DIFFERENTIAL GEOMETRY

John W. Barrett

Typeset by $\mathcal{AMS}$-TEX

Preface

This book is an introduction to aspects of differential geometry in an elementary and novel manner. The main idea is to introduce the concept of a manifold and apply it to the study of Lie groups. The book was developed as a set of lecture notes for a third-year undergraduate lecture course at the University of Nottingham. *Riemannian* geometry is not treated.[1]

The emphasis is on definitions and examples, and the abstract formalism has been reduced to a minimum. Most textbooks take the line that a differentiable manifold is a topological manifold with a differential structure. To eliminate bizarre examples, the topological space is required to be Hausdorff and have a countable base of open subsets. All this is extremely technical, and a theorem of Whitney asserts that any manifold of this type is isomorphic to a submanifold of a Euclidean space, $\mathbb{R}^n$ [1]. Moreover, most examples of manifolds occur naturally in this way.

The definition of a manifold used here is a locally Euclidean subset of $\mathbb{R}^n$. This clear and concise definition was given by John Milnor in a book which was the inspiration for the present work [2].

The sections on the special geometries, and the style generally was influenced by Elmer Rees' book on geometry [3]. Aside from the main development of the book, there are brief excursions into singularity and catastrophe theory, and into the subject of computer vision, based on a specialist book on the subject [4].

There are many exercises distributed in the text. The simpler examples are designed to be done immediately in the lecture room. Students find this gives an opportunity to review and digest what has just been said. Also it gives the lecturer a valuable opportunity for feedback. The longer exercises are there to change the learning from passive to active mode.

References

1. G. de Rham, *Differentiable Manifolds*, Grundlehren der mathematischen Wissenschaften 266, Springer, 1984.
2. J.W. Milnor, *Topology from the differentiable viewpoint*, University Press of Virginia, 1965.
3. E.G. Rees, *Notes on Geometry*, Universitext, Springer, 1983.

---

[1] It overlaps too much with general relativity.

4. J. Mundy and A. Zisserman, eds, *Geometric invariance in computer vision*, MIT Press, 1992.

1. Vector space geometry

Vector spaces. A vector space is a set V, with two operations: addition of vectors, and multiplication by scalars. In this book the scalars are always taken to be the real numbers, $\mathbb{R}$.

The first example of interest is the Euclidean space $V = \mathbb{R}^n$. This has a standard basis set of vectors

$$(1,0,0,\dots,0),(0,1,0,\dots,0),\dots,(0,0,0,\dots,1).$$

A second set of examples of vector spaces arises by taking V to be a linear subspace of $\mathbb{R}^k$. In this example, V does not have a standard, or uniquely specified, basis. Examples of this type turn out to be of major importance in differential geometry. Therefore it is important to understand the precise nature of the difference between the vector space $\mathbb{R}^n$, and these more general examples.

In general, a vector space possesses many different basis sets of vectors. Every basis contains the same number of elements, called the dimension of V.

Linear mappings of vector spaces $\phi\colon V \to W$ are those that preserve the structure, i.e., the addition of vectors and multiplication by scalars. An isomorphism is a linear mapping which has an inverse.

Suppose V is a vector space of dimension n. Choosing an ordered basis set of vectors $e_1, e_2, \dots e_n$ for V is the same thing as specifying a linear isomorphism $\mathbb{R}^n \to V$. Given the basis, the linear isomorphism is defined to be the linear map specified by mapping $(1,0,0,\dots,0) \mapsto e_1$, etc., … , $(0,0,0,\dots,1) \mapsto e_n$. This isomorphism $\mathbb{R}^n \to V$ can also be written

$$(a_1, a_2, \dots, a_n) \mapsto \sum_{i=1}^{n} a_i e_i.$$

The general linear group. Consider the set of all isomorphisms $\mathbb{R}^n \to \mathbb{R}^n$. This is a group, called the general linear group $\mathrm{GL}(n)$. Let us check that $\mathrm{GL}(n)$ is a group.

(1) The composition of two isomophisms ϕ_1 and ϕ_2 is the map $v \mapsto \phi_2(\phi_1(v))$. This linear map is called $\phi_2\phi_1$ and is also an isomorphism.
(2) The composition of maps is associative
(3) The identity mapping $e\colon \mathbb{R}^n \to \mathbb{R}^n$ is the identity element of the group.
(4) The inverse of an element $\phi \in \mathrm{GL}(n)$ is the inverse mapping, i.e.,

$$\phi\phi^{-1} = \phi^{-1}\phi = e.$$

A linear map $\mathbb{R}^n \to \mathbb{R}^n$ has a matrix, and conversely, an $n \times n$ matrix determines a linear map. The linear map determined by the matrix of numbers

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

takes the point $x = (x_1, x_2, \dots x_n) \in \mathbb{R}^n$ to $\phi(x) = (y_1, y_2 \dots y_n)$, where

$$y_i = \sum_{j=1}^{n} \phi_{ij} x_j, \qquad 1 \leq i \leq n.$$

So using this identification of linear maps on $\mathbb{R}^n$ with matrices, $\mathrm{GL}(n)$ can be thought of as the group of invertible $n \times n$ matrices.

Now suppose V is a general finite-dimensional vector space. Then there is an isomorphism $V \to \mathbb{R}^n$, but this is not unique. Suppose ϕ_1, ϕ_2 are any two isomorphisms. If $v \in V$, how are $x = \phi_1(v)$ and $y = \phi_2(v)$ related? Obviously $y = \phi_2 \phi_1^{-1}(x)$, i.e. they are related by a mapping in $\mathrm{GL}(n)$. Any element of $\mathrm{GL}(n)$ could arise in this way. Thus elements of the group $\mathrm{GL}(n)$ relate the *coordinate representations* $x = \phi_1(v)$ and $y = \phi_2(v)$ of V.

Notation. The general convention will be used that if $a \in \mathbb{R}^n$, then the letters $a_1, \dots a_n$ will be used for the coordinates of a, i.e.,

$$a = (a_1, a_2, \dots a_n).$$

Vectors (elements of $\mathbb{R}^n$) are generally written as horizontal row vectors. Sometimes they are written vertically, as column vectors. There is no difference in meaning intended; it is conventional to do this when multiplying a matrix with a vector.

Exercise 1. Define the map $y = \phi(x)$ by

$$(y_1, y_2) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad ad - bc \neq 0.$$

Which equation in y_1 and y_2 defines the image of the points satisfying $x_1^2 + x_2^2 = 1$ under the map ϕ?

Geometry. As a general principle: "The objects of a geometry are those which retain their form under the transformations being considered." Some examples:

(1) Circles are not objects of vector space geometry. As we saw in exercise 1, the image of a circle need not be a circle under a linear isomorphism.

(2) A linear subspace $S \subset \mathbb{R}^n$ has the defining property

$$s_1, s_2 \in S \Rightarrow \lambda s_1 + \mu s_2 \in S$$

This property is preserved under linear mappings ϕ, so the set $\phi(S) = \{\phi(x) \mid x \in S\}$ is a linear subspace if S is. Therefore linear subspaces are objects of vector space geometry.

Solution to exercise 1. The values of x_1 and x_2 for a given point (y_1, y_2) can be calculated using the inverse matrix:

$$(x_1, x_2) = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Substituting these values into the equation $x_1^2 + x_2^2 = 1$ gives

$$\frac{1}{(ad - bc)^2} \left((dy_1 - by_2)^2 + (-cy_1 + ay_2)^2 \right) = 1.$$

This is a more complicated equation than that for (x_1, x_2). In general it will not be a circle but will give an ellipse.

2. Affine Geometry

The affine geometry is the second of the 'special geometries' which we are going to study. As in vector space geometry, there is a group of transformations which characterises affine geometry.

Affine geometry is not too different to vector space geometry. In vector space geometry, the origin plays a distinguished role. However in many applications there are objects in Euclidean space $\mathbb{R}^n$ (e.g. ordinary 'space' of physics) but it does not really matter where the origin is.

Affine subsets of $\mathbb{R}^n$.

An affine subset $A \subset \mathbb{R}^n$ is a subset of $\mathbb{R}^n$ with the property that if $a, b \in A$, then $\lambda a + \mu b \in A$ for all $\lambda, \mu \in \mathbb{R}$ such that $\lambda + \mu = 1$. In other words, the straight line through any two points of A is also in A. Some examples:

(1) A linear subspace $L \in \mathbb{R}^n$. ($\lambda a + \mu b \in L$ without any condition on λ, μ).

(2) The set $x + L = \{x + l \mid l \in L\}$ for a linear subspace L. This follows from the calculation

$$\lambda(x + l_1) + \mu(x + l_2) = x + \lambda l_1 + \mu l_2 \in x + L.$$

The second example is more general than the first because $x+L$ need not contain the origin. It is in fact the most general type of example. All affine subsets are of the form of example 2. To show that this is true, fix an element $a \in A$ and define

$$TA_a = A - a = \{x - a \mid x \in A\}.$$

This is called the tangent space to A at a. What we have to show is that this tangent space is a linear subspace of $\mathbb{R}^n$.

Proof. Suppose $\lambda \in \mathbb{R}$ and $x - a \in TA_a$. Then the scalar multiple $\lambda(x-a) \in TA_a$ because

$$\lambda(x-a) + a = \lambda x + (1-\lambda)a \in A.$$

Now suppose $x - a$, $y - a \in TA_a$. Then the sum $(x-a)+(y-a)$ is in TA_a because

$$x - a + y - a + a = x + y - a = 2\left(\frac{x+y}{2}\right) - a \in A.$$

This shows that affine subsets are just 'linear subspaces with the origin shifted'. This readily suggests that if $a, b \in A$, then $TA_a = TA_b$. This is true because $b - a \in TA_a$, and so if $x - a \in TA_a$, so is

$$(x-a) - (b-a) = x - b.$$

Exercise 1. Show that for each $k \geq 1$,

$$\sum_{i=1}^{k} \lambda_i a_i \in A \qquad \text{if } a_i \in A, \text{ and } \sum_{1}^{k} \lambda_i = 1.$$

Affine maps. An affine map $\mathbb{R}^n \to \mathbb{R}^m$ is defined to be a map satisfying

$$f(\lambda a + \mu b) = \lambda f(a) + \mu f(b), \qquad \lambda + \mu = 1.$$

The affine maps preserve affine subsets, and in particular they map straight lines to straight lines.

If an affine map is invertible, it is called an affine isomorphism. The set of all affine isomophisms $\mathbb{R}^n$ to itself is a group, called the affine group $\mathrm{A}(n)$. This group characterises affine geometry.

It is possible to define a notion of an 'affine space' by giving axioms, in a similar way to the definition of a vector space, such that the affine subsets of $\mathbb{R}^n$ are examples. We do not need to go into this.

There are two important examples of affine isomorphisms of $\mathbb{R}^n$

(1) translations $x \mapsto x + c$, for a constant $c \in \mathbb{R}^n$
(2) linear isomorphisms in $\mathrm{GL}(n)$

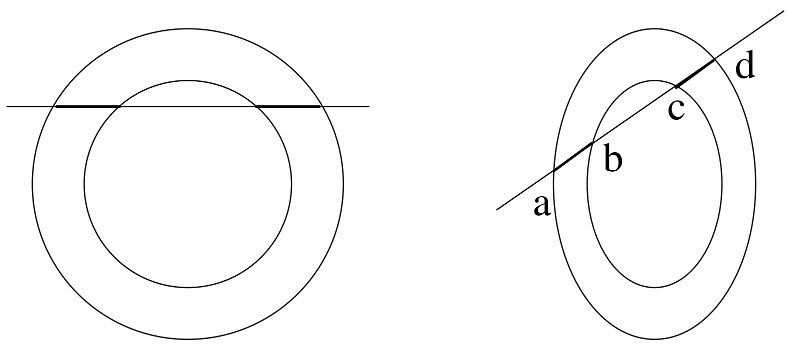
More examples can be made by combining these two. In fact, every affine map $\mathbb{R}^n \to \mathbb{R}^m$ is a linear map $\mathbb{R}^n \to \mathbb{R}^m$ composed with a translation of $\mathbb{R}^m$. The proof is an exercise:

Exercise 2. Prove that if f is an affine map, then the map $\mathrm{d}f$ defined by

$$x \mapsto f(x) - f(0)$$

is linear.

Example. The most general affine map $f\colon \mathbb{R} \to \mathbb{R}$ is given by the formula $f(x) = ax + b$. Then $\mathrm{d}f(x) = ax$. The coefficient a in this linear map is $\frac{\mathrm{d}f}{\mathrm{d}x}$, for any x.



Example. In the figure on the left, there are two concentric circles, $x_1^2 + x_2^2 = \text{constant}$. The line intersects the region between the circles in two segments; it is easy to see that the lengths of the two segments are equal. This is because the diagram is symmetrical about an axis, vertical in the diagram.

Newton wanted to know if the same property is true for the concentric ellipses shown on the right. These are given by $(x_1/e)^2 + x_2^2 = \text{constant}$ for some $e \in \mathbb{R}$. He proved that they are by applying a linear transformation L

$$(x_1, x_2) \mapsto (x_1/e, x_2)$$

to the right hand diagram. Since this transformation is affine, the line is mapped to a line. From the left-hand diagram

$$L(a) - L(b) = L(c) - L(d).$$

Since L is linear, $a - b = c - d$, and so the lengths of the two segments on the right are equal. Newton used this to show that there is no gravitational force inside an ellipsoidal shell of matter.

Solution to exercise 1. There are two strategies for this

(1) Induction. Assume it is true for $k-1$. Then

$$\sum_{i=1}^{k} \lambda_i a_i = (1-\lambda_k) \sum_{i=1}^{k-1} \frac{\lambda_i a_i}{1-\lambda_k} + \lambda_k a_k \in A.$$

The special case $\lambda_k = 1$ can be treated by picking another term to leave out of the sum, since not all λ_i can equal 1. The induction starts with $k = 2$ being the definition.

(2) Tangent space.

$$\sum \lambda_i a_i - b = \sum \lambda_i (a_i - b) \in TA_b,$$

since TA_b is a linear subspace. Hence $\sum \lambda_i a_i \in A$.

Solution to exercise 2.

$$\begin{aligned} \mathrm{d}f(\lambda x + \mu y) &= f(\lambda x + \mu y + (1-\lambda-\mu)0) - f(0) \\ &= \lambda f(x) + \mu f(y) + (1-\lambda-\mu) f(0) - f(0) \\ &= \lambda \mathrm{d}f(x) + \mu \mathrm{d}f(y) \end{aligned}$$

Exercise 3. Let $a, b, c \in \mathbb{R}^2$ be three points which are not collinear (do not lie on a line). Let A, B, C be any three points in $\mathbb{R}^n$. Work out how to define an affine map $\mathbb{R}^2 \to \mathbb{R}^n$ such that

$$a \mapsto A, \qquad b \mapsto B, \qquad c \mapsto C.$$

What can you say about the image of your map?

Exercise 4. Colours can be obtained by mixing red, green and blue in any desired proportions, which are given by numbers adding up to 1. Draw a diagram to represent the possible colours, and plot a point to represent 1/2 blue, 1/4 red, 1/4 green. Explain how the point is plotted.

3. Euclidean Geometry

Euclidean geometry is the study of figures in $\mathbb{R}^n$ using the concept of distance. The length of a vector is $|x| = (x_1^2 + x_2^2 + \ldots x_n^2)^{1/2}$, and the distance between two points x and y is defined to be $d(x,y) = |x-y|$. The latter concept does not require the origin as a distinguished point. To be more precise, the concept of distance is invariant under translations, $d(x+a, y+a) = d(x,y)$.

The space $\mathbb{R}^n$ with the distance function d is called n-dimensional Euclidean space. The 3-dimensional Euclidean space is very familiar as it is a mathematical model of space for the positions of objects in physics.

Euclidean space is a metric space. Mappings $f\colon \mathbb{R}^n \to \mathbb{R}^m$ which preserve the distance function,

$$d(f(x), f(y)) = d(x,y)$$

are called isometries. The motions of rigid bodies in physics are examples of isometries.

An isometry is injective, since $f(x) = f(y)$ implies

$$d(x,y) = d(f(x), f(y)) = 0,$$

which implies that $x = y$. Moreover,

Theorem. *An isometry $\mathbb{R}^n \to \mathbb{R}^m$ is an affine map*

Proof. Let p be an affine linear combination of $x, y \in \mathbb{R}^n$, i.e., $p = \lambda x + \mu y$, with $\lambda + \mu = 1$. Then p is the unique point such that

$$d(x,p) = |\mu| d(x,y) \quad \text{and} \quad d(y,p) = |\lambda| d(x,y).$$

If f is an isometry, then it follows that

$$d(f(x), f(p)) = |\mu| d(f(x), f(y)) \text{ and } d(f(y), f(p)) = |\lambda| d(f(x), f(y)),$$

which shows that $f(p) = \lambda f(x) + \mu f(y)$, i.e., f is affine.

From this theorem, one can easily deduce that any isometry $f\colon \mathbb{R}^n \to \mathbb{R}^n$ must be invertible. Recall that an affine map is a composition of the linear map $\mathrm{d}f$ with a translation. Since translations are always invertible isometries, it hinges on whether $\mathrm{d}f$ is invertible. However $\mathrm{d}f$ is linear and injective, so it must be invertible.

The Euclidean group. The set of all isometries of n-dimensional Euclidean space (to itself) is called the Euclidean group, $\mathrm{E}(n)$. It is a subgroup of $\mathrm{A}(n)$.

There are affine isomorphisms which are not isometries. For example, in 1 dimension, $f(x) = ax + b$ is not an isometry unless $a = \pm 1$. The linear transformation L in Newton's example above, or the transformation in exercise 1 of section 1 are examples in more than one dimension.

Clearly, a Euclidean transformation is a composition of a linear isometry followed by a translation. Hence it is important to characterise the linear isometries. Linear isometries preserve the square of the distance of a point from the origin. This quadratic form can be written in various ways:

$$d(x,0)^2 = |x|^2 = \sum_{i=1}^{n} x_i^2 = x \cdot x$$

Orthogonal group. An orthogonal transformation of $\mathbb{R}^n$ is a linear map $L\colon \mathbb{R}^n \to \mathbb{R}^n$ which preserves distances to the origin, i.e.

$$|L(x)| = |x| \quad \text{for all } x.$$

The set of all linear maps $L\colon \mathbb{R}^n \to \mathbb{R}^n$ which preserve the distance to the origin is a group called the orthogonal group, $\mathrm{O}(n)$.

For linear maps, preserving the distances to the origin actually implies that all distances are preserved, since $|L(x) - L(y)| = |L(x-y)| = |x-y|$. Thus orthogonal transformations are isometries (and hence invertible).

Polarization. Orthogonal transformations actually preserve all dot products. This follows from the 'polarization identity'

$$(x-y)\cdot(x-y) = x^2 + y^2 - 2x \cdot y$$

This shows that $L(x) \cdot L(y) = x \cdot y$.

Summary. We have seen in three cases, vector space geometry, affine geometry and Euclidean geometry, that the geometry is characterised by a group of transformations which preserves the structure of interest (such as linearity or distances).

"The properties of objects in a geometry are those which do not change under the transformations of the geometry."

Exercise 1. Let $x \neq y \in \mathbb{R}^2$. Draw a diagram to show the set of points p satisfying $d(x,p) = |\mu| d(x,y)$ and the set of points p satisfying $d(y,p) = |\lambda| d(x,y)$. Explain why there is only one point p satisfying both equations if $\lambda + \mu = 1$. Be sure to include the cases where λ or μ is negative.

Exercise 2. Let $L\colon \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal map. Write the relation $L(x) \cdot L(y) = x \cdot y$ in terms of the vector components x_i and y_i and the matrix L_{ij}

Show that this orthogonal matrix satisfies

$$\sum_{i=1}^{n} L_{ij} L_{ik} = \delta_{jk},$$

where δ_{ik} is the identity matrix. Conversely, check that any such matrix gives an orthogonal transformation.

Take the determinant of each side of this equation to find the possible values for $\det L$, and give an example of a 3×3 orthogonal matrix exhibiting each of these values.

Exercise 3. Show that all elements of $\mathrm{O}(2)$ are either

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

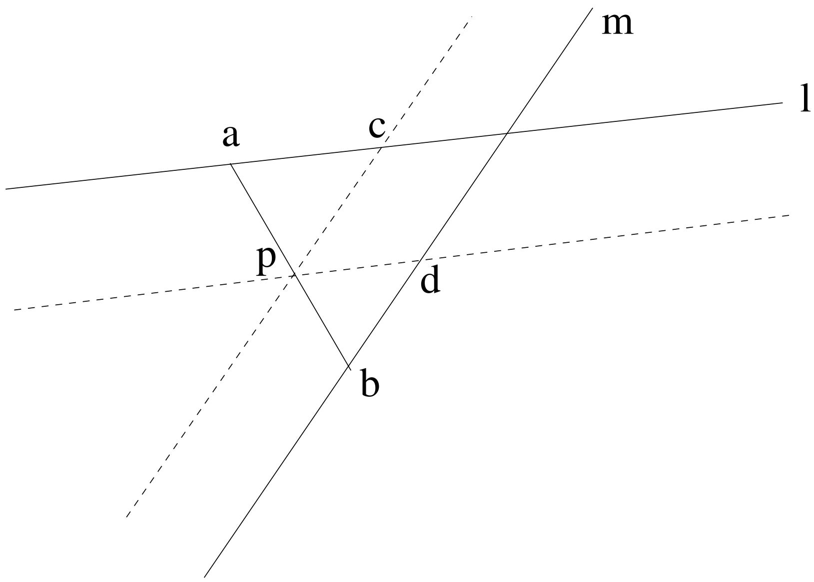
for some value of θ or have another form, which you should find.

Give a geometrical interpretation of both forms. Find an element of $\mathrm{GL}(2)$ with determinant 1 which is not in $\mathrm{O}(2)$.

4. The projective line

The last of the four special geometries studied here is projective geometry. The space for this projective geometry is projective space, which is distinct from Euclidean space, and there is one such space in each dimension. The one-dimensional space is called the projective line.

Motivation. The projective line can be introduced by considering projections, such as arise in perspective drawing. Consider two (straight) lines m and l in $\mathbb{R}^2$, which meet, and a point p which lies on neither line. The projection from line l to line m through the point p takes $a \mapsto b$, as shown in the diagram.



This does not quite define a map $l \to m$ because the point c, lying on a parallel to m through p, does not project onto m. However consider the map

$$l \cup \{\infty\} \to m \cup \{\infty\}$$

which takes

$$\begin{aligned} a &\mapsto b \qquad \text{as shown, for } a \neq c, \infty \\ c &\mapsto \infty \\ \infty &\mapsto d. \end{aligned}$$

This map is a bijection, the inverse being the projection from m to l. Infinity is to be regarded as an extra abstract point, which is added to each line in order to make the projection well-defined.

Example. A concrete example is given by considering p to be the origin, l to be the line $\lambda \mapsto \big(\lambda x_0, (1-\lambda) y_0\big)$, with x_0, y_0 not both 0, and m to be the line $\lambda' \mapsto (\lambda', 1)$, for $\lambda, \lambda' \in \mathbb{R}$.

The projection from l to m is given by

$$\big(\lambda x_0, (1-\lambda) y_0\big) \mapsto \left(\frac{\lambda}{(1-\lambda)} \frac{x_0}{y_0}, 1\right)$$

with also $(x_0, 0) \mapsto \infty$ and $\infty \mapsto (-x_0/y_0, 1)$.

The parameter for each line gives an identification of each line with $\mathbb{R}$. The projection takes parameter λ on line l to parameter

$$\lambda' = \frac{\lambda}{(1-\lambda)} \frac{x_0}{y_0}$$

on line m.

The definition. The projective line P^1 is defined to be the set of equivalence classes of the plane with the origin removed, $\mathbb{R}^2 \setminus \{0\}$, under the relation $v \sim \mu v$, for $\mu \neq 0 \in \mathbb{R}$. If a point $p \in P^1$ is the equivalence class of (x, y), this is written $p = [x : y]$, and is called the ratio of x and y.

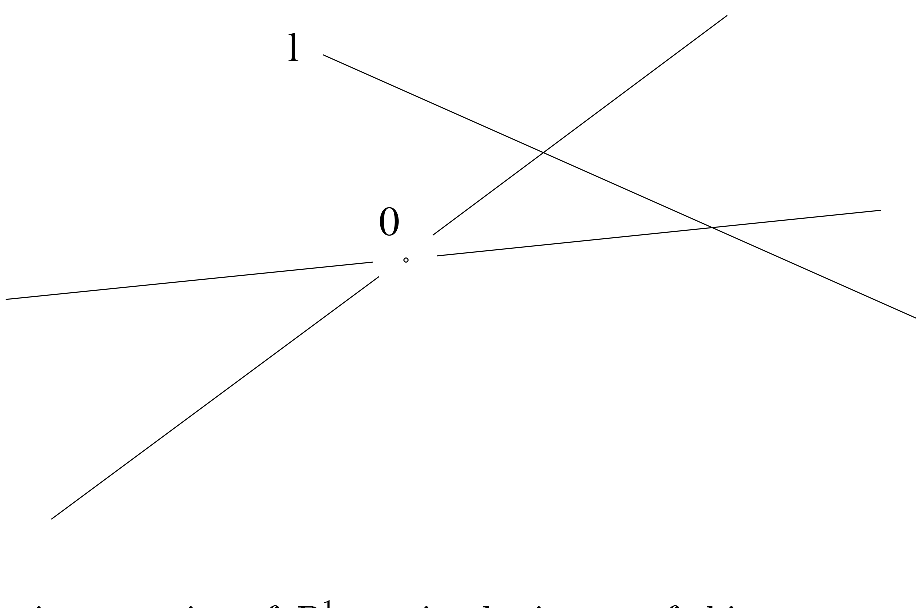
The pair of numbers (x, y) are also called homogeneous coordinates for the points $p \in P^1$. It is worth noting that these are not coordinates in the strict sense to be used later on, because (x, y) is not uniquely determined by a given point $p \in P^1$. However the term 'homogeneous coordinates' is standard in this situation, and so we shall use it.

Each point of the projective line can be identified with a unique line through the origin in $\mathbb{R}^2$, and this is a useful way of thinking of the projective line.

The projective line P^1 can be viewed as $\mathbb{R} \cup \{\infty\}$ in many different ways. Take any two linearly independent vectors u and v. Then the line $l \colon \mathbb{R} \to \mathbb{R}^2$

$$\lambda \mapsto \lambda u + (1 - \lambda) v$$

is not through the origin. This determines a map $\mathbb{R} \to \mathbb{R}^2 \to P^1$ by taking equivalence classes.



There is one point of P^1 not in the image of this map, namely the equivalence class of vectors parallel to l. This point is identified with ∞.

Projective transformations. Elements of GL(2) preserve the equivalence relation $v \sim \mu v$ in the definition of P^1, i.e. if $A \in \mathrm{GL}(2)$, then $Av \sim \mu Av$ is true whenever $v \sim \mu v$. Therefore A determines a transformation of P^1. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then the transformation is

$$[x : y] \mapsto [ax + by : cx + dy]$$

Inhomogeneous coordinates. As noted above, the homogeneous coordinates (x, y) of a point $[x : y]$ in P^1 are not uniquely determined. However the inhomogenous coordinate $z = x/y$ is uniquely determined by $[x : y]$. This is a real number if $y \neq 0$. If $y = 0$, it can be taken to be ∞.

Using the inhomogeneous coordinate to identify P^1 with $\mathbb{R} \cup \{\infty\}$ is the same as considering the identification determined by the line m given by $\lambda \mapsto (\lambda, 1)$ in the example above. This is because

$$[x : y] = [x/y : 1].$$

Using the inhomogeneous coordinate, the formula for a projective transformation becomes

$$\lambda = \frac{x}{y} \mapsto \frac{(ax + by)}{(cx + dy)} = \frac{(a\lambda + b)}{(c\lambda + d)}.$$

Taking the special case $a = x_0/y_0$, $b = 0$, $c = -1$, $d = 1$ gives the formula given in the example.

Question. Why should the formula in the example be a projective transformation? This point has not been made clear so far, but is worth thinking about.

Exercise 1. Which matrices $A \in \mathrm{GL}(2)$ give projective transformations for which $[1 : 0] \mapsto [1 : 0]$? Give the formula for the transformation using the inhomogeneous coordinate. What type of transformation of $\mathbb{R}$ is this? Determine all the projective transformations for which both $[0 : 1] \mapsto [0 : 1]$ and $[1 : 0] \mapsto [1 : 0]$.

5. Projective Geometry

The considerations for the projective line extend to the case of more than one dimension in a straightforward way.

The n-dimensional projective space P^n is defined to be the set of equivalence classes in $\mathbb{R}^{n+1} \setminus \{0\}$ under the relation $v \sim \mu v$, for $\mu \neq 0 \in \mathbb{R}$. The general linear transformations $A \in \mathrm{GL}(n + 1)$ give mappings of $\mathbb{R}^{n+1}$ which respect the equivalence relation, and so determine bijections of P^n. These are called projective transformations, and the set of all projective transformations is called the projective group, PGL$(n+1)$.

More than one element of the group GL$(n+1)$ gives rise to the same projective transformation. This means that PGL$(n+1)$ is not equal to GL$(n+1)$, but is a quotient group. Matrices that give the same projective transformation as $A \in$ GL$(n+1)$ are the scalar multiples, λA, for $\lambda \neq 0$.

A point

$$[x_1 : x_2 : \ldots : x_{n+1}] \in P^n$$

has homogeneous coordinates $(x_1, x_2, \ldots, x_{n+1})$, and inhomogeneous coordinates $(x_1/x_{n+1}, x_2/x_{n+1}, \ldots, x_n/x_{n+1})$ which are valid if $x_{n+1} \neq 0$. The points

$$[x_1 : x_2 : \ldots : x_n : 0]$$

are called 'points at infinity', and there are, for $n > 1$, more than one of them.

Example. In the projective plane P^2, the points $[1:0:0]$ and $[1:1:0]$ are both points at infinity, but $(1,0,0) \neq \mu(1,1,0)$ for any μ, so $[1:0:0] \neq [1:1:0]$.

Equations in projective geometry.

Historically, one of the reasons for the development of projective geometry was its use in simplifying equations. Consider for example the inhomogeneous equation

$$au^2 + buv + cv^2 + du + ev + f = 0$$

whose solutions are points $(u, v) \in \mathbb{R}^2$. This equation can be rewritten in terms of 3 variables by the substitution $u = x/z$, $v = y/z$ as

$$Q(x, y, z) = ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0.$$

This equation is homogeneous, meaning that all the terms are of the same degree. This implies that if (x, y, z) is a solution, then so is $(\mu x, \mu y, \mu z)$. Therefore (x, y, z) can be regarded as the homogeneous coordinates in projective space P^2, and the solutions define a subset of P^2.

The importance of writing the equation in this form is that projective transformations can be applied to homogenous equations, which allows the coefficients $(a, b, c, d, e, f$ in this example) to be simplified.

The function Q in the homogeneous equation is a quadratic form on $\mathbb{R}^3$. A projective transformation of P^2 is given by a linear transformation L of $\mathbb{R}^3$. If the image of (x, y, z) is $(x', y', z') = L(x, y, z)$ the equation for x, y, z can be reexpressed as an equation amongst the new variables x', y', z'. This new equation is also homogeneous of degree two, and is given by a new quadratic form Q' on $\mathbb{R}^3$ obtained from Q by a similarity transformation: $Q'(x', y', z') = Q(L^{-1}(x', y', z')) = 0$. Quadratic forms are similar to a *finite* number of canonical forms.

Example. If the quadratic form Q is positive definite, then it is similar to

$$x^2 + y^2 + z^2 = 0,$$

and the original equation in inhomogeous coordinates becomes

$$u^2 + v^2 + 1 = 0$$

after the projective transformation. In this case there are no solutions for $[x : y : z] \in P^2$.

Exercise 1. The points of $P^n = \{[x_1 : x_2 : \ldots : x_{n+1}]\}$ can be split into two disjoint subsets, as $x_{n+1} \neq 0$ or $x_{n+1} = 0$. Give a bijection of the first subset with $\mathbb{R}^n$. The second subset is called the 'points at infinity'. Find a bijection of the points of infinity with P^{n-1}.

Give a decomposition

$$P^n \cong \mathbb{R}^n \cup \mathbb{R}^{n-1} \cup \mathbb{R}^{n-2} \cup \ldots \mathbb{R}^1 \cup \mathbb{R}^0.$$

Exercise 2. Explain how the equation for a line in $\mathbb{R}^2$

$$ax_1 + bx_2 + c = 0$$

for constants $a, b, c \in \mathbb{R}$ can be written as an equation in P^2. How many points in P^2 solve the corresponding equation which do not correspond to solutions in $\mathbb{R}^2$?

Give the equations of two parallel lines in $\mathbb{R}^2$. Where do these lines meet in P^2?

Exercise 3. Show that the equation of the line joining two distinct points $[a_1 : b_1 : c_1]$, $[a_2 : b_2 : c_2]$ in P^2 is

$$\det \begin{pmatrix} x & y & z \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix} = 0.$$

6. Smooth functions

The functions we are mainly concerned with are those that can be differentiated. One could work with the set of all differentiable functions, but this is inconvenient, because the derivative of a differentiable function need not be itself differentiable. Therefore we shall work with functions which can be differentiated any number of times. These are called smooth functions.

Functions of one variable.

Definition. A function $f\colon \mathbb{R} \to \mathbb{R}$ is smooth if f and each derivative $\frac{\mathrm{d}^n f}{\mathrm{d}x^n}$ exists for every $x \in \mathbb{R}$ and is a continuous function of x.

Often you can show a function is smooth by giving the formula for all the derivatives.

Examples. Most familiar functions from calculus are smooth:

(1) Polynomials are smooth, as $\frac{\mathrm{d}^n f}{\mathrm{d}x^n} = 0$ for n greater than the order of the polynomial.
(2) The standard functions $\sin(x)$, $\cos(x)$, $\exp(x)$ etc., are smooth.
(3) More generally, an analytic function defined on $\mathbb{C}$ defines two smooth functions, its real and imaginary parts, when its domain of definition is restricted to the real axis.

Exercise 1. Are these functions smooth?

(1) $\sqrt{|x|} \qquad x \in \mathbb{R}$
(2) $\cos^{-1} x \qquad -1 < x < 1$
(3)
$$x \mapsto \begin{cases} e^{-1/x^2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

The second example illustrates a use of the inverse function theorem for one variable
$$\frac{\mathrm{d}y}{\mathrm{d}x} = 1/\left(\frac{\mathrm{d}x}{\mathrm{d}y}\right);$$
this will subsequently be generalised to $\mathbb{R}^n$. It is also worth noting that it makes sense to ask about the smoothness of a function defined only on a part of $\mathbb{R}$, namely the open interval $-1 < x < 1$.

The properties of differentiable functions carry over in a straightforward way to the case of smooth functions. Given smooth functions f, g, then the following are smooth:

(1) Linear combinations $x \mapsto \lambda f(x) + \mu g(x)$

(2) Product $fg: x \mapsto f(x)g(x)$
(3) Composite $f \circ g: x \mapsto f(g(x))$

The second of these is proved by using the Leibnitz rule, which gives a formula for the derivative:

$$\frac{\mathrm{d}(fg)}{\mathrm{d}x} = f\frac{\mathrm{d}g}{\mathrm{d}x} + g\frac{\mathrm{d}f}{\mathrm{d}x}.$$

This can be used to give a proof that fg is smooth if f and g are. Iterating Leibnitz rule gives

$$\frac{\mathrm{d}^n(fg)}{\mathrm{d}x^n} = \sum_{k=0}^{n} \binom{n}{k} \frac{\mathrm{d}^k f}{\mathrm{d}x^k} \frac{\mathrm{d}^{n-k} g}{\mathrm{d}x^{n-k}}.$$

The right-hand side is the sum of a product of continuous functions and so is continuous.

The third of these is similarly related to the chain rule, which gives the derivative of a function of a function. Let f be given as a function of y, g be given as a function of x.

$$\frac{\mathrm{d}(f \circ g)}{\mathrm{d}x} = \frac{\mathrm{d}f}{\mathrm{d}y}(g(x))\frac{\mathrm{d}g}{\mathrm{d}x}$$

The notation is potentially confusing. The formula $\frac{\mathrm{d}f}{\mathrm{d}y}(g(x))$ means write f as a function of y, differentiate it with respect to y, then substitute $g(x)$ everywhere for y. For example, if $f(y) = y^3 + y$ and $g(x) = \cos(x)$, then

$$\frac{\mathrm{d}f}{\mathrm{d}y} = y^2 + 1, \qquad \frac{\mathrm{d}f}{\mathrm{d}y}(g(x)) = g(x)^2 + 1 = \cos^2(x) + 1.$$

Functions of several variables.

If $f: \mathbb{R}^n \to \mathbb{R}^m$ is a function, then it has m component functions $f_i: \mathbb{R}^n \to \mathbb{R}$, $i = 1 \ldots m$, given by

$$f(x) = (f_1(x), f_2(x), \ldots, f_m(x)).$$

Definition. A function $f: \mathbb{R}^n \to \mathbb{R}^m$ is smooth if the components $(f_1, f_2, \ldots, f_m)$ of f and each of their partial derivatives

$$f_i, \frac{\partial f_i}{\partial x_j}, \frac{\partial^2 f_i}{\partial x_j \partial x_k}, \ldots \qquad \begin{array}{c} i = 1, \ldots m \\ j, k = 1, \ldots n \end{array}$$

exist and are continuous functions of x.

Exercise 2. Work out the partial derivatives for the components of a linear map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ given by

$$f_i\colon (x_1, \dots, x_n) \mapsto \sum_{j=1}^{n} L_{ij} x_j.$$

Show that a linear map is smooth.

Smooth functions can be combined in a similar way to the one-variable case, by linear combinations $\lambda f + \mu g$ of two functions f, g from $\mathbb{R}^n$ to $\mathbb{R}^m$, by the product fg of two functions $\mathbb{R}^n \to \mathbb{R}$, and by composition of $f\colon \mathbb{R}^n \to \mathbb{R}^m$ with $g\colon \mathbb{R}^m \to \mathbb{R}^k$ to give a smooth function $g \circ f\colon \mathbb{R}^n \to \mathbb{R}^k$.

Exercise 3. Which of the following functions $\mathbb{R}^2 \to \mathbb{R}$ are smooth?

(1) $\sqrt{x^2 + y^2}$
(2) $\sqrt{1 + x^2 + y^2}$
(3) $e^{y \sin x}$

The chain rule for several variables. First, consider a special case. Let $c\colon \mathbb{R} \to \mathbb{R}^m$ and $h\colon \mathbb{R}^m \to \mathbb{R}$ be smooth functions. Then the composite is

$$h \circ c\colon t \mapsto h\big(c_1(t), c_2(t), \dots, c_m(t)\big).$$

The chain rule for differentiating this is

$$\frac{\mathrm{d}(h \circ c)}{\mathrm{d}t} = \sum_{i=1}^{m} \frac{\partial h}{\partial x_i}\big(c(t)\big)\frac{dc_i}{\mathrm{d}t}.$$

This formula is a sum of a product of continuous functions and is therefore a continuous function. By repeated application of this formula, and the Leibnitz rule, one can arrive at a formula for the n-th derivative, thus showing that $h \circ c$ is smooth.

Exercise 4. Put $c(t) = (t^2, t^3)$, and suppose h is a smooth function $\mathbb{R}^2 \to \mathbb{R}$ such that $h(c(t)) = t$.

(1) Differentiate both sides of $h(c(t)) = t$, using the chain rule.
(2) Set $t = 0$. What does this tell you about h? How does this relate to the image of c?

The general case. If $g: \mathbb{R}^n \to \mathbb{R}^m$ is a smooth function of x and $f: \mathbb{R}^m \to \mathbb{R}^k$ a smooth function of y, then

$$\frac{\partial}{\partial x_j}(f \circ g)_i = \frac{\partial}{\partial x_j}(f_i \circ g) = \sum_{l=1}^{m} \frac{\partial f_i}{\partial y_l}(g(x)) \frac{\partial g_l}{\partial x_j}$$

This follows from setting $c(t) = g(x_1, \ldots, x_j + t, \ldots, \ldots, x_n)$, $h = f_i$, and computing

$$\frac{\mathrm{d}}{\mathrm{d}t}(h \circ c)$$

at $t = 0$. The formula can be interpreted as the multiplication of a $k \times m$ matrix with an $m \times n$ matrix.

$$\begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \cdots & & \cdots \\ \frac{\partial f_k}{\partial y_1} & \cdots & \frac{\partial f_k}{\partial y_m} \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \cdots & & \cdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix}$$

Exercise 5. Consider the functions

$$\begin{aligned} &g: U \to \mathbb{R}^2, \qquad U = \{(r,\theta) | r > 0\} \\ &\quad (r,\theta) \mapsto (r\cos\theta, r\sin\theta) \\ &f: \mathbb{R}^2 \to \mathbb{R}^2 \\ &\quad (x,y) \mapsto (x^2 - y^2, xy) \end{aligned}$$

(1) Give reasons to show that f and g are smooth.
(2) Calculate

$$\begin{pmatrix} \frac{\partial g_1}{\partial r} & \frac{\partial g_1}{\partial \theta} \\ \frac{\partial g_2}{\partial r} & \frac{\partial g_2}{\partial \theta} \end{pmatrix}, \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}$$

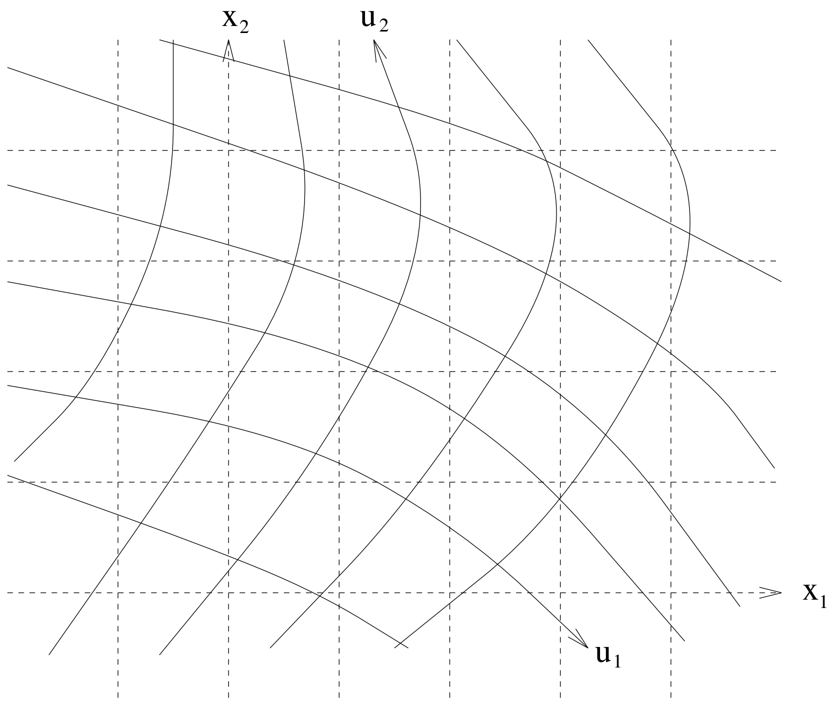
(3) Work out a formula for $f \circ g$, and its matrix of partial derivatives, without using the chain rule. Show the chain rule is satisfied by multiplying matrices.

Diffeomorphisms.

A smooth functions f with a smooth inverse f^{-1} is called a diffeomorphism. This satisfies

$$f \circ f^{-1} = \text{identity} \qquad \text{and} \qquad f^{-1} \circ f = \text{identity}.$$

It can be regarded as defining new coordinates. The figure shows the (x_1, x_2)-coordinate axes of $\mathbb{R}^2$ and the new axes for $(u_1, u_2) = f(x_1, x_2)$, for some diffeomorphism f. Geometric objects defined in terms of the old coordinates can be defined in terms of the new coordinates, and vice versa, using $(x_1, x_2) = f^{-1}(u_1, u_2)$.



Exercise 6. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism, with inverse g. Apply the chain rule to $f \circ g$. What can you say about the matrix of partial derivatives of f?

Solution to exercise 1.

(1) The function is not differentiable at $x = 0$:

$$\lim_{x \to 0} \frac{f(x) - f(0)}{x} = \lim \frac{\sqrt{|x|}}{x} = \pm\infty.$$

(2) If $x = \cos y$, then $\frac{\mathrm{d}x}{\mathrm{d}y} = -\sin y = -\sqrt{1 - x^2} \neq 0$. The inverse function theorem tells us that $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-1}{\sqrt{1-x^2}}$. This formula can be repeatedly differentiated to give formulae for the n-th derivative, since $x^2 < 1$.

(3) All the derivatives of the function $\exp(-1/x^2)$ converge to 0 as $x \to 0$, so the given function is smooth.

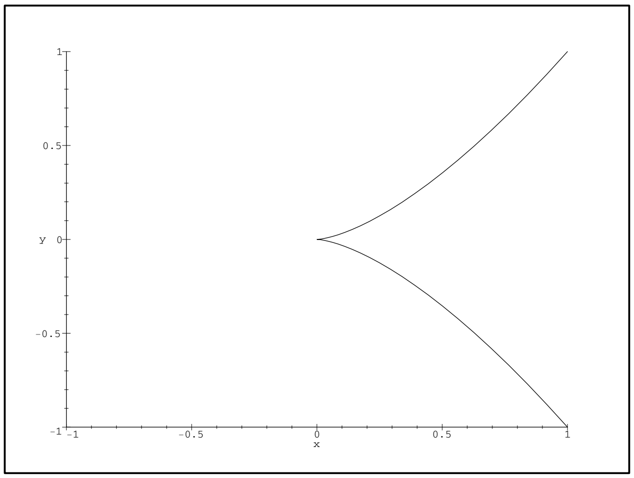
Solution to exercise 2. $\frac{\partial f_i}{\partial x_j} = L_{ij}$, a constant. All further derivatives are zero.

Solution to exercise 4.

(1)

$$2t\frac{\partial h}{\partial x_1} + 3t^2\frac{\partial h}{\partial x_2} = 1.$$

(2) Setting $t = 0$ gives $0 = 1$. Hence h cannot be smooth. (However a continuous h can be found. An example is $h(x_1, x_2) = \sqrt[3]{x_2}$.) The image of c has a cusp at $t = 0$, a point where the curve is not smooth. The curve is called the semicubical parabola.



7. The derivative

A function $f: \mathbb{R}^k \to \mathbb{R}^l$ has a derivative $\mathrm{d}f_x$ at a given point $x \in \mathbb{R}^k$ if there is a linear map $\mathrm{d}f_x: \mathbb{R}^k \to \mathbb{R}^l$ and a continuous function $\epsilon: \mathbb{R}^k \to \mathbb{R}^l$ such that

$$f(x+h) - f(x) = \mathrm{d}f_x(h) + |h|\epsilon(x+h),$$

and

$$\epsilon(x) = 0.$$

The formula can be solved for ϵ as long as $h \neq 0$, and this formula for ϵ is continuous where $h \neq 0$ (assuming f is). So the crucial point of the definition is that the error ϵ converges to 0 as $h \to 0$.

Lemma. *The derivative is unique.*

Proof. Suppose another linear map L also satisfies the definition, with error ϵ'. Then

$$\mathrm{d}f_x(h) - L(h) = |h|\big(\epsilon(x+h) - \epsilon'(x+h)\big).$$

Consider $h = ty$ for a fixed vector y and $t \to 0 \in \mathbb{R}$. Then this becomes

$$t\big(\mathrm{d}f_x(y) - L(y)\big) = t|y|\big(\epsilon(x+ty) - \epsilon'(x+ty)\big).$$

For $t \neq 0$, the number t cancels on both sides, giving

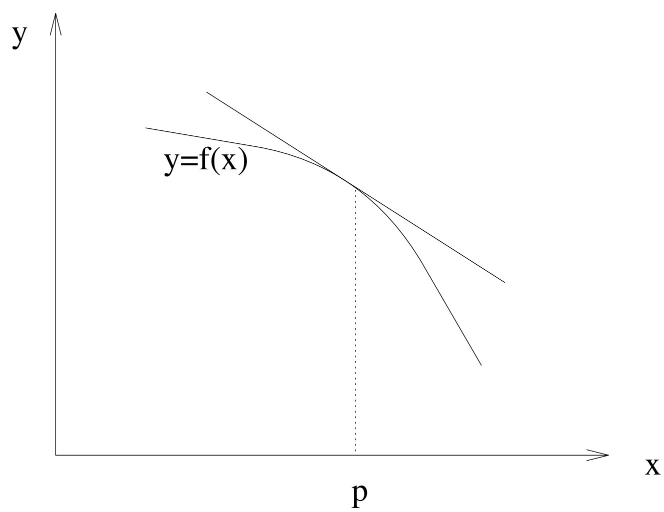
$$\mathrm{d}f_x(y) - L(y) = |y|\big(\epsilon(x+ty) - \epsilon'(x+ty)\big).$$

But as ϵ and ϵ' are continuous, this equation is also true at $t = 0$. At $t = 0$, the right-hand side is zero. It follows that $\mathrm{d}f_x = L$.

For a fixed point p, the map $\mathbb{R}^k \to \mathbb{R}^l$

$$x \mapsto f(p) + \mathrm{d}f_p(x-p)$$

is an affine map which approximates f at p. Clearly, if f is itself affine, then the approximation is exact, and the derivative is the linear map $\mathrm{d}f$ defined earlier for affine maps, and does not depend on p. However if f is not affine, then the linear map $\mathrm{d}f_p$ does vary with p.



The affine approximation to f at p

Exercise 1. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Use the definition to show that $\mathrm{d}f_x = f$, for any x.

Let $h\colon \mathbb{R}^n \to \mathbb{R}^m$ be an affine map. Show that $\mathrm{d}h_x = \mathrm{d}h_y$, for any $x, y \in \mathbb{R}^n$, and that this linear map coincides with the map named $\mathrm{d}h$ in the section on affine maps.

Lemma. *If $f\colon \mathbb{R}^k \to \mathbb{R}^l$ is a smooth function, then the derivative df_x exists for every x, and has as its matrix the partial derivatives*

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \cdots & & \cdots \\ \frac{\partial f_l}{\partial x_1} & \cdots & \frac{\partial f_l}{\partial x_k} \end{pmatrix}.$$

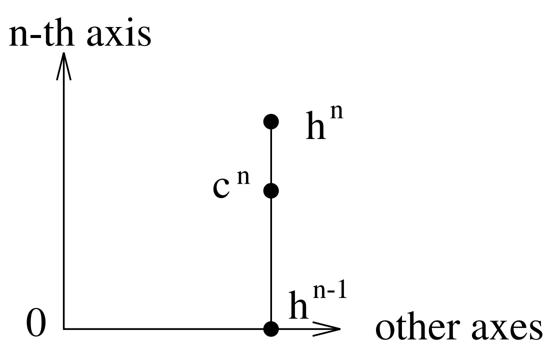
Proof. Given $h \in \mathbb{R}^k$, define a sequence of vectors

$$\begin{aligned} h^0 &= 0 \\ h^1 &= (h_1, 0, 0, \ldots, 0) \\ h^2 &= (h_1, h_2, 0, \ldots, 0) \\ &\vdots \\ h^k &= (h_1, h_2, h_3, \ldots, h_k) = h \end{aligned}$$

The superscript is used to distinguish different vectors h^n, as distinct from the subscript, which denotes the components h_n of h as usual. The sequence of vectors is constructed so that it interpolates between 0 and h, with each successive pair h^n, h^{n-1} differing only in one coordinate, h_n. Then, using the mean value theorem for this coordinate,

$$f_i(x + h^n) - f_i(x + h^{n-1}) = h_n \frac{\partial f_i}{\partial x_n}(x + c^n),$$

where the vector c^n lies on the line between h^n and h^{n-1}.



Summing this equation over n gives

$$f_i(x + h) - f_i(x) = \sum_{n=1}^{k} h_n \frac{\partial f_i}{\partial x_n}(x + c^n).$$

This expression gives the i-th component of a vector obtained by a matrix of partial derivatives acting on the vector h. The partial derivatives are evaluated at points $x + c^n$.

For $|h| \neq 0$, define $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_k)$ by the equation

$$(*) \qquad f_i(x + h) - f_i(x) = \sum_{n=1}^{k} h_n \frac{\partial f_i}{\partial x_n}(x) + |h|\epsilon_i.$$

Then

$$\epsilon_i = \sum_n \frac{h_n}{|h|}\left(\frac{\partial f_i}{\partial x_n}(x+c^n) - \frac{\partial f_i}{\partial x_n}(x)\right)$$

which converges to zero as $h \to 0$, because the difference of partial derivatives converges to zero, and $-1 \le h_n/|h| \le 1$. This shows that the vector ϵ is continuous if $\epsilon(x)$ is defined to be zero.

Equation $(*)$ is just the i-th component of the equation defining the derivative, with the linear map $\mathrm{d}f_x$ given by the matrix of partial derivatives at x.

Chain rule for derivatives. In the light of this lemma, results about the partial derivatives can be transcribed into the new notation. The chain rule becomes

$$\mathrm{d}(f \circ g)_x = \mathrm{d}f_{g(x)} \circ \mathrm{d}g_x.$$

Exercise 2. Let $f\colon \mathbb{R} \to \mathbb{R}$. What is $\mathrm{d}f_1(2)$ in the usual $\frac{\mathrm{d}f}{\mathrm{d}x}$ notation?

Exercise 3. Let $f(x,y) = (x^2+y^2, x^2-y^2)$. Calculate $\mathrm{d}f_{(x,y)}$ using partial derivatives. Work out a formula for ϵ in the definition of the derivative of f, and verify that $\epsilon \to 0$ as $h \to 0$.

Exercise 4. Give a proof of the chain rule by applying the definition of the derivative to $\mathrm{d}(f \circ g)_x$, $\mathrm{d}f_{g(x)}$ and $\mathrm{d}g_x$, and manipulating the three error terms.

8. Open sets

Quite often we need to define smooth mappings not on the whole of $\mathbb{R}^n$, but on certain subsets. For example $x \mapsto \cos x$ is a diffeomorphism of the open interval $0 < x < \pi$ to the open interval $-1 < y < 1$.

Another requirement is to discuss the behaviour of a function is a 'small region around a point'.

Balls. In Euclidean space $\mathbb{R}^n$, the distance $d(x,y) = |x-y|$ can be used to say when points are close. If r is a number greater than zero, then the ball of radius r at $x \in \mathbb{R}^n$ is defined to be

$$B_r(x) = \{y \in \mathbb{R}^n \mid d(x,y) < r\}$$

The ball is the subset of points closer to x than the radius r.

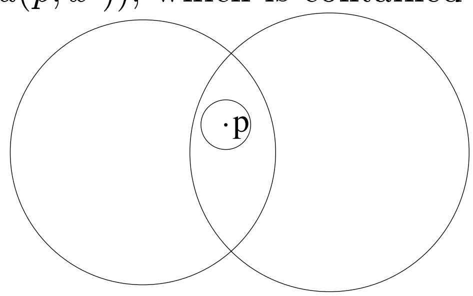
However, as we have already seen, there are many maps of interest, affine maps for example, which do not preserve distances. Therefore it is necessary to have adopt a sophisticated approach to the idea of closeness of points.

Open subsets of $\mathbb{R}^n$. An open subset P of $\mathbb{R}^n$ is defined to be a subset which is a union of balls in $\mathbb{R}^n$, $B_r(x)$, for points $x \in \mathbb{R}^n$, and radii $r > 0$. Note that the union can be over any set of balls, not restricted to finite or countable, and of course x and r vary.

Examples.

(1) A ball
(2) The intersection of two balls
(3) The empty set
(4) The whole of $\mathbb{R}^n$

The second example is an open subset for the following reason. If $p \in P = B_r(x) \cap B_{r'}(x')$, then p is the centre of a ball of radius $\min(r - d(p, x), r' - d(p, x'))$, which is contained in P.



So P is the union of all the balls constructed in this way for each $p \in P$.

A point $y \in Y$ is called an interior point of Y if there is an r such that $B_r(y) \subset Y$. The set Y is an open set if and only if every point of Y is interior.

Exercise 1. Which of the following subsets of the plane $\mathbb{R}^2$ are open sets? Points (x, y) such that:

(1) $x \geq 0$
(2) $x > 0$
(3) $x + 2y = 0$
(4) $x^2 > y^3$
(5) $x^n + y^n > 1$ for every positive even integer n.

If $U \subset \mathbb{R}^n$ is an open subset, then a function defined on U, rather than the whole of $\mathbb{R}^n$, can be differentiated in just the same way for any point $x \in U$. For example, to define a partial derivative of $f\colon U \to \mathbb{R}$,

$$\frac{\partial f}{\partial x_j} = \lim_{h \to 0} \frac{f(x_1, \dots, x_j + h, \dots, x_n) - f(x)}{h}$$

the function f needs to be evaluated at points close to x along a line parallel to the x_j axis. Such points are in the set U, because it contains the whole of a ball centred on x.

This means that all of the previous definitions for differentiation also make sense for functions defined on an open subset of $\mathbb{R}^n$.

Example. The function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $(x, y) \mapsto (x^2, y)$ is a smooth map but is not a diffeomorphism. However if U is the open subset of $\mathbb{R}^2$ defined by $x > 0$, then the same formula gives a diffeomorphism $U \to U$. The inverse function is $(x, y) \mapsto (\sqrt{x}, y)$.

Open subsets of $X \subset \mathbb{R}^n$. Now let X be a subset of $\mathbb{R}^n$, not necessarily an open subset. For example, X could be an affine subset of a lower dimension, or a manifold (see below). An open subset of X is simply any set of the form $X \cap U$, where U is an open subset of $\mathbb{R}^n$.

Now consider $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^k$ and $f\colon X \to Y$ a continuous mapping. If $V \subset Y$, its inverse image, $f^{-1}(V)$ is the set of all points $x \in X$ which map into V.

Continuous mappings respect open sets in the following way:

Proposition. *For $f\colon X \to Y$, the inverse image of an open subset $V \subset Y$ is an open subset of X.*

The situation is simpler if f has a continuous inverse mapping f^{-1}.

Proposition. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^k$. If the continuous function $f\colon X \to Y$ has a continuous inverse f^{-1}, then $U \subset X$ is an open subset of X if and only if $f(U) \subset Y$ is an open subset of Y.*

9. Inverse function theorem

Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism. Then the derivative $\mathrm{d}f_x$ is an invertible linear map for each x. Is the converse true? Suppose f is a smooth map and $\mathrm{d}f_x$ is invertible for all x. Is it a diffeomorphism?

If $n = 1$, this is true. The inverse function theorem for one variable is

Theorem. *If $f\colon \mathbb{R} \to \mathbb{R}$ has non-zero derivative $\frac{\mathrm{d}f}{\mathrm{d}x}$ for every x, then the inverse function exists and is differentiable.*

It is not too hard to extend this to proving that the inverse is smooth.

If $n > 1$, it is not true. Consider the mapping

$$g(x, y) = (e^x \cos y, e^x \sin y).$$

This has derivative

$$\mathrm{d}g_{(x,y)} = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

which is invertible. However, $g(x, y) = g(x, y + 2\pi)$.

This shows the inverse can not exist on the whole of $\mathbb{R}^2$. However, the idea of the definition of the derivative is that a smooth function f is approximated near a point x by an affine map. If x is a regular point, this affine map is invertible, so we expect the original function f to be invertible when its domain and range are restricted to some sufficiently small regions around x and $f(x)$.

Inverse Function Theorem. *If $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^n$ are open subsets, $f\colon U \to V$ is smooth, $x \in U$ and df_x is invertible, then there are open subsets $U' \subset U$, $V' \subset V$ such that $x \in U'$ and f restricted to U' is a diffeomorphism to V'.*

Exercise 1. Are the following functions $\mathbb{R}^2 \to \mathbb{R}^2$ diffeomorphisms? If not, is there an open set containing the origin on which the function is a diffeomophism to its image? Give the reasons for your answers.

(1) $(x, y) \mapsto (x + y^3, y)$
(2) $(x, y) \mapsto (x + x^3, x)$
(3) $(x, y) \mapsto (x + x^3, y)$
(4) $(x, y) \mapsto (x^2 + yx, y)$
(5) $(x, y) \mapsto (x^2 + yx + x, y)$

Exercise 2. Consider $g(x, y) = (e^x \cos y, e^x \sin y)$. Verify that $\mathrm{d}g_{(x,y)}$ is invertible for all $(x, y) \in \mathbb{R}^2$. State how the inverse function theorem applies to the behaviour of g near to the point (x, y). Now verify that the inverse function theorem is true in this case by giving explicit formulae and a domain for the inverse function.

10. Catastrophe theory

Let f be a smooth function from $\mathbb{R}^n$ to $\mathbb{R}^n$. The points $x \in \mathbb{R}^n$ for which the linear map $\mathrm{d}f_x$ is not invertible are called singular points of f. The corresponding $y = f(x)$ are called singular values of f. Points which are not singular are called regular points, and values of y which are not singular values are called regular values of f.

We shall study two basic examples in $\mathbb{R}^2$.

(1) The standard fold

$$(y_1, y_2) = f(x_1, x_2) = (x_1^2, x_2).$$

The matrix of partial derivatives is

$$\begin{pmatrix} 2x_1 & 0 \\ 0 & 1 \end{pmatrix}$$

This is non-invertible (singular) for $x_1 = 0$, giving the singular points. The singular values are on the line $y_1 = 0$. The behaviour of f can be plotted in the y_1-y_2 plane. For each (y_1, y_2), the number of points (x_1, x_2) such that $(y_1, y_2) = f(x_1, x_2)$ are plotted. It is noteworthy that the singular values lie where the number changes. The resulting diagram is called a bifurcation diagram.

(2) The standard cusp

$$(y_1, y_2) = g(x_1, x_2) = (x_1^3 + x_2 x_1, x_2).$$

Exercise 1. Calculate the determinant of the matrix of partial derivatives of the standard cusp g to find an equation in (x_1, x_2) for the singular points of g, and an equation in (y_1, y_2) for the singular values of g.

Verify that the number of points mapping to (y_1, y_2) is the number of roots of the cubic

$$t^3 + y_2 t = y_1.$$

Compute the stationary points of $h(t) = t^3 + y_2 t$, and draw its graph for $y_2 < 0$, $y_2 = 0$ and $y_2 > 0$. For each of the three cases, work out the ranges of y_1 for which $h(t) = y_1$ has one, two or three solutions.

Draw the bifurcation diagram for g.

Exercise 2. Give an example to show that a function can map a regular point to a singular value.

If $g\colon \mathbb{R}^2 \to \mathbb{R}^2$ is the standard cusp, with the coordinates named as $y = g(x)$, then using new coordinates $x' = h^{-1}(x)$ results in a more general form of cusp, $y = g(h(x'))$, with the same bifurcation diagram but different regular points. This follows from the chain rule: the matrix

$$\frac{\partial (g \circ h)_i}{\partial x'_j}$$

is singular at a point x' if and only if the matrix

$$\frac{\partial g_i}{\partial x_j}$$

is at the corresponding value $x = h(x')$. Similarly, one can change the y coordinates, by $y' = l(y)$ which results in the cusp $y' = l(g(x))$ with different singular values, but corresponding smoothly with the singular values for the standard cusp g.

Whitney's theorem. Whitney showed that a generic smooth function $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ has singular points which are all either cusps or folds, expressed in some coordinates x', y', which are related to the standard x, y coordinates of the standard cusp or fold by a diffeomorphism.

The adjective generic refers to a 'typical' function F. This means that if a function f does not obey Whitney's theorem, then there is a function ϵ with arbitrarily small values, such that $F = f + \epsilon$ does.

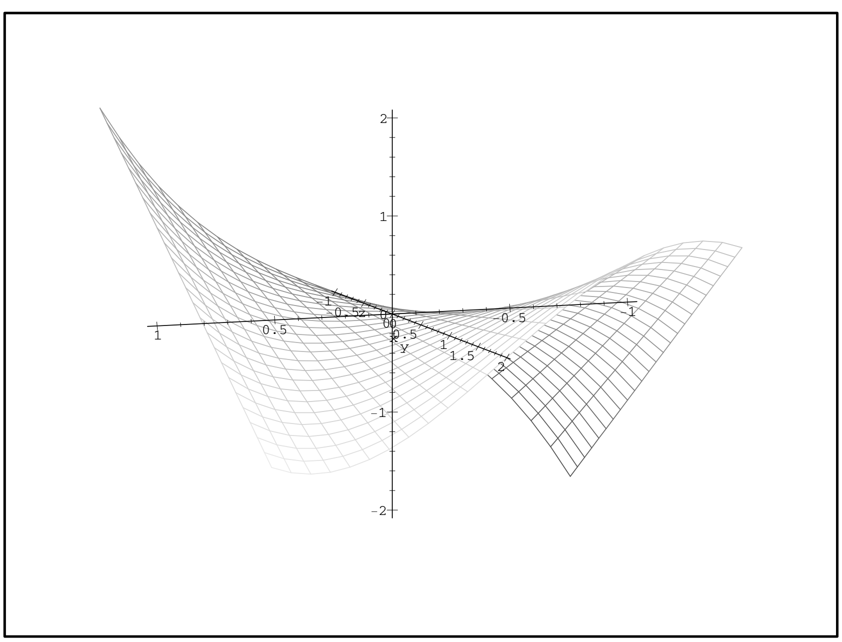
This result is fairly difficult to state precisely, and also hard to prove. We shall just explore its content with examples and applications of the idea.

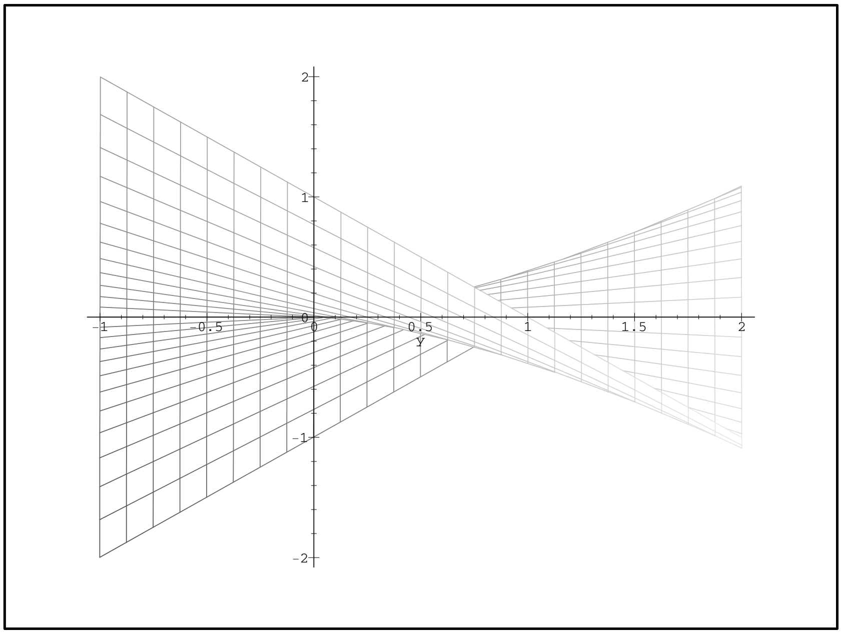
Surface projections. The standard fold and cusp can be viewed as projections of smooth surfaces in three-dimensional space. The fold and cusp are the composite mappings $\mathbb{R}^2 \to \mathbb{R}^3 \to \mathbb{R}^2$

$$(x_1, x_2) \to (x_1^2, x_2, x_1) \to (x_1^2, x_2)$$

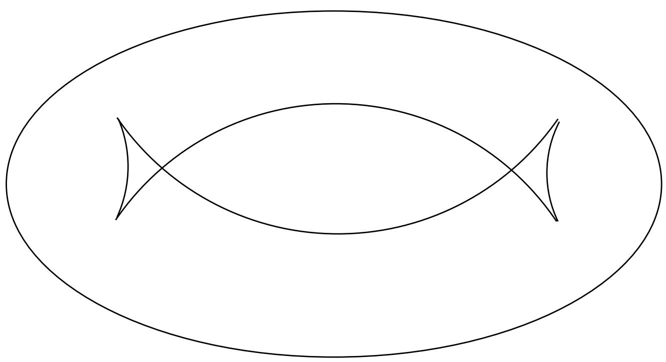
$$(x_1, x_2) \to (x_1^3 + x_2 x_1, x_2, x_1) \to (x_1^3 + x_2 x_1, x_2)$$

The first map in each line parameterises a smooth surface in three-dimensional space, and the second map projects it onto the plane $\mathbb{R}^2$ by ignoring the third coordinate value, as one would see by viewing the surface from a direction along the third coordinate axis. Here are two views of the cusp surface in $\mathbb{R}^3$. The second view is the projection given above.

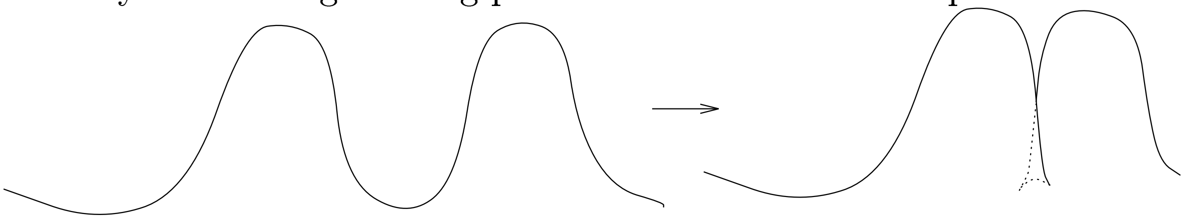




According to Whitney's theorem, the projection of any surface to the plane will generically have singularities of these types. This can be seen by examples, or by experimenting with real objects. The figure shows the projection of a glass torus, in which four cusp points can be seen, with the lines corresponding to folds.



Exercise 3. List a number of qualitatively different ways that the pattern of folds and cusps can change when you rotate an object. Hint: start by considering walking past a camel with two humps.



Other patterns of change can be obtained by viewing the camel humps in different ways.

Exercise 4. Consider the stationary points of the function

$$V(x) = \frac{1}{4}x^4 + \frac{1}{2}ax^2 + bx$$

for constants $a, b \in \mathbb{R}$. How does the number of these vary as the parameters a, b vary? Draw the graph of V for each qualitatively different set of parameters (a, b).

Find the equation satisfied by the stationary points of V. How do these values (a, b, x) relate to the standard cusp? The minimum points can be thought of as places where a ball rolling in the potential $V(x)$ comes to rest. Suppose the ball sits at a minimum point, and the parameters (a, b) are varied slowly and smoothly.

How do the minima of V behave as (a, b) vary? Answer this question by drawing various possible trajectories for a curve in the a-b plane (the bifurcation diagram). Assume that if the minimum point at which the ball sits disappears, then the ball jumps to a new minimum point by rolling downhill.

Show how the following phenomena occur

(1) (Catastrophes) Discontinuous jumps in the position of the ball (as just described).
(2) (Memory) The position of the ball depends on its past history as well as the values of (a, b).
(3) (Hysteresis) Reversing the path of the parameters (a, b) does not reverse the path of the ball.
(4) (Divergence) The final position of the ball depends not only on the initial position and the final parameters, but also on the path taken by the parameters in the a-b plane

Exercise 5. Use the library to find practical applications of the cusp catastrophe described in the previous question. Which of the described phenonema occur in your examples?

11. Manifolds

Examples of surfaces. Some examples of surfaces have already been used, for example in the discussion of the cusp singularity. If $f\colon \mathbb{R}^2 \to \mathbb{R}$ is a smooth function, then the graph of f, namely the set of points $(x, y, z) \in \mathbb{R}^3$ such that $z = f(x, y)$, is a surface. Let us name this set M.

The set M has parameters and coordinates, defined as follows. The function $\pi\colon \mathbb{R}^2 \to M$ defined by $(x, y) \mapsto \bigl(x, y, f(x, y)\bigr)$ is called a parameterisation of M. The function $\phi\colon M \to \mathbb{R}^2$ given by

$$\bigl(x, y, f(x, y)\bigr) \mapsto (x, y)$$

is called a coordinate function (or just coordinates) for M. Each component of this function is called a coordinate, i.e., the functions giving the value of x or y. The parameterisation and coordinate functions just defined are inverses of each other,

$$\mathbb{R}^2 \xrightarrow{\pi} M \xrightarrow{\phi} \mathbb{R}^2$$

is the identity map.

There are more general surfaces than graphs of functions, however. Just one example will suffice for now: the sphere $S^2 \subset \mathbb{R}^3$ defined by

$$\{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}.$$

Clearly, the sphere is not the graph of any function. Also, the sphere does not have a parameterisation in the same way; there is no smooth bijection $\mathbb{R}^2 \to S^2$.

The property the sphere does have is that there are *locally* parameterisations and coordinates. For each point $p \in S^2$, there is an open subset V of the sphere containing p, and an open subset $U \subset \mathbb{R}^2$, and a parameterisation

$$\pi\colon U \to V.$$

Example. The spherical polar parameterisation is the map $U \to V \subset S^2$ defined by

$$(\theta, \phi) \mapsto (\sin\theta \sin\phi, \sin\theta\cos\phi, \cos\theta)$$

and is defined on the set $U \subset \mathbb{R}^2$ given by $0 < \theta < \pi$, $0 < \phi < 2\pi$. The image V is the open subset of the sphere given by excluding the points (x, y, z) such that $x = 0$ and $y > 0$.

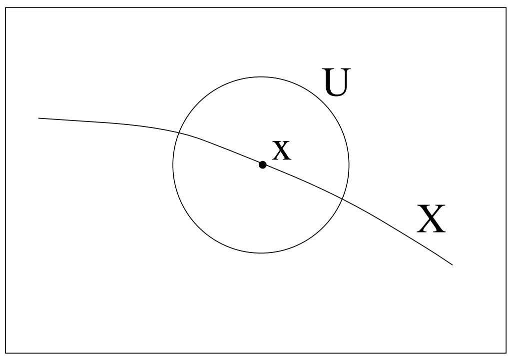
Spherical polar coordinates are the inverse of this mapping, $V \to U$.

The definition of a manifold will generalise this idea. A manifold is a subset of $\mathbb{R}^n$ which has local parameterisations and coordinates from open subsets of $\mathbb{R}^k$. We say that a manifold is a 'locally Euclidean' subset of $\mathbb{R}^n$.

Functions defined on subsets of $\mathbb{R}^n$.

Smooth functions have been defined already when the domain is an open subset of $\mathbb{R}^n$. Now follows the definition for an arbitrary subset.

Definition. Let X be any subset of $\mathbb{R}^n$. A function $f\colon X \to Y \subset \mathbb{R}^m$ is called smooth if for every point $x \in X$ there is an open subset $U \subset \mathbb{R}^n$ so that U contains x, and a smooth function $F\colon U \to \mathbb{R}^m$ which agrees with f on the open subset $V = U \cap X$ of X.



As before, if f is smooth and f^{-1} is also smooth, then f is called a diffeomorphism.

Exercise 1. Let $L \subset \mathbb{R}^n$ be a linear subspace, and $f\colon L \to \mathbb{R}^m$ a linear map. Is f smooth?

Exercise 2. Define $X = \{(x,y) | xy = 1\} \subset \mathbb{R}^2$, and $f\colon X \to \mathbb{R}$ by $(x,y) \mapsto \sqrt{x^2+y^2}$. Is f smooth?

Exercise 3. Define $X = \{(t^2,t^3) \mid t \in \mathbb{R}\}$, and the function $f\colon X \to \mathbb{R}$ by $(t^2,t^3) \mapsto t$. Is f smooth?

Exercise 4. A function is defined on the line $L \subset \mathbb{R}^2$ which passes through two distinct points $a, b \in \mathbb{R}^2$ by

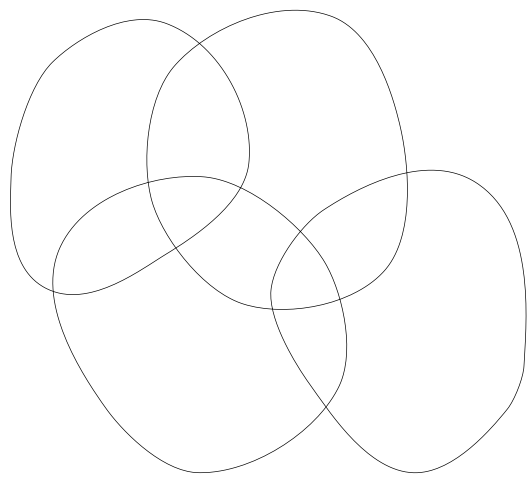
$$f(\lambda a + (1-\lambda)b) = \lambda^2$$

What is the definition of a smooth function defined on L? Construct a function F on the whole plane which agrees with f on L.

Definition of a manifold. A set $X \subset \mathbb{R}^k$ is called a manifold of dimension n if for every $x \in X$ there is an open subset $U \subset X$ containing x, and an open subset $V \subset \mathbb{R}^n$, such that there is a diffeomorphism from U to V.

This diffeomorphism is called a coordinate function, and its inverse a parameterisation of X.

Terminology. The open subsets $U \subset X$ in the definition of a manifold, together with the diffeomorphism to $V \subset \mathbb{R}^n$, are often called coordinate charts, or just charts. A collection of charts is said to cover X if every point $x \in X$ is in at least one of them. Such a collection is called an atlas for X. Coordinate charts are also called coordinate patches.



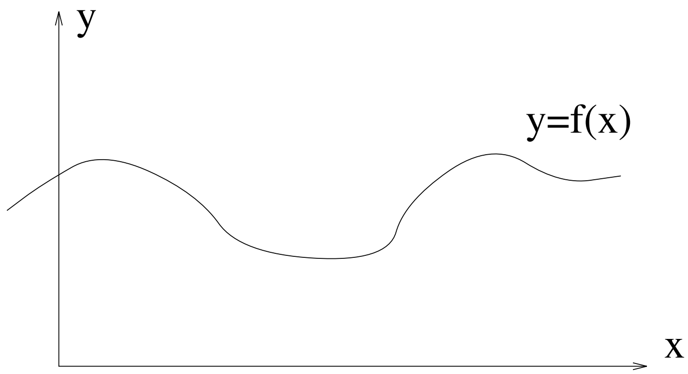
Part of an atlas

Examples of manifolds.

Open subsets. The most elementary examples of manifolds are given by taking $X \subset \mathbb{R}^n$ to be an open subset. Then $U = V = X$ and the diffeomorphism in the definition is the identity map. These examples include $\mathbb{R}^n$ itself.

Graphs. A large class of examples of manifolds are given by the graph of a smooth function defined on an open subset $W \subset \mathbb{R}^n$, $f\colon W \to \mathbb{R}^m$. This is the set $X = \{(w, f(w)) | w \in W\} \subset \mathbb{R}^{n+m}$. The diagram shows a one-dimensional example.

To show that this set is a manifold, take $U = X$ and $V = W$ in the definition of a manifold. The parameterisation is $x \mapsto (x, f(x))$ and its inverse is $(x, f(x)) \mapsto x$.



The graph of a function

Example. The upper hemisphere $U^n \subset \mathbb{R}^{n+1}$ is the subset

$$\sum_{i=1}^{n+1} x_i^2 = 1, \qquad x_{n+1} > 0.$$

This is the graph of the function $B_1(0) \to \mathbb{R}$ defined by

$$(x_1, \dots x_n) \mapsto \sqrt{1 - \sum_{i=1}^{n} x_i^2}.$$

Spheres. The n-dimensional sphere $S^n \subset \mathbb{R}^{n+1}$ is defined as the set of all points satisfying

$$\sum_{i=1}^{n+1} x_i^2 = 1.$$

S^n is a manifold. This is because any point $x \in S^n$ is contained in a hemisphere for one of the axes. For example, if $x_{n+1} > 0$ then $x \in U^n$. If $x_{n+1} < 0$, then x is contained in the lower hemisphere $L^n = -U^n$. If $x_{n+1} = 0$, there is another coordinate, x_k, which is not zero, and then x is contained in a hemisphere defined by taking x_k to be the independent variable, instead of x_{n+1}.

Therefore S^n is the union of a number of open subsets U (the hemispheres) which are diffeomorphic to an open subset of $\mathbb{R}^n$, as required in the definition of a manifold.

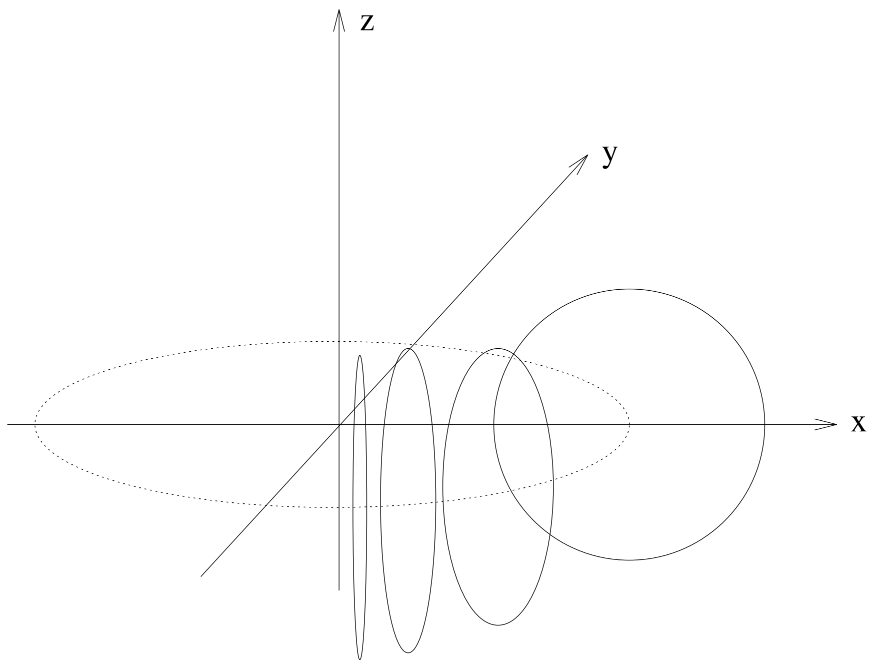
Torus. Manifolds of dimension 2 are called surfaces. These arise most naturally as subsets of $\mathbb{R}^3$, according to our everyday experience. An example we have already met is $S^2 \subset \mathbb{R}^3$. Another example is the torus. This can be drawn by first drawing a circle of radius 2 in the

$x - y$ plane, then taking each of these points as the centre of another circle of radius 1, lying in a plane through the z-axis.

Following this idea through, leads to a definition of $T \subset \mathbb{R}^3$ as the subset of points satisfying

$$\left(\sqrt{x^2+y^2}-2\right)^2+z^2=1$$

It is possible to show that T is a manifold directly, by finding a set of coordinate charts which cover T, as was done for the sphere. It also follows from an exercise below.



Groups. The groups which arise in the special geometries have sets of elements which are labelled by continuous parameters. In this situation, it is natural to ask whether these sets are manifolds. The group $\mathrm{GL}(n)$ is the set of all invertible matrices; thus the matrix entries can be taken as n^2 coordinates. This is a map $\mathrm{GL}(n) \to \mathbb{R}^{n^2}$ given by

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \mapsto (a_{11}, a_{12}, \ldots, a_{nn})$$

It is often convenient to regard $\mathrm{GL}(n)$ as a subset of $\mathbb{R}^{n^2}$, as the re-arrangement of the matrix entries as a vector is of little consequence.

The condition on the matrix A for it to be invertible is $\det A \neq 0$, which is a polynomial equation in the matrix entries. For example, for $n = 2$, this is

$$a_{11}a_{22} - a_{12}a_{21} \neq 0.$$

This defines an open subset of $\mathbb{R}^{n^2}$. In general, this is because the map $\det\colon \mathbb{R}^{n^2} \to \mathbb{R}$ is continuous, and $\mathrm{GL}(n)$ is the inverse image of $\mathbb{R} \setminus \{0\}$, an open set. Thus $\mathrm{GL}(n)$ is an n^2-dimensional manifold.

Various subgroups of $\mathrm{GL}(n)$ are also manifolds. For example, the group of $n \times n$ orthogonal matrices, $\mathrm{O}(n)$, is a manifold of dimension $\frac{1}{2}n(n-1)$.

Alternative definition. Many books use an alternative definition of manifold, as a set X which is a topological space, and an atlas of coordinate charts for X. Thus X is not regarded as a subset of $\mathbb{R}^n$. There are conditions on the atlas for this definition to make sense, and the main disadvantage of this method is that these conditions require much more technical effort to explain. There is no more generality in this as the set of manifolds obtained by the alternative definition is equivalent to the set of manifolds defined here.

Exercise 5. Let $U \subset \mathbb{R}^3$ be the upper hemisphere

$$\{(x, y, z) | x^2 + y^2 + z^2 = 1, z > 0\}$$

The map $f\colon U \to \mathbb{R}^2$ is defined by $(x, y, z) \mapsto (x/z, y/z)$. Give a formula for the inverse function f^{-1}. Is f^{-1} smooth?

Exercise 6. The spiral $S \subset \mathbb{R}^2$ is the set of points of the form

$$(e^\theta \cos\theta, e^\theta \sin\theta),$$

for $\theta \in \mathbb{R}$. Sketch S. Now consider the function $f\colon S \to \mathbb{R}$ defined by $(e^\theta \cos\theta, e^\theta \sin\theta) \mapsto \theta$. Consider a point $x \in S$, and suppose the ball $B_r(x)$ does not contain the origin ($r \leq |x|$). Show that there is a map $F\colon B_r(x; \mathbb{R}^2) \to \mathbb{R}$ which agrees with f by giving an explicit formula. What goes wrong if $r > |x|$?

Is the original function f smooth? Is S a manifold?

Exercise 7. Give a set of parameters for the elements of the affine group $A(n)$. Hence describe $A(n)$ as a subset of $\mathbb{R}^k$ for some k.

Exercise 8. In complex analysis, it is often convenient to add a single extra point ‘at infinity’ to the complex plane, so that any set of points of increasing radius from the origin converges to the extra point ∞. What manifold do think $\mathbb{C} \cup \{\infty\}$ should be?

Maps of manifolds.

If $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ are manifolds, then we have essentially already defined a smooth map $M \to N$: it is a smooth map $M \to \mathbb{R}^n$, as defined previously, such that the image lies in N. If $f: M \to N$ is a smooth map with a smooth inverse, then it is called a diffeomorphism, and M and N are said to be diffeomorphic. For example, S^2 is diffeomorphic to an ellipsiod

$$\{(x, y, z) \mid ax^2 + by^2 + cz^2 = 1\},$$

determined by constants $a, b, c > 0$. The diffeomorphism is the linear map

$$(x, y, z) \mapsto \left(\frac{x}{\sqrt{a}}, \frac{y}{\sqrt{b}}, \frac{z}{\sqrt{c}}\right).$$

The torus and the sphere are an example of two manifolds which are not diffeomorphic. This is intuitively obvious as the torus has a hole through the middle but the sphere does not, but proving it requires some thought. Here is an argument which can be made into a proof. Suppose $f: T \to S^2$ is a diffeomorphism. The torus T has two circles on it which cross at only one point (e.g., given by $z = 0$ and $y = 0$). The images of these circles would be two circles on the sphere which cross at only one point. Draw one circle on a sphere. It is 'obvious' that one cannot draw a second circle to cross it at only one point.

Product of manifolds. If $M \subset \mathbb{R}^k$ and $N \subset \mathbb{R}^l$ are manifolds, then so is $M \times N \subset \mathbb{R}^{k+l}$. As an example, $S^1 \times S^1$ is the subset of $\mathbb{R}^4$ given by

$$x^2 + y^2 = 1, \qquad z^2 + t^2 = 1.$$

Exercise 9. How can $S^1 \times S^1 \subset \mathbb{R}^4$ be parameterised? Find a diffeomophism $S^1 \times S^1 \to T$.

Exercise 10. Give a diffeomorphism between the manifold $D \subset \mathbb{R}^5$ given by

$$w^2 + x^2 + y^2 + z^2 - t^2 = 1$$

and $S^3 \times \mathbb{R}$.

Solution to exercise 1. Let L' be a complementary subspace, so that $\mathbb{R}^n = L + L'$. Decompose a vector $\xi \in \mathbb{R}^n$ as $\xi = l + l'$, $l \in L$, $l' \in L'$, and set $F(\xi) = f(l)$. This map is defined on the whole of $\mathbb{R}^n$ (certainly an open subset), agrees with f, and is smooth. Note that since there are many choices of complementary subspaces L', there are many linear extensions F. There are also choices of F which are not linear.

12. Discrete Quotients

Let $M \subset \mathbb{R}^m$, $N \subset \mathbb{R}^n$ and $f\colon M \to N$ a smooth map. The maps we are interested in are maps which are isomorphisms on a sufficiently small scale.

Definition. A local diffeomorphism $f\colon M \to N$ is a map such that for each $x \in M$ there is an open subset $U \subset M$ containing x and an open subset $V \subset N$, so that f restricts to a diffeomorphism $U \to V$.

If the local diffeomorphism f is onto, we say that N is a discrete quotient of M. In the rest of this section, this is abbreviated to quotient, though the term quotient can apply to more general onto maps. If $y \in N$, we say that f identifies the points $f^{-1}(y) \subset M$ in the quotient.

Example. The simplest example is the map $\mathbb{R} \to S^1$ given by

$$\theta \mapsto (\cos\theta, \sin\theta).$$

The points $\theta + 2\pi n$, $n \in \mathbb{Z}$, are identified to a single point in S^1.

There is a purely topological notion of a quotient space. Namely, given a topological space X and an onto map f to a set Y, then there is the quotient topology of Y defined by making $V \subset Y$ open whenever $f^{-1}(V)$ is. This could give a possibly conflicting notion of quotient, so it is important to show that these coincide.

Theorem. *Let f be a local diffeomorphism of M onto N. Then N has the quotient topology.*

Proof. This is proved by showing that f maps open sets to open sets. This is called an open mapping. Then the result follows easily from this. The details are in the following two lemmas

Lemma. *A local diffeomorphism is an open mapping.*

Proof. For each $x \in M$, let U_x be the open set containing x on which f is a diffeomorphism. Then if $U \subset M$ is any open set,

$$U = \cup_{x \in U} U_x \cap U$$

and

$$f(U) = \cup_{x \in U} f(U_x \cap U).$$

Since f is a homeomorphism $U_x \to f(U_x)$, $f(U_x \cap U)$ is an open subset of $f(U_x)$ and hence an open subset of N. Hence $f(U)$ is an open subset of N.

Lemma. *If $f: M \to N$ is an open mapping and is onto, then N has the quotient topology.*

Proof. Let $V \subset N$ be a subset such that $U = f^{-1}(V)$ is open. Then $f(U)$ is open since f is an open mapping, and $V = f(U)$ since f is onto. Hence a subset of N which is open in the quotient topology is open in the given topology of N. The converse follows from the fact that f is continuous.

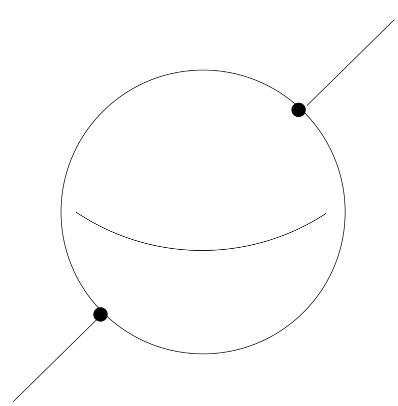
However the condition of local diffeomorphism is much stronger than being a purely topological statement. For example $x \mapsto x^3$ is a homeomorphism but not a local diffeomorphism. But the topological notion of a quotient space is a useful half-way house to give a description of N.

Definition. A fundamental domain for a quotient $f: M \to N$ is a closed subset $D \subset M$ such that $f(D) = N$ but f only identifies points on the boundary ∂D of D.

For example, the interval $[0, 2\pi]$ is a fundamental domain for the map $\mathbb{R} \to S^1$ above. The boundary is the endpoints of the interval, $\{0, 2\pi\}$ and these two points are identified to one point in the quotient.

This example can be generalised to products of the circle, for example $\mathbb{R}^2 \to S^1 \times S^1$. A fundamental domain is a square with opposite edges identified, and all four corners identified to one point.

Projective space. So far, P^n has been described simply as a set. Now it can be described as a manifold, obtained as a quotient of the sphere. Recall that P^n is defined as the set of lines through the origin in $\mathbb{R}^{n+1}$. Each line intersects the sphere S^n in exactly two points, $\pm x$.



So to construct projective space as a manifold, it is sufficient to find a quotient of the sphere, $f: S^n \to Q^n \subset \mathbb{R}^k$. This map to $\mathbb{R}^k$ has the property that $f(x) = f(-x)$. So to construct such a map, a good starting point is to find a set of functions $S^n \to \mathbb{R}$ with this property.

Let $x = (x_1, \dots)$ denote points on S^n. Then the functions $x_j x_k$ for any choice $1 \le j, k \le n+1$ give a function with same value on $\pm x$. Define $f\colon S^n \to \mathbb{R}^{\frac{1}{2}(n+1)(n+2)}$ to be the map

$$(x_1, x_2, \dots, x_{n+1}) \mapsto (x_1^2, x_1x_2, x_1x_3, \dots, x_{n+1}^2),$$

and define Q^n to be the image of f.

Now a number of questions need to be settled to show this construction works.

(1) Is f a local diffeomorphism?
(2) Is Q^n a manifold?
(3) Are any other points identified, beside $\pm x \in S^n$?

The answer to (1) is yes due to the following construction. The map f can be locally inverted by explicit formulae, which will be given here for $n = 2$. The function $f\colon S^2 \to Q^2$ is the map

$$(x_1, x_2, x_3) \mapsto (x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2) = (y_1, y_2, y_3, y_4, y_5, y_6).$$

Assume that $x_3 > 0$. Then $x \in S^2$ can be calculated from

$$x_1 = \frac{x_1x_3}{\sqrt{x_3^2}}, \quad x_2 = \frac{x_2x_3}{\sqrt{x_3^2}}, \quad x_3 = \sqrt{x_3^2}.$$

Therefore the inverse of f is locally

$$(y_1, y_2, y_3, y_4, y_5, y_6) \mapsto \left(\frac{y_3}{\sqrt{y_6}}, \frac{y_5}{\sqrt{y_6}}, \sqrt{y_6}\right).$$

This formula clearly extends to points in an open subset of $\mathbb{R}^6$ and so indeed provides the smooth inverse to f for the hemisphere $x_3 > 0$.

The answer to the question (2) is yes by the following general result.

Lemma. *Let M be a manifold, and N a subset of $\mathbb{R}^k$. If there is a local diffeomorphism $f\colon M \to N$, then N is manifold.*

Proof. Suppose f restricts to a diffeomorphism $U \to V$ for open subsets U and V. Let $x \in U$. Then there is a coordinate chart defined on an open subset $U' \subset M$, with $x \in U'$. Of course, U and U' need not coincide, but one can restrict both maps to the intersection, $U'' = U \cap U'$. Then U'' is diffeomorphic to an open subset of $\mathbb{R}^n$, and to an open subset $f(U'')$ of N. This provides a coordinate chart around the point $f(x) \in N$.

Finally, it is an algebraic exercise to settle question (3).

Exercise 1. Show that the only points identified in the map $S^n \to Q^n$ are $\pm x$. What is a fundamental domain for this quotient? How are the boundary points in your fundamental domain identified?

Exercise 2. Show that the inhomogeneous coordinates for P^n give coordinates for the manifold Q^n.

These calculations show that Q^n really 'is' P^n, viewed as a manifold. Calculations can be done using the inhomogeous coordinates, as previously, but now with the understanding that these are coordinate charts on a manifold.

Exercise 3. Draw the subset $Q^1 \subset \mathbb{R}^3$.

Exercise 4. Find a map $S^1 \to S^1$ which is a local diffeomorphism, but is not a bijection. Is this possible for a map $\mathbb{R} \to \mathbb{R}$?

Exercise 5. Define $M \subset P^4$ by the equation

$$x^2 + y^2 + z^2 - t^2 - w^2 = 0.$$

Find an onto map $S^2 \times S^1 \to M$, and describe the quotient of $S^2 \times S^1$ this defines.

13. The tangent space

Tangent vectors. A curve is a map $c\colon I \to \mathbb{R}^k$, where $I \subset \mathbb{R}$ is an open interval, say $\{a < t < b\}$. Its tangent vector at a given parameter t is

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \left(\frac{\mathrm{d}c_1}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}c_k}{\mathrm{d}t}\right).$$

Now suppose $f\colon \mathbb{R}^k \to \mathbb{R}^l$ is any smooth function. It carries the curve c into a curve c' in $\mathbb{R}^l$, namely $c'(t) = f(c(t))$. The tangent vector for curve $c' = f \circ c$ can be calculated by the chain rule by

$$\mathrm{d}c'_t = \mathrm{d}f_{c(t)} \circ \mathrm{d}c_t,$$

so that

$$\frac{\mathrm{d}c'}{\mathrm{d}t} = \mathrm{d}c'_t(1) = \mathrm{d}f_{c(t)}\left(\mathrm{d}c_t(1)\right) = \mathrm{d}f_{c(t)}\left(\frac{\mathrm{d}c}{\mathrm{d}t}\right).$$

Therefore the derivative of f, $\mathrm{d}f_x$, can be interpreted as a linear mapping of tangent vectors, for curves passing through the point $x \in \mathbb{R}^k$.

Tangent space. Let $M \subset \mathbb{R}^k$ be a manifold, and $x \in M$. Then a vector $v \in \mathbb{R}^k$ is said to be tangent to M at x if there is a curve $c\colon I \to M$ such that $c(t) = x$ for some $t \in I$, and $\frac{\mathrm{d}c}{\mathrm{d}t} = v$.

Definition. The tangent space at $x \in M$, denoted TM_x, is the subset of $\mathbb{R}^k$ consisting of all vectors which are tangent to M at x.

Exercise 1. Show that $0 \in TM_x$.

Lemma. *$TM_x \subset \mathbb{R}^k$ is a linear subspace.*

Proof. Let $p\colon U \to M$ be a parameterisation, for an open subset $U \subset \mathbb{R}^m$, such that $u \in U$ is mapped to x. It will be shown that

$$TM_x = \mathrm{Image}\,(\mathrm{d}p_u)\,,$$

which is a linear subspace.

Firstly, if c is a curve in M with $c(0) = x$, then for values of t sufficiently close to 0, $c(t) = p \circ p^{-1} \circ c(t)$, so at $t = 0$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \mathrm{d}p_u \left(\frac{\mathrm{d}(p^{-1} \circ c)}{\mathrm{d}t} \right) \in \mathrm{Image}\,(\mathrm{d}p_u)\,.$$

Conversely, if $v \in \mathrm{Image}\big(\mathrm{d}p_u\big)$, then $v = \mathrm{d}p_u(\xi)$ for some $\xi \in \mathbb{R}^m$. The curve $c(t) = p(u + t\xi)$ has tangent vector

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \mathrm{d}p_u(\xi) = v$$

at the parameter value $t = 0$.

Exercise 2. Use the chain rule to show that the dimension of TX_x is the dimension of X. Use the fact that a parameterisation has an inverse.

Examples of tangent spaces. The tangent space to $\mathbb{R}^n$ at any point is $\mathbb{R}^n$ itself, as the identity map is a parameterisation, and the derivative of this is also the identity map.

The tangent space to an affine subset $A \subset \mathbb{R}^k$ at any point $x \in A$ is the subspace $TA = \{a - x | a \in A\}$ introduced earlier. It is independent of x. This follows because an affine subset can be parameterised by an affine map $\phi\colon \mathbb{R}^n \mapsto A \subset \mathbb{R}^k$. Then $\mathrm{d}\phi_x$ is just the map

$$\mathrm{d}\phi\colon y \mapsto \phi(y) - \phi(x)$$

introduced earlier. The image of this map is the set of points $\{\phi(y) - \phi(x)\}$ for a fixed x and all $y \in \mathbb{R}^n$, which is the definition of TA.

Manifolds $M \subset \mathbb{R}^k$ can occur as the solutions to an equation $F = 0$, where $F\colon \mathbb{R}^k \to \mathbb{R}^l$. Then $\mathrm{d}F_x(v) = 0$ for any vector $v \in TM_x$. If $v = \frac{\mathrm{d}c}{\mathrm{d}t}$, then $F \circ c = 0$, so $\frac{\mathrm{d}F \circ c}{\mathrm{d}t} = \mathrm{d}F_x(v) = 0$. This gives a method of determining linear equations for TM_x.

Example. The sphere is given by $F: \mathbb{R}^{n+1} \to \mathbb{R}$

$$F = \left(\sum_{i=1}^{n+1} x_i^2\right) - 1 = 0.$$

Then $\mathrm{d}F_x(v) = 2(\sum x_i v_i) = 2x \cdot v = 0$. This is one linear equation in $\mathbb{R}^{n+1}$ and hence determines an n-dimensional linear subspace. Since TS^n_x also has dimension n, it follows that these two spaces are equal. Hence TS^n_x is the set of all vectors satisfying $x \cdot v = 0$.

Exercise 3. Show that the maps $S^3 \to \mathbb{R}^4$ given by

$$u: (x_1, x_2, x_3, x_4) \mapsto (-x_4, -x_3, x_2, x_1)$$

$$v: (x_1, x_2, x_3, x_4) \mapsto (x_3, -x_4, -x_1, x_2)$$

$$w: (x_1, x_2, x_3, x_4) \mapsto (-x_2, x_1, -x_4, x_3)$$

determine three vector fields on S^3. Do $u(x)$, $v(x)$, $w(x)$ form a basis for TS^3_x?

Differentiation on manifolds.

Definition. Let $M \subset \mathbb{R}^k$, $N \subset \mathbb{R}^l$ be manifolds, and $f: M \to N$. Then the derivative of f at $x \in M$ is the linear map $\mathrm{d}f_x: TM_x \to TN_{f(x)}$ which satisfies

$$\mathrm{d}f_x\left(\frac{\mathrm{d}c}{\mathrm{d}t}\right) = \frac{\mathrm{d}(f \circ c)}{\mathrm{d}t}.$$

where c is any curve through x.

In this definition, it is necessary to check that a linear map with these properties exists. Suppose F is defined on an open subset of $\mathbb{R}^k$ and agrees locally with f. Then we can use the chain rule for $\mathbb{R}^k$ on F

$$\mathrm{d}f_x(v) = \frac{\mathrm{d}(f \circ c)}{\mathrm{d}t} = \frac{\mathrm{d}(F \circ c)}{\mathrm{d}t} = \mathrm{d}F_{c(t)}\left(\frac{\mathrm{d}c}{\mathrm{d}t}\right) = \mathrm{d}F_x(v).$$

This shows that the derivative in this more general situation (manifolds) is just the restiction of $\mathrm{d}F_x$ to TM_x. This proves that a linear map with these properties exists. One might worry that this depends on the choice of F; however the definition of $\mathrm{d}f_x$ determines its values on all vectors tangent to M uniquely without reference to F.

Exercise 4. Show that the chain rule holds for maps of manifolds.

Vector fields. A vector field on a manifold is a choice of a tangent vector at each point of M. More precisely, a vector field on a manifold $M \subset \mathbb{R}^k$ is a smooth map $v\colon M \to \mathbb{R}^k$ such that $v(x) \in TM_x$ for each $x \in M$.

The tangent bundle. The tangent spaces for the different points $x \in M$ are generally different subspaces of $\mathbb{R}^k$.

In general, the tangent spaces can be 'glued together' to form the tangent bundle of a manifold, $TM \subset \mathbb{R}^{2k} = \mathbb{R}^k \times \mathbb{R}^k$. This is defined to be the set of all points (x, v), for $x \in M$ and $v \in TM_x$. A vector field can be described as a map $M \to TM$, as

$$x \mapsto (x, v(x)).$$

Exercise 5. Describe the tangent bundle $TS^1 \subset \mathbb{R}^4$ explicitly by giving two equations for the subset. Give a vector field on S^1 which is nowhere zero, i.e., the tangent vector at every point is not 0.

If V is a one-dimensional vector space, explain how a choice of vector in V determines a linear isomorphism $\mathbb{R} \to V$. Use your vector field to give a diffeomorphism $S^1 \times \mathbb{R} \to TS^1$.

Exercise 6. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function. Show that the graph $G \subset \mathbb{R}^{n+m}$ of f is a manifold.

For a point $g \in G$, which function is TG_g the graph of?

Explain how you would show that the subset of $\mathbb{R}^3$ given by

$$x^4 + y^4 + z^4 = 1$$

is a manifold.

Give an explicit description of the tangent space at the point $x = y = \frac{1}{\sqrt[4]{2}}$, $z = 0$, with numerical coefficients in the equation that you use.

Exercise 7. A robot arm in the plane has its elbow at $x \in \mathbb{R}^2$ and hand at $y \in \mathbb{R}^2$. These are constrained by $|x| = 1$ and $|y - x| = 1$. Let $X = \{(x, y)\} \subset \mathbb{R}^4$ be the set of configurations for the arm. Which standard manifold is X diffeomorphic to?

The arm is controlled by a motor which sets the angle of the upper arm x relative to a fixed axis, and a second motor which sets the angle of the lower arm $(y - x)$ relative to the upper arm. Explain how the hand y can be moved in a given direction in the plane given by a tangent vector v by giving a direction in the space of angles controlled by the motors. At what points y does this control mechanism fail to work for some tangent v?

How would you draw a circle of very small radius at the point $y = (1, 0)$? Explain why this does not work at the origin $y = 0$.

Exercise 8. Two solid bodies touch at a single point p. Assume the boundaries of the solid bodies can be modelled as smooth surfaces. What can you say about the relation between the tangent spaces of the two boundary surfaces at p? Give at least one concrete example.

14. Lie Groups

A group is a set G with maps $\mu: G \times G \to G$ and $\sigma: G \to G$ giving the multiplication and inverse of group elements, and an element $e \in G$, the identity, all satisfying the usual axioms. The usual notation is $\mu(a, b) = ab$ and $\sigma(a) = a^{-1}$.

Definition. A Lie group is a group in which G is a manifold and μ and σ are smooth maps.

In the same way, one can define 'Lie' versions of all the elementary definitions in group theory. For example, a subgroup $H \subset G$ which is also a Lie group is called a Lie subgroup of G. A homomorphism of Lie groups $F \to G$ is a group homomorphism which is also a smooth map.

Examples of Lie Groups.

The group $\mathrm{GL}(n)$ is a Lie group. The coordinates for $\mathrm{GL}(n)$ are the matrix entries. The map μ is smooth because a matrix entry for the product $\mu(a, b)$ is a polynomial in the matrix entries for a and b. Also, σ is smooth because the matrix entries for a^{-1} are polynomials divided by $\det a$, which is never zero for elements of $\mathrm{GL}(n)$.

The vector space $\mathbb{R}^n$ is a Lie group, with $\mu(a, b) = a + b$, $\sigma(a) = -a$. This group is called the translation group, $\mathrm{T}(n)$.

Exercise 1. Show $\mathrm{A}(n)$ is a Lie group.

Further examples of Lie groups arise as subgroups of $\mathrm{GL}(n)$. These will be discussed later. The Euclidean groups $\mathrm{E}(n)$ and the projective groups $\mathrm{PGL}(n)$ are also Lie groups.

Actions of Lie groups.

Let M be a manifold, and G a Lie group. An action of G on M is a map $\lambda: G \times M \to M$ satisfying

(1) $\lambda\left(\mu(g, h), x\right) = \lambda\left(g, \lambda(h, x)\right)$
(2) $\lambda(e, x) = x$

Each element $g \in G$ provides a smooth map

$$\lambda_g : x \mapsto \lambda(g, x),$$

called the action of g. The conditions (1) and (2) can be written

$$\lambda_{(gh)} = \lambda_g \circ \lambda_h$$
$$\lambda_e = \text{identity}\,.$$

Since $\lambda_e = \lambda_{g^{-1}g} = \lambda_{g^{-1}} \circ \lambda_g$, the action of g^{-1} is the inverse of the action of g. Therefore, the action of $g \in G$ is a diffeomorphism. The conditions satisfied by an action can be stated alternatively as saying that there is a homomorphism from G to the group of all diffeomorphisms of M.

Classification of actions. There are three properties an action may have

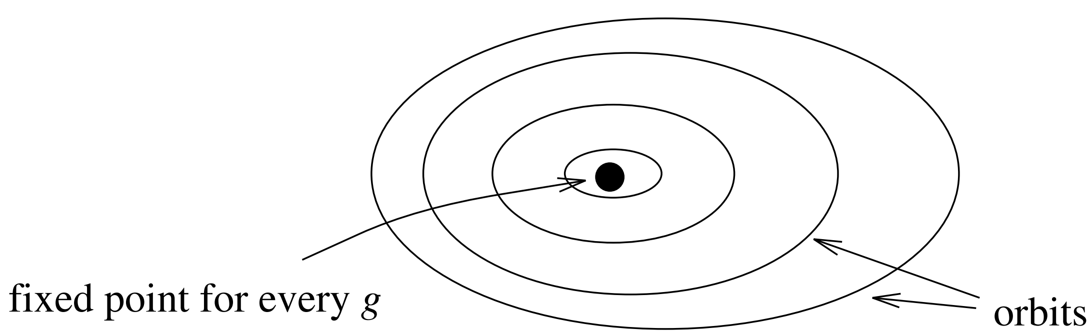
(1) Effective. If $g \in G$ is such that $\lambda_g =$ identity, then $g = e$.
(2) Free. If $\lambda(g, x) = x$ for some $x \in M$, then $g = e$.
(3) Transitive. For all $x, y \in M$, there exists $g \in G$ such that $\lambda(g, x) = y$.

The first property, effective, is that every element of G except the identity does 'something somewhere'. This property is automatic for any action defined as the set of all transformations of a manifold of a particular kind. The action of $\mathrm{GL}(n)$ on P^{n-1} is not effective because all multiples of the identity in $\mathrm{GL}(n)$ act as the identity in P^{n-1}.

A fixed point for an element $g \in G$ is a point $x \in M$ such that $\lambda(g, x) = x$. For example, the rotations of a sphere about the z-axis have the north and south poles as fixed points.

An action is free if every element except e has no fixed points. For example, the action of $\mathrm{T}(n)$ on $\mathbb{R}^n$ by translations is free. By contrast, the action of $\mathrm{O}(3)$ on S^2 is not, because of the fixed points for rotations just mentioned.

The orbit of a point $x \in M$ is the set of all points $\{\lambda(g, x) | g \in G\}$. An action is transitive if there is only one orbit. For example, the action of $\mathrm{T}(n)$ on $\mathbb{R}^n$ by translations is transitive, as is the action of $\mathrm{O}(3)$ on S^2. However if you take the subgroup of rotations about the z-axis, this is not transitive.



Exercise 2. Are the following actions effective, free or transitive?

(1) $\mathrm{GL}(n)$ acting on $\mathbb{R}^n$

(2) The group of rotations about the z axis in $\mathbb{R}^3$ acting on the sphere, S^2.
(3) A Lie group G acting on G by multiplication in the group: $\lambda(g,h) = gh$.
(4) A Lie group G acting on G by conjugation: $\lambda(g,h) = ghg^{-1}$.

15. Flows

A flow is an action of the group T(1) on a manifold. Let $\lambda: \mathbb{R} \times M \to M$ be the action. Then each point $x \in M$ gives a curve

$$\lambda^x : \mathbb{R} \to M$$
$$t \mapsto \lambda(t,x).$$

Since $\lambda^x(0) = \lambda(0,x) = x$, the curve gives all points on the orbit of x under the flow. The curve has tangent vector

$$v(x) = \frac{\mathrm{d}\lambda^x}{\mathrm{d}t}(0) \in TM_x.$$

A vector field, called the velocity vector field of the flow is defined by the function $x \mapsto v(x)$.

Example. A river flows smoothly (of course) along a waterway M. The function $\lambda(t,x)$ gives the position at time t of the molecule of water which is at the point x at time 0. The vector $v(x)$ gives the velocity of the water passing the point x at any value of the time parameter.

This follows from the property of a group action that

$$\lambda^x(t) = \lambda^y(T+t) \qquad \text{for } x = \lambda(T,y),$$

i.e., the curve through x is the same as the curve through y with the parameter shifted by $t \mapsto T+t$. Then

$$\frac{\mathrm{d}\lambda^y}{\mathrm{d}t}(T) = \frac{\mathrm{d}\lambda^x}{\mathrm{d}t}(0) = v(x).$$

Ordinary differential equation. Given a vector field v on a manifold and a point $x \in M$, the problem is to find a curve $c: I \to M$ such that $c(0) = x$, $I \subset \mathbb{R}$ is an open interval containing 0, and the tangent vector at any parameter t agrees with the vector field, i.e.

$$\frac{\mathrm{d}c}{\mathrm{d}t}(t) = v\big(c(t)\big).$$

Such a curve is called an integral curve of the ordinary differential equation. If it exists, the integral curve is unique. (The proof of this, not given here, involves some analysis). If the vector field v is the velocity of a flow, then the solution is given by $c(t) = \lambda^x(t)$. Therefore, the velocity vector field of a flow specifies the flow uniquely.

Example. The differential equation on $\mathbb{R}$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = ac,$$

with $a \in \mathbb{R}$ a constant, has solution $c(t) = xe^{at}$, which is determined by the flow $\lambda(t, x) = e^{at}x$.

Exercise 1. Solve the equation

$$\frac{\mathrm{d}c}{\mathrm{d}t} = ac^2$$

and show that the solutions do not determine a flow.

Example. In mechanics, Newton's equations for the position of n particles $x \in \mathbb{R}^{3n}$ can be written as the ordinary differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = p, \qquad \frac{\mathrm{d}p}{\mathrm{d}t} = f(x).$$

The function f gives the forces on the particles as a function of the positions, and is determined by the particular problem. This is an ordinary differential equation in a subset of $\mathbb{R}^{6n}$.

In Newton's theory of gravity, there is a formula for f determined by the inverse square law. The integral curves of this equation account for the orbits of the planets, amongst other phenomena. For $n \geq 4$, the integral curves cannot always be defined for all $t \in \mathbb{R}$. There is, suprisingly, a configuration of 4 bodies for which the orbits become progressively more violent, and one of the bodies 'reaches infinity' in a finite interval of time.

Vector field as an operator. Suppose $M \subset \mathbb{R}^k$ is a manifold, and v is a vector field on M. If $\phi\colon M \to \mathbb{R}$ is a function, then the vector field can be regarded as a differential operator which acts on ϕ to give a new function, called $\mathrm{D}_v\phi$.

$$\mathrm{D}_v\phi(x) = \mathrm{d}\phi_x\bigl(v(x)\bigr) = \sum_{i=1}^{k} \frac{\partial\phi}{\partial x_i} v_i(x).$$

This operation can be thought of as differentiating ϕ in the direction given by v.

This can be substantiated as follows. Let $\lambda\colon \mathbb{R} \times M \to M$ be a map, which could be a flow. Then it defines a vector field by

$$v(x) = \frac{\mathrm{d}\lambda^x}{\mathrm{d}t}(0)$$

as before. Now differentiating ϕ along each curve λ^x, one finds

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\phi \circ \lambda^x\right)(0) = \mathrm{d}\phi_x\left(\frac{\mathrm{d}\lambda^x}{\mathrm{d}t}\right) = \mathrm{D}_v\phi.$$

Exercise 2. Take $M = S^2$, and the rotation $\lambda\colon \mathbb{R} \times S^2 \to S^2$

$$(t, x_1, x_2, x_3) \mapsto (x_1 \cos t + x_2 \sin t, -x_1 \sin t + x_2 \cos t, x_3)$$

Calculate the vector field v and an expression for $\mathrm{D}_v\phi$.

Partial differential equation.

Suppose $\phi\colon M \to \mathbb{R}$ is a function which is constant along a flow λ, i.e., $\phi\big(\lambda^x(t)\big) = \phi(x)$ for each x, t. This gives the first order partial differential equation

$$\sum_{i=1}^{k} v_i \frac{\partial \phi}{\partial x_i} = 0.$$

This equation can be used to solve the ordinary differential equation given by v. The integral curves of the ordinary differential equation must lie in the subset $\phi(x) =$ constant, for each solution ϕ of the partial differential equation.

Example. In $\mathbb{R}^2$, the equation is

$$v_1(x_1, x_2)\frac{\partial \phi}{\partial x_1} + v_2(x_1, x_2)\frac{\partial \phi}{\partial x_2} = 0.$$

In most cases, a solution to this equation will give a 1-manifold as the set of points $\phi(x_1, x_2) =$ constant, which can be taken as the image of a curve, given by this implicit equation. The difference between a solution to this equation and a solution to the ordinary differential equation is that no parameterisation of the curve is specified.

Symmetries of a differential equation. Lie groups of symmetries for a differential equation can often be used to reduce the number of equations or independent variables. In the simplest cases, the equations will reduce to an equation in one variable which can be solved by integration.

A transformation $\tau: M \to M$ is a symmetry of an equation if it transforms solutions to solutions. For an ordinary differential equation, this means that if the curve $c: \mathbb{R} \to M$ is a solution, then so is $\tau \circ c$. Suppose that these solutions arise from a flow, so that $c = \lambda^x$. Then $\tau \circ c$ must be the curve determined by the flow through the point $\tau(x)$, i.e., $\lambda^{\tau(x)}$. This condition can be written

$$\tau(\lambda(t,x)) = \lambda\big(t, (\tau(x))\big)$$

for all x, t, or

$$\tau \circ \lambda_t = \lambda_t \circ \tau.$$

Now suppose τ is itself any one of the transformations of a second flow λ'. Then

$$\lambda'_s \circ \lambda_t = \lambda_t \circ \lambda'_s.$$

for all $(s,t) \in \mathbb{R}^2$. Two flows which satisfy this condition are said to commute. The map on either side of this equality can be taken to define an action Λ of the group $\mathrm{T}(2) = (\mathbb{R}^2, +)$ on M.

Exercise 3. Show that $\Lambda_{(s,t)}\Lambda_{(u,v)} = \Lambda_{(s+u,t+v)}$, a condition for this to be an action.

Exercise 4. Write down two flows on $\mathbb{R}^2$ which are distinct and

(1) commute
(2) do not commute

Solution to exercise 1. The equation has the solution $c(t) = x/(1 - atx)$, which is defined only as long as $at < 1/x$. Clearly $c(t)$ is infinitely large as this limit is reached, and the solution does not exist for all t. Hence the solutions are not determined by a flow.

16. One-parameter subgroups

A one-parameter subgroup of a Lie group is a homomorphism $h: \mathbb{R} \to G$, i.e., a curve where $h(0) = e$ and $h(s)h(t) = h(s+t)$. Since $h(t) = \big(h(t/n)\big)^n$, the one-parameter subgroup is determined by its values for t arbitrarily close to 0. This section will show that it is in fact determined by its tangent vector at $h(0) = e$.

One-parameter subgroups play a central role in the theory of Lie groups. For example, if G acts on a manifold M by $\lambda\colon G \times M \to M$, then a one-parameter subgroup determines a flow by

$$(t, x) \mapsto \lambda\bigl(h(t), x\bigr).$$

For a fixed $x \in M$, define $\lambda^x\colon G \to M$ by $g \mapsto \lambda(x, g)$. Then the integral curve of the flow through x is $c(t) = \lambda^x\bigl(h(t)\bigr)$. The velocity vector is

$$\frac{\mathrm{d}c}{\mathrm{d}t}(0) = \mathrm{d}\lambda^x_e\left(\frac{\mathrm{d}h}{\mathrm{d}t}(0)\right) \in TM_x.$$

This vector field on M is determined completely by the tangent vector

$$\frac{\mathrm{d}h}{\mathrm{d}t}(0) \in TG_e.$$

Lemma. *A one-parameter subgroup h is determined uniquely by its tangent vector dh/dt at the origin.*

Proof. Let $M = G$ and the action $\lambda(g, x) = gx$ be the group multiplication. Then the flow is $(t, x) \mapsto h(t)x$ and the integral curve through e is just h itself. However, the flow is determined uniquely by its velocity vector field, which by the preceding argument is determined by $\mathrm{d}h/\mathrm{d}t(0)$.

Tangents to $\mathrm{GL}(n)$. Recall that $\mathrm{GL}(n) \subset \mathbb{R}^{n^2}$, the latter regarded as the set of all $n \times n$ matrices. Since it is an open subset, $T\,\mathrm{GL}(n)_x = \mathbb{R}^{n^2}$. A tangent vector can likewise be regarded as an $n \times n$ matrix.

For example, if $c\colon \mathbb{R} \to \mathrm{GL}(n)$ is a curve, with

$$c(t) = \begin{pmatrix} c_{11} & c_{12} & \cdots \\ c_{21} & \cdots & \cdots \\ \vdots & & \vdots \end{pmatrix}$$

then

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \begin{pmatrix} \frac{\mathrm{d}c_{11}}{\mathrm{d}t} & \frac{\mathrm{d}c_{12}}{\mathrm{d}t} & \cdots \\ \frac{\mathrm{d}c_{21}}{\mathrm{d}t} & \cdots & \cdots \\ \vdots & & \vdots \end{pmatrix} \in T\,\mathrm{GL}(n)_{c(t)}.$$

For $\mathrm{GL}(n)$, it is easy to show the converse of the preceding lemma

Lemma. *Any vector $A \in T\,\mathrm{GL}(n)_e$ is tangent to a one-parameter subgroup.*

Proof. The one-parameter subgroup is given by

$$h(t) = \exp(tA) = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}.$$

In this formula, A^n refers to the matrix product. The sum is easily seen to converge, and

$$\frac{\mathrm{d}h}{\mathrm{d}t}(t) = \sum_{n=1}^{\infty} \frac{nt^{n-1}}{n!} A^n = A \sum_{n=1}^{\infty} \frac{t^{n-1}}{(n-1)!} A^{n-1} = Ah(t),$$

so that $\mathrm{d}h/\mathrm{d}t = A$ when $t = 0$.

Exercise 1. Multiply the exponential series to show that $h(s)h(t) = h(s+t)$.

This shows that for $\mathrm{GL}(n)$, the one-parameter subgroups are in $1-1$ correspondence with elements of the tangent space at $e \in \mathrm{GL}(n)$.

17 Subgroups of $\mathrm{GL}(n)$

The following subgroups are all Lie groups.

Special linear group. The special linear group $\mathrm{SL}(n) \subset \mathrm{GL}(n)$ is the subgroup of matrices with determinant equal to one. Since

$$\det(\exp(A)) = \exp(\mathrm{trace}(A))$$

for any matrix A, then $\exp(A) \in \mathrm{SL}(n)$ if and only if $\mathrm{trace}(A) = 0$. Hence $T\,\mathrm{SL}(n)_e \subset T\,\mathrm{GL}(n)_e$ is the linear subspace given by the linear condition

$$\mathrm{trace}(A) = 0.$$

Orthogonal group. The orthogonal group $\mathrm{O}(n)$ is the group of $n \times n$ matrices which satisfy the condition

$$M^T M = e.$$

Suppose $c\colon \mathbb{R} \to \mathrm{O}(n)$ is a curve with $c(0) = e$ and tangent equal to A at e. Then at $t = 0$,

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\left(c(t)^T c(t)\right) = \left(\frac{\mathrm{d}c}{\mathrm{d}t}\right)^T c(0) + c^T(0)\frac{\mathrm{d}c}{\mathrm{d}t} = A^T e + eA = A^T + A,$$

so that A is an antisymmetric matrix. Conversely, if A is an antisymmetric matrix, then

$$(\exp A)^T = \exp A^T = \exp(-A) = (\exp A)^{-1},$$

so $\exp A$ is orthogonal. This shows that $T\,\mathrm{O}(n)_e \subset T\,\mathrm{GL}(n)_e$ is the linear subspace given by the condition $A^T + A = 0$.

The special orthogonal group, $\mathrm{SO}(n)$, is defined to be the intersection $\mathrm{O}(n) \cap \mathrm{SL}(n)$, the orthogonal matrices of determinant 1. Since det is continuous, and the only values it takes in $\mathrm{O}(n)$ are ± 1, any curve c which passes through the point e must have $\det(c(t)) = 1$ for all t. Therefore, $T\,\mathrm{SO}(n)_e = T\,\mathrm{O}(n)_e$.

The group $\mathrm{SO}(3)$ is called the rotation group. To justify this name, we prove the following

Theorem. *Every rotation has an axis.*

Proof. Let $M \in \mathrm{SO}(3)$. The characteristic polynomial of M has at least one real root, so that M has an eigenvector v. Since M is an isometry, the eigenvalue is ± 1. If the eigenvalue is 1, then v is the axis. The matrix M acts in the plane orthogonal to v. In an orthonormal basis which includes v, M is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}.$$

Since the determinant of M is 1, so is the determinant of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, which gives an element of $\mathrm{SO}(2)$, for which the formulae give explicitly a rotation in this plane.

If the eigenvalue is -1 then M acts in the orthogonal plane again, but with determinant -1. This gives a reflection in this plane, and the explicit formulae show that there is a reflection axis, $Mv = v$. This returns to the previous case with v the axis.

Unitary group.

By taking real and imaginary parts of the components of a complex vector, $\mathbb{C}^n$ can be regarded as $\mathbb{R}^{2n}$. Thus an invertible matrix with complex entries determines an element of $\mathrm{GL}(2n)$.

Example. An invertible 1×1 matrix is just a non-zero complex number $c = c_1 + ic_2$. This acts on $x = x_1 + ix_2 \in \mathbb{C}$ by multiplication of complex numbers. This amounts to the formula

$$(x_1, x_2) \mapsto \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

which defines an element of GL(2).

The unitary group, U(n), is defined to be the group of unitary $n \times n$ complex matrices. These satisfy $U^{-1} = \bar{U}^T$.

The special unitary group SU(n) $\subset$ U(n) is the subgroup of unitary matrices with determinant one.

Exercise 1. What complex numbers correspond to elements of U(1)?

Exercise 2. Show that elements of SU(2) can be written in the form $\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$, where a and b are complex numbers. Give a diffeomorphism SU(2) $\to S^3$.

18 The commutator

The condition that two flows commute can be written entirely in terms of their velocity vector fields. Let v and v' be the velocity vector fields of flows λ, λ' on a manifold M.

Definition. Let $M \subset \mathbb{R}^k$, and $v, v'\colon M \to \mathbb{R}^k$ be vector fields on M. The commutator of v and v' is defined to be the function $[v, v']\colon M \to \mathbb{R}^k$ given by

$$\mathrm{D}_v v' - \mathrm{D}_{v'} v.$$

The coordinate expression is that the j-th component of $[v, v']$ is

$$\sum_i v_i \frac{\partial v'_j}{\partial x_i} - v'_i \frac{\partial v_j}{\partial x_i}.$$

Theorem. *The two flows λ and λ' commute if and only if the commutator $[v, v']$ of their velocity vector fields is zero.*

Proof. Pick $x \in M$, and let $y(s,t) = \lambda'(s, \lambda(t,x))$. The behaviour of y for small values of s and t near 0 is determined by the second derivative $\frac{\partial^2 y}{\partial s \partial t}$ at $(0,0)$. This is calculated in the following way.

$$\frac{\partial y}{\partial s}(0,t) = v'\left(\lambda(t,x)\right),$$

and so

$$\frac{\partial^2 y}{\partial t \partial s}(0,0) = \mathrm{d}v'_x\left(\frac{\mathrm{d}\lambda}{\mathrm{d}t}\right) = \mathrm{d}v'_x\left(v(x)\right)$$

Likewise if $z(s,t) = \lambda(t, \lambda'(s,x))$, then

$$\frac{\partial^2 z}{\partial s \partial t} = \mathrm{d}v_x\left(v'(x)\right).$$

If the flows commute, then $y(s,t) = z(s,t)$. Since these mixed second order partial derivatives are equal, the result that $[v, v'] = 0$ follows.

Conversely, assume that $[v, v'] = 0$. According to a previous argument, it is sufficient to show that each transformation λ'_s is a symmetry of the ordinary differential equation determined by the vector field v,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\lambda'_s \circ \lambda^x\right) = v \circ \lambda'_s \circ \lambda^x, \quad \text{for all } s.$$

Note that at $s = 0$, this equation reduces to the defining equation for v, namely

$$\frac{\mathrm{d}\lambda^x}{\mathrm{d}t} = v \circ \lambda^x,$$

and so certainly holds.

Using the chain rule, the condition is equivalent to

$$\mathrm{d}\left(\lambda'_s\right)_x\left(v(x)\right) = v\left(\lambda'_s(x)\right), \quad \text{for all } s \in \mathbb{R}, x \in M.$$

Consider

$$\psi(s) = \mathrm{d}\left(\lambda'_s\right)_x\left(v(x)\right) - v\left(\lambda'_s(x)\right)$$

Some differentiation, and using the hypothesis $[v, v'] = 0$ shows that

$$\frac{\mathrm{d}\psi}{\mathrm{d}s} = \mathrm{d}v'_{\lambda'(s,x)}\left(\psi(s)\right).$$

Since $\psi(0) = 0$, this ordinary differential equation has the unique solution $\psi(s) = 0$ for all $s \in \mathbb{R}$.

Example. For the flow given by rotations of S^2 about the x_3-axis, the vector field is $v(x_1, x_2, x_3) = (x_2, -x_1, 0)$. Consider a second flow given by rotations about the x_1-axis. This vector field is $v'(x_1, x_2, x_3) = (0, x_3, -x_2)$. The commutator is given by

$$\left(\sum_i v_i \frac{\partial v'_1}{\partial x_i} - v'_i \frac{\partial v_1}{\partial x_i}, \sum_i v_i \frac{\partial v'_2}{\partial x_i} - v'_i \frac{\partial v_2}{\partial x_i}, \sum_i v_i \frac{\partial v'_3}{\partial x_i} - v'_i \frac{\partial v_3}{\partial x_i}\right) = (-x_3, 0, x_1).$$

This is not zero, so the flows do not commute.

The preceding theorem gives an interpretation of the vanishing of the commutator of two vector fields, in the case when they generate flows. The commutator of two vector fields is also important when it does not vanish. The main fact is

Theorem. *The commutator of two vector fields on M is also a vector field on M.*

This fact will be proved below. Note that in the example, calculating $\mathrm{D}_v v'(x)$ gives $(0, 0, x_1)$ and $\mathrm{D}_{v'} v(x) = (x_3, 0, 0)$, neither of which are tangent to S^2. Only the difference of these gives a vector field on S^2.

The commutator has an interpretation in terms of vector fields acting on functions. If v and v' are thought of as operators acting on functions, then taking an arbitrary function $\phi\colon M \to \mathbb{R}$, we calculate the difference of v' acting followed by v acting and v acting followed by v' acting. This is also called a commutator, namely the commutator of the first order differential operators given by v and v'.

$$\begin{aligned}\mathrm{D}_v(\mathrm{D}_{v'}\phi) - \mathrm{D}_{v'}(\mathrm{D}_v\phi) &= \sum_{ij} v_i \frac{\partial}{\partial x_i}\left(v'_j \frac{\partial \phi}{\partial x_j}\right) - v'_i \frac{\partial}{\partial x_i}\left(v_j \frac{\partial \phi}{\partial x_j}\right) \\ &= v_i \frac{\partial v'_j}{\partial x_i}\frac{\partial \phi}{\partial x_j} - v'_i \frac{\partial v_j}{\partial x_i}\frac{\partial \phi}{\partial x_j} = \mathrm{D}_{[v,v']}\phi\end{aligned}$$

as the terms involving second derivatives of ϕ cancel. So the commutator of the first order differential operators acting on ϕ is just $[v, v']$ acting on ϕ.

This gives an argument about why the commutator is a vector field. If the manifold is defined by an equation $\phi = 0$, then certainly $\mathrm{D}_{v'}\phi = 0$ and $\mathrm{D}_v(\mathrm{D}_{v'}\phi) = 0$, and so $\mathrm{D}_{[v,v']}\phi = 0$. Hence $[v, v']$ is tangent to M. For S^2, $\phi = x \cdot x - 1$, which explains the example.

More generally, we could take $M \subset \mathbb{R}^k$ to be given by an equation $\phi = 0$ for some $\phi\colon \mathbb{R}^k \to \mathbb{R}^l$. This would give a proof if we knew that all manifolds can be defined in this way (which they can, at least locally). A slight modification of this idea gives the proof.

Proof of theorem. For each point $x \in M \subset \mathbb{R}^k$, there is locally a coordinate function f from M to $\mathbb{R}^m$, and its inverse, a parameterisation π, from $\mathbb{R}^m$ to M. ('Locally' means that these are defined on open subsets around x or its image.) Let $\psi\colon \mathbb{R}^k \to M$ be $\pi \circ F$, where F is any local extension of f to $\mathbb{R}^k$. Then a vector v is a tangent vector to M if and only if

$$\mathrm{d}\psi_x(v) = v$$

This is because ψ is the identity map on M, and TM_x is the image of $\mathrm{d}\pi_x$. This equation is equivalent to writing

$$\mathrm{D}_v\psi = v$$

for a vector field defined locally (i.e., on the open subset of M where ψ is defined).

Then

$$\mathrm{D}_{[v,v']}\psi = \mathrm{D}_v(\mathrm{D}_{v'}\psi) - \mathrm{D}_{v'}(\mathrm{D}_v\psi) = \mathrm{D}_v v' - \mathrm{D}_{v'}v = [v,v'],$$

so that $[v,v'](x) \in TM_x$.

The mapping formula.

Let v be a vector field on M, and $f\colon M \to N$ be a mapping, and w a vector field on N.

Then w is said to be f-related to v if $w(f(x)) = \mathrm{d}f_x\big(v(x)\big)$. For example, if f is a diffeomorphism, then given v, there is a unique f-related vector field on N called the induced vector field

$$w(y) = \mathrm{d}f_x\big(v(x)\big), \quad \text{where} \quad x = f^{-1}(y).$$

Exercise. Show that if $g\colon N \to P$ is another diffeomorphism and z is the vector field induced on P from w, then z is equal to the vector field induced by $g \circ f$ from v.

The mapping formula for the commutator is:

Lemma (Mapping formula). *If $f\colon M \to N$ is a mapping, v, v' are vector fields on M, and w on N is f-related to v, w' f-related to v', then $[w,w']$ is f-related to $[v,v']$.*

Proof.

$$\mathrm{d}f_x([v,v'](x)) = \mathrm{D}_{[v,v']}f$$

evaluated at the point x. But this function is

$$\mathrm{D}_{[v,v']}f = \mathrm{D}_v(\mathrm{D}_{v'}f) - \mathrm{D}_{v'}(\mathrm{D}_v f) = \mathrm{D}_v(w' \circ f) - \mathrm{D}_{v'}(w \circ f)$$

At the point x, the right-hand side is

$$\begin{aligned}
&\mathrm{d}(w' \circ f)_x(v(x)) - \mathrm{d}(w \circ f)_x(v'(x)) \\
&\qquad = \left(\mathrm{d}w'_{f(x)} \circ \mathrm{d}f_x\right)(v(x)) - \left(\mathrm{d}w_{f(x)} \circ \mathrm{d}f_x\right)(v'(x)) \\
&\qquad\qquad = [w,w'](f(x)).
\end{aligned}$$

19. Computer vision

An optical image is the projection of a three-dimensional scene under a smooth map $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$. A variety of possible π's can occur, depending on the camera. It is assumed that various features (points, lines, curves, surfaces, corners, smooth singularities, etc.) can be recognised in the image. The '3-d recovery problem' is to say where these features are in $\mathbb{R}^3$ which give rise to the optical image.

An assumption is usually made about the nature of the scene, as a hypothesis which can be then given a mathematical formulation. For example, it could be that the scene is a face, a microscope slide containing cells, an aerial photograph, or a stack of books to be counted. From this assumption, you have a hypothesis about the scene containing a number of continuous parameters.

For example, if the scene is a cell which is a assumed to be spherical, then the parameters might be the radius of the cell and the position of its centre. If the cell is not assumed spherical, then additional parameters would be needed for its shape and its angular orientation in space.

The 3-d recovery problem can be thought about in two ways

(1) 3-d Euclidean geometry. Find the set of all objects in $\mathbb{R}^3$ which could give rise to the image. For example, if the scene consists of rigid bodies (bodies for which the distance between the constituent parts does not change) then the Euclidean group $E(3)$ acts on the set of all possible positions for each body, and the recovery problem would reduce to finding the possible Euclidean transformations which take each object from a standard position to its actual position.

(2) 2-d Non-Euclidean geometry. For each possible three-dimensional object, find the range of possible images. See which of these fits the given image. For example, if the camera gives a projection along straight lines, then the image can be regarded as a parameterisation of part of the projective plane, P^2. Projective transformations can be applied to the images.

More information can be gained if the image varies with time. One of the ways of converting this information into an easily usable form is to look at the velocity vector for each point in the scene, or image. The velocity vector of the motion of the points in the image is called the optical flow. The geometry of the scene and the camera projection places constraints on the set of possible optical flows.

The 3-d recovery problem in this instance is the determination of the scene at time t given the image at time t and the optical flow at time t. This problem is often called 'shape from motion'. More information can be gained about the scene if one knows the optical flow at a point in time as well as just the image.

In practice, the calculations would be done with a computer. In all but the simplest situations, the equations are exceedingly complicated and do not have a simple solution which can be obtained on paper. A computer program might also take into account other attributes for the image, such as colour, texture, shading, shadows, or statistical data.

Shape from motion.

The motion of a rigid body in in $\mathbb{R}^3$ is given by transformations of $\mathbb{R}^3$ in the Euclidean group $E(3)$. This is

$$x \mapsto Mx + a,$$

where $M \in \mathrm{O}(3)$ and $a \in \mathbb{R}^3$. Suppose that these vary with time $t \in \mathbb{R}$, such that $M =$ identity, and $a = 0$ at $t = 0$. Clearly, a curve in the Euclidean group $E(3)$ is equivalent to a curve $M(t)$ in $\mathrm{O}(3)$ and a curve $a(t)$ in $\mathbb{R}^3$. The curve through point x is $c(t) = M(t)x + a(t)$, which has tangent

$$\frac{\mathrm{d}c}{\mathrm{d}t}(0) = \left(\frac{\mathrm{d}M}{\mathrm{d}t}(0)\right) x + \frac{\mathrm{d}a}{\mathrm{d}t}(0).$$

Define the matrix

$$\Omega = \frac{\mathrm{d}M}{\mathrm{d}t}(0)$$

and the vector

$$\xi = \frac{\mathrm{d}a}{\mathrm{d}t}(0).$$

As Ω is a tangent to $\mathrm{O}(3)$ at e it is an antisymmetric matrix. As there is a curve in $O(3)$ with any antisymmetric matrix as tangent, Ω can be any antisymmetric matrix. Likewise, ξ is tangent to $\mathbb{R}^3$ and can be any vector. Thus the vector fields which can arise as velocity vector fields for rigid body motions are $v\colon \mathbb{R}^3 \to \mathbb{R}^3$

$$v\colon x \mapsto \Omega x + \xi.$$

Exercise 1.

(1) Write $\Omega = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}$. Write out Ωx in components and show it is the vector cross product of ω and x.

The scene is a surface in $\mathbb{R}^3$, the plane (an affine subset) given by

$$Z = pX + qY + r,$$

and the camera is the projection $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$(X, Y, Z) \mapsto (X, Y).$$

(2) The points in the plane move by a rigid body motion depending on a parameter t, time. Explain why the plane remains a plane for all times t.

It is assumed that the plane can always be described by $Z = pX + qY + r$.

(3) Does this assumption place any restriction on the rigid body motions?

The points in the plane can be parameterised by the corresponding points in the optical image, $\mathbb{R}^2$, by a map $\phi\colon \mathbb{R}^2 \to \mathbb{R}^3$, so that

$$\pi(\phi(X, Y)) = (X, Y).$$

(4) Write an explicit formula for ϕ.

As the points in the plane move, so do the points in the optical image. This is given by

$$(X, Y) \mapsto \pi\big(M\phi(X, Y) + a\big).$$

(5) Explain why this is the correct formula. Differentiate this expression with respect to t, at $t = 0$, assuming as above the $M(0) = $ identity, and $a(0) = 0$, and obtain a vector field w on $\mathbb{R}^2$, the optical flow.

As a special case, you should get for $r = 0$

$$(X, Y) \mapsto (\xi_1, \xi_2) + \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ pX + qY \end{pmatrix}.$$

(6) Explain why it should be impossible to determine r from the optical image or the optical flow. Why does r appear in your formula?

Now make the simplifying assumption that $r = 0$ for all time. From measuring the optical flow you can determine the parameters A, B, C, D, E, F in an optical flow

$$(X, Y) \mapsto (E, F) + \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

(7) Look at the four diagrams of an optical flow. Give parameters which give formulae for these optical flows.

(8) Express A, B, C, D, E, F in terms of p, q and ω. Show that these equations are solved by

$$\begin{aligned} \omega_3 &= \frac{1}{2}\left(R \pm \sqrt{|S|^2 - T^2}\right) \\ \omega_1 + i\omega_2 &= k \exp i \left(\frac{\pi}{4} + \frac{1}{2}\arg(S) - \frac{1}{2}\arg(2\omega_3 - R - iT)\right) \\ p + iq &= \frac{1}{k} S \exp i \left(\frac{\pi}{4} - \frac{1}{2}\arg(S) + \frac{1}{2}\arg(2\omega_3 - R - iT)\right), \end{aligned}$$

where

$$\begin{aligned} T &= A + D \\ R &= C - B \\ S &= (A - D) + i(B + C), \end{aligned}$$

k is indeterminate, and there are two solutions, $\pm$, for each choice of k.

(9) Interpret R, S, T in the four diagrams of optical flow.

(10) Suppose the mapping of the surface to the optical image had not been non-singular. Would you expect the optical flow to be a smooth vector field?