3P2 Manifolds

Michael D. Alder

April 16, 2003

Contents

Chapter 1

Introduction

1.1 A Cautionary Tale

I try to start every course of lectures with an overview or outline of the material in the course. Before I do that, a (true) cautionary tale.

While I was an undergraduate doing Mathematics, I told my Calculus lecturer, a celebrated mathematician called Ian Porteous:

"I have been getting the strong sensation in your course of being dragged at high speed down a narrow track in a jungle. There have been dimly sensed paths off to the sides but we have been galloping after you and have no time to explore these possibilities."

I went on:" What I would like, is to be taken up to a high place and given a view over the country through which we have been and that where we are headed, an overview of the subject."

He replied with an amused grin,

"Wouldn't we all."

He elaborated the point, which was that in order to have an overview of a subject in Mathematics, you have to crawl all over it on your hands and knees, and then you move next door and do the same with another mathematical subject, and then, if the subjects are related, you may get a higher level view of the two bits you have sorted out in detail. And then you can do some more detailed understanding of some other bits and link those, and maybe get a higher order insight linking the links. But a higher order view from a height without doing the detailed work is not possible.

The reason it is not possible is that you are looking at ideas. You can only develop a higher level idea by understanding the base level. We could tell you the words for the higher level ideas, but they wouldn't mean anything. For this reason, course outlines in Mathematics are intelligible only after you have done the course, not before. This is extremely maddening, particularly to philosophers, journalists, post-modernists and others with similar intellectual handicaps, but that's the way things are.

It follows that explaining what the course is about in a general way is a waste of time and can't really be done, but I shall do it anyway.

1.2 Years of Calculus

In first year at University Mathematics you did again what you did at school, you studied differential and Integral calculus of functions from $\mathbb{R}$ to $\mathbb{R}$. [1]

In second year you did the Differential and Integral Calculus for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ You were introduced to differential forms and integrated vector fields or differential 1-forms over curves, and 2-forms over surfaces. You found that the trick usually was to reduce each case to some collection of cases you had done before, so, for example, integrating a function over a disc was really doing two ordinary first-year style integrals and quoting Fubini's Theorem.

In third year we continue with the development. Now we are going to be concerned with doing Calculus on Manifolds, that is looking at what it means to have a smooth function $f : P^n \longrightarrow Q^m$ where P and Q are *manifolds* (which you might have guessed from the course title), which leaves you wondering what a manifold is, and perhaps a strong suspicion that we are not talking about the insides of cars here. We shall be looking at differential forms and vector fields on Manifolds. So I need to give you some idea of what a manifold is.

[1] This time we hope you got it right.

1.3 Manifolds

The circle, S^1 is a 1 dimensional manifold. The reason it is of dimension 1 is that locally a little bit of it looks like an interval in $\mathbb{R} = \mathbb{R}^1$. The real line $\mathbb{R}$ itself is also a 1-manifold. (Short for 1-dimensional manifold.)

The sphere, $\mathcal{S}^2$ is a 2-manifold. So is T^2 the torus, and $\mathbb{R}^2$ also. Recall

$$S^2 = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1 \right\}$$

We use the term *curve* for a 1-manifold, *surface* for a 2-manifold.

In general there are n-manifolds for any positive integer n. A 0-manifold is just a discrete set of points and therefore not very interesting from the point of view of calculus. $\mathbb{R}^n$ is always an n-manifold. There are lots more.

Manifolds are not studied because we want to play with Möbius strips, although they give us an excuse. They arise naturally in most branches of Science and Engineering.

Example 1.3.1. In Classical Mechanics (which now has to be taken to include the study of the stability of the solar system, or will all the planets fall into the sun, to robotics and control theory, or how to build a mars rover that won't spend all its time trying to push a rock over) we are often much concerned with the issue of what happens when we rotate something in three dimensions. This draws attention to the linear maps from $\mathbb{R}^3$ to $\mathbb{R}^3$ which preserve distances and angles, the 'rigid' linear maps. They are conveniently represented by the 3×3 matrices having columns which are mutually orthogonal and of length 1, and where the determinant of the matrix is 1. It is immediate that there is a subspace of $\mathbb{R}^9$ (the entries in the matrix) consisting of these particular entries coresponding to rotations. Just as we got a two dimensional surface by putting one smooth condition on the three variables of $\mathbb{R}^3$, and just as any[2] smooth condition cuts down the dimension by one (a statement you are familiar with as the Rank-Nullity theorem when the constraints are not just smooth but linear) we have six conditions on the nine numbers and wind up with a space of matrices which has dimension 3. Thus just as the 2-sphere comes sitting in $\mathbb{R}^3$, this three dimensional space is sitting in $\mathbb{R}^9$. The space is called SO(3) and it is not just a manifold it is a group. In fact the operations of inversion and multiplication are continuous

[2] Well almost any. At least, quite often.

and indeed smooth. Thus it is a 3-manifold which is called a *Lie Group*. Lie[3] Groups are named after Sophus Lie. Not to be confused with Sophy Lee. And at least you now know how to pronounce the expression 'Lie Group'.

You are going to have to take my word for most of the claims here for the time being.

Quite generally, we may wish to measure n real numbers to determine the state of a system, but it may often happen that the numbers are not independent. In this situation the state space is generally a manifold.

Example 1.3.2. It is often desired to find lines (and other things) in images. There is a cool way of doing this called the Hough[4] Transform which works as follows: We go from the image to a space of lines in $\mathbb{R}^2$. We may choose to parametrise the lines by slope and intercept so if $y = mx + c$ is the line we work in the $m - c$ space. So a point in the $m - c$ space is a line in $\mathbb{R}^2$. A point $(x_1, y_1)^T$ in the image space (identified as a subset of $\mathbb{R}^2$) determines a set of points in the $m - c$ space, those lines which pass through the point $(x_1, y_1)^T$, namely

$$\left\{(m, c)^T : y_1 = mx_1 + c\right\}$$

It is easy to see that this is a line in the $m - c$ space.

In figure 1.1, I sketch the idea of the image space and the line space.

So points in the image space determine lines in the $m - c$ space. And if two points lie on a single line, the two lines corresponding in the $m - c$ space to the two points will intersect in a unique point in the $m - c$ space, which in turn is the unique line through the two points in the image space. So if we have lots of points lying along a line in the image, we shall get a lot of lines intersecting in the same point in the $m - c$ space. So we find the lines in the image space by taking each point (pixel) that is a possible pixel belonging to the line, generate the corresponding line in the $m - c$ space, and then add the line for the next pixel. We keep a sort of accumulator of the sum of entries of approximate locations in the $m - c$ space, and where we have a peak in the $m - c$ space we have a line in the image space.

There is an obvious problem with the method as described, *viz*, what about vertical lines which don't get represented in the $m - c$ space? Well, we can either choose *two* spaces, one perhaps obtained by turning the image

[3]Pronounced to rhyme with pee and not pie

[4]Rhymes with Rough and Tough, not Bough, Cough, Lough, Through or Sough

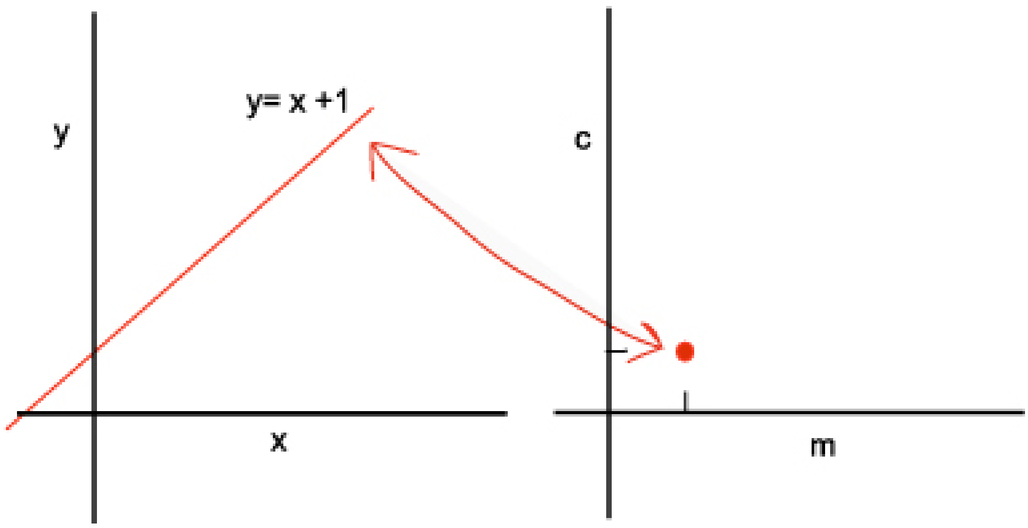

Figure 1.1: Idea of the Hough Transform

sideways, which will give twice the calculations, or we can find a better way of parametrising the space of lines. One way would be to take, for any line in the plane, r the closest distance of the line to the origin, and θ, the angle of the line segment of length r to the X-axis. This means that a point in the image space will now determine a curve (not a straight line) in the $r - \theta$ space, and I leave the inquisitive among you to work out it's shape.

Now any one of the many different possible parametrisations of the space of lines in the plane is always going to have problems associated with it. And the reason is that this space is rather an odd one. It is of dimension two, and is a manifold. You have probably never heard of it before: it is called $\mathbb{RP}^2$.

To give you a description of the space, note that we can give a full specification of the lines in $\mathbb{R}^2$ by writing them as

$$ax + by + c = 0$$

Now this makes it look as though the space is three dimensional, a, b, c, but we quickly realise that multiplying the triple (a, b, c) by any non-zero number will give us the same line. So we have that a point in the line space is a line through the origin in $\mathbb{R}^3$.

We can make this a bit simpler if we insist that $a^2 + b^2 + c^2 = 1$, and say that the space of lines in the plane is the set of points on the unit sphere, but with antipodal points regarded as glued together. Note that when $a = b = 0$ and $c = \pm 1$ we get a point which corresponds to no actual line in $\mathbb{R}^2$. We

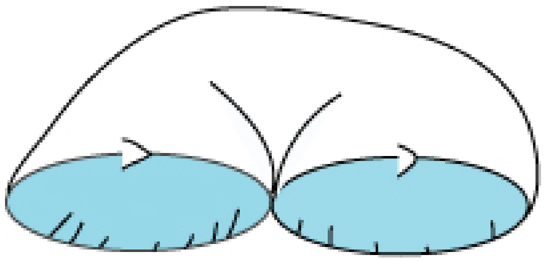

Figure 1.2: Stage in making $\mathbb{RP}^2$

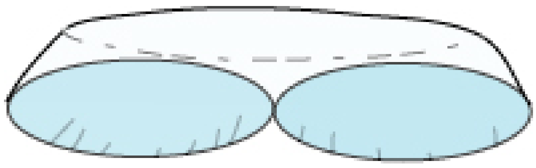

Figure 1.3: Cutting out lines far, far away

call this the *line at infinity* and it turns up in projective geometry.

We can simplify even more by chopping off the lower half of the sphere and saying that the space is now a hemisphere, with opposite points on the boundary glued together. If we glue one pair of opposite points we get a thing like figure 1.2. We now have to glue the two circles together. If one of the circles had the direction in which the gluing was to be done reversed, it would look like a squashed torus, but it doesn't. It is impossible to make the resulting object in three dimensions without it cutting itself.

Since the line at infinity will not appear on the image, and neither will lines close to it, we may confidently take away the north pole of the hemisphere, and indeed cut out a disc around it and throw it away. This gives something like figure 1.3 (which looks a bit like a pair of underpants) and by moving the surface around a bit you can now do the identification of those two circular regions. I leave it to you to work out what reasonably familiar shape you wind up with. You can do it by sewing together the legs of your underpants, although this makes wearing them impractical. The resulting shape is not a manifold, it is a *2-manifold with boundary* along with things like discs and discs with holes in them.

1.4 Calculus on Manifolds

I shall be obliged to skip some of the material listed in the Handbook entry, as covering all of it is impractical. Still, we are going to need to be able to talk about vector fields, and tensor fields on manifolds.

You might like to think about the problems that this will involve. If you take U to be an open subset of $\mathbb{R}^2$ or $\mathbb{R}^3$ then you will have no difficulty defining a vector field on U. But what if U were a 2-sphere? It is intuitively natural to believe we can draw tangent vectors to points on a 2-sphere, but how does one say it in algebra?

In order to get this sorted out we shall need to discuss the idea of the *tangent bundle*, and indeed we shall need to define *tensor bundles* on manifolds.

Also in the course, we learn how to multiply vector fields on manifolds to get another vector field, the Lie bracket. This is used in non-linear control theory as you can see by checking the references. We may investigate Stokes' Theorem on Manifolds and Manifolds with boundary, we look at the Riemannian metric tensor, and we shall also look at some topological results including the Brower Fixed Point Theorem, which says that if you have a matchbox full of plasticene, and if you take out the plasticene, stretch it, tie knots in it, compress it *but do not tear it* and then squash it back in the matchbox, then there is at least one point of the plasticene which is in the same position in the box afterwards as it was before. This result works for any dimension, and a consequence of this is that we can prove an existence result for ordinary differential equations.

There, lots of names you don't recognise and ideas you don't understand. But I did warn you about that.

1.5 Objectives

Since I am a mathematician, my main objective is to point out that we have some interesting ideas here, in the hope that some of you may become Mathematicians, and that those who settle for second best and become Physicists, Chemists or Engineers will at least be able to understand their subject properly, and maybe pay me huge sums of money when, later in life, they come across problems they can't handle but will, as a result of doing this course, recognise as mathematical ones. One of my former students now works on medical image analysis in Silicon Valley and regularly asks questions of a mathematical nature. She keeps me supplied with large cigars as a sort of retainer, and I need more former students like this.

Part of my job then will be getting you to the point where you can read, independently, the advanced text books and research publications you will certainly have to make sense of later on, unless your job is one of mindlessly boring routine.

A large part of my job is to wean you off the idea that an education consists largely of memorising the right things to say in the appropriate circumstances. This may indeed be what most arts students get out of their courses, but it will not cut the mustard in Mathematics. Memorising recipes is not good enough; it may possibly have got you through second year, barely, but it will not work in third year. You will need to follow the logic of complicated arguments and have to produce some of your own. For that reason I see an important element of the course as persuading you to enjoy proofs. I want you to admire their cold clarity and to produce some of your own. The good news is that those you meet will be relatively easy[5]. Mastering the art of proving things is closely related to the skills of cutting through to the core of arguments and ideas and seeing the significant. It should improve your capacity to express yourself with lucidity and force in ordinary language when that may be necessary. As, for example, in persuading your boss to give you a pay rise.

I shall again play fair. The examinations will be based largely on the material given in exercises, just as in 2C2. Only this time I shall not provide solutions to most of them. That will be your job.[6] I shall think it quite fair to give minor variants of exercises where, if you have worked out the solutions yourself you will see the ideas easily, and if you have memorised someone else's solution you won't. I shall not expect much in the way of thought in the examination, because an examination is a stressful environment not much conducive to thought. But I shall test for signs of thought having been done.

I warn strongly against the practice of spending the days taking down lecture notes with your brains out of gear and your nights watching television, cramming all your swotting into the three days before the exam. Doing all the exercises in three days is not feasible. Memorising the notes is pointless.

[5]Mostly.

[6]And in next year's courses, you will have to make up your own exercises. Then you should be well on the way to being able to read *real* books.

They are there to suggest how to do the exercises and to tell you why they are worth doing. They are not incantations to ward off evil.

I gave this warning to an Honours unit in Topology one year: two students got grades in the nineties, the others barely passed. I asked one of the latter group why, and he said he had not done enough work throughout the semester. I pointed out that I had warned him at the beginning, and he replied: ‘Oh yes, but all our lecturers have always told us that, and they were lying. We assumed you were lying too. But you weren’t.’

I’m not now either.

1.6 Useful References

1. *Calculus on Manifolds*, by Michael Spivak, Benjamin Books is an accessible introduction to differential forms and does an excellent job of proving Stokes’ Theorem. It contains a significant fraction of the material I would like to cover but may not be able to, but the meat is in the exercises.

2. *Topology from the Differentiable Viewpoint* by John Milnor, Princeton University Press is a nice cheap book you can get from Amazon which has lots of great material which I shall be using as much as time permits.

3. *Morse Theory* by John Milnor and Michael Spivak gives a beautiful treatment of part of the course and also goes beyond it.

4. *Differential Topology, First Steps* by Andrew Wallace, Benjamin Books, overlaps with all the above books and goes beyond the course, giving pointers to what follows.

5. *Analysis on Real and Complex Manifolds* by Raghavan Narasimhan, Masson et Cie, North-Holland is not for the faint hearted. You should however be able to read it by the end of the course, which will more or less follow the first two chapters. (There are only three chapters in the book!)

6. Lecture notes on Control Theory by Craig Woolsey of Virginia Tech. can be found at:

   `http://www.aoe.vt.edu/~cwoolsey/courses/AOE5984/Lectures/Lectures.htm` and some of the later lectures are relevant and interesting, showing the reason for our studying some of the material. I draw to your particular attention lectures 33-36.

7. *Foundations of Mechanics* by Ralph Abraham and Jerrold Marsden also gives some reasons for studying this material.

8. Marsden has also got some web-notes up for a course called *Control and Dynamical Systems* at

   `http://www.cds.caltech.edu/~marsden/bib_src/mta/Book/`

   which appear to be from the book *Manifolds, Tensors, Analysis, and Applications* Second Edition. Springer-Verlag, 1988; R. Abraham, J. E. Marsden, and T. Ratiu. I don't imagine they will stay there long. His course is similar to the present 3P2 but makes more demands upon the student. Note that this evidently makes the present course one in Applied Mathematics.

9. *Gravitation* by Misner, Thorne and Wheeler is a mammoth tome which deals with the general theory of Relativity, known to mathematicians as differential geometry, and which uses the tools we shall cover, and many others besides.

10. *Applied Exterior Calculus* by Edelen goes way beyond the course but again, you will be able to make some sense of it when the course is finished, and glancing at it will explain to some extent why this material is important. I warn you now that his definition of the Tangent Space is confused with the space of all vector fields in a way which suggests a serious muddle in his thinking; nevertheless he covers some important material.

A glance at the last five items, apart from giving you some insight into how much you don't know, will give you perhaps some reasons for wanting to do the course. There are lots of other reasons, but you can find out about those on your own.

Chapter 2

Topological Preliminaries

2.1 Topological Spaces

I shall put on the website the notes from last year's 2C2 and anyone who wants a copy can download it.

In particular you will recall the definition of a *metric space* as a set X and a map $d : X \times X \longrightarrow \mathbb{R}$ satisfying the rules:

$$\forall\, x, y \in X, \qquad d(x,y) \geq 0 \quad \text{and} \quad d(x,y) = 0 \Rightarrow x = y \tag{2.1}$$

$$\forall\, x, y \in X, \qquad d(x,y) = d(y,x) \tag{2.2}$$

$$\forall\, x, y, z, \in X, \qquad d(x,z) \leq d(x,y) + d(y,z) \tag{2.3}$$

Example 2.1.1. The usual Euclidean metric on $\mathbb{R}^2$ *is* a metric.

Example 2.1.2. Define d on $\mathbb{R}^2$ by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and similarly for $\boldsymbol{y}$.

Exercise 2.1.1. Prove that the above really *are* metrics. You may find the 2C2 book a help here.

Exercise 2.1.2. Prove that if we replace max by min in the last definition we do not get a metric.

You will also, I hope, recall the definition of continuity of a map f from a metric space (X,d) to another (Y,e):

$f : (X, d) \longrightarrow (Y, e)$ is *continuous* at $a \in X$ iff

$$\forall \varepsilon \in \mathbb{R}^+, \exists \delta \in \mathbb{R}^+, \forall x \in X :\ d(x, a) < \delta \Rightarrow e(f(x), f(a)) < \varepsilon$$

And, of course, f is continuous iff it is continuous at a for every $a \in X$. We can rewrite this in slightly more visual terms by defining the *open ball* in a metric space (X,d):

Definition 2.1.1.

$$B(a, \delta) = \{x \in X : d(x, a) < \delta\}$$

is called the *open ball* on a point $a \in X$ of radius δ

Then we have (too trivially for it to demand a proof, but check that what I say is true):

Proposition 2.1.1. *$f : (X, d) \longrightarrow (Y, e)$ is* continuous *at $a \in X$ iff for every open ball V on $f(a)$, there is an open ball U on a such that*

$$f(U) \subseteq V$$

□

The idea of a metric space is that we have extracted exactly the properties that we keep using in arguments about distance and continuity. d is called the *metric* and most students have no great difficulty with the idea once they have got over the shock and horror of realising that they are talking about a few million different things all at once.

This process of abstraction can be carried even further and because the idea of a topological space is so simple and so useful in many different areas I shall explain how we come to it.

Definition 2.1.2. An *open set* in a metric space (X, d) is a subset $U \subseteq X$ such that for every point $a \in U$, there is an open ball on a contained in U.

It follows immediately that every open set is a union of open balls. It also follows that the empty set is open in (X, d), and that the entire space X is open in (X, d).

It is not immediately obvious, but very nearly, that every open ball is an open set:

Proposition 2.1.2. *Every open ball B(a,r) in a metric space (X,d) is open in V.*

Proof:

□

The following propositions, like the last, are easy examples of proofs and you should read them carefully and follow the logic because very soon you will have to make up your own. I have suppressed the d because it can easily be inserted when you feel the need:

Proposition 2.1.3. *If $U_j : j \in J$ is a collection of subsets of a metric space X, and if every U_j is open in X, then the union*

$$\bigcup_{j \in J} U_j$$

is open in X.

Proof:

□

Proposition 2.1.4. *If U, V are open subsets of a metric space X then*

$$U \cap V$$

is open in X.

Proof:

□

Exercise 2.1.3. Prove that the whole space X is open in itself and the empty set $\emptyset$ is open in X

This leads to a simple test for continuity of f:

Proposition 2.1.5. *$f : X \longrightarrow Y$ is continuous iff whenever $V \subseteq Y$ is open in Y, $f^{-1}V$ is open in X*

Proof:

□

Now we observe that two metrics may not be very different. Take d_1 on $\mathbb{R}^2$ to be the Euclidean metric and d_2 to be the metric given in example 2.1.2. Then

Proposition 2.1.6. *Every open set in ($\mathbb{R}^2, d_1$) is open in ($\mathbb{R}^2, d_2$) and vice versa.*

Proof:

□

Thus the identity map is continuous in both directions, and every map from $(\mathbb{R}^2, d_1)$ into any space is continuous iff the same map from $(\mathbb{R}^2, d_2)$ is continuous, and vice-versa, likewise for maps into $\mathbb{R}^2$.

Definition 2.1.3. If $f : X \longrightarrow Y$ is continuous and bijective (1-1 and onto) and its inverse is also continuous, we say f is a *homeomorphism* and that X and Y are *homeomorphic.*

Exercise 2.1.4. Show that the closed interval $[a, b] \subset \mathbb{R}(a < b)$ and the closed interval $[c, d] \subset \mathbb{R}, (c < d)$ are homeomorphic. Show that the open intervals $(a, b), a < b$ and $(c, d), c < d$ are homeomorphic.

Exercise 2.1.5. Show that the open interval $(0, 1) \subset \mathbb{R}$ is homeomorphic to $\mathbb{R}$.

Exercise 2.1.6. Show that if A is homeomorphic to B then B is homeomorphic to A

Exercise 2.1.7. Show that every space is homeomorphic to itself

Exercise 2.1.8. Show that if A is homeomorphic to B and B is homeomorphic to C, then A is homeomorphic to C.

Exercise 2.1.9. Do you believe that $[0, 1]$ is homeomorphic to $\mathbb{R}$? Give reasons.

For many purposes we do not care about the difference between two spaces when they are homeomorphic. It follows that provided the open sets on a space are given, we can do away with the metric! Certainly as far as the quesion of continuity of maps is concerned.

We are led to what strikes many students as a ruthless throwing away of a lot of bathwater which surely must contain a baby or two.

Definition 2.1.4. A *topological space* is a set X and a collection $\mathcal{T}$ of subsets of X such that

1. $\emptyset \in \mathcal{T}$

2. $X \in \mathcal{T}$

3. Whenever $\{U_j : j \in J\}$ is a collection of sets in $\mathcal{T}$ then

$$\bigcup_{j \in J} U_j \in \mathcal{T}$$

4. $\forall\, U, V \subseteq X,\ U, V \in \mathcal{T} \ \Rightarrow U \cap V \in \mathcal{T}$

The sets in $\mathcal{T}$ are called the *open* sets of the topology.

Then we have shown already:

Proposition 2.1.7. *Every metric space is a topological space* □

Remark 2.1.1. We do not do this to be capricious: there are many spaces which are topological spaces which are not in any useful way metric spaces, for example, the space of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. A significant advantage of concerning ourselves with open sets is that the proofs are usually a lot easier. This is because we have stripped down the idea to its bare essentials.

Remark 2.1.2. Students who love babies often feel very uneasy about this level of abstraction, but it is something you get used to. To get used to it quickly and painlessly, work through a couple of proofs. It is very soothing and beats the axioms for a topological space into your head with minimum effort.

Remark 2.1.3. Note how a theorem about things on one level (metric spaces) gets turned into a definition at a higher level of abstraction.

Definition 2.1.5. A subset A of a topological space $X\mathcal{T}$ is *closed* iff its set complement $X \setminus A$ is open.

Definition 2.1.6. For any subset U of a topological space X, the *closure of* U, written $\bar{U}$ is defined to be the intersection of all closed sets containing U

Remark 2.1.4. It is easy to see that since arbitrary unions of open sets are open, then arbitrary intersections of closed sets are closed. Hence the closure of any set is a closed set.

Exercise 2.1.10. Prove the above remark carefully.

2.2 Separation Axioms

One of the first things one should do when one meets a new definition is to construct simple examples, so we first look at topologies on finite sets. First the empty set, $\emptyset$. Is it a topological space? The only subset is the set $\emptyset$ itself, so there is no choice at all for $\mathcal{T}$, it has to contain only the empty set. Now we confirm that the axioms for a topological space are satisfied by

$$X = \emptyset, \mathcal{T} = \{\emptyset\}$$

Axioms 1 and 2 are satisfied immediately (by the same set) and there being only one set in $\mathcal{T}$ makes the question of unions and intersections trivial, so the last two axioms are also satisfied and so we have a Topological space. Big deal.

If we have a one point set, $X = \{0\}$ the axioms are equally satisfied when $\mathcal{T}$ is the set $\{0, \emptyset\}$, and this is the only possible choice.

For a two point set, $X = \{0, 1\}$ we have more choice. We can let $\mathcal{T}$ be the set of all subsets of X,

$$\mathcal{T} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$$

whereupon the axioms must be satisfied, or we can take

$$\mathcal{T} = \{\emptyset, \{0, 1\}\}$$

It is not hard to verify the axioms here.

And there is another topology on the two point set:

$$\mathcal{T} = \{\emptyset, \{0\}, \{0, 1\}\}$$

Exercise 2.2.1. Verify the axioms for this case.

You can see that there is one more topology on the two point set. (!)

For a three point set there are again several distinct topologies, one is:

$$\mathcal{T} = \{\emptyset, \{0\}, \{0, 1\}, \{0, 2\}, \{0, 1, 2\}\}$$

This obviously cannot be made into a metric space, since if the distance between 1 and 2 is some number r, then there would have to be an open ball of radius $\leq r/2$ on 1 containing only the point 1. So there would have to be an open set in the topology containing only the point 1. But there isn't.

Exercise 2.2.2. Find a topology on the three point set which can be derived from a metric, and give a metric from which it is derived. Give another metric which has the same topology.

Obviously we need some way of sorting out the different topologies on sets, and one way is by requiring that topologies satisfy some extra conditions.

Definition 2.2.1. A topological space $(X, \mathcal{T})$ is said to be *hausdorff* iff for any distinct points $x, y \in X$ there are disjoint open sets U, V with $x \in U$ and $y \in V$.

Remark 2.2.1. The topology which has every subset of X in $\mathcal{T}$ is called the *discrete topology.* It is obvious that every finite set with a topology that is derived from a metric must have the discrete topology.

Exercise 2.2.3. Prove the last remark

Exercise 2.2.4. Find a topology on a finite set which is hausdorff but not discrete.

2.3 Compact Spaces: The Heine-Borel Theorem

The closed unit interval and the closed balls in $\mathbb{R}^n$ with the Euclidean metric have nice properties which $\mathbb{R}$ and $\mathbb{R}^n$ do not. In particular, the space of all continuous functions from any closed interval into $\mathbb{R}$ can be made (as we saw in 2C2) into a metric space in several useful ways. The same is not true of open intervals or the whole real line. It would be nice to be able to define an idea which generalises closedness and boundedness for topological spaces. The first is very easy:

Definition 2.3.1. A subset U of a topological space X is *closed* iff $X \setminus U$ is open

That is, U is closed when its complement in X is open.

Note that sets may be neither open nor closed, and may be both. A space where every set is either open or closed is called a *door space*[1]

Boundedness makes excellent sense in a metric space, a set is bounded iff there exists some real number such that the distance between any pair of points in the subset is less than this number.

[1] I am not making this up. Honest.

Exercise 2.3.1. Is the empty set bounded?

Definition 2.3.2. A set of subsets of a topological space X is said to be a *cover* of X iff the union of the subsets is X. It is called an *open cover* if each of the subsets is an open set. If a subset of the set of subsets is also a cover, it is called a *sub-cover*

Definition 2.3.3. A topological space is said to be *compact* iff every open cover has a finite subcover

Example 2.3.1. Let X be the whole real line. I claim it is not compact. To prove this I have to provide an open cover which has no finite subcover. When n is an integer, take the open set $(n - 3/4, n + 3/4)$ which is easily seen to be an open set in $\mathbb{R}$ with the usual topology. The set of all these intervals for all integers is clearly an open cover for $\mathbb{R}$. Removing even one of the intervals gives something which is no longer a cover. Hence $\mathbb{R}$ is not compact.

Example 2.3.2. I claim any closed interval in $\mathbb{R}$ with the usual topology is compact. You are invited to test out your analysis skills by proving this. When you get stuck,check out the *Heine-Borel Theorem*. In fact any closed and bounded subset of $\mathbb{R}^n$ is compact.

Exercise 2.3.2. Prove that if $f : X \longrightarrow Y$ is onto and continuous then if X is compact, so is Y. Deduce that the closed unit interval and $\mathbb{R}$ are not homeomorphic.

2.4 Subspaces

Definition 2.4.1. If X is a topological space with topology $\mathcal{T}$ and if U is a subset of X, then the *subspace topology* on U is just the collection of sets $T \cap U$ for $T \in \mathcal{T}$. With this topology, U is said to be a *(topological)subspace* of X.

Exercise 2.4.1. Prove that the subspace topology *is*[2] a topology for U.

Exercise 2.4.2. Construct a sensible definition of the *interior* of a subspace. (Make it an open set.)

Exercise 2.4.3. Construct a sensible definition of the boundary of a subspace.

[2]Calling it one doesn't make it one!

Exercise 2.4.4. Prove that any compact subspace of a hausdorff space is closed.

Exercise 2.4.5. Prove that any closed subspace of a compact space is compact.

Exercise 2.4.6. Prove that the image of a compact subspace by a continuous map is compact.

Exercise 2.4.7. Define a limit point for a sequence of points in a topological space and show that any compact space must have a limit point for any infinite sequence.

Exercise 2.4.8. Prove that if a sequence is in a subset U of a topological space X then any limit points are in the closure of U.

Remark 2.4.1. The last exercise is taking us into Analysis, a serious diversion.

Remark 2.4.2. Putting almost the whole of this section into the exercises is done for a very good reason. After you have done them you will know what it is.

2.5 Quotient Topologies

You may have discovered in algebra courses that there is are subobjects and quotient objects and both are important.

They are also important in Topology. Subobjects seem more natural in algebra, subgroups and subalgebras and subfields, but quotient objects are easy to understand in Topology, where they can be thought of as the result of gluing things together. Suppose for example I take the unit disc in $\mathbb{R}^2$,

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \right\}$$

I wish to take the boundary and collapse it to a single point. You can picture this as having a sort of bag with a drawstring around the mouth, and I pull the drawstring tight. It should be easy to persuade yourself that the result is a sphere, S^2. Note that no points inside the disc get squashed together, but the bounding circle becomes a single point.

I could do it with an explicit map of D^2 onto S^2. But I can also do it without any reference to an existing space by the following trick: I define a new set Y (called D^2/S^1 actually, but that's too long) which has a point x' for every point x in the interior of D^2, and one extra point say $\star$ (which came from S^1). So far this is not a topological space, but I shall shortly tell you what the open sets are. I do this by means of the obvious map which sends each point x of the interior of D^2 to the corresponding point x' and every point on the boundary of D^2 goes to $\star$. So I have a map from D^2 to the *set* Y which is onto.

I hope you are picturing this clearly.

Now I tell you what the open sets are on Y. I do it so as to make f continuous but only just. To see what that means, observe that if I chose the topology which had only two elements, the empty set and the whole set, then f would certainly be continuous, but Y would look nothing like a 2-sphere which has a lot more open sets. On the other hand if *every* subset of Y were open, then f would not be continuous, in fact it would blast the entire disc into its component points and so would be as discontinuous as you could get.

I therefore define a subset U of Y to be open iff $f^{-1}U$ is open in X. It remains only to be sure that this *is* a topology on Y; certainly if it is, f is continuous.

In general if $f : X \longrightarrow Y$ is an onto map and $\mathcal{T}$ is a topology on X, we say that the set $\mathcal{U}$ of subsets $U \subseteq Y$ such that $f^{-1}U$ is in $\mathcal{T}$ is the *quotient topology* on Y.

Proposition 2.5.1. *With the above definition, the quotient topology really is a topology on* Y

Proof:

□

Exercise 2.5.1. Construct a topological space which is obtained by taking a cylinder and gluing the bounding circles together (a) with the same orientation and (b) with the opposite orientation.

Exercise 2.5.2. Prove that S^2 is homeomorphic to the space D^2/S^1

Remark 2.5.1. So now you can see precisely what I mean when I talk about gluing together the circles when I constructed $\mathbb{R}P^2$ in the Introduction to the course.

Remark 2.5.2. Note that we could have defined S^2 as D^2/S^1. Such a definition is called an *intrinsic* definition of a manifold because it says nothing about any space in which it might be sitting. When we construct a map from D^2/S^1 to $\mathbb{R}^3$ which has image S^2 and is 1-1 and continuous and such that the inverse from the image is also continuous, we are constructing a homeomorphism between D^2/S^1 and S^2, or alternatively we are *embedding* S^2 in $\mathbb{R}^3$. We naturally tend to visualise spaces by means of embeddings, but the space and the embedding are not the same thing. The space can be constructed independent of any particular embedding, and there are lots of different embeddings (when there are any at all. There is no embedding of $\mathbb{RP}^2$ in $\mathbb{R}^3$, or of S^2 in $\mathbb{R}^2$.)

Exercise 2.5.3. Give an intrinsic definition of S^1 and prove that you are right.

2.6 Product Spaces

Definition 2.6.1. If X and Y are sets, the *cartesian product* $X \times Y$ is the set of ordered pairs $\{(x, y) : x \in X, y \in Y\}$

Example 2.6.1. $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$

If X and Y are metric spaces then we can make $X \times Y$ into a metric space.

Exercise 2.6.1. Construct a 'Euclidean product metric' for metric spaces (X, d), (Y, e).

What about the case when X and Y are topological spaces? Is there a natural definition of a product topology. The answer is 'yes' and it goes the way you probably imagine.

Definition 2.6.2. For the cartesian product of two sets $X \times Y$ define the *canonical projections*

$$\pi_x : X \times Y \longrightarrow X, \qquad \pi_y : X \times Y \longrightarrow Y$$

by $\pi_x(x, y) = x$, $\pi_y(x, y) = y$.

Definition 2.6.3. A *base* for a topology is a collection of open sets such that every open set is a union of sets in the base.

Remark 2.6.1. The open balls in $\mathbb{R}^2$ are a base for the topology.

Definition 2.6.4. The *product topology* on $X, \mathcal{T} \times Y, \mathcal{U}$, the cartesian product of topological spaces, has base the sets of the form

$$\pi_x^{-1}T \ \cap \pi_y^{-1}U$$

for $T \in \mathcal{T}$ and $U \in \mathcal{U}$

Remark 2.6.2. Draw a picture.

Definition 2.6.5. A *subbase* for a topology is a collection of sets such that intersections of finitely many sets in the collection for a base for the topology.

Remark 2.6.3. So we could say that the sets $\pi_x^{-1}T$ and $\pi_y^{-1}U$ for $T \in \mathcal{T}$ and $U \in \mathcal{U}$ are a subbase for the product topology

Exercise 2.6.2. Construct a definition of the (multiple) cartesian product of a collection of topological spaces.

Exercise 2.6.3. Prove that the product of two compact sets is compact. Is it true that the product of any non-empty collection of compact topological spaces is compact?

Exercise 2.6.4. Show that the unit ball in $\mathbb{R}^n$ is compact.

Exercise 2.6.5. Prove that the product of two hausdorff spaces is hausdorff.

The following result may make you feel that algebra and tology are not so different after all. I shall express it as a series of easy exercises.

Exercise 2.6.6. Show that $f : X \longrightarrow Y$ a map between tological spaces is continuous iff the inverse image of a set U which is closed in Y is a set which is closed in X.

Exercise 2.6.7. Show that a compact subspace of a hausdorff space is closed in the space.

Exercise 2.6.8. Hence show that if $f : X \longrightarrow Y$ is a map from a compact space to a hausdorff space which is continuous, 1-1 and onto, then its inverse is also continuous.

2.7 A Useful Result

Recall from 2C2 the definition of a Cauchy sequence in a metric space, and the definition of a *complete* metric space as one where every cauchy sequence converges. The following exercise has the merit of being easy and being useful at a later stage:

Exercise 2.7.1. Given a complete metric space (X, d) and a map $f : X \longrightarrow X$, we say that f is a *contraction mapping* iff

$$\exists k \in \mathbb{R},\ 0 < k < 1,\ \ \forall\, x, y \in X, d(f(x), f(y)) \leq kd(x, y)$$

Show that any contraction mapping has a unique fixed point, i.e. $\exists! a \in X : f(a) = a$.

The above result is called *The Contraction Mapping Theorem* and requires only a small amount of ingenuity to prove.

2.8 Partitions of Unity

Like the above result which requires that X be a metric space, this short section may not altogether belong here, but it doesn't belong anywhere else either. So here it comes.

First, a definition:

Definition 2.8.1. If $f : X \longrightarrow \mathbb{R}$ is a map from a topological space X to $\mathbb{R}$ which never takes negative values, the *support* of f, $\text{supp}(f)$, is the subspace of X

$$\{x \in X :\ f(x) > 0\}$$

Second another definition:

Definition 2.8.2. Let U be an open subset of $\mathbb{R}^n$. A collection $\mathcal{F}$ of functions from U to $\mathbb{R}$ is said to be a *partition of unity* iff

1. $\forall f \in \mathcal{F},\ \forall\, x \in U, 0 \leq f(x) \leq 1$
2. for every $x \in U$ there is an open ball $B \subseteq U$ containing x such that only a finite subset of $\mathcal{F}$ have non-zero values on B, and
3. $$\sum_{f \in F} f(x) = 1$$

You will find partitions of unity are useful to us subsequently.

Third, yet another definition: The partition of unity, $\mathcal{F}$ is said to be *subordinate to the open cover* $\mathcal{C}$ iff the support of every $f \in \mathcal{F}$ is contained wholly in some element of $\mathcal{C}$

Finally the proposition we want:

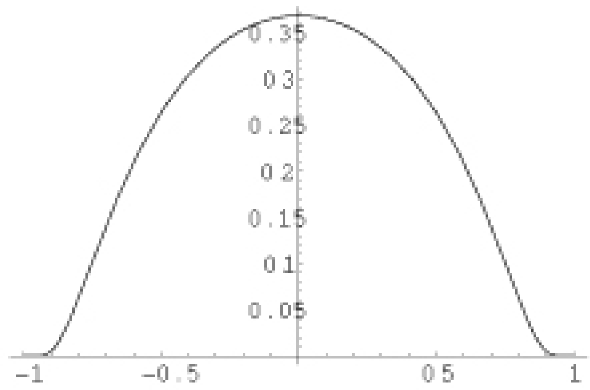

Figure 2.1: A Smooth bump function with compact support

Proposition 2.8.1. *For any open subset $U \subseteq \mathbb{R}^n$ and any open cover $\mathcal{C}$ of U, there is a partition of unity for U of smooth functions which is subordinate to $\mathcal{C}$.*

The proof proceeds through some easy lemmata:

Lemma 2.8.1. There are smooth bump functions.

Proof:

First we define a smooth bump function which is zero outside the unit ball in $\mathbb{R}^1$ and takes only positive values inside it. I choose

$$b(x) = e^{\frac{1}{x^2-1}}$$

The general shape of the curve for values between -1 and $+1$ is shown in figure 2.1; the function is zero outside this open interval.

As you can see by looking at the function, it is infinitely differentiable inside $(-1, 1)$ and outside $[-1, 1]$ it is zero and so also infinitely differentiable. We need to ensure that the derivatives at ± 1 are defined and are all zero and a look at the graph in the vicinity of 1 (figure 2.2) is encouraging. Actually doing a few differentiations of $b(x)$ and taking the limits as $x \to 1$ should convince even the hardened sceptic that b has the required properties on $\mathbb{R}$. To make it work on $\mathbb{R}^n$ I just define $B(\boldsymbol{x} - \boldsymbol{a}) = b(\|\boldsymbol{x} - \boldsymbol{a}\|)$ This gives me the required bump in n-dimensions. Instead of making it at centre $\boldsymbol{a}$ and radius 1, it can be given radius δ for any $\delta \in \mathbb{R}^+$ by writing $B_\delta(\boldsymbol{x}-\boldsymbol{a}) = b(\|\boldsymbol{x}-a\|/\delta)$. Note that inside the ball B is positive and that it is zero at and outside the boundary of the ball. □

Lemma 2.8.2. If A is any compact set in $\mathbb{R}^n$ and $A \subseteq V \subseteq \mathbb{R}^n$ for some open set V, then we can find a smooth function $\hat{A}$ which takes values always

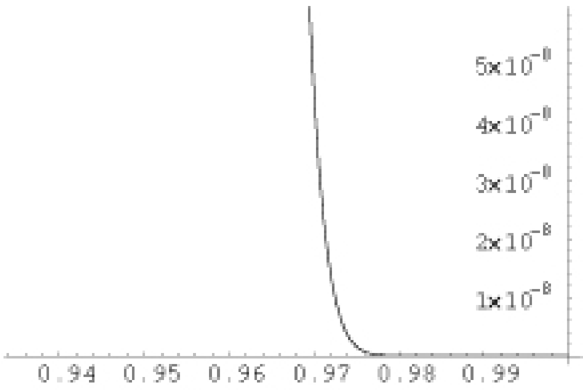

Figure 2.2: The bump function *is* smooth

greater than or equal to zero, has $\hat{A}(x) > 0$ for every $x \in A$ and has support in V.

Proof:

Obviously this is easy if $V = \mathbb{R}^n$, so we are really only interested in the case where V contains A but only just. In this case we can assume there is some smallest distance from $\mathbb{R}^n \setminus V$ to A, say δ. The idea is to cover A with open balls of radius δ, secure in the knowledge that this can be done with a finite number of them (since A is compact), where each open ball is the centre of a bump function of radius δ. If we call them $B_{\delta,1}, B_{\delta,2}, \cdots, B_{\delta,k}$ then

$$\hat{A} = \sum_{j=1,k} B_{\delta,j}$$

will do nicely. □

Now for the main result; suppose we have U and $\mathcal{C}$ given us and we want a partition of unity subordinate to the cover. We can tile $\mathbb{R}^n$ with compact cubes and investigate the situation over any one of them. On each tile, D, we can find a finite subset of the elements of $\mathcal{C}$ which cover D; call them $U_1, U_2, \cdots, U_k$. Take V to be the intersection of their union with an open cube D' containing D of side, let us say, 1.1 times the side of the cube D, and apply the last lemma where A is the intersection of $\bar{U_j} \cap D$ and $\bar{W}$ is the *closure* of W, that is, the intersection of all closed sets containing W. Then we have a finite set of functions, $g_1, g_2, \cdots, g_k$ which are all zero outside the cube D' and at least one of which is greater than zero on any point in D.

Define functions f_j having the same support as the corresponding g_j by

$$f_j = \frac{g_j}{\sum_{(1 \leq j \leq k)} g_j}$$

Extending this to an adjacent tile requires only examination of the overlap which is a compact set covered by a finite number of the f_j, and at worst gives us some redundancy in the set of all functions which cover U. □

Corollary 2.1. *If $U \subseteq \mathbb{R}^n$ is open and A is any closed subset of U then there is a smooth function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ which takes the value 1 on A, and is zero outside U.*

Proof:

We can cover A by compact sets and get a partition of unity for each of them which is zero outside some (small) neighbourhood and sum the functions in the partition of unity. □

Remark 2.8.1. The case where this has to be thought about carefully is when we have something like $U \subset \mathbb{R}^2 = \{(x, y)^T : -e^x < y < e^x\}$ and A is the X axis. The (non-smooth) function which is 1 when $y = 0$ and zero elsewhere can obviously be smoothed out over any bounded interval and go to zero outside the cartesian product of the interval with another interval, the second interval being constrained to lie within U. But this can be done for any such bounded interval, so we only have to make the smoothings agree on the intersections, which is easy enough.

2.9 Summary

Remark 2.9.1. There are lots of books which develop these ideas further, see *An Introduction to Topology and Modern Analysis* by G F Simmons for a nice introduction, and *General Topology* by Kelley for a full blown course. See also Bourbaki for a treatment which tells you everything you might ever want to know about the subject and a great deal you wouldn't.

Chapter 3

Algebraic Preliminaries

3.1 Duality

Last year in 2C2 I introduced you to the dual of a vector space.

Definition 3.1.1. If $\mathcal{V}$ is a vector space, the set of all linear maps from $\mathcal{V}$ to $\mathbb{R}$ is a vector space under pointwise addition and scaling and is called *the dual space*, $\mathcal{V}^\star$, of $\mathcal{V}$.

Maps from a vector space to $\mathbb{R}$ are often called *functionals* on the space. I shall not use this terminology myself but if you come across it in your reading you will know what it means.

In the case where $\mathcal{V}$ is finite dimensional, it is isomorphic to $\mathbb{R}^n$ for some n. It is no surprise therefore that in this case:

Proposition 3.1.1. *If $\mathcal{V}$ is finite dimensional, $\mathcal{V}$ is isomorphic to $\mathcal{V}^\star$*

Proof:

Let $(e_1, e_2, \cdots, e_n)$ be a basis for $\mathcal{V}$. For every $j \in [1..n]$, let $e_j^\star$ denote the linear map from $\mathcal{V}$ to $\mathbb{R}$ which takes e_i to zero if $i \neq j$, and takes e_j to 1. Then since I have told you what $e_j^\star$ does to a basis, it is a unique linear map for every $j \in [1..n]$.

These maps are linearly independent since

$$\sum_{1 \leq j \leq n} a_j e_j^\star = \mathbf{0} \Rightarrow \forall j \in [1..n],\ a_j e_j^\star(e_j) = 0 \Rightarrow \forall j \in [1..n],\ a_j = 0$$

The set of maps $\{e_j^\star, j \in [1..n]\}$ also spans the space $\mathcal{V}^\star$ since any linear map can be defined by what it does to a basis. If the linear map $\alpha \in \mathcal{V}^\star$ takes the basis for $\mathcal{V}$ $(e_1, e_2, \cdots, e_n)$ to $a_1, a_2, \cdots, a_n$ respectively, then it is evident that

$$\alpha = \sum_{1 \leq j \leq n} a_j e_j^\star$$

Hence $(e_1^\star, e_2^\star, \cdots, e_n^\star)$ is a basis for $\mathcal{V}^\star$, and there is an obvious isomorphism defined by $e_j \rightsquigarrow e_j^\star,\ 1 \leq j \leq n$. □

The basis for $\mathcal{V}^\star$ of linear maps $e_j^\star$ is called the *dual basis* to the basis for $\mathcal{V}$ of vectors e_j.

Note that the proof is much clunkier than it need be. We could have argued as follows: Since $\mathcal{V}$ is finite dimensional it is isomorphic to $\mathbb{R}^n$ for some n. Its dual, $\mathcal{V}^\star$ has to be isomorphic to $\mathbb{R}^{n\star}$ since the linear maps from $\mathcal{V}$ to $\mathbb{R}$ are obviously in 1-1 (linear) correspondence with the linear maps from $\mathbb{R}^n$ to $\mathbb{R}$. And $\mathbb{R}^{n\star}$ is isomorphic to $\mathbb{R}^n$.

$$\mathcal{V} \cong \mathbb{R}^n \cong \mathbb{R}^{n\star} \cong \mathcal{V}^\star$$

Although clunky, the first argument is direct and straightforward. It has the advantage that it allowed me to introduce the dual basis, which we shall be using later.

In the particular case of the dual space to $\mathbb{R}^n$, $\mathbb{R}^{n\star}$ will be written from now on as $\mathbb{R}_n$. You will see why this is not altogether a bad notation later on.

Exercise 3.1.1. If $\mathcal{V}$ has isomorphic dual $\mathcal{V}^\star$, make a definition of an inner product in $\mathcal{V}$.

It is a good idea to work through proofs like this in the simplest case which is probably $\mathbb{R}^2$. Then we can identify e_1, e_2 with what in 2C2 we called $\boldsymbol{i}$ and $\boldsymbol{j}$. We then have $e_1^\star$ and $e_2^\star$ are the projection maps

$$\pi_x \begin{bmatrix} x \\ y \end{bmatrix} = x; \quad \pi_y \begin{bmatrix} x \\ y \end{bmatrix} = y$$

In 2C2 I also called these maps dx and dy when they came up in specifying a differential 2-form on $\mathbb{R}^2$. There is nothing very new in here, just jargon, but unlike most jargon this has been carefully designed to make thinking easier.

In the cases we are interested in, the vector space is almost always $\mathbb{R}^n$ for some positive integer n. In this case, we can write a vector in $\mathbb{R}^2$, say, as

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

and an element of $\mathbb{R}_2$ can be written, say, as

$$[2 \;\; 3] \text{ or } [2,3]$$

(with some risk of confusion of the latter with an interval.) This notation has the merit of making it awfully obvious that $\mathbb{R}^n \cong \mathbb{R}_n$.

The point about elements in the dual space is that *they are things which operate on elements of the original space to turn them into numbers (and do so linearly)*, and although a finite dimensional vector space and its dual are isomorphic, and although we have been able to get away with ignoring the difference so far, it will not be practicable to continue to pretend they are the same.

Writing $[2 \;\; 3]$ for a *covector* (an element of the dual space of $\mathbb{R}^2$) and $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ for a vector or point in $\mathbb{R}^2$ raises the question of what happens if you take the dual of a dual space? Or, in simple direct terms, what's a cocovector? [1]

In short, what sort of thing is $\mathcal{V}^{\star\star}$? Technically, if $\mathcal{V}^\star \triangleq L(\mathcal{V}, \mathbb{R})$ is the space of linear maps from $\mathcal{V}$ to $\mathbb{R}$, then

$$\mathcal{V}^{\star\star} = L(L(\mathcal{V}, \mathbb{R}), \mathbb{R})$$

It is a little hard to visualise these things. Happily there is a convincing way to identify $\mathcal{V}^{\star\star}$ with $\mathcal{V}$. Define a map *Double-Dual*

$$DD : \mathcal{V} \longrightarrow \mathcal{V}^{\star\star}$$

which takes $v \in \mathcal{V}$ to the map $\hat{v} : \mathcal{V}^\star \longrightarrow \mathbb{R}$ defined by $\hat{v}(m) = m(v)$ for every $m \in \mathcal{V}^\star$.

This gives a perfectly good map from $\mathcal{V}$ to $\mathcal{V}^{\star\star}$ and it is given without reference to a basis, so is a more fundamental map than the isomorphism between $\mathcal{V}$ and $\mathcal{V}^\star$ when they are both finite dimensional.

Proposition 3.1.2. *For V finite dimensional, DD is an isomorphism of vector spaces.*

Proof:

[1] No jokes please, this is deadly serious stuff, and mathematicians *never* tell jokes. Or lies.

It is easy to see that this map is 1-1; $\hat{v} = \hat{u}$ means that $\forall\ m \in \mathcal{V}^\star,\ \ m(u) = m(v)$. But if $u \neq v$ then there is certainly a projection map which takes u and v to different values in $\mathbb{R}$. So we must have $u = v$ i.e. DD is 1-1.

It is also easy to see it is linear.

$$DD(u+v) = \widehat{u+v} \text{ and } \ \forall m \in \mathcal{V}^\star,\ (\widehat{u+v})(m) = m(u+v) = m(u) + m(v)$$

$$= \hat{u}(m) + \hat{v}(m) = (\hat{u} + \hat{v})(m)$$

$$\Rightarrow DD(u+v) = DD(u) + DD(v)$$

But the argument that provided a dual basis with the same number of elements tells us that the dimension of $\mathcal{V}^{\star\star}$ is also the same as the dimension of $\mathcal{V}$, so the map must be onto as well, and is therefore an isomorphism. □

Exercise 3.1.2. Fill in the proof properly by exhibiting a linear map from $\mathcal{V}$ to $\mathbb{R}$ which gives different values for u, v when $u \neq v$, and by verifying that

$$\forall t \in \mathbb{R}, \forall u \in \mathcal{V},\quad DD(tu) = tDD(u)$$

The *natural* isomorphism between $\mathcal{V}$ and $\mathcal{V}^{\star\star}$ in the finite dimensional case means that there is very little point in distinguishing them, and we shall generally regard them as the same. In particular we shall regard a vector as operating linearly on covectors, by treating the vector v as if it were the same as the cocovector $DD(v)$.

Exercise 3.1.3. Think about the dual space to $\mathcal{L}^2[-\pi, \pi]$ the space of square integrable functions from the closed interval $[-\pi, \pi]$ to $\mathbb{R}$. Notice that there is an element, $\hat{\mathbf{0}}$ of the dual space which takes every function $f \in \mathcal{L}^2[-\pi, \pi]$ and sends it to the number f(0).

I claim that if there were an isomorphism between the space and its dual, then there would have to be a function $\hat{\hat{\mathbf{0}}}$ from $[-\pi, \pi]$ to $\mathbb{R}$ such that for every function $g \in \mathcal{L}^2[-\pi, \pi]$,

$$\int_{-\pi}^{\pi} \hat{\hat{\mathbf{0}}}(x) g(x) dx = g(0)$$

Investigate my claim critically and work out whether there really is a function $\hat{\hat{\mathbf{0}}}$.

3.2 Bivectors: The Exterior Algebra

In order to put some new ideas into the simplest setting I shall start with telling you about the exterior algebra over $\mathbb{R}^2$. Then I shall do it for $\mathbb{R}^3$, and the generalisations will be obvious so I will leave it to you to do it for $\mathbb{R}^n$. I also note that we don't actually need to assume that the vector space is $\mathbb{R}^2$ it could be any two dimensional vector space, (e.g. $\mathbb{R}_2$).

First, an *algebra* is defined as being a vector space where there is a rule for multiplying the elements. $\mathbb{R}$ and $\mathbb{C}$ are examples of real vector spaces where you can multiply the elements in an obvious way. If we take the set of 2×2 matrices, then they form a four dimensional vector space and we can multiply them by ordinary matrix multiplication. If we took the space of piecewise continuous functions from $[-\pi, \pi]$ to $\mathbb{R}$ then they certainly form a vector space and we can multiply the elements pointwise. Of course any $\mathbb{R}^n$ can be made into an algebra by multiplying the components of the vectors.

Multiplying in this case just means that for any two elements, there is a rule which produces a third from them, which distributes over the addition: $a \star (b + c) = a \star b + a \star c$. Any such rule will do[2] We do not require the multiplication to be associative, mainly because in interesting cases it isn't.

Now I take $\mathbb{R}^2$ and let e_1 and e_2 be the standard basis elements (although it will work for any basis in fact). I define a multiplication of e_1 and e_2 to give a new thing called a *bivector*, which in this case is e_{12}.

I now take the four dimensional space of things

$$(a + be_1 + ce_2 + de_{12}), \qquad a, b, c, d \in \mathbb{R}$$

This is a real vector space of dimension 4. We say it has three sorts of things in it, the scalars (a), the vectors (be_1+ce_2) and the bivectors (de_{12}). The rules for multiplying (for which I shall use the symbol $\wedge$) are that $e_1 \wedge e_2 = e_{12}$ and $\forall i, j = 1, 2 \quad e_i \wedge e_j = -e_j \wedge e_i$. The last rule ensures that $e_i \wedge e_i = 0, i = 1, 2$ This with distributivity and some obvious commutativity and associativity is enough to give us the following rule for multiplying two general elements of the algebra:

$$(a + be_1 + ce_2 + de_{12}) \wedge (a' + b'e_1 + c'e_2 + d'e_{12})$$

$$= (aa' + ab'e_1 + ac'e_2 + ad'e_{12}$$

[2]No matter how weird or daft. This is why abstract algebra (the subject) is quite fun in a rather irresponsible way.

$$+a'be_1 + bb'(0) + bc'e_{12} + bd'(0)$$
$$+a'ce_2 - b'ce12 + cc'(0) - cd'(0)$$
$$+a'de_{12} - b'd(0) + 0 + 0)$$

and collecting up terms we get:

$$(aa' + (ab' + a'b)e_1 + (ac' + a'c)e_2 + (ad' + a'd + bc' - b'c)e_{12})$$

We have defined in particular the exterior (or *veck*, or *wedge*) product for two vectors

$$(be_1 + ce_2) \wedge (b'e_1 + c'e_2)$$

to be the bivector

$$(bc' - b'c)e_{12}$$

which we can get from the above formula by putting $a = a' = d = d' = 0$

Exercise 3.2.1. Compute the area of the triangle with vertices at $(2, 1)^T$, $(1, 2)^T$ and the origin by using the exterior product.

Note that the wedge product on vectors is not commutative but is *anticommutative*, that is, $\boldsymbol{a} \wedge \boldsymbol{b} = -\boldsymbol{b} \wedge \boldsymbol{a}$ for any vectors $\boldsymbol{a}, \boldsymbol{b}$, and that the numerical value of the bivector is the determinant of the matrix formed by making columns of the two vectors.

We write

$$\bigwedge(\mathbb{R}^2)$$

for the exterior algebra of dimension 4 over $\mathbb{R}^2$. We also write $\bigwedge^0$ for the subalgebra of *scalars* of this algebra, $\bigwedge^1$ for the subspace of vectors and $\bigwedge^2$ for the subspace of bivectors.

With this structure, $\bigwedge(\mathbb{R}^2)$ is called the *exterior algebra* of $\mathbb{R}^2$, and the binary operator or multiplication $\wedge$ is called the *exterior product.* We write:

$$\bigwedge(\mathbb{R}^2) = \overset{0}{\bigwedge}(\mathbb{R}^2) \oplus \overset{1}{\bigwedge}(\mathbb{R}^2) \oplus \overset{2}{\bigwedge}(\mathbb{R}^2)$$

and $\oplus$ is the same as the cartesian product of vector spaces in this case, giving a decomposition into three subspaces, two of dimension 1 and one of dimension 2.

Remark 3.2.1. Grassmann invented the exterior algebras in 1844, so you are now only about one and a half centuries behind. In first year you were about three and a half centuries behind, so you are catching up.

Remark 3.2.2. There is no pressing need to have a visual picture of a bivector in $\mathbb{R}^2$, but if you really must visualise everything, you can imagine two vectors and the parallelogram formed by them, the origin and their sum. Then if they are $\boldsymbol{u}$ and $\boldsymbol{v}$, and if $\boldsymbol{u}$ comes before $\boldsymbol{v}$ in an anticlockwise movement around the origin, then the bivector $\boldsymbol{u} \wedge \boldsymbol{v}$ can be thought of as the area of the parallelogram. If the order is reversed there will be a minus sign in front, so we can call $\boldsymbol{u} \wedge \boldsymbol{v}$ the *oriented area* of the parallelogram formed by $\mathbf{0}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v}$. Of course we need the numerical value of the area to be flagged as a bivector and not just any old number, so we would write e_{12} after the number.

Exercise 3.2.2. Calculate the bivector $\boldsymbol{u} \wedge \boldsymbol{v}$ when

$$\boldsymbol{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Remark 3.2.3. The only difficulty with this is seeing how little there is in it. Don't look for anything deep, this is *algebra*. It is obviously harmless to construct the definition of $\bigwedge(\mathbb{R}^2)$ as above, about the worst thing you could say is that it might be a waste of time because nothing good will come of it. Actually it will make understanding differential forms much easier.

Exercise 3.2.3. Change the rules of the above multiplication (but keep the same elements) by defining $e_ie_j = -e_je_i, \quad i \neq j$ and $e_ie_i = 1, i = 1, 2$. Try multiplying out the general terms to see what you get. Can you find the Complex Numbers buried in this algebra as a subalgebra? This rather cool little beast is called $\mathcal{C}\ell(2, 0)$, the *Clifford Algebra* on $\mathbb{R}^2$. If you want to learn more about Clifford Algebras, read *Clifford Algebras and Spinors* by Perrti Lounesto.

I am now going to construct $\bigwedge(\mathbb{R}^3)$. I have a 'graded' algebra again with $\bigwedge^0$ is the collection of scalars, $\bigwedge^1$ is the collection of vectors in $\mathbb{R}^3$, $\bigwedge^2$ is the collection of bivectors in $\mathbb{R}^3$ and finally and rather naturally, $\bigwedge^3$ is the collection of trivectors in $\mathbb{R}^3$. The bivectors look like $e_1 \wedge e_2 = e_{12}$, $e_2 \wedge e_3 = e_{23}$ and $e_3 \wedge e_1 = e_{31}$ and a general bivector in $\mathbb{R}^3$ is of the form:

$$ae_{12} + be_{23} + ce_{31}$$

for scalars $a, b, c \in \mathbb{R}$. The trivectors all look like de_{123} for some scalar $d \in \mathbb{R}$ so form a one dimensional subspace. The multiplication has $e_i \wedge e_j = -e_j \wedge e_i, \ \forall \ i, j \in \{1, 2, 3\}$

Exercise 3.2.4. Confirm that the dimension of the algebra $\bigwedge(\mathbb{R}^3)$ is eight (2^3) and write down the product of two general terms. If you don't do this there will always be just a small feeling of mystery about the exterior algebra, if you do you will know almost everything there is to know about them. [3]

Exercise 3.2.5. Compute the area of the triangle with vertices at the origin, at $(3,2,1)^T$ and at $(1,2,3)^T$.

Exercise 3.2.6. Compute the volume of the parallelipiped with vertices at the origin, at $(3,2,1)^T$, at $(1,2,3)^T$, at $(0,3,1)^T$, and at sums of the above distinct vectors.

Exercise 3.2.7. Show that the space $\bigwedge(\mathbb{R}^n)$ has dimension 2^n. You may find writing down Pascal's triangle helpful and there again you may not.

Exercise 3.2.8 (For Algebraists). Write down the multiplication for the[4] Clifford Algebra $\mathcal{C}\ell(3,0)$, using the same rules as before, replacing $\{1,2\}$ by $\{1,2,3\}$ where it seems reasonable, and see if you can find anything of interest about it. Now try the Clifford Algebra $\mathcal{C}\ell(0,3)$, which has $e_ie_i = -1, i = 1,2,3$. (Hint: you might try to set up an obvious isomorphism between e_{ij} and e_k (The Hodge Duality)).

The subspace $\bigwedge^2(\mathbb{R}^n)$ of $\bigwedge(\mathbb{R}^n)$ is a vector space of dimension $C^n_2 = \frac{n!}{(n-2)!2!}$ with basis elements the set $\{e_ij\}$ for $1 \le i < j \le n$, and is hence isomorphic to the space $\mathbb{R}^{n(n-1)/2}$. Similarly the space $\bigwedge^2(\mathbb{R}_n)$ has basis $\{e^\star_{ij} : 1 \le i < j \le n\}$.

We can certainly look at the space $\bigwedge(\mathbb{R}_n)$ and also at the space $\bigwedge^\star(\mathbb{R}^n)$. The latter makes sense since $\bigwedge(\mathbb{R}^n)$ is a perfectly respectable real vector space.

It is immediate that these two spaces are isomorphic since they have the same dimension (2^n). But some thought shows that we can do more than that: the isomorphism can preserve the grading structure, so that if

$$\Phi : \bigwedge(\mathbb{R}_n) \longrightarrow \overset{\star}{\bigwedge}(\mathbb{R}^n)$$

is the isomorphism, we can take $\bigwedge^\star(\mathbb{R}^n)$ and since there are obvious 1-1 homomorphisms $inc_k : k = 0, n$ of the vector subspaces $\bigwedge^k(\mathbb{R}^n)$ into $\bigwedge(\mathbb{R}^n)$ (that have zero values outside the subspaces) then we can take any $f : \bigwedge(\mathbb{R}^n) \longrightarrow \mathbb{R}$ in $\bigwedge^\star(\mathbb{R}^n)$ and by composing with inc_k obtain a linear map from $\bigwedge^k(\mathbb{R}^n)$ to $\mathbb{R}$. And all linear maps from $\bigwedge^k(\mathbb{R}^n)$ to $\mathbb{R}$ can be obtained in this way. So we can arrange Φ to take $\bigwedge^k(\mathbb{R}_n)$ to $\bigwedge^{k\star}(\mathbb{R}^n)$. Both, after all, have the same dimension, $n!/(n-k)!k!$

[3] So do it.

[4] This is entertaining and educational but will have minimal impact on your passing the course.

We can do even more than this. When $k = 0$ we have that $\bigwedge^0(\mathbb{R}_n)$ and $\bigwedge^{0\star}(\mathbb{R}^n)$ are copies of the space $\mathbb{R}$ and of $\mathbb{R}^\star$ respectively, and, since $\mathbb{R}$ is a field with a unique multiplicative identity and the dual vector space $\mathbb{R}^\star$ has the identity map as a special map, $\mathbb{R}^\star$ is hardly worth distinguishing from $\mathbb{R}$. So we can make Φrestricted to $\bigwedge^0(\mathbb{R}_n)$ a rather natural isomorphism of algebras.

When $k = 1$ we have the projections $e_i^\star$ from $\mathbb{R}^n$ to $\mathbb{R}$ in the domain of Φ, and we get the same in the codomain, so we can make Φ the identity map restricted to $\bigwedge^1(\mathbb{R}_n)$.

When $k = 2$ we have *bicovectors*, $e_i^\star \wedge e_j^\star$ in the domain of Φ, and *cobivectors*, $e_{ij}^\star = (e_i \wedge e_j)^\star$, in the range, and these are in 1-1 correspondence so it is natural to make Φ take one to the other. So I define Φ on $\bigwedge^2(\mathbb{R}_n)$ so that it takes $e_i^\star \wedge e_j^\star$ to $e_{ij}^\star = (e_i \wedge e_j)^\star$, for every $i, j : 1 \leq i < j \leq n$. $e_{ij}^\star$ is, of course, the linear map which takes the bivector e_{ij} to 1 and every other basis element to zero. We conclude that bicovectors and cobivectors can be regarded as the same things and no harm will come.

Similarly with dual trivectors, and indeed for larger values of k right up to $k = n$; the subspaces can be identified in a natural manner, although it has to be said that being given a basis for everything stops it being as natural as one might like.

Exercise 3.2.9. I can take $e_i^\star \wedge e_j^\star$ and, bearing in mind the above, regard it, if I wish, as a map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ by writing

$$(e_i^\star \wedge e_j^\star)(\boldsymbol{u}, \boldsymbol{v}) = e_{ij}^\star(\boldsymbol{u} \wedge \boldsymbol{v})$$

Evaluate this map in the case $n = 2, 3, 4$ on any pair of (non-zero) vectors of your choice for a reasonable sample of values of i and j. It is recommended that you write out the vectors both as columns and in basis form, and note the results of the calculation both ways.

Remark 3.2.4. Do the last exercise and you will have gone a long way to understanding a lot about the point of all this algebra, so this exercise really is crucial. Also very easy.

Remark 3.2.5. Last year in 2C2 we defined the wedge product $dx^i \wedge dx^j$

for $dx^i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $1 \leq i \leq n$ the map

$$dx^i \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^i \\ \vdots \\ x^n \end{bmatrix} = x^i$$

and of course dx^j likewise. Now we have renamed dx^i as $e_i^\star$, so we were clearly in the space $\bigwedge(\mathbb{R}_n)$ and multiplying two elements in the subspace $\bigwedge^1(\mathbb{R}_n)$ to get something in $\bigwedge^2(\mathbb{R}_n)$.

You will recall that we could get through most of the course without actually having a definition of what $dx^i \wedge dx^j$ actually *was*, all we needed were the properties. The abstract algebra gives us all the properties, and so there is a sense in which we could stop with $\bigwedge(\mathcal{V}^{n\star})$, the exterior algebra on the dual of a vector space of dimension n, and say it contains all we need. A 2-form on $\mathcal{V}$ is then just an element of $\bigwedge^2(\mathcal{V}^{n\star})$, and a differential 2-form on an open set U in $\mathbb{R}^n$ is a smooth map from U to $\bigwedge^2(\mathbb{R}_n)$. For $U \subseteq \mathbb{R}^3$ for example, a differential 2-form on U assigns a particular 2-form

$$P(x, y, z)e_{12}^\star + Q(x, y, z)e_{23}^\star + R(x, y, z)e_{31}^\star$$

to each point of U, where P, Q, R are smooth functions on U so the 2-form changes smoothly as we move over U, which except for a change of notation is just

$$Pdx \wedge dy + Qdy \wedge dz + Rdz \wedge dx$$

We could define the exterior derivative on differential k-forms just as before, and no harm would come. In fact everything we need is in here, with one exception: everything given here is tied down to a basis. And in future we are going to be changing bases and therefore would like to have an invariant description which does not depend on a choice of basis until we want to do some sums. Just as the physicists rely extensively on Einstein's form of the invariance principle which says that the laws of nature do no depend on your choice of a coordinate system, and just as we preferred as more 'natural' the isomorphism between a space and its double dual, so we want basis free (coordinate free) descriptions of everything we deal with. For that reason I shall now 'find' the exterior algebra sitting in a bigger (but basis free) object called the tensor algebra of a vector space.

3.3 Multilinear Algebra:Tensors

Definition 3.3.1. For $\mathcal{V}$ a real vector space, a *bilinear map*

$$f : \mathcal{V} \times \mathcal{V} \longrightarrow \mathbb{R}$$

is a map such that:

$$\forall u, v, w \in \mathcal{V}, \forall\ s, t \in \mathbb{R}, \quad f(su + tv, w) = sf(u, w) + tf(v, w) \text{ and}$$

$$\forall u, v, w \in \mathcal{V}, \forall\ s, t \in \mathbb{R}, \quad f(u, sv + tw) = sf(u, v) + tf(u, w)$$

This can be summarised by saying that '*f is linear in each variable separately*'.

Example 3.3.1. multiplication is a bilinear map from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$. This tells you that bilinearity and linearity are different kinds of thing. Obviously, multiplication is not linear, or (since one times one is one) two twos would be two, not four as is customarily believed.

Example 3.3.2. The dot or inner product on $\mathbb{R}^2$, $\boldsymbol{\cdot}: \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}$ is bilinear. It is also part of the definition of an inner product on any vector space that it should be bilinear, and so it comes as no surprise to see that the dot product is bilinear on $\mathbb{R}^n$.

Example 3.3.3. Det: $\mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}$ is bilinear. This is obviously different from the dot product which is symmetric: $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$ We know that the determinant on $\mathbb{R}^2$ is produced by the exterior product, and $\boldsymbol{u} \wedge \boldsymbol{v} = -\boldsymbol{v} \wedge \boldsymbol{u}$. Such a bilinear map is called *antisymmetric* or *alternating.*

Exercise 3.3.1. Verify by bashing through the definition of Det for 2×2 matrices, or alternatively the definition of the exterior product, that $\wedge$ regarded as a map from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ is bilinear.

It is important to do the above exercise since the determinant will turn out to be central to the calculation of the alternating tensors which are the whole idea behind differential k-forms. So do the next one too:

Exercise 3.3.2. Show that for any choice of $i, j, 1 \le i < j \le n$, $e^\star_{ij} = e^\star_i \wedge e^\star_j$, regarded as a map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ is an alternating bilinear map. (Hint: Show that it pulls out the i^{th} and j^{th} rows of the two vectors when written as columns, and takes the determinant of the resulting 2×2 matrix. (This was the main fact you should have got from exercise 3.2.9).

The above exercises show that there are at least two kinds of bilinear maps, the symmetric and the antisymmetric.

Exercise 3.3.3. construct a bilinear map which is neither.

It is easy to see that the bilinear maps from $\mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ form a vector space of dimension n^2, where $\mathcal{V}$ has dimension n. The rules for adding and scaling such maps are just what you'd expect.

Exercise 3.3.4. Prove that for $\mathcal{V}$ a real vector space of dimension n, the set of bilinear maps from $\mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ forms a vector space of dimension n^2.

We can now define trilinear maps from $\mathcal{V} \times \mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ and show it is a vector space of dimension n^3. And so on, with quadrilinear, quintrilinear and more to follow. Or we could cut the silliness with:

Definition 3.3.2. The (k-fold) cartesian product of the n-dimensional vector space $\mathcal{V}$ with itself k times will be written $\mathcal{V}^k$. A map f from $\mathcal{V}^k$ to $\mathbb{R}$ is said to be *multilinear* iff for each $i, 1 \leq i \leq k$ we have

$$\forall\ s, t \in \mathbb{R}, \forall\ \boldsymbol{u}_i, \boldsymbol{v}_i \in \mathcal{V} \quad f(\boldsymbol{u}_1, u_2, \cdots, s\boldsymbol{u}_i + t\boldsymbol{v}_i, \cdots, \boldsymbol{u}_k) =$$

$$sf(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_i, \cdots, \boldsymbol{u}_k) + tf((\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{u}_k)$$

Exercise 3.3.5. Prove that the k-multilinear maps form a vector space of dimension n^k. Is this still true when $k = 0$? Please explain.

Definition 3.3.3. A multilinear map $f : \mathcal{V}^k \longrightarrow \mathbb{R}$ is called a *k-tensor* on $\mathcal{V}$. We denote by $\mathcal{T}_k(\mathcal{V})$ the set of all these maps.

Remark 3.3.1. You have to allow that this sure saves ink. Tensors were found in nature by people who had a terrible job working out how to write them down effectively, and others who found out a way of writing them down which made doing sums with them just about possible, but understanding what you were doing almost impossible. A little thought shows that any k-tensor needs n^k numbers to specify it with respect to any basis. This gets to be rather a lot of numbers rather quickly. For $k = 2, n = 3$ we are looking at nine numbers which could be arranged in a 3×3 matrix without much fuss. And we could call one something like a_{ij} where i, j run from 1 to 3. And a k-tensor on an n-dimensional space with a basis would be written perhaps as $a_{i_1, i_2, \cdots i_k}$ where each i_j runs from 1 to n. It is hard to love things as ugly as this unless you gave birth to them.

It gets worse. The above is called a *covariant* k-tensor.

Definition 3.3.4. A *contravariant* k-tensor is a multilinear map

$$g : \mathcal{V}^{\star k} \longrightarrow \mathbb{R}$$

We write one as

$$a^{i_1,i_2,\cdots i_k}$$

We write $\mathcal{T}^k$ for the space of contravariant k-tensors.

Remark 3.3.2. It gets even worse. We can have a tensor of mixed variances:

Definition 3.3.5. A *tensor of type* $\begin{pmatrix} r \\ s \end{pmatrix}$, otherwise known as a tensor contravariant of order r and covariant of order s, is a $r+s$-multilinear map

$$h : \mathcal{V}^\star \times \mathcal{V}^\star \times \mathcal{V}^\star \times \cdots \times \mathcal{V}^\star \times \mathcal{V} \times \mathcal{V} \times \cdots \times \mathcal{V} \longrightarrow \mathbb{R}$$

where there are r copies of $\mathcal{V}^\star$ and s copies of $\mathcal{V}$.

The space of these maps is written $\mathcal{T}_s^r(\mathcal{V})$

We write one of the little sweeties in terms of a suitably chosen basis as:

$$a^{i_1,i_2,\cdots i_r}_{j_i,j_2,\cdots j_s}$$

and it is hard to believe that even the parents of these things loved them.

The good news is that we shall not have to deal with them much in this course in any practical detail, and for those who have to do so at any time in the future, there is always MATLAB.

Exercise 3.3.6. Prove that the space of tensors of type $\begin{pmatrix} r \\ s \end{pmatrix}$, form a vector space of dimension n^{r+s}. Say what a 'suitably chosen basis' is.

Remark 3.3.3. I shall consider only covariant tensors from now on. And I promise not to get involved with any sums, and to conduct everything in a coordinate free way as befits a pure mathematician. [5]

Definition 3.3.6. If f is a (covariant) k-tensor, and g is a (covariant) ℓ-tensor, then we can define a (covariant) $k+\ell$ tensor $f \otimes g$ by

$$f \otimes g(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_k, \boldsymbol{v}_{k+1} \cdots \boldsymbol{v}_{k+\ell}) = f(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_k) \cdot g(\boldsymbol{v}_{k+1} \cdots \boldsymbol{v}_{k+\ell})$$

where $\cdot$ denotes multiplication in $\mathbb{R}$. This new tensor is called the *tensor product* of f and g.

Note that the tensor product is not commutative. Note also that it makes sense to take the tensor product of mixed variance tensors. (But we won't.)

[5] Although I may use a basis inside a proof of something easy.

Exercise 3.3.7. Verify the following:

$$(f_1 + f_2) \otimes g = f_1 \otimes g + f_2 \otimes g$$

$$f \otimes (g_1 + g_2) = f \otimes g_1 + f \otimes g_2$$

$$\forall\ t \in \mathbb{R}, (tf) \otimes g = f \otimes (tg) = t(f \otimes g)$$

$$(f \otimes g) \otimes h = f \otimes (g \otimes h)$$

In view of the last, associativity, we may feel free to write $f \otimes g \otimes h$ without the parentheses and indeed such expressions as $f_1 \otimes f_2 \otimes \cdots \otimes f_q$.

Proposition 3.3.1. *$\mathcal{T}_1(\mathcal{V})$ is just the dual space $\mathcal{V}^\star$. Given a basis $e_1, \cdots, e_n$ for $\mathcal{V}$, there is a standard basis for $\mathcal{T}_1(\mathcal{V})$ $e_1^\star, \cdots, e_n^\star$ and we can express any tensor in $\mathcal{T}_1(\mathcal{V})$ by writing it as:*

$$f = \sum_{1 \leq i \leq n} a_i e_i^\star$$

We can express any tensor in $\mathcal{T}_k(\mathcal{V})$ by writing it as

$$f_1 \otimes f_2 \otimes \cdots \otimes f_k$$

for some choice of 1-tensors $f_i, 1 \leq i \leq k$. The set of all terms

$$e_{i_1}^\star \otimes e_{i_2}^\star \cdots e_{i_k}^\star$$

as the e_{i_j} run through all possible values in a basis for $\mathcal{V}^\star$ is a basis for $\mathcal{T}_k(\mathcal{V})$

Proof:

Instead of giving a proof I shall demonstrate this for $\mathcal{T}_2(\mathbb{R}^2)$ which is more illuminating, and leave you to write down the formal proof.

Suppose then that f is a 2-tensor on $\mathbb{R}^2$, that is f is a bilinear map $f : \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}$. Put:

$$\begin{aligned}
f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) &= a_{11} \\
f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) &= a_{12} \\
f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) &= a_{21} \\
f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) &= a_{11}
\end{aligned}$$

Then

$$f\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix}\begin{bmatrix} y_1 \\ 0 \end{bmatrix}\right) = a_{11}x_1y_1$$

by multilinearity, which we can write as:

$$f\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix}\begin{bmatrix} y_1 \\ 0 \end{bmatrix}\right) = a_{11}e_1^\star \otimes e_1^\star\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$

Similarly we have

$$\begin{aligned} f\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix}\begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right) &= a_{12}e_1^\star \otimes e_2^\star\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) \\ f\left(\begin{bmatrix} 0 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ 0 \end{bmatrix}\right) &= a_{21}e_2^\star \otimes e_1^\star\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) \\ f\left(\begin{bmatrix} 0 \\ x_2 \end{bmatrix}\begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right) &= a_{22}e_2^\star \otimes e_2^\star\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) \end{aligned}$$

Now we observe that:

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = f\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) + f\left(\begin{bmatrix} 0 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$

$$= f\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix}\begin{bmatrix} y_1 \\ 0 \end{bmatrix}\right)+f\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix}\begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right)+f\left(\begin{bmatrix} 0 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ 0 \end{bmatrix}\right)+f\left(\begin{bmatrix} 0 \\ x_2 \end{bmatrix}\begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right)$$

Or

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \sum_{1\leq i,j,\leq 2} a_{ij}e_i^\star \otimes e_j^\star\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$

So

$$f = \sum_{1\leq i,j,\leq 2} a_{ij}e_i^\star \otimes e_j^\star$$

You can see that if we change the dimension from 2 to n but stick with 2-tensors we just get

$$f = \sum_{1\leq i,j,\leq n} a_{ij}e_i^\star \otimes e_j^\star$$

by the same argument. And if instead of looking at 2-tensors we look at k-tensors, we get

$$f = \sum_{1\leq i_1,i_2,\cdots i_k,\leq n} a_{ij}e_{i_1}^\star \otimes e_{i_2}^\star \otimes \cdots e_{i_k}^\star$$

□

This justifies the representation of a tensor as the object

$$a_{i_1,i_2,\cdots,i_k}$$

which is a whole raft of n^k different numbers.

Remark 3.3.4. It may help to think of a tensor as associating a number, the coefficient, with every possible choice of k things from n (repetitions allowed). We choose something from the first n, say i_1 and write down $e^\star_{i_1}$. Now we make a choice from the second n and write down $e^\star_{i_2}$. When we have made our k choices we have

$$e^\star_{i_1} \otimes e^\star_{i_2} \otimes \cdots e^\star_{i_k}$$

With this particular selection we associate the number $a_{i_1,i_2,\cdots,i_k}$. Now we repeat for every possible choice-sequence. There are obviously n^k of these.

When $k = 2$ we can conveniently represent one of these things as an $n \times n$ matrix of numbers. For $k = 3$ we could (inconveniently) represent it as a cube of cells, n cells on each side, with a number in each cell. For $k > 3$, use MATLAB.

One of the problems students face when using tensors comes from the property of the dual space $\mathcal{V}^\star$. If

$$f : \mathcal{U} \longrightarrow \mathcal{V}$$

is a linear map between vector spaces, then there is induced a map

$$f^\star : \mathcal{V}^\star \longrightarrow \mathcal{U}^\star$$

defined by

$$\forall\ g \in \mathcal{V}^\star,\ f^\star(g) = g \circ f$$

This reversal of the direction worries some into imagining that $f^\star$ is an inverse. It isn't. The picture:

$$\mathcal{U} \xrightarrow{f} \mathcal{V} \xrightarrow{g} \mathbb{R}$$

shows what is happening. It gives one more reason for not confusing $\mathcal{V}$ and $\mathcal{V}^\star$.

The same thing happens with covariant tensors for the same reason: If

$$f : \mathcal{U} \longrightarrow \mathcal{V}$$

is the same linear map between vector spaces, there is induced a map

$$f^\star : \mathcal{T}_k(\mathcal{V}) \longrightarrow \mathcal{T}_k(\mathcal{U})$$

(with the same reversal of direction) defined by

$$\forall\ g \in \mathcal{T}_k(\mathcal{V}), \quad (f^\star g)(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots \boldsymbol{u}_k) = g(f(\boldsymbol{u}_1), f(\boldsymbol{u}_2), \cdots, f(\boldsymbol{u}_k))$$

It is easy to see that $f^\star(g)$ is a k-tensor in $\mathcal{T}_k(\mathcal{U})$.

Exercise 3.3.8. Prove the last claim.

Do not ever confuse $f^\star$ with an inverse to f. In the case where we have a matrix representing a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, the matrix representing the induced map from $\mathbb{R}^{m\star}$ to $\mathbb{R}_n$ (with respect to the dual basis to whatever basis we used to represent the original linear map) is the *transpose* of the matrix representing the original linear map.

Definition 3.3.7. If U is an open subset of $\mathbb{R}^n$, a *smooth k-tensor field* on U is a smooth (infintely differentiable) map

$$T : U \longrightarrow \mathcal{T}_k(\mathbb{R}^n)$$

The idea can be thought of as follows: To attach a tensor to a point $u \in U$ is just to provide for u some k-tensor $a_{i_1,i_2,\cdots i_k}$. This stack of n^k definite numbers depends on u. As we move about in U, the numbers change, smoothly. So we can write the tensor field as

$$a_{i_1,i_2,\cdots i_k}(u), \quad u \in U$$

where our stack of numbers is now a stack of functions from $U \subseteq \mathbb{R}^n$ to $\mathbb{R}$. These functions are all required to be infinitely differentiable for the tensor field to be smooth.

There, now you can tell your mummies and daddies that you now know what a tensor field on an open subset of $\mathbb{R}^n$ is. Pray that they don't ask you to give a brief explanation.

Note that a contravariant tensor field of order 1 is otherwise known as a vector field, and a covariant tensor field of order 1 is a differentiable 1-form, or *covector* field. So we have generalised the idea of a vector field rather a lot.

Particular cases are the *stress tensor* (a 2-tensor on $\mathbb{R}^3$) which is esential for the study of elasticity and deformations of solids, and hence necessary for many engineers and physicists, the *stress-energy* tensor is used in Physics, the Riemann curvature tensor in Physics and Mathematics, the Riemann metric tensor and the Einstein tensor in Geometry (hence Physics). These are all tensor fields and they are mostly of low order and defined on the space (or space-time) in which we live. If I have time I shall discuss how the state spaces of physical systems are often more naturally represented as covectors than as vectors.

3.4 Alternating Tensors

Alternating tensors, such as the 2-tensor Det on $\mathbb{R}^2$, are a very important special case so get a section to themselves. Most tensors are not alternating, so there are not too many of those that are, and they are all obtained by messing about with determinants, so are not hard to understand, although the notation sometimes obscures this. Again, doing the exercises will remove the mystery and make you feel comfortable about them. First we give a definition:

Definition 3.4.1. A k-tensor f on $\mathcal{V}$, an n-dimensional vector space,

$$f : \mathcal{V} \times \mathcal{V} \times \cdots \times \mathcal{V} \longrightarrow \mathbb{R}$$

is said to be *alternating* iff

$$\forall\ i, j,\ 1 \leq i < j \leq k,\ f(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_i, \cdots \boldsymbol{v}_j \cdots \boldsymbol{v}_k)$$

$$= -f(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_j, \cdots \boldsymbol{v}_i \cdots \boldsymbol{v}_k)$$

In other words, swapping any two (column) vectors reverses the sign of f.

For the record,

Definition 3.4.2. A k-tensor f on $\mathcal{V}$, an n-dimensional vector space,

$$f : \mathcal{V} \times \mathcal{V} \times \cdots \times \mathcal{V} \longrightarrow \mathbb{R}$$

is said to be *symmetric* iff

$$\forall\ i, j,\ 1 \leq i < j \leq k,\ f(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_i, \cdots \boldsymbol{v}_j \cdots \boldsymbol{v}_k)$$

$$= f(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_j, \cdots \boldsymbol{v}_i \cdots \boldsymbol{v}_k)$$

The example of Det regarded as a bilinear map from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ is a special case of an alternating tensor. We could write this 2-tensor as

$$0(e_1^\star \otimes e_1^\star) + 1(e_1^\star \otimes e_2^\star) + -1(e_2^\star \otimes e_1^\star) + 0(e_2^\star \otimes e_2^\star)$$

or just

$$(e_1^\star \otimes e_2^\star) - (e_2^\star \otimes e_1^\star)$$

for short.

It follows from the definition of an alternating tensor that if two (column) vectors are the same, then f must take the value zero on that particular element. Consequently, when we write out the coefficients, we must have zero for the coefficient corresponding to any repetition of one of the e_{i_j}. For if not, we could evaluate the alternating tensor on a suitably chosen set of k vectors and get the wrong answer; for example with the above expression for Det, if we had $a_{11} \neq 0$ then evaluating it on e_1, e_1 would give the result a_{11}. And reversing the order gives the same result and not its negative. So $a_{11} = 0$. In the same way, we cannot have a non-zero coefficient for any choice which repeats, for any $j \in [1..k]$ the term $e_j^\star$; for such a choice, say

$$e_{i_1}^\star \otimes e_{i_2}^\star \otimes \cdots e_j^\star \otimes \cdots \otimes e_j^\star \otimes \cdots \otimes e_{i_k}^\star$$

could be evaluated on

$$e_{i_1}, e_{i_2}, \cdots e_j, \cdots , e_j, \cdots , e_{i_k}$$

with the two repeated e_j subsequently reversed, and since the answer must be zero, the coeficient corresponding to the choice

$$e_{i_1}^\star \otimes e_{i_2}^\star \otimes \cdots e_j^\star \otimes \cdots \otimes e_j^\star \otimes \cdots \otimes e_{i_k}^\star$$

must be zero.

It follows that the non-zero coefficients can only come from making a choice of k things from n when we never make the same selection twice. (See Remark 3.3.4.)

It follows in particular that any alternating 2-tensor on $\mathbb{R}^2$ must have zero for the coefficient of $e_1^\star \otimes e_1^\star$ and also for the coefficient of $e_2^\star \otimes e_2^\star$.

It is easy to see that the coefficient of $e_1^\star \otimes e_2^\star$ must be the negative of the coefficient for $e_2^\star \otimes e_1^\star$. This guarantees that reversing the order of the two vectors on which the alternating tensor operates really does reverse the sign. In other words, the only alternating 2-tensors on $\mathbb{R}^2$ are scalar multiples of Det. The space of alternating 2-tensors on $\mathbb{R}^2$ is one dimensional and hence looks a lot like $\mathbb{R}$. Since we know $\mathbb{R}$ of old, this is very cheering.

Definition 3.4.3. The set of alternating k-tensors on $\mathcal{V}$ is written $\Omega_k(\mathcal{V})$.

Exercise 3.4.1. Prove $\Omega_k(\mathcal{V})$ is a vector space.

Looking now at the alternating 2-tensors on $\mathbb{R}^n$, we can see that the possibilities are limited: we can choose any pair $e^\star_{i_1}, e^\star_{i_2}$ provided $i_1 \neq i_2$, and assign a non-zero number to $e^\star_{i_1} \otimes e^\star_{i_2}$. There are $n(n-1)$ ways of doing this. But we also have to ensure that the coefficient for $e^\star_{i_1} \otimes e^\star_{i_2}$ is the negative of the coefficient for $e^\star_{i_2}, e^\star_{i_1}$. So the number of free numbers (the dimension of $\Omega^2(\mathbb{R}^n)$) is just $n(n-1)/2$. We can easily write down a basis for the space: it is the set

$$e^\star_{i_1} \otimes e^\star_{i_2} - e^\star_{i_2} \otimes e^\star_{i_1}$$

where $1 \leq i_1 < i_2 \leq n$.

Generalising this to alternating k-tensors on $\mathbb{R}^n$ is a bit muddling to think about, but you can see that we are allowed n different choices for the first $e^\star_{i_1}$ and $n-1$ choices for the second term $e^\star_{i_2}$, and so on until we have $n-k+1$ choices for the k^{th} and last term to give us:

$$e^\star_{i_1} \otimes e^\star_{i_2} \otimes \cdots \otimes e^\star_{i_k}$$

This gives us $n!/(n-k)!$ different choices. But just as swapping any two terms in the case $n = k = 2$ meant we had to swap the sign, so does it here. Consequently once we have a coefficient for any choice sequence, then the coefficient for any permutation of those choices is determined. If one permutation differs from another by one (or an odd number of) swap of terms, then the coefficient of one is the negative of the coefficient of the other. If it differs by two (or an even number of) of swaps of terms, then the coefficients must be the same. So the number of possible distinct coefficients we are free to choose is

$$\frac{n!}{(n-k)!k!}$$

since the number of permutations of k things is, of course, k!

This tells us that $\Omega_k(\mathbb{R}^n)$ is a subspace of $\mathcal{T}_k(\mathbb{R}^n)$ and that it has dimension $n!/(n-k)!k!$ It also shows that a basis for the space can be obtained by writing down every choice of k different projection maps $e^\star_j$, and given a choice we then write down every possible permutation of it with a plus sign when the permutation is even and a minus sign when it is an odd permutation (i.e. obtained by an odd number of swaps). The set of all these choices forms a basis for $\Omega_k(\mathbb{R}^n)$.

Exercise 3.4.2. Write down the basis for $\Omega_3(\mathbb{R}^3)$. Evaluate the single basis element on any three different vectors in $\mathbb{R}^3$. Do you recognise it? (The answer had better be 'yes' or you might do well to leave home and University and run away to sea or take up something easier, basketweaving perhaps.)

Exercise 3.4.3. Write down the basis for $\Omega_2(\mathbb{R}^3)$. Evaluate each of the basis elements on any pair of vectors in $\mathbb{R}^3$ of your choice. Do you recognise each of the maps?

You can see that for k small and $n \leq 4$ this is not too bad and it only starts to get out of hand for large dimensions and large orders. Of course, if $k > n$ life is very simple and $\Omega_k(\mathbb{R}^n)$ collapses to the vector space containing only the zero element with zero basis elements. This is because the number of ways of choosing 35 different things from a set of 29 objects is zero.

There is a nice map which makes a tensor f symmetric: all it does is to take the average of f over all the permutations:

$$\text{Sym} : \mathcal{T}_k(\mathcal{V}) \longrightarrow \mathcal{T}_k(\mathcal{V})$$

and

$$\text{Sym } f(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_k) = \frac{1}{k!} \sum_{\sigma \in S_k} f(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}, \cdots, \boldsymbol{v}_{\sigma(k)})$$

In this notation, S_k is the set (actually, *group*) of permutations on the k subscripts in $[1..k]$. It should be obvious now that for any $f \in \mathcal{T}_k(\mathcal{V})$, Sym$(f)$ is a symmetric tensor which does not change its value when we swap any two vectors in its argument.

Symmetrising tensors is not a particularly pressing need, but making them alternate is.

Fortunately there is also a nice map which 'makes a tensor alternate', i.e. a map

$$\text{Alt} : \mathcal{T}_k(\mathcal{V}) \longrightarrow \Omega_k(\mathcal{V})$$

What it does is to take a tensor f and then Alt(f) on a sequence $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k$ is obtained by taking the average value of the tensor f on every permutation of this sequence, *signs being taken into account*, so that an odd permutation counts negative. Formally:

$$(\text{Alt}(f))(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma) f(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}, \cdots, \boldsymbol{v}_{\sigma(k)})$$

S_k is again the set of permutations on the k subscripts in $[1..k]$, and if σ is one of them, then if it can be obtained by an even number of permutations, $\text{sgn}(\sigma) = 1$ and if by an odd number of permutations, $\text{sgn}(\sigma) = -1$. I leave those of you who haven't done any group theory to experiment to ensure that no permutation is both even and odd, and that every permutation is one or the other.

Exercise 3.4.4. Take the two tensor f on $\mathbb{R}^3$ given by

$$2e_1^\star \otimes e_1^\star + 3e_1^\star \otimes e_2^\star + 4e_2^\star \otimes e_1^\star + 5e_2^\star \otimes e_2^\star$$

Calculate Alt(f) and give it in terms of the basis you have found in earlier questions.

It is easy to prove that:

Proposition 3.4.1.

$$\forall\ f \in \mathcal{T}_k(\mathbb{R}^n), \text{Alt}(f) \in \Omega_k(\mathbb{R}^n)$$

Proof:

□

It is also easy to see that Alt takes an alternating tensor to itself:

Proposition 3.4.2.

$$\forall\ f \in \Omega_k(\mathbb{R}^n), \qquad \text{Alt}(f) = f$$

Proof:

When we add up all the permutations we are adding up $k!$ numbers which are all the same, $f(\boldsymbol{v}_1, v_2, \cdots, \boldsymbol{v}_k)$ (since two minus signs become a plus sign when the permutation is odd) and this gives therefore a sum of $k!\ f(\boldsymbol{v}_1, v_2, \cdots, \boldsymbol{v}_k)$ and on dividing by $k!$ we get the result. □

Remark 3.4.1. It follows that Alt is an idempotent operator on the exterior algebra, doing it once is the same as doing it any number of times.

We can now define a wedge product of two alternating tensors. It is obtained by taking the tensor product, but of course the tensor product of two alternating tensors isn't going to be alternating. So we Alt it, which makes sense, but we stick a weird looking coefficient in front of it:

Definition 3.4.4.

$$\forall f \in \Omega_k(\mathcal{V}),\ \forall g \in \Omega_\ell(\mathcal{V}) \quad f \wedge g \triangleq \frac{(k+\ell)!}{k!\ell!}\mathrm{Alt}(f \otimes g) \in \Omega_{k+\ell}(\mathcal{V})$$

Exercise 3.4.5. Calculate $e_1^\star \wedge e_2^\star$ in $\Omega_2(\mathbb{R}^3)$. Can you see what the weird coefficient does that makes the result definitely more cool than it would otherwise be?

Exercise 3.4.6. When $f = 2e_1^\star + 3e_2^\star + 3e_3^\star \quad \in \Omega_1(\mathbb{R}^3)$ and $g = 3e_1^\star + 2e_2^\star - 1e_3^\star \quad \in \Omega_1(\mathbb{R}^3)$, calculate $f \wedge g \in \Omega_2(\mathbb{R}^3)$ Do you recognise the result? (The answer 'no' is wrong.

Exercise 3.4.7. Calculate $e_1 \star \wedge e_2^\star \wedge e_3^\star$ on $\mathbb{R}^3$ in three different ways. First do it by calculating $e_1^\star \wedge e_2^\star$ and then 'vecking' it with $e_3^\star$. Then by calculating $e_2^\star \wedge e_3^\star$ and vecking $e_1^\star$ by it, and finally by applying Alt to $e_1^\star \otimes e_2^\star \otimes e_3^\star$ with a suitable coefficient to make it come good. Or not.

Exercise 3.4.8. Verify the following hold for all alternating k-tensors f and alternating ℓ-tensors g, any alternating tensor h, and all real numbers a:

1. $$(f+g) \wedge h = f \wedge h + g \wedge h$$

2. $$f^(g+h) = f \wedge g + f \wedge h$$

3. $$af \wedge g = f \wedge ag = a(f \wedge g)$$

4. $$f \wedge g = (-1)^{k\ell}(g \wedge f)$$

Exercise 3.4.9. Verify that

$$f^\star(g \wedge h) = f^\star g \wedge f^\star h$$

where g, h are alternating tensors on a vector space $\mathcal{V}$ and

$$f : \mathcal{U} \longrightarrow \mathcal{V}$$

is a linear map between vector spaces.

Exercise 3.4.10. Find a 2-tensor f on $\mathbb{R}^3$ which has $\text{Alt}(f) = 0$. Take any nonzero 1-tensor g on $\mathbb{R}^3$ and compute $f \otimes g$ and $\text{Alt}(f \otimes g)$. Now compute $g \otimes f$ and Alt of that.

Exercise 3.4.11. Show that if f is a k-tensor and g an ℓ-tensor and $\text{Alt}(f) = 0$, then $\text{Alt}(f \otimes g) = \text{Alt}(g \otimes f) = 0$. Likewise if $\text{Alt}(g)$=0.

Proposition 3.4.3. $\text{Alt}(f \otimes \text{Alt}(g \otimes h)) = \text{Alt}(f \otimes g \otimes h)$

Proof:

$\text{Alt}(\text{Alt}(g \otimes h) - g \otimes h) = \text{Alt}(g \otimes h)$ - $\text{Alt}(g \otimes h) = 0$.

Hence by the preceding exercise,

$$0 = \text{Alt}(f \otimes (\text{Alt}(g \otimes h) - g \otimes h)$$

$$\Rightarrow \text{Alt}(f \otimes \text{Alt}(g \otimes h)) = \text{Alt}(f \otimes g \otimes h)$$

□

Exercise 3.4.12. Show that $f \wedge (g \wedge h) = f \wedge g \wedge h$ follows from the last proposition.

We note that the space of alternating n-tensors on any n-dimensional vector space is 1, and we know that on $\mathbb{R}^n$ any such alternating tensor is some number times Det. The following result should therefore come as no surprise:

Proposition 3.4.4. *Let $(\boldsymbol{v}_1, v_2, \cdots \boldsymbol{v}_n)$ be a basis for an n-dimensional vector space $\mathcal{V}$, and $T : \mathcal{V} \longrightarrow \mathcal{V}$ a linear map, represented relative to the given basis by the matrix $[T]$. If $\omega \in \Omega_n(\mathcal{V})$ and $\forall\ i \in [1..n], \boldsymbol{u}_i = T\boldsymbol{v}_i$, then*

$$\omega(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots \boldsymbol{u}_n) = Det[T]\omega(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n)$$

Proof:

For any $n \times n$ matrix $A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$ (where the $\boldsymbol{a}_j$ are (column) vectors in $\mathbb{R}^n$) define the alternating n-tensor α by

$$\alpha(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) = \omega(\boldsymbol{a}_1^\star(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n), \boldsymbol{a}_2^\star(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n), \cdots, \boldsymbol{a}_n^\star(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n))$$

Since ω is alternating, so is α. Since α is an alternating n-tensor on $\mathbb{R}^n$, it is some multiple of Det, i.e. $\alpha = t \cdot Det$. Taking the case where A is the identity matrix we get

$$tDet[I_n] = t = \omega(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n)$$

Then since $\alpha([T]) = \omega(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n)$ the result follows. □

There is an interesting and important consequence of the last proposition: suppose we have two distinct bases for $\mathcal{V}$, a vector space of dimension n. Then there is an invertible linear map T which goes from one to the other. Any non-zero alternating n-tensor will have its evaluation on the two bases related by a non-zero determinant. The sign of the determinant of the map is the same as the sign of its inverse, and cannot be zero. So we can obtain, for any basis, that another basis is either in the same class as it when the determinant of the map T with respect to the starting basis is positive, or in the opposite camp when the determinant is negative. In other words, the bases for $\mathcal{V}$ are divided into two camps. This division of the bases into two camps does not depend on the n-tensor ω.

One such camp is called an *orientation* of $\mathcal{V}$. Two bases in the same camp have the same orientation. We cannot call one positive and the other negative with any assurance that someone else won't have chosen a different ω and assigned a different value of positive and negative, but at least we will always agree on when two orientations are the same, i.e. if two bases are in the same camp.

In $\mathbb{R}^n$ we can go further and define $(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)$ to be the positive or usual orientation.

This will turn out to be related to the orientability of manifolds.

You can see that I have taken the exterior algebra for a vector space and embedded it in the space of multilinear tensors on the same vector space, in fact as the space of alternating tensors. Given a basis for the space, the alternating tensors can be represented by arrays of arrays ... of arrays of numbers. This makes them fairly definite objects we can do sums with. The classical Tensor Calculus did exactly this. One can only admire the fortitude of these guys.[6]

[6] Of course, this was before television when people had more time on their hands. And if Australian television gets much worse, maybe Oz will turn out some great mathematicians of a computational temperament, but I expect they will still use MATLAB.

3.5 Groups and Group Actions

And now for something really easy. A *group* is a collection of objects which it makes sense to add and subtract. Or to multiply and divide, since there isn't much difference at this level of abstraction.

Example 3.5.1. The integers under addition, $\mathbb{Z}, +$ is a group.

Example 3.5.2. The real numbers under addition, $\mathbb{R}, +$ is another group. The first group is a *subgroup* of this one

Example 3.5.3. $\mathbb{Z}_{12}, +$ is the 'clock arithmetic group where addition is done mod 12.

Example 3.5.4. $\mathbb{Z}_2$ is the simpler group consisting of just two elements with addition table:

+	0	1
0	0	1
1	1	0

This is not a subgroup of $\mathbb{Z}_{12}$, although $\{0, 6\}$ is a subgroup which is isomorphic to $\mathbb{Z}_2$, a term we shall define soon, and which definition you are invited to guess.

Remark 3.5.1. The real numbers under multiplication is *NOT* a group.

Definition 3.5.1. A *group* is a pair $(G, \star)$ consisting of a set G and a *binary operation* $\star : G \times G \longrightarrow G$ such that

1. $\forall\ a, b, c \in G, \quad a \star (b \star c) = (a \star b) \star c$
2. $\exists\ e \in G,\ \forall\ \ a \in G, \quad a \star e = e \star a = a$
3. $\forall\ a \in G,\ \exists\ a^{-1} \in G, \quad a \star a^{-1} = a^{-1} \star a = e$

If in addition:

4. $\forall\ a, b \in G, \quad a \star b = b \star a$

Then G is said to be *abelian* or *commutative.*

Example 3.5.5. The above examples of groups are all abelian. We tend to use $+$ as the name of the operator in this case. The group of all rotations about the origin in $3-$space (with composition as the binary operator) is not abelian. Note that any vector space is an abelian group.

As usual, if we define a new collection of entities, we define maps which preserve the structure on those entities:

Definition 3.5.2. A *homomorphism* $f : (G, \star) \longrightarrow (H, \diamond)$ is a map $f : G \longrightarrow H$ such that

$$\forall\ a, b \in G,\ \ f(a \star b) = f(a) \diamond f(b)$$

Usually we drop the sign for the operation and instead of $a \star b$ we write ab. We also confuse the two operations $\star$ and $\diamond$ despite their being on different sets, and so the definition of a homomorphism is $f(ab) = f(a)f(b)$ which is confusing at first.

Do not confuse *homomorphisms* with *homeomorphisms* because, like weasels and stoats, they is stoatally different.

Definition 3.5.3. If $f : G \longrightarrow H$ is a homomorphism of groups and there exists $g : H \longrightarrow G$ which is also a homomorphism and such that $f \circ g$ is the identity on H and $g \circ f$ is the identity map on G, then f is said to be an *isomorphism* and the two groups are said to be *isomorphic*

Remark 3.5.2. It is sufficient for f to be an isomorphism that it be a homomorphism which is 1-1 and onto. The corresponding statement for continuous maps between topological spaces is false.

Exercise 3.5.1. Prove both of the above claims.

Remark 3.5.3. It has been said that an isomorphism of groups (or anything else) is just a change of name for what is basically the same underlying object. Well, sometimes the actual objects matter, but the idea is a good one.

Definition 3.5.4. A *subgroup* of a group $(G, \star)$ is a subset $H \subseteq G$ which is also a group under $\star$.

Example 3.5.6. $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$.

Example 3.5.7. The group of 2×2 matrices which have columns orthogonal and of norm 1 are a subgroup of the set of all invertible 2×2 matrices under matrix multiplication.

Definition 3.5.5. If $(G, \star)$ and $(H, \diamond)$ are groups, the *product group* is the cartesian product $G \times H$ with the binary operation defined by

$$(g_1, h_1)\spadesuit(g_2, h_2) \triangleq (g_1 \star g_2, h_1 \diamond h_2)$$

Usually we just forget all about having different symbols for the different binary operations and write:

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1, h_2)$$

Example 3.5.8. $(\mathbb{R}^2, +)$ is the product of $\mathbb{R}, +$ with itself.

Exercise 3.5.2. Confirm that.

Groups are important elements in a huge number of applications. They have a lot to do with symmetries, as the next thread of definitions and examples should make clear:

Definition 3.5.6. A *Group Action* on a set X is a map

$$m : G \times X \longrightarrow X$$

such that:

1. $\forall \ x \in X, \quad m(e, x) = x$
2. $\forall \ \boldsymbol{x} \in X, \ \forall \ a, b \in G, \quad m(a, m(b, x)) = m(ab, x)$

Example 3.5.9. The group $\mathbb{Z}_2$ acts on the space $\mathbb{R}^n$ with $\forall \boldsymbol{x} \in \mathbb{R}^n, \ m(0, \boldsymbol{x}) = \boldsymbol{x}$ and $\forall \ \boldsymbol{x} \in \mathbb{R}^n, \quad m(1, \boldsymbol{x}) = -\boldsymbol{x}$. It is clear that the action of $1 \in \mathbb{Z}_2$ is to swap every vector to its negative.

Example 3.5.10. $\mathbb{Z}_2$ acts on the space of complex numbers, $\mathbb{C}$ by 0 sending every number to itself and 1 sending every number to its complex conjugate.

Example 3.5.11. $\mathbb{R}$ acts on the space $\mathbb{R}^n$ with $m(t, \boldsymbol{x}) = t\boldsymbol{x}$.

Example 3.5.12. Let G be the group with four elements and multiplication table:

$\star$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

Let X be the set consisting of the square in $\mathbb{R}^2$ with vertices at ± 1 which I shall call the 4-square to distinguish it from the unit square.

Let the group action be given by e leaves everything as is, a rotates by a right angle in the anticlockwise sense, b by two right angles, same sense, c by three right angles anticlockwise or, what is equivalent, by one right angle clockwise.

Exercise 3.5.3. Verify that this gives a group action of G on X.

This should indicate what groups have to do with symmetries of objects.

One point which by now should be obvious: Instead of taking a group action to be a map from $G \times X$ to X, we could equally regard it as a map which takes each element of G and assigns it to a particular map from X to X. If Y^X denotes the set of all maps from X to Y, then there is a natural equivalence between $Z^{X \times Y}$ and $(Z^Y)^X$.

Exercise 3.5.4. Prove that $\mathbb{R}^2 \cong \mathbb{R}^2$, where $2 = \{0, 1\}$ and the first $\mathbb{R}^2$ is the set of maps from 2 into $\mathbb{R}$ and the second $\mathbb{R}^2$ is short for $\mathbb{R} \times \mathbb{R}$.

This leads to the following thread of ideas:

Definition 3.5.7. For any set X, $\mathrm{Aut}(X)$ is the set of bijective (1-1 and onto) maps of X into itself.

Remark 3.5.4. For any non-empty set, $(\mathrm{Aut}(X), \circ)$ is a group.

Exercise 3.5.5. Prove the above claim.

We can now give an alternative and more intuitive definition of a group action of a group G on a set X,

Definition 3.5.8. A Group Action of a group G on a set X is a homomorphism from G to the group $(\mathrm{Aut}(X), \circ)$.

Exercise 3.5.6. Prove the two defintions are equivalent.

So we have that G picks out some useful subgroup of the set of all automorphisms of X. And if X is not just any old set but, say, a vector space, then we might have a subgroup of the group of isomorphisms of the space with itself.

Remark 3.5.5. There is a lot to Group Theory, and teaching the subject to you is not my job, but I hope you can discern something of its charm from the tiny bit of it you have seen here.

3.6 Topological Groups

A set may be both a topological space and a group. In fact any set can be given a topology and any non-empty set can be made into a group, so this is true in a not very useful sense. It is only when the group operations have some connection with the topology that anybody cares. This leads to:

Definition 3.6.1. A *topological group* is a non empty set G which is (a) a group, that is there exists a binary operation $\star : G \times G \longrightarrow G$, which is associative, has an identity and has inverses, and (b) is a topological space, that is has a collection of subsets which includes the empty set, G, and which is closed under finite intersections and any unions, such that

$$\star : G \times G \longrightarrow G$$

is continuous (with the product topology on $G \times G$) and

$$Inv : G \longrightarrow G$$

the map which takes each $a \in G$ to $a^{-1} \in G$ is also continuous with respect to the given topology.

Example 3.6.1. $\mathbb{R}, +$ is a topological group with the usual addition and with the usual topology derived from the usual metric, $d(x, y) = |x - y|$.

Exercise 3.6.1. Prove it.

Example 3.6.2. The group of invertible 2×2 matrices under multiplication with the metric given by

$$d(a_{ij}, b_{ij}) = max_{1 \leq i,j \leq 2} \{|a_{ij} - b_{ij}|\}$$

Exercise 3.6.2. Prove it.

When the topological group is a manifold then it makes sense to look at the multiplication and inverse maps and to ask if these are not only continuous but *differentiable*. The only manifolds you have met so far which qualify are $\mathbb{R}^n, +$ and, with some optimism on my part, the group of invertible $n \times n$ matrices. If in addition to being continuous the operations of multiplication and inversion are differentiable, the topological group is called a *Lie Group*. Obviously this doesn't make sense for most topological groups, but it does for some of the more obvious ones.

Exercise 3.6.3. Prove that $GL(n, \mathbb{R})$ defined as the group of $n \times n$ invertible matrices with real entries is a Lie Group.

3.7 Category Theory

This is not some new subject to intimidate you, it is a bit of terminology to assist you in making sense of the rather bewildering collection of things you are meeting.

Definition 3.7.1. A *Category* is a class of things called *objects* and another class of things called *maps*. Each object A is associated with a unique map I_A called the *identity map* for A. Each map f is associated with an ordered pair of objects, one called $\mathrm{Dom}(f)$ and the other called $\mathrm{Codom}(f)$, short for Domain of f and Codomain of f respectively. $\mathrm{Dom}(I_X) = \mathrm{Codom}(I_X) = X$ for every identity map. Two maps f, g can be composed to give $f \circ g$ precisely when $\mathrm{Codom}(g) = \mathrm{Dom}(f)$ and whenever $f \circ I_X$ makes sense it is f, and when $I_X \circ f$ is defined it is f. We write $A \xrightarrow{f} B$ when f has $\mathrm{Dom}(f) = A$, $\mathrm{Codom}(f) = B$.

You may recognise a few of these:

Example 3.7.1. The category of topological spaces and continuous maps

Example 3.7.2. The category of vector spaces and linear maps

Example 3.7.3. The category of groups and homomorphisms

Example 3.7.4. The category of normed linear spaces and continuous linear maps

Example 3.7.5. The category of metric spaces and continuous maps.

Just as we defined groups and then their structure preserving maps:

Definition 3.7.2. A *Covariant Functor* between categories written

$$F : \mathcal{C}_1 \longrightarrow \mathcal{C}_2$$

assigns to each object in $\mathcal{C}_1$ an object in $\mathcal{C}_2$ and to each map in $\mathcal{C}_1$ a map in $\mathcal{C}_2$, in such a way that

$$A \xrightarrow{f} B \quad \rightsquigarrow \quad F(A) \xrightarrow{Ff} F(B)$$

which satisfies the condition $F(g \circ f) = Fg \circ Ff$ for all g, f such that $g \circ f$ exists, and

$$\forall A \in \mathcal{C}, F(I_A) = I_{F(A)}$$

I note that there are *contravariant functors* which reverse the direction of the maps.

Example 3.7.6. There is a functor called Forgetful_1 from vector spaces to groups which just forgets that the vectors can also be scaled by real numbers.

Example 3.7.7. There is another forgetful functor Forgetful_2 from metric spaces to topological spaces that forgest that the topology comes from a metric.

Exercise 3.7.1. Make a list of all the forgetful functors you know of and all the categories you know of and draw a humungous Venn diagram where forgetful functors are represented by inclusions. I have drawn a bit of this in figure 3.1 where T is the category of topological spaces, G is the category of groups, V is the category of vector spaces and B is the category of normed vector spaces.

If $\mathcal{V}$ is the category of real vector spaces and linear maps, then there is a contravariant functor $[-, \mathbb{R}]$ which takes every vector space V to its dual space the elements of which are the linear maps from V to $\mathbb{R}$, $[V, \mathbb{R}]$. The map $f : V \longrightarrow W$ between vector spaces is taken by the functor $[-, \mathbb{R}]$ to the map

$$f^\star : W^\star \longrightarrow V^\star$$

defined by

$$g \rightsquigarrow g \circ f$$

Exercise 3.7.2. Verify that this is indeed a contravariant functor.

We can define a number of things in abstract categories and then find them cropping up in particular categories. These include subobjects, quotient objects, products, and isomorphisms, together with the idea of maps being 1-1 and onto.

As an easy and useful example, If A and B are two objects in any category, we say that $A \xrightarrow{f} B$ is an *isomorphism* iff there is $B \xrightarrow{g} A$ in the category such that $f \circ g = I_B$ and $g \circ f = I_A$. In which case, A and B are said to be *isomorphic*. Unfortunately, the isomorphisms tend to be called other things in particular categories, for example calling them homeomorphisms in the case of the category of topological spaces and continuous maps. As usual, the particular categories had to exist before anyone thought of abstracting them. In the category of smooth manifolds and smooth maps the isomorphisms are called smooth *diffeomorphisms*. For you at the moment this makes sense on the only things you can be sure are smooth manifolds, namely open subsets of $\mathbb{R}^n$.

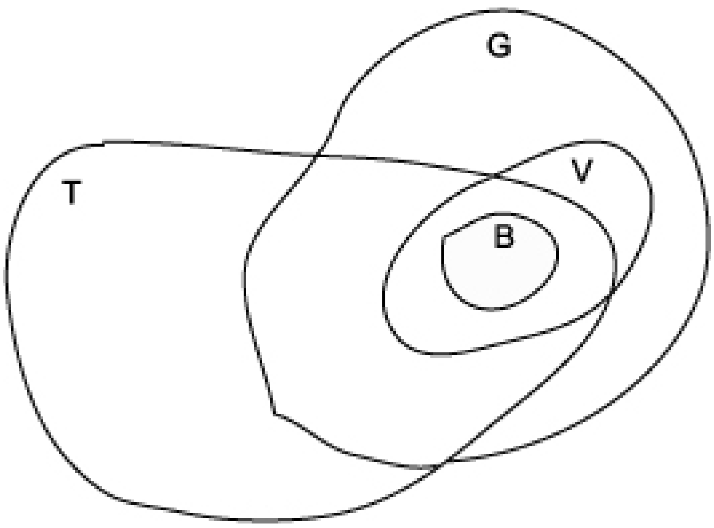

Figure 3.1: Some relations between four categories you know about.

Chapter 4

Analysis Preliminaries

4.1 Tangency

The notion of differentiating a function tends to be tied in, for most of you, with the computations you did in first or second year, and consequently you may be very hazy about the underlying ideas. This makes for trouble later, in fact now, so it would be a good idea to get the basics clear. Anybody can be trained to do computations. Hey, that's what computers are for. Mathematicians on the other hand are insterested in ideas.

The fundamental idea is that two maps can be *tangent* at a point. For this to be the case they have to actually take the same value at the point, but this is clearly not enough, as figure 4.1 indicates for maps from $\mathbb{R}$ to $\mathbb{R}$, the orange and the green are tangent, the red and blue are not. What we need is that even if we look at the crossing point under arbitrarily high magnification, the two curves look close to each other near the crossing point. This leads to:

Definition 4.1.1. Two maps $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are *tangent* at $\boldsymbol{a} \in \mathbb{R}^n$ iff $f(\boldsymbol{a}) = g(\boldsymbol{a})$ and

$$\lim_{\|\boldsymbol{h}\|\to 0} \frac{\|f(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a}+\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$$

This says, in English, that the distance between the values of f and g near $\boldsymbol{a}$ is small, *even compared with* $\|\boldsymbol{h}\|$*, the size of the region we are looking at,* and is indeed getting even smaller as the size of $\boldsymbol{h}$ also gets smaller.

Tangency is more fundamental than the idea of the derivative of a map, since we shall see that it makes sense to decide if two maps from, for example, the

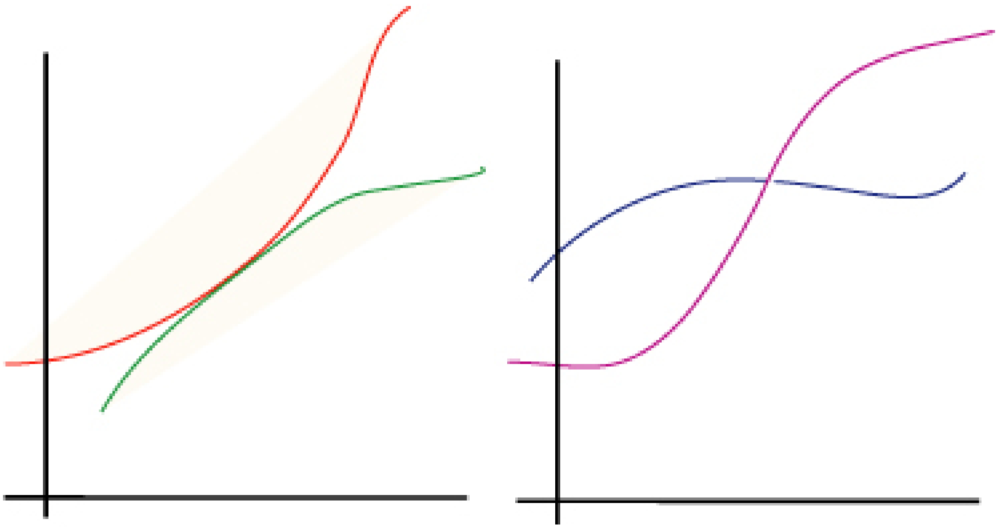

Figure 4.1: Two maps that are tangent at a point (and two that aren't)

circle S^1 to the sphere S^2 are tangent at a point, but it makes little sense to talk of an affine map from S^1 to S^2 since these are not linear spaces.

Tangency is an *equivalence relation* on the set of continuous maps taking a given value at $\boldsymbol{a}$, which is to say that any map is tangent to itself at $\boldsymbol{a}$, that if f is tangent to g at $\boldsymbol{a}$, then g is tangent to f at $\boldsymbol{a}$, and finally that if f is tangent to g at $\boldsymbol{a}$ and g is tangent to h at $\boldsymbol{a}$, then f is tangent to h at $\boldsymbol{a}$. As with any equivalence relation, this divides the set of maps with a given value of $f(\boldsymbol{a})$ into *equivalence classes*.[1]

When looking at maps from a normed vector space to another, we are fortunate to have a special class, the *affine* maps, which are of the form $f(\boldsymbol{x}) = T(\boldsymbol{x}) + \boldsymbol{b}$ for T a linear map and $\boldsymbol{b}$ some particular element of $\mathbb{R}^m$. Linear maps between finite dimensional vector spaces can be represented by matrices and are therefore real things we can do sums with. It is the case that for every point $\boldsymbol{a}$ in $\mathbb{R}^n$ and quite a lot of tangency classes of maps,

[1] Think of it this way: We have a bunch of cows in a field. If two cows have the same colour they are equivalent. Obviously, if Sally is a cow in the field then she is of the same colour as herself; if Sally and Betsy are cows of the same colour then Betsy and Sally have the same colour, and finally if Sally is the same colour as Betsy and Betsy is the same colour as Milly, then Sally is the same colour as Milly. Consequently, being of the same colour is an equivalence relation, and if we are tidy minded we would make sure all the brown cows were in one corner, all the white cows in another, the red cows in a third, the blue cows in a fourth, and so on. And what we can do for cows we can do for maps tangent at a point.

there is an affine map in the class.

Definition 4.1.2. If the map $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is tangent to an affine map at $\boldsymbol{a} \in \mathbb{R}^n$, then f is said to be *differentiable* at $\boldsymbol{a}$, and the linear part of the affine map is called the *derivative* of f at $\boldsymbol{a}$.

Remark 4.1.1. We can write this as:

Definition 4.1.3. The map $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *differentiable at* $\boldsymbol{a}$ *in* $\mathbb{R}^n$ iff there is a map $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ which is *linear*, such that

$$\lim_{\|\boldsymbol{h}\| \to 0} \frac{\|f(\boldsymbol{a}+\boldsymbol{h}) - (T(\boldsymbol{h}) + f(\boldsymbol{a}))\|}{\|\boldsymbol{h}\|} = 0$$

The map T is the *derivative* (sometimes *differential*) of f at $\boldsymbol{a}$.

If such a T exists, it is unique. The next sequence of exercises will establish this important point.

Exercise 4.1.1. For T a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, define

$$\|T\| = \sup_{\|\boldsymbol{x}\|=1} \|T(\boldsymbol{x})\|$$

Show that the definition ensures that every such T has a norm, $\|T\|$, and that $\|T\| = 0 \Rightarrow T = \mathbf{0}$.

Exercise 4.1.2. Show that $\forall\ \boldsymbol{x} \in \mathbb{R}^n, \quad \|T(\boldsymbol{x})\| \leq \|T\|\|\boldsymbol{x}\|$

Exercise 4.1.3. Prove that if two affine maps are tangent at a point they are identical.

Exercise 4.1.4. Show that if $f : U$
R^m is tangent to $g : U \longrightarrow$ at $\boldsymbol{a}inU$ for U open in $\mathbb{R}^n$, and if f is continous, then g is also continuous.

Exercise 4.1.5. Prove that a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous at the origin. (Hint: use compactness of the unit ball.)

Exercise 4.1.6. Prove that a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous everywhere.

This gives us

Proposition 4.1.1. *If $f : U \longrightarrow \mathbb{R}^m$ is differentiable at $\boldsymbol{a} \in U$ for U some open subset of $\mathbb{R}^n$, then f is continous at $\boldsymbol{a}$*

Proof:

Since $f(\boldsymbol{a}) = A(\boldsymbol{a})$ for some affine map A and A is continuous, the requirement

$$\lim_{\|\boldsymbol{h}\|\to 0} \frac{\|f(\boldsymbol{a}+\boldsymbol{h}) - A(\boldsymbol{a}+\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$$

implies that f is also continuous. □

In the context of maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ (or indeed between any two normed vector spaces) it is then legitimate to define the linear part of the unique affine map which is tangent to f at $\boldsymbol{a}$ (when it exists) as the *derivative* of f at $\boldsymbol{a}$ and we can write it as $Df(\boldsymbol{a})$ with a clear conscience.

There is a problem involved in doing this on a manifold such as a sphere, where linear maps are not to be found. How we get around this will be seen in the next chapter.

It would be easy to go on and define partial derivatives and to show how, if we use the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, we get the matrix of partial derivatives representing the linear part of the affine map tangent to f. Try it yourselves if you don't believe me.

Intuitively, the derivative of f at $\boldsymbol{a}$ is the linear part of the best affine approximation to f at $\boldsymbol{a}$. Picturing the difference between f and an affine map as being negligible when we are close enough to $\boldsymbol{a}$ is the main theme behind a lot of the arguments we shall be using.

4.2 The Inverse Function Theorem

Although you have officially had this result proved last year, I shall go over it again because it is instructive to have a clear picture of it in your heads. Since I want to make you love proving theorems, it is only fair to give you a serious theorem that needs a fair amount of work. By your standards, at least.

First the one dimensional case:

Proposition 4.2.1. *If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuously differentiable at $a \in \mathbb{R}$ and if $f'(a) \neq 0$, then there is an interval U containing a such that $f|U$ is 1-1 and has an inverse, f^{-1}, that is differentiable in the interval $f(U)$with, for any $y \in f(U)$,*

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$$

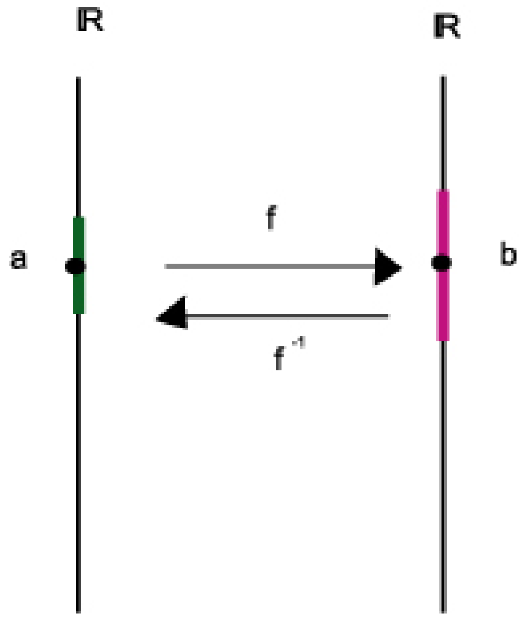

Figure 4.2: Inverse Functions

Proof:

Figure 4.2 shows the situation where I have put $f(a) = b$ to save ink. We may assume without loss of generality that $f'(a) > 0$. Then since it is continuous, there is some open interval U on a such that $f'(x) > 0$ for all $x \in U$. This makes f monotone increasing on U by the Mean Value Theorem.

Let $f(U)$ (purplish in the figure) be the image of U (green in the figure). Then I claim that $f(U)$ is an open interval on b. It is clearly an interval having no holes in it, since f being differentiable must be continuous and the Intermediate Value Theorem tells us that there cannot be a point missing between any two points of $f(U)$.

But it might be a closed interval or maybe contain its supremum or infimum. This doesn't happen: suppose the supremum of $f(U)$ were in $f(U)$. Then we would have to have some $u \in U$ with $f(u)$ equal to this supremum. But there are points bigger than u in U, and these would have to go to points less than the supremum, which contradicts f being monotone increasing. Consequently the interval $f(U)$ is open at the top and by the same argument it is also open at the bottom, so $f(U)$ is an open interval.

By the same argument, any open sub-interval of U is sent to an open sub-interval of $f(U)$.

$f|U$ is 1-1. Suppose not; then if $\exists\ x < y \in U,\ f(x) = f(y)$, then there is, by the Mean Value Theorem, some point z with $x < z < y$ and $f'(z) = 0$, contradiction. So no such x, y exist. $f|U$ is onto its image $f(U)$ by definition. So there is an inverse, f^{-1} from $f(U)$ to U, and since any open interval in U is sent to an open interval in $f(U)$ by f, it is pulled back into one by f^{-1} so f^{-1} is continuous.

Now I show that f^{-1} is differentiable at $b = f(a)$. Put $\varepsilon = f(a+h) - f(a)$ for some h small enough to ensure that $a + h \in U$. Then

$$\frac{f^{-1}(b+\varepsilon) - f^{-1}(b)}{\varepsilon} = \frac{a + h\ -\ a}{f(a+h) - f(a)} = \frac{h}{f(a+h) - f(a)}$$

This is the reciprocal of

$$\frac{f(a+h) - f(a)}{h}$$

The limit of the latter as $h \to 0$ is given to exist and is $f'(a) > 0$, and since the reciprocal of a function is continuous when the function is continuous and not zero, the limit of

$$\frac{f^{-1}(b+\varepsilon) - f^{-1}(b)}{\varepsilon}$$

as $\varepsilon \to 0$ must also exist, is the required derivative of f^{-1} at $f(a)$, and is $1/f'(a)$.

This shows f^{-1} is continuous in an interval and differentiable at a point in it. But the argument does not depend on the point a and works for any point in $f(U)$. Moreover, since the derivative of the inverse is the reciprocal of the derivative, and since this is a continuous operation when it can be done at all, then f^{-1} is *continuously* differentiable in $f(U)$. □

Remark 4.2.1. Recall that a *diffeomorphism* is a map that is differentiable and has an inverse which is also differentiable. It is an isomorphism in the category of open subsets of $\mathbb{R}^n$ for various positive integers n and smooth maps. We can summarise the above theorem by saying that any smooth map from $\mathbb{R}$ to $\mathbb{R}$ is a diffeomorphism in some neighbourhood of any point for which the derivative is non-zero. Drawing the graphs of some smooth maps suggests that except in the case where a map is constant over an interval, the set of points where the derivative is zero (critical points of the map) is not, as a rule, a very big set. I shall return to this point later.

The generalisation to higher dimensions is best motivated, I hope, by a simple example. Look at the map $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by

$$\begin{bmatrix} x \\ y \end{bmatrix} \rightsquigarrow \begin{bmatrix} x^2 \\ y^2 \end{bmatrix}$$

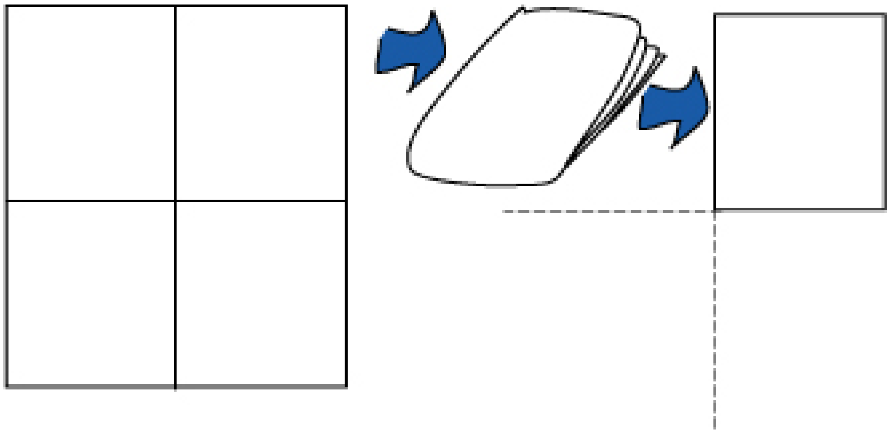

Figure 4.3: Picture of a map from $\mathbb{R}^2$ to $\mathbb{R}^2$

You can think of this as folding over both the x and y axes to give a (mostly) fourfold covering of the positive quadrant. It stops being fourfold on the axes where it is two fold except at the origin where it is 1-1. Note that the derivative of f is

$$\begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix}$$

which is non-degenerate (non-singular, invertible) except along the axes. This is not a coincidence. My picture, figure 4.3, may give a geometric picture of f.

Note that a very little open ball U on the point $(1,1)^T$ gets stretched to approximately an open ball of twice the radius on $(1,1)^T$ by f. It won't be quite a ball because away from $(1,1)^T$ it gets stretched by more than two in the x and y directions, for points further from the origin, and by less than two for points closer to the origin. All the same, it gets sent to some open set in $\mathbb{R}^2$ and provided the ball U does not cut either axis, the image $f(U)$is an open set and $f|U$ is clearly 1-1 and onto, so has an inverse for every point in $f(U)$. In fact if f is continuously differentiable then the inverse is differentiable and indeed continuously so, and the derivative of the inverse is the inverse of the derivative. (That last clause sounds cute but it needs to be investigated for meaning!)

The proof of the n-dimensional case of the inverse function theorem is a bit more complicated than the one dimensional case and needs a preliminary lemma:

Lemma 4.2.1. If U is an open ball in $\mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^n$ is continuously differentiable and for every $\boldsymbol{x} \in U$ the j^{th} partial derivative of the i^{th}

coordinate map evaluated at $\boldsymbol{x}$,

$$D_j(f^i)(\boldsymbol{x}) \triangleq \frac{\partial f^i}{\partial x_j}(x)$$

has $|D_j(f^i)$ bounded above by M for all $i, j, \boldsymbol{x}$, then

$$\forall\ \boldsymbol{x}, \boldsymbol{y} \in U, \quad \|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \leq n^2 M \|\boldsymbol{x} - \boldsymbol{y}\|$$

Proof:

This is copied almost verbatim from the proof in Spivak's *Calculus on Manifolds*, p35.

We have

$$f^i(\boldsymbol{y}) - f^i(\boldsymbol{x}) = \sum_{j=1,n} [f^i(y^1, y^2, \cdots y^j, x^{j+1}, \cdots, x^n) - f^i(y^1, y^2, \cdots, y^{j-1}, x^j, \cdots, x^n)]$$

Applying the mean value theorem we obtain

$$f^i(y^1, \cdots, y^i, x^{j+1}, \cdots x^n) - f^i(y^1, \cdots, y^{j-1}, x^j, \cdots, x^n) = (y^j - x^j) D_j f^i(z_{ij})$$

for some z_{ij} on the line segment joining the points

$$(y^1, \cdots, y^j, x^{j+1}, \cdots x^n) \text{ and } (y^1, \cdots, y^{j-1}, x^j, \cdots, x^n)$$

The expression $(y^j - x^j) D_j f^i(z_{ij})$ has absolute value less than or equal to $M|y^j - x^j|$. Thus

$$|f^i(\boldsymbol{y}) - f^i(\boldsymbol{x})| \leq \sum_{j=1,n} |y^j - x^j| M \leq nM\|\boldsymbol{y} - \boldsymbol{x}\|$$

since $\forall\ j \in [1..n]\ |y^j - x^j| \leq \|\boldsymbol{y} - \boldsymbol{x}\|$ Putting these together for all the n component functions f^i we get

$$\|f(\boldsymbol{y}) - f(\boldsymbol{x})\| \leq \sum_{i=1,n} |f^i(y) - f^i(x)| \leq n^2 M \|\boldsymbol{y} - \boldsymbol{x}\|$$

□

Remark 4.2.2. The condition that there exist some K such that

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq K|\boldsymbol{x} - \boldsymbol{y}|$$

is known as a *Lipschitz* condition on f and if f is continuously differentiable in some open ball, then its partial derivatives are certainly bounded on some (possibly smaller) open ball, and hence we conclude that if f is continuously differentiable (also known as 'of class $\mathcal{C}^1$') on an open set then there is an open ball on which it satisfies a Lipschitz condition. Again this is merely jargon, but if you have seen it before it is not nearly so terrifying when you meet it again.

Exercise 4.2.1. Try to prove the lemma on your own. It is intuitively rather natural if you think of the partial derivatives as giving local length stretching factors.

Theorem 4.2.1 (Inverse Function Theorem). If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a continuously differentiable map and $\exists\ \boldsymbol{a} \in \mathbb{R}^n, \quad Df(\boldsymbol{a})$ is invertible, then there is an open ball U on $\boldsymbol{a}$ having $f(U)$ open in $\mathbb{R}^n$ such that f^{-1} exists on $f(U)$ and $Df^{-1}(f(\boldsymbol{x}))$ is the inverse of $Df(\boldsymbol{x})$ for all $\boldsymbol{x} \in U$.

We start off by noting the condition of invertibility of the matrix of partial derivatives $Df(\boldsymbol{x})$ is that its determinant should be non-zero, and since the determinant is a polynomial function, which by now we know well, it is certainly continuous. Hence the composite $\mathrm{Det}(Df(\boldsymbol{x}))$ is a continuous map from $\mathbb{R}^n$ to $\mathbb{R}$, and there is therefore some open set V containing $\boldsymbol{a}$ in which $\mathrm{Det}(Df(\boldsymbol{x}))$ does not change sign. Without loss of generality, suppose then that $\mathrm{Det}(Df(\boldsymbol{x})) > 0, \quad \forall x \in V$.

$Df(\boldsymbol{a})$ is defined to be the (unique) linear map such that

$$\lim_{\|\boldsymbol{h}\|\to 0} \frac{\|f(\boldsymbol{a}+\boldsymbol{h}) - (Df(\boldsymbol{a})(\boldsymbol{h}) + f(\boldsymbol{a}))\|}{\|\boldsymbol{h}\|} = 0$$

Recall that this makes the affine map $Df(\boldsymbol{a})(\boldsymbol{x}) + f(\boldsymbol{a})$ *tangent* to f at $\boldsymbol{a}$.

Without loss of generality, we may take $Df(\boldsymbol{a})$ to be the identity map, since we can otherwise consider instead the map f followed by the inverse of $Df(\boldsymbol{a})$. This, by the chain rule will certainly have derivative at $\boldsymbol{a}$ the identity. But if we can prove it for the new map. then it must be true for the original one as well.

The idea behind the proof of the theorem is that if the derivative $Df(\boldsymbol{a})$ is the identity, then within some neighbourhood of $\boldsymbol{a}$ the derivative (which is continuous by assumption) will not differ much from the identity, and hence will not compress or stretch distances between points very much. The point of the lemma is that since this holds for differential stretching of lengths in all directions, it holds for all pairs of points. So open balls will neither be compressed nor stretched (under the mapping f) by a number very different from 1. We shall phrase this idea more formally below and use it throughout the proof.

Since the derivative $Df(\boldsymbol{a})$ is continuous we may also take it that we can find some open set V' containing $\boldsymbol{a}$ such that for every $\boldsymbol{x} \in V'$,

$$\left\| \frac{\partial f^i}{\partial x^j}(\boldsymbol{x}) - \frac{\partial f^i}{\partial x^j}(\boldsymbol{a}) \right\| < \frac{1}{2n^2} \quad \forall i, j \in [1..n]$$

This is the claim that the matrix which is the derivative of f at $\boldsymbol{x}$ is not too different from the identity matrix in a neighbourhood of $\boldsymbol{a}$. It is an immediate consequence of the derivative being continuous, since all its entries have to be continuous.

The map $g(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{x}$ has a matrix of partial derivatives of the form $Df(\boldsymbol{x}) - I_n$ where I_n is the identity matrix. At $\boldsymbol{a}$, this gives the zero matrix. Since Dg is also continuously differentiable, we can take another open set V'' such that

$$\|Dg(\boldsymbol{x})\| < \frac{1}{2n^2}, \quad \forall\ \boldsymbol{x} \in V''$$

Taking the intersection of the three open sets $V \cap V' \cap V''$ gives us one which I might just as well call V. Then for every $\boldsymbol{x}, \boldsymbol{y} \in V$ the following three conditions hold:

$$\mathrm{Det} Df(\boldsymbol{x}) > 0 \tag{4.1}$$

$$\forall\ i, j \in [1..n], \quad \left| \frac{\partial f^i}{\partial x^j}(\boldsymbol{x}) - \frac{\partial f^i}{\partial x^j}(\boldsymbol{a}) \right| < \frac{1}{2n^2} \tag{4.2}$$

$$\|f(\boldsymbol{x}) - \boldsymbol{x} - (f(\boldsymbol{y}) - \boldsymbol{y})\| < \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\| \tag{4.3}$$

The last follows by applying the lemma 4.2.1 to g. By some elementary manipulations we obtain from 4.3 that

$$\forall\ \boldsymbol{x}, \boldsymbol{y} \in V, \quad \|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \geq \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\| \tag{4.4}$$

$$\forall\ \boldsymbol{x}, \boldsymbol{y} \in V, \quad \|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \leq \frac{3}{2}\|\boldsymbol{x} - \boldsymbol{y}\| \tag{4.5}$$

These last inequalities make sure that the amount of stretching of the distance between pairs of points that goes on in V as a result of f cannot be either very great or very small, it is going to be reasonably close to 1 always; certainly pairs of points will not be 'stretched' by more than 3/2 nor compressed by less than 1/2. This idea is the key to the rest of the proof, as advertised earlier.

Inequality 4.4 clearly forces f to be 1-1 in V and hence f^{-1} is defined on $f(V)$.

Inequality 4.5 and 4.4 between them ensure that if I take an open ball of radius δ on any point $\boldsymbol{x}$ in V, its image contains the open ball of radius $\delta/2$ on $f(\boldsymbol{x})$ and is in turn contained in the open ball on $f(\boldsymbol{x})$ of radius $3\delta/2$. We use this to show that f^{-1} is continuous:

I claim that the image by f of any open subset of V is open in $\mathbb{R}^n$. First I observe that the inequalities 4.5 and 4.4 can be run backwards to give:

$$\forall \;\; \boldsymbol{u}, \boldsymbol{v} \in f(V), \quad \frac{2}{3}\|\boldsymbol{u}-\boldsymbol{v}\| \leq \;\; \|f^{-1}(\boldsymbol{u}) - f^{-1}(\boldsymbol{v})\| \;\; \leq 2\|\boldsymbol{u}-\boldsymbol{v}\| \quad (4.6)$$

This can be stated informally as 'If f never stretches pairs of points by more than one and a half, nor compresses them to less than one half, then f^{-1} cannot stretch points by more than two, nor compress them by less than two thirds'. If you doubt this, prove it carefully.

If $B \subseteq V$ is an open ball on $\boldsymbol{x}$ and $\boldsymbol{z} \in f(B)$, then I shall have established my claim if I can find an open ball on $\boldsymbol{z}$ which is wholly in $f(B)$. But the open ball on $\boldsymbol{z}$ of radius $\delta/2$ is contained in $f(B)$ whenever the open ball on $f^{-1}(\boldsymbol{z})$ of radius δ is contained in B, by one of the last inequalities, and clearly there is some such δ since B is open.

This shows that f^{-1} is continuous on $f(V)$. It remains to show it is differentiable.

First I prove the derivative of f^{-1} exists at $\boldsymbol{b} = f(\boldsymbol{a})$. This is just to make the idea very stark, because in this case f' is the identity matrix:

I know that

$$\lim_{\|\boldsymbol{h}\|\to 0} \frac{\|f(\boldsymbol{a}+\boldsymbol{h}) - (\boldsymbol{h} + f(\boldsymbol{a}))\|}{\|\boldsymbol{h}\|} = 0$$

$$\Rightarrow \lim_{\|\boldsymbol{h}\|\to 0} \frac{2(\|f(\boldsymbol{a}+\boldsymbol{h}) - (\boldsymbol{h}+\boldsymbol{b})\|)}{\|\boldsymbol{h}\|} = 0$$

$$\Rightarrow \lim_{\|\boldsymbol{h}\|\to 0} \frac{\|f^{-1}(\boldsymbol{b}+\boldsymbol{h}) - (f^{-1}(\boldsymbol{b}) + \boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$$

by inequality 4.6, which shows that f^{-1} has derivative the identity matrix.

If now $Df(\boldsymbol{x}) = T$, an $n \times n$ matrix which is invertible, we have from the definition of the derivative that

$$\lim_{\|\boldsymbol{h}\| \to 0} \frac{\|f(\boldsymbol{x}+\boldsymbol{h}) - (T(\boldsymbol{h}) + f(\boldsymbol{x}))\|}{\|\boldsymbol{h}\|} = 0$$

$$\Rightarrow \lim_{\|\boldsymbol{h}\| \to 0} \frac{2\|f(\boldsymbol{x}+\boldsymbol{h}) - (T(\boldsymbol{h}) + f(\boldsymbol{x}))\|}{\|\boldsymbol{h}\|} = 0$$

(since just multiplying by 2 preserves limits and $2 \times 0 = 0$)

$$\Rightarrow \lim_{\|\boldsymbol{h}\| \to 0} \frac{\|f^{-1}(\boldsymbol{y} + T(\boldsymbol{h})) - (\boldsymbol{x}+\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$$

where I have put $\boldsymbol{y} = f(\boldsymbol{x})$ and am using the inequalities 4.6. Putting $\boldsymbol{k} = T(\boldsymbol{h})$ we get

$$\lim_{\|\boldsymbol{k}\| \to 0} \frac{\|f^{-1}(\boldsymbol{y}+\boldsymbol{k}) - (f^{-1}(\boldsymbol{y}) + T^{-1}(\boldsymbol{k}))\|}{\|\boldsymbol{k}\|} = 0$$

since $\|\boldsymbol{h}\| \to 0 \Rightarrow \|\boldsymbol{k}\| \to 0$, since $\|T^{-1}(\boldsymbol{h})\| \leq 2\|\boldsymbol{h}\|$.

This establishes that $Df^{-1}(f(\boldsymbol{x})) = (Df(x))^{-1}$ throughout V. Now let the required open ball U on $\boldsymbol{a}$ be any one that is inside V.

Since the derivative of f^{-1} is the inverse of the derivative of f which is continuous, and since matrix inversion is a continuous operation when it can be done at all, it follows that the derivative of f^{-1} is also continuous on $f(U)$. □

Remark 4.2.3. This can conveniently be framed in the terms:
If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is continuously differentiable in a neighbourhood of a point $\boldsymbol{a}$, and if the derivative at $\boldsymbol{a}$ is non-degenerate, then there is a neighbourhood of $\boldsymbol{a}$ on which f is a diffeomorphism.

Exercise 4.2.2. Let $f : (X,d) \longrightarrow (Y,e)$ be a map between metric spaces which satisfies the conditions:

$$\forall\, \boldsymbol{x}, \boldsymbol{y} \in X,\ \frac{1}{2}d(\boldsymbol{x},\boldsymbol{y}) \leq e(f(\boldsymbol{x}), f(\boldsymbol{y})) \leq \frac{3}{2}d(\boldsymbol{x},\boldsymbol{y})$$

(1) Show that f is continuous. (2) Show that f is 1-1 (3) Show that f is open and (4) Deduce that f is a homeomorphism between X and the image of f.

Remark 4.2.4. This is a classical theorem of some note, and I have presented a rather verbose argument so as to encourage you to see the principal ideas-that at a point where the derivative is non-degenerate, we can assume without loss of generality that it is represented by the identity matrix, that, in a neighbourhood of this point, continuity of the derivative ensures that the derivatives (represented by their matrices) will never be too far from the identity, certainly they won't be degenerate, and that this ensures that we can put upper and lower bounds on the extent to which pairs of points are expanded or compressed. The rest is working from the definitions of f being 1-1, its inverse being continuous, and its inverse being differentiable.

The above argument is far from the sort of proof that Mathematicians love, they prefer conciseness to intelligibility, and if you have to work in order to follow the argument, why, that's fine. You should certainly be able to provide a proof. Do not try to memorise it, remember the sequence of ideas and make sure your argument holds water. Now that you have had the ideas explained to you, you might like to prove the result in the fewest number of lines and leaving out all the English.

You can then give it to a second year student as an example of a proof you have just been doing in third year. As a way of driving them to suicide or doing computer science[2], this is hard to beat.

4.3 The Implicit Function Theorem

You will recall[3] the rank-nullity theorem from Linear Algebra. It says that if $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear map and has rank r, and if the kernel has dimension k, then $k + r = n$. You probably used this in the form: *Given r independent linear equations in n variables, there is an $(n - r)$ dimensional subspace of solutions to the set of equations.* I hope that (a) you recall the theorem and (b) you made the connection with sets of equations.

You may have noticed when constructing implicit representations of curves and surfaces that this seems to generalise: If we have r independent smooth conditions on n variables, we get an $n-r$ dimensional space of solutions that is a manifold in some obvious cases..

Example 4.3.1. Let $n = 2$ and let the (non-linear) constraint be $x^2+y^2-1 = 0$. Then we get a one-dimensional space of solutions, otherwise known as a circle. We know this is a manifold from the introduction, though not why.

[2] Much the same.
[3] Or else.

Example 4.3.2. Let $n = 3$ and take the one condition $x^2 + y^2 + z^2 - 1 = 0$ to get a 2-dimensional space, in fact a 2-sphere. We know this is a manifold too. Put on the extra condition $x + y + z = 0$ and we get a plane cutting the sphere to give a 1-dimensional circle.

Example 4.3.3. Take $n = 3$ again and also $x^2 + y^2 + z^2 - 1 = 0$ and the new condition $(x - 3)^2 + (y - 5)^2 + (z - 4)^2 - 1 = 0$ and we get the empty set because the two spheres are too far apart to intersect. But if they *did* intersect, they'd be sure to do so in a circle. Unless they just touched in a point.

Oh well, two out of three ain't bad.

There seems to be something going on here that there ought to be a theorem about. And it seems not too far fetched to believe that quite a lot of the time, a single constraint ought to cut down the dimension by one, and that we should be able to show this at a point by approximating the function by its derivative and using the rank-nullity theorem. It looks plausible that in the vicinity of a point the graph of the derivative when restricted to some neighbourhood is diffeomorphic to the graph of the function, so if it works for affine maps it should work for smooth maps too.

The vague feeling that there ought to be something in this, is the stuff of conjectures. And in time, some conjectures turn into theorems.[4]

There is indeed something in this and the most basic thing in it is the *Implicit Function Theorem* which again I shall prove in rather verbose terms, extracting the key ideas in order to show you how people (a) remember and (b) invent proofs.

First I do it in the simplest case. Think $f(x, y) = x^2 + y^2 - 1$. I extend the definition of the kernel of a map from linear maps:

Definition 4.3.1. Given a map $f : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$,

$$\ker(f) \triangleq \{\boldsymbol{x} \in \mathbb{R}^2 : f(\boldsymbol{x}) = 0\}$$

Proposition 4.3.1 (Baby Implicit Function Theorem). *If the function $f : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ is continuously differentiable at (a, b) and if $\partial f/\partial y \neq 0$ there, then there is a neighbourhood U of a in $\mathbb{R}$ and a 1-1 map $g : U \longrightarrow \mathbb{R}$ which is a continuously differentiable on U and such that $f(x, g(x)) = 0$. This makes the graph of g the same set as ker$(f) \cap (U \times \mathbb{R})$.*

[4] And some just die, killed by a nasty counter-example.

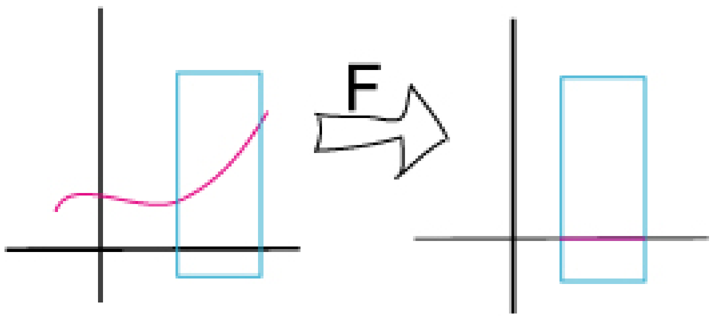

Figure 4.4: Straightening out the curve

Proof:

The idea here is to extend f to something we can use the inverse function theorem on. The figure 4.4 shows how we can use F to 'straighten out' the curve and send it to the x-axis in a neighbourhood of any point where $\partial f/\partial y \neq 0$.

Define

$$\begin{array}{rcl} F : \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{R} \times \mathbb{R} \\ (x, y) & \rightsquigarrow & (x, f(x, y)) \end{array}$$

Note that we have $f = e_2^\star \circ F$; in English, doing F and then projecting down onto the second component gives us f.

Note also that since f is differentiable, so is F with derivative:

$$\begin{bmatrix} 1 & \frac{\partial f}{\partial x} \\ 0 & \frac{\partial f}{\partial y} \end{bmatrix}$$

and that since $\partial f/\partial y \neq 0$ by assumption, the determinant of F is non zero.

Then the inverse function theorem tells us that at $F(a, b)$ there is a neighbourhood of (a, b) say $U \times V$, and a differentiable inverse function F^{-1}. Moreover this inverse is continuously differentiable on $F(U \times V)$ and gives a diffeomorphism between $U \times V$ and $F(U \times V)$.

It is obvious that we can write

$$F^{-1}(u, v) = (u, h(u, v))$$

for some differentiable map h, with $u \in U$ and $v \in V$.

Now define $g(x) = h(x, 0)$. Clearly g is also a differentiable map since it is the composite of (differentiable) h with the (differentiable) inclusion map which sends x to $(x, 0) \in \mathbb{R} \times \mathbb{R}$. And for the same reasons it is continuously differentiable.

We need to verify that $f(x, g(x)) = 0$:

$$f(x, g(x)) = f(x, h(x, 0)) = f \circ F^{-1}(x, 0) = e_2^\star \circ F \circ F^{-1}(x, 0) = e_2^\star(x, 0) = 0$$

Finally, I note that $g(x)$ is the only point in V satisfying $f(x, g(x)) = 0$, since F is 1-1 in $U \times V$ and $F^{-1}(x, 0)$ can only give $(x, g(x))$. So g is 1-1. □

Corollary 4.1. *The map $x \rightsquigarrow (x, g(x))$ is a diffeomorphism from U to the set $V \cap S$, where $S = \{(x, y) \in \mathbb{R} \times \mathbb{R} : f(x, y) = 0\}$.*

Proof:

If g is differentiable on an open set U, then the map $x \rightsquigarrow (x, g(x))$ is certainly a diffeomorphism from U to the graph of g in $\mathbb{R} \times \mathbb{R}$. But we have just shown that the graph of g is $V \cap S$. □

Remark 4.3.1. Anyone doubting that $x \rightsquigarrow (x, g(x))$, taking an interval to it's graph is a diffeomorphism is invited to prove it carefully.

Remark 4.3.2. Note that we can interpret this as saying that if we are given a curve implicitly as $\ker(f)$, we can find a parametric representation of part of the curve where the partial derivative $\partial f/\partial y \neq 0$. We can also find a parametric representation where $\partial f/\partial x \neq 0$ by swapping x and y. So we can represent a curve parametrically over most of it. You may find it amusing to investigate when we can guarantee to parametrise the whole curve.

Remark 4.3.3. This tells us that, locally, the set $\ker(f) = \{(x, y) \in \mathbb{R} \times \mathbb{R} : f(x, y) = 0\}$ is a curve, given a checkable condition on f. Well, we knew that. Didn't we?

Exercise 4.3.1. Find an f for which $\ker(f)$ isn't a curve.

Now we do the full theorem. The generalisation is not very different in this case. We note that for a map $f : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$, the matrix of partial derivatives has m rows and $n+m$ columns. The last m columns gives a square matrix of partial derivatives which I shall call the $m \times m$ sub-derivative of f

Theorem 4.3.1 (Implicit Function Theorem). If $f : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is continuously differentiable and at $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^n \times \mathbb{R}^m$ the $m \times m$ subderivative matrix is invertible and $f(\boldsymbol{a}, \boldsymbol{b}) = \mathbf{0}$, then there is an open ball U on $\boldsymbol{a}$ in

$\mathbb{R}^n$, an open ball V on b in $\mathbb{R}^m$, and a differentiable map $g : U \longrightarrow \mathbb{R}^m$ with $g(\boldsymbol{a}) = \boldsymbol{b}$, such that with

$$S \triangleq \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^m : f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0} \in \mathbb{R}^m\}$$

we have

$$\forall\ \boldsymbol{x} \in U, \quad (\boldsymbol{x}, g(\boldsymbol{x})) \in S$$

and the element $g(\boldsymbol{x}), \boldsymbol{x} \in U$ is the *only* point in V such that $(\boldsymbol{x}, g(\boldsymbol{x})) \in S$

Proof:

Define

$$\begin{array}{rcl} F : \mathbb{R}^n \times \mathbb{R}^m & \longrightarrow & \mathbb{R}^n \times \mathbb{R}^m \\ (\boldsymbol{x}, \boldsymbol{y}) & \rightsquigarrow & (\boldsymbol{x}, f(\boldsymbol{x}, \boldsymbol{y})) \end{array}$$

Then we have

$$f = \pi_2 \circ F$$

where

$$\begin{array}{rcl} \pi_2 : \mathbb{R}^n \times \mathbb{R}^m & \longrightarrow & \mathbb{R}^m \\ (\boldsymbol{x}, \boldsymbol{y}) & \rightsquigarrow & \boldsymbol{y} \end{array}$$

and the derivative of F is a square $(n+m) \times (n+m)$ matrix with the identity in the top left $n \times n$ rows and columns, zeros in the top right m columns and n rows, and the derivative of f in the lower m rows. The determinant of $Df(\boldsymbol{a}, \boldsymbol{b})$ is the determinant of the $m \times m$ subderivative, which we are told is non-zero, and hence by the inverse function theorem, there is an open ball W on $(\boldsymbol{a}, \boldsymbol{b})$ and the map F^{-1} exists and is differentiable on $F(W)$. We can take without loss of generality, W to be an open set of the form $U \times V$ where U and V are open balls on $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively. The map F^{-1} can be written

$$F^{-1}(\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{u}, h(\boldsymbol{u}, \boldsymbol{v}))$$

for some differentiable map $h : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$.

Now define

$$\begin{array}{rcl} g : \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \\ \boldsymbol{x} & \rightsquigarrow & h(\boldsymbol{x}, \boldsymbol{0}) \end{array}$$

Then $\forall\ \boldsymbol{x} \in U$:

$$f(\boldsymbol{x}, g(\boldsymbol{x})) = f(\boldsymbol{x}, h(\boldsymbol{x}, \boldsymbol{0})) = f \circ F^{-1}(\boldsymbol{x}, \boldsymbol{0}) = \pi_2 \circ F \circ F^{-1}(\boldsymbol{x}, \boldsymbol{0}) = \boldsymbol{0}$$

which establishes that there is indeed a differentiable map from U to S. Again there is a unique element of $S = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^m : f(\boldsymbol{x}, \boldsymbol{y}) = 0\}$ corresponding to $\boldsymbol{x} \in U$, □

Remark 4.3.4. Just as for the Baby case, there is a diffeomorphism from U to the intersection $U \times V$ with the graph of g, and hence to the intersection of S with $U \times V$, which establishes that locally S is diffeomorphic to an open ball of $\mathbb{R}^n$.

Remark 4.3.5. In the case of the baby theorem we can easily see that the function g, although it is differentiable, might be a nasty one to write down. Consider, for example, $f(x, y) = x^2 + y^3 + \sin(xy)$. Trying to get y as a function of x is obviously not going to be very easy in this case. We have, however, an easy rule for obtaining $g'(x)$. We note however that

$$DF(x, y) = F'(x, y) = \begin{bmatrix} 1 & \frac{\partial f}{\partial x} \\ 0 & \frac{\partial f}{\partial y} \end{bmatrix}$$

and so $DF^{-1} =$

$$\begin{bmatrix} 1 & -\frac{\partial f}{\partial x} / \frac{\partial f}{\partial y} \\ 0 & 1/\frac{\partial f}{\partial y} \end{bmatrix}$$

by matrix inversion, and the top right hand entry is $\partial h/\partial x$ which is the same as g' In other words we have shown that:

$$\frac{dg}{dx} = \frac{-\partial f/\partial x}{\partial f/\partial y}$$

Exercise 4.3.2. Show that the result generalises to the (big beefy) Implicit Function Theorem:

$$Dg = (D_2 f(\boldsymbol{x}, g(\boldsymbol{x})))^{-1} \circ D_1 f(\boldsymbol{x}, g(\boldsymbol{x}))$$

where I leave it to your ingenuity to find suitable meanings for $D_2 f$ and $D_1 f$.

We can rewrite the Implicit Function Theorem without the clumsy term *subderivative*. For $m < n$ let $\pi_{(n-m)} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be the projection

$$(x^1, x^2, \cdots, x^n) \rightsquigarrow (x^{n-m+1}, \cdots x^n)$$

that is we project on the last m components of each vector. Then:

Proposition 4.3.2. *If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a continuously differentiable map with $n > m$, and if there is a $\mathbf{c} \in \mathbb{R}^n$ with $f(\boldsymbol{c}) = 0$ and $Df(\boldsymbol{a})$ has rank m, then there is an open ball U on $\mathbf{c}$ in $\mathbb{R}^n$ and a subset S of U such that S is diffeomorphic to an open ball in $\mathbb{R}^{(n-m)}$ and is the intersection with U of*

$$\mathrm{Ker(f)} \triangleq \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) = \mathbf{0}\}$$

Proof:

We can arrange that $\mathbb{R}^n$ is decomposed into $\mathbb{R}^{n-m} \times \mathbb{R}^m$ by taking the first $(n-m)$ items in any vector in $\mathbb{R}^n$. This splits $\boldsymbol{c}$ into $(\boldsymbol{a}, \boldsymbol{b})$ and makes it look like the Inverse Function theorem. Now the fact that $Df(\boldsymbol{c})$ has rank m means that if we look at the matrix of partial derivatives, we can find m independent columns. We now permute these columns so that they are all at the right hand side of the matrix, giving us the required condition for the derivative. The permutation of columns is, of course, a non-singular linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$, call it A. Then $f \circ A$ satisfies all the conditions for the Implicit Function Theorem, so there is a neighbourhood $U \times V$ in $\mathbb{R}^{(n-m)} \times \mathbb{R}^m$ of $A(\boldsymbol{c})$ and a map $g : U \longrightarrow \mathbb{R}^m$ the graph of which is the set $\text{Ker}(f \circ A) \cap V$, where $\text{Ker}(f \circ A) = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{(n-m)} \times \mathbb{R}^m : f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}\}$ □

Remark 4.3.6. In particular, take $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $f(x, y) = x^2 + y^2 - 1$. Then $D(f) = [2x, 2y]$ This has rank one except where $x = y = 0$ which is not in the set $\ker(f)$. Hence we deduce that the set $\ker(f)$ is locally one dimensional and sufficiently small neighbourhoods of any point look like intervals of $\mathbb{R}$. In other words it is a 1-manifold.

4.4 Sard's Theorem

If you reflect on the assumptions about f that go into the Implicit Function Theorem in the baby case, you will note that we need $\partial f/\partial y$ to be non-zero in order to ensure that the tangent to the curve given implicitly doesn't go vertical. When you draw the set $S^1 = \{(x, y)^T \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ you observe that there are only two points where things go blooey, namely $x = \pm 1, y = 0$. Likewise there are two points where $\partial f/\partial x$ is zero, at the top and bottom of the circle. It would be nice if the places where things go blooey were always a nice small set like this. For a start, it would mean that finding the maximum of a differentiable function, which involves checking out the critical points one at a time, would be something we could hope to do before the sun goes cold.

Some definitions to make talking about these things easier:

Definition 4.4.1. A *critical point* of a map $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a point $\boldsymbol{a} \in \mathbb{R}^n$ where $df(\boldsymbol{a})$ has rank $< m$ or is not defined, and a *regular point* of f is a point $\boldsymbol{b} \in \mathbb{R}^n$ where the rank of $df(b) = m$. A point in the image of f is called a *critical value* if it is the image of a critical point. Points in the image of f which are not critical values are called *regular values.*

Example 4.4.1. The only critical point of $f(x, y) = x^2 + y^2 - 1$ is at the origin since $df(x, y) = [2x, 2y]$ is a linear map (for any given point (x, y)) which is onto $\mathbb{R}$ except when $x = y = 0$. The set of critical values is the singleton set consisting of -1.

Example 4.4.2. The critical points of the function $\cos(x)$ are at $x = n\pi,\ n \in \mathbb{Z}$. The critical value set has only two things in it, ± 1.

Exercise 4.4.1. Construct a differentiable function $f : \mathbb{R} \longrightarrow \mathbb{R}$ which has an infinite number of critical values in the unit interval. Is there a smooth such function?

Exercise 4.4.2. Construct a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ for suitable n, m which has *every* point a critical point.

Generalising, it looks as though the set of critical points of smooth maps from $\mathbb{R}$ to $\mathbb{R}$ is usually on the small side, although it can be infinite as you will have worked out by doing the above exercises. In fact, as already noted, the standard method of finding the maxima and minima of functions of one or more variables relies on finding the places where the derivative goes to zero, and then inspecting each of these. If the set of these points was usually humungous, then we'd have problems using this method.

One of the awful possibilities that disturbs the dreams of potential mathematicians is that the set of critical values of a map might be so big that it contains some open subset of the image space. In fact there are such maps: you have so far been introduced only to very nicely behaved functions (except for the Dirichlet function which you met last year). You will be glad however to hear that these nasty functions are not smooth. Smooth maps are not so likely to give you nightmares.

If we have to say that the set of critical values of a map 'is not horribly big', there are several possible senses of 'not horribly big'. One sense of 'horribly big' is 'being infinite'. But infinite sets that can be put into one-one correspondence with the natural numbers are not regarded as 'horribly big' by Mathematicians, who dispose of several such sets before breakfast as a matter of daily routine.

Another, worse, possibility is that the set can be put into one-one correspondence with the set of all real numbers, But almost any map from $\mathbb{R}$ to $\mathbb{R}^2$ will have rather a lot of critical points and also critical values. The fact remains that the uncountable infinite set of critical values looks like a curve in $\mathbb{R}^2$, which is a rather thin sort of set. We are inclined therefore to be dismissive of sets which have area zero in $\mathbb{R}^2$, volume zero in $\mathbb{R}^3$, or *Lebesgue Measure* zero in $\mathbb{R}^n$.

Definition 4.4.2. The unit n-cube in $\mathbb{R}^n$ is the set

$$I^n = \left\{ \boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n : \forall\, j \in [1..n], 0 \leq x_j \leq 1 \right\}$$

It has (n)-Lebesgue measure 1, by definition. Moving it by adding any fixed vector to each of the points in the n-cube doesn't change its measure, in fact the measure of any set (that has one) is unchanged by such shifts. Rotating a set doesn't change the measure either. Scaling it along any axis by a factor a multiplies the measure by $|a|$. The result of doing this to the unit cube with any choice of axes is called an *n-cuboid*, or *cuboid* for short. In fact it is still a cuboid if we multiply by different numbers along all the axes. A shifted or rotated cuboid is also a cuboid. If every side of a cuboid is scaled by the same factor a, the measure of the cuboid is multiplied by a^n.

Exercise 4.4.3. Go through the above in dimensions two and three making the appropriate translations– 'cuboid' becomes 'rectangle', 'measure' becomes 'area', *et cetera*– and confirm that the statements are all true.

Definition 4.4.3. A subset $A \subset \mathbb{R}^n$ has *Lebesgue Measure Zero* iff for every $\varepsilon \in \mathbb{R}^+$ it is possible to cover A by n-cuboids such that the sum of the measures of the cuboids is less than ε.

Exercise 4.4.4. Prove that the area of the unit circle is zero. (Anyone who thinks it is π will be expected to prove that $\pi = 0$.)

An important rule for calculating Lebesgue measure of sets that are not cubes is that if $(U_1, U_2, \cdots U_n, \cdots)$ is a countable collection of disjoint sets all of which have measure, then the measure of the union exists and is the limit of the partial sums

$$\sum_{j=1,\infty} \mu(U_j)$$

where $\mu(A)$ is the Lebesgue measure of the set A and where $+\infty$ is an allowable value for a measure. It should be obvious that the infinite sum always exists (although it may be $+\infty$). It should be at the very least plausible that the n-Lebesgue measure of a set $A \times B$ with $A \subseteq \mathbb{R}^k$ and $B \subseteq \mathbb{R}^{(n-k)}$ is the k-lebesgue measure of A multiplied by the $(n-k)$-Lebesgue measure of B. It obviously works for cuboids, and by the above rule about countably infinite decompositions, it must work for any set which can be expressed as a countably infinite union of cuboids.

Exercise 4.4.5. Prove that the unit ball on the origin in $\mathbb{R}^n$ can be expressed as a countable union of cuboids. Use this to get a recurrence relation between the measure of an n ball in terms of an $n-1$ ball. Hence calculate the Lebesgue measure of the unit ball in $\mathbb{R}^n$ for $1 \leq n \leq 7$, and deduce the power of π which occurs in the measure of the unit ball in $\mathbb{R}^n, n \geq 1$. For what (integral) value of the dimension is the measure of the unit ball a maximum?

The result known as Sard's Theorem has a number of variant forms. I shall give the simplest. In English, it says that for continuously differentiable maps from $\mathbb{R}^n$ to $\mathbb{R}^n$, the set of critical values has measure zero. (Not the measure of the critical points, which may be the whole space!)

Theorem 4.4.1 (Sard's Theorem). If $U \subseteq \mathbb{R}^n$ is open in $\mathbb{R}^n$ and the map $g : U \longrightarrow \mathbb{R}^n$ is continuously differentiable, and $B \subseteq U$ is the set

$$\{\boldsymbol{x} \in U : \text{ Det } (g') = 0\}$$

then g(B) has measure zero.

Proof:

We first prove the result for the set of critical points in a closed cube A of side ℓ in U. If we take some positive integer N we can divide each side into ℓ/N intervals and hence divide the cube A into N^n subcubes. Suppose such a subcube, C, contains a point $\boldsymbol{x}$ such that $\text{Det}(Df(\boldsymbol{x})) = 0$. Then the image of the entire subcube C by $Df(\boldsymbol{x}))$ is in some affine space of dimension less than n. Without loss of generality, take this to be a hyperplane V of dimension $(n-1)$.

From the definition of the derivative, definition 4.1.3, we can take N big enough to ensure that, for any $\varepsilon \in \mathbb{R}^+$, for any $\boldsymbol{y} \in C$,

$$\|Dg(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{y}) - (g(\boldsymbol{x}) - g(\boldsymbol{y}))\| < \varepsilon\|\boldsymbol{x}-\boldsymbol{y}\|$$

It follows that the set $\{g(\boldsymbol{y}) : \boldsymbol{y} \in C\}$ lies within $\varepsilon\sqrt{n}(\ell/N)$ of the hyperplane $V + g(\boldsymbol{x})$, since the diameter of the cube of side ℓ/N is just $\sqrt{n}\ell/N$. I have drawn a picture of this when $n = 3$. The orange blob is the set $f(C)$, the hyperplane V runs down the middle of the thing shaped like a book. The hemispherical object I am just coming to.

We also have that there is, by lemma 4.2.1, some constant M such that

$$\|g(\boldsymbol{x}) - g(\boldsymbol{y})\| < M\|\boldsymbol{x}-\boldsymbol{y}\| \leq M\sqrt{n}\ell/N$$

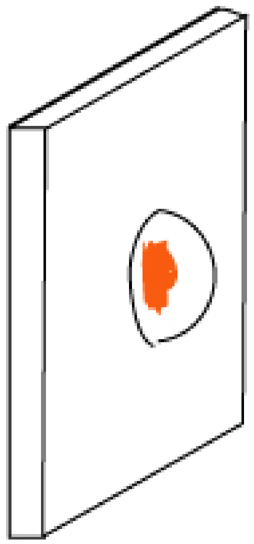

Figure 4.5: The image of the subcube C by f

Thus the image of C by g lies in both the book shaped object with V running through the middle page and also the ball of radius $M\sqrt{n}\ell/N$. This intersection in turn lies inside a short fat cylinder of length $2\varepsilon\sqrt{n}\ell/N$ and radius of the $(n-1)$-sphere $M\sqrt{n}\ell/N$ The $n-1$ measure of the sphere is some constant K (probably involving π but who cares) times the radius raised to the power $(n-1)$. So the measure of $g(C)$ is certainly less than

$$k\varepsilon(\frac{\ell}{N})^n$$

for another constant k. And since this holds for every subcube C and there are N^n of them, we conclude that the measure of the critical values of $g|A$ is less than

$$k\varepsilon\ell^n$$

Since this can be made as small as we like by choosing ε as small as we like, it follows that the measure of the critical values of $g|A = 0$.

But we can cover all of $\mathbb{R}^n$ by a countable collection of cubes like A, and thus by the countable additivity property of a measure, we conclude that $g(B)$ has Lebesgue measure zero. □

4.5 Autonomous Systems of ODEs

4.5.1 Systems of ODEs and Vector Fields

Consider the system of linear ordinary differential equations:

$$\begin{aligned} \dot{x} &= -y \qquad x(0) = 1 \\ \dot{y} &= x \qquad y(0) = 0 \end{aligned}$$

We can write this as a two dimensional problem:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

or more succinctly:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} \tag{4.7}$$

where A is the above matrix.

The matrix A defines a vector field on $\mathbb{R}^2$ by taking the location $\boldsymbol{x}$ to the vector $A(\boldsymbol{x})$. We are now used to the idea of a vector field on $\mathbb{R}^2$ both visually in terms of lots of little arrows stuck on the space as with the Mathematica examples from 2C2, and algebraically as a map from $\mathbb{R}^2$ to $\mathbb{R}^2$ sending locations to arrows (with their tails attached to those locations).

Such a system of ordinary differential equations is called *autonomous*, meaning that the vector field specified by the system doesn't change in time. Consequently we can either refer to an *Autonomous System of Ordinary Differential Equations defined on an open set* $U \subseteq \mathbb{R}^n$, or we can talk about a *Smooth Vector Field* on U. The second is much shorter and easier to think about.

If we draw the vector field in the above case, we get arrows which go around the space in a positive direction as in figure 4.6

A *solution* to the system of differential equations, or an *integral curve for the vector field* is a map $f : \mathbb{R} \longrightarrow \mathbb{R}^2$, usually written

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

with the property that $\dot{x}$ and $\dot{y}$ satisfy the given system of equations. What this means is that we think of a point moving in $\mathbb{R}^2$ so that it's velocity at

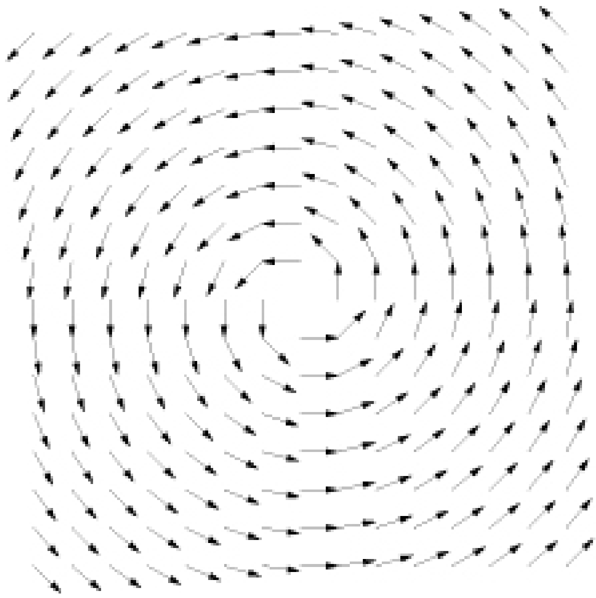

Figure 4.6: A vector field or system of ODEs in $\mathbb{R}^2$

any point is just the vector attached to that point. So the solution curve has to have the vector field tangent to it always.

It is possible to learn to solve autonomous systems of differential equations without ever understanding that they are all about vector fields which give the velocity of a moving point, and that a solution is simply a function which says where the moving point is at any time, and which agrees with the given vector field in what the velocity vector is. This is a pity.

In the above case, you can see by looking at the system what the solution is: obviously the solution orbits are circles, and given the initial condition where at time $t = 0$ we start at the point $(1, 0)^T$, the solution can be written down as

$$x = \cos(t), \qquad y = \sin(t)$$

and it is easy to verify that this works.

Exercise 4.5.1. Do it.

Obviously, solving initial value ODE problems for more complicated vector fields isn't going to be so easy, and doing it in dimensions greater than three by the 'look at it and think' method also looks doomed. So it is desirable to have a general rule for getting out the solution. Fortunately this is easy enough for linear vector fields in principle, although the calculations can be messy in preactice. But again, that's what computers are for.

4.5.2 Exponentiation of *Things*

If you write down the usual series for the exponential function you get:

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \frac{x^n}{n!} + \cdots$$

Now think about this and ask yourself what x has to be for this to make sense. You are used to x being a real number, but it should be obvious that it could equally well be a complex number. After all, what do you do with x? Answer, you have to be able to multiply it by itself lots of times, and you have to be able to scale it by a real number, and you have to be able to add the results of this. You also have to have an identity to represent x^0. Oh, and you need to be able to take limits of these things. So it will certainly work for x a real or a complex number. But it also makes sense if x is a square matrix. Or, with any system where the objects can be added and scaled and multiplied by themselves. And have limits of sequences of these things.

The name of a system of objects which can be added and scaled by real numbers is a vector space, and a vector space where the vectors can also be multiplied is called an *algebra*. We can do exponentiation in any algebra which has a norm and a multiplicative identity. (And it would be a help if it was *complete* in that norm, i.e. limits of cauchy sequences exist.) The square $n \times n$ matrices form such an algebra. We can also hope to take sequences of them and maybe have them converge to some matrix. So we can exponentiate square matrices.

Exercise 4.5.2. Exponentiate the matrix A in equation 4.7. Now exponentiate the matrix tA. Do you recognise the result?

It should be obvious that we could, in principle, calculate the exponential of a matrix to some number of terms, and if the infinite sum makes sense and the sequence of partial sums converges, then we could always get some sort of estimate of $\exp(A)$ for any matrix A by computing enough terms. We would hope that multiplying A by itself n times would give some reasonable sort of matrix, and when we divided all the entries by $n!$ we would get something pretty close to the zero matrix. If this happened for all the n past some point, then we could optimistically suppose that $\exp(A)$ was some matrix which we could at least get better and better approximations to, which after all is exactly what we have with $\exp(x)$ for x a real number.

Exercise 4.5.3. Define the norm of an $n \times n$ matrix A to be

$$\|A\| = \sup_{\|\boldsymbol{x}\|=1} \|A(\boldsymbol{x})\|$$

as in an earlier problem, and show that $\|A^2\| \leq (\|A\|)^2$. Hence prove that the function exp is always defined for any $n \times n$ matrix.

Exercise 4.5.4. If $e^{tA} \triangleq \exp(tA)$ denotes a map from $\mathbb{R}$ to the space of $n \times n$ matrices, show that its derivative is Ae^{tA}.

There are other algebras where a bit of exponentiation makes sense, so be prepared for them.

4.5.3 Solving Linear Autonomous Systems

In principle this is now rather trivial:

Proposition 4.5.1. *If* $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ *is an autonomous linear system of ODEs with* $\boldsymbol{x}(0) = \boldsymbol{a}$*, then*

$$\boldsymbol{x} = e^{tA}\boldsymbol{a}$$

is the solution.

Proof:

Differentiating e^{tA} gives Ae^{tA} by the last exercise and since $exp^{\mathbf{0}} = I$ the identity matrix, the initial value $\boldsymbol{x}(0) = \boldsymbol{a}$ is satisfied. So it is certainly a solution. □

If this looks a bit like a miracle and in need of explanation, you are thinking sensibly and merely need to do more of it. It may help to note that the exponential function is the unique function with slope at a point the same as the value at the point, and that this leads to the general solution for the linear ODE in dimension one, and that this goes over to higher dimensions with no essential changes. In effect, the exponential function was invented to solve all these cases. It actually goes deeper than this, see Vladimir Arnold's book *Ordinary Differential Equations.*

4.5.4 Existence and Uniqueness

Could you have two different solutions (or more)? No, not for linear systems, but this requires thought. Certainly the 1-dimensional ODE given by

$$\dot{x}(t) = 3x^{2/3}, \qquad x(0) = 0$$

has the solution $x(t) = x^{1/3}$ but also the solution $x(t) = 0$ It also has infinitely many other solutions. (Can you find some?) Of course this is not a linear

ODE, but it is clear that some sort of conditions will need to be imposed before we can look at vector fields which are not linear and expect them to have solutions. Happily, there is a simple one which guarantees at least local existence and uniqueness:

Theorem 4.5.1. If $f : U \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a continuously differentiable vector field, then for any point $\boldsymbol{a}$ in U there is a neighbourhood $W \subseteq U$ of $\boldsymbol{a}$ containing a solution to the system of equations $\dot{\boldsymbol{x}} = f(\boldsymbol{x})$ with $\boldsymbol{a}$ as initial value, and the solution is unique. Moreover, there is a continuously differentiable map $F : W \times J \longrightarrow \mathbb{R}^n$ for some interval $J = (-a, a)$ on $0 \in \mathbb{R}$ such that for all $\boldsymbol{b}$ in W, the map $F_b : J \longrightarrow \mathbb{R}^n$ is the solution for initial value $\boldsymbol{b}$ at $t = 0$.

There is a proof in Hirsch and Smale's *Differential Equations, Dynamical Systems and Linear Algebra*, pages 163 to 169.

There is a better proof in Arnold's book on page 213. It is actually the same proof but much better explained. It is given for the general (non-autonomous) case. Both arguments use the contraction mapping theorem from chapter 2. You should read through it if you have not already done a proof in your ODEs course. Assuming you did one.

The results follow easily from a more basic result sometimes called *The Straightening Out Theorem* (In Arnold *The basic theorem of the theory of ordinary differential equations* or *the rectification theorem*. See chapter 2). The theorem says that in a neighbourhood U of a point of $\mathbb{R}^n$ where the (continuously differentiable) vector field is non-zero, we can find a one-one differentiable map from U to $W \subseteq \mathbb{R}^n$ with a differentiable inverse, such that the transformed vector field on W is uniform and constant.

Given that we can do that, we could also make the vectors all have length one and lie along the x^1 axis in $\mathbb{R}^n$ with a rotation and scaling. The system of ODEs then would be, in this transformed region W, the rather boring system:

$$\begin{aligned} \dot{x_1} &= 1 \\ \dot{x_2} &= 0 \\ &\vdots \\ \dot{x_n} &= 0 \end{aligned}$$

with the solution

$$x_1(t) = t + a_1;\ x_2(t) = a_2;\ \cdots \boldsymbol{x}_n(t) = a_n$$

If you believe in the Straightening Out Theorem, then it is obvious that any continuously differentiable vector field has at any point where the vector field is non-zero a solution which is unique in some neighbourhood of the point and which depends smoothly on the point. All we have to do is to map the straight line boring solution(s) back by the differentiable inverse.

Exercise 4.5.5. Prove the last remark.

When the vector field is zero at a point, the solution is the constant function taking all of $\mathbb{R}$ to the point. So there is a unique solution here too.

Remark 4.5.1. You will find a proof of the straightening out theorem in Arnold. I shan't prove it in this course on the grounds that this isn't a course on ODEs. At least, I don't think it is.

4.6 Flows

I rather slithered over one important point, which is the question of whether we always get a solution for all time, past and future. It is not hard to see that the vector field $X(x) = x^2, X(0) = 1$ on $\mathbb{R}$ has a solution

$$x(t) = \frac{1}{1-t}$$

which goes off to infinity in finite time. From which we deduce that it is not in general possible to ensure that there is a solution for all time, and this explains the cautious statement of the last theorem. The best we can hope to do, the theorem tells us, for a smooth vector field at a point is to find a neighbourhood of the point in which there is a parametrised curve, $\boldsymbol{x}(t) :\ t \in (-a, a)$ where if we are lucky a will be ∞ and if we aren't it will be some possibly rather small positive number.

Definition 4.6.1. A vector field on $U \subseteq \mathbb{R}^n$ is said to be *complete* if any solution can be extended to the whole real line.

Exercise 4.6.1. Show that if a vector field has compact support then it is complete.

Exercise 4.6.2. Show that if U is an open ball in $\mathbb{R}^n$ centred on the origin and X is a smooth vector field on U, then if X is complete, and if $\text{Proj}(X(\boldsymbol{x}), \boldsymbol{x})$ is the projection of $X(\boldsymbol{x})$ on $\boldsymbol{x}$, then

$$\lim_{\|\boldsymbol{x}\| \to 1} \text{Proj}(X(\boldsymbol{x}), \boldsymbol{x}) = \mathbf{0}$$

Remark 4.6.1. It should be obvious that there are not many physical situations where things go belting off to infinity in finite time, and for that reason I shall restrict myself from now on to complete vector fields. If I forget to put the word in, put it in yourself. Also put the word 'smooth' in front of the term 'vector field' whenever it occurs since I shall not consider any other sort.

The business of getting a solution is going to work not just for the point we selected as our starting point but also for neighbouring points provided we don't go too far away. In the happy case where the vector field has solutions for all time, the space U on which the vector field is defined is decomposable as a set of integral curves, since solutions can't intersect each other, or themselves, although they can, of course, be closed loops. This statement follows from the uniqueness of a solution. Hence we deduce that a vector field gives rise to what is called a *foliation* of the space into integral curves. You can, perhaps, guess that partial differential operators more complicated than vector fields will give rise to higher dimensional foliations, decomposing the space into surfaces and other manifolds.

Exercise 4.6.3. Describe the foliation of $\mathbb{R}^2$ by the vector field

$$-y\,\frac{\partial}{\partial x} + x\,\frac{\partial}{\partial y}$$

Recall that we discussed the idea of groups acting on sets and came to the conclusion that they were conveniently seen as homomorphisms from a group G into the group $\mathrm{Aut}(V)$ of maps from the set V into itself. Then a complete smooth vector field X on $U \subseteq \mathbb{R}^n$ gives rise to an action of the group $\mathbb{R}$ on U as follows:

$$\begin{aligned} \boldsymbol{x} : \mathbb{R} \times U &\longrightarrow U \\ (t, \boldsymbol{x}_0) &\rightsquigarrow \boldsymbol{x}(t) \end{aligned}$$

where $\boldsymbol{x}(t)$ is the integral curve of X with $\boldsymbol{x}(0) = \boldsymbol{x}_0$.

To prove this is indeed a group action, we need to show that $\boldsymbol{x}(0, \boldsymbol{x}_0) = \boldsymbol{x}_0$ for every x_0 which follows immediately from my definition of $\boldsymbol{x}$. (Since the additive identity of $\mathbb{R}$ is 0.) We also need to show that

$$\forall\ s, t \in \mathbb{R}, \forall\ \boldsymbol{x}_0 \in \mathbb{R}^n,\ \boldsymbol{v}(s, \boldsymbol{v}(t, \boldsymbol{x}_0)) = \boldsymbol{v}(s+t, \boldsymbol{x}_0)$$

which merely means that if you travel for time t from $\boldsymbol{x}_0$ along the solution curve, and then go on for time s, this gives the same result as travelling for time $s + t$ from the starting point $\boldsymbol{x}_0$, which is, after all, what we expect a solution curve to do.

If we fix t and look to see what the group action does, it is a map from $\mathbb{R}^n$ to itself. Well, we knew that, it comes from chapter three. It is a truth that this map is always a smooth diffeomorphism. The old fashioned way of saying this is that the solutions depend smoothly upon the initial conditions, but I much prefer the modern way of saying it. You should be able to see that all we are doing is taking each point as input, and outputting the point it will get to after time t.

Proposition 4.6.1. *For a complete smooth vector field X on U open in $\mathbb{R}^n$, for any $t \in \mathbb{R}$, the map $\boldsymbol{x}_t : U \longrightarrow U$, which sends $\boldsymbol{x}_0$ to $\boldsymbol{x}(t, \boldsymbol{x}_0)$ is a diffeomorphism of U*

Proof:

The map $\boldsymbol{x}_t$ certainly has an inverse, $\boldsymbol{x}_{-t}$. And the theorem on existence of solutions to an ODE establishes that the map is continuously differentiable when X is. So if X is smooth, so is $\boldsymbol{x}_t$. □

Remark 4.6.2. The set of diffeomorphisms $\{\boldsymbol{x}_t : t \in \mathbb{R}\}$, or in other words the map $\boldsymbol{x} : \mathbb{R} \times U \longrightarrow U$, is called in old fashioned books a *one-parameter group of diffeomorphisms*. I shall simply say that the map $\boldsymbol{x}$ obtained from the vector field X is the *flow* of X.

Remark 4.6.3. Given a flow $\boldsymbol{x}$ on $U \subseteq \mathbb{R}^n$ we can always recover the vector field by simple taking any point, $\boldsymbol{a}$ and differentiating the map $\boldsymbol{x}_{\boldsymbol{a}} : \mathbb{R} \longrightarrow U$ which sends t to $\boldsymbol{x}(t, \boldsymbol{a})$ at $t = 0$. This must give us the required vector field from which the flow can be derived. So there is a correspondence between flows and vector fields.

You now have three ways of thinking about vector fields. They are bunches of arrows tacked onto a space; they are autonomous systems of ordinary differential equations. And they are also flows. This demonstrates that vector fields are more interesting and complicated than you might have supposed. Now we consider a fourth way of thinking about them.

4.7 Vector Fields as Operators

Vector fields have more structure than we have really noticed so far in our young lives, and we shall need to know something about this structure. If

$U \subseteq \mathbb{R}^n$ is an open subset, we shall be concerned with two things in this section, first the set of smooth maps from U to $\mathbb{R}$ and second vector fields on U. So some definitions to make things easier to talk about:

Definition 4.7.1. A map $f : U \longrightarrow \mathbb{R}^m$ for U open in $\mathbb{R}^n$ is said to be *smooth* iff it is infinitely differentiable, that is to say, all partial derivatives of all orders exist.

Remark 4.7.1. This ensures, of course, that f is continuous and continuously differentiable.

Definition 4.7.2. For U an open subset of $\mathbb{R}^n$, $\mathcal{F}(U)$ denotes the set of smooth maps from U to $\mathbb{R}$. I shall write $\mathcal{F}$ for short when U is fixed.

The set is in fact a vector space by pointwise addition of functions, and an algebra if we take into account pointwise multiplication. It is also a *ring*, which is to say it has both addition and multiplication like the ring of integers $\mathbb{Z}$. On the other hand, it is not a *field* like $\mathbb{R}$ or $\mathbb{C}$ since we cannot usually do division: that goes wrong when the map we are trying to divide by has a zero, as so many maps do.

I shall for the moment stick with the idea of a vector field (covector field) on U as a map $X : U \longrightarrow \mathbb{R}^n$ ($X' : U \longrightarrow \mathbb{R}_n$) where the elements on the output side of the map are regarded as arrows or covectors to be attached to points on the input side. This is going to have to be made precise subsequently because we need to generalise to the case where U is on some manifold which is not a subspace of a linear space, for example, a sphere.

Definition 4.7.3. For U an open subset of $\mathbb{R}^n$, $\mathcal{V}(U)$ denotes the set of smooth vector fields on U. And we already have $\Omega_1(U)$ for the set of covector fields.

Remark 4.7.2. In older books, a covector field is called a *contravariant* vector field and a vector field is called a *covariant* vector field. See for example, Mackey's *Theoretical Foundations of Quantum Mechanics.*

Again, $\mathcal{V}(U)$ and $\Omega_1(U)$ are vector spaces since we can add any two vector fields pointwise. I shall write $\mathcal{V}$ or Ω_1 for short when U is fixed for some discussion.

Now maps in $\mathcal{F}(U)$ act on vector fields in $\mathcal{V}(U)$ in an obvious way. Any $h \in \mathcal{F}$ and any $X \in \mathcal{V}$ can be multiplied as follows: for each point $\boldsymbol{a} \in U$, $h(\boldsymbol{a})$ is a real number and $X(\boldsymbol{a})$ is an "arrow" in $\mathbb{R}^n$ and we just scale the vector by the number. This gives a longer or shorter vector pointing in the same direction, or maybe with its direction reversed.

Definition 4.7.4. $\forall\ h \in \mathcal{F},\ \forall X \in \mathcal{V},\ hX \in \mathcal{V}$ is defined by

$$\forall \boldsymbol{a} \in U,\ hX(\boldsymbol{a}) = h(\boldsymbol{a})X(\boldsymbol{a})$$

Remark 4.7.3. It may be noted that we have something like a vector space over $\mathcal{F}$ here. Only since $\mathcal{F}$ is not a field, the set of vector fields over $\mathcal{F}$ can't be a vector space. The technical term is *module* over the ring $\mathcal{F}(U)$. For another example of a module over a ring, think of $\mathbb{Z} \times \mathbb{Z}$, an infinite two dimensional grid of points. This is a module over $\mathbb{Z}$. You can see how elements in it can be added and scaled by integers, but it is not a vector space. In the same way, vector fields on U can be added and scaled by elements of $\mathcal{F}(U)$, but it is not a vector space. Nearly, but no cigar.

Remark 4.7.4. I do not much care for dropping jargon on innocents, but some of the books use it and it is as well not to be intimidated by it.

A more interesting fact is that vector fields act on maps, specifically there is an operation which any $X \in \mathcal{V}$ does to maps in $\mathcal{F}$. If you think of $n = 2$ so that U is some open set in the plane, we can think of a map $h \in \mathcal{F}$ by its graph, and imagine $X \in \mathcal{V}$ as a stack of little arrows attached to points of U.

Now at each point $\boldsymbol{a}$ of U, we can take the directional derivative of h *in the direction of* $X(\boldsymbol{a})$, evaluate it at $\boldsymbol{a}$ and multiply by the length of $X(\boldsymbol{a})$. This will give us a new number to associate with the point $\boldsymbol{a}$, that is to say, we have a new map from U to $\mathbb{R}$.

Example 4.7.1. Take the vector field $-y\boldsymbol{i} + x\boldsymbol{j}$ on $\mathbb{R}^2$. What does it do to the map $h(x, y) = x^2 + y^2$?

Solution: We can write the directional derivative of h in the direction of $\begin{bmatrix} -y \\ x \end{bmatrix}$ multiplied by the length of this vector as

$$[2x, 2y] \begin{bmatrix} -y \\ x \end{bmatrix} = -2xy + 2xy = 0$$

so the effect of this X on this h is to kill it stone dead and send it to the zero map.

Remark 4.7.5. Xh is an element of $\mathcal{F}$ again, so we can think of a vector field X on U as an operator $\hat{X}$ on $\mathcal{F}(U)$.

$$\begin{aligned} \hat{X} : \mathcal{F}(U) &\longrightarrow \mathcal{F}(U) \\ h &\rightsquigarrow \hat{X}h \end{aligned}$$

where $\hat{X}h(\boldsymbol{a}) = Dh(\boldsymbol{a})X(\boldsymbol{a})$. Note that $\hat{X}$ is a *linear* operator;

Exercise 4.7.1. Confirm that $\hat{X}$ is a linear operator on $\mathcal{F}$.

Remark 4.7.6. **NEVER** confuse $\hat{X}h$ with hX. The former is a smooth map from U to $\mathbb{R}$, the latter is a vector field on U. Of course we can also talk about $h\hat{X}$, which is an operator form of a vector field on U.

Remark 4.7.7. The above example can be turned upside down: Find a map $h \in \mathcal{F}(\mathbb{R}^2)$ such that it satisfies the partial differential equation:

$$-y\,\frac{\partial h}{\partial x} + x\,\frac{\partial h}{\partial y} = \mathbf{0}$$

Then we have that $h(x,y) = x^2+y^2$ is a solution. In other words, vector fields regarded as operators on $\mathcal{F}$ give rise to (linear) partial differential equations. This has a lot to do with the interest of the area.

Remark 4.7.8. Old fashioned language is as follows: If $\hat{X}h = \mathbf{0}$ then h is said to be a *first integral* for X. Note that the solution to $\hat{X}h = \mathbf{0}$ and the solution of integral curves to the vector field are different kinds of things, although we see at once that if $\hat{X}h = \mathbf{0}$ then h must be constant along integral curves.

Exercise 4.7.2. Prove that last remark.

Example 4.7.2. The constant vector field $1\boldsymbol{i} + 0\boldsymbol{j}$ acts on $\mathcal{F}(\mathbb{R}^2)$ to simply give $\partial h/\partial x$ for any $h \in \mathcal{F}$. Similarly, the vector field $0\boldsymbol{i} + 1\boldsymbol{j}$ acts on $\mathcal{F}$ by sending each h to $\partial h/\partial y$.

In the light of the above example, I shall stop writing a vector field on $\mathbb{R}^2$ in the notation $P\boldsymbol{i} + Q\boldsymbol{j}$ Instead I shall write the vector field as

$$P(x,y)\frac{\partial}{\partial x} + Q(x,y)\frac{\partial}{\partial y}$$

This makes it clear (or should) that the vector field on U is sitting there, waiting to pounce on any map in $\mathcal{F}(U)$ and do some serious partial differentiating, in directions determined by the vector field.

Definition 4.7.5. The *Operator form* for a Vector field on an open set $U \subseteq \mathbb{R}^n$, expresses it (with respect to the standard basis) as

$$P_1(\boldsymbol{x})\frac{\partial}{\partial x^1} + P_2(\boldsymbol{x})\frac{\partial}{\partial x^2} + \cdots P_n(\boldsymbol{x})\frac{\partial}{\partial x^n} \quad \text{or} \quad \sum_{j=1,n} P_j \frac{\partial}{\partial x^j}$$

Remark 4.7.9. Many writers write L_X for the operator form of the vector field when they insist on thinking of the vector field as lots of little arrows. This is harmless but why bother? Why not just use X for both forms? (Actually there is a reason. We can think of L_X being an operator on $\mathcal{F}$ and also by defining it to act on the vector fields $\mathcal{X}$ by $L_X(X) = [X, X]$, see a little later. It can then be extended to act on the whole tensor algebra, where it is called the *Lie Derivative with respect to* X. See Abraham, *Foundations of Mechanics* pages 47-53.) I shall not,however, distinguish X from $\hat{X}$ from now on except where confusion might arise.

Remark 4.7.10. Some writers use the *Einstein Summation Convention* whereby a repeated dummy index/subscript (sometimes insisting that a subscript match an index) implies a summation. So the above may be written variously as:

$$P_j \frac{\partial}{\partial x^j} \quad \text{or} \quad P_j D_j$$

This is confusing at first but you get used to it. I shall save your poor little brains by not using it myself, but many books *do* use it. I suggest you translate into the longer notation until translation becomes superfluous.

Observe that the zero vector field kills every map $h \in \mathcal{F}$, that is, it sends them all to the zero function. Observe also that:

$$\forall\ h_1, h_2 \in \mathcal{F},\ \forall\ X \in \mathcal{V},\ (h_1 + h_2)X = h_1 X + h_2 X$$

$$\forall\ h_1 \in \mathcal{F},\ \forall\ X, Y \in \mathcal{V},\ h_1(X + Y) = h_1 X + h_1 Y$$

$$\forall\ h_1, h_2 \in \mathcal{F},\ \forall\ X \in \mathcal{V},\ X(h_1 + h_2) = X h_1 + X h_2$$

$$\forall\ h_1 \in \mathcal{F},\ \forall\ X, Y \in \mathcal{V},\ (X + Y)h_1 = X h_1 + Y h_1$$

Since X operates on $\mathcal{F}$, we can do it twice and get $XX(h)$ for any h and also XYh for any $X, Y \in \mathcal{V}$. Thus we can in a sense multiply vector fields. In general the result is a perfectly good operator on $\mathcal{F}$, but some calculations will rapidly convince you that XY is not, in general, a vector field operator but something much nastier.

Example 4.7.3. Let $V = -y\ \partial/\partial x + x\ \partial/\partial y$ and $W = x\ \partial/\partial x + y\ \partial/\partial y$ Then VWh is

$$-xy\frac{\partial^2 h}{\partial x^2} - y\frac{\partial h}{\partial x} - y^2\frac{\partial^2 h}{\partial x \partial y} - 0 + x^2\frac{\partial^2 h}{\partial y \partial x} + x\frac{\partial h}{\partial y} + xy\frac{\partial^2 h}{\partial y^2} + 0$$

and $WVh =$

$$-xy\frac{\partial^2 h}{\partial x^2} + 0 + x^2\frac{\partial^2 h}{\partial y \partial x} + x\frac{\partial h}{\partial y} - y^2\frac{\partial^2 h}{\partial x \partial y} - \frac{\partial h}{\partial x} + xy\frac{\partial^2 h}{\partial y^2} + 0$$

Neither of these look like a vector field operating on h. If however we take the difference, $VW - WV$ we get some happy cancellation and wind up with

$$(-y\frac{\partial h}{\partial x} + x\frac{\partial h}{\partial y}) - (x\frac{\partial h}{\partial y} - y\frac{\partial h}{\partial x}) = 0$$

which *is* a vector field although not a very interesting one.

Exercise 4.7.3. Write down another pair of vector fields V, W and compute $VW - WV$. Check to see if you always get the zero vector field. What is it telling you about the vector fields when $VW - WV$=0? (Some intelligent conjectures would be of interest but only if supported by evidence not used in framing the conjecture.)

Exercise 4.7.4. If $X = P(x,y)\partial/\partial x + Q(x,y)\partial/\partial y$ and $W = R(x,y)\partial/\partial x + S(x,y)\partial/\partial y$, calculate $XW - WX$ and verify that is is a vector field.

Exercise 4.7.5. Compute $XW - WX$ for $X, W \in \mathcal{V}(\mathbb{R}^n)$ and show it is a vector field in $\mathcal{V}(\mathbb{R}^n)$ Show that this also holds for $\mathcal{V}(U)$ for any open set $U \subseteq \mathbb{R}^n$.

Definition 4.7.6. The *Lie Bracket* or *Poisson Bracket* of two vector fields X, W in $\mathcal{X}(U)$, for $U \subseteq \mathbb{R}^n$ is written $[X, W]$ and defined by

$$[X, W] \triangleq XW - WX$$

It is a multiplication on the vector space of Vector fields on U.

Exercise 4.7.6. Do some simple calculations preferably for $U \subseteq \mathbb{R}^1$ and convince yourself that the Lie bracket multiplication is *not* in general associative but does satisfy the *Jacobi Identity*:

$$\forall X, Y, Z \in \mathcal{X}(U), \quad [X, [Y, Z]] + [Y[X, Z]] + [Z, [X, Y]] = 0$$

Exercise 4.7.7. Prove that the Jacobi Identity is *always* satisfied for Vector Fields.

Exercise 4.7.8. What is $X(h_1 h_2)$, the effect of operating with X on the product of two smooth maps from U to $\mathbb{R}$?

The Lie bracket makes the vector space of vector fields on U, an open subset of $\mathbb{R}^n$, into an *algebra*, which you will recall is merely a vector space where the vectors can be multiplied, and where the multiplication distributes over addition.

Exercise 4.7.9. Prove $[X,(Y+Z)] = [X,Y]+[X,Z]$ and $[(X+Y),Z] = [X,Z]+[Y,Z]$ Prove also that $\forall\ a \in \mathbb{R},\ [aX,Y] = a[X,Y]$ and $[X,aY] = a[X,Y]$.

Remark 4.7.11. The above properties you will recognise as bilinearity.

Any vector space with a multiplication (distributing over addition) which satisfies the Jacobi Identity, is bilinear, and anticommutative, $[X,Y] = -[Y,X]$ is called a *Lie Algebra*. We shall find whole families of them quite soon.

Exercise 4.7.10. Investigate the relation between $[hX,Y]$, $[X,hY]$ and $h[X,Y]$.

It should be apparent that although the calculations tend to be messy and provide great scope for making errors, they are not essentially difficult. A natural candidate for a good symbolic algebra package, you might say.

Exercise 4.7.11. Is there a multiplicative identity for the Lie Bracket operation on vector spaces? That is, is there a vector field J such that for every other vector field,X, $[J,X] = X$? (Hint: what is $[J,J]$?)

Recall that we accidentally found a solution, $h(x,y) = x^2+y^2$ to the PDE

$$-y\,\frac{\partial h}{\partial x} + x\,\frac{\partial h}{\partial y} = \mathbf{0}$$

Now this is one solution, and finding a single solution is very nice, but we usually want the general solution. In this particular case you can probably guess it. But in general, if we have some linear partial differential operator L acting on $\mathcal{F}$ and we want the set of all solutions of $Lh = 0$, then it will usually be a lot harder to find them. This process is aided by the following idea: The set of solutions of L is going to be a linear subspace of $\mathcal{F}$, by definition of the term linear operator. Call it $\mathcal{F}_0$. Now a *symmetry* of the solution space of the operator L, often called a symmetry of the operator L, is some vector field operator X such that X takes $\mathcal{F}_0$ into itself, ie. if whenever h is a solution to $Lh = \mathbf{0}$, so is Xh. If we know the collection of all symmetry operators for L and we have a solution, then we can find all the other solutions. In trivial cases this will amount to no more than adding in arbitrary constant functions, but in non-trivial cases it will do a whole lot more than this. So it would be a good idea to be able to find, for a given L, the set of all symmetries X for L. It is clear that the Poisson-Lie bracket can be used for any pair of linear operators, not just vector fields. The following observation goes some way to explaining our interest in them:

Proposition 4.7.1. *If $[L, X] = wL$ for some function $w \in \mathcal{F}$, then X is a symmetry of L.*

Proof:

We need to show that $\forall\ h \in \mathcal{F}, \quad L(Xh) = \mathbf{0}$ *Now*

$$LX - XL = gL \Rightarrow LX = gL + XL$$

and

$$\forall h \in \mathcal{F}, (gL + XL)h = gLh + XLh = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

□

Exercise 4.7.12. Prove the converse, that if X is a (vector space) symmetry of L, then $[L, X] = gL$ for some $g \in \mathcal{F}$.

Now it is possible to prove that the set of all vector space symmetries of an operator L is itself a Lie Algbra. Which is one reason for wanting to know more about them.

Some students last year wanted to know why it was that the partial differential equations we looked at all had their variables separable: does this happen for all possible PDEs and why does it work for these cases? The answer to this question is rather long and may be found in Volume 4 of the Encylopedia of Mathematics and Its Applications, *Symmetry and Separation of Variables* by Willard Miller. It has a lot to do with Lie Algebras.

It is now possible to state properly a problem I invited you to think about earlier.

Going back to the idea of flows, it makes sense to discover whether flows commute. For a suitable pair of flows, $x, y : \mathbb{R} \times U \longrightarrow \mathbb{R}^n$ we can start off from $\boldsymbol{a} \in U$ and go by flow x for a time s and then by flow y for time t. This will get us to some point in U, written naturally enough as $y_t \circ x_s(\boldsymbol{a})$. Or we could go the other way around, first by y and then by x to get $x_s \circ y_t(\boldsymbol{a})$. If we always wind up at the same point for any starting point and any pair of times s, t then we may say that the flows commute.

Then when the flows x, y correspond to the vector fields X, Y, we have the following result: x and y commute iff $[X, Y] = \mathbf{0}$. You can see that this works for the case of the two vector fields V, W in Exercise 4.7.3.

At present we lack the machinery to prove this result economically, so I shall develop it in the next two sections.

4.8 Tangent Spaces

At any point in $\mathbb{R}^2$ there are a lot of possible vectors we might attach to the point. In fact it is clear that the collection of all of them is just $\mathbb{R}^2$ again. Of course we maintain the idea that the two spaces are of different things, the first is a space of locations and the second a space of arrows. The fact that we use the same name for these should not blind us to the fact that they are collections of different kinds of objects. In fact you learnt in first year that it was useful to flip between the two interpretations: When we write the equation for a line as $\boldsymbol{a} + t\boldsymbol{b}$, we think of this as the line through the point $\boldsymbol{a}$ in the direction $\boldsymbol{b}$. But if $\boldsymbol{b}$ is a point, how come it *has* a direction? Some of you may have found this confusing at first, but you learnt to go along with it in order to solve problems and pass exams. We tend to take it that if two things are isomorphic they are the same with just the names changed. Now we have the opposite problem where the names are the same but the interpretation is different.

All this will become much clearer on manifolds, but for the time being I shall persist in this wretched convention. At least you are used to it, even if it is fundamentally muddled.

Anyway, I shall say that we attach the entire space $\mathbb{R}^2$ (arrows) to the point $\boldsymbol{a}$ in $\mathbb{R}^2$ (locations). This is called the *tangent space* at $\boldsymbol{a}$, which is a sensible name as it is a space of all possible vectors which might be in some conceivable vector field at the point $\boldsymbol{a}$. We write this as $T_a(\mathbb{R}^2)$.

I can do this for every $\boldsymbol{a} \in \mathbb{R}^2$, and if I do I get the space $\mathbb{R}^2 \times \mathbb{R}^2$ where the first is the space of locations and the second is the space of arrows. We write this space as $T(\mathbb{R}^2)$. I also have a natural projection

$$\begin{array}{rcl} \pi : \mathbb{R}^2 \times \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ (\boldsymbol{a}, \boldsymbol{v}) & \rightsquigarrow & \boldsymbol{a} \end{array}$$

which sends the entire tangent space over $\boldsymbol{a}$ to the point $\boldsymbol{a}$ to which it is attached.

With this structure, the object is called the *tangent bundle* over $\mathbb{R}^2$ The space of locations is called the *base space* of the bundle, and the tangent space over any point is called the *fibre* of the bundle. It is useful to picture this as figure 4.7

Since I couldn't draw the four dimensional space $\mathbb{R}^2 \times \mathbb{R}^2$, I drew a few of the planes in it, each one projecting down to a different point in the plane

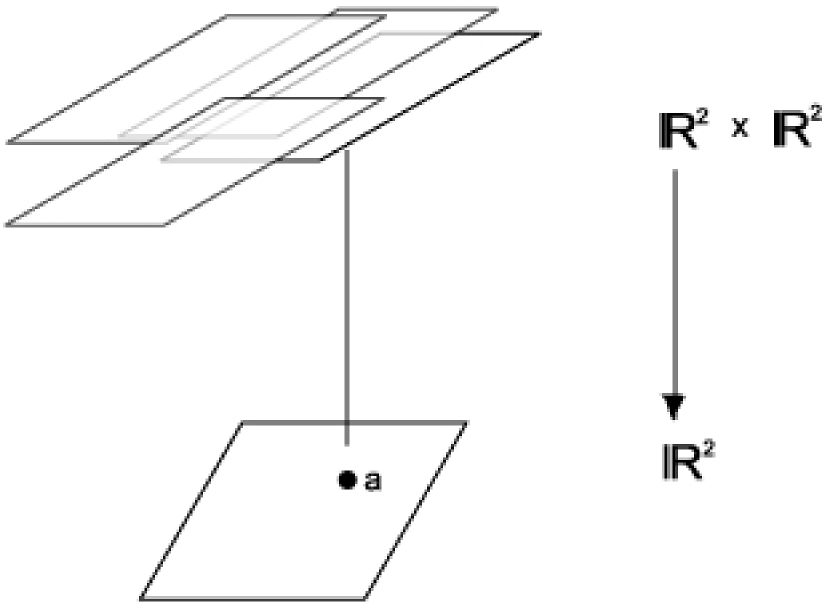

Figure 4.7: The tangent bundle over $\mathbb{R}^2$

at the bottom.

We are going to have to worry later about the space of tangents to a 2-sphere, where the base space is S^2. The tangent bundle will again be some four dimensional space of possible tangents, and again drawing it will be impossible. It is another four dimensional manifold. Its structure will tell us about what sorts of vector fields one can have on a sphere.

Definition 4.8.1. A *trivial vector bundle* over a topological space B is a triple $(B \times V, \pi, B)$ where V is some vector space and $\pi : B \times V \longrightarrow B$ is the projection map $\pi(\boldsymbol{b}, \boldsymbol{v}) = \boldsymbol{b}$ for every $(\boldsymbol{b}, \boldsymbol{v}) \in B \times V$. The vector space V is called the *fibre* of the bundle.

As you might gather, there are some non-trivial vector bundles. For an example, take the Möbius bundle over the circle S^1 with fibre $\mathbb{R}$. If you limit the fibre to be the interval $[-1, 1]$ it is just a Möbius strip. The whole thing is a bit too big to embed in $\mathbb{R}^3$ (at least if we want the fibre to be embedded in an affine way) and you will note that for intervals J of the circle that are too small to go all the way around the bundle over the interval is just $J \times \mathbb{R}$. So it is *locally trivial*. Over all however, it is intuitively clear that it is *not* $S^1 \times \mathbb{R}$ I should tell you that the tangent bundle to a 2-sphere is also locally trivial but not globally trivial. Let me say what this means:

Definition 4.8.2. A *locally trivial vector bundle* is a triple (E, π, B) where

$\pi : E \longrightarrow B$ is a continuous map between topological spaces, and there is a vector space V, the fibre, and a cover of B by open sets, such that for every open set U in the cover, $\pi^{-1}(U)$ is $U \times V$.

Example 4.8.1. The Möbius bundle mentioned above is the archetype example of a locally trivial but not globally trivial vector bundle. You should think of locally trivial vector bundles as a collection of product bundles glued together in some way that might introduce a twist. We shall find that the tangent bundle to a manifold is always a locally trivial vector bundle, but usually not globally trivial, that is, it is not generally the product of the manifold with $\mathbb{R}^n$

There is a category of things called fibre bundles or just bundles, complete with bundle maps. See Husemoller, *Fibre Bundles.* They exist, like groups, manifolds, algebras, and for that matter numbers, because people have found a use for them. Since the only cases we shall be concerned with are things like the tangent bundle, I shan't say any more about them. But the idea of a space being a whole lot of other spaces, the fibres, glued together is a natural and appealing one, and I hope you will find the idea interesting.

Definition 4.8.3. A *section* of a vector bundle (E, π, B) is a map $s : B \longrightarrow E$ such that $\pi \circ s = I_B$ where I_B is the identity on B.

If you think about what this means, you will see that s is selecting, for each $\boldsymbol{b} \in B$, some vector from V. Think of it as a vector which is attached to $\boldsymbol{b}$.

I can now define a vector field on $\mathbb{R}^2$ in a way which is related to but more informative than the earlier definition you met last year.

Definition 4.8.4. A vector field on an open set $U \subseteq \mathbb{R}^n$ is a section of the tangent bundle.

That makes five ways of thinking about vector fields. Don't ask me which is the right way, they all are.

Note that the tangent bundle over $\mathbb{R}^n$ is a trivial vector bundle. In fact every vector bundle over $\mathbb{R}^n$ for any n is trivial. They stop being trivial very quickly when we look at vector fields on any thing else however, as you have observed, the Möbius bundle is not trivial.

The tangent bundle over S^2 is more like the Möbius bundle than a trivial bundle. To persuade yourself of this, note that if the tangent bundle over S^2 were trivial, then it would make sense to have a constant vector field on it which is non-zero everywhere. You could do this on the circle S^1 for example. If you try to draw a vector field on a 2-sphere by putting lots of little arrows

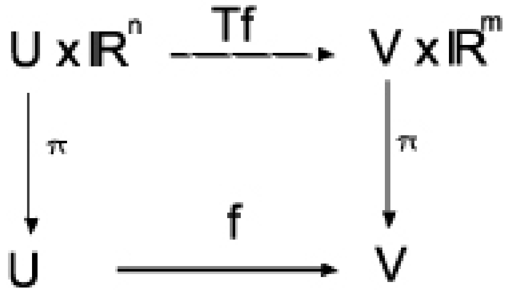

Figure 4.8: Exercise: Join the dots.

on it, you will soon discover that you cannot fill in the whole sphere with a smooth vector field. Of course your inability to do this doesn't show it can't be done, and in fact we can prove the result that every smooth vector field on $\mathbb{R}^2$ has a zero vector somewhere by studying the structure of the tangent bundle.

Now suppose the base of the tangent bundle is some open $U \subseteq \mathbb{R}^n$ and we have a differentiable map $f : U \longrightarrow V$, for V an open set in $\mathbb{R}^m$. Then there is a tangent bundle over V which is $V \times \mathbb{R}^m$. And there is a sort of gap on the tangent spaces which looks to me as if it ought to be filled in. If you have been following this carefully, you will immediately see how to fill in the dotted line in figure 4.8, that is you should be able to define Tf

This idea is really very cool, and maybe it should be taught in first year courses on calculus: the function f takes a point $\boldsymbol{a}$ to a point $\boldsymbol{b} = f(\boldsymbol{a})$. And the derivative of f at $\boldsymbol{a}$ is a linear map, *which we think of as taking any tangent vector at* $\boldsymbol{a}$ *to a tangent vector at* $f(a)$

Example 4.8.2. Take $f : \mathbb{R} \longrightarrow \mathbb{R}$ to be the map $f(x) = x^2$ Then $f(1) = 1$ and the derivative of f at 1 is the linear map from $\mathbb{R}$ to $\mathbb{R}$ which takes a tangent vector v to $2v$. So Tf is the map

$$\begin{aligned} Tf : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \times \mathbb{R} \\ (x, v) &\rightsquigarrow (x^2, 2xv) \end{aligned}$$

Example 4.8.3. Take $f : \mathbb{R}^2 \times \mathbb{R}^3$ to be any differentiable map then

$$\begin{aligned} Tf : \mathbb{R}^2 \times \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \times \mathbb{R}^3 \\ (\boldsymbol{x}, \boldsymbol{v}) &\rightsquigarrow \left(f(\boldsymbol{x}), \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} \end{bmatrix}_{\boldsymbol{x}} (\boldsymbol{v})\right) \end{aligned}$$

Definition 4.8.5. For U open in $\mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^m$ a smooth map

$$(Tf)(\boldsymbol{a}, \boldsymbol{v}) = (f(\boldsymbol{a}), f_{\star \boldsymbol{a}}(\boldsymbol{v}))$$

where $f_{\star \boldsymbol{a}}(\boldsymbol{v}) = Df_{\boldsymbol{a}}(\boldsymbol{v})$ is a convention that is commonly used.

You can see that Tf incorporates both f and its derivative in a way which is conceptually much cleaner than the usual approach to the definition of the derivative. In other words, I could present first year students with the idea that the real line has a tangent space sitting at every point of it. This is hard to draw because each vector gets in the way of the other points, not to mention the other vectors, although there is a way of doing it. Of course, it can be argued that it is rather difficult to visualise the real line itself. It is infinitely thin but has numbers written on it, each to an infinite number of decimal places in general. And it sticks out of the ends of the universe. Believing three impossible things before breakfast ought to be child's play to you by now.

If I can convince first year students that the real line makes sense, then I ought to be able to convince them that there is also a space of tangents to it, and that differentiable functions are functions which take pairs consisting of points of the real line and tangent vectors to them, to points and tangent vectors to *them*. Then with a bit of work I could get the notation for the derivative as the second part, the bit that maps (linearly) tangent vectors to tangent vectors.

Well, maybe it needs more than a bit of work. But I hope that you see the idea is simple. The notation that we are used to is not designed to make the ideas of calculus transparent but to make the calculations quick. Now we have computers, that is not such an issue as it used to be, whereas making the computations intelligible is morre important than ever.

I said that there is a way of drawing the tangent space to $\mathbb{R}$ and it is rather obvious if you think about it for thirty seconds. I can also draw the tangent space to S^1. All I do is to turn the tangent vectors through a right angle:

I have drawn the tangent bundle in figure 4.9. You will have to take my word for it that this *is* the tangent space to the circle, but if you draw the section which takes $\theta \in \mathcal{S}^1$ to $(\theta, 1) \in S^1 \times \mathbb{R}$, you should be able to see that it corresponds to something like figure 4.10, which looks the way a constant vector field on the circle ought to look.

Exercise 4.8.1. Sketch a version of figure 4.10 which corresponds to the section $s(\theta) = (\theta, \sin(\theta))$

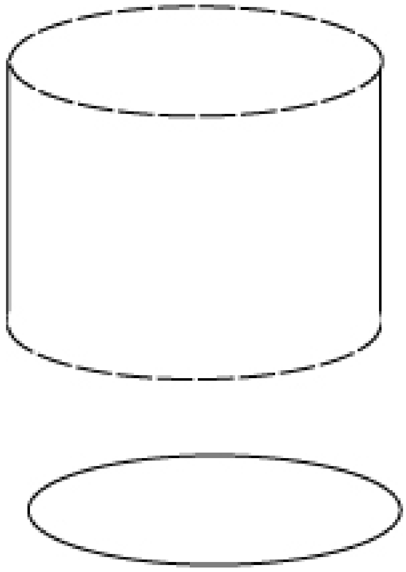

Figure 4.9: Some of the tangent space to a circle.

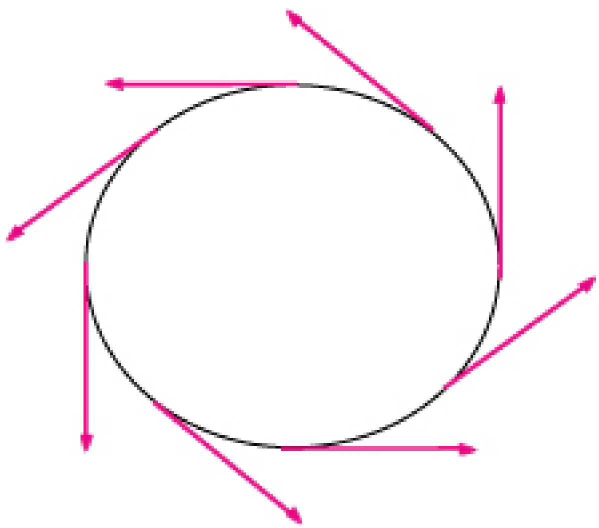

Figure 4.10: The unit vector field on a circle.

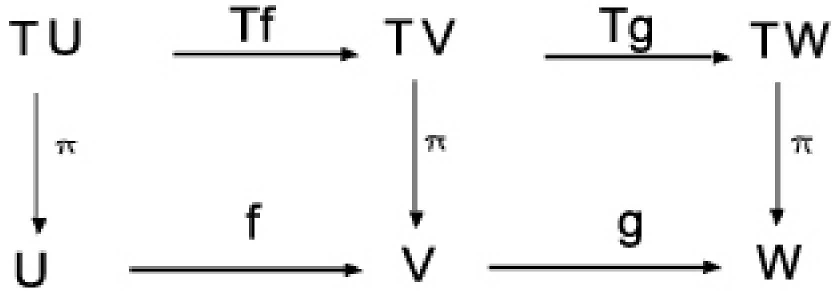

Figure 4.11: The Chain Rule.

Why do I write Tf for the function between the tangent spaces? Because T is a *functor* which I define on suitable objects U to give a new object, the tangent bundle over U, and for suitable maps between U and V to give a sensible map between TU and TV. A 'suitable object' will turn out to be a differential manifold, which things we will discuss in the next chapter.

Note that the chain rule holds; figure 4.11 shows what happens:

We have the chain rule embodied in the observation that $T(g \circ f) = Tg \circ Tf$ which is part of the definition of a functor.

4.9 Moving Vector Fields

4.9.1 One move

Definition 4.9.1. When U, V are open subsets of $\mathbb{R}^n$, I write $\mathcal{D}(U, V)$ for the space of all diffeomorphisms from U to V. Of course, there may not be any. I write $\mathcal{D}(U)$ for the space $\mathcal{D}(U, U)$. There is always at least one, the identity. Note that $\mathcal{D}(U)$ is a *group* under composition.

Exercise 4.9.1. Prove that last remark.

Remark 4.9.1. We shall be obliged to show at some point that $\mathbb{R}^n$ and $\mathbb{R}^m$ are diffeomorphic if and only if $n = m$. In fact we can replace diffeomorphic by homeomorphic and the result is still true. It is however not easy to prove, although your geometric intuitions should assure you that $\mathbb{R}$ and $\mathbb{R}^2$ are not homeomorphic.

Exercise 4.9.2. Prove that $\mathbb{R}$ and $\mathbb{R}^2$ are not homeomorphic.

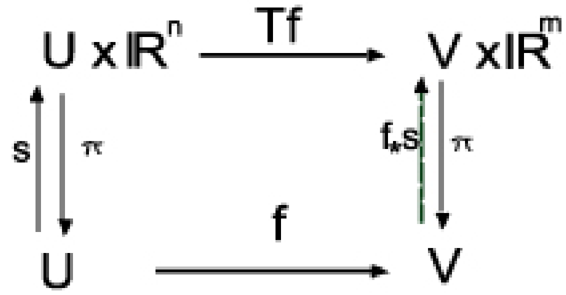

Figure 4.12: Moving vector fields.

Remark 4.9.2. Proving that $\mathbb{R}^2$ and $\mathbb{R}^3$ are not homeomorphic is harder. Can you think of a method?

Note that I am running out of letters to describe different things, and you cannot tell (unless I say so) whether f is a map from U to $\mathbb{R}$ or a map from U to V for U in $\mathbb{R}^n$, $V \subseteq \mathbb{R}^m$, and other possibilities suggest themselves. So I shall always tell you what set of maps any given map is taken from.

Suppose we have $f \in \mathcal{D}(U, V)$ and a vector field V on U, open in $\mathbb{R}^n$. Then we can use f to *move* the vector field to V. This is most easily seen by the section definition of a vector field, see figure 4.12

If the section of the tangent bundle which specifies the vector field on U is s, then the section over V which represents the result of using f to 'move' s is represented by the map $f_\star s$ (shown dotted in the figure) which is defined by

$$(Tf) \circ s \circ f^{-1} \tag{4.8}$$

In other words, take a point $\boldsymbol{b}$ of V, go back to see where it came from in U, say $\boldsymbol{a}$, find the vector asociated with $\boldsymbol{a}$ by the vector field, and finally use Tf to get to a tangent vector over $\boldsymbol{b}$. Observe that f has to be a diffeomorphism to ensure that the 'moved' vector field is defined and smooth.

Exercise 4.9.3. Prove that last remark.

Regarding a vector field as a section of the tangent bundle certainly makes it easy to write down the effect of a shift of it by a diffeomorphism. It is harder, but not too hard, to write down the definition of moving a vector field by a diffeomorphism in the case where the vector field is specified in the old fashioned way as a map from U to $\mathbb{R}^n$, the space of possible arrows on U.

If $X : U \longrightarrow \mathbb{R}^n$ is the vector field in old-fashioned notation, and we want $f_\star X : X \longrightarrow \mathbb{R}^n$ (same n in light of earlier remarks) where f is a diffeomorphism in $\mathcal{D}(U, V)$, then

$$f_\star X(\boldsymbol{b}) = (Df(f^{-1}(\boldsymbol{b})))(X(f^{-1}(\boldsymbol{b})) \tag{4.9}$$

The first outer parentheses give a linear operator on vectors on U, and the second outer parentheses give a suitable vector on U for it to act on. Alternatively you can think of the matrix representations of the former as an $n \times n$ matrix, and the second as a column matrix with n entries.

Exercise 4.9.4. Satisfy yourself that the two definitions of using a diffeomorphism to move a vector field are the same.

That was definitely clunkier than the section definition, but tolerable. It could be improved by writing it as:

$$f_\star X(f(\boldsymbol{a})) = Df(\boldsymbol{a})X(\boldsymbol{a}) \tag{4.10}$$

We have now expressed the same idea in two forms of the several ways of looking at vector fields. We next need to see the form this takes in the operator notation for a vector field. I shall go back to distinguishing X from $\hat{X}$ so as to reduce the risk of muddle.

Suppose we take $\boldsymbol{u}$ to be the vector $X(\boldsymbol{a})$ and we want to move it by f to get $\boldsymbol{v} = f_\star(\boldsymbol{u})$ at $\boldsymbol{b} = f(\boldsymbol{a})$. Write $\hat{X} : \mathcal{F}(U) \longrightarrow \mathcal{F}(U)$ for the operator form of X; it sends $g \in \mathcal{F}(U)$ to the function $\hat{X}(g)$, which takes at $\boldsymbol{a} \in U$ the value

$$Dg(\boldsymbol{a}) \cdot X(\boldsymbol{a})$$

where $\cdot$ is just matrix multiplication, or the action of a covector on a vector. I have put it in just to help you parse the string properly. This will be 'moved' to the operator that takes any $h \in \mathcal{F}(V)$ to $Dh(\boldsymbol{b}) \cdot \boldsymbol{v}$, where $\boldsymbol{b} = f(\boldsymbol{a})$ and $\boldsymbol{v}$ is $(f_\star X)(\boldsymbol{b}) = Df(\boldsymbol{a})\boldsymbol{u}$. Think of this as taking place upstairs in the Tangent space.

In other words, the result of operating on $\hat{X}$ is to give $f_\star(\hat{X})$ which operates on $h : V \longrightarrow \mathbb{R}$ to produce the function which at $\boldsymbol{b} \in V$ has the value

$$Dh(\boldsymbol{b}) \cdot Df(f^{-1}(\boldsymbol{b})) \cdot X(f^{-1}(\boldsymbol{b}))$$

where again $\cdot$ denotes matrix multiplication or evaluation. The left hand term is a row matrix, the middle term an $n \times n$ square matrix and the last term is a column matrix in the usual representation.

The above can be written by the chain rule as

$$D(h \circ f)(f^{-1}(\boldsymbol{b})) \;\cdot\; X(f^{-1}(\boldsymbol{b}))$$

Now $\hat{X}$ is the map which sends $g(\boldsymbol{a})$ to $Dg(\boldsymbol{a}) \cdot X(\boldsymbol{a})$ or, better, sends g to $Dg \cdot X$. So it sends $h \circ f$ to $D(h \circ f) \cdot X$

We can therefore write

$$\forall\ h \in \mathcal{F}(X), \quad (f_\star \hat{X})(h)(\boldsymbol{b}) = (\hat{X}(h \circ f))(f^{-1}(\boldsymbol{b}))$$

or more tersely:

$$\forall\ h \in \mathcal{F}(X), \quad (f_\star \hat{X})h = (\hat{X}(h \circ f)) \circ f^{-1} \tag{4.11}$$

Which is pretty cool and establishes rather elegantly the necessity for both f and f^{-1} to exist and be smooth.

Using s for the section form of a vector field, X for the arrow form and $\hat{X}$ for the operator form when they are the same thing is going to use up letters even faster. From now on I shall use the same symbol for each form. This will test your skill at avoiding muddle and confusion.

The final form of the shifting of vector fields by means of diffeomorphisms to be considered is the flow form of a vector field. If $\boldsymbol{x} : \mathbb{R} \times U \longrightarrow U$ is the flow on U corresponding to the vector field X, and if $f : U \longrightarrow X$ is a diffeomorphism, then f exports $\boldsymbol{x}$ to X to be a map $f_\star \boldsymbol{x} : \mathbb{R} \times V \longrightarrow V$ defined by

$$f_\star \boldsymbol{x} = f \circ \boldsymbol{x} \circ (I, f)^{-1} \tag{4.12}$$

where

$$\begin{aligned} (I, f) : \mathbb{R} \times U \quad &\longrightarrow \quad \mathbb{R} \times V \\ t, \boldsymbol{a} \quad &\rightsquigarrow t, f(\boldsymbol{a}) \end{aligned}$$

and obviously $(I, f)^{-1} = (I, f^{-1})$.

I shall not confound the flow $\boldsymbol{x}$ with the vector field X since this might confuse me, which would never do. I shall however write $\boldsymbol{x}_t$ for the flow to indicate that it can be regarded as a group of diffeomorphisms of U.

Exercise 4.9.5. Work out an ODE form of the shifting of a vector field. I suggest you do it for a concrete case on $\mathbb{R}^2$ first.

Exercise 4.9.6. construct some examples of vector fields on $\mathbb{R}^2$ and some diffeomorphisms (choose simple ones!) and use the diffeomorphisms to shift the vector fields back to $\mathbb{R}^2$. Do it using every form of a vector field, and confirm the equivalence of the five forms of equations 4.8, 4.9, 4.11, 4.12 and your very own ODE form.

Definition 4.9.2. If $f : U \longrightarrow U$ is a diffeomorphism of an open subset of $\mathbb{R}^n$ and X is a vector field on U, we say that f leaves X *invariant* whenever $f_\star X = X$.

Exercise 4.9.7. For your own choice of examples of vector fields and diffeomorphisms, which vector fields are invariant under which diffeomorphisms? If none, find at least two.

4.9.2 Doing it with Video

If we have some definite vector field G on U (complete and smooth, as usual), then we get a flow g_t which is a whole stack of elements in $\mathcal{D}(U)$, and each of these diffeomorphisms allows us to shift any vector field X to a new one:

Definition 4.9.3. If G is a vector field on U, open in $\mathbb{R}^n$ with flow $g_t : t \in \mathbb{R}$, and if X is another vector field on U, then for every $t \in \mathbb{R}$, the flow takes X to the vector field X_t defined by

$$X_t(g_t(\boldsymbol{a})) \triangleq g_{t\star,\boldsymbol{a}}(X(\boldsymbol{a})) \triangleq Dg_t(\boldsymbol{a})X(\boldsymbol{a})$$

Remark 4.9.3. The idea here is that the flow g carries X around the space U, moving the vector field X, this time smoothly with t instead of just once and for all for a single fixed diffeomorphism. Video, so to speak, instead of fixed images.

At $t = 0$, $X_0 = X$. I have described the vector field X_t in terms of some vector being assigned to a point, the first and oldest form of what a vector field is. I could, again, have described it using the idea that it is a section of the tangent bundle. In this case, X is a section of the bundle over U and we have the alternative definition:

$$X_t = T(g_t) \circ X \circ g_{-t}$$

Exercise 4.9.8. Confirm that this agrees with the case of a fixed diffeomorphism.

I could also have used the operator form for a vector field X where $X : \mathcal{F}(U) \longrightarrow \mathcal{F}(U)$ in which case I would have that the effect of the flow on X is to obtain the operator $X_t : \mathcal{F}(U) \longrightarrow \mathcal{F}(U)$ with

$$X_t(h) = X(h \circ g_t) \circ g_{-t}$$

since g_{-t} is the inverse to g_t. This only means we have $X_t = g_{t\star}X$ in the earlier notation.

Exercise 4.9.9. Confirnm that this too is consistent with the notation for a single diffeomorphism.

Definition 4.9.4. A vector field X is said to be *invariant* under the flow g_t iff $\forall\ t \in \mathbb{R}, X_t = X$.

Remark 4.9.4. We defined invariance under a single diffeomorphism; the case of invariance under a flow is more interesting.

Now for some exercises to nail these ideas down firmly.

My standard list of examples of vector fields includes:

1. The constant vector field $1\boldsymbol{i} + 0\boldsymbol{j}$ acts on $\mathcal{F}(\mathbb{R}^2)$ to simply give $\partial h/\partial x$ for any $h \in \mathcal{F}$. Call it EW
2. Similarly, the constant vector field $0\boldsymbol{i} + 1\boldsymbol{j}$ acts on $\mathcal{F}$ by sending each h to $\partial h/\partial y$. Call it NS.
3. There is also the (sum) constant vector field $\boldsymbol{i} + \boldsymbol{j}$, call it $NWSE$.
4. From example 4.7.3 we have: $X = -y\ \partial/\partial x + x\ \partial/\partial y$ and
5. $W = x\ \partial/\partial x + y\ \partial/\partial y$

You should be able to think of a few more.

Exercise 4.9.10. . For the above vector fields, find the associated flows and test each pair to see if they commute. Calulate the Lie bracket for the vector fields and verify that the Lie Bracket is zero iff the flows commute.

Exercise 4.9.11. For the above vector fields, find out which are invariant with respect to each of the others.

Exercise 4.9.12. Is a vector field always invariant under its own flow? If not always, ever or when?

Exercise 4.9.13. Sketch the results of using a flow to shift another flow around. Coloured pencils might come in handy.

Sliding vector fields around under the action of a flow is a reasonable idea, but note that there is another vector field we can get in this situation.

In operator terms, for any $h \in \mathcal{F}$, $X_t(h)$ will depend on t and we can therefore take

$$\frac{d}{dt} X_t(h)$$

Exercise 4.9.14. For some examples of vector fields X and flows g_t, compute $d/dt\ X_t(h)$. Write them down carefully and sketch the arrow form.

Remark 4.9.5. In the case where the vector field X is invariant under the flow, we have that $\forall\ t \in \mathbb{R}, X_t = X$ and so

$$\frac{d}{dt} X_t = 0$$

Example 4.9.1. Let G be the system of ODEs

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

The vector field G (I shall use the same symbol for other forms of the vector field) gives rise to the flow

$$g_t = e^{tG}$$

which, by direct calculation or a bit of drawing gives us:

$$g_t \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

Since g_t is a linear map for every $t \in \mathbb{R}$, we can observe that

$$\begin{aligned} T(g_t) : \mathbb{R}^2 \times \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (\boldsymbol{a}, \boldsymbol{x}) &\rightsquigarrow (g_t(\boldsymbol{a}), g_t(\boldsymbol{x})) \end{aligned}$$

If we apply this flow to the vector field G itself, we have that

$$G_t \begin{bmatrix} b^1 \\ b^2 \end{bmatrix} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} b^1 \\ b^2 \end{bmatrix}$$

which is

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b^1 \\ b^2 \end{bmatrix}$$

which is the same as the action of G. (Since rotations in the plane commute!) So for every $t \in \mathbb{R}, G_t = G$.

We see immediately that g_t leaves G invariant and it follows that

$$\frac{d}{dt} G_t \;=\; \mathbf{0}$$

since G_t does not depend on t at all.

Example 4.9.2. With the same G and g take X to be the vector field

$$\frac{\partial}{\partial x}$$

Then X assigns the same vector $\left[\begin{array}{c} 1 \\ 0 \end{array}\right]$ to every point. X_t must assign to every point the result of applying g_t to the vector $\left[\begin{array}{c} 1 \\ 0 \end{array}\right]$ which is clearly the vector $\left[\begin{array}{c} \cos(t) \\ \sin(t) \end{array}\right]$.

So the effect of $g_{t\star}$ is to steadily rotate the constant vector field anticlockwise. This seems reasonable.

The derivative of the flow,

$$\frac{d}{dt} X_t$$

can be worked out several ways; if we take it in operator form we have

$$\forall\ h \in \mathcal{F}(\mathbb{R}^2), \quad X_t(h) = \cos(t)\frac{\partial h}{\partial x} + \sin(t)\frac{\partial h}{\partial y}$$

so the derivative with respect to t gives the field

$$-\sin(t)\ \frac{\partial}{\partial x} + \cos(t)\ \frac{\partial}{\partial y}$$

This is a constant vector field (at any time; I mean constant in space) which is orthogonal to the vector field X_t.

Proposition 4.9.1. *For any vector field G on U with flow g_t, we have in operator notation with $h \in \mathcal{F}(U)$*

$$Gh = \frac{d}{dt} h \circ g_t = \lim_{t \to 0} \frac{h \circ g_t - h}{t} \tag{4.13}$$

Proof:

We have

$$\forall\ \boldsymbol{a} \in U,\ \forall\ h \in \mathcal{F},\quad \hat{G}h(\boldsymbol{a}) = Dh(\boldsymbol{a})G(a)$$

where I have gone back to hatting vector fields to turn them into operators. Now $G(\boldsymbol{a}) = Dg(0, \boldsymbol{a})$ by the definition of the flow g for the vector field G. So we can rewrite the above as

$$\forall\ \boldsymbol{a} \in U,\ \forall\ h \in \mathcal{F},\quad \hat{G}h(\boldsymbol{a}) = Dh(\boldsymbol{a})Dg_{\boldsymbol{a}}(0)$$

where $g_a : \mathbb{R} \longrightarrow U$ is the map g restricted to fixed $\boldsymbol{a}$ and has derivative at $t = 0$ precisely the vector $G(a)$.

The above equation can be written:

$$\forall\ \boldsymbol{a} \in U,\ \forall\ h \in \mathcal{F},\quad \hat{G}h(\boldsymbol{a}) = D(h \circ g_a)(0)$$

and the right hand side is defined to be

$$\lim_{t \to 0} \frac{h \circ g_a(t) - h \circ g_a(0)}{t}$$

which is

$$\lim_{t \to 0} \frac{h \circ g_a(t) - h(\boldsymbol{a})}{t}$$

giving us:

$$\hat{G}h(\boldsymbol{a}) = \lim_{t \to 0} \frac{h \circ g_t(\boldsymbol{a}) - h(\boldsymbol{a})}{t}$$

which is the required result in only slightly different language. □

Exercise 4.9.15. Verify this for $U = \mathbb{R}$ and G the constant vector field which assigns the arrow $(1, 0)^T$ to every point. Construct an example where $U = \mathbb{R}^2$.

There is an interesting and useful relationship between X_t, G, the source of the flow, so to speak, and $\boldsymbol{d}/dt\ X_t$:

Proposition 4.9.2.

$$\frac{d}{dt}\ X_t = [X_t, G]$$

Proof:

Define Z_t by

$$Z_t = \frac{d}{dt} X_t$$

Then

$$\begin{aligned}
\forall h \in \mathcal{F}(U), \quad Z_0(h) &= \lim_{t\to 0} \frac{X_t(h) - X(h)}{t} \\
&= \lim_{t\to 0} \frac{(X(h \circ g_t)) \circ g_{-t} - X(h)}{t} \\
&= \lim_{t\to 0} \frac{(X(h \circ g_t) - X(h) \circ g_t) \circ g_{-t}}{t} \\
&= \lim_{t\to 0} \frac{(X(h \circ g_t) - X(h)) - (X(h) \circ g_t - X(h)) \circ g_{-t}}{t} \\
& \quad \text{And since } g_{-t} \text{ tends to the identity as } t \to 0 \\
Z_0(h) &= \lim_{t\to 0} \frac{X(h \circ g_t) - X(h)}{t} - \lim_{t\to 0} \frac{X(h) \circ g_t - X(h)}{t} \\
&= X \lim_{t\to 0} \frac{h \circ g_t - h}{t} - \lim_{t\to 0} \frac{X(h) \circ g_t - X(h)}{t}
\end{aligned}$$

Since X is linear. It follows immediately that

$$Z_0(h) = XG(h) - GX(h) = [X_0, G](h) \text{ by Proposition 4.13.}$$

This shows that the result is true for $t = 0$, Now $g_{t\star}$ is a linear map and $g_{t\star}(Z_0) = Z_t$; we apply $g_{t\star}$ to both sides of the equation

$$Z_0 = [X_0, G]$$

to get the conclusion. □

4.10 Commuting Vector Fields

I still need to prove the result I mentioned earlier, that the flows x_t, y_t commute iff the vector fields X, Y corresponding to them satisfy the condition $[X, Y] = \mathbf{0}$.

First I observe that if $f : U \longrightarrow U$ is any diffeomorphism and X, Y are two vector fields on U, and if they are moved by f to $f_\star X$, and $f_\star Y$ respectively, then

$$f_\star[X, Y] = [f_\star X, f_\star Y]$$

This is a trivial exercise in definitions that was used in the last proof.

Exercise 4.10.1. Prove the preceding remark.

Next I prove a lemma that assures us that if a vector field is invariant under a diffeomorphism, then so is the flow. And also the converse, but that is rather obvious.

Lemma 4.10.1. If $f : U \longrightarrow U$ is a diffeomorphism for U open in $\mathbb{R}^n$, And if G is a (complete, smooth) vector field on U with flow g_t, then

$$\forall\ t \in \mathbb{R},\quad f \circ g_t = g_t \circ f\ \Leftrightarrow \forall\ \boldsymbol{a} \in U,\quad G(f(\boldsymbol{a})) = Df(\boldsymbol{a})(G(\boldsymbol{a}))$$

Proof:

I can define a flow on U by

$$\forall t \in \mathbb{R},\quad \hat{g}_t = f \circ g_t \circ f^{-1}$$

I claim this is the flow associated with the vector field $f_\star(G)$. This is easy to prove in operator notation. Observe that I can write, in view of proposition 4.13,

$$\forall\ h \in \mathcal{F}(U),\quad G(h) = \frac{d}{dt}(h \circ g_t)\bigg|_0$$

So

$$f_\star G(h) = G(h \circ f) \circ f^{-1} = (\frac{d}{dt}(h \circ f \circ g_t\bigg|_0) \circ f^{-1}$$

$$= \frac{d}{dt}(h \circ f \circ g_t \circ f^{-1}\bigg|_0$$

since f^{-1} is just a change of variables which does not depend on t.

$$= \frac{d}{dt} h \circ \hat{g}_t\bigg|_0$$

We have immediately that if

$$\forall\ \boldsymbol{a} \in U,\quad f_{\star\boldsymbol{a}}(G(\boldsymbol{a})) = G(f(\boldsymbol{a})\ (G \text{ is invariant under } f)$$

then f moves the whole flow into itself and

$$\forall\ t \in \mathbb{R},\quad f \circ g_t = g_t \circ f$$

Conversely, if the flow is invariant then by differentiating we deduce that the vector field is also invariant. □

Now I can prove:

Proposition 4.10.1. *If X, Y are complete, smooth vector fields on U, for U an open subset of $\mathbb{R}^n$ and if x_t, y_t are the associated flows, then*

$$\forall\ s, t \in \mathbb{R}, \quad x_y \circ y_s = y_s \circ x_t \Leftrightarrow [X, Y] = \mathbf{0}$$

Proof:

For x_t and y_t to commute we have that X is invariant under y_t, by the last lemma, and using the earlier notation for X_t as the result of moving X by y_t

$$\frac{dX_t}{dt} = \mathbf{0} = [X_t, Y] = [X, Y]$$

Conversely, if $[X, Y] = \mathbf{0}$ we have

$$\forall\ t \in \mathbb{R}, \quad \frac{dX_t}{dt} = (x_t)_\star[X, Y] = 0$$

and the result follows from the last lemma. □

The last few results have been very geometric in character and you should be sure that you can interpret them quickly when you see them in algebra, and also go in the other direction and first see geometric properties of flows and vector fields from contemplating scruffy drawings, and then say them in algebra. This is guaranteed to impress the peasants and create the impression of colossal intellect at very little expense.

4.11 Tensor Fields

Just as we can take an m-dimensional vector space of arrows (casually referred to as $\mathbb{R}^m$) and assign an arrow to each point of an open set $U \subseteq \mathbb{R}^m$ by having a map from U to the space of arrows, so we can 'attach' other things. And just as we can attach arrows more or less smoothly by ensuring the map is smooth, so we can hope to attach other more complicated objects.

Physics and engineering give us a whole of things that can usefully be attached to a space.

Numbers. As when we have what is called a *scalar field*, for example, the temperature at a point of a region U of the space we live in. Of course it is rather easy to attach a number at each point. We do it with a function from U to $\mathbb{R}$, in other words we have been dealing with them for many a long yonk.

Arrows. These things are called *vector fields*, surprise surprise. Typically we have them in classical mechanics representing forces, and the arrows are attached to points in a phase space of positions and momenta having dimension six. In Robotics we may have nastier spaces occurring, involving perhaps the special orthogonal group. Again, we attach the arrows to each point of some open set $U \subseteq \mathbb{R}^m$ by a (usually smooth) map from U to the vector space of arrows, once somewhat casually identified as $\mathbb{R}^m$, but better thought of as a the part of the tangent bundle over the point.

Matrices If you take a solid object and deform it then you can represent the deformation by a map $f : U \longrightarrow \mathbb{R}^3$ where U is an open subset of $\mathbb{R}^3$ representing the solid object. And at each point of U the map f has a derivative which is a 3×3 matrix saying something about the local nature of the deformation. So we quite naturally have a matrix attached to each point of U. If we apply a force to the boundary of U, this force will be transmitted, with some changes, to the interior of U and the way in which the material deforms will be related to both the external force and to the properties of the medium, which may be different in different directions. The way in which an external force on the boundary is transmitted and turned into a deformation of the region U and the amount of fight that U puts up to being deformed, are determined by elasticity properties of U which also take a matrix (at least) to specify them, and which can in principle change from point to point.

Matrices may also be used to specify the inner product on a space; the ordinary dot product on $\mathbb{R}^2$ can be represented by the identity matrix:

$$\begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = [x, y] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

If instead we had some other (symmetric, positive definite) matrix, we would have merely a different inner product. Given an inner product we can derive a norm and a metric. Now we can imagine a space in which the inner product changed from place to place, and we could make sense of this by attaching the relevant matrix at different points of the space, as you move smoothly through the space, the matrix changes smoothly too. When we do this on a manifold it is called the *Riemannian Metric Tensor*, although it is really not just a metric but a whole lot more. Of course, the way to use such a thing is to calculate the inner product of vectors in the tangent space at the point where the matrix is attached. In other words, we need a local vector bundle with fibre the vector space of linear maps from V to V where V is the tangent space.

bivectors Also cobivectors, *et cetera*. All the things described in the Algebraic Preliminaries, Chapter Three, can be attached. In particular if we take the vector space of k-multilinear maps from $V \times V \times V \cdots \times V$ to $\mathbb{R}$, where there are k copies of the tangent space, V then we can attach this vector space to each point of U just as we took the tangent space over U. It is just another locally trivial vector bundle over U, this time of dimension n^k. We can do this for all the k at once, to give a locally trivial vector bundle over U with fibre a whole stack of vector spaces and inclusions and maps between them. All sitting over each and every single point of the space U. If the mind doesn't boggle, at least a bit, you clearly don't understand it. Then to get a tensor field, you take a *section* of this tensor bundle, each point of U is assigned something in the tensor bundle sitting over the point. Note that everything assigned has to be from some sort of vector space, though it might have some humungous dimension.

The above are called covariant tensor fields. We also want to have contravariant tensor fields where we have s-multilinear maps from $V^\star \times V^\star \times \cdots \times V^\star$ to $\mathbb{R}$ where there are s copies of $V^\star$, the dual space to V, the space of tangents to U. And finally we want to do it for *mixed* variances as discussed in Chapter Three. The entire tensor bundle for an open set $U \subseteq \mathbb{R}^m$ is just the cartesian product of U with the humungous vector space of all those k, s multilinear maps.

All of these things are called *tensor fields*, scalar fields and vector fields being among them. They can all be thought of in terms of (k, s)-tensors for some natural numbers k and s.

Exercise 4.11.1. Show that the scalar fields and vector fields and covector fields can be regarded as tensor fields for particular choices of k, s.

Exercise 4.11.2. Show that the Riemannian Metric Tensor is a tensor field for particular choices of k, s.

4.12 Summary

We started off by looking at *tangency* of maps f from $\mathbb{R}^n$ to $\mathbb{R}^m$. This is a more fundamental idea than the derivative, regarded as the linear map which is part of the best affine approximation to f, because the latter does not make sense when we are not in a linear space but on, say a sphere. Or some more horrible space. We proved the Inverse Function Theorem and the Implicit Function Theorem, with, I fervently hope, some insight into what is going on, and Sard's Theorem, which in the easy case considered here is no more than making some simple estimates of the measure of degenerate cuboids.

We then started on discovering more things about vector fields than you would have thought possible. You saw them in five different and complementary ways:

1. As a whole lot of little arrows stuck onto a space U.
2. As an autonomous system of Ordinary Differential Equations on U.
3. As a *flow*, which is a (1-parameter) subgroup of the diffeomorphisms of U into itself, the flow being the family of all possible solutions to the system of ODEs.
4. As an operator on the space of all smooth functions from U to $\mathbb{R}$. This led us naturally to turning vector fields on U into an algebra, with a non-associative multiplication called the *Lie Bracket*
5. As a section of the tangent bundle over U.

We looked at the interplay between vector fields and diffeomorphisms of U and how one can move a vector field by a diffeomorphism, and (looking at the video case) use a flow to move a vector field about smoothly. We looked at commuting vector fields and how they relate to Lie (or Poisson) Brackets. Finally we made a few rather vague remarks about tensors and tensor fields designed to put your brain on the right planet when it comes to handling these things later.

Chapter 5

Manifolds

5.1 Definitions

A *manifold* is something that is locally just a piece of $\mathbb{R}^n$ for some n. We need to rule out some rather obscure possibilities allowed by that rather shoddy definition. Let us make it a bit more respectable:

Definition 5.1.1 (Trial). A *Manifold* is a topological space M and for every $\boldsymbol{a} \in M$, $\boldsymbol{a}$ is in an open set U in M, and there is a map $f : U \longrightarrow \mathbb{R}^m$ such that f is a homeomorphism between U and the open set $f(U) \subseteq \mathbb{R}^m$.

I have gone to dimension m for M in order to distinguish clearly the dimension of a manifold and the dimension of any space in which it is sitting. Thus S^1 has dimension one although it is sitting in $\mathbb{R}^2$.

The first dificulty is the apparent claim that if U is open then so is $f(U)$; it seems at first sight to follow from f^{-1} being continuous– but that only guarantees that $f(U)$ is open in $f(U)$, not that it is open in $\mathbb{R}^m$. So there is a problem here.

Another problem occurs when we consider the quotient space $\mathbb{W}$eird$\mathbb{R}$, which is obtained by taking $\mathbb{R} \times S^0$, two copies of $\mathbb{R}$, and declaring that

$$p : \mathbb{R} \times S^0 \longrightarrow \mathbb{W}\text{eird}\mathbb{R}$$

has $p(t, -1) = p(t, 1)$ for every t except zero. So $\mathbb{W}$eird$\mathbb{R}$ is the set of points t for $t \in \mathbb{R} \setminus \{0\}$, together with the pair of points $(0, 1), (0, -1)$. It looks like a copy of $\mathbb{R}$ except that it has two zeros. Now according to the trial definition given above, this is a manifold. I have a bad feeling about this one also. I figure we could look at vector fields on manifolds and find that such spaces play hell with completeness of the vector fields. But it is easily fixed, we can eliminate it as a possibility by inserting the word 'hausdorff' before the words 'topological space' in the definition.

Finally, if the space is not connected, we have something that looks a bit unlikely because it hasn't got a unique dimension: we could take the disjoint union of $\mathbb{R}$ and $\mathbb{R}^2$ and this would be a manifold. This too gives me a bad feeling. I would like to say it is two manifolds, not one. I could insist on a manifold being connected, but quite sensible contenders such as $S^0 \times \mathbb{R}$ and O_3 the orthogonal group are not connected. So instead I shall change the definition:

Definition 5.1.2. A *Manifold* is a hausdorff topological space M such that there is an integer m called its *dimension*, and for every $\boldsymbol{a} \in M$ there is an open set $U \subseteq M$, with $\boldsymbol{a} \in U$, and a homeomorphism c from U to an open set $c(U) \subseteq \mathbb{R}^m$.

Recall that we claimed that $\mathbb{R}^n$ homeomorphic to $\mathbb{R}^m$ occurs only when $n = m$. It is a lot harder to prove this than you might think, although you can distinguish the case when $n = 1$ as being easy.

It now becomes possible to prove that some obvious things are manifolds:

Example 5.1.1. S^1 is a manifold. How we show this depends on our definition of S^1, we saw that we could construct it as a quotient space of the closed interval I or as a subset of $\mathbb{R}^2$. Taking the latter definition, we have that it is

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : \quad x^2 + y^2 - 1 = 0 \right\}$$

To see that is is a manifold, we merely use the implicit function theorem which says in this case that for any point other than $x = \pm 1$ we have the partial derivative for y is non-zero and so for all other points there is a neighbourhood diffeomorphic to an open interval in $\mathbb{R}$, and hence certainly homeomorphic. And for the points $x = \pm 1$ we need only swap x and y to get a diffeomorphism of any neighbourhood of either point with an open interval of the y axis. Alternatively we can use proposition 4.3.2 and observe that the rank of $\mathrm{D}(f)$ is always 1.

The space is a subspace of a metric space so must be hausdorff.

This gives the result without me having to find explicit homeomorphisms, but of course it is easy to do that too.

Exercise 5.1.1. Supply a collection of four homeomorphisms $c_j : 1 \leq j \leq 4$ from suitable open subsets of S^1 to $\mathbb{R}$ which suffice to make S^1 a manifold.

Exercise 5.1.2. Supply a different collection of two homeomorphisms from open subsets of S^1 to $\mathbb{R}$ which suffice to make it a manifold.

Since these homeomorphisms need to be talked about rather often, I shall name them:

Definition 5.1.3. A *chart* (sometimes: *coordinate chart*) is a homeomorphism $c_j : U_j \longrightarrow W_j \subseteq \mathbb{R}^m$ from an open subset of a manifold M to an open subset W_j of $\mathbb{R}^m$.

Then we can say that a manifold is a hausdorff space in which every point is in the domain of a chart into $\mathbb{R}^m$. Same m for every chart. Note that a chart is the inverse of a parametrisation. If we insist that our parametrisations of a space should be homeomorphisms, then parametrising the unit circle takes at least two.

Definition 5.1.4. An *atlas* of charts for a manifold M is some set of charts such that every point of M is in the domain of some chart in the atlas, that is, the domains of the charts are an open cover of M.

Exercise 5.1.3. Prove that the cartesian product of two manifolds is a manifold and hence that the torus $T^2 \cong S^1 \times S^1$ is a two dimensional manifold.

Exercise 5.1.4. Prove that S^2 is a manifold. (Hint: Use proposition 4.3.2)

Exercise 5.1.5. Prove that S^n is a manifold for every positive integer n.

Exercise 5.1.6. Prove that the space consisting of the unit interval with the end points identified, widely believed to be S^1, is a manifold. (Hint: Try mapping it to S^1 by a 1-1 onto map and confirm that the topology is right for the map to be a homeomorphism. Then charts on S^1 correspond to charts on $I/\partial I$.)

Exercise 5.1.7. Prove that O_2, the space of 2×2 matrices having columns orthogonal and of length 1 is a manifold, by explicitly constucting a map from $\mathbb{R}^4$ to $\mathbb{R}^3$ of which it is the kernel, and showing the conditions of the implicit function theorem can be satisfied.

Exercise 5.1.8. Do the same for O_3.

Exercise 5.1.9. Make manifolds into a category by defining the maps between them.

Exercise 5.1.10. Define the term submanifold.

Exercise 5.1.11. Now look at a map from $\mathbb{R}$ into $S^1 \times S^1$ sending t to $\cos(t), cos(\pi t/3)$. Is the image a submanifold in your definition? Should it be?

Exercise 5.1.12. Show that SO(3), the subset of O_3 having determinant $+1$, is a submanifold.

Note that we could have written the implicit function theorem in the form:

Proposition 5.1.1. *If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a smooth map with kernel*

$$S = \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) = 0\}$$

and if Df has rank m on S, then S is a manifold of dimension $n - m$

Proof:

See proposition 4.3.2 □

Remark 5.1.1. Note that in this form it is a nice generalisation of the rank-nullity theorem of first year linear algebra. Since all linear maps are smooth, being their own derivatives, the rank-nullity theorem is just the special case where f is linear.

5.2 Submanifolds

Earlier I asked you to define the term *submanifold*, and I expected you to produce something like " M_1 is a submanifold of M_2 iff restricting the charts from an atlas for M_2 to M_1 gives an atlas of charts for M_1." The example of the line inserted into the torus might have given you occasion to think that this might fail to capture some of the intuitive feeling for what a submanifold ought to be. For example, S^1 as usually defined is, I should say, a submanifold of $\mathbb{R}^2$. But the problem with the line on the torus is that it winds infinitely often about the torus and is dense in it, that is, for every point of the torus and for every open set containing that point, the line intersects that open set. The subset of the torus not covered by the image of the line has measure zero. If this is to be a submanifold it is not a very nice one.

Definition 5.2.1. A *submanifold* of a Manifold M_2 is a subset which is the image by a continuous 1-1 map of a manifold M_1.

Definition 5.2.2. A map $f : X \longrightarrow Y$ between topological spaces is said to be *proper* iff the inverse image of a compact set is compact.

Exercise 5.2.1. Show that a map of the line $\mathbb{R}$ into the torus T^2 which has dense image is not a proper map.

Definition 5.2.3. a map from a manifold to another which is one-one, continuous has a continuous inverse from its image and is proper, is called an *embedding.*

Remark 5.2.1. The map $f : [0, 2\pi) \longrightarrow S^1$ which has

$$f(t) = \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}$$

is clearly continuous and 1-1 and proper. But the inverse is not continuous. So it is not a homeomorphism, nor an embedding.

Definition 5.2.4. The image of an embedding of a manifold M_1 in a manifold M_2 is called a *closed submanifold*

Exercise 5.2.2. Prove that a closed submanifold is a manifold. Is a submanifold that is not closed a manifold?

Exercise 5.2.3. Give a condition on a submanifold that it be closed.

5.3 Smooth Manifolds

We want to be able to talk about smooth vector fields on manifolds, and indeed to recover all the material covered in the analysis preliminaries chapter, this time on manifolds. Intuitively, it feels reasonable to talk about a smooth map from S^1 to S^2, where a moving point moves smoothly around a loop on the surface of the sphere. As defined, we see that it makes sense to have a continuous map from S^1 to S^2, but not a smooth one. We have no reason to expect the charts we pick for a manifold to be smooth, and indeed it doesn't even make sense to require them to be, since all we know about M is that it is a hausdorff topological space. On the other hand, it is possible to think of a manifold as being a collection of open subsets of $\mathbb{R}^m$ glued together on the overlap of the domains of the charts. In figure 5.1 I show two of the charts and their overlap getting sent to generally different parts of $\mathbb{R}^m$. If I were to take the open sets in $\mathbb{R}^m$, cut them out, and then glue them together on the overlap by identifying points which get carried to each other by the homeomorphism, and if I do this for all the open sets, I reconstruct the manifold.

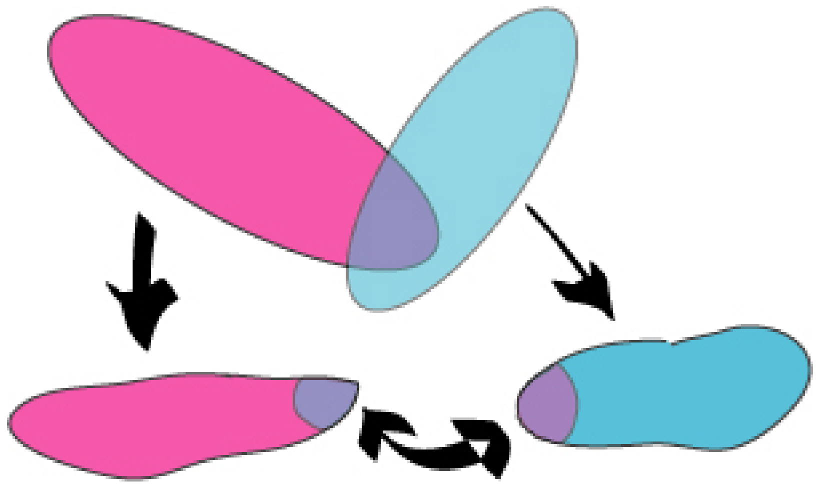

Figure 5.1: Charts as gluing devices

Exercise 5.3.1. Show that the quotient space of the disjoint union of the open sets in $\mathbb{R}^m$ by the identification of points produced by the homeomorphisms, is indeed homeomorphic to the manifold.

This gives us a way of putting a *differential structure* on M. The intuitive idea of a differential structure on a manifold is that it is enough to make sense of maps between such manifolds being smooth or differentiable.

The important part of this is the composite homeomorphisms on the image of the overlap: if $\phi_i : U_i \longrightarrow \mathbb{R}^m$ is one of the charts and $\phi_j : U_j \longrightarrow \mathbb{R}^m$ is another then we have a map

$$\phi_j \circ \phi_i^{-1} : V_{ij} \longrightarrow W_{ij}$$

where $V_{ij} = \phi_i(U_i \cap U_j)$ and $W_{ij} = \phi_j(U_i \cap U_j)$. This is a composite of homeomorphisms and inclusion maps so is a homeomorphism from an open subset of $\mathbb{R}^m$ to an open subset of $\mathbb{R}^m$.

The idea is to stipulate that these homeomorphisms should actually be *diffeomorphisms*. This makes sense because the range and domain of each such map is an open subset of $\mathbb{R}^m$.

I defined an *atlas* for a manifold M as a set of charts the domains of which covered the manifold. Obviously there are a lot of possible charts which one could use for such a thing as a circle, S^1, or a sphere, S^2.

Suppose we have a manifold and an atlas of charts

$$\mathcal{A} = \{\phi_i : U_i \longrightarrow \mathbb{R}^m\}$$

satisfying the condition that $\phi_j \circ \phi_i^{-1} : V_{ij} \longrightarrow W_{ij}$ is a smooth diffeomorphism whenever it is defined, i.e. whenever $U_i \cap U_j \neq \emptyset$. We shall say that $\mathcal{A}$ is a *smooth* atlas for M. Suppose now another chart, not in the atlas, is given, $\psi : U_k \longrightarrow \mathbb{R}^m$. We shall say that the chart is *compatible with the atlas* $\mathcal{A}$ if adjoining the chart to $\mathcal{A}$ gives another smooth atlas. We can keep on doing this for rather a lot of charts.

Definition 5.3.1. A *Maximal Smooth Atlas* for a manifold M is a smooth atlas which contains every possible compatible chart. It is also called a (smooth) *Differential Structure* for M. With this atlas, M is said to be a *smooth differential manifold.*

Remark 5.3.1. The above construction of a manifold is *intrinsic* in that it gives you a way of constructing a manifold without any reference to any other space in which it might be sitting. It is like the definition of a circle as the unit interval with the end points glued together, as contrasted with defining S^1 as a subset of $\mathbb{R}^2$. Yet manifolds arise in nature (so to speak) by putting conditions on n measured variables. Thus manifolds are very commonly *embedded* as subsets of $\mathbb{R}^n$, and this is how we normally meet them. Apart from the universe we live in, which has the structure of a three dimensional manifold (crossed, by the cartesian product, with Time) all the manifolds we shall deal with can be thought of as submanifolds of $\mathbb{R}^n$ for some n. There wouldn't seem to be much sense in thinking of the universe sitting in some higher dimensional space since there is no way of getting outside the universe and checking up on it. So in Physics we definitely need to have intrinsic definitions of things. It is also true that all the stuff we did on vector fields can be shifted across to manifolds by taking charts. It is also true that all smooth m-manifolds can be smoothly embedded in $\mathbb{R}^n$ for n not more than $2m + 1$. So we are not gaining any generality by defining things intrinsically. I had to decide whether to go through the intrinsic definition above or to simply consider subsets of $\mathbb{R}^n$ defined by smooth conditions on the n variables. The second would have saved some time, but you would have been unable to read most of the books in the area, so I went the long way around.

5.4 The Tangent Bundle

I shall now show that there is a sensible category of smooth differential manifolds and smooth maps. This means I need to talk about the tangent space on a manifold.

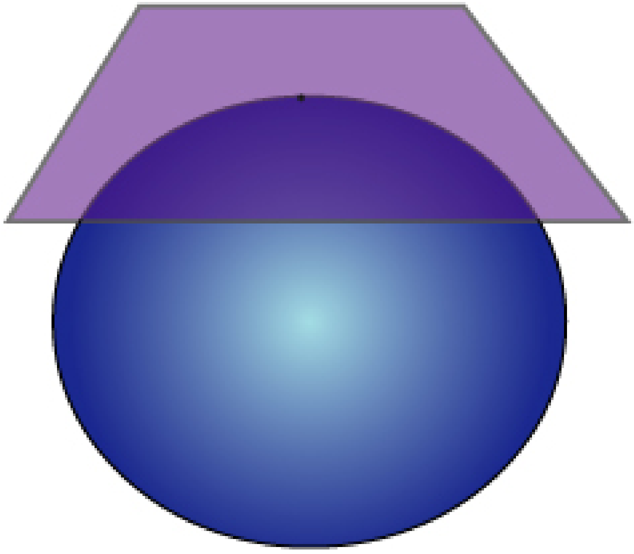

Figure 5.2: A bit of the tangent plane to a 2-sphere at the north pole

Intuitively, the tangent space on a sphere is simpler to grasp than the idea of the tangent space on $\mathbb{R}^2$. If N is the north pole of the 2-sphere, the point $(0, 0, 1)^T$, we can visualise the tangent plane on it as in figure 5.2.

The tangent plane doesn't get confused with the plane itself as it would for some open set $U \subseteq \mathbb{R}^2$. For the whole tangent space, we would have to draw tangent planes to every point on the sphere, and the trouble with this is that they would intersect each other. That is a result of trying to embed the tangent space in $\mathbb{R}^3$, which we can guess ought to be impossible, since the tangent space $T(S^2)$ has to be a four dimensional manifold.

The problem then becomes one of defining the tangent space at a point of a smooth differential manifold, given that we do not have it embedded in $\mathbb{R}^n$. We need to do this because as yet we do not in any case have the idea of the manifold being smoothly embedded properly defined.

We have to do this through charts. Recall the definition of tangency from early in chapter three:

Definition 5.4.1. Two maps $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are *tangent* at $\boldsymbol{a} \in \mathbb{R}^n$ iff $f(\boldsymbol{a}) = g(\boldsymbol{a})$ and

$$\lim_{\|\boldsymbol{h}\| \to 0} \frac{\|f(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a}+\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$$

And the definition of differentiable:

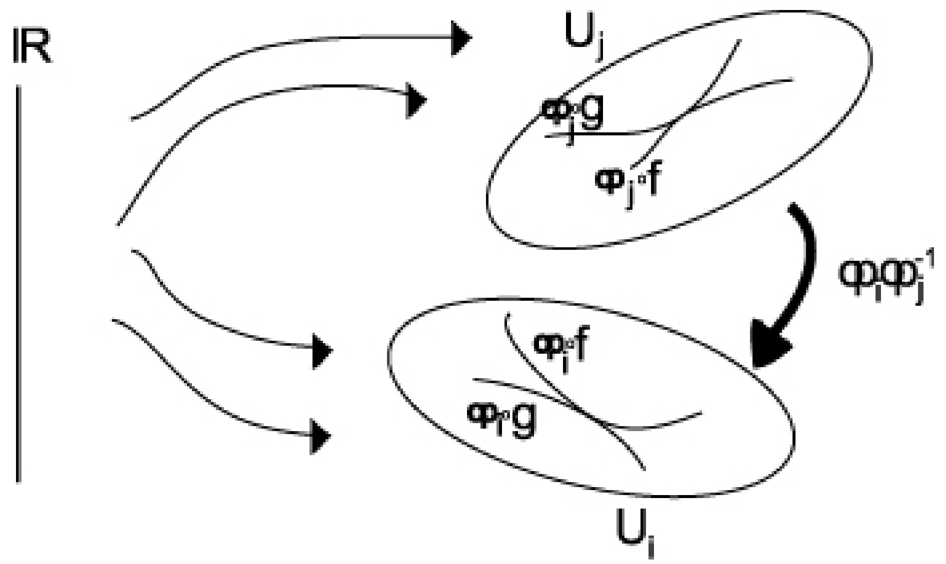

Figure 5.3: Tangency is chart independent

Definition 5.4.2. If the map $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is tangent to an affine map at $\boldsymbol{a} \in \mathbb{R}^n$, then f is said to be *differentiable* at $\boldsymbol{a}$, and the linear part of the affine map is called the *derivative* of f at $\boldsymbol{a}$.

Now given continuous maps $f, g : \mathbb{R} \longrightarrow M$ with $f(0) = g(0) = \boldsymbol{a}$, and a chart $\phi_j : U_j \longrightarrow \mathbb{R}^m$ which has $\boldsymbol{a} \in U_j$, we can decide whether f and g are tangent at 0 by testing to see if $\phi_j \circ f$ is tangent to $\phi \circ g$ at 0, since both are just continuous maps from $\mathbb{R}$ to $\mathbb{R}^m$. We can also test to see if there is an affine map which is tangent to both. If so, we could say that f and g are both differentiable with respect to the chart ϕ_j and have the same derivative. What however about the choice of ϕ_j? $\boldsymbol{a}$ may well be in the domain of other charts. Let us suppose that it is in the domain U_i of $\phi_i : U_i \longrightarrow \mathbb{R}^m$. Since we have that $\phi_i \circ \phi_j^{-1}$ is a diffeomorphism by the assumption of a differentiable structure on M, we have the diagram of figure 5.3

We have by the assumption that $\phi_j \circ f, \phi_j \circ g$ are tangent to some common affine map $h : \mathbb{R} \longrightarrow \mathbb{R}^m$ at 0. By the chain rule it follows that $\phi_i \circ \phi_j^{-1}$ composed with all these three maps takes the derivative to the derivative of $\phi_i \circ \phi_j^{-1}$ at $\boldsymbol{a}$ composed with h, and that this is the derivative of both $\phi_i \circ \phi_j^{-1} \circ \phi_j \circ f$ and of $\phi_i \circ \phi_j^{-1} \circ \phi_j \circ g$, that is, f and g are both differentiable with respect to the chart ϕ_i and have the same derivative.

Thus we have shown:

Proposition 5.4.1. *Tangency equivalence and differentiability of maps from $\mathbb{R}$ to a smooth manifold M are independent of the choice of chart.* □

Exercise 5.4.1. Prove that the same holds true for maps from M to $\mathbb{R}$.

The requirement that we deal with maps from $\mathbb{R}$ into the manifold is sometimes inconvenient, and we will often require only that the maps be defined on some interval containing 0. Obviously this will suffice for the definition of tangency, and since any open interval in $\mathbb{R}$ is homeomorphic to the whole of $\mathbb{R}$ by a map which is smooth and has a smooth inverse this is not a restriction.

Remark 5.4.1. Note that although it makes sense to say that f and g, maps from $\mathbb{R}$ to M, are differentiable, it makes less sense to say what the derivative actually *is*. Any choice of affine approximation can only hold good in a particular chart: change the chart and the best affine approximation is replaced by another. We get around this by the approach Alexander used on the Gordian knot, we define the derivative of f to be the entire (tangency) equivalence class of maps from $\mathbb{R}$ to M which are tangent to f at 0. In any particular chart there will be some particular affine map from $\mathbb{R}$ to $\mathbb{R}^m$. We can think of the linear part of this as a little arrow in $\mathbb{R}^m$. And I can also think of it as a little arrow in the tangent space when M is suitable embedded in $\mathbb{R}^n$ as we shall soon see.

Definition 5.4.3. A *tangent arrow* at $\boldsymbol{a} \in M$ is the class of all maps $f : \mathbb{R} \longrightarrow M$ which have $f(0) = \boldsymbol{a}$ and which are such that $\phi_j \circ f$ is tangent to some affine map $h : \mathbb{R} \longrightarrow \mathbb{R}^m$ at 0 for any chart ϕ_j containing $\boldsymbol{a}$ in its domain.

Remark 5.4.2. The tangent arrows at $\boldsymbol{a}$ can be thought of as velocity vectors for trajectories through $\boldsymbol{a}$.

Proposition 5.4.2. *The tangent arrows at $\boldsymbol{a} \in M$ comprise a vector space of dimension m.*

Proof:

We scale a tangent arrow at $\boldsymbol{a}$ by observing that in any chart, we can take the scalar multiple of the linear part of the affine representative of the arrow. Likewise we can add the linear parts to get another affine map. This new affine map is itself the representative, with respect to the chart, of a tangency equivalence class. But the tangency equivalence class does not depend on the chart. Hence scalar multiplication and addition of tangent arrows is well defined. Since we are adding linear maps from $\mathbb{R}$ to $\mathbb{R}^m$, the dimension of the space of tangent arrows in m, and all the axioms for a vector space are satisfied. □

Remark 5.4.3. This means we can regard $TM_{\boldsymbol{a}}$ as having a hausdorff topology; just take a basis for it and map it to $\mathbb{R}^m$ by an isomorphism, then take the standard norm in $\mathbb{R}^m$, the derived metric and the derived topology. Move it across to $TM_{\boldsymbol{a}}$. It is a fact that for finite dimensional vector spaces there is essentially only one topology which makes all linear maps continuous and which makes all one dimensional linear subspaces homeomorphic to $\mathbb{R}$ (with its usual topology).

Exercise 5.4.2. Define a *cotangent* space at $\boldsymbol{a} \in M$ by looking at tangency equivalence classes of maps from M to $\mathbb{R}$. Show that there is a well defined addition and scaling of *cotangents* which also form a linear space of dimension m. Show that the space of cotangents at $\boldsymbol{a}$ is dual to the space of tangent arrows at $\boldsymbol{a}$.

Definition 5.4.4. If M is a smooth differential manifold, I define $TM_{\boldsymbol{a}}$ for any $\boldsymbol{a} \in M$ to be the space of tangent arrows at $\boldsymbol{a}$. I define

$$TM = \bigcup_{\boldsymbol{a} \in M} TM_{\boldsymbol{a}}$$

TM is called the *tangent bundle* of M

Proposition 5.4.3. *TM is a locally trivial vector bundle and a manifold of dimension $2m$*

Proof:

The bundle is the triple (TM, π, M) with the map $pi : \bigcup TM_{\boldsymbol{a}} \longrightarrow M$ taking every element of $TM_{\boldsymbol{a}}$ to $\boldsymbol{a}$. The fibre is therefore a vector space of dimension m. It is locally trivial because over the domain of any chart $\phi_j : U_j \longrightarrow V_j \subseteq \mathbb{R}^m$ there is a bundle homeomorphism between the restriction of TM to U_j and $V_j \times \mathbb{R}^m$. This is just a matter of sending $\boldsymbol{a}$ in the base space to $\phi_j(\boldsymbol{a})$ and choosing a basis for the space of tangent arrows to map the fibre to $\mathbb{R}^m$. So TM is a locally trivial vector bundle.

To show it is a manifold of dimension $2m$ we have to supply charts from it to open subsets of $\mathbb{R}^{2m}$. We can use the basis we provided for $TM_{\boldsymbol{a}}$ to map $U_j \times TM_{\boldsymbol{a}}$ to $V_j \times \mathbb{R}^m$, going by ϕ_j on the first component and by the isomorphism of vector spaces on the second. It is easy to verify that this gives an atlas of charts for TM. $\square$

Exercise 5.4.3. Verify the last remark.

At the cost of some abstraction which will probably make you uneasy at first, we have produced a formal way of defining what we mean by a tangent arrow at a point of a manifold. Some calculations provide a soothing reassurance that the abstraction is nothing to be worried about.

Exercise 5.4.4. The exponential map

$$t \rightsquigarrow \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}$$

for $t \in (-\pi, \pi)$ has an inverse which is a chart on S^1, and the same map with $t \in\ (0, 2\pi)$ likewise has an inverse which is a chart. The domains of the two charts cover S^1. If the charts are ϕ_1, ϕ_2 respectively show that they define a differential structure on S^1. Define a tengent vector to S^1 at the north pole of the circle by specifying a map from $(-0.1, 0.1)$ to S^1 which takes 0 to the north pole.

5.5 Smooth Maps

If $M_1 \xrightarrow{f} M_2$ is a map between manifolds, and if $\boldsymbol{v} : (-\delta, \delta) \longrightarrow M_1$ is a map sending 0 to $\boldsymbol{a} \in M_1$ which defines a tangent arrow on M_1, then $f \circ v$ should, in any reasonable world, define a tangency class, that is a tangent arrow on M_2 In such a case we can say that f is differentiable at $\boldsymbol{a}$. The idea is straightforward, all that remains is to beat the world into being reasonable.

Exercise 5.5.1. Write down a map from S^1 to S^2 which you feel ought to be a differentiable map and then find an atlas of two charts for each space with respect to which the map *is* differentiable.

What works for differentiable also works for infinitely differentiable, i.e. smooth maps on smooth manifolds.

In the light of the last exercise we can say that $f : M_1 \longrightarrow M_2$, a map between manifolds, is smooth at $\boldsymbol{a} \in M_1$ with respect to charts $\phi_1 : U_1 \longrightarrow R^{m_1}$ for $U \subseteq M_1$ and $\boldsymbol{a} \in U$ and $\psi_1 : V_1 \longrightarrow \mathbb{R}^{m_2}$ for $V \subseteq M_2$ and $f(\boldsymbol{a}) \in V$, iff $\psi_1 \circ f \circ \phi_1^{-1}$ is smooth at $\phi_1(\boldsymbol{a})$. Had we chosen different charts ϕ_2 and ψ_2 we would certainly have had $\psi_2 \circ \psi_1^{-1}$ and $\phi_2 \circ \phi_1^{-1}$ smooth maps, so f is smooth at $\boldsymbol{a}$ whatever (compatible) charts we use. Consequently we can say whether or not f is smooth at $\boldsymbol{a}$ by any choice of chart, and the smoothness is independent of our choice. Since we can do this at every point of M_1 we can tell whether any map is smooth. We could have taken any $\boldsymbol{v} : (-\delta, \delta) \longrightarrow M_1$ to define the tangency class of an arrow (with $\boldsymbol{v}(0) = \boldsymbol{a}$) and then $f \circ \boldsymbol{v}$ defines a tangency class at $f(\boldsymbol{a})$, and providing that when we carry this to maps from $\mathbb{R}^{m_1}$ to $\mathbb{R}^{m_2}$ everything is differentiable, this defines what the derivative of f does to the tangency class.

We therefore have a category of smooth manifolds and smooth maps between them.

5.6 Vector Fields on Manifolds

I can define the *tangent functor* just as I did for maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ by saying what $Tf : TM_1 \longrightarrow TM_2$ does:

$$(\boldsymbol{a}, \boldsymbol{v}) \ \in TM_{1\boldsymbol{a}}, \qquad Tf(\boldsymbol{a}, \boldsymbol{v}) = (f(\boldsymbol{a}), f \circ \boldsymbol{v}) \in T_{2f(\boldsymbol{a})}$$

where $\boldsymbol{v}$ is a tangency class of functions from a neighbourhood of 0 in $\mathbb{R}$ to $\boldsymbol{a} \in M_1$ and where I compose each element v of the class $\boldsymbol{v}$ with f to get a tangency class in M_2. This gives a particular linear map from $TM_{1\boldsymbol{a}}$ to $TM_{2f(\boldsymbol{a})}$ which could be called $Df(\boldsymbol{a})$ or $df_{\boldsymbol{a}}$ or $f_\star(\boldsymbol{a})$ according to taste. All these forms occur in the literature. It takes the tangency class $\boldsymbol{v}$ containing the map v to the tangency class of maps tangent to $f \circ v$.

Exercise 5.6.1. Prove this is a linear map from the fibre over $\boldsymbol{a}$ to the fibre over $f(\boldsymbol{a})$. Maps between locally trivial vector bundles which are linear on the fibres are called vector bundle maps. They are the right maps in the category of locally trivial vector bundles.

Definition 5.6.1. The map $f_\star(\boldsymbol{a})$ is called the *derivative* of f at $\boldsymbol{a}$.

Exercise 5.6.2. Show that when M_1 and M_2 are open subsets of $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$ respectively, the derivative in the new sense is the same as the derivative in the old sense.

Remark 5.6.1. You might have some entertainment out of preparing an imaginary syllabus for first year students in an ideal university, where instead of defining the derivative of maps from $\mathbb{R}$ to $\mathbb{R}$ in the usual way you first define the tangent bundle and then the tangent functor just as we have defined them for manifolds, and then define the derivative this way. Imagining explaining this to first year students is a good way of clarifying your ideas. Since the tangent bundle is just $\mathbb{R}\times\mathbb{R}$ which you can draw, it is not particularly difficult to make the ideas fairly transparent. The chain rule still assures us that the Tangent Functor really is a functor from the category of smooth manifoldsm to itself.

Exercise 5.6.3. Show that the tangent functor takes the identity to the identity and verify that the chain rule assures us that T is a functor.

Exercise 5.6.4. There is a nice smooth map from the unit circle to itself which doubles the angle from the starting point, say $(1,0)^T$, and hence wraps S^1 around itself twice. Show that such a map f is differentiable, and that tangent vectors get doubled in length by the map Tf.

Remark 5.6.2. Note that in going from first year calculus on $\mathbb{R}$ to second year calculus on $\mathbb{R}^n$, what really happened was that everything reduced to repeated ordinary one dimensional differentiations or integration. In going from doing Calculus on flat spaces like $\mathbb{R}^m$ to manifolds, again we reduce it to what we have done before. This time we choose charts to get back to $\mathbb{R}^m$, or open subsets of it. So there are no calculations that are essentially new, we just have to deal with a stack of them. Some issues, however, remain to be clarified.

Exercise 5.6.5. With the differential structure of exercise 5.4.4, define a *unit* tangent vector at the north pole on S^1. Since you can draw this tangent vector quite easily and specify its length, this must be possible. Would it still be a unit tangent vector in a different chart if I had chosen some other atlas?

Exercise 5.6.6. Specify, using the differential stucture above, the vector field on S^1 which has a unit arrow at each point and gives rise to the rotation flow in the positive sense on S^1. How does the definition of unit vector depend on the Atlas? How is it that we can talk about a unit vector field on the circle S^1 and yet the definition of what vectors are unit vectors seems to depend on the chart? How is it that we can decide when any two tangent vectors at different places on S^1 are the same length in a way that does not seem to depend on charts? How much sense would it make on S^2?

You can see that we seem to have some strange problems here. We have proved that the tangent space at a point $\boldsymbol{a}$ of a manifold M is a vector space, but we have not proved that it is a normed vector space. This is because it isn't, not in any unique way. So any choice of a unit vector or more generally a basis would seem to be arbitrary. Moreover, the tangent space at any other point would have to have a quite separate basis provided for it, since the tangent bundle is just a union of the tangent spaces at points. Given a chart, we can certainly say that over the domain of the chart,U_i, the tangent bundle is the trivial bundle with fibres all the same space, the space of arrows in the open set $\phi_i(U_i) \subseteq \mathbb{R}^m$. So for points $\boldsymbol{a}, \boldsymbol{b}$ of M both in U_i we can at least assign some meaning to two arrows, one at each point, being 'the same length' or 'the same direction'. Unfortunately this depends on the chart: if both points are also in U_j and $\phi_j : U_j \longrightarrow \mathbb{R}^m$ is a different chart, then we do not have that the arrows at $\boldsymbol{a}$ and $\boldsymbol{b}$ are the same length in both charts.

Example 5.6.1. Take two charts for the right hand half of S^1, one of which, ϕ_1, sends the point on S^1 making an angle θ with respect to the point $(1,0)^T$ with the usual sense to the real number θ, and the other, ϕ_2 sends the same point to $\sin(\theta)$. Now look at a 'unit vector' at the point $\boldsymbol{a} = (1,0)^T$ in these coordinate systems. We have that the map $v_{\boldsymbol{a}} : (-0.1, 0.1) \longrightarrow \mathbb{R}^2$ given by

$$t \rightsquigarrow \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}$$

when composed with ϕ_1 gives the identity map, $\phi_1 \circ v_{\boldsymbol{a}}(t) = t$ which means that we have a unit vector on the circle at the point $\boldsymbol{a}$. If we now look at a tangent at the point $\boldsymbol{b} = (1/\sqrt{2}, 1/\sqrt{2})$ by defining

$$v_{\boldsymbol{b}}(t) = \begin{bmatrix} \cos(t + \pi/4) \\ \sin(t + \pi/4) \end{bmatrix}$$

then we easily see that $\phi_1 \circ v_{\boldsymbol{b}}(t) = t + \pi/4$ and the derivative at $t = 0$ again gives a unit vector.

If we take $\phi_2 \circ v_{\boldsymbol{a}}(t) = \sin(t)$ and the derivative at $t = 0$ again gives a unit vector. However $\phi_2 \circ v_{\boldsymbol{b}}(t) = \sin(t + \pi/4)$ and the derivative at $t = 0$ is $\cos(\pi/4) = \sqrt{2}/2$ which is not a unit vector.

The answer to the puzzle 'what is going on here?' is that we can draw unit vectors to points on the unit circle for two reasons. One is because S^1 is sitting in $\mathbb{R}^2$ and the vectors can be identified with particular tangent vectors in $\mathbb{R}^2$ and these are casually identified with elements of $\mathbb{R}^2$ complete with the usual inner product and norm. In other words, a lot of the machinery we have been tacitly using comes with the particular embedding of S^1 in $\mathbb{R}^2$ and is not in fact a property of the manifold itself. The second reason is that it is possible to show that the space TS^1 is diffeomorphic to $S^1 \times \mathbb{R}$, with a diffeomorphism that is a vector bundle map, that is, linear on the fibres.

This gives us a way of defining a constant unit vector field on S^1. Just take it that we can map $\boldsymbol{a}$ to, for example, $(\boldsymbol{a}, 1) \in S^1 \times \mathbb{R}$. Of course, this depends on the particular bundle diffeomorphism between TS^1 and $S^1 \times \mathbb{R}$.

However we cannot define a unit vector field on S^2 despite the embedding in $\mathbb{R}^3$ which we usually have as part of the definition. Try it if you don't believe me! And if it makes sense to move vectors around while keeping their length the same it, this is definitely a property of the embedding or of a chart.

Thus we need to be a bit careful when we use our geometric intuitions here. Nevertheless you can see that we can define a vector field on a manifold in an intuitively satisfactory manner, and when the manifold is embedded in $\mathbb{R}^n$, it is easy to visualise for $n \leq 3$.

Definition 5.6.2. A *smooth vector field* on a smooth manifold M is a smooth section of the tangent bundle TM, that is a map V from M to TM which has the property that $\pi \circ V$ is the identity on M where $\pi : TM \longrightarrow M$ is the projection that sends $(\boldsymbol{a}, \boldsymbol{v})$ to $\boldsymbol{a}$.

Requiring V to be smooth is both sensible and intelligible, although we could look at non-smooth vector fields if we felt an overpowering urge to do so. I don't.

Thus a vector field assigns to each point $\boldsymbol{a} \in M$ a pair consisting of the point $\boldsymbol{a}$ and a tangent arrow $\boldsymbol{v}$ at $\boldsymbol{a}$. It makes perfectly good sense when the manifold is some open subset $U \subseteq \mathbb{R}^m$, and gives a more sensible way of specifying vector fields than the traditional one of just giving a map from U to $\mathbb{R}^m$.

We had rather a lot of different ways of thinking about vector fields on $\mathbb{R}^m$ and we now look to see how these work on Manifolds.

Definition 5.6.3. A *smooth diffeomorphism* between manifolds M_1 and M_2 is a smooth map $M_1 \stackrel{f}{\longrightarrow} M_2$ which has a smooth inverse. It is just an isomorphism in the category of smooth manifolds and smooth maps. From now on I shall call it just a diffeomorphism.

Definition 5.6.4. $D(M)$ is the set of diffeomeorphisms from M to itself. With composition as the binary operation it is a group.

Exercise 5.6.7. Prove the last remark.

Definition 5.6.5. A *flow* on a smooth Manifold M is a group homomorphism from the topological group $\mathbb{R}, +$ to the group $D(M)$ of diffeomeorphisms of M to itself.

Remark 5.6.3. This is just a group action of $\mathbb{R}$ on the manifold.

Remark 5.6.4. It may cross your mind to wonder whether $D(M)$ can be made into a *topological* group and the answer is yes, and this is a very good thing to do, but I shan't do it.

A flow $m : \mathbb{R} \longrightarrow D(M)$ gives, for each $\boldsymbol{a} \in M$, and each $t \in \mathbb{R}$, $m(t, \boldsymbol{a}) \in M$ by a mild abuse of language, with $m(0, \boldsymbol{a}) = \boldsymbol{a}$. Writing this as $m_{\boldsymbol{a}} : \mathbb{R} \longrightarrow M$

by another mild abuse of language we have that since m is smooth, we can take it that $m_{\boldsymbol{a}}$ determines a tangent arrow at $\boldsymbol{a}$ and since this holds at every $\boldsymbol{a} \in M$ we have a smooth vector field specified by differentiating the flow.

Conversely, any smooth complete vector field on a manifold determines a flow. This is a slight problem because I haven't said what a complete vector field is on a manifold, so I need to state a more modest result:

Proposition 5.6.1. *If M is a smooth manifold and if $V : M \longrightarrow TM$ is a smooth vector field on M, then for any point $\boldsymbol{a} \in M$ there is an interval $(-\delta, \delta) \subseteq \mathbb{R}$ containing zero and a map*

$$m : (-\delta, \delta) \longrightarrow D(M)$$

which is a homomorphism where that makes sense, that is, $m(s+t) = m(s) \circ m(t)$ for $-\delta < s, t, s+t < \delta$,

and

$$m_{\boldsymbol{a}} : (-\delta, \delta) \longrightarrow M$$

is in the tangency class of $V(\boldsymbol{a})$

Proof:

If you believe the Straightening Out Theorem from the last chapter,you should believe this, since all we have to do is take a chart which straightens the vector field out in the neighbourhood of a point where the vector is non-zero. This takes the vector field to a constant one with vectors of length one and a rather trivial solution, and the inverse of the chart takes the solution back to a (local) solution on M. The curves go in with their parametrisations, so join up properly: if it takes time s to get from $\boldsymbol{a}$ to $\boldsymbol{b}$ and time t to get from $\boldsymbol{b}$ to $\boldsymbol{c}$ then it takes time $s + t$ to get from $\boldsymbol{a}$ to $\boldsymbol{c}$ even if a and c are in the domains of different charts. If the vector field is zero then of course the global constant solution occurs. The only nicety here is that two different chart inverses could take the local line solution back to different curves which intersected but which disagreed at one or more points. This can easily happen in a non-hausdorff space such as $\mathbb{W}eird\mathbb{R}$, where one solution curve goes through one of the two zeros and another goes through the other one. It is impossible in a hausdorff space: if c_1 and c_2 are two such integral curves passing through $\boldsymbol{a} \in M$ and the intersection of the domains is I, let K be the set of $t \in I$ such that $c_(t) = c_2(t)$. Then it is easy to show that K is a subset of I which is both closed in I and open in I and hence is equal to I. See Abraham *Foundations of Mechanics* p39. □

This solution exists and is unique but is only a local solution, and we have no guarantee that it will extend to a solution for all time, that will determine a flow. If it can be so extended, we say the vector field is *complete*. It is then rather tautologous to observe that every smooth, complete vector field on a smooth manifold gives rise to a unique flow on the manifold.

5.7 Vector Fields as Operators

Since the definition of the operation of a smooth vector field on the space of smooth functions to $\mathbb{R}$ was local, it goes over to vector fields on manifolds with only minor changes. If $\mathcal{F}(M)$ denotes the space of smooth maps from M to $\mathbb{R}$, we want to take any function in it and at each point of M differentiate it in the direction of the vector at that point and multiply by the length of the vector to get a number. The problem is that the vector on a manifold cannot be said to have a length, since the vector space of tangents is not normally given an inner product or norm. This is not a problem. The tangent arrow $\boldsymbol{v}$ at $\boldsymbol{a} \in M$ is a tangency equivalence class of smooth maps from $(-\delta, \delta)$ to M which takes 0 to $\boldsymbol{a}$. Take any member of this class, v, and compose it with $f \in \mathcal{F}(M)$ to get $f \circ v : (-\delta, \delta) \longrightarrow \mathbb{R}$ and differentiate it at 0 to get a number. Then verify that (a) the numbers will be the same no matter which v we choose and (b) when we take M to be $\mathbb{R}^m$ we get the same result as we got for the action of a vector field V on a function $f \in \mathcal{F}(\mathbb{R}^m)$

Definition 5.7.1. $\mathcal{X}(M)$ denotes the set of smooth complete vector fields on M and $\mathcal{F}(M)$ denotes the set of smooth maps from M to $\mathbb{R}$. Then

$$\forall\ V \in \mathcal{X}(M),\ \forall\ f \in \mathcal{F}(M),\ \forall \boldsymbol{a} \in M,\ Vf(\boldsymbol{a}) = \frac{d}{dt}(f \circ v)\bigg|_0$$

where v is any representative of $V(\boldsymbol{a})$

Exercise 5.7.1. Show that Vf is well defined.

Exercise 5.7.2. Show that this is equivalent to the earlier definition when we choose M to be an open subset of $\mathbb{R}^m$.

Exercise 5.7.3. Show that for any $X \in \mathcal{X}(M)$, and any $f, g \in \mathcal{F}(M)$,

$$X(fg) = fXg + gXf$$

Remark 5.7.1. For algebraists. If R is a ring, then a map $X : R \longrightarrow R$ is said to be a *derivation* iff X is a homomorphism on the additive group and

$\forall\ f, g \in R,\ X(fg) = fXg + gXf$. If you don't know what a ring is and you want to know, look it up. If you do know, verify that $\mathcal{F}(M)$ is a ring under pointwise multiplication and addition of functions.

Exercise 5.7.4. Define the rotation flow of the 2-sphere (regarded as a subset of $\mathbb{R}^3$) as a positive rotation at unit speed around the z-axis. Write down the system of ODEs which has this as the resulting flow. Specify the vector field which corresponds to this system (a) by giving vectors in $\mathbb{R}^3$, (b) by choosing a suitable atlas and specifying the vector field via the atlas and the Straightening Out Theorem and (c) by specifying what the vector field does to smooth maps from S^2 to $\mathbb{R}$. In case (c) test out your result by checking what happens to the function $f(x, y, z) = x^2 + y^2 + z^2$ and some other functions of your own devising.

Exercise 5.7.5. Define a torus by identifying opposite sides of a square and define a vector field on the square that survives the identification process to become one on the torus. What do the vector fields which are *constant* on the square become on the torus? (There are essentially two different types: say what they are.)

For a manifold M and a chart $\phi_j : U_j \longrightarrow \mathbb{R}^m$, we can take the vector $\boldsymbol{e}_1$ at a point $\phi_1(\boldsymbol{a})$ and identify it with any map $v : (-\delta, \delta) \longrightarrow M$ which has $(\phi \circ v)'_0 = 1$. Now if $f : M \longrightarrow \mathbb{R}$ is a smooth map, then we can take the map $f \circ \phi_j^{-1}$ and it makes sense to write

$$\frac{\partial f \circ \phi_j^{-1}}{\partial x^1}$$

for the operation of $\boldsymbol{e}_1$ on f at $f(\boldsymbol{a})$. And in general we can take

$$\frac{\partial}{\partial x^i}, \qquad 1 \leq i \leq m$$

to be a basis for the tangent space- relative to any chart. That is, any chart gives us a basis for the tangent space at each point in the domain of the chart.

Since we have vector fields as operators on manifolds, all the material about Lie Brackets makes sense and the theorem about commuting vector fields X and Y having $[X, Y] = \mathbf{0}$ still holds. In fact we can call the theory a 'local' theory when it holds for any open set $U \subseteq \mathbb{R}^m$, and 'global' when it works on any manifold. Then we can transfer quite a lot of things from local theories to global theories if we verify that the results do not depend on any specific choice of chart. Thus it makes sense to say when a vector field on a manifold is smooth, because this will not depend on any chart, but it makes no sense in general to say that a vector field on a manifold is constant, or that all vectors have the same length, because although this makes sense in $\mathbb{R}^m$, it will depend on the chart, and we can easily find charts which disagree about whether two vectors at different places on the manifold have the same length or the same direction.

Exercise 5.7.6. I define a smooth map $f : M_1 \longrightarrow M_2$ to be a *local diffeomorphism* iff for every point $\boldsymbol{a} \in M_1$ there is an open set in M_1 containing $\boldsymbol{a}$ and $f|U$ is a diffeomorphism onto its image, an open set in M_2. Show that f is a local diffeomorphism iff the map Tf is 1-1 and onto always.

Now we can define a few useful terms:

Definition 5.7.2. A smooth *immersion* of a manifold M_1 in a manifold M_2 is a smooth map

$$f : M_1 \longrightarrow M_2$$

which has Tf 1-1 everywhere.

Definition 5.7.3. A *smooth embedding* is a smooth immersion which is 1-1.

Exercise 5.7.7. Give an example of smooth immersion which is not an embedding.

5.8 Tensor Fields on Manifolds

In order to be explicit and accessible I shall discuss only tensor fields of low rank on manifolds of low dimension, never more than three in either case, and I shall save comments on differential forms for the next chapter. In general suppose the manifold has dimension m.

Recall from Chapter Three or elsewhere, that if V is a (real) vector space then there is a space $V^\star$ of linear maps from V to $\mathbb{R}$. (In second year I wrote elements of $\mathbb{R}^n$ as vertical columns and elements of $\mathbb{R}_n = \mathbb{R}^{n\star}$ as horizontal rows, and the effect of the latter on the former then came out as matrix multiplication.) In what follows, $V = V_{\boldsymbol{a}}$ is going to be the space of tangents at a point $\boldsymbol{a}$ on a manifold. Note that I must be careful to avoid assuming that V is $\mathbb{R}^m$ for any m, since that space comes with an inner product, distinguished basis elements and other such baggage. My space $V_{\boldsymbol{a}}$ has elements which are tangency classes of maps from some open interval containing 0 to the manifold M, which take 0 to $\boldsymbol{a}$. I can add them by means of a chart and get a tangency class for the sum, and this will not depend on which chart I used or which elements of the tangency classes. So although I know that, for example $V_{\boldsymbol{a}}$ and $V_{\boldsymbol{b}}$ are isomorphic vector spaces, I have no particular isomorphism provided. So I cannot compare vectors in one with vectors in the other.

As well as $V^\star$, there is a space of *bilinear* maps from $V \times V$ to $\mathbb{R}$, and this is a vector space also, having dimension m^2 Without revisiting Chapter Three in too much detail, I point out that I need to specify for each combination of pairs of basis elements $\boldsymbol{e}_i, \boldsymbol{e}_j, \quad 1 \leq i, j \leq m$ what a bilinear map does, and then I have specified it completely; I can work out its value on any pair $(\boldsymbol{u}, \boldsymbol{v}) \in V \times V$.

I may also be interested in the vector space of bilinear maps from $V^\star \times V$ to $\mathbb{R}$, the vector space of bilinear maps from $V \times V^\star$ to $\mathbb{R}$ and the bilinear maps from $V^\star \times V^\star$ to $\mathbb{R}$. These all have the same dimension and they are all different kinds of fish and rabbits. Since there is however a natural isomorphism between the space $\mathrm{Bil}(U \times V, W)$ of bilinear maps from $U \times V$ to W for any three real vector spaces U, V, W and the vector space $\mathrm{Lin}(U, \mathrm{Lin}(V, W))$ of linear maps from U to the space of linear maps from V to W, and since we can take the double dual of V, $V^{\star\star}$, to be naturally isomorphic to V, we can take the special case of the bilinear maps from $V \times V^\star$ to $\mathbb{R}$ and regard it as the same as the space of linear maps from V to V.

Exercise 5.8.1. Show that 'there is however a natural isomorphism between the space $\mathrm{Bil}(U \times V, W)$ of bilinear maps from $U \times V$ to W for any three real vector spaces U, V, W and the vector space $\mathrm{Lin}(U, \mathrm{Lin}(V, W))$ of linear maps from U to the space of linear maps from V to W'.

Exercise 5.8.2. Show that 'we can take the double dual of V, $V^{\star\star}$, to be naturally isomorphic to V'.

An *inner product* on a vector space V is a bilinear map from $V \times V$ to $\mathbb{R}$ which is symmetric and positive definite, that is to say

$$\forall\ \boldsymbol{u}, \boldsymbol{v} \in V, h(\boldsymbol{u}, \boldsymbol{v}) = h(\boldsymbol{v}, \boldsymbol{u}) \ \text{ and } \ h(\boldsymbol{u}, \boldsymbol{u}) \geq 0 \ \text{ and } \ h(\boldsymbol{u}, \boldsymbol{u}) = 0 \Rightarrow \boldsymbol{u} = \boldsymbol{0}$$

It follows that it is a particular kind of tensor of rank 2. Similarly there is a symmetric positive definite tensor on the dual space, a bilinear map from $V^\star \times V^\star$ to $\mathbb{R}$. We can now envisage a tensor field on a manifold having such tensors attached to each point of the space. Such a thing is called a *Riemannian Metric Tensor*, although it would be better to call it a Riemannian Inner Product tensor field. Certainly however we can get a local metric from it by the usual means. In the following example I have been sloppy with notation so as to relate it to what you already know more easily.

Example 5.8.1. On a suitable open set in $\mathbb{R}^2$ I define a new metric by saying that locally it is given by the matrix

$$\begin{bmatrix} 1+xy & 0 \\ 0 & x^2+y^2 \end{bmatrix}$$

Find the length of the curve along the parabola $y = x^2$ from the origin to $x = y = 1$ in this metric.

Solution:

The ordinary formula for the curve is that it is $\int_{\boldsymbol{c}} d\ell$ where $\boldsymbol{c}$ is the curve and $d\ell^2 = dx^2 + dy^2$ is the 'infinitesimal path length'. We can write this as

$$d\ell^2 = [dx, dy] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix}$$

Our new and improved inner product changes from place to place but it gives rise to a norm just as the old one does, and it is a norm on the cotangent space. We therefore have

$$d\ell_1^2 = [dx, dy] \begin{bmatrix} 1+xy & 0 \\ 0 & x^2+y^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix}$$

for the new way of measuring the differential path length and so the length of the path along the parabola, with $x = t, y = t^2$ is

$$\int_0^1 \sqrt{(1+t^3).1 + (t^2+t^4)(4t^2)}\; dt \approx 1.49958$$

where the approximation is done using Mathematica. This compares with about1.29361 using the standard metric.

Example 5.8.2. Find the path length of the spiral $r = \theta$ for $0 \le \theta \le 2\pi$ in the metric on $\mathbb{R}^2$ given by

$$d\ell_2^2 = [d\theta, dr] \begin{bmatrix} r^2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d\theta \\ dr \end{bmatrix}$$

Solution: This is just the usual metric on $\mathbb{R}^2$ disguised by using polar coordinates since $d\ell_2^2 = (rd\theta)^2 + (dr)^2$ is the usual way of calculating the 'infinitesimal path length' and the answer is

$$\int_0^{2\pi} \sqrt{t^2+1}\; dt \approx 21.2563$$

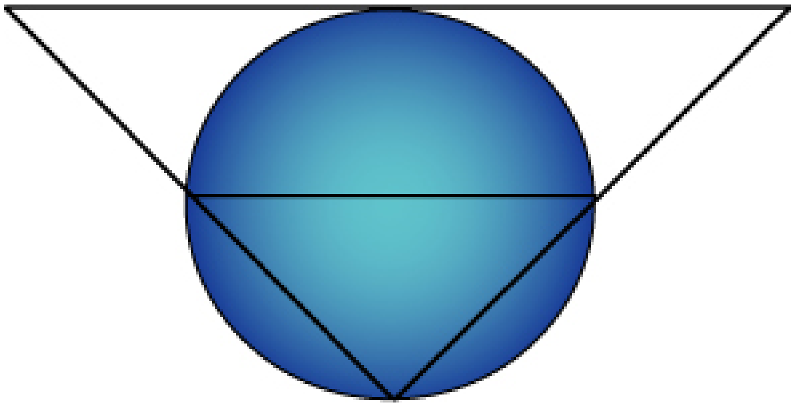

Figure 5.4: Putting a metric on the Northern Hemisphere

This compares with $2\sqrt{2}\pi \approx 8.885765876$ in the euclidean metric on the θ, r space. Well, in that space the curve is a straight line.

Exercise 5.8.3. Find the path length of the above spiral using the metric given by

$$d\ell_2^2 = [d\theta, dr] \begin{bmatrix} r^2 & 0 \\ 0 & r^4 \end{bmatrix} \begin{bmatrix} d\theta \\ dr \end{bmatrix}$$

Exercise 5.8.4. There is a rather obvious metric on S^2, as it comes with the standard embedding in $\mathbb{R}^3$, which measures the distance between two points along great circles. This is how airlines measure the distance between the capital cities of the world, and it makes a fair amount of sense to crows, air-passengers, and the blokes who have to check that there is enough fuel in the tanks to get you from here to wherever. Suppose now I give you the disk D^2 and tell you that I slapped it on the northern hemisphere of S^2 so that it went over stereographically after first having its radius doubled, as in figure 5.4, where we just draw a line from the South pole through the northern hemisphere until it meets a point on the disc sitting in the tangent space to the North pole. Then the two points where the line cuts, on the northern hemisphere and on the disc of radius 2 units, are mapped together.

This gives a chart on the northern hemisphere. Write its inverse out as an explicit function from D^2 to $\mathbb{R}^3$. Now find a metric tensor field on the disc D^2 which maps to the crow-and-airline flying metric on the 2-sphere by the inverse of the chart.

Exercise 5.8.5. Check the last result by computing the path length of a straight line from the centre of the disc to its boundary in this metric. It had better be $\pi/2$ since it goes to a quarter of a great circle of radius 1.

Exercise 5.8.6. Now take the parametrisation of the 2-sphere which takes the rectangle with s going from 0 to 2π and t going from -1 to 1, and first wraps it around the cylinder by $x = \cos(s),\ y = \sin(s),\ z = t$, and then 'shrink wraps' the cylinder onto the sphere by

$$x = \sqrt{1-t^2}\cos(s)\ y = \sqrt{1-t^2}\sin(s), z = t$$

This gives a chart the domain of which covers S^2 except at the poles and the meridian $s = 0$. Calculate the metric tensor for the crow-airline metric on S^2 in this coordinate system. Geography atlases use this chart, and it plays hell with Greenland. You have a chance to see precisely how much hell it plays.

Remark 5.8.1. In the last two or three exercises, you have been calculating things by taking charts or local coordinate systems. You should see however that the underlying object, a metric tensor, exists on the manifold in a sense which is independent of any particular chart, and the charts are matters of convenience (or inconvenience as the case may be). The metric tensor is a section of a tensor bundle, which is at each point the set of linear maps from $V^s tar \times V^\star$ to $\mathbb{R}$, where V is the space of tangent arrows to the manifold at that point and $V^\star$ is the cotangent space, identified with the dual space to V. This may make your head spin at first, but as you can see the calculations are nothing special.

Exercise 5.8.7. For those who have read some of the chapter on Algebraic Prelimiaries, Chapter 3, and are feeling brave, rewrite the above calculations in the correct form, explaining exactly what dx and dy are.

Exercise 5.8.8. For those feeling really brave, using the metric of exercise 5.8.4 on D^2, find the path between $(\theta = 0, r = 1/2)$ and $(\theta = 1, r = 2/3)$ which has the shortest path length. Such paths are called *geodesics* and when the metric tensor is constant in $\mathbb{R}^m$, these come out to be straight lines.

5.9 Summary

We can define a manifold as a hausdorff space which is locally $\mathbb{R}^m$ for some fixed $m \in \mathbb{Z}^+$, and the implicit function theorem ensures that we get lots of them by putting smooth conditions on n variables for $n > m$. The machinery of atlases of charts or *local coordinate systems* means that we can picture a topological manifold as a collection of open sets in $\mathbb{R}^n$ glued together by charts. By making the gluing process differentiable or smooth. we get differential structure on the manifold. Whereupon differential manifolds are the natural setting for looking at differentiable maps.

The idea of the derivative has been generalised from the old school-day picture of it as a tangent line to a curve. First we went over to thinking of it as the linear part of the best affine approximation to a function, now we think of it as being the tangent functor, which operates as a map on the points and also as a linear map on the tangent space at that point, taking arrows to arrows. A tangent arrow on a manifold is part of a space of all possible arrows, represented by maps from $\mathbb{R}$ into the manifold, the arrow being a tangency equivalence class which, within any chart, is differentiable. Whether two maps are tangent does not depend on the choice of chart, and whether a function is differentiable also does not depend on the chart. On the other hand the norm of a vector does depend on the chart, as does the notion of a vector being moved about on the manifold. The tangent space at a point on a manifold is a vector space of dimension the same as the manifold, so a tangent vector *is* a vector. A vector field on a manifold is a section of the tangent bundle. Just as in the local theory (where manifolds aree open subsets of $\mathbb{R}^m$), the interpretation of a vector field as an operator makes sense.

We also have a space of cotangents, maps from the manifold to $\mathbb{R}$ also forming tangency classes. More generally if $V_{\boldsymbol{a}}$ is the vector space of tangent arrows at a point $\boldsymbol{a}$ of a manifold, we can take the vector space of bilinear maps from $V_{\boldsymbol{a}} \times V^{\star}_{\boldsymbol{a}}$ to $\mathbb{R}$ and this is an example of a tensor bundle over the same point $\boldsymbol{a}$. The fibre is a new vector space (of dimension m^2 where m is the dimension of M) and we can take a section of that to get a tensor field, in this case if we ensure that the bilinear map is always symmetric and positive definite we can regard the section of the tensor bundle as giving a local metric on the manifold. This is one example of a tensor field.

Chapter 6

Differential Forms on Manifolds

I am skipping this in 2003, and haven't the time to write it up this year. The basic idea is that we can define tensor operations on each fibre of the tangent bundle, since it's a respectable finite dimensional vector space. That is to say, for $T_{\boldsymbol{a}}M$ the tangent bundle over $\boldsymbol{a} \in M$ it makes sense to call it V and take multilinear maps from $V \times V \times \cdots \times V \times V^\star \times V^\star \times \cdots \times V^\star$ to $\mathbb{R}$, where the number of copies of V is r and the number of copies of $V^\star$ is s and any such map is an (r, s) tensor. This is certainly a vector space, of dimension n^{r+s}, and we can imagine it sitting there over the point a just as V does. In fact we can imagine the whole tensor algebra for every possible r and s, sitting over the point $\boldsymbol{a}$ of the manifold. And we do this for every point $\boldsymbol{a}$ to get the *Tensor Bundle*. We can take sections of the tangent bundle to get vector fields, sections of the cotangent bundle to get covector fields or contravariant vector fields, and in general sections of the tensor bundles as tensor fields on the manifold. Alternating tensors are a particularly significant sort of tensor, and alternating tensor fields are known as differential forms on the manifold.

These ideas are a bit mind boggling at first, but as with vector fields, they correspond to quite natural things as the examples of the metric tensor field shows. Just as in the case of vector fields (a) there is more structure to tensor fields than may at first appear and (b) the best place to meet them first time is probably in the local theory, i.e. for open subsets of $\mathbb{R}^m$. They get globalised to manifolds without any special surprises (we need partitions of unity to get the integration of differential forms over submanifolds working properly) and we get Stokes' Theorem on Manifolds. We can also work on Manifolds with Boundary, examples of which are D^2, the closed unit ball in $\mathbb{R}^2$.

Chapter 7

Differential Topology

7.1 The Fundamental Theorem of Algebra

I remarked that if $\mathbb{R}^n$ and $\mathbb{R}^m$ were homeomorphic, then $n = m$. This is quite hard to prove. However it is easy to see that if $\mathbb{R}^n$ and $\mathbb{R}^m$ are *diffeomorphic* then $n = m$. This follows because a diffeomorphism f has a tangent functor which takes not just $\boldsymbol{a} \in \mathbb{R}^n$ to $f(\boldsymbol{a}) \in \mathbb{R}^m$ but also takes the tangent space at $\boldsymbol{a}$ to the tangent space at $\boldsymbol{b}$ by a linear map, the derivative of f. And the dimension of the tangent space at $\boldsymbol{a}$ is n and the dimension of the tangent space at b is m. And the inverse map f^{-1} also exists and is differentiable and its tangent functor is an inverse to the tangent functor of f. So the linear (tangent) spaces are isomorphic and hence (from elementary linear algebra) have the same dimension. So $n = m$. Alternatively, if there is a diffeomorphism from some open set $U \subseteq \mathbb{R}^n$ to some open set $V \subseteq \mathbb{R}^m$ then the derivative at any point has an inverse and is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ so it must be an isomorphism of vector spaces which ensures they have the same dimension.

I have proved:

Proposition 7.1.1. *Diffeomorphic spaces have the same dimension.*

□

Exercise 7.1.1. Make the above argument rigorous.

Definition 7.1.1. Given $M_1 \xrightarrow{f} M_2$ a smooth map between manifolds of dimensions m_1, m_2 respectively, we say $\boldsymbol{a} \in M_1$ is a *regular point* of f iff rank($f_\star(\boldsymbol{a}) = m_2$), where $f_\star(\boldsymbol{a}) = Df(\boldsymbol{a}) = df_{\boldsymbol{a}}$ is the linear map between the tangent spaces which we have called the derivative of f at $\boldsymbol{a}$.

Exercise 7.1.2. For this to make sense, the map $f_\star$ has to be defined as a linear map between vector spaces, which it is, and we have to believe that the rank of a linear map makes sense. We agree that the rank of a matrix makes sense, and it is not hard to show that it does not change as we change bases for the domain and codomain vector spaces. Which I hope you did in Linear Algebra. If not do it now. Since no charts were used in defining these things, we conclude that the rank of the derivative is a properly defined number.

Definition 7.1.2. If a point $\boldsymbol{a}$ is not a regular point of $M_1 \xrightarrow{f} M_2$ it is said to be a *critical* point of f, and $f(\boldsymbol{a})$ is said to be a critical value.

Remark 7.1.1. This is not different in any interesting way from the case where M_1 and M_2 are in fact open sets of $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$ and for very good reasons. Obviously, at a critical point $\boldsymbol{a}$ we have rank$(f_\star(\boldsymbol{a})) < m_2$

Exercise 7.1.3. Prove the last remark.

Remark 7.1.2. Milnor defines his regular values of f as all those points of M_2 which are not critical values. This includes the points which are not values of f at all, if f is not onto, and although I hesitate to diverge from such an authority as John Milnor, I hesitate even more to follow him on this one. So be warned that Milnor and I differ in the case when the map is not onto.

Let $M_1 \xrightarrow{f} M_2$ be a smooth map between manifolds having the same dimension, m. If $\boldsymbol{b}$ is a point in the image of f which is a regular value, and if M is compact, then the set $f^{-1}(\boldsymbol{b})$ is finite. This is because at each point of $f^{-1}(\boldsymbol{b})$ f is a local diffeomorphism so the set has to be discrete and since $f^{-1}(\boldsymbol{b})$ is closed (because $\boldsymbol{b}$ is) it is compact and so the discrete set is finite. It may of course be empty, which counts as finite since $0 \in \mathbb{N}$. Let $\#(\boldsymbol{b})$ be the number of elements in $f^{-1}(\boldsymbol{b})$. Then there is an open set V containing $\boldsymbol{b}$ and no critical values of f, such that the function $\#(\boldsymbol{y})$ is constant on V.

Exercise 7.1.4. Prove the above.

As a corollary we can derive the Fundamental Theorem of Algebra which says that every polynomial over the complex numbers has a zero:

Proposition 7.1.2. *If $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots a_1 z + a_0$ is a polynomial function from $\mathbb{C}$ to $\mathbb{C}$ with coefficients a_j complex numbers, and if $n \geq 1$ then* $\exists z \in \mathbb{C} : p(z) = 0$

Proof:

Extend p from the complex plane to S^2 by sterographic projection from the plane regarded as the tangent space to the North pole of S^2 to everywhere except the South pole,after first doubling the size of the tangent plane, and then note that if we call this map from $\mathbb{C}$ to S^2 h_+, then define $\tilde{p} : S^2 \longrightarrow S^2$ by

$$\tilde{p}(\boldsymbol{a}) = h_+ \circ p \circ h_+^{-1}$$

on $S^2 \setminus \{\mathbf{S}\}$, where S is the South pole, and $\tilde{p}(S) = S$.

Then $\tilde{p}$ is a smooth map from S^2 to S^2. It is easy to see that this holds for every point except the South pole, since polynomials are smooth and so are h and h^{-1}. To see it is also smooth at the South pole, which can be thought of as the point ∞, look at $1/p$ and make that defined everywhere except at the north pole, and show it is smooth at the South pole, using the same stereographic projection from the South pole with h_-. Note that $h_+ \circ h_-(z) = 1/z$ and observe that the corresponding effect on the sphere S^2 is a reflection in the plane of the equator.

Now the smooth map $\tilde{p}$ must have the number of points $\#\tilde{p}^{-1}(\boldsymbol{y})$ a constant except possibly at the critical values of $\tilde{p}$. These are at the zeros of p', and again since we have a smooth map from S^2 to itself and S^2 is compact, there are only finitely many of these. The rest of S^2 has the function $\#\tilde{p}^{-1}(\boldsymbol{y})$ constant. Moreover the constant, k, can't be zero except in the rather special case where the polynomial is a constant map. In particular, there are k points z for which $p(z) = 0$ unless 0 is a critical value of p, i.e. the north pole is a critical value of $\tilde{p}$. Either way, something gets mapped to $0 \in \mathbb{C}$. $\square$.

Exercise 7.1.5. Fill in the gaps in the above argument until you believe it.

Remark 7.1.3. There are a large number of proofs of the Fundamental Theorem of Algebra and all the simplest are actually topological.

7.2 Sard's Theorem for Manifolds

One possibility for a manifold is that it could be an uncountable number of copies of $\mathbb{R}^n$. Such a thing is evidently rather on the large side. I could even make it connected by having bridges between any pair of copies which also look like $\mathbb{R}^n$, say gluing the left hand half of one to the right hand half of another, where left and right are measured with respect to the x axis.

Exercise 7.2.1. Produce a formal definition of Mike's Monster Manifold and show it *is* a manifold.

I wish to exclude this kind of monster. The following definition helps:

Definition 7.2.1. A manifold is said to be *second countable* iff the topology has a countable base.

Remark 7.2.1. Recall that a set is countable iff it can be put in a 1-1 onto correspondence with $\mathbb{N}$ the set of natural numbers.

Exercise 7.2.2. Show that $\mathbb{R}^n$ has a countable base and that MMM does not. Show that S^n has a countable base.

I have managed to avoid saying anything about any measure on a manifold which corresponds to Lebesgue Measure on $\mathbb{R}^m$, but it should not be hard to believe that the notion of Lebesgue measure zero is intelligible:

Definition 7.2.2. If M is a smooth manifold of dimension m then a subset $A \subset M$ has *Lebesgue measure zero* in M iff the measure of $phi_j(A)$ in $\mathbb{R}^m$ is zero for every chart ϕ_j.

Exercise 7.2.3. Show that this does not depend on the choice of chart.

Exercise 7.2.4. Does this definition make sense for topological manifolds?

The argument for Sard's theorem for maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ then goes over without essential change.

Theorem 7.2.1. Let $M_1 \xrightarrow{f} M_2$ be a smooth map between second countable manifolds. The the set of critical values in M_2 has Lebesgue measure zero.

Exercise 7.2.5. Produce a proof by looking at the local argument of Sard's theorem for subsets of $\mathbb{R}^m$ and using second countability to make it work for the whole manifold.

Exercise 7.2.6. Show that without second countability, Sard's Theorem for manifolds fails.

From now on, all manifolds will be assumed to be second countable.

Definition 7.2.3. A subset A of a topological space X is *dense* in X iff every open set in X contains at least one point of A.

Example 7.2.1. The rationals $\mathbb{Q}$ are dense in $\mathbb{R}$. So are the irrationals. The set of points $(x, y)^T$ in $\mathbb{R}^2$ where both x and y are rational is dense in $\mathbb{R}^2$. You will have used some such fact in order to show $\mathbb{R}^2$ is second countable. I hope.

A useful corollary to Sard's theorem can now be stated:

Proposition 7.2.1. *If* $M_1 \xrightarrow{f} M_2$ *is a smooth map between manifolds which is onto, then the set of regular values of* f *is dense in* M_2

Proof:

If there were an open set in M_2 not containing a regular value it would have to consist entirely of critical values and would have a chart taking it to an open set in $\mathbb{R}_{m_2}$ which has non-zero measure, contradicting Sard's theorem. □

If we take the height function from S^2 to $\mathbb{R}$ we see that the subset of S^2 which has all its points at the same height is a circle of constant latitude, and a submanifold of S^2. This is generally the case:

Proposition 7.2.2. *If* $M_1 \xrightarrow{f} M_2$ *is a smooth map between manifolds of dimensions* m_1, m_2*, and if* y *is a regular value of* f*, then* $f^{-1}(y)$ *is a submanifold of dimension* $m_1 - m_2$*.*

Proof:

This is just the implicit function theorem. □

Exercise 7.2.7. Prove that the above claim is correct by taking charts.

7.3 Manifolds with Boundary

While a smooth $n-k$-dimensional manifold occurs naturally, together with a smooth embedding in $\mathbb{R}^n$, as the kernel of a smooth map $\mathbb{R}^n \xrightarrow{f} \mathbb{R}^k$ which has full rank for its derivative on the kernel, manifolds with boundary arise naturally from inequalities. The simplest examples of manifolds with bounrary are the closed balls in $\mathbb{R}^n$. Thus in particular, D^2 is a manifold with boundary and the boundary is S^1. You should be prepared to believe that the boundary of a manifold with boundary is a manifold (without boundary) of dimension one less than the dimension of the manifold with boundary.

Definition 7.3.1. The space H^m is defined to be the subset of $\mathbb{R}^n$ with $x^1 \geq 0$, with the subspace topology, where $\mathbb{R}^m$ has its usual topology induced from the standard inner product. In algebra:

$$H^n \triangleq \left\{ \begin{bmatrix} x^1 \\ \vdots \\ x^m \end{bmatrix} \in \mathbb{R}^m : x^1 \geq 0 \right\}$$

The *boundary* of H^m is the subset of H^m having $x^1 = 0$.

Definition 7.3.2. A *smooth manifold of dimension m with boundary* is a hausdorff topological space M and a maximal compatible atlas of compatible charts to H^m which take open sets in M to sets open in H^m, such that every $\boldsymbol{x} \in M$ is in the domain of such a chart. If any chart takes such an $\boldsymbol{x}$ to the boundary of H^m, then $\boldsymbol{x}$ is said to be an element of the boundary of M. Points of M not on the boundary are said to be *interior* points.

Remark 7.3.1. . If any chart takes $\boldsymbol{x}$ to the boundary of H^m then any other chart having $\boldsymbol{x}$ in its domain must also take $\boldsymbol{x}$ to the boundary.

Exercise 7.3.1. Prove the last remark.

Remark 7.3.2. When we say that the charts are *compatible* we mean that the composites $\phi_i \circ \phi_j^{-1}$ are smooth where they are defined.

Exercise 7.3.2. Prove that D^2 is a smooth manifold with boundary of dimension 2.

Exercise 7.3.3. Prove that if $g : M \longrightarrow \mathbb{R}$ is a smooth map from a manifold M and y is a regular value of g then

$$\{\boldsymbol{x} \in M : g(\boldsymbol{x}) \geq 0\}$$

is a manifold with boundary and the set of points with $g(\boldsymbol{x}) = 0$ is the boundary.

Remark 7.3.3. you might decide to do the above exercises in reverse order. Then again, you might not.

Exercise 7.3.4. Prove that every closed ball in $\mathbb{R}^n$ is a manifold with boundary.

Definition 7.3.3. We write ∂M for the boundary of the manifold with boundary M. Obviously a manifold without boundary is a manifold-with-boundary M that has $\partial M = \emptyset$.

Proposition 7.3.1. *If $M_1 \xrightarrow{f} M_2$ is a map from a manifold of dimension m_1 with boundary onto a manifold M_2 of dimension m_2, and if $\boldsymbol{y} \in M_2$ is a regular point both of f and $f|\partial M_1$, then $f^{-1}(\boldsymbol{y})$ is an $m_1 - m_2$ dimensional manifold with boundary.*

Proof:

□

Exercise 7.3.5. Provide a proof. You need to use the implicit function theorem on the interior and boundary points separately.

Proposition 7.3.2. *In the case of the last proposition, the boundary of $f^{-1}(y)$ is the intersection of $f^{-1}(y)$ with ∂M_1.*

Proof:

Without loss of generality, take f to be defined from an open subset U of H^{m_1} to an open subset V in $\mathbb{R}^{m_2}$, and ensure that U is small enough so that f is regular on U. Let $\pi_1(x^1, x^2, \cdots, x_1^m)^T = x^1$ for all $\boldsymbol{x} = (x^1, x^2, \cdots x_1^m)^T \in f^{-1}(\boldsymbol{y})$. Then by a recent exercise, we see that $f^{-1}(\boldsymbol{y}) = \pi^{-1}(\mathbb{R})$ and its boundary is $\pi_{-1}(0)$ which is the intersection of $f^{-1}(\boldsymbol{y})$ with ∂M_1. □

Definition 7.3.4. If $A \subset B$ is a subset of a topological space, with $A \xrightarrow{i} B$ the inclusion, a *retraction* of B to A is a map $B \xrightarrow{r} A$ such that $r \circ i$ is the identity, that is, A is left fixed and B is mapped onto A.

Proposition 7.3.3. *There is no smooth retraction $D^n \xrightarrow{r} S^{n-1}$.*

Proof:

Suppose there were such a map, r then let y be a regular value of r. Then it is certainly a regular value on $\partial D^n = S^{n-1}$, where r is the identity. The set $r^{-1}(y)$ is therefore a smooth 1-dimensional manifold with boundary the intersection of $r^{-1}(y)$ with S^{n-1}. This can contain only the point $\boldsymbol{y}$ and no other point of S^{n-1}. But the inverse image of a point $\boldsymbol{y}$ by a continuous map has to be a closed subset of a compact set and is compact. So we have a one dimensional compact manifold with boundary, where the boundary consists of a single point. Now the one dimensional compact manifolds with non-empty boundary are the finite unions of compact intervals. And each of these has an even number of boundary points. Hence the described situation is impossible and hence r does not exist. □

Remark 7.3.4. You may want to see a proof that the only compact 1-manifolds with (non-empty) boundary are finite unions of closed and bounded intervals. The appendix to Milnor's *Topology From the Differentiable Viewpoint* gives one.

Remark 7.3.5. The result generalises to say that there is never a retraction from any manifold with boundart onto the boundary.

Exercise 7.3.6. Prove it.

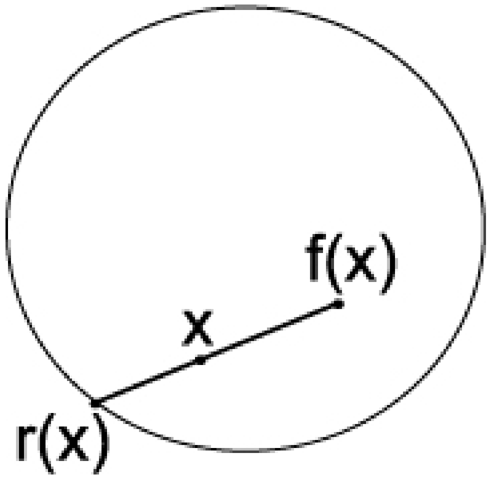

Figure 7.1: Turning a fixed point-free map into a retraction

7.4 Brouwer's Fixed Point Theorem

We can now derive the following important theorem, the Brouwer Fixed Point Theorem. It has been described as follows in dimension three: If you pack a matchbox full of plasticene and then take the plasticene out, twist it, tie knots in it, but do *not* tear it, then push it back into the box, there is at least one point of the plasticene which is in the same position as it was before.

Theorem 7.4.1. If D^n is the closed unit ball in $\mathbb{R}^n$ and $D^n \xrightarrow{f} D^n$ is any continuous map, then there exists an $\boldsymbol{x} \in D^n$ such that $f(\boldsymbol{x}) = \boldsymbol{x}$.

Proof:

We prove it first for the case where the map f is smooth. Suppose there were a smooth map $D^n \xrightarrow{f} D^n$ which had no fixed point. Then as indicated in figure 7.1 we could construct a smooth retraction $D^n \xrightarrow{r} S^{n-1}$ by taking the straight line from $f(\boldsymbol{x})$ to $\boldsymbol{x}$ and extending it until it meets S^{n-1} in $r(\boldsymbol{x})$.

It is not hard to see that if f is smooth so is r.

But we know that no such retraction exists, hence we know that every smooth map $D^n \xrightarrow{f} D^n$ must have at least one fixed point.

Finally we show that there is no continuous map f which is fixed point free.

First we note that the Weierstrass Approximation Theorem says that any continuous map on a compact space can be uniformly approximated as closely as we wish by a polynomial function.

Suppose that a continuous map f exists from D^n to D^n which is fixed point free, then the map $\|f(\boldsymbol{x}) - \boldsymbol{x}\|$ to $\mathbb{R}$ has some minimum value since it is continuous on a compact set and hence has compact image, hence the infimum is attained at some point, and is non-zero. Call this infimum ε. Now we can ensure that the continuous map f is approximated by a smooth (polynomial) map p to within less than ε uniformly. So p must be also a fixed point free map. But this is impossible. Hence f does not exist, and every continuous map $D^n \xrightarrow{f} D^n$ has at least one fixed point. □

7.5 Summary

This last chapter has been a necessarily brief introduction to Differential Topology and it should be possible now for you to read Milnor's book as cited earlier, also his *Morse Theory* which is another fine presentation of some useful Mathematics. I wish it had been possible to go further, but there are severe limitations on how much can be expected of third year students.