \mathbb{R}^p$ — they cannot possibly be! Moral: before embarking on a proof that two manifolds are transverse, it is important to make sure you choose the smallest ambient

manifold possible — i.e. that contains them both.

(5) Let M be a smooth curve in $\mathbb{R}^2$ and let $(x, v) \in TM$. Under what circumstances does *equality* hold in (3) above?

(6) Let $N \subset \mathbb{R}^p$ be a manifold, and let $f : M \to \mathbb{R}$ be a smooth function. How might one define a gradient vector $\nabla f(x) \in T_x M$ at each point $x \in M$? □

2.2 Tubular neighbourhoods

Vector bundles play an important rôle in the theory of manifolds. For now, we use $\nu(M, N)$ to construct a *tubular neighbourhood* of M in N.

Definition 2.7. Let $M \subset N \subset \mathbb{R}^n$ be manifolds. A *tubular neighbourhood of* M *in* N is a neighbourhood U of M in N together with a smooth submersion $\pi : U \to M$ such that $\pi_{|M} = 1_M$.

Theorem 2.8. Tubular neighbourhood theorem. *Let $M \subset N \subset \mathbb{R}^n$ be manifolds. Then*
(i) M has a tubular neighbourhood U_M in $\mathbb{R}^n$.
(ii) M has a tubular neighbourhood $U_{M,N}$ in N.

Proof. Consider the *exponential map of the normal bundle* $n : \nu(M, \mathbb{R}^n) \to \mathbb{R}^n$ given by $n(x, v) = x + v$. I claim that

1. At every point $(x, 0)$ of $M \times \{0\} \subset \nu(M, \mathbb{R}^n)$, $d_{(x,0)}n$ is an isomorphism, and
2. n is $1-1$ on $M \times \{0\}$.

The second of these is obvious. The first holds, because

1. $T_{(x,0)}\nu(M, \mathbb{R}^n) = T_x M \oplus (T_x M)^\perp$ (**Exercise:** show this, by differentiating suitable curves in $\nu(M, \mathbb{R}^n)$), and
2. $d_{(x,0)}n(\hat{x}, 0) = \hat{x}, \quad d_{(x,0)}n(0, \hat{v}) = \hat{v}$, since n is the restriction to $T_{(x,0)}\nu(M, \mathbb{R}^n)$ of a linear map (and therefore its derivative is the derivative of that linear map, which is, of course, the map itself),

so that the image of $d_{(x,0)}n$ contains $T_x M$ and $(T_x M)^\perp$. This means that it is all of $\mathbb{R}^n$, and $d_{(x,0)}n$ is an isomorphism. From these two facts it follows, by Theorem 1.50, that there are neighbourhoods W of $M \times \{0\}$ in $\nu(M, \mathbb{R}^n)$ and U_M of M in $\mathbb{R}^n$ such that $n : W \to U$ is a diffeomorphism. Now define $\pi : U \to M$ by the composite of $n^{-1} : U \to W$ with the natural projection $W \to M$.

(ii) Let U_M be a tubular neighbourhood of M in $\mathbb{R}^n$. The intersection $U_M \cap N$ is a neighbourhood of M in N. The projection $\pi : U_M \to M$ restricts to give a projection $\pi_{M,N} : U_M \cap N \to M$. All that is lacking is that this is a submersion. It *is* a submersion at each point of M, since $\pi_{|M}$ is the identity. The set of points where any map is a submersion is open (this follows directly from the local normal form for a submersion, 1.27). So M has a neighbourhood inside $U_M \cap N$ on which $\pi_|$ is a submersion. We take this, together with the restriction of π, as our tubular neighbourhood $U_{M,N}$. $\square$

Remark 2.9. In the tubular neighbourhood of M in $\mathbb{R}^n$ constructed in the proof, for each point $u \in U$ the line $\pi(u) - u$ lies in $T_{\pi(u)}M^\perp$. Hence for each $u \in U$, $\pi(u)$ is a critical point of the "distance-squared" function $\rho_u : M \to \mathbb{R}$ defined by $\rho_u(x) = \|x - u\|^2$ (Exercise — use the fact that $\|x - u\|^2 = (x - u) \cdot (x - u)$ to obtain a simple expression for $d_x\rho_u$). In fact I claim that if M is closed in $\mathbb{R}^n$, and the map n of the proof of the Tubular Neighbourhood Theorem 2.25 defines a dffeomorphism from a neighbourhood W of $M \times \{0\}$ in $\nu(M, \mathbb{R}^n)$ to a neighbourhood U of M in $\mathbb{R}^n$, then $\pi(u)$ is the *closest* point to u on M, the point where ρ_u attains its absolute minimum. This is left as an exercise.

Exercise 2.10. Suppose $M \subset \mathbb{R}^n$ and U is a neighbourhood of M in $\mathbb{R}^n$. Show that if M is compact then there exists $\varepsilon > 0$ such that

$$\bigcup_{x \in M} B(x, \varepsilon) \subset U.$$

Show, by an example, that if M is not compact, such an ε may not exist. (In fact there exists a strictly positive function $\eta : M \to \mathbb{R}$ such that

$$\bigcup_{x \in M} B(x, \eta(x)) \subset U,$$

but to prove this one needs (I think) to use partitions of unity, which we have not yet introduced.)

Exercise 2.11. (i) Write down an explicit formula for the composite $S^1 \times \mathbb{R} \to \nu(S^1, \mathbb{R}^2) \xrightarrow{n} \mathbb{R}^2$ (where the map $S^1 \times \mathbb{R} \to \nu(S^1, \mathbb{R})$ is the diffeomorphism you found in Exercise 2.6(1)(ii), and n is the "exponential map" used in the proof of 2.8).
(ii) Find a neighbourhood of W of $S^1 \times \{0\}$ in $\nu(S^1, \mathbb{R}^2)$, and a neighbourhood U of S^1 in $\mathbb{R}^2$, such that $n : W \to U$ is a diffeomorphism. You may find it helpful to replace $\nu(S^1, \mathbb{R}^2)$ by $S^1 \times \mathbb{R}$ (because you can draw it) and use the composite map you found in (i) instead of n.
(iii) Write down a formula for the projection $\pi : U \to S^1$ that you get as the composite

$$U \xrightarrow{n^{-1}} W \to S^1 \times \{0\} = S^1.$$

It should be something very familiar!
(iv) Generalise the answer to (iii) to $S^n \subset \mathbb{R}^{n+1}$. $\square$

Remark 2.12. The diffeomorphism $W \simeq U_M$ in the proof of 2.8(i) shows that the way that M sits inside $\mathbb{R}^n$ is just the same (up to diffeomorphism) as the way that $M = M \times \{0\}$ sits inside the normal bundle space $\nu(M, \mathbb{R}^n)$. Our cheap proof of part (ii) of 2.8 avoided the construction of an "exponential map" $\nu(M,N) \to N$, and thus did not give us a diffeomorphism from a neighbourhood of $M \times \{0\}$ in $\nu(M,N)$ to a neighbourhood of M in N. However, it is possible to show the exixtence of such a diffeomorphism, using the existence of tubular neighbourhoods of M in $\mathbb{R}^n$ and of N in $\mathbb{R}^n$ in a simple variant of the proof of 2.8(i) - see Exercises II.

2.3 Sard's theorem

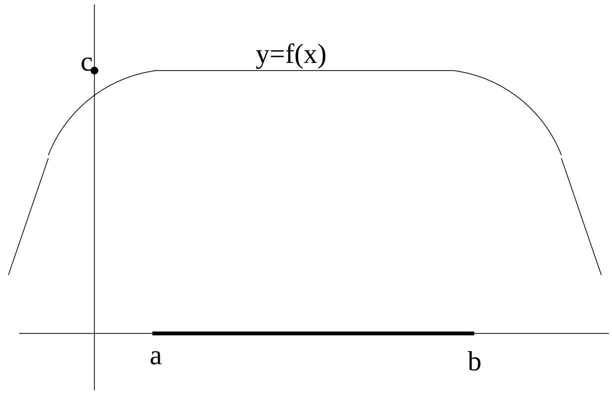
Tubular neighbourhoods have many uses in differential topology. The first use we will put them to, is as a means of constructing perturbations of maps. It will be extremely important to be able to perturb a given map $f : M \to N$ an arbitrarily small amount so that it behaves "well" with respect to some predetermined criterion. The main example of this will be in our treatment of intersection theory: given a map $f : M \to N$ and a submanifold Z of N, we would like to be able to perturb f an arbitrarily small amount so that it becomes transverse to Z. And we would like the perturbed map to be homotopic to f. In the construction of such perturbations and homotopies, the linear structure of the ambient space $\mathbb{R}^n$ is extremely useful: the easiest way to deform or perturb a map $f : M \to N$ is to add to it some other map $\varepsilon p(x)$ (ε for "small', p for "perturbation"). The problem is that in a general manifold N we have no additive structure, so the addition $f(x) + \varepsilon p(x)$ can only be done in the ambient $\mathbb{R}^n$. The role of the tubular neighbourhood $U_N \subset \mathbb{R}^n$ and its projection $\pi : U_N \to N$ is to allow us to project the perturbed map $f(x) + \varepsilon p(x)$ back into N, to get a new map $f_\varepsilon : M \to N$.

Theorem 2.13. Weak Transversality Theorem *Given a smooth map of manifolds $f : M \to N$, and a submanifold Z of N, then for every $\varepsilon > 0$ there exists a smooth map $f_\varepsilon : M \to N$, homotopic to f, such that for all $x \in M$, $\|f(x) - f_\varepsilon(x)\| < \varepsilon$, and such that f_ε is transverse to Z.*

Here the distance we are using is the Euclidean distance, in the ambient Euclidean space of N. The proof of 2.13 relies on two ingredients: the first is Sard's theorem, a moderately deep theorem of analysis whose proof we will skip (though see Milnor's little book for an accessible treatment). The second is a very clever lemma of René Thom, which puts Sard's theorem to spectacular use.

Theorem 2.14. Sard's Theorem *Let $f : M \to N$ be a smooth map of smooth manifolds. The set of critical values of f has measure 0.* □

Notice first that it says that the set of critical *values* has measure zero (we will explain what that means in a moment), not the set of critical *points*. The latter set can be very big, as in the picture below, which shows the graph of a smooth function $f : \mathbb{R} \to \mathbb{R}$.



The set of critical points is the interval $[a, b]$, whereas the set of critical values has just one point, c.

Definition 2.15. The set $X \subset \mathbb{R}^m$ has *measure zero in* $\mathbb{R}^m$ (or just *measure zero* if the ambient space is clear) if for every $\varepsilon > 0$ it is possible to find a covering of X by an enumerable set of rectangles $\{R_j\}_{j\in\mathbb{N}}$ such that

$$\sum_j \mathrm{vol}(R_j) < \varepsilon.$$

Here "rectangle" means a subset of $\mathbb{R}^m$ of the form $[a_1, b_1] \times \cdots \times [a_m, b_m]$, and the volume $\mathrm{vol}(R)$ of such a rectangle is $(b_1 - a_1) \times \cdots \times (b_m - a_m)$. A rectangle is by definition closed and bounded, and hence compact.

Note that the notion of measure zero is relative: the real interval $[0, 1]$ does not have measure zero in $\mathbb{R}$, but if we consider $[0, 1]$ as a subset of $\mathbb{R}^2$ by means of the inclusion of $\mathbb{R}$ as $\mathbb{R} \times \{0\}$, it does have measure zero in $\mathbb{R}^2$.

Proposition 2.16.

1. *The union of a countable set of sets $X_k \subset \mathbb{R}^m$ of measure zero in $\mathbb{R}^m$ has measure zero in $\mathbb{R}^m$.*
2. *If $X \subset \mathbb{R}^m$ has measure zero in $\mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^m$ is smooth, then $f(X)$ has measure zero in $\mathbb{R}^m$. Ditto if f is defined only on some open set $U \subset \mathbb{R}^m$, replacing $f(X)$ by $f(X \cap U)$.*
3. *No open subset of $\mathbb{R}^m$ can have measure zero in $\mathbb{R}^m$.*

Proof : 1. This is easy to prove directly from the definition (just find a countable cover of X_k with total volume $\varepsilon/2^k$, and put these covers together).

2. This requires a little more work. Since this is Analysis, I feel entitled to be a bit sketchy:
Step 1 A smooth map is locally Lipschitz: for every compact set $K \subset U$, there is a constant C such that for $x_1, x_2 \in K$, $\|f(x_1) - f(x_2)\| \leq C\|x_1 - x_2\|$. We call C a *Lipschitz constant*

for f on K.
Step 2 Any open set $U \subset \mathbb{R}^m$ is union of a countable collection of rectangles $\bar{U}_k$. And any set $X \subset \mathbb{R}^m$ is the union of the countable collection $X \cap \bar{U}_k$.
Step 3 Suppose that f has Lipschitz constant C on the rectangle R. Then $f(R)$ is contained in a rectangle of volume less than or equal to $C^m \text{vol}(R)$.
Step 4 Suppose that f has Lipschitz constant C_k on $\bar{U}_k$. Given $\varepsilon > 0$, choose a cover of $X \cap \bar{U}_k$ *by rectangles contained in* $\bar{U}_k$ whose total volume is less than

$$\frac{\varepsilon}{2^k (C_k)^n}.$$

Step 5 Then $f(X)$ is contained in a countable union of rectangles of total volume $< \varepsilon$. Thus $f(X)$ has measure zero.

3. Any open set contains a rectangle R. If the countable union of rectangles R_k contains R then $\sum_k \text{vol}(R_k) \geq \text{vol}(R)$. □

Definition 2.16. (continued). Let $M^m \subset \mathbb{R}^n$ be a manifold. We say the set $X \subset M$ has *measure zero in M* if for every chart $\phi : U \to V$ on M, $\phi(U \cap X)$ has measure zero in $\mathbb{R}^m$. If a property is held by all points of M except for the members of some set of measure zero, we say that the property holds for *almost all $x \in M$.*

In order to check that X has measure zero in M it is enough to check that $\phi(U \cap X)$ has measure zero in $\mathbb{R}^m$ for each of a collection of charts ϕ whose domains together cover M. I leave the proof of this as an exercise (use (i) and (ii) of 2.16).

The use we will make of the property of having measure zero will generally go via the following rather weak consequence:

Proposition 2.17. *If $X \subset M$ has measure zero then $M \setminus X$ is dense in M.*

Proof If not, then some open subset of M is contained in X. Intersecting this set with the domain of some chart, we can transport it to $\mathbb{R}^m$, where it is open. But no non-empty open set in $\mathbb{R}^n$ can have measure zero. □

Example 2.18. Applications of Sard's Theorem

1. If $f : M^m \to N^n$ is a smooth map and $m < n$, then the set of critical values of f is exactly the image of M. hence, by Sard's Theorem, the image of f has measure zero in N. In particular, the image of a *smooth* map $\mathbb{R} \to \mathbb{R}^2$ has measure zero in $\mathbb{R}^2$. Contrast this with the Peano curve, a surjective continuous map from the unit interval to the unit square.

2. Sard's Theorem has an application in "classification problems". For example, any triple of straight lines through $(0, 0)$ in the plane can be transformed by a linear isomorphism to a given fixed triple, say the triple $\{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_1 = x_2\}$.

Proposition 2.19. *The corresponding statement is no longer true if we replace triples of lines through* $(0,0)$ *by quadruples.*

Proof Let $X_1 = \{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_1 = x_2\} \cup \{x_1 = -x_2\}$. The set of quadruples of straight lines through $(0,0)$ contains the set Q of quadruples of the form $\{x_2 = ax_1\} \cup \{x_2 = bx_1\} \cup \{x_2 = cx_1\} \cup \{x_2 = dx_1\}$. There is evidently a bijection $s : Q \to \mathbb{R}^4$. The set $\mathrm{Gl}_2(\mathbb{R})$ of invertible matrices

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$$

is also a 4-dimensional manifold. The subset G' of $\mathrm{Gl}_2(\mathbb{R})$ such that $A(X_1) \in Q$ is easily identifiable — it is the open set consisting of those matrices in which $\alpha_{12} \neq 0, \alpha_{11} + \alpha_{12} \neq 0, \alpha_{11} - \alpha_{12} \neq 0, \alpha_{11} \neq 0$. The composite of the map $G' \to Q$ with the map $s : Q \to \mathbb{R}^4$ is smooth. We will call it f. Because both the source and the target of f have the same dimension, 4, one might hope that f would be surjective — showing that every quadruple in Q is isomorphic to the quadruple X_1. However, the derivative $d_A f : T_A G' \to T_{f(A)}\mathbb{R}^4$ is *never* surjective. For every matrix in the 1-parameter family of matrices $A(t) = \{(1+t)A : t \in \mathbb{R}\}$ takes X_1 to the *same* quadruple of lines; in other words $f(A(t)) = f(A(0))$ for all t, and hence $d_A f(A'(0)) = 0$. As $A'(0) = A \neq 0$, $d_A f$ has non-trivial kernel and therefore is not surjective. We conclude by Sard's theorem that the image of f has measure zero in $\mathbb{R}^4$ — only rather special quadruples of lines are isomorphic to the quadruple X_1. □

□

Classification problems of this type increase in complexity when we go up a dimension. Clearly not *every* quadruple of planes through $(0,0,0)$ in $\mathbb{R}^3$ is linearly isomorphic to the quadruple $X_0 = \{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_3 = 0\} \cup \{x_1 + x_2 + x + 3 = 0\}$: each triple of the four planes making up X_0 intersect only at $(0,0,0)$, and the same must be true of any quadruple of planes to which X_0 is linearly isomorphic.

Exercise 2.20. (i) Show that X_0 *is* linearly isomorphic to any quadruple of planes with this property.

(ii) Let $X_1 = X_0 \cup \{x_1 + 2x_2 + 3x_3 = 0\}$. Each triple of the five planes making up X_1 meets only at $(0,0,0)$. Let $\mathscr{G}$ be the set of quintuples of planes with this property. Is every quintuple in $\mathscr{G}$ linearly isomorphic to X_1? □

2.4 The construction of transverse perturbations

Now we turn to our main theme, the density of transversality.

Theorem 2.21. Thom's Transversality lemma *Suppose that* $F : M \times P \to N$ *is a smooth map, that* $Z \subset N$*, and that* $F \pitchfork Z$*. For* $y \in P$*, denote by* f_y *the map* $M \to N$ *defined by* $f_y(x) = F(x, y)$*. Then for almost all* $y \in P$*,* $f_y \pitchfork Z$*.*

Proof The key to the proof is the following lemma:

Lemma 2.22. *Under these circumstances,* $f_y \pitchfork Z$ *if and only if* y *is a regular value of the projection* $\pi : F^{-1}(Z) \to P$*.*

Proof of lemma: This is Exercise 1.55(iii), and most of the details were covered in Example 1.54 in the case where $M = \mathbb{R}^m$, $P = \mathbb{R}^p$ and $N = \mathbb{R}^n$, so we give just the outline of the proof now.

Step 1:

$$y \text{ is a regular value of } \pi \quad \Leftrightarrow \quad \big(M \times \{y\}\big) \pitchfork F^{-1}(Z).$$

Step 2:

$$\big(M \times \{y\}\big) \pitchfork F^{-1}(Z) \quad \Leftrightarrow \quad f_y \pitchfork Z.$$

Details of steps 1 and 2 are given in Example 1.54(ii) and (i) respectively.

Proof of Theorem: By Sard's theorem, almost all points $y \in P$ are regular values of $\pi_| : F^{-1}(Z) \to P$. Hence, by the lemma, for almost all y, $f_y \pitchfork Z$ □

Now we can prove the elementary transversality theorem. We want to perturb $f : M \to N$ an arbitrarily small amount to make it transverse to Z. To do this, we try to "deform" f by adding to it some map depending on new variables.

Example 2.23. Suppose that $M \subset \mathbb{R}^n$ and that $f : M \to \mathbb{R}$. One way we might get a family F of functions, deforming f, is by taking $\mathbb{R}^n$ as parameter space and defining $F : M \times \mathbb{R}^n \to \mathbb{R}$ by

$$F(x, a) = f(x) + a \cdot x.$$

Clearly $F(x, 0) = f(x)$, or, in other words, $f = f_0$. We can think of F as a family of functions $M \to \mathbb{R}$, parametrised by $a \in \mathbb{R}^n$. □

In this example the target space $\mathbb{R}$ is a vector space and so we can add on to $f(x)$ the term $a \cdot x$. When the target space is not $\mathbb{R}$ but some manifold N without any additive structure, we have to be more ingenious. This is where the tubular neighbourhood theorem comes in.

Example 2.24. Suppose that $f : M \to N$ is a smooth map, and that U_N is a tubular neighbourhood of N in its ambient Euclidean space $\mathbb{R}^p$. Consider the map $G : M \times \mathbb{R}^p \to \mathbb{R}^p$ defined by $G(x, a) = f(x) + a$. In general $G(x, a)$ does not lie in N. However, suppose that U_N is a tubular neighbourhood of N in $\mathbb{R}^p$, with projection $\pi : U_N \to N$. If $G(x, a) \in U_N$ then by applying π we get a point in N. For example, suppose that N is compact, and that $\varepsilon > 0$ is so small that $U_N \supset \cup_{y \in N} B(y, \varepsilon)$. Then for $x \in M$ and $a \in B(0, \varepsilon)$, $G(x, a) \in U_N$. Thus we can define a family

$$F : M \times B(0, \varepsilon) \to N,$$

deforming f, by

$$F(x,a) = \pi(f(x)+a).$$

If N is not compact, we can make use of a function $\eta : N \to \mathbb{R}$ as in Exercise 2.10, such that

$$\bigcup_{y \in N} B(y, \eta(y)) \subset U_N,$$

and define $F : M \times B(0,1) \to N$ by

$$F(x,a) = \pi(f(x) + \eta(f(x))a).$$

These two deformations are by no means the only ways to deform a given function or map, but they illustrate a general approach. In fact they do more: $\square$

Theorem 2.25. *Let $f : M \to N$ be a smooth map, let Z be a submanifold of N, and let $F : M \times B(0,1) \to N$ be the family constructed in Example 2.24. Then*
(i) F is a submersion;
(ii) for almost all $a \in B(0,1)$, the map $f_a : M \to N$ defined by $f_a(x) = F(x,a)$ is transverse to Z.

Proof We already know that $\pi : U_N \to N$ is a submersion, by definition of tubular neighbourhood. To prove that F is a submersion it is enough to show that the map $G : M \times B(0,1) \to U_N$, $G(x,u) = f(x) + \eta(f(x))a$, is a submersion. But this is obvious, if we look closely: with respect to the variable a, G is just a translation plus a dilation by a non-zero scale factor $\eta(f(x))$. So, for any $v \in \mathbb{R}^p$, we have

$$d_{(x,a)}G\left(0, \frac{1}{\eta(f(x))}v\right) = v.$$

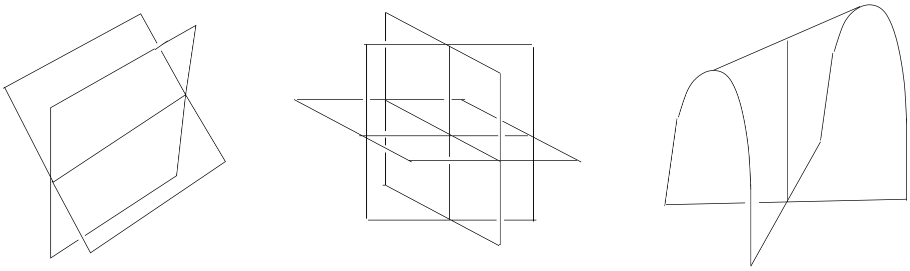
This shows that $F = \pi \circ G$ is a submersion.
(ii) As F is a submersion, it is transverse to Z. It follows by 2.21 that for almost all $a \in B(0,1)$, $f_a \pitchfork Z$. $\square$

Corollary 2.26. *There exists a map arbitrarily close to f, and homotopic to it, which is transverse to Z.*

Proof By 2.25 and 2.16(iii), in every neighbourhood of 0 in $B(0,1)$ there are points a such that $f_a \pitchfork Z$. Choose any such a, as close as to 0 as is desired, and, using F as in the proof of 2.25, define a homotopy $H : M \times [0,1] \to N$ by $H(x,t) = F(x,ta)$. Then $H(x,0) = f(x)$ and $H(x,1) = f_a(x)$. $\square$

In fact 2.26 contains the Weak Transversality Theorem 2.13. So we have achieved the goal of this section.

There are in fact many circumstances where transversality theorems rather like 2.25 are used. Ours is the Weak Transversality Theorem because it is the easiest, only addressing the behaviour of f with respect to a single submanifold. The full strength version is Thom's Transversality Theorem. It assures us that by an arbitrarily small perturbation we can deform a given map to behave well in any one of a potentially unlimited number of different ways. For example, we can deform it so that all the critical points of the deformed map are non-degenerate; or, if we begin with a map $f : M^2 \to N^3$, we can deform it so that in a neighbourhood of each of its singular points the image is diffeomorphic to one of the singular surfaces shown here:



The method always uses the idea that "transversality is dense", but it is an auxiliary map (the *jet extension map* of f) that is required to be transverse to certain submanifolds of an auxiliary space, the *jet space*. You might gain at least some idea of what this is about from

Exercise 2.27. The Whitney umbrella is the image of the map $\mathbb{R}^2 \to \mathbb{R}^3$ $f(x_1, x_2) = (x_1, x_2^2, x_1x_2)$. For each point $x \in \mathbb{R}^2$, the derivative d_xf is a linear map $\mathbb{R}^2 \to \mathbb{R}^3$. We can view this family of linear maps as a (smooth) map $df : \mathbb{R}^2 \to L(\mathbb{R}^2, \mathbb{R}^3)$, where $L(\mathbb{R}^2, \mathbb{R}^3)$ is the space of linear maps from $\mathbb{R}^2$ to $\mathbb{R}^3$. Show

1. f is an immersion except at 0.
2. d_0f lies in the submanifold $\sum^1$ of $L(\mathbb{R}^2, \mathbb{R}^3)$ consisting of linear maps of rank 1.
3. $df : \mathbb{R}^2 \to L(\mathbb{R}^2, \mathbb{R}^3)$ is transverse to $\sum^1$.
4. If $g : \mathbb{R}^2 \to \mathbb{R}^3$ is a polynomial map such that $g(0) = 0$ and d_0g has rank 1, then after a change of co-ordinates in $\mathbb{R}^2$ and $\mathbb{R}^3$, the first degree part of f can be brought to the form $(x_1, x_2) \mapsto (x_1, 0, 0)$.
5. (Harder) If also $dg \pitchfork \sum^1$, then after a further change of coordinates in source and target, the second-degree part of g can be brought to the form $(x_1, x_2) \mapsto (x_1, x_2^2, x_1x_2)$.

□

In the 1940's Hassler Whitney proved a number of theorems about the dimension of Euclidean spaces in which a manifold might be embedded or immersed.

Theorem 2.28. *If M^m is a smooth manifold then there exists an embedding $M \to \mathbb{R}^{2m+1}$.*

Theorem 2.29. *If M^m is a smooth manifold then there exists an embedding $M \to \mathbb{R}^{2m}$.*

Theorem 2.30. *If M^m is a smooth manifold then there exists an immersion $M \to \mathbb{R}^{2m-1}$.*

The first of these is very easy to prove, using Sard's Theorem with a little ingenuity, whereas the second and third are much harder. In fact the first can be strengthened, if we allow some terms we have not defined: in the space of smooth, proper maps $M \to \mathbb{R}^{2m+1}$, embeddings form an open dense set. See Exercises II for a guide to the construction of a proof. In contrast, in the image of a map $f : M^m \to \mathbb{R}^{2m}$, we may encounter transverse self-intersections, where two parts of the image coming from distant parts of M meet one-another transversely, as in Example 1.37. To remove such a self-intersection would require a large modification of the map. So there is no embedding nearby, and embeddings are not dense in the space of proper maps $M \to \mathbb{R}^{2n}$. Because of this, the proof of Whitney's second theorem is much harder than the first.

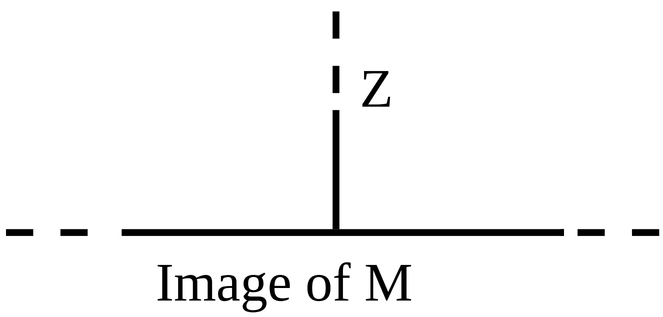
A similar phenomenon, the persistance of transverse intersection, explains the difficulty of the third of these theorems. For example, a map from $M^2 \to \mathbb{R}^{2\times 2-1} = \mathbb{R}^3$ may have non-immersive points like the Whitney umbrella discussed in Exercise 2.27. Because the derivative of such a map is transverse to the set of linear maps of rank 1, the points where its meets with this set (i.e. the non-immersive points of the map) cannot be removed by an arbitrarily small perturbation. So immersions are not dense in the space of maps $M^2 \to \mathbb{R}^3$.

2.5 The stability of transverse intersections

Suppose that $f : M \to N$ is smooth, and transverse to the submanifold Z of N. Suppose that there is a smooth family of maps $F : M \times P \to N$ deforming f, i.e. there is a $y_0 \in P$ such that $f_{y_0} = f$ (notation as in Theorem 2.21). We would like to show that for y in some neighbourhood of y_0 in P,

1. f_y continues to be transverse to Z, and
2. $f_y^{-1}(Z)$ is diffeomorphic to $f^{-1}(Z)$.

Neither is true in quite the simple way one would like. In the following picture $N = \mathbb{R}^3$, $M = \mathbb{R}$ and Z is an open interval situated along the vertical axis.



As both M and Z are 1-dimensional, the only way they can be transverse is by not meeting. In the picture, they do not meet. But if we move the image of M up an arbitrarily small amount, they will meet, and transversality will fail. It is clear that we need to ask for Z to be closed in N, otherwise this sort of situation can always be manufactured.

Now interchange M and Z. Now Z *is* closed, but transversality will still fail if we move M an arbitrarily small amount. So we need to impose some kind of closedness on the image of M also. The simplest thing to ask for is that $f : M \to N$ be proper.

Exercise 2.31. Suppose that $F : M \times P \to N$ is a smooth map and Z is a submanifold of N. For each $y \in P$ we denote by f_y the map $M \to N$ defined by $f_y(x) = F(x, y)$. Let

$$B := \{(x, y) \in M \times P : f_y \text{ is not transverse to } Z \text{ at } x\}.$$

(a) Find B when

(i) $M = (0, \infty) \subset \mathbb{R}$, $P = \mathbb{R}$, $N = \mathbb{R}^3$ and $Z = \{0\} \times \mathbb{R} \times \{0\} \subset N$, and $F : M \times P \to N$ is defined by $F(x, t) = (x - t, 0, 0)$

(ii) $M = \mathbb{R}$, $N = \mathbb{R}^3$, $P = \mathbb{R}$ and $Z = (0, \infty) \times \{(0, 0)\} \subset N$, and $F : M \times P \to N$ is defined by $F(x, t) = (t, x, 0)$.

(b) Check that in both cases, the set $\{y \in P : f_y$ is not tranverse to Z$\}$ is not closed in P. Note that this set is just the projection to P of the set B. □

Exercise 2.32. (i) Suppose that $Z \subset N$ is closed in N, and $F : M \times P \to N$ is smooth. Show that the set B defined in the previous exercise is closed in $M \times P$. Hint: let C be the complement of B. Because Z is closed in N, for every point $w \in N$ there is a neighbourhood U of w in N in which Z has a regular equation g. [3] Then $C \cap F^{-1}(U) = \{(x, y) \in F^{-1}(U) : g \circ f_y \text{ is submersion at } x\} \cup F^{-1}(U \setminus Z)$, and thus is open in $F^{-1}(U)$. Every point $(x, y) \in M \times P$ lies in such an $F^{-1}(U)$, so C is open.

(ii) Show that if Z is closed in N and M is compact then the restriction to B of the projection $M \times P \to P$ is proper, and conclude that the set

$$\{y \in P : f_y \pitchfork Z\}$$

is open in P. □

Proving the second part of the stability of transverse intersections, that they remain essentially unchanged when they are perturbed, is more subtle. How do we show that

[3] This is not the case if Z is not closed. For example, the point $(0, 0, 0)$ in Exercise 2.31(ii) has no such neighbourhood. If Z is closed then points *on* Z have such neighbourhoods, by Corollary 1.31, and points *not on* Z have neighbourhoods whose intersection with Z is empty, in which case as regular equation we can take any equation which is never satisfied.

$f_{y_1}^{-1}(Z) \simeq f_{y_2}^{-1}(Z)$ if y_1 is near y_2 and $f_{y_1} \pitchfork Z$? As yet we have no tools for producing such diffeomorphisms. In fact there is a theorem that assures us that precisely this is the case, but its proof is beyond the scope of these notes[4].

Theorem 2.33. Ehresmann's Fibration Theorem *Suppose that $f : X \to P$ is a proper submersion. Then f is a locally trivial fibre bundle.*

A locally trivial fibre bundle is a map $\pi : X \to P$ with the property that over some neighbourhood of each point in P, π is diffeomorphic to the projection of a product. More precisely, for each $y \in P$ there is a neighbourhood U of y in P and a diffeomorphism $\pi^{-1}(U) \xrightarrow{\phi_U} U \times \pi^{-1}(y)$ such that the diagram

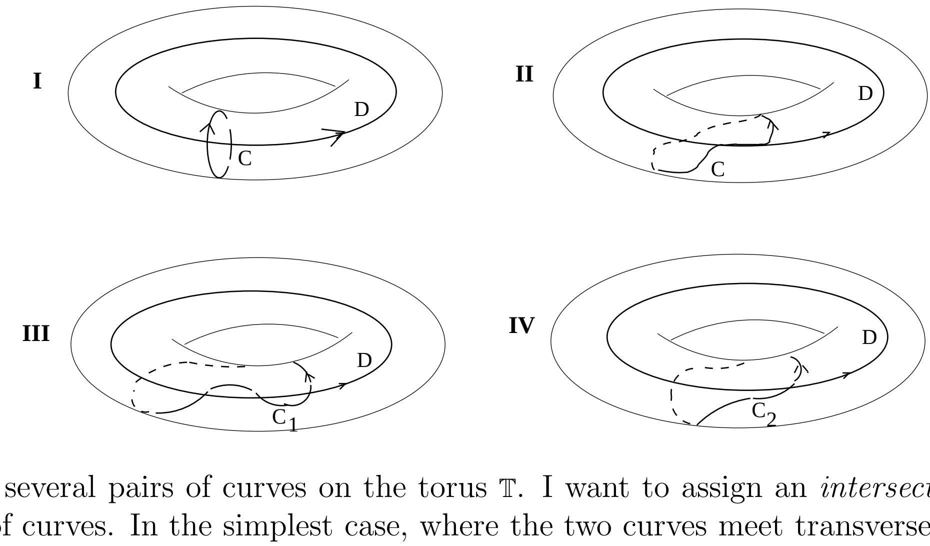
$$\begin{tikzcd} \pi^{-1}(U) \arrow[rr, "\phi_U"] \arrow[dr, "\pi"'] & & U \times \pi^{-1}(y) \arrow[dl, "\pi_1"] \\ & U & \end{tikzcd}$$

where π_1 is projection to the first factor, is commutative.

Exercise 2.34. Suppose that $F : M \times P \to N$ is smooth, and that $f_{y_0} \pitchfork Z$. Suppose that M is compact and Z is closed in N. Prove that there is a neighbourhood U of y in P such that $\pi_| : F^{-1}(Z) \cap (M \times U) \to U$ is a locally trivial fibre bundle, and in particular that for all $y \in U$, $f_y \cap Z$ and $f_y^{-1}(Z) \simeq f_y^{-1}(Z)$. $\square$

3 Oriented Intersection Theory

Consider the following pictures:



We see several pairs of curves on the torus $\mathbb{T}$. I want to assign an *intersection index* to each pair of curves. In the simplest case, where the two curves meet transversely, I want to do it essentially by counting intersection points, as in pictures I, III and IV. But I want it

[4]A sketch of a proof is given in Section 4 of my Lecture Notes for MA5N0 Cohomology, Connections, Curvature and Characteristic Classes.

to be defined for *every* pair of curves - even the pair in II, which meet at infinitely many points. This last seems problematic - until we recall that we have just proved that we can deform one of the curves an arbitrarily small amount, in a homotopy, so that it becomes transverse to the other. So in picture II, we could deform C to C_1, as shown in III, and then count intersection points. But we could also deform it to C_2, as in IV, and then count — getting a different answer. The way to make the answers the same in III and IV is to count each intersection point with a sign — i.e. as ± 1 — and then add. We have to have some convention to count with signs; this is what *orientation* is about.

In the case of curves meeting on a surface, it's not too hard. We agree on

1. an orientation for C and for D
2. an order in which we calculate the intersection — i.e. C and D or D and C, and
3. a rule which determines, at every point on the surface, a "positive direction of rotation".

Then the procedure for assigining a sign to each transverse intersection point is as follows:

Rule 3.1. let v_1 and v_2 be forward pointing tangent vectors to C and D at x. Because $C \pitchfork D$ at x, v_1 and v_2 are not parallel. Hence there is a shortest rotation taking v_1 to v_2. If this is in the positive direction, count $+1$. If it is in the negative direction, count -1.

The reason this rule is so good is that where two intersection points can be brought together and annihilate one another in a homotopy, then it turns out that *they must have opposite sign*, and so the sum of the "oriented intersection numbers" is the same before and after the homotopy. This will be shown below.

An orientation on a curve is unproblematic — it's just an arrow, a direction, and there are two possibilities. But what about on a surface, or in higher dimensions still? The general definition is initially a little surprising, and we have to work to reconcile it with our intuition.

Definition 3.2. (i) Let V be a vector space over $\mathbb{R}$. Say two ordered bases $v_1, \ldots, v_m$ and $w_1, \ldots, w_m$ are *equivalent* if the change of basis matrix [5] from one to the other has positive determinant. There are thus two equivalence classes of bases. An *orientation* of V is the choice of one of these two, called the class of *positive bases*. So there are two possible orientations of V. Note that to specify an orientation it is enough to declare one ordered basis to be positive.

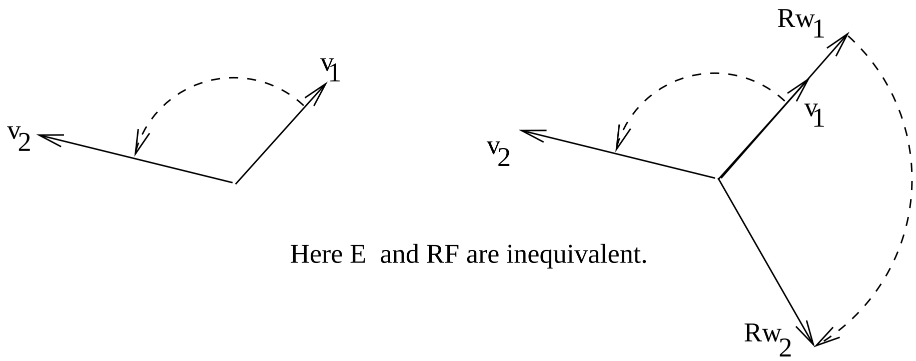
Example 3.3. 1. On $\mathbb{R}^n$ we define the *standard orientation* by declaring that the basis $v_1, \ldots, v_n$ is positive if and only if $\det(v_1, \ldots, v_n)$ is positive. In all of what follows, we will orient $\mathbb{R}^n$ in this way.

[5] If $E = e_1, \ldots, e_m$ and $E' = e'_1, \ldots, e'_m$ are ordered bases for the same vector space V, the change of basis matrix $[I]^E_{E'}$ has as its j'th column the expression of the vector e_j with respect to the basis E'. The crucial property of this matrix is that if $v \in V$ and we denote by $[v]_E$ and $[v]_{E'}$ the columns of coefficients of v with respect to bases E and E', then $[v]_{E'} = [I]^E_{E'}[v]_E$.

2. Let V be any 3-dimensional vector space and let $E := e_1, e_2, e_3$ be any basis. As *ordered* bases, E and $E' := e_3, e_1, e_2$ are different: however, the change of basis matrices $[I]_{E'}^{E}$ and $[I]_{E}^{E'}$ are
$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$
with determinant 1, so the bases E and E' are equivalent. On the other hand, if E'' is the ordered basis e_1, e_3, e_2, then the change of basis matrices $[I]_{E''}^{E}$ and $[I]_{E}^{E''}$ are both equal to
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$
with determinant -1. So E and E'' are inequivalent.

3. If V is a 1-dimensional space, an orientation, in the sense just defined, allows us to draw an arrow: it points in the direction of the (unique) vector of a positive basis. Conversely, an arrow allows us to determine an orientation: a basis v is positive if v points in the direction of the arrow.

4. If V is a 2-dimensional space, an orientation allows us to determine a "positive" direction of rotation at each point: it is the direction of shortest rotation from the first vector of a positive ordered basis to the second. This choice of direction does not depend on the positive ordered basis chosen. For let $E = v_1, v_2$ and $F = w_1, w_2$ be two ordered bases. Let R be a rotation taking w_1 to a positive multiple of v_1. The basis $RF := Rw_1, Rw_2$ is equivalent to the basis w_1, w_2, since a rotation has positive determinant. So
$$E \sim F \qquad \Leftrightarrow \qquad E \sim RF.$$
The change of basis matrix $[I]_{RF}^{E}$ has the form
$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ 0 & \lambda_{22} \end{pmatrix}.$$
As $\lambda_{11} > 0$, this matrix has positive determinant if and only if $\lambda_{22} > 0$. But it is clear from the diagram below that this is also the condition that the shortest rotation from Rw_1 to Rw_2 should be in the same direction as the shortest rotation from v_1 to v_2.



From now on we will use "basis" to denote "ordered basis". □

Definition 3.2. (continued) (ii) Let W_1 and W_2 be oriented vector spaces. An isomorphism $\ell : W_1 \to W_2$ *preserves orientation* if for some (and hence for any) positive basis $e_1, \ldots, e_n$ of W_1, $\ell(e_1), \ldots, \ell(e_n)$ is a positive basis for W_2. If ℓ does not preserve orientation, then it *reverses* it.

Exercise 3.4. (a) Show that ℓ preserves orientation if and only if for some (and hence for any) positive bases E_1 of W_1 and E_2 of W_2, $\det[\ell]_{E_2}^{E_1} > 0$.
(b) If $W_1 = W_2$ then show that the sign of $\det[\ell]_E^E$ is independent of the choice of basis E. □

Definition 3.2. (continued)(iii) Let V_1 and V_2 be open sets in $\mathbb{R}^m$. A diffeomorphism $h : V_1 \to V_2$ *preserves orientation* at a point y in V_1 if $d_y h : \mathbb{R}^m \to \mathbb{R}^m$ preserves orientation. If h preserves orientation at every point $y \in V_1$ then we simply say that it preserves orientation.

Now, finally, we are in a position to define an orientation on a manifold. Essentially, an orientation on a manifold M is an orientation for each tangent space $T_x M$, but which must, in some sense, 'vary continuously' from one tangent space point to another. There are several equivalent ways of making this vague requirement precise.

For the first, think of an orientation on each tangent space as a rule for deciding whether a basis of that space is positive or negative. We don't know yet what it means to say that a *rule* varies continuously from one tangent space to another. Less problematic is the notion of a continuous family of bases. For example, for each point x in the hemisphere $\{(x_1, x_2, x_3) \in S^2 : x_3 > 0\}$, the vectors

$$v_1(x) = (-x_3, 0, x_1), v_2(x) = (0, -x_3, x_2)$$

form a basis for $T_x S^2$. Evidently v_1 and v_2 vary continuously, so it is reasonable to say that the basis $E(x)$ consisting of $v_1(x), v_2(x)$ is a continuous family of bases. An orientation of a manifold is an orientation for each tangent space $T_x M$, which "knit together" in the sense that they *all give the same sign to each basis in a continuous family of bases.* More precisely,

Definition 3.5. Let $M \subset \mathbb{R}^n$ be a manifold.
(i) A *tangent vector field* (usually abbreviated to *vector field*) on $U \subset M$ is a map $v : U \to \mathbb{R}^n$

such that for every $x \in U$, $v(x) \in T_xM$.

(ii) A *continuous family of bases*, or *tangent frame field*, on a set $U \subset M$ is an m-tuple of continuous tangent vector fields on U, $v_1, \ldots, v_m$, such that for every $x \in U$, $v_1(x), \ldots, v_m(x)$ is a basis for T_xM.
Example Suppose $\phi : U \to V$ is a chart on M^m. Let $e_1, \ldots, e_m$ be the standard basis for $\mathbb{R}^m$. For each j and for each $x \in U$, let

$$v_j(x) := (d_x\phi)^{-1}(e_j).$$

Then v_j is a tangent vector field on U.

Moreover, since $e_1, \ldots, e_m$ are a basis for $\mathbb{R}^m$, $v_1(x), \ldots, v_m(x)$ is a basis for T_xM for each $x \in U$. Thus $v_1, \ldots, v_m$ is a continuous (even smooth) tangent frame field on U.

(iii) An orientation of a manifold M is a choice of orientation for each T_xM with the following property: whenever $U \subset M$ is connected and $v_1, \ldots, v_m$ is a continuous tangent frame field on U, then *either* for all x in U, $v_1(x), \ldots, v_m(x)$ is a positive basis for T_xM, *or* for all x in U $v_1(x), \ldots, v_m(x)$ is a negative basis for T_xM.

Remark 3.6. If $v_1, \ldots, v_m$ and $v'_1, \ldots, v'_m$ are tangent frame fields on the connected open set $U \subset M$, let $E(x)$ denote the basis $v_1(x), \ldots, v_m(x)$ for T_xM, and let $E'(x)$ denote the basis $v'_1(x), \ldots, v'_m(x)$. The determinant $\det[I]^{E(x)}_{E'(x)}$ varies smoothly, and so always has the same sign on U. Hence, we need only check that (iii) holds for *one* frame field on U — if it does, then it holds for every frame field.

Consider once again the sphere $S^2 \subset \mathbb{R}^3$. We define an orientation on each T_xS^2 by the following rule:

the basis v_1, v_2 of T_xS^2 is positive if the basis x, v_1, v_2 of $\mathbb{R}^3$ is positive[6].

Let us see that this meets the requirement of 3.5(iii). Suppose that $U \subset S^2$ is connected and that v_1, v_2 are continuous vector fields on U such that for all $x \in U$, $v_1(x), v_2(x)$ is a basis for T_xS^2. Then the function $\det(x, v_1(x), v_2(x))$ is continuous, and nowhere zero on U. Hence it has the same sign everywhere on U. If it is everywhere positive, then $v_1(x), v_2(x)$ is a positive basis of T_xS^2 for all x in U. If it is everywhere negative, then $v_1(x), v_2(x)$ is a negative basis of T_xS^2 for all x in U. Thus, we have succeeded in orienting S^2.

Not every manifold can be oriented. Below we will see that the Möbius strip cannot. We say M is *orientable* if it can be oriented, and *non-orientable* otherwise.

Exercise 3.7. Here is a slight variant on the preceding definition. It is based on the intuitively appealing idea that tangent spaces at nearby points are themselves close enough to one another that orthogonal projection from one to the other is an isomorphism.

[6] with respect to the standard orientation defined in 3.3(1)

(i) Show that if $M \subset \mathbb{R}^n$ is smooth, then for each $x_0 \in M$, there is a connected neighbourhood U_{x_0} of x_0 such that for all $x \in U_{x_0}$, the orthogonal projection

$$\pi_{x_0} : T_x M \to T_{x_0} M$$

is an isomorphism.
(ii) Show that an assignment of orientations to each tangent space $T_x M$ satisfies the compatibility condition of 3.5(iii) if and only if for all $x_0 \in M$ and $x \in U_{x_0}$, each of the linear isomorphisms $\pi_{x_0} : T_x M \to T_{x_0} M$ preserves orientation. □

Proposition 3.8. *(i) Let M^m be a manifold. Then M is orientable if and only if there is an atlas $\{\phi_\alpha : U_\alpha \to V_\alpha \subset \mathbb{R}^m\}$ such that whenever $U_\alpha \cap U_\beta \neq \emptyset$, the crossover diffeomorphism*

$$\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$$

preserves orientation.

We will call an atlas with this property an *orienting atlas*.
Proof Suppose first that M is oriented. Take any atlas $\{\phi_\alpha : U_\alpha \to V_\alpha\}$ in which all of the U_α are connected. Let E be the standard basis $e_1 = (1, 0, \ldots, 0), \ldots, e_m = (0, \ldots, 0, 1)$ for $\mathbb{R}^m$. For each α, the vectors $d_x\phi^{-1}(e_1), \ldots, d_x\phi^{-1}(e_m)$ form a frame field, E_α, on U_α. Thus, either they are a positive basis for $T_x M$ for all $x \in U_\alpha$, or a negative basis for $T_x M$ for all $x \in U_\alpha$. In the second case, replace ϕ_α by $R \circ \phi_\alpha$, where $R : \mathbb{R}^m \to \mathbb{R}^m$ is a reflection. Now *all* of the frame fields E_α are everywhere positive. Note that $[d_x\phi_\alpha]_E^{E_\alpha}$ is the identity matrix I_n. Since

$$\begin{aligned} [d_{\phi_\alpha(x)}(\phi_\beta \circ \phi_\alpha^{-1})]_E^E &= [d_x\phi_\beta]_E^{E_\beta}[I]_{E_\beta}^{E_\alpha}[d_x\phi_\alpha]^{-1} \\ &= [I]_{E_\beta}^{E_\alpha} \end{aligned}$$

it follows that $\phi_\beta \circ \phi_\alpha^{-1}$ preserves orientation. Thus our atlas is an orienting atlas.

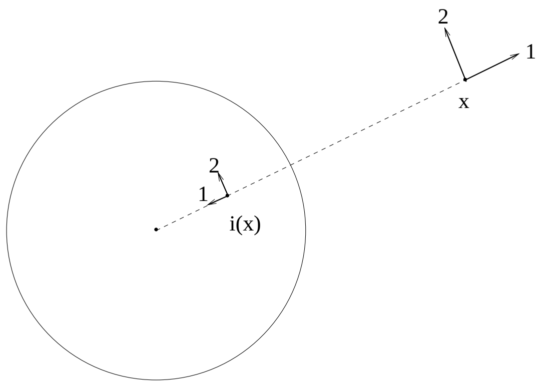
Conversely, given an orienting atlas, orient each $T_x M$ so that $d_x\phi_\alpha : T_x M \to T_x\mathbb{R}^m$ preserves orientation. Because all of the crossover maps preserve orientation, this instruction is unequivocal. Now each frame field E_α, is everywhere positive. By remark 3.6, this shows that our orientations satisfy the local compatibility condition. □

Example 3.9. Consider the atlas for S^2 consisting of the two stereographic projections $\phi_N : S^2 \setminus \{N\} \to \mathbb{R}^2$ and $\phi_S : S^2 \setminus \{S\} \to \mathbb{R}^2$ discussed in Example 1.2. The crossover map $\phi_S \circ \phi_N^{-1} : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}^2 \setminus \{0\}$ is inversion i in the unit circle, given given in coordinates by $i(x_1, x_2) = (\frac{x_1}{x_1^2+x_2^2}, \frac{x_2}{x_1^2+x_2^2})$. This map does not preserve orientation: its derivative at (x_1, x_2) has matrix

$$\frac{1}{x_1^2 + x_2^2}\begin{pmatrix} x_1^2 - x_2^2 & -2x_1x_2 \\ -2x_1x_2 & x_2^2 - x_1^2 \end{pmatrix}$$

with negative determinant. So $\{\phi_N, \phi_S\}$ is not an orienting atlas. If we compose ϕ_N with a reflection such as $r(x_1, x_2) = (-x_1, x_2)$, we get a new atlas, $\{r \circ \phi_N, \phi_S\}$ which *is* an orienting atlas.

The fact that inversion in the unit circle reverses orientation can be understood geometrically: if x grows radially, $x/\|x\|$ shrinks radially. If x moves anticlockwise around the origin, so does $x/\|x\|$. So the image, under the derivative $d_x i$, of the positive basis shown, is as shown in the following diagram.



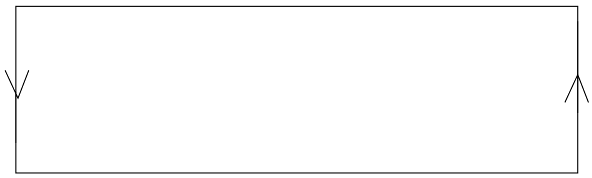
Philosophical Remark 3.10. Formally, the way we obtain an orientation from an orienting atlas in the second paragraph of this proof has some similarity with our definition of *angle* on the sphere S^n, in the third part of Example 1.2. In both cases we define something on the manifold by means of charts. For this definition to be unequivocal, we have to show that we get the same answer whichever chart we use. In Example 1.2, we defined "angle" between two curves on the sphere S^n by using stereographic projection (ϕ_N or ϕ_S) to transport the two curves to $\mathbb{R}^n$, where we measure the angle between them in the usual way. This made sense because the crossover diffeomorphisms $\phi_N \circ \phi_S^{-1}$ and $\phi_S \circ \phi_N^{-1}$ preserve angles, so that the answer we get applying ϕ_N is the same as the answer we get applying ϕ_S. Of course, there we couldn't use just *any* charts: our definition relies on the choice of two particular charts. We say that our atlas $\{\phi_N, \phi_S\}$ equips S^n with a *conformal structure*, a structure in which we can measure angles. Returning to orientation: once we have an atlas in which all crossover diffeomorphisms preserve orientation, we can use the charts to orient the manifold in an unequivocal way. We might say that such an atlas equips M with an "oriented structure", though in practice the term is not used.

Example 3.11. (i) Suppose that $f : \mathbb{R}^{m+1} \to \mathbb{R}$ has $t \in \mathbb{R}$ as regular value, and $f^{-1}(t) \neq \emptyset$. Then we can orient $M := f^{-1}(t)$ as follows: the basis $v_1, \ldots, v_m$ of T_xM is positive if the basis $\nabla f(x), v_1, \ldots, v_m$ of $\mathbb{R}^{m+1}$ is positive. Note that this generalises the method used above to orient S^2.

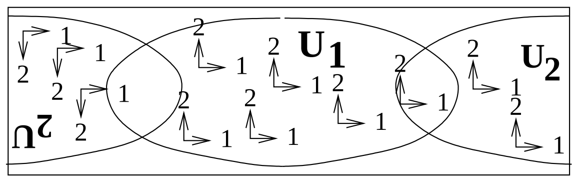
(ii) More generally, if $f : \mathbb{R}^{m+c} \to \mathbb{R}^c$ has t as a regular value and $f^{-1}(t) \neq 0$ then we can orient $M := f^{-1}(t)$ as follows: the basis $v_1, \ldots, v_m$ of T_xM is positive if the basis $\nabla f_1(x), \ldots, \nabla f_c(x), v_1, \ldots, v_m$ of $\mathbb{R}^{m+c}$ is positive.

(iii) The Möbius strip is not orientable. To see this, think of it as obtained from a rectangle by gluing together opposite ends after a half-twist, as shown here. Suppose it is orientable. Because of the half twist, the smooth frame field shown on the connected set U_2 in (iii) must appear as shown in (ii) when U_2 is depicted on M. But now we have a contradiction: the frame fields on U_1 and U_2 must each be everywhere positive or everywhere negative. However, the bases they provide in T_xM are equivalent for x in the right hand overlap of U_1 and U_2, but inequivalent for x in the left hand overlap.

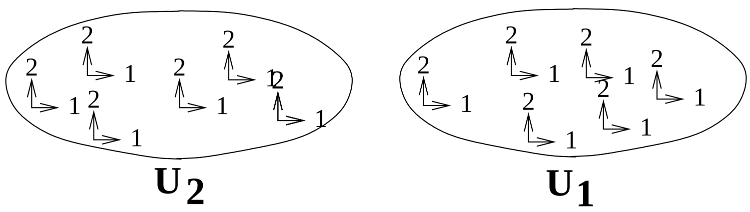
(i)



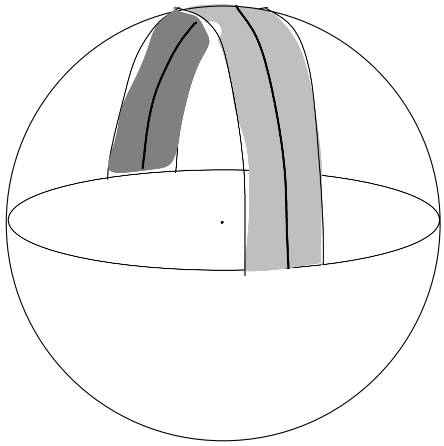
(ii)



(iii)



(iv) The non-orientability of the Möbius strip has an interesting consequence for the manifold M we obtained as the image of the quadratic map $S^2 \to \mathbb{R}^6$ in Example 1.41(ii): it is impossible to define it by a single set of regular equations. That is, it is impossible to find a set U open in $\mathbb{R}^6$ containing M on which there is a smooth map g to $\mathbb{R}^4$ such that $M = g^{-1}(0)$ and 0 is a regular value of g. For I claim that M contains a Möbius strip, and is therefore non-orientable. The claim is essentially proved by the following picture.



For since f identifies antipodal points on the sphere, we get all of M as the image of the closed upper hemisphere. The map f is 1-1 on the interior of the hemisphere, and identifies antipodal boundary points. Therefore the image of the strip shown is a Möbius strip in M. Because M is therefore not orientable, by (ii) above it cannot be globally defined by regular equations.

(vi) **Product Orientation** If M and N are oriented manifolds, the product $M \times N$ acquires an orientation as follows. First, for $(x, y) \in M \times N$,

$$T_{(x,y)}(M \times N) = T_xM \times T_yN.$$

We determine an orientation of $T_{(x,y)}(M \times N)$ by taking a positive basis $E = v_1, \ldots, v_m$ for T_xM and a positive basis $F = w_1, \ldots, w_n$ for T_yN and declaring that the basis

$$(v_1, 0), \ldots, (v_m, 0), (0, w_1), \ldots, (0, w_n)$$

(which we denote by (E, F)) is a positive basis for $T_{(x,y)}M \times N$. If we chose different positive bases E' for T_xM and F' for T_yN, then the change of basis matrix $[I]^{(E,F)}_{(E',F')}$ satisfies

$$[I]^{(E,F)}_{(E',F')} = \begin{pmatrix} [I]^E_{E'} & 0 \\ 0 & [I]^F_{F'} \end{pmatrix}$$

and thus

$$\det[I]^{(E,F)}_{(E',F')} = \det[I]^E_{E'} \det[I]^F_{F'} > 0.$$

So (E, F) and (E', F') are equivalent to one another, and the product orientation is well defined. $\square$

Exercise 3.12. Prove that the procedures defined in (i) and (ii) of Example 3.11 do indeed define orientations. $\square$

Remark 3.13. It is important to reconcile our definition of orientation with prior intuition. The Möbius strip is famously "one-sided". What does this have to do with its non-orientability? Example 3.11(i) suggests a link. If a surface $S \subset \mathbb{R}^3$ is orientable, we can use the orientation to distinguish between the two sides of the surface. We do this by defining a nowhere vanishing normal vector field n, which we use to distinguish the side it points into from the side it points away from. The vector field n is simply defined to be the unique unit vector in $T_xS^\perp$ such that if v_1, v_2 is a positive basis then $n(x), v_1, v_2$ is a positive basis for $\mathbb{R}^3$. Equally, $n(x) = v_1 \times v_2 / \|v_1 \times v_2\|$. Reciprocally, if S is "two-sided", then selecting one of the two sides, we define a nowhere-vanishing normal vector field by choosing at each point the unit normal pointing into the chosen side. This can then be used to orient the surface, playing the same role as the gradient vector ∇f in Example 3.11(i).

Recall that a zero-dimensional manifold is just a collection of discrete points.

Definition 3.14. *An orientation of a zero-dimensional manifold is simply the assignation of a sign, $+$ or $-$, to each point.*

Thus, exactly as for a connected positive-dimensional orientable manifold, a connected 0-dimensional manifold has two possible orientations. This definition of orientation is a bit difficult to reconcile *propter hoc* with the definition of orientation for a positive dimensional manifold, but we will shortly see that it fits beautifully into the intersection theory we will develop.

Now, returning to the original theme of this chapter, we will define the oriented intersection number of two compact oriented submanifolds of complementary dimension which meet transversely in an oriented manifold.

Definition 3.15. (i) Suppose U and V are complementary subspaces of a finite dimensional real vector space W. Suppose that all three spaces are oriented. If $u_1, \ldots, u_m$ is a positive basis for U and $v_1, \ldots, v_{n-m}$ is a positive basis for V, then we say that U and V have *positive intersection* if $u_1, \ldots, u_m, v_1, \ldots, v_{n-m}$ is a positive basis for W, and *negative intersection* if it is a negative basis.

(ii) Suppose that M^m and Z^{n-m} are oriented submanifolds of the oriented manifold N^n, with M compact and Z closed. Suppose also that $M \pitchfork Z$. Then we define the *oriented intersection number of M and Z at x*, $(M \cdot Z)_x$, as follows:

$$(M \cdot Z)_x = \begin{cases} +1 & \text{if } T_xM \text{ and } T_xZ \text{ have positive intersection in } T_xN \\ -1 & \text{if } T_xM \text{ and } T_xZ \text{ have negative intersection in } T_xN \end{cases} \tag{3.1}$$

We define the *oriented intersection number of M and Z in N*, $(M \cdot Z)_N$, as

$$(M \cdot Z)_N = \sum_{x \in M \cap Z} (M \cdot Z)_x.$$

(iii) A mild extension of the previous definition: we do not suppose that M^m is a submanifold of N^n, but simply that $f : M \to N$ is a smooth map which is transverse to Z^{n-m}

(with M, N and Z oriented, and M compact and Z closed as before). As the dimensions are complementary, for each $x \in f^{-1}(Z)$ the derivative $d_x f : T_x M \to T_{f(x)} N$ must be an injection. We define

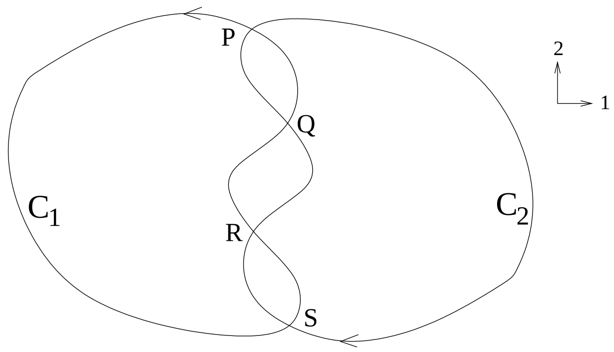
$$(f \cdot Z)_x = \begin{cases} +1 & \text{if } d_x f(T_x M) \text{ and } T_{f(x)} Z \text{ have positive intersection in } T_{f(x)} N \\ -1 & \text{if } d_x f(T_x M) \text{ and } T_{f(x)} Z \text{ have negative intersection in } T_{f(x)} N \end{cases} \quad (3.2)$$

We define the oriented intersection number of f and Z, $(f \cdot Z)_N$, as

$$(f \cdot Z)_N = \sum_{x \in f^{-1}(Z)} (f \cdot Z)_x.$$

Observe that (ii) is the special case of (iii) in which f is the inclusion of the submanifold M into N.

Remark 3.16. Recall that an oriented 0-dimensional manifold is just a collection of discrete points with a sign ($+$ or $-$) attached to each one. In the circumstance described in 3.15(ii), $X \cap Z$ acquires a natural orientation - simply assign to each point the sign defined by (3.1). Thus, for example in the following picture, in which $\mathbb{R}^2$ has the standard orientation, $C_1 \cap C_2 = -P + Q - R + S$, and $(C_1 \cdot C_2)_{\mathbb{R}^2} = 0$. Note that $C_2 \cap C_1 = P - Q + R - S$; with this convention, intersection has become non-commutative!



By means of a similar convention, using (3.2), we can orient $f^{-1}(Z)$ in 3.15(iii).

Exercise 3.17. (i) Show, as we must do in order for the last parts of Definition 3.15 to make sense, that with the given hypotheses $f^{-1}(Z)$ is finite.

(ii) Equally necessary, show also that in part (i) of Definition 3.15, the sign of the basis $u_1, \ldots, u_m,\ v_1, \ldots, v_{n-m}$ of W is independent of which positive bases $u_1, \ldots, u_m$ of U and $v_1, \ldots, v_{n-m}$ of V we choose. One way to do this is the following: suppose that E_U and E'_U are two bases for U, and E_V and E'_V are two bases for V. Denote by E the basis for W obtained by putting together E_U and E_V (as described in part (i) of the definition), and by E' the basis of W obtained by putting together E'_U and E'_V. Express the matrix $[I]^E_{E'}$ in terms of the matrices $[I]^{E_U}_{E'_U}$ and $[I]^{E_V}_{E'_V}$. $\square$

It remains to extend our definitions to cover the case where M and Z are *not* transverse. As I said earlier, the idea is to move M in a homotopy, to a position where it is transverse to Z, and then use definition 3.15(ii). In fact the language of (iii) allows a more flexible description. In these terms, we define an oriented intersection number $(f \cdot Z)_N$ even when f is not transverse to Z, by perturbing f in a homotopy to a map f_1 which is transverse to Z, calculating $(f_1 \cdot Z)_N$ using part (iii) of Definition 3.15, and then defining $(f \cdot Z)_N$ to be $(f_1 \cdot Z)_N$. But for this to make sense, we have to show that it doesn't matter which transverse perturbation of f we take. The relation of homotopy is transitive, so if f_1 and f_2 are both homotopic to f, then f_1 is homotopic to f_2, and thus it will be enough to show

Theorem 3.18. *If $f_1 : M \to N$ is homotopic to $f_2 : M \to N$, and both f_1 and f_2 are transverse to Z, then $(f_1 \cdot Z)_N = (f_2 \cdot Z)_N$.*

The proof of this will occupy some time. As a first step, recall that a homotopy between $f_1 : M \to N$ and $f_2 : M \to N$ is a smooth map $F : M \times [0, 1] \to N$ such that $F(x, 0) = f_1(x)$ and $F(x, 1) = f_2(x)$. In particular, its domain is *not* a manifold. We begin by slightly enlarging the definition of manifold to incorporate such objects.

Definition 3.19. (i) H^m is the half-space $\{(x_1, \ldots, x_m) \in \mathbb{R}^m : x_m \geq 0\}$. The *boundary* of H^m, ∂H^m, is the set $\{(x_1, \ldots, x_m) \in \mathbb{R}^m : x_m = 0\}$.

(ii) A subset $M \subset \mathbb{R}^n$ is an m-dimensional *manifold with boundary* if every point in M has a neighbourhood diffeomorphic to an open set in H^m. If M^m is an m-dimensional manifold with boundary, the *boundary* of M, which we denote by ∂M, is the set of points $x \in M$ such that in some chart $\phi : U \to H^m$, $\phi(x) \in \partial H^m$.

Example 3.20. (i) The closed unit ball $\bar{B}^m := \{x \in \mathbb{R}^m : \|x\| \leq 1\}$ is an m-dimensional manifold with boundary, with $\partial \bar{B}^m = S^{m-1}$. The intervals $[a, b]$ and $[a, b)$ are 1-dimensional manifolds with boundary. $\partial[a, b] = \{a, b\}$, and $\partial[a, b) = \{a\}$.

(ii) Every manifold is a manifold-with-boundary (though with empty boundary), since every open set in $\mathbb{R}^m$ is diffeomorphic to an open set in H^m. On the other hand, a manifold with boundary is only a manifold if its boundary is empty.

(iii)**Exercise** If M^m is a smooth manifold and $f : M \to \mathbb{R}$ has t_0 as regular value, then $f^{-1}((-\infty, t_0])$ and $f^{-1}([t_0, \infty))$ are both manifolds with boundary, and each has boundary equal to $f^{-1}(t_0)$. Hint: $f^{-1}((-\infty, t_0)$ is open in M, so each point of $f^{-1}((-\infty, t_0)$ has a neighbourhood diffeomorphic to an open set in $\mathbb{R}^m$, and thus to an open set in H^m. For points in $f^{-1}(t_0)$, use the local normal form of a submersion.

(iii) Since $\mathbb{R}^m \times H^n = H^{m+n}$, the cartesian product of an m-dimensional manifold and an n-dimensional manifold with boundary is an $(m + n)$-dimensional manifold with boundary. In particular this applies to $M \times [0, 1]$. $\square$

Proposition 3.21. *(i) Suppose that M^m is a manifold with boundary and that for some chart $\phi : U \to V \subset H^m$, $\phi(x) \in \partial H^m$. Then for every other chart ψ on M whose domain contains x, $\psi(x) \in \partial H^m$.*

(ii) ∂M, if not empty, is a manifold of dimension $m-1$.

Proof (i) If for some chart ϕ, $\phi(x)$ is not in ∂H^m, then there is a neighbourhood of $\phi(x)$ in H^m which is in fact open in $\mathbb{R}^m$. By the inverse function theorem, if U is open in $\mathbb{R}^m$ and $\phi : U \to V \subset \mathbb{R}^m$ is a diffeomorphism then V is open in $\mathbb{R}^m$. It follows that for *any* chart ψ, $\psi(x)$ has a neighbourhood in H^m which is open in $\mathbb{R}^m$. Hence $\psi(X) \notin \partial H^m$.

(ii) For any chart $\phi : U \to V \subset H^m$, $\phi(U \cap \partial M) = V \cap \partial H^m$ (we are using (i) here). As $\phi_{|U\cap\partial M}$ has inverse $(\phi^{-1})_{|V\cap\partial H^m}$, and $V \cap \partial H^m$ is an open set in $\mathbb{R}^{m-1}$, each point in ∂M has a neighbourhood diffeomorphic to an open set in $\mathbb{R}^{m-1}$. □

Notice that we have shown quite clearly that ∂M is a manifold *without boundary.*

Example 3.22. (i) If M is a manifold and N is a manifold with boundary then $M \times N$ is a manifold with boundary, and $\partial(M \times N) = M \times \partial N$.

(ii) If both M and N are manifolds with (non-empty) boundary then $M \times N$ is *not* a manifold with boundary. Consider for example the product of two closed intervals. □

If M is an oriented manifold with boundary then ∂M inherits an orientation from M, the *boundary orientation.* The construction of the boundary orientation is very similar to the construction of an orientation using the gradient vector of a defining equation in Example 3.11(i).

Definition 3.23. (i) Let $y \in \partial H^m$. We set

$$T_y H^m = T_y \mathbb{R}^m$$

$$T_y^{\text{out}} H^n = \{(v_1, \ldots, v_n) \in T_y H^n : v_n < 0\}.$$

(ii) If M^m is a manifold with boundary and $x \in \partial M$, let ϕ be a chart on M around x and let $\phi(x) = y$. Define $T_x M = (d_y \phi^{-1})(T_y H^m)$ and $T_x^{\text{out}} M = (d_y \phi^{-1})(T_y^{\text{out}} H^m)$. We call the vectors of $T_x^{\text{out}} M$ *outward pointing tangent vectors.*

Note that here ϕ^{-1} is *not* smooth (old) at the points of $V \cap \partial H^m$ (where the domain of ϕ^{-1} contains no set open in $\mathbb{R}^m$), so we are cheating a little when we speak of its derivative $d_y \phi^{-1}$. In order to speak of $d_y \phi^{-1}$ we really need to make use of a local smooth (old) extension of ϕ^{-1} on the ambient $\mathbb{R}^m$, and use *its* derivative [7] at y. This is OK, provided this derivative

[7] Of course, in practice ϕ^{-1} will usually be given in terms of local ambient coordinates, and so we will not have to struggle to find such a local extension.

is independent of the choice of local extension. In fact this is easy to see. For every vector $v \in T_y\mathbb{R}^m = \mathbb{R}^m$, either v or $-v$ is equal to

$$\lim_{t \to 0+} \frac{\gamma(t) - \gamma(0)}{t} \tag{3.3}$$

(one sided limit), for some smooth curve $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ taking 0 to y and taking $[0, \varepsilon)$ to V. Thus if F is any smooth (old) extension of ϕ^{-1} around y, then either $d_yF(v)$ or $d_yF(-v)$ is equal to

$$\lim_{t \to 0} \frac{F(\gamma(t)) - F(\gamma(0))}{t} = \lim_{t \to 0+} \frac{F(\gamma(t)) - F(\gamma(0))}{t} = \lim_{t \to 0+} \frac{\phi^{-1}(\gamma(t)) - \phi^{-1}(\gamma(0))}{t}$$

and thus is independent of choice of F. As $d_yF(-v) = -d_yF(v)$, for every vector v the value of $d_yF(v)$ does not depend on the choice of F.

In fact this still leaves the question of whether our definition is independent of the choice of chart ϕ, exactly as in the definition of T_xM for a manifold (i.e. without boundary), in 1.11. But this can be dealt with by the same argument as in 1.11, adapted by the trick we've just used to deal with half-spaces.

With the definition of the outward pointing tangent vectors, once again there is an issue of well-definedness: this is resolved by

Lemma 3.24. *Suppose that $x \in \partial M$ and that ϕ is a chart on M around x, with $\phi(x) = y \in \partial H^m$. Then*

$$(d_y\phi^{-1})(T_y^{out}H^m) = \{-\gamma'(0+) : \gamma : [0, \varepsilon) \to M \text{ is such that } \gamma(0) = x\} \setminus T_x\partial M.$$

(where $\gamma'(0+)$ means the one-sided limit as in (3.3)). □

Definition/Proposition 3.25. Suppose M is an oriented manifold with boundary. The *boundary orientation* on ∂M is defined by the criterion

$v_1, \ldots, v_{m-1}$ is a positive basis for $T_x\partial M$ if $n, v_1, \ldots, v_{m-1}$ is a positive basis for T_xM

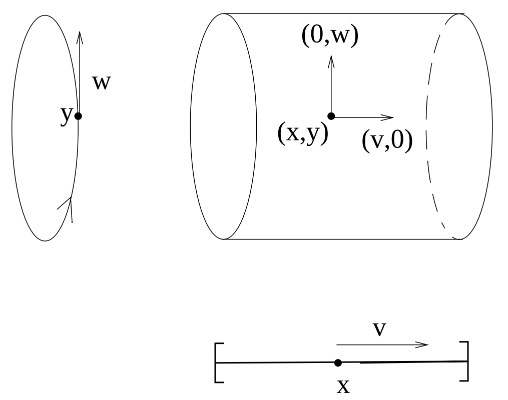
where $n \in T_x^{\text{out}}M$ is any outward pointing vector. This criterion is independent of which outward pointing vector we choose.

Proof If n_1 and n_2 are outward pointing, then for $t \in [0, 1]$ so is $(1 - t)n_1 + tn_2$. Hence for $t \in [0, 1]$

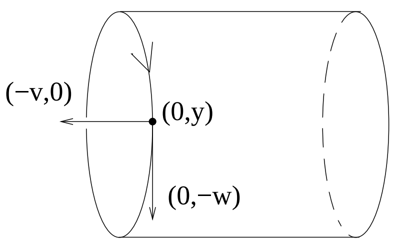
$$(1 - t)n_1 + tn_2, v_1, \ldots, v_{m-1}$$

is a basis for T_xM. By the continuity of the determinant function, this basis cannot change sign as t goes from 0 to 1. □

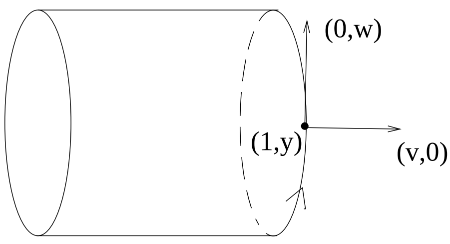
Example 3.26. Consider the cylinder $[0, 1] \times S^1$, which we give the product orientation (Example 3.11(v)) coming from the orientations of $[0, 1]$ and S^1 shown. As $v = 1$ is a positive basis for $T_x[0,1]$ for any $x \in [0,1]$, and w is a positive basis for T_yS^1, it follows that $(v,0),(0,w)$ is a positive basis for $T_{(x,y)}[0,1]\times S^1$.



In the two pictures below, we use the boundary orientation we have just defined on $[0,1]\times S^1$ to orient $\partial([0,1]\times S^1) = \{0\}\times S^1 \bigcup \{1\}\times S^1$. In the left hand picture, we see that $(-v,0)$ is an outward-pointing vector. Moreover, $(-v,0),(0,-w)$ is a positive basis for $T_{(0,y)}[0,1]\times S^1$: it is equivalent to the positive basis $(v,0),(0,w)$, since the change of basis matrix is the diagonal matrix with both diagonal entries -1, and has positive determinant. Hence $(0,-w)$ is a positive basis for $T_{(0,y)}S^1$. On the other hand, by a similar argument, $(0,w)$ is a *positive* basis for $T_{(1,y)}\partial([0,1]\times S^1)$, for now $(v,0)$ is an *outward* pointing vector.



Boundary orientation at (0,y)



Boundary orientation at (1,y)

□

Notice that the arrows on the right and left hand edges of the cyclinder *point in opposite directions*. This is a general phenomenon.

If M is any manifold, then $\partial([0,1]\times M) = \{0\}\times M \bigcup \{1\}\times M$. Let $i_0 : M \to \partial([0,1]\times M)$ and $i_1 : M \to \partial([0,1]\times M)$ be the maps $i_0(x) = (0,x)$, $i_1(x) = (1,x)$. Evidently i_0 is a diffeomorphism onto $\{0\} \times M$ and i_1 is a diffeomorphism onto $\{1\} \times M$.

Proposition 3.27. *Let M be an oriented manifold (without boundary) and let $[0,1]$ have the usual orientation, in which $1 \in T_t[0,1]$ is a positive basis for all t. Let $[0,1] \times M$ have the product orientation, and then give $\partial([0,1] \times M) = \{0\} \times M \bigcup \{1\} \times M$ the boundary orientation. Then i_1 preserves orientation and i_0 reverses orientation.*

Proof Let $v_1, \ldots, v_m$ be a positive basis for T_xM. By definition of the product orientation,

$$(1,0), (0,v_1), \ldots, (0,v_m) \ \text{ is a positive basis for } \ T_{(t,x)}[0,1] \times M \ \text{ for all } \ t. \tag{3.4}$$

We have

$$T^{\text{out}}_{(0,x)}([0,1] \times M) = T^{\text{out}}_0[0,1] \times T_xM.$$

So

$$-(1,0) \in T^{\text{out}}_{(0,x)}([0,1] \times M).$$

Thus

$$d_xi_0(v_1), \ldots, d_xi_0(v_m) = (0,v_1), \ldots, (0,v_m) \quad \text{is a } \textit{negative} \text{ basis for } \ T_{(0,x)}\partial([0,1] \times M);$$

by definition of the boundary orientation, it would only be positive if

$$(-1,0), (0,v_1), \ldots, (0,v_m) \quad \text{were a positive basis for } T_{(0,x)}\big([0,1] \times M\big),$$

and this contradicts (3.4). Hence i_0 reverses orientation.

On the other hand

$$(1,0) \in T^{\text{out}}_{(1,x)}([0,1] \times M),$$

so, again by (3.4),

$$d_xi_1(v_1), \ldots, d_xi_1(v_m) = (0,v_1), \ldots, (0,v_m) \quad \textit{is a} \ \text{ positive basis for } \ T_{(1,x)}\partial([0,1] \times M).$$

Hence i_1 preserves orientation. □

Remark 3.28. The previous result can be restated as

$$\partial\big([0,1] \times M\big) = \{1\} \times M - \{0\} \times M. \tag{3.5}$$

Here we orient $\{0\} \times M$ and $\{1\} \times M$ via their natural identification with M, and interpret the minus sign to indicate "opposite orientation". A special case of all this is the statement that

$$\partial[a,b] = b - a, \tag{3.6}$$

in which a and b are not to be regarded as numbers but as points in the manifold $\mathbb{R}$, and the minus sign refers to the orientation of the point a, as in Definition 3.14. From (3.5) and (3.6) we see that in this "arithmetic of manifolds", if M is a manifold (without boundary) then

$$\partial\big([0,1] \times M\big) = \big(\partial[0,1]\big) \times M. \tag{3.7}$$

Now we've defined the boundary orientation, we can state and prove the main theorem on oriented intersection numbers, from which great consequences, and in particular Theorem 3.18, will promptly follow.

Theorem 3.29. *Suppose that W^{m+1} is a manifold with boundary, $F : W \to N^n$ is smooth and Z^{n-m} is a submanifold of the manifold N, with W compact, Z closed in N, and all manifolds oriented. Give ∂W the boundary orientation, and denote by ∂F the restriction of F to ∂W. Suppose that $\partial F \pitchfork Z$. Then $(\partial F \cdot Z)_N = 0$.*

Before proving this, we re-state 3.18 as

Corollary 3.30. *Suppose that f_0 and f_1 are smooth maps from $M^m \to N^n$, both transverse to $Z^{m-n} \subset N$. Suppose that M is compact and Z is closed in N, and that all manifolds are oriented. Then*

$$(f_0 \cdot Z)_N = (f_1 \cdot Z)_N.$$

Proof of corollary: Let $F : [0,1] \to N$ be a homotopy from f_0 to f_1. Note that $f_0 = \partial F \circ i_0$ and $f_1 = \partial F \circ i_1$. Apply 3.29, taking $W = [0,1] \times M$. We have

$$0 = (\partial F \cdot Z)_N = (\partial F_{|\{1\}\times M} \cdot Z)_N + (\partial F_{|\{0\}\times M} \cdot Z)_N.$$

Since $i_1 : M \to \{1\} \times M$ is an orientation preserving diffeomorphism,

$$(\partial F_{|\{1\}\times M} \cdot Z)_N = ((\partial F \circ i_1) \cdot Z)_N = (f_1 \cdot Z)_N.$$

Since $i_0 : M \to \{0\} \times M$ is an orientation reversing diffeomorphism,

$$(\partial F_{|\{0\}\times M} \cdot Z)_N = -((\partial F \circ i_0) \cdot Z)_N = -(f_0 \cdot Z)_N.$$

Therefore

$$(f_1 \cdot Z)_N - (f_0 \cdot Z)_N = 0.$$

□

Proof of 3.29: Since ∂F is already transverse to Z, it is possible to perturb F so that it becomes transverse to Z *without changing* ∂F. The proof of this is a simple modification of the proof of the Weak Transversality Theorem 2.13. It involves one technique we have not yet described, and will be left as an exercise when this technique (gluing maps using partitions of unity) is introduced.

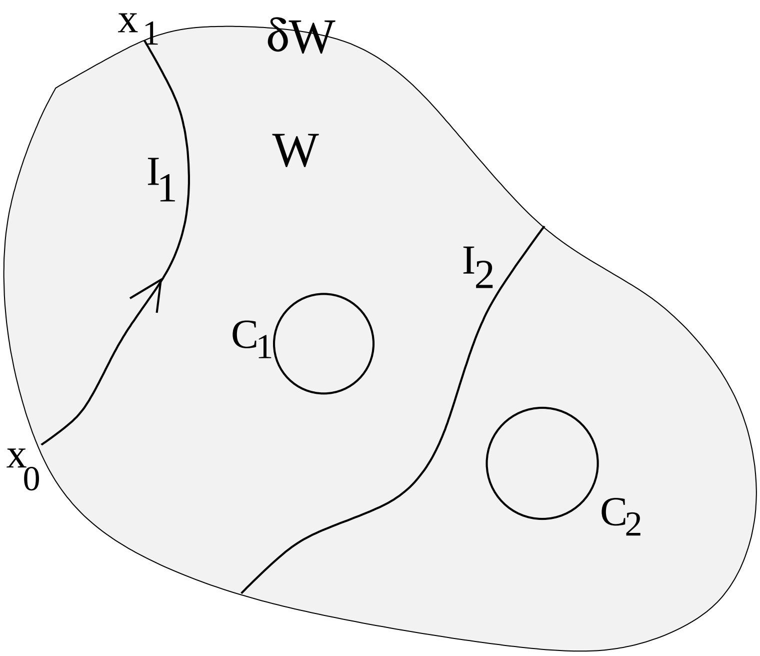
Now we need

Lemma 3.31. *Suppose $F : W^p \to N^n$ is a smooth map, and Z is a submanifold of N of codimension k. If $F \pitchfork Z$ and $\partial F \pitchfork Z$, then $F^{-1}(Z)$ is a $p-k$-dimensional manifold with boundary, and*

$$\partial(F^{-1}(Z)) = (\partial F)^{-1}(Z).$$

Proof of lemma Exercise (see Exercises III) □

By the lemma, $F^{-1}(Z)$ is a compact 1-dimensional manifold with boundary equal to $F^{-1}(Z) \cap \partial W = (\partial F)^{-1}(Z)$. Every compact connected 1-dimensional manifold with boundary is diffeomorphic either to $[0, 1]$ (if it has non-empty boundary) or to S^1 (if its boundary is empty) [8]. Thus every component of $F^{-1}(Z)$ is either a circle or an interval. The situation is illustrated in the following picture.



The key step in the proof is to show that at points x_0 and x_1 of $F^{-1}(Z) \cap \partial W$ which lie on the same connected component of $F^{-1}(Z)$ (e.g. as shown in the picture), we have

$$(\partial F \cdot Z)_{x_0} = -(\partial F \cdot Z)_{x_1}.$$

This is carried out by defining a *preimage orientation* on $F^{-1}(Z)$, and showing that with respect to this orientation, for every $x \in F^{-1}(Z) \cap \partial W$,

$$(F^{-1}(Z) \cdot \partial W)_x = (\partial F \cdot Z)_x \tag{3.8}$$

Once this is done, the theorem will follow. For observe that if I is a component of $F^{-1}(Z)$ with $I \cap \partial W = \{x_0, x_1\}$, then however I is oriented,

$$(I \cdot \partial W)_{x_0} = -(I \cdot \partial W)_{x_1}.$$

Suppose, for example, I is oriented, as is I_1 in the picture, so that the positive direction runs from x_0 to x_1. Let v_1 be a positive basis for $T_{x_0}I$, and v_1 a positive basis for $T_{x_1}I$. Because

[8]Though extremely plausible, this statement (the "classification of compact 1-manifolds with boundary") requires proof: see e.g. the Appendix to "Differential Topology" by Guillemin and Pollack. It's not hard.

$F^{-1}(Z) \pitchfork \partial W$ (this is a consequence of having $\partial F \pitchfork Z$), neither vector is tangent to ∂W. In fact, $-v_0 \in T^{\text{out}}_{x_0} W$ and $v_1 \in T^{\text{out}}_{x_1} W$. It follows, by definition of oriented intersection numbers and of boundary orientation, that

$$(I \cdot \partial W)_{x_0} = -1, \quad (I \cdot \partial W)_{x_1} = +1. \tag{3.9}$$

If the opposite orientation is chosen for I, then both signs are reversed, but in either case the two intersections have opposite signs.
Thus,

$$(\partial F \cdot Z)_N = \sum_{x \in \partial F^{-1}(Z)} (\partial F \cdot Z)_x = \sum_{x \in F^{-1}(Z) \cap \partial W} (F^{-1}(Z) \cdot \partial W)_x. \tag{3.10}$$

The points of $F^{-1}(Z) \cap \partial W$ come in pairs, at opposite ends of the same connected component of $F^{-1}(Z)$. By (3.9), the contributions of two such points to the sum in (3.10) cancel one another out. Thus, $(\partial F \cdot Z) = 0$.

It remains to define preimage orientation and prove (3.8). We do this in a separate proposition, which follows. □

To define the preimage orientation, we introduce the notion of *short exact sequence* of vector spaces.

Definition 3.32. (i) A sequence of linear maps between vector spaces

$$\cdots \to V_{k-1} \to V_k \to V_{k+1} \to \cdots$$

is *exact* if the kernel of each map is equal to the image of its predecessor. A *short exact sequence* is an exact sequence of the form

$$0 \to A \to B \to C \to 0.$$

Since the domain of the first map is 0, so is its image, so the second map is injective. Similarly, the kernel of the last map is all of C, so the third map must be surjective.

Example 3.33. Exact sequences occur naturally in the theory of manifolds. If y is a regular value of the smooth map $f : M \to N$, and $f^{-1}(y) \neq \emptyset$, then for each $x \in f^{-1}(y)$,

$$0 \to T_x f^{-1}(y) \to T_x M \xrightarrow{d_x f} T_y N \to 0$$

is exact. □

(ii) Now suppose that in the short exact sequence $0 \to A \xrightarrow{i} B \xrightarrow{j} C \to 0$, all three spaces are oriented. We define the *sign* of the sequence as follows. Let E be a positive basis for A, and let $F = c_1, \ldots, c_p$ be a positive basis for C. Choose vectors $b_1, \ldots, b_p \in B$ such that $j(b_k) = c_k$ for $k = 1, \ldots, p$, and write $F' = b_1, \ldots, b_p$. Then $i(E), F'$ is a basis for B. Call the sequence *positive* if E, F' is a positive basis, and *negative* otherwise.

It is easy to see that

1. Given E and F, the sign of $i(E), F'$ is independent of the choices of the b_j such that $j(b_i) = c_i$. For if F'' is another set of vectors such that $j(F'') = F$, then the members of F'' differ from those of F' by elements of $\ker(j)$, and so by linear combinations of the elements of $i(E)$. Thus the change of basis matrix $[I]_{i(E),F'}^{i(E),F''}$ can be transformed to the identity matrix by adding multiples of the first m rows to the last p. Hence it has determinant $+1$.

2. If F_1 and F_2 are bases for C, then the matrix of change of basis $[I]_{F_2}^{F_1}$, applied to the vectors of F_1', transforms them into vectors which differ from those of F_2' by elements of $\ker(j)$. Hence, as in the previous case, the bases $i(E), F_1'$ and $i(E), F_2'$ are equivalent.

3. If E_1 and E_2 are equivalent bases of A, then the bases $i(E_1), F'$ and $i(E_2), F'$ are equivalent.

Thus the definition of positive sequence is unambiguous.

(iii) Given a short exact sequence in which two of the three spaces are oriented, there is a unique orientation of the third making the sequence positive. Thus, a short exact sequence, together with orientations on any two of its members, can be used to orient the third.

(iv) Suppose $f : M \to N$ has y as regular value, and $f^{-1}(y) \neq \emptyset$. Here is how we define the preimage orientation on $f^{-1}(y)$. Let $x \in f^{-1}(y)$, and write $f(x) = y$. There is a short exact sequence

$$0 \to T_x f^{-1}(y) \to T_x M \xrightarrow{d_x f} T_{f(x)} N \to 0.$$

Use the orientations on $T_x M$ and $T_{f(x)} N$ to orient $T_x f^{-1}(y)$ according to the recipe in (iii).

(v) Now suppose $f : M \to N$ is transverse to Z, and $x \in f^{-1}(Z)$. Let $y = f(x)$. The procedure for orienting $T_x f^{-1}(Z)$ is only a little more complicated. There are short exact sequences

$$0 \to T_y Z \to T_y N \to T_y N / T_y Z \to 0$$

and

$$0 \to T_x f^{-1}(Z) \to T_x M \to T_y N / T_y Z \to 0.$$

Exactness of the first is trivial; exactness of the second is a consequence of equation (1.4) in the definition of transversality, together with the fact that $T_x f^{-1}(Z) = (d_x f)^{-1}(T_y Z)$.

Given orientations of $T_x M$, $T_y Z$ and $T_y N$, we use the first short exact sequence to orient $T_y N / T_y Z$ as explained in (iii), and then use the second short exact sequence to orient $T_x f^{-1}(Z)$.

This last definition might seem frighteningly non-implementable (and therefore, for the operationally inclined[9], incomprehensible). In fact it can be used in practice, but for our purposes it can perfectly well remain purely theoretical, because of the following result.

[9]such as me

Proposition 3.34. *If $F : W \to N$ and $Z \subset N$ are as in 3.29, and we give ∂W the boundary orientation, and $F^{-1}(Z)$ the preimage orientation, then for $x \in \partial F^{-1}(Z)$,*

$$(\partial F \cdot Z)_x = (F^{-1}(Z) \cdot \partial W)_x.$$

Proof Because $T_xF^{-1}(Z)$ and $T_x(\partial W)$ are transverse, for dimensional reasons their intersection is 0. Thus every vector $v \in T_xW$ can be written *uniquely* as a sum $v = v_1 + v_2$ with $v_1 \in T_xF^{-1}(Z)$ and $v_2 \in T_x(\partial W)$. Denote by p the map $v = v_1 + v_2 \mapsto v_2$. Thus there is a short exact sequence

$$0 \to T_xF^{-1}(Z) \to T_xW \xrightarrow{p} T_x(\partial W) \to 0; \tag{3.11}$$

and we can re-state our definition of oriented intersection number (3.15) as: $(F^{-1}(Z)\cdot\partial W)_x = \pm 1$ according to whether the exact sequence (3.11) is positive or negative.

There is an isomorphism (induced by $d_x(\partial F)$)

$$\bar{d} : T_x(\partial W) \to T_yN/T_yZ;$$

we have

$$(\partial F \cdot Z)_x = \pm 1$$

according to whether $\bar{d}$ preserves or reverses orientation.

Suppose we compose the map $p : T_xW \to T_x\partial W$ with $\bar{d}$. We obtain the exact sequence

$$0 \to T_xF^{-1}(Z) \to T_xW \to T_yN/T_yZ \to 0. \tag{3.12}$$

which is what we use, following Example 3.33(v), to determine the transverse preimage orientation of $F^{-1}(Z)$: each tangent space $T_xF^{-1}(Z)$ is oriented so that this sequence is positive. The two short exact sequences are shown together in the following diagram.

$$\begin{tikzcd}
 & & & & 0 & \text{(3.11)} \\
 & & & T_x\partial W \arrow[ur] \arrow[dd, "\bar{d}"] & & \\
0 \arrow[r] & T_xF^{-1}(Z) \arrow[r] & T_xW \arrow[ur, "p"] \arrow[dr, "\bar{d}\circ p"'] & & & \\
 & & & T_yN/T_yZ \arrow[dr] & & \\
 & & & & 0 & \text{(3.12)}
\end{tikzcd}$$

Because of the way in which (3.12) is obtained from (3.11) by inserting the isomorphism $\bar{d}$, which may preserve or reverse orientation, we have

$$1 = \text{sign}(3.12) = \text{sign}(3.11) \quad \times \quad \text{sign}(\bar{d})$$

$$= \text{sign}(3.11) \quad \times \quad (\partial F \cdot Z)_x.$$

It follows that

$$\text{sign}(3.11) \quad = \quad (\partial F \cdot Z)_x,$$

in other words

$$(F^{-1}(Z) \cdot \partial W)_x = (\partial F \cdot Z)_x.$$

□

This proposition completes the proof of Theorem 3.29, and thus of the homotopy-invariance of transverse intersection numbers. Now that we have done this, we can legitimately define non-transverse oriented intersection numbers, as described after Exercise 3.17. We state this as a formal definition:

Definition 3.35. (i) If $f : M^m \to N^n$ is smooth, Z^{n-m} is a submanifold of N, all three manifolds are oriented, and M is compact and Z is closed in N, then we define the *intersection number of f and Z in N*, $(f \cdot Z)_N$ by finding f_1 homotopic to f and transverse to Z, and setting $(f \cdot Z)_N = (f_1 \cdot Z)_N$.

(ii) If, in (i), M is actually a submanifold of N and $i : M \to N$ is the inclusion, we define the intersection number of M and Z in N, $(M \cdot Z)_N$, by

$$(M \cdot Z)_N = (i \cdot Z)_N.$$

Having made this definition, we can reststate 3.29 without the hypothesis that ∂F be transverse to Z:

Theorem 3.36. *Suppose that W^{m+1} is a manifold with boundary, $F : W \to N^n$ is smooth and Z^{n-m} is a submanifold of the manifold N, with W compact, Z closed in N, and all manifolds oriented. Give ∂W the boundary orientation, and denote by ∂F the restriction of F to ∂W. Then $(\partial F \cdot Z)_N = 0$.* □

Proposition 3.37. *In the situation of 3.35(ii),*

$$(M \cdot Z)_N = (-1)^{dim_M \ dim_Z}(Z \cdot M)_N.$$

Proof The proof is easy if $M \pitchfork Z$. Suppose $x \in M \cap Z$, $v_1, \ldots, v_m$ is a basis of T_xM and $w_1, \ldots, w_d$ is a basis of T_xZ. Then by the assumptions on dimension and transversality, both

$$v_1, \ldots, v_m, w_1, \ldots, w_d$$

and

$$w_1, \ldots, w_d, v_1, \ldots, v_m$$

are bases of T_xN. Each basis can be obtained from the other by md transpositions, and it follows that the signs of the two bases are related by a factor of $(-1)^{md}$. Since this holds for each point $x \in M \cap Z$, we have

$$(M \cdot Z)_N = \sum_{x \in M \cap Z} (M \cdot Z)_x = (-1)^{md} \sum_{x \in M \cap Z} (Z \cdot M)_x = (-1)^{md} (Z \cdot M)_N.$$

In case M is not transverse to Z, there is a perturbation i_1 of i which is transverse to Z, and which is still an embedding (i_1 can be chosen as close as we like to i, and being an embedding is an open condition). Now apply the previous argument to the image M_1 of i_1. □

We end this section with one minor extension of the main theorem:

Proposition 3.38. *The oriented intersection number $(f \cdot Z)_N$ of $f : M \to N$ and $Z \subset N$ is invariant under homotopies of the inclusion of Z in N.*

Proof $(f \cdot Z)_N$ equal to $(M \times Z \cdot \mathrm{graph}(f))_{M \times N}$. A homotopy of the embedding of Z in N induces a homotopy of the embedding of $M \times Z$ in $M \times N$. Now apply 3.30. □

It's time to harvest some of the consequences of our hard work!

4 Applications of Oriented Intersection Numbers

Let M^n and N^n be oriented manifolds of the same dimension, with M compact, and let $f : M \to N$ be a smooth map. For a regular value y of f, define

$$\deg(f; y) = \sum_{x \in f^{-1}(y)} \mathrm{sgn}(x),$$

where

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{if } d_x f \text{ preserves orientation} \\ -1 & \text{if } d_x f \text{ reverses orientation} \end{cases}$$

Proposition 4.1. *For any two regular values y_0, y_1 of f,*

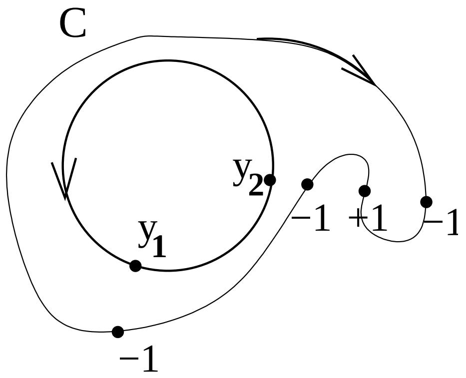
$$deg(f; y_0) = deg(f; y_1).$$

Proof $\deg(f; y)$ is equal to the oriented intersection number in $M \times N$ of $M \times \{y\}$ and the graph of f, $\mathrm{gr}(f)$ (**Exercise**). Choose a path γ in N joining y_0 and y_1. The map $\Gamma : [0,1] \times M \to M \times N$ defined by $\Gamma(t, x) = (x, \gamma(t))$ is a homotopy between the inclusion i_0 of M in $M \times N$ as $M \times \{y_0\}$ and the inclusion i_1 of M as $M \times \{y_1\}$. Hence

$$\deg(f; y_0) = (M \times \{y_0\} \cdot \mathrm{gr}(f)) = (i_0 \cdot \mathrm{gr}(f)) = (i_1 \cdot \mathrm{gr}(f)) = (M \times \{y_1\} \cdot \mathrm{gr}(f)) = \deg(f; y_1)$$

by the homotopy invariance of oriented intersection numbers. □

Example 4.2. The following diagram shows a curve C wrapping around the circle. Radial projection from the centre of the circle defines a map $\rho : C \to S^1$. Two regular values, y_1 and y_2, are shown, as are their preimages in C. Whereas y_1 has three preimages in C, y_2 has only one. The sign of each preimage is indicated in the diagram with a $+1$ or a -1. It is easy to determine visually: first determine the direction of movement of $\rho(x)$ as x moves through the preimage point in the positive direction on C, then compare this with the positive direction on S^1.



By inspection,

$$\deg(\rho; y_1) = -1 = \deg(\rho; y_2),$$

bearing out what is proved in 4.1.

In the light of 4.1 we can make the folowing definition:

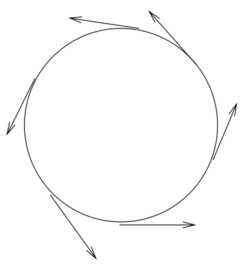
Definition 4.3. *Let M^n and N^n be oriented manifolds of the same dimension, with M compact, and let $f : M \to N$ be a smooth map. Define deg(f) to be deg$(f; y)$ for any regular value y.*

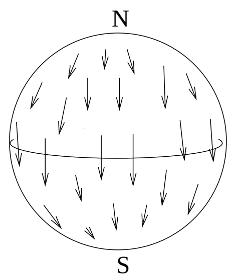
By the homotopy-invariance of oriented intersection numbers, we have

Proposition 4.4. *Homotopic maps have the same degree.* □

4.1 Vector fields on spheres

It is easy to construct a nowhere vanishing vector field [10] on S^1, as in the left-hand picture.





[10]We mean tangent vector fields. It is usual to omit the word 'tangent', unless other types of vector fields — e.g. normal vector fields— are also being discussed.

The field drawn is given by $v(x_1, x_2) = (-x_2, x_1)$. A similar formula,

$$v(x_1, x_2, x_3, x_4) = (-x_2, x_1, -x_3, x_4),$$

gives a nowhere-vanishing vector field on S^3, and the same idea gives a nowhere vanishing vector field on S^n for every odd n. Does there exist a nowhere-vanishing vector-field on S^2? It's easy to find a vector field with just two zeros: a vector field vanishing only at N and S is shown in the right-hand picture. This question can be recast as a problem in hairdressing: given a sphere with a hair growing from every point, is it possible to comb them all in such a way that all the hairs lie flat on the surface of the sphere? As you will observe on the heads of your friends, this problem is hard to solve even on a hemisphere.

Theorem 4.5. *If n is even, every vector field on the n-sphere S^n must vanish at least at one point.*

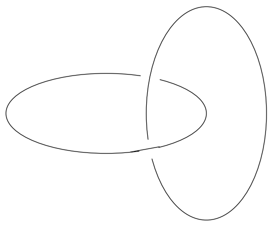
Proof
Step 1: If n is even, the antipodal map $a : S^n \to S^n$ has degree -1. (**Exercise**).

Step 2: If v is a nowhere vanishing vector field on S^n, then the antipodal map $a : S^n \to S^n$ is homotopic to the identity map. To construct a homotopy, we first obtain a unit vector field $\tilde{v}$ by dividing v by its length. Then we define $F : [0, 1] \times S^n \to S^n$ by $F(t, x) = \cos(\pi t)x + \sin(\pi t)\tilde{v}(x)$. Clearly $F(0, x) = x, F(1, x) = -x$, and $F(t, x) \in S^n$ for all t. Since the identity map has degree 1, and degree is a homotopy-invariant, n cannot be even. □

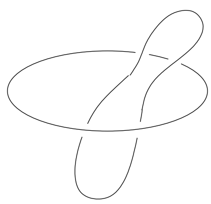
Exercise 4.6. How many *everywhere independent* vector fields can you find on S^3? Can you find a frame field defined on all of S^3? □

4.2 Linking Numbers

Consider the two pairs of loops of wire shown in the following picture.



(i)

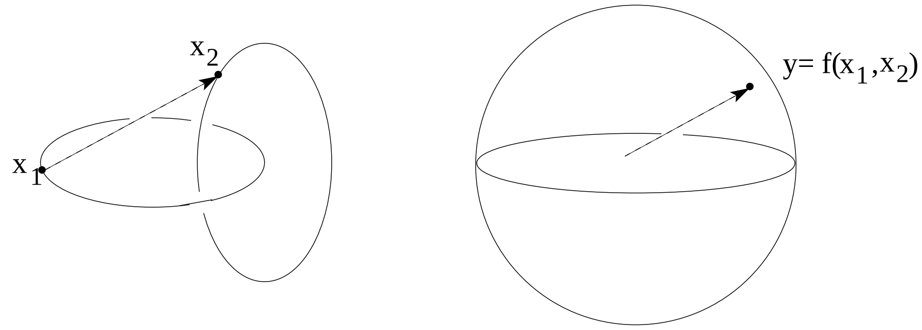


(ii)

It's clear that the two in the left cannot be separated, whereas the two on the right can be. Can we give a mathematical description of the way they sit relative to one another, which accounts for this difference? Using the degree of a suitable smooth map, we can.

Let C_1 and C_2 be disjoint smooth closed curves in $\mathbb{R}^3$. Define a map $C_1 \times C_2 \to S^2$ by

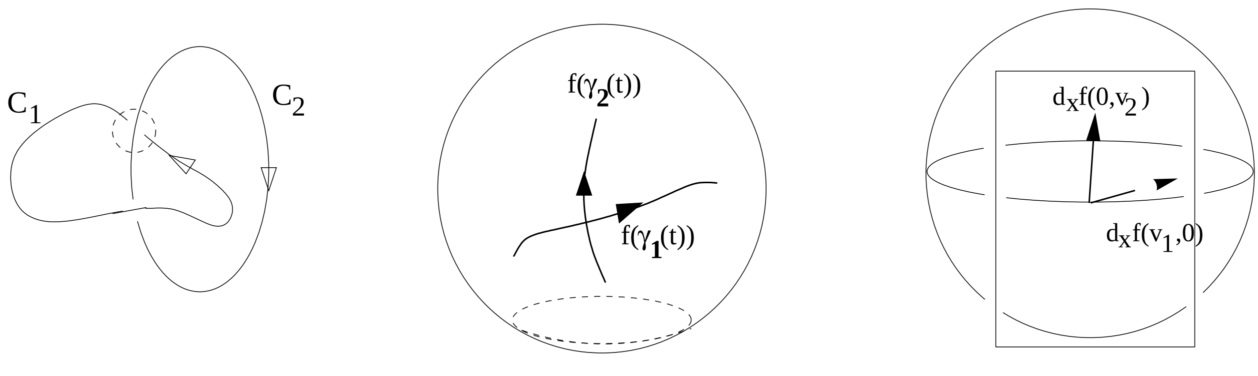
$$(x_1, x_2) \mapsto \frac{x_2 - x_1}{\|x_2 - x_1\|}.$$



If C_1 and C_2 are oriented, we can calculate the degree of f, by giving $C_1 \times C_2$ the product orientation, and giving S^2 its usual orientation as the boundary of the unit ball. Define the *linking number* of C_1 and C_2, $\ell(C_1, C_2)$, by

$$\ell(C_1, C_2) = \deg(f).$$

To see that this *does* capture something of the difference between (i) and (ii) in the first picture, we calculate the two linking numbers. Remarkably (we cannot say this without a glow of satisfaction) the techniques we have developed make this very simple. To calculate the degree of the map f, we have to select a point y in S^2 and count its preimages, with their sign. The point (x_1, x_2) in $C_1 \times C_2$ is a preimage of y if the line from x_1 to x_2 points in the direction of y. While it looks a little difficult to count the preimages of the point y in the picture, if instead we choose as y the unit vector orthogonal to the plane of the page and pointing towards us, then the preimages of y are simply pairs (x_1, x_2) where the curve C_2 appears to cross over the curve C_1. In the picture below, there is just one such point, which I have ringed.



Now we compute its sign. If v_1 and v_2 are positive bases for $T_{x_1}C_1$ and $T_{x_2}C_2$ then $(v_1, 0), (0, v_2)$ is a positive basis for $T_{(x_1,x_2)}C_1 \times C_2$. Suppose that γ_1, γ_2 are parametrisations of C_1 and C_2 around x_1 and x_2 such that $\gamma_1'(0) = v_1,\ \gamma_2'(0) = v_2$. Then $(v_1, 0)$ and $(0, v_2)$ are the tangent vectors to the curves $\tilde{\gamma}_1(t) = (\gamma_1(t), 0)$ and $\tilde{\gamma}_2(t) = (0, \gamma_2(t))$. Moreover

$$d_{(x_1,x_2)}f(v_1, 0) = (f \circ \tilde{\gamma}_1)'(0), \quad d_{(x_1,x_2)}f(0, v_2) = (f \circ \tilde{\gamma}_2)'(0).$$

We only have to decide whether $d_{(x_1,x_2)}f(v_1, 0), d_{(x_1,x_2)}f(0, v_2)$ is a positive or a negative basis for T_yS^2, so we can use a fairly rough and ready approach (which will be justified below).

Imagine a unit pencil pointing from x_1 to $\gamma_2(t)$. It's easy to see when it's transplanted and placed with the back end at the centre of the unit ball, then its tip draws a curve on the sphere something like the one shown, and with the orientation shown. Similarly, $f \circ \tilde{\gamma}_1(t)$ appears on the sphere as shown. The right hand picture shows the tangent vectors to these two curves, and since they make a positive basis for T_yS^2, we conclude that $\ell(C_1, C_2) = 1$.

When we calculate $d_{(x_1,x_2)}f(0, v_2)$, we move the tip of the pencil in the direction of v_2. From this it follows that $d_{(x_1,x_2)}f(0, v_2)$ is a *positive* multiple of the orthogonal projection of v_2 to the plane $T_{f(x_1,x_2)}S^2$ (which is equal to $(x_2 - x_1)^{\perp}$). When we calculate $d_{(x_1,x_2)}f(0, v_2)$, we move the other end of the pencil in the direction of v_1. From this it follows that $d_{(x_1,x_2)}f(v_1, 0)$ is a *negative* multiple of the orthogonal projection of v_2 to the plane $T_{f(x_1,x_2)}S^2$.

These two observations lead to the following method for computing the linking numbers of two curves C_1 and C_2 by counting apparent crossings of their orthogonal projections $\pi(C_1), \pi(C_2)$ to a plane P. Let n be a unit vector orthogonal to P. Then n determines an orientation of P — v_1, v_2 is a positive basis if n, v_1, v_2 is a positive basis for $\mathbb{R}^3$. Suppose that $x_1 \in C_1$, $x_2 \in C_2$ and $\pi(x_1) = \pi(x_2)$, so that at this point, which we call p, $\pi(C_1)$ and $\pi(C_2)$ cross.

- Call this point p a *crossing of C_2 over C_1*, or *overcrossing*, if the vector $x_2 - x_1$ is a positive multiple of n (because $\pi(x_1) = \pi(x_2)$, $x_2 - x_1$ must be a multiple of n, although possibly negative). Note that overcrossings and undercrossings are interchanged if we interchange the order of C_1 and C_2.

- Call an overcrossing *transverse* if $\pi_{|C_1}$ and $\pi_{|C_2}$ are immersions at x_1 and x_2, so that $\pi(C_1)$ and $\pi(C_2)$ are manifolds in some neighbourhood of p, and if moreover $\pi(C_1)$ and $\pi(C_2)$ are transverse at p.

- Call the overcrossing p *positive* if $d_{x_1}\pi(v_1), d_{x_2}\pi(v_2)$ is a *negative* basis for P, and *negative* if $d_{x_1}\pi(v_1), d_{x_2}\pi(v_2)$ is a *positive* basis for P, and write $\mathrm{sgn}(p) = 1$ if p is a positive overcrossing and -1 if p is a negative overcrossing.

Proposition 4.7. *(i) If $\pi(C_1)$ and $\pi(C_2)$ are transverse to one another at each overcrossing, then the vector n is a regular value of the map f.*
(ii) In this case,

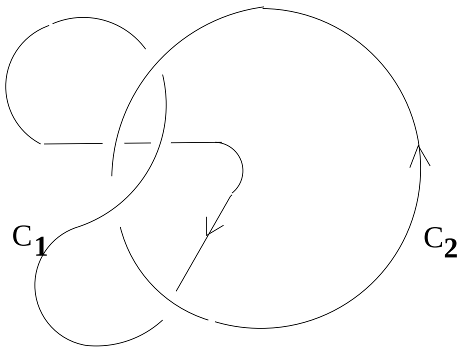
$$\ell(C_1, C_2) = \sum_p sgn(p) \tag{4.1}$$

where the sum is over overcrossings.

Example 4.8. The linking number of the two curves in picture (ii) on page 76 is 0, in whichever order we place them and however we orient them. There are two overcrossings, and they have opposite sign. □

Exercise 4.9. What happens to $\ell(C_1, C_2)$ if we interchange C_1 and C_2? □

Exercise 4.10. Find the linking numbers of the following pair of curves.



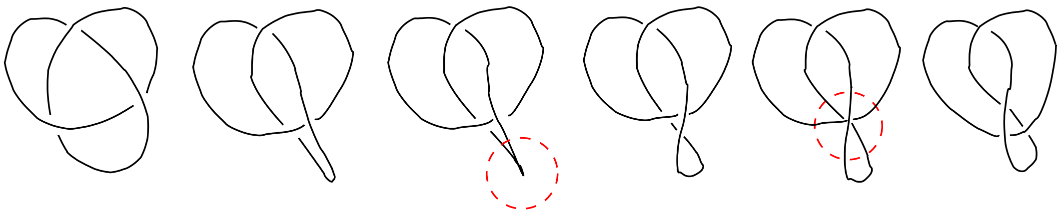
□

Exercise 4.11. Consider the map $S^1 \to S^1 \times S^1$ defined by $\theta \mapsto (p\theta, q\theta)$. Composing with the diffeomorphism $S^1 \times S^1 \to \mathbb{T}$ described in 1.5, we get a curve C on the torus $\mathbb{T}$. It is known as a (p, q) curve (because it winds around the torus p times in one direction and q times in the other). Let $C(\varepsilon)$ be the image on $\mathbb{T}$ of the curve $S^1 \to S^1 \times S^1$ given by θ to $(p\theta + \varepsilon, q\theta)$ for $\varepsilon > 0$.
(i) Show that if $\varepsilon > 0$ is small then $C \cap C(\varepsilon) = \emptyset$.
(ii) For small ε, find $\ell(C, C(\varepsilon))$. □

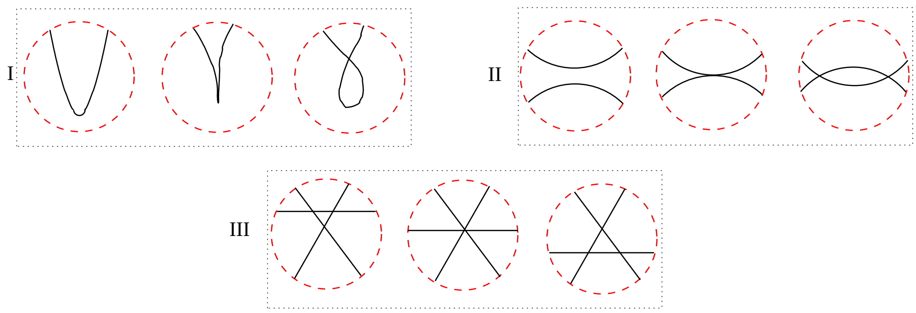
Exercise 4.12. (i) Show that if C_1 and C_2 are disjoint closed oriented curves in $\mathbb{R}^3$ then $\ell(C_1, C_2)$ is unchanged if we deform C_1 or C_2 in a homotopy during which the two curves remain at all times disjoint.
(ii) We say that we can *separate* C_1 and C_2 if it possible, by means of a homotopy during which the curves remain at all times disjoint, to place them on opposite sides of a plane. Show that if it is possible to separate C_1 and C_2 then $\ell(C_1, C_2) = 0$. Hint: suppose they are separated by the plane H. Then it is possible to choose a vector $y \in S^2$ such that $f^{-1}(y) = \emptyset$ (where f is the map $C_1 \times C_2 \to S^2$ whose degree we measure as our definition of $\ell(C_1, C_2)$).
□

Remark 4.13. One might ask, given that there is a simple algorithm to determine the linking number of two closed curves, why do we need the rather sophisticated definition in terms of the degree of a smooth map? After all, the very definition of degree took rather a lot of work. Some authors do prefer to define the linking number by means of the algorithm. But they then have the problem of showing that
(i) $\ell(C_1, C_2)$ is an invariant of the two curves, and does not depend on the planar projection chosen, and
(ii) $\ell(C_1, C_2)$ does not change if we move the curves without passing one through the other. In fact both can be shown by a completely different procedure from ours. Consider the following sequence of pictures. It begins with one plane projection of a trefoil knot, and ends with another. There are two points at which something topologically non-trivial happens,

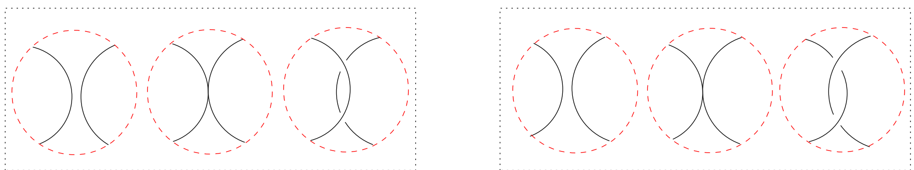
ringed with a dashed circle.



The local views in the neighbourhood of the special points immediately before, during and after these changes are shown (as moves I and III) in the following table of *Reidemeister moves*. It is possible to show that given any two transverse planar projections of the pair of curves C_1, C_2, one projection can be deformed to the other through a sequence of such moves. Similarly, if the pair C_1, C_2 can be deformed to the pair C_1', C_2' without at any moment passing through one another, the deformation can be represented by means of planar projections which undergo only these same changes.



None of the three Reidemeister moves change the value of the sum in (4.1). In the case of the move of type I, this is because the overcrossing it introduces (or cancels) is of one curve over itself, and therefore does not contribute to (4.1). In the case of the type II move, if the two crossings shown in the picture do appear in (4.1) then they do so with opposite sign, as in the following picture on the left, and again do not change its value.

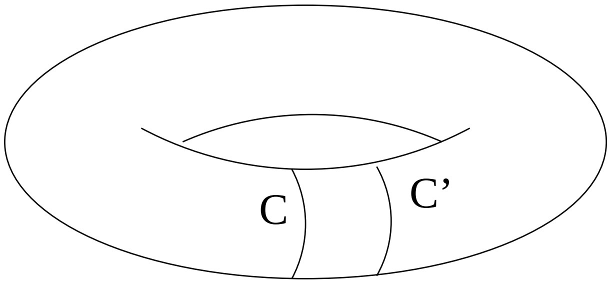


Possible sequence *Impossible sequence*

Exercise 4.14. Show that a type III move does not change the sum (4.1).

4.3 Self-intersection and the Euler Characteristic

Let N^{2m} be an oriented manifold of even dimension and let M^m be a compact oriented submanifold. We can define the *self-intersection* number $(M \cdot M)_N$ by the procedure of 3.35: we imagine given two copies of M, and then, leaving one fixed, we perturb the other, in a homotopy, so that it becomes transverse to the first. For example, the self-intersection number of the curve $C \subset \mathbb{T}$ is zero, since it can be slid right off itself.



Let M be a compact oriented manifold. Let $\Delta \subset M \times M$ be the diagonal. Give $M \times M$ the product orientation, and orient Δ by the diffeomorphism $M \to \Delta$ sending x to (x, x). The *Euler characteristic* $\chi(M)$ is the self-intersection number of Δ in $M \times M$, $(\delta \cdot \Delta)_{M \times M}$.

Proposition 4.15. *If $\chi(M) \neq 0$, and the map $f : M \to M$ is homotopic to the identity map, then f must have a fixed-point.*

Proof **Exercise**.

From 4.15 we can deduce a weak version of the Poincaré-Hopf theorem.

Theorem 4.16. *If $\chi(M) \neq 0$ then every tangent vector field on M must have at least one zero.*

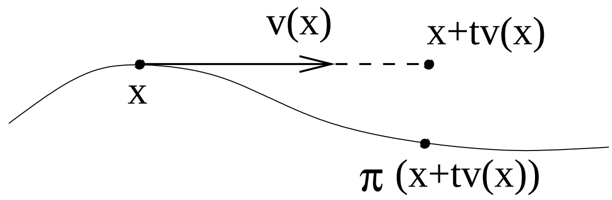
Proof Let U_M be a tubular neighbourhood of M. Given a tangent vector v on M, define a homotopy $F : [0, \varepsilon] \times M \to M$, by

$$F(t, x) = \pi(x + tv(x)).$$

For ε small enough and $t \in [0, \varepsilon]$,then $x + tv(x) \in U_M$ and thus F is well-defined. If v has no zero then f_t has no fixed point (see the Exercise below). But f_t is homotopic to the identity map. □

Exercise 4.17. Show that if M is compact, oriented and of odd dimension then $\chi(M) = 0$. Hint: use 3.37. □

Exercise 4.18. Show that, as claimed in the proof of 4.16, the map $f_t(x) = F(t, x)$ has no fixed point. Hint: the *picture* should be clear enough:



How to make this into a precise argument? For which points $x' \in U_M$ is $\pi(x')$ equal to x? $\square$

The full strength Poincaré Hopf Theorem gives more precise information about the zeros of a vector field, with only isolated zeros, on a compact oriented manifold. First, the *index* $\text{ind}_x(v)$ of an isolated zero x of a vector field v is defined; then

Theorem 4.19. *If v is a vector field with only isolated zeros on the compact oriented manifold M, then*

$$\sum_{x \text{ a zero of } v} ind_x(v) = \chi(M).$$

$\square$

We will not prove this theorem here; some of the details of the proof, and in particular the definition of the index of an isolated zero, are given in Exercises 4.

4.4 Brouwer's Fixed Point Theorem

Let $Y \subset X$. A *retraction* $X \to Y$ is a continuous map such that for all $y \in Y, r(y) = y$.

Proposition 4.20. *There can be no smooth retraction from a compact oriented manifold with boundary, W, onto its boundary.*

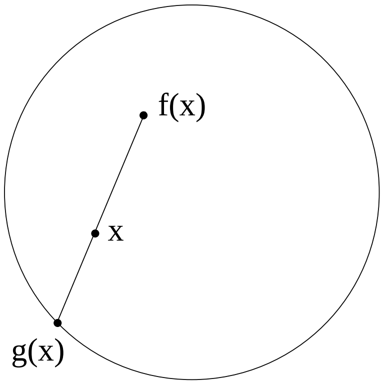
Proof **Exercise** Hint: if r is a smooth retraction, what is the degree of ∂r? Use Theorem 3.29 $\square$

Theorem 4.21. Brouwer's Fixed Point Theorem *Let $f : \bar{B}^n \to \bar{B}^n$ be a continuous map (where $\bar{B}^n$ is the unit ball in $\mathbb{R}^n$). Then f has a fixed point.*

Proof Suppose first that f is smooth. Define a smooth map $g : \bar{B}^n \to S^{n-1}$ by

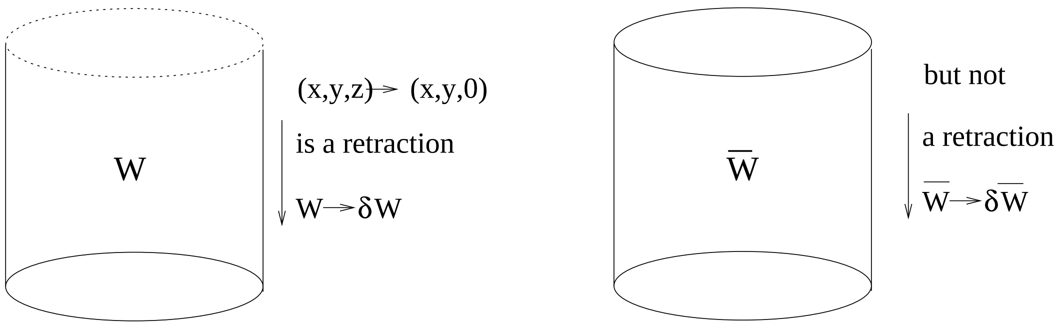
$$x \mapsto \text{the point where the line from } f(x) \text{ to } x \text{, continued, meets } S^{n-1}.$$

Evidently this is a retraction from $\bar{B}^n$ to $\partial \bar{B}^n$, contradicting 4.20.



Now consider the case where f is merely continuous, and has no fixed point. As $\bar{B}^n$ is compact, there exists $\varepsilon > 0$ such that for all $x \in \bar{B}^n$, $\|x - f(x)\| \geq \varepsilon$. Choose smooth $\tilde{f} : \bar{B}^n \to \bar{B}^n$ such that $\|\tilde{f}(x) - f(x)\| < \varepsilon$ for all x. Then also $\tilde{f}$ has no fixed point. Now apply the previous argument to $\tilde{f}$. □

The hypothesis of compactness in 4.21 is vital. It is easy to give examples of non-compact manifolds W with retractions $W \to \partial W$. For example, take W to be the semi-infinite cylinder $S^1 \times [0, 1)$ (so that $\partial W = S^1 \times \{0\}$), and let $r(x, y, z) = (x, y, 0)$. As soon as we modify W to make it compact, for example by taking its closure in $\mathbb{R}^3$, $\overline{W} = S^1 \times [0, 1]$, the previously defined retraction fails. In this case, $r(x, y, z) = (x, y, 0)$ is still a smooth map $\overline{W} \to \partial\overline{W}$, but it is no longer the identity map on all of ∂W.



The proof of 4.21 relied upon Theorem 3.29, which says that if W^{n+1} is a compact oriented manifold with boundary and $f : W \to N^n$ is smooth, with N oriented, then $\deg(\partial f) = 0$.

Exercise 4.22. What can you say about the degree of ∂r in the example of $r : W \to \partial W$ and $r : \overline{W} \to \partial\overline{W}$ just discussed? □

4.5 The Fundamental Theorem of Algebra

Theorem 4.23. *Let $p(z)$ be a polynomial of degree $N > 0$, with complex coefficients, in the complex variable z. Then p has a zero.*

Proof Let the degree of p be N. After dividing through by the leading coefficient, we can assume that $p(z)$ has the form

$$p(z) = z^N + c_{N-1}z^{N-1} + \cdots + c_0.$$

Step 1 Consider the family of polynomials, parametrised by $t \in [0, 1]$,

$$p_t(z) = z^N + (1 - t)(c_{N-1}z^{N-1} + \cdots + c_0).$$

Because the highest power eventually dominates, there is a real number R such that for all $t \in [0, 1]$, every zero of p_t (if there are any) is contained inside the disc $B^2(R)$. In other words, such that if $|z| \geq R$ then $p_t(z) \neq 0$.

Step 2 Let S_R be the circle of radius R centred at 0. For each $t \in [0, 1]$, consider the map $f_t : S_R \to S^1$ defined by

$$z \mapsto \frac{p_t(z)}{|p_t(z)|}.$$

These maps are all smooth, by Step 1. As they are all homotopic, all have the same degree. The degree of f_1 is N; hence, so is the degree of f_0.

Conclusion If $p = p_0$ had no zero inside the disc of radius R, then the same formula would define an extension of the map $f_0 : S_R \to S^1$ to a smooth map $F : \bar{B}^2(R) \to S^1$. But then by Theorem 3.29, the degree of $f_0 = \partial F$ would be zero. $\square$

5 Abstract Manifolds

5.1 Definition and examples

In this section we enlarge the range of objects that we study, by defining manifolds which are not necessarily subsets of some Euclidean space. We will do it in stages, working our way towards a sensible definition.

In Example 1.41 and the discussion following it, we showed that the image of the map $f : S^2 \to \mathbb{R}^6$ defined by $f(x_1, x_2, x_3) = x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3)$ is a smooth manifold homeomorphic to the space $S^2/\sim$, where $\sim$ is the equivalence relation identifying antipodal points. Thus $S^2/\sim$ can be thought of as a smooth manifold. How to do this without going to the trouble of finding a 2-1 map $S^2 \to \mathbb{R}^6$, and then hunting for its image?

Definition 5.1. (i) A *topological manifold* of dimension m is a topological space with the property that every point has a neighbourhood homeomorphic to an open set in $\mathbb{R}^m$. It is usual to insist also that M be Hausdorff and second-countable (i.e. having a countable dense subset), and we will impose this requirement. We will also impose the requirement that M be *paracompact*: every open cover has a locally finite refinement [11] Every subset of $\mathbb{R}^N$ is Hausdorff, second countable and paracompact, so these requirements hold automatically for the manifolds contained in $\mathbb{R}^N$ that we have been studying up to now.

(ii) A homeomorphism $\phi : U \to V$, where U is open in M and V is open in $\mathbb{R}^m$, is called a *chart*, and a collection $A = \{\phi_\alpha : U_\alpha \to V_\alpha\}$ of charts such that $M = \bigcup_\alpha U_\alpha$ is called an *atlas* for M.

(iii) An atlas $\{\phi_\alpha : U_\alpha \to V_\alpha\}$ is *smooth* if whenever $U_\alpha \cap U_\beta \neq \emptyset$, the crossover homeomorphism

$$\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$$

is smooth.

(iv) If M is a manifold equipped with a smooth atlas A, then a map $f : M \to \mathbb{R}^k$ is *smooth with respect to* A at a point $x \in M$ if $f \circ \phi_\alpha^{-1}$ is smooth, where $\phi_\alpha \in A$ is a chart defined on some neighbourhood of x. And a map $f : \mathbb{R}^k \to M$ is smooth with respect to A at $y \in \mathbb{R}^k$ if $\phi_\alpha \circ f$ is smooth at y, for some chart ϕ_α whose domain contains $f(y)$.

Because of the smoothness of the crossover maps of charts in A, the criteria for smoothness described in (iv) are independent of the choice of chart $\phi_\alpha \in A$ used to verify them.

When $M \subset \mathbb{R}^N$ is a manifold in the sense defined in Section 1, any two of its charts (which we required to be smooth (new)) automatically enjoyed the crossover property described in 5.1(iii), and so an atlas of such charts was automatically smooth in the sense of 5.1(iii). Here, in our new more abstract situation, it makes no sense to speak of smooth (new or old) maps

[11] A *refinement* of an open cover $\{U_\alpha\}$ is an open cover $\{V_\beta\}$ such that each V_β lies entirely in some U_α.

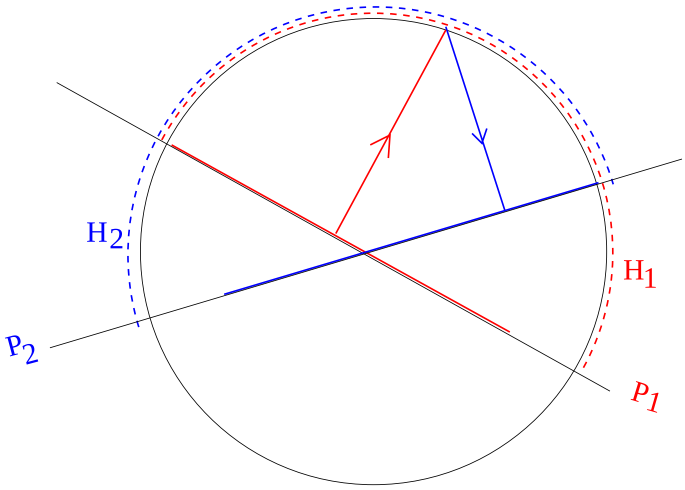
from M to $\mathbb{R}^k$ without reference to the charts of a smooth atlas, because M is not embedded in any bigger space already equipped with a notion of differentiability.

Example 5.2. We define an atlas on $\mathbb{RP}^n = S^n/\sim$ as follows. Each $(n+1)$-dimensional vector subspace $P \subset \mathbb{R}^{n+1}$ divides S^n into two open hemispheres. Let H be one of them. As H contains no pairs of antipodal points, the map $q_| : H \to \mathbb{RP}^n$ is 1-1; since q is continuous and open (see Example 1.41(ii)), it follows that $q(H)$ is open in $\mathbb{RP}^n$ and $q : H \to q(H)$ is a homeomorphism. Also H is homeomorphic to an open set in $\mathbb{R}^n$: for instance, the orthogonal projection $\mathbb{R}^{n+1} \to P$ maps H homeomorphically to the open unit ball B in P. We take the composite

$$q(H) \xrightarrow{q^{-1}} H \xrightarrow{\pi} B$$

as a chart on $\mathbb{RP}^n$, and call it ϕ. As hemispheres H cover S^n, their images $q(H)$ cover $\mathbb{RP}^n$, so we have an atlas.

This atlas is smooth. This is most easily seen with a diagram for the case $n = 1$. We take two hyperplanes P_1 and P_2, and two hemispheres (semicircles in the picture) that they determine, H_1 and H_2.



Denote by $\pi_1 : H_1 \to B_1$ and $\pi_2 : H_2 \to B_2$ the orthogonal projections used in the construction of the charts. We have

$$\phi_i\big(q(H_1) \cap q(H_2)\big) = \pi_i(H_1 \cap H_2)$$

for $i = 1, 2$, and

$$\phi_2 \circ \phi_1^{-1} = \pi_2 \circ \pi_1^{-1};$$

the diagram below shows the composite

$$\pi_2 \circ \pi_1^{-1} : \pi_1(H_1 \cap H_2) \to \pi_2(H_1 \cap H_2).$$

Because of its geometrical origin, it's easy to believe that this map is smooth.

In fact a precise argument is straightforward. After an ambient rotation, P_1 is the subspace $\{x_{n+1} = 0\}$, and $\pi_1^{-1} : B_1 \to H_1$ is given by the formula

$$(x_1, \ldots, x_n) \to \left(x_1, \ldots, x_n, \sqrt{1 - (x_1^2 + \cdots + x_n^2)} \right).$$

As π_2 is a linear projection, the composite $\pi_2 \circ \pi_1^{-1}$ is smooth. $\square$

Example 5.3. Now here is another atlas on $\mathbb{RP}^n$. This time we use the fact that $\mathbb{RP}^n$ can also be thought of as the quotient of $\mathbb{R}^{n+1} \setminus \{0\}$ by the equivalence relation $x \sim_1 \lambda x$ where $\lambda \in \mathbb{R} \smallsetminus \{0\}$. The equivalence classes here are just straight lines through zero, and each meets S^n in two antipodal points, so there is a natural bijection between the set of these equivalence classes, $\mathbb{R}^{n+1}/\sim_1$ and the quotient of S^n by the antipodal identification. It follows that there are mutually inverse continuous maps between the two quotients, and they can be identified with one another. Thus, we think of $\mathbb{RP}^n$ ss the space of lines ℓ through 0 in $\mathbb{R}^{n+1}$.

We will denote the equivalence class in $\mathbb{RP}^n$ of a point $(x_1, \ldots, x_{n+1})$ in $\mathbb{R}^{n+1}$ by $[x_1, \ldots, x_{n+1}]$. Thus for any $\lambda \neq 0$,

$$[x_1, \ldots, x_{n+1}] = [\lambda x_1, \ldots, \lambda x_{n+1}]. \tag{5.1}$$

For a point $P = [x_1, \ldots, x_{n+1}] \in \mathbb{RP}^n$, the x_i are its so-called *homogeneous coordinates*. Of course, they are not co-ordinates at all: the value of x_i is not well defined, by (5.1). What *is* well defined is the ratio x_i/x_j, provided $x_j \neq 0$. We define

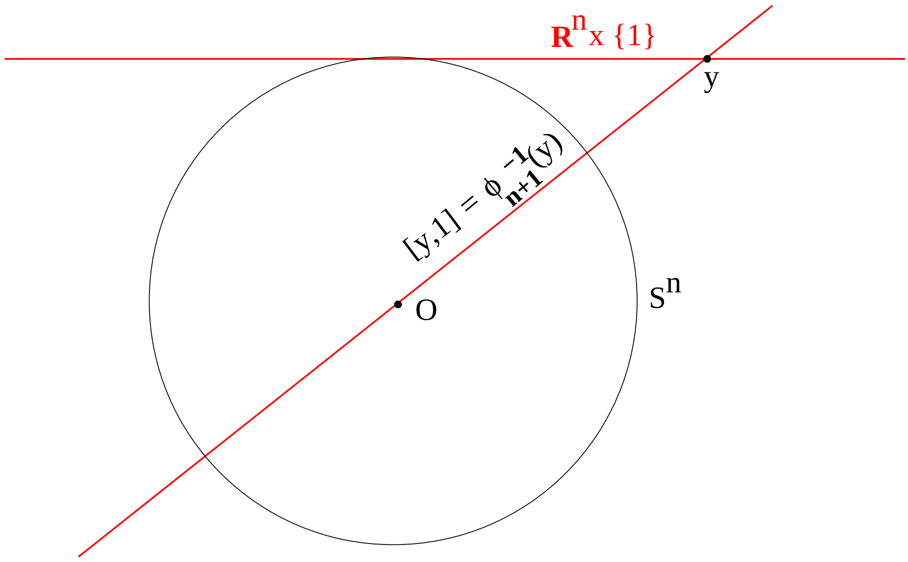
$$U_j = \{[x_1, \ldots, x_{n+1}] \in \mathbb{RP}^n : x_j \neq 0\}.$$

If $P \in U_j$, by dividing its homogeneous co-ordinates through by x_j we get a representation in which the j-th homogeneous coordinate is equal to 1. Every point in U_j has a unique representation of this form. We use this to define a bijection $\phi_j : U_j \to \mathbb{R}^n$:

$$\phi_j([x_1, \ldots, x_{n+1}]) = \left(\frac{x_1}{x_j}, \ldots, \frac{x_{n+1}}{x_j} \right)$$

where the resulting 1 in the j-th place is omitted.

The inverse of this map has a geometrical interpretation which deserves mention, and is illustrated in the following diagram, in which we take $j = n + 1$.



For each $y \in \mathbb{R}^n$, there is a unique line through 0 in $\mathbb{R}^{n+1}$ containing the point $(y, 1)$. This line is of course the linear span of $(y, 1)$, and is thus the equivalence class of $(y, 1)$ under $\sim_1$. It already has a name: $[y, 1]$. It is the image of y under ϕ_{n+1}^{-1}. Notice that every line not contained in the horizontal plane $\mathbb{R}^n \times \{0\}$ meets $\mathbb{R}^{n+1} \times \{1\}$, and is thus in the image of ϕ_{n+1}^{-1}.

The crossover maps in this atlas are easier to compute than in the previous atlas: if $i < j$ then $\phi_j(U_i \cap U_j) = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : y_i \neq 0\}$, and

$$\phi_i \circ \phi_j^{-1}(y_1, \ldots, y_n) = \phi_i\big([y_1, \ldots, y_{j-1}, 1, y_j, \ldots, y_n]\big) = \left(\frac{y_1}{y_i}, \ldots, \frac{y_{j-1}}{y_i}, \frac{1}{y_i}, \frac{y_j}{y_i}, \ldots, \frac{y_n}{y_i}\right) \quad (5.2)$$

where the 1 in the i-th place is omitted. This is obviously smooth, since $y_i \neq 0$ in its domain.

To distinguish this atlas from the one introduced in Example 5.2, we will call it $\mathcal{A}'$. □

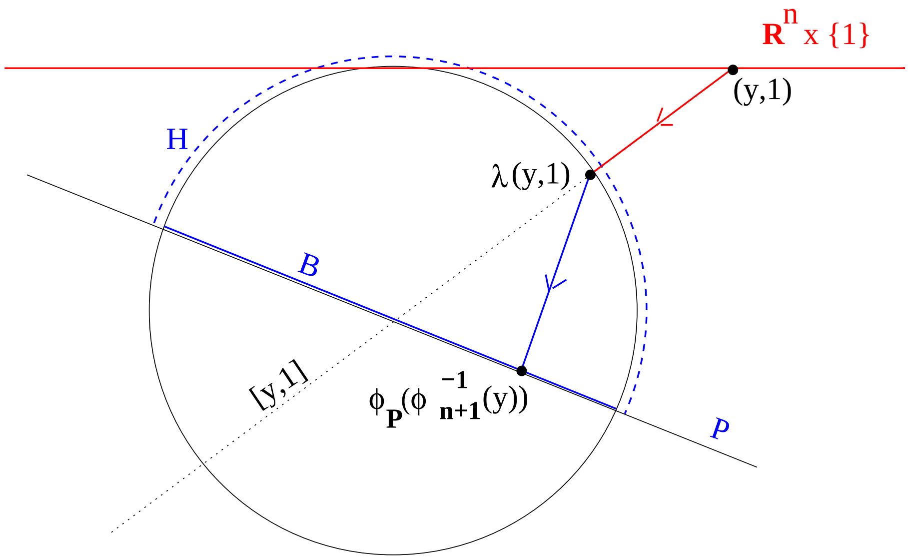
Lemma 5.4. *The two atlases $\mathcal{A}$ and $\mathcal{A}'$ on $\mathbb{RP}^n$ defined in Examples 5.2 and 5.3 determine the same criterion of smoothness.*

Proof Denote charts in $\mathcal{A}$ by ϕ_P and charts in $\mathcal{A}'$ by ϕ_j. It is enough to show that each composite $\phi_P \circ \phi_j^{-1}$ and $\phi_j \circ \phi_P^{-1}$, where defined, is smooth. For then e.g. if $f : \mathbb{RP}^n \to \mathbb{R}^k$ is smooth with respect to $\mathcal{A}$, we have

$$f \circ \phi_j^{-1} = (f \circ \phi_P^{-1}) \circ (\phi_P \circ \phi_j^{-1})$$

and so f is smooth with respect to $\mathcal{A}'$.

Showing smoothness of $\phi_P \circ \phi_j^{-1}$ is straightforward, though it may be guided by the following picture.



The picture shows a map from the domain $\mathbb{R}^n$ of ϕ_j^{-1} to the range B of ϕ_P. By ϕ_{n+1}^{-1} we map the point y to the point $[y, 1]$ in $\mathbb{RP}^n$, which can also be thought of as the line through $(y, 1)$ and 0 shown. This line meets S^n is two antipodal points, one of which, $\lambda(y, 1)$, lies in H. Its image $q(\lambda(y, 1))$ is equal to the point $[y, 1]$ under the identification of $S^2/\sim$ with $(\mathbb{R}^{n+1} \setminus \{0\})/\sim_1$. Thus

$$\phi_P \circ \phi_j^{-1}(y) = \pi \circ q^{-1})(q(\lambda(y, 1)) = \pi(\lambda(y, 1))$$

as shown. In fact $\lambda = \pm 1/\|(y, 1)\|$, and so depends smoothly on y, and thus the map $\mathbb{R}^n \to H$ is smooth. As the composite of this map with the linear projection π, $\phi_P \circ \phi_{n+1}^{-1}$ is also smooth. □

Definition 5.5. Two smooth atlases $\mathcal{A}$ and $\mathcal{A}'$ on the same manifold M are *equivalent* if they determine the same criterion of smoothness

The discussion in the proof of the last lemma suggests the following alternative characterisations of equivalence of atlases.

Proposition 5.6. *The following statements are equivalent to one another:*
(1) The atlases $\mathcal{A}$ and $\mathcal{A}'$ are equivalent.
(2) Every chart in $\mathcal{A}$ is smooth with respect to $\mathcal{A}'$, and vice-versa.
(3) $\mathcal{A} \cup \mathcal{A}'$ is itself a smooth atlas. □

Definition 5.7. *(i) A smooth structure on a topological manifold M is an equivalence class of smooth atlas.*

(ii) A smooth manifold M is a topological manifold together with a choice of smooth structure.

In order to specify a smooth structure on a manifold, it is of course enough to specify just one smooth atlas, or, equivalently, a criterion for smoothness. In effect, this is what we did when we introduced manifolds in Section 1, as subsets of $\mathbb{R}^N$. Subsets M of $\mathbb{R}^N$ meeting the conditions of the old definition 1.6 also meet the conditions of the new definition 5.7: any atlas of smooth (new) diffeomorphisms on M determines a smooth structure, and 1.4, together with 5.6, shows that all of these smooth structures are the same.

Although the atlases $\mathcal{A}$ and $\mathcal{A}'$ of 5.2 and 5.3 define the same smooth structure on $\mathbb{RP}^n$, $\mathcal{A}'$ is preferable to $\mathcal{A}$. There are two reasons: a reason of convenience, that the crossover maps are much simpler in $\mathcal{A}'$ than in $\mathcal{A}$, and a more important reason, namely that the atlas $\mathcal{A}'$ allows us to endow $\mathbb{RP}^n$ with a geometric structure as well as a criterion for smoothness. Suppose that $L \subset \mathbb{R}^n$ is any affine subspace (i.e. a translation of a linear subspace).

Exercise 5.8. *In this case $\phi_j \circ \phi_i^{-1}(L)$ is also an affine subspace, of the same dimension as L.* □

Hint: first read on for a few lines.

In view of 5.8, we can speak of *lines* and *planes* in $\mathbb{RP}^n$, and indeed of higher dimensional 'linear' subspaces. The word 'linear' is in quotes because $\mathbb{RP}^n$ is not a vector space. In fact we make a slightly wider definition:

Definition 5.9. *A k-dimensional linear subspace of $\mathbb{RP}^n$ is the space obtained as the image, under the quotient map $\mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{RP}^n$, of a $k+1$-dimensional vector subspace V of $\mathbb{R}^{n+1}$.*

Since any V^{k+1} is linearly isomorphic to $\mathbb{R}^{k+1}$, and the image $[V]$ of V is just the set of lines through 0 in V, it is itself a projective space of dimension k. The great virtue of the charts $\phi_j \in \mathcal{A}'$ is that if $[V]$ is a linear subspace of $\mathbb{RP}^n$ then $\phi_j([V])$ is an affine subspace of $\mathbb{R}^n$. In fact it is useful to think of the inverse to (say) ϕ_1,

$$(y_1, \ldots, y_n) \mapsto (1, y_1, \ldots, y_n)$$

as an inclusion of $\mathbb{R}^n$ in $\mathbb{RP}^n$, and to view each affine subspace in $\mathbb{R}^n$ as the intersection of $\mathbb{R}^n$ with a projective subspace of $\mathbb{RP}^n$.

Example 5.10. In Example 1.41(ii) we showed that the image X of S^2 under the quadratic map $S^2 \to \mathbb{R}^6$

$$f(x_1, x_2, x_3) = (x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3)$$

is a manifold homeomorphic to $S^2/\sim$. Now we construct a smooth map $g : X \to \mathbb{RP}^2$ making the diagram

$$\begin{tikzcd}
S^2 \arrow[dd, "p"'] \arrow[dr, "f"] & \\
& X \arrow[dl, "g"] \\
\mathbb{RP}^2 &
\end{tikzcd}$$

commute, where $p : S^2 \to \mathbb{RP}^2$ is the quotient map $p(x_1, x_2, x_3) = [x_1, x_2, x_3]$.

The procedure is simple. We want g to map $(x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3)$ to the point in $\mathbb{RP}^2$ represented in homogeneous coordinates by $[x_1, x_2, x_3]$. Denote the coordinates on $\mathbb{R}^6$ by $y_1, \ldots, y_6$. If $x_1 \neq 0$ then $[x_1, x_2, x_3] = [x_1^2, x_1x_2, x_1x_3]$. so we can define g by the formula

$$g(y_1, \ldots, y_6) = [y_1, y_4, y_5].$$

This is valid provided $y_1 \neq 0$, because $f^{-1}(\{y_1 \neq 0\}) = \{x_1 \neq 0\}$.

If $y_1 = 0$ we cannot use this formula to define g; instead, we use

$$g(y_1, \ldots, y_6) = \begin{cases} [y_4, y_2, y_6] & \text{if } y_2 \neq 0 \\ \\ [y_5, y_6, y_3] & \text{if } y_3 \neq 0 \end{cases}$$

Every point in the image of f lies in at least one of $\{y_1 \neq 0\}$, $\{y_2 \neq 0\}$ or $\{y_3 \neq 0\}$, so g is well defined. Composition of g with any of the three standard charts on $\mathbb{RP}^2$ gives a smooth map, so g is smooth.

Exercise 5.11. (i) Show that g is a diffeomorphism by finding a smooth inverse.
(ii)* Show that if M is a manifold and $f : S^2 \to M$ is a smooth map such that $f(x) = f(-x)$ for all $x \in S^2$, then f passes to the quotient to define a smooth map $\bar{f} : \mathbb{RP}^2 \to M$ making the diagram

$$\begin{tikzcd} S^2 \arrow[dd, "p"'] \arrow[dr, "f"] & \\ & M \\ \mathbb{RP}^2 \arrow[ur, "\bar{f}"'] & \end{tikzcd}$$

commute.

Example 5.12. (i) $\mathbb{RP}^3$ is homeomorphic to the Lie group SO(3). There is a simple argument to prove this.
Step 1: $\mathbb{RP}^3$ is the image of a *closed* hemisphere $H \subset S^3$ under the quotient map q; under q, the only identifications in H which take place are those of antipodal points on ∂H. Orthogonal projection to the equatorial hyperplane determines a homeomorphism (but not a diffeomorphism) $\pi : H \to B$, where B is a closed ball in $\mathbb{R}^3$. Thus $q \circ \pi^{-1}$ maps B onto $\mathbb{RP}^3$, in the process identifying antipodal points on ∂B. This map passes to the quotient to give a map — in fact a homeomorphism — from the quotient space $B/\sim$ to $\mathbb{RP}^3$, where $\sim$ is the equivalence relation which identifies antipodal points on ∂B.
Step 2: SO(3) is also homeomorphic to $B/\sim$. For SO(3) is the set of rotations about axes in $\mathbb{R}^3$. A point $x \in B$ determines such a rotation; unless it is the centre of the ball, 0, x lies on a unique line ℓ_x through 0, which we can take as our axis of rotation. We use the norm of x to determine the angle of rotation, and we use the position of x on the line ℓ_x to determine a sense of rotation. In detail: provided $x \neq 0$, it determines an orientation of the plane $x^\perp$, and a rotation through $\|x\|\pi$ in the positive sense in this plane. As $\mathbb{R}^3 = \ell_x \oplus x^\perp$, we get a

well defined rotation about ℓ_x. If $x = 0$, it does not matter that we have not defined an axis of rotation, since we intend only to rotate through 0 radians, and a rotation through 0 about one axis is the same as a rotation through 0 about any other. Finally, a rotation through π in one sense is the same thing as a rotation about π in the opposite sense. So *antipodal points on ∂B determine the same rotation.* Thus, we have a surjection $B \to \mathrm{SO}(3)$ which is 1-1 except that it identifies antipodal points on ∂B.

Together, Steps 1 and 2 show that there is a homeomorphism $\mathbb{RP}^3 \to \mathrm{SO}(3)$.

(ii) In fact $\mathbb{RP}^3$ and $\mathrm{SO}(3)$ are *diffeo*morphic. Rather than trying to improve the map in (i), we use a different approach. First we define a smooth map $\rho : S^3 \to \mathrm{SO}(3)$, and then we show that it passes to the quotient to define a smooth map $\mathbb{RP}^3 \to \mathrm{SO}(3)$. To define ρ, we think of S^3 as the multiplicative group of unit quaternions:

$$(a, b, c, d) \leftrightarrow a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

and look for an orthogonal representation of the unit quaternions on $\mathbb{R}^3$. There is an ingenious way of doing this. We identify points in $\mathbb{R}^3$ with quaternions with zero real part:

$$(x, y, z) \leftrightarrow x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

and then let $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ act on the space of such quaternions by conjugation:

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \mapsto (a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k})^{-1}(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) \tag{5.3}$$

(note that the right hand side of (5.3) has zero real part). As a linear map in (x, y, z), this map has matrix

$$\begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2(ad + bc) & 2(bd - ac) \\ 2(bc - ad) & a^2 - b^2 + c^2 - d^2 & 2(cd + ab) \\ 2(ac + bd) & 2(cd - ab) & a^2 - b^2 - c^2 + d^2 \end{pmatrix}. \tag{5.4}$$

It is not completely obvious that this matrix is in $\mathrm{SO}(3)$, although it is straightforward to check that it is. Note that since ψ is smooth and a group homomorphism, it is obvious that the image is a 3-dimensional compact subgroup of $\mathrm{Gl}_3(\mathbb{R})$.

Exercise 5.13. (i) Show that (5.4) is indeed the matrix of the linear map (5.3).
(ii) Check that $(5.4) \in \mathrm{SO}(3)$.
(iii) Show that the map $S^3 \to \mathrm{SO}(3)$ defined by sending $(a, b, c, d) \in S^3$ to the matrix (5.4) passes to the quotient to define a smooth map $\mathbb{RP}^3 \to \mathrm{SO}(3)$.
(iv) Show that the map $\mathbb{RP}^3 \to \mathrm{SO}(3)$ you found in (iii) is a a diffeomorphism.

Exercise 5.14. Write down an explicit formula for the map $B^3 \to \mathbb{RP}^3$ of Example 5.12(i) Step 1, with respect to the standard coordinates on B^3 and homogeneous coordinates on $\mathbb{RP}^3$. Is it smooth?

Exercise 5.15. The map $f : B^3 \to \mathrm{SO}(3)$ described in Example 5.12(i) above can be written explicitly in coordinates, though it is not absolutely straightforward to do so. As a first step, it is relatively easy to find the restriction of f to the boundary S^2 of B^3. Recall that $f(x, y, z)$ consists of a rotation through an angle of $\|(x, y, z)\|\pi$ about the line spanned by (x, y, z). If $\|(x, y, z)\| = 1$ then this rotation is just multiplication by -1 on the orthogonal complement $(x, y, z)^\perp$. If v_1, v_2 is *any* basis of $(x, y, z)^\perp$, the matrix of this rotation with respect to the basis $E = (x, y, z), v_1, v_2$ is therefore

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

To get its expression with respect to the standard basis of $\mathbb{R}^3$ we conjugate by the change of basis matrix:

$$f(x, y, z) = [I]_N^E \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} [I]_E^N.$$

(Recall that $[I]_N^E$ is the matrix whose columns are the expressions of $(x, y, z), v_1$ and v_2 with respect to the standard basis (so the first column is $(x, y, z)^t$), and $[I]_E^N$ is its inverse.)
To do:(i) Provided $y \neq 0$, we can take $v_1 = (-y, x, 0), v_2 = (0, -z, y)$ as basis for $(x, y, z)^\perp$. Using this basis, compute the matrix $f(x, y, z)$.
(ii) Check that, after simplification, the expression you have obtained is valid even if $y = 0$. Can you explain why?
(iii) Show that the map $S^2 \to \mathrm{SO}(3)$ you have obtained passes to the quotient to define a smooth map $\mathbb{RP}^2 \to \mathrm{SO}(3)$.
(iv) Find explicit formula(e) for its inverse.

Example 5.16. We can play the same game with the group $\mathrm{Gl}_n(\mathbb{R})$ as with $\mathbb{R}^{n+1} \setminus \{0\}$. That is, we identify two matrices A_1 and A_2 if one is a scalar multiple of the other. The quotient of $\mathrm{Gl}_n(\mathbb{R})$ by this equivalence relation is called $\mathbb{P}\mathrm{Gl}_n(\mathbb{R})$. We denote the equivalence class of a matrix A by $[A]$. If we think of $\mathrm{Mat}_{n\times n}(\mathbb{R})$ as $\mathbb{R}^{n^2}$, then clearly $\mathrm{Gl}_n(\mathbb{R}) \subset \mathbb{R}^{n^2} \setminus \{0\}$. The equivalence relation we have just defined on $\mathrm{Gl}_n(\mathbb{R})$ is the same as the quivalence relation used to obtain $\mathbb{RP}^{n^2-1}$ from $\mathbb{R}^{n^2} \setminus \{0\}$. Thus $\mathbb{P}\mathrm{Gl}_n(\mathbb{R})$ is an (open) subset of $\mathbb{RP}^{n^2-1}$, and in particular a manifold of dimension $n^2 - 1$. It is also a group: $A \sim B$ if there exists $\lambda \neq 0$ such that $\lambda I_n A = B$, and thus the equivalence classes are the cosets of the normal subgroup S of $\mathrm{Gl}_n(\mathbb{R})$ consisting of scalar matrices

$$S := \{\lambda I_n : 0 \neq \lambda \in \mathbb{R}\}.$$

So $\mathbb{P}\mathrm{Gl}_n(\mathbb{R})$ is the quotient of $\mathrm{Gl}_n(\mathbb{R})$ by S. What this amounts to is that if $A' = \lambda_1 A$ and $B' = \lambda_2 B$ then $A'B' = \lambda_1\lambda_2 AB$, so that *the group operation in $Gl_n(\mathbb{R})$ (multiplication) passes to the quotient to define a group operation in $\mathbb{P}Gl_n(\mathbb{R})$.* □

Proposition 5.17. *$\mathbb{P}Gl_n(\mathbb{R})$ is a Lie group.*

Proof Denote the product operation on $\mathbb{P}\mathrm{Gl}_n(\mathbb{R})$ by m, and the operation of inversion by i. The chart domains in $\mathbb{P}\mathrm{Gl}_n(\mathbb{R})$ are the sets

$$U_{ij} := \{[A] \in \mathbb{P}\mathrm{Gl}_n(\mathbb{R}) : a_{ij} \neq 0\},$$

and the chart ϕ_{ij} is defined by

$$\phi_{ij}([A]) = \begin{pmatrix} a_{11}/a_{ij} & a_{12}/a_{ij} & \cdots & \alpha_{1n}/a_{ij} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}/a_{ij} & a_{n2}/a_{ij} & \cdots & \alpha_{nn}/a_{ij} \end{pmatrix}.$$

The matrix on the right has a 1 in the i, j position: we regard it as an element of $\mathbb{R}^{n^2-1}$. To ease problems of notation, we denote this copy of $\mathbb{R}^{n^2-1}$ by V_{ij}. Elements of V_{ij} will be written in the form $(x) = (x_{\alpha\beta})$, where it is understood that $x_{ij} = 1$.

Now we show that $m : \mathbb{P}\mathrm{Gl}_n(\mathbb{R}) \times \mathbb{P}\mathrm{Gl}_n(\mathbb{R}) \to \mathbb{P}\mathrm{Gl}_n(\mathbb{R})$ is smooth. Suppose that $[A] \in U_{ij}$, $[B] \in U_{k\ell}$ and $[AB] \in U_{pq}$. Then around $(\phi_{ij}([A]), \phi_{k\ell}([B])) \in V_{ij} \times V_{k\ell}$, the map

$$\phi_{pq} \circ m \circ (\phi_{ij}^{-1}, \phi_{k\ell}^{-1})$$

takes the form

$$((x), (y)) \mapsto \Big(\frac{\sum_\beta x_{\alpha\beta} y_{\beta\gamma}}{\sum_\beta x_{p\beta} y_{\beta q}}\Big).$$

This is evidently smooth wherever the denominator is non-zero, i.e. on

$$(\phi_{ij} \times \phi_{k\ell}) \, \big(m^{-1}(U_{pq})\big),$$

as required.

I leave to the reader the task of verifying that inversion i also determines a smooth map $\mathbb{P}\mathrm{Gl}_n(\mathbb{R}) \to \mathbb{P}\mathrm{Gl}_n(\mathbb{R})$. $\square$

In Example 2.18(2) above, we used Sard's theorem to show that not all quadruples of lines through $(0, 0)$ in $\mathbb{R}^2$ are linearly isomorphic to one another, by looking at the way that the group $\mathrm{Gl}_2(\mathbb{R})$ acts on quadruples of lines. We can now think of this slightly differently. The space $\mathcal{Q}$ of quadruples of distinct lines though $(0, 0)$ is an open set in $(\mathbb{RP}^1)^4$. Thus it is a manifold of dimension 4. The action of $\mathrm{Gl}_2(\mathbb{R})$ on $\mathcal{Q}$ is smooth, and passes to the quotient to define a smooth action of $\mathbb{P}\mathrm{Gl}_2(\mathbb{R})$ on $\mathcal{Q}$. The equivalence class of a given quadruple (say the quadruple X_1 of Example 2.18) is the image of the group $\mathbb{P}\mathrm{Gl}_2(\mathbb{R})$ under the smooth map $[A] \mapsto A \cdot X_1$. Since $\dim \mathbb{P}\mathrm{Gl}_2(\mathbb{R}) = 3$ and $\dim \mathcal{Q} = 4$, this image has measure zero in $\mathcal{Q}$, by Sard's Theorem.

5.2 An important example - the Grassmanian

Now we give another example of abstract manifold, generalising $\mathbb{RP}^n$ in an obvious way. The *Grassmanian* $G_{k,n}$ is the space of all k-dimensional linear subspaces of $\mathbb{R}^n$. In particular $G_{1,n+1} = \mathbb{RP}^n$. To save breath, we call the elements of $G_{k,n}$ *k-planes* in $\mathbb{R}^n$. At the outset, $G_{k,n}$ doesn't even have a topology, though most likely we would agree that we have an intuitive notion of the relative closeness of different linear subspaces to one another. We give it a topology and a $k(n-k)$-dimensional smooth structure at the same time, by defining bijections between certain subsets of $G_{k,n}$ and $\mathbb{R}^{k(n-k)}$, and then declaring that their domains are open and that the bijections are homeomorphisms, and calling them charts. For this to define a topology, the crossover maps between the bijections must be homeomorphisms. This will follow from the first thing we prove, which will be that they are actually diffeomorphisms. This is the serious part of the task. Everything else is just what we define it to be; but the fact that it is possible to choose rather natural charts on $G_{k,n}$ such that all the crossover maps are smooth, is a fact of nature, and not just a matter of definition.

The smooth structure we will define has the property that if $v_1(t), \ldots, v_k(t)$ are vectors in $\mathbb{R}^n$, depending smoothly on a real variable t, and are independent for all t, then the k-plane they span, $P(t)$, will also vary smoothly in $G_{k,n}$. Indeed, it is the unique smooth structure with this property.

Construction of the bijections Choose a basis $e_1, \ldots, e_n$ for $\mathbb{R}^n$. We will define a chart on $G_{k,n}$ for each subset I of $\{1, 2, \ldots, n\}$ containing k elements. Let I be such a subset, let J be its complement in $\{1, \ldots, n\}$, and let $W_J = \mathrm{Sp}\{e_j : j \in J\}$. We take, as the domain of our chart, the set

$$U_I = \{V \in G_{k,n} : V \cap W_J = \{0\}\}.$$

For example, if $n = 3$, $k = 2$, and $I = \{1, 2\}$ then U_I is the set of 2-planes which meet the x_3-axis only at 0.

Lemma 5.18. *Let $I = \{i_1, \ldots, i_k\}$, let $V \in G_{n,k}$ and let $v_1 = (v_{11}, \ldots, v_{1,n}), \ldots, v_k = (v_{k1}, \ldots, v_{kn})$ be a basis of V. Let M be the $n \times k$ matrix*

$$\begin{pmatrix} v_{11} & \cdots & v_{k1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ v_{1n} & \cdots & v_{kn} \end{pmatrix}$$

Then
(i) $V \in U_I$ if and only if the $k \times k$ submatrix M_I of M, formed by rows $i_1, \ldots, i_k$, is invertible.

(ii) Each $V \in U_I$ has a unique basis $v_1^I, \ldots, v_k^I$ for which M_I is the $k \times k$ identity matrix. $\square$

For each $V \in U_I$, denote by $M^I(V)$ the matrix of 5.18 for the basis $v_1^I, \ldots, v_k^I$ of 5.18(ii). We define $\varphi_I : U_I \to \mathrm{Mat}_{(n-k)\times k} \simeq \mathbb{R}^{k(n-k)}$ by $\varphi_I(V) = M_J^I(V)$, where $M_J^I(V)$ is the $(n-k) \times k$

matrix formed by rows $j_1, \ldots, j_{n-k}$ of $M^I(V)$. We recover the matrix $M^I(V)$ from $M^I_J(V)$ by interspersing, among the rows of $M^I_J(V)$, the k rows of the $k \times k$ identity matrix, so that in the resulting $n \times k$ matrix they occupy positions $i_1, \ldots, i_k$. Applying this procedure to an arbitrary $(n-k) \times k$ matrix A gives us a $n \times k$ matrix which we denote by $\psi_I(A)$, whose columns span a k-plane V; this k-plane is $\varphi_I^{-1}(A)$.

Now we consider the crossover maps

$$\varphi_{I_1} \circ \varphi_{I_2}^{-1} : \varphi_{I_2}(U_{I_1} \cap U_{I_2}) \to \varphi_{I_1}(U_{I_1} \cap U_{I_2}),$$

where I_1 and I_2 are two different k-element subsets of $\{1, \ldots, n\}$.

Lemma 5.19. *(i) Let $V \in U_{I_1}$. Then $V \in U_{I_2}$ also if and only if $M^{I_1}_{I_2}(V)$ is invertible.*

(ii) If $V \in U_{I_1} \cap U_{I_2}$ then $M^{I_2}(V) = M^{I_1}(V)(M^{I_1}_{I_2}(V))^{-1}$.

(iii) $\varphi_{I_1} \circ \varphi_{I_2}^{-1}$ is a diffeomorphism.

Proof (i) is just 5.18(i). For (ii), simply observe that in the product $M^{I_1}(V)(M^{I_1}_{I_2})^{-1}$, the $k \times k$ submatrix formed by the rows corresponding to I_2 is the $k \times k$ identity matrix (as it is just $(M^{I_1}_{I_2}(V))^{-1}M^{I_1}_{I_2}(V)$). Thus $M^{I_1}(V)(M^{I_1}_{I_2}(V))^{-1} = M^{I_2}(V)$, since this property characterises $M^{I_2}(V)$.

(iii) Given an $(n-k)\times k$ matrix A, we obtain $\varphi_{I_1}(\varphi_{I_2}^{-1}(A))$ by first forming the matrix $\psi_{I_2}(A)$, then right-multiplying it by the inverse of its submatrix consisting of its i_1'st, ..., i_k'th rows, and then throwing away those rows (which are now the rows of the identity matrix) from the product. In symbols,

$$\varphi_{I_1} \circ \varphi_{I_2}^{-1}(A) = \Big(\psi_{I_2}(A)_{I_1}^{-1}\big(\psi_{I_2}(A)\big)\Big)_J$$

where J is the complement of I. This map is clearly smooth. Since it has inverse $\varphi_{I_2} \circ \varphi_{I_1}^{-1}$, which is also smooth, by the same argument, it is a diffeomorphism. □

We give $G_{k,n}$ a topology by declaring that each of the U_I is open, and that each of the φ_I is a homeomorphism. In other words, the open sets in $G_{k,n}$ are, first, the preimages under the maps φ_I of open sets in $\mathrm{Mat}_{(n-k)\times k}$, and second, arbitrary unions of these preimages. This procedure does indeed define a topology on $G_{k,n}$. Notice that we need to know that $\varphi_{I_1} \circ \varphi_{I_2}^{-1}$ is a homeomorphism for this to work: for it establishes that a subset of $G_{k,n}$ which is open with respect to one chart is also open with respect to any other.

So far, all of our construction has been with respect to a fixed basis E of $\mathbb{R}^n$. Is the choice of basis important? If we use a different basis, do we get the same smooth structure? Let E' be another basis for $\mathbb{R}^n$, and let I' be a k-element subset of $\{1, \ldots, n\}$. Denote by $\varphi'_{I'} : U'_{I'} \to \mathrm{Mat}_{(n-k)\times k}$ the chart corresponding to I', constructed using the basis E'.

Proposition 5.20. *Each such map $\varphi'_{I'} : U'_{I'} \to Mat_{(n-k)\times k}$ is a diffeomorphism, with respect to the smooth structure defined using the charts φ_I constructed using the basis E.*

Proof This is a consequence of the next lemma. □

Lemma 5.21. *The map*

$$\varphi_I \circ \varphi'^{-1}_{I'} : \varphi'_{I'}(U'_{I'} \cap U_I) \to \varphi_I(U'_{I'} \cap U_I)$$

is smooth.

Proof Exercise. Use the fact that if $v_1, \ldots, v_k$ is a basis for the k-plane V, and M' is the matrix of 5.18, whose columns are the expression of the vectors $v_1, \ldots, v_k$ with respect to the basis E', then $M' = [I]^E_{E'}M$, where $[I]^E_{E'}$ is the change of basis matrix (see Appendix 1, 7.1). □

Now it is clear, by 5.20, that we can include *all* maps $\varphi'_{I'}$ as charts in our atlas for $G_{k,n}$. The choice of basis E is completely immaterial.

It remains to show that $G_{k,n}$ is Hausdorff and second countable. Second countability is an immediate consequence of the fact that a finite number of chart domains U_I cover $G_{k,n}$, and that each of them is itself second countable. To show that $G_{k,n}$ is Hausdorff, observe that if V_1 and V_2 are any two k-planes, there is an $(n-k)$-plane W complementary to them both. Choose a basis E' for $\mathbb{R}^n$ such that $W = \mathrm{Sp}\{e_{k+1}, \ldots, e_n\}$, and let $I' = \{1, \ldots, k\}$. Then V_1 and V_2 both belong to $U'_{I'}$, which is homeomorphic, via φ'_I, to $\mathrm{Mat}_{(n-k)\times k}$. As $\mathrm{Mat}_{(n-k)\times k}$ is Hausdorff, so is U_I, and we may choose disjoint neighbourhoods O_1, O_2 of V_1 and V_2, open in $U'_{I'}$. As $U'_{I'}$ is open in $G_{k,n}$, so are O_1 and O_2.

Exercise 5.22. Show that if $v_1(t), \ldots, v_k(t)$ are vectors in $\mathbb{R}^n$, each depending smoothly on a parameter t, and are linearly independent for all t, then the map

$$\mathbb{R} \to G_{k,n}$$

sending t to $\mathrm{Sp}\{v_1(t), \ldots, v_k(t)\}$ is smooth.

Exercise 5.23. The projective space $\mathbb{RP}^n$ is the special case $k = 1$ of the Grassmannian $G_{k,n+1}$. We have described in detail how to construct charts on $\mathbb{RP}^{n+1}$ and on $G_{k,n+1}$. Do they give the same charts when $k = 1$? □

5.3 Complex manifolds

Example 5.24. The *complex projective space* $\mathbb{CP}^n$ is defined in close analogy to $\mathbb{RP}^n$: it is the space of complex lines through 0 in $\mathbb{C}^{n+1}$, and thus the quotient of $\mathbb{C}^{n+1} \setminus \{0\}$ by the equivalence relation

$$(z_1, \ldots, z_{n+1}) \sim (\lambda z_1, \ldots, \lambda z_{n+1})$$

for $\lambda \in \mathbb{C} \setminus \{0\}$. An atlas $\{\phi_j : U_j \to \mathbb{C}^n\}$ is defined by exactly the same procedure as for $\mathbb{RP}^n$, and the crossover maps are given by the same formula (5.2), with respect to complex coordinates on $\mathbb{C}^n = \mathbb{R}^{2n}$. The crossover maps are not just smooth, but actually holomorphic — that is, differentiable, in the complex sense, with respect to each complex variable y_j. In particular, the derivative $d_y(\phi_j \circ \phi_i^{-1}) : \mathbb{R}^{2n} = \mathbb{C}^n \to \mathbb{C}^n = \mathbb{R}^{2n}$ is complex-linear. $\square$

Definition 5.25. Identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ in some fixed way, e.g.

$$(x_1 + iy_1, \ldots, x_n + iy_n) \to (x_1, y_1, x_2, y_2, \ldots, x_n, y_n).$$

An even dimensional smooth manifold M^{2n} equipped with an atlas A is an n-dimensional *complex manifold* if all the crossover maps of charts in A are holomorphic, with respect to this identification.

The Grassmanian $G_{k,n}(\mathbb{C})$ of k-dimensional complex subspaces of $\mathbb{C}^n$ is a complex manifold of (complex) dimension $k(n-k)$. For the linear algebra involved in the construction of the bijections φ_I works over *any* field, not just $\mathbb{R}$, and inspection of the crossover maps $\varphi_{I_1} \circ \varphi_{I_2}^{-1}$ shows that they are in fact rational maps (each component function is a quotient of polynomials in the $k(n-k)$ complex variables) and therefore holomorphic.

Projective spaces play a central role in algebraic geometry. They are natural homes to many manifolds that do not naturally embed in Euclidean space. For example, there is a well-known embedding, due to Plücker, of the Grassmannian $G_{k,n}$ into a suitable projective space (in fact, $\mathbb{RP}^{\binom{n}{k}-1}$) -see 5.27 below. For complex manifolds the advantage is even more dramatic: a compact complex manifold of dimension > 0 cannot be embedded as a complex submanifold of $\mathbb{C}^N$ for any N. The reason is simply that if $M \hookrightarrow \mathbb{C}^N$ is a holomorphic embedding then each complex coordinate function on $\mathbb{C}^N$ restricts to a holomorphic function f_j on M. As M is compact, $|f_j|$ must attain its maximum value somewhere on M. Let ϕ be a chart around this point. Because $|f_j \circ \phi^{-1}|$ has a local maximum, the maximum modulus principal of the theory of complex variables implies that $f_j \circ \phi^{-1}$ is constant. So by analytic continuation $f_j = z_j$ is locally constant on all of M, and thus M is a collection of discrete points.

On the other hand, very many compact complex manifolds do embed as complex submanifolds of $\mathbb{CP}^n$ — for example, the complex Grassmanian, via the Plücker embedding.

Exercise 5.26. (1) Recall that $\mathrm{SO}(n)$ is the Lie group of $n \times n$ orthogonal real matrices with positive determinant. Let $V_0 = \mathrm{Sp}\{e_1, \ldots, e_k\} \subset \mathbb{R}^n$, and define a map $f : \mathrm{SO}(n) \to G_{k,n}$ by $f(A) = A \cdot V_0$; that is, $f(A)$ is the k-plane to which V_0 is mapped by A. Show
(i) f is surjective;
(ii) f is smooth;
(iii) $G_{k,n}$ is compact.

(2) Do the same with the complex Grassmanian $G_{k,n}(\mathbb{C})$, using $U(n)$ in place of $\mathrm{SO}(n)$.

Exercise 5.27. In this exercise we define and study the Plücker embedding of $G_{k,n}$ in a certain projective space.

(1) There are $\binom{n}{k}$ k-element subsets of $\{1, \ldots, n\}$. Define a map $P : (\mathbb{R}^n)^k \to \mathbb{R}^{\binom{n}{k}}$ as follows: given $(v_1, \ldots, v_k) \in (\mathbb{R}^n)^k$, form the $n \times k$ matrix M as in 5.18. Map $(v_1, \ldots, v_k)$ to the k-tuple consisting of the determinants of all the distinct $k \times k$ submatrices of M. Show

1. $v_1, \ldots, v_k$ are linearly independent if and only if $P(v_1, \ldots, v_k) \neq 0$.
2. $P(v_1, \ldots, v_i, \ldots, v_j + \lambda v_i, \ldots, v_k) = P(v_1, \ldots, v_k)$.
3. If $v_1, \ldots, v_k$ span the same k-plane as $w_1, \ldots, w_k$ then $P(v_1, \ldots, v_k)$ is a scalar multiple of $P(w_1, \ldots, w_k)$.
4. P passes to the quotient to give rise to a smooth map $p : G_{k,n} \to \mathbb{RP}^{\binom{n}{k}-1}$. More precisely, if $(\mathbb{R}^n)^k_0$ is the open subset of $(\mathbb{R}^n)^k$ consisting of linearly independent k-tuples of vectors, there is a smooth map p making the diagram

$$\begin{array}{ccc} (\mathbb{R}^n)^k & \xrightarrow{P} & \mathbb{R}^{\binom{n}{k}} \setminus \{0\} \\ \downarrow & & \downarrow \\ G_{k,n} & \xrightarrow{p} & \mathbb{RP}^{\binom{n}{k}-1} \end{array}$$

   commute, where the left vertical arrow takes $(v_1, \ldots, v_k)$ to $\mathrm{Sp}\{v_1, \ldots, v_k\}$ and the right vertical arrow is the usual quotient map.

5. p is injective.
6. p is differentiable.
7. p is an embedding (a diffeomorphism onto its image).

(2) Check that if the constructions are made with $\mathbb{C}$ in place of $\mathbb{R}$ then we obtain a holomorphic embedding of $G_{k,n}(\mathbb{C})$ into $\mathbb{CP}^{\binom{n}{k}-1}$.

For more information on the Plücker embedding, see e.g. [4].

Terminology A smooth manifold is really a pair, consisting of a topological manifold and a smooth structure. It becomes slightly tedious to refer always to smooth manifolds in this way, however, and so we will usually leave the smooth structure unnamed, and refer simply to a smooth manifold M. This is, after all, what we do when we refer to a group G: we mean a pair consisting of a set together with a multiplication. When we speak of smooth manifolds in this way, we will speak of smooth charts, meaning charts which are smooth diffeomorphisms with respect to the given (but unnamed) smooth structure. However, one must be careful: the same manifold (as topological space) may have many different smooth structures. This phenomenon was first described by Milnor in the 1960's: he showed that there are 28 different smooth structures on the sphere S^7.

5.4 Quotient spaces as manifolds

We've already observed the versatility of the notion of the quotient of a topological space by an equivalence relation. The abstract definition of manifold explored in this chapter allows us to recover some of this versatility in the case where the equivalence relation is induced by a smooth group action.

Let M be a manifold and let G be a group of diffeomorphisms of M. Define a relation $\sim$ on M by

$$x_1 \sim x_2 \text{ if there exists } \varphi \in G \ \text{ such that } \varphi(x_1) = x_2.$$

It's easy to check that this is an equivalence relation. We denote the set of equivalence classes by M/G, and the quotient map $M \to M/G$ by q. It's useful to write $q(x) = [x]$.

Example 5.28. 1. $M = S^2, G = \{\mathrm{id}_{S^2}, \text{antipodal map}\}$. Thus $x_1 \sim x_2$ if $x_1 = \pm x_2$. This is the case studied in Example 1.41(ii). The quotient is $\mathbb{RP}^2$.

2. $M = \mathbb{R}^{n+1} \smallsetminus \{0\}$, $G =$ group of scalar maps $\{x \mapsto \lambda x : \lambda \in \mathbb{R}^*\}$. So $x_1 \sim x_2$ if there exists a non-zero $\lambda \in \mathbb{R}$ such that $x_1 = \lambda x_2$. The quotient is $\mathbb{RP}^n$.

3. $M = \mathbb{R}^2$, $G =$ group of translations $T_{m,n}(x, y) = (x + m, y + n)$ with $m, n \in \mathbb{Z}$. So $(x_1, y_1) = (x_2, y_2)$ if $(x_1 - x_2, y_1 - y_2) \in \mathbb{Z} \times \mathbb{Z}$.

4. $M = \mathbb{R}^2$, G is the group generated by the two maps

$$\varphi_1(x, y) = (x, y + 1), \quad \varphi_2(x, y) = (x + 1, 1 - y).$$

In each of the first three examples, the abstract structure of the group$G$ is evident: it is, respectively, $\mathbb{Z}/2$ (in its incarnation as $\{\pm 1\}$), $\mathbb{R}^\times$ (the multiplicative group of non-zero real numbers) and $\mathbb{Z} \times \mathbb{Z}$. Where the abstract structure of the group is clear, and where we know how each element of G operates as a diffeomorphism of M, we often use the abstract name of G; so Examples 5.28(i),(ii),(iii) become $S^2/\mathbb{Z}_2$ or $S^2/\{\pm 1\}$, $\mathbb{R}^{n+1} \smallsetminus \{0\}/\mathbb{R}^\times$ and $\mathbb{R}^2/\mathbb{Z} \times \mathbb{Z}$ respectively.

Proposition 5.29. *If X is a topological space and G is a group of homeomorphisms of X, then the quotient map $q : X \to X/G$ is an open map.*

Proof By definition of the quotient topology, a subset W of X/G is open if its preimage in X, $q^{-1}(W)$, is open. Let $U \subset X$ be open. We want to show that $q(U)$ is open in X/G. Now $q^{-1}(q(U)) = \cup_{\varphi \in G} \varphi(U)$. As each φ is a homeomorphism, each set $\varphi(U)$ is open. So $q^{-1}(q(U))$ is a union of open sets, and therefore open. □

When is M/G, with the quotient topology, a smooth manifold? More precisely, when can we construct, in a natural way, a smooth structure on M/G? The smooth structure on

M/G should evidently bear some relation to the smooth structure on M, so for example the map q should be smooth with respect to the smooth structures on M and on M/G. It is also desirable that the following usinversal property should hold:

every time there is a smooth map $f : M \to N$ such that $f(x_1) = f(x_2)$ if $x_1 \sim x_2$, then there exists a unique smooth map $\bar{f} : M/G/ \to N$ (the "quotient map") making the diagram

$$\begin{tikzcd} M \arrow[d, "q"'] \arrow[dr, "f"] & \\ M/G \arrow[r, "\bar{f}"'] & N \end{tikzcd}$$

commutative.

It is one of the features of the definition of quotient topological space that the corresponding property holds, with "continuous" in place of "smooth".

The notion of manifold is more delicate than that of topological space. It turns out that not every quotient M/G is a manifold.

Example 5.30. $M = \mathbb{R}$, $G = \{\pm 1\}$ acting by multiplication. So $x_1 \sim x_2$ if $x_1 = \pm x_2$. The quotient space is homeomorphic to the half-space $\mathbb{R}_{\geq 0}$ (exercise), and the point $[0]$ has no neighbourhood in $\mathbb{R}/G$ homeomorphic to an open set in $\mathbb{R}$.

Definition 5.31. *The group G of homeomorphisms of the topological space X acts* **properly discontinuously** *if for every $x \in X$ there is a neighbourhood W of x in X such that $\varphi(W) \cap W = \emptyset$ if $\varphi \in G$ and $\varphi \neq id_X$.*

In Examples 5.28(i),(iii),(iv), G acts properly discontinuously. In Examples 5.28(ii) and 5.30 it does not.

Proposition 5.32. *If the group G of homeomorphisms of X acts properly discontinuously then $q : X \to X/G$ is a local homeomorphism.*

Proof The statement means that for all $x \in X$, there exists a neighbourhood W of x such that $q_| : W \to q(W)$ is a homeomorphism. Because q is continuous and open, it's only necessary to show that there is a neighbourhood W of x on which q is 1-1. But this holds if W is as in Definition 5.31. □

We impose one further condition:

Definition 5.33. *The group G of homeomorphisms of the topological space X acts* **with property (*)** *if for every $x_1, x_2 \in X$ with x_1 not equivalent to x_2, there are neighbourhoods U_1, U_2 of x_1, x_2 such that*

$$\textit{for all } \varphi_1, \varphi_2 \in G,\ \varphi_1(U_1) \cap \varphi_2(U_2) = \emptyset. \tag{5.5}$$

By taking $\varphi_1 = \varphi_2 = \mathrm{id}_X$, we see that X must be very nearly Hausdorff for it to be possible for a group of diffeomorphisms to act with property (*): inequivalent points, at least, must have disjoint neighbourhoods.

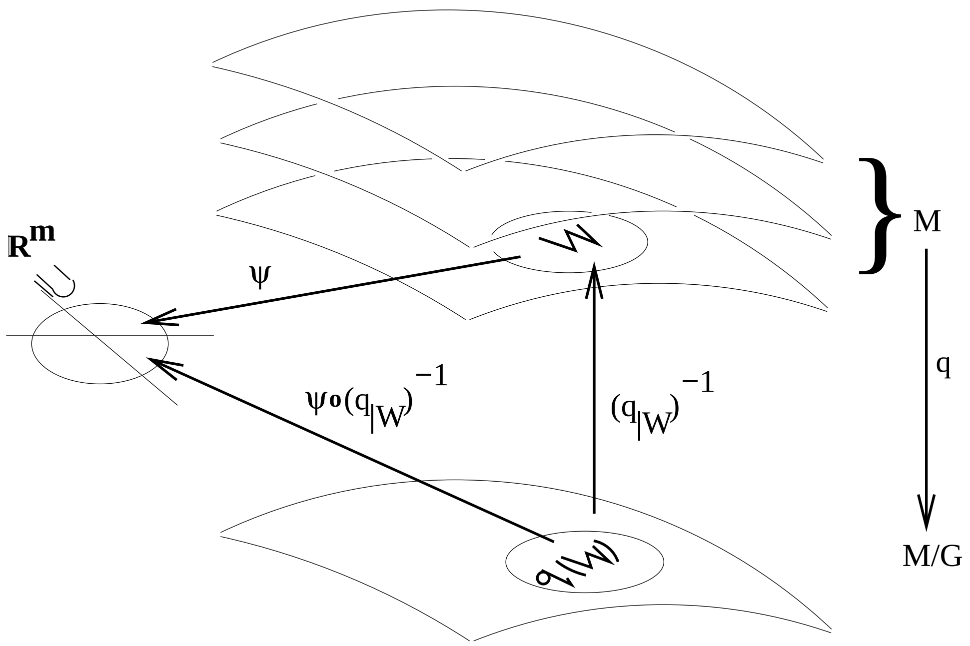
Proposition 5.34. *If G acts on X with property (*) then X/G is Hausdorff.*

Proof Let $[x_1] \neq [x_2] \in X/G$. Then the points x_1, x_2 in X are inequivalent. Choose neighbourhoods U_1 and U_2 of x_1 and x_2 satisfying (5.5). Then $q(U_1)$ and $q(U_2)$ are disjoint neighbourhoods of $[x_1]$ and $[x_2]$. □

Proposition 5.35. *(i) If X is 2nd countable then X/G is 2nd countable.*
(ii) If X is paracompact then X/G is paracompact.

Proof (i) 2nd countability just means having a countable dense set. This property is obviously inherited by a quotient.
(ii) `In progress.` □

Let the group G act properly discontinuously on the manifold M. The inverses to the local homeomorphisms of Proposition 5.32, followed by charts on M, make up an atlas on M/G, so at least M/G is a topological manifold. Of course, we have to shrink the open set W until it lies in the domain of a chart on M, but such shrunken W's still cover M, and their images in M/G cover M/G.

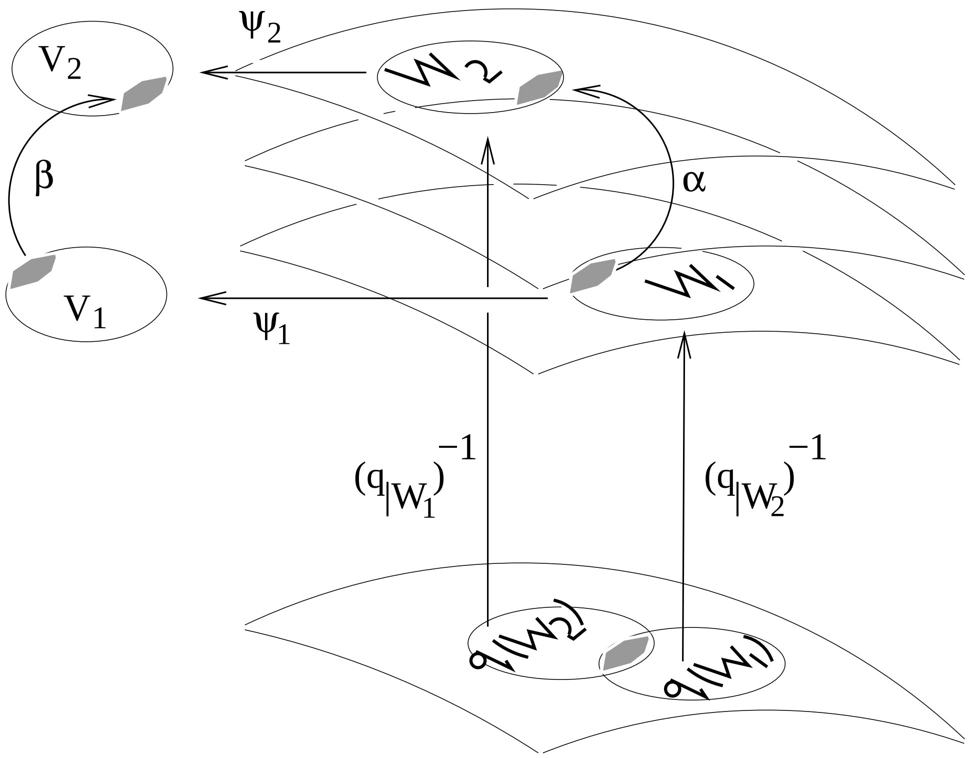


Theorem 5.36. *If M is a smooth manifold and G is a group of diffeomorphisms of M, and acts properly discontinuously on M, then the atlas just constructed is smooth.*

Proof Suppose that $W_1 \xrightarrow{q} q(W_1)$ and $W_2 \xrightarrow{q} q(W_2)$ are homeomorphisms, and that $q(W_1) \cap q(W_2) \neq \emptyset$. Write $U_1 = q(W_1), U_2 = q(W_2)$. Let $\psi_1 : W_1 \to V_1$ and $\psi_2 : W_2 \to V_2$ be charts on M. We have to show that

$$(\psi_2 \circ q^{-1}_{|W_2}) \circ (\psi_1 \circ q^{-1}_{|W_1})^{-1}$$

is smooth on its domain of definition, $\psi_1(q_{|W_1}^{-1}(U_1 \cap U_2))$. In the following diagram, this map is marked β.



We begin by showing that

$$q_{|W_2}^{-1} \circ (q_| W_1) : q_{|W_1}^{-1}(U_1 \cap U_2) \to q_{W_2}^{-1}(U_1 \cap U_2)$$

is smooth. This is the curved arrow marked α in the diagram. Suppose $x_1 \in q_{|W_1}^{-1}(U_1 \cap U_2)$, and let x_2 be its image under $q_{W_2}^{-1} \circ q_{|W_1}$. As $[x_1] = [x_2]$ in M/G, there is a $\varphi \in G$ such that $\varphi(x_1) = x_2$. *I claim that in some neighbourhood of x_1, $q_{|W_2}^{-1} \circ q_{|W_1}$ coincides with φ.* From this it follows at once that $q_{|W_2}^{-1} \circ q_{|W_1}$ is smooth at x_1, which is all we need to prove.
In fact the claim is easy to prove. The required neighbourhood of x_1 is just $\varphi^{-1}(W_2) \cap W_1$. For if $x' \in \varphi^{-1}(W_2) \cap W_1$, then both $\varphi(x')$ and $q_{|W_2}^{-1} \circ q_{W_1}(x')$ lie in W_2, and both have the same image under q. As q is 1-1 on W_2, they must coincide.

Now let $\psi_1 \to V_1 \subset \mathbb{R}^m$ and $\psi_2 : W_2 \to V_2 \subset \mathbb{R}^m$ be charts on M. Then the the fact that the map α, $q_{|W_2}^{-1} \circ q_{W_1}$, is smooth means nothing other than that the map β is smooth: for

$$\beta = \psi_2 \circ \alpha \circ \psi_1^{-1},$$

and by the definition of smoothness of a map between manifolds (Definition 5.1(iv)), α is smooth if and only if β is. $\square$

5.5 Submanifolds

Now we make a definition which places our earlier notion of manifold in a proper relation to the new notion. We begin by using a smooth structure on a smooth manifold to define a notion of smoothness for maps defined on arbitrary subsets of the manifold. This might seem familiar.

Definition 5.37. *(i) Let N^n be a smooth manifold, and let X be a subset of N. A map $f : X \to \mathbb{R}^p$ is smooth (newer) if for each $x \in X$ there is a neighbourhood U of x in N and a map $F : U \to \mathbb{R}^p$, smooth with respect to the smoothness criterion on N, such that on $U \cap X$, f and F coincide*[12].

(ii) If M and N are smooth manifolds and $X \subset M, Y \subset N$, then a map $f : X \to Y$ is a diffeomorphism if it is smooth (newer) and has a smooth (newer) inverse.

(iii) Let N be a smooth manifold. A subset $X \subset N$ is a smooth submanifold of N of dimension m if for all $x \in M$ there is a neighbourhood U of x in M, an open set $V \subset \mathbb{R}^m$, and a diffeomorphism (newer) $\phi : U \to V$. is called a chart on M around x.

Of course, this is just a repeat of definitions 1.3 and 1.6. And what we now see is that the smooth manifolds we defined in Section 1 are in fact *smooth submanifolds of* $\mathbb{R}^n$.

5.6 Embedding manifolds in Euclidean space

Although we have gone to some trouble to give a definition of manifold without an ambient space, in fact every smooth manifold is diffeomorphic to a submanifold of some Euclidean space $\mathbb{R}^N$. We will give a proof only for compact manifolds. The proof is not really difficult, though it can seem fussy and complicated. The central idea is that each chart $\phi : U \to V \subset \mathbb{R}^m$ embeds at least part of M into $\mathbb{R}^m$. If M has a finite atlas $\{\phi_j : U_j \to V_j : j = 1, \ldots, N\}$, we try to piece these embeddings together to get an embedding into $V_1 \times \cdots \times V_N$. The first problem is that each chart ϕ_j is defined only on a subset of M. To construct our embedding of M, we have to extend the domain of each chart to all of M. This is done using *bump functions*. A bump function on a manifold M is a smooth, non-negative function which takes the value 1 on a prescribed region and the value 0 on another. A typical application of such a function is as follows: suppose we are given a function f defined on some open set $U \subset M$, and taking values in $\mathbb{R}$. We wish to extend it to a function defined on all of M. As posed, the problem may be insoluble, since $f(x)$ may tend to infinity as x approaches the boundary of U. However, if we are allowed to alter f in the complement of some closed set $W \subset U$, then we can use a bump function, $\rho : M \to \mathbb{R}$, satisfying

$$\rho(x) = \begin{cases} 1 & \text{if } x \in W \\ 0 & \text{if } x \notin U \end{cases} .$$

[12] We will retain the adjective 'newer' for at least one paragraph.

The function ρf coincides with f in W, and $(\rho f)(x)$ now tends to 0 as x approaches the boundary of U. Thus we can extend it to a smooth function $\tilde{f} : M \to \mathbb{R}$ defined by

$$\tilde{f}(x) = \begin{cases} (\rho f)(x) & \text{if } x \in W \\ 0 & \text{if } x \notin U \end{cases} . \tag{5.6}$$

We will use bump functions in very much this way, to extend charts on M so that they are defined on all of M. We will need only a very simple kind of bump function:

Lemma 5.38. *Let $B(a,R)$ be a ball in $\mathbb{R}^m$, with and let $0 < \delta$. There exists a smooth function $\rho : \mathbb{R}^m \to [0,1]$ such that*

$$\rho(x) = \begin{cases} 1 & \text{if } x \in B(0,R) \\ 0 & \text{if } x \notin B(0,R+\delta) \end{cases}$$

In order not to interrupt the discussion, I will leave some details to Exercises V, and indicate here only the outline of a

Proof

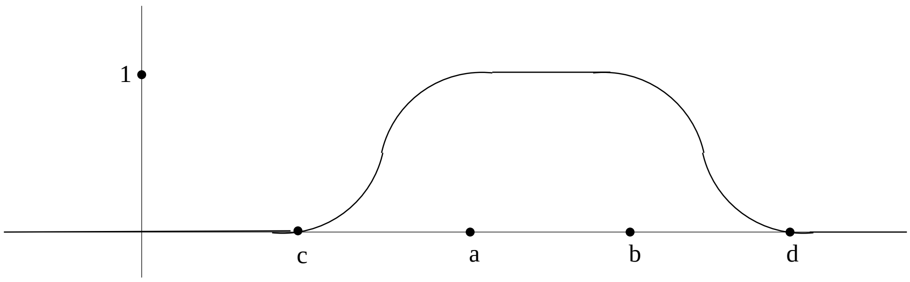
The function

$$f(x) = \begin{cases} e^{-1/x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is everywhere smooth, and all its derivatives vanish at 0. Thanks to this last property, we can use it to"glue together" arbitrary smooth functions. For example, we can glue together the function f, on the positive half of the real line, with the function 0 on the negative half, to get the smooth function

$$g(x) = \begin{cases} e^{-1/x^2} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

With a bit of ingenuity we can use this idea to produce, for any interval $[a,b]$ and any $\delta > 0$ a smooth bump function $\rho : \mathbb{R} \to [0,1]$ taking the value 1 in $[a,b]$ and the value 0 outside $(a-\delta, b+\delta)$.



To prove the lemma, take a bump function $\sigma : \mathbb{R} \to [0,1]$ such that $\sigma(t) = 1$ for $t \leq R^2$, $\sigma(t) = 0$ for $t \geq (R+\delta)^2$. Define $\rho : \mathbb{R}^m \to [0,1]$ by $\rho(x) = \sigma(\|x-a\|^2)$. □

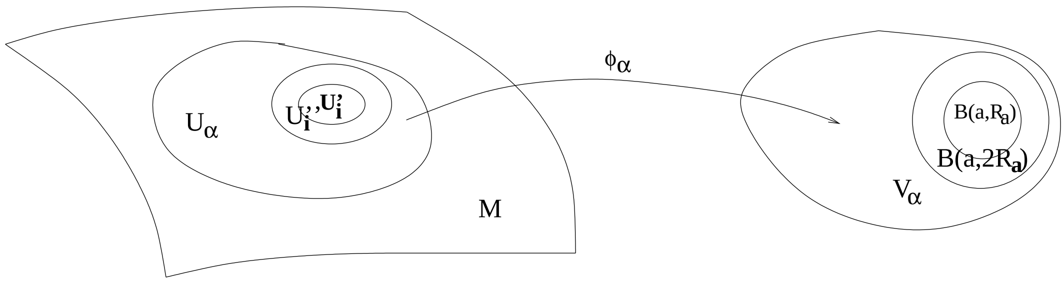
Theorem 5.39. *Suppose that M is a smooth compact manifold. Then for some T there exists an embedding f of M into $\mathbb{R}^T$.*

Proof We approach the conclusion by iterated approximation.
Step 1 Because we want to use the elementary bump functions constructed in the lemma, we first construct an atlas whose domains are all diffeomorphic to Euclidean balls. Let $\{\phi_\alpha : U_\alpha \to V_\alpha\}$ be a smooth atlas on M. For each $a \in V_\alpha$ there exists $R_a > 0$ such that

$$B(a, 2R_a) \subset V_\alpha. \tag{5.7}$$

The open balls $B(a, R_a)$ cover V_α, so their preimages $\phi_j^{-1}(B(a, R_a))$ cover U_α.

The collection of all of these balls, for all α, covers M, so as M is compact there is a finite subcover. Denote the members of this subcover by $U_1', \ldots, U_N'$, and for each i let $U_i'' = \phi_\alpha^{-1}(B(a, 2R_a))$ for the appropriate chart ϕ_α. Denote by ϕ_j the restriction to U_j'', and to U_j', of ϕ_α. Then $\{\phi_j : U_j' \to B(a_j, R_j)\}$ is an atlas, equivalent to the original atlas.



Step 2 Now we extend its charts to all of M. Choose $0 < \delta < R_a$ and define a bump function $\sigma_j : \mathbb{R}^m \to [0, 1]$ such that $\sigma_j(x) = 1$ for $x \in B(a, R_a)$ and $\sigma_j(x) = 0$ for $x \notin B(a, R_a + \delta)$. Write $\rho_j = \sigma_j \circ \phi_\alpha$. Then

$$\rho_j(x) = \begin{cases} 1 & \text{if } x \in U_j' \\ 0 & \text{if } x \text{ is outside some closed neighbourhood of } U_j' \text{ contained in } U_j'' \end{cases}$$

The map $\rho_j \phi_j$ coincides with $\phi_j = \phi_\alpha$ on U_j', and is identically 0 outside some closed neighbourhood of U_j' contained in U_j''. It can thus be extended smoothly to all of M, by giving it the value 0 outside U_j''.

Step 3: We define a smooth map $f : M \to \mathbb{R}^{Nm}$ by

$$f(x) = \big(\rho_1(x)\phi_1(x), \ldots, \rho_N(x)\phi_N(x)\big). \tag{5.8}$$

This is very nearly good enough. It is an immersion, because $\rho_j \phi_j = \phi_j$ on U_j', ϕ_j is an immersion (indeed a diffeomorphism), and the U_j' cover M. Now we have to ensure that it is injective.

Clearly $\rho_j \phi_j$ is injective on U_j', and so f is injective on U_j'. We have therefore to devise a way of distinguishing, with our map, between points in U_j' and points not in U_j'. Keeping a record of the values of the $\rho_j, j = 1, \ldots, N$ is almost good enough: redefine f by

$$f(x) = (\rho_1\phi_1(x), \ldots, \rho_N\phi_N(x), \rho_1(x), \ldots, \rho_N(x)).$$

With only a tiny modification to our constructions, this map is 1-1. The point is to ensure that if the last N components of f agree on x_1 and x_2 then x_1 and x_2 must both lie in the same U_j'. So modify ρ_j so that it takes the value 1 *only* on U_j'. Because ϕ_α transforms $U_j' \subset U_j''$ into a pair of concentric balls, this can indeed be arranged. Now, as the U_j' cover M, x_1 must lie in some U_j'. So $\rho_j(x_1) = 1$. If $f(x_1) = f(x_2)$ then we must have $\rho_j(x_2) = 1$ also, so $x_2 \in U_j'$, and now it follows from the injectivity of $\rho_j\phi_j$ on U_j' that $x_1 = x_2$.

To spare the notation, write $T = mN + N$.

Step 4: As M is compact and f is 1-1 and continuous, it is a homeomorphism onto its image. I claim that its image is a manifold and its inverse is smooth. Suppose $y \in f(M)$ is the image of $x \in U_j'$. Then $V_j' := f(U_j')$ is a neighbourhood of y in $f(M)$. By 1.48, applied to $f \circ \phi_j^{-1}$, V_j' is a submanifold of $\mathbb{R}^T$, and $f \circ \phi_j^{-1} : \phi_j(U_j) \to U_j'$ is a diffeomorphism. This shows that $f(M)$ is a submanifold of $\mathbb{R}^T$. Moreover, on V_j', $f^{-1} = \phi_j^{-1} \circ (f \circ \phi_j^{-1})^{-1}$, and so is smooth. Since the V_j' cover $f(M)$, this completes the proof. $\square$

For later use, we now give a rather sophisticated consequence of the existence of elementary bump functions.

Theorem 5.40. *Let M be a smooth manifold, and let $\{U_\alpha\}$ be an open cover of M. There exists a collection of smooth, non-negative functions ρ_α on M such that*
(i) for each α, the support of ρ_a is contained in U_α

(ii) For all $x \in M$, $\sum_\alpha \rho_\alpha(x) = 1$

(iii) Local finiteness: every $x \in M$ has a neighbourhood U such that for all except a finite number of α, $\rho_\alpha \equiv 0$ on U.

The collection $\{\rho_\alpha\}$ is called a *partition of unity subordinate to the open cover* $\{U_\alpha\}$.

We will not prove the theorem here. Proofs can be found in e.g. Guillemin and Pollack, *Differential Topology*, and Madsen and Tornehave, *From Calculus to Cohomology*

5.7 Further Structure on Abstract Manifolds

Let M^m and N^n be smooth manifolds, let $f : M \to N$ be a smooth map, let $x \in M$ and let $y = f(x) \in N$. Let ϕ be a chart on M around x, and let ψ be a chart on N around y, with $\phi(x) = w, \psi(y) = z$. Since the composite $\psi \circ f \circ \phi^{-1}$ is a smooth (old) map, we can consider its derivative

$$d_w(\psi \circ f \circ \phi^{-1}) : \mathbb{R}^m \to \mathbb{R}^n. \tag{5.9}$$

If we choose different charts ϕ' and ψ' around x and y, we get a different derivative, but one that is related to the first in a simple way: there is a commutative diagram

$$\begin{tikzcd}
\mathbb{R}^m \arrow[r, "d_w(\psi\circ f\circ\phi^{-1})"] \arrow[d, "d_w(\phi'\circ\phi^{-1})"'] & \mathbb{R}^n \arrow[d, "d_z(\psi'\circ\psi^{-1})"] \\
\mathbb{R}^m \arrow[r, "d_{w'}(\psi'\circ f\circ\phi'^{-1})"] & \mathbb{R}^n
\end{tikzcd}$$

The vertical maps are isomorphisms, so the ranks of the top and bottom horizontal maps are equal. Thus, we can unambiguously define f to be a submersion, immersion or local diffeomorphism at x according to whether the derivative (5.9) is surjective, injective or an isomorphism.

The proofs of the Inverse Function Theorem for manifolds, 1.19, and of the Local Normal Forms for submersions and immersions, 1.27 and 1.26, go through unchanged, and their consequences — 1.29, 1.31, 1.40 and 1.48, follow. With a bit of ingenuity, we can even define transversality by means of charts, and recover 1.52 for maps transverse to submanifolds of abstract manifolds.

It seems curious that we can refer to what is really a property of the derivative — its surjectivity, injectivity or isomorphiusm — before we even have a notion of tangent space T_xM, let alone of the derivative d_xf.

So do we really need a notion of tangent vector and tangent space for abstract manifolds? After all, we managed to prove a significant theorem, that a compact abstract manifold is diffeomorphic to a submanifold of Euclidean space, without needing tangent spaces.

However, we will need them. Several different approaches are possible. Here I will sketch two.

5.7.1 A first approach to tangent spaces

Let us be guided by the principle that what is important about a mathematical object is not what it is, but what it does. When we studied smooth submanifolds of $\mathbb{R}^n$, their tangent spaces were made up of vectors which already *existed* in the ambient space. What did they *do*? They differentiated functions — that is, we could differentiate functions along them.

Let M be a submanifold of $\mathbb{R}^N$, $x \in M$ and $v \in T_xM$. If f is a smooth function defined on some neighbourhood of the point x, we can measure the derivative of f along the vector v, $d_xf(v)$. Because now we are thinking of v as the subject of the action we are trying to describe, and f as its object, we switch their roles in the notation and rewrite $d_xf(v)$ as $v \cdot f$. The sum and product rules for differentiation lead easily to the following properties:

1. $v \cdot (f + g) = v \cdot f + v \cdot g$
2. $v \cdot (\lambda f) = \lambda\, v \cdot f$
3. $v \cdot (fg) = f(x)\, v \cdot g + g(x)\, v \cdot f$.

The first two of these properties express the fact that $v\cdot$ is an $\mathbb{R}$-linear operator

$$C^\infty(M)_x \to \mathbb{R},$$

where $C^\infty(M)_x$ is the set of smooth functions defined on some neighbourhood of x. The third property is known as the Leibniz property. Any map $C^\infty_x(M) \to \mathbb{R}$ with these three properties is called a *derivation*. We have seen that every $v \in T_xM$ determines a derivation $v\cdot$. In fact the converse is true: *every* derivation arises in this way. For a moment, denote the set of derivations $C^\infty_x(M) \to \mathbb{R}$ by D_xM.

Proposition 5.41. *Let M^m be a submanifold of $\mathbb{R}^N$.*
(i) D_xM is a real vector space of dimension m.
(ii) The map $T_xM \to D_xM$ sending $v \in T_xM$ to $v\cdot$ is an isomorphism.

Proof (i) First consider the case $M = \mathbb{R}^m$, and let $\ell \in D_aM$. By the Leibniz property, we have $\ell(1) = \ell(1 \times 1) = 1\ell(1) + 1\ell(1) = 2\ell(1)$ (where inside the brackets 1 is the constant function with value 1). Hence $\ell(1) = 0$. By the linearity property (2), it follows that if h is any constant function then $\ell(h) = 0$.

I claim that the values $\ell(x_j)$, for $j = 1, \dots, m$, completely determine ℓ. To see this, we need a minor result, the so-called Hadamard's lemma, proved in Exercises IV:

Lemma 5.42. *Every function $f \in C^\infty_a(\mathbb{R}^m)$ can be written in the form*

$$f(x) = f(a) + \sum_j (x_j - a_j) g_j(x) \tag{5.10}$$

for some functions $g_j \in C^\infty_a(\mathbb{R}^m)$.

In fact this is really a form of Taylor's Theorem. In particular, by differentiating with respect to x_j and taking $x = a$, we see that $g_j(a) = \partial f/\partial x_j(a)$.

Let $f \in C^\infty_a(\mathbb{R}^m)$. By applying the lemma twice, once to f and then to the g_j in the resulting equation (5.10), we get the expression

$$f(x) = f(a) + \sum_j g_j(a)(x_j - a_j) + \sum_{i,j} (x_i - a_i)(x_j - a_j) g_{ij}(x). \tag{5.11}$$

The functions $(x_1 - a_1), \dots, (x_m - a_m)$ all vanish at a, so by the Leibniz property, $\ell((x_i - a_i)(x_j - a_j)) = 0$. Indeed, so also is $\ell((x_i - a_i)(x_j - a_j)g_{ij})$ equal to 0. Since $\ell(f(a)) = 0$, from (5.11) we get

$$\ell(f) = \sum_j g_j(a)\ell(x_j - a_j) = \sum_j \partial f/\partial x_j(a)\ell(x_j). \tag{5.12}$$

Now if $v = (\ell(x_1), \ldots, \ell(x_m)) \in \mathbb{R}^m = T_a\mathbb{R}^m$, then

$$v \cdot f = d_a f(v) = \sum_j \ell(x_j)\partial f/\partial x_j(a) = \ell(f). \tag{5.13}$$

This proves (ii) for $\mathbb{R}^m$, and it follows for arbitrary smooth manifold M by taking charts.

Since D_aM is clearly a vector space, and the map $T_xM \to D_xM$ is a linear injection, (i) follows also. $\square$

The identification of T_xM with D_xM is so natural that D_xM is best regarded simply as an alternative way of viewing T_xM. Its advantage is that it is defined as soon as we have a notion of smooth function. *We do not need an ambient space.*

Definition 5.43. *(i) Let M be a smooth manifold. We define T_xM to be D_xM. A tangent vector to M at x is an element of D_xM.*

(ii) If $f : M \to N$ is a smooth map of smooth manifolds, then $d_xf : T_xM = D_xM \to D_{f(x)}N = T_{f(x)}N$ is defined as follows: for each $g \in C^\infty_{f(x)}N$, $g \circ f \in C^\infty_x(M)$. Given $v \in D_xM$, define $d_xf(v) \in D_{f(x)}N$ by

$$d_xf(v) \cdot g = v \cdot (g \circ f).$$

In fact this way of thinking of tangent vectors is rather fruitful. It leads us immediately to the possibility of viewing a vector field v as a rule for transforming each function f into a new function $v \cdot f$, whose value at x is $d_xf(v(x))$. Viewing v as an operator

$$C^\infty(M) \to C^\infty(M)$$

in this way, it satisfies

1. $v \cdot (f + g) = v \cdot f + v \cdot g$
2. $v \cdot (\lambda f) = \lambda v \cdot f$
3. $v \cdot (fg) = f\, v \cdot g + g\, v \cdot f.$

This in turn leads to the crucially important notions of the *Lie bracket* of two vector fields, and the *Lie algebra* of a Lie group. These are explored in Exercises V.

5.7.2 A second approach to tangent spaces

For a physicist, there is no reality beyond the measurements (for example of the positions and velocities of a number of objects) he or she takes. The measurements taken by another observer will be different, but one set of measurements can be transformed into the other by agreed rules of transformation. Each set of measurements might be thought of as a chart, but the physicist has no psychological need for an absolute object of which the charts are partial views. The only important thing is to know how to transform one set of measurements into another.

The parallels with the notion of manifold, and of smooth atlas, are clear. They give rise to a viewpoint on the notion of manifold in which a manifold *is* its charts (and their transformation rules). The reality of the manifold is expressed in its local co-ordinates (i.e. via its charts). Indeed, we had no difficulty speaking of immersions and submersions of abstract manifolds, by means of charts. We did not need a notion of tangent space independent of the charts.

This view encourages the definition of the tangent space T_xM as being $T_{\phi(x)}\mathbb{R}^m$, for any chart ϕ around x, subject to the identification that if $v_1 \in T_{\phi_1(x)}\mathbb{R}^m$ and $v_2 \in T_{\phi_2(x)}\mathbb{R}^m$, then v_1 and v_2 are "the same" if $d_{\phi_1(x)}(\phi_2 \circ \phi_1^{-1})(v_1) = v_2$. This is made precise with the following slightly ponderous definition:

Definition 5.44. *(i) Let M be a manifold with smooth atlas $\mathscr{A}$. Then T_xM is the quotient of the disjoint union $\coprod_{\phi_\alpha} T_{\phi_\alpha(x)}\mathbb{R}^m$ (over all $\phi_\alpha \in \mathscr{A}$ in whose domain x lies), by the equivalence relation*

$$v_1 \in T_{\phi_\alpha(x)}\mathbb{R}^m \sim v_2 \in T_{\phi_\beta(x)}\mathbb{R}^m \quad \textit{if} \quad d_{\phi_\alpha(x)}(\phi_\beta \circ \phi_\alpha^{-1})(v_1) = v_2.$$

(ii) If $f : M^m \to N^n$ is a smooth map of smooth manifolds, then $d_xf : T_xM \to T_{f(x)}N$ is defined as follows: T_xM is identified with $T_{\phi(x)}\mathbb{R}^m$, and $T_{f(x}N$ is identified with $T_{\psi(f(x))}\mathbb{R}^n$, where ϕ and ψ are charts on M and N around x and $f(x)$ respectively. Then d_xf is identified with $d_{\phi(x)}(\psi \circ f \circ \phi^{-1})$.

Thus, the familiar diagram

$$\begin{tikzcd}
T_xM \arrow[r, "d_xf"] \arrow[d, "d_x\phi"'] & T_{f(x)}N \arrow[d, "d_{f(x)}\psi"] \\
\mathbb{R}^m \arrow[r, "d_{\phi(x)}(\psi\circ f\circ\phi^{-1})"] & \mathbb{R}^n
\end{tikzcd}$$

reappears, this time with the vertical maps as *identifications* rather than isomorphisms.

5.7.3 A third approach to tangent spaces

In Section 1 we defined tangent vectors to M as the derivatives of curves in M. This led to an especially convenient description of the derivative of a map $f : M \to N$:

$$d_xf(\gamma'(0)) = (f \circ \gamma'(0)).$$

This motivates a third view of T_xM. Each smooth parametrised curve $\gamma : (\mathbb{R}, 0) \to (M, x)$ gives us a way of differentiating functions $f : (M, x) \to (\mathbb{R}, 0)$, by associating to f the derivative $(f \circ \gamma)'(0)$. Say that two curve-germs $\gamma_1 : (\mathbb{R}, 0) \to (M, x)$ and $\gamma_2 : (\mathbb{R}, 0) \to (M, x)$ are *equivalent* if for all smooth functions f,

$$(f \circ \gamma_1)'(0) = (f \circ \gamma_2)'(0). \tag{5.14}$$

Proposition 5.45. *(i)* $\gamma_1 \sim \gamma_2$ *if and only if for some chart* ϕ *on* M *around* x*,* $(\phi\circ\gamma_1)'(0) = (\phi\circ\gamma_2)'(0)$*.* $\square$

When M is a smooth submanifold of a Euclidean space $\mathbb{R}^N$, we can identify an equivalence class of curve germs $\gamma : (\mathbb{R}, 0) \to (M, x)$ with their common tangent vector $\gamma'(0) \in \mathbb{R}^N$. When M is an abstract smooth manifold, we take such an equivalence class *as our definition* of tangent vector:

Definition 5.46. *(i)* T_xM *is the set of equivalence classes of smooth curve germs* $\gamma : (\mathbb{R}, 0) \to (M, x)$ *by the equivalence relation of (5.14).*
(ii) If $f : M \to N$ *is a smooth map of smooth manifolds, then* $d_xf : T_xM \to T_{f(x)}N$ *is defined by the formula*

$$d_xf([\gamma]) = [f \circ \gamma].$$

The map $d_x\phi : T_xM \to T_{\phi(x)}\mathbb{R}^m$ is a bijection, and its target is naturally identified with $\mathbb{R}^m$. Using $d_x\phi$ we can give T_xM the structure of an m-dimensional vector space. This structure is independent of the choice of chart ϕ, by the smoothness of each crossover maps and the consequent linearity of its derivative.

5.8 You don't have to choose

All three views of T_xM are equivalent. Each has its advantages in certain contexts. The approaches we have used stem from the three approaches to the derivative we described in Section 1.

6 Differential Forms and Integration on Manifolds

6.1 First examples

In this chapter we begin to study analysis on manifolds. There are two reasons for doing this. The first is that manifolds, such as the universe, are home to many objects — forces, electric and magnetic fields, flows, etc etc — which are well-modelled by the differential forms we will shortly define. The second, slightly less obvious, is that the methods of analysis, and in particular the integration of differential forms, lead to new approaches to questions we have already been thinking about, such as the degrees of smooth maps, oriented intersection numbers, etc.

We begin with some physical motivation which leads to the definition of differential forms. If the physics doesn't help you appreciate the mathematics, ignore it.

Example 6.1. In physics one often meets expressions like

$$\int_a^b F(\gamma(t)) \cdot \gamma'(t)dt \tag{6.1}$$

where F is a vector field on $\mathbb{R}^3$ and $\gamma : [a, b] \to \mathbb{R}^2$ is a parametrisation of a smooth curve C. This integral is a measurement of the amount of work done in moving an object along the trajectory γ while it is subject to the force F. In so called conservative systems, the force F is the gradient of a potential function f (e.g. gravitational potential or electrical potential); then the integral can be written

$$\int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)dt \tag{6.2}$$

But notice that the gradient vector ∇f is a way of considering the derivative $d_x f$ as a vector, whereas $d_x f$ is really a covector - something which eats vectors and spits out numbers. We can rewrite (6.2) as

$$\int_a^b (d_{\gamma(t)} f)(\gamma'(t))dt. \tag{6.3}$$

Written like this, it is clearly the integral of the derivative of $f \circ \gamma$, and thus is equal to $f(\gamma(b)) - f(\gamma(a))$. This expresses a version of the law of conservation of energy: in a conservative system, the work done is equal to the change in the value of the potential from one end of the trajectory to the other.

The derivative $d_{(x_1,x_2)} f$ has matrix

$$\left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right]$$

with respect to the basis of $\mathbb{R}^2$ dual to the coordinate functions x_1, x_2. It is useful to develop a notation which avoids having to write a matrix.

Define dx_i as the covector (i.e. linear form on $\mathbb{R}^m$) whose matrix with respect to the basis dual to the coordinates $x_1, \ldots, x_m$ is

$$\begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \qquad \text{(1 in the } i\text{-th place)}$$

More simply, dx_i is the derivative of the coordinate function x_i. The dx_i provide a useful way of representing the derivative $d_x f$ of a function f on $\mathbb{R}^n$:

$$d_x f = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n.$$

In this notation, the force $F = (F_1, F_2, F_3)$ that we considered in (6.1) can be replaced by the expression

$$F_1 dx_1 + F_2 dx_2 + F_3 dx_3,$$

which can be thought of as defining a smoothly varying family of covectors [13]. Such a family is known as a 1-form, and for some reason is often denoted by the symbol ω ("omega").

Now suppose that $\sigma : [c, d] \to \mathbb{R}^2$ is another parametrisation of the curve C of (6.1). I claim that for any 1-form ω,

$$\int_c^d \omega(\sigma'(s))ds = \pm \int_a^b \omega(\gamma'(t))dt. \tag{6.4}$$

The proof is very simple: both parametrisations of C are diffeomorphisms, so we can write

$$\gamma = \sigma \circ (\sigma^{-1} \circ \gamma).$$

Write $\sigma^{-1} \circ \gamma =: h$, so $\gamma = \sigma \circ h$. Then

$$\int_a^b \omega(\gamma'(t))dt = \int_a^b \omega\big(\sigma'(h(t))h'(t)\big)dt = \int_a^b \omega\big(\sigma'(h(t))\big)h'(t)dt$$

Now use integration by substitution; write $s = h(t)$ (as indeed it is); then $ds = h'(t)dt$, and

$$\int_a^b \omega\big(\sigma'(h(t))\big)h'(t)dt = \int_{h(a)}^{h(b)} \omega(\sigma'(s))ds = \begin{cases} \int_c^d \omega(\sigma'(s))ds & \text{if } h \text{ is increasing} \\ \int_d^c \omega(\sigma'(s))ds & \text{if } h \text{ is decreasing.} \end{cases}$$

Thus $\int_c^d \omega(\sigma'(s))ds$ is equal to $\int_a^b \omega(\gamma'(t))dt$ if h preserves orientation, and to $-\int_a^b \omega(\gamma'(t))dt$ if h reverses orientation.

This allows us to define the integral of a 1-form over a compact oriented curve in $\mathbb{R}^n$, $\int_C \omega$:

[13] I am not suggesting that force is really a covector rather than a vector, merely that in our integral this is how it is functioning

Definition 6.2. *Let C be a compact oriented curve (i.e. a 1-manifold) in $\mathbb{R}^n$ and let ω be a 1-form on $\mathbb{R}^n$. Choose an orientation-preserving diffeomorphism $\gamma : [a, b] \to C$. Then*

$$\int_C \omega := \int_a^b \omega(\gamma'(t))dt.$$

The calculation we have just made shows that the integral is independent of the choice of orientation-preserving diffeomorphism (parametrisation), and thus is well-defined.

Returning to the example in which the 1-form was the derivative of some function, $\omega = df$: suppose that C begins at P and ends at Q; then we have

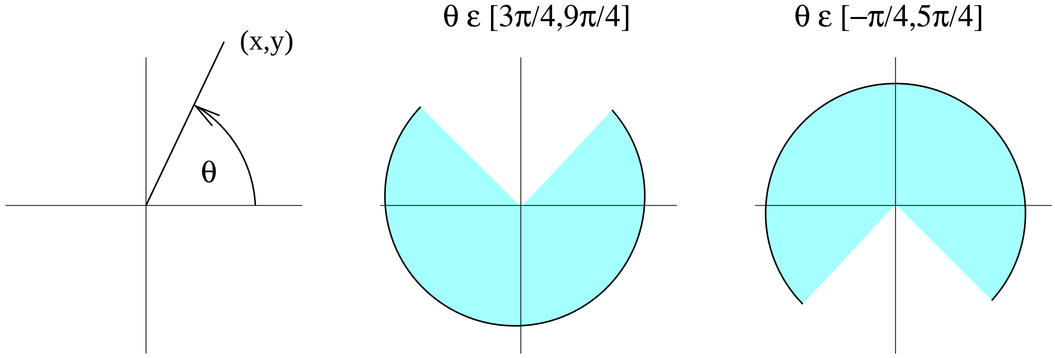
$$\int_C df = f(Q) - f(P). \tag{6.5}$$

The points P and Q together make up the boundary of C, ∂C. Thus (6.5) can be written, somewhat tendentiously, as

$$\int_C df = \int_{\partial C} f. \tag{6.6}$$

It is not clear at this point why we should write $f(Q)-f(P)$ as an integral. This will become clear later. One thing that is clear, however, is that if $P = Q$ (i.e. if C is a closed curve), or in other words, if C has no boundary, then $\int_C df = 0$. □

Example 6.3. A less trivial example of a 1-form on $\mathbb{R}^2 \setminus \{0\}$ is provided by the derivative of the polar coordinate function θ shown. The angle θ is not a well defined smooth function on all of $\mathbb{R}^2 \setminus \{0\}$. In any sector of angle less than 2π, there is a smooth determination of θ, but it is not unique, and determinations of θ in overlapping regions will in general differ.



However, any two determinations θ_1, θ_2 of θ differ by a multiple of 2π. It follows that the derivatives $d\theta_1$ and $d\theta_2$ are equal. Thus, they piece together to define a single, well-defined, smooth 1-form on $\mathbb{R}^2 \setminus \{0\}$, which we denote by $d\theta$. It is easy to calculate: if $x > 0$ then one determination of θ is as $\arctan(y/x)$; thus

$$d\theta = \frac{-y}{x^2} \times \frac{1}{1+(y/x)^2}dx + \frac{1}{x} \times \frac{1}{1+(y/x)^2}dy = \frac{-ydx + xdy}{x^2+y^2}. \tag{6.7}$$

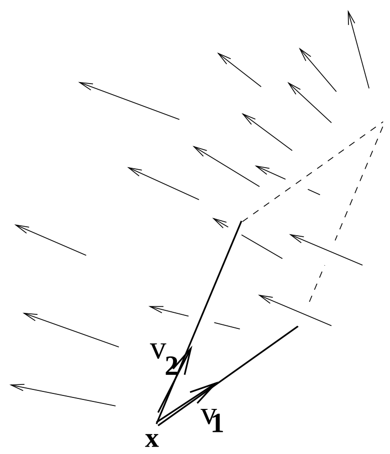
If $y > 0$ then another determination of θ is as $\text{arccotan}(x/y)$; you should check that its derivative has the same formula (6.7).

What is the significance of this 1-form? If C is an oriented curve in $\mathbb{R}^2 \setminus \{0\}$, what is the geometrical meaning of $\int_C d\theta$? If C is a curve contained in a sector in which θ is well defined, then by (6.5), $\int_C d\theta$ measures the difference between the value of θ at one end of C and the other. But what if C is not contained in such a sector? What about $\int_{S^1} d\theta$?

Exercise 6.4. *Calculate* $\int_{S^1} d\theta$. □

□

Example 6.5. Let X be the velocity field of a fluid flow in a region U of $\mathbb{R}^3$. For each point $x \in U$, each pair of vectors u, v, and each positive real ε, imagine a parallelogram P_ε based at x and spanned by εu and εv. Note that the area of P_ε, divided by ε^2, is independent of ε.



Let

$$\omega(v_1, v_2) = lim_{\varepsilon \to 0} \frac{\text{fluid flow through } P_\varepsilon \text{ per unit time}}{\varepsilon^2}. \tag{6.8}$$

We give this a sign: flow through P_ε is positive if it goes in the same direction as $v_1 \times v_2$, and negative if it goes in the opposite direction. The area of the parallelogram P_ε is $\varepsilon^2 \|u \times v\|$. The unit normal vector in the positive direction is $u \times v / \|u \times v\|$. Thus the fluid flow through P_ε is approximated by

$$X(x) \cdot \frac{u \times v}{\|u \times v\|} \text{Area of } P_\varepsilon = \varepsilon^2 X(x) \cdot u \times v.$$

As $\varepsilon \to 0$ this approximation improves, and indeed (one can check that)

$$\omega(u, v) = X(x) \cdot u \times v.$$

Let $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$; then this last expression is equal to

$$X_1(x)(u_2 v_3 - u_3 v_2) + X_2(x)(u_3 v_1 - u_1 v_3) + X_3(x)(u_1 v_2 - u_2 v_2)$$

$$= X_1(x) \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - X_2(x) \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + X_3(x) \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}. \tag{6.9}$$

For each $x \in U$, this is a bilinear form — it is linear in each of u and v – and moreover it is *alternating*: if we interchange the order of u and v, it changes sign.

Definition 6.6. *A 2-form on $\mathbb{R}^m$ is a family of bilinear alternating maps depending smoothly on $x \in \mathbb{R}^m$.*

Thus ω defined in (6.8) is a smooth 2-form on $\mathbb{R}^3$. □

Fussy remark 6.7. In (6.9), the subindices of the X_j are rather neatly ordered. Each determinant $u_i v_j - u_j v_i$ involves all of the subindices *except* one, and the subindex of its coefficient X_k is the missing index. In fact this is the most symmetric way of assigning subindices to the variables and coefficients in this formula. Taking care of this kind of symmetry can be very helpful in understanding, or even noticing, the structure of complicated calculations, and in getting them right, and it is worth while spending a little time on it.

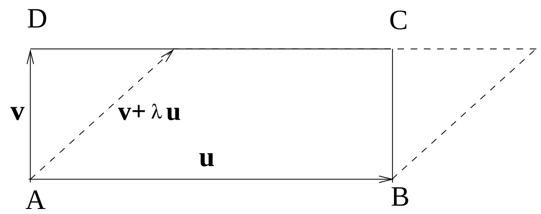
6.2 Determinants, volume and change of variable in multiple integration

Exercise 6.8. The appearance of alternating multilinear forms when we are considering area or volume is natural and inevitable. Let $P(u, v)$ be the parallelogram spanned by vectors u and v, and let $|P(u, v)|$ be its area. The formula

$$\text{Area} = \text{base} \times \text{height}$$

makes clear that if we add a multiple of u to v, the area of the parallelogram they span does not change:

$$|P(u, v + \lambda u)| = |P(u, v)|.$$

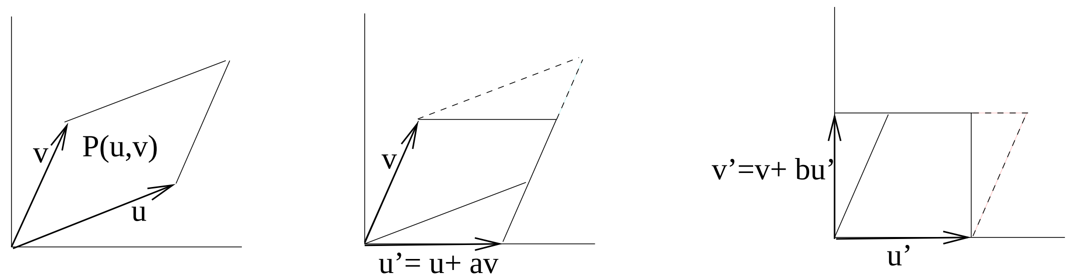


Show that if $b : V \times V \to \mathbb{R}$ is a bilinear form on the vector-space V, then the following are equivalent:

1. for all $u, v \in V$ and $\lambda \in \mathbb{R}$, $b(u + \lambda v, v) = b(u, v)$
2. for all $v \in V$, $b(v, v) = 0$
3. for all $u, v \in V$, $b(v, u) = -b(u, v)$.

□

Now consider the following sequence of pictures:



We transform the parallelogram $P(u, v)$ spanned by vectors u, v into a rectangle $P(u', v')$ with its edges along the coordinate axes. Denote the area of $P(u, v)$ by $|P(u, v)|$.

$$|P(u', v)| = |P(u, v)|$$

since $P(u', v)$ is obtained from $P(u, v)$ by subtracting from it a triangle (shown with dashed edges) and adding to it another, congruent, triangle. Similarly

$$|P(u', v')| = |P(u', v)|.$$

It is clear, in fact, that by adding to either vector a multiple of the other, we do not alter the area of the parallelogram $P(u, v)$.

In this respect, $|P(u, v)|$ behaves identically to $\det[u, v]$, where $[u, v]$ is the matrix with columns u and v. The coincidence does not end there: since, in our diagram, the two vectors u' and v' lie on the coordinate axes, it is easy to check that

$$|P(u', v')| = \det[u', v'].$$

Because the process of transforming u, v to u', v' has altered neither the area of the paralleogram, nor the value of the determinant, we deduce that

$$|P(u, v)| = \det[u, v].$$

This is true in complete generality, if we take the *absolute value* of the determinant, since interchanging u and v changes the sign of the determinant but leaves the area unchanged. In fact the corresponding statement is true in all dimensions:

Theorem 6.9. *For any vectors $u_1, \ldots, u_n \in \mathbb{R}^n$,*

$$|P(u_1, \ldots, u_n)| = \big|\det[u_1, \ldots, u_n]\big|.$$

Here we mean, by $P(u_1, \ldots, u_n)$ the *parallelipiped* spanned by $u_1, \ldots, u_n$. The theorem can easily be proved by essentially the same procedure as for the 2-dimensional case. We know that $det[u_1, \ldots, u_n]$ is unaltered by adding to one vector a multiple of one of the others, and show that $|P(u_1,\ldots,u_n)|$ has the same property. To do this without a drawing, one needs a precise definition of $P(u_1,\ldots,u_n)$ and of $P(u_1,\ldots,u_i+\lambda u_j,\ldots,u_n)$; equipped with such a precise definition, it is easy to show that the regions

$$P(u_1,\ldots,u_n)\setminus P(u_1,\ldots,u_i+\lambda u_j,\ldots,u_n)$$

and

$$P(u_1,\ldots,u_i+\lambda u_j,\ldots,u_n)\setminus P(u_1,\ldots,u_n)$$

are congruent (by a translation, in fact, just as in the pictures above, where these two regions were the triangles that are lost and gained in each of the two transformations). From this it follows that $P(u_1,\ldots,u_n)$ has the same volume as $P(u_1,\ldots,u_i+\lambda u_j,\ldots,u_n)$. By operations of this type it is possible to transform $u_1,\ldots,u_n$ into vectors $u_1',\ldots,u_n'$ lying along the coordinate axes. Then

$$|P(u_1,\ldots,u_n)| = |P(u_1',\ldots,u_n')| = \big|\det[u_1',\ldots,u_n']\big| = \big|\det[u_1,\ldots,u_n]\big|.$$

Further details of the steps in this argument are given in Exercises V. $\square$

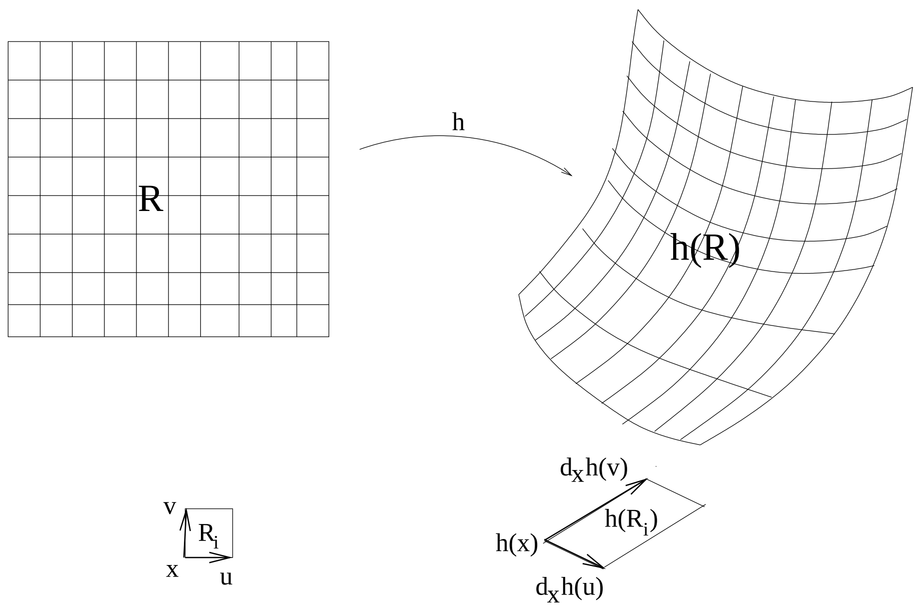
Theorem 6.9 is the key to the proof of the change-of-variable formula in multiple integration. We will need this formula when we prove a higher dimensional analogue of (6.4).

Theorem 6.10. *Suppose that R is a region in $\mathbb{R}^m$, $h : R \to h(R) \subset \mathbb{R}^m$ is a diffeomorphism, and $f : R \to \mathbb{R}$ is an integrable function. Then*

$$\int_{h(R)} f = \int_R (f\circ h)|J(h)|$$

where $J(h) = \det[dh]$.

Sketch proof.



The diagram shows a partition of R (drawn as a rectangle, though this is not necessary) into rectangles R_i. For any function g on R, the integral $\int_R g$ is defined by means of such partitions. If the supremum of g on R_i is achieved at x_i,

$$\sum_i g(x_i)\mathrm{vol}(R_i) \tag{6.10}$$

is an 'approximation from above' to the value of $\int_R g$; the integral itself is the infimum, over all partitions, of these approximations from above. (Of course this is not how one *calculates* it, in general.) Now consider the function $g = f \circ h$. Given any partition of R, by means of the diffeomorphism h we obtain a partition of $h(R)$ (though not into rectangles). Suppose that on R_i, $f \circ h$ achieves its supremum at x_i. Let $y_i = h(x_i)$. Then on $h(R_i)$, f achieves its supremum at y_i. Consider the sum

$$\sum_i f(y_i)\mathrm{vol}(h(R_i)) \tag{6.11}$$

where $y_i = h(x_i)$. This is an approximation from above of $\int_{h(R)} f$. Suppose $R_i = x_i + P(v_1, \ldots, v_m)$. If R_i is very small,

$$h(R_i) \overset{\text{approx}}{=} y_i + P\big(d_{x_i}h(v_1), \ldots, d_{x_i}h(v_m)\big). \tag{6.12}$$

The parallelipiped on the right has volume $|\det[d_{x_i}h(v_1), \ldots, d_{x_i}h(v_m)]|$. If $v_i = \lambda_i e_i$, then

$$\big|\det\big[d_{x_i}h(v_1), \ldots, d_{x_i}h(v_m)\big]\big| = \big|\det\big[\lambda_1 d_{x_i}h(e_1), \ldots, \lambda_m d_{x_i}h(e_m)\big]\big| =$$

$$= \lambda_1 \cdots \lambda_m \big|\det\big[\frac{\partial h}{\partial x_1}, \ldots, \frac{\partial h}{\partial x_m}\big]\big| = \mathrm{vol}(R_i)|\det[d_x h]|.$$

Therefore

$$\sum_i f(y_i)\mathrm{vol}(h(R_i)) \overset{\text{approx}}{=} \sum_i f(h(x_i))\mathrm{vol}(R_i)\,|\det[d_{x_i}h]\,|. \tag{6.13}$$

When we take the infimum over all choices of partitions $\{R_i\}$ of R, on the left hand side we get $\int_{h(R)} f$; on the right hand side we get $\int_R (f \circ h)|\det[df]|$. The infimum of the set of approximations from above is approached by taking finer and finer partitions, and as the partition gets finer, the approximation (6.13) becomes closer. In the limit, the two infima coincide. □

Exercise 6.11. *Suppose that $A : (\mathbb{R}^m)^k \to \mathbb{R}$ is k-linear and alternating. Show that A shares the following properties with* det*:*

(i) If $v_i = v_j$ for some $i \neq j$ then $A(v_1, \ldots, v_k) = 0$. Ditto if $v_i = \lambda v_j$.

(ii) $A(v_1, \ldots, v_i + \lambda v_j, \ldots, v_k) = A(v_1, \ldots, v_i, \ldots, v_k)$. (Hint: use linearity in the i-th argument to write the left hand side as a sum of two terms.)

(iii) For any permutation σ of k objects, $A(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = sign(\sigma)A(v_1, \ldots, v_k)$ (Hint: write σ as the composite of a sequence of transpositions.) □

Exercise 6.12. *Suppose that $A : (\mathbb{R}^m)^m \to \mathbb{R}$ is m-linear and alternating. Show that A is a scalar multiple of* det. Hint: The properties listed in the previous exercise enable you to reduce any m-tuple of vectors $v_1, \ldots, v_m$ of $\mathbb{R}^m$ to something like row echelon form without altering the value of $A(v_1, \ldots, v_m)$. This is essentially what we did to prove Theorem 6.9, although that proof used only property (ii) of 6.11, and moreover contained, as a last step, the calculation that the scalar is ± 1. □

6.3 Differential forms

We are ready for the definition of differential forms.

Definition 6.13. *A differential k-form on $\mathbb{R}^m$ is a smooth family of alternating k-linear maps $\mathbb{R}^m \times \cdots \times \mathbb{R}^m \to \mathbb{R}$. Smoothness here means the following: a k-form ω on $U \subset \mathbb{R}^m$ is smooth if, whenever $v_1, \ldots, v_k$ are smooth vector fields on U, then the function $\omega(v_1, \ldots, v_k)$ is smooth.*

The set of all k-forms defined on an open set $U \subset \mathbb{R}^m$ is denoted $\Omega^k(U)$. It is useful (and consistent) to regard functions on U as 0-forms, and thus denote the set of all smooth functions on U as $\Omega^0(U)$.

Remark 6.14. There is a minor issue of notation here. Suppose that U is an open set in $\mathbb{R}^m$ and $\omega \in \Omega^k(U)$. For each point $x \in U$, we get a k-linear alternating form $(\mathbb{R}^m)^k \to \mathbb{R}$. We could in priniciple denote it by $\omega(x)$, since it depends on x. Since it is to be evaluated on k-tuples of vectors $v_1, \ldots, v_k$ in $\mathbb{R}^m$, we would then have to write

$$\omega(x)(v_1, \ldots, v_k).$$

The universally observed convention is to drop reference to x, and simply write $\omega(v_1, \ldots, v_k)$. This is slightly uncomfortable when first encountered. Later, we will work with differential forms on manifolds. If $\omega \in \Omega^k(M)$, then for each $x \in M$ we get a k-linear alternating map $(T_xM)^k \to \mathbb{R}$. Here, the tangent space comes with the label x, and so the uncomfortable imprecision is removed. But in practice it will not be a problem even when we are working with k-forms on open sets in Euclidean spaces.

We have already seen what is in some sense the most important example, the determinant, which is a (constant) m-form on $\mathbb{R}^m$ — constant in the sense that it doesn't vary with x.

Lemma 6.15. *If $U \subset \mathbb{R}^m$, then every $\omega \in \Omega^m(U)$ is of the form $\omega = L\det$, where L is a smooth function on U.*

Proof This is just Exercise 6.12 again. At each point, ω is a scalar multiple of det. The form is smooth if and only if the scalar varies smoothly with x. Thus, $\omega = L\det$ for some smooth function L. □

We introduce now some notation which makes representing forms simpler.

Definition 6.16. *Suppose that ω_1 and ω_2 are two 1-forms. Define a 2-form $\omega_1 \wedge \omega_2$ (this reads "ω_1 wedge ω_2") by the formula*

$$\omega_1 \wedge \omega_2(u, v) = \omega_1(u)\omega_2(v) - \omega_2(u)\omega_1(v). \tag{6.14}$$

It is very easy to check that $\omega_1 \wedge \omega_2$ is indeed alternating and bilinear, and varies smoothly with x provided ω_1 and ω_2 do so. For example,

$$dx_1 \wedge dx_2\,(u, v) = dx_1(u)dx_2(v) - dx_2(u)dx_1(v) = u_1v_2 - u_2v_1.$$

And (6.9) can be rewritten as

$$(X_1\, dx_2 \wedge dx_3 + X_2\, dx_3 \wedge dx_1 + X_3\, dx_1 \wedge dx_2)(u, v).$$

As another example, observe that if $e_1, \ldots, e_m$ form a basis for the m-dimensional vector space V, and if $\ell_1, \ldots, \ell_m$ is the dual basis for the dual vector space V^* (so that $\ell_j(e_i) = \delta_{ij}$), then for $i \neq j$,

$$\begin{aligned} \ell_i \wedge \ell_j(e_i, e_j) &= 1 \\ \ell_i \wedge \ell_j(e_j, e_i) &= -1 \\ \ell_i \wedge \ell_j(e_s, e_t) &= 0 \text{ if } \{i, j\} \neq \{s, t\} \end{aligned}$$

.

Lemma 6.17. *(i) Let $e_1, \ldots, e_m$ be a basis for the m-dimensional vector space V, and let $\ell_1 \ldots, \ell_m$ be the dual basis for V^* Then every alternating bilinear map $A : V \times V \to \mathbb{R}$ can be written uniquely in the form*

$$A = \sum_{1 \le i < j \le m} \lambda_{ij}\, \ell_i \wedge \ell_j.$$

where the λ_{ij} are real numbers.

(ii) Let $U \subset \mathbb{R}^m$ and $\omega \in \Omega^2(U)$. Then ω can be written uniquely in the form

$$\omega = \sum_{1 \le i < j \le m} L_{ij} dx_i \wedge dx_j,$$

where the L_{ij} are smooth functions on U.

Proof (i) Let A be an alternating bilinear map $V \times V \to \mathbb{R}$. We have

$$A(u, v) = A(\sum_i u_i e_i, \sum_j v_j e_j) = \sum_{i,j} u_i v_j A(e_i, e_j)$$

by bilinearity, and

$$\sum_{i,j} u_i v_j A(e_i, e_j) = \sum_{i<j} u_i v_j A(e_i, e_j) + \sum_{i>j} u_i v_j A(e_i, e_j).$$

Interchange the summation indices i and j in the second summand on the right: it becomes $\sum_{i<j} u_j v_i A(e_j, e_i)$, and so is equal to $-\sum_{i<j} u_j v_i A(e_i, e_j)$, since A is alternating. Hence

$$A(u, v) = \sum_{i<j} (u_i v_j - u_j v_i) A(e_i, e_j) = \sum_{i<j} A(e_i, e_j) \ell_i \wedge \ell_j \, (u, v)$$

This shows that

$$A = \sum_{i<j} A(e_i, e_j) \, \ell_i \wedge \ell_j.$$

(ii) By (i), at each $x \in U$, ω is a (unique) linear combination of the constant 2-forms $dx_i \wedge dx_j$ for $1 \leq i < j \leq m$. In order that ω be a smooth 2-form, these coefficients must vary smoothly with x, giving the functions L_{ij}. □

It will be useful to define the wedge product of k-forms and ℓ-forms. Unfortunately it cannot be defined by means of as simple a formula as (6.14), for two reasons; first, we have to ensure that the object we define really is alternating, and second, we would like the associative law to hold: we want

$$\omega_1 \wedge (\omega_2 \wedge \omega_3) = (\omega_1 \wedge \omega_2) \wedge \omega_3,$$

for forms $\omega_1, \omega_2, \omega_3$ of any degree. The result of these requirements is a rather complicated formula. Before stating it, let me just point out, in reassurance, that it encompasses the most important example, namely det: it will turn out that on $\mathbb{R}^m$,

$$\det = dx_1 \wedge \cdots \wedge dx_m,$$

and thus the wedge product in the end brings about an important simplification.

Definition 6.18. *Let $U \subset \mathbb{R}^m$ and let $\omega_1 \in \Omega^k(U)$ and $\omega_2 \in \Omega^\ell(U)$. Define*

$$\omega_1 \wedge \omega_2 \in \Omega^{k+\ell}(U)$$

by

$$(\omega_1 \wedge \omega_2)(v_1, \ldots, v_{k+\ell}) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} sign(\sigma) \omega_1\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big) \omega_2\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)}\big).$$

For example, let us evaluate $dx_1 \wedge (dx_2 \wedge dx_3)$ on the triple of vectors (u, v, w). We have

$$dx_1 \wedge (dx_2 \wedge dx_3)\,(u, v, w) =$$

$$\frac{1}{2!}\big\{dx_1(u)\big(dx_2 \wedge dx_3\,(v, w)\big) - dx_1(u)\big(dx_2 \wedge dx_3\,(w, v)\big) + dx_1(v)\big(dx_2 \wedge dx_3\,(w, u)\big)$$
$$-dx_1(v)\big(dx_2 \wedge dx_3\,(u, w)\big) + dx_1(w)\big(dx_2 \wedge dx_3\,(u, v)\big) - dx_1(w)\big(dx_2 \wedge dx_3\,(v, u)\big)\big\}$$
$$= dx_1(u) \frac{dx_2 \wedge dx_3\,(v, w) - dx_2 \wedge dx_3\,(w, v)}{2!}$$

$$dx_1(v)\frac{dx_2\wedge dx_3\,(w,u) - dx_2\wedge dx_3\,(u,w)}{2!}$$
$$+dx_1(w)\frac{dx_2\wedge dx_3\,(u,v) - dx_2\wedge dx_3\,(v,u)}{2!}$$

Each of the fractions in this expression is a difference of 2! terms. Because the form $dx_i \wedge dx_j$ is alternating, the second is (-1) times the first. Thus

$$\begin{aligned}
&dx_1 \wedge (dx_2\wedge dx_3)\,(u,v,w)\\
&= dx_1(u)\,dx_2\wedge dx_3\,(v,w)\\
&\quad +dx_1(v)\,dx_2\wedge dx_3\,(w,u)\\
&\quad\quad +dx_1(w)\,dx_2\wedge dx_3\,(u,v)\\
&= \begin{vmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{vmatrix}
\end{aligned}$$

In fact exactly the same calculation, with arbitrary 1-forms ω_1, ω_2 and ω_3 in place of dx_1, dx_2 and dx_3, shows that

$$\omega_1\wedge(\omega_2\wedge\omega_3)(u,v,w) = \begin{vmatrix} \omega_1(u) & \omega_1(v) & \omega_1(w) \\ \omega_2(u) & \omega_2(v) & \omega_2(w) \\ \omega_3(u) & \omega_3(v) & \omega_3(w) \end{vmatrix}$$

Permutation of the ω_i results in the same sign-changes on both sides of this expression, and from this, associativity of the wedge product, at least among 1-forms, follows.

Exercise 6.19. *(i) Suppose that $\omega_1, \ldots, \omega_k$ are 1-forms, and $v_1, \ldots, v_k$ are vectors. Show that*

$$\omega_1\wedge(\omega_2\wedge(\cdots\wedge\omega_k)\cdots)\,(v_1, \ldots, v_k) = \begin{vmatrix} \omega_1(v_1) & \cdots & \omega_1(v_k) \\ \cdots & \cdots & \cdots \\ \omega_k(v_1) & \cdots & \omega_k(v_k) \end{vmatrix} \tag{6.15}$$

Hint: induction on k. The form of the induction step is suggested by the previous calculation.

(ii) Conclude that the parentheses on the left hand side of (6.15) are unnecessary.

□

Exercise 6.20. *Let U be an open subset of $\mathbb{R}^m$. Show that every form $\omega \in \Omega^k(U)$ can be written uniquely as*

$$\omega = \sum_{1\le i_1<\cdots<i_k\le m} L_{i_1\cdots i_k} dx_{i_1}\wedge\cdots\wedge dx_{i_k},$$

where the $L_{i_1\cdots i_k}$ are smooth functions on U. Hint: this is a straightforward generalisation of 6.17. Notice, however, that we have used 6.19 to dispense with any parentheses in the expression $dx_{i_1}\wedge\cdots\wedge dx_{i_k}$. □

Exercise 6.21. *Prove that if* $\omega_1 \in \Omega^k(U), \omega_2 \in \Omega^\ell(U)$ *and* $\omega_3 \in \Omega^m(U)$ *then*

$$\omega_1 \wedge (\omega_2 \wedge \omega_3) = (\omega_1 \wedge \omega_2) \wedge \omega_3.$$

That is, the wedge product of forms of arbitrary degree is associative. Hint: use 6.20 and 6.19(ii). □

Exercise 6.22. *Let* $\omega_1 \in \Omega^k(U)$ *and* $\omega_2 \in \Omega^\ell(U)$. *Show that*

$$\omega_1 \wedge \omega_2 = (-1)^{k\ell} \omega_2 \wedge \omega_1.$$

□

6.4 Integration of differential forms

Just as we could integrate 1-forms on 1-manifolds, we can integrate k-forms on $\mathbb{R}^m$ over compact oriented k-dimensional submanifolds of $\mathbb{R}^m$. Let us first define the integral of a k-form ω over a submanifold Z which is diffeomorphic to a region $R \subset \mathbb{R}^k$ with smooth boundary. Suppose that

$$\psi : R \to Z$$

is an orientation-preserving diffeomorphism, and suppose that $\omega \in \Omega^k(U)$ for some open set U containing Z. Let $e_1 = (1, \ldots, 0), \ldots, e_k = (0, \ldots, 1)$ be the standard basis for $\mathbb{R}^k$. Then for each $x \in R$, we apply ω to the k-tuple of vectors $d_x\psi(e_1), \ldots, d_x\psi(e_k)$. This gives us a number. We define $\int_Z \omega$ by

$$\int_Z \omega := \int_R \omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big) \tag{6.16}$$

Note that the vector $d_x\psi(e_j)$ can also be writen as $\partial\psi/\partial x_j$.

The integral over the product region R can of course be evaluated as an iterated integral

$$\int \cdots \int \omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big) dx_1 \ldots dx_k.$$

Here, the symbols dx_j make another appearance. In some sense they are the same as the 1-forms we discussed in Example 6.1, but here their only role is to guide us in the order of integration (first integrate with respect to x_1, then with respect to x_2, etc.)

In order that (6.16) can be said to define $\int_Z \omega$, we need

Lemma 6.23. *If* $\phi : R' \to Z$ *is another orientation-preserving diffeomorphism, then*

$$\int_R \omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big) = \int_{R'} \omega\big(d_y\phi(e_1), \ldots, d_y\phi(e_k)\big). \tag{6.17}$$

This is exactly what we showed in the 1-dimensional case, in the discussion leading up to 6.2. And essentially the same proof works here, although because we allow $k > 1$ we have to use the formula for a change of variable in multiple integration, 6.10, where in the 1-dimensional case we used integration by substitution.

Proof Because ψ and ϕ are both diffeomorphisms, so is $\phi^{-1}\circ\psi := h$, and we have $\psi = \phi\circ h$. Then

$$\omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big) = \omega\big(d_y\phi(d_xh(e_1)), \ldots, d_y\phi(d_xh(e_k))\big), \tag{6.18}$$

where $y = h(x)$, by the chain rule. Write

$$d_xh(e_i) = \sum_j \alpha_{ij}e_j \tag{6.19}$$

(so $\alpha_{ij} = \partial h_j/\partial x_i$).

Lemma 6.24. *If $A : \mathbb{R}^k \times \cdots \times \mathbb{R}^k \to \mathbb{R}$ is any alternating k-linear map, and $v_1, \ldots, v_k$ are any k vectors in $\mathbb{R}^m$, then*

$$A\big(\sum_j \alpha_{1j}v_j, \ldots, \sum_j \alpha_{kj}v_j\big) = \begin{vmatrix} \alpha_{11} & \cdots & \alpha_{k1} \\ & & \\ \alpha_{1k} & \cdots & \alpha_{kk} \end{vmatrix} A(v_1, \ldots, v_k).$$

Proof This is really a small modification of Exercise 6.12. Define a new map $B : (\mathbb{R}^k)^k \to \mathbb{R}$ by

$$B\left(\begin{bmatrix} \alpha_{11} \\ \cdot \\ \alpha_{1k} \end{bmatrix}, \cdots, \begin{bmatrix} \alpha_{k1} \\ \cdot \\ \alpha_{kk} \end{bmatrix}\right) = A\big(\sum_j \alpha_{1j}v_1, \ldots, \sum_j a_{kj}v_j\big).$$

Because A is alternating, so is B. Hence, by 6.12, $B = \lambda \det$ for some $\lambda \in \mathbb{R}$. To find the value of λ, we evaluate B on some k-tuple of vectors where we already know the answer. Because $B = \lambda \det$, we have

$$\lambda = \lambda \det\left(\begin{bmatrix} 1 \\ \cdot \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ \cdot \\ 1 \end{bmatrix}\right) = B\left(\begin{bmatrix} 1 \\ \cdot \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ \cdot \\ 1 \end{bmatrix}\right) = A(v_1, \ldots, v_k).$$

where the last equality is by the definition of B. Hence

$$A\big(\sum_j \alpha_{1j}v_j, \ldots, \sum_j \alpha_{kj}v_j\big) = B\left(\begin{bmatrix} \alpha_{11} \\ \cdot \\ \alpha_{1k} \end{bmatrix}, \cdots, \begin{bmatrix} \alpha_{k1} \\ \cdot \\ \alpha_{kk} \end{bmatrix}\right) = A(v_1, \ldots, v_k)\begin{vmatrix} \alpha_{11} & \cdots & \alpha_{k1} \\ & & \\ \alpha_{1k} & \cdots & \alpha_{kk} \end{vmatrix}.$$

$\square$

Now we continue the proof of 6.23. By (6.18) and (6.19),

$$\omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big) = \omega\big(\sum_j \alpha_{1j}d_y\phi(e_j), \ldots, \sum_j \alpha_{kj}d_y\phi(e_j)\big).$$

Apply Lemma 6.24 with $A = \omega$ and $v_j = d_y\phi(e_j)$ for $j = 1, \ldots, k$. We get

$$\omega\big(\sum_j \alpha_{1j} d_y\phi(e_j), \ldots, \sum_j \alpha_{kj} d_y\phi(e_j)\big) = \begin{vmatrix} \alpha_{11} & \cdots & \alpha_{k1} \\ \\ \alpha_{1k} & \cdots & \alpha_{kk} \end{vmatrix} \omega\big(d_y\phi(e_1), \ldots, d_y\phi(e_k)\big).$$

Since $\alpha_{ij} = \partial h_j / \partial x_i$, the determinant here is the jacobian determinant of the diffeomorphism h, $J(h)$. Thus

$$\omega\big(\sum_j \alpha_{1j} d_y\phi(e_j), \ldots, \sum_j \alpha_{kj} d_y\phi(e_j)\big) = J(h)\ \omega\big(d_y\phi(e_1), \ldots, d_y\phi(e_k)\big)$$

and

$$\int_R \omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big) = \int_R J(h)\ \omega\big(d_y\phi(e_1), \ldots, d_y\phi(e_k)\big) \tag{6.20}$$

where, again, $y = h(x)$. But the right hand side here is just

$$\int_{R'} \omega\big(d_y\phi(e_1), \ldots, d_y\phi(e_k)\big)$$

by the change-of-variables formula 6.10. □

The expression $\omega\big(d_x\psi(e_1), \ldots, d_x\psi(e_k)\big)$ appearing in the definition (6.16) of $\int_Z \omega$ is rather clumsy. Here is a more flexible notation.

Definition 6.25. *Suppose that U and V are open sets in Euclidean spaces, $\omega \in \Omega^k(U)$, and $h : V \to U$ is a smooth map. Define a new k-form, $h^*(\omega) \in \Omega^k(V)$ (the pull back of ω by h), by*

$$h^*(\omega)(v_1, \ldots, v_k) = \omega(d_x h(v_1), \ldots, d_x h(v_k)).$$

Note that if ω is a 0-form — i.e. a smooth function f — then this definition says that $h^*(\omega) = f \circ h$.

It is, once again, slightly uncomfortable that the x appears on the right hand side and not on the left. See Remark 6.14 for an expression of regret.

Example 6.26. (i) Let $f : \mathbb{R}^m \to \mathbb{R}$, and let t be a coordinate on $\mathbb{R}$. Then for $v \in \mathbb{R}^m$,

$$f^*(dt)(v) = dt(d_x f(v)) = dt\big(\sum_i \frac{\partial f}{\partial x_i} v_i\big) = \sum_i \frac{\partial f}{\partial x_i} v_i = \sum_i \frac{\partial f}{\partial x_i} dx_i\ (v). \tag{6.21}$$

Thus $f^*(dt) = df$. The third equality on the previous line is also slightly uncomfortable. The expressions on either side of it are different because on the left we are thinking of $d_x f(v)$ as a vector in $T_{f(x)}\mathbb{R}$, while on the right we are thinking of $d_x f(v)$ as a number in $\mathbb{R}$. The 1-form dt intervenes to turn the vector into a number. Identifying $T_{f(x)}\mathbb{R}$ with $\mathbb{R}$ and thinking of

$d_x f(v)$ as a number, we cause occasional notational confusion.

(ii) Let $f : \mathbb{R}^n \to \mathbb{R}^2$. Then for vectors $v_1, v_2 \in \mathbb{R}^n$, we have

$$\begin{aligned}
f^*(dx_1{\wedge}dx_2)(v_1, v_2) &= dx_1{\wedge}dx_2\ (d_x f(v_1), d_x f(v_2)) \\
&= dx_1(d_x f(v_1))\, dx_2(d_x f(v_2)) - dx_1(d_x f(v_2))\, dx_2(d_x f(v_1)) \\
&= d_x f_1(v_1) d_x f_2(v_2) - d_x f_1(v_2) d_x f_2(v_1) \\
&= df_1{\wedge}df_2\ (v_1, v_2)
\end{aligned}$$

So $f^*(dx_1{\wedge}dx_2) = df_1{\wedge}df_2$. In view of (i), this can also be written $f^*(dx_1{\wedge}dx_2) = f^*(dx_1){\wedge}f^*(dx_2)$. □

Slightly less immediate, but nonetheless easy to prove directly from the definitions of wedge product and pull-back, is the more general result

Lemma 6.27.

$$f^*(\omega_1{\wedge}\omega_2) = f^*(\omega_1) \wedge f^*(\omega_2).$$

□

It is an immediate consequence of the chain rule that

$$(g \circ f)^*(\omega) = f^*(g^*(\omega)).$$

The definition of pull-back enables us to write the integral in (6.16) as

$$\int_R \psi^*(\omega)(e_1, \ldots, e_k) \tag{6.22}$$

Now the k-form $\psi^*(\omega)$ is equal to $f \det$ for some smooth function f on R, by 6.15, and thus to $f dx_1{\wedge}\cdots{\wedge}dx_k$, by 6.19. So $\psi^*(\omega)(e_1, \ldots, e_k) = f dx_1{\wedge}\cdots{\wedge}dx_k(e_1, \ldots, e_k) = f$.

Definition 6.28. *(i) Let R be a region in $\mathbb{R}^k$. We define*

$$\int_R f dx_1{\wedge}\cdots{\wedge}dx_k = \int_R f(x) dx_1 \cdots dx_k.$$

Lemma 6.23 can now be restated as

Proposition 6.29. *Suppose that R_1 and R_2 are regions in $\mathbb{R}^k$ and $h : R_1 \to R_2$ is an orientation-preserving diffeomorphism. Then for any k-form ω on R_2,*

$$\int_{R_1} h^*(\omega) = \int_{R_2} \omega. \tag{6.23}$$

And in the light of this, the definition, in (6.16), of the integral of a k-form over a k-dimensional oriented submanifold of $\mathbb{R}^m$. can be rewritten as

Definition 6.30. *(i) Let Z be an oriented submanifold of $\mathbb{R}^N$, diffeomorphic, by an orientation-preserving diffeomorphism, to a region R in $\mathbb{R}^k$. Let ω be a k form defined on some set U containing Z. Choose an orientation-preserving diffeomorphism $\psi : R \to Z$. Then*

$$\int_Z \omega := \int_R \psi^*(\omega). \tag{6.24}$$

The proof that this does not depend on the choice of R or ψ now runs as follows: if $\psi : R_1 \to Z$ and $\phi : R_2 \to Z$ are orientation-preserving diffeomorphisms, then $\phi^{-1} \circ \psi =: h$ is an orientation preserving diffeomorphism $R_1 \to R_2$. So

$$\int_{R_1} \psi^*(\omega) = \int_{R_1} (\phi \circ h)^*(\omega) = \int_{R_1} h^*(\phi^*(\omega)) = \int_{R_2} \phi^*(\omega),$$

where the last equality is (6.23) applied to the form $\phi^*(\omega)$.

Definition 6.30. *(continued) (ii) Let Z be any compact submanifold of $\mathbb{R}^N$, and let ω be a k-form defined on some set U containing Z. Choose a finite atlas $\{\phi_i : U_i \to V_i\}_{\{i=1,\dots,n\}}$ for Z, and a partition of unity $\{\rho_i\}$ subordinate to the open cover $\{U_i\}$. (Recall that this means a collection of smooth functions $\rho_i : Z \to [0,1]$ such that supp$(\rho_i) \subset U_i$, and such that for all $x \in Z$, $\sum_i \rho_i(x) = 1$.) Define*

$$\int_Z \omega = \sum_i \int_{U_i} \rho_i \omega. \tag{6.25}$$

Exercise 6.31. *Show that the value of the right hand side of (6.25) is independent of the atlas and partition of unity chosen.* Actually this is quite easy; if $\{\psi_j : U'_j \to V'_j\}_{\{j=1,\dots,s\}}$ is another atlas and σ_j is a partition of unity subordinate to the cover $\{U'_j\}$, then $\{U_i \cap U'_j\}_{\{i=1,\dots,r,\, j=1,\dots,s\}}$ is also an open cover and $\{\rho_i \sigma_j\}$ is a partition of unity subordinate to it. □

Remark 6.32. This is a one of those definitions which may seem utterly impractical, since although partitions of unity exist, one rarely gets one's hands on one explicitly. There are two answers to this objection. The first is that in many of the cases where we really want to calculate some integral $\int_Z \omega$, we can use a map $\psi : R \to Z$ which is a diffeomorphism outside a set of measure zero, and then integrate $\psi^*(\omega)$ over R. For example, this is the case if Z is a torus $S^1 \times \cdots \times S^1$: there is an obvious parametrisation $R := [0,1]^k \to (S^1)^k$ which is a diffeomorphism on the interior of R. The second answer is Patience! increasingly as one goes on in mathematics, definitions do not immediately yield a practical procedure for calculation. Sometimes it is only after introducing an apparently impractical definition that one can develop methods which make calculation possible. In fact this is the case with the integral of a function of one variable: without the fundamental theorem of calculus, which tells you that you can calculate a definite integral by evaluating an anti-derivative of the integrand f at the end-points, the definition of the Riemann integral (as the supremum of the set of integrals of step functions approximating f from below) would also look rather impractical.

Another answer is that definitions are not always made with the purpose you might expect. In particular, it may not be intended to calculate *all* the things they define.

We will shortly see, by means of Stokes's Theorem, which is itself a generalisation of the fundamental theorem of calculus, that the integral of a certain class of forms is always zero, a result which might seem rather disappointing as a calculation, but has important theoretical consequences, and enables us to calculate the integrals of the members of another class of forms rather easily, in a way which makes their topological significance clear.

We now introduce an $\mathbb{R}$-linear map, the *exterior derivative*, transforming k-forms into $(k+1)$-forms.

Definition 6.33. *(i) Let U be an open set in $\mathbb{R}^N$, and let $\omega \in \Omega^k(U)$. Define a $(k+1)$-form $d\omega$ (the exterior derivative of ω), as follows: ω can be written uniquely as*

$$\omega = \sum_{i_1<\cdots<i_k} L_{i_1\cdots i_k} dx_{i_1}\wedge\cdots\wedge dx_{i_k}$$

where each $L_{i_1\cdots i_k}$ is a smooth function on U. Then

$$d\omega = \sum_{i_1<\cdots<i_k} dL_{i_1\cdots i_k}\wedge dx_{i_1}\wedge\cdots\wedge dx_{i_k}.$$

Notice that this coincides with the already existing notion of the derivative when $k=0$ and our k-form ω is just a smooth function.

Example 6.34. (i) $d(dx_{i_1}\wedge\cdots\wedge dx_{i_k}) = d(1 dx_{i_1}\wedge\cdots\wedge dx_{i_k}) = d1\wedge dx_{i_1}\wedge\cdots\wedge dx_{i_k} = 0.$

(ii) If $\omega = x_1 dx_2$ then $d\omega = dx_1\wedge dx_2$. So for example $d(x_1dx_2 - x_2dx_1) = 0$.

(iii) If $\omega = f_1\, dx_2\wedge dx_3 + f_2\, dx_3\wedge dx_1 + f_3\, dx_1\wedge dx_2$ then

$$d\omega = \Big(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3}\Big) dx_1\wedge dx_2\wedge dx_3. \tag{6.26}$$

(iv) If $\omega = f_1\, dx_1 + f_2\, dx_2 + f_3\, dx_3$ then

$$d\omega = \Big(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\Big) dx_1\wedge dx_2 + \Big(\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}\Big) dx_3\wedge dx_1 + \Big(\frac{\partial f_3}{\partial x_2}\frac{\partial f_2}{\partial x_3}\Big) dx_2\wedge dx_3. \tag{6.27}$$

Classically, one meets the expressions

$$\mathrm{div}(f_1, f_2, f_3) = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3}$$

the *divergence* of (f_1, f_2, f_3)) and

$$\mathrm{curl}(f_1, f_2, f_3) = \Big(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2},\ \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1},\ \frac{\partial f_3}{\partial x_2}\frac{\partial f_2}{\partial x_3}\Big)$$

(the *curl*, or *rotational*, of (f_1, f_2, f_3).) The divergence figures in Gauss's Theorem

$$\int_{\partial R} f \cdot \mathbf{n} = \int_R \operatorname{div}(f)$$

where R is a region in $\mathbb{R}^3$ and $\mathbf{n}$ the unit outward pointing normal to ∂R, and the curl in Stokes's Theorem

$$\int_{\partial S} f = \int_S \operatorname{curl}(f)$$

where S is an oriented surface in $\mathbb{R}^3$, and the integral on the left is a contour integral, i.e the integral over ∂S of the 1-form $f_1dx_1 + f_2dx_2 + f_3dx_3$. Both theorems are special cases of the general version of Stokes's theorem we prove below.

(v) If $\omega = df$ for some function f, then $d\omega = 0$. For

$$d(df) = \sum_j d\Big(\frac{\partial f}{\partial x_j}\Big)\wedge dx_j = \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i \wedge dx_j$$

$$= \sum_{i<j} \Big(\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 f}{\partial x_j \partial x_i}\Big) dx_i \wedge dx_j,$$

and this is equal to zero, by the equality of mixed partial derivatives.

(vi) If ω is the 1-form $d\theta$, then $d\omega = 0$, even though $d\theta$ is not globally the derivative of a function. It is enough that this should be true locally for the calculation in (v) to go through.

(vii) **Exercise** If $\omega \in \Omega^k(U)$ for any k, then $d(d\omega) = 0$. The proof of this is a generalisation of the calculation in (v), involving equality of mixed partial derivatives. It requires only good book-keeping to keep track of the signs.

(viii) **Exercise** If $\omega_1 \in \Omega^k(U)$ then

$$d(\omega_1 \wedge \omega_2) = (d\omega_1)\wedge\omega_2 + (-1)^k \omega_1 \wedge d\omega_2.$$

□

The exterior derivative and the pull-back of forms are related in a very gratifying way:

Proposition 6.35. *If U_1 and U_2 are open in $\mathbb{R}^n$ and $\mathbb{R}^p$ respectively, and $h : U_1 \to U_2$ is a smooth map, then for $\omega \in \Omega^k(U_2)$,*

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

Proof We have already seen, in 6.26(i), that this is true for 0-forms (functions). To prove it for $k > 0$, by the $\mathbb{R}$-linearity of f^* and d we need only prove it for the k-form $\omega = L dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, where L is a smooth function. We have

$$h^*(d\omega) = h^*\big(d(L\, dx_{i_1} \wedge \cdots \wedge dx_{i_k})\big) = h^*(dL \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k})$$

$$= h^*(dL) \wedge dh_{i_1} \wedge \cdots \wedge dh_{i_k} = d(h^*(L)) \wedge dh_{i_1} \wedge \cdots \wedge dh_{i_k}$$

$$= d\big(h^*(L) \wedge dh_{i_1} \wedge \cdots \wedge dh_{i_k}\big) = d(h^*(\omega)).$$

For the penultimate equality we have used 6.34(viii), taking $\omega_1 = L$ and $\omega_2 = dh_{i_1} \wedge \cdots \wedge dh_{i_k}$ (so that $d\omega_2 = 0$). $\square$

6.5 Stokes's Theorem

The final theorem we will prove is *Stokes's Theorem*, which provides a fundamental link between

$$\begin{array}{c} d : \{\text{forms}\} \to \{\text{forms}\} \\ \omega \mapsto d\omega \end{array} \qquad \text{and} \qquad \begin{array}{c} \partial : \{\text{Manifolds}\} \to \{\text{Manifolds}\} \\ M \mapsto \partial M \end{array}$$

There is a suggestive similarity between the operators d and ∂:

$$d(d\omega) = 0 \qquad \text{and} \qquad \partial(\partial M) = \emptyset.$$

On the other hand, d *raises* degree, whereas ∂ *lowers* dimension. But they fit together beautifully:

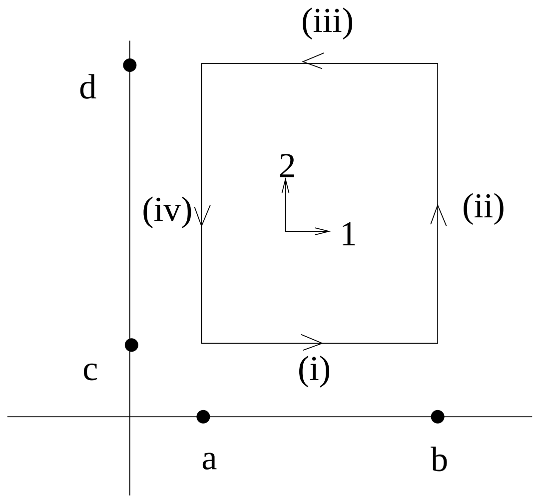
Theorem 6.36. *Let $M^m \subset \mathbb{R}^N$ be a compact oriented manifold with boundary ∂M, and let $\omega \in \Omega^{m-1}(M)$. Then*

$$\int_M d\omega = \int_{\partial M} \omega.$$

Corollary 6.37. *If M^m is a compact oriented manifold without boundary and $\omega = d\sigma \in \Omega^m(M)$ then $\int_M \omega = 0$.* $\square$

Pre-proof of Stokes's Theorem When $m = 1$, Stokes's theorem is just the fundamental theorem of calculus - provided that we supply the appropriate interpretation for the term "boundary orientation" in this case — compare (6.5) and (6.6).

Let's look in detail at the case $M = [a_1, b_1] \times [a_2, b_2]$. Admittedly, this is not a manifold with boundary in the sense we have defined, because of its corners. But as an example it is very instructive.



The orientation chosen in M induces the boundary orientation shown by the arrows. Let $\omega = pdx + qdy$. Then $d\omega = (\partial q/\partial x - \partial p/\partial y)dx\wedge dy$, and

$$\int_M d\omega = \int_{[a,b]\times[c,d]} (\partial q/\partial x - \partial p/\partial y)dx\wedge dy$$

By definition of the integral of an m-form over a region in $\mathbb{R}^m$, this is just

$$\int_{[a,b]\times[c,d]} (\partial q/\partial x - \partial p/\partial y)dxdy$$

We can evaluate this multiple integral as an iterated integral in two ways: integrating first with respect to x and then with respect to y or vice versa. It is most convenient to use one order for the first term and the other for the second:

$$\int_M d\omega = \int_c^d \int_a^b \frac{\partial q}{\partial x}(x,y)dx\,dy - \int_a^b \int_c^d \frac{\partial p}{\partial y}(x,y)dy\,dx$$

By the fundamental theorem of calculus,

$$= \int_a^b \frac{\partial q}{\partial x}(x,y)dx = q(b,y) - q(a,y)$$

and

$$= \int_c^d \frac{\partial p}{\partial y}(x,y)dy = p(x,d) - p(x,c).$$

So

$$\int_M d\omega = \int_c^d \Big(q(b,y) - q(a,y)\Big)dy - \int_a^b \Big(p(x,d) - p(x,c)\Big)dx \tag{6.28}$$

In the picture I have divided up ∂M into four parts, marked (i) — (iv). They can be parametrised as follows:

$$\begin{array}{llll}
\text{(i)} & \gamma_1 : [a,b] \to \mathbb{R}^2 & \gamma_1(x) = (x,c) \\
\text{(ii)} & \gamma_2 : [c,d] \to \mathbb{R}^2 & \gamma_2(y) = (b,y) \\
\text{(iii)} & \gamma_3 : [a,b] \to \mathbb{R}^2 & \gamma_3(x) = (x,d) \\
\text{(iv)} & \gamma_4 : [c,d] \to \mathbb{R}^2 & \gamma_2(y) = (a,y)
\end{array}$$

Of these, γ_1 and γ_2 preserve orientation and γ_3 and γ_4 reverse it. Hence

$$\int_{\partial M} \omega = \int_{\text{(i)}} \omega + \int_{\text{(ii)}} \omega + \int_{\text{(iii)}} \omega + \int_{\text{(iv)}} \omega$$

$$= \int_a^b \omega(\gamma_1'(x))dx + \int_c^d \omega(\gamma_2'(y))dy - \int_a^b \omega(\gamma_3'(x))dx - \int_c^d \omega(\gamma_4'(y))dy$$

Because the γ_i have such a simple form, their derivatives are all equal to basis vectors: $\gamma_1'(x) = \gamma_3'(x) = (1,0)$ and $\gamma_2'(y) = \gamma_4'(y) = (0,1)$. Thus

$$\int_{\partial M} \omega = \int_a^b p((g_1(x))dx + \int_c^d q(\gamma_2(y))dy - \int_a^b p(\gamma_3(x))dx - \int_c^d q(\gamma_4(y))dy$$

$$= \int_a^b p(x,c)dx + \int_c^d q(b,y)dy - \int_a^b p(x,d)dx - \int_c^d q(a,y)dy.$$

$$= \int_a^b \Big(p(x,c) - p(x,d)\Big)dx + \int_c^d \Big(q(b,y) - q(a,y)\Big)dy.$$

And this is the same as (6.28).

A calculation very similar to this one works in m dimensions, and is in some ways more satisfying (to me) than the proof which follows, in which the rather simple idea at the heart of Stokes's Theorem is muddied by the presence of the bump functions of a partition of unity. I suggest it as an exercise.

Proof of Stokes's Theorem: Let M be a compact oriented manifold with boundary. Let $\{\phi_i : U_i \to V_i\}_{\{i=1,\dots,T\}}$ be an atlas of orientation-preserving charts, with each V_i an open subset of the half-space H^m. Let $\{\rho_i\}$ be a partition of unity subordinate to the open cover $\{U_i\}$. Then

$$\int_M d\omega = \int_M d\Big(\sum_i \rho_i \omega\Big) = \sum_i \int_{U_i} d(\rho_i\omega) = \sum_i \int_{V_i} \phi_i^{-1*}\big(d(\rho_i\omega)\big)$$

$$= \sum_i \int_{V_i} d\big(\phi_i^{-1*}(\rho_i\omega)\big)$$

and similarly

$$\int_{\partial M}\omega = \sum_i \int_{\partial V_i}\phi_i^{-1*}(\omega).$$

From a comparison of these two expressions, it is clear that it is enough to prove, for any compactly supported $(m-1)$-form σ and open set V in H^m, that

$$\int_V d\sigma = \int_{\partial V}\sigma.$$

We have

$$\sigma = \sum_j L_j dx_1\wedge\cdots\wedge\widehat{dx_j}\wedge\cdots\wedge dx_m,$$

so

$$d\sigma = \sum_j \frac{\partial L_j}{\partial x_j}dx_j\wedge dx_1\wedge\cdots\wedge\widehat{dx_j}\wedge\cdots\wedge dx_m = \sum_j (-1)^{j-1}\frac{\partial L_j}{\partial x_j}dx_1\wedge\cdots\wedge dx_m.$$

Choose $R>0$ such that $\mathrm{supp}(\sigma)\subset[-R,R]^m\cap V$. Because $\mathrm{supp}(\sigma)\subset V\cap[-R,R]^m$, we can write

$$\int_V d\sigma = \int_{[-R,R]^{m-1}\times[0,R]}\sum_j (-1)^{j-1}\frac{\partial L_j}{\partial x_j}dx_1\wedge\cdots\wedge dx_m$$

It is convenient to integrate the j-th summand first with respect to x_j. For $j<m$, this gives

$$(-1)^{j-1}\int_{[-R,R]^{m-2}\times[0,R]}\Big(\int_{-R}^{R}\frac{\partial L_j}{\partial x_j}dx_j\Big)dx_1\cdots\widehat{dx_j}\cdots dx_m.$$

But here the inner integral is zero, since by the fundamental theorem of calculus it is equal to

$$L_j(x_1,\ldots,R,\ldots,x_m) - L_j(x_1,\ldots,-R,\ldots,x_m)$$

and L_j is identically zero outside (and therefore on the boundary of) $[-R,R]^m$. For $j=m$, we get

$$(-1)^{m-1}\int_{[-R,R]^{m-1}}\Big(\int_0^R \frac{\partial L_m}{\partial x_m}dx_m\Big)dx_1\ldots dx_{m-1}$$

and, once again because σ vanishes outside $[-R,R]^m$, this is equal to

$$= (-1)^m\int_{[-R,R]^{m-1}} L_m(x_1,\ldots,x_{m-1},0)dx_1\ldots dx_{m-1}.$$

Finally, because $dx_m(v)=0$ for every vector v tangent to ∂V, if $j<m$ the $(m-1)$ form $dx_1\wedge\cdots\wedge\widehat{dx_j}\wedge\cdots\wedge dx_m$ is identically zero on ∂V, and so on ∂V σ reduces to $L_m dx_1\wedge\cdots\wedge dx_{m-1}$. The boundary orientation of $\mathbb{R}^{m-1}\times\{0\} = \partial H^m$ is $(-1)^{m-1}$ times the orientation of $\mathbb{R}^{m-1}\times\{0\}$ coming from its identification with $\mathbb{R}^{m-1}$. Therefore

$$\int_{\partial V}\sigma = \int_V d\sigma.$$

□

6.6 Concluding Remarks

For reasons of time, our treatment of differential forms on manifolds only covers manifolds embedded in Euclidean space, and so is seriously incomplete. Moreover, even here, we spoke only of differential forms defined on open sets of the ambient Euclidean space. This is fine when we are concerned only with integrating forms over submanifolds, but forms play another, less obvious role, in the study of cohomology, and for this one needs to make sense of the idea of a differential form on a manifold M, without reference to an ambient space. The space of all smooth k-forms on a manifold M is denoted $\Omega^k(M)$, and the collection of spaces and linear maps

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

(where $n = \dim M$) is known as the *de Rham complex* of M. From this complex important information about the cohomology of M can be extracted by purely algebraic procedures. A good place to read about this is in the final chapters of [3], though the treatment there is still restricted to manifolds embedded in Euclidean space. More advanced treatments, covering abstract manifolds as well, and developing de Rham cohomology, can be found in the classic text [1] and in [5]. The book of Conlon, [2], is, despite its name, more encyclopaedic, though a very useful reference.

7 Appendices

7.1 Appendix A: Linear Maps and Matrices

Let U be an n-dimensional real vector space. A choice of basis determines a bijection

$$\{\text{Linear maps } U \to U\} \simeq \{n \times n \text{ real matrices}\}$$

and an isomorphism

$$\{\text{Invertible linear maps } U \to U\} \simeq \mathrm{Gl}(n, \mathbb{R}),$$

but since in general one basis is no better than another, there are many equally good bijections and isomorphisms. More generally, if U and V are vector spaces of dimensions n and p respectively then choices of basis E for U and F for V determine a bijection

$$\{\text{Linear maps } U \to V\} \to \{p \times n \text{ real matrices}\}.$$

What is the relation between different bijections obtained in this way? To sort this all out, we introduce some notation which may seem rather ponderous at first, but works like a Rolls Royce once you put it on the road.

Remember that a choice of basis gives us a way of converting vectors in our n-dimensional space U into n-tuples of numbers. Let $E = e_1, \ldots, e_n$ be a basis for U. If $u \in U$ is equal to $\alpha_1 e_1 + \cdots + \alpha_n e_n$, then we write

$$[u]_E = \begin{bmatrix} \alpha_1 \\ \cdots \\ \cdots \\ \alpha_n \end{bmatrix}.$$

That is, $[u]_E$ *is the expression of* u *with respect to the basis* E, as a column vector.
Suppose that U and V are vector spaces with bases $E = e_1, \ldots, e_n$ and $F = f_1, \ldots, f_m$ respectively, and let $T : U \to V$ be a linear map. Let us denote the matrix of T with respect to the bases E in the source, U, and F in the target V by

$$[T]_F^E.$$

By definition,

$$[T]_F^E = \begin{bmatrix} [T(e_1)]_F & \cdots & [T(e_n)]_F \end{bmatrix} \tag{7.1}$$

— its columns are the expressions of the vectors $T(e_1), \ldots, T(e_n)$ with respect to the basis F. Multiplication of matrices is defined so that multiplication by $[T]_F^E$ converts the expression for a vector u with respect to the basis E, into the expression for $T(u)$ with respect to the basis F. That is,

$$[T(u)]_F = [T]_F^E [u]_E.$$

Once you know the rule for multiplying a column vector by a matrix, this last equality is actually obvious:

$$[T]_F^E \begin{bmatrix} \alpha_1 \\ \cdots \\ \cdots \\ \alpha_n \end{bmatrix} = \alpha_1 \cdot \text{column } 1 \text{ of } [T]_F^E + \ \cdots + \alpha_n \cdot \text{column } n \text{ of } [T]_F^E$$

$$\begin{aligned} &= \alpha_1[T(e_1)]_F + \ \cdots \ + \alpha_n[T(e_n)]_F \\ &= [\alpha_1 T(e_1) + \cdots + \alpha_n T(e_n)]_F \\ &= [T(\alpha_1 e_1 + \cdots + \alpha_n e_n)]_F \\ &= [T(u)]_F. \end{aligned}$$

From this everything else follows. For example, suppose that $U \xrightarrow{S} V$ and $V \xrightarrow{T} W$ are linear maps, and let E be a basis for U, F a basis for V and G a basis for W. Then

$$[T]_G^F \, [S]_F^E = [T]_G^F \, \begin{bmatrix} [S(e_1)]_F & \cdots & [S(e_n)]_F \end{bmatrix};$$

since

$$[T]_G^F [v]_F = [T(v)]_G$$

for any $v \in V$, in particular

$$[T]_G^F [S(e_i)]_F = [T(S(e_i))]_G$$

and thus

$$\begin{aligned} [T]_G^F [S]_F^E &= \begin{bmatrix} [T(S(e_1))]_G & \cdots & [T(S(e_n))]_G \end{bmatrix} \\ &= [T \circ S]_G^E. \end{aligned}$$

The equality we have just deduced,

$$[T]_G^F [S]_F^E = [T \circ S]_G^E, \tag{7.2}$$

expresses the fundamental relation between multiplication of matrices and composition of linear transformations; it figures as Theorem 7.3 in Derek Holt's Linear Algebra Lecture Notes.

The matrix for a change of basis can also be conveniently represented in this way. Suppose that E and F are now both bases for the same space U. Denote the identity map from U to U by I. Then of course $[I]_E^E$ and $[I]_F^F$ are both "identity matrices": matrices with 1's down the diagonal and zeros everywhere else: for example

$$\begin{aligned} [I]_E^E &= \begin{bmatrix} [I(e_1)]_E & \cdots & [I(e_n)]_E \end{bmatrix} \\ &= \begin{bmatrix} [e_1]_E & \cdots & [e_n]_E \end{bmatrix} \end{aligned}$$

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

On the other hand, $[I]_F^E$ and $[I]_E^F$ are *not* identity matrices; for example

$$[I]_F^E = \left[\ [e_1]_F \quad \cdots \quad [e_n]_F \ \right]$$

and of course $[e_i]_F$ is not the i-th column of the identity matrix unless $e_i = f_i$.

However, $[I]_F^E$ and $[I]_E^F$ are mutually inverse:

$$[I]_F^E [I]_E^F = [I \circ I]_F^F = [I]_F^F$$

is the identity matrix. And in particular if $T : U \to U$ we have

$$[I]_E^F \ [T]_F^F \ [I]_F^E = [I \circ T]_E^F \ [I]_F^E$$

$$= [I \circ F \circ I]_E^E = [T]_E^E.$$

In other words,

$$[T]_E^E = \left(\ [I]_F^E \ \right)^{-1} \ [T]_F^F \ [I]_F^E,$$

and $[T]_E^E$ is the conjugate of $[T]_F^F$ by $[I]_F^E$.

7.2 Appendix B: Some Topics in Linear Algebra

Experience in previous years' exams shows that students often find difficulties with basic linear algebra questions of the following type:

Exercise 7.1. Let W be a vector space, and let U and V be subspaces of W. In the product space $W \times W$, let D be the diagonal, $D = \{(w, w) : w \in W\}$. Show that

$$U + V = W \quad \text{if and only if} \quad (U \times V) + D = W \times W.$$

Exercise 7.2. Let U, V and W be vector spaces, and let $S : U \to W$ and $T : V \to W$ be linear maps. Let

$$X = \{(u, v) \in U \times V : S(u) = T(v)\}.$$

Let $p : X \to U$ be the restriction of the usual projection $U \times V \to U$.

(a) Show that p is injective if and only if T is injective.

Now suppose also that $S(U) + T(V) = W$ (i.e. the set $\{S(u) + T(v) : u \in U, v \in V\}$ is all of W).

(b) Show that p is surjective if and only if T is surjective.

Try these questions now! Neither involves much more than basic set theory.

The notion of *quotient space* is very useful. Students tend to meet it only in group theory (quotient groups) and not in linear algebra (quotient spaces) though the latter is really a special case of the former.

Let V be a vector space and let W be a subspace. Declare vectors $v_1, v_2 \in V$ to be equivalent modulo W if their difference $v_1 - v_2$ lies in W. This is easily checked to be an equivalence relation. The quotient space V/W is defined to be the set of equivalence classes. It is useful to denote the equivalence class of v by $[v]$.

Exercise 7.3. (i) Show that in these circumstances $[v] = v + W (= \{v + w : w \in W\})$. Because the expression $v + W$ is more transparent than $[v]$ (for a start, it tells you what the equivalence relation is), it is more usual.
(ii) Let $V = \mathbb{R}^2$ and let W be the subspace $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\}$. Make a drawing showing W and showing the cosets $v + W$ for $v = (1, 0), (0, 1)$ and $(1, 2)$.
(iii) Let $U = W^\perp = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = -x_2\}$. Add U to your drawing for (ii), and then find a "natural" bijection $U \to V/W$.
(iv) Show that the operation of addition defined in the quotient V/W (in general, not just in the previous example) by

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W$$

(or, in the other notation, by $[v_1] + [v_2] = [v_1 + v_2]$) is well defined.
(v) How is multiplication by scalars defined in V/W?
(vi) Is the bijection you found in (iii) an isomorphism?
(vii) Suppose that $e_1, \ldots, e_n$ is a basis for V, and that the first k of them, $e_1, \ldots, e_k$, form a basis for the subspace $W \subset V$. Show that $e_{k+1} + W, \ldots, e_n + W$ is then a basis for V/W.
(viii) Let V be a vector space with subspaces U and W. Under what circumstances is the map $U \to V/W$ defined by

$$u \mapsto u + W$$

(a) injective
(b) surjective
(c) an isomorphism?

7.3 Appendix C: Derivatives

The other thing that people struggle with is the notion of the derivative of a smooth map. This is the key definition in the course MA225 Differentiation, but unfortunately many people clearly only feel comfortable with the notion of derivative they learned about at A-level. The derivative of a smooth map $f : \mathbb{R}^n \to \mathbb{R}^p$ at a point $x \in \mathbb{R}^n$ is not a number, but a linear map $d_x f : \mathbb{R}^n \to \mathbb{R}^p$, which depends on x. It is an easy consequence of the definition of the derivative that

$$d_x f(\hat{x}) = \lim_{h \to 0} \frac{f(x + h\hat{x}) - f(x)}{h}. \tag{7.3}$$

Incidentally, the statement that f is differentiable at x is little more than the statement that the limit in (7.3) depends *linearly* on the vector $\hat{x}$. It is not hard to find examples of (non-differentiable) maps f and points x such that the limit in (7.3) exists for all $\hat{x}$, but is *not* linear in $\hat{x}$. However, in this course, and in the rest of this appendix, we are concerned only with maps which are everywhere differentiable. For completeness I give the formal definition:

Definition 7.4. *Let U be an open set in $\mathbb{R}^n$. The map $f : U \to \mathbb{R}^p$ is differentiable at x if there is a linear map $L : \mathbb{R}^n \to \mathbb{R}^p$ such that*

$$f(x + \hat{x}) = f(x) + L(\hat{x}) + \|\hat{x}\| E(x, \hat{x}),$$

where the "error term" E tends to 0 as $\hat{x}$ tends to zero. In other words, f is differentiable at x if there is a linear map $L : \mathbb{R}^n \to \mathbb{R}^p$ such that

$$\lim_{\hat{x} \to 0} \frac{f(x + \hat{x}) - \big(f(x) + L(\hat{x})\big)}{\|\hat{x}\|} = 0.$$

In this case the linear map L is called the **derivative** *of f at x, and denoted by $d_x f$.*

Exercise 7.5. Show that if f is differentiable at x then (7.3) holds for all vectors $\hat{x}$.

Theorem 7.6. *(i) If all first order partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$ exist in some neighbourhood of the point x, and are all continuous at x, then f is differentiable at x.*
(ii) The map $f : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at x if and only if all of its component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are differentiable at x. □

Exercise 7.7. Suppose that f is differentiable at x. Show, using (7.1) and (7.3) above, that in the matrix of $d_x f$ with respect to the usual bases of $\mathbb{R}^n$ and $\mathbb{R}^p$, the entry in row i and column j is the partial derivative of the i'th component f_i with respect to x_j, $\frac{\partial f_i}{\partial x_j}$, evaluated at x.

In many cases it is not practicable to calculate all these partial derivatives, and you have to make use of the notion of the derivative itself, as a linear map. The simplest way to describe the derivative $d_x f$ is to give a formula for $d_x f(\hat{x})$, where x is the point we are calulating the derivative and $\hat{x}$ is the vector to which we are applying it. It is really important that you have a good understanding of this notion. It is one of those ideas that ultimately makes mathematicss simpler rather than more complicated, although it may cost some effort to master it at first. It ties together into a single concept all $n \times p$ first order partials of f, and, for example, makes remembering the chain rule absolutely straightforward:

Theorem 7.8. The chain rule *If $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^p$ are open sets, and $f : U \to V$ and $g : V \to \mathbb{R}^q$ are differentiable at x and at $f(x)$ respectively, then $g \circ f$ is differentiable at x and*

$$d_x(g \circ f) = d_{f(x)} g \circ d_x f.$$

The statement of the chain rule in terms of partial derivatives is then an easy consequence of (7.2) in Appendix A, whereas a statement of the chain rule purely in terms of partial derivatives is quite complicated and easy to memorise wrongly.

The derivative plays a central role in the subject of differentiable manifolds.

Exercise 7.9. 1) Find the derivative at (x,y) of the map $f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ sending (x,y) to the vector product $x \times y$. 'Find the derivative' means: give an explicit formula for $d_{(x,y)}f(a,b)$. Important: don't write everything in coordinates. Expand the expression $(x+ha)\times(y+hb)$ using the fact that the vector product is bilinear (i.e. linear in each of its arguments). Write $d_{(x,y)}f(a,b)$ as a sum of vector products.

2) Ditto, for the map $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ sending (x,y) to the scalar product $x \cdot y := x_1y_1 + \cdots + x_ny_n$. Important: write your answer as a sum of scalar products.

3) Find the derivative of the map $\mathrm{Mat}_{2\times 2}(\mathbb{R}) \to \mathbb{R}$ sending the 2×2 matrix M to $\det M$. Here $\mathrm{Mat}_{2\times 2}(\mathbb{R})$ is the vector space of all 2×2 matrices. Important: write $d_A \det(B)$ as a sum of determinants.

4) Find the derivative of the map $\det : \mathrm{Mat}_{3\times 3}(\mathbb{R}) \to \mathbb{R}$ sending the 3×3 matrix M to $\det M$. Essential: write $d_A \det(B)$ as a sum of determinants.

5) On the basis of (3) and (4), guess a formula for $d_A \det(B)$ for the general case where we consider $\det : \mathrm{Mat}_{n\times n}(\mathbb{R}) \to \mathbb{R}$. Can you prove your formula?

6) What is the kernel of $d_I \det$, where I is the identity matrix? If you've done (3) and (4) but not (5), do this exercise just for 2×2 matrices and 3×3 matrices, and then guess the answer for $n \times n$ matrices.

7) Let

$$p : \mathrm{Mat}_{n\times p}(\mathbb{R}) \times \mathrm{Mat}_{p\times q}(\mathbb{R}) \to \mathrm{Mat}_{n\times q}(\mathbb{R})$$

be the product map, $p(X,Y) = XY$. Write down a formula for

$$d_{(X,Y)}p(A,B).$$

8) Inside $\mathrm{Mat}_{n\times n}(\mathbb{R})$ is the open set $\mathrm{Gl}_n(\mathbb{R})$ consisting of invertible matrices. Consider the *inversion map* $i : \mathrm{Gl}_n(\mathbb{R}) \to \mathrm{Gl}_n(\mathbb{R})$ sending A to A^{-1}.
i) Find $d_I i$ when $n = 2$ (where I is the identity matrix). That is, find an expressions for $d_I i(A)$, where A is a matrix. *You must write your anwer in terms of matrices.* How does this compare with the case $n = 1$?
ii) What is the answer for $n = 3$? For general n? Guesswork, on the basis of examples, is as good a route as any for arriving at the answer, but of course a proof is eventually required.

iii) There is a very simple proof which makes clever use of the chain rule, by applying it to the composition

$$\mathrm{Gl}_n(\mathbb{R}) \xrightarrow{j} \mathrm{Gl}_n(\mathbb{R}) \times \mathrm{Gl}_n(\mathbb{R}) \xrightarrow{i\times\mathrm{id}} \mathrm{Gl}_n(\mathbb{R}) \times \mathrm{Gl}_n(\mathbb{R}) \xrightarrow{p} \mathrm{Gl}_n(\mathbb{R}), \tag{7.4}$$

where $j(X) = (X, X)$, $i \times \mathrm{id}$ is the map $(X, Y) \mapsto (X^{-1}, Y)$, and p is the product map of part (7) of this exercise. The composite of the maps in (7.4) maps everything to I,

$$X \mapsto (X, X) \xrightarrow{i\times 1} (X^{-1}, X) \xrightarrow{p} X^{-1}X = I$$

so its derivative at any point (matrix) $X \in \mathrm{Gl}_n(\mathbb{R})$ is zero. Use the chain rule, and your knowledge of $d_{(X,Y)}p$, to obtain (and indeed prove) a simple formula for $d_I i(A)$ and, more generally, for $d_X i(A)$. How does your answer to the last part compare with the case $n = 1$? What is the crucial difference?

8. Consider the map $f : \mathrm{Mat}_{n\times n}(\mathbb{R}) \to \mathrm{Mat}_{n\times n}\mathbb{R}$ sending $X \to X^3$. What is $d_X f(A)$? Be careful! Find your answer by *carefully* computing the limit (7.3).

9. The *exponential map* for $n \times n$ matrices is defined by the same formula as when $n = 1$:

$$\exp(X) = I + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots.$$

Find $d_X \exp(A)$. Find $d_I \exp(A)$.

At a more basic level, it is also extremely important that you have a clear idea of how to write down the matrix of a linear map $L : V \to W$ with respect to chosen bases of the vector spaces V and W.

8) Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ and $B \in \mathrm{Mat}_{p\times p}(\mathbb{R})$. Define linear maps

$$R_A : \mathrm{Mat}_{n\times p}(\mathbb{R}) \to \mathrm{Mat}_{n\times p}(\mathbb{R})$$

and

$$L_B : \mathrm{Mat}_{n\times p}(\mathbb{R}) \to \mathrm{Mat}_{n\times p}(\mathbb{R})$$

by $R_A(X) = XA$, $L_B(X) = BX$.
i) In the case where $n = 2$ and $p = 1$, choose a basis for $\mathrm{Mat}_{n\times p}(\mathbb{R})$ and write down the matrices of R_A and L_B with respect to this basis.
ii) Ditto when $n = p = 2$.
iii) What is the relation between the determinants of the matrices you have found in (i) and (ii) and $\det A$ and $\det B$?
iv) Find the answer to (iii) for general n, p.
v) Suppose that $\det A = 0$. What is the kernel of R_A? If $\det B = 0$, what is the kernel of L_B?

7.4 Appendix D: Quotient spaces in topology

"The circle S^1 is obtained from the interval $[0,1]$ by joining the end-points." What does this mean? In mathematics, gluing and joining are carried out by means of the quotient topology. If X is a topological space and $\sim$ is an equivalence relation on X, then the quotient of X by $\sim$, denoted $X/\sim$, is the topological space defined as follows:

- its points are the equivalence classes of points in X. Thus, equivalent points in X are "identified with one another" or "glued together" in $X/\sim$.
- the map sending $x \in X$ to its equivalence class $[x] \in X/\sim$ is denoted by $q : X \to X/\sim$.
- a subset $U \subset X/\sim$ is open if and only if $q^{-1}(U)$ is open in X.

Example 7.10. To glue the endpoints of $[0,1]$ to one another, we define a suitable equivalence relation. Simply declare

$$x \sim x' \quad \text{if} \quad \begin{cases} x = x', & \text{or} \\ x = 0, x' = 1, & \text{or} \\ x' = 0, x = 1 \end{cases}$$

This might seem artificial. It is! We are creating a topological space to suit our purpose.

The topology on $X/\sim$ (the "quotient topology") is defined precisely in order for the following proposition to hold:

Proposition 7.11. *Let $\sim$ be an equivalence relation on the topological space X and let $X/\sim$ be the quotient space. Then*

1. $q : X \to X/\sim$ *is continuous.*
2. *Suppose $f : X \to Y$ is a continuous map with the property that for all $x, x' \in X$, if $x \sim x'$ then $f(x) = f(x')$. Then there is a unique continuous map $\bar{f} : X/\sim \to Y$ such that the diagram*

$$\begin{tikzcd} X \arrow[d, "q"'] \arrow[dr, "f"] & \\ X/\sim \arrow[r, "\bar{f}"] & Y \end{tikzcd}$$

commutes — that is, such that $\bar{f} \circ q = f$. [14]

The condition we impose on f in order that $\bar{f}$ be defined is often summarised by saying that f *respects the equivalence relation* $\sim$. The existence of $\bar{f}$ is described by saying that f *passes to the quotient to define $\bar{f}$. The map $\bar{f}$ is the key object in this game.*

[14] We do not prove this here, though it's very easy. At this point it's more important to learn to use the quotient topology than to prove its properties. It's also often true that proving that an object or construction has certain properties becomes easier after one has become familiar with using it.

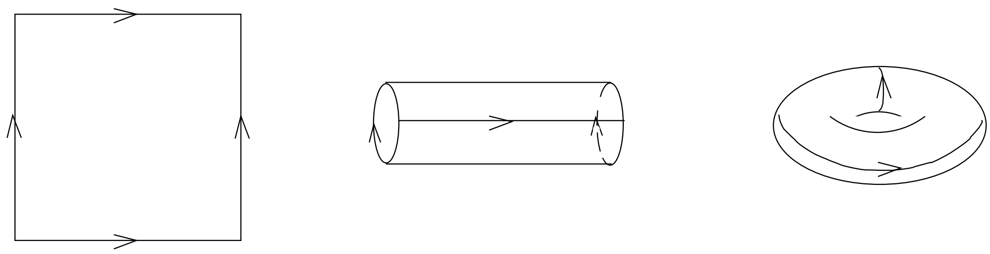
Exercise 7.12. Show that if there exists a map $\bar{f} : X/\sim \; \to Y$ such that $\bar{f} \circ q = f$, then f must respect the equivalence relation.

Example 7.13. In Example 7.10 we constructed the space $[0,1]/\sim$. Now we will see that it is homeomorphic to S^1. This is fundamental - but easy, because we have set things up right. First, define $f : [0,1] \to Y$, taking care that f respects the equivalence relation. In this case all that this requirement means is that $f(0) = f(1)$. A natural choice for f is the exponential map $f(x) = \exp(2i\pi x)$. Because f respects the equivalence relation, we have a continuous map $\bar{f} : X/\sim \; \to S^1$. Now we show that $\bar{f}$ is 1-1. Suppose that $\bar{f}([x_1]) = \bar{f}([x_2])$. Then $f(x_1) = f(x_2)$, i.e. $e^{2i\pi x_1} = e^{2i\pi x_2}$. Therefore $e^{2i\pi(x_1 - x_2)} = 1$, and so $x_1 - x_2$ is an integer. The only way this can happen when $x_1, x_2 \in [0,1]$ is if x_1 and x_2 are the opposite endpoints. Thus $[x_1] = [x_2]$, and $\bar{f}$ is 1-1.

Exercise 7.14. Suppose that $f : X \to Y$ respects the equivalence relation $\sim$ on X. Show that if f identifies (i.e. maps to the same image) *only* points which are equivalent to one another, then $\bar{f}$ is 1-1.

To conclude that $[0,1]/\sim$ is homeomorphic to S^1 is now easy. Because q is continuous, $q([0,1]/\sim)$ is compact. As $S^1 \subset \mathbb{R}^2$, it is Hausdorff. As f is surjective, so is $\bar{f}$. Thus $\bar{f} : [0,1]/\sim \; \to S^1$ is a continuous bijection from a compact space to a Hausdorff space. It follows that it is a homeomorphism[15]

Exercise 7.15. On the unit square $[0,1] \times [0,1]$, define an equivalence relation by identifying boundary points as indicated by the arrows in (i) of the following diagram.



(i) (ii) (iii)

— for each $y \in [0,1]$, the boundary point $(0,y)$ is identified with $(1,y)$, and for each $x \in [0,1]$ the boundary point $(x,0)$ is identified with $(x,1)$. Parts (ii) and (iii) of the diagram below give an intuitive image of how as a result of these identifications, the square becomes a torus: first we glue the upper and lower edges, forming a cylinder. In the process, the left and right hand edges have become the boundary circles of the cylinder. Then we glue these two circles, forming a torus.

[15]Since the domain is compact, any closed subset is compact too, so its image in the target is compact, and therefore closed (a consequence of the target's being Hausdorff). So the map sends closed sets to closed sets, and being 1-1, must send open sets to open sets. So its inverse is continuous.

To do: Show that the quotient $\big([0,1]\times[0,1]\big)/\sim$ is homeomorphic to the product $S^1\times S^1$. Hint: use the idea of Example 7.13.

Exercise 7.16. Let S_+ be the closed upper hemisphere in S^2. Define an equivalence relation $\sim_1$ on S_+ by identifying diametrically opposite points on the boundary. That is,

$$x \sim_1 x' \quad \text{if} \quad \begin{cases} x = x', \text{ or} \\ x, x' \in \partial S_+ \text{ and } x = -x' \end{cases}$$

Define an equivalence relation $\sim_2$ on the *whole* 2-sphere S^2 by

$$x \sim_2 x' \quad \text{if} \quad \begin{cases} x = x', \text{ or} \\ x = -x' \end{cases}$$

Show that the inclusion $i : S_+ \hookrightarrow S^2$ induces a homeomorphism $S_+/\sim_1 \;\to S^2/\sim_2$. Hint: the map we want is shown dashed in the diagram below:

$$\begin{tikzcd} S_+ \arrow[r, "i"] \arrow[d, "q_1"'] \arrow[dr, "q_2\circ i"] & S^2 \arrow[d, "q_2"] \\ S_+/\sim_1 \arrow[r, dashed] & S^2/\sim_2 \end{tikzcd}$$

Exercise 7.17. On the unit disc D, define an equivalence relation $\sim_3$ by the same recipe as $\sim_1$ in Exercise 7.16: diametrically opposite points on the boundary are identified with one another. Show that $D/\sim_3$ is homeomorphic to $S_+/\sim_1$ and thus to $S^2/\sim_2$.

Exercise 7.18. Let L_2 be the set of straight lines through the origin in $\mathbb{R}^3$. Show that there is a natural bijection $L_2 \to S^2/\sim_2$.

Exercise 7.19. (i) Show that the map $q : S^2 \to S^2/\sim_2$ is not only continuous but open.
(ii) Let X be a topological space and let G be a group acting continuously on X [16]. Define an equivalence relation on X by

$$x \sim x' \quad \text{if there exists } g \in G \text{ such that } g\cdot x = x'.$$

Show that $q : X \to X/\sim$ is not only continuous but open.

Exercise 7.20. Prove Proposition 7.11.

Underlying all of this discussion is the deeper issue of how we define structures on the objects we wish to study. One way, followed to a very large extent in these notes (in the first four chapters, to be precise), is to make use of already-existing structures by describing our objects as *sub-objects* of pre-existing and well understood objects. The manifolds studied in the first four chapters are all embedded in some $\mathbb{R}^n$, and our notion of smoothness relies on the pre-existing notion of the smoothness of a map defined on an open set in $\mathbb{R}^n$. By the

[16] That is, for each $g \in G$ there is a continuous map $X \to X$, (necessarily a homeomorphism), which we denote by $x \mapsto g\cdot x$, such that for all $x \in X$, and $g_1, g_2 \in G$, $g_2\cdot(g_1\cdot x) = (g_1 g_2)\cdot x$.

same token, in group theory is is often easier to study some group as a subgroup of a known group (e.g. as a subgroup of $\mathrm{Gl}_n(\mathbb{R})$ or of the group of isometries of the plane or 3-space) than as an "abstract" group, defined, for example, by means of generators and relations. This is why the theory of (linear) representations of groups is such a powerful tool in group theory.

The alternative approach, often regarded as more abstract, is to view the objects we want to describe as *quotients* of familiar objects. For example, $\mathbb{Z}/m$ is the quotient of $\mathbb{Z}$ by the equivalence relation in which $n \sim n'$ if $n - n'$ is a multiple of m. The definition of a group by means of generators and relations is another example of this approach: it defines a group as the quotient of the free group on the given generators by the normal subgroup determined by the given relations. Although sub-objects are usually easier to understand than quotients, definitions by means of quotients are sometimes simpler than definitions as sub-objects. For example, in Example 1.41(ii), we studied the image of the 2-sphere under the map to $\mathbb{R}^6$ defined by all the quadratic monomials. We showed that it is a smooth manifold, and is homeomorphic to the quotient of S^2 by the equivalence relation identifying antipodal points. Finding equations for this image, and showing that it is a manifold, is rather lengthy. It's much easier to think of it, at least as a topological space, as a quotient of S^2.

References

[1] R. Bott and L.Tu, *Differential forms in algebraic topology*, Graduate Texts in Math. 82, Springer Verlag, 1982

[2] Lawrence Conlon, *Differentiable Manifolds A First Course*, Birkhäuser Advanced Texts Vol. 5, Birkhäuser, Boston, 1993

[3] V.Guillemin and A.Pollack, *Differential Topology*, Prentice-Hall, 1974

[4] J. Harris, *Algebraic Geometry, a First Course* Springer Graduate Texts in Mathematics 133, Springer Verlag, 1992

[5] I.Madsen and J.Tornehave, *From calculus to cohomology*, Cambridge University Press, 1997.

[6] J.Milnor, *Topology from the differentiable viewpoint*, The University Press of Virginia, Charlottesville, 1965