

# Differential Geometry

Jacob Denson

October 24, 2024

# Table Of Contents

<b>I</b>	<b>Foundations</b>	<b>2</b>
<b>1</b>	<b>Topological Considerations</b>	<b>4</b>
1.1	Manifolds and Atlases . . . . .	4
1.2	Constructing Manifolds: The Grassmanian . . . . .	7
1.3	Basic Properties of Manifolds . . . . .	11
1.4	Construction in the Category of Manifolds . . . . .	14
1.5	Euclidean Neighbourhoods are Open . . . . .	16
1.6	Equivalence of Regularity Properties . . . . .	17
1.7	Boundaries . . . . .	19
1.8	Invariance of Domain via Jordan Curves . . . . .	20
<b>2</b>	<b>Differentiable Structures</b>	<b>22</b>
2.1	Defining Differentiability . . . . .	22
2.2	Smooth Functions on a Manifold . . . . .	28
2.3	Partial Derivatives . . . . .	29
2.4	The Differential Map . . . . .	31
2.5	The Rank Theorems . . . . .	32
2.6	Immersions, Submersions, and Covers . . . . .	34
2.7	Submanifolds . . . . .	35
2.8	Manifolds and Measure . . . . .	42
2.9	Partitions of Unity . . . . .	46
2.10	Differentiable Manifolds with Boundary . . . . .	49
<b>3</b>	<b>The Tangent Bundle</b>	<b>51</b>
3.1	First . . . . .	51
3.2	A Smooth Structure on the Tangent Bundle . . . . .	51
3.3	Vector Bundles . . . . .	54
3.4	The Space of Derivations . . . . .	55
3.5	Curves as Vectors . . . . .	59
3.6	Sections and Vector Fields . . . . .	60
3.7	Tangents on Manifolds with Boundary . . . . .	63
3.8	Universality of the Tangent Bundle . . . . .	64
3.9	Orientation . . . . .	66

3.10	Whitney's Embedding Theorem . . . . .	71
<b>4</b>	<b>The Cotangent Bundle</b>	<b>75</b>
4.1	The Dual of a Vector Bundle . . . . .	76
4.2	The Method of Lagrangian Multipliers . . . . .	81
4.3	Tensors . . . . .	84
4.4	Local Operators and Derivational Viewpoints . . . . .	86
4.5	The Classical Viewpoint . . . . .	88
<b>5</b>	<b>Lie Derivatives and Differential Equations</b>	<b>90</b>
5.1	Parameter Curves . . . . .	98
5.2	Distributions and Integrability Conditions . . . . .	99
<b>6</b>	<b>Differential Forms and Integration</b>	<b>110</b>
6.1	Scalar Densities . . . . .	111
6.2	Oriented Integration . . . . .	112
6.3	Lower Dimensional Integration . . . . .	113
6.4	Alternating Tensors . . . . .	114
6.5	Differential Forms . . . . .	118
6.6	Integration of Differential Forms . . . . .	119
6.7	The Differential Operator . . . . .	120
6.8	Stoke's Theorem . . . . .	126
6.9	Basic De-Rham Cohomology . . . . .	127
<b>7</b>	<b>Jets</b>	<b>142</b>
<b>8</b>	<b>Lie Groups</b>	<b>144</b>
8.1	A Basic Example: $SO(3)$ . . . . .	144
8.2	General Theory . . . . .	145
<b>II</b>	<b>Riemannian Geometry</b>	<b>152</b>
<b>9</b>	<b>Riemannian Manifolds</b>	<b>153</b>
9.1	The Gradient Vector . . . . .	155
9.2	Unit Tangent Bundle . . . . .	156
9.3	The Volume Form . . . . .	157
<b>10</b>	<b>Connections</b>	<b>158</b>
10.1	Covariant Derivative Operators . . . . .	162
10.2	Affine Connections . . . . .	163

<b>11 Geodesics</b>	<b>165</b>
11.1 The First Variation For Arclength and Energy . . . . .	166
11.2 The Euler-Lagrange Equation For Geodesics . . . . .	168
11.3 Geodesic Completeness . . . . .	172
11.4 The Second Variation Formula For Energy . . . . .	173
11.5 TODO: Separate Jacobi Field Notes . . . . .	176
<b>12 Extra</b>	<b>177</b>
12.1 Connections . . . . .	177
12.2 Riemannian Submanifolds . . . . .	178
12.3 Jacobi Fields . . . . .	179
12.4 Nonpositive Curvature . . . . .	180
12.5 March 22nd . . . . .	181
<b>13 Finsler Manifolds</b>	<b>183</b>
13.1 Curvature Forms . . . . .	189
<b>14 Variation Formulas for Finsler Manifolds</b>	<b>190</b>
<b>15 Finsler Manifolds</b>	<b>194</b>
 <b>III Differential Topology</b>	 <b>196</b>
<b>16 Intersection Theory</b>	<b>198</b>
16.1 Intersection Theory Modulo Two . . . . .	205
16.2 Oriented Intersection Theory . . . . .	207
16.3 Degree and Winding Numbers . . . . .	212
16.4 Euler Characteristic . . . . .	216
16.5 Lefschetz Fixed Point Theory . . . . .	216
 <b>IV Differential Geometry</b>	 <b>221</b>
<b>17 Curves</b>	<b>223</b>
17.1 Planar Curves . . . . .	223
<b>18 Moving Frames</b>	<b>227</b>
<b>19 Differential Topology Notes</b>	<b>228</b>
19.1 November 2nd . . . . .	228
19.2 The Variation of Curvature Formula . . . . .	228

<b>20 Homology of Vector Bundles</b>	<b>230</b>
20.1 Vector Valued Differential Forms . . . . .	231
20.2 Algebraic Understand of Connections . . . . .	231

# **Part I**

## **Foundations**

The main ideas of differential geometry began with the beautiful analysis of curves in the plane, which illustrated the careful balance between the study of local and global properties of mathematical shapes. We will use the study of curves to illustrate the main strategies for the understanding of the ‘differential shapes’ we will eventually come to call manifolds. Starting with the basic definition of a curve in the plane, we proceed to define the basic concepts of curvature and its consequences.

# Chapter 1

## Topological Considerations

In some form of mathematical heaven, all objects would exist in the linear realm, where problems are easy to solve. Unfortunately, we live in the real world. When a physicist describes the motion of a robot's arm, rigidity forces the joints to move along curves bound to a sphere, forced never to move in a linear fashion. When an algebraic geometer studies the curve described by the equation  $x^2 + y^3 = 5$ , he must analyze a shape which bends and curves, never straight. Differential geometry gives the mathematician tools to cheat; in many cases, the shapes we study may not be non-linear, but are at least *locally linear*, so around each point we can, at least locally, try and recover the methods used in the linear case. The challenge is to obtain these locally linear variants, as well as to extend these methods to obtain *global* results, revealing the geometry of the shape we study.

### 1.1 Manifolds and Atlases

Topology attempts to describe the properties of space invariant under continuous stretching and squashing. Differential geometry extends this description to spatial properties constant when space is stretched and squashed, but not 'bent'. Four centuries of calculus have established differentiability in the Euclidean spaces  $\mathbb{R}^n$ . A basic environment to extend the notions of differentiability to topological spaces are those which are locally similar to  $\mathbb{R}^n$ . A *topological manifold* is a Hausdorff topological space  $M$  such that for each point  $p \in M$ , there exists an open neighbourhood  $U \subset M$  containing  $p$ , and a non-negative integer  $n \geq 0$ , such that  $U$  is homeomorphic to an open subset of  $\mathbb{R}^n$ . If the same  $n$  works for all points  $p$ , we say  $M$  is *n dimensional* and write  $\dim M = n$ . Often, we introduce the dimension implicitly by stating 'let  $M^n$  be a manifold', which means that  $M$  is an  $n$  dimensional manifold.

*Remark.* We assume our manifolds are Hausdorff because non-Hausdorff manifolds are far and few between in practice in applications of manifold theory to other areas of mathematics, and it makes some parts of the theory more manageable.

**Example.** Any open subset of  $\mathbb{R}^n$  is a manifold, since an open ball in  $\mathbb{R}^n$  is homeomorphic to  $\mathbb{R}^n$ . More generally, the same kind of argument show that any open subset of a manifold is a manifold. These manifolds will be called open submanifolds of  $M$ .



**Example.** If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is continuous, then we can consider it's graph

$$\Gamma f = \{(x, f(x)) : x \in \mathbb{R}^n\}.$$

The map  $\pi : \Gamma f \rightarrow \mathbb{R}^n$  defined by setting  $\pi(x, y) = x$  is a homeomorphism, with inverse  $\pi^{-1}(x) = (x, f(x))$ , so  $\Gamma f$  is a manifold. Later on, we will see that locally, all hypersurfaces in a manifold locally look like graphs like this.

*Remark.* As the above examples show, any topological space homeomorphic to a topological manifold is also a topological manifold. This is a bad omen, because we want to discuss properties of space which remain invariant under differentiable maps, and differentiability should certainly be a stronger concept than continuity. This indicates we must add additional structure to a manifold in order to distinguish smooth maps from continuous maps. We introduce this structure in the next chapter.

Geometrically a homeomorphism  $x : U \rightarrow \mathbb{R}^n$  from an open set  $U$  on a manifold can be seen as a way of assigning coordinates to points on the manifold, because we associate with each geometric point  $p \in U$  a sequence of numbers  $x(p) = (x^1(p), \dots, x^n(p)) \in \mathbb{R}^n$ . One way to see the study of manifolds is as an extension of analytic geometry to non-planar topological systems. Indeed, the technologies developed have immediate applications to projective, hyperbolic, and elliptic geometries, and the language of manifolds has become the common language of most modern day geometers.

**Example.** Consider the circle

$$S^1 = \{x \in \mathbb{R}^2 : |x| = 1\}.$$

For any proper open subset  $I \subset S^1$ , we define an angle function to be a continuous map  $\theta : I \rightarrow \mathbb{R}$  such that  $e^{i\theta(x)} = x$  for all  $x \in I$ . Then  $\theta$  is a topological embedding of  $I$  in  $\mathbb{R}$ , with continuous inverse  $\theta^{-1}(t) = e^{it}$ . Angle functions exist on any proper open subset of  $S^1$ , and therefore cover  $S^1$ . Thus  $S^1$  is a 1 dimensional manifold.

As a manifold, the circle is distinct from an open subset of  $\mathbb{R}^n$  because we cannot put coordinates over the whole space at once; instead, we must analyze the circle piece by piece to determine the structure on the whole space. This is the main trick to manifold theory – a manifold might be a big strange object globally, but at least locally we can coordinatize, and assume we are working on an open subset of  $\mathbb{R}^n$ .

**Example.** The method of stereographic projection gives another system of coordinates on the circle, which generalizes to higher dimension shapes  $S^n$ . We will project the open subset  $S^1 - \{(1, 0)\}$  to the line  $\{-1\} \times \mathbb{R}$ , by taking the intersection of the line between  $p$  and  $(1, 0)$  and the line  $\{-1\} \times \mathbb{R}$ . A formula for the projection from  $S^1 - \{(1, 0)\}$  to the line  $\{-1\} \times \mathbb{R}$  is given by the continuous function

$$f(x, y) = \frac{2y}{1 - x}.$$

Another calculation shows the inverse is given by

$$f^{-1}(y) = \left(1 - \frac{8}{4 + y^2}, \frac{4y}{4 + y^2}\right) = \left(\frac{y^2 - 4}{y^2 + 4}, \frac{4y}{y^2 + 4}\right).$$

Similar calculations show that on the  $n$ -dimensional spheres  $S^n$ , we have a projection map from  $S^n - \{e_1\}$  onto  $\{-1\} \times \mathbb{R}^n$  by the same process, and one can calculate the formula for this projection is given by

$$f(x_1, \dots, x_n) = \left(\frac{2x_2}{1 - x_1}, \dots, \frac{2x_n}{1 - x_1}\right).$$

and the inverse takes the form

$$\begin{aligned} f^{-1}(y_2, \dots, y_n) &= \left(1 - \frac{8}{4 + |y|^2}, \frac{4y_2}{4 + \|y\|^2}, \dots, \frac{4y_n}{4 + \|y\|^2}\right) \\ &= \left(\frac{|y|^2 - 4}{|y|^2 + 4}, \frac{4y_2}{|y|^2 + 4}, \dots, \frac{4y_n}{|y|^2 + 4}\right). \end{aligned}$$

If we project from the point  $-e_1$  to  $\{1\} \times \mathbb{R}^{n-1}$  instead of  $(1, 0, \dots, 0)$ , then the homeomorphism defined on  $S^1 - \{(-1, 0, \dots, 0)\}$  is calculated to be

$$g(x_1, \dots, x_n) = \left(\frac{x_2}{1 + x_1}, \dots, \frac{x_n}{1 + x_1}\right)$$

with inverse

$$g^{-1}(y_2, \dots, y_n) = \left(\frac{4 - |y|^2}{4 + |y|^2}, \frac{4y_2}{4 + |y|^2}, \dots, \frac{4y_n}{4 + |y|^2}\right).$$

Since any element of  $S^n$  is contained in the domain of one of these projections, we conclude  $S^n$  is a manifold.

When we analyze manifolds, it is convenient to consider not only homeomorphisms onto  $\mathbb{R}^n$ , but also maps onto open subsets of  $\mathbb{R}^n$ . For a manifold  $M$ , we call a pair  $(x, U)$  a *coordinate chart* if  $U$  is an open subset of  $M$ , and  $x$  is a homeomorphism from  $U$  to an open subset of  $\mathbb{R}^n$ . A *coordinate ball* is a chart  $(x, U)$ , where  $x$  is a homeomorphism of  $U$  and the open unit ball in  $\mathbb{R}^n$ . Similarly, we can define coordinate cubes, or other shapes. Letters like  $x, y$  and  $z$  are often used for charts, so that one can pretend one is actually working with coordinates  $x = (x^1, \dots, x^n)$  for a point in  $\mathbb{R}^n$  rather than a function on an abstract manifold. It is often very useful to trick your brain into the coordinate way of thinking, and the only disadvantage is that things can be slightly ambiguous when working in Euclidean space itself.

**Example.** Consider the set  $M(n) = M(n; \mathbb{R})$  of  $n \times n$  matrices with entries in the real numbers. We can identify  $M(n)$  with  $\mathbb{R}^{n \times n}$  by considering only the coefficients of the matrix, and this tells us  $M(n)$  is a topological manifold of dimension  $n^2$ . More generally, the set

$M(n, m)$  of  $n \times m$  real matrices has dimension  $nm$ . The determinant map  $\det : M(n) \rightarrow \mathbb{R}$  acts as a polynomial in the entries of the matrix, so the function is continuous, and so the general linear group  $GL(n) = \{M \in M(n) : \det(M) \neq 0\}$  is an open submanifold of  $M(n)$ . The set  $M(n; \mathbb{C})$  of  $n \times n$  complex matrices is also a  $2n$  dimensional manifold, as is the set  $GL(n; \mathbb{C})$  of invertible complex-coefficient matrices.

**Example.** Let  $M(n, m; k) \subset M(n, m)$  be the set of  $n \times m$  matrices of rank  $k$ . For any matrix  $M \in M(n, m; k)$ , there are permutation matrices  $P$  and  $Q$  such that

$$PMQ = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where  $A \in GL(k)$ . We can then define  $L(N) \in M(n, m)$  by setting

$$L(N) = PNQ = \begin{pmatrix} A(N) & B(N) \\ C(N) & D(N) \end{pmatrix}.$$

The map  $L$  is rank preserving and continuous. The continuity implies that there exists  $U \subset M(n, m)$  such that  $A(N)$  is invertible for all  $N \in U$ . The matrix

$$\begin{pmatrix} I_k & 0 \\ -C(N)A(N)^{-1} & I_{n-k} \end{pmatrix}$$

is invertible, and so  $L(N)$  has the same rank as

$$\begin{aligned} & \begin{pmatrix} I_k & 0 \\ -C(N)A^{-1}(N) & I_{n-k} \end{pmatrix} \begin{pmatrix} A(N) & B(N) \\ C(N) & D(N) \end{pmatrix} \\ &= \begin{pmatrix} A(N) & B(N) \\ 0 & D(N) - C(N)A^{-1}(N)B(N) \end{pmatrix}. \end{aligned}$$

Thus each  $N \in U$  has rank  $k$  if and only if  $D(N) = C(N)A^{-1}(N)B(N)$ , i.e.

$$U \cap M(n, m; k) = \{N \in U : D(N) = C(N)A^{-1}(N)B(N)\}.$$

In particular, we can identify elements of  $U \cap M(n, m; k)$  with an open subset of  $GL(k) \times M(k, n-k) \times M(k, m-k)$ , i.e. with triples  $(A, B, C)$ . Since we can cover  $M(n, m; k)$  by open sets of the form  $U$ , it follows that  $M(n, m; k)$  is a manifold with dimension  $k^2 + k(n-k) + k(m-k) = k(n+m-k)$ .

## 1.2 Constructing Manifolds: The Grassmanian

Sometimes, you may wish to place a topological structure on a set, with the purpose of making it into a manifold, without having any topological structure to begin with. In most cases, the set is contained in a bigger space with some topological structure, and you can just supplant the relative topology on the subspace. But in certain cases, there is no canonical topological structure to start working with. The next theorem provides a useful strategy for placing a topology on a space and simultaneously showing it is a manifold. We will use this construction technique to define the *Grassmanian manifold*.

**Theorem 1.1.** *Let  $M$  be a set together with a cover  $\{U_\alpha\}$  and a family of bijections  $\{x_\alpha\}$ , where  $x_\alpha$  maps  $U_\alpha$  onto an open subset of a manifold  $M_\alpha$ . Suppose also that*

- *For any  $\alpha, \beta$ ,  $x_\alpha(U_\alpha \cap U_\beta)$  is an open subsets of  $M_\alpha$ , and  $x_\beta \circ x_\alpha^{-1}$  is a continuous map from  $x_\alpha(U_\alpha \cap U_\beta)$  into  $M_\beta$ .*
- *If  $a, b \in M$ , then either there is a set  $U_\alpha$  with  $a, b \in U_\alpha$ , or there are disjoint sets  $U_\alpha, U_\beta$  such that  $a \in U_\alpha, b \in U_\beta$ .*

*Then there is a unique topological structure on  $M$  making it into a Hausdorff manifold, such that each  $(x_\alpha, U_\alpha)$  is a chart.*

*Proof.* Without loss of generality, we may assume  $M_\alpha$  is an open subset of Euclidean space for all  $\alpha$ . Consider the topological structure on  $M$  generated by all sets of the form  $x_\alpha^{-1}(V)$ , where  $V$  is an open subset of  $x_\alpha(U_\alpha)$ . The first condition of this theorem implies that these sets form a base for a topology; if  $W_\alpha$  is an open set of  $x_\alpha(U_\alpha)$ , and  $W_\beta$  is an open subset of  $x_\beta(U_\beta)$ , then set  $V_\alpha = x_\alpha^{-1}(W_\alpha)$  and  $V_\beta = x_\beta^{-1}(W_\beta)$ . Then

$$x_\alpha(V_\alpha \cap V_\beta) = W_\alpha \cap (x_\alpha \circ x_\beta^{-1})(x_\beta(U_\alpha) \cap W_\beta)$$

is an intersection of open sets, and is therefore open, and so  $W_\alpha \cap W_\beta = x_\alpha^{-1}(V_\alpha \cap V_\beta)$  is an element of the basis. Note this calculation shows that the maps  $x_\alpha$  are open in the topology induced by this basis, hence the maps  $x_\alpha$  are homeomorphisms. The space  $M$  is also Hausdorff by the second property, since two points  $p, q$  in a common set  $U_\alpha$  can be separated as in Euclidean space, or separated by two disjoint charts  $U_\alpha$  and  $U_\beta$ . Thus  $M$  is a manifold.  $\square$

*Remark.* If the family  $\{U_\alpha\}$  has a countable subcover, and each of the manifolds  $M_\alpha$  is second countable, then the manifold  $M$  constructed above will be second countable.

*Remark.* In the next chapter, we will define the notion of a smooth structure on a manifold. If  $x_\alpha \circ x_\beta^{-1}$  is smooth in addition to being continuous, then it is easy to see that the manifold  $M$  constructed above also has a unique smooth structure such that each  $(x, U)$  is a smooth coordinate chart.

**Example.** *Let  $V$  be an  $n$  dimensional vector space. If  $E = (e_1, \dots, e_n)$  is an ordered basis of  $V$ , then we can associate a bijective map  $x_E : V \rightarrow \mathbb{R}^n$  such that*

$$x_E^{-1}(a_1, \dots, a_n) = a_1 e_1 + \dots + a_n e_n.$$

*If  $F = (f_1, \dots, f_n)$  is another ordered basis, then  $x_F \circ x_E^{-1}$  is an invertible linear map, and therefore a homeomorphism. Thus there is a unique topology and smooth structure on  $V$  such that each map  $x_E$  is a diffeomorphism from  $V$  to  $\mathbb{R}^n$ . This topology is of course, the same topology that makes all the vector space operations on  $V$  continuous.*

Now we've gone through a basic example, let's consider a more sophisticated one. The last example shows that for any finite dimensional vector spaces  $V$  and  $W$ , the space  $L(V, W)$  of linear maps from  $V$  to  $W$  has a natural topology and smooth structure making it homeomorphic to Euclidean space. We now consider the *Grassmann manifold*, which parameterizes subspaces of a fixed dimension in  $\mathbb{R}^n$ .

Let  $G(k, n)$  be the family of all  $k$  dimensional subspaces of  $\mathbb{R}^n$ . For each  $W \in G(n - k, n)$ , let  $U_W$  be the set of all  $V \in G(k, n)$  such that  $V \cap W = \{0\}$ , i.e. the set of all  $k$  dimensional subspaces of  $\mathbb{R}^n$  which complement  $W$ . For each  $W_0 \in G(n - k, n)$  and  $V_0 \in U_{W_0}$ , we will define a bijection

$$x_{V_0, W_0} : U_{W_0} \rightarrow L(V_0, W_0),$$

Using the result above, we will give  $G(k, n)$  the structure of a topological space such that each  $U_{W_0}$  is open, and  $x_{V_0, W_0}$  is a homeomorphism of  $U_{W_0}$  with  $L(V_0, W_0)$ .

To define  $x_{V_0, W_0}$ , such that for any  $V \in U_{W_0}$ , the linear transformation  $T = x_{V_0, W_0}(V)$  is the unique linear map such that

$$V = \{v_0 + Tv_0 : v \in V_0\}.$$

That such a map exists follows from the fact that if  $\pi : \mathbb{R}^n \rightarrow V_0$  is the projection induced by the decomposition  $\mathbb{R}^n = V_0 \oplus W_0$ , then  $\pi|_V : V \rightarrow V_0$  is an isomorphism because  $\text{Ker}(\pi|_V) = V \cap W_0 = \{0\}$ , and so we can define  $Tv_0 = (\pi|_V)^{-1}v_0 - v_0$ . The uniqueness of  $T$  follows because any such  $T$  should be defined by this formula. If  $T$  is given then  $v_0 + Tv_0 \in V$ , and since  $\pi|_V(v_0 + Tv_0) = v_0$ ,  $Tv_0 = (\pi|_V)^{-1}(v_0) - v_0$ . For any  $T \in L(V_0, W_0)$ , the vector space  $V$  defined above lies in  $U_{W_0}$ , so  $x_{V_0, W_0}$  really does give a bijection between  $U_{W_0}$  and  $L(V_0, W_0)$ .

It now suffices to show that the maps  $\{x_{V_0, W_0}\}$  are compatible with one another, so that they really do give  $G(k, n)$  the structure of a manifold. So let  $(V_0, W_0)$  and  $(V_1, W_1)$  be pairs of complementary subspaces as above. Let  $\pi : \mathbb{R}^n \rightarrow V_1$  be the projection induced by the direct sum decomposition  $\mathbb{R}^n = V_1 \oplus W_1$ , and let  $\eta : W_1 \rightarrow \mathbb{R}^n$  be the inclusion map:

- To begin with, it suffices to show that  $x_{V_0, W_0}$  maps  $U_{W_0} \cap U_{W_1}$  onto an open subset of  $L(V_0, W_0)$ . If  $V \in U_{W_0}$ , and  $T = x_{V_0, W_0}(V)$ , then  $V \in U_{W_1}$  if and only if  $\pi \circ (1 + T)$  is an invertible map from  $V_0$  to  $V_1$ . Thus

$$x_{V_0, W_0}(U_{W_0} \cap U_{W_1}) = \{T \in L(V_0, W_0) : \pi \circ (1 + \eta \circ T) \text{ is invertible}\}.$$

But the latter is clearly an open subset of  $L(V_0, W_0)$ , i.e. because the map  $T \mapsto \pi \circ (1 + \eta \circ T)$  is continuous from  $L(V_0, W_0)$  to  $L(V_0, V_1)$ , and the space of invertible maps is an open subset of  $L(V_0, V_1)$ .

- To show that the map  $x_{V_1, W_1} \circ x_{V_0, W_0}^{-1}$  is continuous between the images of  $U_{W_0} \cap U_{W_1}$ , suppose  $V \in U_{W_0} \cap U_{W_1}$ ,  $T = x_{V_0, W_0}(V)$ , and  $S = x_{V_1, W_1}(V)$ . Fix  $v_1 \in V_1$ , and suppose that  $v_0 \in V_0$  is the unique vector such that

$$v_1 = \pi(v_0 + Tv_0) = [\pi \circ (1 + \eta \circ T)]v_0.$$

Thus  $v_0 = [\pi \circ (1 + \eta \circ T)]^{-1}v_1$ . Now it follows from our definition that  $v_0 + Tv_0 = v_1 + Sv_1$ , and so we can write

$$\begin{aligned} Sv_1 &= v_0 + Tv_0 - v_1 \\ &= [(1 + \eta \circ T)(\pi \circ (1 + \eta \circ T))^{-1}v_1] - v_1 \\ &= ((1 + \eta \circ T) \circ (\pi \circ (1 + \eta \circ T))^{-1} - 1)v_1. \end{aligned}$$

But this implies that

$$S = (1 - \pi) \circ ((1 + \eta \circ T) \circ (\pi \circ (1 + \eta \circ T))^{-1} - 1).$$

From this formula we see that  $T \mapsto S$  is a continuous function, given that composition and inversion of this kind is continuous provided that  $T$  is restricted to lie in  $U_{W_1}$ , which we assumed.

Finally, it suffices to verify that the sets  $\{U_W : W \in G(n - k, n)\}$  cover  $G(k, n)$ , and give a Hausdorff structure to  $G(k, n)$ . That the sets cover  $G(k, n)$  follow from the fact that any subspace of a finite dimensional vector space has a complement. In fact, a finite family of charts suffice to cover  $G(k, n)$ , since the next lemma shows that any finite dimensional subspace has an axis oriented complement, which shows that  $G(k, n)$  is second countable. It is actually *compact*, as we will see later.

**Lemma 1.2.** *If  $V$  is a finite dimensional subspace of  $\mathbb{R}^n$ , there there exists an axis oriented complement  $W$  to  $V$ .*

*Proof.* If  $V$  is  $k$  dimensional, let  $v_1, \dots, v_k$  be a basis for  $V$ . Then the matrix with these vectors as rows has rank  $k$ , and thus has an invertible  $k \times k$  minor. Let  $i_1, \dots, i_k$  be the columns of this minor. We can then set  $W$  to be the span of  $\{e_1, \dots, e_n\} - \{e_{i_1}, \dots, e_{i_k}\}$ .  $\square$

To verify the Hausdorff condition, we now show that any two finite dimensional subspaces of  $\mathbb{R}^n$  have a common complement.

**Lemma 1.3.** *Suppose  $V_0, V_1 \in G(k, n)$ . Then there exists  $W \in G(n - k, n)$  such that  $V_0 \cap W = V_1 \cap W = \{0\}$ .*

*Proof.* Suppose  $V_0 \cap V_1$  is  $m$  dimensional. Let  $U_0$  and  $U_1$  be complements of  $V_0 \cap V_1$  in  $V_0$  and  $V_1$  respectively. Then  $U_0$  and  $U_1$  are both  $k - m$  dimensional, and so there exists an isomorphism  $T : U_0 \rightarrow U_1$ . If  $A$  is a complement of  $V_0 + V_1$ , which is  $n + m - 2k$  dimensional, then we claim that

$$A \oplus (1 + T)U_0$$

complements both  $V_0$  and  $V_1$ :

- Fix  $a \in A$  and  $u_0 \in U_0$ , and suppose that  $a + (1 + T)u_0 \in V_0$ . Since  $A$  complements  $V_0 + V_1$ , this implies that  $a = 0$ , and so  $Tu_0 \in V_0$ . But  $Tu_0$  is also in  $U_1$ , and  $V_0 \cap U_1 = \{0\}$ , so  $Tu_0 = 0$ , and thus  $u_0 = 0$ .

- Fix  $a \in A$ , and  $u_0 \in U_0$ , and suppose that  $a + (1 + T)u_0 \in V_1$ . As above, we then have  $a = 0$ , and so  $(1 + T)u_0 = u_0 + Tu_0 \in V_1$ . Since  $Tu_0 \in U_1 \subset V_1$ ,  $u_0 \in V_1$ . But  $U_0 \cap V_1 = \{0\}$ , so  $u_0 = 0$ .

Thus we have verified that  $A \oplus (1 + T)U_0$  is a common complement.  $\square$

We have proved all the required facts to prove that  $G(k, n)$  has the structure of a second countable,  $k(n - k)$  dimensional manifold. It is simple to see that the maps above are also smoothly compatible, so that  $G(k, n)$  is a smooth manifold. We will finish off our introduction of the Grassmanian by showing that  $G(k, n)$  is actually *compact*.

**Lemma 1.4.**  *$G(k, n)$  is a compact manifold.*

*Proof.* Let  $A \subset M(n)$  be the subset of linear operators consisting of orthogonal projections onto elements of  $G(k, n)$ . The map  $P \mapsto \text{Im}(P)$  gives a bijection  $f$  between  $A$  and  $G(k, n)$ . It is clear that  $A$  is a closed and bounded subset of  $M(n)$ , and is therefore compact, so it suffices to show that  $f$  is continuous. If  $W_0 \in G(n - k, n)$  corresponds to an orthogonal projection operator  $P_0 : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , then  $f^{-1}(U_{W_0})$  is equal to the set of all rank  $k$  orthogonal projection operators  $P$  such that  $\det(P - P_0) \neq 0$ , and thus an open subset  $U_0$  of  $A$  since  $T \mapsto \det(T - P_0)$  is continuous. Now fix  $P_0 \in A$ , let  $V_0 = \text{Im}(P_0)$ , and let  $W_0$  be the orthogonal complement to  $V_0$ . If  $\{P_n\}$  is a sequence in  $A$  converging to  $P_0$ , continuity will follow if we can show that  $f(P_n)$  converges to  $f(P_0)$ . Since  $U_0$  is open, we may assume without loss of generality that  $\{P_n\} \subset U_0$ . It thus suffices to show that the sequence of matrices  $T_n = (x_{V_0, W_0} \circ f)(P_n)$  converges to zero. If  $V_n = \text{Im}(P_n)$ , we know that  $T_n = (P_0|_{V_n})^{-1} - 1$ . But

$$T_n = (P_0|_{V_n})^{-1} - 1 = ((P_0 - P_n)|_{V_n} + P_n|_{V_n})^{-1} - 1 = ((P_0 - P_n)|_{V_n} + 1)^{-1} - 1.$$

The operator norm of  $(P_0 - P_n)|_{V_n}$  is upper bounded by the operator norm of  $P_0 - P_n$ , which converges to zero. Thus the quantity on the right hand side converges to  $(0 + 1)^{-1} - 1 = 0$ . Thus  $f$  is continuous, and so  $G(k, n)$  is compact.  $\square$

*Remark.* If  $V$  is a finite dimensional vector space, then the construction above, which is essentially basis invariant, can be transferred over to show the space  $G_k(V)$  of  $k$  dimensional subspaces of  $V$  is also a manifold. A special case is the space  $G_1(V)$  of lines through the origin in  $V$ , known as the *projectivization* of  $V$  and denoted by  $\mathbf{P}(V)$ . The space  $\mathbf{P}^n = \mathbf{P}(\mathbb{R}^{n+1})$  is known as  *$n$  dimensional real projective space*, and will be given a more simple construction later on in this chapter.

### 1.3 Basic Properties of Manifolds

Many proofs about manifolds use a reliable trick. First, we conjure forth local homeomorphisms to  $\mathbb{R}^n$ . Then we transport nice properties of  $\mathbb{R}^n$  across the homeomorphism, thereby inducing the properties on the manifold. In fact, the general philosophy of manifold theory is that most properties of  $\mathbb{R}^n$  will carry across to arbitrary spaces that look locally like  $\mathbb{R}^n$ . Thus we can perform linear algebra on spaces that are not really linear!

**Theorem 1.5.** *Every manifold is locally compact.*

*Proof.* Let  $M$  be a manifold, let  $p \in M$  be a point, and let  $U$  be a neighbourhood of  $p$ . Then there is a coordinate chart  $(x, V)$ , where  $V \subset U$  and  $p \in V$ . If  $B$  is a closed ball around  $x(p)$  in  $x(U)$ , then  $B$  is a compact neighbourhood of  $x(p)$ , hence  $x^{-1}(B)$  is a compact neighbourhood of  $p$ , contained in  $U$ . Thus  $M$  has a basis of pre-compact coordinate balls, and is therefore locally compact.  $\square$

The same method shows that every manifold is locally path-connected, and thus locally connected. The next problem requires more foresight on the reader, though the basic technique used is exactly the same.

**Theorem 1.6.** *A connected manifold is path-connected.*

*Proof.* Let  $M$  be a connected manifold, and fix  $p \in M$ . Let

$$U = \{q \in M : p \text{ and } q \text{ are path connected}\}.$$

Since  $M$  is locally path connected,  $U$  is open. Now suppose  $q$  is a limit point of  $U$ . Take some coordinate ball  $(x, V)$  around  $q$ . Then  $V$  contains some  $r \in U$ , which is path connected to  $p$ , and we can then patch this path with a path from  $r$  to  $q$  to obtain a path from  $p$  to  $q$ . Thus  $U$  is open, closed, and non-empty, so  $U = M$ .  $\square$

Since every manifold is locally connected, any manifold can be split up into the disjoint sum of its connected components. It therefore often suffices to study connected manifolds, since any manifold is the disjoint union of connected manifolds.

**Example.**  $GL(n)$  is not connected, since  $\det(GL(n))$  is disconnected. We now show  $GL(n)$  consists of two path connected components, those matrices with positive determinant, and those matrices with negative determinant. To do this, we shall construct paths in  $GL(n)$  reducing invertible matrices to certain canonical forms. In our proof, we shall use the fact that  $GL_n(\mathbb{R})$  can be described as tuples of  $n$  linearly independent vectors in  $\mathbb{R}^n$ , which simplifies notation. If  $v_1, \dots, v_n$  are linearly independent, consider adding one vector to another.

$$(v_1, \dots, v_p, \dots, v_q, \dots, v_n) \mapsto (v_1, \dots, v_p + v_q, \dots, v_q, \dots, v_n)$$

These vectors are path connected in  $GL_n(\mathbb{R})$  by the path

$$t \mapsto (v_1, \dots, v_p + tv_q, \dots, v_q, \dots, v_n)$$

Similarly, we may subtract rows from one another. Next, consider multiplying a row by a scalar  $\gamma > 0$ ,

$$(v_1, \dots, v_p, \dots, v_n) \mapsto (v_1, \dots, \gamma v_p, \dots, v_n)$$

These matrices are path connected by

$$t \mapsto (v_1 \quad \dots \quad [(1-t) + t\gamma]v_p \quad \dots \quad v_n)^t$$



We cannot perform this technique if  $\gamma < 0$ , because then  $(1 - t) + t\gamma = 0$  for some choice of  $t$ , and the resulting vectors become linearly dependent. We should not expect to find a path when  $\gamma < 0$ , since multiplying by a negative number reverses the sign of the determinant, and we know from the continuity of the determinant that the sign of the determinant determines at least two connected components. The same reasoning shows we can't necessarily swap two rows. Fortunately, we don't need these operations – we may use the path-connected elementary matrices to reduce any matrix to a canonical form. A modification of the Gauss Jordan elimination algorithm (left to the reader as a simple exercise) shows all matrices can be path-reduced to a matrix of the form

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \dots & 0 & \pm 1 \end{pmatrix}$$

One matrix has determinant greater than zero, the other has determinant less than zero. Thus  $GL(n)$  consists of two homeomorphic path-connected components: the matrices with determinant greater than zero, and the component with determinant less than zero, reflecting the linear transformations which preserve orientation, and the ones which reverse orientation. This is quite a different situation from the space  $GL(n, \mathbb{C})$  of invertible matrices over the complex numbers, which is connected for any  $n$ .

Another useful fact of a manifold, if the dimension of the manifold is the same, is that all the local parts of space look identical to one another. We say a topological space  $X$  is *homogenous* if, for every two points  $p, q \in X$ , there is a homeomorphism  $f : X \rightarrow X$  with  $f(p) = q$ .

**Theorem 1.7.** *Every connected manifold is homogenous.*

*Proof.* Begin by considering a closed unit disk  $\mathbf{D}^n \subset \mathbb{R}^n$ . Let  $p, q$  lie in the interior of the disk. Then there is a homeomorphism  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  which fixes all points outside of the interior of  $\mathbf{D}^n$ . Let  $\psi$  be a smooth, non-negative function supported on a closed set contained in the interior of the unit disk, with  $\psi(x) = 1$  for all points  $x$  on the line segment between  $p$  and  $q$ . Consider the vector field  $v(x) = \psi(x)(q - p)$ . Then there is a one parameter family of diffeomorphisms  $\{\varphi_t\}$  such that  $\varphi_0$  is the identity map, and for each  $x \in \mathbb{R}^n$  and  $t \in \mathbb{R}$ ,

$$\frac{d\varphi_t(x)}{dt} = \psi(x)(q - p).$$

In particular, for each  $x$  on the line segment between  $p$  and  $q$ ,

$$\frac{d\varphi_t(x)}{dt} = q - p,$$

so  $\varphi_1(p) = q$ . Thus we can set  $f = \varphi_1$ .

Now let  $M$  be a connected manifold, fix  $p \in M$ , and consider

$$U = \{q \in M : \text{there is a homeomorphism } f : M \rightarrow M \text{ such that } f(p) = q.\}$$

If  $q \in U$ , let  $(x, U)$  be a coordinate chart centered at  $q$  and with  $x(U) = \mathbb{R}^n$ . If  $B$  is the open unit disk in  $\mathbb{R}^n$ , then for any  $r \in B$ , there is a homeomorphism  $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  with  $\tilde{f}(0) = r$ . We can then define a homeomorphism  $f : M \rightarrow M$ , by setting  $f(x) = x$  for all  $x \in M - x^{-1}(B)$ , and setting  $f = x^{-1} \circ \tilde{f} \circ x$  for all  $x \in U$ , then  $f(q) = r$ , and we can compose this homeomorphism with a homeomorphism  $g : M \rightarrow M$  with  $g(p) = q$ , which gives  $r \in U$ . Thus  $U$  is open. A similar argument shows also that  $U$  is closed, thus  $U = M$ .  $\square$

*Remark.* If  $M$  is a connected, smooth manifold, then the same argument shows that one can pick  $f : M \rightarrow M$  in the argument above to be a diffeomorphism.

## 1.4 Construction in the Category of Manifolds

The category  $\mathbf{Man}$  of topological manifolds has useful constructions which are good to know as tools for constructing more manifolds from simpler ones. For instance, if  $M$  and  $N$  are manifolds, then the disjoint union  $M \cup N$  is a manifold. Here are some more constructions.

**Theorem 1.8.** *If  $M$  and  $N$  are manifolds, then  $M \times N$  is a manifold.*

*Proof.* Products of homeomorphisms onto open subsets of  $\mathbb{R}^n$  suffice.  $\square$

**Example.** Since  $S^1$  is a one dimensional manifold, the manifold  $\mathbb{T}^2 = S^1 \times S^1$  is a two dimensional manifold, known as the torus. It is most natural to picture this manifold via the embedding  $f : \mathbb{T}^2 \rightarrow \mathbb{R}^3$  given in coordinates by

$$f(\theta_1, \theta_2) = (\cos(\theta_1)[2 + \cos(\theta_2)], \sin(\theta_1)[2 + \cos(\theta_2)], \sin(\theta_2)),$$

whose image is a donut in three dimensional space. More generally, the  $n$  torus  $\mathbb{T}^n = S^1 \times \cdots \times S^1$  is an  $n$  manifold.

**Example (Surfaces of Revolution).** If  $M$  is a 1-manifold in the subplane of  $\mathbb{R}^3$  defined by

$$X = \{(x, y, 0) : x, y > 0\}$$

Then the space obtained by rotating  $M$  around the  $y$  axis is a 2-manifold, known as a surface of revolution, and homeomorphic to  $M \times S^1$ .

Under some restrictions, we can consider quotient spaces of manifolds which are manifolds. In the most simplest case, if  $f : M \rightarrow N$  is a locally injective open surjective map, and  $M$  is a manifold, then  $N$  is a manifold, because sufficiently small charts on  $M$  can be pushed forward giving charts on  $N$ . Later, we will see more interesting examples of quotient manifolds, generated by the actions of Lie groups.

**Example (The Möbius Strip).** Consider the group action of  $\mathbb{Z}$  on  $(-\infty, \infty) \times \mathbb{R}$  by letting  $n \cdot (x, y) = (x + n, (-1)^n y)$ . The quotient space is the Möbius strip  $\mathbf{M}$ . The projection map  $\pi : (-\infty, \infty) \times \mathbb{R} \rightarrow \mathbf{M}$  is open and locally injective, so  $\mathbf{M}$  is a manifold. By throwing away points, we find  $M$  can also be obtained as a quotient of  $[-1, 1] \times (-1, 1)$  by identifying  $(-1, x)$  with  $(1, -x)$ , for each  $x \in (0, 1)$ .

**Example (Projective Space).** Consider the group action of  $\{-1, 1\}$  on  $S^n$  given by reflection. We can then consider a quotient space obtained by identifying opposite sides of the sphere. The projection is open and locally injective, so the resulting quotient space, denoted  $\mathbb{R}\mathbf{P}^n$ , known as real projective space. Another way to describe the space is as  $\mathbb{R}^{n+1} - \{0\}$ , where  $x$  and  $\lambda x$  are identified, for  $\lambda \neq 0$ . To obtain explicit charts on  $\mathbb{R}\mathbf{P}^n$ , define  $x : S^n \rightarrow \mathbb{R}^n$  by

$$x(a_1, \dots, a_{n+1}) = \frac{1}{a_i}(a_1, \dots, \hat{a}_i, \dots, a_{n+1})$$

This map is continuous whenever  $a_i \neq 0$ . For all points  $p$ ,  $x(p) = x(-p)$ , so the chart descends to a map on  $\mathbb{R}\mathbf{P}^n$  instead. It even has a continuous inverse

$$x^{-1}(b_1, \dots, b_n) = [b_1, \dots, 1, \dots, b_n]$$

Since our maps cover the space,  $\mathbb{R}\mathbf{P}^n$  is an  $n$  dimensional manifold.

**Example.** A similar space can be formed in the case of complex scalars, by viewing  $\mathbb{R}^{2n}$  as  $\mathbb{C}^n$ . We then identify  $z$  with  $\lambda z$ , for  $\lambda \in \mathbb{C} - \{0\}$  to form the quotient space  $\mathbb{C}\mathbf{P}^n$  from  $\mathbb{C}^{n+1} - \{0\}$ . For  $n = 1$ ,  $\mathbb{C}\mathbf{P}^1$  is just the Riemann sphere, a 2 manifold. In general,  $\mathbb{C}\mathbf{P}^n$  is a  $2n$  dimensional real manifold, or  $n$  ‘complex’ dimensions. We may consider the chart

$$x(z_1, \dots, z_{n+1}) = \frac{1}{z_i}(z_1, \dots, \hat{z}_i, \dots, z_{n+1})$$

with inverse

$$x^{-1}(w_1, \dots, w_n) = [w_1, \dots, 1, \dots, w_n]$$

where we consider complex multiplication instead of real multiplication. The space formed is known as complex projective space.

Even though  $\mathbb{R}\mathbf{P}^2$  is locally trivial, the space is very strange globally, and cannot be embedded in  $\mathbb{R}^3$ . Nonetheless, the geometry our eyes percieve is modelled very accurately by the spherical construction of projective space. We don’t see the really weird part of  $\mathbb{R}\mathbf{P}^2$ , since our eye cannot see the full circumference of vision, but these topological problems occur in the algorithms involved in patching portions of vision across the entire circumference.

**Example (Gluing Surfaces).** Let  $M$  and  $N$  be connected  $n$ -manifolds. We shall define the connected sum  $M\#N$  of the two manifolds. There are two sets  $B_1$  and  $B_2$  in  $M$  and  $N$  respectively, both homeomorphic to the closed unit ball in  $\mathbb{R}^n$ . Then there is a homeomorphism  $h : \partial B_1 \rightarrow \partial B_2$ , and we may define the connected sum as

$$M\#N = (M - B_1^\circ) \cup_h (N - B_2^\circ)$$

*The topological structure formed is unique up to homeomorphism, but this requires some tough topology to show for general manifolds. An example application is the construction of the  $n$ -holed torus  $T \# T \# \dots \# T \# T$ , which is a surface embeddable in  $\mathbb{R}^3$ .*

## 1.5 Euclidean Neighbourhoods are Open

In these notes, we consider a neighbourhood as in the French school, as any subset containing an open set, regardless of whether it is open or not. Nonetheless, let  $M$  be a manifold, and take a point  $p$  with neighbourhood  $U$  homeomorphic to  $\mathbb{R}^n$ , let's say, by some continuous function  $f : U \rightarrow \mathbb{R}^n$ . Then  $U$  contains an open set  $V$ , and  $f(V)$  is open in  $\mathbb{R}^n$ , so that  $f(V)$  contains an open ball  $W$  around  $f(x)$ . But then  $W$  is homeomorphic to  $\mathbb{R}^n$ , and  $f^{-1}(W)$  is a neighbourhood of  $x$  open in  $V$  (and therefore open in  $M$ ) homeomorphic to  $\mathbb{R}^n$ . This complicated discussion stipulates that we may always choose open neighbourhoods in the definition in a manifold. Remarkably, it turns out that all neighbourhoods homeomorphic to  $\mathbb{R}^n$  must be open; to prove this, we require an advanced theorem of algebraic topology.

**Theorem 1.9** (Invariance of Domain). *If  $f : U \rightarrow \mathbb{R}^n$  is a continuous, injective function, where  $U$  is an open subset of  $\mathbb{R}^n$ , then  $f(U)$  is open, and  $f$  is an embedding of  $U$  in  $\mathbb{R}^n$ .*

Classically, analysts called a connected open set a domain, and so this theorem shows that the property of being a domain is invariant under continuous injective maps from  $\mathbb{R}^n$  to itself. In multivariate calculus, the inverse function theorem shows this for differentiable mappings with non-trivial Jacobian matrices across its domain; invariance of domain stipulates that the theorem in fact holds for any such continuous map  $f$  on an open domain. The theorem can be proven in an excursion in some basic algebraic topology. In an appendix to this chapter, we shall prove the theorem based on the weaker assumption of the Jordan curve theorem, which is slightly more intuitive than invariance of domain, but still requires strong techniques in algebraic topology to prove.

**Lemma 1.10.** *If  $U \subset \mathbb{R}^n$  and  $V \subset \mathbb{R}^m$  are open, then  $U \cong V$  implies  $n = m$ .*

*Proof.* If  $n < m$ , consider the projection  $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$

$$\pi(x_1, \dots, x_n) = (x_1, \dots, x_n, 0, \dots, 0)$$

Clearly no subset of  $\pi(\mathbb{R}^n)$  is open. But if  $f : V \rightarrow U$  is a homeomorphism, then  $\pi \circ f : V \rightarrow \mathbb{R}^m$  is continuous and injective, so  $\pi(V) \subset \pi(\mathbb{R}^n)$  is open by invariance of domain.  $\square$

The *dimension* of a point on a manifold is the dimension of the euclidean space which is locally homeomorphic to a neighbourhood of the point. When a manifold is connected, one can show simply that the dimension across the entire space is invariant, and we may call this the *dimension of the manifold*. As shorthand, we let  $M^n$  denote a manifold  $M$  which is  $n$  dimensional.

**Corollary 1.11.** *The dimension of a point on a manifold is unique.*

*Proof.* Let  $U$  and  $V$  be two non-disjoint neighbourhoods of a point homeomorphic to  $\mathbb{R}^n$  and  $\mathbb{R}^m$  by  $f : U \rightarrow \mathbb{R}^n$  and  $g : V \rightarrow \mathbb{R}^m$ . Then  $U \cap V$  is also open, and homeomorphic to open sets of  $\mathbb{R}^n$  and  $\mathbb{R}^m$ . We conclude  $n = m$ .  $\square$

**Theorem 1.12.** *Any subset of a manifold locally homeomorphic to Euclidean space is open in the original topology.*

*Proof.* Let  $M$  be a manifold, and  $U \subset M$  homeomorphic to  $\mathbb{R}^n$  by a function  $f$ . Let  $x \in U$  be arbitrary. There is an open neighbourhood  $V$  of  $x$  that is homeomorphic into  $\mathbb{R}^n$  by a function  $g$ . Since  $V$  is open in  $M$ ,  $U \cap V$  is open in  $U$ , so  $f(U \cap V)$  is open in  $\mathbb{R}^n$ . We obtain a one-to-one continuous function from  $f(U \cap V)$  to  $g(U \cap V)$  by the function  $g \circ f^{-1}$ . It follows by invariance of domain that  $g(U \cap V)$  is open in  $\mathbb{R}^n$ , so  $U \cap V$  is open in  $V$ , and, because  $V$  is open in  $M$ ,  $U \cap V$  is open in  $M$ . In a complicated manner, we have shown that around every point in  $U$  there is an open neighbourhood contained in  $U$ , so  $U$  itself must be open.  $\square$

Really, this theorem is just a generalized invariance of domain for arbitrary manifolds – since the concept of a manifold is so intertwined with Euclidean space, it is no surprise we need the theorem for  $\mathbb{R}^n$  before we can prove the theorem here.

## 1.6 Equivalence of Regularity Properties

Many important results in differentiable geometry require spaces with more stringent properties than those that are merely Hausdorff. At times, we will want to restrict ourselves to topological manifolds with these properties. Fortunately, most of these properties are equivalent.

**Theorem 1.13.** *For any manifold, the following properties are equivalent:*

- (1) *Every component of the manifold is  $\sigma$ -compact.*
- (2) *Every component of the manifold is second countable.*
- (3) *The manifold is metrizable.*
- (4) *The manifold is paracompact (so every compact manifold is metrizable).*

**Lemma 1.14** (1)  $\rightarrow$  (2). *Every  $\sigma$ -compact, locally second countable space is globally second countable.*

*Proof.* Let  $X$  be a locally second countable space, equal to the union of compact sets  $\bigcup_{i=1}^{\infty} A_i$ . For each  $x$ , there is an open neighbourhood  $U_x$  with a countable base  $C_x$ . If, for some  $A_i$ , we consider the set of  $U_x$  for  $x \in A_i$ , we obtain a cover, which therefore must have a finite subcover  $U_{x_1}, U_{x_2}, \dots, U_{x_n}$ . Taking  $\bigcup_{i=1}^n C_{x_i}$ , we obtain a countable base  $C_i$  for all points in a neighbourhood of  $A_i$ . Then, taking the union  $\bigcup_{i=1}^{\infty} C_i$ , we obtain a countable base for  $X$ .  $\square$

**Lemma 1.15** (2)  $\rightarrow$  (3). *If a manifold is second countable, then it is metrizable.*

*Proof.* This is a disguised form the Urysohn metrization theorem, proved in a standard course in general topology. If you do not have the background, you will have to have faith that this lemma holds. All we need show here is that a second countable manifold is regular, and this follows because every locally compact Hausdorff space is Tychonoff.  $\square$

**Lemma 1.16** (3)  $\rightarrow$  (1). *Every connected, locally compact metrizable space is  $\sigma$ -compact.*

*Proof.* Consider any connected, locally compact metric space  $(X, d)$ . For each  $x$  in  $X$ , let

$$r(x) = \frac{\sup\{r \in \mathbb{R} : \bar{B}_r(x) \text{ is compact}\}}{2}$$

Since  $X$  is locally compact, this function is well defined and positive for all  $x$ . If  $r(x) = \infty$  for any  $x$ , then  $\{\bar{B}_n(x) : n \in \mathbb{Z}\}$  is a countable cover of the space by compact sets. Otherwise,  $r(x)$  is finite for every  $x$ . Suppose that

$$d(x, y) + r' < 2r(x)$$

By the triangle inequality, this tells us that  $\bar{B}_{r'}(y)$  is a closed subset of  $\bar{B}_{r(x)}(x)$ , which is hence compact. This shows that, when  $d(x, y) < r(x)$ ,

$$r(y) \geq r(x) - \frac{d(x, y)}{2}$$

Put more succinctly, this equation tells us that the function  $r : X \rightarrow \mathbb{R}$  is continuous:

$$|r(x) - r(y)| < \frac{d(x, y)}{2}$$

This has an important corollary. Consider a compact set  $A$ , and let

$$A' = \bigcup_{x \in A} \bar{B}_{r(x)}(x)$$

We claim that  $A'$  is also compact. Consider some sequence  $\{x_i\}$  in  $A'$ , and let  $\{a_i\}$  be elements of  $A$  for which  $x_i \in \bar{B}_{r(a_i)}(a_i)$ . Since  $A$  is compact, we may assume  $\{a_i\}$  converges to some  $a$ . When  $d(a_i, a) < r(a)/2$ ,

$$r(a_i) < r(a) + r(a)/4$$

and so

$$d(a, x_i) \leq d(a, a_i) + d(a_i, x_i) < r(a)/2 + [r(a) + r(a)/4] = 7r(a)/4$$

Since we chose  $r(a)$  to be half the supremum of compact sets, the sequence  $x_k$  will eventually end up in the compact ball  $B_{3r(a)/4}(a)$ , and hence will converge.

If  $A$  is a compact set, we will let  $A'$  be the compact set constructed above. Let  $A_0$  consist of an arbitrary point  $x_0$  in  $X$ , and inductively, define  $A_{k+1} = A'_k$ , and  $A = \bigcup_{i=0}^{\infty} A_k$ . Then  $A$  is the union of countably many compact sets.  $A$  is obviously open. If  $x$  is a limit point of  $A$ , then there is some sequence  $\{x_i\}$  in  $A$  which converges to  $x$ , so  $r(x_i) \rightarrow r(x)$ . If  $|r(x_i) - r(x)| < \varepsilon$ , and also  $d(x_i, x) < r(x) - \varepsilon$ , then  $x$  is contained in  $\bar{B}_{r(x_i)}(x_i)$ , and hence if  $x_i$  is in  $A_k$ , then  $x$  is in  $A_{k+1}$ . Thus  $A$  is non-empty and clopen, so  $X = A = \bigcup A_k$  is  $\sigma$ -compact.  $\square$

**Lemma 1.17** (4)  $\rightarrow$  (1). *A connected, locally compact, paracompact space is  $\sigma$  compact.*

*Proof.* Consider a locally-finite cover  $\mathcal{C}$  of precompact neighbourhoods in a space  $X$ . Fix  $x \in X$ . Then  $x$  intersects finitely many elements of  $\mathcal{C}$ , which we may label  $U_{1,1}, U_{1,2}, \dots, U_{1,n_1}$ . Then

$$U_1 = \overline{U_{1,1}} \cup \overline{U_{1,2}} \cup \dots \cup \overline{U_{1,n_1}}$$

intersects only finitely more elements of  $\mathcal{C}$ , since the set is compact, and we need only add finitely more open sets  $U_{2,1}, \dots, U_{2,n_2}$ , obtaining

$$U_2 = \overline{U_{2,1}} \cup \dots \cup \overline{U_{2,n_2}}$$

Continuing inductively, we find an increasing sequence of compact neighbourhoods. Then  $U = \bigcup U_i$  is open because a neighbourhood of  $y \in U_k$  is contained in  $U_{k+1}$ . If  $y$  is a limit point of  $U$ , take a neighbourhood  $V \in \mathcal{C}$ , which must intersect some  $U_k$ . Then  $y \in U_{k+1}$ , so  $U$  is closed. We conclude  $X = U$  is  $\sigma$  compact.  $\square$

**Lemma 1.18** (1)  $\rightarrow$  (4). *A  $\sigma$  compact, locally compact Hausdorff space is paracompact.*

*Proof.* Let  $X = \bigcup C_i$  be a locally compact,  $\sigma$ -compact space. Since  $C_1$  is compact, it is contained in an open precompact neighbourhood  $U_1$ . Similarly,  $C_2 \cup \overline{U_1}$  is contained in a precompact neighbourhood  $U_2$  with compact closure. We find  $U_1 \subset U_2 \subset \dots$ , each with compact closure, and which cover the entire space. Now let  $\mathcal{U}$  be an arbitrary open cover of  $X$ . Each  $V_k = U_k - \overline{U_{k-2}}$  (letting  $U_{-2} = U_{-1} = U_0 = \emptyset$ ) is open, and its closure  $\overline{V_k}$  is a closed subset of compact space, hence compact. Since  $\mathcal{U}$  covers  $\overline{V_k}$ , it has a finite subcover  $U_1, \dots, U_n$ , and we let

$$\mathcal{V}_1 = (U_1 \cap V_1), (U_2 \cap V_1), \dots, (U_n \cap V_1)$$

be a collection of refined open sets which cover  $V_1$ . Do the same for each  $V_k$ , obtaining  $\mathcal{V}_2, \mathcal{V}_3, \dots$ , and consider  $\mathcal{V} = \bigcup \mathcal{V}_i$ . Surely this is a cover of  $X$ , and each point is contained only in some  $\mathcal{V}_k$  and  $\mathcal{V}_{k+1}$ , so this refined cover is locally finite.  $\square$

## 1.7 Boundaries

There is an additional family of ‘manifolds with sides’ which often occurs in the theory of differential geometry. A *manifold with boundary* is a topological space  $M$  such that for each  $x \in M$ , there exists a neighbourhood  $U$  of  $x$  and a homeomorphism  $f : U \rightarrow \mathbb{R}^n$ , or a homeomorphism  $f : U \rightarrow \mathbf{H}^n$ , where

$$\mathbf{H}^n = \{x \in \mathbb{R}^n : x_1 \geq 0\}.$$

Invariance of domain can be used to show that only one of the two homeomorphisms can be constructed around a point. Given a manifold with boundary  $M$ , we let  $M^\circ$  denote the *interior of the manifold*, the family of all points which have neighbourhoods homeomorphic

to  $\mathbb{R}^n$ , and  $\partial M$  denote the *boundary of the manifold*, the family of all points which have neighbourhoods homeomorphic to  $\mathbf{H}^n$ . Of course, we can consider *coordinate half balls*  $(x, U)$ , where  $x$  is a homeomorphism between  $U$  and the  $B \cap \mathbf{H}^n$ , where  $B$  is the open unit ball centered at the origin.

**Theorem 1.19.** *If  $M$  is an  $n$  dimensional manifold with boundary, then  $\partial M$  is a manifold of dimension  $n - 1$ .*

*Proof.* Let  $x$  be a point in  $\partial M$ , and let  $U$  be a neighbourhood homeomorphic to  $\mathbf{H}^n$  by a map  $f : U \rightarrow \mathbf{H}^n$ . Consider the points in  $U$  that map to the boundary plane under  $f$ ,

$$V = \{y \in U : f(y) = (0, x_2, \dots, x_n)\}$$

We contend that  $V = U \cap \partial M$ , so that  $V$  is a neighbourhood of  $x$  in the relative topology, and since  $V$  is homeomorphic to  $\mathbb{R}^{n-1}$ , this will show that  $\partial M$  is an  $n - 1$  dimensional manifold. It is easy to see that  $V \subset U \cap \partial M$ . Conversely, the other points in  $U - V$  have a neighbourhood homeomorphic to  $\mathbb{R}^n$ , so that they do not lie in  $\partial M$ . Thus  $U - V \subset M^\circ \cap U$ , and this completes the proof.  $\square$

**Example.**  $\mathbf{H}^n$  is the easiest example of a manifold with boundary. Its boundary consists of  $\{0\} \times \mathbb{R}^{n-1}$ , which is an  $n - 1$  manifold. Another manifold with boundary is the unit disc  $\mathbf{D}^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ . We have already shown that the disc's boundary,  $\partial \mathbf{D}^n = S^{n-1}$ , is an  $n - 1$  manifold.

## 1.8 Invariance of Domain via Jordan Curves

For this section, we will prove invariance of domain, relying on two unproved (but ‘obviously true’) theorems.

**Theorem 1.20** (The Generalized Jordan Curve Theorem). *Every subspace  $X$  of  $\mathbb{R}^n$  homeomorphic to  $S^{n-1}$  splits  $\mathbb{R}^n - X$  into two components, and  $X$  is the boundary of each.*

**Theorem 1.21.** *If a subspace  $Y$  of  $\mathbb{R}^n$  is homeomorphic to the unit disc  $\mathbf{D}^n$ , then  $\mathbb{R}^n - Y$  is connected.*

We’ll put on the finishing touches to Invariance of Domain now. Hopefully this will give you intuition to why the theorem is true.

**Lemma 1.22.** *One of the components of  $\mathbb{R}^n - X$  is bounded, and the other is unbounded. We call the bounded component the inside of  $X$ , and the unbounded component the outside.*

*Proof.* Since  $X$  is homeomorphic to  $S^n$ , it is a compact set, and therefore contained in some ball  $B$ .  $\mathbb{R}^n - B$  is connected, so therefore one component of  $\mathbb{R}^n - X$  is contained within  $B$ . Since  $B$  is bounded, this component is bounded. If both components are bounded, we conclude that the union of the two components plus  $X$  is bounded, a contradiction. Therefore the other component is unbounded.  $\square$



**Lemma 1.23.** *If  $U \subset \mathbb{R}^n$  is open,  $A \subset U$  is homeomorphic to  $S^n$ ,  $f : U \rightarrow \mathbb{R}^n$  is one-to-one and continuous, and  $A \cup (\text{inside of } A)$  is homeomorphic to  $\mathbf{D}^n$ , then  $f(\text{inside of } A) = \text{inside of } f(A)$ .*

*Proof.* Since  $f$  is continuous,  $f(\text{inside of } A)$  is connected, and is therefore contained either entirely within the outside of  $f(A)$  or the inside of  $f(A)$ . The same is true of  $f(\text{outside of } A)$ . The difference is that, due to compactness,  $f(A \cup (\text{inside of } A))$  is homeomorphic to  $A \cup (\text{inside of } A)$ , and in connection, homeomorphic to  $\mathbf{D}^n$ . Therefore  $\mathbb{R}^n - f(A \cup \text{inside of } A)$  is connected. It follows that  $f(\text{inside of } A)$  is a component of  $\mathbb{R}^n - f(A)$ , so it is equal to either the inside or outside of space. Since  $f(\text{inside of } A)$  is contained within a bounded ball, we conclude that it is equal to the inside.  $\square$

**Theorem 1.24** (Invariance of Domain). *If  $f : U \rightarrow \mathbb{R}^n$  is an injective continuous function, where  $U$  is an open subset of  $\mathbb{R}^n$ , then  $f(U)$  is open, and therefore  $f$  is homeomorphic onto its image.*

*Proof.* Let  $V$  be an arbitrary open subset of  $U$ . We must show  $f(V)$  is also open. Let  $x \in V$  be arbitrary, and consider a closed ball  $\bar{B}$  containing  $x$ , and contained in  $V$ . The boundary of  $\bar{B}$  is homeomorphic to  $S^{n-1}$ , and the interior  $B$  is equal to the inside of  $\bar{B}$ . By the lemma (2.4) above, we conclude that

$$f(B) = \text{inside of } f(\partial B)$$

Since  $\partial B$  is closed in  $\mathbb{R}^n$ , the inside is open, hence  $f(B)$  is open. By an extension of this argument, we have shown the the image of any open set is open, so invariance of domain is proved.  $\square$

The unproved theorems we rely on here require quite advanced techniques in algebraic topology, which we will not discuss in these notes for quite some time.

# Chapter 2

## Differentiable Structures

As a topological space, we know when a map between manifolds is continuous, but when is a map differentiable? What we seek is a definition abstract enough to work on any manifold, yet possessing the same properties of differentiable functions on  $\mathbb{R}^n$ .

### 2.1 Defining Differentiability

Let us be given a map  $f : M \rightarrow N$  between manifolds. Given a correspondence  $b = f(a)$ , a reasonable inquiry would be to consider two charts  $(x, U)$  and  $(y, V)$ , where  $U$  is a neighbourhood of  $a$  and  $V$  is a neighbourhood of  $b$ . We obtain a map  $y \circ f \circ x^{-1}$ , defined between open subsets of Euclidean space. We have thus ‘expressed  $f$  in coordinates’, and we can consider  $f$  differentiable at  $p \in M$  if  $y \circ f \circ x^{-1}$  is differentiable at  $x(a)$ . Unfortunately, without additional structure this idea is doomed to fail, since charts on a topological manifold are only defined up to a homeomorphism.

**Example.** Consider the chart  $y : \mathbb{R} \rightarrow \mathbb{R}$ ,  $y(t) = t^3$ , and let  $x$  be the identity chart. If  $f(x) = \sin(x)$ , then  $x \circ f \circ x^{-1} = f$  is differentiable, yet

$$(x \circ f \circ y^{-1})(t) = \sin(t^{1/3})$$

is not differentiable at the origin.

If we are to stick with this definition, we must identify additional structure on topological manifolds. Our method will be to identify ‘differentiable charts’, which can be used to determine whether a map is differentiable, so we can ignore ‘non-differentiable’ charts. Two charts  $(x, U)$  and  $(y, V)$  are  $C^\infty$  related, if either  $U \cap V = \emptyset$ , or  $U \cap V$  is nonempty, and the two functions

$$y \circ x^{-1} : x(U \cap V) \rightarrow y(U \cap V)$$

$$x \circ y^{-1} : y(U \cap V) \rightarrow x(U \cap V)$$

are  $C^\infty$  functions. One can see a chart as laying a blanket down onto a manifold. Two charts are  $C^\infty$  related if, when we lay them down over each other, they contain no creases! The

fact that manifolds do not have a particular preference for coordinates is both a help and a hindrance. On one side, it forces us to come up with elegant, coordinate free approaches to geometry. On the other end, these coordinate free approaches can become quite abstract.

A *smooth* or  $C^\infty$  *atlas* for a manifold  $M$  is a family of  $C^\infty$  related charts  $\{(x_\alpha, U_\alpha)\}$ , such that  $\{U_\alpha\}$  is a cover of  $M$ . A maximal atlas is called a *smooth structure* on a manifold, and a manifold together with a smooth structure is called a *smooth* or *differentiable manifold*. The coordinate charts in the atlas of a smooth manifold are referred to as smooth, or curvilinear. In the literature, each map  $y \circ x^{-1}$  is known as a *transition map*, so that we say an atlas for a manifold has  $C^\infty$  transition maps. From now on, when we mention a chart on a differentiable manifold, we implicitly assume the chart is the member of the smooth structure of the manifold.

Let  $f : M \rightarrow N$  be a map between two smooth manifolds. Then  $f$  is *differentiable* at  $p \in M$  if it is continuous at  $p$ , and if for some chart  $(x, U)$  on  $M$  and  $(y, V)$  on  $N$  with  $p \in U$  and  $f(p) \in V$ , the map

$$y \circ f \circ x^{-1} : x(f^{-1}(V) \cap U) \rightarrow \mathbb{R}^m$$

is differentiable at  $x(p)$ . The function  $f$  itself is *differentiable* if it is differentiable at every point on its domain, or correspondingly, if  $y \circ f \circ x^{-1}$  is differentiable for any two charts  $x$  and  $y$ . We say the function is  $C^k$ , if all partial derivatives of  $y \circ f \circ x^{-1}$  of order  $\leq k$  are continuous, and  $C^\infty$ , or *smooth* if partial derivatives of all orders are continuous. If  $f : M \rightarrow N$  is bijective, and both  $f$  and  $f^{-1}$  are  $C^k$  differentiable maps, then we say  $f$  is a  $C^k$  *diffeomorphism*. In these notes, a diffeomorphism will always refer to a smooth diffeomorphism, i.e. a  $C^\infty$  diffeomorphism, unless otherwise specified. Since differentiability is a *local* condition on Euclidean spaces, and manifolds are locally Euclidean spaces, the smooth structure, which preserves the differentiability across all local charts, allows us to define differentiability on arbitrary manifolds.

**Lemma 2.1.** *Every atlas extends to a unique smooth structure.*

*Proof.* Let  $\mathcal{A}$  be an atlas for a manifold  $M$ , and consider the set  $\mathcal{A}^*$ , which is the union of all atlases containing  $\mathcal{A}$ . We shall show that  $\mathcal{A}^*$  is also an atlas, and therefore necessarily the unique maximal one. Let  $(x, U)$  and  $(y, V)$  be two charts in  $\mathcal{A}^*$ , with  $U \cap V \neq \emptyset$ . For each  $p \in U \cap V$ , let  $(z, W)$  be a chart in  $\mathcal{A}$  with  $p \in W$ . Then for  $a \in y(U \cap V \cap W)$ , we have

$$(x \circ y^{-1})(a) = [(x \circ z^{-1}) \circ (z \circ y^{-1})](a).$$

By assumption,  $x \circ z^{-1}$  and  $z \circ y^{-1}$  are  $C^\infty$  maps on  $U \cap V \cap W$ , since  $x$  and  $y$  are in some atlas containing  $z$ . Thus  $x \circ y^{-1}$  is  $C^\infty$  on  $U \cap V \cap W$ . Since  $p$  was arbitrary, we can take unions over all neighbourhoods  $W$  of points to conclude that  $x \circ y^{-1}$  is smooth on  $U \cap V$ . Of course, this argument also implies  $y \circ x^{-1}$  is smooth, so  $x$  and  $y$  are  $C^\infty$  related. Thus  $\mathcal{A}^*$  is also an atlas.  $\square$

This implies, in particular, that we may specify a smooth structure by just giving a family of  $C^\infty$  related transition maps covering a manifold, which is almost always easier to specify than giving a complete maximal atlas.

**Corollary 2.2.** *If  $x$  is a chart defined on a differentiable manifold  $M$ , and is  $C^\infty$  related to each map in a generating atlas  $\mathcal{A}$ , then  $x$  is in the smooth structure generated by  $\mathcal{A}$ .*

**Lemma 2.3.** *Let  $f : M \rightarrow N$  be a map, and suppose there are smooth charts  $(x_1, U_1)$  and  $(y_1, V_1)$  on  $M$  and  $N$ , with  $p \in U_1 \cap f^{-1}(V_1)$ , such that  $y_1 \circ f \circ x_1^{-1}$  is differentiable at  $x_1(p)$ . Then if  $(x_2, U_2)$  and  $(y_2, V_2)$  are smooth charts on  $M$  and  $N$ , with  $p \in U_2 \cap f^{-1}(V_2)$ , then  $y_2 \circ f \circ x_2^{-1}$  is differentiable at  $x_2(p)$ .*

*Proof.* We find that for  $a \in x_2(U_1 \cap U_2 \cap f^{-1}(V_1) \cap f^{-1}(V_2))$ ,

$$(y_2 \circ f \circ x_2^{-1})(a) = [(y_2 \circ y_1^{-1}) \circ (y_1 \circ f \circ x_1^{-1}) \circ (x_1 \circ x_2^{-1})](a).$$

Clearly  $x_1 \circ x_2^{-1}$  is differentiable at  $x_2(p)$ ,  $y_1 \circ f \circ x_1^{-1}$  is differentiable at  $x_1(p)$ , and  $y_2 \circ y_1^{-1}$  is differentiable at  $y_1(f(p))$ , so the chain rule implies  $y_2 \circ f \circ x_2^{-1}$  is differentiable at  $x_2(p)$ .  $\square$

For simplicity, in these notes we assume almost all the objects we consider are smooth, i.e.  $C^\infty$ . This is mostly a convenience, since it means we can freely differentiate as many times as we desire, and do not have to constantly list the number of times an object must be differentiated before we can apply a certain technique. Transferring these results to less smooth objects is often just a matter of checking how many times an object is differentiable.

However, sometimes it is useful to only ascribe a ' $C^k$  structure' to a manifold, for  $k \geq 1$ , rather than a ' $C^\infty$  structure'. Two charts  $(x, U)$  and  $(y, V)$  are  $C^k$  related if  $U \cap V = \emptyset$ , or  $U \cap V \neq \emptyset$  and  $y \circ x^{-1}$  is a  $C^k$  map on Euclidean space. Thus a  $C^k$  manifold is a manifold equipped with a  $C^k$  atlas, and on a  $C^k$  manifold we can consider differentiable maps, and moreover,  $C^i$  maps for all  $i \leq k$ . On the other hand, it does not make sense to talk about  $C^i$  maps for  $i > k$ , or  $C^\infty$  maps. Topologically speaking, the family of  $C^k$  manifolds is no more general than a  $C^\infty$  manifold, since a deep result of differential topology shows any  $C^k$  manifold can be given a  $C^\infty$  structure. But sometimes we can equip a manifold with a  $C^k$  atlas naturally, but it is not natural to equip it with a smooth structure. Our first example of this occurs later on in this chapter during our discussion of submanifolds.

We can also consider an even more rigid structure than a  $C^\infty$  atlas. If we require transition maps between charts to be (real) analytic, and we build up an atlas of charts compatible with respect to this requirement, known as a  $C^\omega$  atlas, we obtain a *real-analytic manifold*. If the dimension of the manifold is equal to  $2n$ , and we identify  $\mathbb{R}^{2n}$  with  $\mathbb{C}^n$ , then we can require our transition maps to be holomorphic, and by doing so, we obtain the family of *complex manifolds*.

**Example.** *Let  $M$  be a manifold, and  $U$  an open submanifold. Then we can define an atlas on  $U$ , by considering all smooth charts  $(x, V)$  on  $M$ , with  $V \subset U$ . This is a maximal atlas, and is the unique smooth structure such that the following properties hold:*

1. *If  $f : M \rightarrow N$  is differentiable, then  $f|_U : U \rightarrow N$  is differentiable.*
2. *The inclusion map  $i : U \rightarrow M$  is differentiable.*
3. *If  $f : N \rightarrow M$  is differentiable, and  $f(N) \subset U$ , then  $f : N \rightarrow U$  is differentiable.*

Later on, we will see these results can also be extended to lower dimensional ‘submanifolds’ of  $M$ .

**Example.** Consider the manifold  $\mathbb{R}^n$ , and define a smooth structure by considering the generating atlas containing only the identity map. This defines a smooth structure on  $\mathbb{R}^n$ , such that

1.  $(x, U)$  is a chart on  $\mathbb{R}^n$  if and only if  $x : U \rightarrow \mathbb{R}^n$  is a diffeomorphism onto an open subset of  $\mathbb{R}^n$  in the classical sense.
2. A map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable in the sense of a manifold if and only if it is differentiable in the classical sense.
3. A map  $f : M \rightarrow \mathbb{R}^n$  is differentiable if and only if each coordinate map  $f^i : M \rightarrow \mathbb{R}$  is differentiable.
4. A chart  $x : U \rightarrow \mathbb{R}^n$  is a diffeomorphism from  $U$  to  $x(U)$ .

Our definition thus contains the classical calculus as a special case.

We note that the transition maps given to define all the topological manifolds in the previous chapter are all  $C^\infty$  related to one another. It is left as an exercise to check this is true. Thus all the topological manifolds in the previous chapter are also differentiable manifolds. There are some manifolds which do not have a continuous structure – the creases in the manifold cannot be evened out. But these occur in certain pathological situations that we won’t touch upon in these introductory notes.

**Example.** On  $\mathbb{R}^2 - \{0\}$ , we can consider the polar coordinate system. A polar coordinate chart  $((\theta, r), U)$  is a chart such that  $r(z) = |z|$  for all  $z \in U$ , and on  $U$ ,

$$x = r \cos(\theta) \quad \text{and} \quad y = r \sin(\theta).$$

Assuming elementary knowledge of differential forms, the two equations above imply that on  $U$ ,

$$dr = \frac{xdy + ydx}{(x^2 + y^2)^{1/2}},$$

and

$$d\theta = \frac{xdy - ydx}{x^2 + y^2}.$$

The forms on the right hand side are invariant of the particular function  $r$  or  $\theta$  chosen, and define a differential form on  $\mathbb{R}^2 - \{0\}$ . A simple computation verifies that

$$d\left(\frac{xdy - ydx}{x^2 + y^2}\right) = 0.$$

By Poincare's lemma, for any simple connected open set  $U \subset \mathbb{R}^2 - \{0\}$ , there exists a smooth function  $\theta : U \rightarrow \mathbb{R}$  such that

$$d\theta = \frac{xdy - ydx}{x^2 + y^2}.$$

We let  $r : U \rightarrow (0, \infty)$  be defined by  $r(z) = |z|$ . Another purely algebraic argument shows that  $dx = d(r \cos \theta)$  and  $dy = d(r \sin \theta)$ , and so there exists  $A, B \in \mathbb{R}$  such that  $x = r \cos(\theta) + A$  and  $y = r \sin(\theta) + B$ . The fact that  $r^2 = x^2 + y^2$  implies that we must actually have  $A = B = 0$ . Thus we find that  $(\theta, r)$  is a polar coordinate chart on  $\mathbb{R}^2 - \{0\}$ , and moreover, such systems exist on any simply connected subset of  $\mathbb{R}^2 - \{0\}$ .

**Example.** On  $\mathbb{R}^3 - \mathbb{R}e_3$ , we can consider spherical coordinate systems, i.e. coordinate charts  $((r, \phi, \theta), U)$  such that for each  $(x, y, z) \in U$ ,

$$\begin{aligned} x &= r \cos(\theta) \cos(\phi) \\ y &= r \cos(\theta) \sin(\phi) \\ z &= r \sin(\theta). \end{aligned}$$

One can verify that the map  $(a, b, c) \mapsto (a \cos(b) \cos(c), a \cos(b) \sin(c), a \sin(b))$  has derivative

$$\begin{pmatrix} \cos(b) \cos(c) & -a \sin(b) \cos(c) & -a \cos(b) \sin(c) \\ \cos(b) \sin(c) & -a \sin(b) \sin(c) & a \cos(b) \cos(c) \\ \sin(b) & a \cos(b) & 0 \end{pmatrix}.$$

The determinant of this, after simplification, is

$$-a^2 \cos(b)(\cos^2(b) \cos^2(c) + \sin^2(b) \cos^2(c) + \sin^2(c)).$$

Thus the map has full rank precisely when  $b \notin \pi/2 + n\mathbb{Z}$ , and thus the inverse function theorem guarantees the existence of a spherical coordinate system locally about any point in  $\mathbb{R}^3 - \{0\}$  except those points of the form  $(0, 0, t)$ . As in the last example, the Poincare lemma will guarantee the existence of a spherical coordinate system on any simply connected subset of  $\mathbb{R}^3 - \mathbb{R}e_3$ .

**Example.** On  $\mathbb{R}^3 - \mathbb{R}e_3$ , we can also consider cylindrical coordinate systems, i.e. triples  $(r, \theta, z)$  such that

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \\ z &= z \end{aligned}$$

One verifies by the inverse function theorem that cylindrical coordinates exist on any simply connected subset of  $\mathbb{R}^3 - \mathbb{R}e_3$ .

**Example.** More generally, on the set  $\mathbb{R}^n - \text{span}(e_3, \dots, e_d)$ , we can consider generalized spherical coordinate systems, given by maps  $(r, \theta_1, \dots, \theta_{d-1})$ , such that

$$\begin{aligned} x_1 &= r \cos(\theta_1) \dots \cos(\theta_n) \\ x_2 &= r \cos(\theta_1) \dots \cos(\theta_{n-1}) \sin(\theta_n) \\ x_3 &= r \cos(\theta_1) \dots \cos(\theta_{n-2}) \sin(\theta_{n-1}) \\ &\vdots \\ x_d &= r \sin(\theta_1). \end{aligned}$$

Generalized spherical coordinate systems exist on any simply connected subset of  $\mathbb{R}^n - \text{span}(e_3, \dots, e_d)$ .

**Example.** The differentiable structure on  $S^n$  is defined by the stereographic projection maps. Equivalently, we may also define this structure on  $S^1$  by the angle functions. The generalized spherical coordinates also specify this structure, but require a rotation in higher dimensions due to singularities when  $n \geq 2$ , so results get exponentially more complicated. To verify that the angle functions are differentiable, we note that if  $(\theta, U)$  and  $(\psi, V)$  are angle functions, then  $\theta \circ \psi^{-1}$  is just a translation by a multiple of  $2\pi$  on each connected component of  $\psi(U \cap V)$ , hence  $C^\infty$ . Indeed, if

$$\psi^{-1}(t) = e^{it} = \theta^{-1}(t')$$

then  $t = t' + 2\pi n$  for some unique integer  $n$ . Define  $f(t) = n$ , giving us a map  $f : U \cap V \rightarrow \mathbb{Z}$ . This map is continuous because if  $t_i \rightarrow t$ , then

$$(\theta \circ \psi^{-1})(t_i) \rightarrow (\theta \circ \psi^{-1})(t)$$

so  $t_i + 2\pi f(t_i) \rightarrow t + 2\pi f(t)$ , hence  $f(t_i) \rightarrow f(t)$ . The continuity of  $f$  implies that  $f$  is constant on every connected component of  $U \cap V$ .

**Example (Differentiable Product).** If  $M$  and  $N$  are differentiable manifolds, we may consider an atlas on  $M \times N$  with the differentiable structure generated by all maps  $x \times y$ , where  $x$  is a chart on  $M$  and  $y$  is a chart on  $N$ . From this definition, we find  $f \times g : X \rightarrow M \times N$  is differentiable if and only if  $f : X \rightarrow M$  and  $g : X \rightarrow N$  are differentiable. This is the unique differentiable structure on  $M \times N$  which has this property, since the property determines the charts on the manifold.

**Example (Differentiable Quotients and  $\mathbf{P}^n$ ).** If  $N$  is a quotient space of a differentiable manifold  $M$  whose projection  $\pi : M \rightarrow N$  is locally injective, then we may ascribe a differentiable structure to it. We take all smooth charts  $x : U \rightarrow \mathbb{R}^n$  on  $M$  such that  $U$  is homeomorphic to  $\pi(U)$  by  $\pi$ . We may then push the chart onto  $N$ , and all the charts placed down on  $N$  will be  $C^\infty$  related. As a covering, this can be extended to a maximal atlas. In fact, this is the unique structure on  $N$  which causes  $f : N \rightarrow X$  to be differentiable if and only if  $f \circ \pi : M \rightarrow X$  is differentiable. For instance, this gives the projective plane  $\mathbf{P}^n$  and the Möbius strip  $\mathbf{M}$  a canonical smooth structure.

**Example.** Smooth structures on manifolds are not unique. Let  $\mathbb{R}_1$  be the canonical smooth manifold on  $\mathbb{R}$ . Let  $\mathbb{R}_2$  be the smooth structure on  $\mathbb{R}$  generated by the map  $x$ , such that  $x(t) = t^3$ . Then  $\mathbb{R}_1$  and  $\mathbb{R}_2$  are diffeomorphic. Let  $x : \mathbb{R}_2 \rightarrow \mathbb{R}_1$  be our diffeomorphism. It is surely bijective. Let  $y$  be a chart on  $\mathbb{R}_2$ . We must verify that  $y = z \circ x$ , where  $z$  is a chart on  $\mathbb{R}_1$ . We may show this by verifying that  $y \circ x^{-1} = z$ , and  $x \circ y^{-1} = z^{-1}$  is  $C^\infty$  on  $\mathbb{R}_1$ . But this was exactly why  $y$  was a chart on  $\mathbb{R}_2$  in the first place, hence the map is a diffeomorphism. Later, we will show that most manifolds which lie in Euclidean space have a natural smooth structure, however, and this will be taken as the canonical smooth structure in any of these cases.

## 2.2 Smooth Functions on a Manifold

The set of all scalar-valued differentiable maps defined on a manifold  $M$  form an algebra  $C^\infty(M)$ . Note that a continuous map  $f : M \rightarrow N$  induces an algebra homomorphism  $f^\# : C(N) \rightarrow C(M)$  defined by  $f^\#(g) = g \circ f$ . If  $f$  and  $g$  are  $C^\infty$ , then  $g \circ f$  is  $C^\infty$ , and so we may restrict  $f^\#$  to a map from  $C^\infty(N)$  to  $C^\infty(M)$ . We see therefore that the ‘map’  $C^\infty$  defines a contravariant functor from the category of differential manifolds to the category of algebras over the real numbers.

**Lemma 2.4.** A continuous map  $f : M \rightarrow N$  between manifolds is smooth if and only if  $f^\#(C^\infty(N)) \subset C^\infty(M)$ .

*Proof.* Suppose  $f^\#(C^\infty(N)) \subset C^\infty(M)$ . Let  $(y, V)$  be a chart on  $N$  at a point  $q$ , and let  $(x, U)$  be a chart on  $M$  at  $p \in f^{-1}(q)$ . By assumption, each  $y^i \circ f = f^\#(y^i)$  is differentiable, so that  $y^i \circ f \circ x^{-1}$  is differentiable. But this implies  $y \circ f \circ x^{-1}$  is differentiable, so  $f$  is differentiable.  $\square$

**Theorem 2.5.** A homeomorphism  $f : M \rightarrow N$  is a diffeomorphism if and only if  $f^\#$  is an isomorphism between  $C^\infty(N)$  and  $C^\infty(M)$ .

*Proof.* Given  $g = f^{-1}$ , we note that  $(g \circ f)^\# = f^\# \circ g^\#$  is just the identity map. Thus if  $f$  is a diffeomorphism, then  $f^\# \circ g^\#$  is the identity, so that  $f^\#$  is invertible, and hence an isomorphism. Conversely, if  $f^\#$  is a bijection between  $C^\infty(N)$  and  $C^\infty(M)$ , then  $f^\#(C^\infty(N)) = C^\infty(M) \subset f^\#(C^\infty(M))$ , so  $f$  is differentiable, and  $g^\#(C^\infty(M)) = C^\infty(N) \subset C^\infty(N)$ , so  $g$  is also differentiable, hence  $f$  is a diffeomorphism.  $\square$

*Remark.* More generally,  $f$  is  $C^k$  if and only if  $f^\#(C^k(N)) \subset C^k(M)$ , and a  $C^k$  diffeomorphism if and only if  $f$  maps  $C^k(N)$  isomorphically to  $C^k(M)$ .

Suppose we know  $C^\infty(M)$  for all manifolds  $M$ . Then we may recover the smooth structure on  $M$ , which is the set of diffeomorphisms from open subsets of  $M$  to open subsets of euclidean space. We can actually define a  $C^\infty$  manifold in a completely algebraic way. Given some topological space  $X$ , suppose that from the sheaf of continuous functions on  $X$ , we consider a subsheaf  $C^\infty(U)$ , such that every point  $p \in M$  has a neighbourhood  $U$  and



$x^1, \dots, x^n \in C^\infty(U)$  such that  $x = (x^1, \dots, x^n)$  is a homeomorphism of  $U$  with an open subset of  $\mathbb{R}^n$ , and  $f \in C^\infty(U)$  if and only if  $f \circ x^{-1}$  is a  $C^\infty$  function. Then there is a unique  $C^\infty$  structure on  $X$  such that the  $C^\infty$  real-valued maps on any open set  $U$  are precisely the elements of  $C^\infty(U)$ . These facts constitute the foundation of the algebraic viewpoint of function theory, which attempts to uncover the nature of manifolds solely by determining how geometric influences the algebraic properties of the commutative algebra  $C^\infty(M)$ . You can actually get pretty far with this approach: Nestruev's book "Smooth Manifolds and Observables" attempts to introduce differential geometry solely in this manner, but we prefer to introduce the geometric and algebraic approach simultaneously for maximal insight.

## 2.3 Partial Derivatives

In calculus, when a function is differentiable, we obtained a derivative, a measure of a function's local change. On manifolds, determining an analogous object is difficult due to the lack of a specified coordinate system. For now, we shall stick to structures corresponding to some particular set of coordinates. Consider a differentiable map  $f \in C^1(M)$ . We have no conventional coordinates to consider partial derivatives on, but if we fix some chart  $x : U \rightarrow \mathbb{R}^n$  on  $M$ , we obtain a differentiable map  $f \circ x^{-1}$ , and we define, for a point  $p \in U$ ,

$$\left. \frac{\partial f}{\partial x^k} \right|_p = D_k(f \circ x^{-1})(x(p))$$

Geometrically, this is the change in the function  $f$  when we trace the function along the coordinate lines from the map  $x$ ; literally, if we define a curve  $c(t) = (f \circ x^{-1})(x(p) + te_k)$ , then

$$c'(0) = \left. \frac{\partial f}{\partial x^k} \right|_p$$

Sometimes we use the notation  $\partial_{x^k} f$  for the partial derivative in the  $k$ 'th direction, which simplifies the notation in heavy calculations.

**Theorem 2.6.** *If  $(x, U)$  and  $(y, V)$  are charts at a point  $p$ , and  $f : M \rightarrow \mathbb{R}$  is differentiable at  $p$ , then*

$$\left. \frac{\partial f}{\partial x^i} \right|_p = \sum_j \left. \frac{\partial y^j}{\partial x^i} \right|_p \left. \frac{\partial f}{\partial y^j} \right|_p$$

*Proof.* We just apply the chain rule in Euclidean space.

$$\begin{aligned} \left. \frac{\partial f}{\partial x^i} \right|_p &= D_i(f \circ x^{-1})(x(p)) = D_i((f \circ y^{-1}) \circ (y \circ x^{-1}))(x(p)) \\ &= \sum_j D_j(f \circ y^{-1})(y(p)) D_i(y_j \circ x^{-1})(x(p)) = \sum_j \left. \frac{\partial f}{\partial y^j} \right|_p \left. \frac{\partial y_j}{\partial x^i} \right|_p. \quad \square \end{aligned}$$

**Example.** Let us compute the laplacian on  $\mathbb{R}^2$  in polar coordinates.

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

To do this, we need to relate partial derivatives by the chain rule. If  $(r, \theta)$  is the polar coordinate chart, and  $(x, y)$  the standard chart on  $\mathbb{R}^2$ , then

$$x = r \cos(\theta) \quad y = r \sin(\theta)$$

(note that this is a relation between functions, and can be applied pointwise at any point on the charts). Thus the matrix of partial derivatives is

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

We can invert this matrix to obtain the partial derivatives with respect to  $x$  and  $y$ . We have

$$\begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r \cos(\theta) & r \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\frac{\sin(\theta)}{r} & \frac{\cos(\theta)}{r} \end{pmatrix}$$

Now we apply the chain rule. We have

$$\frac{\partial}{\partial x} = \cos(\theta) \frac{\partial}{\partial r} - \frac{\sin(\theta)}{r} \frac{\partial}{\partial \theta} \quad \frac{\partial}{\partial y} = \sin(\theta) \frac{\partial}{\partial r} + \frac{\cos(\theta)}{r} \frac{\partial}{\partial \theta}$$

So

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= \left( \cos(\theta) \frac{\partial}{\partial r} - \frac{\sin(\theta)}{r} \frac{\partial}{\partial \theta} \right) \left( \cos(\theta) \frac{\partial f}{\partial r} - \frac{\sin(\theta)}{r} \frac{\partial f}{\partial \theta} \right) \\ &= \cos^2(\theta) \frac{\partial^2 f}{\partial r^2} + \frac{\cos(\theta) \sin(\theta)}{r^2} \frac{\partial f}{\partial \theta} - \frac{\cos(\theta) \sin(\theta)}{r} \frac{\partial^2 f}{\partial r \partial \theta} \\ &\quad + \frac{\sin^2(\theta)}{r} \frac{\partial f}{\partial r} - \frac{\sin(\theta) \cos(\theta)}{r} \frac{\partial^2 f}{\partial \theta \partial r} + \frac{\sin(\theta) \cos(\theta)}{r^2} \frac{\partial f}{\partial \theta} + \frac{\sin^2(\theta)}{r^2} \frac{\partial^2 f}{\partial \theta^2} \\ \frac{\partial^2 f}{\partial y^2} &= \left( \sin(\theta) \frac{\partial}{\partial r} + \frac{\cos(\theta)}{r} \frac{\partial}{\partial \theta} \right) \left( \sin(\theta) \frac{\partial f}{\partial r} + \frac{\cos(\theta)}{r} \frac{\partial f}{\partial \theta} \right) \\ &= \sin^2(\theta) \frac{\partial^2 f}{\partial r^2} - \frac{\cos(\theta) \sin(\theta)}{r^2} \frac{\partial f}{\partial \theta} + \frac{\cos(\theta) \sin(\theta)}{r} \frac{\partial^2 f}{\partial r \partial \theta} \\ &\quad + \frac{\cos^2(\theta)}{r} \frac{\partial f}{\partial r} + \frac{\sin(\theta) \cos(\theta)}{r} \frac{\partial^2 f}{\partial \theta \partial r} - \frac{\sin(\theta) \cos(\theta)}{r^2} \frac{\partial f}{\partial \theta} + \frac{\cos^2(\theta)}{r^2} \frac{\partial^2 f}{\partial \theta^2} \end{aligned}$$

All the nastiness cancels out by use of the trigonometric identities, and we find

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r} \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \theta^2}$$

Which gives a simple radial description of the Laplacian. It makes sense that such a radial expansion exists, because the Laplacian describes the second averages of  $f$  on balls around a point.

Partial derivatives also satisfy a nice ‘derivation’ property, which we leave to the reader to calculate. It is essentially the product rule for partial derivatives. We shall see later that this property characterizes the partial derivative maps in the space of linear maps on  $C^\infty(M)$ , which relates to the ‘intrinsic’ tangent space of a manifold.

**Lemma 2.7.** *If  $f$  and  $g$  are differentiable maps from a manifold  $M$  to  $\mathbb{R}$ , then*

$$\left. \frac{\partial fg}{\partial x^i} \right|_p = f(p) \left. \frac{\partial g}{\partial x^i} \right|_p + g(p) \left. \frac{\partial f}{\partial x^i} \right|_p$$

The partial derivatives provide immediate quantities to measure the rate of change of a real-valued differentiable function on a manifold. Next, we will consider a ‘coordinate-independent’ way to measure this rate of change for any two functions, by the introduction of ‘infinitesimals’ on manifolds.

## 2.4 The Differential Map

The coordinate operators allow us to extend differentiation to functions in  $C^\infty(M)$ , for any manifold  $M$ . However, defining the derivative of a differentiable map  $f : M \rightarrow N$  between arbitrary manifolds is more tricky. We could define the derivative coordinatewise at a point  $p$  by considering the derivative  $D(y \circ f \circ x^{-1})(x(p))$ , for some coordinate systems  $x$  and  $y$ . The trouble is that in this way we can only talk of properties of the derivative that are invariant under the coordinate systems chosen. The trouble here is that there is no ‘definite’ space the operator is defined over – as we change the coordinate systems, the space changes. To obtain a ‘universal’ coordinate independent map, we need to form a ‘universal’ space which represents all coordinates at the same time, upon which the derivative operators are invariant. This space is known as the tangent bundle. There are deep and elegant constructions for this bundle, but for now, we only require the absolute basics.

Vectors  $v$  in  $\mathbb{R}^n$  are often pictured as starting at the origin, and ending at the point with the same coordinates as  $v$ . But it is often convenient to picture these vectors as starting at a different beginning point than the origin. We introduce the notation  $v_p$ , where  $p$  is a point in Euclidean space, and  $v$  is a vector, to be the vector beginning at  $p$  and ending at  $p + v$ . One way to think of these values are as points  $p$  in  $\mathbb{R}^n$  with an added ‘first order’ infinitesimal shift in the direction  $v$ . If  $U$  is an open subset of  $\mathbb{R}^n$ , we can define the *tangent bundle* of  $U$  to be the space  $TU = \{v_p : p \in U, v \in \mathbb{R}^n\}$ . At each point  $p \in U$ , the *fibre* at  $p$ , denoted  $T_p U$ , consists of all tangent vectors beginning at  $p$ , which can be made into a vector space by defining  $v_p + w_p = (v + w)_p$  and  $\lambda v_p = (\lambda v)_p$ . We now generalize this process to arbitrary manifolds not necessarily lying in Euclidean space.

There are many complex and elegant ways to form the tangent bundle on a differentiable manifold. We will eventually discuss them in time, but we now construct the space in the most quick and dirty way. On a differentiable manifold  $M$ , we locally specify the tangent bundle in charts, and then patch them together into a reasonable structure on the entire manifold. For each point  $p$ , we let the fibre  $T_p M$  consist of equivalence classes of tuples of the

form  $(x, v)_p$ , where  $v \in \mathbb{R}^n$ , and  $(x, U)$  is a chart with  $p \in U$ , where we identify  $(x, v)_p$  and  $(y, w)_p$  if  $w = D(y \circ x^{-1})(p)(v)$ . In other words,  $w$  is identified by  $v$  if it is obtained from  $v$  by a change in coordinates. Putting this altogether gives the required tangent vectors  $[x, v]_p$ . The fibres  $T_p M$  are still vector spaces, obtained by defining  $[x, v]_p + [x, w]_p = [x, v + w]_p$ , and  $\lambda[x, v]_p = [x, \lambda v]_p$ . The set  $TM$  is then constructed as the union of the vector spaces  $T M_p$ , and is the natural generalization of the tangent bundle to an open subset of Euclidean space.

If  $f : U \rightarrow V$  maps an open subset  $U \subset \mathbb{R}^n$  to  $V \subset \mathbb{R}^m$ , and is differentiable, then we can consider the derivatives  $Df(p)$  at  $p \in U$ , which describe how the function acts locally around  $p$ . Since a differentiation indicates that the space is ‘locally linear’, then the derivative should act on tangent vectors by the actual linear map. In other words, we can put all the derivatives together to define the *covariant derivative* of  $f$ , denoted  $f_*$ , by  $f_*(v_p) = (Df(p)(v))_{f(p)}$ . The linear map obtained by restriction of  $f_*$  to the domain  $T_p U$  and codomain  $T_{f(p)} V$  is denoted  $f_*|_p$ .

For a differentiable map  $f : M \rightarrow N$ , we can define  $f_* : TM \rightarrow TN$  by mapping  $[x, v]_p$  to  $[y, w]_{f(p)}$ , where  $w = D(y \circ f \circ x^{-1})(p)(v)$ . This is easily shown to be independent of the coordinates chosen. Thus we obtain linear maps  $f_*|_p$  from  $T M_p$  to  $T N_{f(p)}$ . The chain rule for derivative can be succinctly represented by the fact that if  $f : M_1 \rightarrow M_2$  and  $g : M_2 \rightarrow M_3$  are differentiable maps, then  $(g \circ f)_* = g_* \circ f_*$ . We say that  $f$  has *rank*  $k$  at a point  $p$  if  $f_*|_p$  is a rank  $k$  linear map. The next section will show the rank gives essentially all the local information about a differentiable map if we are free to choose arbitrary smooth coordinate systems.

## 2.5 The Rank Theorems

The rank theorems provide the existence of coordinates on a manifold which simplify how the maps operate immensely. They are essentially an extension of the Euclidean inverse and implicit function theorems to general smooth functions on manifolds.

**Theorem 2.8.** *If  $f : M \rightarrow N$  is a smooth map with rank  $k$  at a point  $p$ , there is a coordinate chart  $(x, U)$  at  $p$  and a coordinate chart  $(y, V)$  at  $f(p)$  such that for each  $a \in x(U \cap f^{-1}(V))$ ,*

$$(y \circ f \circ x^{-1})(a_1, \dots, a_n) = (a_1, \dots, a_k, *, \dots, *),$$

where the asterixes denote arbitrary smooth functions of  $a$ .

*Proof.* Let  $(x, U)$  and  $(y, V)$  be arbitrary coordinate systems around  $p$  and  $f(p)$ . By a permutation the coordinates of  $x$  and  $y$ , we may guarantee that the matrix

$$\left( \frac{\partial y^i \circ f}{\partial x_j} \Big|_p \right)_{i,j=1}^k$$

is invertible. Define a map  $z : U \cap f^{-1}(V) \rightarrow \mathbb{R}^n$  around  $p$  by setting  $z^i = y^i \circ f$ , for  $1 \leq i \leq k$ , and  $z^i = x^i$  otherwise. The matrix

$$D(z \circ x^{-1})(p) = \begin{pmatrix} \left( \left( \frac{\partial y^i \circ f}{\partial x^j} \right) \Big|_p \right)_{i,j=1}^k & X \\ 0 & I \end{pmatrix}$$

is invertible, hence, by the inverse function theorem,  $z \circ x^{-1}$  is a diffeomorphism in a neighbourhood of  $x(p)$ . It follows that for some neighbourhood  $W$  of  $p$ ,  $(z, W)$  is a smooth coordinate system at  $p$ . For  $1 \leq i \leq k$ ,  $(y_i \circ f \circ z^{-1})(a_1, \dots, a_n) = a_i$ , which completes the proof.  $\square$

**Corollary 2.9.** *If  $f : M \rightarrow N$  is smooth, and has rank  $k$  in a neighbourhood of a point  $p$ , then we may choose coordinate systems  $(x, U)$  and  $(y, V)$  around  $p$  and  $f(p)$  such that if  $a \in x(U \cap f^{-1}(V))$ ,*

$$(y \circ f \circ x^{-1})(a_1, \dots, a_n) = (a_1, \dots, a_k, 0, \dots, 0).$$

*Proof.* Choose  $x$  and  $y$  as in the theorem above. Then

$$D(y \circ f \circ x^{-1})(p) = \begin{pmatrix} I & 0 \\ * & \left( \frac{\partial y_i \circ f}{\partial x_j} \right) \Big|_p \end{pmatrix}$$

Since  $f$  is rank  $k$ , the matrix in the bottom right corner must vanish in a neighbourhood of  $p$ . Therefore, for  $a$  in a small neighbourhood of  $x(p)$ , the value  $(y \circ f \circ x^{-1})(a)$  depends only on  $a_1, \dots, a_k$ . Write

$$(y \circ f \circ x^{-1})(a) = (a_1, \dots, a_k, \psi_{k+1}(a_1, \dots, a_k), \dots, \psi_n(a_1, \dots, a_k)).$$

Then we can define a chart  $(z, V)$  around  $f(p)$  by setting  $z^i = y^i$ , for  $i \leq k$ , and for  $i > k$ ,

$$z^i(q) = y^i(q) - \psi_{k+1}(y^1(q), \dots, y^k(q)).$$

Then for  $q$  in this neighbourhood, we find

$$D(z \circ y^{-1})(q) = \begin{pmatrix} I & 0 \\ * & I \end{pmatrix},$$

So  $z$  is a coordinate system, and

$$z \circ f \circ x^{-1}(a_1, \dots, a_n) = (a_1, \dots, a_k, 0, \dots, 0). \quad \square$$

In particular, this theorem enables us to understand full rank maps.

**Corollary 2.10.** *If  $f : M^n \rightarrow N^m$  is rank  $m$  at  $p$ , then for any coordinate system  $y$  around  $f(p)$ , there exists a coordinate system  $x$  at  $p$  such that*

$$y \circ f \circ x^{-1}(a_1, \dots, a_n) = (a_1, \dots, a_m)$$

*Proof.* Applying the last theorem, we note in the beginning of the proof of the theorem that we need not rearrange the coordinates of  $y$  in the case that the matrix is rank  $m$ , just the coordinates of  $x$ .  $\square$

**Corollary 2.11.** *If  $f : M^n \rightarrow N^m$  is rank  $n$  at  $p$ , then for any coordinate system  $x$  at  $p$ , there exists a coordinate system  $y$  at  $f(p)$  such that*

$$y \circ f \circ x^{-1}(a_1, \dots, a_n) = (a_1, \dots, a_n, 0, \dots, 0)$$

*Proof.* If  $f$  is rank  $n$  at a point, then it is rank  $n$  on a neighbourhood, since the set of full rank matrices is open. Choose coordinate systems  $u$  and  $v$  such that  $u \circ f \circ v^{-1}(a^1, \dots, a^n) = (a^1, \dots, a^n, 0, \dots, 0)$ . Define a map  $\lambda$  on  $\mathbb{R}^m$  by  $\lambda(a^1, \dots, a^m) = (x \circ v^{-1}(a^1, \dots, a^n), a^{n+1}, \dots, a^m)$ . Then  $\lambda$  is a diffeomorphism, hence  $\lambda \circ y$  is a coordinate system, and

$$\begin{aligned} (\lambda \circ y) \circ f \circ x^{-1}(a^1, \dots, a^n) &= \lambda \circ (y \circ f \circ v^{-1}) \circ (v \circ x^{-1})(a^1, \dots, a^n) \\ &= \lambda(v \circ x^{-1}(a^1 \dots a^n), 0 \dots 0) \\ &= (a^1 \dots a^n, 0 \dots 0) \end{aligned}$$

and we have found the chart required.  $\square$

## 2.6 Immersions, Submersions, and Covers

The charts above are all locally constructed, but we can use global topological properties to infer properties of an entire map from the rank of it's covariant derivative at points. Call a smooth map  $f : M \rightarrow N$  an *immersion* if  $f_*|_p$  is injective for all points  $p$ , and a *submersion* if  $f_*|_p$  is always surjective. In terms of rank,  $f$  is an immersion if the rank of  $f$  at  $p \in M$  is the dimension of  $T_p M$ , and a submersion if the rank of  $f$  is the dimension of  $T_{f(p)} N$ .

**Theorem 2.12.** *Let  $f : M \rightarrow N$  be a map between smooth manifolds. Then*

- (a) *If  $f$  is injective and has locally constant rank, then  $f$  is an immersion.*
- (b) *If  $f$  is surjective, has constant rank,  $M$  is second-countable, and  $N$  has the same dimension throughout, then  $f$  is a submersion.*

*Proof.* If  $p$  is a point, if the dimension of  $N$  is  $m$  at  $f(p)$ , and if  $f$  has rank  $k < m$  in a neighbourhood of  $p$ , then there are coordinate systems  $x$  and  $y$  such that

$$(y \circ f \circ x^{-1})(a^1, \dots, a^n) = (a^1, \dots, a^k, 0, \dots, 0)$$

which clearly isn't injective if  $k < n$ . This proves (a). To prove (b), consider a cover of  $f$  by charts  $(x_\alpha, U_\alpha)$ , where  $f(U_\alpha) \subset V_\alpha$  for some chart  $(y, V_\alpha)$  on  $N$ . Since  $M$  is second-countable, it is also Lindelöf, and we may therefore assume the cover is countable. Let  $N$  have dimension  $m$ , and let  $f$  have rank  $k < n$ . The coordinate systems can be chosen so that

$$(y \circ f \circ x^{-1})(a^1, \dots, a^n) = (a^1, \dots, a^k, 0, \dots, 0)$$

This shows us that  $(y \circ f \circ x_\alpha^{-1})(x(U_\alpha))$  is nowhere dense in  $y(V_\alpha)$ , and thus,  $f(U_\alpha)$  is nowhere dense in  $N$ . It then follows by the Baire category theorem for locally compact topological spaces that  $\bigcup f(U_\alpha) = f(M)$  is nowhere dense on  $N$ , which implies  $f$  is not surjective, a contradiction.  $\square$

Variants of this proof can be adapted to various other cases, where  $f$  may not have constant rank, and  $N$  has non-constant dimension. We avoid discussing them in detail, because they are complicated to state, but obvious to adapt to any particular situation by working out the details there.

**Theorem 2.13.** *If  $f : M^n \rightarrow N^m$  is a submersion, then  $f$  is an open map.*

*Proof.* If  $p \in M$ , pick a neighbourhood  $x$  around  $p$  and  $y$  around  $f(p)$  such that

$$x \circ f \circ y^{-1}(a^1, \dots, a^n) = (a^1, \dots, a^m)$$

This map is open, showing  $f$  is locally open, and thus open on its entire domain, since openness is a local property.  $\square$

**Corollary 2.14.** *A submersion from a compact manifold to a connected manifold with the same dimension is surjective.*

*Proof.* If  $M$  is compact, and  $f : M \rightarrow N$  an immersion, then  $f(M)$  is compact, hence closed, and since  $f$  is open,  $f(M)$  is also an open subset of  $N$ . But then  $f(M) = N$ , since it is an open, closed, non-empty set.  $\square$

The idea of a submersive map  $f : M \rightarrow N$  is very closely related to its family of smooth sections  $s : N \rightarrow M$ , maps such that  $f \circ s = \text{id}_N$ . More precisely, a submersive map is closely related to its local smooth sections over neighbourhoods of  $N$ .

**Theorem 2.15.** *A smooth map  $f : M \rightarrow N$  is a submersion if and only if every  $p \in M$  is in the image of some local section.*

*Proof.* If  $f$  is a submersion, then there are local coordinates  $x$  and  $y$  around every point where  $(y \circ f \circ x^{-1})(a^1, \dots, a^n) = (a^1, \dots, a^m)$ . Given any point  $p_0 \in M$ , consider a point  $q_0 \in N$  with  $y^1(q_0) = x^1(p_0), \dots, y^m(q_0) = x^m(p_0)$ , and define the map  $s(q) = x^{-1}(y(q), x^{m+1}(p), \dots, x^n(p))$ , then  $s$  is the required section. Conversely, if  $s : U \rightarrow M$  is a local section, then  $\text{id} = f_* \circ s_*$ , and therefore the rank of  $f_*$  at  $p$  must be at least the rank of  $\text{id}$ , so it has full rank.  $\square$

## 2.7 Submanifolds

All this discussion of ranks and immersions is most useful to studying submanifolds. consider two smooth manifolds  $M$  and  $N$ , and suppose  $i : N \rightarrow M$  is an injective immersion. With this structure, we say  $N$  is an *immersed submanifold* of  $M$ . If  $i$  is also a *topological*

*embedding*, then we say  $N$  is a *submanifold* of  $M$ . For a smooth submanifold, the smooth structure on  $N$  making  $i$  an immersion is then unique. This is because if  $(x, U)$  is any chart on  $N$ , then there is a chart  $(y, V)$  on  $M$  such that  $(y \circ i \circ x^{-1})(a) = (a, 0)$ . Conversely, if  $(y, U)$  is a chart on  $N$  such that  $U \cap i(N) = \{p \in U : y^{k+1}(p) = \dots = y^n(p) = 0\}$ , and we set  $y = (\tilde{y}, 0)$ , then  $\tilde{y} \circ i$  is a diffeomorphism, and therefore a chart on  $N$ , which uniquely specifies the atlas.

*Remark.* If  $N \subset M$  is a  $k$  dimensional manifold under the relative topology, then  $N$  has a smooth structure making the inclusion map  $i : N \rightarrow M$  an immersion if and only if for each  $p \in N$ , there exists a neighbourhood  $U$  of  $p$  in  $M$ , and a chart  $(x, U)$  on  $M$ , such that

$$U \cap N = \{p \in U : x^{k+1}(p) = \dots = x^n(p) = 0\}.$$

The first  $k$  coordinates can then be taken as a chart as in the last paragraph. Thus  $N$  has a natural  $C^\infty$  structure. If we can only choose a  $C^k$  chart  $(x, U)$  above instead of a  $C^\infty$  chart, then  $N$  only has a natural  $C^k$  structure. Thus, for instance, it wouldn't be natural to consider

$$M = \{(x, y) \in \mathbb{R}^2 : y = x^2 \sin(1/x)\}$$

as a  $C^2$  manifold, but only a  $C^1$  manifold, even though one can extend the natural  $C^1$  structure on  $M$  to a  $C^\infty$  structure. Most of the manifolds we have specified are  $C^\infty$  submanifolds of  $\mathbb{R}^n$  for some  $n$ , and one can verify that all the differentiable structure we've given to these manifolds is the only canonically induced by the submanifold structure.

**Example.** *The biggest utility is when these manifolds lie in Euclidean space. Classically, a  $k$ -dimensional  $C^m$  manifold was a subset  $M$  of some Euclidean space  $\mathbb{R}^n$ , such that at ever  $x \in M$ , there is an open subset  $U$  of Euclidean space containing  $x$ , an open subset  $V$  of  $\mathbb{R}^n$ , and a  $C^m$  diffeomorphism  $f : U \rightarrow V$  such that  $f(U \cap M) = V \cap \mathbb{R}^k \times \{0\}$ . All these diffeomorphisms, restricted to  $M$ , are  $C^m$  related to one another, because they are the restricted of diffeomorphisms on the whole of  $\mathbb{R}^n$ , and give a  $C^m$  differentiable structure onto  $M$ . Conversely, any submanifold of Euclidean space has this property because of the rank theorems, because if  $i : M \rightarrow \mathbb{R}^n$  is an immersion, then we can extend any coordinate system on  $M$  to a coordinate system on an open subset of  $\mathbb{R}^n$ , such that  $M$  is flat in this coordinate system. The natural choice of a differentiable structure on any manifold lying in Euclidean space is this differentiable structure, and, as you can check, this is the differentiable structure we have chosen for all the manifolds we have used in past examples, which naturally can be embedded in Euclidean space.*

**Example.** *Our knowledge of immersion can now be used to show every submanifold of  $\mathbb{R}^K$  can be identified locally as a 'graph' of a function. Let  $M$  be a  $k$  dimensional submanifold of  $\mathbb{R}^K$ , and fix  $p \in M$ . Then  $T_p M$  can be identified with a  $k$ -dimensional subspace of  $T_p \mathbb{R}^K$ . If  $T_p M$  is spanned by  $(x_1)_p, \dots, (x_k)_p$  for  $x_1, \dots, x_k \in \mathbb{R}^K$ , then we can find  $k$  distinct indices  $i_1, \dots, i_k$  such that for each  $j \in \{1, \dots, k\}$ ,  $(x_j)_{i_j} \neq 0$ . We consider the linear map  $y : \mathbb{R}^k \rightarrow \mathbb{R}^K$  given by  $y(x) = (x_{i_1}, \dots, x_{i_k})$ . Since  $y$  is linear,  $y_*(x_p) = y(x)_{y(p)}$ , and our choice of basis implies that  $y_*|_p$  is injective when restriction to  $T_p M$ . Thus  $y|_M$  is an immersion in*



a neighbourhood  $U$  of  $p$  in  $M$ . If  $U$  is further restricted, the rank theorem also implies that  $y$  restricts to a coordinate chart on  $U$ . If we let  $V = y(U)$ . Then there exists a function  $g : V \rightarrow \mathbb{R}^K$  such that for each  $p \in U$ ,  $g(y(p)) = p$ . For simplicity, if we permute coordinates such that  $i_1 = 1, \dots, i_k = k$ , then this means that

$$U = \{(y^1, \dots, y^k, g^{k+1}(y), \dots, g^K(y) : y \in V)\}.$$

Thus  $M$  is locally the graph of the function  $g$ .

An important method of finding submanifolds of  $\mathbb{R}^n$  is by specifying the submanifolds as a level set of Euclidean space. For instance, the sphere  $S^n$  can be defined as the level set  $f^{-1}(1)$  of the map  $f(x) = \|x\|$ . Provided that  $\nabla f$  does not vanish on the domain level set, the level set is always a manifold, and this fact can be generalized to maps between arbitrary manifolds.

**Theorem 2.16.** *If  $f : M^n \rightarrow N^m$  has constant rank  $k$  in a neighbourhood of the points mapping to  $q \in N$ , then  $f^{-1}(q)$  is a closed  $n - k$  submanifold of  $M$ .*

*Proof.* If  $f(p) = q$ , and  $f$  has rank  $k$  at  $p$ , then we may write

$$y \circ f \circ x^{-1}(a_1, \dots, a_n) = (a_1, \dots, a_k, 0, \dots, 0)$$

for some charts  $(x, U)$  and  $(y, V)$ , with  $y(q) = 0$ . This means that

$$U \cap f^{-1}(q) = \{r \in U : x^1(r) = \dots = x^k(r) = 0\},$$

so  $f^{-1}(q)$  is specified by  $n - k$  dimensional slices in  $M$ , and can therefore be given a  $n - k$  submanifold structure.  $\square$

**Example.** *If  $M$  is a 1-submanifold of  $\mathbb{R}^2$ , specified as the level set of a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $M = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$ , then the surface of revolution about the  $z$ -axis related to  $M$  can be identified as the space*

$$N = \{(x, y, z) \in \mathbb{R}^3 : f(\sqrt{x^2 + y^2}, z) = 0\}$$

*It is a surface, because the differential of the defining level set mapping has rank 1 at any point in a neighbourhood of  $N$  (because we can use the chain rule, and the fact that the map  $(x, y, z) \mapsto (\sqrt{x^2 + y^2}, z)$  has full rank at every point).*

**Example.** *The special linear group  $SL(n) = SL_n(\mathbb{R})$  is the set of invertible matrices with determinant one. Since the determinant is a multilinear function, we can find the determinant via the formula*

$$D(\det)(v_1, \dots, v_n)(w_1, \dots, w_n) = \sum_{k=1}^n \det(v_1, \dots, w_k, \dots, v_n)$$

where  $M(n, n)$  is identified with  $(\mathbb{R}^n)^n$ . Then for any  $(v_1, \dots, v_n) \in GL_n(\mathbb{R})$ ,

$$D(\det)(v_1, \dots, v_n)(v_1, \dots, v_n) = n \det(v_1, \dots, v_n) \neq 0.$$

Thus  $\det$  has rank 1 at every point (full rank), and  $SL(n)$  is a (closed) submanifold of  $GL(n)$  dimension  $n^2 - 1$ .

**Example.** The orthogonal group  $O(n)$  is the set of matrices  $M$  such that  $MM^t = I$ . Then  $O(n)$  is closed, for the map  $\psi : M \mapsto MM^t$  is continuous, and  $O(n) = \psi^{-1}(I)$ .  $\psi$  maps  $M(n, n)$  into the set of symmetric matrices, which is a subspace of  $M(n, n)$  of dimension  $n(n+1)/2$ . If we take the  $i$ 'th diagonal entry of  $MM^t = I$ , we obtain that

$$v_{i1}^2 + v_{i2}^2 + \cdots + v_{in}^2 = 1$$

This implies that  $M$  lies on  $(S^{n-1})^n \subset \mathbb{R}^{n^2}$ , so  $O(n)$  is closed and bounded, and therefore compact. Consider the diffeomorphism  $R_A : B \mapsto BA$  of  $GL(n)$ , for a fixed  $A \in GL(n)$ . We also have  $\psi \circ R_A = \psi$ , for  $A \in O(n)$ . We conclude that

$$D(\psi)(A) \circ D(R_A)(I) = D(\psi \circ R_A)(I) = D(\psi)(I)$$

Since  $R_A$  is a diffeomorphism,  $D(\psi)(A)$  has the same rank as  $D(\psi)(I)$ . Let us find this rank. Explicitly, we may write the projections of  $\psi$  as

$$\psi^{ij}(M) = \sum_{k=1}^n M_{ik} M_{jk}$$

Then

$$\left. \frac{\partial \psi^{ij}}{\partial M^{lk}} \right|_M = \begin{cases} 2M_{ik} & i = j = l \\ M_{ik} & j = l \\ M_{jk} & i = l \\ 0 & \text{elsewise} \end{cases}$$

In particular, at the identity,

$$\left. \frac{\partial \psi^{ij}}{\partial M^{lk}} \right|_I = \begin{cases} 2 & i = j = k = l \\ 1 & j = k = l \\ 1 & i = k = l \\ 0 & \text{elsewise} \end{cases}$$

It follows that  $\psi_*(TGL(n)_I)$  is the space of all symmetric matrices in  $TGL(n)_I$ , which has dimension  $n(n+1)/2$ . Thus  $\psi$  has constant rank  $n(n+1)/2$ , and we find the space of orthogonal matrices has dimension

$$n^2 - n(n+1)/2 = n(n-1)/2$$

as a differentiable manifold.

**Example.** Every orthogonal matrix has determinant  $\pm 1$ . The special orthogonal group  $SO(n)$  is the set of orthogonal matrices of determinant one, and is an open submanifold of  $O(n)$ . It's actually a connected component. To see this, consider the obvious (Lie-group) action of  $SO(n)$  on  $S^{n-1}$ . For  $n \geq 2$ , it is easy to see that the action is transitive: if  $v_1 \in S^{n-1}$

is given, we can extend  $v_1$  to an orthonormal basis  $(v_1, \dots, v_n)$ ; possibly permuting this basis, the basis can be assumed oriented, and then the orthogonal matrix mapping  $e_i$  to  $v_i$  is in  $SO(n)$ . For a fixed  $p \in S^{n-1}$ , the action  $f(M) = Mp$  is surjective. If  $N$  is fixed,  $g(M) = NM$ , and  $h(p) = Np$ , then we have a commutative diagram. The vertical maps are diffeomorphisms, hence, taking differentials of all maps in the diagram, we conclude that  $h_* \circ f_* = f_* \circ g_*$ , and since  $h_*$  and  $g_*$  are rank preserving, we conclude that the rank of  $f_*$  at any  $M$  is the same as it's rank at  $NM$  (this argument works for any 'differentiable group', and any 'differentiable action', a notion we will define precisely later when we introduce Lie groups). Because  $SO(n)$  is second countable, we conclude that  $f$  is a submersion, hence it is open. In the one dimensional case, the map isn't surjective, but in this case  $SO(1)$  and  $S^0$  are discrete spaces, so the map  $M \mapsto Mp$  is trivially open. In any dimension  $\geq 2$ , the stabilizer of the unit vector  $e_n$  is the subgroup of matrices of the form

$$\begin{pmatrix} X & 0 \\ 0 & 1 \end{pmatrix}$$

where  $X \in SO(n-1)$ , so the stabilizer is actually diffeomorphic to  $SO(n-1)$ . Using the orbit stabilizer theorem, we find that for  $n \geq 2$ ,  $S^{n-1}$  is homeomorphic to  $SO(n)/SO(n-1)$ . Using induction, we may assume that  $SO(n-1)$  is connected. But for  $n \geq 2$ ,  $S^{n-1}$  is connected. The theorem then follows from the general fact that if  $H$  is a closed subgroup of  $G$ , and  $H$  and  $G/H$  are connected, then  $G$  is connected.

Lower dimensional matrix groups often have special coordinate systems:

- On  $SO(2) = \mathbb{T}$ , the standard coordinate system is obtained by taking angles.
- In  $SO(3)$  it is easy to prove using the spectral theorem that any rotation  $R$  is given by an oriented rotation about the plane perpendicular to a vector  $x$  with  $Rx = x$ . Thus a rotation is given by picking a line  $l \in \mathbf{RP}^2$  and an angle  $\theta \in SO(1)$  to rotate perpendicularly about the line. Thus we have a natural covering map  $\mathbf{RP}^2 \times \mathbb{T} \rightarrow SO(3)$ , which can be used to perform computations.
- More concretely, we have a map  $\mathbb{T}^3 \rightarrow SO(3)$ , which gives the theory of *Euler angles*. In three dimensions, every rotation can be given first by a rotation about the  $z$  axis, known as the *yaw*, a rotation about the  $y$  axis, the *pitch*, and a rotation about the  $x$  axis, a *roll*. Thus we can describe almost every rotation by three angles  $\phi$ ,  $\theta$ , and  $\psi$ , which describes the rotation

$$\begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

However, this map is *not* a covering map, as can be verified by looking at the derivatives, which become singular when  $\theta = 0$ . Indeed, the singularities that prevent this map from being a covering map result in computational glitches, known as *gimbal locks*, when implemented to represent three dimensional rotations on computers in coordinates.

These coordinate systems are often more useful in performing calculations than the more general results given above.

When talking about manifolds with boundary, symmetry in coordinates is lost, which prevents quite as nice a theory of constant rank maps. There are two cases where things are simple. If  $f : M^n \rightarrow N^m$  is a *immersion* at a point  $p \in M^\circ$ , then there exists a chart  $(x, U)$  around  $p$  and a chart  $(y, V)$  around  $f(p)$  such that

$$(y \circ f \circ x^{-1})(t_1, \dots, t_n) = (t_1, \dots, t_n, 0, \dots, 0).$$

Similarly, if  $f : M^n \rightarrow N^m$  is a *submersion* at a point  $p \in M$ , and  $f(p) \in N^\circ$ , then there exists a chart  $(x, U)$  around  $p$ , and a chart  $(y, V)$  around  $f(p)$  such that

$$(y \circ f \circ x^{-1})(t_1, \dots, t_n) = (t_1, \dots, t_m).$$

This follows because we have complete freedom in our choice of the coordinate systems around the manifold with boundary.

**Theorem 2.17.** *If  $a < b$  are real numbers,  $M$  is a smooth manifold with  $\partial M = \emptyset$ , and  $f \in C^\infty(M)$  is a smooth map with  $a$  and  $b$  as regular values, then  $f^{-1}([a, b])$  is a submanifold of  $M$  with boundary, with the same dimension as  $M$ .*

*Proof.* The set  $f^{-1}((a, b))$  is a open subset of  $M$ , so it is a smooth manifold without boundary at any point  $p$  with  $f(p) \in (a, b)$ . If  $f(p) \in \{a, b\}$ , then applying the rank theorem shows that there are coordinates  $(x, U)$  around  $p$  such that

$$(f \circ x^{-1})(t_1, \dots, t_n) = t_1.$$

It is then clear that  $f^{-1}((a, b)) \cap U$  is diffeomorphic to a relatively open subset of  $\mathbb{H}^n$ , which gives a boundary chart at  $p$ .  $\square$

There is another important situation which becomes most important in the intersection theory of smooth manifolds.

**Theorem 2.18.** *Let  $f : M^n \rightarrow N^m$  be a smooth map with  $\partial N = \emptyset$ , and let  $q \in N$  be a regular value for both the map  $f : M \rightarrow N$  and the restriction map  $\partial f : \partial M \rightarrow N$ . Then  $f^{-1}(q)$  is a manifold with boundary of dimension  $n - m$ , and moreover,  $\partial(f^{-1}(q)) = (\partial f)^{-1}(q)$ .*

*Proof.* It is certainly true that  $f^{-1}(q) \cap M^\circ$  is a manifold without boundary with dimension  $n - m$ . Now consider  $p \in f^{-1}(q) \cap \partial M$ . Our goal is to show there are coordinate system  $(x, U)$  and  $(y, V)$  with  $x(p) = 0$  and  $y(q) = 0$ , such that

$$(y \circ f \circ x^{-1})(t_1, \dots, t_n) = (t_{n-m+1}, \dots, t_n).$$

It then follows that  $f^{-1}(q) \cap M$  is diffeomorphic to  $\mathbb{H}^{n-m}$  by taking the first  $n - m$  coordinates as a boundary chart. To obtain such a coordihart, it clearly suffices to consider a map  $f :$

$U \rightarrow \mathbb{R}^m$ , where  $U$  is a relatively open neighbourhood of 0 in  $\mathbb{H}^n$ ,  $f(0) = 0$ , such that there exists indices  $k_1, \dots, k_m > 1$  such that the matrix

$$\left( \frac{\partial f^i}{\partial x^{k_j}} \right)$$

is invertible. By permuting coordinates in  $\{2, \dots, n\}$ , which is certainly valid in  $\mathbb{H}^n$ , we may assume  $k_1 = n - m + 1, \dots, k_m = n$ . Then the inverse function theorem guarantees that  $z = (x^1, \dots, x^{n-m}, f^1, \dots, f^m)$  is a coordinate system in a neighbourhood of the origin. And for such a coordinate system, we have

$$(f \circ x^{-1})(t_1, \dots, t_n) = (t_{n-m+1}, \dots, t_n),$$

so we have proved what was required.  $\square$

Immersed submanifolds behave almost as well as submanifolds, apart from the odd inconsistency of dropping differentiable maps to subdomains. Let  $i : N_0 \rightarrow N$  be an injective immersion between two manifolds, and let  $g : M \rightarrow N$  be a differentiable function with  $g(M) \subset N_0$ . Then we can consider the induced function  $\tilde{g} : M \rightarrow N_0$  such that  $g = i \circ \tilde{g}$ . It is natural to ask whether the map  $g$  is differentiable.

**Example.** Immerse  $(-\pi, \pi)$  in  $\mathbb{R}^2$  via the lemniscate map  $f(t) = (\sin 2t, \sin t)$ . The map  $g : (-\pi, \pi) \rightarrow (-\pi, \pi)$  defined by

$$g(x) = \begin{cases} \pi + x & : x < 0 \\ 0 & : x = 0 \\ x - \pi & : x > 0 \end{cases}$$

is not even continuous, yet  $f \circ g$  is differentiable.

Continuity is all that can go wrong in this situation.

**Theorem 2.19.** Let  $i : N_0 \rightarrow N$  be an injective immersion of  $M$  in  $N$ , and suppose  $f : M \rightarrow N$  is differentiable, with  $f(M) \subset N_0$ . If the induced map  $\tilde{f} : M \rightarrow N_0$  is continuous, then  $\tilde{f}$  is differentiable.

*Proof.* Fix  $p \in M$ , and let  $f(p) = q$ . Choose a coordinate chart  $(\tilde{y}, U_0)$  on  $N_0$  with  $q \in U_0$ , and then find a coordinate chart  $(y, U)$  on  $N$ , such that  $U_0 = U \cap i(N_0)$ , and  $(y \circ i \circ \tilde{y}^{-1})(a) = (a, 0)$ , so  $\pi \circ y \circ i = \tilde{y}$ , where  $\pi$  is projection onto the first set of coordinates. Since  $\tilde{f}$  is continuous,  $\tilde{f}^{-1}(U_0)$  is an open neighbourhood of  $p$ , and so we can choose a coordinate system  $(x, V)$  around  $p$ , with  $V \subset \tilde{f}^{-1}(U_0)$ . We know  $y \circ f \circ x^{-1}$  is smooth, but this means that

$$\begin{aligned} \tilde{y} \circ \tilde{f} \circ x^{-1} &= (\pi \circ y \circ i) \circ \tilde{f} \circ x^{-1} \\ &= \pi \circ y \circ f \circ x^{-1}, \end{aligned}$$

which is smooth, so  $\tilde{f}$  is smooth.  $\square$

In particular, this is never a problem for submanifolds.

**Theorem 2.20.** *The following are sufficient to conclude that an injective immersion  $f : M \rightarrow N$  is an embedding.*

- (a)  $f$  is open or closed.
- (b)  $f$  is proper (the inverse image of compact sets are compact).
- (c)  $M$  is compact.
- (d) The dimension of  $M$  is equal to the dimension of  $N$ .

*Proof.* An injective open or closed continuous map is a topological embedding, so (a) is fairly trivial. If  $M$  is compact, then  $f$  is closed, and if the dimension of  $M$  is equal to the dimension of  $N$ ,  $f$  is open (by looking at the coordinate charts), so (c) and (d) follow from (a). Similarly, if  $f$  is proper and its codomain is locally compact, so  $f$  is closed, and hence an embedding. To see why, we fix a closed set  $C \subset M$ . Suppose  $y \notin f(C)$ . Pick a precompact neighbourhood  $V$  of  $y$ . Then  $C \cap f^{-1}(\overline{V})$  is compact, so  $f(C) \cap \overline{V}$  is compact. In particular, since  $y \notin f(C) \cap \overline{V}$ , we can find an open neighbourhood  $W$  of  $y$  not containing any points in  $f(C) \cap \overline{V}$ . But this means  $V \cap W$  is an open neighbourhood of  $y$  not containing any points in  $f(C)$ , so  $f(C)$  is closed.  $\square$

One can think of a proper map  $f : M \rightarrow N$  as being one which maps ‘points near infinity’ on  $M$  to ‘points near infinity’ on  $N$ , this being certainly true for the model case of proper maps  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

## 2.8 Manifolds and Measure

The Lebesgue measure on  $\mathbb{R}^n$  gives us a nice way to calculate the volume of various sorts of objects. A subset  $E$  has measure zero if, for any  $\varepsilon$ , there is a cover of  $E$  by open sets  $U_1, U_2, \dots$  with  $\sum |U_k| < \varepsilon$ . Suppose that  $\mu$  is a measure on a differentiable manifold  $M$  behaving locally like the Lebesgue measure, in the sense that for any chart  $(x, U)$ , the pushforward measure  $x_*(\mu)$  is absolutely continuous with respect to the Lebesgue measure. If  $\mu(E) = 0$ , then  $x_*(\mu)(x(E)) = 0$  for every chart  $x$ , and conversely, assuming  $M$  is second countable, if this is true for every chart, then  $\mu(E) = 0$ . Thus it makes sense, without any measure on the manifold, to say a set  $E$  has *measure zero* if it can be covered by countably many charts  $(x_i, U_i)$ , where  $x_i(E \cap U_i)$  has measure zero for each  $i$ .

**Lemma 2.21.** *If  $f \in C^1(M, N)$ , where  $M$  and  $N$  are smooth manifolds with the same dimension, and if  $E \subset M$  has measure zero, then  $f(E)$  has measure zero.*

*Proof.* Begin with the case  $M = N = \mathbb{R}^n$ . Let  $E$  be a compact rectangle in  $\mathbb{R}^n$ , and suppose that  $\|\partial_i f^j\|_{L^\infty(E)} \leq K$  for all  $i$  and  $j$ . Then for  $x, y \in E$ , by tracing the path from  $x$  to  $y$  along

the coordinate lines,

$$|f(x) - f(y)| \leq \sum_{i=1}^n |f^i(x) - f^i(y)| \leq K \sum_{i=1}^n \sum_{j=i}^n |x^j - y^j| \leq Kn^2|x - y|$$

If  $E$  has measure zero, then, by the  $\sigma$  compactness of  $\mathbb{R}^n$ , we may assume  $E$  is compact, so that  $E$  is contained in some rectangle, and suppose on this rectangle that the partial derivative bound holds. If  $E$  is covered by rectangles  $U_1, U_2, \dots$  with  $\sum |U_i| \leq \varepsilon$ , then each  $U_i$  is mapped into a ball of radius  $Kn^2 \text{diam}(U_i)$ , whose volume agrees with  $|U_i|$  up to a constant  $C$  depending on  $K$  and  $n$ , and so  $f(E)$  is covered by balls with total measure  $\leq C\varepsilon$ , and we may let  $\varepsilon \rightarrow 0$ . Now in general on any manifold, let  $E$  be covered by charts  $(x_1, U_1), (x_2, U_2), \dots$  with  $x_k(E \cap U_k)$  a set of measure zero, in such a way that there are charts  $(y_1, V_1), (y_2, V_2), \dots$  with  $U_k \subset f^{-1}(V_k)$  (we may take an initial covering of  $E$  by charts, and then duplicate the charts, suitably small to be contained in the  $V_k$ ). Then each  $(y_k \circ f \circ x^{-1})(x(E \cap U_k)) = y_k(f(E \cap U_k))$  has measure zero, and so  $f(E) = \bigcup f(E \cap U_k)$  has measure zero.  $\square$

If the rank of a map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  has rank  $< m$ , then intuitively, it maps small neighbourhoods of points to sets close to hyperplanes, which have measure zero. We will show this idea shows the image of the set of low rank points always has measure zero on a manifold. Call a point  $p$  on  $M$  *critical* for a map  $f : M \rightarrow N$  if  $f$  is not a submersion at  $p$ , and *regular* if the map is a submersion at  $p$ . A point  $q \in N$  is then called a *regular value* if all points in  $f^{-1}(q)$  are regular. Otherwise,  $q$  is known as a *critical value*.

**Theorem 2.22 (Sard).** *If  $f : M^n \rightarrow N^n$  is  $C^1$ , and  $M$  is second countable, then the set of critical values is a set of measure zero in  $N$ .*

*Proof.* We begin with the case  $N = \mathbb{R}^n$ , and  $M$  an open subset of  $\mathbb{R}^n$ . Let  $C$  be a closed cube in  $M$  with side lengths  $l$ , and let  $B$  denote the critical values in  $C$ . Then  $f$  is uniformly continuously differentiable on  $C$ , and so for any  $\varepsilon$ , for large enough  $N$ , we can divide  $C$  into  $N^n$  cubes  $S_k$  of side length  $l/n$ , such that for  $x, y \in S_k$ ,

$$|f(y) - f(x) - Df(x)(y - x)| < \varepsilon|y - x| \leq \varepsilon \frac{l\sqrt{n}}{N}$$

and we also know that  $f$  is Lipschitz on  $C$ , so there is  $M$  such that

$$|f(y) - f(x)| < M|y - x| < \frac{lM\sqrt{n}}{N}$$

If  $B$  intersects  $S_k$ , then there exists  $x \in S_k$  such that  $Df(x)(y - x)$  lies in an  $n - 1$  dimensional subspace of  $\mathbb{R}^n$ , and so the two inequalities above imply that  $f$  maps  $S_k$  into a ‘ $n$  dimensional cylinder’, formed by taking a ball of radius  $lM\sqrt{n}/N$  in  $n - 1$  dimensional space as the base of the cylinder, and a height of length  $2\varepsilon l\sqrt{n}/N$ . This shape has volume equal, up to a constant depending on  $n$ , by

$$\left(\frac{2\varepsilon l\sqrt{n}}{N}\right) \left(\frac{lM\sqrt{n}}{N}\right)^{n-1} = 2\varepsilon \frac{l^n n^{n/2} M^{n-1}}{N^n}$$

But we have  $N^n$  cubes, which implies the set  $f(B \cap C)$  can have volume at most

$$2\varepsilon l^n n^{n/2} M^{n-1}$$

and we may then let  $\varepsilon \rightarrow 0$  to conclude  $f(B \cap C)$  is a set of measure zero. But since  $\mathbb{R}^n$  is the union of countably many cubes  $C$ , we conclude that  $B$  itself has measure zero. The general case for second countable manifolds follows from taking a countable set of coordinate charts covering  $M$ , and then applying the Euclidean Sard's theorem.  $\square$

Sard's theorem has a more general form which is very useful in differential topology.

**Theorem 2.23 (Sard).** *If  $f : M^n \rightarrow N^m$  is  $C^k$ , and  $k \geq \max(1, n - m + 1)$ , then the set of critical values is a set of measure zero in  $N$ .*

Our case occurs when  $n = m$ . It is easy to prove the case where  $m > n$ . The tricky case is where  $m < n$ .

**Theorem 2.24.** *If  $f : M^n \rightarrow N^m$  is  $C^1$ ,  $M^n$  is connected, and  $n < m$ , then  $f(M)$  has measure zero in  $N^m$ .*

*Proof.* Consider the map  $g : M \times \mathbb{R}^{m-n} \rightarrow N$ , defined by  $g(p, x) = f(p)$ . Then all values of  $g$  are critical, and so  $g(M \times \mathbb{R}^{m-n}) = f(M)$  has measure zero by the previous case of Sard's theorem we proved.  $\square$

Let us now prove the difficult case where  $m < n$ . We shall only prove the theorem for smooth maps. The proof for  $C^k$  maps with  $k < \infty$  is *essentially* the same proof, but involving some more advanced technology, i.e. the Whitney extension theorem, which enables us to boost the smoothness of functions in certain situations.

**Theorem 2.25.** *If  $n > m$ , and  $f : M^n \rightarrow N^m$  is a  $C^{n-m+1}$  map, then the set of critical values has measure zero in  $N$ .*

In this situation, we cannot simply rely on the fact that neighbourhoods of critical points are mapped to neighbourhoods of lower dimensional sets. Thus we must rely on the higher order smoothness in some way. It clearly suffices to prove the theorem when  $M = U$  is an open subset of  $\mathbb{R}^n$ , and  $N = \mathbb{R}^m$ . We let  $C \subset U$  be the set of all critical points. We must show  $f(C)$  has measure zero. For each  $s \in \{1, \dots, n\}$ , we let  $C_s$  be the set of all critical points  $x$  such that  $f_\alpha(x) = 0$  for all  $|\alpha| \leq s$ .

**Lemma 2.26.** *If  $s > n/m - 1$ , then  $f(C_s)$  has measure zero.*

*Proof.* Let  $I \subset U$  be a closed cube with sides parallel to the axis. It suffices to show  $f(C_s \cap I)$  has measure zero. By Taylor's theorem and compactness, we can write  $f(y) = f(x) + R(x, y)$ , where for  $x \in B_s \cap I$  and  $y \in I$ ,  $|R(x, y)| \lesssim |x - y|^{s+1}$ . Choose a large integer  $N$ , and divide  $I$  into  $N^n$  cubes with sidelength  $\sim 1/N$ . Then for any such subcube  $J$  that intersects  $C_s$ ,  $f(J)$  is contained in a ball of radius  $O(1/N^{s+1})$ . Thus  $f(J)$  has measure at most  $O(1/N^{m(s+1)})$ . But this means that  $f(B_s)$  has measure at most  $O(N^{n-m(s+1)})$ . If  $s > n/m - 1$ , this is  $o(1)$  as  $N \rightarrow \infty$ .  $\square$



The inequality  $n/m - 1 < n - m + 1$  is certainly true for  $m = 1$  and  $m = 2$ . For  $m \geq 3$ , the inequality is equivalent to the inequality

$$n < m \left( \frac{m-2}{m-1} \right) = m \left( 1 - \frac{1}{m-1} \right)$$

But this clearly implies  $n < m$ . Thus  $f(C_{n-m+1})$  has measure zero.

We now prove Sard's theorem by induction on  $n$ , the case  $n = 1$  holding by application of previous cases. We write

$$C = (C - C_1) \cup (C_1 - C_2) \cup \cdots \cup (C_{n-m} - C_{n-m+1}) \cup C_{n-m+1}.$$

We know  $f(C_{n-m+1})$  has measure zero, so it suffices to prove  $f(C_s - C_{s+1})$  has measure zero for  $s < r$ , and that  $f(C - C_1)$  has measure zero.

We can write  $C - C_1 = \bigcup_{i,j} A_{ij}$ , where  $A_{ij}$  is the set of critical points  $x$  with  $D_i f_j(x) \neq 0$ . Let us consider the case  $A_{11}$ , for simplicity in notation. Then for each  $x_0 \in A_{11}$ , the implicit function theorem guarantees that there exists a neighbourhood  $V$  of  $x_0$ , a neighbourhood  $W$  of  $\mathbb{R}^n$ , and a diffeomorphism  $u : W \rightarrow V$  such that for  $(t, y) \in U$ ,  $f_1(u(t, y)) = f_1(x_0) + t$ . For each  $t$ , we let  $W_t = \{y \in \mathbb{R}^{n-1} : (t, y) \in W\}$  and define  $h_t : W_t \rightarrow V$  by setting

$$h_t(y) = (f_2(u(t, y)), \dots, f_n(u(t, y))).$$

Then  $y \in W_t$  is a critical point for  $f \circ h_t$  if and only if  $(t, y)$  is a critical point for  $f$ . Applying Sard's theorem in the case  $n - 1$ , we conclude that  $f(A_{11}) \cap \{t\} \times \mathbb{R}^{m-1} \cap V$  has measure zero. But then by Fubini's theorem, we conclude that  $f(A_{11} \cap V)$  has measure zero. But since  $A_{11}$  can be covered by countably many such neighbourhoods  $V$ , this means that  $f(A_{11})$  has measure zero. The same technique shows  $f(A_{ij})$  has measure zero for all indices  $i, j$ , the proof only requiring a permutation of coordinates.

Now we address the analysis of  $C_s - C_{s+1}$ , with  $s < r$ . We write  $C_s - C_{s+1} \subset \bigcup_{\alpha, i} C(\alpha, i)$ , where  $\alpha$  is a multiindex with  $|\alpha| = s$ ,  $i, j \in \{1, \dots, n\}$ , and  $C(\alpha, i, j)$  is the set of points  $x$  with  $D_\alpha f(x) = 0$ , but  $D_i D_\alpha f_j(x) \neq 0$ . For simplicity, we shall show  $f(C(\alpha, 1, 1))$  has measure zero, the general case obtained by permuting coordinates. By the implicit function theorem, for each  $x_0 \in C(\alpha, i)$ , there exists a neighbourhood  $V$  of  $x_0$ , an open set  $W \subset \mathbb{R}^n$ , and a diffeomorphism  $g : W \rightarrow V$  such that for each  $(t, y) \in W$ ,

$$f_\alpha(g(t, y), y) = (t, z_2(t, y), \dots, z_m(t, y)).$$

Define  $W_0 = \{y : (0, y) \in W\}$ . Clearly  $C(\alpha, 1, 1) \cap V$  is contained in the set of points  $(g(0, y), y)$ , with  $y \in W_0$ , such that  $y$  is critical for the map  $h_0 : W_0 \rightarrow \mathbb{R}^m$  given by setting  $h_0(y) = f(g(0, y), y)$ . But by induction we know the image of the critical points for  $h_0$  form a set of measure zero, from which we obtain that  $f(C(\alpha, 1, 1) \cap V)$  has measure zero. Since  $C(\alpha, 1, 1)$  can be covered by countably many neighbourhoods  $V$ , we conclude  $f(C(\alpha, 1, 1))$  has measure zero. The general case for  $C(\alpha, i, j)$  is treated by very similar methods. This completes our proof of Sard's theorem.

*Remark.* The only change that results in the tight statement of Sard's theorem is replacing the application of the implicit function theorem with something which 'boosts smoothness' in the case where  $f$  is  $C^k$ , with  $k < \infty$ , since repeated applications of the implicit function theorem will reduce smoothness in this case.

## 2.9 Partitions of Unity

The use of  $C^\infty$  functions relies on the fact that manifolds possess them in plenty. The following theorem gives us our first plethora. First, we detail some  $C^\infty$  functions on  $\mathbb{R}^n$ .

1. The map  $f : \mathbb{R} \rightarrow \mathbb{R}$ , defined by

$$g(t) = \begin{cases} e^{-t} & : t > 0 \\ 0 & : \text{elsewhere} \end{cases}$$

is  $C^\infty$ . We have  $0 < f(t) < 1$  on  $(0, \infty)$ , and  $f^{(n)}(0) = 0$  for all  $n$ .

2. The  $C^\infty$  map  $g(t) = f(t-1)f(t+1)$  is positive on  $(-1, 1)$ , and zero everywhere else. Similarly, for any  $\varepsilon$ , there is a map  $g_\varepsilon$  which is positive on  $(-\varepsilon, \varepsilon)$  and zero elsewhere.

3. The map

$$l(t) = \begin{cases} (\int_{-\varepsilon}^t g_\varepsilon) / (\int_{-\varepsilon}^\varepsilon g_\varepsilon) & : t \in (0, \infty) \\ 0 & : \text{elsewhere} \end{cases}$$

is  $C^\infty$ , is zero for negative  $t$ , increasing on  $(0, \varepsilon)$ , and one on  $[\varepsilon, \infty)$ .

4. There is a differentiable map  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  defined by  $h(x_1, \dots, x_n) = g(x_1)g(x_2) \dots g(x_n)$  which is positive on  $(-1, 1)^n$ , and zero elsewhere.

With these nice functions in hand, we may form them on arbitrary manifolds.

**Theorem 2.27.** *If  $M$  is a differentiable manifold, and  $C$  is a compact set contained in an open set  $U$ , then there is a differentiable function  $f : M \rightarrow \mathbb{R}$  such that  $f(x) = 1$  for  $x$  in  $C$ , and whose support  $\overline{\{x \in M : f(x) \neq 0\}}$  is contained entirely within  $U$ .*

*Proof.* For each point  $p$  in  $C$ , consider a chart  $(x, V)$  around  $p$ , with  $\overline{V} \subset U$ , and  $x(V)$  containing the open unit square  $(-1, 1)^n$  in  $\mathbb{R}^n$ . We may clearly select a finite subset of these charts  $(x_k, V_k)$  such that the  $x_k^{-1}((-1, 1)^n)$  cover  $C$ . We may define a map  $f_k : V_k \rightarrow \mathbb{R}$  equal to  $h \circ x_k$ , where  $h$  is defined above. It clearly remains  $C^\infty$  if we extend it to be zero outside of  $V_k$ . Then  $\sum f_k$  is positive on  $C$ , and whose closure is contained within  $\bigcup \overline{V_k} \subset U$ . Since  $C$  is compact, and the function is continuous,  $\sum f_k$  is bounded below by  $\varepsilon$  on  $C$ . Taking  $f = l \circ (\sum f_k)$ , where  $l$  is defined above, we obtain the map needed.  $\square$

To enable us to define  $C^\infty$  functions whose support lie beyond this range, we need to consider a technique to extend  $C^\infty$  functions defined locally to manifolds across the entire domain. A *partition of unity* on a manifold  $M$  is a family of  $C^\infty$  functions  $\{\phi_i : i \in I\}$ , and such that the following two properties hold:

1. The supports of the functions forms a locally finite set.
2. For each point  $p \in M$ , the finite sum  $\sum_{i \in I} \phi_i(p)$  is equal to 1.

If  $\{U_i\}$  is an open cover of  $M$ , then a partition of unity is subordinate to this cover if it also satisfies (3):

3. The closure of each function is contained in some element of the cover.

It is finally our chance to use the topological ‘niceness’ established in the previous chapter.

**Lemma 2.28** (The Shrinking Lemma). *If  $M$  is a paracompact manifold, and  $\{U_i\}$  is an open cover, then there exists a refined cover  $\{V_i\}$  such that for each  $i \in I$  there exists  $i'$  such that  $\overline{V_i} \subset U_{i'}$ .*

*Proof.* Without loss of generality, we may assume  $\{U_i\}$  is locally finite, and  $M$  is connected. Then  $M$  is also  $\sigma$ -compact,  $M = \bigcup C_i$ . Since  $C_i$  is compact, and each  $p \in C_i$  locally intersects only finitely many  $U_i$ , then  $C_i$  intersects only finitely many  $U_i$ . Therefore  $\bigcup C_i$  intersects only countably many  $U_i$ , and thus our locally finite cover is countable. Consider an ordering  $\{U_1, U_2, \dots\}$  of  $\{U_i\}$ . Let  $C_1$  be the closed set  $U_1 - (U_2 \cup U_3 \cup \dots)$ . Let  $V_1$  be an open set with  $C_1 \subset V_1 \subset \overline{V_1} \subset U_1$ . Inductively, let  $C_k$  be the closed set  $U_k - (V_1 \cup \dots \cup V_{k-1} \cup U_{k+1} \cup \dots)$ , and define  $V_k$  to be an open set with  $C_k \subset V_k \subset \overline{V_k} \subset U_k$ . Then  $\{V_1, V_2, \dots\}$  is the desired refinement.  $\square$

**Theorem 2.29.** *Any cover on a paracompact manifold induces a subordinate partition of unity.*

*Proof.* Let  $\{U_i\}$  be an open cover of a paracompact manifold  $M$ . Without loss of generality, we may consider  $\{U_i\}$  locally finite. Suppose that each  $U_i$  has compact closure. Choose  $\{V_i\}$  satisfying the shrinking lemma. Apply Theorem (2.13) to  $\overline{V_i}$  to obtain functions  $\psi_i$  that are 1 on  $\overline{V_i}$  and zero outside of  $U_i$ . These functions are locally finite, and thus we may define  $\phi_i$  by

$$\phi_i(p) = \frac{\psi_i(p)}{\sum_j \psi_j(p)}$$

Then  $\phi_i$  is the partition of unity we desire.

This theorem holds for any  $\{U_i\}$  provided Theorem (2.13) holds on any closed set, rather than just a compact one. Let  $C$  be a closed subset of a manifold, contained in an open subset  $U$ , and for each  $p \in C$ , choose an open set  $U_p$  with compact closure contained in  $U$ . For each  $p \in C^c$ , choose an open subset  $V_p$  contained in  $C^c$  with compact closure. Then our previous compact case applies to this cover, and we obtain a subordinate partition of unity  $\{\zeta_i\}$ . Define  $f$  on  $M$  by

$$f(p) = \sum_{\overline{\zeta_i} \subset U_p} \zeta_i(p)$$

Then the support of  $f$  is contained within  $U$ , and  $f = 1$  on  $C$ .  $\square$

Partitions of unity allow us to extend local results on a manifold to global results. The utility of these partitions is half the reason that some mathematicians mandate that manifolds must be paracompact – otherwise many nice results are lost.

**Theorem 2.30.** *In a  $\sigma$ -compact manifold  $M$ , there exists a smooth, real-valued function  $f$  such that  $f^{-1}((-\infty, t])$  is compact for each  $t$ .*

*Proof.* Let  $M$  be a  $\sigma$ -compact manifold with  $M = \bigcup B_i$ , Where  $\overline{B_i}$  is compact,  $B_i$  is diffeomorphic to a ball, and the  $B_i$  are a locally finite cover. Consider a partition of unity  $\{\psi_i\}$  subordinate to  $\{B_i\}$ , and take the sum

$$f(x) = \sum k\psi_k(x)$$

Then  $f$  is smooth, since locally it is the finite sum of smooth functions. If  $x \notin B_1, \dots, B_n$ , then

$$f(x) = \sum_{k=1}^{\infty} k\psi_k(x) = \sum_{k=n+1}^{\infty} k\psi_k(x) \geq (n+1) \sum_{k=n+1}^{\infty} \psi_k(x) = (n+1)$$

Therefore if  $f(x) < n$ ,  $x$  is in some  $B_i$ . Thus  $f^{-1}((-\infty, n])$  is a closed subset of  $\overline{B_1} \cup \dots \cup \overline{B_n}$ , and is therefore compact.  $\square$

Other existence proofs also follow naturally.

**Lemma 2.31.** *If  $A$  is a closed subset of a paracompact manifold  $M$ , there is a differentiable function  $f : M \rightarrow [0, 1]$  with  $f^{-1}(0) = A$ .*

*Proof.* Let us begin proving this for  $M = \mathbb{R}^n$ . Let  $\{U_i\}$  be a countable cover of  $\mathbb{R}^n - A$  by open unit balls. For each  $U_i$ , pick a smooth  $f_i : \mathbb{R}^n \rightarrow [0, 1]$  positive on  $U_i$ , and equal to zero on  $\mathbb{R}^n - U_i$ . Define

$$\alpha_j = \min \left\{ \left\| \frac{\partial^n f_i}{\partial x_{i_1} \dots \partial x_{i_n}} \right\|_{\infty} : i \leq j, n \leq j \right\}$$

Each  $\alpha_j$  is well defined, because  $f_i$  is  $C^{\infty}$  and tends to zero as we leave  $U_i$ . Define

$$f = \sum_{k=1}^{\infty} \frac{f_k}{\alpha_k 2^k}$$

Then  $f$  is differentiable, since if  $k \geq n$ ,

$$\frac{1}{\alpha_k 2^k} \frac{\partial^n f_k}{\partial x_{i_1} \dots \partial x_{i_n}}$$

so the sums of all partial derivatives converge uniformly, and  $f^{-1}(0) = A$  because the function is positive for some coefficient everywhere else.

To address the general case, let  $M$  be paracompact, and find a cover of coordinate balls  $\{B_{\alpha}\}$  for  $M - A$ . Then we may find a partition of unity  $\{\psi_{\alpha}\}$  for this family, and find smooth  $f_{\alpha} : M \rightarrow \mathbb{R}$  with  $f_{\alpha}^{-1}(0) = A \cap B_{\alpha}$ . But then  $f = \sum \psi_{\alpha} f_{\alpha}$  is smooth (by local finiteness), and satisfies  $f^{-1}(0) = A$ .  $\square$

**Corollary 2.32.** *If  $A$  and  $B$  are disjoint closed subsets of a paracompact manifold  $M$ , then there is a function  $h : M \rightarrow [0, 1]$  with  $h^{-1}(0) = A$ ,  $h^{-1}(1) = B$ .*

*Proof.* Modifying the function obtained in the last theorem, we can find  $f : M \rightarrow [0, 1]$  with  $f^{-1}(1) = B$ . Denote  $f^{-1}(0)$  by  $C$ . If  $g : M \rightarrow [0, 1/2]$  satisfies  $g^{-1}(0) = A$ , and if  $\psi : M \rightarrow [0, 1]$  is a bump function on  $B$ , vanishing on  $A \cup C$ , then

$$h = \psi f + (1 - \psi)g$$

satisfies  $h^{-1}(1) = B$ , because for  $p \notin B$ ,  $f(p), g(p) < 1$ . Now certainly  $h(p) = 0$  for  $p \in C$ . If  $p \in A - C$ ,  $\psi(p) = 0$ , so  $h(p) = g(p) > 0$ , and if  $p \notin A \cup C$ ,  $f(p), g(p) > 0$ , so  $h(p) > 0$ . Thus  $h^{-1}(0) = A$ , and we have shown  $h$  is the needed function in the theorem.  $\square$

## 2.10 Differentiable Manifolds with Boundary

Recall that we may extend differentiability on subsets of Euclidean space in the following way; a map  $f : A \rightarrow B$  between arbitrary subsets of Euclidean space is differentiable if  $f$  can be extended to a differentiable map on an open subset of Euclidean space containing  $A$ .

**Theorem 2.33.** *If  $f : \mathbf{H}^n \rightarrow \mathbb{R}$  is locally differentiable (every point has a neighbourhood on which  $f$  can be locally extended to be differentiable), then  $f$  is differentiable in the sense defined above.*

*Proof.* Let  $\{U_\alpha\}$  be a open cover of  $\mathbf{H}^n$  in  $\mathbb{R}^n$  with smooth functions  $g_\alpha : U \rightarrow \mathbb{R}$  agreeing with  $f$  where the two are jointly defined. Consider a partition of unity  $\{\Phi_\alpha\}$  subordinate to  $\{U_\alpha\}$ . Define  $g = \sum g_k \Phi_k$ , defined on  $\bigcup U_\alpha$ , a open extension of  $\mathbf{H}^n$ . Each pair  $g_k$  and  $\Phi_k$  are differentiable, so the multiplication of the two is differentiable. Since these functions are locally finite, we also have  $g$  differentiable across its domain. If  $p \in \mathbf{H}^n$ , then  $g_k(p) = f(p)$ . Thus

$$g(p) = \sum g_k(p) \Phi_k(p) = \sum f(p) \Phi_k(p) = f(p)$$

since the  $\Phi_k$  sum up to one. Thus  $g|_{\mathbf{H}^n} = f$ , and we have extended  $f$  to be differentiable.  $\square$

This allows us to define a notion of differentiable structure for a manifold with boundary. We can extend the notion of two charts being  $C^\infty$  consistent, because we have extended differentiability on non open subsets of  $\mathbf{H}^n$  to a non-local criterion. Similarly, a map  $f : M \rightarrow N$  can also be considered differentiable if  $y \circ f \circ x^{-1} : x(U) \rightarrow y(V)$  can be extended to be a differentiable function on an open subset of Euclidean space. Thus manifolds with boundary have effectively the same theory as manifolds without boundary.

**Example** (Differentiable structures on the boundary of a manifold). *Given a differentiable manifold with boundary  $M$ , we can assign a unique differentiable structure to  $\partial M$  such that the inclusion map is an imbedding. We can generate it from the atlas consisting of the restriction of charts on  $M$  to the boundary, projected into an  $(n - 1)$  dimensional subspace of  $\mathbb{R}^n$ . The transition maps are easily verified to be  $C^\infty$ .*

One issue with manifolds with boundary is that the rank theorem does not hold. The problem is that, at the boundary, we are restricted in how we reapply coordinates – we can only consider smooth automorphisms of  $\mathbf{H}^n$  rather than  $\mathbb{R}^n$ . However, for immersions  $f$  of manifolds with boundary in manifolds without, we do have coordinate charts for which

$$(y \circ f \circ x^{-1})(a_1, \dots, a_n) = (a_1, \dots, a_n, 0, \dots, 0)$$

because in the theorem we currently have, we need not change the coordinates on the manifold with boundary, only on the manifold without. Thus a submanifold with boundary  $N$  can be identified in manifolds without boundary  $M$  as those subsets which have ‘half slices’, i.e. for each  $p \in N$ , there is a neighbourhood  $U$  of  $p$  in  $M$  and a coordinate chart  $(x, U)$  centred at  $p$  such that either

$$U \cap N = \{q \in U : x^{k+1}(q), \dots, x^n(q) = 0, x^k(q) \geq 0\},$$

in which case  $p \in \partial N$ , or

$$U \cap N = \{q \in U : x^{k+1}(q), \dots, x^n(q) = 0\},$$

in which case  $p \in N^\circ$ .

# Chapter 3

## The Tangent Bundle

Historically, calculus was the subject of infinitesimals, differentiable functions which are ‘infinitesimally linear’. It took over 200 years to make precise the analytical notions defining the field; in the process, infinitesimals vanished from sight, replaced by linear approximations, epsilons and deltas. On manifolds, we cannot discuss global linear approximations, since the space is not globally linear. Thus we must reconcile the use of infinitesimals, since a manifold is ‘infinitesimally linear’. The construction of the tangent bundle of infinitesimals is therefore integral to the development of modern geometry. One can describe modern differential geometry as studying manifolds with additional structure ascribed to the tangent bundle, and so a further analysis of the tangent bundle we constructed in the last chapter is essential.

### 3.1 First

### 3.2 A Smooth Structure on the Tangent Bundle

Recall that if  $M$  is a manifold, then the tangent bundle  $TM$  is the union of the tangent spaces  $T_pM$ , for all  $p \in M$ , where  $T_pM$  is the equivalence class of vectors in various charts on the manifold, where  $[x, v]_p$  is identified with  $[y, w]_p$  if

$$w = D(y \circ x^{-1})(x(p))(v)$$

We view elements of  $TM$  as vectors whose base points emerge from the manifold at every point. Thus there is a projection map  $\pi : TM \rightarrow M$ , which projects a vector to its basepoint, i.e. we set  $\pi[x, v]_p = p$ . On each of the ‘tangent spaces’  $\pi^{-1}(p)$ , often denoted  $T_pM$  or  $M_p$ , we have a vector space structure defined by setting

$$\lambda[x, v]_p + \gamma[x, w]_p = [x, \lambda v + \gamma w]_p$$

The linearity of derivatives makes this construction well defined. One of the nicest parts about the tangent bundle is that we can express the collection of all derivatives of a differentiable map  $f : M \rightarrow N$  in all coordinate systems as a single mathematical object, i.e. via the

map  $f_* : TM \rightarrow TN$ , defined by setting

$$f_*[x, v]_p = [y, D(y \circ f \circ x^{-1})(x(p))(v)]_{f(p)}.$$

For a fixed  $p$ , we let  $f_*|_p$  denote the map from  $T_p M$  to  $T_{f(p)} N$ . This map is linear, since it is really just the linear transformation given by the derivative, but expressed in a coordinate independent fashion. Our goal is to give  $TM$  a smooth structure, such that for each smooth function  $f$ , the function  $f_*$  is smooth.

As an initial, simple example, we consider  $T\mathbb{R}^n$ . Since the space here can be covered by a single coordinate system, we can identify  $T\mathbb{R}^n$  with  $\mathbb{R}^n \times \mathbb{R}^n$ , where we identify  $(p, v)$  with  $[\text{id}, v]_p$ . Often the vector  $(p, v)$  is denoted  $v_p$ , so we remember which part of the product is the ‘vector’, and which part is the base point. Under the identification of  $T\mathbb{R}^n$  with  $\mathbb{R}^n \times \mathbb{R}^n$ , the tangent bundle on  $\mathbb{R}^n$  possesses both topological and smooth structure, which agrees with our intuitions about how tangent vectors change over time. If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is smooth, then  $f_* : T\mathbb{R}^n \rightarrow T\mathbb{R}^m$  is also a smooth map between the tangent spaces, now viewed as smooth manifolds, since the entries of the derivative matrix change smoothly as we vary the base point.

The most geometric way to obtain a smooth manifold structure for an arbitrary manifold  $M$  is to assume  $M$  occurs as a smooth submanifold of  $\mathbb{R}^n$ , for some  $n$ . Then the inclusion map  $i : M \rightarrow \mathbb{R}^n$  is an imbedding, and therefore induces an injective map  $i_* : TM \rightarrow T\mathbb{R}^n$ . Thus we can identify the vectors in  $TM$  with the subset  $i_*(TM)$  of  $\mathbb{R}^n \times \mathbb{R}^n$ . Now  $i_*(TM)$  is a smooth submanifold of  $T\mathbb{R}^n$ . To see this, we let  $(x, U)$  be a chart on  $\mathbb{R}^n$ , such that  $U \cap i_*(M) = \{p \in U : x^{k+1}(p) = \dots = x^n(p) = 0\}$ . Then it is easy to see that

$$(U \times \mathbb{R}^n) \cap i_*(TM) = \{v_p \in U : x^{k+1}(p) = dx^i(v_p) = 0 \text{ for all } i \in \{k+1, \dots, n\}\}.$$

so  $x_*$  give a ‘slice chart’ for  $i_*(TM)$ . It is only slightly technical to show this smooth structure does not depend on the particular embedding of  $M$  in Euclidean space, so, assuming our manifold already occurs in Euclidean space, this solves our problem of giving the bundle a smooth structure. More generally, this argument shows that if  $N$  is a submanifold of  $M$ , then  $TN$  is naturally identified as a subset of  $TM$ .

**Example.** Consider the smooth manifold  $S^1 \subset \mathbb{R}^2$ . Then  $TS^1$  can be identified as a subset of  $\{v_x : x \in S^1, v \in \mathbb{R}^2\}$ . At each point  $x \in S^1$ ,  $T_x S^1$  is one dimensional, and is thus spanned by a single nonzero vector  $v_x \in T_x(\mathbb{R}^2)$ . In fact, we claim that if  $x = (x_1, x_2)$ , then  $T_x S^1$  is spanned by  $(-x_2, x_1)_x$ . Indeed, if  $\theta : U \rightarrow I$  is an angle chart on  $S^1$ , where  $U \subset S^1$  is open and  $I$  is an open interval in  $\mathbb{R}$ , then  $\theta^{-1}(t) = (\cos(t), \sin(t))$ . Thus for each  $s \in \mathbb{R}$ ,

$$\theta_*^{-1}(s_t) = s \cdot (-\sin(t), \cos(t)).$$

But this implies  $T_x S^{-1}$  is spanned by  $(-\sin(\theta(x)), \cos(\theta(x))) = (-x_2, x_1)$ .

**Example.** More generally, we claim that for each  $x \in S^n$ ,  $TS^n$  can be identified with the set of all  $v_x \in T\mathbb{R}^{n+1}$  such that  $x \in S^n$  and  $v \cdot x = 0$ . To see this claim, we fix  $x \in S^n$ . Then we



may choose  $x_2, \dots, x_{n+1} \in \mathbb{R}^{n+1}$  such that if we set  $x_1 = x$ , then  $\{x_1, \dots, x_n\}$  is an orthogonal basis. Then

$$S^n = \{t_1 x_1 + \dots + t_{n+1} x_{n+1} : t_1^2 + \dots + t_{n+1}^2 = 1\}.$$

We consider the open set  $U \subset S^n$  defined by

$$U = \{t_1 x_1 + \dots + t_{n+1} x_{n+1} : t_1 > 0, t_1^2 + \dots + t_n^2 = 1\}.$$

Then there is a smooth chart  $y : U \rightarrow B$ , where  $B$  is the open unit ball in  $\mathbb{R}^n$ , given by setting  $y(t_1 x_1 + \dots + t_{n+1} x_{n+1}) = (t_2, \dots, t_{n+1})$ . The map is certainly smooth, satisfies  $y(x) = 0$ , and the inverse is given by

$$y^{-1}(t) = (1 - |t|^2)^{1/2} x_1 + t_2 x_2 + \dots + t_{n+1} x_{n+1}.$$

But computing the derivative map at  $t = 0$  gives

$$D(y^{-1})(0)(v_2, \dots, v_{n+1}) = v_2 x_2 + \dots + v_{n+1} x_{n+1}.$$

In particular, this means that  $T_x S^n$  is spanned by  $(x_2)_x, \dots, (x_{n+1})_x$ , which was what we needed to show.

This construction is, in low dimensions, the geometrically natural way to view the smooth structure of the tangent space. Since all manifolds can be imbedded in some dimension of Euclidean space, this is certainly a sufficient method to define the tangent space on all manifolds. Nonetheless, it is not entirely elegant because the imbedding is not unique, so there are many different candidates for ‘the’ tangent bundle (though all the candidates will essentially be equivalent), and it is not clear which imbedding will make the tangent bundle easiest to work with.

A similar, but slightly more abstract way to obtain a canonical smooth structure is to use the coordinate charts on a manifold  $M$ . Given any chart  $(x, U)$  on the manifold, the map  $x : U \rightarrow x(U)$  is a smooth map, and as such we can consider  $x_* : TU \rightarrow Tx(U)$ . We claim that the  $x_*$  can be chosen as a consistent family of charts on  $TM$ . Recall the way we construct abstract smooth manifolds from chapter 2. If  $(y, V)$  is another coordinate chart on the manifold, then  $x_*(TV \cap TU) = x(U \cap V) \times \mathbb{R}^n$  is an open subset of Euclidean space, and by definition,

$$(y \circ x^{-1})_*(v_p) = (D(y \circ x^{-1})(p)(v))_{(y \circ x^{-1})(p)}$$

Since  $y \circ x^{-1}$  is  $C^\infty$ , not only does the map change over points smoothly, but the choice of  $D(y \circ x^{-1})$  changes smoothly as  $p$  ranges, which means the entire map, viewed as a map between open subsets of  $\mathbb{R}^n \times \mathbb{R}^n$ , to itself, is smooth. Thus these maps are  $C^\infty$  related, and combine to give a smooth structure on  $TM$ , and we have abstractly defined a smooth structure on  $TM$ . The coordinates  $x_*$  into  $\mathbb{R}^{2n}$  on  $TM$  are often denoted by  $(x, dx)$ , so  $dx^i[x, v]_p = v^i$ .

### 3.3 Vector Bundles

We're not finished with our tangent bundle discussion yet. There are many different perspectives with which we can view the tangent bundle of a manifold, and all are useful at one point or another. The unity in description is best achieved with the concept of a vector bundle. The abstract definition of a vector space emanating from points on a topological space comes from the theory of vector bundles. A *vector bundle*  $\xi$  over a topological space  $X$  is a topological space  $E_\xi$  together with a continuous, surjective projection  $\pi_\xi : E \rightarrow X$ , such that for each  $x \in X$ , the set  $E_x = \pi^{-1}(x)$  has the structure of a finite dimensional vector space, and such that we have local triviality; for each point  $x \in X$ , there is a neighbourhood  $U$  and a homeomorphism  $\phi : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  which is a linear isomorphism on each fibre for  $x \in U$ .

The function  $\dim(E_x)$  is a locally constant function of  $x$ , so if  $X$  is connected,  $\dim(E_x)$  is a constant function of  $x$ , equal to some integer  $n$ , and we refer to  $\xi$  as a  *$n$  dimensional vector bundle*. If  $X$  and  $E$  are both differentiable manifolds, and the projection maps  $\pi$  and triviality maps  $\phi$  are differentiable, then we shall call  $\xi$  a *smooth vector bundle*.

**Example.** For any topological space  $X$ , we have trivial  $n$  dimensional vector bundles  $\varepsilon^n(X)$ , given by the space  $X \times \mathbb{R}^n$  and the projection map  $\pi(v_x) = x$ . In some sense, every vector bundle on  $X$  is obtained from 'twisting' this trivial bundle globally, while keeping the trivial structure locally.

**Example.** The Möbius strip  $\mathbf{M}$  can be seen as a vector bundle over the circle  $S^1$ . Indeed,  $S^1$  can be identified with the orbit space of  $\mathbb{R}$  under the group action  $n \cdot x = (x + n)$ , and  $\mathbf{M}$  as  $\mathbb{R}^2$  under the group action  $n \cdot (x, y) = (x + n, (-1)^n y)$ . Thus the projection map  $\pi(x, y) = x$  filters through the group action and induces a vector bundle structure.

**Example.** If  $\xi$  and  $\eta$  are bundles on two spaces respectively onto two spaces  $X$  and  $Y$ , we can define a bundle  $\xi \times \eta$  on  $X \times Y$  by taking the product of the bundle spaces.

If  $X$  and  $Y$  are spaces with bundles  $(\xi, E)$  and  $(\eta, F)$  on two spaces  $X$  and  $Y$ , then a *bundle map* is a pair of continuous maps  $f : X \rightarrow Y$  and  $f_\# : E \rightarrow F$ , where each  $f_p^\# = (f_\#)|_{X_p}$  is linear from  $X_p$  to  $Y_{f(p)}$ . An isomorphism in the category of bundle maps is known as an *equivalence* of bundles. If  $X$  and  $Y$  are the same space, it is often assumed that  $f$  must be the identity map, so  $f_\#$  maps  $\pi^{-1}(p)$  to  $\pi^{-1}(p)$  linearly for each point  $p$  in the base space. Since  $f$  can be easily obtained from  $f_\#$ , we can describe a bundle map solely by the map  $f_\#$  between bundles, or more particularly by the linear maps  $(f_\#)_p : X_p \rightarrow Y_p$ , which can be viewed as a continuous parameterization of linear maps over  $p$ . Note that now we have defined an equivalence of bundles, we can work backwards and *define* the local triviality conditions of bundles by saying that every vector bundle is locally equivalent to the trivial bundle over suitably small neighbourhoods.

**Theorem 3.1.** Let  $\xi$  and  $\eta$  be bundles over the same base space  $X$ , then a bundle map  $f : \xi \rightarrow \eta$  which is an isomorphism on each fibre is a bundle equivalence.

*Proof.* Let  $(\xi, E)$  and  $(\eta, F)$  be bundles. The map  $f : E \rightarrow F$  is clearly injective, so all that remains is to check the inverse is continuous. But locally at any point  $p \in X$ , we can consider a neighbourhood  $U$  for which there are trivializations  $\phi : \pi_\xi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  and  $\eta : \pi_\eta^{-1}(U) \rightarrow U \times \mathbb{R}^n$ , and then  $g = \eta \circ f \circ \phi^{-1}$  is a bundle map from  $U \times \mathbb{R}^n$  to itself which is an isomorphism on each fibre. Since being a homeomorphism is a local condition, it now suffices to prove that a bundle map from a trivial bundle to itself which is an isomorphism on each fibre is an equivalence.

Note that the map  $g$  induces a continuous choice of linear isomorphisms  $M : U \rightarrow GL(n)$  such that  $g(p, v) = (p, M(p)v)$ . Clearly such  $M(p)$  must exist, and be unique, and then the inverse of  $g$  is given by  $g(p, v) = (p, M(p)^{-1}v)$ . Since the choice of isomorphism  $M^{-1} : U \rightarrow GL(n)$  is continuous, we conclude that  $g$  is an equivalence. Of course, since inversion of matrices is also smooth, it follows that a smooth bundle map which is an isomorphism on each fibre is a smooth bundle equivalence.  $\square$

### 3.4 The Space of Derivations

The algebraists found another characterization of the tangent bundle, which is elegant, but much more abstract than the definition by coordinates. Note that on  $\mathbb{R}^n$ , the space of vectors  $v \in T\mathbb{R}_p^n$  can be identified with the space of directional derivatives, functionals  $D_v$  on  $C^\infty(M)$  defined by

$$D_v(f)(p) = \lim_{h \rightarrow 0} \frac{f(p + hv) - f(p)}{h}.$$

This map satisfies the product rule

$$D_v(fg)(p) = f(p)D_v(g)(p) + g(p)D_v(f)(p).$$

The idea of the algebraic tangent bundle is to identify tangent vectors with certain functionals on  $C^\infty(M)$ . This is compatible with the interpretation of tangent vectors as velocities over the manifold. If we are moving across a surface, then the velocity we are travelling at determines how the surface below us changes, and the measurement of this change can be identified in the velocity we are travelling at.

Since  $C^\infty(M)$  is a vector space, we may consider the dual space  $C^\infty(M)^*$ , which is a monstrous vector space consisting of all linear functionals from  $C^\infty(M)$  to  $\mathbb{R}$ . A derivation at a point  $p \in M$  is a linear functional  $\lambda \in C^\infty(M)^*$  satisfying  $\lambda(fg) = f(p)\lambda(g) + g(p)\lambda(f)$ . A  $C^\infty$  map  $f : M \rightarrow N$  induces a linear map  $f^* : C^\infty(N) \rightarrow C^\infty(M)$ , defined by  $f^*(g) = g \circ f$ , which further induces a map  $f_* : C^\infty(M)^* \rightarrow C^\infty(N)^*$ , defined by  $[f_*\lambda](g) = \lambda(f^*(g))$ . If  $\lambda$  is a derivation at  $p$ , then

$$\begin{aligned} f_*(gh) &= \lambda(f^*(gh)) = \lambda((gh) \circ f) = \lambda((g \circ f)(h \circ f)) \\ &= g(f(p))\lambda(h \circ f) + h(f(p))\lambda(g \circ f) \\ &= g(f(p))f_*(h) + h(f(p))f_*(g) \end{aligned}$$

so  $f_*(\lambda)$  is a derivation at  $f(p)$ . We can directly calculate that  $(f_* \circ g_*) = (f \circ g)_*$ , so that if  $f$  is a diffeomorphism,  $f_*$  is an isomorphism. We shall soon find that we can identify the space of derivations at a point  $p$  as the space of tangent vectors at  $p$ . The identification will match the  $f_*$  defined here with the  $f_*$  defined on the tangent space, providing an elegant algebraic definition of the covariant derivative.

The differential operators

$$\left. \frac{\partial}{\partial x^1} \right|_p, \dots, \left. \frac{\partial}{\partial x^n} \right|_p$$

are all derivations. We will show, in fact, that these operators span the space of all derivations, so that the space is  $n$  dimensional, and can be identified with the tangent bundle under the map

$$\sum v^i \left. \frac{\partial}{\partial x^i} \right|_p \mapsto [x, v]_p$$

This means the set of all derivations *is* the tangent bundle, up to a change in notation. In other words, a ‘tangent’ on the manifold can be thought of a way of measuring change, by moving along the manifold at a certain velocity.

**Lemma 3.2.** *If  $f \in C^\infty(M)$  is constant,  $\lambda(f) = 0$  for any derivation  $\lambda$ .*

*Proof.* By scaling, assume without loss of generality that  $f = 1$  for all  $p$ . Then

$$\lambda(f) = \lambda(f^2) = \lambda(f) + \lambda(f) = 2\lambda(f)$$

We then just subtract  $\lambda(f)$  from both sides of the equation. □

**Lemma 3.3.** *If  $\lambda$  is a derivation, and  $f(p) = g(p) = 0$ , then  $\lambda(fg) = 0$ .*

*Proof.*

$$\lambda(fg) = f(p)\lambda(g) + g(p)\lambda(f) = 0 + 0 = 0$$

We verified the proof by direct calculation. □

We will show this space, though lying in a very high dimensional vector space, is actually very low dimensional.

**Lemma 3.4.** *If  $f \in C^\infty(\mathbb{R}^n)$  and  $f(0) = 0$ , then there exists functions  $f_i \in C^\infty(\mathbb{R}^n)$ , such that  $f(x) = \sum f_i(x)x^i$ , and  $f_i(0) = \partial_i f(0)$ .*

*Proof.* Define  $g(t) = f(tx)$ . Then  $g'(t) = \sum x^i \partial_i f(tx)$

$$f(x) = \int_0^1 g'(t) dt = \sum x^i \int_0^1 \partial_i f(tx) dt = \sum x^i f_i(x)$$

and each  $f_i$  is verified to be  $C^\infty$ , with  $f_i(0) = \partial_i f(0)$ . □

**Theorem 3.5.** *The space of derivations at the origin on  $\mathbb{R}^n$  is  $n$  dimensional, with basis*

$$\left. \frac{\partial}{\partial x^1} \right|_0, \dots, \left. \frac{\partial}{\partial x^n} \right|_0$$

*Proof.* Let  $f \in C^\infty(\mathbb{R}^n)$ . Then  $\lambda(f) = \lambda(f - f_0)$ , and so we can write  $f - f_0 = \sum x^i f_i$ , and then

$$\lambda(f - f_0) = \sum \lambda(x^i) \frac{\partial f}{\partial x^i}(0)$$

Thus

$$\lambda = \sum \lambda(x^i) \frac{\partial}{\partial x^i}$$

The independence of each partial derivative derivation is left to the reader.  $\square$

We can extend this theorem to arbitrary smooth manifolds.

**Lemma 3.6.** *If  $\lambda$  is a derivation at  $p$ , and  $f$  and  $g$  are equal in a neighbourhood of  $p$ , then  $\lambda(f) = \lambda(g)$ .*

*Proof.* We shall prove that if  $f = 0$  in a neighbourhood  $U$  of  $p$ , then  $\lambda(f) = 0$ . Consider a bump function  $\psi \in C^\infty(M)$  such that  $\psi = 1$  at  $p$ , and  $\psi = 0$  outside of  $U$ . Then  $\psi f = 0$ , and

$$0 = \lambda(0) = \lambda(\psi f) = \psi(p)\lambda(f) + f(p)\lambda(\psi) = \lambda(f)$$

hence  $\lambda(f) = 0$ .  $\square$

If  $f$  is only defined in a neighbourhood  $U$  of  $p$ , we may still compute a well-defined value  $\lambda(f)$ . Consider a function  $\psi = 1$  in  $V \subset U$ , and equal to zero outside of  $U$ . Then  $\psi f \in C^\infty(M)$ , and  $\lambda(\psi f)$  is invariant of the bump function chosen, by the last lemma. This implies that we can identify the space of derivations at  $p \in U$  in  $C^\infty(U)$  with  $C^\infty(M)$ , and we find that derivations really act on the *germ* of functions defined in a neighbourhood of  $p$ . When we are working with analytic or complex manifolds, we no longer have bump functions to work with, so we must begin by working on the space of derivations defined on germs of analytic or holomorphic functions from the very beginning of the theory.

**Theorem 3.7.** *The space of derivations at  $p$  is  $n$ -dimensional, and if  $(x, U)$  is a chart centered at  $p$ , then a basis for the space are the partial derivatives*

$$\left. \frac{\partial}{\partial x^1} \right|_p, \dots, \left. \frac{\partial}{\partial x^n} \right|_p$$

*Proof.* By restriction, we can identify derivations at  $p$  on  $M$  with derivations at  $p$  on  $U$ , which can be identified by diffeomorphism with the derivations at the origin on  $x(U) \subset \mathbb{R}^n$ , and therefore with the space of derivations at the origin in  $\mathbb{R}^n$ . This identification maps the partial derivatives at  $p$  with respect to the coordinates  $x$  onto the partial derivatives at  $x(p)$ , and this is all that is needed to verify the proof.  $\square$

If we collect all derivations at all points on a manifold together, we obtain a structure corresponding exactly to  $TM$ . The correspondence is

$$[x, v]_p \mapsto \sum_{k=1}^n v_k \frac{\partial}{\partial x^k} \Big|_p$$

which induces a topology (and smooth structure) on derivations making the correspondence a homeomorphism (diffeomorphism). We will rarely distinguish between the two sets, and alternate between the two notations depending on which is convenient. For instance, we will often speak of a tangent vector operating on functions, or of a derivation as an element of the tangent space. Ultimately, they are the same mathematical objects, viewed from two different lenses.

*Remark.* We can enlarge the space  $C^\infty(M)$  of real-valued smooth functions to the complex vector space of complex-valued smooth functions on  $M$ . The linear functionals  $\lambda$  in the complex dual space  $C^\infty(M)^*$  such that  $\lambda(f) \in \mathbb{R}$  if  $f$  is a real-valued smooth function can then be identified with the original dual space. In the same way, the functionals  $\lambda$  such that  $\lambda(f)$  is purely imaginary when  $f$  is real-valued and smooth can also be identified with the original dual space. If  $\lambda$  is any complex linear functional, and we define

$$\lambda_1(u + iv) = \operatorname{Re}(\lambda(u)) + i \operatorname{Re}(\lambda(v))$$

$$\lambda_2(u + iv) = \operatorname{Im}(\lambda(u)) + i \operatorname{Re}(\lambda(v))$$

Then  $\lambda_1$  and  $\lambda_2$  are complex linear,  $\lambda = \lambda_1 + i\lambda_2$ , and  $\lambda_1$  and  $\lambda_2$  are the unique such functions decomposing  $\lambda$  such that  $\lambda_1(f)$  is real and  $\lambda_2(f)$  is purely imaginary when  $f$  is real-valued. If  $d$  is a derivation on the space of real-valued functions, and if we define  $d(u + iv) = du + idv$ , then  $d$  is also a derivation on complex-valued functions by a trivial calculation. On the other hand, if  $d$  is a complex-valued derivation, then

$$d_1(uv) = \operatorname{Re}(u(p)dv + v(p)du) = u(p)d_1v + v(p)d_1u$$

so  $d_1$ , and, left as an exercise,  $d_2$ , are easily seen to be derivations. Thus every complex-valued derivation can be split into two real-valued derivations, and so it follows that every complex-valued derivation at  $p$  can be written as

$$\sum a_i \frac{\partial}{\partial x_i} \Big|_p$$

where  $a_i \in \mathbb{C}$ . This is sometimes more natural to work with than the original tangent space over the real numbers, and is a kind of ‘complexification’ of the original tangent space.

*Remark.* If  $I$  is the ideal of  $C^\infty(M)$  consisting of all functions  $f$  with  $f(p) = 0$ , and  $I^2$  is the subspace of  $I$  generated by all products of functions in  $I$ , then the lemma above shows that any derivation on  $C^\infty(M)$  vanishes on  $I^2$ . Conversely, if there is a functional  $\lambda : I \rightarrow \mathbb{R}$

which vanishes on  $I^2$ , then it can be extended to a unique derivation on  $C^\infty(M)$ , by defining, for arbitrary  $f \in C^\infty(M)$ ,

$$\lambda(f) = \lambda(f - f(p))$$

an equation which must hold for any derivation which extends  $\lambda$ . It then follows that  $\lambda$  is a linear operator, and  $\lambda$  annihilates constant functions, from which it then follows that

$$\begin{aligned}\lambda(fg) &= \lambda(fg - f(p)g(p)) \\ &= \lambda([f - f(p)][g - g(p)]) + f(p)\lambda'(g) + g(p)\lambda'(f) - 2f(p)g(p)\lambda(1) \\ &= f(p)\lambda'(g) + g(p)\lambda'(f).\end{aligned}$$

Thus derivations on  $C^\infty(M)$  can be identified with  $(I/I^2)^*$ , which we have verified to be finite dimensional, so  $V/W$  is finite dimensional. This is related to the fact that the *cotangent bundle* of  $M$  at  $p$ , defined in the next chapter, can be identified with  $I/I^2$ . This is often the way to construct the tangent bundle to a more varied class of geometric spaces, in particular, those that occur in algebraic geometry. Unfortunately, it is a little analytically unstable. If  $I$  is the ideal of  $C^k(\mathbb{R})$  consisting of functions vanishing at the origin, then  $I/I^2$  is infinite dimensional. For a given  $f$ , we define

$$\text{ord}(f) = \sup \left\{ \alpha > 0 : \lim_{x \rightarrow 0} f(x)|x|^{-\alpha} = 0 \right\}$$

By Taylor's theorem, if  $f \in C^k(\mathbb{R})$ , and  $f(0) = 0$  we can write

$$f(x) = a_1x + \cdots + a_{k-1}x^{k-1} + g(x)x^k,$$

where  $g$  is a continuous function with  $g(0) = 0$ . This means every element  $f$  of  $I^2$  in  $C^1(\mathbb{R})$  can be written as  $\sum_{i=2}^k a_i x^i + x^{k+1}g(x)$  for some continuous  $g$ , and so it follows that  $\text{ord}(f) \geq k + 1$ , if all the  $a_i$  vanish, or otherwise, is an integer between 1 and  $k$ . But this means that none of the functions  $f_\varepsilon(x) = x^{1+\varepsilon}$ , for  $\varepsilon \in (0, 1)$ , lie in  $I^2$ , yet  $f_\varepsilon \in I$ . Moreover,  $\{f_\varepsilon\}$  are linearly independent modulo  $I^2$ , since  $\text{ord}(\sum a_i f_{\varepsilon_i}) = k + \min(\varepsilon_i)$ , which lies in  $(k, k + 1)$ . This means that the coordinate-based tangent space structure should be used in this setting. Alternatively, it is a difficult, but proven theorem that every  $C^k$  manifold has a compatible smooth structure, and we can use this to define the space of derivations on the manifold.

### 3.5 Curves as Vectors

There is a third important view of the tangent bundle on a manifold, which is perhaps the most geometrically visual. To construct  $M_p$ , we consider curves  $c : (a, b) \rightarrow M$  which pass through  $p$  at some time  $t$ . Given our previous construction of the tangent bundle, we can consider the tangent to the curve  $c_*(1_t)$  in  $M_p$ , which represents the speed of the curve passing through the point; any tangent vector can be put in this form for some curve  $c$ . Classically,  $c_*(1_t)$  is often denoted in Leibnizian fashion as  $dc/dt$ , where the additional

parameter  $t$  is obscured. We could have defined  $M_p$  by these curves, provided we identify  $c$  and  $c'$  if  $dc/dt = dc'/dt$ . Of course, without the original tangent bundle to work with, stating this isn't so elegant. We have to fix some coordinate system  $(x, U)$  at  $p$ , and identify two curves  $c$  and  $c'$  with  $c(t) = p$ ,  $c'(t') = p$  if  $(x \circ c)'(t) = (x \circ c')'(t')$ . Our new tangent bundle is obviously equivalent to our original tangent bundle. This is the closest intrinsic way to visualize the tangent vectors on a manifold. If a manifold does not lie in  $\mathbb{R}^n$ , it isn't really fair to see tangent vectors as vectors lying 'off' the manifold, because the tangent vectors don't really 'point' to anything. For instance, in Einsteinian physics, we consider the universe as a 4-dimensional manifold, and seeing the tangent bundle as vectors lying 'outside of the universe' seems particularly strange. But the directions we can travel *are* visualizable from inside the manifold, and curves describe the way these curves travel.

### 3.6 Sections and Vector Fields

A *section* on a vector bundle  $(\xi, E)$  is a continuous map  $f : X \rightarrow E$  for which  $\pi \circ f$  is the identity map. One can think of a section as a continuous choice of a vector assigned to each point in space. Because of the local triviality property, the space of continuous sections, denoted  $\Gamma(E)$  forms a module over the space of real-valued continuous functions on  $X$ . To begin with, sections provide an interesting way to study the topological behaviour of a vector bundle. A *frame* on an  $n$  dimensional vector bundle  $\pi : E \rightarrow B$  is a family of sections  $s_1, \dots, s_n$  such that  $\{s_1(p), \dots, s_n(p)\}$  is a basis at each point.

**Theorem 3.8.** *There exists a frame  $s_1, \dots, s_n : U \rightarrow \pi^{-1}(U)$  on a set  $U$  if and only if there exists a trivialization  $\phi : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$ .*

*Proof.* If  $\phi : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  is a trivialization, we can define a *local frame*  $s_1, \dots, s_n$  on  $U$  by setting  $s_k(p) = \phi^{-1}(p, e_k)$ . On the other hand, let  $s_k$  be a local frame defined on a common domain  $U$ , upon which there is a local trivialization  $\psi : \pi^{-1}(V) \rightarrow V \times \mathbb{R}^n$  for some  $V \subset U$ , then the maps  $\psi \circ s_k$  can be seen as a series of maps from  $V$  to  $\mathbb{R}^n$ , which can be stacked together, row by row, to form a single map  $F : V \rightarrow GL(n)$ . If we define a map  $\phi : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  by  $\phi(\sum a^i s_i(p)) = (p, a)$ , then we find  $(\psi \circ \phi^{-1})(p, v) = (p, F(p)(v))$ , which is a homeomorphism since it is a bundle equivalence. But by patching together over a covering of  $U$  by neighbourhoods  $V$ , we conclude that  $\phi$  is an equivalence everywhere.  $\square$

Sections also provide easy ways to form of *subbundles*  $F \subset E$  of bundles  $\pi : E \rightarrow B$ , which are subsets which also form vector bundles in the sense that  $F$  has the same vector space structure as  $E$  ( $F_p$  is a subspace of  $E_p$  at each point), and under the same topology  $F$  has local trivializations.

**Theorem 3.9.** *A subset  $F \subset E$  is an  $m$  dimensional subbundle of  $E$  if and only if there are  $m$  linearly independent sections  $s_1, \dots, s_m$  into  $F$  in a neighbourhood of each point in the base space  $B$ .*



*Proof.* The trivialization gives us the linearly independent sections, and conversely, a modification of our argument as to why frames give local trivializations show  $F$  has a local trivialization. Instead of arguments about  $GL(n)$ , we need to use arguments about constant rank matrices instead, which we leave to the reader to modify from our previous discussion as to why  $M(n, m; k)$  is a manifold.  $\square$

**Example.** Let  $f : \xi \rightarrow \eta$  be a bundle map between two vector bundles, such that the map  $p \mapsto \text{rank}(f_p)$  is locally constant on the base space of  $\xi$ , then  $\text{Ker}(f)$ , the union of all the kernels of  $f_p$  on the fibres of  $\xi$ , forms a subbundle of  $\xi$ , whose dimension at each point  $p$  is the different in dimensions between the dimension of  $f$  at  $p$  and the rank of  $f_p$ . By locality, it suffices to prove this for bundles maps  $f : \varepsilon^n(X) \rightarrow \varepsilon^m(Y)$  on trivial bundles with a constant rank  $k$ . Notice that in this case  $f$  corresponds to a choice of a continuous function from  $X$  to  $Y$ , as well as a continuous choice  $M : X \rightarrow M(n, m; k)$  such that  $f(p, v) = (f(p), M(p)(v))$ . Slight modifications to our arguments about  $M(n, m; k)$  in the first chapter guarantee that there exists a continuous choice of matrix  $N(p) \in GL(n)$  locally around each point  $p$  such that

$$M(p)N(p) = \begin{pmatrix} A(p) & 0 \\ B(p) & 0 \end{pmatrix}$$

where  $A \in GL(k)$ , and  $B \in M(n - k, k)$ . But then the sections  $s_j(p) = (p, N(p)(e_{k+j}))$ , for  $j \leq n - k$ , are linearly independent and parameterize the kernel of  $f$  locally. It follows from the last theorem that  $\text{Ker}(f)$  is a subbundle of  $\xi$ . Notice that these ideas can also be used to prove that the image of  $f$  in  $\eta$  is a subbundle, because if we choose a continuous  $N(p) \in GL(n)$  such that

$$N(p)M(p) = \begin{pmatrix} A(p) & B(p) \\ 0 & 0 \end{pmatrix}$$

where  $A(p) \in GL(k)$  and  $B(p) \in M(n, n - k)$ , then  $s_j(p) = (p, N(p)M(p)(e_j))$  are linearly independent and constitute a frame for the image of  $f$ .

**Example.** If  $f : \xi \rightarrow \eta$  is a bundle map injective as a map over base spaces, and the rank of  $f$  is locally constant, then we can also make the union  $\text{coker}(f)$  of the cokernels of the  $f_p$ , formed by the union of the cokernels of each  $f_q$  (well defined because  $f^{-1}(q)$  consists of at most one point) and given the quotient topology, into a vector bundle over the base space of  $\eta$ . To see this, it suffices to note the image of  $f$  is a subbundle of the extension space of  $\eta$ , so we may choose a frame  $s_1, \dots, s_k$  locally mapping into the image. This frame may then be extended locally to a frame on entire fibres of  $\eta$ , and these added sections, once projected into the quotient space, constitute a trivialization of the cokernel bundle.

**Example.** If  $(\xi, E)$  is a bundle on  $X$ , and  $f : Y \rightarrow X$  is continuous, we can define an induced bundle  $f^*(\xi)$  on  $Y$ , by letting the extended space be the fibre product

$$Y \times_X E = \{(y, v) \in Y \times E : f(y) = \pi_\xi(v)\}$$

If  $s_1, \dots, s_n : U \rightarrow E$  is a local frame for  $\xi$ , then the maps  $t_k(x) = (x, (s_k \circ f)(x))$  is a local frame on  $f^*(\xi)$ , showing  $f^*(\xi)$  is locally trivial. If  $(\eta, F)$  is another bundle on

$Y$ , with a bundle map  $(\phi, \phi_\#) : \eta \rightarrow \xi$ , such that  $\phi_\# = f$ , then there is a unique map  $\phi \times_X \pi_\eta : F \rightarrow Y \times_X E$  with the standard commuting properties, and one can check fairly easily that this is a bundle map, which gives  $f^*(\eta)$  a universal property. This universal property can be used fairly simply to check that  $g^*(f^*(\pi_\xi))$  is equivalent to  $(f \circ g)^*(\pi_\xi)$ .

**Example.** If  $N$  is a submanifold of another manifold  $M$ , there are two natural vector bundles on  $N$ . The first is the tangent bundle  $N$ . The second is the tangent bundle  $TM|_N$ . The map  $i : N \rightarrow M$  induces an injective map  $i_* : TN \rightarrow TM|_N$  which means we can view  $TN$  as a subbundle of  $TM|_N$ . We call the quotient  $TM|_N/TN$ , denoted  $(TN)^\perp$ , the normal bundle to  $N$  in  $M$ , since vectors in the quotient can in the case where  $M = \mathbb{R}^n$  (or a Riemannian manifold) be identified with the bundle formed by vectors in  $TM$  forming the orthogonal complement to  $TN$  at each point.

**Example.** If  $\xi$  and  $\eta$  are bundles over the same base space  $X$ , then we can define a new bundle  $\xi \oplus \eta$  over  $X$  called the Whitney sum of the bundles, which using our previous exposition can be most easily described as  $\Delta^*(\pi \times \xi)$ , where  $\Delta : X \rightarrow X \times X$  is the diagonal map  $\Delta(x) = (x, x)$ . Thus it is easy to see it is a vector bundle, and it has the unique property that for any bundle maps  $\nu \rightarrow \xi$  and  $\nu \rightarrow \eta$ , there is a unique map  $\nu \rightarrow \xi \oplus \eta$ .

**Example.** If  $\xi$  is a subbundle of  $\eta$ , we can form the quotient bundle  $\eta/\xi$  by taking the quotient vector space on each fibre. If we take a local frame  $s_1, \dots, s_n$  for  $\xi$ , and extend it to a local frame on  $\eta$  by taking additional sections  $t_1, \dots, t_n$ . Then the projections  $\pi \circ t_i$  constitute a local frame for the quotient. The construction of these sections also shows that  $\eta$  is isomorphic to the Whitney sum  $\eta/\xi \oplus \xi$ .

The main sections we see in differential geometry are sections of the tangent bundle on the manifold, and we call these sections *vector fields*. On a smooth vector bundle, we can consider smooth sections, and the smooth vector fields are the vector fields we will really care about. On vector bundles, we denote sections by lower case letters like  $s$  or  $t$ , whereas a vector field is often denoted by capital letters like  $X$ ,  $Y$ , or  $Z$ , and the value of a vector field  $X$  on a manifold  $M$  at a point  $p \in M$  is  $X_p \in T_p$ . Vector fields are so important that we often denote a single vector in the tangent bundle as  $X$ , as well, which could get confusing if we forget to distinguish the two, but often vector fields act just like vectors in the same way that real valued functions ‘act’ like numbers, so this doesn’t cause a problem. Vector fields form a vector space. Locally, around a chart  $(x, U)$ , we may express the vector field in terms of the basis  $X_p = \sum a^i(p) \partial / \partial x^i(p)$ . This vector field is differentiable or continuous if and only if the functions  $a^i$  are smooth or continuous.

The space  $\Gamma(TM)$  of all smooth vector fields is itself a vector space, an algebra over  $C^\infty(M)$ . If  $X \in \Gamma(TM)$ , and  $f \in C^\infty(M)$ , then we define a new function  $Xf \in C^\infty(M)$  by setting  $(Xf)(p) = X_p(f)$ . What’s more, we have the elegant equation  $X(fg) = fX(g) + gX(f)$ , which expresses the pointwise derivation property globally. In general, a *derivation* is a map  $F : A \rightarrow A$  on an algebra such that  $F(ab) = aF(b) + bF(a)$ . If  $F : C^\infty(M) \rightarrow C^\infty(M)$  is any derivation, then we may then define a vector field  $X$  such that  $X_p$  is the unique vector satisfying  $X_p(f) = F(f)(p)$ , and it is easily verified that  $X_p \in T_p$ , and that  $X$  itself is

a smooth vector field. This vector field is the unique field which generates a derivation on  $C^\infty(M)$ , for if locally,  $X = \sum a^i \frac{\partial}{\partial x^i}$ , then  $a^i(p) = X_p(x^i) = F(x^i)(p)$ . The derivation corresponding to  $X$  is  $F$ , so  $C^\infty$  vector fields and derivations on  $C^\infty(M)$  are in one to one correspondence.

### 3.7 Tangents on Manifolds with Boundary

Given a manifold  $M$  with boundary, it is easy to define the tangent bundle on the interior of the manifold, but it is less clear what the fibres of the bundle should look like on the boundary; should they have the same dimension as on the interior, or one dimension less? We shall find it is most convenient, suprisingly, to make the tangent space at the boundary as the dimension of the space on the interior.

The safest option is to look at the space of derivations at points on the boundary of the manifold, which are defined exactly the same as on any manifold without boundary. The locality properties apply, and so it suffices to determine the space  $V$  of derivations on  $\mathbf{H}^n$ . Surely we have the partial derivatives  $\partial/\partial x^1, \dots, \partial/\partial x^{n-1}$ , so the space is at least  $n-1$  dimensional. But perhaps suprisingly, the partial derivative operator  $\frac{\partial}{\partial x^n}$  is also still well defined – to calculate it, we take some  $f \in C^\infty(M)$ , and extend to some differentiable  $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$ , and then take partial derivatives. The value we obtain is independent of extension, because we can also calculate the partial derivative as the limit of the quantities  $(f \circ x^{-1})(te_n)/t$ , as  $t \downarrow 0$  from above. Arguing the same way as in our calculation of the space of derivations in  $\mathbb{R}^n$ , we may take Taylor series to determine that these partial derivatives span the space of all derivations. Thus, viewing  $TM$  as the space of derivations over every point, we find that  $T_p M$  is  $n$  dimensional at the boundary.

An additional feature of the tangent bundle on the boundary of a manifold is we can identify some vectors as ‘outward’ pointing, ‘inward’ pointing, and parallel to the manifold at the boundary. For a chart  $(x, U)$  at the boundary, we can write any derivation at  $p \in U$  as

$$\sum a_i \frac{\partial}{\partial x^i}$$

Recalling that  $\mathbf{H}^n = \{x : x_n \geq 0\}$ , we say the derivation is outward pointing if  $a_n < 0$ , and inward pointing if  $a_n > 0$ , and parallel to the manifold if  $a_n = 0$ . This is well defined even when we vary coordinate systems, because if  $y \circ x^{-1} : U \rightarrow V$  is a diffeomorphism, where  $U, V \subset \mathbf{H}^n$ , then by invariance of domain we know that

$$y \circ x^{-1}(t_1, \dots, t_{n-1}, 0) = (f^1(t_1, \dots, t_{n-1}), \dots, f^{n-1}(t_1, \dots, t_{n-1}), 0)$$

so that for  $p \in \partial M$ ,

$$\left. \frac{\partial y^n}{\partial x^i} \right|_p = \begin{cases} 0 & : i \neq n \\ \text{positive} & : i = n \end{cases}$$

where the fact for  $i = n$  follows because  $y^n$  is always non-negative, and  $(y^n \circ x^{-1})(x(p) + te^n) \rightarrow 0$  as  $t \downarrow 0$ , so that the function  $t \mapsto (y^n \circ x^{-1})(p + te_n)$  must be an increasing function

in a suitably small half neighbourhood of the origin. Thus we find that if a vector  $v$  can be written

$$\sum a_i \frac{\partial}{\partial x^i} = \sum b_j \frac{\partial}{\partial y^j}$$

then

$$b_n = \sum a_k \frac{\partial y^n}{\partial x_k} = a_n \frac{\partial y^n}{\partial x^n}$$

and we see that the sign of  $b_n$  and  $a_n$  agree.

*Remark.* In particular, we note that one can use this process to find a globally defined smooth vector field  $X$  on  $TM|_{\partial M}$  which points outward at each point, since we can form a field locally in each coordinate system and add that up using a partition of unity, noting that outward pointing vectors do not cancel one another out.

### 3.8 Universality of the Tangent Bundle

For a map  $f : M \rightarrow N$ , the map  $f_* : TM \rightarrow TN$  is meant to be a sufficient generalization of the derivative operator on Euclidean space. The fact that this is a ‘universal’ generalization can in fact be proved, once we introduce a categorical viewpoint. Note that the association of  $M$  with  $TM$  and  $f$  with  $f_*$  is a *functor* from the category of  $C^\infty$  manifolds to the category of smooth vector bundles, such that

- The bundle  $T\mathbb{R}^n$  is isomorphic to  $\mathcal{E}^n(\mathbb{R}^n)$  by a trivialization  $t_n$  such that for any map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the diagram commutes, where old  $f_*$  is defined by  $(p, v) \mapsto (f(p), Df(p)(v))$ .
- If  $U \subset M$  is an open submanifold, there are equivalences

$$u_{U,M} : TU \rightarrow (TM|_U)$$

such that if  $i : U \rightarrow M$  is the inclusion map, then commutes, and for any differentiable  $f : M \rightarrow N$ , we have a diagram

The first bullet says that the functor, restricted to Euclidean spaces, is naturally equivalent to the functor which associates  $\mathbb{R}^n$  with  $\mathcal{E}^n(\mathbb{R}^n)$  and  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  with the old definition of the differential. The second says that the family of maps  $u_{U,M}$  is a natural transformation between the tangent functor and the restriction of the functor to open submanifolds.

We shall verify that any functor satisfying these properties is unique up to a natural equivalence. We shall first explore the properties of functors satisfying the properties above. So until necessary, we will let  $M \rightarrow TM$  stand for such a functor. These properties will be obvious for the standard tangent functor, but as long as we don’t use any facts about our constructed tangent bundle, the theorem will remain true for any of the other functors.

Given a chart  $(x, U)$  on a manifold  $M$ , we have a chain of isomorphisms

$$(TM)|_U \xleftarrow{u_{U,M}} TU \xrightarrow{x_*} Tx(U) \xrightarrow{u_{x(U),\mathbb{R}^n}} (T\mathbb{R}^n)|_{x(U)} \xrightarrow{t_n|_{x(U)}} \mathcal{E}^n(x(U))$$

Define  $\alpha_x : (TM)|_U \rightarrow \mathcal{E}^n(x(U))$  to be the chain of compositions. We shall try and understand the properties of  $\alpha_x$  independent of the properties of our tangent bundle, because if we have some other functor  $M \mapsto T'M$ , with other equivalences  $t'_n$  and  $u'_{U,M}$ , then we have  $\beta_x : (T'M)|_U \rightarrow \mathcal{E}^n(x(U))$  defined exactly as  $\alpha_x$  is defined, and we can consider  $\beta_x^{-1} \circ \alpha_x : (TM)|_U \rightarrow (T'M)|_U$ . Provided that this map is independent of  $x$ , we can put the maps together for all  $x$ , and obtain an equivalence between  $TM$  and  $T'M$ .

**Lemma 3.10.** *If  $V \subset U$  is open, and  $y = x|_V$ , then  $\alpha_y = (\alpha_x)|_V$ .*

*Proof.* Denote the inclusion maps by  $i : U \rightarrow M$ ,  $\iota : V \rightarrow M$ ,  $j : V \rightarrow U$ , and  $k : y(V) \rightarrow x(U)$ . Then consider the diagram We verify commutativity by verifying the commutativity of each square. The first square follows by breaking the diagram into triangles. These subtriangles commutes by the universal properties of the functor, and these gives us the commutativity of the square. The second square follows by functoriality, because  $x \circ j = k \circ y$ . The third square commutes for the same reason the first square commutes, and the last square obviously commutes.  $\square$

**Lemma 3.11.** *If  $A \subset \mathbb{R}^n$  and  $B \subset \mathbb{R}^m$  are open, and  $f : A \rightarrow B$  is  $C^\infty$ , then the diagram commutes.*

*Proof.* First, assume that  $f$  can be extended to a map  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Denoting inclusions by  $i : A \rightarrow \mathbb{R}^n$  and  $j : B \rightarrow \mathbb{R}^m$ . Then we have a huge diagram, whose subtriangles and subsquares all obviously commute. and this diagram contains the other diagram as a subdiagram, since we know that old  $g_*$  and old  $f_*$  agree on  $A$  (since  $A$  is an open subset), so that the theorem is proved in this case.

In general, we might not be able to extend the entire map  $f$  to all of  $\mathbb{R}^n$ , but we can at least find a function  $g$  for each  $p \in A$  such that  $g$  agrees with  $f$  on a neighbourhood  $A' \subset A$  of  $p$ . Then we have a commutative diagram where  $i : A' \rightarrow A$  is the inclusion map. Every subshape but the top left and bottom rectangle obviously commutes. First, note that the bottom rectangle is just an instance of this theorem, but where we can extend our map to all of  $\mathbb{R}^n$ , so we have already argued it's commutativity. To see that the top left square commutes, we extend it to a larger diagram. Defining  $j : A \rightarrow \mathbb{R}^n$  to be the inclusion, we have a diagram To prove that  $u_{A,\mathbb{R}^n} \circ i_* = u_{A',\mathbb{R}^n}$ , it suffices to prove that  $j_* \circ i_* = u_{A',\mathbb{R}^n}$ , because  $j_*$  is equal to  $u_{A,\mathbb{R}^n}$  when viewed as functions without a specified codomain. But  $j \circ i$  is just the inclusion of  $A'$  in  $\mathbb{R}^n$ , so this fact is obvious. We have verified enough commutativity to conclude that the composition

$$TA \xrightarrow{f_*} TB \xrightarrow{u_{B,\mathbb{R}^m}} (T\mathbb{R}^m)|_B \xrightarrow{t_m|_B} \mathcal{E}^m(\mathbb{R}^m)|_B$$

is equal to the composition

$$TA \xrightarrow{u_{A,\mathbb{R}^n}} (T\mathbb{R}^n)|_A \xrightarrow{t_n|_A} \mathcal{E}^n(\mathbb{R}^n)|_A \xrightarrow{\text{old } g_*} \mathcal{E}^m(\mathbb{R}^m)|_B$$

on  $TA|_{A'}$ , and since old  $g_*$  is equal to old  $f_*$  on  $(TA)|_{A'}$ , we have verified the theorem over  $A'$ . Since  $A'$  was arbitrary, we obtain commutativity over all of  $A$ .  $\square$

**Lemma 3.12.** *If  $(x, U), (y, V)$  are two coordinates charts with  $p \in U \cap V$ , then  $\beta_y^{-1} \circ \alpha_y$  and  $\beta_x^{-1} \circ \alpha_x$  agree at  $p$ .*

*Proof.* We may assume  $U = V$ , because our first lemma shows the theorem is true in general otherwise. Assuming this is true, we consider the diagram. The triangle obviously commutes, and the rectangle commutes by the second lemma. This implies that  $\alpha_y = \text{old}(y \circ x^{-1})_* \circ \alpha_x$ , and  $\beta_y = \text{old}(y \circ x^{-1})_* \circ \beta_x$ . The desired result follows immediately.  $\square$

Putting together all  $\beta_x^{-1} \circ \alpha_x : (TM)|_U \rightarrow (T'M)|_U$ , we now have a well defined equivalence  $e_M$  from  $TM$  to  $T'M$ . Now we need only prove that this is in fact a natural equivalence – that is, for any  $f : M \rightarrow N$ ,  $e_N \circ f_* = f_\# \circ e_M$ . Let's prove this first for  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . First, note that  $e_{\mathbb{R}^n} = (t'_n)^{-1} \circ t_n$ , because most of the maps involved in the construction of the equivalence are trivial. The properties of the tangent map imply that the squares and triangles of the diagram commute, and this shows the naturality equation holds. Second, note that  $e_M \circ i_* = i_\# \circ e_U$ , where  $i : U \rightarrow M$  is the inclusion of an open submanifold of  $M$ , where  $(x, U)$  is a chart, because  $e_M$  is essentially formed by putting together all  $i_\# \circ e_M \circ i_*^{-1}$ . Similarly, we find that if  $(x, U)$  is a chart, then  $e_{x(U)} \circ x_* = x_\# \circ e_U$ . Putting all this together, we prove that, given  $f : M \rightarrow N$ ,  $e_N \circ f_* = f_\# \circ e_M$  holds at a neighbourhood of each point. Fixing some point  $p$ , we choose a chart  $(x, U)$  containing the point, and a chart  $(y, V)$  containing  $f(p)$ , such that  $f(U) \subset V$ . We may assume  $x(U) = \mathbb{R}^n$  and  $y(V) = \mathbb{R}^m$ . If  $g = y \circ f \circ x^{-1}$ , then we have justified that all the squares in the diagram but the middle one commute, as do the top and bottom triangles, and the rectangle as a whole (in the sense that  $e_{\mathbb{R}^m} \circ g_* = g_\# \circ e_{\mathbb{R}^n}$ ). But a final diagram chase justifies that  $f_\# \circ e_M$  is equal to  $e_N \circ f_*$  on  $TM|_U$ . Since  $(x, U)$  was arbitrary, we have shown that the property is a natural equivalence in full. Thus

**Theorem 3.13.** *There is a unique functor from the category of  $C^\infty$  manifolds to the category of vector bundles which satisfies the bullet point properties we denoted at the beginning of the section.*

The beauty of this categorical proof is that it depends on very little of the structure of  $C^\infty$  manifolds. The proof easily extends to  $C^k$  manifolds, and even to  $C^\omega$  manifolds. It also shows uniqueness of association to  $C^\infty$  vector bundles, and effectively settles the question of how well the tangent bundle represents the linear property of differentiable maps on spaces.

## 3.9 Orientation

The key idea of differential geometry is that classical geometric concepts can be given a rigorous standing when reinterpreted as some structure on the tangent bundle. These are normally just simple extensions of linear algebraic constructions, applied over each fibre of the tangent bundle. The first, easiest concept to introduce, is orientation. In a *real* vector space  $V$  of positive dimension, each tuple  $(v_1, \dots, v_n)$  gives rise to a linear isomorphism  $T : \mathbb{R}^n \rightarrow V$  such that  $T(e_i) = v_i$ . Given another basis  $(w_1, \dots, w_n)$ , we obtain another isomorphism  $S : \mathbb{R}^n \rightarrow V$  with  $S(e_i) = v_i$ , and therefore a linear operator  $T \circ S^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ .

The determinant of this operator is non-zero and therefore positive or negative. We say these tuples are *equally oriented* if the determinant of this operator is positive, and *oppositely oriented* if the determinant is negative. This divides the bases of  $V$  into two equivalence classes, and an *orientation* for  $V$  is a choice of one of these classes. For each orientation  $\mu$ , we let  $-\mu$  denote the opposite orientation to  $\mu$ . Thus an ordered basis  $(v_1, \dots, v_n)$  on  $V$  is *positively oriented* if  $[v_1, \dots, v_n] = \mu$ , and *negatively oriented* if  $[v_1, \dots, v_n] = -\mu$ .

*Remark.* There is a problem with degeneracy when  $V$  has dimension zero. Then there is only a single basis for  $V$ , namely, the empty set, and so only one equivalence class. It is often convenient, however, to distinguish between ‘positively orienting’ this empty basis, and ‘negatively orientating’ this empty basis. Thus in the case when  $V$  is zero dimensional, we consider two different orientations, which either give the empty set a positive orientation, or a negative orientation. The advantage in the zero dimensional case over the higher dimension case is that there is always a natural choice of orientation,  $V$  is *positively orientated* if  $\emptyset$  is a positively oriented basis, and *negatively oriented* if  $\emptyset$  is a negatively oriented basis. We might also say  $V$  has orientation  $+1$  or orientation  $-1$ .

Given two vector spaces  $V$  and  $W$  with fixed orientations  $\mu$  and  $\nu$ , an isomorphism  $T : V \rightarrow W$  is *orientation preserving* if, given an oriented ordered basis  $(v_1, \dots, v_n)$ , i.e.  $\mu = [v_1, \dots, v_n]$ ,  $\nu = [T(v_1), \dots, T(v_n)]$ . Every isomorphism is either orientation preserving or orientation reversing.

The key reason that orientation exists is that the determinant of an operator is always positive or always negative. We see the same phenomenon occur when considering equivalences  $f : \mathcal{E}^n(X) \rightarrow \mathcal{E}^n(X)$ , which can be written

$$f(p, \nu) = \left( p, \sum a_{ij}(p) \nu^j e_j \right)$$

where the  $a_{ij} : X \rightarrow \mathbb{R}$  change continuously, and since the matrix  $(a_{ij})$  is always invertible, it’s determinant is either always positive or always negative. Given a vector bundle  $\xi$  corresponding to the projection map  $\pi : E \rightarrow B$ , we define an orientation for  $\xi$  to be a *smooth* choice of orientation  $\mu_p$  on  $B_p$  for each  $p \in B$  in the sense that if  $t : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  is a local trivialization of  $\pi$ , then the maps  $t_p : B_p \rightarrow \{p\} \times \mathbb{R}^n$  are either all orientation preserving or orientation reversing. To verify that a particular orientation is ‘consistent’, we need only verify it for a set of connected open sets which cover the bundle, because trivializations are already orientation reversing or preserving when they are locally trivialized, as we just calculated noted.

A bundle is called orientable if it has an orientation, and a differentiable manifold is called orientable if its tangent bundle is orientable. An oriented manifold is a manifold with a fixed orientation on its tangent bundle, and if  $f : M \rightarrow N$  is a local diffeomorphism between oriented manifolds of the same dimension, we say  $f$  is orientation preserving if  $f_* : TM \rightarrow TN$  is orientation preserving on each fibre.

**Example.** Since  $T\mathbb{R}^n$  is equivalent to  $\mathcal{E}^n(\mathbb{R}^n)$ , Euclidean space is an orientable manifold, with a canonical orientation induced by the canonical choice of basis, the unit vectors  $\mu_p = [(e_1)_p, \dots, (e_n)_p]$ . If  $U$  is an open subset of  $\mathbb{R}^n$ , then  $TU$  can be embedded in  $T\mathbb{R}^n$ , and we

can define an orientation  $\mu_p$  in the same way. More generally, if  $\pi : E \rightarrow B$  is an orientable bundle, and if  $F \subset E$ , then the bundle  $\pi|_F$  is also orientable.

**Example.** The spheres  $S^n$  are all orientable – to obtain the orientation, we note that, viewing  $TS^n$  as a subset of  $\mathcal{E}^{n+1}(\mathbb{R}^{n+1})$ , we see that  $p_p \notin TS^n$  for any  $p \in S^n$ . We may then define an orientation on  $S^n$  by letting  $[v_1, \dots, v_n]$  be an oriented basis at  $p$  if  $[p, v_1, \dots, v_n]$  is oriented in  $\mathbb{R}^n$ . More generally, if  $M$  is a manifold with boundary with orientation  $\mu$ , then  $\partial M$  has a natural orientation by saying a basis  $(v_1, \dots, v_{n-1})$  of  $\partial M_p$  is orientable if  $[w, v_1, \dots, v_{n-1}] = \mu$  for any outward pointing vector  $w \in M_p$ . If we consider the orientation on  $\mathbb{R}^{n-1}$  as a subset of  $\mathbf{H}^n$ , we see that we obtain  $(-1)^n$  times the standard orientation. The reason for this choice will become clear when we talk about integration on manifolds, in which orientation plays a key role.

**Example.** For any manifold  $M$ ,  $TM$  is always orientable as a differentiable manifold. We calculate the transition map between two charts induced by the coordinate maps  $(x, U)$  and  $(y, V)$  on  $M$  to be

$$\begin{aligned} & (y^1, \dots, y^n, \dot{y}^1, \dots, \dot{y}^n) \circ (x^1, \dots, x^n, \dot{x}^1, \dots, \dot{x}^n)^{-1}[p, v] \\ &= \left[ (y \circ x^{-1})(p), \sum_{i,j} v^j \frac{\partial y^i}{\partial x^j} \Big|_p e_i \right] \end{aligned}$$

and so the matrix of partial derivatives is

$$M = \begin{pmatrix} \left( \frac{\partial y^i}{\partial x^j} \right) & 0 \\ X & \left( \frac{\partial y^i}{\partial \dot{x}^j} \right) \end{pmatrix} = \begin{pmatrix} N & 0 \\ X & N \end{pmatrix}$$

Essentially, we have found that

$$\frac{\partial \dot{y}^i}{\partial \dot{x}^j} = \frac{\partial y^i}{\partial x^j}$$

Now

$$\begin{pmatrix} N^{-1} & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} N & 0 \\ X & N \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & N^{-1} \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ X & I_n \end{pmatrix}$$

Taking determinants on both sides, we find

$$\frac{\det(M)}{\det(N)^2} = 1$$

and so  $\det(M) = \det(N)^2 > 0$ . This implies that the family of charts  $(x, \dot{x})$  have consistent orientations, and so together they give us an orientation on  $T(TM)$ . Here we have used the fact that if we have a family of chart  $\{(x_\alpha, U_\alpha)\}$  covering a manifold  $M$ , and

$$\det \left( \frac{\partial x_\beta^i}{\partial x_\alpha^j} \right)$$

is always positive, then the orientations induced from the  $(x_\alpha)_*$  combine to give an orientation on  $TM$ .



We can view orientation, in another language, through the tool of frames, left to the reader to prove. A choice  $\mu_p$  of orientation on a bundle  $\pi : E \rightarrow B$  is consistent if and only if every local frame is either always orientation preserving or orientation reversing on each fibre (we assume the frame has some particular ordering). This is just a restatement of the duality between local frames and trivializations.

**Example.** If  $V$  and  $W$  are oriented vector spaces, there is a natural orientation on  $V \oplus W$  obtained by declaring that if  $(v_1, \dots, v_n)$  and  $(w_1, \dots, w_m)$  are bases for  $V$  and  $W$ , then  $(v_1, \dots, v_n, w_1, \dots, w_m)$  is an oriented basis for  $V \oplus W$ . If  $M$  and  $N$  are oriented manifolds, there is a natural orientation associated with the product  $M \times N$ . At each pair of points  $(p, q) \in M \times N$ ,  $T_p(M \times N)$  is naturally isomorphic to  $T_p M \oplus T_q N$ , and since  $T_p M$  and  $T_q N$  have orientations,  $T_p M \oplus T_q N$  naturally has an orientation. One can verify quite simply that if  $t : \pi_M^{-1}(U) \rightarrow U \times \mathbb{R}^n$  and  $s : \pi_N^{-1}(V) \rightarrow V \times \mathbb{R}^n$  are orientation preserving trivializations of  $TM$  and  $TN$ , then  $t \times s$  is an orientation preserving trivialization of  $T(M \times N)$ , and there are enough of these trivializations of this sort to cover  $T(M \times N)$ , so the orientation we have given to  $M \times N$  is consistent.

**Example.** If  $\pi : E \rightarrow B$  is a bundle with an orientation  $\mu$ , and  $f : X \rightarrow E$ , then  $f^*(\pi) = \xi : F \rightarrow X$  is orientable, which we can obtain by making the bundle map  $g : F \rightarrow E$  orientation preserving on each fibre. Of course,  $f^*(\pi)$  might be orientable even if  $\pi$  is not orientable. But, in the special case where we consider  $\pi^*(\pi)$ , which is a bundle over  $E$ , then  $\pi^*(\pi)$  is orientable if and only if  $\pi$  is, because we can view the bundle  $\pi$  as the restriction of the bundle  $\pi^*(\pi)$  over the set  $\{v \in E : v \text{ is a zero vector}\}$ , which is isomorphic to  $B$ .

**Example.** If  $\pi : E \rightarrow B$  and  $\xi : F \rightarrow B$  are two orientable vector bundles, then the Whitney sum  $\pi \oplus \xi : K \rightarrow B$  is also orientable, because the bundle can be covered by local trivializations which are just products of trivializations on each factor, and so we can easily define an orientation here by saying that if  $s_k : U \rightarrow E$ ,  $t_k : U \rightarrow F$  combine to give a local frame of  $\pi \oplus \xi$ , then this frame is oriented if either both  $s_k$  and  $t_k$  are oriented in each bundle, or if bundle is not oriented. On the other hand, if  $\pi$  is orientable, but  $\xi$  is non-orientable, then  $\pi \oplus \xi$  is non-orientable. If  $s_k : U \rightarrow F$  is a frame, then, after perhaps shrinking the neighbourhood, we may find an oriented frame  $t_k : U \rightarrow E$ , and then the frames combine to give a frame of  $\pi \oplus \xi$ . We could say  $s_k$  is oriented if the frame on  $\pi \oplus \xi$  is oriented. However, if  $\pi$  is a bundle, then  $\pi \oplus \pi$  is always orientable, because for any finite dimensional vector space  $V$ ,  $V \oplus V$  has a natural orientation such that, if  $(v_1, \dots, v_n)$  is any basis of  $V$ , then  $(v_1, 0), \dots, (v_n, 0), (0, v_1), \dots, (0, v_n)$  is an oriented basis of  $V \oplus V$ . This allows us to give a natural orientation to  $\pi \oplus \pi$  on each fibre, and it is not difficult to check this orientation is consistent in trivializations.

An assignment of orientation can also be seen as a ‘section’, not of the tangent bundle, but of an associated ‘orientation bundle’. Orientations on a vector space do not form a vector space, so this orientation bundle is not a vector bundle, but instead a member of a more general family called ‘fibre bundles’. Given a topological space  $M$ , known as a *model space*, and a *base space*  $X$ , we consider the family of all  $M$  bundles, which are topological spaces

$E$  together with projection maps  $\pi : E \rightarrow X$  such that each  $x \in X$  has a neighbourhood  $U$  such that  $\pi^{-1}(U)$  is homeomorphic to  $M \times U$  which preserves the basepoint in  $U$ . Of course, we can form the *trivial  $M$  bundle*  $\varepsilon(X, M) = M \times X$ , but there are certainly more interesting bundles.

The orientation bundle associated with any vector bundle  $\xi$  over a space  $X$  is a fibre bundle with model space  $M = \{-1, 1\}$ . We let  $O(\xi)$  have the extension space  $X \times \{-1, 1\}$ . Let the extension space of  $\xi$  be denoted  $E$ , and the projection map  $\pi : E \rightarrow X$ . We let  $\pi_O : X \times \{-1, 1\}$  be the projection map on  $O(\xi)$ . For each  $p$ , we consider an arbitrary linear isomorphism  $S_p : E_p \rightarrow \mathbb{R}^n$ . The topological structure on  $X \times \{-1, 1\}$  is then induced such that, if  $t : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  is a trivialization of  $\xi$ , then  $(t_*) : U \times \{-1, 1\}$  is a trivialization of  $O(\xi)$ , given by

$$(t_*)(x_p) = \text{sgn}(\det(t_p \circ S_p^{-1})) \cdot x_p.$$

Given another trivialization  $s : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$ , we have

$$((s_*) \circ (t_*)^{-1})(x_p) = \text{sgn}(\det(s_p \circ t_p^{-1})) \cdot x_p,$$

which is continuous because  $\text{sgn}(\det(s_p \circ t_p^{-1}))$  is actually a value independent of  $p$ , since  $s_p \circ t_p^{-1}$  is a continuously varying family of elements of  $GL(n)$ , hence the determinant is either always positive or always negative. Thus we obtain a topological structure on  $X \times \{-1, 1\}$  which gives it the structure of a fibre bundle. One can view a choice of orientation on  $\xi$  as a continuous section of  $O(\xi)$ , which results in the following theorem.

**Theorem 3.14.** *If  $X$  is locally connected, then  $\xi$  is orientable if and only if there exists a global section  $s : X \rightarrow O(\xi)$ , in which case  $O(\xi)$  is equivalent to  $\varepsilon(X, M)$ , corresponding to the two choices of orientation on  $\xi$ .*

*Proof.* Suppose  $u : X \rightarrow O(\xi)$  is a global section of the orientation bundle. Let  $p \in X$  be a point, and consider a trivialization  $t : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$ . Then  $t_* \circ x : X \rightarrow X \times \{-1, 1\}$  is continuous, and therefore there exists  $a \in \{-1, 1\}$  such that  $(t_* \circ u)(p) = a_p$  for all  $p \in U$ . We let  $t$  be orientation preserving if  $a = 1$ , and orientation reversing if  $a = -1$ . If  $t$  is an orientation preserving trivialization on  $U$ , and  $s$  is another orientation preserving trivialization, then  $s$  is oriented if and only if  $\det(s_p \circ t_p^{-1}) > 0$ , because

$$(s_* \circ u) = [(s_*) \circ (t_*)^{-1}] \circ [(t_*) \circ u],$$

and  $(s_*) \circ (t_*)^{-1}$  is the map given by multiplication by  $\text{sgn}(\det(s_p \circ t_p^{-1}))$ . Thus we see that the orientation we have chosen is consistent.

Conversely, if  $\mu$  is an orientation of  $\xi$ , we can define a section  $u : X \rightarrow O(\xi)$  such that if  $t : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  is an oriented trivialization, then we let  $u(x) = (t_*)^{-1}(1_x)$ . This is certainly well defined, since if  $s$  is any other oriented trivialization, then  $s_* \circ t_*^{-1}$  is the identity map, so

$$(s_*)^{-1}(1_x) = (s_*)^{-1} \circ (s_* \circ (t_*)^{-1})(1_x) = (t_*)^{-1}(1_x).$$

Thus we have a global section of  $O(\xi)$ . □

**Example.** We have already seen that  $O(\varepsilon^n(X))$  is homeomorphic to  $X \times \{-1, 1\}$ , so this provides an argument as to why  $\varepsilon^n(X)$  is orientable.

**Example.** If  $\xi$  is the Möbius bundle over  $S^1$ , with the Möbius strip  $\mathbf{M}$  as the extension space, and the projection map  $\pi : \mathbf{M} \rightarrow S^1$ . Then  $O(\xi)$  is equivalent to the fibre bundle  $\eta$  with extension space  $S^1$  and projection map  $\pi(t) = 2t$ . We essentially argued that  $\xi$  was not orientable by tracing this fibre bundle around the manifold.

### 3.10 Whitney's Embedding Theorem

We now use our results about tangent spaces and differentiability to show that all compact manifolds can be embedded in low dimensional space. This is a easy version of the general Whitney's embedding theorem, which shows this result is true for all manifolds.

**Lemma 3.15.** *Any compact manifold is imbeddable in some Euclidean space.*

*Proof.* Let  $M$  be a compact manifold. Then there is a finite set of charts

$$(x_1, U_1), \dots, (x_n, U_n)$$

covering  $M$ . Since the  $U_k$  is locally finite, by the shrinking lemma we may choose open sets  $V_k \subset U_k$  with  $\overline{V_k} \subset U_k$  which cover  $M$ . Consider a partition of unity  $\psi_k$  equal to 1 on  $V_k$ , and subordinate to  $U_k$ , and define

$$f(p) = (\psi_1(p)x_1(p), \dots, \psi_n(p)x_n(p), \psi_1(p), \dots, \psi_n(p)) = (f_1(p), \dots, f_n(p))$$

then  $f$  is an immersion of  $M$ , because if  $p \in V_k$ , then in a neighbourhood of  $p$  we have  $f_k(p) = x_k(p)$ , so the partial derivative matrix

$$\left( \frac{\partial f^i}{\partial x_k^j} \right)$$

contains the identity as a submatrix, so the map has full rank. The map is also injective, because if  $p \in V_m$ , and  $\psi_k(p)x_k(p) = \psi_k(q)x_k(q)$  and  $\psi_k(p) = \psi_k(q)$  for some  $q$ , then  $\psi_m(q) = \psi_m(p) = 1$ , so  $q \in U_m$ , and  $x_m(p) = x_m(q)$ , so  $p = q$ . Since  $M$  is compact, the immersion is an imbedding, so  $M$  is imbeddable in Euclidean space.  $\square$

This theorem already has an application, for we can now use it to show all compact manifolds are embeddable in a fairly low dimensional Euclidean space.

**Theorem 3.16** (Whitney's Embedding Theorem). *A compact submanifold  $M^n$  of  $\mathbb{R}^N$  can be embedded in  $\mathbb{R}^{2n+1}$ .*

*Proof.* We say a *chord* of  $M$  is a vector in  $\mathbb{R}^N$  of the form  $p - q$ , for distinct  $p, q \in M$ . If  $U$  is the open subset of  $(p, q) \in M \times M$  with  $p \neq q$ , then the map  $f : U \rightarrow S^{N-1}$  given by

$$f(p, q) = \frac{p - q}{|p - q|}$$

is  $C^\infty$ , and since  $2n < N - 1$ ,  $f(U)$  has measure zero since all values are critical. Similarly, if  $V$  is the open set of vectors  $v \in TM \subset T\mathbb{R}^N$  with  $|v| \neq 0$ , then we have a map  $g : V \rightarrow S^{N-1}$  defined by  $g(v_p) = v/|v|$ , and again, since  $2n < N - 1$ ,  $g(V)$  has measure zero. It follows that  $f(U) \cup g(V)$  has measure zero, hence  $S^{N-1} - f(U) - g(V)$  is non-empty, and so there exists a vector  $v$  not parallel to any chord in  $\mathbb{R}^N$ , and not parallel to any tangent space. It follows that the projection of  $M$  onto a hyperplane perpendicular to  $v$  is an injective immersion, and since  $M$  is compact, it is an imbedding. Continuing this process, we may imbed  $M$  into  $\mathbb{R}^{2n+1}$ .  $\square$

*Remark.* If the submanifold is smooth, the embedding is smooth, but more generally, if the submanifold is only  $C^k$  for  $k \geq 1$ , the embedding is still  $C^k$ .

We have proven that all compact manifolds can be embedded in some Euclidean space, so that all compact  $n$  dimensional manifolds can be embedded in  $\mathbb{R}^{2n+1}$ . For non-compact manifolds, this argument only shows that these manifolds can be *immersed* in  $\mathbb{R}^{2n+1}$ , not imbedded. We can also easily turn this theorem into an *approximation result*.

**Theorem 3.17.** *Let  $M$  be a compact,  $C^k$  manifold. Then for any  $C^k$  map  $f : M^n \rightarrow \mathbb{R}^m$ , with  $m \geq 2n + 1$ , and for any  $\varepsilon > 0$ , there is a  $C^k$  embedding  $g : M^n \rightarrow \mathbb{R}^m$  such that  $\|f - g\|_{L^\infty(M)} \leq \varepsilon$ .*

*Proof.* Without loss of generality, we may assume  $M \subset \mathbb{R}^K$  for some  $K$ . Then we can identify  $M$  with the graph of  $f$ , as a subset of  $\mathbb{R}^K \times \mathbb{R}^m$ . Clearly, it suffices to approximate the projection map  $\pi : \mathbb{R}^{K+m} \rightarrow \mathbb{R}^m$  with an embedding. But this is precisely what we have done in the previous theorem.  $\square$

With some additional technicalities, the theorem above generalizes to non-compact manifolds.

**Lemma 3.18.** *For any manifold  $M$ , there exists a proper map  $\rho : M \rightarrow [0, \infty)$ .*

*Proof.* Let  $\{U_k\}$  be a cover of precompact sets, and let  $\phi_k$  be a partition of unity for this cover. We then define

$$\rho = \sum_{k=1}^{\infty} k\phi_k.$$

If  $x \in M$  and  $\rho(x) \leq N$ , then there must exist some  $k \leq N$  such that  $x \in U_k$ . Thus

$$\rho^{-1}[0, N] \subset \bigcup_{k=1}^N U_k,$$

The set  $\bigcup_{k=1}^N U_k$  has compact closure, and since  $\rho^{-1}[0, N]$  is closed in  $M$ , it is actually compact. If  $K \subset \mathbb{R}$  is any compact set, then  $K \cap [0, \infty)$  is compact, hence closed and bounded, and so there is  $N$  such that  $K \cap [0, \infty) \subset [0, N]$ . But then  $\rho^{-1}(K) = \rho^{-1}(K \cap [0, \infty))$  is a closed subset of  $\rho^{-1}([0, N])$ , which is compact, hence  $\rho^{-1}(K)$  is compact.  $\square$

**Theorem 3.19.** *Every  $n$  dimensional manifold  $M$  is embeddable in  $\mathbb{R}^{2n+1}$ .*

*Proof.* The compact argument case certainly gives an injective *immersion*  $f_0 : M \rightarrow \mathbb{R}^{2n+1}$ . Composing this argument with a diffeomorphism of  $\mathbb{R}^{2n+1}$  with its unit ball, we may assume  $|f_0(x)| < 1$  for all  $x$ . Let  $\rho : M \rightarrow [0, \infty)$  be a proper map, and consider the map  $f(x) = f_0(x) + a\rho(x)$  from  $M$  to  $\mathbb{R}^{2n+2}$ . Now fix a unit vector  $a \in \mathbb{R}^{2n+2}$ , and consider the resulting projection  $\pi_a \circ f : M \rightarrow \mathbb{R}^{2n+2}$ , i.e. the map

$$(\pi_a \circ f)(x) = f(x) - (f(x) \cdot a) \cdot a$$

Suppose  $(\pi_a \circ f)_*$  is not injective at  $x \in M$ , and let  $y \in \pi_a \circ f$ . We have

$$(\pi_a \circ f)_*(v_x) = f_*(v_x) - (f_*(v_x) \cdot a_y)a_y.$$

Thus if  $(\pi_a \circ f)_*(v_x) = 0$ , then  $f_*(v_x)$  is a scalar multiple of  $a_x$ . But  $TM$  is a manifold with dimension  $2n$ , hence by Sard's theorem, this is impossible for almost every  $a \in S^{2n+1}$ . In particular, this means we may choose  $a$  such that  $\pi_a \circ f$  is an immersion, and moreover, we may assume that  $a$  is not a pole of this  $S^{2n+1}$ , i.e.  $a \neq (0, \dots, 0, \pm 1)$ . But then  $\pi_a \circ f$  is now actually a *proper map*. This follows because if  $a = (a_0, t)$ , with  $a_0 \in \mathbb{R}^{2n+1}$  and  $t \in \mathbb{R}$ , the last coordinate of  $(\pi_a \circ f)(x)$  has magnitude

$$|\rho(x) - [(f_0(x) \cdot a_0) + \rho(x)t]| = |(1 - t^2) \cdot \rho(x) - t \cdot (f_0(x) \cdot a_0)| \geq (1 - t^2)\rho(x) - t|a_0|.$$

Thus we conclude

$$|(\pi_a \circ f)(x)| \geq (1 - t^2)\rho(x) - t|a_0| \geq (1 - t^2)\rho(x) - 1,$$

and so if  $|(\pi_a \circ f)(x)| \leq M$ , then

$$\rho(x) \leq \frac{M + 1}{1 - t^2}.$$

In particular, this enables us to conclude that there exists a constant  $A$ , and for sufficiently large  $R$ , if  $B_R$  is the closed ball of radius  $R$  at the origin, then  $(\pi_a \circ f)^{-1}(B_R)$  is a closed subset of  $\rho^{-1}([0, B_{AM}])$ , and in particular, is compact. From this, we can conclude  $\pi_a \circ f$  is proper. But then  $\pi_a \circ f$  is a proper and injective, and is therefore a homeomorphism onto its image.  $\square$

*Remark.* For a given manifold  $M^n \subset \mathbb{R}^K$ , viewing  $TM$  as a subset of  $T\mathbb{R}^K$ , we set

$$S(M) = \{v_x \in TM : |v| = 1\}.$$

It is an easy argument that  $S(M)$  is a  $2n - 1$  dimensional submanifold of  $T\mathbb{R}^K$ . The map  $f : S(M) \rightarrow S^{K-1}$  given by  $f(v_x) = v$  is smooth, and therefore by Sard's theorem, forms a set of measure zero if  $2n - 1 < K - 1$ , i.e.  $2n < K$ . If  $a \in S^{K-1}$  is not in the image of this map, and it then follows that if  $\pi_a$  is given by projection onto the hyperplane orthogonal to  $a$ , then  $\pi_a$  is an immersion of  $M$  in  $\mathbb{R}^{K-1}$ . In particular, since we can embed any  $n$  manifold  $M$  in  $\mathbb{R}^{2n+1}$ , we can also *immerse*  $M$  in  $\mathbb{R}^{2n}$ . Generalizing the approximation property established from the embedding theorem, we find that if  $M$  is compact, and  $K \geq 2n$ , then for any smooth map  $f : M \rightarrow \mathbb{R}^K$  and any  $\varepsilon > 0$ , there exists a smooth immersion  $g : M \rightarrow \mathbb{R}^K$  with  $\|f - g\|_{L^\infty(M)} \leq \varepsilon$ .

Whitney's embedding theorem can be improved to show any manifold  $M^n$  can be embedded in  $\mathbb{R}^{2n}$ , and *immersed* in  $\mathbb{R}^{2n-1}$ . However, the approximation theorem above is tight - there are maps  $M^n \rightarrow \mathbb{R}^{2n}$  that cannot be approximated by an embedding. Let us indicate some ideas behind this result.

**Theorem 3.20.** *Let  $M^n$  be a compact manifold. Then there exists a map  $f : M^n \rightarrow \mathbb{R}^{2n-1}$  which is an immersion except at finitely many points.*

*Proof.* Let  $f : M \rightarrow \mathbb{R}^{2k}$  be an immersion. Then we obtain a map  $f_* : TM \rightarrow T\mathbb{R}^{2k}$ . In  $\pi(v_x) = v$  is the projection map from  $T\mathbb{R}^{2k}$  to  $\mathbb{R}^{2k}$ , then we obtain a map  $F = \pi \circ f_*$  from  $TM$  to  $\mathbb{R}^{2k}$ .

Let  $a \in \mathbb{R}^{2k}$  be a regular value of this map. We claim that there are finitely many vectors in  $F^{-1}(a)$ . Fix  $a \in \mathbb{R}^{2k}$  and suppose we can find an infinite sequence  $v_i$  based at points  $p_i \in M$ . By compactness, we may assume that the sequence  $p_i$  converges to some  $p \in M$ . But locally,  $f$  is an immersion, so we can switch to a coordinate system  $(x, U)$  around  $p$  such that for sufficiently large  $i$ ,

$$v_i = \left. \frac{\partial}{\partial x^1} \right|_{p_i}.$$

But this means that the sequence  $v_i$  converges to the vector  $v = \left. \frac{\partial}{\partial x^1} \right|_p$ , for which  $F(v) = a$ . Since  $F$  is not injective in a neighbourhood of  $v$ , it cannot be true that  $F$  is an immersion at  $v$ , so  $a$  is not a regular value.

If  $a$  is a regular value, and  $\pi_a$  is the projection map from  $\mathbb{R}^{2k}$ , onto the hyperplane orthogonal to  $a$ , then  $\pi_a \circ f$  is an immersion except at the base-points of  $F^{-1}(a)$ , completing the proof.  $\square$

The exceptional points to the immersion above are known as *cross caps*. Proving further embedding theorems thus requires a knowledge of this caps.

# Chapter 4

## The Cotangent Bundle

We now begin the real task of differentiable geometry, which is to equip the tangent space of a differentiable manifold with enough structure to begin doing some actual geometry. The first idea is that any natural operation on vector spaces can be applied to vector bundles. The essential idea is phrased more precisely using category theory. A covariant endofunctor  $F$  on the category of finite dimensional vector spaces associates with any two vector spaces  $V$  and  $W$  a map from  $\text{Hom}(V, W)$  to  $\text{Hom}(F(V), F(W))$ , and  $F$  is called a *continuous* functor if the map from  $\text{Hom}(V, W)$  to  $\text{Hom}(F(V), F(W))$  is continuous for each  $V$  and  $W$ .

**Theorem 4.1.** *If  $F$  is a continuous endofunctor, and  $\xi = \pi_0 : E \rightarrow B$  is any vector bundle, then there is a bundle  $F(\xi) = \pi_1 : E' \rightarrow B$  for which  $E'_p = F(E_p)$ , and such that every trivialization  $t : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$  corresponds to a trivialization  $F(t) : \pi_1^{-1}(U) \rightarrow U \times F(\mathbb{R}^n)$ .*

*Proof.* Given  $\xi$ , define  $E' = \bigcup F(E_p)$ , with  $\pi_1 : E' \rightarrow B$  projecting  $F(E_p)$  onto  $p$ . A trivialization  $t$  corresponds to a set of linear maps  $t_p : E_p \rightarrow \mathbb{R}^n$ , for each  $p$  in some open set  $U$ , and we define  $s$  as the map corresponding to the set of maps  $F(t_p) : F(E_p) \rightarrow F(\mathbb{R}^n)$ . We declare a topology on  $F(\xi)$  by letting the maps  $F(t)$  be trivializations. If  $t$  and  $s$  are two trivializations on a common open set  $U$ , then  $F(t) \circ F(s)^{-1} = F(t \circ s^{-1})$  corresponds to a map from  $U \times F(\mathbb{R}^n)$  to itself, which is a homeomorphism if and only if the map  $p \mapsto F((t \circ s^{-1})_p)$  from  $U$  to  $GL(F(\mathbb{R}^n))$  is continuous. But we know that the map  $t \mapsto (t \circ s^{-1})_p$  is continuous, and so the continuity of the corresponding map is also continuous because the functor is continuous. By choosing some isomorphism of  $F(\mathbb{R}^n)$  with Euclidean space once and for all, we obtain a vector bundle  $F(\xi)$ . If  $f : \xi \rightarrow \eta$  is a bundle map, then it is given by maps  $f_p : E_p \rightarrow F_p$ , for each  $p$ , and we can obtain a map  $F(f) : F(\xi) \rightarrow F(\eta)$  by taking the union of the  $F(f_p)$ . The continuity of the functor again guarantees that  $F(f)$  remains continuous, and is therefore a bundle map.  $\square$

Thus a continuous functor on the category of vector spaces extends uniquely to a functor on the category of vector bundles which acts as the original functor on the fibres of bundle maps. *Smooth functors* can be considered similarly, we induces a functor on the category of smooth vector bundles. A similar process can be carried out when  $F$  is a functor of multiple variables, or is contravariant, or is only restricted to manifolds of a fixed dimension (e.g. a

result required for the construction of scalar densities). We will take these generalizations for granted in the sequel.

## 4.1 The Dual of a Vector Bundle

The first object we can equip the tangent bundle is with cotangent vectors, which often operate as differentials in certain geometric arguments. Recall that if  $V$  is a vector space, then there is a natural vector space  $V^*$  associated with it, known as the *dual space* of  $V$ , which is the space of all linear functionals  $f : V \rightarrow \mathbb{R}$ . If  $f : V \rightarrow W$  is a linear map, we can define another linear map, the dual  $f^* : W^* \rightarrow V^*$ , by setting  $f^*(g) = g \circ f$ . Thus the dual can be viewed as a *contravariant* functor in the category of vector spaces. If  $V$  is finite dimensional, and has some basis  $(v_1, \dots, v_N)$ , then we can construct a *dual basis*  $(v_1^*, \dots, v_N^*)$  on  $V^*$  defined by  $v_n^*(v_m) = \delta_{nm}$ . This implies the functor  $V \mapsto V^*$  is smooth, for if a map  $f$  is given in some pair of bases by a matrix  $M$ , then  $f^*$  will be given by the matrix  $M^T$  with respect to the dual bases, and the map  $M \mapsto M^T$  is smooth. Thus, given any vector bundle  $\xi$ , we can associate with it the *dual bundle*  $\xi^*$ , whose elements consist of functionals over a particular fibre.

For any finite dimensional vector space  $V$ ,  $V^*$  is isomorphic to  $V$ . But this is an artificial fact emerging from the fact that  $V^*$  has the same dimension as  $V$ , and there is no ‘canonical’ way to identify  $V$  with its dual. This makes it difficult to naturally identify  $\xi$  with  $\xi^*$  (in the theory of *complex* vector bundles, one cannot necessarily identify  $\xi$  with  $\xi^*$  since these bundles may not be isomorphic). However, since  $V^{**}$  is naturally isomorphic to  $V$ , we can always identify  $\xi^{**}$  with  $\xi$ . This example can be generalized, and the proof of the equivalence is no more complicated.

**Theorem 4.2.** *Let  $F$  and  $G$  be continuous functors on finite dimensional vector spaces naturally isomorphic to one another. Then the extensions of  $F$  and  $G$  to functors on vector bundles are naturally isomorphic to one another.*

*Proof.* Let  $\eta$  be a natural isomorphism between  $F$  and  $G$ . Given a bundle  $(\xi, E)$ , we can define a map  $\eta_\xi : F(\xi) \rightarrow G(\xi)$ , which is an isomorphism on each fibre, by putting together the equivalences  $\eta_{F(E_p)} : F(E_p) \rightarrow G(E_p)$ , for each point  $p$ . The only tricky part is to verify this map is continuous. For any trivialization  $t$  of  $\xi$  on  $U \subset B$ , we have a commutative diagram

$$\begin{array}{ccccc}
 & & \eta_{F(\xi^n(U))} & & \\
 & \swarrow & \text{arc} & \searrow & \\
 U \times F(\mathbb{R}^n) & \xleftarrow{F(t)} & F(\pi)^{-1}(U) & \xrightarrow{\eta_\xi} & G(\pi)^{-1}(U) & \xrightarrow{G(t)} & U \times G(\mathbb{R}^n)
 \end{array}$$

The diagram above implies that in order to prove  $\eta_\xi$  is continuous, it suffices to prove that  $\eta_{F(\xi^n(U))}$  is continuous for each set  $U$ . But this is just  $\eta_{F(\xi^n(U))}(v_p) = \eta_{F(\mathbb{R}^n)}(v)_p$ , which is obviously continuous, because it is a constant choice of a linear map. Thus  $\eta$  truly does extend to a bundle equivalence, and the naturality is clear.  $\square$



The most important case of the dual bundle occurs when we study the tangent bundle  $TM$ , the dual bundle  $(TM)^*$  will be known as the *cotangent bundle* of  $M$ , which we also denote by  $T^*M$ . It is a smooth vector bundle, so we can consider smooth sections, which we call *covector fields*. Unlike elements of  $TM$ , it is unwise to think of elements of  $T^*M$  as vectors pointing in a particular direction. Instead, covectors *act* on vectors; in particular, given a smooth covector field  $\omega$ , and a vector field  $X$ , we can define a function  $\omega(X)$  by  $\omega(X)(p) = \omega_p(X_p)$ . If  $\omega$  and  $X$  are both smooth fields, then  $\omega(X)$  will also be smooth.

**Example.** For any smooth real-valued function  $f : M \rightarrow \mathbb{R}$ , we can define a covector field  $df : M \rightarrow T^*M$ , such that  $df_p(v) = v(f)$  for each  $v \in T_pM$ . The covector field  $df$  is known as the *differential* of  $f$ . If  $X$  is a vector field, then  $df(X) = X(f)$  is a smooth function.

If  $(x, U)$  is a coordinate system on  $M$ , then the  $x^i$  are smooth, and so we can consider the local differentials  $dx^i$ . Now we calculate

$$dx_p^i \left( \frac{\partial}{\partial x^j} \Big|_p \right) = \frac{\partial x^i}{\partial x^j} \Big|_p = \delta_j^i.$$

This calculation implies that the family of vectors  $\{dx_p^i\}$  are precisely the dual basis of the basis  $\{\partial/\partial x^i|_p\}$ . This means every covector field  $\omega$  can be locally written as

$$\omega = \sum \omega \left( \frac{\partial}{\partial x^i} \right) dx^i = \sum \omega_i dx^i.$$

Just like the fact that a coordinate system  $(q, U)$  on  $M$  induces a coordinate system  $(q, dq)$  on  $TU$  (the  $dq^i$  here are precisely the differentials above, acting on tangent vectors to get coordinates), it also induces a coordinate system  $(q, p)$  on  $T^*U$ , where for each  $\omega \in T^*U$ ,  $\omega = \sum p_i(\omega) dq^i$ . The field  $\omega$  is continuous/smooth if on this chart if and only if the  $\omega_i$  are smooth/continuous. In particular, since  $df(\partial_i) = \partial_i(f)$ , we obtain the classical formula

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i,$$

which holds on  $U$ . If  $\sum \omega_i dx^i = \sum \eta_j dy^j$ , then

$$\sum \eta_j dy^j = \sum \eta_j \frac{\partial y^j}{\partial x^i} dx^i.$$

Thus we find

$$\omega_i = \sum \eta_j \frac{\partial y^j}{\partial x^i}.$$

A smooth covector field can be seen as an assignment of  $n$  functions to each coordinate system on a manifold, which are related to one another by the equation above, complicated by the resulting topology of the manifold. The advantage of the modern approach using the tangent bundle is that these complications are summarized in the topology of  $TM$  rather than in the relations between the various coordinate systems.

**Example. TODO:** Move to a later section? In Lagrangian mechanics, one considers Lagrangians  $L : \mathbb{R}_x^N \times \mathbb{R}_{\dot{x}}^N \rightarrow \mathbb{R}$ , and one is lead to study functions  $a : I \rightarrow \mathbb{R}^N$  which satisfy a second order ordinary differential equation of the form

$$\frac{d}{dt} \left\{ \frac{\partial L}{\partial \dot{x}^i}(a, \dot{a}) \right\} = \frac{\partial L}{\partial x^i}(a, \dot{a}),$$

for  $1 \leq i \leq N$ , known as the system of Euler-Lagrange equations. For instance, if  $L(x, \dot{x}) = K(\dot{x}) - V(x)$ , where  $K(\dot{x}) = m|\dot{x}|^2/2$  and  $F(x) = -\nabla V(x)$ , then one obtains Newton's equation  $m\ddot{x} = F(x)$ . The reason to study Lagrange's equation instead of Newton's equation is that the equation is coordinate independent. Indeed, suppose we have two coordinate systems  $(x, U)$  and  $(y, U)$ , where  $U$  is an open subsets of  $\mathbb{R}^N$ . We claim that a function  $a : I \rightarrow U$  satisfies the Euler-Lagrange equations in the  $x$  coordinates if and only if it satisfies the Euler-Lagrange equations in the  $y$  coordinates. Indeed, suppose that

$$\frac{d}{dt} \left\{ \frac{\partial L}{\partial \dot{y}^i}(a, \dot{a}) \right\} = \frac{\partial L}{\partial y^i}(a, \dot{a}),$$

for  $1 \leq i \leq N$ . Then the fact that  $\partial \dot{y}^j / \partial \dot{x}^i = \partial y^j / \partial x^i$  and  $\partial y^j / \partial \dot{x}^i = 0$  and the change of variables formula implies that

$$\begin{aligned} \frac{\partial L}{\partial \dot{x}^i} &= \sum_j \frac{\partial L}{\partial y^j} \frac{\partial y^j}{\partial \dot{x}^i} + \frac{\partial L}{\partial \dot{y}^j} \frac{\partial \dot{y}^j}{\partial \dot{x}^i} \\ &= \sum_j \frac{\partial L}{\partial y^j} \cdot 0 + \frac{\partial L}{\partial \dot{y}^j} \frac{\partial y^j}{\partial x^i} \\ &= \sum_j \frac{\partial L}{\partial \dot{y}^j} \frac{\partial y^j}{\partial x^i}. \end{aligned}$$

Using the fact that

$$\begin{aligned} \frac{d}{dt} \left\{ \frac{\partial y^j}{\partial x^i}(a) \right\} &= \sum_k \frac{\partial y^j}{\partial x^i \partial x^k}(a) \frac{d(x^k \circ a)}{dt} \\ &= \left( \sum_k \frac{\partial y^j}{\partial x^i \partial x^k} \dot{x}^k \right) (a, \dot{a}) \\ &= \frac{\partial}{\partial x^i} \left( \sum_k \frac{\partial y^j}{\partial x^k} \dot{x}^k \right) (a, \dot{a}) \\ &= \frac{\partial \dot{y}^j}{\partial x^i}(a, \dot{a}), \end{aligned}$$

we conclude, again from change of variables, that

$$\begin{aligned} \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{x}^i} \right) &= \sum_j \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{y}^j} \right) \frac{\partial y^j}{\partial x^i} + \frac{\partial L}{\partial \dot{y}^j} \frac{d}{dt} \left( \frac{\partial y^j}{\partial x^i} \right) \\ &= \frac{\partial L}{\partial y^j} \frac{\partial y^j}{\partial x^i} + \frac{\partial L}{\partial \dot{y}^j} \frac{\partial \dot{y}^j}{\partial x^i} = \frac{\partial L}{\partial x^i}. \end{aligned}$$

To study constrained mechanical systems, i.e. a pendulum acting under gravity, but constrained to lie on an arc, one therefore studies a variant of Lagrangian mechanics on a general manifold  $M$ , induced by a Lagrangian  $L : TM \rightarrow \mathbb{R}$ , and with dynamics induced by the Euler-Lagrange equations in coordinates.

We now consider the passage from Lagrangian mechanics to Hamiltonian mechanics, where the cotangent space enters the picture. The generalized momenta of a Lagrangian system  $L : \mathbb{R}_q^N \times \mathbb{R}_{\dot{q}}^N \rightarrow \mathbb{R}$  is given by

$$p_i = \frac{\partial L}{\partial \dot{q}^i}.$$

It is often the case that the matrix

$$\left( \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \right)$$

is invertible, which enables us to use  $(p, q)$  as a coordinate system locally on  $TM$ . We define the Hamiltonian  $H : \mathbb{R}_p^N \times \mathbb{R}_q^N$  in coordinates such that

$$H(p, q) = \sum p_i \dot{q}^i - L(q, \dot{q}).$$

The Euler-Lagrange equations then become

$$\frac{dp_i}{dt} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} = \frac{\partial L}{\partial q^i} = -\frac{\partial H}{\partial q^i},$$

and

$$\frac{dq^i}{dt} = \dot{q}^i = \frac{\partial H}{\partial p_i}.$$

Thus we have converted Lagrange's equations, which are second order in the  $q$  and  $\dot{q}$  variables, into a system of two first order ordinary differential equations in the  $p$  and  $q$  variables, which is often much simpler to analyze.

The main problem we have now is using a Lagrangian formulation on a general manifold, because the generalized momenta  $p$  are not invariantly defined in coordinates. The indices, however, should hint that a global, coordinate independent formulation of the generalized momenta should involve covectors in some way. But the fact that  $p$  is a function of  $q$  and  $\dot{q}$  means it won't quite be a covector field. Given a Lagrangian  $L : TM \rightarrow \mathbb{R}$ , we define a function  $\rho : TM \rightarrow T^*M$  mapping  $T_q M$  to  $T_q^* M$  for each  $q \in M$ , such that in coordinates,

$$\rho(q, \dot{q}) = \sum_i p_i(q, \dot{q}) dq^i.$$

It is simple to check that that  $\rho$  is coordinate invariant, and thus globally defined. The condition that

$$\left( \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \right)$$

is invertible is now equivalent to  $\rho$  being a diffeomorphism onto an open set  $U$  of  $T^*M$ . If we define the Hamiltonian  $H : U \rightarrow \mathbb{R}$  such that

$$H(\eta) = \eta(\rho^{-1}(\eta)) - L(\rho^{-1}(\eta)),$$

which is just a fancy way to write the definition we have given above invariantly. If we consider the Poincaré 2-form  $\omega$  on  $T^*U$ , then we can consider a vector field  $X_H$  on  $TU$  such that  $\omega(X_H, Y) = dH(Y)$ , then Hamilton's equation are now invariant defined by the vector field  $X_H$  on  $TU$ .

All this formalism simplifies considerably if we assume that  $L = T(q, \dot{q}) - V(q)$ , where for each fixed  $q$ ,  $T$  is a positive definite symmetric quadratic form on  $T_qM$ , i.e. we can write

$$T(q, \dot{q}) = \frac{1}{2} \sum g_{ij}(q) \dot{q}^i \dot{q}^j.$$

The quantities  $\{g_{ij}\}$  are known as the masses of the system. Then

$$p_k = \frac{\partial T}{\partial \dot{q}^k} = \sum g_{ki}(q) \dot{q}^i.$$

The quantity  $T$  is induced by a 2-tensor field, which gives  $M$  the structure of a Riemannian manifold. And the equation above thus shows that  $p$  is precisely the musical isomorphism of the velocity  $\dot{q}$  induced by the Riemannian structure on  $M$ . Thus velocity and momentum are duals of one another with respect to the kinetic energy tensor. The Hamilton then becomes

$$\begin{aligned} H(p, q) &= \sum_i p_i \dot{q}^i - \frac{1}{2} \sum_{i,j} g_{ij}(q) \dot{q}^i \dot{q}^j + V(q) \\ &= (1/2) \sum g^{ij}(q) p_i p_j + V(q) \\ &= (1/2) \cdot \langle p, p \rangle + V(q). \end{aligned}$$

If  $F(q) = -\nabla_q V(q)$ , then Hamilton's equation then read

$$\frac{dp}{dt} = F(q)$$

and if  $M = \{g^{ij}\}$ , then

$$\frac{dq}{dt} = Mp.$$

Let  $\phi(x, u)$  be an equation on  $\mathbb{R}^n \times \mathbb{R}^m$  homogenous of degree two in the variable  $u$ , i.e.  $\phi(x, \lambda u) = \lambda^2 \phi(x, u)$ . This implies

$$\begin{aligned} \langle \nabla_u \phi(x, u), u \rangle &= \lim_{\lambda \rightarrow 0} \frac{\phi(x, (\lambda + 1)u) - \phi(x, u)}{\lambda} = \phi(x, u) \lim_{\lambda \rightarrow 0} \frac{(\lambda + 1)^2 - 1}{\lambda} \\ &= 2\phi(x, u) \end{aligned}$$

If we define  $\rho_i = \partial_{u^i}(\phi)$ , then  $2\phi = \sum \rho_i u^i$ . Thus taking differentials gives  $2d\phi = \sum \rho_i du^i + u^i d\rho_i$ . But we also have  $d\phi = \sum \rho_i du^i + \partial_{x^i} \phi dx^i$ , so subtracting these equations gives  $d\phi = u^i d\rho_i - (\partial_{x^i} \phi) dx^i$ . But we also have

$$d\phi = \frac{\partial \psi}{\partial x^i} dx^i + \frac{\partial \psi}{\partial \rho_i} d\rho_i$$

so by uniqueness of the expansion we conclude

$$u^i = \frac{\partial \psi}{\partial \rho_i} \quad - \quad \frac{\partial \phi}{\partial x^i} = \frac{\partial \psi}{\partial x^i}$$

These equations are very useful in mechanics, where  $\phi(x, u) = \sum a_{ij} u^i u^j$  is a measure of kinetic energy, and  $\rho_i = \sum a_{ij} u^j$  is some form of generalized momentum.

Classical differential geometers were not afraid to describe  $df$  as the ‘infinitesimal change’ with respect to  $f$  as the  $x_i$  change. Eventually, it was realized that an infinitesimal change can be described as an action on infinitesimal lengths, or tangent vectors. Indeed, we find

$$f(x + y) = f(x) + \sum y^i \left. \frac{\partial f}{\partial x^i} \right|_x + o(|y|^2) = f(x) + df_x(y_x) + o(|y|^2).$$

Thus the infinitesimal changes  $df$  and  $dx^i$  became functions on tangent vectors, which enables the classical notation to be preserved almost unchanged.

If  $f : M \rightarrow N$  is smooth, then we obtain a differential map  $f_* : TM \rightarrow TN$ . For a fixed  $p \in M$ ,  $f_*|_p$  is a linear map from  $T_p M$  to  $T_{f(p)} N$ , and we can thus take adjoints to obtain a dual map  $(f_*|_p)^*$  from  $T_{f(p)}^* N$  to  $T_p^* M$ . The fact that  $f_*|_p$  is not injective prevents us from putting all these maps together to form a bundle map  $f^* : T^* N \rightarrow T^* M$  (though if  $f$  is an embedding this bundle map is definable). On the other hand, given a covector field  $\omega$  on  $N$ , we can define a *pullback* covector field  $f^* \omega$  on  $M$  by setting  $(f^* \omega)_p = (f_*|_p)^*(\omega_{f(p)})$ . Note that for vector fields, there is no way to *pushforward* a vector field  $X$  on  $M$  to a vector field  $f_* X$  on  $N$ , which makes the analysis of covector fields quite different than that of vector fields.

## 4.2 The Method of Lagrangian Multipliers

The theory we have developed is enough to obtain an incredibly practical theorem, which has usage almost everywhere in applied mathematics, and has theoretical applications as well. Let  $M$  be a differentiable manifold, and let  $N$  be a  $C^\infty$  submanifold. Given a function  $f \in C^\infty(M)$ , we may wish to consider maximizing  $f$  over  $N$ . One method of finding the maximum over  $f$  is to find a cover of  $N$  by coordinate charts  $(x_\alpha, U_\alpha)$ , and then find points where

$$\frac{\partial f}{\partial x_\alpha^i} = 0$$

for all  $i$ , which are candidates for extrema. In terms of our notation, if  $h = f|_N$ , then candidates for extrema are points  $p$  for which  $dh(p) = 0$ . This may prove impractical, especially when the equations for the partial derivatives are cumbersome to solve. The method of Lagrangian multipliers provides an alternate method of proof, applying when  $N$  is specified as the level set of some specifying function  $g : M \rightarrow M'$ , where  $N = g^{-1}(p)$ . We begin by trying to identify  $TN$  and  $T^*N$  as subbundles of  $TM$  and  $T^*M$  using the map  $g$ .

**Theorem 4.3.** *Let  $g : M_1 \rightarrow M_2$  be a map such that  $g$  has locally constant rank on a neighbourhood of  $g^{-1}(p)$ . Then the tangent bundle  $TN$  can be identified with the subbundle of  $TM|_N$  consisting of all vectors  $v$  with  $g_*(v) = 0$ .*

*Proof.* If  $g : M \rightarrow M'$  is rank  $k$  at  $p$ , then choose some coordinate system  $(x, U)$  on  $M$  and  $(y, V)$  centred at  $f(p)$  such that

$$(y \circ g \circ x^{-1})(t^1, \dots, t^n) = (t^1, \dots, t^k, 0, \dots, 0)$$

Then  $(x^{n-k+1}, \dots, x^n)$  is a coordinate system on  $N$ , and

$$g_* \left( \sum a^i \frac{\partial}{\partial x^i} \right) = \sum_{i,j} a^i \frac{\partial y^j \circ g}{\partial x^i} \frac{\partial}{\partial y^j} = \sum_{i=1}^k a^i \frac{\partial}{\partial y^i}$$

and so  $g_* \left( \sum a^i \frac{\partial}{\partial x^i} \right) = 0$  if and only if  $a^i = 0$  for  $i = 1, \dots, k$ , in which case  $\sum a^i \frac{\partial}{\partial x^i}$  is a vector in  $TN$ .  $\square$

**Example.** *This is the most natural way to think of tangent vectors on  $S^n$ , since  $S^n$  is  $f^{-1}(1)$ , where  $f(x) = |x|^2$ . Since  $Df(x)(v) = 2(x \cdot v)$ ,  $TS^n$  can be identified with the subbundle*

$$\{v_p \in T\mathbb{R}^{n+1} |_{S^n} : p \cdot v = 0\}.$$

Consider the special case where  $g : M^n \rightarrow \mathbb{R}^n$  has constant rank  $k$  in a neighbourhood of  $g^{-1}(x)$ . Then we can consider the covector fields  $dg^1, \dots, dg^n$ . Since

$$g_*(v_p) = (dg^1(v), \dots, dg^n(v))_{g(p)},$$

the space  $TN$  can be identified with the subbundle

$$\{v \in TM|_N : dg^1(v) = \dots = dg^n(v) = 0\}.$$

At each point  $p \in N$ ,  $dg^1(p), \dots, dg^n(p)$  span a  $k$  dimensional subset of  $T^*M$ . The adjoint map  $\omega_p \mapsto \omega_p|_{T_p N}$  has rank  $n - k$ , since  $N$  is  $n - k$  dimensional, which implies that  $dg^1(p), \dots, dg^n(p)$  span the kernel of the reduction map.

Now we return our original problem, suppose that  $N$  is specified as the level set of a differentiable map  $g : M \rightarrow \mathbb{R}^n$ , and that  $p \in M$  is a critical point for a function  $f$  on  $N$ . This means that  $df(p)(v) = 0$  for all vectors  $v \in T_p N$ . But this means there must exist scalars  $\lambda_i$  such that

$$df(p) = \sum \lambda_i dg^i(p).$$

Thus we have a coordinate independant way to find a critical points on the set. In the special case where  $g : M \rightarrow \mathbb{R}$ , we need only find  $\lambda$  such that  $df(p) = \lambda \cdot dg(p)$ .

**Example.** Consider finding extrema of the function  $f(x, y) = 5x - 3y$ , subject to the constraint  $x^2 + y^2 = 136$ . Then the constraint function is  $g(x, y) = x^2 + y^2$ , and

$$df = 5dx - 3dy \quad dg = 2xdx + 2ydy$$

Extrema occur at points  $(x, y)$  such that

$$5dx - 3dy = 2\lambda xdx + 2\lambda ydy$$

Which implies that  $2\lambda x = 5$ ,  $2\lambda y = -3$ , and in order for the point  $(x, y)$  to occur in the constraint region, we require  $25/4\lambda^2 + 9/4\lambda^2 = 17/2\lambda^2 = 136$ , so  $1/16 = \lambda^2$ , and so  $\lambda = \pm 1/4$ . If  $\lambda = 1/4$ , then we have an extrema  $(10, -6)$  and  $(-10, 6)$ . Since the constraint region is compact, we must have a maxima and minima, and since the function is non-constant, one of these points must be the maximum, and one the minimum. Since  $f(10, -6) = 68$ ,  $f(-10, 6) = -68$ ,  $(10, -6)$  is the maxima of  $f$ , and  $(-10, 6)$  the minima.

**Example.** Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a self-adjoint operator, and consider maximizing  $\langle Tx, x \rangle$  over  $S^{n-1}$ . If  $f(x) = \langle Tx, x \rangle$ , then

$$f(x) = \sum_{i,j} T_{ij} x_i x_j$$

so we may take differentials,

$$df(x) = \sum_{i=1}^n \left( \sum_{j \neq i} (T_{ij} + T_{ji}) x_j + 2T_{ii} x_i \right) dx^i = 2 \sum_{i,j} T_{ij} x_j dx^i$$

and the constraint region is defined by the function  $g(v) = \sum x_i^2$ , so  $dg(x) = 2 \sum x_i dx^i$ . Since  $S^{n-1}$  is compact, extrema exist, and at this extremum point  $x$  there is  $\lambda$  such that

$$\sum_{i,j} T_{ij} x_j dx^i = \lambda \sum x_i dx^i$$

and so  $Tx = \lambda x$ , implying that  $x$  is an eigenvector of  $T$ , with eigenvalue  $\lambda$ . If  $V$  is the orthogonal complement of the span of  $x$ , then  $T(V) \subset V$ , because if  $\langle w, v \rangle = 0$ , then  $\langle Tw, v \rangle = \langle w, Tv \rangle = \lambda \langle w, v \rangle = 0$ , and  $T : V \rightarrow V$  is still self-adjoint. This implies that if  $V$  is non-trivial, we may find another eigenvector in  $V$ . Continuing this process, we find a sequence of orthogonal eigenvectors  $v_1, \dots, v_n$  for  $T$ , which diagonalizes  $T$ .

The fact that all orthogonal matrices can be diagonalized gives us a general decomposition results for matrices in  $GL(n)$ , which is analogous to the fact that a nonzero complex number can be written as  $rz$  where  $r > 0$  and  $|z| = 1$ . In general, the result is known as the polar decomposition theorem.

**Theorem 4.4.** Any invertible matrix  $M$  can be uniquely decomposed as  $M_1 M_2$ , where  $M_1$  is orthogonal and  $M_2$  is symmetric and positive definite. In fact,  $GL(n)$  is diffeomorphic to  $O(n) \times \mathbb{R}^{n(n+1)/2}$ , where we identify  $\mathbb{R}^{n(n+1)/2}$  with the space of symmetric positive definite matrices under the exponential map on the family of symmetric matrices.

*Proof.* The uniqueness is easy, because if  $M = M_1 M_2 = M'_1 M'_2$ , then

$$(M'_2)^2 = M'_2 (M'_1)^{-1} M'_1 M'_2 = M'^t M = M_2 M_1^{-1} M_1 M_2 = M_2^2$$

But since  $M_2$  and  $M'_2$  are positive definite, this fact implies  $M_2 = M'_2$ , and therefore that  $M_1 = M'_1$ . Existence is a bit more tricky. Given  $M \in GL(n)$ ,  $MM^t$  is self adjoint and positive definite, so that there is a positive definite matrix  $N$  such that  $MM^t = N^2$  (If  $MM^t$  was diagonal, then  $N$  would be easy to find, because we could take square roots along the diagonal, but in general we can just diagonalize  $M$  and then take square roots). Then  $N^{-1}M$  is orthogonal, because  $N^{-1}$  is also positive definite, so

$$(N^{-1}M)^t(N^{-1}M) = M^t N^{-2} M = M^t (MM^t)^{-1} M = I$$

and so  $M = N(N^{-1}M)$ .

Now we claim the association  $M \mapsto (M_1, M_2)$  is continuous. Since  $M_2 = M_1^{-1}M$ , it suffices to prove that the map  $M \mapsto M_1$  is continuous. Suppose that  $M^1, M^2, \dots$  is a sequence in  $GL(n)$  converging to  $GL(n)$ . Then since  $O(n)$  is compact, some subsequence  $M_1^{i_k}, M_1^{i_2}, \dots$  converges to an orthogonal matrix  $N$ . But now  $M_2^{i_k} = (M_1^{i_k})^{-1} M^{i_k}$  converges as well, to  $N^{-1}M$ , and the space of self adjoint matrices is closed so  $N^{-1}M$  is self adjoint. Since  $M^{i_k} = M_1^{i_k} M^{i_k}$  converges to  $N(N^{-1}M)$ , it follows that  $N = M_1$  and  $N^{-1}M = M_2$ , hence  $M^{i_k} \rightarrow M_1$ . But now we conclude that in general  $M_1^k \rightarrow M_1$  because this argument can be adapted to show every subsequence contains a further subsequence converging to  $M_1$ . Since the map  $(M_1, M_2) \rightarrow M_1 M_2$  is continuous, this verifies the decomposition is actually a homeomorphism. In fact, the map  $(M_1, M_2) \mapsto M_1 M_2$  is obviously smooth, as is its inverse, since it is actually analytic in its matrix entries. So the correspondence is a diffeomorphism.  $\square$

### 4.3 Tensors

We now consider tensors, which are a generalization of covectors as linear functionals to multilinear functionals. If  $V$  is a vector space, then we let  $T^n(V)$  denote the space of maps  $f : V^n \rightarrow \mathbb{R}$  which are *multilinear*, or linear in each variable. Of course,  $T^1(V)$  is just our ordinary dual space  $V^*$ . If  $T$  is an  $n$  tensor, and  $S$  is an  $m$  tensor, then the *tensor product*  $(T \otimes S)(v, w) = T(v)S(w)$  is an  $n + m$  tensor, and the tensor product is an associative, distributive operation which turns the space  $\bigoplus_{n=1}^{\infty} T^n(V)$  into a graded algebra (by convention, we can let  $T^0(V)$  denote real numbers, and define  $\lambda \otimes T = \lambda T$ , so that  $T(V) = \bigoplus_{n=0}^{\infty} T^n(V)$  is a graded algebra with identity. In particular, if  $f_1, \dots, f_n$  is a basis for  $V^*$ , then a basis for  $T^m(V)$  is given by the elements  $f_{i_1} \otimes f_{i_2} \otimes \dots \otimes f_{i_m}$ , with  $i_1, \dots, i_m \in [n]$ , so  $T^m(V)$  has dimension  $n^m$ . Given a linear map  $f : V \rightarrow W$ , we can define a map  $f^* : T^n(W) \rightarrow T^n(V)$  by setting  $f^*(\omega)(v_1, \dots, v_n) = \omega(f(v_1), \dots, f(v_n))$ . Thus we obtain a continuous tensor functor, so for each vector bundle  $\xi$ , we can define the bundle  $T^n(\xi)$  of  $n$  tensors over  $\xi$ . In particular, we obtain the bundle  $T^n(TM)$  of tensors over the tangent bundle, which we also denote by  $T^n M$ . Given a coordinate system  $x$  around a point  $p \in M$ , a basis for  $T_p^1 M$  is given by  $dx_p^1, \dots, dx_p^n$



for some coordinate system  $x$  around  $p$ , a basis for  $T_p^m M$  is given by  $dx_p^{i_1} \otimes \cdots \otimes dx_p^{i_m}$ . Thus an  $m$  tensor field  $\omega$  on  $M$  can be locally written as

$$\omega = \sum a_{i_1 \dots i_m} dx^{i_1} \otimes \cdots \otimes dx^{i_m}$$

and  $\omega$  is continuous/smooth if and only if the functions  $a_{i_1 \dots i_m}$  are. If

$$\sum a_{i_1 \dots i_m} dx^{i_1} \otimes \cdots \otimes dx^{i_m} = \sum b_{j_1 \dots j_m} dy^{j_1} \otimes \cdots \otimes dy^{j_m},$$

then we have the horrendous conversion formula

$$a_{i_1 \dots i_m} = \sum b_{j_1 \dots j_m} \frac{\partial y^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial y^{j_m}}{\partial x^{i_m}},$$

where the products here are just ordinary products of functions.

Though not as often occurring as covariant tensors, contravariant tensor fields are defined similarly, as multilinear functionals operating on  $V^*$ , and the space of  $n$  contravariant tensors is denoted  $T_n(V)$ .  $T_1(V)$  is just  $V^{**}$ , and is naturally isomorphic to  $V$ , and we can take tensor products to obtain all elements of  $T_n(V)$ , so if  $V$  has a basis  $v_1, \dots, v_n$ , then a basis for  $T_m(V)$  is given by  $v_{i_1} \otimes \cdots \otimes v_{i_m}$ . Given  $f : V \rightarrow W$ , we obtain  $f^* : T_n(V) \rightarrow T_n(W)$  in the obvious way, and this functor gives us  $T_n(\xi)$  for any vector bundle  $\xi$ . Putting covariant and contravariant tensors gives us the space  $T_n^m(V)$  of mixed tensors, multilinear functions over  $m$  copies of  $V$  and  $n$  copies of  $V^*$ , and we can therefore form  $T_n^m(\xi)$ . Over the tangent bundle  $TM$ , we can form the space  $T_n^m(TM)$ , also denoted by  $T_n^m M$ , and vector fields over these sets can be locally written as

$$\sum a_{i_1 \dots i_n}^{j_1 \dots j_m} dx^{i_1} \otimes \cdots \otimes dx^{i_n} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_m}},$$

where if

$$\begin{aligned} \sum a_{i_1 \dots i_n}^{j_1 \dots j_m} dx^{i_1} \otimes \cdots \otimes dx^{i_n} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_m}} \\ = \sum b_{\alpha_1 \dots \alpha_n}^{\beta_1 \dots \beta_m} dy^{\alpha_1} \otimes \cdots \otimes dy^{\alpha_n} \otimes \frac{\partial}{\partial y^{\beta_1}} \otimes \cdots \otimes \frac{\partial}{\partial y^{\beta_m}}, \end{aligned}$$

then

$$a_{i_1 \dots i_n}^{j_1 \dots j_m} = \sum_{\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m} b_{\alpha_1 \dots \alpha_n}^{\beta_1 \dots \beta_m} \frac{\partial y^{\alpha_1}}{\partial x^{i_1}} \cdots \frac{\partial y^{\alpha_n}}{\partial x^{i_n}} \frac{\partial x^{j_1}}{\partial y^{\beta_1}} \cdots \frac{\partial x^{j_m}}{\partial y^{\beta_m}}.$$

We call an element of  $T_n^m(V)$  an  $(m, n)$  tensor, or a tensor with  $m$  covariant indices and  $n$  contravariant indices.

There are all kinds of algebraic tricks one can do on mixed tensors. Consider the endofunctor which associates a vector space  $V$  with the vector space  $\text{End}(V)$ , and for each isomorphism  $f : V \rightarrow W$ , we consider  $f_* : \text{End}(V) \rightarrow \text{End}(W)$  by setting  $f_*(T) = f \circ T \circ f^{-1}$ . This functor is continuous, so we can consider the bundle  $\text{End}(\xi)$  for each vector bundle  $\xi$ . For this space it is only *equivalences*  $f : \xi \rightarrow \phi$  which extend to maps  $f_* : \text{End}(\xi) \rightarrow \text{End}(\phi)$ .

Now the mixed tensor space  $T_1^1(V)$  is isomorphic to  $\text{End}(V)$ , where  $S : V \rightarrow V$  corresponds to the tensor  $\omega_S(v, \lambda) = \lambda(S(v))$ . If  $f : V \rightarrow W$  is an isomorphism, and  $S : W \rightarrow W$  is an endomorphism, then

$$\omega_{f_*S}(v, \lambda) = \lambda((f_*S)(v)) = \lambda((f \circ S \circ f^{-1})(v)) = (f^*\omega_S)(v, \lambda)$$

Thus we have found a natural isomorphism between  $T_1^1(V)$  and  $\text{End}(V)$ . More generally, we have a natural isomorphism between  $V^* \otimes W$  and  $\text{Hom}(V, W)$ . Thus  $\text{End}(\xi)$  is naturally equivalent to  $T_1^1(\xi)$ , and so any operation we can perform on  $\text{End}(V)$  can be transferred naturally to  $T_1^1(V)$ . For instance, we can define the *trace*, or *contraction* of a tensor  $\omega \in T_1^1(V)$  to be the trace of the endomorphism  $S$  corresponding to  $\omega$ . If we fix a basis, and write  $\omega = \sum a_i^j e_i^* \otimes e_j$ , then  $\text{trace}(\omega) = \sum a_i^i$ . In particular, a smooth vector field on  $T_1^1(TM)$  can be contracted pointwise to produce a smooth function. Similarly, we can always contract a bottom and top index on an  $(n, m)$  tensor to form a  $(n - 1, m - 1)$  tensor.

## 4.4 Local Operators and Derivational Viewpoints

We previously identified the smooth vector fields on a manifold  $M$  with the derivations on  $C^\infty(M)$ . We now identify tensor fields as certain operators on vectors fields. Given an  $n$  tensor field  $\omega$  and  $n$  vector fields  $X_1, \dots, X_n$ , we can define a smooth function

$$\omega(X_1, \dots, X_n)(p) = \omega(p)(X_1(p), \dots, X_n(p)),$$

and  $\omega$  operates as a multilinear map on  $\Gamma(TM)^n$ , where the map is linear not only over real numbers, but also  $C^\infty$  functions, since the map is ‘defined pointwise’. We find that all such multilinear maps are just tensor fields in disguise.

**Theorem 4.5.** *If  $T : \Gamma(TM)^n \rightarrow C^\infty(M)$  is a  $C^\infty(M)$  multilinear map, there is a unique smooth  $n$  tensor field  $\omega$  on  $M$  such that  $T(X_1, \dots, X_n) = \omega(X_1, \dots, X_n)$ .*

*Proof.* We first note that if  $X_k = Y_k$  in a neighbourhood of  $p$ , then

$$T(X_1, \dots, X_n)(p) = T(Y_1, \dots, Y_n)(p)$$

To see this, we find a smooth bump functions  $f_1, \dots, f_n$  equal to one in a neighbourhood of  $p$  such that  $f_k X_k = f_k Y_k$ , and then

$$\begin{aligned} T(X_1, \dots, X_n)(p) &= f_1(p) \dots f_n(p) T(X_1, \dots, X_n)(p) \\ &= T(f_1 X_1, \dots, f_n X_n)(p) \\ &= T(f_1 Y_1, \dots, f_n Y_n)(p) \\ &= f_1(p) \dots f_n(p) T(Y_1, \dots, Y_n)(p) \\ &= T(Y_1, \dots, Y_n)(p) \end{aligned}$$

Thus  $T$  is locally defined. To show that  $T(X_1, \dots, X_n)(p)$  depends only on  $X_1(p), \dots, X_n(p)$ , it suffices to show that  $T(X_1, \dots, X_n)(p) = 0$  if  $X_k(p) = 0$ , for some  $k$ . If  $(x, U)$  is a coordinate system around  $p$ , then we can write

$$X_n = \sum b_n^i \frac{\partial}{\partial x^i}.$$

But then

$$T(X_1, \dots, X_k, \dots, X_n)(p) = \sum b_1^{i_1}(p) \dots b_n^{i_n}(p) T(\partial_{i_1}, \dots, \partial_{i_n})(p)$$

which shows that  $T$  is defined pointwise. We can now set

$$\omega(p)(v_1, \dots, v_n) = T(X_1, \dots, X_n)(p)$$

where  $X_k(p) = v_k$ , and we have verified this is well defined.  $\omega$  is smooth because

$$\omega = \sum \omega_{i_1 \dots i_n} dx^{i_1} \otimes \dots \otimes dx^{i_n}.$$

Thus  $\omega_{i_1 \dots i_n} = T(dx^{i_1}, \dots, dx^{i_n})$  is a smooth function.  $\square$

Because of this, we do not distinguish tensor fields from their corresponding operators on vector fields. Similar results hold for contravariant tensors, which operate on covector fields.

If  $E$  and  $F$  are vector bundles on some manifold  $M$ , we say a map  $f : \Gamma(E) \rightarrow \Gamma(F)$  is a *local operator* if whenever a section  $s \in \Gamma(E)$  vanishes in a neighbourhood of a point  $p$ , then  $f(s)(p) = 0$ . We say it is a *pointwise operator* if  $f(s)(p) = 0$  whenever  $s(p) = 0$ .

**Example.** The space  $C^\infty(M)$  can be considered as the space of sections over the trivial bundle  $\varepsilon^1(M)$ . Vector fields  $X$ , viewed as first order linear operators, operate as linear maps from  $C^\infty(M)$  to itself. These operators are local, since if  $f$  vanishes around a neighbourhood of a point  $p$ , then  $X(f)(p) = 0$ . But they are not point operators since a function can surely vanish at a point, whereas its derivative might not.

**Example.** We will soon define the exterior derivative operator  $d$ , which maps alternating tensor fields  $\Gamma(\Omega^k(M))$  to  $\Omega^{k+1}(M)$ . The derivative operator is local, but not pointwise defined.

Because of the existence of bump functions, a local map  $f : \Gamma(E) \rightarrow \Gamma(F)$  naturally extends to a sheaf morphism from sections of  $E$  to sections of  $F$ . We have essentially verified in the argument above that  $C^\infty(M)$  linear maps are pointwise operators. On the other hand, if  $T : \Gamma(E) \rightarrow \Gamma(F)$  is a pointwise operator, we can define a bundle map from  $E$  to  $F$  by letting  $T(v_p) = T(s)(p)$ , where  $s \in \Gamma(E)$  has  $s(p) = v_p$ . Thus every pointwise operator is  $C^\infty(M)$  linear. In general, a  $C^\infty(M)$  linear operator can therefore be thought as an operator defined pointwise. This is a basic instance of the *Serre Swan correspondence*, which says that module structures over  $C^\infty(M)$  can be identified with vector bundle structures on  $M$ .

## 4.5 The Classical Viewpoint

The classical viewpoint of vectors in the tangent bundle, and tensors in general, is obsessed with the coefficients associated to these objects and the way they transform with respect to coordinate systems. We previously mentioned that for a coordinate chart  $(x, U)$ , then for any  $(n, m)$  tensor field  $\omega$ , there are functions

$$a_{i_1 \dots i_n}^{j_1 \dots j_m} \in C^\infty(U)$$

such that, on  $U$ ,

$$\omega = \sum_{\substack{i_1, \dots, i_n \\ j_1, \dots, j_m}} a_{i_1 \dots i_n}^{j_1 \dots j_m} dx^{i_1} \otimes \dots \otimes dx^{i_n} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{j_m}}$$

If  $(y, V)$  is another coordinate system, then there are functions

$$b_{i_1 \dots i_n}^{j_1 \dots j_m} \in C^\infty(V)$$

such that on  $V$ ,

$$\omega = \sum_{\substack{i_1, \dots, i_n \\ j_1, \dots, j_m}} b_{i_1 \dots i_n}^{j_1 \dots j_m} dy^{i_1} \otimes \dots \otimes dy^{i_n} \otimes \frac{\partial}{\partial y^{j_1}} \otimes \dots \otimes \frac{\partial}{\partial y^{j_m}}.$$

Then, restricted to  $U \cap V$ , we have the monstrous transformation rule

$$a_{i_1 \dots i_n}^{j_1 \dots j_m} = \sum_{\substack{\alpha_1, \dots, \alpha_n \\ \beta_1, \dots, \beta_m}} b_{\alpha_1 \dots \alpha_n}^{\beta_1 \dots \beta_m} \frac{\partial y^{\alpha_1}}{\partial x^{i_1}} \dots \frac{\partial y^{\alpha_n}}{\partial x^{i_n}} \frac{\partial x^{j_1}}{\partial y^{\beta_1}} \dots \frac{\partial x^{j_m}}{\partial y^{\beta_m}}.$$

Classically, one *defines* a  $(n, m)$  tensor on a  $k$  dimensional manifold as an assignment of  $k^{n+m}$  functions to each chart  $(x, U)$ , defined on  $U$ , such that the transformation above is satisfied on  $U \cap V$ , where  $(y, V)$  is another coordinate system. An *invariant* is a  $(0, 0)$  tensor, or a smooth function on the entire manifold. Let us consider some examples of this terminology.

**Example.** A classical differential geometer will tell you that the Kronecker delta function  $\delta_i^j$  is a tensor. What they mean by this is, if we assign the same  $n^2$  functions  $\delta_i^j$  to every coordinate system, then for any two coordinate systems  $x$  and  $y$ ,

$$\delta_i^j = \frac{\partial x^j}{\partial x^i} = \sum \frac{\partial x^j}{\partial y^\alpha} \frac{\partial y^\alpha}{\partial x^i} = \sum \delta_\alpha^\beta \frac{\partial x^j}{\partial y^\beta} \frac{\partial y^\alpha}{\partial x^i}$$

so, mystically, these coordinate systems really do define a smooth tensor field  $A$  over  $T_1^1(TM)$ . To figure out what  $A$  really is, we calculate

$$A(X, \omega) = \sum \omega_i X^i = \omega(X)$$

so  $A$  is just the tensor which operates as evaluation of a functional on a vector field. In terms of the identification of  $\text{End}(TM)$  with  $T_1^1(TM)$ , the tensor  $A$  corresponds to the identity map.

It is a useful skill to be able to translate the classical language into modern language, and vice versa, and be able to prove things in both settings. Thus in the following, we state theorems in their classical context, prove them by classical means, then reexpress them into modern theorems and prove them in this context.

**Theorem 4.6.** *If  $\sum v_i \mu^i$  is an invariant for any covector field  $\sum v_i dx^i$ , then  $\sum \mu^i \frac{\partial}{\partial x^i}$  is a well defined vector field.*

*Proof.* For any covector field  $\sum v_i dx^i = \sum (v')_i dy^i$ , we have

$$\sum \mu^i v_i = \sum \mu'^i v'_i$$

and we have the relation  $(v')_i = \sum v_j \frac{\partial x^j}{\partial y^i}$ . If we choose  $v_\alpha = 1$ , and  $v_i = 0$  otherwise, we find

$$\mu^\alpha = \sum \mu^i v_i = \sum_i \mu'^i v'_i = \sum_{i,j} \mu'^i \frac{\partial x^j}{\partial y^i} v_j = \sum_i (\mu')^i \frac{\partial x^\alpha}{\partial y^i}$$

so

$$\sum_i \mu^i \frac{\partial}{\partial x^i} = \sum_{i,j,k} (\mu')^k \frac{\partial x^i}{\partial y^k} \frac{\partial y^j}{\partial x^i} \frac{\partial}{\partial y^j} = \sum_{j,k} (\mu')^k \frac{\partial y^j}{\partial y^k} \frac{\partial}{\partial y^j} = \sum_k (\mu')^k \frac{\partial}{\partial y^k}$$

is a well defined vector field, and is uniquely determined since we can determine the value of  $\mu^i$  at any point  $x$  by choosing a smooth vector field  $v$  such that  $v_i(x) = 1$ ,  $v_j(x) = 0$ , and then  $\mu^i(x) = \sum \mu^i(x) v_i(x)$ .  $\square$

*Remark.* In modern terms, the functions  $v_i$  correspond to a  $C^\infty(M)$  linear map  $T\mu = \sum \mu^i v_i$  from  $\Gamma(T^*M)$  to  $C^\infty(M)$ . Such a map is an element of  $\Gamma(T^{**}M)$ , which is naturally isomorphic to  $\Gamma(TM)$ , so the linear map corresponds to a unique smooth vector field  $v$ , and the  $v_i$  are obviously the coordinates of  $v$ .

## Chapter 5

# Lie Derivatives and Differential Equations

So far, we've managed to construct a manifold, and a coordinate-independent way to measure derivatives on the manifold via the tangent bundle. However, a problem results when we wish to measure the rate of change of a rate of change. But it's difficult to measure the rate of change of vector field  $X$ , because we cannot compare  $X_p$  and  $X_q$  for  $p \neq q$ , as the elements lie in different vector spaces. We can of course switch to a coordinate system for close enough  $p$  and  $q$ , associating with each  $X$  a function  $F(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  we can differentiate, but this doesn't really apply to a global perspective in any way, and one cannot naively patch together these maps so that the derivative of a vector field is a vector field. The idea here is to provide a direction in which to take a derivative, which in this case is induced by a vector field. We recall the main existence and uniqueness result from ordinary differential equations.

**Theorem 5.1.** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a locally Lipschitz function, then at any point  $p \in \mathbb{R}^n$ , there is a suitably small  $\varepsilon$  and a unique curve  $x : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$  with  $x(0) = p$ , and  $x' = f \circ x$ , and if we extend  $x$  to have the maximum possible interval domain, then either  $x$  is defined everywhere, or  $x$  is defined on some interval  $(a, b)$ , and  $\lim_{t \rightarrow a} |x(t)| = \lim_{t \rightarrow b} |x(t)| = \infty$ .*

The theorem has the following interpretation in the theory of differentiable manifolds. Given any differentiable vector field  $X : M \rightarrow TM$ , and  $p \in M$ , there is a suitably small  $\varepsilon$  and a unique differentiable curve  $x : (-\varepsilon, \varepsilon) \rightarrow M$  such that  $x'(t) = X_{x(t)}$ . We shall call a curve satisfying this equation an *integral curve* for  $X$ . This result obviously follows because we can switch to any particular coordinate system. The unique curves imply that we can define a function  $\phi$  on some open subset of  $M \times \mathbb{R}$  by setting  $\phi_t(p) = x(t)$ , where  $x$  is the unique integral curve for  $X$  with  $x(0) = p$ . Often  $\{\phi_t\}$  is also denoted by  $\{\exp(tX)\}$  or  $\{e^{tX}\}$ .

If  $\phi_{t+h}(p)$ , and  $\phi_t(\phi_h(p))$  are all well defined, then  $\phi_t(\phi_h(p)) = \phi_{t+h}(p)$  because if  $x : (-\varepsilon_0, \varepsilon_0) \rightarrow M$  is an integral curve satisfying  $x(0) = p$ , and  $y : (-\varepsilon_1, \varepsilon_1) \rightarrow M$  is an integral curve with  $y(0) = x(h)$ , then the uniqueness theorem tells us that  $y(u) = x(h+u)$  where these functions are defined, because  $z(u) = x(h+u)$  is also an integral curve with  $z(0) = x(h)$ .  $\phi$  is defined on an open submanifold of  $M \times \mathbb{R}$ , and it is a very difficult theorem of functional

analysis to show that if  $X$  is a  $C^k$  vector field, then  $\phi$  is  $C^k$ . We will take this for granted in the sequel. Since  $\phi_{-t} = \phi_t^{-1}$  on a small enough neighbourhood of every domain, we conclude that each  $\phi_t$  is ‘almost’ a diffeomorphism, and is certainly a local diffeomorphism. For compact manifolds, we can obtain a global family of diffeomorphisms.

**Theorem 5.2.** *If  $X$  is compactly supported, then  $\phi$  is defined everywhere.*

*Proof.* If  $K$  is the support of  $X$ , then  $\phi_t$  is defined on  $K^c$  for all  $t$ , because if  $p \notin K$ , then  $x(p) = p$  is an integral curve of  $X$ . If we cover  $K$  by finitely many coordinate charts  $(x_1, U_1), \dots, (x_n, U_n)$ , such that  $\phi_t$  is defined for  $|t| < \varepsilon_i$  on  $U_i$ , then we have defined  $\phi_t$  on  $M$  for all  $|t| < \min(\varepsilon_1, \dots, \varepsilon_n)$ , and we can then define  $\phi$  for arbitrarily large real numbers by considering the composition  $\phi_t^n = \phi(nt)$ .  $\square$

One way to interpret the local uniqueness and existence of vector fields is that all vector fields are the same locally up to a change of coordinates, when the vector field is nonvanishing.

**Theorem 5.3.** *If  $X$  is defined on  $M$  with  $X_p \neq 0$ , then there is a coordinate chart  $(x, U)$  around  $p$  such that for  $q \in U$ ,*

$$X_q = \left. \frac{\partial}{\partial x^1} \right|_q.$$

*Proof.* It obviously suffices to prove this for  $p = 0$ , and  $M = \mathbb{R}^n$ . We may also start with a coordinate chart such that  $X_0 = \partial_1$ . Consider the induced diffeomorphism  $\phi$  from the vector field  $X$ , and take the map  $\alpha : (-\varepsilon, \varepsilon) \times U \rightarrow \mathbb{R}^n$ , where  $U \subset \mathbb{R}^{n-1}$  is a neighbourhood defined by  $\alpha(t, x) = \phi_t(0, x)$ . Clearly the map is  $C^\infty$ , and if  $f : x(U) \rightarrow \mathbb{R}$  is arbitrary, then

$$\alpha_* (\partial_i(0, a)) (f) = \frac{\partial(f \circ \alpha)}{\partial x^i}(0, a) = \frac{\partial f}{\partial x_i}(a)$$

Thus  $\alpha_* \partial_i(0, a) = \partial_i(0, a)$ . We also calculate

$$\alpha_* (\partial_t(0, a)) (f) = \lim_{h \rightarrow 0} \frac{f(\phi_h(a)) - f(a)}{h} = X_a(f)$$

so  $\alpha_* (\partial_t(0, a)) = X_a$ . We conclude that  $\alpha_*$  is nonsingular in a neighbourhood of  $(0, a)$ , and so locally  $\alpha$  is invertible, and the inverse, which we call  $y$ , is a coordinate system. But we know that

$$\frac{\partial}{\partial y^1}(a) = \alpha_* \left( \frac{\partial}{\partial t}(a) \right) = X_a,$$

which completes the proof.  $\square$

Note that here we have used the fact that

$$X_p f = \lim_{t \rightarrow 0} \frac{f(\phi_t(p)) - f(p)}{t}.$$

This limit suggests that the translation maps  $\phi_h$  can be used to measure the rate of change along the vector  $X_p$ . As foreshadowed at the beginning of the chapter, we can use this process to measure the rate of change of many other objects defined on a differentiable manifold.

First, we switch notations, and denote  $X(f)$  by  $L_X(f)$ , and call it the *Lie derivative* of  $f$  along  $X$ . Using this notation, we can also consider the derivatives of covariant vector fields  $\omega$ , obtaining a new covariant vector field

$$(L_X\omega)_p = \lim_{t \rightarrow 0} \frac{(\phi_t^*\omega)_p - \omega_p}{t}.$$

Thus for any vector  $v \in M_p$ ,

$$\omega((\phi_t)_*v) = \omega(v) + t(L_X\omega)(v) + o(t).$$

Similarly, we consider the derivatives of a vector field  $Y$  by a vector field  $X$ , defined by

$$(L_XY)_p = \lim_{t \rightarrow 0} \frac{Y_p - ((\phi_t)_*Y)_p}{t}.$$

Here the order of terms is switched because  $((\phi_t)_*Y)_p$  is really pushing forward the point on the vector field at  $Y_{\phi_{-t}}$ , and therefore is looking backwards in time rather than forwards. Thus  $((\phi_{-t})_*Y)_p = Y_p + t(L_XY)_p + o(t)$ . Both derivatives exist whenever the fields in question are smooth, because they are just partial derivatives of a suitably defined smooth function on  $M \times \mathbb{R}$ . Moreover, we will now calculate the derivative in coordinates, which also provides a proof of existence. It will be easiest to compute this once we verify the linearity of the derivative, as well as the product rule.

**Theorem 5.4.** *If  $f \in C^\infty(M)$ ,  $\omega \in \Gamma(T^*M)$ , and  $X, Y \in \Gamma(TM)$ , then*

$$(i) \quad L_X(f\omega) = (L_Xf)\omega + f(L_X\omega).$$

$$(ii) \quad L_X(fY) = (L_Xf)Y + f(L_XY).$$

$$(iii) \quad L_X(Y(f)) = (L_XY)(f) + Y(L_Xf).$$

$$(iv) \quad L_X(\omega(Y)) = \omega(L_XY) + \omega(L_XY).$$

*Proof.* We write

$$\begin{aligned} L_X(f\omega)(p)(v) &= \lim_{h \rightarrow 0} \frac{\phi_h^*(f\omega)(p)(v) - (f\omega)(p)(v)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(\phi_h(p))\omega(\phi_h(p))((\phi_h)_*(v)) - f(p)\omega(p)(v)}{h} \\ &= (FG)'(0) \end{aligned}$$

Where  $F(t) = f(\phi_t(p))$ , and  $G(t) = \omega(\phi_t(p))((\phi_t)_*(v))$ . Since  $F(0)' = (L_Xf)(p)$ , and  $G(0)' = (L_X\omega)(p)(v)$ , we find that

$$L_X(f\omega)(p)(v) = f(p)(L_X\omega)(p)(v) + (L_Xf)(p)\omega(p)(v)$$

and this completes the proof of (i). The propositions (ii), (iii), and (iv) are proved in essentially the same way.  $\square$



It will be easiest to begin computing the Lie derivative of a covariant vector field. Given  $X$ , with induced flow  $\phi_t$ , we calculate that

$$((\phi_t)^* df)(v_p) = ((\phi_t)_*(v_p))(f) = v_p(f \circ \phi_t) = d(f \circ \phi_t)(v_p)$$

Thus using the smoothness of the values involved, we conclude that

$$L_X(df) = \lim_{t \rightarrow 0} \frac{d(f \circ \phi_t)(p) - df(p)}{t} = \lim_{t \rightarrow 0} d\left(\frac{f \circ \phi_t - f}{t}\right)(p) = d(Xf)$$

In particular, this means that if  $X = \sum a^i \partial_i$ , then

$$L_X dx^i = d(X(x^i)) = d(a^i) = \sum \frac{\partial a^i}{\partial x^j} dx^j.$$

Thus if  $\omega = \sum b_i dx^i$ , then

$$L_X(\omega) = \sum X(b_i) dx^i + b_i L_X(dx^i) = \sum a^j \frac{\partial b_i}{\partial x^j} dx^i + b_i \frac{\partial a^i}{\partial x^j} dx^j$$

Thus we see from the coefficients that the Lie derivative involves both the derivatives of the coefficients of  $\omega$  and the coefficients of  $X$ . This makes sense, since  $L_X(\omega)$  is neither  $C^\infty(M)$  linear in  $X$  nor in  $\omega$ . Now we can use the ‘product rule’ to calculate the Lie derivative of a vector field.  $X = \sum a^i \partial_i$  and  $Y = \sum b^i \partial_i$ , then we know that

$$\begin{aligned} \sum a^j \frac{\partial b^i}{\partial x^j} &= L_X(b^i) \\ &= L_X(dx^i(Y)) \\ &= (L_X dx^i)(Y) + dx^i(L_X Y) \\ &= \sum \frac{\partial a^i}{\partial x^j} b^j + dx^i(L_X Y). \end{aligned}$$

Thus we find

$$L_X Y = \sum dx^i (L_X Y) \partial_i = \sum \left( a^j \frac{\partial b^i}{\partial x^j} - \frac{\partial a^i}{\partial x^j} b^j \right) \frac{\partial}{\partial x^i}$$

This very complicated expressions can be interpreted in a very simple way, namely, for  $f \in C^\infty(M)$ ,

$$(L_X Y) = X(Yf) - Y(Xf).$$

Thus we often define the right hand side as the *Lie bracket*, or *commutator* of  $X$  and  $Y$ , denoted  $[X, Y]$ . We verify quite simply that

$$\begin{aligned} [X, Y]_p(fg) &= X(Y(fg)) - Y(X(fg)) \\ &= X(gY(f) + fY(g)) - Y(gX(f) + fX(g)) \\ &= X(g)Y(f) + gX(Y(f)) + X(f)Y(g) + fX(Y(g)) \\ &\quad - Y(g)X(f) - gY(X(f)) - Y(f)X(g) - fY(X(g)) \\ &= g[X, Y]_p(f) + f[X, Y]_p(g) \end{aligned}$$

so second order terms cancel out, and so the difference is really a derivation at  $p$ . There is an easier, coordinate way to do this. The idea is that if  $X$ , and  $f$  are fixed, with flow  $\phi_t$ , then there is a smooth family of functions  $g_t$  such that  $f \circ \phi_t = f + tg_t$ , and  $g_0 = Xf$ . We then calculate that

$$\begin{aligned} ((\phi_t)_* Y)_p(f) &= (\phi_t)_*(Y_{\phi_{-t}(p)})(f) = Y_{\phi_{-t}(p)}(f \circ \phi_t) \\ &= (Yf)(\phi_{-t}(p)) + t(Yg_t)(\phi_{-t}(p)) \end{aligned}$$

and so

$$(L_X Y)(f) = \lim_{t \rightarrow 0} \frac{(Yf)(p) - (Yf)(\phi_{-t}(p))}{t} - (Yg_t)(\phi_{-t}(p))$$

The first term is easily as  $X_p(Yf)$ , and the second is easily seen to converge to  $(Yg_0)(p) = Y_p(Xg)$ .

The commutator reveals certain properties of the Lie derivative which are not obvious from a first glance. For instance,  $[X, Y] = -[Y, X]$ , and so as a consequence,  $L_X(Y) = -L_Y(X)$ , and  $L_X(X) = 0$ . Secondly, we see that  $[aX_1 + bX_2, Y] = a[X_1, Y] + b[X_2, Y]$ , so  $L_{aX_1 + bX_2} = aL_{X_1} + bL_{X_2}$ , giving linearity in the lower argument. Finally, we see the important Jacobi identity

$$[X, [Y, Z]] = [[X, Y], Z] + [Y, [X, Z]],$$

which can be seen as a kind of ‘product rule’ for the Lie bracket.

A useful interpretation of the Lie bracket of vector fields is as a measure of ‘noncommutativity’ of two vector fields. We begin by noting that if  $X$  is a vector field on  $M$  generating a flow  $\{\phi_t\}$ , and  $\alpha : M \rightarrow N$  is a diffeomorphism, then  $\alpha_* X$  generates the flow  $\{\alpha \circ \phi_t \circ \alpha^{-1}\}$ .

**Lemma 5.5.** *If  $\alpha : M \rightarrow M$  is a diffeomorphism,  $X \in \Gamma(TM)$  generates a flow  $\{\phi_t\}$ , then  $\alpha_* X = X$  if and only if  $\phi_t \circ \alpha = \alpha \circ \phi_t$  for all  $t$ .*

*Proof.* For then  $\alpha_* X = X$  generates the flow  $\{\alpha \circ \phi_t \circ \alpha^{-1}\}$ , which means that  $\alpha \circ \phi_t \circ \alpha^{-1} = \phi_t$ . Conversely, if  $\phi_t = \alpha \circ \phi_t \circ \alpha^{-1}$ , then  $\alpha_* X$  generates the flow  $\{\phi_t\}$ , so  $\alpha_* X = X$ .  $\square$

**Theorem 5.6.** *Let  $X, Y \in \Gamma(TM)$ , generating flows  $\{\phi_t\}$  and  $\{\psi_s\}$ . Then  $[X, Y] = 0$  if and only if  $\phi_t \circ \psi_s = \psi_s \circ \phi_t$  for all  $t, s$  for which this equation makes sense.*

*Proof.* If  $\phi_t \circ \psi_s = \psi_s \circ \phi_t$ , then the last lemma implies that  $(\phi_t)_* Y = Y$  for all  $t$ . But this means that for all  $p \in M$ ,

$$L_X(Y)_p = \lim_{t \rightarrow 0} \frac{Y_p - ((\phi_t)_* Y)_p}{t} = 0,$$

so  $[X, Y] = 0$ . Conversely, if  $[X, Y] = 0$ , fix  $p \in M$ , and consider a curve  $c(t) = ((\phi_t)_* Y)_p$  in  $M_p$ . Then

$$\begin{aligned} c'(t) &= \lim_{s \rightarrow 0} \frac{((\phi_{t+s})_* Y)_p - ((\phi_t)_* Y)_p}{s} \\ &= (\phi_t)_* \left( \lim_{s \rightarrow 0} \frac{((\phi_s)_* Y)_{\phi_{-t}(p)} - Y_{\phi_{-t}(p)}}{s} \right) \\ &= (\phi_t)_* L_X(Y)_{\phi_{-t}(p)} = 0. \end{aligned}$$

Thus  $c(t) = c(0) = Y_p$  for all  $t$ . Since  $p$  was arbitrary, this implies  $(\phi_t)_* Y = Y$ , and we can apply the lemma to conclude  $\phi_t \circ \psi_s = \psi_s \circ \phi_t$  for all  $s$ . Since  $t$  was arbitrary, this completes the proof.  $\square$

**Corollary 5.7.**  $X^1, \dots, X^k$  are vector fields on  $M$  such that  $X^1(p), \dots, X^k(p)$  are linearly independent for each  $p \in M$ , and if  $[X^i, X^j] = 0$  for all  $1 \leq i, j \leq k$ , then around each point  $p \in M$ , there exists a coordinate system  $(x, U)$  with  $p \in U$  and such that

$$X^i = \frac{\partial}{\partial x^i}.$$

*Proof.* For the flows  $\{\phi_t^1\}, \dots, \{\phi_t^n\}$ ,  $\phi_t^i \circ \phi_s^j = \phi_s^j \circ \phi_t^i$  for all indices  $i, j$  and times  $t, s$ . Define

$$\alpha(t^1, \dots, t^k) = (\phi_{t^1}^1 \circ \dots \circ \phi_{t^k}^k)(p).$$

Then for each  $1 \leq i \leq k$ , and for each  $f \in C^\infty(M)$ ,

$$\alpha_* \left( \frac{\partial}{\partial t^i} \Big|_p \right) (f) = \lim_{s \rightarrow 0} \frac{f(\phi_s^i(p)) - f(p)}{s} = X_p^i(f).$$

Thus

$$\alpha_* \left( \frac{\partial}{\partial t^i} \Big|_p \right) = X_p^i.$$

In particular, we conclude  $\alpha$  has full rank in a neighbourhood of 0, and so there exists a coordinate system  $(x, U)$  about  $p \in M$  such that  $(x^i \circ \alpha)(t) = t^i$  for  $1 \leq i \leq k$ , and  $(x^i \circ \alpha)(t) = 0$  for  $i > k$ . But this means that for  $1 \leq i \leq k$ ,

$$X_p^i = \alpha_* \left( \frac{\partial}{\partial t^i} \Big|_p \right) = \frac{\partial}{\partial x^i} \Big|_p.$$

Thus we have found the required coordinate system.  $\square$

Thus the Lie bracket measures the amount that two vector fields can ‘fail’ to commute. It is much more technical, but we can also obtain a more quantitative version of this statement. Given two vector fields  $X$  and  $Y$  generating flows  $\phi$  and  $\psi$  respectively, and given  $p \in M$ , we can consider the curve

$$c(t) = (\psi_{-t} \circ \phi_{-t} \circ \psi_t \circ \phi_t)(p),$$

which travels on the integral curve of  $X$  for time  $t$ , then the integral curve of  $Y$  for time  $t$ , then backwards along the integral curve of  $X$  for time  $t$ , then backwards along the integral curve of  $Y$  for time  $t$ . If  $\phi$  and  $\psi$  commute, then  $c$  is constant. In general, all flows commute ‘up to first order’.

**Lemma 5.8.**  $c'(0) = 0$ .

*Proof.* Let

$$\alpha(x, y, z, w) = (\psi_x \circ \phi_y \circ \psi_z \circ \phi_w)(p).$$

Then the chain rule implies that

$$c'(0) = \alpha_* \left( \frac{\partial}{\partial w} \Big|_0 + \frac{\partial}{\partial z} \Big|_0 - \frac{\partial}{\partial y} \Big|_0 - \frac{\partial}{\partial x} \Big|_0 \right).$$

But

$$\alpha^* \left( \frac{\partial}{\partial w} \Big|_0 \right) (f) = \frac{\partial(f \circ \alpha)}{\partial w} \Big|_0 = \lim_{s \rightarrow 0} \frac{f(\phi_s(p)) - f(p)}{s} = X_p(f).$$

$$\alpha^* \left( \frac{\partial}{\partial z} \Big|_0 \right) (f) = \frac{\partial(f \circ \alpha)}{\partial z} \Big|_0 = \lim_{s \rightarrow 0} \frac{f(\psi_s(p)) - f(p)}{s} = Y_p(f).$$

$$\alpha^* \left( \frac{\partial}{\partial y} \Big|_0 \right) (f) = \frac{\partial(f \circ \alpha)}{\partial y} \Big|_0 = \lim_{s \rightarrow 0} \frac{f(\psi_s(p)) - f(p)}{s} = Y_p(f).$$

$$\alpha^* \left( \frac{\partial}{\partial x} \Big|_0 \right) (f) = \frac{\partial(f \circ \alpha)}{\partial x} \Big|_0 = \lim_{s \rightarrow 0} \frac{f(\phi_s(p)) - f(p)}{s} = X_p(f).$$

Thus

$$c'(0) = X_p + Y_p - X_p - Y_p = 0. \quad \square$$

Thus the non-commutativity of  $c$  must be encoded in the higher derivatives. *Normally*, for a curve  $c$  on a manifold, the second derivative  $c''(0)$  is difficult to interpret in a coordinate independent manner, required the theory of jets. However, in the special case where  $c'(0) = 0$ , we can interpret  $c''(0)$  as a tangent vector. More generally, given a smooth function  $f : M \rightarrow N$ , and a point  $p \in M$  which is critical for  $f$ , we can define a bilinear map  $f_{**} : M_p \times M_p \rightarrow N_{f(p)}$ , such that if  $X_p \in M_p$  and  $Y_p \in M_p$ , then  $Z_{f(p)} = f_{**}(X_p, Y_p)$  is defined such that

$$Z_{f(p)}\{g\} = X_p(Y\{g \circ f\}),$$

where  $Y$  is *any* extension of  $Y_p$ . Using the fact that  $f_*$  vanishes at  $p$ , one can show this formula defines a derivation, and is independent of the extension of  $Y$ , i.e. in coordinates, if  $X_p = \sum a^i \partial_{x_i}$  and  $Y_p = \sum b^j \partial_{x_j}$ , we have

$$f_{**}(X_p, Y_p) = \sum a^i b^j \frac{\partial^2 f_l}{\partial x^j \partial x^k} \frac{\partial}{\partial y^l}.$$

In the special case of a curve, we can define  $c''(0) = f_{**}(1_0, 1_0)$ , a definition that has the property that  $c''(0)\{g\} = (g \circ c)''(0)$ .

**Theorem 5.9.**  $c''(0) = 2[X, Y]_p$ .

*Proof.* This is a crazy, but simple calculation using the chain rule. To prevent things from getting excessively crazy, we introduce some clever notation. Define

$$\alpha_1(t, h) = \psi_t(\phi_h(p)),$$

$$\alpha_2(t, h) = \phi_{-t}(\psi_h(\phi_h(p))),$$

and

$$\alpha_3(t, h) = \psi_{-t}(\phi_{-h}(\psi_h(\phi_h(p)))).$$

Then for any  $f \in C^\infty(M)$ ,

$$\frac{\partial(f \circ \alpha_1)}{\partial t} = Yf \circ \alpha_1,$$

$$\frac{\partial(f \circ \alpha_2)}{\partial t} = -Xf \circ \alpha_2,$$

and

$$\frac{\partial(f \circ \alpha_3)}{\partial t} = -Yf \circ \alpha_3.$$

By the chain rule, we find

$$c''(0) = \frac{\partial^2(f \circ \alpha_3)}{\partial t^2} \Big|_0 + 2 \frac{\partial^2(f \circ \alpha_3)}{\partial t \partial h} \Big|_0 + \frac{\partial^2(f \circ \alpha_3)}{\partial h^2} \Big|_0.$$

We calculate easily that

$$\frac{\partial^2(f \circ \alpha_3)}{\partial t^2} \Big|_0 = - \frac{\partial(Yf \circ \alpha_3)}{\partial t} \Big|_0 = (Y^2 f \circ \alpha_3)(0, 0) = (Y^2 f)(p).$$

Next, we calculate by repeated applications of the chain rule that

$$\begin{aligned} \frac{\partial^2(f \circ \alpha_3)}{\partial t \partial h} \Big|_0 &= \frac{\partial(-Yf \circ \alpha_3)}{\partial h} \Big|_0 \\ &= \frac{\partial(-Yf \circ \alpha_2)}{\partial t} \Big|_0 + \frac{\partial(-Yf \circ \alpha_2)}{\partial h} \Big|_0 \\ &= (XYf \circ \alpha_2)(0, 0) + \frac{\partial(-Yf \circ \alpha_1)}{\partial t} \Big|_0 + \frac{\partial(-Yf \circ \alpha_1)}{\partial h} \Big|_0 \\ &= (XYf)(p) - (Y^2 f \circ \alpha_1)(0, 0) - (XYf)(p) \\ &= (XY - Y^2 - XY)(f)(p). \end{aligned}$$

Some more liberal applications of the chain rule show

$$\frac{\partial^2(f \circ \alpha_3)}{\partial h^2} \Big|_0 = \frac{\partial^2(f \circ \alpha_2)}{\partial t^2} \Big|_0 + 2 \frac{\partial^2(f \circ \alpha_2)}{\partial t \partial h} \Big|_0 + \frac{\partial^2(f \circ \alpha_2)}{\partial h^2} \Big|_0.$$

We then calculate

$$\frac{\partial^2(f \circ \alpha_2)}{\partial t^2} \Big|_0 = (X^2 f \circ \alpha_2)(0, 0) = (X^2 f)(p)$$

$$\begin{aligned}
\left. \frac{\partial^2(f \circ \alpha_2)}{\partial t \partial h} \right|_0 &= \left. \frac{\partial(-Xf \circ \alpha_2)}{\partial h} \right|_0 \\
&= \left. \frac{\partial(-Xf \circ \alpha_1)}{\partial t} \right|_0 + \left. \frac{\partial(-Xf \circ \alpha_1)}{\partial h} \right|_0 \\
&= (-YXf \circ \alpha_1)(0,0) - (X^2f)(p) = (-YX - X^2)(f)(p),
\end{aligned}$$

and

$$\begin{aligned}
\left. \frac{\partial^2(f \circ \alpha_2)}{\partial h^2} \right|_0 &= \left. \frac{\partial^2(f \circ \alpha_1)}{\partial t^2} \right|_0 + 2 \left. \frac{\partial^2(f \circ \alpha_1)}{\partial t \partial h} \right|_0 + \left. \frac{\partial^2(f \circ \alpha_1)}{\partial h^2} \right|_0 \\
&= (Y^2f)(p) + 2 \left. \frac{\partial(Yf \circ \alpha_1)}{\partial h} \right|_0 + (X^2f)(p) \\
&= (Y^2 + 2XY + X^2)(f)(p).
\end{aligned}$$

Putting these together, we conclude

$$\begin{aligned}
\left. \frac{\partial^2(f \circ \alpha_3)}{\partial h^2} \right|_0 &= (X^2f)(p) - 2(YX + X^2)(f)(p) + (Y^2 + 2XY + X^2)(f)(p) \\
&= (Y^2 + 2[X, Y])(f)(p).
\end{aligned}$$

In conclusion, this means that

$$\begin{aligned}
c''(0) &= (Y^2f)(p) + 2(XY - Y^2 - XY)(f)(p) + (Y^2 + 2[X, Y])(f)(p) \\
&= 2[X, Y](f)(p) = 2[X, Y]_p(f).
\end{aligned}$$

This completes the calculation. □

## 5.1 Parameter Curves

If  $X_1, \dots, X_k$  are vectors fields with  $[X_i, X_j]_p \neq 0$  for some  $i, j$ , it is not possible to find a coordinate system  $(x, U)$  with  $p \in U$  such that on  $U$ ,

$$X_i = \frac{\partial}{\partial x^i}.$$

On the other hand, it may be possible to find functions  $a_1, \dots, a_k$  such that

$$[X_i/a_i, X_j/a_j] = 0,$$

in which case we can find a coordinate system  $(x, U)$  such that for each  $i$ , on  $U$  we have

$$X_i = a_i \frac{\partial}{\partial x_i}.$$

The integral curves of  $X_i$  thus lie along the coordinate lines of the system  $(x, U)$ . In two dimensions, this is always possible.

**Theorem 5.10.** *If  $X, Y$  are linearly independent vector fields on a 2-manifold, then around any point  $p$  we can find coordinates  $(x, U)$  such that  $X = a\partial_{x^1}$  and  $Y = b\partial_{x^2}$  for  $a, b \in C^\infty(U)$ .*

*Proof.* Without loss of generality, we may assume we are working on an open subset  $U$  of  $\mathbb{R}^2$ , that  $p = 0$ , that  $X = \partial/\partial x^1$ , and that  $Y_0 = \partial/\partial x^2$ . If  $\phi$  is the flow generated by  $Y$ , then there is a smaller open neighbourhood  $V \subset U$  such that for each  $x \in U$ , there is a unique  $a(x)$  and  $t(x)$  such that  $\phi_{t(x)}(a(x), 0) = x$ . The map  $x \mapsto (a(x), x^2) = (y^1, y^2)$  is  $C^\infty$ , and we verify that its Jacobian at  $x = 0$  is the identity, so the map gives a coordinate system  $y$  in a neighbourhood of the identity. If we fix  $y^2$ , and vary  $y^1$ , we travel parallel to the  $x$  axis, hence along the integral curves of  $X$ . Conversely, if we fix  $y^1$ , and vary  $y^2$ , then we travel along the integral curves of  $Y$ .  $\square$

*Remark.* If  $(x, U)$  gives parameter curves for two vector fields  $X$  and  $Y$ , then so too does  $(y, U)$ , where  $y = (\alpha, \beta) \circ x$  for any diffeomorphisms  $\alpha$  and  $\beta$  of  $\mathbb{R}$ . In particular, for any smooth curve in  $M$  with  $c(0) = p$  and  $c'(t)$  never a multiple of  $X$  or  $Y$ , we can find a coordinate system  $(x, U)$  giving parameter curves for  $X$  and  $Y$  with  $x(c(t)) = (t, t)$ .

On the other hand, there exists vector fields  $X, Y$ , and  $Z$  in  $\mathbb{R}^3$  such that no coordinate system can locally parameterize integral curves of  $X, Y$ , and  $Z$ . So two dimensions is the only case where we are guaranteed a successful theory.

**Example.** Let  $Y(x, y, z) = (0, 1, 0)$ ,  $Z(x, y, z) = (0, 0, 1)$ , and let  $X(x, y, z) = (1, 0, yx)$ . Then the integral curves of  $Y$  are lines parallel to the  $Y$  axis, the integral curves of  $Z$  are lines parallel to the  $Z$  axis, and the integral curves of  $X$  are the  $x$  axis, and parabolas in planes parallel to the  $xz$  plane, of varying eccentricity in  $y$ . If  $(x, U)$  is a coordinate chart around the origin, such that as  $x^1$  varies we travel along the integrals of  $X$ , as  $x^2$  varies we travel along the integral curves of  $Y$ , and as  $x^3$  varies we travel along the integral curves of  $Z$ . If we assume without loss of generality that  $x(0) = 0$ , then on one hand we conclude that  $x^{-1}(a, b, 0)$  lies in the  $xy$  plane, since as we vary the  $x^1$  coordinate from  $(0, 0, 0)$  to  $(a, 0, 0)$ , we travel along the  $x$ -axis, and then as we vary  $x^2$  from  $(a, 0, 0)$  to  $(a, b, 0)$ , we travel on a line parallel to the  $y$  axis. On the other hand, if we vary  $x^1$  first from  $(0, 0, 0)$  to  $(0, b, 0)$ , then we travel on the  $y$ -axis, and then we travel along a parabola as we vary  $x^1$  from  $(0, b, 0)$  to  $(a, b, 0)$ .

## 5.2 Distributions and Integrability Conditions

There is one way to switch our point of view from vector fields so that global existence results may be obtained. Consider a one-dimensional subbundle  $\Delta$  of  $TM$ . Thus for each point  $p \in M$ , we associate a one-dimensional subspace  $\Delta_p \subset T_p M$  such that around each point  $p \in M$ , we can find a neighbourhood  $U$  and a smooth vector field  $X \in \Gamma(TU)$  such that  $\Delta_q$  is spanned by  $X_q$  for each  $q \in U$ . Our goal is now to find a one-dimensional immersed submanifold  $i : N \rightarrow M$  such that  $i_*(T_p N) = \Delta_{i(p)}$  for each  $p \in N$ , known as an integral curve. The existence and uniqueness theory of differential equations gives the local existence of the

manifold  $N$ , but the advantage of this approach is that we can join overlapping solutions. Thus  $M$  can be uniquely written as the disjoint union of connected manifolds  $N$  forming integral curves. The function  $\Delta$  is known as a *one dimensional distribution*, and  $N$  is called a *foliation* of  $\Delta$ . Our goal in this section is to extend this result to higher dimensional settings.

A smooth  $m$ -dimensional subbundle  $\Delta$  of  $TM$  is known as an  *$m$  dimensional distribution* (these have nothing at all to do with the distributions of modern analysis). A useful fact for our purposes are that for each  $p \in M$ , we can find a neighbourhood  $U$  of  $p$  and  $m$  vector fields  $X_1, \dots, X_m \in \Gamma(TU)$  which span  $\Delta$  at each point. Our goal, given an  $m$  dimensional distribution  $\Delta$ , is to find immersed  $m$  dimensional submanifolds  $i : N \rightarrow M$  such that  $i_*(T_p N) = \Delta_{i(p)}$  for each  $p \in N$ . We call  $N$  an *integral manifold* to  $\Delta$ .

In general, the problem of finding integral manifolds is not, in general, possible. The simplest example is given in  $\mathbb{R}^3$  by the distribution  $\Delta$  expressed by the equation  $dz = ydx$ , i.e. such that

$$\Delta_{(x,y,z)} = \{v_{(x,y,z)} : dz(v) = ydx(v)\}.$$

Thus  $\Delta_{(x,y,z)}$  is spanned by  $\partial_y$  and  $\partial_x + y\partial_z$ . If  $N$  was an integral surface to  $\Delta$  at the origin, then we may locally write  $z = f(x, y)$  for some  $C^\infty$  function  $f$ , such that

$$\frac{\partial f}{\partial x} = y \quad \text{and} \quad \frac{\partial f}{\partial y} = 0.$$

The second equation implies  $f$  is a function solely depending on  $x$ , but that cannot possibly be true since the partial derivatives of  $f$  depend on  $y$ . Thus  $f$ , and thus the surface  $N$ , cannot possibly exist.

To see what causes integrability to fail here, let us consider the more general situation where  $\Delta_{(x,y,z)}$  is specified by the equation  $dz = f(x, y)dx + g(x, y)dy$ , for  $C^\infty$  functions  $f$  and  $g$ , i.e. such that

$$\Delta_{(x,y,z)} = \{v_{(x,y,z)} : dz(v) = f(x, y)dx(v) + g(x, y)dy(v)\}.$$

As in the previous case, we can locally write  $z = u(x, y)$  for some  $C^\infty$  function  $u$ , and the constraints of this function are that  $u_x(x, y) = f(x, y)$  and  $u_y(x, y) = g(x, y)$ . Since mixed partials of smooth functions are equal, we conclude

$$f_y = u_{xy} = u_{yx} = g_x.$$

It is well known that if  $f$  and  $g$  are such functions for which this equation holds, then such a function  $u$  can be found, and thus an immersed submanifold exists.

**Lemma 5.11.** *If  $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$  are smooth functions such that, in a neighbourhood of 0,  $f_y = g_x$ . Then for any  $z_0 \in \mathbb{R}$ , there exists a function  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  defined in a neighbourhood of zero such that  $u_x = f$ ,  $u_y = g$ , and  $u(0, 0) = z_0$ .*

*Proof.* Let  $u(0, 0) = z_0$ . Then define  $u(x, 0) = \int_0^x f(t, 0) dt$ , and finally define

$$u(x, y) = u(x, 0) + \int_0^y g(x, t) dt = \int_0^x f(t, 0) dt + \int_0^y g(x, t) dt.$$



Then

$$\begin{aligned} u_x(x, y) &= f(x, 0) + \int_0^y g_x(x, t) dt \\ &= f(x, 0) + \int_0^y f_y(x, t) dt = f(x, 0) + (f(x, y) - f(x, 0)) = f(x, y). \end{aligned}$$

and  $u_y(x, y) = g(x, y)$  by the fundamental theorem of calculus.  $\square$

Let us try and come up with a necessary and sufficient condition for integral curves to the more general distribution  $\Delta$  given by the equation

$$dz = f(x, y, z)dx + g(x, y, z)dy.$$

Our goal is to write  $z = u(x, y)$ , where  $u$  is a function such that  $u_x(x, y) = f(x, y, u(x, y))$  and  $u_y(x, y) = g(x, y, u(x, y))$ . Trying to apply equality of mixed derivatives, we calculate that

$$u_{xy}(x, y) = f_y(x, y, u) + f_z(x, y, u)u_y(x, y) = f_y(x, y, u) + f_z(x, y, u)g(x, y, u)$$

and

$$u_{yx}(x, y) = g_x(x, y, u) + g_z(x, y, u)u_x(x, y, u) = g_x(x, y, u) + g_z(x, y, u)f(x, y, u).$$

But this implies that  $f_y + f_z g = g_x + g_z f$ . This equation turns out to be sufficient to obtain the existence of  $u$ , but we need not prove this specific case, because we can treat a much more general situation.

We consider  $C^\infty$  functions  $f_i : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  for  $i \in \{1, \dots, m\}$ , where we use  $t$  to denote points in  $\mathbb{R}^m$  and  $x$  to denote points in  $\mathbb{R}^n$ . We then fix  $x_0 \in \mathbb{R}^n$ , and try and find functions  $u : \mathbb{R}^m \rightarrow \mathbb{R}^n$  such that  $u(0, 0) = x_0$ , and  $u_{t^i}(t) = f_i(t, u(t))$  for each  $i$ . Setting mixed partials in  $t$  equal to one another, we find a necessary condition for this result to hold, and this is also a sufficient condition.

**Theorem 5.12.** *Let  $U$  and  $V$  be open subsets of  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , where  $U$  is a neighbourhood of zero, and let  $f_i : U \times V \rightarrow \mathbb{R}^n$  be  $C^\infty$  functions for  $i \in \{1, \dots, m\}$ . Moreover, suppose that for each  $i, j \in \{1, \dots, m\}$ ,*

$$\frac{\partial f_j}{\partial t^i} - \frac{\partial f_i}{\partial t^j} + \sum_{k=1}^n \frac{\partial f_j}{\partial x^k} f_i^k - \frac{\partial f_i}{\partial x^k} f_j^k = 0.$$

*Then for each  $x_0 \in V$ , there exists a unique function  $u : W \rightarrow V$ , where  $W$  is a neighbourhood of zero, such that*

$$\frac{\partial u}{\partial t^i}(t) = f_i(t, u(t)).$$

*Proof.* Uniqueness will be obvious from our proof of existence. We define  $u(0) = x_0$ . Then we define  $u(t, 0, \dots, 0)$  such that

$$\frac{\partial u}{\partial t^1}(t, 0, \dots, 0) = f_1(t, 0, \dots, 0, u(t, 0, \dots, 0)).$$

Such a definition exists, and is unique, since it is an ordinary differential equation. For a fixed  $t^1$ , we then define  $u(t^1, t, 0, \dots, 0)$  such that

$$\frac{\partial u}{\partial t^2}(t^1, t, 0, \dots, 0) = f_2(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)).$$

Now define

$$g(t) = \frac{\partial u}{\partial t^1}(t^1, t, 0, \dots, 0) - f_1(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)).$$

We calculate that

$$\begin{aligned} g'(t) &= \frac{\partial^2 u}{\partial t^2 \partial t^1}(t^1, t, 0, \dots, 0) \\ &\quad - \frac{\partial f_1}{\partial t^1}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \\ &\quad - \sum_{k=1}^n \frac{\partial f_1}{\partial x^k}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \frac{\partial u^k}{\partial t^2}(t^1, t, 0, \dots, 0) \\ &= \frac{\partial f_2(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0))}{\partial t^1} \\ &\quad - \frac{\partial f_1}{\partial t^1}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \\ &\quad - \sum_{k=1}^n \frac{\partial f_1}{\partial x^k}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \frac{\partial u^k}{\partial t^2}(t^1, t, 0, \dots, 0) \\ &= \frac{\partial f_2}{\partial t^1}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \\ &\quad + \sum_{k=1}^n \frac{\partial f_2}{\partial x^k}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \frac{\partial u^k}{\partial t^1}(t^1, t, 0, \dots, 0) \\ &\quad - \frac{\partial f_1}{\partial t^1}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \\ &\quad - \sum_{k=1}^n \frac{\partial f_1}{\partial x^k}(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)) \frac{\partial u^k}{\partial t^2}(t^1, t, 0, \dots, 0) = 0. \end{aligned}$$

Since  $g(0) = 0$ , this means that

$$\frac{\partial u}{\partial t^1}(t^1, t, 0, \dots, 0) = f_1(t^1, t, 0, \dots, 0, u(t^1, t, 0, \dots, 0)).$$

By a nasty calculation, we can continue defining  $u$  along coordinate lines, using the integrability conditions to verify that  $u$  satisfies the required partial differential equations at each step.  $\square$

**Example.** Let  $f : \mathbb{C}^2 \rightarrow \mathbb{C}$  be a complex analytic function. This means that  $f(z^1, z^2) = u(z^1, z^2) + iv(z^1, z^2)$ , where for  $k \in \{1, 2\}$ , if we set  $z^k = x^k + iy^k$ , then the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x^k} = \frac{\partial v}{\partial y^k} \quad \text{and} \quad \frac{\partial v}{\partial x^k} = -\frac{\partial u}{\partial y^k}$$

hold. We will show the ‘holomorphic differential equation’  $\phi'(z) = f(z, \phi(z))$  is locally solvable by a complex analytic function  $\phi$ . Let  $\phi = \phi^1 + i\phi^2$ . The Cauchy-Riemann equations cannot be expressed as a partial differential equation like the one above, but we can ‘cheat’ the Cauchy-Riemann equation by using the complex differentiable nature of  $f$ . Note that if  $\phi$  solves the system of four PDEs

$$\begin{aligned} \frac{\partial \phi^1}{\partial x} &= u(z, \phi(z)) & \frac{\partial \phi^2}{\partial x} &= v(z, \phi(z)) \\ \frac{\partial \phi^1}{\partial y} &= -v(z, \phi(z)) & \frac{\partial \phi^2}{\partial y} &= u(z, \phi(z)), \end{aligned}$$

then  $\phi$  automatically solves the Cauchy-Riemann equations, and  $\phi'(z) = f(z, \phi(z))$ . Setting mixed derivatives equal here, shows integrability conditions are implied by the Cauchy-Riemann equations for  $f$ , and thus a function  $\phi$  exists locally in a neighbourhood of any point.

A basic fact about many of the basic theorems of differential geometry is that results can be guaranteed if certain ‘integrability conditions’ are satisfied, which essentially amounts to saying appropriate mixed derivatives are equal to one another. But the important point here is that ‘integrability conditions’ can always be expressed in a much more concise, geometric manner, without mentioning partial derivatives at all. So we now prove essentially the same formula, but using a more geometric approach.

If  $f : M \rightarrow N$  is  $C^\infty$ , then for two vector fields  $X \in \Gamma(TM)$ ,  $Y \in \Gamma(TN)$ , we say  $X$  and  $Y$  are  $f$  related if for each  $p \in M$ ,

$$Y(f(p)) = f_*(X(p)).$$

This means precisely that for  $g \in C^\infty(N)$ ,

$$X(g \circ f) = Yg \circ f$$

Given a vector field  $X \in \Gamma(TM)$ , there need not be a vector field  $Y \in \Gamma(TN)$  which is  $f$  related to  $X$ , and vice versa, except in one simple case, where  $f$  is an immersion.

**Lemma 5.13.** Let  $f : M \rightarrow N$  be an immersion, and let  $Y \in \Gamma(TN)$  be a vector field with  $Y(f(p)) \in f_*|_p(T_p M)$  for each  $p \in M$ . Then there exists a unique vector field  $X \in \Gamma(TM)$  which is  $f$ -related to  $Y$ .

*Proof.* Clearly, we must define  $X(p)$  to be the unique vector such that  $f_*|_p(X(p)) = Y_{f(p)}$ . It remains to show that  $X$  is smooth. To do this, we work locally, for each  $p \in M$  finding coordinate systems  $(x, U)$  with  $p \in U$  and  $(y, V)$  with  $U \subset V$  such that

$$(y \circ f \circ x^{-1})(a^1, \dots, a^n) = (a^1, \dots, a^n, 0, \dots, 0).$$

Then there are functions  $\alpha^i \in C^\infty(U)$  such that for each  $q \in U$ ,

$$Y(f(q)) = \sum_{k=1}^n \alpha^k(q) \left. \frac{\partial}{\partial y^k} \right|_{f(q)}.$$

But then it is easy to see that for each  $q \in U$ ,

$$X(q) = \sum_{k=1}^n \alpha^k(q) \left. \frac{\partial}{\partial x^k} \right|_q,$$

so  $X$  is  $C^\infty$ . □

For us, the most important property is the relation to the Lie bracket.

**Theorem 5.14.** *If  $X_1$  is  $f$ -related to  $Y_1$ , and  $X_2$  is  $f$ -related to  $Y_2$ , then  $[X_1, X_2]$  is  $f$ -related to  $[Y_1, Y_2]$ .*

*Proof.* We calculate that for any  $g \in C^\infty(N)$  and any  $p \in M$ ,

$$\begin{aligned} f_*([X_1, X_2]_p)(g) &= [X_1, X_2]_p(g \circ f) \\ &= X_1(p)(X_2\{g \circ f\}) - X_2(p)(X_1\{g \circ f\}) \\ &= X_1(p)((Y_2g) \circ f) - X_2(p)((Y_1g) \circ f) \\ &= \{Y_1(Y_2g) \circ f - (X_2(X_1g)) \circ f\}_p \\ &= \{[Y_1, Y_2]g\}_{f(p)}. \end{aligned}$$

Since  $g$  was arbitrary, we conclude that  $f_*[X^1, X^2]_p = [Y^1, Y^2]_{f(p)}$ . □

Given a distribution  $\Delta$  on  $M$ , we say a vector field  $X \in \Gamma(TM)$  belongs to  $\Delta$  if  $X_p \in \Delta_p$  for each  $p \in M$ . Suppose an integral manifold  $N$  of  $M$  exists, given by an immersion  $i : N \rightarrow M$ . If  $Y_1$  and  $Y_2$  are two vector fields which belong to  $\Delta$ , then we have shown there exist vector fields  $X_1$  and  $X_2$  on  $N$  which are  $f$  related to  $Y_1$  and  $Y_2$ . Then  $[X_1, X_2]$  is  $f$ -related to  $[Y_1, Y_2]$ . But this means that for each  $p \in N$ ,  $f_*([X_1, X_2]_p) = [Y_1, Y_2]_{f(p)}$  is an element of  $\Delta_{f(p)}$ . If there is an integral manifold through each point on  $M$ , this implies that  $[Y_1, Y_2]_q \in \Delta_q$  for each  $q \in M$ . Thus we conclude that if  $Y_1$  and  $Y_2$  belong to  $\Delta$ , then  $[Y_1, Y_2]$  belongs to  $\Delta$ . We say that a distribution  $\Delta$  is *integrable*, or *involutive* if the set of vector fields belonging to  $\Delta$  is closed under the Lie bracket. We have proved this is a necessary condition for integral manifolds to exist, and we will soon show this is sufficient.

**Lemma 5.15.** *If  $X_1, \dots, X_m$  are vector fields spanning a distribution  $\Delta$  in a neighbourhood  $U$ , then  $\Delta$  is integrable on  $U$  if and only if there exists smooth functions  $a_{ij}^k$  such that*

$$[X_i, X_j] = \sum a_{ij}^k X_k.$$

*Proof.* Such functions clearly exist if  $\Delta$  is integrable. Conversely, if  $Y_1$  and  $Y_2$  are vector fields belonging to  $\Delta$ , then we can find smooth functions  $\{\alpha_k\}$  and  $\{\beta_k\}$  such that  $Y_1 = \sum_{k=1}^m \alpha_k X_k$  and  $Y_2 = \sum_{k=1}^m \beta_k X_k$ . Since

$$[\alpha_i X_i, \beta_j X_j] = \alpha_i \beta_j [X^i, X^j] + \alpha_i X^i(\beta_j) X^j - \beta_j X^j(\alpha_i) X^i,$$

we can certainly write  $[Y_1, Y_2]$  as a linear combination of  $X_1, \dots, X_m$  with  $C^\infty(M)$  coefficients.  $\square$

*Remark.* If, on  $\mathbb{R}^3$ ,  $X = \partial_x + f\partial_z$ , and  $Y = \partial_y + g\partial_z$ , then

$$[X, Y] = (g_x - f_y + fg_z - gf_z) \partial_z.$$

If  $\Delta_{(x,y,z)}$  is spanned by  $\partial_x + f\partial_z$  and  $\partial_y + g\partial_z$  at each point, then  $\Delta$  is integrable if and only if  $g_x - f_y + fg_z - gf_z = 0$ , which is the integrability condition we saw before. The next theorem thus addresses the PDE solution we gave before in a geometric setting.

We now address the ‘solutions to PDEs’ approach we took before in a geometric setting. Our proof, however, looks quite different.

**Theorem 5.16** (Frobenius Integrability Theorem). *If  $\Delta$  is an integrable,  $m$  dimensional distribution on a manifold  $M$ , and  $p \in M$ , then there is a coordinate system  $(x, U)$  with  $p \in U$  such that for each  $a^{m+1}, \dots, a^n$ , the set*

$$\{q \in U : x^{m+1}(q) = a^{m+1}, \dots, x^n(q) = a^n\}$$

*is an integral manifold of  $\Delta$ .*

*Proof.* Without loss of generality, we may assume  $M = \mathbb{R}^n$ ,  $p = 0$ , and that  $\Delta_0$  is spanned by  $\partial_1|_0, \dots, \partial_m|_0$ . Let  $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and define

$$\pi(a) = (a^1, \dots, a^m).$$

Then  $\pi_*$  is locally injective when restricted as a map from  $\Delta$  to  $T\mathbb{R}^m$ . So in a neighbourhood of the origin in  $M$ , for each  $q$ , we can find smooth vector fields  $X^1(q), \dots, X^m(q) \in \Delta_q$  such that for each  $i \in \{1, \dots, m\}$ ,

$$\pi_*(X^i(q)) = \partial_i|_{\pi(q)}.$$

Thus the vector fields  $X^i$  and  $\partial_i$  are  $\pi$ -related in a neighbourhood of the origin. But this means that for each  $i, j \in \{1, \dots, m\}$ ,  $[X^i, X^j]$  are  $\pi$ -related to  $[\partial_i, \partial_j] = 0$ . Thus  $\pi_*([X^i, X^j]_p) = 0$  for each  $p$  in a neighbourhood of the origin. But since  $\pi_*$  is injective on  $\Delta_p$ , this implies

$[X^i, X^j]_p = 0$ . But this means that  $[X^i, X^j] = 0$ , so in particular, we find a coordinate system  $(x, U)$  such that for each  $i \in \{1, \dots, m\}$ ,

$$X^i = \frac{\partial}{\partial x^i}.$$

But this means that integral manifolds are obtained by fixing the coordinates  $x^{m+1}, \dots, x^n$ , and varying  $x^1, \dots, x^m$ . If  $N$  is any connected integral submanifold obtained by an immersion  $i : N \rightarrow \mathbb{R}^n$  and with  $i(p) = 0$ , then for each  $1 \leq k \leq n - m$ , and for each tangent vector  $X_q$  of  $N$ ,

$$d(x^{m+k} \circ i)(X_q) = i_*(X_q)(x^{m+k}) = 0,$$

since  $i_*(X_q)$  is an element of  $\Delta_{i(q)}$ . So  $x^{m+k}$  is constant on  $i^{-1}(I) \cap N$ , which gives the required uniqueness result.  $\square$

As we mentioned, the advantage of distribution theory is it enables us to state simply global results about differential equations on manifolds. We say a manifold  $N$  is an  $m$  dimensional *foliation* of  $M$  if it is immersed in  $M$  by a map  $i : N \rightarrow M$  which is injective, surjective, and such that around each  $p \in M$ , there exists a coordinate system  $(x, U)$  with  $p \in U$  and such that the components of  $N \cap i^{-1}(U)$  are given by

$$\{q \in N : x^{m+1}(i(q)) = a^{m+1}, \dots, x^n(i(q)) = a^n\},$$

for fixed coefficients  $a^{m+1}, \dots, a^n$ . Each component of  $N$  is known as a *leaf*, or *folium* of the foliation.

**Theorem 5.17.** *Let  $\Delta$  be a smooth integrable distribution on  $M$ . Then  $M$  is foliated by an integral manifold  $N$  of  $\Delta$ , each component of  $N$  being called a maximal integral manifold of  $\Delta$ . If  $M$  is metrizable, then  $N$  is metrizable.*

*Proof.* We can cover  $M$  by a family of coordinate systems  $(x_i, U_i)$  such that for each of the slices

$$S_i(a) = \{p \in U_i : x_i^{m+1}(p) = a^{m+1}, \dots, x_i^n(p) = a^n\}$$

for  $a = (a^{m+1}, \dots, a^n)$ , are integral manifolds of  $\Delta$ . We let  $N = M$ , but with a different topological and smooth structure which we will now define. We let  $\{S_i(a)\}$  be a cover of  $N$ , and for each  $S_i(a)$ , we associate the chart  $(x_{i,a}, U \cap S_i(a))$  where  $x_{i,a}(p) = (x_i^1(p), \dots, x_i^m(p))$ . If a slice  $S_i(a)$  intersects a slice  $S_j(b)$ , then  $x_{i,a} \circ x_{j,b}^{-1}$  is certainly smooth, so this gives  $N$  a topological and smooth manifold structure. It is clear the set theoretic inclusion  $i : N \rightarrow M$  gives a foliation of  $M$  by  $N$ . It now suffices to show that if  $M$  is metrizable, then  $N$  is metrizable. We may assume  $M$  is connected, and thus that the family of coordinate systems  $(x_i, U_i)$  is countable. For each  $i, j$ , and  $a \in \mathbb{R}^m$ ,  $S_i(a) \cap U_j$  has at most countably many components in  $N$ , and each component is contained in a slice of  $U_j$ , so  $S_i(a) \cap U_j$  is contained in countably many slices of  $U_j$ . We say two slices  $S_i(a)$  and  $S_j(b)$  if there is a sequence  $i_0, \dots, i_n$  and  $a_0, \dots, a_n$  with  $i_0 = i$ ,  $i_n = j$ ,  $a_0 = a$ , and  $a_n = b$ , such that  $S_{i_k}(a_k) \cap S_{i_{k+1}}(a_{k+1}) \neq \emptyset$ . The union of slices joined to a particular slice forms a component of  $N$ , and countably many slices are joined to one another, hence the union is second countable. Thus  $N$  is metrizable.  $\square$

If  $(x, U)$  is a coordinate chart considered in the proof, then two slices of the chart may belong to the same leaf of the foliation. However, only at most countably many slices can belong to the same leaf, for otherwise the manifold  $N$  will not be metrizable. Here is a simple consequence.

**Theorem 5.18.** *Let  $M$  be a  $C^\infty$  manifold, and let  $M_1$  be a leaf of the foliation  $i : N \rightarrow M$  determined by some distribution  $\Delta$ . If  $f : L \rightarrow M$  is smooth and  $f(L) \subset i(M_1)$ , then  $f$  is smooth considered as a map into  $M_1$ .*

*Proof.* It suffices to show that  $f$  is continuous as a map into  $M_1$ . Given  $p \in L$ , choose a coordinate system  $(x, U)$  around  $f(p)$  such that the slices

$$\{q \in U : x^{m+1}(q) = a^{m+1}, \dots, x^n(q) = a^n\}.$$

Since  $f$  is continuous as a map into  $M$ , there exists a coordinate system  $(y, V)$  around  $p$  such that  $f(V) \subset U$ . Without loss of generality, we may assume that  $V$  is connected. Then  $f(V)$  is connected. If  $f(V)$  was not contained in a single slice of the chart  $(x, U)$ , then it would pass over uncountably many slices, which is impossible since  $f(V) \subset M_1$ , and  $M_1$  intersects at most countably many slices. Thus  $V$  is contained in a single slice, and it is therefore clear that  $f$  is continuous restricted as a map from  $V$  to  $M_1$ , and thus continuous as a map into  $M_1$  at  $p$ . Since  $p$  is arbitrary, this shows  $f$  is continuous.  $\square$

We can also restate the Frobenius integrability theorem in a ‘dual form’ in terms of differential forms. Given a distribution  $\Delta$  of rank  $m$ , we can consider, for each  $k$ , the subspace  $I^k(\Delta)$  of  $\Omega^k(M)$  consisting of  $k$ -forms  $\omega$  such that for any  $k$  vector fields  $X_1, \dots, X_k$  belonging to  $\Delta$ ,  $\omega(X_1, \dots, X_k) = 0$ . This are simply sections of the sub-bundle of  $\Omega^k(TM)$  consisting of  $k$ -forms vanishing on sections of  $\Delta$ . We then let  $I(\Delta) = I^0(\Delta) + \dots + I^n(\Delta) \subset \Omega(M)$ . Then  $I(\Delta)$  is actually an ideal of  $\Omega(M)$ , because if  $\omega \in I^i(\Delta)$  and  $\eta \in \Omega^j(M)$ , then  $\omega \wedge \eta \in I^{i+j}(\Delta)$ .

If  $\Delta$  is an  $m$  dimensional distribution, then for each  $p \in M$ , there is a chart  $(x, U)$  containing  $p$  such that, restricted to  $U$ , the ideal  $I(\Delta)$  is locally generated by  $n - m$  one-forms  $\omega^{m+1}, \dots, \omega^n$ . To see why, we note that for a suitably small neighbourhood  $U$  there are  $m$  vector fields  $X_1, \dots, X_m$  which span the distribution  $\Delta$  on  $U$ . We may also choose a coordinate chart  $(x, U)$  such that for each  $i \in \{1, \dots, m\}$ ,

$$X_i = \frac{\partial}{\partial x^i} \Big|_p.$$

It then follows that

$$(dx^1(p) \wedge \dots \wedge dx^m(p))(X_1|_p, \dots, X_m|_p) = 1.$$

By continuity,  $(dx^1 \wedge \dots \wedge dx^m)(X_1, \dots, X_m)$  is non-zero in a neighbourhood of  $p$ , which, by thinning  $U$  if necessary, we may assume to be all of  $U$ . It therefore follows that  $dx^1(q), \dots, dx^m(q)$  are linearly independent linear functionals on  $\Delta_q$  for all  $q \in U$ . Since  $\Delta_q$  is  $m$  dimensional, this implies

$$dx^1(q), \dots, dx^m(q)$$

spans  $\Delta_q^*$ , and, in particular,  $dx^1, \dots, dx^m$  span the  $C^\infty(U)$  module  $\Gamma(\Delta|_U^*)$ . In particular there exist smooth functions  $a_{ij} \in C^\infty(U)$  for each  $i \in \{1, \dots, m\}$  and  $j \in \{m+1, \dots, n\}$  such that

$$dx^j = \sum_{i=1}^m a_{ij} dx^i.$$

If we define, for each  $j \in \{m+1, \dots, n\}$ , the 1-form

$$\omega_j = dx^j - \sum_{i=1}^m a_{ij} dx^i,$$

then  $\omega_j(X) = 0$  for each  $X$  belonging to  $\Delta$ . Then  $dx^1, \dots, dx^m, \omega_{m+1}, \dots, \omega_{m+n}$  form a basis for  $\Gamma(T^*U)$ , and so if  $\omega$  is any 1-form which vanishes on all vector fields belonging to  $\Delta$ , then we can write it as a  $C^\infty(U)$  linear combination of  $\omega_{m+1}, \dots, \omega_n$ . The space of all alternating  $k$  tensors on  $\Delta_p$  has dimension

$$\frac{m!}{k!(m-k)!}.$$

Similarly, the space of all alternating  $k$  tensors on  $T_p M$  has dimension

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

The space of  $k$  forms in the ideal generated by  $\omega_{m+1}, \dots, \omega_{m+n}$  has dimension

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

But this means that any alternating  $k$  tensor on  $T_p M$  which vanishes when restricted to a tensor on  $\Delta_p$  is an element of the ideal generated by the forms  $\omega_{m+1}, \dots, \omega_n$ , so we conclude  $I(\Delta|_U) = (\omega_{m+1}, \dots, \omega_n)$ .

*Remark.* The last argument also established the fact that, as an ideal of  $\Gamma(TM)$ ,  $I(\Delta)$  is locally generated by the elements of  $I^1(\Delta)$ . A partition of unity argument then shows that  $I(\Delta)$  is generated by  $I^1(\Delta)$ .

To prove the following restatement of the Frobenius integrability theorem, we recall that for any  $k$  form  $\omega$ , the exterior derivative  $d\omega$  satisfies

$$\begin{aligned} d\omega(X_1, \dots, X_{k+1}) &= \sum_{i=1}^{k+1} (-1)^{i+1} X_i(\omega(X_1, \dots, \hat{X}_i, \dots, X_{k+1})) \\ &\quad + \sum_{1 \leq i < j \leq k} (-1)^{i+j} \omega([X_i, X_j], X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_{k+1}). \end{aligned}$$

The exterior derivative extends to a linear map on  $\Gamma(\Omega(TM))$  by linearity.

**Theorem 5.19** (The Frobenius Integrability Theorem). *A distribution  $\Delta$  on a manifold  $M$  is integrable if and only if  $d(I(\Delta)) \subset I(\Delta)$ .*



*Proof.* Suppose  $\Delta$  is integrable. If  $\omega$  is a  $k$  form vanishing on  $\Delta$ , then for any  $k + 1$  vector fields  $X_1, \dots, X_{k+1} \in \Gamma(\Delta)$ , we have  $[X_i, X_j] \in \Gamma(\Delta)$ , and thus  $d\omega(X_1, \dots, X_{k+1}) = 0$  by the formula above. On the other hand, if  $d(I(\Delta)) \subset I(\Delta)$ , then in particular, for any one form  $\omega \in I(\Delta)$ , and any  $X, Y \in \Gamma(\Delta)$ ,

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]) = -\omega([X, Y]) = 0.$$

Thus  $\omega([X, Y]) = 0$ . But since  $\omega$  was arbitrary, this means  $[X, Y] \in \Gamma(\Delta)$ , which shows  $\Delta$  is integrable.  $\square$

*Remark.* The condition  $d(I(\Delta)) \subset I(\Delta)$  can be verified by checking that  $d(I^1(\Delta)) \subset I^2(\Delta)$ , since  $I(\Delta)$  is generated by the one-forms it contains.

TODO: Interpret meaning of next corollary (something symplectic?).

**Corollary 5.20.** *Let  $\omega^{m+1}, \dots, \omega^n$  be linearly independant one-forms. Then there exists a neighbourhood  $U$  around each point such that there are smooth functions  $f_{m+1}^\alpha, \dots, f_n^\alpha, g^{m+1}, \dots, g^n$  for each  $\alpha \in \{m+1, \dots, n\}$  such that for each  $\alpha$ , on  $U$ ,*

$$\omega^\alpha = f_{m+1}^\alpha dg^{m+1} + \dots + f_n^\alpha dg^n,$$

*if and only if there exists 1-forms  $\theta_\beta^\alpha$  for  $\alpha, \beta \in \{m+1, \dots, n\}$  such that for each  $\alpha$ ,*

$$d\omega^\alpha = \theta_{m+1}^\alpha \wedge \omega^{m+1} + \dots + \theta_n^\alpha \wedge \omega^n.$$

*Proof.* The one-forms  $\omega^{m+1}, \dots, \omega^n$  generate a  $m$  dimensional distribution  $\Delta$  by considering the intersection of the kernels of the one-forms at each point. Then  $I(\Delta)$  is generated by  $\omega^{m+1}, \dots, \omega^n$ . If we can write

$$\omega^\alpha = f_{m+1}^\alpha dg^{m+1} + \dots + f_n^\alpha dg^n,$$

then  $dg^{m+1}, \dots, dg^n$  are linearly independant, hence we may write  $dg^\alpha = \sum h_\beta^\alpha \omega^\beta$  for some smooth functions  $h_\beta^\alpha$ . It follows that for each  $\alpha$ ,

$$d\omega^\alpha = \sum_{i=m+1}^n df_i^\alpha \wedge dg^i = \sum_{j=m+1}^n \sum_{i=m+1}^n h_j^i df_i^\alpha \wedge \omega^j.$$

Conversely, if

$$d\omega^\alpha = \sum_{i=m+1}^n \theta_i^\alpha \wedge \omega^i,$$

then  $d(I(\Delta)) \subset I(\Delta)$ . It follows from the Frobenius theorem that  $\Delta$  is integrable, so around each point  $p$ , there exists a coordinate system  $(x, U)$  such that the integral manifolds of  $\Delta$  on  $U$  are given by the slices

$$\{p \in U : x^{m+1}(p) = a^{m+1}, \dots, x^n(p) = a^n\},$$

for fixed constants  $a^{m+1}, \dots, a^n$ . But then  $dx^{m+1}, \dots, dx^n$  span  $I^1(\Delta)$  on  $U$ , so there exists  $f_\beta^\alpha$  such that on  $U$ ,

$$\omega^\alpha = \sum_{\beta=m+1}^n f_\beta^\alpha dx^\beta. \quad \square$$

# Chapter 6

## Differential Forms and Integration

We wish to define a ‘coordinate-independent’ way to define integration, so that we may extend the notion of integration from open subsets of  $\mathbb{R}^n$  to more general manifolds. Recall the change of variables formula, which says that if  $y : U \rightarrow V$  is a diffeomorphism, and  $f \in L^1(U)$ , then

$$\int_V \left| \det \left( \frac{\partial y^i}{\partial x^j} \right) \right| \cdot (f \circ y)(x) dx = \int_U f(y) dy. \quad (6.1)$$

The objects we integrate on a manifold therefore must satisfy some transformation law so that the integrals defined in each coordinate system are independent. We cannot simply integrate scalar functions on a manifold since the integral of the function in different coordinate systems can give different answers. Of course, there is no natural choice of measure on a manifold, so we should not expect to be able to integrate pure functions. Whatever objects  $\omega$  we integrate must contain additional information which substitutes for a choice of measure on the manifold. If, in a certain coordinate system  $(x, U)$ , the object  $\omega$  is associated with a smooth function  $\omega_x : U \rightarrow \mathbb{R}$ , then in order to integrate  $\omega$  in a coordinate independent way based on (6.1), we would expect

$$\omega_y = \omega_x \cdot \left| \det \left( \frac{\partial y^i}{\partial x^j} \right) \right|. \quad (6.2)$$

Then, for any two coordinate systems  $(x, U)$ ,  $(y, V)$ , if  $\omega_x$  is compactly supported on  $U \cap V$ , then (6.1) indicates that

$$\int_{x(U)} (\omega_x \circ x^{-1}) dx = \int_{y(V)} (\omega_y \circ y^{-1}) dy. \quad (6.3)$$

If  $\omega$  is *compactly supported*, in the sense that there exists a compact set  $K \subset M$  such that  $\omega$  vanishes on any coordinate system disjoint from  $K$ , then we can cover  $K$  by a finite family of coordinate systems  $\{(x_1, U_1), \dots, (x_N, U_N)\}$ , as well as considering a smooth partition of unity  $\{\phi_1, \dots, \phi_N\}$  on  $U_1 \cup \dots \cup U_N$ , with  $\phi_i$  compactly supported on some compact set  $K_i \subset U_i$ . Then it is easy to verify from (6.3) that the quantity

$$\int_M \omega = \sum_{i=1}^N \int_{x_i(K \cap K_i)} \phi_i \cdot \omega_{x_i} \circ x_i^{-1} \quad (6.4)$$

is well defined and independent of the cover and partition of unity chosen. The objects  $\omega$  will be known as *scalar densities*.

## 6.1 Scalar Densities

It should be clear from what we have discussed previously how we will determine the scalar densities on a manifold  $M$ : we will find a vector bundle over  $M$ , and the scalar densities will then be precisely the sections of this vector bundle. We will first do this in a somewhat coordinate dependent way.

Consider the functor  $\text{Vol}_n$ , defined as follows:

- The domain of  $\text{Vol}_n$  is the family of  $n$  dimensional vector spaces, with the functor mapping each such space to  $\mathbb{R}$ .
- For each  $n$  dimensional vector space  $V$ , we fix an isomorphism  $T_V : V \rightarrow \mathbb{R}^n$ . For each linear map  $f : V \rightarrow W$ , we then have a map  $T_W \circ f \circ T_V^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . We define  $\text{Vol}_n(f)$  to be the linear map on  $\mathbb{R}$  given by multiplication of the *determinant* of the map  $T_W \circ f \circ T_V^{-1}$ .

The functor  $\text{Vol}_n$  is certainly smooth, and so given any  $n$  dimensional manifold  $M$ , we can associate a line bundle  $\text{Vol}(TM)$ . The family of isomorphisms  $\{T_V\}$  chosen is pretty much irrelevant to the construction of the line bundle  $\text{Vol}(TM)$ , i.e. since the resulting functor  $\text{Vol}_n$  will be naturally isomorphic.

A section of  $\text{Vol}(TM)$  will be called a *scalar density*. If  $\omega$  is a scalar density, and  $(x, U)$  induces a trivialization  $\text{Vol}(x_*) : \text{Vol}(TU) \rightarrow U \times \mathbb{R}$ , then we can associate with  $\omega$  a function  $\omega_x : U \rightarrow \mathbb{R}$  such that

$$\text{Vol}(x_*)\{\omega(p)\} = (p, \omega_x(p)).$$

If  $(y, V)$  is another coordinate system, then  $(y \circ x^{-1})_* : \mathcal{E}^n(U \cap V) \rightarrow \mathcal{E}^n(U \cap V)$  is given by  $(p, v) \mapsto (p, D(y \circ x^{-1})(x(p)))$ . Thus  $\text{Vol}((y \circ x^{-1})_*) : \mathcal{E}^1(U \cap V) \rightarrow \mathcal{E}^1(U \cap V)$  is given by

$$t_p \mapsto \left| \det(D(y \circ x^{-1})(x(p))) \right| \cdot t_p.$$

In particular, if  $\omega$  is associated the functions  $\omega_x$  and  $\omega_y$ , then

$$\omega_y = \omega_x \cdot \left| \det \left( \frac{\partial y^i}{\partial x^j} \right) \right|$$

Thus sections of  $\text{Vol}(TM)$  transform in coordinates in precisely the way required.

Given a section  $\omega$  of  $\text{Vol}(TM)$ , we can consider a pointwise-map

$$\omega : \Gamma(TM) \times \cdots \times \Gamma(TM) \rightarrow C^\infty(M)$$

such that if  $p \in U$ , and  $(x, U)$  is a coordinate system, and  $Y_j(p) = x_*(X_j|_p)$ , then

$$\omega(X_1, \dots, X_n)(p) = \omega_x(p) \cdot \left| \det(Y_1, \dots, Y_n) \right|.$$

If  $A \in \Gamma(\text{End}(TM))$  is a smooth family of endomorphisms of the tangent bundle, then

$$\omega(AX_1, \dots, AX_n) = |\det(A)| \cdot \omega(X_1, \dots, X_n). \quad (6.5)$$

Conversely, if  $\omega : \Gamma(TM) \times \dots \times \Gamma(TM) \rightarrow C^\infty(M)$  is a pointwise-defined map satisfying (6.5), then we can define a section  $\omega$  of  $\text{Vol}(TM)$  such that for each coordinate system  $(x, U)$ , and  $p \in U$ , we have

$$\omega_x(p) = \omega\left(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right).$$

Then (6.5) shows that the family of functions  $\{\omega_x\}$  agree with one another in such a way to induce a section of  $\text{Vol}(M)$ . In particular, we see that a scalar density can be viewed as a way to assign areas to parallelograms in the tangent space of a manifold. Thus a scalar density is a ‘ruler’ which can be used to measure infinitesimal volume on a manifold.

This property leads to a more coordinate independent definition of the tangent bundle  $\text{Vol}(M)$ . Given a vector space  $V$  of dimension  $n$ , we define a *degree one density*  $\omega$  on  $V$  to be a function on the  $n$ -fold wedge product  $V \wedge \dots \wedge V$  satisfying

$$\omega(cv) = |c| \cdot \omega(v)$$

for all  $v \in V \wedge \dots \wedge V$  and  $c \in \mathbb{R}$ . The set of degree one densities on a vector space  $V$  form a one-dimensional vector space; if  $V$  has a basis  $\{x_1, \dots, x_n\}$ , then any degree-one density is a multiple of the density given by

$$(\sum A_{1j}x_j, \dots, \sum A_{nj}x_j) \mapsto \text{Det}(A).$$

A map  $f : V \rightarrow W$  between vector spaces of the same dimension pulls back densities to densities. And so this construction gives rise to a smooth functor, which is easily checked to be naturally isomorphic to the functor we defined above. And thus we have an alternate, basis independent construction of the space of scalar densities.

## 6.2 Oriented Integration

There is a simple trick, that will prove immensely useful, related to the construction of scalar densities on an *oriented* manifold  $M$ . Instead of considering objects  $\omega$  which undergo the transformation law

$$\omega_y = \omega_x \cdot \left| \det \left( \frac{\partial y^i}{\partial x^j} \right) \right|,$$

we consider objects which undergo the transformation law

$$\omega_y = \omega_x \cdot \det \left( \frac{\partial y^i}{\partial x^j} \right).$$

For the same reasons as before, we can associate with each manifold  $M$  a one-dimensional bundle  $\Omega^n(M)$  (the notation will become clear shortly) whose sections transform in the manner above, and are known as *signed scalar densities*. To do this, we consider the functor  $\Omega^n$

which associates with each vector space  $V$  the real numbers, such that if  $f : V \rightarrow W$  is a linear map, then  $\Omega^n(f) : \mathbb{R} \rightarrow \mathbb{R}$  is given by multiplication by

$$\det(T_W \circ f \circ T_V^{-1})$$

rather than multiplication by the absolute value of this determinant.

If the functors  $\Omega^n$  and  $\text{Vol}$  are considered as functors on the family of *oriented vector spaces*, then they are actually isomorphic. Indeed, if  $V$  and  $W$  are oriented, and  $f : V \rightarrow W$  is an oriented isomorphism, then both  $\text{Vol}(f)$  and  $\Omega^n(f)$  are given by multiplication by  $\det(T_W \circ f \circ T_V^{-1})$ , since this determinant is positive. If  $M$  is an oriented manifold, we can therefore obtain an isomorphism between  $\text{Vol}(M)$  and  $\Omega^n(M)$  by taking a cover of  $M$  by oriented coordinate charts, in which case the correspondence is then trivial. Thus signed and unsigned scalar densities are, in a sense, the same category of objects on a manifold. On non-orientable manifolds, this is not the case. For instance, if  $\omega \in \Gamma(\Omega^n(M))$  is an everywhere non-zero section, then it automatically gives an orientation of  $M$  by letting those coordinate charts  $(x, U)$  be orientable if  $\omega_x$  is a positive function. On the other hand,  $\text{Vol}(M)$  always has an everywhere non-zero section, since the sign of a scalar density at a point does not depend on the coordinate system, and so we can form a partition of unity to add up positive scalar densities without worrying about cancellation.

## 6.3 Lower Dimensional Integration

It may appear strange why we introduced the family of signed scalar densities. But it turns out that signed scalar densities are much more amenable than unsigned scalar densities on *lower dimensional submanifolds*. Of course, we can certainly define a  $k$ -dimensional scalar density; we simply consider a pointwise-defined map

$$\omega : \Gamma(TM) \times \cdots \times \Gamma(TM) \rightarrow C^\infty(M),$$

which takes  $k$  vector fields  $X_1, \dots, X_k \in \Gamma(TM)$ . Restricted to the span of these vector fields,  $\omega$  should operate like the scalar densities we previously defined. In particular, this means that for any  $a_{ij} \in C^\infty(M)$ , for  $i, j \in \{1, \dots, k\}$ ,

$$\omega \left( \sum_{i=1}^k a_{1i} X_i, \dots, \sum_{i=1}^k a_{ki} X_i \right) = |\det(a_{ij})| \cdot \omega(X_1, \dots, X_k).$$

Thus  $\omega$  is a way of measuring the area of  $k$ -dimensional parallelograms. If  $N$  is a  $k$  dimensional submanifold of  $M$ , then the restriction of  $\omega$  to vectors fields in  $\Gamma(TN)$  induces a scalar density on  $N$ . Thus, if  $\omega$  is a  $k$ -dimensional scalar density and  $\text{supp}(\omega) \cap N$  is compact, then the integral

$$\int_N \omega$$

is well defined. The only problem with this definition is that the family of all scalar densities are very complex. In particular, the bundle  $\xi$  associated with the family of scalar densities is

infinite dimensional at each point. For instance, even when  $k = 1$ ,  $\xi_p$  would have to contain all the seminorms on  $T_p M$ , which is certainly an infinite dimensional family.

The advantage of considering *signed  $k$  dimensional densities*, is that there is an obvious class of simple signed densities. Instead of focusing on all functions  $\omega$  such that

$$\omega\left(\sum a_{1i}X_i, \dots, \sum a_{ki}X_i\right) = \det(a_{ij}) \cdot \omega(X_1, \dots, X_k),$$

we can focus on those maps  $\omega$  which are *multilinear*. The multilinear  $k$  dimensional scalar densities will be known as *differential  $k$ -forms*. Though these densities can now only on *oriented  $k$  dimensional submanifolds*, they have an elegant theory which makes them much more useful than their more general counterparts. In particular, the theory of multilinear maps satisfying the signed transformation law have a rich algebraic theory, which we now detail.

## 6.4 Alternating Tensors

A  $k$  tensor  $T \in T^k(V)$  is called *alternating* if

$$T(v_1, \dots, v_k) = 0$$

if  $v_i = v_j$  for some distinct indices  $i$  and  $j$ . This is equivalent to the tensor being *skew-symmetric*, i.e.

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -T(v_1, \dots, v_j, \dots, v_i, \dots, v_k),$$

because we are working over a field which is of a characteristic not equal to two. We let  $\Omega^k(V)$  denote the family of all alternating tensors. If  $T$  is alternating, then for any coefficients  $A = a^{ij}$ ,

$$T\left(\sum_{j=1}^k a^{1j}v_j, \dots, \sum_{j=1}^k a^{kj}v_j\right) = \det(A) \cdot T(v_1, \dots, v_k),$$

which is what we want from our signed scalar densities. We let  $\Omega^k(V)$  denote the family of all alternating tensors.

To understand this family better, for a  $k \times k$  matrix  $A = (a^{ij})$ , and a tensor  $T \in T^k(V)$ , we define a tensor  $T \cdot A \in T^k(V)$  by defining

$$(T \cdot A)(v_1, \dots, v_k) = T\left(\sum_j a^{1j}v_j, \dots, \sum_j a^{kj}v_j\right).$$

Then for any alternating tensor  $T$ ,

$$T \cdot A = \det(A)T$$

In particular, if we specialize this operation, defining, for  $\sigma \in S_k$ , and  $T \in T^k(V)$ ,

$$(T \cdot \sigma)(v_1, \dots, v_k) = T(v_{\sigma(1)}, \dots, v_{\sigma(k)}),$$

then this gives a right representation of  $S_k$  on  $T^k(V)$ , and for any alternating tensor  $T \in \Omega^k(V)$ ,  $T \cdot \sigma = \text{sgn}(\sigma)T$ . If we now define, for  $T \in T^k(V)$ ,

$$\text{Alt}(T) = \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma) \cdot (T \cdot \sigma),$$

then  $\text{Alt}(T)$  is an alternating tensor for all  $k$  tensors  $T$ , and if  $T$  is already alternating, then  $\text{Alt}(T) = T$ . Thus  $\text{Alt}$  is a projection map from  $T^k(V)$  to  $\Omega^k(V)$ .

If  $\omega \in \Omega^k(V)$ , and  $\nu \in \Omega^l(V)$ , it is rarely true that  $\omega \otimes \nu \in \Omega^{k+l}(V)$ . To obtain a natural substitute for the tensor product, we introduce the *wedge product* of alternating tensors, i.e. defining an alternating  $k + l$  tensor by setting

$$\begin{aligned} \omega \wedge \nu &= \frac{(k+l)!}{k!l!} \text{Alt}(\omega \otimes \eta) \\ &= \frac{1}{k!l!} \sum_{\sigma \in S_{k+l}} \text{sgn}(\sigma) (\omega \otimes \eta) \cdot \sigma. \end{aligned}$$

The factorials here are necessary in order to ensure the wedge product is an associative operation. The wedge product is also bilinear, and ‘anticommutative’, in the sense that  $\omega \wedge \nu = (-1)^{kl}(\nu \wedge \omega)$ . If  $f : W \rightarrow V$  is a linear map, then  $f^*(\omega \wedge \nu) = f^*(\omega) \wedge f^*(\nu)$ .

*Remark.* For computations, it is often handy to take sums over *cross sections* of  $S_{k+l}$ , rather than over the entire sum of  $S_{k+l}$ . We let  $G$  be the subgroup of  $S_{k+l}$  which map  $\{1, \dots, k\}$  and  $\{k+1, \dots, k+l\}$  into themselves. A cross section is then just a family of permutations containing exactly one element from each left coset of  $S_{k+l}/G$ . If  $K$  is a cross section of  $S_{k+l}$ , then for alternating tensors  $\omega$  and  $\nu$ , it is simple to verify that

$$\omega \wedge \nu = \sum_{\sigma \in K} \text{sgn}(\sigma) [(\omega \otimes \nu) \cdot \sigma].$$

The advantage of this approach is that we need only sum over  $(k+l)!/k!l!$  permutations rather than  $(k+l)!$  different permutations. An example of a cross section is the collection of all *shuffle permutations*, i.e. permutations  $\sigma \in S_{k+l}$  such that  $\sigma(1) < \sigma(2) < \dots < \sigma(k)$ , and  $\sigma(k+1) < \dots < \sigma(k+l)$ . Indeed, it is easy to see the shuffle permutations contain a representative from each left coset, since multiplication on the right by an element of  $G$  can be used to reorder each of the sets  $\{1, \dots, k\}$  and  $\{k+1, \dots, k+l\}$  so that the new permutation is a shuffle permutation. And no shuffle permutation is related to any other, since multiplying a shuffle permutation on the right by a nontrivial element of  $G$  will disrupt the ordering of the shuffle.

**Example.** Let  $\omega_1$  and  $\omega_2$  be linear functionals on  $V$ . Then for any two vector  $v \in V$ , we can associate the vector  $\omega(v) = (\omega_1(v), \omega_2(v)) \in \mathbb{R}^2$ . Then for  $v, w \in V$ ,

$$(\omega_1 \wedge \omega_2)(v, w) = \omega_1(v)\omega_2(w) - \omega_1(w)\omega_2(v) = \det(\omega(v), \omega(w)),$$

which is the signed area of the parallelogram spanned by  $\omega(v)$  and  $\omega(w)$ .

It takes some work, but let's prove associativity.

**Lemma 6.1.** *If  $S \in T^k(V)$ ,  $T \in T^l(V)$ , and  $\text{Alt}(S) = 0$ , then*

$$\text{Alt}(S \otimes T) = \text{Alt}(T \otimes S) = 0.$$

*Proof.* We calculate

$$(k+l)! \cdot \text{Alt}(S \otimes T) = \sum_{\sigma \in S_{k+l}} \text{sgn}(\sigma) \cdot (S \otimes T) \cdot \sigma.$$

If we let  $G$  be the subgroup of  $S_{k+l}$  consisting of permutations which fix  $\{k+1, \dots, k+l\}$ . Then

$$\sum_{\sigma \in G} (S \otimes T) \cdot \sigma = k! \cdot \text{Alt}(S) \otimes T = 0.$$

If we let  $\{\sigma_1, \dots, \sigma_N\}$  denote representatives of the cosets  $G \backslash S_{k+l}$ , then

$$\sum_{\sigma \sigma_i \in G \sigma_i} (S \otimes T) \cdot \sigma \cdot \sigma_i = \left( \sum_{\sigma \in G} (S \otimes T) \cdot \sigma \right) \cdot \sigma_i = 0 \cdot \sigma_i = 0.$$

Thus summing over each representative completes the calculation in the case  $\text{Alt}(S \otimes T)$ . The calculation of  $\text{Alt}(T \otimes S)$  is treated similarly.  $\square$

**Lemma 6.2.** *For any alternating tensors  $\omega \in \Omega^k(V)$ ,  $\nu \in \Omega^l(V)$ ,*

$$\text{Alt}(\text{Alt}(\omega \otimes \eta) \otimes \theta) = \text{Alt}(\omega \otimes \eta \otimes \theta) = \text{Alt}(\omega \otimes \text{Alt}(\eta \otimes \theta)).$$

*Proof.* Clearly

$$\text{Alt}(\text{Alt}(\omega \otimes \eta) - \omega \otimes \eta) = 0.$$

Thus the last lemma implies

$$\text{Alt}(\text{Alt}(\omega \otimes \eta) \otimes \theta - \omega \otimes \eta \otimes \theta) = 0.$$

But rearranging completes the proof of one side of the equality to be proved. The other side is treated similarly.  $\square$

**Theorem 6.3.** *If  $\omega \in \Omega^k(V)$ ,  $\eta \in \Omega^l(V)$ , and  $\theta \in \Omega^m(V)$ , then*

$$(\omega \wedge \eta) \wedge \theta = \omega \wedge (\eta \wedge \theta) = \frac{(k+l+m)!}{k!l!m!} \text{Alt}(\omega \otimes \eta \otimes \theta).$$

*Proof.* We calculate

$$\begin{aligned} (\omega \wedge \eta) \wedge \theta &= \frac{(k+l+m)!}{(k+l)!m!} \text{Alt}((\omega \wedge \eta) \otimes \theta) \\ &= \frac{(k+l+m)!}{k!l!m!} \text{Alt}(\text{Alt}(\omega \otimes \eta) \otimes \theta) \\ &= \frac{(k+l+m)!}{k!l!m!} \text{Alt}(\omega \otimes \eta \otimes \theta). \end{aligned}$$

A similar argument shows that the same is true for  $\omega \wedge (\eta \wedge \theta)$ , so we have proved associativity.  $\square$



*Remark.* More generally, we find that if  $\omega_i \in \Omega^{k_i}(V)$ , then

$$\omega_1 \wedge \cdots \wedge \omega_n = \frac{(k_1 + \cdots + k_n)!}{k_1! \cdots k_n!} \text{Alt}(\omega_1 \otimes \cdots \otimes \omega_n).$$

If we choose representatives  $\sigma_1, \dots, \sigma_N$  from  $G \backslash S_{k_1 + \cdots + k_n}$ , where  $G$  is the subgroup consisting of permutations which fix  $\{1, \dots, k_1\}, \{k_1 + 1, \dots, k_1 + k_2\}, \dots, \{k_1 + \cdots + k_{n-1} + 1, \dots, k_1 + \cdots + k_n\}$ , then

$$\begin{aligned} \omega_1 \wedge \cdots \wedge \omega_n \\ = \text{sgn}(\sigma_1) \cdot (\omega_1 \otimes \cdots \otimes \omega_n) \cdot \sigma_1 + \cdots + \text{sgn}(\sigma_N) \cdot (\omega_1 \otimes \cdots \otimes \omega_n) \cdot \sigma_N. \end{aligned}$$

*Remark.* If  $\omega_1, \dots, \omega_k \in V^*$ , and for  $v_1, \dots, v_k, (\omega_1 \wedge \cdots \wedge \omega_k)(v_1, \dots, v_k)$  is the area of the parallelepiped spanned by  $\omega(v_1), \dots, \omega(v_k) \in \mathbb{R}^k$ , i.e.

$$(\omega_1 \wedge \cdots \wedge \omega_k)(v_1, \dots, v_k) = \det(\omega(v_1), \dots, \omega(v_k)).$$

where  $\omega(v) = (\omega_1(v), \dots, \omega_k(v))$  for each  $v \in V$ .

The wedge product therefore turns  $\bigoplus_k \Omega^k(V)$  into a graded algebra, which has an identity if we let  $\Omega^0(V)$  denote the real numbers, with the wedge product defined as  $a \wedge \omega = a\omega$ . Unlike  $T(V)$ ,  $\Omega(V)$  is finite dimensional if  $V$  is finite dimensional, since we will soon show that  $\Omega^k(V) = (0)$  if  $k > \dim(V)$ .

*Remark.* We note that even without the  $(k + l)!/k!l!$ , the wedge product would still be associative. The reason the coefficients have been included is so that  $v_1, \dots, v_n$  are a basis for  $V$ , and  $\phi_1, \dots, \phi_n$  are the dual basis, then  $(\phi_1 \wedge \cdots \wedge \phi_n)(v_1, \dots, v_n) = 1$ .

**Theorem 6.4.** *If  $\{v_1, \dots, v_n\}$  is a basis for  $V$ , and  $\{\phi_1, \dots, \phi_n\}$  the dual basis for  $V^*$ , then the set of all*

$$\phi_I = \phi_{i_1} \wedge \cdots \wedge \phi_{i_k}$$

where  $I = \{1 \leq i_1 < \cdots < i_k \leq n\}$ , forms a basis for  $\Omega^k(V)$ .

*Proof.* If  $\omega = \sum a_{i_1 \dots i_k} \phi_{i_1} \otimes \cdots \otimes \phi_{i_k}$ , then

$$\begin{aligned} \omega &= \text{Alt}(\omega) \\ &= \sum a_{i_1 \dots i_k} \text{Alt}(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}) \\ &= \sum \frac{a_{i_1 \dots i_k}}{k!} (\phi_{i_1} \wedge \cdots \wedge \phi_{i_k}). \end{aligned}$$

Thus  $\Omega^k(V)$  is spanned by  $\{\phi_I\}$ . But if  $j_1 < \cdots < j_n$ , then  $\phi_I(v_{j_1}, \dots, v_{j_k}) = 1$  if and only if  $i_r = j_r$  for each  $k \in \{1, \dots, k\}$ , so the family of vectors  $\{\phi_I\}$  must be linearly independent.  $\square$

**Corollary 6.5.** *Covectors  $\omega_1, \dots, \omega_k \in \Omega^1(V)$  are linearly independent if and only if  $\omega_1 \wedge \cdots \wedge \omega_k \neq 0$ .*

*Proof.* If the vectors are linearly dependant, then certainly  $\omega_1 \wedge \cdots \wedge \omega_k = 0$ . If the vectors are linearly independent, they can be completed to a dual basis, and then  $\omega_1 \wedge \cdots \wedge \omega_k$  is an element of the basis.  $\square$

## 6.5 Differential Forms

We now return to the study of differential forms from an algebraic perspective. Given a linear function  $f : V \rightarrow W$ , we find that for the induced map  $f^* : T^k(W) \rightarrow T^k(V)$ ,  $f^*(\Omega^k(W)) \subset \Omega^k(V)$ . Thus this functor is smooth, and so given any smooth vector bundle  $\xi$ , we can associate another vector bundle  $\Omega^k(\xi)$ . We will of course want to focus on the alternating tensor bundle  $\Omega^k(TM)$ , which is a subbundle of  $T^k(TM)$ , and which we also denote by  $\Omega^k(M)$ . A section of this bundle is known as a *differential form*, the space of all such sections forming a  $C^\infty(M)$  module. By taking wedge products locally, we can take the direct sum of bundles  $\Omega(M) = \bigoplus \Omega^k(M)$  into a bundle of algebras, and the sections of this space form a  $C^\infty(M)$  algebra, denoted  $\Omega^*(M)$ . As we have seen in the last section, if  $(x, U)$  is a coordinate system, then any  $\omega \in \Gamma(TM)$  can locally be written as

$$\sum_I a_I dx^I,$$

where  $I$  ranges over all monotonic multi-indexes.

**Theorem 6.6.** *For any  $n$ , a manifold  $M^n$  is orientable if and only if  $\Omega^n(TM)$  has a nowhere vanishing section.*

*Proof.* If  $\omega$  is a nowhere zero differential  $n$  form, we can define an orientation  $\mu$  by letting  $(e_1, \dots, e_n) \in \mu_p$  if  $\omega(p)(e_1, \dots, e_n) > 0$ . In coordinates  $(x, U)$ ,  $\omega = f dx^1 \wedge \dots \wedge dx^n$ , and we must either have  $f > 0$  everywhere, or  $f < 0$  everywhere, since  $\omega$  doesn't vanish. This shows the choice of orientation is locally trivial. Conversely, if  $M^n$  is orientable, we can cover  $M$  by a family of orientated coordinate charts  $(x_\alpha, U_\alpha)$ . If  $\varphi_\alpha$  is a partition of unity subordinate to these charts, and we choose a nonvanishing differential form  $\omega_\alpha$  on  $U_\alpha$ , then we can consider the form

$$\omega = \sum \varphi_\alpha \omega_\alpha$$

and then  $\omega(p) \neq 0$  nowhere, since for a given oriented basis  $e_1, \dots, e_n$  of  $M_p$ ,  $\omega_\alpha(p)(e_1, \dots, e_n) > 0$ , hence  $\omega(p)(e_1, \dots, e_n) > 0$ .  $\square$

Before we move on, we note that if  $f : M \rightarrow N$  is a smooth map between manifolds, then we can pull back covariant tensor fields under the dual map  $f^*$ . We note that if  $\omega$  is a  $k$ -form on  $N$ , then  $f^*\omega$  is a  $k$ -form on  $M$ , i.e.  $f^*\omega$  is alternating if  $\omega$  is alternating. The dual map is particularly amenable to the operations on forms. In particular,  $f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta$ . One can thus feasibly compute the action of  $f^*$  in coordinates. In particular if we consider coordinates  $(x, U)$  on  $M$  and  $(y, V)$  on  $N$ , then

$$f^*(dy^i)(X) = dy^i(f_*X) = (f_*X)(y^i) = X(y^i \circ f),$$

so in particular,

$$f^*(dy^i) = \sum_j \frac{\partial(y^i \circ f)}{\partial x_j} dx^j.$$

Taking wedge products will allow us to compute the complete action of  $f^*$  on  $k$ -forms for any  $k$ . We perform a complete calculation only on  $n$ -forms.

**Lemma 6.7.** *Let  $f : M \rightarrow N$  be a smooth map between  $n$  dimensional manifolds. Then if  $(x, U)$  and  $(y, V)$  are coordinate systems on  $M$  and  $N$  respectively, then*

$$f^*(a \, dy^1 \wedge \cdots \wedge dy^n) = (a \circ f) \cdot \det \left( \frac{\partial y^i \circ f}{\partial x^j} \right) dx^1 \wedge \cdots \wedge dx^n.$$

*Proof.* We calculate that  $f^*(a \, dy^1 \wedge \cdots \wedge dy^n) = b \, dx^1 \wedge \cdots \wedge dx^n$  for some smooth function  $b$ , where

$$\begin{aligned} b(p) &= f^*(a \, dy^1 \wedge \cdots \wedge dy^n)_p \left( \frac{\partial}{\partial x^1} \Big|_p, \dots, \frac{\partial}{\partial x^n} \Big|_p \right) \\ &= (a \, dy^1 \wedge \cdots \wedge dy^n)_{f(p)} \left( f_* \left( \frac{\partial}{\partial x^1} \Big|_p \right), \dots, f_* \left( \frac{\partial}{\partial x^n} \Big|_p \right) \right) \\ &= (a \circ f)(p) \cdot (dy^1 \wedge \cdots \wedge dy^n)(f(p)) \\ &\quad \cdot \left( \sum \frac{\partial y^i \circ f}{\partial x^1} \Big|_p \frac{\partial}{\partial y^i} \Big|_{f(p)}, \dots, \sum \frac{\partial y^i \circ f}{\partial x^n} \Big|_p \frac{\partial}{\partial y^i} \Big|_{f(p)} \right) \\ &= (a \circ f)(p) \cdot \det \left( \frac{\partial y^i \circ f}{\partial x^j} \Big|_p \right). \end{aligned} \quad \square$$

## 6.6 Integration of Differential Forms

As we alluded to in the previous discussion in this chapter, if  $N$  is an oriented  $k$ -dimensional submanifold of a manifold  $M$ , and  $\omega$  is a  $k$ -form on  $M$ , then  $\omega$  induces a scalar density on  $N$ . Thus if  $\text{supp}(\omega) \cap N$  is compact, then the integral  $\int_N \omega$  is well defined. Of course, we are being as restrictive as possible here to ensure everything is concretely defined. One can weaken compactness assumptions given additional criteria about the form  $\omega$ , or extend the definition of  $N$  to more general domains. For instance, it is relatively easy to define the integral of a  $k$ -form on an oriented  $k$ -dimensional manifold with boundary  $N$  since the boundary of the manifold is a ‘set’ of measure zero, so we can set

$$\int_N \omega = \int_{N^\circ} \omega.$$

We also want to consider the integral of differential forms over shapes like triangles and cubes, which are neither manifolds nor manifold with boundary, since they have ‘corners’. So we define a *manifold with corners* to be a topological space  $M$  such that each point  $p \in M$  has a neighbourhood homeomorphic either to an open subset of  $\mathbb{R}^n$ , an open subset of a half space  $\mathbb{H}^n$  with the image of  $p$  on the boundary of  $\mathbb{H}^n$ , or an open subset of

$$\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_1, \dots, x_n \geq 0\},$$

with the image of  $p$  equal to 0. Topologically, a manifold with corners is the same as a manifold with boundary, but we can consider smooth structures on a manifold with corners which distinguish the two categories, and we will of course consider smooth manifolds with corners. Unlike smooth manifolds with boundary, the boundary points  $\partial M$  formed by points which do not have charts diffeomorphic to open subsets of  $\mathbb{R}^n$  does not form a smooth manifold with boundary, though the interior  $M^\circ$  is a smooth manifold. This is because  $\partial M$  still has corners. On the other hand, we can define  $(\partial M)^\circ$  to be the set of points  $p$  with charts diffeomorphic to  $\mathbb{H}^n$ , and this set will form a smooth manifold. Of course, on manifolds with corners we can also define tangent bundles, cotangent bundles, tensor bundles, and so on. In particular, if  $\omega$  is a  $k$ -form on an oriented smooth manifold with corners  $M$ , we will define

$$\int_M \omega = \int_{M^\circ} \omega.$$

If  $M$  is an oriented manifold, then  $(\partial M)^\circ$  has a natural orientation obtained by outward pointing vectors, and so for a  $k-1$  form on  $M$ , we define

$$\int_{\partial M} \omega = \int_{(\partial M)^\circ} \omega.$$

Other generalizations are obvious to the imagination, but hard to specify precisely, so we leave these for another time.

## 6.7 The Differential Operator

A section of  $\Omega^0(TM)$ , or a differential form of order 0, is just a smooth function on  $M$ . If  $f$  is such a function, we defined its differential as

$$df = \sum \frac{\partial f}{\partial x^\alpha} dx^\alpha$$

Notice that we can now say that ‘the differential of a zero form is a one form’. In this section we extend this definition, defining an operator  $d$  taking  $k$  forms to  $k+1$  forms. Our main goal thereafter will be to prove Stoke’s theorem, which says that for a  $k$ -form  $\omega$  on a  $k+1$  dimensional manifold  $M$  with boundary,

$$\int_M d\omega = \int_{\partial M} \omega.$$

The operator  $d$  is defined such that  $d\omega$  satisfies this equation pointwise on an ‘infinitesimal surface’.

The idea of extending a signed-area on  $k$ -dimensional shapes to a signed-area on  $k+1$ -dimensional shapes is quite sneaky. For a fixed  $p \in M$ , after switching to an oriented coordinate system  $(x, U)$  with  $x(p) = 0$ , we consider  $k+1$  tangent vectors  $X_1|_p, \dots, X_{k+1}|_p \in T_p M$ . We consider vectors  $a_i \in \mathbb{R}^n$  such that

$$X_i|_p = \sum_{j=1}^n a_i^j \frac{\partial}{\partial x^j} \Big|_p.$$

Then for suitably small  $\varepsilon > 0$ , we can consider a ‘curvilinear paralleliped’  $\Sigma(\varepsilon) \subset M$  which is the inverse image under  $x$  of the paralleliped generated by the vectors  $\varepsilon a_1, \dots, \varepsilon a_{k+1}$ . We also extend the vectors  $X_1|_p, \dots, X_{k+1}|_p$  to vector fields  $X_1, \dots, X_{k+1}$  on  $U$  such that  $x_*(X_i|_q) = (a_i)_{x(q)}$  for each  $q \in U$ . Then  $\Sigma(\varepsilon)$  has a natural orientation such that  $X_1, \dots, X_{k+1}$  is oriented. In particular,  $\partial\Sigma(\varepsilon)$  also has a natural orientation. The  $k+1$  form  $d\omega$  is then defined such that as  $\varepsilon \rightarrow 0$ ,

$$\int_{\partial\Sigma(\varepsilon)} \omega = \varepsilon^{k+1} \cdot d\omega_p(X_1, \dots, X_k) + o(\varepsilon^{k+1}) = \int_{\Sigma(\varepsilon v_1, \dots, \varepsilon v_k)} d\omega + o(\varepsilon^{k+1}).$$

Thus  $d\omega$  is defined so that Stoke’s theorem holds ‘infinitesimally’.

To show such a limit exists, we note that  $\partial\Sigma(\varepsilon)$  is the union of  $2k$  faces, which can be paired together and identified with elements of  $i \in \{1, \dots, k\}$ . Thus we consider the pair of faces  $\Sigma_i(\varepsilon)^-$  and  $\Sigma_i(\varepsilon)^+$ , the first being the inverse image of the  $k$ -dimensional paralleliped spanned by  $\{\varepsilon v_1, \dots, \widehat{\varepsilon v_i}, \dots, \varepsilon v_{k+1}\}$ , and the second being the inverse image of the same paralleliped shifted by  $\varepsilon v_i$ . Now  $X_i$  is an outward pointing vector field to  $\Sigma_i(\varepsilon)^+$ , and  $-X_i$  is an outward pointing vector to  $\Sigma_i(\varepsilon)^-$ . Thus the vector fields  $X_1, \dots, \widehat{X_i}, \dots, X_{k+1}$  are oriented on  $\Sigma_i(\varepsilon)^-$  if and only if the basis  $-X_i, X_1, \dots, \widehat{X_i}, \dots, X_{k+1}$  are oriented, i.e. if  $i$  is even, and  $X_1, \dots, \widehat{X_i}, \dots, X_{k+1}$  is an oriented basis on  $\Sigma_i(\varepsilon)^+$  if and only if  $X_i, X_1, \dots, \widehat{X_i}, \dots, X_k$  is oriented, i.e. if  $i$  is odd. Thus expanding definitions, and then applying the mean value theorem, we conclude

$$\begin{aligned} & \int_{\Sigma_i^+(\varepsilon)} \omega + \int_{\Sigma_i^-(\varepsilon)} \omega \\ &= (-1)^{i+1} \int_{[0, \varepsilon]^k} dy^1 \dots \widehat{dy^i} \dots dy^{k+1} \\ & \quad \left( \omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1})(x^{-1}(\varepsilon a_i + y_1 a_1 + \dots + y_{k+1} a_{k+1})) \right. \\ & \quad \left. - \omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1})(x^{-1}(y^1 a_1 + \dots + y^{k+1} a_{k+1})) \right) \\ &= (-1)^{i+1} \varepsilon \cdot \int_{[0, \varepsilon]^k} \left( X_i(\omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1}))(p) + o(1) \right) dy \\ &= (-1)^{i+1} \varepsilon^{k+1} X_i(\omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1}))(p) + o(\varepsilon^{k+1}). \end{aligned}$$

Thus we have computed that we should expect

$$d\omega_p(X_1|_p, \dots, X_{k+1}|_p) = \sum_{i=1}^{k+1} (-1)^{i+1} \cdot X_i(\omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1}))(p).$$

Our only problem is that the right hand side depends on the extension of the vector fields in a neighbourhood of  $p$ , and thus may not locally depend on the vector fields at the point (we also do not even know if this extension is independent of the coordinate system we chose

at  $p$ ). We will show this *does not* depend on the extension of the vector field, and thus not on the coordinate system, by giving a slightly more invariant definition of the differential operator, which works when applied to vector fields  $X_1, \dots, X_{k+1}$  that may not necessarily commute, unlike those given above, by introducing some Lie brackets.

**Theorem 6.8.** *Let  $\omega$  be a  $k$ -form on a manifold  $M$ , then there exists a  $k + 1$ -form  $d\omega$  on  $M$  such that for any vector fields  $X_1, \dots, X_{k+1} \in \Gamma(TM)$ ,*

$$\begin{aligned} d\omega(X_1, \dots, X_{k+1}) &= \sum_{i=1}^{k+1} (-1)^{i+1} X_i(\omega(X_1, \dots, \hat{X}_i, \dots, X_{k+1})) \\ &\quad + \sum_{1 \leq i < j \leq k+1} (-1)^{i+j} \omega([X_i, X_j], X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_{k+1}). \end{aligned}$$

*Proof.* It suffices to show that the definition is  $C^\infty(M)$  linear in the vector fields  $X_1, \dots, X_{k+1}$ . Write the expression above as  $\Sigma_1(X_1, \dots, X_{k+1}) + \Sigma_2(X_1, \dots, X_{k+1})$ . It is obvious that  $d\omega$  is  $\mathbb{R}$  linear in the vector fields. And for any function  $f \in C^\infty(M)$  and  $j \in \{1, \dots, k+1\}$ , we find by the product rule that

$$\begin{aligned} \Sigma_1(X_1, \dots, fX_j, \dots, X_{k+1}) &= \sum_{i \neq j} (-1)^{i+1} X_i(\omega(X_1, \dots, \hat{X}_i, \dots, fX_j, \dots, X_{k+1})) \\ &\quad + (-1)^{j+1} \cdot f \cdot X_j(\omega(X_1, \dots, \widehat{fX_j}, \dots, X_{k+1})) \\ &= f \cdot \Sigma_1(X_1, \dots, X_{k+1}) \\ &\quad + \sum_{i \neq j} (-1)^{i+1} X_i(f) \cdot \omega(X_1, \dots, \hat{X}_i, \dots, X_j). \end{aligned}$$

$$\begin{aligned} \Sigma_2(X_1, \dots, fX_j, \dots, X_{k+1}) &= f \cdot \Sigma_2(X_1, \dots, X_{k+1}) \\ &\quad + \sum_{i < j} (-1)^{i+j} \cdot X_i(f) \cdot \omega(X_j, X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_{k+1}) \\ &\quad + \sum_{i > j} (-1)^{i+j+1} \cdot X_i(f) \cdot \omega(X_j, X_1, \dots, \hat{X}_j, \dots, \hat{X}_i, \dots, X_{k+1}) \\ &= f \cdot \Sigma_2(X_1, \dots, X_{k+1}) \\ &\quad + \sum_{i \neq j} (-1)^i X_i(f) \cdot \omega(X_1, \dots, \hat{X}_i, \dots, X_j). \end{aligned}$$

Thus summing up, we conclude

$$d\omega(X_1, \dots, fX_j, \dots, X_{k+1}) = fd\omega(X_1, \dots, X_{k+1}),$$

hence the operator is  $C^\infty(M)$  linear.  $\square$

It is a bit of magic to define the differential in terms of vector fields, and then find that this definition is really pointwise by proving  $C^\infty(M)$  linearity. But it is necessary magic; even though  $d\omega$  does not depend on any properties of the vector fields except those at a point, it certainly depends on the behaviour of  $\omega$  in a neighbourhood of a point, and this must enter the picture somehow.

*Remark.* We note that  $d$  is not  $C^\infty(M)$  linear, only  $\mathbb{R}$  linear, which reflects the fact that  $d$  is not defined pointwise; really only very basic operators can be defined pointwise, since in coordinates, they only are obtained by linear combinations of the values defining the tensor.

We note that the exterior derivative defined above does generalize the definition  $df$  defined for smooth functions, which we think of as 0-forms on the manifold. The definition above is precisely how we defined  $df$ , namely

$$df(X) = X(f).$$

In coordinates, we have

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i.$$

It is incredibly convenient to work with the higher order exterior derivative in coordinates, which we now do, after studying the differential.

If  $X_1, \dots, X_{k+1}$  are vector fields with  $[X_i, X_j] = 0$  for each  $i, j$ , we find

$$\begin{aligned} d(a\omega)(X_1, \dots, X_{k+1}) &= \sum_{i=1}^{k+1} (-1)^{i+1} X_i \left( (a\omega)(X_1, \dots, \hat{X}_i, \dots, X_{k+1}) \right) \\ &= ad\omega(X_1, \dots, X_{k+1}) \\ &\quad + \sum_{i=1}^{k+1} (-1)^{i+1} X_i(a) \cdot \omega(X_1, \dots, \hat{X}_i, \dots, X_{k+1}) \\ &= ad\omega(X_1, \dots, X_{k+1}) \\ &\quad + \sum_{i=1}^{k+1} (-1)^{i+1} da(X_i) \cdot \omega(X_1, \dots, \hat{X}_i, \dots, X_{k+1}) \\ &= ad\omega(X_1, \dots, X_{k+1}) + (da \wedge \omega)(X_1, \dots, X_{k+1}). \end{aligned}$$

Next, we calculate that for any index set  $I$ ,  $d(dx^I) = 0$ . We find

$$\begin{aligned} d(dx^I) \left( \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_k}} \right) &= \sum_{i=1}^{k+1} (-1)^{i+1} \frac{\partial}{\partial x^{j_i}} \left( dx^I \left( \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_i}}, \dots, \frac{\partial}{\partial x^{j_k}} \right) \right) \\ &= 0. \end{aligned}$$

Thus, for any index set  $I$  and smooth function  $a$ ,  $d(adx^I) = da \wedge dx^I$ . Thus

$$d(adx^I) = \sum_{i=1}^n \frac{\partial a}{\partial x^i} dx^i \wedge dx^I,$$

which gives a simple calculation to compute the exterior derivative.

**Theorem 6.9.** For any  $k$ -form  $\omega$  and  $l$ -form  $\eta$ ,

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

*Proof.* We work locally in coordinates. By linearity, it suffices to deal with the case  $\omega = adx^I$  and  $\eta = bdx^J$ . Since there is  $\alpha$  such that  $dx^I \wedge dx^J = (-1)^\alpha dx^{I \cup J}$ , so  $d(dx^I \wedge dx^J) = 0$ . Since  $\omega \wedge \eta = abdx^I \wedge dx^J$ , we find

$$d(\omega \wedge \eta) = d(ab) \wedge dx^I \wedge dx^J = (adb + bda) \wedge dx^I \wedge dx^J.$$

On the other hand,

$$d\omega \wedge \eta + (-1)^k \omega \wedge d\eta = bda \wedge dx^I \wedge dx^J + (-1)^k adx^I \wedge db \wedge dx^J = (adb + bda) \wedge dx^I \wedge dx^J. \quad \square$$

**Theorem 6.10.** For any  $k$  form  $\omega$ ,  $d(d\omega) = 0$ . In short,  $d^2 = 0$ .

*Proof.* If  $\omega = adx^I$ , then

$$\begin{aligned} d(d\omega) &= d(da \wedge dx^I) \\ &= \sum_{i=1}^n d\left(\frac{\partial a}{\partial x^i} dx^i \wedge dx^I\right) \\ &= \sum_{i=1}^n \frac{\partial a}{\partial x^i} d(dx^i \wedge dx^I) + \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 a}{\partial x^j \partial x^i} dx^j \wedge dx^i \wedge dx^I. \end{aligned}$$

The first sum is zero because  $d(dx^i \wedge dx^I) = 0$  for all  $i$  and  $I$ . The second sum is zero because if  $i = j$ ,  $dx^j \wedge dx^i \wedge dx^I = 0$ , and if  $i \neq j$  we can pair up the term with the term obtained by swapping  $i$  and  $j$ , and since mixed partials are equal, and  $dx^j \wedge dx^i = -dx^i \wedge dx^j$ , the two terms cancel one another out.  $\square$

The exterior derivative is, in some senses, uniquely defined. It is the unique linear operator  $d$  such that  $d^2 = 0$ ,  $df$  is defined as usual for functions, and such that  $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$ . This is easy to see by working in a coordinate system, as we saw before. But the derivative is some sense, even more natural, since it satisfies a naturality condition.

**Theorem 6.11.** Given  $f : M \rightarrow N$  and a  $k$ -form  $\omega$  on  $N$ ,  $d(f^*\omega) = f^*(d\omega)$ .

*Proof.* We prove the result by induction. For the case  $k = 0$ , and a smooth function  $g \in C^\infty(M)$ , if  $(x, U)$  is a coordinate system on  $M$ , and  $(y, V)$  a coordinate system on  $N$ ,

$$\begin{aligned} f^*(dg) &= f^*\left(\sum_{i=1}^m \frac{\partial g}{\partial y^i} dy^i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial g}{\partial y^i} \circ f\right) \frac{\partial y^i \circ f}{\partial x^j} \cdot dx^j. \end{aligned}$$



On the other hand, by the chain rule

$$\begin{aligned}
d(f^*g) &= d(g \circ f) \\
&= \sum_{j=1}^m \frac{\partial(g \circ f)}{\partial x^j} dx^j \\
&= \sum_{j=1}^m \sum_{i=1}^n \left( \frac{\partial g}{\partial y^i} \circ f \right) \frac{\partial y^i \circ f}{\partial x^j} \cdot dx^j.
\end{aligned}$$

Thus the result is true for 0-forms. Next, we consider the case of 1-forms, but only when  $\omega = dy^j$  for some  $j$ . Then clearly  $f^*(d^2y^j) = 0$ . On the other hand,

$$f^*(dy^j) = \sum_{\alpha=1}^n \frac{\partial y^j \circ f}{\partial x^\alpha} dx^\alpha.$$

Thus

$$d(f^*(dy^j))(X) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \left( \frac{\partial^2 y^j \circ f}{\partial x^\alpha \partial x^\beta} \right) dx^\alpha \wedge dx^\beta.$$

But this is equal to zero since mixed partial cancel out. Now we consider the general case by induction. Assuming the formula for  $k - 1$  forms, if  $\omega = adx^i \wedge dx^I$ , we calculate that for any index  $i$  and size  $k - 1$  index set  $I$ , by induction we find

$$\begin{aligned}
d(f^*\omega) &= d(f^*(adx^I)) \\
&= d(f^*(dx^i) \wedge f^*(adx^I)) \\
&= d(f^*(dx^i)) \wedge f^*(adx^I) - f^*(dx^i) \wedge d(f^*(adx^I)) \\
&= f^*(d^2x^i) \wedge f^*(adx^I) - f^*(dx^i) \wedge f^*(d(adx^I)) \\
&= -f^*(dx^i \wedge d(adx^I)) \\
&= f^*(adx^i \wedge dx^I).
\end{aligned}$$

□

For each fixed  $k$ , the operator  $d$  is the unique linear map, up to scalar multiplication, which takes  $k$  forms to  $k + 1$  forms, such that for each  $f : M \rightarrow N$ ,  $f^*(d\omega) = d(f^*\omega)$ . This is a result of R.S Palais (Natural Operations on Differential Forms, 1959). This might explain why the differential operator turns out to be so important.

**Example.** One nice thing about the differential is it generalizes the differential operators found in the three dimensional vector calculus found in a standard multivariate class. If  $f$  is a function, then

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

Thus  $df$  carries the same information as the gradient of  $f$ . If  $\omega = a_1 dx + a_2 dy + a_3 dz$ , then

$$d\omega = \left( \frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} \right) dx \wedge dy + \left( \frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} \right) dy \wedge dz + \left( \frac{\partial a_1}{\partial z} - \frac{\partial a_3}{\partial x} \right) dz \wedge dx.$$

Thus if we write  $Z = (a_1, a_2, a_3)$ , then we verify

$$d\omega(X, Y)(p) = (\nabla \times Z) \cdot (X \times Y).$$

Thus  $d\omega$  carries the same information as the curl of the vector field  $Z$ . If

$$\omega = a_1 dy \wedge dz + a_2 dz \wedge dx + a_3 dx \wedge dy$$

then

$$d\omega = \left( \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z} \right) (dx \wedge dy \wedge dz).$$

Thus if  $W = (a_1, a_2, a_3)$ , then  $d\omega(X, Y, Z) = (\nabla \cdot W)((X \times Y) \cdot Z)$ . So  $d\omega$  carries the same information as the divergence of the vector field  $W$ . In particular, the equation  $d^2 = 0$  implies the three classical equations

$$\nabla \times \nabla f = \nabla \cdot (\nabla \times X) = 0$$

for any smooth functions  $f$  and vector fields  $X$ . In the next section, we prove a generalization of Stoke's theorem which encompasses Green's theorem, Gauss' theorem, and Stoke's theorem.

## 6.8 Stoke's Theorem

We now show that Stoke's theorem holds for manifolds. The proof, after switching to simple coordinate systems, is a simple application of the fundamental theorem of calculus.

**Theorem 6.12.** *If  $M$  is an oriented  $n$  dimensional manifold, and  $\omega$  is an  $n - 1$  form with compact support, then*

$$\int_{\partial M} \omega = \int_M d\omega.$$

*Proof.* Suppose first that  $\omega$  is supported on  $[0, 1]^n$ , but compactly supported on the interior. Write

$$\omega = \sum_{i=1}^n a_i \cdot dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n.$$

Then

$$d\omega = \sum_{i=1}^n (-1)^{i+1} \frac{\partial a_i}{\partial x^i} \cdot (dx^1 \wedge \cdots \wedge dx^n).$$

Thus we calculate from Fubini's theorem and the fundamental theorem of calculus that

$$\begin{aligned} & \sum_{i=1}^n (-1)^{i+1} \int_{[0,1]^n} \frac{\partial a_i}{\partial x^i}(x) dx^1 \cdots dx^n \\ &= \sum_{i=1}^n (-1)^{i+1} \int_{[0,1]^{n-1}} [a_i(x_1, \dots, 1, \dots, x_n) - a_i(x_1, \dots, 0, \dots, x_n)] dx^1 \cdots \widehat{dx^i} \cdots dx^n \\ &= 0. \end{aligned}$$

Thus the theorem is obvious in this case. Next, we address the case where  $\omega$  is supported on a compact subset of  $[0, 1]^n$ , but  $\text{supp}(\omega) \cap \partial[0, 1]^n \subset \{x \in [0, 1]^n : x_1 = 0\}$ . Again, the fundamental theorem of calculus and Fubini's theorem enables us to conclude that

$$\int_M d\omega = - \int_{[0,1]^{n-1}} a_1(0, x_2, \dots, x_n) \cdot dx^2 \dots dx^n.$$

On the other hand, because  $\{e_2, \dots, e_n\}$  is *not oriented correctly* on  $\partial M$ ,

$$\int_{\partial M} \omega = - \int_{[0,1]^{n-1}} a_1(0, x_2, \dots, x_n) \cdot dx^2 \dots dx^n,$$

thus Stoke's theorem is, again, obvious. But now if  $M$  is an arbitrary manifold, and  $\omega$  is compactly supported, then  $\text{supp}$  can be covered by finitely many coordinate systems  $(x_1, U_1), \dots, (x_N, U_N)$  like the one above, and a partition of unity  $\phi_1, \dots, \phi_N$ , such that  $(x^{-1})^*(\phi_i \omega)$  fits into one of the two cases above. Then we see that for each  $i$ ,

$$\int_M d(\phi_i \omega) = \int_{\partial M} \omega.$$

We also know

$$\int_M d\omega = \sum_{i=1}^N \int_M \phi_i d\omega.$$

We calculate that  $d(\phi_i \omega) = d(\phi_i) \wedge \omega + \phi_i d\omega$ . Since  $\phi_1 + \dots + \phi_N = 1$ ,  $d\phi_1 + \dots + d\phi_N = 0$ . Thus

$$\sum_{i=1}^N d(\phi_i \omega) = \sum_{i=1}^N \phi_i d\omega.$$

Thus

$$\sum_{i=1}^N \int_M \phi_i d\omega = \sum_{i=1}^N \int_M d(\phi_i \omega) = \sum_{i=1}^N \int_{\partial M} \phi_i \omega = \int_{\partial M} \omega. \quad \square$$

Just as the fundamental theorem of calculus is essential to understand analysis, Stokes' theorem is essential to modern manifold theory. As we shall see in the next chapter, it provides a connection between the analysis of the 'boundaries' of manifolds to the analysis of differential forms, which continues the connection between the topology of a manifold and its smooth structure.

## 6.9 Basic De-Rham Cohomology

Our goal in this chapter will be study when a  $k$ -form  $\omega$  on a manifold  $M$  is *exact*, that is, when there exists a  $k - 1$  form  $\nu$  such that  $\omega = d\nu$ . In some senses, this is like forming an 'antiderivative' to the form  $\omega$ . In particular, this is exactly the case when considering a

1-form  $f dx$  on  $\mathbb{R}$ , in which case if  $dF = f$ , then  $F$  is an antiderivative to  $f$ . We will find that the ability to find such antiderivatives depends largely on the ‘shape’ of the manifold  $M$ , i.e. its homology class. The most fundamental result in this regard is the Poincare lemma. We recall that a smooth manifold  $M$  is *smoothly contractible* if there exists a smooth map  $H : [0, 1] \times M \rightarrow M$  such that  $H(0, p) = p$  for all  $p \in M$ , and there exists  $p_0 \in M$  such that  $H(1, p) = p_0$  for all  $p \in M$ . If  $\omega$  is an exact  $k$ -form, then certainly  $\omega$  is *closed*, i.e.  $d\omega = 0$ . 12e manifold, this is also sufficient.

**Lemma 6.13** (Poincare Lemma). *If  $M$  is a smoothly contractible manifold and  $k > 0$ , all closed  $k$  forms on  $M$  are exact.*

The trick to the proof of this statement is to analyze  $[0, 1] \times M$ . For each  $t \in [0, 1]$  we define  $i_t : M \rightarrow [0, 1] \times M$  by setting  $i_t(p) = (t, p)$ . If  $\omega$  is a form on  $[0, 1] \times M$ . We begin by analyzing the one forms on  $[0, 1] \times M$ .

**Lemma 6.14.** *If  $\omega$  is a closed 1-form on  $[0, 1] \times M$ , then  $i_1^*\omega - i_0^*\omega$  is an exact one-form on  $M$ .*

*Proof.* Let us begin by working in coordinates. Consider the function  $t : [0, 1] \times M \rightarrow [0, 1]$  given by projection onto the first factor. If  $(x, U)$  is a coordinate system on  $M$ , then  $dx^1, \dots, dx^n, dt \in \Gamma(T^*([0, 1] \times U))$  form a frame, and  $(x^1, \dots, x^n, t)$  form a coordinate system in a neighbourhood of each point. Thus given a closed 1-form  $\omega$ , there are smooth functions  $a_1, \dots, a_n, f \in C^\infty([0, 1] \times U)$  such that on  $[0, 1] \times U$ ,

$$\omega = a_1 dx^1 + \dots + a_n dx^n + f dt.$$

For each  $\alpha \in [0, 1]$  and  $p \in M$ ,

$$i_\alpha^*(\omega)(p) = a_1(\alpha, p) dx^1 + \dots + a_n(\alpha, p) dx^n.$$

Now  $d\omega = 0$  implies that for each  $i \in \{1, \dots, n\}$ ,

$$\frac{\partial a_i}{\partial t} = \frac{\partial f}{\partial x^i}.$$

Thus

$$a_i(1, p) - a_i(0, p) = \int_0^1 \frac{\partial a_i}{\partial t}(s, p) ds = \int_0^1 \frac{\partial f}{\partial x^i}(s, p) ds.$$

In particular, if we define a smooth function  $g$  on  $U$  by setting

$$g(p) = \int_0^1 f(s, p) ds,$$

then

$$\frac{\partial g}{\partial x^i} = \int_0^1 \frac{\partial f}{\partial x^i}(s, p) ds = a_i(1, p) - a_i(0, p).$$

It thus follows that  $dg = i_1^* \omega - i_0^* \omega$ . This does not complete the proof, since we worked only locally in a coordinate system. But the function  $f$  does not really depend on the coordinate system, since the vector field

$$X = \frac{\partial}{\partial t},$$

is globally defined on  $[0, 1] \times M$ , and  $f = \omega(X)$ . Thus the local definition of  $g$  extends globally to all of  $M$ .  $\square$

The general case of  $k$ -forms works in essentially the same way for any  $k > 0$ .

**Lemma 6.15.** *If  $\omega$  is a closed  $k$ -form, then  $i_1^* \omega - i_0^* \omega$  is exact.*

*Proof.* Write  $\pi(t, p) = p$  for the smooth projection map  $\pi : [0, 1] \times M \rightarrow M$ . We can write  $\omega = \omega_1 + (dt \wedge \eta)$ , when  $\eta$  is a  $k - 1$  form, and for each  $(t, p) \in [0, 1] \times M$  and  $X_1, \dots, X_k \in T_{(t,p)}([0, 1] \times M)$ , we have  $\omega_1(X_1, \dots, X_k) = 0$  if there exists  $i \in \{1, \dots, k\}$  such that  $\pi_*(X_i) = 0$ , with  $\eta$  also satisfying the analogous property for  $k - 1$  forms. We define a  $k - 1$  form  $I\omega$  on  $M$  such that for  $X_1, \dots, X_{k-1} \in T_{(t,p)}([0, 1] \times M)$ ,

$$(I\omega)_p(X_1, \dots, X_{k-1}) = \int_0^1 \eta(s, p)((i_s)_* X_1, \dots, (i_s)_* X_{k-1}) ds$$

We claim that for any (not necessarily closed)  $k$ -form  $\omega$ ,

$$i_1^* \omega - i_0^* \omega = d(I\omega) + I(d\omega),$$

which would complete the proof in general. The advantage of this formula is that both sides of the equation are linear in  $\omega$ , and so we can work in coordinates, and moreover, assume  $\omega$  is a monomial. We consider two cases:

- $\omega = f dx^I$  for some index set  $I$ . Then

$$d\omega = * + \frac{\partial f}{\partial t} dt \wedge dx^I.$$

It is easy to see that

$$I(d\omega)_p = \int_0^1 \frac{\partial f}{\partial t}(p, s) ds = (i_1^* \omega)_p - (i_0^* \omega)_p.$$

On the other hand,  $I\omega = 0$ , so the proof is complete.

- $\omega = f dt \wedge dx^I$ . Then  $i_1^* \omega = i_0^* \omega = 0$ . Now

$$\begin{aligned} I(d\omega)_p &= I \left( - \sum_{\alpha=1}^n \frac{\partial f}{\partial x^\alpha} dt \wedge dx^\alpha \wedge dx^I \right)_p \\ &= - \sum_{\alpha=1}^n \left( \int_0^1 \frac{\partial f}{\partial x^\alpha}(s, p) ds \right) dx^\alpha \wedge dx^I. \end{aligned}$$

On the other hand, we have

$$(I\omega)_p = \left( \int_0^1 f(s, p) ds \right) dx^I.$$

And then

$$\begin{aligned} d(I\omega) &= \sum_{\alpha=1}^n \frac{\partial}{\partial x^\alpha} \left( \int_0^1 f(s, p) ds \right) dx^\alpha \wedge dx^I \\ &= \sum_{\alpha=1}^n \left( \int_0^1 \frac{\partial f}{\partial x^\alpha}(s, p) ds \right) dx^\alpha \wedge dx^I. \end{aligned}$$

Adding the two terms completes this argument.

Thus if  $\omega$  is closed, then  $i_1^* \omega - i_0^* \omega = d(I\omega)$  is exact.  $\square$

*Proof.* To complete the proof of the Poincare lemma, we consider a smooth contraction  $H : [0, 1] \times M \rightarrow M$ . Given any  $k$ -form  $\omega$  on  $M$ , we consider the  $k$ -form  $H^* \omega$  on  $[0, 1] \times M$ . If  $\omega$  is closed, then  $H^* \omega$  is closed. Thus  $i_1^*(H^* \omega) - i_0^*(H^* \omega)$  is exact. But  $i_0^*(H^* \omega) = 0$ , and  $i_1^*(H^* \omega) = \omega$ .  $\square$

In the special case where  $U$  is an open subset of  $\mathbb{R}^n$  which is *star shaped*, in the sense that there is a point  $x_0 \in U$  such that if  $x \in U$ , then any point on the line between  $x$  and  $x_0$  lies in  $U$ , there exists a  $k$ -form  $\omega$  we can find a *explicit* formula for a  $k - 1$  form  $\eta$  such that  $d\eta = \omega$ .

**Theorem 6.16.** *If  $U$  is star-shaped around the origin, define  $H(t, x) = tx$ . If*

$$\omega = \sum_{i_1 < \dots < i_k} a_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k},$$

*then*

$$\begin{aligned} I(H^* \omega) &= \sum_{i_1 < \dots < i_k} \sum_{\alpha=1}^k (-1)^{\alpha-1} \left( \int_0^1 t^{k-1} a_{i_1 \dots i_k}(tx) dt \right) \\ &\quad x^{i_\alpha} dx^{i_1} \wedge \dots \wedge \widehat{dx^{i_\alpha}} \wedge \dots \wedge dx^{i_k}. \end{aligned}$$

*In particular, if  $\omega$  is closed, then the formula above gives a form  $\eta$  such that  $\omega = d\eta$ .*

*Proof.* We calculate that

$$\begin{aligned} (H^* \omega)(t, x) &= \sum_{i_1 < \dots < i_k} a_{i_1 \dots i_k}(tx) \bigwedge_{\alpha=1}^k (x^{i_\alpha} dt + t dx^\alpha) \\ &= \sum_{i_1 < \dots < i_k} a_{i_1 \dots i_k}(tx) \left( t^k dx^{i_1} \wedge \dots \wedge dx^{i_k} \right. \\ &\quad \left. + \sum_{\alpha=1}^k (-1)^{\alpha-1} t^{k-1} x^{i_\alpha} \cdot dt \wedge (dx^{i_1} \wedge \dots \wedge \widehat{dx^{i_\alpha}} \wedge \dots \wedge dx^{i_k}) \right). \end{aligned}$$

Thus

$$\begin{aligned} I(H^*\omega)_x &= \sum_{i_1 < \dots < i_k} \sum_{\alpha=1}^k (-1)^{\alpha-1} \left( \int_0^1 a_{i_1 \dots i_k}(tx) t^\alpha \right) x^{i_\alpha} (dx^{i_1} \wedge \dots \wedge \widehat{dx^{i_\alpha}} \wedge \dots \wedge dx^{i_k}). \end{aligned}$$

This completes the calculation.  $\square$

Just as topological information about a manifold indicates information about antidifferentiation, the Poincare lemma, combined with Stokes' theorem, also allows us to transfer information about integration into topological information.

**Theorem 6.17.** *Any compact oriented manifold is not smoothly contractible.*

*Proof.* Let  $M$  be a compact oriented manifold of dimension  $n$ . Working locally in coordinates, we can find an  $n$ -form  $\omega$  such that

$$\int_M \omega \neq 0.$$

Since  $d\omega$  is an  $n+1$  form, it is trivial that  $d\omega = 0$ , so  $\omega$  is closed. If there existed  $\eta$  such that  $d\eta = \omega$ , then  $\eta$  is trivially compactly supported because it is defined on a compact manifold, so we may apply Stokes' theorem to conclude

$$\int_M \omega = \int_M d\eta = \int_{\partial M} \eta = 0.$$

Thus there exists a closed form on  $M$  which is not exact, so  $M$  cannot be smoothly contractible.  $\square$

*Remark.* This tactic clearly should not work on a non compact oriented manifold. Indeed, if we consider any non-negative, smooth, compactly supported function  $\phi$  on the real line, and set  $\omega = \phi dx$ , then  $\omega$  is closed *and exact*, because  $\omega = d\eta$ , where

$$\eta(t) = \int_{-\infty}^t \phi(s) ds.$$

In this case,  $\eta$  is not compactly supported, so we cannot directly apply Stokes' theorem to obtain a contradiction.

We now want to form the ability to 'count' how many closed forms there are on a manifold which are not exact. This will make it much easier for us to characterize the behaviour of integrals of differential forms. If we let  $Z^k(M) \subset \Omega^k(M)$  denote the class of closed  $k$ -forms on  $M$ , and  $B^k(M)$  the class of exact  $k$  forms, then we can form the quotient space  $H^k(M) = Z^k(M)/B^k(M)$ , known as the  *$k$ 'th De Rham cohomology vector space*. The dimension of  $H^k(M)$  then precisely counts the maximal number of linearly independent non-exact closed  $k$  forms on  $M$ . This is natural from the perspective of integration theory; if  $N$  is an

oriented  $k$ -dimensional compact submanifold of  $M$  without boundary, and  $\eta$  is a  $k - 1$  form, then Stoke's theorem implies that, since  $\partial N = \emptyset$ ,

$$\int_N d\eta = \int_{\partial N} \eta = 0.$$

Thus for any  $k$ -form  $\omega$ ,

$$\int_N (\omega + d\eta) = \int_N \omega.$$

Thus the family of all 'possible integrals' on  $N$  given by closed forms  $\omega$  is characterized by the De Rham cohomology  $H^k(M)$ . We use the computation of De Rham groups as a chance to hone our techniques for integrating functions on manifolds.

**Example.** *If  $M$  is smoothly contractible, then  $H^k(M)$  is trivial for  $k \geq 0$ . Thus there are no interesting ways to integrate closed submanifolds on Euclidean space.*

**Example.** *The space  $B^0(M)$  is trivial, so  $H^0(M)$  is really just the space of all smooth real-valued functions  $f$  on  $M$  such that  $df = 0$ . Another way of describing this is the space of all functions on  $M$  which vanish on each component of  $M$ , so the dimension of  $H^0(M)$  gives the number of connected components of  $M$ .*

**Example.** *We have shown that if  $M$  is oriented, compact, and  $n$ -dimensional, then  $H^n(M)$  is nontrivial.*

We need to do some more heavy lifting before we can calculate the homology of anything non-contractible. The next simplest example for us will be obtained from introducing a hole to a contractible example, i.e.  $\mathbb{R}^n - \{0\}$ . To calculate this De-Rham cohomology, we introduce a generalized system of polar coordinates on  $\mathbb{R}^n - \{0\}$  for all  $n > 0$ .

On  $S^{n-1}$ , there is a natural choice of an 'orientation form' on  $S^{n-1}$ . We define an  $n - 1$  form  $\sigma$  on  $\mathbb{R}^n - \{0\}$  by setting

$$\sigma_x(X_1, \dots, X_{n-1}) = \det(x, X_1, \dots, X_{n-1}).$$

Then  $\sigma$  restricts to a form on  $S^{n-1}$  such that for any oriented vectors  $X_1, \dots, X_{n-1} \in T_x S^{n-1}$ ,  $\sigma_x(X_1, \dots, X_{n-1}) > 0$ . In standard coordinates,

$$\sigma = \sum_{i=1}^n (-1)^{i-1} x^i dx^1 \wedge \dots \wedge \widehat{dx^i} \wedge \dots \wedge dx^n.$$

This form is not closed on  $\mathbb{R}^n$ , i.e.  $d\sigma = n \cdot dx^1 \wedge \dots \wedge dx^n$ , but its restriction  $\sigma|_{S^{n-1}}$  to  $S^{n-1}$  is closed, simply because it is an  $n - 1$  form on an  $n - 1$  dimensional manifold.

We now consider the retraction map  $r : \mathbb{R}^n - \{0\} \rightarrow S^{n-1}$  given by setting  $r(x) = x|x|^{-1}$ , and consider the form  $\eta = r^*(\sigma|_{S^{n-1}})$ . Then  $\eta$  is closed, since

$$d\eta = r^*(d\sigma|_{S^{n-1}}) = r^*(0) = 0.$$



But  $\eta$  is certainly not exact, because if we had  $\eta = d\psi$ , then since  $i^*(\eta) = \sigma|_{S^{n-1}}$ , we would have

$$d(i^*(\psi)) = i^*(d\psi) = i^*(\eta) = \sigma|_{S^{n-1}},$$

which is impossible since  $\int_{S^{n-1}} \sigma > 0$  (it is positive on any oriented basis).

It will be useful for us to calculate  $\eta$  in cartesian coordinates. But this is simple, since we find that for each  $x \in \mathbb{R}^n - \{0\}$ ,

$$\eta_x = |x|^{-n} \sigma_x.$$

To prove this claim, we consider  $n - 1$  vectors  $v_1, \dots, v_{n-1} \in \mathbb{R}^n$ , and show that for each  $x \in \mathbb{R}^n - \{0\}$ ,

$$\sigma_{r(x)}(r_*((v_1)_x), \dots, r_*((v_{n-1})_x)) = \sigma_x((v_1)_x, \dots, (v_{n-1})_x).$$

We note that  $r_*(x_x) = 0$ , since the curve passing through  $x$  and travelling in the direction of  $x$  becomes constant in the image of the retraction map  $r$ . This implies the left hand side of the equation vanishes if  $v_i$  is a multiple of  $x$ , for any  $i$ . Thus to prove this claim, we may assume that the  $n - 1$  vectors  $v_1, \dots, v_{n-1}$  all lie in the plane perpendicular to  $x$ . The proof will be complete if we can show for such vectors,  $r_*((v_i)_x) = |x|^{-1} v_i$ . We may assume  $v_i$  is a unit vector, but then this is obvious, since by Pythagoras' theorem,

$$\begin{aligned} r(x + tv_i) &= \frac{x + tv_i}{|x + tv_i|} \\ &= \frac{x + tv_i}{\sqrt{|x|^2 + t^2}} \\ &= (x + tv_i) \left( \frac{1}{|x|} + O(t^2) \right) \\ &= r(x) + t(v_i/|x|) + O(t^2), \end{aligned}$$

which implies that  $D_{v_i}(r)(x) = v_i/|x|$ . Thus we conclude

$$\eta_x = \frac{1}{|x|^n} \sum_{i=1}^n (-1)^{i-1} x^i (dx^1 \wedge \dots \wedge \widehat{dx^i} \wedge \dots \wedge dx^n).$$

A simple corollary is that we've proved a 'polar coordinate integration' formula in  $\mathbb{R}^n$ .

**Theorem 6.18.** *Let  $f : B \rightarrow \mathbb{R}$ , where  $B$  is the closed unit ball in  $\mathbb{R}^n$ . Define a function  $g : S^{n-1} \rightarrow \mathbb{R}$  by setting*

$$g(x) = \int_0^1 r^{n-1} f(rx) dr.$$

*Then*

$$\int_B f dx^1 \wedge \dots \wedge dx^n = \int_{S^{n-1}} g \sigma.$$

*Proof.* If  $s : B - \{0\} \rightarrow (0, \infty)$  is given by setting  $s(x) = |x|$ , then we claim that

$$ds \wedge \eta = |x|^{-(n-1)} \cdot dx^1 \wedge \cdots \wedge dx^n.$$

Indeed, we have

$$ds = |x|^{-1} \sum_{i=1}^n x^i dx^i$$

so

$$\begin{aligned} ds \wedge \eta &= \frac{1}{|x|^{n+1}} \sum_{i=1}^n (-1)^{i-1} (x^i)^2 dx^i \wedge (dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n) \\ &= \frac{dx^1 \wedge \cdots \wedge dx^n}{|x|^{n-1}}. \end{aligned}$$

If we consider the diffeomorphism  $g : (0, 1) \times S^{n-1} \rightarrow B - \{0\}$  given by  $g(s, x) = sx$ , then  $g^*(ds) = ds$ , and  $g^*(\eta) = \sigma|_{S^{n-1}}$  since  $(r \circ g)(s, x) = x$ , and  $\eta = r^* \sigma|_{S^{n-1}}$ . Thus we conclude that if  $f$  is compactly supported on the interior of  $B$ , away from the origin, then

$$\begin{aligned} \int_B f dx^1 \wedge \cdots \wedge dx^n &= \int_B f s^{n-1} ds \wedge \eta \\ &= \int_{S^{n-1}} \left( \int_0^1 f(sx) s^{n-1} ds \right) \sigma. \end{aligned}$$

An approximation argument then establishes this result in general.  $\square$

We would like to reduce our computations to coordinate neighbourhoods, but in order to do this, we must introduce another quotient class of forms. We restrict our knowledge to the class  $\Omega_c^k(M)$  of differential forms with compact support, with the corresponding spaces  $Z_c^k(M)$  and  $B_c^k(M)$ , which is the class of forms which is the differential of a form *with compact support*. We therefore get an induced cohomology  $H_c^k(M)$ , known as the *De Rham cohomology with compact support*. Of course, if  $M$  is compact, then  $H_c(M) = H(M)$ .

**Example.** The space  $B_c^k(M)$  is not the same as the class of all exact differentials with compact support. On  $\mathbb{R}^n$ , consider a compactly supported function  $f$  with  $f(x) \geq 0$ , and  $f(x_0) > 0$  at some point  $x_0$ . Now the form  $\omega = f dx^1 \wedge \cdots \wedge dx^n$  is exact, but it is not a differential of an  $n - 1$  form  $\eta$  with compact support, because if this were true, then by Stoke's theorem

$$0 < \int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}^n} d\eta = 0$$

Thus  $H_c^n(\mathbb{R}^n)$  is non trivial, even though  $H^n(\mathbb{R}^n)$  is trivial. A similar argument shows that if  $M$  is any oriented  $n$  manifold, then  $H_c^n(M)$  is nontrivial.

In fact, we now prove  $H_c^n(M)$  is always one dimensional.

**Theorem 6.19.** *If  $M$  is oriented and connected, the map*

$$\omega \mapsto \int_M \omega$$

*induces an isomorphism of  $H_c^n(M)$  to  $\mathbb{R}$ .*

*Proof.* First, take  $M = \mathbb{R}$ . If  $\omega$  is a closed one form on  $\mathbb{R}$ , there exists a function  $f$  such that  $df = \omega$ . If  $\omega$  vanishes outside of  $[a, b]$ , then  $f$  must be constant on  $(-\infty, a]$  and  $[b, \infty)$ . Since

$$f(b) - f(a) = \int_a^b \omega = \int_{\mathbb{R}} \omega,$$

if  $\int_{\mathbb{R}} \omega = 0$ , then  $f(b) = f(a)$ , and so if  $g(x) = f(x) - f(a)$ , then  $g$  has compact support and  $dg = \omega$ . Thus we have proved the theorem in this special case.

Now we prove that if the theorem is true for all  $n - 1$  manifolds, then it must be true for  $\mathbb{R}^n$ . Consider a closed  $n$  form  $\omega$  with compact support on  $\mathbb{R}^n$ . For simplicity we may assume  $\omega$  is supported on the interior of the unit ball centered at the origin. We claim that there exists an  $n - 1$  form  $\eta$  such that  $d\eta = \omega$ . In particular, if  $\omega = f dx^1 \wedge \cdots \wedge dx^n$ , we can actually define  $\eta$  by the formula

$$\begin{aligned} \eta(x) &= \sum_{i=1}^n (-1)^{i-1} \left( \int_0^1 t^{n-1} f(tx) dt \right) x^i \cdot (dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n) \\ &= \left( \int_0^{|x|} t^{n-1} f(t \cdot r(x)) dt \right) \cdot |x|^{-n} \sum_{i=1}^n (-1)^{i-1} x^i \cdot (dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n) \\ &= \left( \int_0^{|x|} t^{n-1} f(t \cdot r(x)) dt \right) \cdot r^*(\sigma|_{S^{n-1}}). \end{aligned}$$

However,  $\eta$  does not have compact support, which we will fix. Define  $g : S^{n-1} \rightarrow \mathbb{R}$  by setting

$$g(x) = \int_0^1 t^{n-1} f(tx) dt.$$

If  $|x| > 1$ ,  $f(x) = 0$ . Thus for such  $x$  we have

$$\eta(x) = g(r(x)) \cdot r^*(\sigma|_{S^{n-1}}) = r^*(g\sigma|_{S^{n-1}}).$$

By the polar coordinates formula,

$$\int_{S^{n-1}} g\sigma = \int f dx^1 \wedge \cdots \wedge dx^n = \int_{\mathbb{R}^n} \omega = 0.$$

By induction, we know that this implies  $g\sigma|_{S^{n-1}}$  is exact, so we can write  $g\sigma|_{S^{n-1}} = d\lambda$  for some  $n - 2$  form  $\lambda$  on  $S^{n-1}$ . But this means that

$$\eta = r^*(d\lambda) = d(r^*\lambda)$$

If we choose  $h \in C^\infty(M)$  such that  $h(x) = 1$  for  $|x| \geq 1$  and  $h(x) = 0$  in a neighbourhood of the origin, then

$$\omega = d\eta = d(\eta - d(hr^*\lambda)),$$

The form  $\eta - d(hr^*\lambda)$  has compact support since for  $|x| > 1$ ,

$$\eta - d(hr^*\lambda) = \eta - d(r^*\lambda) = 0.$$

Thus we have established the result in this case.

Finally we show that if the theorem is true for  $\mathbb{R}^n$ , then the theorem is true for all connected oriented  $n$  dimensional manifolds. Let  $M$  be an  $n$  dimensional manifold, and consider some  $n$ -form  $\omega$  on  $M$  with

$$\int_M \omega \neq 0,$$

and such that  $\omega$  is compactly supported on some open set  $U \subset M$  diffeomorphic to  $\mathbb{R}^n$ . We must show that for any other compactly supported  $n$  form  $\omega'$ , there exists a constant  $c$  and a compactly supported  $n-1$  form  $\lambda$  such that  $\omega = \omega' + d\lambda$ . Using a partition of unity argument, we may assume without loss of generality that  $\omega'$  is also compactly supported on an open set  $V$  diffeomorphic to  $\mathbb{R}^n$ . Since  $M$  is connected, there is a sequence of open sets  $U_0, \dots, U_N$  with  $U_i \cap U_{i+1} \neq \emptyset$  for each  $i$ , with  $U_0 = U$  and  $U_N = V$ . We may then choose forms  $\omega_i$  supported on  $U_i \cap U_{i+1}$  with  $\int_{U_i} \omega_i \neq 0$ . Then since we are assuming the theorem holds on  $\mathbb{R}^n$ , we can find constants  $c_i$  such that

$$\omega_0 - c_0\omega = d\eta_0$$

for  $i \in \{1, \dots, N\}$ ,

$$\omega_i - c_i\omega_{i-1} = d\eta_i,$$

and such that  $\omega' - c_{N+1}\omega_N = d\eta_{N+1}$ . Adding up these inequalities gives the desired result.  $\square$

One can use the chain of sets argument at the end of the last proof to establish another result.

**Theorem 6.20.** *If  $M$  is a connected non-orientable  $n$  manifold,  $H_c^n(M) = 0$ .*

*Proof.* Let  $\omega$  be a compactly supported  $n$ -form on  $M$ , supported on an open set  $U$  diffeomorphic to  $\mathbb{R}^n$  by a map  $x_0 : U \rightarrow \mathbb{R}^n$ , and such that

$$\int_U \omega > 0;$$

the integral being defined after choosing an orientation for  $U$  induced by  $x_0$ . It suffices to show that  $\omega = d\eta$  for some form  $\eta$  with compact support. If  $M$  is non-orientable, there is a sequence  $U_0, \dots, U_N$  with  $U_0 = U_N = U$ ,  $U_i \cap U_{i+1} \neq \emptyset$  for each  $i \in \{1, \dots, N-1\}$ , and each  $U_i$  is diffeomorphic to  $\mathbb{R}^n$  by a map  $x_i : U_i \rightarrow \mathbb{R}^n$  such that  $x_{i+1} \circ x_i^{-1}$  is orientation

preserving for each  $i \in \{0, \dots, N-1\}$ , and such that  $x_1 \circ x_N^{-1}$  is orientation reversing. Without loss of generality, we may assume  $\omega$  is supported on  $U_0 \cap U_1$ , and that

$$\int_{U_0} \omega > 0$$

We set  $\omega_0 = \omega$ , set  $\omega_N = -\omega$ , and for each  $i \in \{1, \dots, N-1\}$ , we pick an  $n$ -form  $\omega_i$  supported on  $U_i \cap U_{i+1}$  such that

$$\int_{U_i} \omega_i > 0,$$

where  $U_i$  is given the orientation induced by  $x_i$ . Since  $x_{i+1} \cap x_i^{-1}$  is orientation preserving for each  $i \in \{0, \dots, N\}$ , we find

$$\int_{U_{i+1}} \omega_i = \int_{U_i} \omega_i > 0.$$

For each  $i \in \{0, \dots, N-1\}$ , since  $H_c^n(U_i) = H_c^n(\mathbb{R}^n) = 0$ , we can find  $c_i \in \mathbb{R}$  and a compactly supported  $n-1$  form  $\eta_i$  such that  $\omega_{i+1} = c_i \omega_i + d\eta_i$ , and

$$c_i = \frac{\int_{U_{i+1}} \omega_{i+1}}{\int_{U_{i+1}} \omega_i} = \frac{\int_{U_{i+1}} \omega_{i+1}}{\int_{U_i} \omega_i} > 0.$$

Thus we conclude that there is  $c > 0$  and a compactly supported  $n-1$  form  $\eta$  such that

$$-\omega = c\omega + d\eta$$

But this means that  $\omega$  is exact, because  $\omega = d(-\eta/(c+1))$ .  $\square$

**Theorem 6.21.** *If  $M$  is a connected non-compact  $n$ -manifold,  $H^n(M) = 0$ .*

*Proof.* Let  $\omega$  be an  $n$ -form. We begin by assuming the support of  $\omega$  is compactly contained in an open set  $U$  diffeomorphic to  $\mathbb{R}^n$ . We consider a special cover of  $M$ , consisting of a sequence of open sets  $\{U_i\}$  such that each  $U_i$  is diffeomorphic to  $\mathbb{R}^n$ ,  $U_i \cap U_{i+1} \neq \emptyset$  for each  $i$ , and for any compact set  $K$ , it is eventually true that  $K \cap U_i = \emptyset$ . We set  $\omega_0 = \omega$ , and then choose  $n$ -forms  $\omega_i$  compactly supported on  $U_i \cap U_{i+1}$ , with  $\int_{U_i} \omega_i \neq 0$ . Then there are constants  $c_i$  and  $n-1$  forms  $\eta_i$  each compactly supported in  $U_i$  for each  $i$  such that  $\omega_i = c_i \omega_{i+1} + d\eta_i$ . By induction, it therefore follows that

$$\omega = \sum_{i=0}^{\infty} c_0 \dots c_{i-1} d\eta_i = d \left( \sum_{i=0}^{\infty} c_0 \dots c_{i-1} \eta_i \right),$$

which is well defined by local compactness, since  $d\eta_i$  eventually vanishes on every compact set. But now a partition of unity argument shows that this type of argument works for all forms, not necessarily just compactly supported forms.  $\square$

We have thus computed the De Rham cohomology in a few useful situations:

- If  $M$  is a connected, smoothly contractible manifold, then  $H^k(M)$  is trivial for  $k > 0$ , and  $H^0(M)$  is one dimensional.
- If  $M$  is a connected  $n$ -dimensional manifold, then  $H^0(M)$  is one dimensional, and

$$\dim(H^n(M)) = \begin{cases} 1 & : \text{if } M \text{ is compact and orientable,} \\ 0 & : \text{if } M \text{ is non-compact or non-orientable.} \end{cases}$$

and

$$\dim(H_c^n(M)) = \begin{cases} 1 & : \text{if } M \text{ is orientable,} \\ 0 & : \text{if } M \text{ is non-orientable.} \end{cases}$$

Since one only ever integrates  $n$  dimensional forms over a manifold, these results are the most useful for applying De-Rham cohomology to integration theory. In the next section, we apply this result to obtain a useful formula for integration of pullbacks of differential forms, known as the *degree formula*.

A smooth map  $f : M \rightarrow N$  between two manifolds induces a map

$$f^* : \Gamma(\Omega^k(TN)) \rightarrow \Gamma(\Omega^k(TM))$$

for each  $k$ . Since  $f^*(d\eta) = d(f^*\eta)$ ,  $f^*(Z^k(N)) \subset Z^k(M)$ , and  $f^*(B^k(N)) \subset B^k(M)$ . In particular,  $f^*$  descends to a map from  $H^k(N)$  to  $H^k(M)$ .

**Example.** Consider the simplest example where  $k = 0$ . Then  $H^0(N)$  can be identified with the family of functions on  $N$  which are constant on each component. If  $f : M \rightarrow N$  is smooth, and  $u \in C^\infty(N)$  is constant on each fibre, then  $f^*u = u \circ f$  is constant on each fibre, and is therefore an element of  $H^0(M)$ . Knowledge of the behaviour of  $f^*$  from  $H^0(M)$  to  $H^0(N)$  thus gives us the description of which components of  $M$  map to which components of  $N$ .

Let us suppose that  $M$  and  $N$  are both compact, connected, oriented  $n$ -manifolds, and we consider the map  $f^* : H^n(N) \rightarrow H^n(M)$ . Since  $H^n(N)$  and  $H^n(M)$  are both one dimensional, and  $H^n(M)$  and  $H^m(N)$  can be naturally identified with  $\mathbb{R}$  under the maps

$$\omega \mapsto \int_M \omega \quad \text{and} \quad \omega \mapsto \int_N \omega,$$

there must exist a constant  $\alpha$ , depending only on the map  $f$  such that for each closed  $n$ -form  $\omega$  on  $N$ ,

$$\int_M f^*(\omega) = \alpha \int_N \omega.$$

This constant  $\alpha$  is known as the *degree* of the map  $f$ , denoted  $\deg(f)$ . Thus for any  $n$ -form  $\omega$  on  $N$ ,

$$\int_M f^*(\omega) = (\deg f) \int_N \omega.$$

We now show that, perhaps surprisingly,  $\deg(f)$  is *always* an integer.

**Theorem 6.22.** Let  $f : M \rightarrow N$  be a proper map between two compact, connected, oriented  $n$  manifolds, and let  $q \in N$  be a regular value of  $f$ . For each  $p \in f^{-1}(q)$ , let

$$\text{sgn}_p(f) = \begin{cases} 1 & : f_*|_p : T_p M \rightarrow T_p N \text{ is orientation preserving} \\ -1 & : f_*|_p : T_p M \rightarrow T_p N \text{ is orientation reversing.} \end{cases}$$

Then

$$\deg(f) = \sum_{p \in f^{-1}(q)} \text{sgn}_p(f),$$

where the right hand side is equal to zero if  $f^{-1}(q) = \emptyset$ , and is well defined since  $f^{-1}(q)$  is finite because  $M$  is compact. In particular, this quantity is invariant of the regular value chosen (and by Sard's theorem, regular values always exist).

*Proof.* Let  $f^{-1}(q) = \{p_1, \dots, p_N\}$ . Since  $q$  is a regular value, we may choose  $N$  disjoint coordinate charts  $(x_1, U_1), \dots, (x_N, U_N)$  with  $p_i \in U_i$  for each  $i$ , and a chart  $(y, V)$  containing  $q$  such that  $y \circ f \circ x_i^{-1} : x_i(U_i) \rightarrow y(V)$  is a diffeomorphism for each  $i$ . Set  $\omega = g dy^1 \wedge \dots \wedge dy^n$ , where  $g$  is a non-negative function with compact support contained in  $V$ . Then

$$\int_M f^* \omega = \sum_{i=1}^N \int_{U_i} f^* \omega.$$

But for each  $i$ , because  $f$  is a diffeomorphism,

$$\int_{U_i} f^* \omega = \begin{cases} \int_V \omega & : f \text{ is orientation preserving} \\ -\int_V \omega & : f \text{ is orientation reversing} \end{cases}. \quad \square$$

**Example.** Consider the antipodal map  $f : S^n \rightarrow S^n$  given by setting  $f(x) = -x$ . Then  $f$  is a diffeomorphism, and hence either orientation preserving or orientation reversing. Since  $f_*(v_x) = (-v)_{-x}$ , and so  $f_*$  takes the oriented basis  $\{(e_2)_{e_1}, \dots, (e_{n+1})_{e_1}\}$  to the basis  $\{(-e_2)_{-e_1}, \dots, (-e_{n+1})_{-e_1}\}$ . Since

$$\det(-e_1, -e_2, \dots, -e_{n+1}) = (-1)^{n+1},$$

we conclude  $f$  is orientation preserving if  $n$  is odd, and orientation reversing if  $n$  is even. Since  $f^{-1}(p)$  consists of a single point for each  $p \in S^n$ , this means that the degree of  $f$  is  $-1$  if  $n$  is even, and  $1$  if  $n$  is odd.

Two maps  $f, g : M \rightarrow N$  are *smoothly homotopic* if there exists a smooth map  $H : [0, 1] \times M \rightarrow N$  such that  $H(0, p) = f(p)$ , for each  $p \in M$ , and  $H(1, p) = g(p)$  for each  $p \in N$ . Notice that  $M$  is *smoothly contractible* if and only if the identity map  $i : M \rightarrow M$  is smoothly homotopic to a constant map. Recall that for each  $k$ -form  $\omega$  on  $[0, 1] \times M$ , we constructed a  $k-1$  form  $I\omega$  on  $M$  such that

$$i_1^* \omega - i_0^* \omega = d(I\omega) + I(d\omega).$$

This result implies a more general result.

**Theorem 6.23.** *If  $f, g$  are smoothly homotopic, then for each  $k$ , the two maps  $f^*, g^* : H^k(N) \rightarrow H^k(M)$  are equal.*

*Proof.* Then if  $\omega$  is any closed  $k$ -form on  $N$ ,

$$g^* \omega - f^* \omega = d(I\omega),$$

so  $g^* \omega$  is equal to  $f^* \omega$  in  $H^k(M)$ .  $\square$

**Corollary 6.24.** *If  $M$  and  $N$  are compact oriented  $n$  manifolds, and  $f, g : M \rightarrow N$  are smoothly homotopic, then  $\deg(f) = \deg(g)$ .*

**Corollary 6.25.** *If  $n$  is even, there are no globally non-zero vector fields on  $S^n$ .*

*Proof.* Since the degree of the identity map on  $S^n$  is equal to one, and the degree of the antipodal map  $x \mapsto -x$  is equal to  $-1$ , this implies the identity map is not smoothly homotopic to the antipodal map. If  $X$  is an everywhere non-zero vector field on  $S^n$ , then we can construct a homotopy  $H : [0, 1] \times M \rightarrow M$  between the identity map and the antipodal map by setting  $H(t, p)$  to be the point  $\pi t$  radians along the unique great circle connecting  $p$  to  $-p$  which travels in the same direction as  $X_p$ . This gives a contradiction.  $\square$

*Remark.* If  $n$  is odd, we can construct an everywhere non-zero vector field  $X$  on  $S^n$  such that if  $p = (x_1, \dots, x_{n+1})$ ,

$$X_p = (-x_2, x_1, -x_4, x_3, \dots, -x_{n+1}, x_n).$$

In particular, the antipodal map is homotopic to the identity map if  $n$  is odd.

As another application, consider the retraction  $r : \mathbb{R}^n - \{0\} \rightarrow S^{n-1}$  given by setting  $r(x) = x/|x|$ . Then if  $i : S^{n-1} \rightarrow \mathbb{R}^n - \{0\}$  is the inclusion, then  $r \circ i$  is the identity map in  $S^{n-1}$ . Conversely,  $i \circ r$  is *not* the identity map on  $\mathbb{R}^n - \{0\}$ . On the other hand, it is homotopic to the identity, which we can define by setting

$$H(p, t) = tp + (1 - t)r(p).$$

A retraction  $r$  such that  $i \circ r$  is homotopic to the identity map is known as a *deformation retraction*. Thus  $(r \circ i)^*$  and  $(i \circ r)^*$  act as the identity maps on  $H^k(S^{n-1})$  and  $H^k(\mathbb{R}^n - \{0\})$  for each  $k$ . Thus  $r^*$  and  $i^*$  are inverses of one another, so in particular,  $H^k(S^{n-1})$  and  $H^k(\mathbb{R}^n - \{0\})$  are *isomorphic* to one another. In particular, this implies  $H^{n-1}(\mathbb{R}^n - \{0\})$  is 1-dimensional - a generator for this space being the closed form  $r^* \sigma$  defined earlier in this chapter.

**Theorem 6.26.** *For  $0 < k < n - 1$ ,  $H^k(\mathbb{R}^n - \{0\}) = H^k(S^{n-1}) = 0$ .*

*Proof.* We prove the theorem by induction on  $n$ , the claim being obvious for  $n < 3$ . For  $n = 3$ , we claim  $H^1(\mathbb{R}^3 - \{0\})$  is trivial. Let  $\omega$  be a closed one-form on  $\mathbb{R}^3 - \{0\}$ . Let

$$A = \mathbb{R}^3 - \{(0, 0) \times (-\infty, 0]\} \quad \text{and} \quad B = \mathbb{R}^3 - \{(0, 0) \times [0, \infty)\}.$$

Then  $A$  and  $B$  are both contractible (they are both star shaped), so there exists smooth functions  $f_A$  and  $f_B$  on  $A$  and  $B$  respectively such that  $\omega = df_A$  on  $A$ , and  $\omega = df_B$  on  $B$ . In



particular,  $df_A - df_B = 0$  on  $A \cap B$ . Since  $A \cap B = (\mathbb{R}^2 - \{0\}) \times \mathbb{R}$ , which is connected, this implies  $f_A - f_B$  is a constant  $c$  on  $A \cap B$ . But this means that  $\omega$  is exact, for  $\omega = d(f_A - c)$  on  $A$ , and  $\omega = df_B$  on  $B$ , and  $f_A - c$  and  $f_B$  agree on  $A \cap B$ .

In general, consider  $\mathbb{R}^n - \{0\}$ . The case of one-forms is similar. We define

$$A = \mathbb{R}^n - \{\{0\} \times (-\infty, 0]\} \quad \text{and} \quad B = \mathbb{R}^n - \{\{0\} \times [0, \infty)\}.$$

Use contractibility to show closed one-form are the differentials of functions on  $A$  and  $B$ , and then shift by a constant. If  $k \geq 2$ , and  $\omega$  is a  $k$ -form on  $\mathbb{R}^n - \{0\}$ , since  $A$  and  $B$  are star-shaped, there exists two  $k-1$  forms  $\eta_A$  and  $\eta_B$  on  $A$  and  $B$  such that  $\omega = d\eta_A$  on  $A$ , and  $\omega = d\eta_B$  on  $B$ . Since  $A \cap B = (\mathbb{R}^{n-1} - \{0\}) \times \mathbb{R}$ , which is homotopic to  $\mathbb{R}^{n-1} - \{0\}$ , or to  $S^{n-2}$ , we have by induction that  $H^{k-2}(S^{n-2})$  is zero dimensional. Since  $d(\eta_A - \eta_B) = 0$  on  $A \cap B$ , there exists a  $k-2$  form  $\lambda$  on  $A \cap B$  such that  $\eta_A - \eta_B = d\lambda$ .

Unlike the previous case, we cannot extend  $\eta_A - d\lambda$  to a  $k-2$  form on  $A$ , since  $\lambda$  is only defined on  $A \cap B$ . To circumvent this, we consider a partition of unity  $\{\phi_A, \phi_B\}$  subordinate to  $\{A, B\}$ . Then the form  $\phi_A \lambda$  and  $\phi_B \lambda$  extend to forms on  $A$  and  $B$ . On  $A \cap B$ , we have

$$\begin{aligned} \eta_A - d(\phi_B \lambda) &= \eta_A - d\phi_B \cdot \lambda - \phi_B d\lambda \\ &= (\eta_B + d\lambda) + d\phi_A \cdot \lambda - (1 - \phi_A)d\lambda \\ &= \eta_B + d\phi_A \cdot \lambda + \phi_A \cdot d\lambda \\ &= \eta_B + d(\phi_A \lambda). \end{aligned}$$

Thus we can define a form  $\eta$  by setting it equal to  $\eta_A - d(\phi_B \lambda)$  on  $A$ , and equal to  $\eta_B + d(\phi_A \lambda)$  on  $B$ . Then  $d\eta = d\eta_A = \omega$  on  $A$ , and  $d\eta = d\eta_B = \omega$  on  $B$ . So  $\omega$  is exact, and so we conclude that  $H^k(\mathbb{R}^n - \{0\})$  is trivial.  $\square$

We end this chapter by finishing our calculation of the homology groups  $H_c^k(\mathbb{R}^n)$ .

**Theorem 6.27.** *For  $0 \leq k < n$ , we have  $H_c^k(\mathbb{R}^n) = 0$ .*

*Proof.* We begin by showing  $H_c^0(\mathbb{R}^n) = 0$ . Indeed, if  $f$  is a compactly supported function on  $\mathbb{R}^n$  with  $df = 0$ , then  $f$  is constant, and since the function is compactly supported,  $f = 0$ . So there are no nontrivial compactly supported one-forms. For  $0 < k < n$ , we consider a closed compactly supported  $k$ -form  $\omega$  on  $\mathbb{R}^n$ . There certainly exists a  $k-1$  form  $\eta$  such that  $d\eta = \omega$ . We just do not know if we can find a compactly-supported  $k-1$  form with this property. Let  $B$  be a closed ball containing  $\omega$ . Then we know  $d\eta = 0$  on  $\mathbb{R}^n - B$ , which is homotopic to  $\mathbb{R}^n - \{0\}$ . This means that on  $\mathbb{R}^n - B$ ,  $\eta = d\lambda$  for some  $k-2$  form  $\lambda$ . If  $f : \mathbb{R}^n \rightarrow [0, 1]$  is a smooth function with  $f(x) = 0$  on a neighbourhood of  $B$ , and  $f(x) = 1$  on  $\mathbb{R}^n - 2B$ , where  $2B$  is the ball with the same centre of  $B$  but twice the radius, then  $f\lambda$  is a well-defined form on  $\mathbb{R}^n$ , and on  $\mathbb{R}^n - 3B$ ,  $d(f\lambda) = d(\lambda) = \eta$ . Thus if we define  $\eta' = \eta - d(f\lambda)$ , then  $\eta'$  is compactly supported on  $3B$ , and  $d(\eta') = d\eta = \omega$ .  $\square$

# Chapter 7

## Jets

If  $(x, U)$  is a chart on  $M$ , which identifies  $TM|_U$  with  $U \times \mathbb{R}^n$ , i.e.  $(x, v)$  coordinates for  $TM|_U$ . Given some other chart  $(y, U)$ , with induced coordinates  $(y, w)$  on  $TM|_U$ , if  $\tau = y \circ x^{-1}$ , then we have

$$y = \tau(x) \quad \text{and} \quad w = (D\tau)(x)\{v\}.$$

If we now consider the induced identification of  $T(TM)|_U$  with  $U \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$  induced by  $(x, v)$ , i.e. a coordinate system  $(x, v, \dot{x}, \dot{v})$ , then given some other coordinate system  $(y, w, \dot{y}, \dot{w})$ , if we define a matrix  $(D^2\tau)(x, v)$  by letting  $(D^2\tau)_{jk} = \sum (\partial_\alpha \partial_k \tau^j) v_\alpha$ , then

$$\begin{bmatrix} \dot{y} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} (D\tau)(x) & 0 \\ (D^2\tau)(x, v) & (D\tau)(x) \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix}.$$

Given  $p \in M$ , and two tangent vectors  $v, w \in T_p M$ , we can define a tangent vector  $\text{vlift}_v(w)$  in  $T_{v_p}(TM)$ , the *vertical lift* of  $w$  over  $v$ , to be the tangent vector to the curve  $\gamma(t) = v + tw$  when  $t = 0$ . In coordinates, if  $v$  and  $w$  are given in coordinates by  $(x, a)$  and  $(x, b)$ , then the vertical lift is given in coordinates by  $(x, a, 0, b)$ , and it is simple to check this definition is independent of the coordinate system because the top right of the matrix above vanishes. A *vector field*  $X$  on  $M$  induces a flow  $\{\Phi_t^X\}$ , and we define the *tangent lift* of  $X$  at  $v_p \in T_p M$  to be the tangent vector to the curve  $\gamma(t) = (\Phi_t^X)_* \{v_p\}$  when  $t = 0$ . In coordinates, if the vector field  $X$  is given in coordinates by  $(x, a(x))$ , then the tangent lift is given in coordinates by  $(x, v, a(x), Da(x)\{v\})$ .

It follows that if two vectors in  $T_p M$  are given in coordinates by  $(x(p), v)$  and  $(x(p), w)$ , then the vector  $(x(p), v, 0, w)$  is independent of the coordinate system, and is called the *vertical lift* of  $w$  over  $v$ .

$$(x(p), v, 0, w)\{f\} = (D_2 f)\{w\}$$

The vector field on  $TM|_U$  given in coordinates  $(x, v) \mapsto (x, v, 0, v)$  is thus independent of the coordinate system, and we therefore have a natural vector field on  $TM$ , called the *vertical lift* of  $TM$ .

In this chapter, we study the *jet bundle* of a manifold  $M$ , which is a bundle which allows us to capture the  $k$ th order behaviour of functions in  $C^\infty(M)$  at points. The first order behaviour of a function  $f \in C^\infty(M)$  is captured by the *cotangent bundle* of  $M$ , i.e. the first

order behaviour is captured by the section  $df \in \Gamma(TM)$ . We wish to extend this correspondence to higher order behaviour. It turns out it is more natural to do this by capturing the behaviour of a function *up to order  $k$*  rather than the  $k$ th order behaviour, since this latter behaviour is difficult to capture in a coordinate invariant manner for  $k > 1$ .

For  $k \geq 0$ , we will define a bundle  $J^k(M)$ , such that each smooth function  $f \in C^\infty(M)$  is associated with a smooth section  $j^k f \in \Gamma(J^k M)$ , and such that  $J^k f(p) = J^k g(p)$  if and only if  $f$  and  $g$  agree up to order  $k$  at  $p$ . This immediately indicates how we can define the bundle. We let  $I$  be the ideal of all functions vanishing at  $p$ , and define

$$J_p^k(M) = C^\infty(M)/I^{k+1}.$$

It is easy to verify that  $J_p^k(M)$  is a finite dimensional vector space by using a Taylor expansion; indeed, in coordinates  $(x, U)$ , elements are given by sums of the form

$$\sum_{|\alpha| \leq k} c_\alpha dx^\alpha,$$

where we let  $dx^\alpha$  denote the section of  $J_p^k(M)$  corresponding to the equivalence class of the function  $(x - x(p))^\alpha$ . The  $k$  jet of a smooth function  $f$  is then precisely given by

$$j^k f(p) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} \cdot \partial_x^\alpha f(p) \cdot dx^\alpha.$$

If we give  $J^k(M)$  the structure of a smooth bundle such that each of these expressions in coordinates are smooth trivializations, then  $j^k f$  is then a smooth section of  $J^k(M)$ .

One use of jets is to define a coordinate independent theorem of partial differential equations on a smooth manifold. Indeed, a partial differential operator  $L$  of order at most  $k$  can be identified with a map  $\mathcal{L} : J^k(M) \rightarrow \varepsilon^1(M)$  preserving basepoints – we can then consider  $L : C^\infty(M) \rightarrow C^\infty(M)$  by setting  $Lf = \pi \circ \mathcal{L} \circ j^k f$ , where  $\pi : \varepsilon^1(M) \rightarrow \mathbb{R}$  is the natural projection. The operator  $L$  is linear precisely when  $\mathcal{L}$  is a bundle morphism, for then, in coordinates  $(x, U)$ , we can find smooth functions  $c_\alpha \in C^\infty(U)$  such that

$$\mathcal{L}(dx^\alpha) = c_\alpha,$$

and then

$$Lf(p) = \sum_{\alpha} \frac{c_\alpha(p)}{\alpha!} (\partial_x^\alpha f)(p).$$

We note that the family of linear partial differential operators can be identified with the dual bundle to  $J^k(M)$ .

We note that it is difficult to define the  $k$ th order part of a jet, since this part is not stable under coordinate changes. However, it is more natural to do this on a *Riemannian manifold*.  
TODO: WHY?

# Chapter 8

## Lie Groups

A *Lie group* is a group whose multiplication and inversion operations are smooth. Examples include the group  $\mathbb{R}^n$  under addition, the circle group  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ , and more generally, the toral groups  $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ . The main noncommutative examples occur as matrix groups, like  $GL_n(\mathbb{R})$ ,  $SL_n(\mathbb{R})$ ,  $O_n(\mathbb{R})$  and  $SU_n(\mathbb{R})$ .

An important fact about  $M(n)$  is that matrix multiplication is a differentiable operation in the entries of the matrices (it is a polynomial in the entries), hence continuous. By Cramer's rule, the operation mapping  $M$  to  $M^{-1}$  is also continuous and differentiable function in the entries of the matrix, because it is a rational function of the matrix elements, which doesn't have any singularities in  $GL(n)$ .

### 8.1 A Basic Example: $SO(3)$

In classical physics, the most important Lie group is the rotation group  $SO(3)$ , the space of  $3 \times 3$  matrices  $A$  such that  $A^T = -A$ . The tangent spaces of this function have an interesting algebraic structure. Consider a curve  $A(t) \in SO(3)$ . Then, differentiating the identity  $A(t)A(t)^T = I$ , we conclude that

$$A'(t)A(t)^T + A(t)A'(t)^T = 0.$$

Thus  $A'(t) = -A(t)A'(t)^T A(t)$ , and moreover,  $A(t)A'(t)^T$  is a skew symmetric matrix. Thus the tangent space to  $SO(3)$  at the identity matrix is identified with the space of all skew symmetric matrices, the collection of which we denote by  $\mathfrak{so}(3)$ . More generally, the tangent space to  $SO(3)$  at a general element  $A \in SO(3)$  can be naturally identified with  $\mathfrak{so}(3) \cdot A$ . The space  $\mathfrak{so}(3)$  is no longer a group. But the tangent space structure gives it a vector space structure. We will soon give it an algebraic multiplication structure which 'infinitesimally' characterizes the group  $SO(3)$ .

## 8.2 General Theory

Any subgroup of a Lie group which is also a submanifold is a Lie group, because the group structure on the subgroup is just the restriction of the group structure on the entire group, which is differentiable. This is essentially the argument we used to show that  $SL_n(\mathbb{R})$ ,  $O_n(\mathbb{R})$ , and  $SU_n(\mathbb{R})$  are Lie groups in previous parts of the notes. We could have also used the fact that  $S^1$  is a subgroup of the multiplicative group of non-zero complex numbers to show it was a Lie group.  $S^3$  is a Lie group, because it is a subgroup of the Lie group of quaternions, consisting of elements of norm one. More generally, we define a *Lie subgroup* of a Lie group to be a subgroup, with some  $C^\infty$  structure making the operations of the subgroup differentiable, and such that the  $C^\infty$  structure makes the inclusion of the subgroup an immersion. As an example of a subgroup that is not an imbedded submanifold, consider the set of points  $(x, cx) \in \mathbb{T}^2$ , where  $c$  is an irrational number.

Lie groups are incredibly useful in geometry, because we often want to consider some symmetries which occur in a problem, which often turn out to be a group with some Lie structure. As an example, suppose we are discussing the metric structure of  $\mathbb{R}^n$ . In this situation, it is natural to discuss the Euclidean group  $E_n$ , which is the collection of all isometries of  $\mathbb{R}^n$ . The easiest case to analyze is the group  $E_1$ . For each  $x \in \mathbb{R}$ , and  $z \in \{-1, 1\}$ , define  $T_{xz}(t) = x + zt$ . Every  $T \in E_1$  can be written as  $T_{xz}$  for some  $x$  and some  $z$ . To see this, let  $x = T(0)$ . Since  $T$  is an isometry, either  $T(1) = x + 1$  or  $T(1) = x - 1$ , because  $|T(1) - x| = 1$ . If  $T(1) = x + 1$ , then  $T(y) = x + y$ , because  $x + y$  is the only number satisfying

$$|T(y) - x| = |y| \quad |T(y) - (x + 1)| = |y - 1|$$

Similarly, if  $T(1) = y - 1$ , then  $T(x) = y - x$ . Since

$$T_{x_0 z_0} \circ T_{x_1 z_1} = T_{(x_0 + z_0 x_1)(z_0 z_1)}$$

and so  $E_1$  can be described as the *semidirect product* of the multiplicative group  $\{-1, 1\}$  and  $\mathbb{R}$  under the representation  $\rho : \{-1, 1\} \rightarrow \text{Aut}(\mathbb{R})$  defined by  $\rho(x)(y) = -y$ . For  $E_2$ , we note that if  $T : \mathbb{C} \rightarrow \mathbb{C}$  is an isometry such that  $T(0) = 0$ , and  $T(1) = 1$ , then  $|T(i)| = 1$  and  $|T(i) - 1| = \sqrt{2}$ , so either

- $T(i) = 1$ , which implies  $T(z) = z$  for all  $z \in \mathbb{C}$ , because  $z$  is uniquely specified by the values  $|z|$ ,  $|z - 1|$ , and  $|z - i|$ .
- $T(i) = -1$ , which implies  $T(z) = \bar{z}$  for all  $z \in \mathbb{C}$ , because  $\bar{z}$  is the unique point with  $|\bar{z}| = |z|$ ,  $|\bar{z} - 1| = |z - 1|$ , and  $|\bar{z} - (-i)| = |z - i|$ .

If  $z \in \mathbb{C}$ , and  $w \in \mathbb{T}$ , we define  $T_{zw}$  to be the isometry  $T(u) = z + wu$ . If  $T \in E_2$  is arbitrary, and if  $T(0) = u$ ,  $T(1) = w$ , then  $T_{uw}^{-1} \circ T$  maps 0 to 0, and 1 to 1, hence either  $T = T_{uw}$  or  $T = \overline{T_{uw}}$ . Note that  $\overline{T_{uw}}(z) = T_{\overline{uw}}(\bar{z})$ , so that the set of all  $T_{uw}$  is a normal subgroup of  $E_2$ . Since the group of all  $T_{uw}$  is isomorphic to the semidirect product  $\mathbb{C} \rtimes \mathbb{T}$ , because  $T_{u_0 w_0} \circ T_{u_1 w_1} = T_{(u_0 + w_0 u_1)(w_0 w_1)}$ , and therefore  $E_2$  is isomorphic to the semidirect product

$(\mathbf{C} \rtimes \mathbb{T}) \rtimes \{-1, 1\}$  with multiplication law

$$(z_0, w_0, t_0)(z_1, w_1, t_1) = \begin{cases} (z_0 + \overline{w_0 z_1}, w_0 \overline{w_1}, t_0 t_1) & t_0 = -1 \\ (z_0 + w_0 z_1, w_0 w_1, t_0 t_1) & t_0 = 1 \end{cases}$$

The fact that  $\{-1, 1\}$  is isomorphic to  $O_1$ , and  $\{-1, 1\} \rtimes \mathbb{T}$  is isomorphic to  $O_2$  hints at a more general fact about the structure of the Euclidean groups, but we need to know some structure of the metric of  $\mathbb{R}^n$  first. Say a point  $x$  lies *between* two points  $y$  and  $z$  if  $x = \lambda y + (1 - \lambda)z$  for  $0 \leq \lambda \leq 1$ . This holds if and only if  $\|y - x\| + \|x - z\| = \|y - z\|$ , because if  $x = \lambda y + (1 - \lambda)z$ , then

$$\|y - x\| + \|x - z\| = \|(1 - \lambda)y - (1 - \lambda)z\| + \|\lambda y - \lambda z\| = \|y - z\|$$

and the Cauchy Schwartz inequality implies that this inequality occurs only when  $y - x = \lambda(x - z)$  for some  $\lambda > 0$ , in which case we find

$$x = \left( \frac{1}{1 + \lambda} \right) y + \left( \frac{\lambda}{\lambda + 1} \right) z$$

We say  $x, y, z$  are colinear if one point lies between the other pair of points. This occurs if and only if  $y - x$  and  $z - x$  are linearly dependant, because if  $\lambda(y - x) = (z - x)$ , then

- If  $\lambda < 0$ , then  $\|y - x\| + \|x - z\| = \|y - z\|$ , so  $x$  lies between  $y$  and  $z$ .
- If  $0 < \lambda < 1$ , then  $z$  lies between  $x$  and  $y$ , because  $z = \lambda y + (1 - \lambda)x$ .
- If  $\lambda > 1$ , then  $y$  lies between  $x$  and  $z$ , because

$$y = \frac{1}{\lambda} z + \frac{\lambda - 1}{\lambda} x$$

This tells us that an isometry maps straight lines to straight lines, because betweenness and colinearity are purely metric conditions, hence preserved by an isometry, and a line can be described by a set of points such that any triple of points is colinear. Furthermore, an isometry maps planes to planes, because a plane can be described as the smallest set containing a triple of non-colinear points  $x, y, z$ , and also containing the line generated by any points in the set. We claim that this plane is the set of points

$$\{x + \lambda(y - x) + \gamma(z - x) : \lambda, \gamma \in \mathbb{R}\}$$

If  $x + \lambda_0(y - x) + \gamma_0(z - x)$  and  $x + \lambda_1(y - x) + \gamma_1(z - x)$  are two points in this plane, then the set of points on the line between these two points is exactly  $x + (\lambda_0 + t\lambda_1)(y - x) + (\gamma_0 + t\gamma_1)(z - x)$ , as  $t$  ranges over all real numbers, and these points all lie in the set above. Conversely, if  $X$  is any colinearly closed set containing  $x, y$ , and  $z$ , then  $x + t_0(y - x)$  and  $x + t_1(z - x)$  are elements of  $X$ , for all  $t \in \mathbb{R}$ , and therefore

$$x + t_0(y - x) + t_2(t_1(z - x) - t_0(y - x)) = x + t_0(1 - t_2)(y - x) + t_2 t_1(z - x)$$

are also points in  $X$ , for all  $t_0, t_1, t_2 \in \mathbb{R}$ . This implies that the set of all points  $x + \lambda(y - x) + \gamma(z - x)$  are contained in  $X$ . Since colinearity is a metric notion, an isometry maps planes to planes.

Now suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is an isometry, with  $T(0) = 0$ . Our discussion implies that  $T$  maps lines through the origin to lines through the origin. Thus  $T(cx) = cT(x)$ , because  $cT(x)$  is the only point on the line through the origin and  $x$  which lies at a distance  $|c|\|x\|$  from the origin and a distance  $|c - 1|\|x\|$  from  $x$ . Similarly, for a fixed  $x, y \in \mathbb{R}^n$ , if we assume that  $T$  maps the plane generated by  $x$  and  $y$  to itself, then since  $T(0) = 0$  we find  $T(x + y) = T(x) + T(y)$ . Otherwise, we consider a linear isometry  $S$  which projects the plane generated by  $T(x)$  and  $T(y)$  to the plane generated by  $x$  and  $y$ , and then it follows that  $(S \circ T)(x + y) = S(Tx + Ty)$ , hence  $T(x + y) = Tx + Ty$ , because  $S$  is a bijection. It follows that  $T(0) = 0$  holds if and only if  $T$  is an element of the orthogonal group of isometric linear transformations  $O_n$ . If  $T \in E_n$  is any linear transformation, and if  $T(0) = x$ , then the isometry  $T_x^{-1} \circ T$  maps zero to zero, hence  $T_x^{-1} \circ T \in O_n$ , and we find that we can write any Euclidean transformation as a rotation and a translation, and by normality we find the Euclidean group is actually the semidirect product of  $O_n$  and  $\mathbb{R}^n$ , since if  $M, N \in O_n$ ,

$$(T_x \circ M) \circ (T_y \circ N) = T_{x+My} \circ MN$$

$E_n$  can be given the structure of a Lie group if we take the topology corresponding to  $\mathbb{R}^n \times O_n$ , in which case

$$(x, M)(y, N)^{-1} = (x, M)(-Ny, N^{-1}) = (x - MNx, MN^{-1})$$

which is differentiable, since the multiplication map on  $O_n$  is differentiable, and the map  $x - MNx$  is differentiable since the action of  $M_n$  on  $\mathbb{R}^n$  defined by  $(M, x) \mapsto Mx$  is differentiable.

The left and right translation maps  $L_x(y) = xy$  and  $R_x(y) = yx$  are diffeomorphisms on any Lie group  $G$ , so they induce bundle equivalences  $(L_x)_* : TG \rightarrow TG$  and  $(R_x)_* : TG \rightarrow TG$ . We say a vector field  $X$  is *left-invariant* if  $(L_x)_*X = L_x \circ X$  for all  $x \in G$ , i.e. if  $(L_x)_*(X_y) = X_{xy}$ . It suffices to show that  $(L_x)_*(X_e) = X_x$ , because then

$$(L_x)_*(X_y) = (L_x \circ L_y)_*(X_e) = (L_{xy})_*(X_e) = X_{xy}$$

Given any  $v \in G_e$ , we can define a unique left invariant vector field  $X$  with  $X_e = v$  by setting  $X_x = (L_x)_*(v)$ .

**Theorem 8.1.** *Any left-invariant vector field is automatically  $C^\infty$ .*

*Proof.* We need only verify that the map  $X_p = (L_p)_*(v)$  is  $C^\infty$  for any  $v \in G_e$ , and it suffices to prove this in a neighbourhood of the origin. Let  $(x, U)$  be a chart around a neighbourhood of the origin. Let  $V \subset U$  be a neighbourhood chosen such that  $ab^{-1} \in U$  if  $a, b \in U$ . Then

$$Xx_i$$

□

**Corollary 8.2.** *A Lie group always has trivial tangent bundle.*

Since a Lie group is differentiable, we should be able to ‘linearly approximate’ the group multiplication action.

To remedy this fact, we are required to introduce some complicated machinery, which can be skimmed at a first reading. First, the elementary theory of Lie groups tells us that every tangent vector at the identity can be extended to a left-invariant smooth vector field on the Lie group. Given a vector  $X_e \in \mathfrak{g}$ , we can consider it as a base-point of a left-invariant vector field  $X$ , and use the field to generate a unique curve  $\phi : \mathbb{R} \rightarrow G$  satisfying  $\phi_0 = e$  and  $d\phi_t/dt = X_{\phi_t}$ . It turns out that  $\phi_{t+u} = \phi_t \phi_u$ , so that  $\phi$  is actually a *homomorphism* from  $\mathbb{R}$  to  $G$  (a ‘one-parameter subgroup of  $G$ ’), and we let the *exponential map* from  $\mathfrak{g}$  to  $G$  by letting  $e^X = \phi_1$ . It turns out that  $e^{tX} = \phi_t$ , so the exponential map models all curves emerging from infinitesimals at the identity.

**Example.** Over the multiplicative group  $\mathbf{C}^\times$  of non-zero complex numbers, we can identify the tangent bundle  $T\mathbf{C}^\times$  with  $\mathbf{C}^\times \times \mathbf{C}$ , and the left-invariant vector fields take the form  $X_z = wz$  for some  $w \in \mathbf{C}$ . This tells us that the exponential on this space is the unique solution to the differential equation

$$\frac{dz}{dt} = wz$$

and this is just the standard exponential map  $z(t) = e^{tw}$ , so that the exponential on  $\mathbf{C}^\times$  is just the normal exponential function. Since the exponential descends consistently to Lie subgroups, this tells us that the exponential map on the multiplicative group  $\mathbb{R}^\times$  is just the standard exponential as well.

**Example.** Over the group  $GL_n(\mathbb{R})$ , we can identify the tangent space at each point  $M \in GL_n(\mathbb{R})$  with the space  $M_n(\mathbb{R})$  of all  $n \times n$  matrices. The left-invariant vector fields on  $GL_n(\mathbb{R})$  are of the form  $X_M = NM$  for some  $N \in M_n(\mathbb{R})$ , and the unique solution to the system of linear equations

$$\frac{dM}{dt} = NM$$

is the matrix exponential

$$e^M = \sum_{n=0}^{\infty} \frac{M^n}{n!}$$

hence the exponential of Lie groups generalizes many versions of the exponential defined in analysis.

**Example.** Over the additive group of real numbers  $\mathbb{R}$ ,  $T\mathbb{R}$  is just the trivial tangent bundle  $\mathbb{R} \times \mathbb{R}$ , so elements of the tangent space at the identity can be identified with real numbers, and the left-invariant vector fields are just the vector fields with a constant velocity. Recalling the unique solutions to the differential equation

$$\frac{dx}{dt} = c$$

We find that the for a tangent vector  $t \in \mathbb{R}$  at the identity,  $e^t = t$  is just the identity map. More generally, the exponential related to the additive group  $\mathbb{R}^n$  is just the identity map under a suitable identification of the tangent bundle.



Some elementary facts about the exponential, proved in an elementary introduction to differentiable manifolds are that

- For any tangent vector  $X$ ,

$$\left. \frac{de^{tX}}{dt} \right|_{t=0} = X$$

so as  $t$  ranges over all real-numbers,  $e^{tX}$  is just the one-parameter subgroup of  $G$  corresponding to the map  $\phi$  used to construct the exponential. Thus  $e^{(t+u)X} = e^{tX}e^{uX}$ .

- If  $\phi : G \rightarrow H$  is a homomorphism, then  $e^{\phi_*(X)} = \phi(e^X)$ .
- $\exp : \mathfrak{g} \rightarrow G$  is a diffeomorphism in a neighbourhood of the identity. It is not always surjective, even if  $G$  is connected, but it is surjective for the group  $GL_n(\mathbf{C})$ , or any compact, connected Lie group.

Returning to our discussion of constructing homomorphisms of Lie groups from operations on the tangent space, we see that  $e^{\phi_*(X)} = \phi(e^X)$  gives a much stronger condition on the linear map  $\phi_*$  restricting which linear maps can be differentials of homomorphisms. In particular, since the image of  $\exp$  always contains a neighbourhood of the identity, given any linear map  $\phi_*$  such that  $\phi_*(X) = \phi_*(Y)$  if  $e^X = e^Y$ , we can define a map  $\phi$  on a neighbourhood of the identity of  $G$  by letting  $\phi(e^X) = e^{\phi_*(X)}$ . Provided that  $\phi(gh) = \phi(g)\phi(h)$  where defined, we can try to extend  $\phi$  to a homomorphism on the entire space by using the fact that a neighbourhood of the identity in the Lie group generates the entire space. The sufficient condition for this method to work is that the Lie group is simply connected, and this is not too much of a problem because we can always swap a connected Lie group  $G$  with its simply connected cover  $\tilde{G}$ , and the underlying space of infinitesimals will be the same.

Thus we are left with determining the conditions on  $\phi_*$  such that  $\phi(gh) = \phi(g)\phi(h)$ . This is where the Lie bracket enters the picture. If  $X$  and  $Y$  are arbitrary smooth vector fields on  $G$ , then they induce integral curves  $\phi : \mathbb{R} \times G \rightarrow G$  and  $\psi : \mathbb{R} \times G \rightarrow G$  respectively, and the Lie bracket operation  $[X, Y]$  (turning the space of smooth vector fields into an infinite dimensional Lie algebra) defines a smooth vector field with

$$[X, Y]_p = (1/2) \left. \frac{d^2(\psi_{-t} \circ \phi_{-t} \circ \psi_t \circ \phi_t)}{dt} \right|_p$$

which effectively means that for any  $C^\infty$  function  $f$ ,

$$[X, Y]_p(f) = (1/2) \left. \frac{d^2(f \circ \psi_{-t} \circ \phi_{-t} \circ \psi_t \circ \phi_t)}{dt} \right|_p$$

The Lie bracket essentially measures the commutivity of  $X$  and  $Y$ .

It turns out that if  $X$  and  $Y$  are both left-invariant vector fields on a Lie group  $G$ , then  $[X, Y]$  is also a left-invariant vector field, hence the Lie bracket descends to an operation on the tangent space  $\mathfrak{g}$ , where if  $X, Y \in \mathfrak{g}$ , then  $[X, Y] \in \mathfrak{g}$  satisfies

$$[X, Y](f) = (1/2) \frac{d^2}{dt} f(e^{tX} e^{tY} e^{-tX} e^{-tY})$$

Viewing  $\mathfrak{g}$  as the space of curves through the origin identified up to first order, this implies that for any two curves  $\lambda, \gamma$ ,

$$f(\lambda_t \gamma_t \lambda_t^{-1} \gamma_t^{-1}) = f(e) + t^2[\lambda, \gamma](f) + o(t^3)$$

Thus the Lie bracket expresses the second order coefficients of conjugation on the Lie group. Surprisingly, the second order terms are sufficient to characterize the Lie group operation. First note that if  $\phi : G \rightarrow H$  is a group homomorphism, then in two different ways, we calculate that

$$\begin{aligned} f(\phi(\lambda_t \gamma_t \lambda_t^{-1} \gamma_t^{-1})) &= f(e) + t^2[\lambda, \gamma](f \circ \phi) + o(t^3) \\ &= f(e) + t^2(\phi_*[\lambda, \gamma])(f) + o(t^3) \\ f(\phi(\lambda_t \gamma_t \lambda_t^{-1} \gamma_t^{-1})) &= f((\phi \circ \lambda)_t (\phi \circ \gamma)_t (\phi \circ \lambda)_{-t} (\phi \circ \gamma)_{-t}) \\ &= f(e) + t^2[\phi_*(\lambda), \phi_*(\gamma)](f) + o(t^3) \end{aligned}$$

It follows that  $\phi_*[\lambda, \gamma] = [\phi_*(\lambda), \phi_*(\gamma)]$ , so all linear maps on infinitesimals induced by homomorphisms of groups preserve the Lie bracket operation. We shall find that we can ‘almost’ always find a Lie group homomorphism from a linear map on the space of infinitesimals preserving the Lie bracket.

Since  $\exp$  is locally a diffeomorphism in a neighbourhood of both identities, we can find a differentiable inverse  $\log$ , and the Baker-Hausdorff formula implies that there is a neighbourhood of the origin in  $\mathfrak{g}$  and universal constants  $a_\alpha$  such that

$$\log(e^X e^Y) = X + Y + \sum_{|\alpha| > 1} a_\alpha(X, Y)^\alpha$$

where  $\alpha$  are multi-indexes, and if  $\alpha = (\alpha_1, \dots, \alpha_n)$ , then

$$(X, Y)^\alpha = \underbrace{[X, [X, [\dots, [X, [Y, [\dots, [Y, \dots]]]]]]]}_{\alpha_1 \text{ times}} \underbrace{[\dots]}_{\alpha_2 \text{ times}}$$

A nasty formula which implies that given a linear map  $\phi_* : \mathfrak{g} \rightarrow \mathfrak{h}$ , the sufficient condition for the map  $\phi : G \rightarrow H$  defined by  $\phi(e^X) = e^{\phi_*(X)}$  to satisfy  $\phi(gh) = \phi(g)\phi(h)$  where defined, a sufficient condition for this to hold is that

$$\log(e^{\phi_*(X)} e^{\phi_*(Y)}) = \phi_*(\log(e^X e^Y))$$

and in terms of the Baker-Hausdorff formula, we find that this means

$$\phi_*(X) + \phi_*(Y) + \sum a_\alpha(\phi_*(X), \phi_*(Y))^\alpha = \phi_*(X) + \phi_*(Y) + \sum a_\alpha \phi_*((X, Y)^\alpha)$$

It can be proved by induction that for any map  $\phi_*$  with  $\phi_*[X, Y] = [\phi_*X, \phi_*Y]$ ,  $\phi_*((X, Y)^\alpha) = (\phi_*(X), \phi_*(Y))^\alpha$ , so a sufficient condition for the homomorphism condition  $\phi(gh) = \phi(g)\phi(h)$  is that  $\phi_*$  preserves the Lie bracket.

Now that this property holds, if  $U$  is the neighbourhood upon which the homomorphism property holds, then any  $g \in G$  can be written as  $h_1 h_2 \dots h_n$  for some  $h_i \in U$ , and if  $\phi$  can be extended to a homomorphism on all of  $G$ , then we can let  $\phi(g) = \phi(h_1) \dots \phi(h_n)$ , and provided this is well defined,  $\phi$  will be a homomorphism that is differentiable at the identity, hence differentiable everywhere, and  $\phi_*$  is the differential of  $\phi$ . The only problem is that this extension of  $\phi$  won't necessarily be well defined everywhere, and in order for this to work, we will need to switch to studying simply connected Lie groups.

**Example.** Recall the Grassmanian space  $G(k, n)$  of  $k$  dimensional subspaces of  $\mathbb{R}^n$ . There is another way to view the topology of  $G(k, n)$  which is very useful for constructing continuous maps on the space. Consider the family  $M(n, k; k)$  of full rank  $n$  by  $k$  matrices, which may be identified by separating columns with the family of linearly independent tuples  $(v_1, \dots, v_k)$  of vectors in  $\mathbb{R}^n$ . We obtain a surjective map  $f : M(n, k; k) \rightarrow G(k, n)$  by mapping a matrix corresponding to the tuples  $(v_1, \dots, v_k)$  to the vector space  $\text{span}(v_1, \dots, v_k)$ . We claim that this map is a submersion, hence it is a continuous open map identifying the topological structure of  $G(k, n)$  as a quotient space of  $M(n, k; k)$ , by identifying vector tuples which generate the same subspace. The utility of this is that continuous map  $g$  with domain  $G(k, n)$  can be constructed as continuous maps with domain  $M(n, k; k)$  which have the same value on vector tuples that generate the same subspace. To prove that the map is a submersion, we consider the open set

$$U = \left\{ \begin{pmatrix} A \\ B \end{pmatrix} : A \in GL_k(\mathbb{R}) \right\}$$

If we let  $V = \text{span}(e_1, \dots, e_k)$ , and  $V' = \text{span}(e_{k+1}, \dots, e_n)$ , then  $f(U)$  is a subset of  $A_{V'}$ , and if we let  $(y, A_{V'})$  be the standard coordinates corresponding to the set  $A_{V'}$ , identifying  $L(V, V')$  with  $M(k, n - k)$ , then

$$(y \circ f) \begin{pmatrix} A \\ B \end{pmatrix} = BA^{-1}$$

which is the linear transformation mapping the  $i$ 'th column of  $A$  to the  $i$ 'th column of  $B$ . The map is now easily seen to be differentiable. For a fixed  $A$ , the map  $B \mapsto BA^{-1}$  is an invertible linear map, and therefore the map is a submersion everywhere, because we can always permute the coordinates so that a given matrix is in  $U$  (this corresponds to a diffeomorphism of  $M(n, k; k)$ ), and when we take the span of these vectors, the way to get back to the original span is to unpermute the coordinates, so we can always assume our matrices have the nice form above. In other words, the group  $GL_n(\mathbb{R})$  acts transitively on  $M(n, k; k)$  by left multiplication, and transitively on  $G(k, n)$  by the action  $M \cdot V = \{Mv : v \in V\}$ , and we find that  $f$  is a  $GL_n(\mathbb{R})$  morphism, because  $f(MN) = Mf(N)$ .

# **Part II**

## **Riemannian Geometry**

# Chapter 9

## Riemannian Manifolds

On  $\mathbb{R}^n$ , an inner product enables us to measure the lengths of vectors, and angles between them. By considering inner products of the tangent space of a manifold, we can find a way to measure lengths of smooth curves, and angles between curves passing through a point. This is the theory of riemannian geometry.

First, we recall the necessary linear algebra. A bilinear form on a real vector space  $V$  is a bilinear map  $\beta : V \times V \rightarrow \mathbb{R}$ . This form is *symmetric* if  $\beta(v, v') = \beta(v', v)$  for all  $v, v' \in V$ , *positive-definite* if  $\beta(v, v) > 0$  for all  $v \in V - \{0\}$ , and *non-degenerate* if for all  $v \in V - \{0\}$ , there is  $v' \in V$  with  $\beta(v, v') \neq 0$ . Given a non-degenerate form on  $V$ , we can naturally define a map  $v \mapsto v^\flat$  from  $V$  to  $V^*$  via the equation  $v^\flat(w) = \beta(v, w)$ . This map is injective precisely when  $\beta$  is non-degenerate, and thus naturally identifies  $V$  with  $V^*$  if  $V$  is finite dimensional.

A *Riemannian manifold* is a manifold  $M$  equipped with a *Riemannian metric*, a smooth assignment of a positive-definite symmetric bilinear form  $\langle \cdot, \cdot \rangle_p$  on the tangent spaces  $T_p M$ , for each point  $p \in M$ . We can characterize this smoothness in several equivalent ways:

- We can view the Riemannian metric as a section of the  $(0, 2)$  tensor bundle  $T^2 M$ , and such sections naturally have a definition of smoothness.
- The Riemannian metric will be smooth if and only if for any two smooth vector fields  $X$  and  $Y$  on  $M$ ,  $\langle X, Y \rangle$  is a smooth function on  $M$ .
- In any coordinate system  $x : U \rightarrow M$ , we can write the Riemannian metric uniquely as  $\sum g_{ij}(x) dx^i \otimes dx^j$ , where  $g_{ij} = g_{ji}$ . Then the metric is smooth if and only if the functions  $g_{ij}$  are smooth.

A smooth choice of *non-degenerate* bilinear forms is called a *pseudo Riemannian metric* on  $M$ , and gives  $M$  the structure of a *pseudo Riemannian manifold*.

**Example.**  $\mathbb{R}^n$  is a Riemannian manifold, with the Riemannian tensor field

$$\sum dx^i \otimes dx^i$$

Often, the field is just denoted by the Kronecker delta  $\delta$ , since the coordinates  $g_{ij}$  associated with this metric is just the Dirac delta function  $\delta_{ij}$ .

**Example.** For some more explicit Riemannian metrics defined in coordinates, it is useful to consider the Riemannian space forms on  $\mathbb{R}^n$ . Let  $s_\lambda$  be the unique solution to the differential equation  $s''_\lambda = -\lambda s_\lambda$  with  $s_\lambda(0) = 0$  and  $s'_\lambda(0) = 1$ . Then define  $\psi(x) = [s_\lambda(|x|)/|x|]^2$ , and set

$$g_{jk} = (1 - \psi) \frac{x_j x_k}{|x|^2} + \psi \delta_{jk}.$$

This defines a metric on  $\mathbb{R}^d$  for  $\lambda \leq 0$ , and on an open ball of radius  $\pi/\sqrt{\lambda}$  for  $\lambda > 0$ . Using more advanced terminology, this definition here is precisely the metric in normal coordinates on a Riemannian manifold with constant curvature  $\lambda$ . *TODO: Use Jacobi fields to prove this.*

*TODO: Merge with previous example. The metric structure of the hyperbolic plane  $\mathbf{H}^2$  can be modelled as the upper half plane in  $\mathbb{R}^2$ , where the metric is given by*

$$g = \frac{dx^2 + dy^2}{y^2}.$$

*This metric pushes the  $x$  axis ‘off to infinity’ by stretching the distance between points near the axis. This is the Poincare model of the hyperbolic plane. More generally, we can define the higher dimensional hyperbolic spaces*

$$\mathbf{H}^n = \{(x, y) : x \in \mathbb{R}^{n-1}, y > 0\}$$

*with a metric  $g = (dx_1^2 + \dots + dx_n^2 + dy^2)/y^2$ .*

**Example.** A surface  $\Sigma$  in  $\mathbb{R}^3$  inherits a Riemannian metric from the ambient space it lies in by restricting the inner product on  $T_p \mathbb{R}^3$  to  $T_p \Sigma$  for each  $p \in \Sigma$ . More generally, an immersed submanifold of a Riemannian manifold naturally obtains a Riemannian structure. If  $i : M \rightarrow N$  is an immersion, and  $N$  has a Riemannian metric, then we can define the Riemannian metric on  $M$  by the equation

$$\langle X, Y \rangle_M = \langle i_* X, i_* Y \rangle_N$$

*The same construction does not necessarily work in the pseudo Riemannian setting, since the restriction of a non-degenerate bilinear form to a subspace need not be non-degenerate.*

Since every smooth manifold can be immersed in Euclidean space, every smooth manifold has a Riemannian metric. Alternatively, we can construct a Riemannian metric by locally defining a positive definite form, and then extending it using a partition of unity.

An *isometry* between two Riemannian manifolds is a diffeomorphism which acts as an isometry on each tangent space. The study of Riemannian geometry can be considered the study of properties of Riemannian manifolds which are invariant under isometries.

Pseudo Riemannian metrics often occur in physical theories such as relativity theory. Since the eigenvalues corresponding to the symmetric operator change smoothly, and cannot be zero by non-degeneracy, the number of negative and positive eigenvalues do not change across a connected pseudo-Riemannian manifold. The number of negative eigenvalues for a pseudometric is known as the *index* of the manifold. An index 0 metric is precisely a Riemannian metric, and an index 1 metric is known as a *Lorentz metric*.

**Example.** The classical example of a Lorentz metric occurs in special relativity, known as the Minkowski metric on  $\mathbb{R} \times \mathbb{R}^n$ , given by the metric

$$g = dx_1^2 + \cdots + dx_n^2 - dt^2.$$

This metric is natural in the study of special relativity, since the symmetries of that theory are precisely the isometries of this space, the Lorentz transforms. More generally, in the study of general relativity, the  $\mathbb{R}^n \times \mathbb{R}$  by a general  $n + 1$  dimensional manifold, and the Minkowski metric by an arbitrary Lorentz metric, satisfying certain physical equations known as the Einstein equations, which control how the Lorentz metric behaves on the space.

## 9.1 The Gradient Vector

Since a non-degenerate form enables us to identify a vector space  $V$  with its dual, a metric enables us to identify  $TM$  and  $T^*M$  in a natural way. This identification is also called the *musical isomorphism*, since we use  $v \mapsto v^\flat$  to denote the map from  $TM$  to  $T^*M$ , and  $\omega \mapsto \omega^\sharp$  to denote the map from  $T^*M$  to  $TM$ . In coordinates, for a vector  $v = \sum v^j \partial_{x^j}$  and  $\omega = \sum \omega_j dx^j$ , we have

$$v^\flat = \sum g_{ij} v^i dx^j \quad \text{and} \quad \omega^\sharp = \sum g^{ij} \omega_i \frac{\partial}{\partial x^j},$$

where  $g^{ij}$  denotes the entries of the inverse of the matrix with coefficients  $g_{ij}$ , i.e. so that  $\sum g^{ij} g_{jk} = \delta_k^i$  for each fixed  $i$  and  $k$ . The entries  $g^{ij}$  then transform as a  $(2, 0)$  tensor, and thus give rise to an inner product on each of the cotangent spaces  $T_p^*M$ .

For any scalar valued function  $f : M \rightarrow \mathbb{R}$ , the differential  $df$  is a section of  $T^*M$ . Taking the dual using the bilinear form, we obtain the *gradient*  $\nabla f$ , which is a section of  $TM$ , defined by the equation

$$\langle \nabla f, X \rangle = df(X).$$

In coordinates, we have  $\nabla f = \sum g^{ij} (\partial_{x^i} f) \partial_{x^j}$ . Over Euclidean space, with the standard metric, we obtain the standard gradient.

**Example.** On the sphere  $S^2$ , we have spherical coordinates

$$x = r \cos \phi \cos \theta, \quad y = r \sin \phi \cos \theta, \quad z = r \sin \theta.$$

We then have

$$\begin{aligned} dx &= (\cos \phi \cos \theta) dr + (-r \sin \phi \cos \theta) d\phi + (-r \cos \phi \sin \theta) d\theta \\ dy &= (\sin \phi \cos \theta) dr + (r \cos \phi \cos \theta) d\phi + (-r \sin \phi \sin \theta) d\theta \\ dz &= (\sin \theta) dr + (r \cos \theta) d\theta \end{aligned}$$

and so the metric is

$$\begin{aligned} g &= dx \otimes dx + dy \otimes dy + dz \otimes dz \\ &= dr \otimes dr + (r^2 \cos^2 \theta) d\phi \otimes d\phi + (r^2) d\theta \otimes d\theta, \end{aligned}$$

Thus the basis given by spherical coordinates is orthogonal, but not orthonormal. We have

$$\left| \frac{\partial}{\partial r} \right| = 1 \quad \left| \frac{\partial}{\partial \phi} \right| = r |\cos \theta| \quad \left| \frac{\partial}{\partial \theta} \right| = r$$

Thus

$$\nabla f = \frac{\partial f}{\partial r} \frac{\partial}{\partial r} + \frac{1}{r^2 \cos^2 \theta} \frac{\partial f}{\partial \phi} \frac{\partial}{\partial \phi} + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \frac{\partial}{\partial \theta}.$$

**Example.** In Minkowski space, we have a time coordinate  $t$ , and three spatial coordinates  $x$ ,  $y$ , and  $z$ . If  $t$  is chosen as the 0th coordinate, the Minkowski metric is chosen such that

$$g_{ij} = \begin{cases} 1 & i = j > 0 \\ -c^2 & i = j = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then if  $f$  is a function on Minkowski space, we find

$$\nabla f = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} - \frac{1}{c^2} \frac{\partial f}{\partial t}.$$

The gradient  $\nabla f$  has the same interpretation as in Euclidean space, i.e. the direction in the tangent space that gives the steepest ascent on the manifold relative to the underlying metric. If  $v \in M_p$  is any vector with  $|v| = 1$ , then the Schwarz inequality implies

$$|v(f)| = |df(v)| = |\langle v, \nabla f \rangle| \leq |\nabla f|$$

If  $f$  is real valued, then the tangent space of the level sets of the function  $f$  are precisely the vectors orthogonal to the gradient vector.

## 9.2 Unit Tangent Bundle

If  $M$  is a Riemannian manifold, we consider a submanifold  $SM$  of  $TM$  defined by the equation  $\langle v, v \rangle = 1$ . It is a submanifold of dimension  $2n - 1$ , if  $M$  has dimension  $n$ .

**Example.** If  $w$  is a unit tangent vector to  $S^2$  at some point  $v \in S^2$ , then  $(v, w, v \times w)$  is a right handed orthonormal coordinate frame. Thus we find that the space of unit tangent vectors is diffeomorphic to  $SO(3)$ .

On any manifold, we can define  $SM$  by taking the *projectivization* of the tangent space  $TM$ , but then this manifold is not so naturally embedded as a submanifold of  $TM$ .



### 9.3 The Volume Form

Any Riemannian manifold has a natural notion of volume. Tensorizing the definition of the metric, we can obtain a natural inner product on all of the tensor bundles of  $M$ ; it is precisely the unique inner product that agrees with the Riemannian metric on tangent and cotangent vectors, and satisfies

$$\langle \omega_1 \otimes \eta_1, \omega_2 \otimes \eta_2 \rangle = \langle \omega_1, \omega_2 \rangle \langle \eta_1, \eta_2 \rangle.$$

In particular, if  $M$  is  $n$  dimensional, then we naturally obtain an inner product on  $\Omega^n(TM)$ . Since  $\Omega^n(TM)$  is a one dimensional bundle, there exists exactly two locally defined  $n$ -forms that have magnitude one with respect to this inner product. If we are working in a coordinate system  $x$ , and the metric has coefficients  $g_{ij}$  in these coordinates forming a matrix  $G$ , then we can calculate (e.g by induction and cofactor expansions) that

$$\langle dx^1 \wedge \cdots \wedge dx^n, dx^1 \wedge \cdots \wedge dx^n \rangle = \det(G).$$

Thus the norm one elements of  $\Omega^n(M)$  are precisely

$$\pm \frac{dx^1 \wedge \cdots \wedge dx^n}{\det(G)^{1/2}}.$$

This means that the resulting scalar density  $dV = |dx^1 \wedge \cdots \wedge dx^n| |\det(G)^{-1/2}|$  is invariantly defined on  $M$ , which we call the *volume element* on  $M$ . If  $M$  is *oriented*, we can choose an oriented choice of the norm one elements, and thus obtain a globally defined  $n$  form, the *volume form*. Thus for any Riemannian manifold  $M$ , we can define it's *volume*  $\text{Vol}(M)$  as

$$\text{Vol}(M) = \int_M dV.$$

Any compact manifold will have finite volume, but non-compact manifolds can have finite or infinite volume.

# Chapter 10

## Connections

Let  $c : I \rightarrow M$  be a smooth curve into a manifold  $M$ . Without additional structure, it does not make sense to state that  $c$  has *constant velocity*. Indeed, on  $\mathbb{R}^d$  constant velocity curves travel on straight lines, but under an arbitrary diffeomorphism these curves can be mapped onto curves that behave in an arbitrary fashion. We begin by discussing the analogous problem with such definitions infinitesimally on the tangent bundle, and then discuss how the concept of a *connection* allows one to define constant velocity curves, which in differential geometry are called *autoparallels*, and which we will relate to the study of *geodesics* in later sections.

Consider the bundle  $T(TM)$ , the tangent bundle to the tangent bundle. For any coordinate system  $x$  on  $U \subset M$ , we obtain an induced coordinate system  $(x, \dot{x})$  on  $\pi_{TM}^{-1}(U) \subset TM$ , where  $x(v)$  is just the coordinates of the base point of  $v$ , and  $\dot{x}(v) \in \mathbb{R}^d$  gives the coefficients of the expansion of  $v$  in the basis  $\partial_{\dot{x}^i}$ . Thus we can write every element of  $T_v(TM)$  as

$$\sum a^i \frac{\partial}{\partial \dot{x}^i} \Big|_v + b^i \frac{\partial}{\partial x^i} \Big|_v.$$

In another coordinate system  $y$  on  $U$ , we have

$$\dot{y}^i(p) = \sum \frac{\partial y^i}{\partial x^j} \Big|_p \dot{x}^j(p).$$

Thus

$$\frac{\partial \dot{y}^i}{\partial \dot{x}^j} \Big|_v = \frac{\partial y^i}{\partial x^j} \Big|_p \quad \text{and} \quad \frac{\partial \dot{y}^i}{\partial x^j} \Big|_v = \sum \frac{\partial^2 y^i}{\partial x^k \partial x^j} \Big|_p \dot{x}^k(p).$$

But this means that

$$\frac{\partial}{\partial \dot{x}^i} \Big|_v = \sum \frac{\partial y^j}{\partial x^i} \Big|_p \frac{\partial}{\partial \dot{y}^j} \Big|_v + \sum \frac{\partial^2 y^j}{\partial x^i \partial x^k} \Big|_p \dot{x}^k(p) \frac{\partial}{\partial \dot{y}^j} \Big|_v$$

and

$$\frac{\partial}{\partial \dot{x}^i} \Big|_v = \sum \frac{\partial y^j}{\partial x^i} \Big|_p \frac{\partial}{\partial \dot{y}^j} \Big|_v.$$

Setting  $A(x)$  to be the matrix with entries  $A_j^i = \partial y^i / \partial x^j$ , and  $B(v)$  the matrix with entries  $B_j^i = \sum \partial^2 y^i / \partial x^j \partial x^k \dot{x}^k$ , then if

$$\sum a^i \frac{\partial}{\partial x^i} \Big|_v + b^i \frac{\partial}{\partial \dot{x}^i} \Big|_v = \sum c^i \frac{\partial}{\partial y^i} \Big|_v + d^i \frac{\partial}{\partial \dot{y}^i} \Big|_v,$$

then

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} A & 0 \\ A & B \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

Thus if  $a = 0$ , then  $c = 0$ , and so the span of the vectors  $\{\partial_{x^i}\}$  at each point generates a vector bundle  $V(TM)$  which is invariant under coordinate changes in  $M$ . We can see this in a coordinate invariant way also, because  $V(TM)$  is the kernel of the bundle map  $\pi_* : T(TM) \rightarrow TM$  induced by  $\pi : TM \rightarrow M$ . We call  $V(TM)$  the *vertical bundle*.

On the other hand, the span of the vectors  $\{\partial_{x^i}\}$  is only invariant if  $B = 0$ , i.e. under *linear coordinate changes*. This makes sense from our original discussion. If we could form a vector bundle  $H(TM)$  in a coordinate independent way from the span of vectors  $\{\partial_{x^i}\}$ , then from  $c : I \rightarrow M$ , we can differentiate and obtain a curve  $c' : I \rightarrow TM$  and again differentiate to obtain  $c'' : I \rightarrow T(TM)$ . In coordinates, we would have

$$c''(t) = \sum \frac{d(x^i \circ c)}{dt} \frac{\partial}{\partial x^i} \Big|_{c'(t)} + \frac{d^2(x^i \circ c)}{dt^2} \frac{\partial}{\partial \dot{x}^i} \Big|_{c'(t)},$$

and thus we could then declare  $c$  to have constant velocity if  $c''(I) \subset H(TM)$ , and this would then be a coordinate independent property, which we have seen is not possible.

A connection is a structure on a manifold that allows us to obtain some definition of a constant velocity curve. It is as easy to define this for a general vector bundle  $\pi : E \rightarrow M$ , which is often helpful in computations, so we might as well do so. For any vector bundle  $\pi : E \rightarrow M$ , we have a short exact sequence

$$0 \rightarrow VE \rightarrow TE \rightarrow \pi^*(TM) \rightarrow 0.$$

A connection is a *splitting* of this short exact sequence. This requires us to perform one of the three equivalent processes:

- We could form a decomposition  $TE = VE \oplus HE$  for some bundle  $HE$ , known as the *horizontal bundle* of the connection.
- We could *split on the left*, defining a connection as a morphism  $\omega : TE \rightarrow VE$ , known as the *connection form*.
- We could *split on the right*, defining a connection as a morphism  $l : \pi^*(TM) \rightarrow TE$ , known as the *lifting map*.

If one performs any of these three properties, there is a canonical way to obtain the other three maps. The connection form has  $HE$  as a kernel, and the lifting map has  $HE$  as an

image. Given a manifold whose tangent bundle equipped with a connection, an *autoparallel* is then a curve  $c : I \rightarrow M$  such that  $c''(I) \subset H(TM)$ , which we think of as having constant velocity with respect to the connection.

Suppose  $\pi : E \rightarrow M$  is a vector bundle, and consider an open set  $U \subset M$  such that we have both a coordinate system  $x : M \rightarrow \mathbb{R}^d$ , and sections  $e_\alpha : M \rightarrow E$  given a frame on  $E$ . Then  $(x, a)$  forms a coordinate system on  $\pi^{-1}(U)$ , where  $x(v) = x(\pi(v))$  is the basepoint coordinate, and  $a(v)$  gives the coefficient expansion of  $v$  in the frame  $\{e_\alpha\}$ . Thus  $TE \rightarrow E$  has a natural frame, which we denote by  $\{e_i\} \cup \{\dot{e}_\alpha\}$ . In coordinates, we can write

$$\omega \left( \sum a^i e_i + b^\alpha \dot{e}_\alpha \right) = \sum (a^i N_i^\alpha + b^\alpha) \dot{e}_\alpha$$

for smooth functions  $N_i^\alpha$  on  $E$ . They are called the *(nonlinear) connection coefficients*. In general, the functions are only homogeneous as functions on  $TE$ , and do *not* transform tensorially under a change of coordinates (the formula for the transformation is somewhat technical, see the remark below). However, if the coefficients  $N_i^\alpha$  are always *linear* in  $v$ , then we call the connection *linear*, and we can then write  $N_i^\alpha(\sum v^\beta e_\beta) = \sum \Gamma_{i\beta}^\alpha v^\beta$  for smooth coefficients  $\Gamma_{i\beta}^\alpha$  on  $M$ , which also do not transform tensorially. We call these the *linear connection coefficients*. The tangent vectors

$$\frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - \sum N_i^\alpha \dot{e}_\alpha,$$

which we also denote by  $\delta e_i$ , then form a frame for  $HE$ , and the associated lifting map  $l : \pi^*(TM) \rightarrow H(TM)$  is then given by  $l(\sum a^i e_i) = \sum a^i \delta e_i$ . The vectors  $\delta e_i$  transform as

$$\frac{\delta}{\delta x^j} = \sum \frac{\partial y^i}{\partial x^j} \frac{\delta}{\delta y^i},$$

just like the usual tangent vectors. On the cobundle  $E^*$ ,  $VE^\perp$  can be defined without reference to a connection, and has a frame  $\{dx^j\}$ . A basis in coordinates to  $H(TM)^\perp$  can be defined by introducing the covectors  $\delta(\dot{e}^\alpha)^* = \sum N_i^\alpha dx^i + d\dot{e}^\alpha$ .

A connection gives rise to *parallel transports along curves*.

**Theorem 10.1.** *Let  $E \rightarrow M$  be a vector bundle equipped with a connection. Then for any curve  $c : I \rightarrow M$ , and any  $t_0$  and  $t_1$ , there exists a bijective map  $\Gamma(c)_{t_0}^{t_1} : E_{c(t_0)} \rightarrow E_{c(t_1)}$  called the parallel transport map, such that for each  $v \in E_{c(t_0)}$ , if we consider the curve  $\gamma : I \rightarrow E$  given by*

$$\gamma(t) = \Gamma(c)_{t_0}^t \{v\},$$

*then  $\gamma$  is the unique curve with  $\gamma(t) \in E_{c(t)}$  for all  $t$  and  $\gamma'(t) \in HE$  for all  $t \in I$ . The curve  $\gamma$  is called the parallel transport of  $v$  along  $c$ . If the connection is a linear connection, then the parallel transport maps are linear.*

*Proof.* The proof is fairly simple. Define  $F = c^*E$ , a vector bundle on  $[0, 1]$ . The pull back of any vector bundle with a connection naturally has a connection (we simply pull back the

horizontal subbundle). The connection on  $F$  gives us a lifting map  $l : \pi^*(TI) \rightarrow TF$ , which induces a vector field  $X \in \Gamma(TF)$  by the formula  $X_v = l(\partial_t|_v)$ . The integral curves of this vector field are precisely the parallel transport maps needed above. The calculations in the following argument verify that if the connection is linear, then the parallel transport maps are linear.  $\square$

*Remark.* Suppose  $\gamma : I \rightarrow E$  is a parallel transport curve, and we can write  $\gamma$  in a frame as

$$\gamma(t) = \sum \gamma^\alpha(t) e_\alpha(c(t)).$$

Our constraints on  $\gamma$  amounts to  $\omega(\gamma'(t)) = 0$ , which we can write as

$$\omega \left( (c^i)'(t) \frac{\partial}{\partial x^i} \Big|_{\gamma(t)} + (\gamma^\alpha)'(t) \dot{e}_\alpha(\gamma(t)) \right) = 0,$$

or

$$(\gamma^\alpha)' = - \sum (c^i)' N_i^\alpha(\gamma).$$

Thus if we set  $\Gamma(c)_0^t(v) = \sum A^\alpha(t, v) e_\alpha$ , then  $A$  satisfies the system of nonlinear differential equations

$$(\partial_t A^\alpha) = - \sum (c^i)' N_i^\alpha(A),$$

such that  $A(0, v) = v$ . If  $N_i^\alpha(v) = \sum \Gamma_{i\alpha}^\beta \dot{e}_\beta$ , then we obtain a linear differential equation

$$(\partial_t A^\alpha) = - \sum (c^i)' \Gamma_{i\beta}^\alpha(c) A^\beta.$$

This in particular implies that  $A(c)_0^t$  is then a linear function of  $v$ . We can thus write  $A^\alpha = \sum A_\beta^\alpha v^\beta$ , and then  $(A_\beta^\alpha)' = - \sum (c^j)' \Gamma_{j\lambda}^\alpha(c) A_\beta^\lambda$ .

*Remark.* In the next remark, we calculation the change of variables formula for the nonlinear connection coefficients. If in a coordinate system  $y$  we have connection coefficients  $\tilde{N}_k^j(y, w)$ , and we write

$$A_a^i = \frac{\partial y^i}{\partial x^a} \quad B_{ab}^i = \frac{\partial^2 y^i}{\partial x^a \partial x^b} \quad C_i^a = \frac{\partial x^a}{\partial y^i} \quad D_{ij}^a = \frac{\partial^2 x^a}{\partial y^i \partial y^j},$$

then

$$\begin{aligned} \omega \left( \frac{\partial}{\partial x^a} \right) &= \omega \left( \sum A_a^i \frac{\partial}{\partial y^i} + B_{ab}^j \dot{x}^b \frac{\partial}{\partial \dot{y}^j} \right) \\ &= \sum (A_a^i \tilde{N}_i^j + B_{ac}^j \dot{x}^c) \frac{\partial}{\partial \dot{y}^j} \\ &= \sum (A_a^i \tilde{N}_i^j + B_{ac}^j \dot{x}^c) C_j^b \frac{\partial}{\partial \dot{x}^b}. \end{aligned}$$

Thus

$$N_a^b = \tilde{N}_i^j A_a^i C_j^b + B_{ac}^j \dot{x}^c C_j^b$$

Inverting this equation, we obtain that

$$\tilde{N}_i^j = N_a^b C_i^a A_b^j + D_{ik}^b \dot{y}^k A_b^j$$

If the connection is linear, we can also directly express a relationship between linear connection coefficients  $\Gamma_{il}^j$  and  $\tilde{\Gamma}_{il}^j$ , i.e.

$$\Gamma_{ac}^b = \tilde{\Gamma}_{ik}^j A_a^i A_c^k C_j^b + B_{ac}^j C_j^b \quad \text{and} \quad \tilde{\Gamma}_{ik}^j = (C_i^a \Gamma_{ac}^b C_k^c + D_{ik}^b) A_b^j.$$

Note that these quantities *only transform tensorially* under linear coordinate changes, due to the presense of the additional correction factor appearing in the formulas.

## 10.1 Covariant Derivative Operators

On a general manifold, it is possible to differentiate a smooth function  $f : M \rightarrow \mathbb{R}$  at a point  $p$  in a direction  $v \in T_p M$ . It is *not* possible to define the derivative of a smooth vector field  $X \in \Gamma(TM)$  at a point  $p$  in a direction  $v$ , because in general it is impossible to compare tangent spaces at two different points. We do have a Lie derivative, which allows us to define derivatives if we have a flow in a manifold, but this only allows us to define the derivative of a smooth vector field  $X$  in the direction of another vector field  $Y$ , and so the derivative of  $X$  with respect to  $Y$  also involves derivatives of  $Y$ , which is undesirable in many contexts.

Nonetheless, we now see it *is* possible to define derivatives once we have defined a connection on  $M$ . More generally we can define the derivative of sections of any vector bundle  $E$  equipped with a connection, where the derivative is defined with respect to a vector in the tangent bundle  $TM$ . Given  $s : M \rightarrow E$  and  $v \in T_p M$ , we consider a curve  $c : [0, 1] \rightarrow M$  with  $c'(0) = v$ , and set

$$(\nabla_v s)(p) = \lim_{t \rightarrow 0} \frac{\Gamma(c)_t^0 \{s(c(t))\} - s(c(0))}{t}.$$

To see this operator is independent of the choice of  $c$ , we consider a frame  $\{e_\alpha\}$  on  $E$ , writing

$$s = \sum s^\alpha e_\alpha \quad \text{and} \quad \Gamma(c)_t^0(s) = \sum A_t^\alpha(s) e_\alpha(c(t)).$$

Then one can verify that

$$\Gamma(c)_t^0 \{s(c(t))\} = \sum A_t^\alpha(s(c(t))) e_\alpha(c(t)),$$

and the derivative at the origin then becomes, via the chain rule, equal to

$$\sum [(\partial_t A^\alpha)(s) + (\partial_i s^\alpha) v^i] e_\alpha,$$

which is invariant of the curve  $c$ . We can further simplify this to

$$\sum [\partial_i s^\alpha + N_i^\alpha(s)] v^i e_\alpha.$$

This is, in general, a *nonlinear* differential equation in  $s$ , since the coefficients depend on  $s$ . If the connection is linear, we can further simplify this to

$$\sum [\partial_i s^\alpha + \Gamma_{i\beta}^\alpha s^\beta] v^i e_\alpha,$$

and so we get a linear differential equation in  $s$ .

For a general connection, we see that for constants  $\lambda_1$  and  $\lambda_2$ , we have

$$\nabla_{\lambda_1 v_1 + \lambda_2 v_2}(s) = \lambda_1 \nabla_{v_1}(s) + \lambda_2 \nabla_{v_2}(s).$$

In addition, if the connection is linear, we have

$$\nabla_v(\lambda_1 s_1 + \lambda_2 s_2) = \lambda_1 \nabla_v(s_1) + \lambda_2 \nabla_v(s_2)$$

and

$$\nabla_v(as) = v(a)s + a\nabla_v(s).$$

In fact, if you have a linear operator  $\nabla$  satisfying these properties, you can define coefficients

$$\Gamma_{i\beta}^\alpha = e_\alpha^*(\nabla_{\partial_i}(e_\beta))$$

and check these coefficients define a linear connection, so one can take a covariant derivative operator as an equivalent definition of a connection. **TODO:** Equivalence of existence of this operator with definition of connection, and check whether nonlinear connection derivative determines the connection.

If  $c : I \rightarrow M$  is a curve, and a bundle  $E$  on  $M$  equipped with a connection  $\nabla$ , we can pullback the connection to get a connection on  $c^*E$ , which, given that  $TI$  is one dimensional, we should really think of as a map  $D_t$  defined on sections  $s$  of  $E$  along  $c$ , given by

$$D_t s = \frac{Ds}{dt} = (\nabla_T s)(c(t)),$$

where  $T = \dot{c}$ . A curve  $c : I \rightarrow M$  is an autoparallel if and only if  $D_t s = 0$ . Indeed, we calculate that this is true in coordinates if and only if for all  $i$ ,

$$\frac{d^2 c^i}{dt^2} + \sum N_j^i(c') \frac{dc^j}{dt} = 0.$$

But this implies that  $c''(I) \subset H(TM)$ .

## 10.2 Affine Connections

We shall have need to consider affine connections on a vector bundle  $\pi : E \rightarrow M$ . An *affine connection* on  $E$  is a bilinear map  $\nabla : \Gamma(TM) \times \Gamma(E) \rightarrow \Gamma(E)$ , such that for  $a \in C^\infty(M)$ ,  $X \in \Gamma(TM)$ , and  $Y \in \Gamma(E)$ ,

$$\nabla_{aX}(Y) = a\nabla_X Y \quad \text{and} \quad \nabla_X(aY) = X(a)Y + a\nabla_X Y.$$

We call  $\nabla_X Y$  the *covariant derivative* of  $Y$  in the direction of  $X$ . The linearity in the  $X$  variable over  $C^\infty(M)$  means that  $(\nabla_X Y)(p)$  depends only on  $X_p$ , and so we may think of  $\nabla$  as a map  $\Gamma(E) \rightarrow \Omega^1(TM) \otimes \Gamma(E)$ .

The *fundamental theorem of Riemannian geometry* states that there is a unique affine connection, the *Riemannian connection* on  $M$ , which is *torsion free*, i.e.

$$\nabla_X Y - \nabla_Y X = [X, Y],$$

and *metric compatible*, i.e.

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle.$$

For an affine connection on  $TM$ , we can define the *torsion* of the connection to be the bilinear map  $\text{Tor}(X, Y) = (\nabla_X Y - \nabla_Y X) - [X, Y]$ . Thus the Riemannian connection is *torsion free* in the sense that  $\text{Tor} = 0$ .

TODO: Prove fundamental theorem.

Given a map  $\phi : N \rightarrow M$  and a vector bundle  $E \rightarrow M$ , we can consider the vector bundle  $\phi^*E$  on  $N$ , such that  $(\phi^*E)_p$  is equal to  $E_{\phi(p)}$ , with the smooth structure such that if  $\{e_j\}$  is a smooth frame on  $E$ , then  $e_j \circ \phi$  is a smooth frame on  $\phi^*E$ . If  $\nabla$  is an affine connection on  $E$ , then we naturally have an affine connection  $\phi^*\nabla$  on  $\phi^*E$ , which in a local frame  $\{e_j\}$ , is given by the formula

$$\phi^*\nabla_X(\sum a^j e_j)(p) = \sum [X(a^j)e_j + a^j \nabla_{\phi_*(X_p)} e_j](p).$$

One can check this formula is independent of the particular frame we choose. If  $M$  is a Riemannian manifold, and  $\nabla$  is the Riemannian connection on  $M$ , then the pullback connection satisfies

$$X\langle Y, Z \rangle = \langle \phi^*\nabla_X Y, Z \rangle + \langle Y, \phi^*\nabla_X Z \rangle.$$

Moreover, given vector fields  $X, Y \in \Gamma(TN)$ , we can interpret  $\phi_*Y$  and  $\phi_*Z$  as elements of  $\phi^*E$ , and then

$$\phi^*\nabla_X\{\phi_*(Y)\} - \phi^*\nabla_Y - \phi_*[X, Y] = 0.$$

When we are willing to abuse notation, we will write  $\phi^*\nabla$  as  $\nabla$ . These formulas will be useful when calculating variations associated with geodesics on a manifold.

The *curvature*  $R$  of a connection is a  $(1, 2)$  tensor on  $TM$  given by

$$R_{bc}^a(dx^b \wedge dx^c) \otimes \frac{\partial}{\partial x^a},$$

where  $R_{bc}^a = dx^a \left[ \frac{\delta}{\delta x^b}, \frac{\delta}{\delta x^c} \right]$   
 given by  $R_{bc}^a$   
 is given by

$$R = R_{bc}^a(dx^b \wedge dx^c)$$



# Chapter 11

## Geodesics

On Riemannian manifolds, we can define the length of any curve  $c : I \rightarrow M$  of bounded variation, by using the formula

$$L(c) = \int_I \sqrt{\langle \dot{c}, \dot{c} \rangle} dt.$$

We can thus make any *connected* Riemannian manifold into a metric space by defining

$$d(p, q) = \inf L(c),$$

where the infimum is taken over all piecewise smooth curves  $c : [0, 1] \rightarrow M$  from  $p$  to  $q$ , such that  $c'$  is non-vanishing at all but finitely many points.

It is clear the distance function  $d$  is symmetric. Modifying the concatenation of two paths slightly so that concatenation is smooth also shows the metric satisfies the triangle inequality. The next lemma justifies the final property that guarantees  $d$  is a metric.

**Theorem 11.1.** *For distinct  $p, q \in M$ ,  $d(p, q) \neq 0$ .*

*Proof.* Find a coordinate system  $(x, U)$  centered at  $p$ . If  $x(U)$  contains the closed ball  $B$  of radius  $\varepsilon > 0$  centered at the origin, then by smoothness,

$$|\sum g_{ij}(p)v_i, v_j|^2 \sim \sum v_i^2$$

with implicit constants uniform in  $B$ . Fix  $c : [0, 1] \rightarrow M$  in  $C$ . Let  $t_0$  be the first time at which  $|x(c(t_0))| = \varepsilon$ . We can thus identify the restriction of the curve  $c$  to the interval  $[0, t_0)$  with a curve in  $x(U)$ . Thus

$$L(c) \geq \int_0^{t_0} |c'(t)|_g \gtrsim \int_0^{t_0} \sqrt{\sum c_j'(t)^2} dt.$$

If we define  $r(t) = |c(t)|^2$ , then  $r(0) = 0$ ,  $r(t_0) = \varepsilon^2$ , and  $r'(t) = 2c(t) \cdot c'(t)$ . Thus

$$\varepsilon^2 = \int_0^{t_0} r'(t) = 2 \int_0^{t_0} c(t) \cdot c'(t) \leq 2\varepsilon \int_0^{t_0} |c'(t)| \lesssim L(c).$$

Thus  $L(c) \gtrsim \varepsilon$  with a uniform implicit constant. Thus  $d(p, q) > 0$ . □

*Remark.* A similar argument shows that the metric  $d$  gives  $M$  the same topology as its differentiable structure, i.e. because, locally in coordinates,  $d(p, q) \sim |p - q|$ .

The distance  $d(p, q)$  is not *always* realized by a particular curve. For instance, on  $\mathbb{R}^d - \{0\}$ , with the standard Riemannian metric, the induced metric is the standard one. Thus for all  $v \in \mathbb{R}^d - \{0\}$ ,  $d(v, -v) = 2|v|$ , but there is no curve from  $v$  and  $-v$  in  $\mathbb{R}^d - \{0\}$  with this length. We will see this is a lack of the *completeness* of  $\mathbb{R}^d - \{0\}$ , i.e. the induced metric is not a complete metric, or equivalently, the manifold is not *geodesically complete*, i.e. there are geodesics that cannot be continued indefinitely.

## 11.1 The First Variation For Arclength and Energy

We now analyze the behaviours of the minimizers of the length functional by using the calculus of variations. Fix a curve  $c : I \rightarrow M$  which is piecewise smooth, i.e. smooth away from  $a = t_0 < t_1 < \dots < t_n = b$ . Let  $\square = I \times (-\varepsilon, \varepsilon)$ , and consider a map  $\alpha : \square \rightarrow M$ , which is smooth which restricted to each box  $[t_i, t_{i+1}] \times (-\varepsilon, \varepsilon)$ , and such that if  $c_s(t) = \alpha(t, s)$ , then  $c_0 = c$ . We define the *first variation*  $\delta L(\alpha)$  to be the derivative of  $L(c_s)$  in the  $s$  variable when  $s = 0$ . We say  $c$  is a *geodesic* if  $\delta L(\alpha) = 0$  for all variations  $\alpha$  which *fix endpoints*, i.e. so that  $c_s(0) = c(0)$  and  $c_s(1) = c(1)$  for all  $s \in (-\varepsilon, \varepsilon)$ . We thus think of a geodesic as an extreme point of the length functional  $L$ . Any length minimizing curve must be a geodesic, and we will see that all sufficiently short geodesics are length minimizing.

Given a variation  $\alpha$ , define vector fields  $T$  and  $V$  by setting

$$T(t, s) = \alpha_*(\partial_t|_{(t,s)}) \quad \text{and} \quad V(t, s) = \alpha_*(\partial_s|_{(t,s)}).$$

We now compute a formula for  $\delta L(\alpha)$  in terms of  $T$  and  $V$ . By breaking up the interval  $I$  into pieces, we may assume without loss of generality that  $\alpha$  is smooth on  $I \times (-\varepsilon, \varepsilon)$ . Differentiating under the integral sign, using the metric compatibility of the affine connection on  $M$ , and letting

$$\hat{T} = \langle T, T \rangle^{-1/2} T$$

be the unit vector normalization of the vector field  $T$ , we conclude that

$$\begin{aligned} \delta L(\alpha) &= \frac{d}{ds} \int_I \langle T, T \rangle^{1/2} \\ &= \frac{1}{2} \int_I \frac{V\{\langle T, T \rangle\}}{\langle T, T \rangle^{1/2}} \\ &= \int_I \langle \nabla_V T, \hat{T} \rangle. \end{aligned}$$

Since  $[T, V] = 0$  and the affine connection on  $M$  is torsion free, we have  $\nabla_V T = \nabla_T V$ . Integrating by parts, using the identity

$$\frac{d}{dt} \langle V, \hat{T} \rangle = \langle \nabla_T V, \hat{T} \rangle + \langle V, \nabla_T \hat{T} \rangle,$$

if  $I = [a, b]$ , we conclude that

$$\delta L(\alpha) = \langle V, \hat{T} \rangle \Big|_a^b - \int_I \langle V, \nabla_T \hat{T} \rangle.$$

If our variation is only piecewise smooth, then decomposing into smaller pieces gives the formula

$$\delta L(\alpha) = \sum_i \langle V, \hat{T} \rangle \Big|_{t_i^+}^{t_{i+1}^-} - \int_I \langle V, \nabla_T \hat{T} \rangle.$$

This is the *first variation formula* for arclength.

Any vector field  $V$  on the curve  $c$  can be realized as the variation field for some variation. Moreover, if  $V(0) = V(1) = 0$ , then we can choose this variation to fix endpoints. One choice of such a variation is setting  $\alpha(s, t) = \exp_{c(t)}(sV)$ . But this means that  $\delta L(\alpha) = 0$  for all variations if and only if  $\hat{T}(t_i^-) = \hat{T}(t_i^+)$  for all  $0 < i < n$ , so that  $\hat{T}$  is a continuous vector field on  $I$ , and if  $\nabla_T \hat{T} = 0$  on  $c$ . That  $\hat{T}$  is continuous implies that the only reason  $c$  could fail to be  $C^1$  is if  $\langle \dot{c}, \dot{c} \rangle$  is not continuous.

There is another approach to defining geodesics. Instead of the length functional  $L$  directly, we can instead study the *energy functional*

$$E(c) = \frac{1}{2} \int_0^1 \langle \dot{c}, \dot{c} \rangle dt.$$

The advantage is that the integrand here is smooth at the origin, but the equation is now no longer invariant under reparameterization, and energy is minimized by minimizing changes of velocity. We can see this in two different ways:

- If  $s : [0, 1] \rightarrow [0, 1]$  is a monotone diffeomorphism, then

$$2E(c \circ s^{-1}) = \int_0^1 \frac{|c'(s^{-1}(s))|_g^2}{s'(s^{-1}(s))^2} ds = \int_0^1 |c'(t)|_g^2 s'(t)^{-1} dt.$$

Using the fact that  $\int_0^1 s'(t) dt = 1$ , Cauchy-Schwartz implies that

$$l = \int_0^1 |c'(t)|_g = \int_0^1 [s'(t)^{1/2}] [s'(t)^{-1/2} |c'(t)|_g] \leq \left( \int_0^1 s'(t)^{-1} |c'(t)|_g \right)^{1/2} \leq 2E(c \circ s^{-1}).$$

Moreover, this inequality is an equality if and only if  $s'(t) = \lambda |c'(t)|_g$ . Thus energy is minimized over all reparameterizations of  $c$  by that parameterization that makes  $c$  into a constant speed curve.

- We will now see minimizers of the energy equation are precisely the constant speed geodesics.

To obtain the second point, we compute the first variation  $\delta E(\alpha)$  using an analogous method to that for arclength. Applying the torsion free criterion, and integration by parts tells us that

$$\delta E(\alpha) = \int \langle \nabla_V T, T \rangle = 2 \int \langle \nabla_T V, T \rangle = \sum \langle V, T \rangle \Big|_{t_i^+}^{t_{i+1}^-} - \int \langle V, \nabla_T T \rangle.$$

This is the *first variation formula* for energy. We see now that  $c$  is an extremal for energy if and only if  $\nabla_T T = 0$  and  $T$  is a continuous vector field on  $c$ , i.e. so that  $c$  is  $C^1$ . Note then that

$$\partial_t \langle T, T \rangle = 2 \langle T, \nabla_T T \rangle = 0,$$

so if  $\nabla_T T = 0$ , the curve  $c$  must be constant speed, and then  $\nabla_T \hat{T} = 0$ . Thus the extremals for energy are precisely the *constant speed geodesics*.

## 11.2 The Euler-Lagrange Equation For Geodesics

We can also approach the study of geodesics from the study of Euler Lagrange equations. The Lagrangian we are considering in our length functional is

$$L(x, v) = \frac{1}{2} \sum g_{ij}(x) \dot{x}_i \dot{x}_j.$$

Now

$$\frac{\partial L}{\partial x_l} = (1/2) \sum_{i,j} \frac{\partial g_{ij}}{\partial x_l} v_i v_j,$$

and

$$\frac{\partial L}{\partial v_l} = \sum_i g_{li} v_i.$$

Thus

$$\frac{d}{dt} \left\{ \frac{\partial L}{\partial v_l}(c, c') \right\} = \sum_{i,j} \frac{\partial g_{li}}{\partial x_j} c'_i(t) c'_j(t) + \sum_r g_{lr}(c(t)) c''_r(t).$$

Geodesics must satisfy the Euler-Lagrange equation, which in our case reads that for each  $l$ ,

$$\frac{1}{2} \sum_{i,j} \left[ \frac{\partial g_{ij}}{\partial x_l} - 2 \frac{\partial g_{li}}{\partial x_j} \right] c'_i(t) c'_j(t) - \sum_i g_{li}(c(t)) c''_i(t) = 0.$$

We note that

$$2 \sum_{i,j} \frac{\partial g_{li}}{\partial x_j} = \sum_{i,j} \frac{\partial g_{li}}{\partial x_j} c'_i(t) c'_j(t) + \frac{\partial g_{lj}}{\partial x_i} c'_i(t) c'_j(t),$$

and so if we define

$$[ij, l] = \frac{\partial g_{li}}{\partial x_j} + \frac{\partial g_{lj}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_l},$$

then we conclude that

$$\frac{1}{2} \sum_{i,j} [ij, l] (c(t)) c'_i(t) c'_j(t) + \sum_r g_{lr}(c(t)) c''_r(t) = 0.$$

If we define

$$\Gamma_{ij}^l = \sum_k g^{lk} [ij, k]$$

then, inverting our equation by multiplying by the inverse metric coefficients, geodesics are the solutions to the equation

$$c''_l(t) + \sum_{i,j} \Gamma_{ij}^l(c(t)) c'_i(t) c'_j(t) = 0.$$

This is a *second order* ordinary differential equation with smooth coefficients, and thus has a unique smooth solution for small times, given some initial values  $c(0)$  and  $c'(0)$ . We conclude that *geodesics must be smooth*.

The homogeneity of the, Euler-Lagrange equation tells us that this family of solutions to the Euler-Lagrange equation, and so there exists a map  $\exp$ , defined on an open, star-shaped subset  $\Omega$  of  $TM$ , with codomain  $M$ , such that for any geodesic  $\gamma : [0, 1] \rightarrow M$  with  $\gamma(0) = p$  and  $\gamma(0)' = v$ ,  $\gamma(t) = \exp(p, tv)$ . We call  $\exp$  the *exponential map* of the manifold  $M$ .

The exponential map is a diffeomorphism in the tangent variables in a neighborhood of the origin. More precisely, if we define  $\phi(x, v) = (x, \exp(x, v))$ , then by the initial conditions condition on the exponential map,  $D_v \phi(0, 0)$  is the identity. Since  $\phi(x, 0) = (x, x)$ ,  $D_x \phi(0, 0) = I + I$ . This means that for any  $x_0$ , there exists an open neighborhood  $W$  of  $(0, 0)$  and an open set  $U \times U \subset M \times M$  such that  $\exp$  is a diffeomorphism from  $W$  to  $U \times U$ . This means that for any  $x_0, x_1 \in U$ , there exists a unique vector  $v$  such that  $(x_0, v) \in W$  and  $\phi(x_0, v) = (x_0, x_1)$ .

It is often handy to take the Euler-Lagrange equations, which are a family of *second order differential equations* on  $M$ , and convert them into a system of *first order differential equations* on  $T^*M$ . This is analogous to the conversion of Lagrangian mechanics to Hamiltonian mechanics in classical physics. Given a geodesic  $c(t)$ , consider the curve  $\xi(t) = c'(t)^b$  in  $T^*M$ , i.e. in coordinates, setting

$$\xi_i(t) = \sum_j g_{ij} c'_j(t).$$

If we substitute  $\xi$  into the Euler-Lagrange equation, we see that

$$\begin{aligned}
\frac{\partial L}{\partial x_l}(c(t), c'(t)) &= \frac{1}{2} \sum \frac{\partial g_{ij}}{\partial x_l} c'_i(t) c'_j(t) \\
&= \frac{1}{2} \sum_{i,j,a,b} \frac{\partial g_{ij}}{\partial x_l} g^{ia} \xi_a(t) g^{jb} \xi_b(t) \\
&= \frac{1}{2} \sum_{a,b} \left( \frac{\partial}{\partial x_l} \left\{ \sum_{i,j} g^{ia} g_{ij} g^{jb} \right\} \right. \\
&\quad \left. - \sum_{i,j} \frac{\partial g^{ia}}{\partial x_l} g_{ij} g^{jb} - \sum_{i,j} g^{ia} g_{ij} \frac{\partial g^{jb}}{\partial x_l} \right) \xi_a(t) \xi_b(t) \\
&= \frac{1}{2} \sum_{a,b} \left( \frac{\partial g^{ab}}{\partial x_l} - 2 \frac{\partial g^{ab}}{\partial x_l} \right) \xi_a(t) \xi_b(t) \\
&= -\frac{1}{2} \sum_{a,b} \frac{\partial g^{ab}}{\partial x_l} \xi_a(t) \xi_b(t).
\end{aligned}$$

We also calculate that

$$\frac{\partial}{\partial v_l}(c(t), c'(t)) = \sum_i g_{li} c'_i(t) = \xi_l(t),$$

Thus the Euler-Lagrange equation is equivalent to the equation

$$\frac{d\xi_l}{dt} = -\frac{1}{2} \sum_{a,b} \frac{\partial g^{ab}}{\partial x_l}(c) \xi_a \xi_b.$$

Together with the differential equation

$$\frac{dc_l}{dt} = \sum_a g^{la}(c) \xi_a,$$

we obtain a system of first order differential equations equivalent to the Euler-Lagrange equation. If we set

$$p(x, \xi) = \frac{1}{2} \sum g^{jk} \xi_j \xi_k,$$

and consider the *Hamiltonian vector field*

$$H_p = \frac{\partial p}{\partial \xi} \frac{\partial}{\partial x} - \frac{\partial p}{\partial x} \frac{\partial}{\partial \xi},$$

then  $H_p$  generates the system of differential equations above, and so we conclude (either by general Hamiltonian theory, or by a direct calculation) that  $p$  is *constant* on geodesics, which has the consequence that if  $c$  is a geodesic, then the quantity

$$\sum g_{ij}(c(t)) c'_i(t) c'_j(t)$$

is independent of  $t$ .

For each  $p \in M$ , the map  $\exp_p$  is a diffeomorphism from an open neighborhood of the point  $0_p \in T_p M$  to an open neighborhood  $U$  of  $p$ . Choosing some orthogonal basis for  $T_p M$ , and thus obtaining a diffeomorphism  $T_p M \rightarrow \mathbb{R}^n$ , we thus obtain a coordinate system  $(x, U)$  on  $M$ , centered at  $p$ , whose image is a ball  $B$  of radius  $\varepsilon > 0$  about the origin in  $\mathbb{R}^n$ . This is called a *normal coordinate system* at  $p$ .

**Lemma 11.2.** (*Gauss Lemma*) *Let  $(x, U)$  be a normal coordinate system centered at  $p$  onto a ball  $B$ . If  $G$  is the positive-definite matrix of coefficients such that*

$$g = \sum g_{ij} dx^i dx^j,$$

*then for all  $x \in B$ ,  $G(x)x = x$ .*

*Proof.* For each  $x$ , the curve  $c(t) = tx$  solves the Euler Lagrange equations. We have seen this implies that  $|c'(t)|_g$  is constant, and since  $G(0)$  is the identity, we find that

$$\sum g_{ij}(x) x_i x_j = |c'(1)|_g = |c'(0)|_g = |x|^2.$$

We have thus show that  $x^T G(x)x = |x|^2$ , and so all the remains to show that  $G(x)x = x$  is to prove that for any vector  $v$  orthogonal to  $x$ ,  $v^T G(x)x = 0$ .

For any such  $v$ , there exists a curve  $\gamma(t)$  such that  $\gamma(0) = x$ ,  $\gamma'(0) = v$ , and  $|\gamma'(t)| = |x|$  for all  $t$ . Define

$$\alpha(u, t) = u\gamma(t).$$

Then it suffices to show that

$$A(u, t) = \sum g_{ij}(\alpha) \frac{\partial \alpha}{\partial u} \frac{\partial \alpha}{\partial t}$$

is zero for all  $u$  and  $t$ . Expanding out derivatives shows that

$$\begin{aligned} \frac{\partial A}{\partial u} &= \sum_j \frac{\partial \alpha_j}{\partial t} \left( \sum_r g_{rj} \frac{\partial^2 \alpha_r}{\partial u^2} + \sum_{i,l} [il, j] \frac{\partial \alpha_i}{\partial u} \frac{\partial \alpha_l}{\partial u} \right) \\ &\quad + \sum_i \frac{\partial \alpha_i}{\partial u} \left( \sum_r g_{ir} \frac{\partial^2 \alpha_r}{\partial u \partial t} + \sum_{j,l} [jl, i] \frac{\partial \alpha_j}{\partial u} \frac{\partial \alpha_l}{\partial t} \right). \end{aligned}$$

Because each term is a geodesic for a fixed  $t$ , the first sum vanishes. We also calculate that if

$$B(u, t) = \sum g_{ij}(\alpha) \frac{\partial \alpha_i}{\partial u} \frac{\partial \alpha_j}{\partial u},$$

then

$$\frac{\partial B}{\partial t} = 2 \sum_i \frac{\partial \alpha_i}{\partial u} \left( g_{ir} \frac{\partial^2 \alpha_r}{\partial u \partial t} + \sum_{j,l} [jl, i] \frac{\partial \alpha_j}{\partial u} \frac{\partial \alpha_l}{\partial t} \right).$$

Thus  $\partial_t B = 2\partial_u A$ . But  $\partial_t B = 0$ , since  $B(u, t)$  is the magnitude of the vector  $\gamma$ , which is equal to  $|x|$  for all  $t$ . Thus  $\partial_u A = 0$ , and so we conclude that  $A$  is independent of  $u$ . So

$$A(u, t) = A(0, t) = \sum g_{ij}(0)\gamma_i(t)(0\gamma'_j(t)) = 0,$$

which completes the proof.  $\square$

There are some geometric tricks we can use to characterize geodesics on certain special manifolds. To do this, we introduce the notion of an *isometry*, which is an injective function  $f : M \rightarrow N$  between two Riemannian manifolds such that  $f^*g_N = g_M$ . An isometry clearly maps geodesics to geodesics. Let's characterize the geodesics on  $S^{n-1}$ . Fix a great circle  $C \subset S^{n-1}$ , and consider two points  $p, q \in C$  which possess a unique minimum length geodesic  $\gamma$  connecting them. The isometry given by reflection about  $C$  fixes  $p$  and  $q$ , and maps  $\gamma$  to another geodesic  $\gamma'$ , with the same length as  $\gamma$ . Thus  $\gamma = \gamma'$ . This can only be the case if  $\gamma$  is contained in  $C$ . Thus the geodesic is just given by travelling along the great circle, and thus we see these are all possible geodesics, since any geodesic must locally agree with a minimum length geodesic. Unless two points are antipodal, using the fact there is a unique great circle passing through these points, we conclude that there are exactly two geodesics between the two points, and one is clearly smaller than the other, so these points have a minimal length geodesic. The other geodesic is a critical point of the Euler-Lagrange equation, but is a 'saddle point' of the functional, i.e. because it does not even locally minimize a path between the points. Antipodal points have a one-dimensional family of geodesics passing between them, and each is a minimum geodesic, so a unique minimizer does not exist.

The geodesics on a cylinder  $S^1 \times \mathbb{R}$  can be given by unravelling the cylinder, i.e. considering an isometry  $I : (S^1 \times \mathbb{R}) - L \rightarrow \mathbb{R}^2$ , where  $L$  is some vertical line on the cylinder. Geodesics in  $S^1 \times \mathbb{R}$  must thus look like straight lines after applying  $I$ , and we thus see from this that there are infinitely many geodesics between any two points, but each pair has a unique minimum geodesic.

## 11.3 Geodesic Completeness

A Riemannian manifold is *geodesically complete* if the exponential map  $\exp$  is globally defined as a map from  $TM$  to  $M$ , i.e. so that geodesics can be extended infinitely. We now show this is the case if and only if the Riemannian metric on  $M$  gives  $M$  the structure of a complete metric space.

**Theorem 11.3.** *A Riemannian manifold  $M$  is geodesically complete if and only if it is a complete metric space. On a geodesically complete Riemannian manifold, any two points are connected by a minimal length geodesic.*

*Proof.* Suppose  $M$  is geodesically complete. Fix two points  $x_0, x_1 \in M$ . Consider a normal coordinate system  $(x, U)$  centered at  $x_0$ , where  $x(U) = 2B$ , where  $B$  is a ball of some small radius  $\varepsilon > 0$ , a small enough ball to guarantee that  $d(x_0, q) = |x(q)|$  for  $q \in U$ . Now let  $S$



denote the inverse image of  $\partial B$ . By compactness, there exists a point  $r \in S$  which minimizes  $d(x_1, r)$ . We claim that the radial geodesic  $\gamma$  obtained from extending the minimum geodesic from  $x_0$  to  $r$  eventually touches  $x_1$ , and moreover, is a minimal length geodesic between  $x_0$  and  $x_1$ . Since every curve from  $x_0$  to  $x_1$  must pass through  $S$ ,

$$d(x_0, x_1) = d(r, x_1) + \varepsilon.$$

Our proof would be completed if we could show that for all  $\varepsilon \leq t \leq d(x_0, x_1)$ ,

$$d(\gamma(t), x_1) = d(x_0, x_1) - t.$$

This is easily seen for  $\varepsilon \leq t \leq 2\varepsilon$  by working in the normal coordinate system around  $x_0$ . To prove the result for  $2\varepsilon \leq t \leq d(x_0, x_1)$ , let  $t_0$  be a supremum of  $t$  such that this equation held. If we have  $t_0 = d(x_0, x_1)$ , then by continuity our result would hold, so assume  $t_0 < d(x_0, x_1)$ . Let  $p = \gamma(t_0)$ , and form a normal coordinate system  $(x_0, U_0)$  around  $p$ , where  $x_0(U_0) = 2B_0$ , where  $B_0$  is a  $\varepsilon_0$  ball satisfying the analogous properties that  $B$  satisfies. Let  $r_0$  be the point on the inverse image  $S_0$  of the boundary of  $B_0$  which minimizes distance to  $x_1$ . Then we find that

$$d(r_0, x_1) = d(p, x_1) - \varepsilon_0 = d(x_0, x_1) - t - \varepsilon_0 = d(x_0, x_1) - d(x_0, p) - \varepsilon_0.$$

By the triangle inequality, we have

$$d(x_0, r_0) \geq d(x_0, x_1) - d(r_0, x_1) \geq \varepsilon_0 + t_0.$$

But this means that the path that travels from  $x_0$  to  $p$ , and then radially from  $p$  to  $r_0$ , is a minimum length path from  $x_0$  to  $r_0$ , which implies this path is a smooth geodesic. This is only the case if  $\gamma$  agree with the radial length geodesic. Thus we obtain a contradiction to the fact  $t_0$  was the upper bound, which concludes the construction of the minimal length geodesic.

We have thus proven that on any geodesically complete manifold, any two points are connected by a length minimizing geodesic. We claim it follows simply from this that  $M$  is a complete metric space. If  $K$  is a closed, bounded subset of  $M$ , fix a point  $p$ , and choose  $R$  such that  $K$  is contained in the open ball of radius  $R$  centered at  $p$ . Now if  $B \subset T_p M$  is the closed ball of radius  $R$  centered at the origin, then  $B$  is compact. By continuity,  $\exp_p(B)$  is compact. But  $K \subset \exp_p(B)$ , so  $K$  is compact. Thus  $M$  has the Heine-Borel property. It follows from this that  $M$  must be a complete metric space, by standard topological arguments.

Conversely, suppose that  $M$  is a complete metric space. Given some geodesic  $\gamma : (a, b) \rightarrow M$ , if  $\{t_i\}$  is a sequence converging to  $b$ , then  $\{\gamma(t_i)\}$  is a Cauchy sequence in  $M$ . Thus this sequence converges to some point  $\gamma(b)$ , and we can now use the uniqueness of geodesics to extend  $\gamma$ . A least upper bound argument shows that  $\gamma$  extends to a globally defined geodesic.  $\square$

## 11.4 The Second Variation Formula For Energy

For a curve  $c$ , we defined the quantity  $\delta E$  as a function on variations  $\alpha$ . But the first variation formula shows that this quantity depends only on the variation vector field  $W$  along  $c$ , and

so we can consider  $\delta E$  as a function of this vector field  $W$ , and then  $\delta E$  is a linear functional on the vector space of all such vector fields along  $c$ .

For a curve  $c : I \rightarrow M$ , we want to consider a second variation of energy. Thus we consider a 2-parameter variation  $\alpha : (-\varepsilon, \varepsilon)^2 \times [a, b] \rightarrow M$  with  $\alpha(0, 0, \cdot) = c(\cdot)$ , thus defining a family of curves  $c_{u_1, u_2}$  for each  $|u_1|, |u_2| < \varepsilon$ . We then define

$$\delta^2 E(\alpha) = \frac{\partial^2 E(c_u)}{\partial u_1 \partial u_2}.$$

The *second variation formula* says that for a variation of a constant speed geodesic if we define  $V_1$  and  $V_2$  to be the variation fields along  $c$  with respect to each parameter, then

$$\delta^2 E(\alpha) = - \sum \langle V_2, D_t V_1 \rangle \Big|_{t_i}^{t_{i+1}} - \int_a^b \langle V_2, D_t^2 V_1 + R(V_1, T)T \rangle$$

As a consequence, this quantity depends only on  $V_1$  and  $V_2$ , and so we write it as  $\delta^2 E(V_1, V_2)$ .

To prove the second variation formula, we can use metric compatibility and the lack of torsion to write

$$\begin{aligned} \delta^2 E(\alpha) &= \int_I \langle \nabla_{V_1} \nabla_{V_2} T, T \rangle + \langle \nabla_{V_2} T, \nabla_{V_1} T \rangle \\ &= \int_I \langle \nabla_{V_1} \nabla_T V_2, T \rangle + \langle \nabla_T V_2, \nabla_T V_1 \rangle. \end{aligned}$$

Now we use the fact that  $\nabla_{V_1} \nabla_T V_2 = \nabla_T \nabla_{V_1} V_2 + R(V_1, T)T$ . An integration by parts, using that  $\nabla_T T = 0$ , yields that

$$\int \langle \nabla_T \nabla_{V_1} V_2, T \rangle = \langle \nabla_{V_1} V_2, T \rangle \Big|_a^b.$$

Thus we have

$$\delta^2 E(\alpha) = \langle \nabla_{V_1} V_2, T \rangle \Big|_a^b + \int_I [\langle \nabla_T V_1, \nabla_T V_2 \rangle - \langle R(V_1, T)T, T \rangle].$$

For piecewise smooth variations, summing up the endpoint terms gives

$$\delta^2 E(\alpha) = \sum \langle \nabla_{V_1} V_2, T \rangle \Big|_{t_i^+}^{t_{i+1}^-} + \int_I [\langle \nabla_T V_1, \nabla_T V_2 \rangle - \langle R(V_1, T)T, T \rangle].$$

If we are consider a *smooth* variation rather than a piecewise smooth variation, and also assume that this variation fixes endpoints, then the endpoint terms completely vanish, and so

$$\delta^2 E(\alpha) = \int_I \langle \nabla_T V_1, \nabla_T V_2 \rangle - \langle R(V_1, T)T, T \rangle.$$

We call this quantity the *index form* of  $V_1$  and  $V_2$ , denoted  $I(V_1, V_2)$ . It is symmetric and bilinear over vector fields. A further integration by parts also allows us to write

$$I(V_1, V_2) = \langle \nabla_T V_1, V_2 \rangle \Big|_{t_i^+}^{t_{i+1}^-} - \int \langle \nabla_T \nabla_T V_1, V_2 \rangle + \langle R(V_1, T)T, T \rangle.$$

If  $\alpha : \square \rightarrow M$  is a variation fixing endpoints, defining a family of curves  $\{c_s\}$ , with  $c_0$  a geodesic and with  $V$  the associated vector field, then

$$E(c_s) = E(c_0) + s^2 \frac{I(V, V)}{2} + O(s^3).$$

It thus follows that if  $I(V, V) < 0$ , then  $E(c_s) < E(c_0)$  for  $s \neq 0$  and suitably close to zero. Thus  $c_0$  is not a minimizer of energy (and as a result, not a length minimizing geodesic). Conversely, if  $I(V, V) < 0$  for some vector field on  $c$  vanishing at endpoints, then we can find a variation of  $c$  under which  $c$  is not a minimum. It will be of use to introduce a notation for the set of all vector fields vanishing at the endpoints on  $c$ : we denote the space of such vector fields by  $\Omega_c$ .

Conversely, if  $I(V, V) \geq 0$  for all vector fields  $V \in \Omega_c$ , we would hope that  $c$  *actually is* a minimum length curve between two points. To do this, we begin by studying ‘critical points’ of the index form, vector fields  $J$  along  $c$  with  $I(J, V) = 0$  for all  $V \in \Omega_c$ . Vector fields with this property are called *Jacobi fields*. The second variation formula shows that a vector field  $J$  is Jacobi if and only if it satisfies the differential equation

$$\nabla_T \nabla_T J + R(J, T)T = 0.$$

This is a system of  $d$  second order ordinary differential equations. Thus the family of all Jacobi fields on a curve  $c$  is  $2d$  dimensional.

Two points  $a$  and  $b$  on a curve are *conjugate* if there exists a Jacobi field  $J$  on  $c$  such that  $J(a) = J(b) = 0$ . The multiplicity of the pair is then the dimension of the vector space of Jacobi fields vanishing at  $a$  and  $b$ . There exists Jacobi fields taking on every possible value of  $J(a)$  by existence and uniqueness of solutions to ODEs, and thus by rank-nullity, the space of Jacobi fields vanishing at  $a$  is  $d$  dimensional. A direct computation shows that the vector field  $J = (t - a)T(t)$  is a Jacobi field vanishing at  $a$ , but non-zero everywhere else. It thus follows that the multiplicity of a pair of conjugate points is *at most*  $d - 1$ .

This is tight. For instance, on the sphere  $S^d$ , antipodal points are conjugate along geodesics, with multiplicity  $d - 1$ . To see this, we note that every Jacobi field  $J$  arises as the variation field of a variation  $\alpha : \square \rightarrow M$  with the property that the resulting family of curves  $\{c_s\}$  are all geodesics. If  $J(a) = J(b) = 0$ , we can choose this family such that  $c_s(a) = c(a)$  for all  $s$ , but not of the endpoints, which only agree up to second order.

Then all geodesics beginning at one antipodal point meet at the other antipodal point, and there are a  $d - 1$  dimensional family of such geodesics, giving rise to  $d - 1$  linearly independent Jacobi fields vanishing at each point.

Now  $(at + b)T$  is a Jacobi field for all  $a$  and  $b$ , as may be computed directly. It follows that every Jacobi field  $J$  can be written uniquely as  $J^\perp + (at + b)T$  for a Jacobi field  $J^\perp$  everywhere perpendicular to  $c$ . In particular, if  $J$  is perpendicular to  $c$  at one point, it is perpendicular to  $c$  at every point.

TODO: Theorem 8.6 of Spivak Volume 4; If a curve  $c : [a, b] \rightarrow M$  contains a point conjugate to  $a$  in its interior, then there exists  $W$  with  $I(W, W) < 0$ , and thus  $c$  is not length minimizing.

## 11.5 TODO: Separate Jacobi Field Notes

Given a vector  $v \in T_p M$ , define a variation  $\alpha(t, s) = \exp_p(t(v + sw))$ . Then the variation field

$$J(t) = (\partial_s \alpha)(t, 0) = (\exp_p)_*(tv)(tw)$$

satisfies the differential equation  $\nabla_T J = \nabla_J T$ , where  $\nabla_r = (\exp_p)_*(tr)(\partial_r)$ .

# Chapter 12

## Extra

### 12.1 Connections

On  $\mathbb{R}^n$ , given two vector fields  $X$  and  $Y$ , we can define the ‘derivative’ of the vector field  $Y$  along  $X$  to be the vector field

$$(D_X Y)_p = \lim_{t \rightarrow 0} \frac{Y_{p+tX} - Y_p}{t}$$

On a manifold, there is no canonical way to differentiate vector fields; the Lie derivative  $L_X Y$  isn’t a satisfying definition because it measures how ‘independent’ two vector fields are, rather than measuring the change of  $Y$  along  $X$ . The main problem is that we are unable to ‘connect’ tangent spaces on a manifold close to one another, like we can in  $\mathbb{R}^n$ . On a Riemannian manifold, there is a canonical way to connect the fibres of a tangent bundle at points close to one another, and this is best explained through the concept of a connection.

Recall the definition of a smooth vector bundle  $\pi : E \rightarrow M$ , and smooth sections  $s : M \rightarrow E$ . We denote the  $C^\infty(M)$  module of all smooth sections by  $\Gamma(M)$ . A *connection* is a map  $\nabla : \Gamma(TM) \times \Gamma(E) \rightarrow \Gamma(E)$  which maps  $(X, s) \rightarrow \nabla_X(s)$ , which is  $C^\infty(M)$  linear in  $X$ ,  $\mathbb{R}$  linear in  $s$ , and satisfies the product rule

$$\nabla_X(fs) = X(f)s + f\nabla_X(s)$$

for any  $f \in C^\infty(M)$ , so that  $\nabla_X$  operates ‘like a derivative’ on  $s$ . An *affine connection* is a connection where  $E = TB$ .

**Example.** *The definition*

$$(\nabla_X Y)_p = D_X(Y)_p = \lim_{t \rightarrow 0} \frac{Y_{p+tX} - Y_p}{t}$$

*is an affine connection on  $\mathbb{R}^n$ .*

**Example.** *The Lie derivative is not an affine connection on a manifold, because it isn’t  $C^\infty(M)$  linear in the first variable.*

The fact that  $\nabla$  is  $C^\infty(M)$  linear in the 1st variable tells us that for a fixed section  $s$ , the map  $X \mapsto \nabla_X s$  acts ‘like a tensor’ in  $X$ . An argument analogous to the tensor characterization lemma reveals that  $(\nabla_X s)(p)$  depends only on the value  $X_p$  of  $X$  at  $p$ , not the entire vector field. On the other hand, for the sections  $s$  we only have  $\mathbb{R}$  linearity, so  $\nabla_X s$  is allowed to depend on more of the behaviour of  $s$ . A bump function type argument, analogous to the proof that derivations on  $C^\infty(M)$  are local, shows that  $(\nabla_X s)(p)$  depends only on the behaviour of  $s$  in a neighbourhood of  $p$ . The introduction of Christoffel symbols will show that, in fact,  $(\nabla_X s)(p)$  depends only on the behaviour of  $s$  on a curve tangent to  $X$ .

Suppose that we consider a trivialization  $U$  from  $\pi^{-1}(U) \rightarrow U \times \mathbb{R}^n$ , some local frame  $s_1, \dots, s_n$  on  $U$ , and a coordinate chart  $(x, U)$ . We know that we can define  $\nabla_X(s_k)$ , even though the sections  $s_k$  are not defined globally, because  $\nabla_X$  depends only on the behaviour of  $s_k$  locally. The *Christoffel symbols* with respect to this setup are functions  $\Gamma_{i\beta}^\alpha$  such that

$$\nabla_{e_i}(s_\beta) = \sum \Gamma_{i\beta}^\alpha s_\alpha$$

where  $e_i = \partial/\partial x_i$ . Then if  $X = \sum a^i e_i$ , and  $s = \sum b^\beta s_\beta$

$$\begin{aligned} \nabla_X(s) &= \sum \nabla_X(b^\beta s_\beta) = \sum X(b^\beta) s_\beta + b^\beta \nabla_X(s_\beta) \\ &= \sum X(b^\beta) s_\beta + b^\beta a^i \nabla_{e_i}(s_\beta) \\ &= \sum X(b^\beta) s_\beta + b^\beta a^i \Gamma_{i\beta}^\alpha s_\alpha \\ &= \sum (X(b^\alpha) + b^\beta a^i \Gamma_{i\beta}^\alpha) s_\alpha \end{aligned}$$

The Christoffel symbols and the  $a^i$  at  $p$  do not depend on  $s$ , and  $b^\beta(p)$  depends only on the values of  $s$  at  $p$ .  $X(b^\alpha)$  depends only on the behaviour of  $s$  tangent to  $X$ , and therefore  $\nabla_X(s)$  depends only on the behaviour of  $s$  locally on a curve tangent to  $X$ .

There are infinitely many degrees of freedom for defining an affine connection on a manifold. Any smooth family of Christoffel symbols gives rise to a connection on  $\mathbb{R}^n$ , and using a partition of unity type decomposition, one can patch together local connections to a global connection. The only thing to note here is that linear combinations of connections need not be a connection, but convex combinations are.

To introduce the natural choice of connection on a Riemannian manifold, we need to discuss how parallel transports arise from a particular connection. It turns out that if we have a connection on a smooth bundle  $\pi : E \rightarrow M$ , and  $\gamma$  is a curve on  $M$  from a point  $p$  to a point  $q$ , then there arises a linear transformation  $P_p^q : E_p \rightarrow E_q$  arising from the curve  $\gamma$ , called a parallel transport. Conversely, a family of parallel transports give rise to a connection.

## 12.2 Riemannian Submanifolds

Recall that if  $M$  is a Riemannian manifold with metric  $g$ , and  $N$  is an immersed submanifold with immersion  $i : M \rightarrow N$ , then we can give  $i^*(TM)$  a metric structure by the pullback metric  $i^*(g)$ . We can of course embed  $TN$  in  $i^*(TM)$ , and obtain a Riemannian metric

structure on  $TN$  by restricting  $i^*(g)$ , but we can also define another interesting bundle from this embedding. We define the *normal bundle* of  $N$  with respect to the embedding to be the orthogonal complement of  $TN$  in  $i^*(TM)$ .

**Theorem 12.1.** *The Levi-Civita connection  $\nabla$  relative to the induced metric on  $N$  is given by*

$$\nabla_X^N Y = P(\nabla_X^M Y)$$

For  $X, Y \in \Gamma(TM)$ , where  $P$  is the orthogonal projection of  $i^*(TM)$  onto  $TN$ .

*Proof.* It is straightforward to check  $\nabla^N$  defines an affine connection on  $TN$ . It suffices to check that the connection is metric compatible and torsion free. We find  $\nabla_X^N Y - \nabla_Y^N X = P(\nabla_X^M Y - \nabla_Y^M X) = P([X, Y]) = [X, Y]$ , since we know that if  $X$  and  $Y$  lie in a subbundle of the tangent bundle, then  $[X, Y]$  also lies in this subbundle. To check metric compatibility, we find

$$X_p(i^*g)(Y_p, Z_p) = X_p g(i_* Y_p, i_* Z_p) = g(\nabla_X^M(i_* Y_p), Z_p) + g(Y, \nabla_X^M Z) = g(\nabla_X^N Y, Z) + g(Y, \nabla_X^N Z)$$

and the uniqueness of the Levi-Civita connection gives our result.  $\square$

The *second fundamental form* of a pair  $N \subset M$  of Riemannian manifold is defined to be  $\mathbf{II}(X, Y) = Q(\nabla_X^M Y)$ , where  $Q$  is the orthogonal projection onto  $TN^\perp$ . It is  $C^\infty(M)$  bilinear and symmetric, and so corresponds to a smooth covariant two tensor field on  $i^*(TM)$ . The equation

$$\nabla_X^M Y = \nabla_X^N Y + \mathbf{II}(X, Y)$$

The manifold  $N$  is *totally geodesic* if  $\mathbf{II} = 0$ . The *mean curvature vector*  $H$  is the trace of  $\mathbf{II}$ , and is the gradient of the volume function on the manifold. A manifold is *minimal* if the mean curvature vector vanishes. Given a unit vector  $\nu \in N_p M$  normal to  $M$ , set  $A^\nu : T_p M \rightarrow T_p M$  by  $A^\nu(X) = -Q(\nabla_X^M \nu)$ , known as the *shape operator*. Then  $A^\nu$  is self-adjoint, and  $\langle \mathbf{II}(X, Y), \nu \rangle = \langle A^\nu(X), Y \rangle$ .

The Gauss equation says that

$$\langle R^N(X, Y)Z, W \rangle = \langle R^M(X, Y)Z, W \rangle + \langle \mathbf{II}(X, W), \mathbf{I}(Y, Z) \rangle - \langle \mathbf{II}(X, Z), \mathbf{II}(Y, W) \rangle$$

## 12.3 Jacobi Fields

Let  $\gamma$  be a geodesic. Then  $\gamma$  is locally minimizing if and only if  $\gamma$  has no conjugate points. Recall that  $I(V, V)$  is the index form

$$I(V, V) = \int_0^t |\nabla_{\partial_t} V|^2 - \langle R(V, \gamma')\gamma', V \rangle$$

A pair of *conjugate points* along a geodesic if there is a nonvanishing Jacobi field between the two points.

**Theorem 12.2.** *If  $\gamma$  has no conjugate points then  $I(\cdot, \cdot)$  is positive definite.*

We proved this theorem this morning, in the lecture I missed. This lecture we focus on the converse. This means that past a conjugate point pair, geodesics can fail to be locally minimizing.

**Theorem 12.3** (Index Inequality). *Let  $\gamma$  be a geodesic with no conjugate points, and let  $J$  be a Jacobi field along  $\gamma$ . If  $V$  is a vector field along  $\gamma$  with  $V(0) = J(0)$  and  $V(l) = J(l)$  then  $I(J, J) \leq I(V, V)$ , with equality if and only if  $J = V$ .*

*Proof.* Since  $\gamma$  contains no conjugate points,  $I(V, V) = 0$  for all  $V$  with  $V(0) = V(l) = 0$ , with  $I(V, V) = 0$  if and only if  $V = 0$ . In particular,  $I(J - V, J - V) = I(J, J - V) - I(V, J - V) = I(V, J - V) \geq 0$ , since  $I(J, J - V) = 0$ . But then  $I(J, J) = I(J, V)$ , which shows  $I(J, J) \leq I(V, V)$ .  $\square$

**Corollary 12.4.** *A geodesic containing an interior point which is conjugate to  $\gamma(0)$  is not locally minimizing.*

*Proof.* Let  $\gamma(t_0)$  be the first conjugate point to  $\gamma(0)$ . Then there is a nonzero Jacobi field  $J$  on  $\gamma_{[0, t_0]}$  with  $J(0) = J(t_0) = 0$ , which can be extended to a vector field  $X$  on all of  $\gamma$  by making the vector field vanish elsewhere. Then  $I(X, X) = 0$  on  $[0, t_0]$ . Note  $X$  is not smooth at  $t_0$ , since  $(\nabla_{\partial_t} J)(t_0) = 0$ , because this would imply  $J$  is zero everywhere. However, if  $\delta$  is sufficiently small, then  $\gamma$  contains no pairs of conjugate points around  $t_0$ , so there exists a local Jacobi field  $W$  with  $W(t_0 - \delta) = J(t_0 - \delta)$  and  $W(t_0 + \delta) = 0$ . The index inequality implies  $I(V, V) < I(X, X)$  on  $[t_0 - \delta, t_0 + \delta]$ . But  $X = V$  outside of the  $[t_0 - \delta, t_0 + \delta]$ , so  $I(V, V) < I(X, X) = 0$  on  $[0, l]$ , and so there is a variation of  $\gamma$  decreasing the length.  $\square$

*Remark.* A quantitative generalization of this theorem is given by the Morse index theorem. The Morse index is the maximal dimension of a subspace of variation fields on which  $I$  is negative definite. We just proved that if there is a conjugate point, the index is at least one. The Morse index theorem says the index is finite and equals the # of interior conjugate points, counted with multiplicity.

## 12.4 Nonpositive Curvature

Recall that given  $p \in M$  and a 2-plane  $\Pi \subset T_p M$ , we define the *sectional curvature*

$$K(\Pi) = \frac{\langle R(V, W)W, V \rangle}{|V|^2|W|^2 - \langle V, W \rangle^2}$$

We set  $K_M > k$  if  $K(\Pi) > k$  for all 2 planes  $\Pi \subset T_p M$ .

**Theorem 12.5.** *If  $K_M \leq 0$  then  $M$  has no conjugate points.*



*Proof.* Let  $\gamma$  be a unit speed geodesic on the manifold, and let  $V$  be a normal vector field along  $\gamma$  with  $V(0) = V(l) = 0$ . Then

$$\left. \frac{d^2 L(\gamma_s)}{ds^2} \right|_{s=0} = \int_0^l [|\nabla_{\partial_t} V|^2 - \langle R(V, \gamma', V) \rangle] \geq \int_0^l |\nabla_{\partial_t} V|^2 dt > 0$$

except when  $\nabla_{\partial_t} V = 0$ , but this only occurs if  $V = 0$  (uniqueness of initial conditions).  $\square$

**Theorem 12.6** (Cartan-Hadamard). *If  $M$  is a complete Riemannian manifold, and  $K_M \leq 0$ , then the exponential is a covering map.*

*Proof.* It suffices to note that if  $M$  is a complete Riemannian manifold, and  $F$  is a local isometry, then  $F$  is a covering map. We know that since  $K_M \leq 0$ ,  $M$  has no conjugate points, which is the set of points where the differential of the exponential map is non-invertible.  $\square$

We shall find that if  $K_M$  is strictly negative, then geodesics are unique.

## 12.5 March 22nd

Now we address problems related to positive curvature. Recall that the Ricci tensor is a symmetric quadratic form which is obtained by contracting the Riemann curvature tensor, i.e.

$$\text{Ric}(X, Y) = \sum_{n=1}^N \langle R(X, e_n) e_n, Y \rangle$$

It is the directional average of sectional curvatures. We say  $\text{Ric}_M \geq K$  if  $\text{Ric}_M(X, X) \geq K$  for all unit vectors  $X$ .

**Theorem 12.7** (Bonnet-Myers). *If  $(M, g)$  is complete, and  $\text{Ric}_M \geq (n-1)K > 0$ , then every geodesic of length  $\geq \pi/\sqrt{K}$  contains conjugate points. Thus the diameter of the manifold is less than or equal to  $\pi/\sqrt{K}$ .*

*Proof.* Let  $\gamma : [0, l] \rightarrow M$  be a unit length geodesic. Suppose  $L(\gamma) = l > \pi/\sqrt{K}$ . We will now show  $\gamma$  is not locally minimizing. Choose an orthonormal basis  $\{E_1, \dots, E_n\}$  of  $T_{\gamma(0)}M$ , and extend the basis by parallel transport. Set  $V_i = \varphi E_i$  where  $\varphi(0) = \varphi(l) = 0$ . We calculate

$$\begin{aligned} \sum I(V_i, V_i) &= - \sum \int_0^l \langle \nabla_{\partial_t} \nabla_{\partial_t} V_i, V_i \rangle + \langle R(V_i, \varphi') \varphi', V_i \rangle dt \\ &= - \int_0^l (n-1) \varphi'' \varphi + \varphi^2 \text{Ric}(\varphi', \varphi') \\ &\leq -(n-1) \int_0^l (\varphi + K\varphi) \varphi dt \end{aligned}$$

Setting  $\varphi(t) = \sin(\pi t/l)$ , which is the first laplacian eigenfunction on  $[0, l]$ , this causes the integral to be bounded by

$$-(n-1) \int_0^l [-(\pi/l)^2 + k] \varphi^2 < 0$$

provided  $l > \pi/\sqrt{K}$ . □

**Corollary 12.8.** *Let  $(M, g)$  be a complete manifold, with  $\text{Ric}_M \geq (n-1)k > ?$ . Then  $M$  is compact with a finite fundamental group.*

*Proof.* Every point in the manifold can be connected to be a geodesic of length at most  $\pi/\sqrt{K}$ , so it follows that the exponential map on the closure of a ball maps surjectively onto the whole space, hence the image is compact. To obtain that the image is a finite fundamental group, the universal cover of this manifold also satisfies the complete curvature condition, hence it is compact, and so the number of sheets is finite, hence the fundamental group is finite. □

**Example.** *Since the fundamental group of  $S^1 \times S^2$  is  $\mathbb{Z}$ , which is infinite, for every metric there is a plane with zero curvature. The Hopf conjecture is whether  $S^2 \times S^2$  has a product metric with  $K \geq 0$ .*

**Theorem 12.9** (Synge). *A compact orientable even dimensional manifold with positive sectional curvature*

*Proof.* Suppose  $M$  is not simply connected. Every nontrivial free homotopy class has a minimizing geodesic □

# Chapter 13

## Finsler Manifolds

Let  $V$  be a vector space. A *strongly convex Minkowski norm* is a function  $F : V \rightarrow [0, \infty)$ , smooth away from the origin, homogeneous of degree one, and such that for each  $v \in V$ , the Hessian of the function  $F^2$  on  $V - \{0\}$  is positive definite. The *indicatrix* of  $F$  is then  $F^{-1}(1)$ , and by homogeneity, uniquely determines  $F$ . A *Finsler metric* on a manifold  $M$  will then be a smooth choice of strongly convex Minkowski norms on each tangent space of  $M$ , and so it is profitable to start by studying such norms.

**Example.** Let  $Q : V \times V \rightarrow \mathbb{R}$  be an inner product on  $V$ . Then  $F(v) = Q(v, v)^{1/2}$  is a strongly convex Minkowski norm on  $V$ . All such inner products are equivalent by a change of basis. But this is not the case with Minkowski norms.

In general, strongly convex Minkowski norms are precisely norms that can be closely approximated up to second order by the square root of an inner product like above. For each  $v \in V - \{0\}$ , define an inner product  $Q_v$  on  $V$  by taking half the Hessian of  $F^2$  at the point  $v$ . One may then check that  $F(w)$  and  $Q_v(w, w)^{1/2}$  agree up to second order at  $v$ . In coordinates, we can write  $Q_v(w, u) = \sum g_{jk}(v) w^j u^k$  for constants

$$g_{jk} = F \partial_j \partial_k F + (\partial_j F)(\partial_k F).$$

We have the *fundamental inequality*  $dF(v)\{w\} \leq F(w)$ , with equality if and only if  $w$  is a scalar multiple of  $v$ . This inequality implies that  $F(v + w) \leq F(v) + F(w)$ , with equality if and only if  $w$  is a scalar multiple of  $v$ . Thus a strongly convex Minkowski norm is really a norm in the sense of functional analysis. One may view the fundamental inequality as a strengthening of Euler's theorem  $dF(v)\{v\} = F(v)$ , which holds for all homogeneous functions. Using the metric, we can convert the fundamental inequality to read

$$Q_v(v, w) \leq F(v)F(w),$$

with equality if and only if  $w$  is a positive scalar multiple of  $v$ , and then the fundamental inequality can be seen as a generalization of Cauchy-Schwartz.

**Example.** Consider the convex limaçon in  $\mathbb{R}^2$  given in polar coordinates by  $r = 3 + \cos \theta$ . It is described in Cartesian coordinates by  $f(x, y) = 0$ , where  $f(x, y) = (x^2 + y^2) -$

$3\sqrt{x^2 + y^2} - x$ . Using Okubo's technique, we can find a homogeneous function  $F$  equal to one on the convex limaçon by solving

$$f(z/F(z)) = 0,$$

and it is given by  $F(x, y) = (x^2 + y^2)/(3\sqrt{x^2 + y^2} + x)$ . This function is strongly convex.

**Example.** The quartic unit ball in  $\mathbb{R}^2$  given by the equation  $x^4 + y^4 = 1$  has a strictly convex interior, but is not the unit ball of a strongly convex metric. The associated function is  $F(x, y) = (x^4 + y^4)^{1/4}$ , but the Hessian becomes singular on the  $x$  and  $y$  axes. On the other hand, the perturbed functions

$$F(x, y) = \sqrt{(x^4 + y^4)^{1/4} + \varepsilon(x^2 + y^2)}$$

for  $\varepsilon > 0$  are strongly convex.

Now we consider the case of manifolds. A *Finsler Metric* on a manifold  $M$  is a homogeneous function  $F : TM \rightarrow [0, \infty)$ , smooth away from the zero section, such that for each  $x \in M$ , the function is a strongly convex Minkowski norm on  $T_x M$ . As a shorthand, we write  $\partial_{\alpha|\beta}$  for  $\partial_{x^\alpha} \partial_{v^\beta}$ , or  $\partial_\alpha$  or  $\bar{\partial}_\beta$  for  $\partial_{x^\alpha}$  and  $\partial_{v^\beta}$ .

**Example.** Let  $M$  be a Riemannian manifold, fix a one form  $\omega \in \Omega^1(M)$  with  $\|\omega(x)\|_g < 1$  for all  $x \in M$ , and define a function  $F(x, v) = \|v\|_g + \omega(v)$ . Then one can check that  $F$  is strongly convex, and thus gives a Finsler metric on  $M$ . To give an explicit example on  $M = \mathbb{R}^2 - \{0\}$ , for each  $x \in M$ , we define  $F$  so that at each  $x$ , the indicatrix of  $F$  is an ellipse of eccentricity  $e(x)$  with one focus at the origin, and such that the directrix of the ellipse associated with that focus given by the line passing through 0 perpendicular to the line segment from 0 to  $x$ . One can check that this metric is given by

$$F(x, v) = \frac{\sqrt{e(x)\rho(x, v) + \lambda(x, v)^2} + \lambda(x, v)}{e(x)},$$

where  $\lambda(x, v) = (x \cdot v)/|x|^2$  and  $\rho(x, v) = |v|^2/|x|^2$ . See (Antonelli, Ingarden, Matsumoto) for more details. If  $e$  is constant, then the geodesics of this Finsler manifold are logarithmic spirals through the origin.

**Example.** Consider the Finsler metric on the ball  $B$  of radius 2 at the origin with constant negative Gaussian curvature  $-1/4$ , of the form

$$F(x, v) = \frac{4|v|}{4 - |x|^2} + \frac{16|x|}{(4 - |x|^2)(4 + |v|^2)} \frac{v_1 + v_2}{|x|}.$$

The Finsler metric distance from the origin to the rim of the disk is infinite, but the distance from the rim of the disk to the origin is  $\log 2$ . We will see the geodesics of this Finsler manifold are precisely the hyperbolic trajectories on  $B$ .

Consider the *sphere bundle*  $\pi : SM \rightarrow M$  and *slit tangent bundle*  $\pi : TM - 0 \rightarrow M$ . The sphere bundle  $SM$  is obtained from  $TM - 0$  by projectivizing, i.e. identifying vectors with positive scalar multiples of themselves, or alternatively if  $M$  has a Finsler metric, by taking the submanifold of  $TM - 0$  of elements with norm one.

We will often work with the vector bundle  $B = \pi^*(TM)$  on  $TM$ , as well as the dual bundle  $B^*$ , which we can identify with  $\pi^*(T^*M)$ . Two natural sections  $l$  and  $\omega$  of  $B$  and  $B^*$  are given in coordinates by

$$l^i = F^{-1}v^i \quad \text{and} \quad \omega_i = \bar{\partial}_i F.$$

The vector field  $l$  is called the *distinguished section*, and  $\omega$  is called the *Hilbert form*, due to Hilbert's observation that for a curve  $c : [0, 1] \rightarrow M$ ,

$$\int_0^1 F\left(c, \frac{dc}{dt}\right) dt = \int_0^1 (\bar{\partial}_i F)\left(c, \frac{dc}{dt}\right) \frac{dc^i}{dt} dt = \int_c \omega.$$

This formula is also called *Hilbert's Invariant Integral*. The bundle  $B$  is also equipped with an inner product  $g$  induced by the Finsler metric on  $M$ , given in coordinates by the formula

$$g_{ij} = \bar{\partial}_{ij}\{F^2\}/2 = F(\bar{\partial}_{jk}F) + (\bar{\partial}_j F)(\bar{\partial}_k F),$$

and which we can also think of as a section of  $B^* \otimes B^*$  (in which case the inner product  $g$  is called the *fundamental tensor*). Conversely,  $B^*$  has an inner product induced by the dual Finsler metric on  $M$ , and the coordinates  $g^{ij}$  in coordinates are given precisely by the entries of the matrix given by taking the inverse of the matrix with coordinates  $g_{ij}$ . With respect to the metric  $g$ ,  $l$  and  $\omega$  are dual, i.e.  $\sum g_{ij}l^j = \omega_i$ , or for any vector  $w$ ,  $g(l, w) = \omega(w)$ .

We also define the *Cartan tensor*

$$A_{ijk} = \frac{F}{2} \bar{\partial}_k g_{ij} = \frac{F}{4} \bar{\partial}_{ijk} F^2.$$

This is the *Cartan tensor*. It is a symmetric tensor, a section of  $B^* \otimes B^* \otimes B^*$ . Sometimes, in the literature the tensor  $C_{ijk} = A_{ijk}/F$  is used as the definition of the Cartan tensor, but it is often more natural to work with tensors that are homogeneous of order zero so that they are naturally defined on the sphere bundle. Roughly speaking,  $g$  is the inner product that approximates  $F$  to second order, and  $A$  somehow provides a measure of how non-Riemannian the function  $F$  is; if  $F$  was induced by a Riemannian metric, then it is a theorem of Deicke, found in a 1953 paper, that  $A_{ijk} = 0$  uniformly if and only if the Finsler metric is Riemannian. Some useful formulas are that  $g(l, l) = \sum g_{ij}l^i l^j = 1$ , and that  $A_{ijk}l^i = 0$ .

We often choose to work with special frames  $\{e_a\}$  for  $B$ , which are  $g$  orthonormal, and satisfy  $e_n = l$ . The dual frame  $\{\omega^a\}$  to  $\{e_a\}$  then has the property that  $\omega^n = \omega$ . If we write  $e_a = \sum e_a^i \partial_i$  and  $\omega^a = \sum \omega_i^a dx^i$ , then  $\sum e_a^i \omega_i^b = \delta_a^b$ . Thus we have  $\partial_i = \sum \omega_i^a e_a$  and  $dx^i = \sum e_a^i \omega^a$ . For  $a, b \neq n$ ,

$$F(\bar{\partial}_{ij}F)e_a^i e_b^j = \delta_{ab},$$

because  $g(e_a, e_b) = \delta_{ab}$ , and  $\sum \bar{\partial}_i F e_a^i = \sum \omega_i^n e_a^i = \delta_a^n$ . Conversely, for  $a, b \neq n$ ,

$$\sum_i \omega_i^a \omega_i^b = F(\bar{\partial}_{ij} F).$$

This follows simply by inverting the last equation.

We define the *formal Christoffel symbols*

$$[jk, s] = \frac{\partial_j g_{sk} + \partial_k g_{sj} - \partial_s g_{jk}}{2},$$

and then define  $\gamma_{jk}^i$  by ‘raising indices’, defining

$$\gamma_{jk}^i = g^{is} [jk, s].$$

In the future, we will raise and lower indices by summing over the metric in this way. We also define the *nonlinear connection*

$$N_j^i = \gamma_{jk}^i v^k - C_{jk}^i \gamma_{rs}^k v^r v^s.$$

To make the nonlinear connection homogeneous of order zero, we consider

$$\tilde{N}_j^i = F^{-1} N_j^i = \gamma_{jk}^i l^k - A_{jk}^i \gamma_{rs}^k l^r l^s.$$

There is another formula, which proves more useful in the analysis of geodesics.

**Lemma 13.1.** *If we set  $G^i = \gamma_{ab}^i v^a v^b$ , then  $N_j^i = (1/2) \bar{\partial}_j G^i$ .*

*Proof.* Using homogeneity, we see that if  $G_a = g_{ab} G^b = \gamma_{ajk} v^j v^k$ , then  $\bar{\partial}_i G_a = 2\gamma_{aij} v^j$ , because

$$\begin{aligned} \bar{\partial}_i \{\gamma_{ajk}\} v^j v^k &= \frac{1}{2} (\partial_{j|i} g_{ak} + \partial_{k|i} g_{aj} - \partial_{a|i} g_{jk}) v^j v^k \\ &= \frac{\partial_{j|aik} \{F^2\} + \partial_{k|aij} \{F^2\} - \partial_{a|ijk} \{F^2\}}{4} v^j v^k = 0. \end{aligned}$$

And so

$$\bar{\partial}_i G_a = 2\gamma_{aij} v^j.$$

But now we see that

$$\begin{aligned} \frac{1}{2} \bar{\partial}_j G^i &= \frac{1}{2} \bar{\partial}_j \{g^{ik} G_k\} \\ &= \frac{1}{2} [\bar{\partial}_j \{g^{ik}\} G_k + g^{ik} \bar{\partial}_j \{G_k\}] \\ &= -\frac{1}{4} g^{it} \bar{\partial}_{tuj} \{F^2\} g^{uk} G_k + g^{ik} \gamma_{kja} v^a \\ &= -C_{uj}^i G^u + \gamma_{ja}^i v^a \\ &= \gamma_{jk}^i v^k - C_{jk}^i G^k \\ &= \gamma_{jk}^i v^k - C_{jk}^i \gamma_{rs}^k v^r v^s = N_j^i. \end{aligned}$$

□

None of the quantities  $\{N_j^i\}$ ,  $\{G^a\}$ , or  $\{\gamma_{jk}^i\}$  transform tensorially. But the quantities  $\tilde{N}_j^i$  have a somewhat nice transformation formula. The quantities  $\tilde{N}_b^a$  in a new coordinates  $\tilde{x}$  are given by the old quantities  $N_j^i$  by the formula

$$\tilde{N}_b^a = \frac{\partial \tilde{x}^a}{\partial x^i} \frac{\partial x^j}{\partial \tilde{x}^b} N_j^i + \frac{\partial \tilde{x}^a}{\partial x^i} \frac{\partial^2 x^i}{\partial \tilde{x}^b \partial \tilde{x}^s} \tilde{v}^s.$$

The coefficients  $N_j^i$  define a (nonlinear) Ehresman connection on  $TM - 0$ , and thus split  $T(TM - 0)$  into two subbundles  $H = H(TM - 0)$  and  $V = V(TM - 0)$ , and conversely split  $T^*(TM - 0)$  into  $H^\perp$  and  $V^\perp$ . We define  $\delta_j = \partial_j - N_j^i \bar{\partial}_i$  and  $\delta v^i = dv^i + N_j^i dx^j$ . Then  $H$  is spanned by  $\{\delta_j\}$ ,  $V$  is spanned by  $\{\bar{\partial}_j\}$ ,  $H^\perp$  is spanned by  $\delta v^i$ , and  $V^\perp$  is spanned by  $dv^i$ . Now  $TM - 0$  has a natural Riemannian metric

$$g_{ij} dx^i \otimes dx^j + g_{ij} \frac{\delta v^i}{F} \otimes \frac{\delta v^j}{F}$$

known as the *Sasaki metric*, which has the property that  $H$  and  $V$  are then orthogonal.

We consider an orthonormal frame  $\{\hat{e}_a\} \oplus \{\hat{e}_{n+a}\}$  and  $\{\hat{\omega}^a\} \oplus \{\hat{\omega}^{n+a}\}$  for  $H \oplus V$  and  $V^\perp \oplus H^\perp$  which are dual to one another. Then we can write

$$\hat{e}_a = \sum \hat{e}_a^i \delta_i \quad \text{and} \quad \hat{e}_{n+a} = \sum \hat{e}_{n+a}^i \bar{\partial}_i$$

and

$$\hat{\omega}^a = \sum \hat{\omega}_i^a dx^i \quad \text{and} \quad \hat{\omega}^{n+a} = \sum \hat{\omega}_i^{n+a} \frac{\delta v^i}{F}.$$

We will make the additional assumption that  $\hat{e}_{n+a}^i = \hat{e}_a^i$ , and thus that  $\hat{\omega}_i^{n+a} = \hat{\omega}_i^a$ . Moreover, we will assume that  $\hat{\omega}_n = \omega$  and  $\hat{e}_n = l$ .

We leave it an exercise to prove that for such a frame,

$$d(\log F) = \omega_i \frac{\delta y^i}{F} = \hat{\omega}^{n+n}.$$

It follows from this that  $d(\log F) \in H^\perp$ , and thus  $F$  is constant along every *horizontal curve* on  $M$ , i.e.  $\delta_i F = 0$  for all  $i$ .

A linear connection  $\nabla$  on the bundle  $B$  is *torsion free* if

$$\nabla_X Y - \nabla_Y X = [X, Y].$$

It is *metric compatible* if

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

There *does not exist* a  $g$ -compatible, torsion free linear connection on the bundle  $B$  unless the Finsler metric on  $M$  is induced by a Riemannian metric. There are thus *several* linear connections on  $B$  which approximate these properties. The *Chern connection* is the *unique* torsion free connection on  $B$  which is *almost metric compatible*, in the sense that

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) + 2A(\nabla_X l, Y, Z),$$

where  $C$  is the Cartan tensor. In coordinates, write  $\nabla \partial_i = \omega_i^j \partial_j$  for one-forms  $\omega_i^j$  (the notation overloads  $\omega$ , which is unfortunate, but we will try and keep these notations distinct by indexing elements of a frame by indices such as  $a$  and  $b$ , and indexing coordinates by indices such as  $i$  and  $j$ ), then the connection is torsion free if and only if for all  $j$ ,

$$\sum dx^j \wedge \omega_j^i = 0,$$

or equivalently, if  $\omega_j^i = \Gamma_{jk}^i dx^k$  for functions  $\Gamma_{jk}^i$  symmetric in  $j$  and  $k$ . If we write  $\Gamma_{i|jk} = g_{is} \Gamma_{jk}^s$ , then metric compatibility is equivalent to an expression of the form

$$\Gamma_{i|jk} = \frac{1}{2} (\delta_j g_{ik} + \delta_k g_{ij} - \delta_i g_{jk}).$$

This should be compared to the formula for  $\gamma_{i|jk}$ .

If we consider a special  $g$ -orthonormal basis for  $B$ , then we can also consider connection coefficients  $\tilde{\omega}_b^a$  such that

$$\nabla e_b = \tilde{\omega}_b^a e_a.$$

Then one can check that

$$\tilde{\omega}_b^a = (de_b^i) \omega_i^a + e_b^j \omega_j^i \omega_i^a.$$

Conversely,

$$\omega_j^i = (d\omega_j^a) e_a^i + \omega_j^b \tilde{\omega}_b^a e_a^i.$$

The torsion free criterion then reads that

$$d\omega^a - \omega^b \wedge \tilde{\omega}_b^a = 0$$

and almost metric compatibility reads that

$$\tilde{\omega}_b^a + \tilde{\omega}_a^b = -2A_{abc} \hat{\omega}^{n+c}.$$

These equations imply that  $\tilde{\omega}_n^n = 0$  (and more generally,  $\tilde{\omega}_a^n = -\tilde{\omega}_n^a$ ), since we have

$$A_{abc} = A_{ijk} e_a^i e_b^j e_c^k = F/4 \bar{\partial}_{ijk} \{F^2\} e_a^i e_b^j e_c^k.$$

Since  $\sum \bar{\partial}_{ijk} \{F^2\} l^i = 0$ , if  $n \in \{a, b, c\}$ , then  $A_{abc} = 0$ . We will mainly use the Chern connection, but there are other options of a linear connection. For instance, the *Cartan connection* is the connection with coefficients  $\omega_j^i + A_{jk}^i \delta v^k / F$ . It is metric compatible, but *has torsion*. We refer to (Bao-Chern) for other connections of common use.

Before we move on, we list several other formulas which prove useful. First,

$$\Gamma_{jk}^i l^j = F^{-1} N_k^i$$

and so if we write  $\nabla l = (\nabla l)^i \partial_i$ , then  $(\nabla l)^i = dl^i + l^j N_j^i$ , then

$$\frac{\delta v^i}{F} = dl^i + l^j \omega_j^i + l^i d(\log F) = (\nabla l)^i + l^i d(\log F).$$



Thus

$$\hat{\omega}^{n+a} = \omega_i^a \frac{\delta v^i}{F} = \omega^a(\nabla l) + \delta_n^a d(\log F).$$

But now we get that

$$g(l, \nabla l) = \omega(\nabla l) = \hat{\omega}^{n+n} - d(\log F) = 0.$$

For  $a \neq n$ , we have  $\hat{\omega}^{n+a} = \tilde{\omega}_n^a$ , but it is not true that  $\hat{\omega}^{n+n} = \omega_n^n$ . The latter is equal to zero, whereas the former is never zero.

## 13.1 Curvature Forms

The *curvature 2-forms* of the Chern connection are  $\Omega_j^i = d\omega_j^i - \omega_j^k \wedge \omega_k^i$ . They can be expanded as

$$\Omega_j^i = \frac{1}{2} R_{jkl}^i dx^k \wedge dx^l + P_{jkl}^i dx^k \wedge \frac{\delta y^l}{F} + \frac{1}{2} Q_{jkl}^i \frac{\delta y^k}{F} \wedge \frac{\delta y^l}{F}.$$

The objects  $R$ ,  $P$ , and  $Q$  are called the *hh*, *hv*, and *vv curvature tensors*. We will assume without loss of generality that  $R$  and  $Q$  are antisymmetric in the  $k$  and  $l$  index (though we will soon see that  $Q$  vanishes for the Chern connection).

TODO: Bianchi identities.

# Chapter 14

## Variation Formulas for Finsler Manifolds

Let  $\square = [0, r] \times (-\varepsilon, \varepsilon)$ , and consider

$$0 = t_0 < t_1 < \cdots < t_a = r.$$

Consider the space of all regular curves  $c : [0, r] \rightarrow M$  which are piecewise smooth on  $[t_i, t_{i+1}] \times (-\varepsilon, \varepsilon)$ . A *smooth variation* of a curve in  $X$  is a function  $\sigma : \square \rightarrow M$  which is piecewise smooth on each box  $[t_i, t_{i+1}] \times (-\varepsilon, \varepsilon)$  and with  $\sigma(\cdot, 0) = c(\cdot)$ .

Arising from any variation is a pair of vector fields

$$T = \frac{d\sigma}{dt} \quad \text{and} \quad U = \frac{d\sigma}{du},$$

given by differentiation in the first and second variables. They are well defined away from the times  $\{t_j\}$ . The map  $\sigma$  lifts to a map  $\hat{\sigma} : \square \rightarrow TM - 0$ , and thus gives rise to two vector fields

$$\hat{T} = \frac{d\hat{\sigma}}{dt} \quad \text{and} \quad \hat{U} = \frac{d\hat{\sigma}}{du}.$$

We consider a special frame  $\{e_s\}$  for the bundle  $B$  on  $TM - 0$ , and use this frame to write  $T = T^s e_s$  and  $U = U^s e_s$ . We also write

$$\hat{\sigma}^* \omega^s = a^s dt + b^s du \quad \text{and} \quad \hat{\sigma}^* \tilde{\omega}_s^r = A_s^r dt + B_s^r du.$$

Since  $\omega^s \in V^\perp$ ,

$$a^s = (\hat{\sigma}^* \omega^s)(\hat{\partial}_t) = \omega^s(\hat{T}) = \omega^s(T) = F(T)\omega^s(l) = \delta_n^s F(T).$$

Similarly,  $b^s = U^s$ ,  $A_s^r = \tilde{\omega}_s^r(\hat{T})$ ,  $B_s^r = \tilde{\omega}_s^r(\hat{U})$ . Since  $\tilde{\omega}_s^n = -\tilde{\omega}_n^s$ , the almost metric compatibility implies that  $A_n^s = B_n^s = A_s^n + A_n^s = B_s^n + B_n^s = 0$ . If we pull back the torsion free condition, then for  $s \neq n$ ,

$$\begin{aligned} 0 &= \hat{\sigma}^* [d\omega^s - \omega^r \wedge \tilde{\omega}_r^s] \\ &= d(b^s du) - (b^r du) \wedge (A_r^s dt + B_r^s du) - (F(T)dt) \wedge (A_n^s dt + B_n^s du) \\ &= \left( \frac{\partial b^s}{\partial t} + b^r A_r^s - F(T)B_n^s \right) (dt \wedge du), \end{aligned}$$

and for  $s = n$ , noting that  $dF$  lies in  $H^\perp$ ,

$$\begin{aligned}
0 &= \hat{\sigma}^* [d\omega^s - \omega^r \wedge \tilde{\omega}_r^s] \\
&= d(F(T)dt + b^n du) - (b^r du) \wedge (A_r^n dt + B_r^n du) - (F(T)dt) \wedge (A_n^n dt + B_n^n du) \\
&= \left( -dF(\hat{U}) + \frac{\partial b^n}{\partial t} + b^r A_r^n \right) (dt \wedge du) \\
&= \left( -\frac{\partial a^n}{\partial u} + \frac{\partial b^n}{\partial t} + b^r A_r^n \right).
\end{aligned}$$

Now the length  $L(u)$  of the curve  $t \mapsto \sigma(t, u)$  is given by

$$\sum_i \int_{t_i}^{t_{i+1}} F(T) dt = \sum_i \int_{t_i}^{t_{i+1}} a^n dt.$$

Thus

$$\begin{aligned}
L'(u) &= \sum_i \int_{t_i}^{t_{i+1}} \frac{\partial a^n}{\partial u} \\
&= \sum_i \int_{t_i}^{t_{i+1}} \left[ \frac{\partial b^n}{\partial t} + b^r A_r^n \right] dt \\
&= \sum_i b^n \Big|_{t_i}^{t_{i+1}} + \int_{t_i}^{t_{i+1}} b^r A_r^n dt.
\end{aligned}$$

If a curve is an extreme point among the space of functions under which we are considering variations, then  $L'(0) = 0$  for any variation. We may isolate the variation so that  $b^r = 0$  outside of a single interval  $[t_i, t_{i+1}]$ . It then follows that for any choice of functions  $\{b^r\}$  vanishing at  $t_i$  and  $t_{i+1}$ ,

$$\int_{t_i}^{t_{i+1}} b^r A_r^n dt = 0.$$

This is only possible if  $A_r^n = 0$  for all  $r$ . If a curve is  $C^1$  on  $[0, r]$  and  $A_r^n = 0$  for all  $r$ , then we will call that curve a *geodesic*. Noting that  $b^n = \omega(U) = g(U, T)/F(T)$ , it follows that for any variation of a geodesic curve, not necessarily fixed at the endpoints,

$$L'(u) = \sum_i \frac{g(U, T)}{F(T)} \Big|_{t_i}^{t_{i+1}}.$$

To work out a system of equations that a geodesic must satisfy, for  $r \neq n$  we write

$$A_r^n = \tilde{\omega}_r^n(\hat{T}) = -\tilde{\omega}_r^n(\hat{T}) = -\hat{\omega}^{n+r}(\hat{T}).$$

These quantities vanish if and only if  $\hat{T}$  lies in the span of  $\hat{e}_a$  for  $1 \leq a \leq n$ , i.e. if  $\hat{T}$  is a horizontal vector with respect to the Ehresman connection. Now we know that

$$\hat{T} = \sum \frac{dc^i}{dt} \partial_i + \frac{d^2 c^i}{dt^2} \bar{\partial}_i.$$

This vector can only be in the span of  $\hat{e}_a$  if

$$\hat{T} = \sum \frac{dc^i}{dt} \omega_i^a \hat{e}_a = Fl^i \delta_i.$$

Thus  $(c, c')$  is an integral curve of the vector field  $Fl^i \delta_i$ . The vector field  $Fl^i \delta_i$  is  $C^\infty$ , and thus it follows that geodesics are  $C^\infty$ . Note then that

$$\frac{d^2 c^k}{dt^2} = -Fl^i N_i^k = \frac{dc^i}{dt} N_i^k.$$

Recalling the function  $G$ , which is homogeneous of order two, we have

$$\frac{d^2 c^k}{dt^2} = -\frac{1}{2} \frac{dc^i}{dt} \bar{\partial}_i G^k = -G^k.$$

Thus geodesics are given by solutions to the equation  $\ddot{c} = -G(c, \dot{c})$ .

There is an alternate approach we can take to the problem, involving vector fields and covariant derivatives rather than differential forms. But there is some technicalities. In the usual, Riemannian approach to the problem, for a vector field  $W$  along a curve  $c : [0, 1] \rightarrow M$  with velocity vector  $T$ , we would define the covariant derivative as

$$D_T W = \left( \frac{dW^i}{dt} + W^j T^k \Gamma_{jk}^i \right) \partial_i,$$

The problem on a Finsler manifold is *which* coefficients  $\Gamma_{jk}^i$  to use; two natural options are to use  $\Gamma_{jk}^i(T)$  or  $\Gamma_{jk}^i(W)$ . We call the first the covariant derivative  $D_T W$  with reference vector  $T$ , and the second the covariant derivative  $D_T W$  with reference vector  $W$ . Now consider a variation of a curve  $c$ , with associated vector fields  $T$  and  $U$ .

**Lemma 14.1.**  $D_T U = D_U T$ , if  $T$  is used as the reference vector in both covariant derivatives.

*Proof.* The first vector field is

$$\left( \frac{dU^i}{dt} + U^j T^k \Gamma_{jk}^i(T) \right) \partial_i$$

and the second is

$$\left( \frac{dT^i}{du} + T^j U^k \Gamma_{jk}^i(T) \right) \partial_i$$

By equality of mixed derivatives,  $dU^i/dt = dT^i/du$ , and

$$U^j T^k \Gamma_{jk}^i(T) = T^j U^k \Gamma_{jk}^i(T)$$

because  $\Gamma_{jk}^i(T) = \Gamma_{kj}^i(T)$ . □

Differentiating the equation

$$L(u) = \int_I g_T(T, T)^{1/2}$$

and using the metric compatibility condition that

$$\partial_u \{g_T(T, T)\} = U\{g_T(T, T)\} = 2g(D_U T, T),$$

where  $D_U T$  is taken with reference vector  $T$ , we find that

$$L'(u) = \int_I \frac{g_T(D_U T, T)}{g_T(T, T)^{1/2}} = \int_I g_T(D_T U, l).$$

Using Metric compatibility again gives that

$$\partial_t \{g_T(U, l)\} = g_T(D_T U, l) + g_T(U, D_T l).$$

where covariant derivatives are taken with reference vector  $T$ . Integration by parts thus yields

$$L'(u) = \sum g_T(U, l)|_{t_i}^{t_{i+1}} - \int_I g_T(U, D_T l).$$

This quantity can only be zero for all variations (and thus all vector fields  $U$  fixed at the endpoints) if  $D_T l = 0$ . Working in a special orthogonal basis, we have  $D_T l = \tilde{\omega}_n^a(T) e_a$ , and so we recover the condition we obtained by working with differential forms. We now that if the curve has constant speed  $F(T) = A$ , then  $D_T l = A^{-1} D_T T$ , and so such a curve is a geodesic if and only if  $D_T T = 0$ . Conversely, if  $D_T T = 0$ , then metric compatibility implies that  $\partial_t \{g_T(T, T)\} = 0$ , i.e. so that the curve has constant speed, and then  $D_T l = 0$  follows immediately.

Now we consider the *second variation formula*. Consider a variation as above, using the same notation. Now pulling back the identity

$$d\omega_j^i - \omega_j^k \wedge \omega_k^i = \frac{1}{2} R_{jkl}^i \hat{\omega}^k \wedge \hat{\omega}^l + P_{jkl}^i \hat{\omega}^k \wedge \hat{\omega}^{n+l}.$$

Pulling back this identity gives that

$$\frac{\partial B_j^i}{\partial t} - \frac{\partial A_j^i}{\partial u} = A_j^k B_k^i - B_j^k A_k^i + R_{jnl}^i a^n b^l + \sum_{l \neq n} P_{jnl}^i a^n B_n^l - P_{jkl}^i b^k A_n^l.$$

If  $i = n$  and  $j \neq n$ , then the Bianchi identities imply that  $P_{inj}^n = 0$ , and so

$$\frac{\partial A_j^n}{\partial u} = \frac{\partial B_j^n}{\partial t} - A_j^k B_k^n + B_j^k A_k^n + \sum_{l \neq n} -R_{jnl}^n a^n b^l + P_{jkl}^n b^k A_n^l$$

Substituting this into the equation for length gives that

$$L''(0) = \left( \frac{\partial b^n}{\partial u} - \sum_i b^i B_n^i \right) \Big|_0^r + \int_0^r a^n (\sum_i B_n^i B_n^i - b^i R_{ij} b^j),$$

where  $R_{ij} = l^j R_{jkl}^i l^L$ .

# Chapter 15

## Finsler Manifolds

Riemann, in his famous 1854 lecture, began exploring the concepts of higher dimensional differential geometry that would influence the next century of development, by proposing a metric geometry of length obtained by measuring the length of infinitesimal tangent vectors, so as to measure the lengths of curves by integration. If the length of tangent vectors on a manifold is defined in terms of an inner product, we obtain the modern concept of a *Riemannian manifold*. If we instead consider a more general notion of length, namely, a *norm* on a vector space, then we obtain the theory of *Finsler manifolds*. Riemann mainly restricts attention to Riemannian manifolds in his lecture for the purposes of simplicity, but discusses the possibility of extending the results discussed to the more general setting. Here we explore the more general theory.

We begin with the theory of norms on finite dimensional vector spaces. Let  $V$  be a vector space. Then for a function  $G : V \rightarrow \mathbb{R}$ , we can define the *Hessian* of  $G$  at  $v \in V$  to be the quadratic form  $\text{Hess}G(v) : V \rightarrow \mathbb{R}$  given by

$$\text{Hess}G(v)(w) = \partial_t^2 \big|_{t=0} \{G(v + tw)\}.$$

With respect to a basis for  $V$ , the Hessian is the usual quadratic form induced by the matrix with entries  $\partial_{jk}G$ . We can also interpret  $\text{Hess}G$  as a bilinear form, i.e. by setting

$$\text{Hess}G(v)(w_1, w_2) = \partial_{t,s}^2 \big|_{t=0, s=0} \{G(v + sw_1 + tw_2)\}.$$

Let us suppose that  $F : V \rightarrow \mathbb{R}$  is a homogeneous function of order one, smooth away from the zero section of  $V$ , and *strongly convex*, so that for each  $v \in V - \{0\}$ , the Hessian  $G(v) = \text{Hess}\{F^2/2\}(v)$  is a positive-definite quadratic form. The convexity condition on  $F$  implies that  $F$  is then a Minkowski norm on  $V$ , i.e. satisfying the triangle inequality  $F(v + w) \leq F(v) + F(w)$ , with equality if and only if  $v$  and  $w$  are scalar multiples of one another. In this setting we can rephrase the triangle inequality in differential form, normally called the *fundamental inequality*, that  $DF(v)\{w\} \leq F(w)$ , with equality if and only if  $w$  is a scalar multiple of  $v$ . That  $DF(v)\{v\} = F(v)$  follows from Euler's Homogeneous Function Theorem. The fundamental inequality also implies a version of the Cauchy-Schwartz inequality for Minkowski metrics, i.e. that  $G(v)(v, w) \leq F(v)F(w)$ . We can see the proof in

coordinates by homogeneity; since  $\partial_k\{F^2/2\}$  is homogeneous of order one,

$$\sum_j \partial_j \partial_k \{F^2/2\}(v) v^j = \partial_k \{F^2/2\}(v) = F(v)(\partial_k F)(v).$$

Then summing in  $k$  gives that

$$\sum_{j,k} \partial_j \partial_k \{F^2/2\}(v) v^j w^k = F(v) \sum_k (\partial_k F)(v) w^k \leq F(v)F(w).$$

The inequality implies that the function  $F$  gives rise to a *family* of inner products on  $V$ , one for each  $v \in V - \{0\}$ , given by the quadratic form  $G(v)$ .

Now we introduce Finsler manifolds, which are manifolds  $M$  equipped with a function  $F : TM \rightarrow [0, \infty)$ , smooth away from the zero section of  $TM$ , such that  $F_p : T_p M \rightarrow [0, \infty)$  is homogeneous of order one and strongly convex. The function  $F$  is then known as a *Finsler metric* on  $M$ . The metric is *reversible* if  $F(-v) = F(v)$ . All Riemannian manifolds are Finsler manifolds, since we can use the inner product structure and define  $F(v) = \langle v, v \rangle^{1/2}$ . But the theory of Minkowski metrics is much smoother, i.e. there exist infinitely many non-isometric Minkowski metrics on a vector space, whereas all inner product structures on a vector space are isometric.

Given an oriented  $C^1$  curve  $c : [a, b] \rightarrow M$  on a Finsler manifold, we can then define its length by the usual formula, i.e.

$$\int F(c'(t)) dt,$$

and then the length is invariant under increasing diffeomorphisms of  $[a, b]$ , which follows from the homogeneity of  $F$ . If  $M$  is a *reversible* Finsler manifold, then the length of  $c$  is independent of the orientation, though in general the length of a curve traversed backwards need not have the same length as a curve traversed forwards. Taking the infimum of the length of paths, we can then define a quasi-metric  $d : M \times M \rightarrow [0, \infty)$ , which will be symmetric, and thus a metric, when  $M$  is reversible.

# **Part III**

## **Differential Topology**



Differential Topology is a subject closely aligned with differential geometry. The goal of differential geometry is to study various geometric structures that can be added to a manifold, i.e. a Riemannian metric, Lie group structure, and so forth. The goal of differential topology is to determine what geometric structures tell us about the topological structure of the underlying manifold. Just like in algebraic topology, we will naturally consider properties of manifolds which are homotopy invariant. Of course, since we will deal with smooth maps, we will also want to consider smooth homotopies, i.e. a smooth map  $H : [0, 1] \times M \rightarrow N$  for two functions  $f_0 : M \rightarrow N$  and  $f_1 : M \rightarrow N$  such that  $f_0(p) = H(0, p)$  and  $f_1(p) = H(1, p)$  for each  $p \in M$ . But there are various equivalences that allow us to reduce topological situations to the case of smooth maps.

# Chapter 16

## Intersection Theory

It is natural, given two manifolds  $M$  and  $N$  lying in an ambient space, to try and understand the geometric structure of the intersection  $M \cap N$ . In general, it is not possible to study this intersection from the perspective of differential geometry, since  $M \cap N$  can become quite pathological.

**Example.** Consider the plane  $\Sigma_0 = \mathbb{R}^k \times \{0\}$  in  $\mathbb{R}^d$ . We note the fact that if  $C \subset \mathbb{R}^k$  is any closed set, there exists a scalar-valued function  $f \in C^\infty(\mathbb{R}^k)$  such that  $f^{-1}(0) = C$ . The graph of  $f$  can be identified with a smooth submanifold of  $\mathbb{R}^d$ , given by

$$\Sigma_1 = \{(x, f(x)) \times \{0\} : x \in \mathbb{R}^k\}.$$

It follows that  $\Sigma_0 \cap \Sigma_1 = C \times \{0\}$ . Since all submanifolds of some ambient space locally look like this picture, and closed sets can be quite strange, we can expect arbitrary intersections of manifolds to behave quite pathologically.

Fortunately, this type of intersection only occurs in certain pathological cases; we will see that a ‘generic’ intersection behaves as our intuition might expect. This leads us to the notion of transversality. Recall that rank theorems we established in Chapter 2. Fix two manifolds  $M^n$  and  $N^m$  contained in some ambient manifold  $E^d$ . Given  $p \in M \cap N$ , there exists a chart  $(y, U)$  in  $E$  containing  $p$  such that

$$N \cap U = \{p \in U : y^{m+1}(p) = \cdots = y^d(p) = 0\}.$$

If  $i : M \rightarrow E$  is the inclusion map, and if the map  $g = (y^{m+1}, \dots, y^d) \circ i$  has constant rank, then the rank theorem implies that  $(U \cap M) \cap N = g^{-1}(0)$  is a smooth submanifold of  $M$ . Establishing this for a cover of  $M \cap N$  by charts  $U$  shows  $M \cap N$  is a smooth submanifold.

This seems like a tricky condition to verify, but we can find a simple intrinsic geometric condition in the special case where the map  $g$  is a *submersion* when composed with the coordinates  $(y^{m+1}, \dots, y^d)$ . For each  $p \in M \cap N$ ,  $T_p M$  and  $T_p N$  can be identified as subspaces of  $T_p E$ . We note that  $g$  is a submersion at  $p \in M \cap N$  precisely when

$$\left. \frac{\partial}{\partial y^{m+1}} \right|_p, \dots, \left. \frac{\partial}{\partial y^d} \right|_p \in T_p M.$$

But since

$$T_p N = \text{span} \left( \left. \frac{\partial}{\partial y^1} \right|_p, \dots, \left. \frac{\partial}{\partial y^m} \right|_p \right),$$

this occurs if and only if  $T_p M + T_p N = T_p E$ . In particular, given two manifolds  $M$  and  $N$ , we say  $M$  and  $N$  are *transverse* to one another if for each  $p \in M \cap N$ ,  $T_p M + T_p N = T_p E$ . It then follows that  $M \cap N$  is a smooth submanifold of  $M$  and  $N$ .

*Remark.* If we define the codimension  $\text{codim}_E(L) = \dim(E) - \dim(L)$  for any submanifold  $L$  of  $E$ , if  $M$  and  $N$  intersect transversally, then

$$\text{codim}_E(M \cap N) = \text{codim}_E(M) + \text{codim}_E(N),$$

at least if  $\text{codim}_E(M) + \text{codim}_E(N) \leq d$ ; if  $\text{codim}_E(M) + \text{codim}_E(N) > d$ , then  $M$  and  $N$  can only intersect transversally if  $M \cap N = \emptyset$ .

*Remark.* If we define  $V, W \subset \mathbb{R}^d$  to be transversal if  $V + W = \mathbb{R}^d$ , then two manifolds  $M$  and  $N$  are transversal precisely when  $T_p M$  and  $T_p N$  are transversal at each  $p \in M \cap N$ .

**Example.** Consider the hyperboloid

$$\Sigma_0 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 - z^2 = 1\}.$$

Fix  $a > 0$ , and let  $\Sigma_1$  be the sphere of radius  $a$  at the origin. Let us consider when  $\Sigma_0$  is transversal to  $\Sigma_1$ . Let us set  $f(x, y, z) = x^2 + y^2 - z^2$ , and  $g(x, y, z) = x^2 + y^2 + z^2$ . At each point  $p = (x, y, z) \in \mathbb{R}^3$ , we let

$$V(p) = \{v \in T_p \mathbb{R}^3 : df_p(v) = 0\}$$

and

$$W(p) = \{w \in T_p \mathbb{R}^3 : dg_p(w) = 0\}.$$

Then  $V(p)$  and  $W(p)$  are transversal precisely when  $V(p) \cap W(p)$  is one dimensional, which occurs precisely when  $\text{span}(df_p, dg_p)$  is two dimensional. We calculate that

$$\text{span}(df, dg) = \text{span}(x \cdot dx + y \cdot dy, z \cdot dz).$$

Thus  $V(p)$  and  $W(p)$  fail to be transversal if  $x = y = 0$ , or if  $z = 0$ . If  $p \in \Sigma_0 \cap \Sigma_1$ , then  $T_p \Sigma_0 = V(p)$  and  $T_p \Sigma_1 = W(p)$ . Thus  $\Sigma_0$  and  $\Sigma_1$  are transversal precisely when  $\Sigma_0 \cap \Sigma_1$  contains no points  $p = (x, y, z)$  with  $x = y = 0$ , or  $z = 0$ . But  $\Sigma_0$  contains no points with  $x = y = 0$ , so this situation is not a problem. If  $\Sigma_0 \cap \Sigma_1$  contains a point  $(x, y, 0)$ , then  $x^2 + y^2 = 1$ , and  $x^2 + y^2 = a^2$ , so  $a^2 = 1$ . Thus  $\Sigma_0$  intersects  $\Sigma_1$  transversally whenever  $a \neq 1$ . We find that if  $a < 1$ ,  $\Sigma_0 \cap \Sigma_1 = \emptyset$ , if  $a = 1$ , then  $\Sigma_0 \cap \Sigma_1$  is the unit circle in the  $xy$  axis, and if  $a > 1$ , then  $\Sigma_0 \cap \Sigma_1$  is the union of two circles above and below the  $xy$  axis.

We can also generalize transversality to talk about the intersections between the images of two maps  $f : M \rightarrow E$  and  $g : N \rightarrow E$ . We say  $f$  and  $g$  are *transverse* to each other if, for each  $p \in M$  and  $q \in N$  such that  $f(p) = g(q) = r$ ,  $f_*(T_p M) + g_*(T_q N) = T_r E$ . The map

$f \times g : M \times N \rightarrow E \times E$  is then a submersion at each point  $(p, q)$  such that  $f(p) = g(q)$ . Thus

$$M \times_E N = \{(p, q) \in M \times N : f(p) = g(q)\}$$

is a manifold with codimension equal to the dimension of  $E$ , known as the *pullback* of  $f$  and  $g$ , or the fibre product of  $f$  and  $g$ . The natural map from  $M \times_E N$  to  $E$  is denoted  $f \times_E g$ . It is the same as the categorical fibre product of the two spaces  $M$  and  $N$ . We will prefer to work with maps, rather than sets, because then the study of homotopies becomes natural. But we prefer the geometric intuition given to us by the intersections of manifolds. In the case where  $g$  is the inclusion map of  $N$  in  $E$ , then  $M \times_E N$  is equal to  $f^{-1}(N)$ , which is a manifold with boundary. To keep things ‘between’ concrete intersections and abstract fibre products, we prefer to discuss the fibre product between a map  $f : M \rightarrow E$  which is transverse to a submanifold  $N$  of  $E$  (by this we mean that the maps  $f : M \rightarrow E$  and the inclusion map  $i : N \rightarrow E$  are transverse). Then  $M \times_E N$  is diffeomorphic to  $f^{-1}(N)$ , which is a smooth submanifold of  $M$ . The general theory here isn’t really less general than discussing pairs of transverse maps  $f$  and  $g$ , since if  $f$  and  $g$  are transverse to one another, then  $M \times_E N$  is diffeomorphic to  $(f \times g)^{-1}(\Delta)$ , where  $f \times g : M \times N \rightarrow E \times E$  is the product map, and  $\Delta = \{(p, p) : p \in E\}$  is a submanifold of  $E \times E$  known as the *diagonal* of  $E$ .

Things only become slightly more technical if we deal with the intersection between a manifold with boundary and a manifold without boundary. Recall that if  $f : M \rightarrow N$ , where  $N$  is a manifold,  $M$  is a manifold with boundary, and  $q \in N$  is a regular value of both  $f$  and  $\partial f$ . Then  $f^{-1}(q)$  is a manifold with boundary, and  $\partial(f^{-1}(q)) = (\partial f)^{-1}(q)$ . Thus if  $M$  and  $N$  are submanifolds of some manifold  $E$ , where  $M$  is a submanifold with boundary, and both  $M^\circ$  and  $\partial M$  intersect  $N$  transversally. Then  $M \cap N$  is a manifold with boundary such that

$$\text{codim}_E(M \cap N) = \text{codim}_E(M) + \text{codim}_E(N),$$

and such that  $\partial(M \cap N) = (\partial M) \cap N$ . More generally, if  $M$  is a manifold with boundary, and  $N$  is a submanifold of a manifold  $E$ , and  $f : M \rightarrow E$  is a smooth map such that both  $f$  and  $\partial f$  are transverse to  $N$ , then  $f^{-1}(N)$  is a manifold with boundary, and  $\partial(f^{-1}(N)) = (\partial f)^{-1}(N)$ . The main reason we care so much about transversality is that it is a regularity condition which is essentially ‘generic’, i.e. any pair of manifolds may be deformed ever so slightly so that they intersect transversally, which makes the condition so useful.

**Theorem 16.1.** *Let  $N$ ,  $S$ , and  $E$  be manifolds, and  $M$  a manifold with boundary. Then consider two smooth maps  $H : S \times M \rightarrow E$  and  $g : N \rightarrow E$ , where  $H$  and  $\partial H$  intersects  $g$  transversally. For each  $s \in S$ , let  $f_s : M \rightarrow E$  be the smooth function  $f_s(p) = H(s, p)$ . If  $H$  and  $\partial H$  are transversal to  $N$ , then for almost every  $s \in S$ , both  $f_s$  and  $\partial f_s$  are transversal to  $g$ .*

*Proof.* At each point  $a = (s, p) \in S \times M$ , we know that  $T_a(S \times M)$  can be decomposed into  $T_a(S) \oplus T_a(M)$ . By the transversality assumption, if  $r = H(s, p)$ , and  $g(q) = r$ , then

$$T_r(E) = H_*(T_a(S)) + H_*(T_a(M)) + g_*(T_q(N)).$$

Now  $(f_s)_*(T_p(S)) = H_*(T_a(S))$ , so it suffices to show that for almost all  $s \in S$ , for every  $p \in M$  and  $q \in N$  with  $f_s(p) = g(q)$ ,

$$T_r(E) = H_*(T_a(M)) + g_*(T_q(N)).$$

By transversality,  $W = (S \times M) \times_E N$  is a manifold with boundary, and for each  $b = (s, p, q) \in W$ ,

$$T_b W = \{(v, w, u) \in T_a(S) \times T_a(M) \times T_q(N) : H_*(v) + H_*(w) = g_*(u)\}.$$

Consider the three projection maps  $\pi_S, \pi_M$ , and  $\pi_N$  from  $W$  onto  $S, M$ , and  $N$  respectively, and suppose  $s \in S$  is a regular value of  $\pi_S$ . Then for each  $p \in M$  and  $q \in N$  with  $f_s(p) = g(q)$ ,

$$T_a(S) = (\pi_S)_*(T_b(W)).$$

However, since  $H \circ (\pi_S \times \pi_M) = g \circ \pi_N$ ,

$$H_*(T_a(S) + (\pi_M)_*(T_a(M))) = (g \circ \pi_N)_*(T_b W) \subset g_*(T_q N).$$

But this implies that  $H_*(T_a(S)) \subset g_*(T_q N) + H_*(T_a(M))$ , which implies transversality. A similar argument involving  $\partial H$  and  $\partial \pi_S$  shows that  $\partial f_s$  is transverse for almost every  $s \in S$ .  $\square$

**Example.** Let  $M$  and  $N$  be submanifolds in  $\mathbb{R}^n$ . Then the transversality theorem implies that for almost every  $x \in \mathbb{R}^n$ ,  $M + x$  is transverse to  $N$ . Indeed, we consider the map  $H : M \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ , defined by letting  $H(p, x) = p + x$ . Then  $H$  is transverse to  $N$  since  $H$  is a submersion, so the transversality theorem implies that the map  $f_x : M \times \mathbb{R}^n$  given by letting  $f_x(p) = p + x$  is transverse for almost all  $x \in \mathbb{R}^n$ . But this means exactly that  $M + x$  is transverse to  $N$ .

**Example.** Recall that for each  $n, m, k$ ,  $M(n, m; k)$  is the space  $n \times m$  matrices of rank  $k$ . Then consider the map  $F : \mathbb{R}^k \times M(k, n; k) \rightarrow \mathbb{R}^n$  given by  $f(x, A) = Ax$ . Now

$$\frac{\partial f^i}{\partial a_{ij}} = x_j$$

and if  $j \neq i$ ,

$$\frac{\partial f^i}{\partial a_{jk}} = 0.$$

In particular, if  $x \neq 0$ , then this implies that, if the coordinates  $a_{ij}$  are appropriately ordered, then the derivative of  $f$  at  $x$  takes the form

$$\begin{pmatrix} x & 0 & \dots & 0 \\ 0 & x & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & x \end{pmatrix},$$

and therefore has full rank  $n$ . Thus  $f$  is a submersion. If  $M$  is a fixed submanifold without boundary, then  $F$  is transverse to  $M$ , which implies that for almost all  $A \in M(k, n; k)$ , the map  $x \mapsto Ax$  is transverse to  $M$ . We can interpret this as saying almost all  $k$  dimensional planes are transverse to  $M$ . Indeed, transferring this statement to a more geometric language, we see that for almost every  $V \in G(n; k)$ ,  $V$  is transverse to  $M$ .

To show that transverse maps are generic, it suffices to show that there *exists* a transverse deformation of any smooth map  $f : M \rightarrow E$ . Without loss of generality, by Whitney's approximation theorem, we can assume  $E \subset \mathbb{R}^K$  for some  $K$ . The idea here will be to deform the map in  $\mathbb{R}^K$ , and then correct back to a map into  $E$ . To begin with, we require a 'nonlocal' version of the inverse function theorem.

**Lemma 16.2.** *Let  $f : E \rightarrow F$  be a smooth map, let  $M$  be a submanifold of  $E$ , and suppose  $f$  maps  $M$  diffeomorphically onto  $f(M)$ , such that  $f_*|_p$  is an isomorphism between  $T_p(E)$  and  $T_p(F)$  for each  $p \in M$ . Then there exists an open set  $U$  containing  $M$  such that  $f$  maps  $U$  diffeomorphically onto  $F$ .*

*Proof.* We can of course usually apply the inverse function theorem locally, so we can consider a cover  $\{U_\alpha\}$  of open, precompact sets covering  $M$ , such that, restricted to  $U_\alpha$ ,  $f$  is a diffeomorphism between  $U_\alpha$  and  $f(U_\alpha)$ , which we will denote by  $V_\alpha$ . We may refine the cover  $\{U_\alpha\}$  such that the resultant cover  $\{V_\alpha\}$  is locally finite, each set  $U_\alpha$  is precompact, and each set  $U_\alpha$  satisfies  $f^{-1}(f(M)) \cap U_\alpha = M \cap U_\alpha$  for each  $\alpha$ ; to justify why we may assume the latter property, we note that if we were unable to find a neighbourhood of a point  $p \in M$  with this property, then we could find points a sequence of points  $\{p_k\}$  in  $E - M$  converging to  $p$ , such that  $f(p_k) \in f(M)$  for each  $k$ . Since  $f$  is a diffeomorphism between  $M$  and  $f(M)$ , we can find a secondary sequence of points  $\{p'_k\}$  in  $M$  converging to  $p$ , with  $f(p_k) = f(p'_k)$  for each  $k$ . This clearly contradicts the fact that  $f$  is a diffeomorphism in a neighbourhood of  $p$ .

Now let  $g_\alpha : V_\alpha \rightarrow U_\alpha$  be the inverse to  $f$ , restricted to  $U_\alpha$ . Let  $A$  be the set of all  $x \in \bigcup_\alpha V_\alpha$  such that, if  $x \in V_\alpha \cap V_\beta$  for some indices  $\alpha, \beta$ , then  $g_\alpha(x) = g_\beta(x)$ . Our proof will certainly be complete if we can show  $A$  contains an open neighbourhood of  $f(M)$ . If  $q \in f(M)$ , and  $f(p) = q$ , then  $q$  has a neighbourhood  $V$  which intersects finitely many of the  $V_\alpha$ , say  $V_{\alpha_1}, \dots, V_{\alpha_N}$ . If no neighbourhood of  $q$  is a subset of  $A$ , then we can find a sequence  $\{q_k\}$  converging to  $q$ , and distinct indices  $i, j \in \{1, \dots, N\}$ , such that  $g_{\alpha_i}(q_k) \neq g_{\alpha_j}(q_k)$  for each  $k$ . Since  $U_{\alpha_i}$  and  $U_{\alpha_j}$  are precompact, we may also assume, by thinning the sequence, that the sequences  $\{g_{\alpha_i}(q_k)\}$  and  $\{g_{\alpha_j}(q_k)\}$  converge to points  $p_i$  and  $p_j$  respectively. But  $f(g_{\alpha_i}(q_k)) = f(g_{\alpha_j}(q_k)) = q_k$ , so by continuity,  $f(p_i) = f(p_j) = f(p) = q$ , which implies  $p_i = p_j = p$ . Such a sequence cannot exist because  $f$  is a diffeomorphism in a neighbourhood of  $p$ . Thus  $A$  contains a neighbourhood of  $q$ , and since  $q$  was arbitrary, an open neighbourhood of  $f(M)$ .  $\square$

**Lemma 16.3.** *If  $M \subset \mathbb{R}^K$  is a compact boundaryless manifold, then for suitably small  $\varepsilon > 0$ , if  $M_\varepsilon$  is the  $\varepsilon$  thickening of  $M$  in  $\mathbb{R}^K$ , then for each  $x \in M_\varepsilon$ , there exists a unique  $\pi(x) \in M$*

closest to  $x$ , and moreover, the map  $\pi : M_\varepsilon \rightarrow M$  is a smooth submersion. If  $M$  is non-compact, there exists a function  $\varepsilon : M \rightarrow (0, \infty)$  such that if

$$M_\varepsilon = \{x \in \mathbb{R}^K : d(x, p) < \varepsilon(p) \text{ for some } p \in M\},$$

then each  $y \in M_\varepsilon$  is closest to a unique  $\pi(y) \in M$ , and the map  $\pi : M_\varepsilon \rightarrow M$  is a submersion.

*Proof.* We consider the normal bundle  $N(M)$  to  $M$  in  $\mathbb{R}^K$ , which for each  $p \in M$ , consists of the orthogonal complement to  $T_p(M)$  in  $T_p(\mathbb{R}^K)$ . If  $T(M)$  is locally spanned by a frame  $v_1, \dots, v_n$ , then the Gram-Schmidt orthogonalization process can be used to produce a smooth frame  $w_1, \dots, w_{K-n}$  for  $N(M)$ . Thus  $N(M)$  is a smooth bundle. We now define  $F : N(M) \rightarrow \mathbb{R}^K$  by letting  $F(v_x) = x + v$ . We can identify  $M$  with the subset of  $N(M)$  consisting of  $0_p$ , for all  $p \in M$ , and  $F$  restricts to a diffeomorphism on this set. Since  $F$  has full rank at each point on  $M$ , there exists a neighbourhood  $U$  of  $M$  in  $N(M)$  on which  $F$  is a diffeomorphism. Clearly there exists a function  $\varepsilon : M \rightarrow (0, \infty)$  such that  $F(U)$  contains  $M_\varepsilon$ , or if  $M$  is compact, a single scalar  $\varepsilon > 0$  which works for all  $p \in M$ . The projection map  $\pi : M_\varepsilon \rightarrow M$  is defined by taking the inverse of  $F$  on  $M_\varepsilon$ , and then using the bundle structure of  $N(M)$  to project down to  $M$ . Clearly, this map is a submersion. All that remains now is to show that if  $M$  is compact, so that  $\varepsilon > 0$  is a scalar, then  $\pi(x)$  is the unique closest point on  $M$  to  $x$ . Certainly  $\pi(x)$  is a critical point of the map  $\rho : M \rightarrow [0, \varepsilon)$  defined by setting  $\rho(p) = |p - x|^2$ , since the critical points are precisely the  $p \in M$  such that  $x - p \in N_p(M)$ . Since  $F$  is invertible,  $\pi(x)$  is the unique point such that there exists  $v \in N_{\pi(x)}(M)$  with  $|v| < \varepsilon$  and such that  $\pi(x) + v = x$ . And since  $M$  is compact, a minimum is attained, so  $\pi(x)$  must be precisely this minimum.  $\square$

*Remark.* This is an instance of the *tubular neighbourhood theorem* in Riemannian geometry. Given any submanifold  $M$  of a Riemannian manifold  $E$ , we can consider the normal bundle  $N(M; E)$ , such that at each  $p \in M$ ,  $N_p(M; E) = T_p(N)^\perp$ . Then for each  $p \in M$ , there exists a neighbourhood  $U$  of  $p$  in  $E$ , and a neighbourhood  $V$  of  $0_p \in N(M; E)$ , and a diffeomorphism  $f : V \rightarrow U$  such that  $f(0_q) = q$  for any  $0_q \in V$ . The proof is essentially the same in the more general case.

**Corollary 16.4.** *Let  $f : M \rightarrow N$  be a smooth map, where  $\partial M = \emptyset$ . Then there exists some  $K$  and some smooth map  $F : M \times B \rightarrow N$ , where  $B$  is the open unit ball in  $\mathbb{R}^K$ , such that  $F(p, 0) = f(p)$  for all  $p \in M$ , and for any  $p \in M$ , the map  $s \rightarrow F(p, s)$  is a submersion from  $B$  to  $N$ . In particular, this means both  $F$  and  $\partial F$  are submersions.*

*Proof.* We may assume that  $N$  is a submanifold of  $\mathbb{R}^K$ , and then we set  $F(p, x) = \pi(f(p) + \varepsilon(p)x)$ , where  $\varepsilon : M \rightarrow [0, \infty)$  is a smooth function such that there exists a submersion  $\pi : N_\varepsilon \rightarrow N$  (we can always choose  $\varepsilon$  to be smooth by using a partition of unity type approach). Then  $F$  obviously satisfies the required properties.  $\square$

Thus we can now apply the transversality theorem to conclude that for any submanifold  $L$  of  $N$  without boundary, and for almost all  $x \in B$ , the map  $p \mapsto \pi(f(p) + \varepsilon(p)x)$  is a transverse map both from  $M \rightarrow N$  and from  $\partial M \rightarrow N$ . In particular, we conclude directly that any smooth map between two manifolds is homotopic to a map which is transverse to any given submanifold  $L$ .

**Corollary 16.5.** *Let  $M$  be a compact manifold with boundary,  $E$  a Riemannian manifold, and  $N$  a submanifold of  $E$ . Let  $f : M \rightarrow E$  be smooth. Then for any  $\varepsilon > 0$ , there exists a map  $f_0 : M \rightarrow E$  transverse to  $N$  with  $\|f - f_0\|_{L^\infty(M)} \leq \varepsilon$ .*

Often, we will need to extend ‘partially’ transverse maps to fully transverse maps in a compatible way. Given maps  $f : M \rightarrow E$  and  $g : N \rightarrow E$ , and two sets  $C_1 \subset M$  and  $C_2 \subset N$ , we will say  $f$  is transverse to  $g$  on the pair  $(C_1, C_2)$  if at each point  $p \in C_1$  and  $q \in C_2$  with  $f(p) = g(q) = r$ ,

$$f_*(T_p(M)) + g_*(T_q(N)) = T_r E.$$

We note that  $f$  and  $g$  are transverse on a pair  $(C_1, C_2)$ , then there exists open neighbourhoods  $U_1$  and  $U_2$  containing  $C_1$  and  $C_2$  such that  $f$  and  $g$  are transverse on  $(U_1, U_2)$ .

**Theorem 16.6.** *Let  $N$  and  $E$  be manifolds, and let  $M$  be a manifold with boundary. Let  $f : M \rightarrow E$  and  $g : N \rightarrow E$  be smooth maps, and consider two closed sets  $C_1 \subset M$  and  $C_2 \subset N$ . Suppose  $f$  and  $\partial f$  are transverse to  $g$  on  $(C_1, C_2)$ . Then there exists a smooth map  $f_0 : M \rightarrow E$  homotopic to  $f$  and  $g_0 : N \rightarrow E$  homotopic to  $g$ , such that  $f_0$  and  $\partial f_0$  are transverse to  $g$ , and  $f_0$  and  $g_0$  agree with  $f$  and  $g$  on  $C_1$  and  $C_2$ .*

*Proof.* Let  $f$  and  $g$  be transverse on a pair of open sets  $U_1$  and  $U_2$ . Let  $\gamma_1$  and  $\gamma_2$  be Urysohn functions for  $U_1^c$  and  $U_2^c$  i.e. equal to one on  $U_1^c$  and  $U_2^c$ , but vanishing in a neighbourhood of  $C_1$  and  $C_2$ . Then define a function  $F_0 : M \times B \rightarrow E$  by setting

$$F_0(p, x) = F(p, \gamma_1(p)^2 \cdot x),$$

where  $F : M \times B \rightarrow E$  is the transverse map constructed previously. Similarly, define

$$G_0(p, x) = G(p, \gamma_2(p)^2 \cdot x),$$

where  $G : N \times B \rightarrow E$  is the transverse map constructed previously. We claim that if we consider

$$F \times G : (M \times B) \times (N \times B) \rightarrow E \times E,$$

then  $(F \times G)$  and  $\partial(F \times G)$  is transverse to the diagonal  $\Delta$  in  $E \times E$ . This is certainly true at any point  $(p, x_1, q, x_2)$  where  $\gamma_1(p)$  or  $\gamma_2(q)$  are nonzero, since either  $F$  or  $G$  are submersions at this point. But if  $\gamma_1(p) = \gamma_2(q) = 0$ , then  $p \in U_1$  and  $q \in U_2$ , and  $(F \times G)$  operates as  $(p, x_1, q, x_2) \mapsto (f(p), g(q))$ . Since  $f$  and  $\partial f$  are transverse to  $g$  on  $(U_1, U_2)$ ,  $F \times G$  are transverse to  $\Delta$  here. Thus we conclude that for almost all  $(x_1, x_2) \in B$ , the maps  $f_0(p) = F(p, x_1)$  and  $g_0(q) = G(q, x_2)$  are transverse, which gives the maps we were required to produce.  $\square$

Since  $\partial M$  is always a closed subset of  $M$ , the following is immediate.

**Corollary 16.7.** *If  $f : \partial M \rightarrow E$  is a map transverse to a map  $g : N \rightarrow E$ , then  $f$  extends to a map from  $M$  to  $E$  if and only if it extends to a map from  $M$  to  $E$  which is transverse to  $g$ .*



## 16.1 Intersection Theory Modulo Two

In this chapter, we exploit a simple fact about one dimensional manifolds, which follows from the fact that the only compact one manifolds with boundary up to diffeomorphism are  $S^1$  or  $[0, 1]$ . Thus *any* compact one manifold with boundary  $M$  has an *even number* of boundary points. Though obvious, the theorem itself is sometimes quite powerful. Let us consider a simple example.

**Theorem 16.8.** *If  $M$  is a compact manifold with boundary, then there exists no smooth retraction  $r : M \rightarrow \partial M$ .*

*Proof.* Suppose such a map  $r : M \rightarrow \partial M$  existed. By Sard's theorem, there exists a regular value  $p \in \partial M$ , and thus  $r^{-1}(p)$  is a one manifold with boundary. In particular, this implies  $\partial(r^{-1}(p))$  consists of an even number of points, which is impossible since  $\partial(r^{-1}(p)) = (\partial r)^{-1}(p) = \{p\}$ .  $\square$

**Corollary 16.9** (Brouwer). *Let  $B$  be a closed unit ball in any dimension. Then any smooth map  $f : B \rightarrow B$  must have a fixed point.*

*Proof.* If  $f$  has no fixed point, then we obtain a smooth map  $g : B \rightarrow S^1$  by letting  $g(x)$  be the unique point on the boundary on the line between  $x$  and  $f(x)$ , which gives a contradiction because  $g$  is a contraction map. To verify smoothness, we can write  $g(x) = x + t(x)(f(x) - x)$ , where  $t(x) > 0$  is the unique value such that

$$|t(x)(f(x) - x) + x|^2 = 1.$$

Expanding this out using the inner product, we find

$$t(x)^2 |f(x) - x|^2 + 2t(x)[(f(x) - x) \cdot x] + [|x|^2 - 1] = 0$$

This polynomial can be written as  $A(x)t(x)^2 + B(x)t(x) + C(x) = 0$  for smooth maps  $A, B$ , and  $C$ . Thus

$$t(x) = \frac{-B(x) + \sqrt{B(x)^2 - 4A(x)C(x)}}{2A(x)},$$

which is smooth because  $A(x) = |f(x) - x|^2$  never vanishes.  $\square$

We now exploit our simple fact to give a simple topological invariant relating the intersections between two maps. Suppose  $M, N$ , and  $E$  are manifolds, where  $M$  is compact, and  $N$  is a closed submanifold of  $E$ , and the dimension of  $M$  and  $N$  are *complementary*, in the sense that  $\dim(M) + \dim(N) = \dim(E)$ . If  $f : M \rightarrow E$  is a smooth map which is transverse to  $N$ , then  $f^{-1}(N)$  is a closed, zero dimensional submanifold of  $M$ , hence consisting of finitely many points. We can therefore define  $I_2(f, N) \in \mathbb{Z}_2$  to be the parity of the number of points in  $f^{-1}(N)$ . Then  $I_2(f, N)$  is a homotopy invariant of  $f$ .

**Theorem 16.10.** *If  $f_0, f_1 : M \rightarrow E$  are homotopic maps, both transverse to  $N$ , then  $I_2(f_0, N) = I_2(f_1, N)$ .*

*Proof.* We can find a homotopy  $H : [0, 1] \times M \rightarrow E$  between  $f_0$  and  $f_1$  such that both  $H$  and  $\partial H$  are transverse to  $N$ . Then  $W = H^{-1}(N)$  is a one dimensional manifold with boundary, and thus has an even number of boundary points. But  $\partial W = \{0\} \times f_0^{-1}(N) \cup \{1\} \times f_1^{-1}(N)$ . Thus the difference in the number of points in  $f_0^{-1}(N)$  and  $f_1^{-1}(N)$  must be even, which shows  $I_2(f_0, N) = I_2(f_1, N)$ .  $\square$

*Remark.* If  $M$  and  $N$  are both compact, then for transverse smooth maps  $f : M \rightarrow E$  and  $g : N \rightarrow E$  we can define  $I_2(f, g)$  as the number of points in the compact, zero dimensional manifold  $M \times_E N$ . This is also a homotopy invariant of both  $f$  and  $g$ , and one can verify quite easily that  $I_2(f, g) = I_2(g, f)$ . This type of intersection can be made into a special case of the intersection theory we consider here, since then  $M \times_E N = (f \times g)^{-1}(\Delta)$ .

*Remark.* If  $N$  is *not* a closed submanifold of  $E$ , we might still be able to define  $I_2(f, N)$  if  $f^{-1}(N)$  consists of a finite number of points. But then  $I_2(f, N)$  is *not* necessarily a homotopy invariant of  $f$ ; an example is given by taking  $N = (0, 2)$ ,  $M = \{p\}$  a single point, and letting  $f_t(p) = t$ . Then  $f_0$  and  $f_1$  are both transverse to  $N$ , but  $I_2(f_0, N)$  is even, and  $I_2(f_1, N)$  is odd.

Similarly, if  $M$  is not *compact*, then  $I_2(f, N)$  might not be a homotopy invariant. If  $N \subset \mathbb{R}^2$  is the  $x$  axis,  $M = \mathbb{R}$ , and  $f_t(x) = (x \cos(t), 1 + \sin(t))$ , then  $f_0$  and  $f_{\pi/2}$  are transverse to  $N$ , but  $I_2(f_0, N)$  is even, and  $I_2(f_{\pi/2}, N)$  is odd.

Because homotopy is an equivalence relation, we can even define  $I_2(f, N)$  where  $f$  is *not* transverse to  $N$ , because we can always replace  $f$  with an equivalent, homotopic map which is transverse to  $N$ , and then count the number of points in the intersection modulo two. The last theorem shows this is independant of the homotopic map we select, and so the quantity  $I_2(f, N)$  is well defined.

**Example.** Consider two loops  $M$  and  $N$  on a two dimensional torus that intersect in a single position. This means precisely that  $I_2(M, N) = 1$ . Since the intersection number is invariant under homotopy, no matter how we deform  $M$  and  $N$ , they will still intersect one another, in particular, intersecting an odd number of times.

Since the intersection number is a homotopy invariant, if  $f : M \rightarrow E$  is homotopic to a constant map, then  $I_2(f, N)$  is even for any complementary submanifold  $N$  of  $E$ , except if  $M$  is zero dimensional, consists of an odd number of points, and  $N = E$ . If  $E$  is contractible, then  $I_2(f, N)$  is even for any map  $f : M \rightarrow N$ , except if  $M$  is zero dimensional.

**Theorem 16.11.** *No compact manifold is contractible (unless  $M$  is a point).*

*Proof.* Let  $M$  be a compact manifold. If  $p \in M$  is a point, then  $I(M, p) = 1$ . But this is impossible if  $M$  is contractible.  $\square$

It is curious to consider the case when  $M$  is compact, and  $f : M \rightarrow E$  is a smooth map with  $\dim(M) = \dim(E)/2$ . Then we can consider the quantity  $I_2(f, f)$ , or the *self intersection number* of  $f$  modulo two.

**Example.** *The loop travelling through the middle of the Möbius strip has intersection number one with itself. In particular, no matter how you deform this loop, it will still intersect the position it originally lied in; it is impossible to completely separate from its original position.*

Suppose  $M$  and  $N$  are closed, transverse submanifolds of  $E$  with complementary dimension, and  $M$  is the boundary of a compact manifold  $W$ . Then, intuitively, we should expect  $I_2(M, N) = 0$ , since any intersection between  $M$  and  $N$  either enters, or leaves,  $W$ , and each entrance must be matched with an exit. The next theorem shows our intuition is accurate.

**Theorem 16.12.** *Suppose  $M$  is a manifold which forms the boundary of a compact manifold  $W$ . Let  $N$  be a submanifold of  $E$  with dimension complementary to  $M$ , and suppose  $f : M \rightarrow E$  is a smooth map which extends to a smooth map  $\tilde{f} : W \rightarrow E$ . Then  $I_2(f, N) = 0$ .*

*Proof.* Without loss of generality, up to homotopy, we may assume that  $f$  and  $\tilde{f}$  are both transverse to  $g$ . Then  $\tilde{f}^{-1}(N)$  is a closed compact, one dimensional submanifold of  $W$  with boundary equal to  $f^{-1}(N)$ , which therefore consists of an even number of points.  $\square$

*Remark.* If  $W$  is non-compact, then this theorem is not necessarily true. For instance, if

$$W = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, z \geq 0\},$$

and  $M = \{(x, y, 0) \in \mathbb{R}^3 : x^2 + y^2 = 1\}$ , then a line  $N$  can pass through  $M$  without intersecting any other points on  $W$ .

Two compact submanifolds  $M$  and  $N$  of  $E$  with the same dimension are *cobordant* in  $E$  if there exists a compact manifold with boundary  $W \subset E \times [0, 1]$  such that  $\partial W = M \times \{0\} \cup N \times \{1\}$ . If  $M$  can be deformed into  $N$ , then  $M$  and  $N$  are cobordant, but this need not be the case - a pair of trousers shows that a single disk in  $\mathbb{R}^2$  is cobordant to two disks.

**Theorem 16.13.** *If  $M$  and  $N$  are cobordant in  $E$ , then for any compact manifold  $L$  in  $E$  with dimension complementary to  $M$  and  $N$ ,  $I_2(M, L) = I_2(N, L)$ .*

*Proof.* Let  $W \subset E \times [0, 1]$  be a compact manifold with  $\partial W = \{0\} \times M \cup \{1\} \times N$ . Then let  $f : W \rightarrow E$  be the map  $f(x, t) = x$ . Then the boundary theorem implies that  $I_2(\partial W, L) = 0$ .  $\square$

## 16.2 Oriented Intersection Theory

In the last section, we began with the observation that the boundary of any compact one dimensional manifold with boundary has an even number of points, and used this observation to build up a large number of topological invariants corresponding to counting parities of points associated with smooth maps. In this chapter, we do much the same, but using orientation rather than parities to count points, which gives more detailed topological information about the point.

Recall that any point  $p$ , viewed as a zero dimensional manifold, can either be given an orientation number equal to  $+1$ , or an orientation  $-1$ . If  $M$  is a compact, oriented one dimensional manifold with boundary, then each point  $p$  in  $\partial M$  has a natural orientation, equal to  $+1$  if outward pointing vectors at  $p$  have positive orientation, or  $-1$  if the outward pointing vectors are negatively oriented. The simple idea here is that, since any compact, oriented one manifold with boundary can be decomposed into a unit of lines and circles, the sum of the induced orientation numbers on the boundary points is always equal to zero.

To discuss an oriented intersection number, it clearly suffices to give a natural orientation to  $S = f^{-1}(N)$ , whenever  $f : M \rightarrow E$  is a smooth map transverse to some closed, submanifold  $N$  of  $E$ , where  $M, N$ , and  $E$  are all oriented manifolds, and only  $M$  has boundary. Given any point  $p \in M$  with  $f(p) = q$ , consider some complementary subspace  $V_p \subset T_p M$  to  $T_p S$ . Then  $f_*|_p$  gives an isomorphism between  $V_p$  and  $W_q = f_*(V_p)$ , which is a subspace of  $T_q E$  complementary to  $T_q N$ . We give  $W_q$  the orientation such that  $W_q \oplus T_q N$  is oriented the same way that  $T_q E$  is, give  $V_p$  the orientation such that  $f_*$  is orientation preserving between  $V_p$  and  $W_q$ , and then give  $T_p S$  the orientation such that  $V_p \oplus T_p S$  is orientated the same way as  $T_p M$ . Thus we orient  $T_p S$  such that

$$f_*(V_p) \oplus T_q N = T_q E \quad \text{and} \quad V_p \oplus T_p(M \cap N) = T_p M$$

**Lemma 16.14.** *The orientation given to  $T_p S$  is invariant of the choice of  $V_p$ .*

*Proof.* Suppose  $V_p$  and  $V'_p$  are both complementary to  $T_p S$ . Then we have natural projection maps  $\pi : T_p M \rightarrow V_p$  and  $\pi' : T_p M \rightarrow V'_p$ , and  $\pi|_{V'_p}$  and  $\pi'|_{V_p}$ . If we let  $W_p = f_*(V_p)$  and  $W'_p = f_*(V'_p)$ , then we have projection maps  $\eta : T_q E \rightarrow W_q$  and  $\eta' : T_q E \rightarrow W'_q$ , and a commutative diagram

$$\begin{array}{ccc}
 T_p S & \xrightarrow{\quad} & T_q N \\
 \downarrow & & \downarrow \\
 T_p M & \xrightarrow{\quad} & T_q E \\
 \swarrow \quad \searrow & & \swarrow \quad \searrow \\
 V_p & \xrightarrow{\cong} & V'_p \quad W_q \xrightarrow{\cong} W'_q \\
 \searrow \quad \swarrow & & \searrow \quad \swarrow \\
 & \xrightarrow{\cong} & 
 \end{array}$$

By construction, the curved arrows are orientation preserving isomorphisms. Now let  $K$  be the dimension of  $E$ , let  $k$  be the codimension of  $N$  in  $E$ . If  $(w_1, \dots, w_k)$  is an oriented basis for  $W_q$ , and  $(w_{k+1}, \dots, w_K)$  is an oriented basis for  $T_q N$ , then  $(w_1, \dots, w_K)$  is an oriented basis for  $T_q E$ . Now for each  $1 \leq i \leq k$ , find  $w'_i \in W'_q$ , as well as  $a_{ij}$  for  $k+1 \leq j \leq K$ , such that

$$w_i = w'_i + \sum_{j=k+1}^K a_{ij} w_j.$$

Then since

$$\det \begin{pmatrix} I_k & A \\ 0 & I_{n-k} \end{pmatrix} = 1,$$

we find that

$$[(w_1, \dots, w_K)] = [(w'_1, \dots, w'_k, w_{k+1}, \dots, w_K)].$$

Thus we find that the isomorphism between  $W_q$  and  $W'_q$ , and thus the isomorphism between  $V_p$  and  $V'_p$ , are both orientation preserving. But this implies that the orientation induced on  $T_p S$  is the same regardless of whether it is induced by  $V_p$  or by  $V'_p$ .  $\square$

To verify that this gives a consistent orientation to the manifold  $S$ , fix  $p \in S$  such that  $f(p) = q \in N$ . If  $\dim(E) = K$ ,  $\text{codim}_E(N) = k$ , and  $\dim(M) = n$ , then we can find coordinate systems  $(x, U)$  around  $p$ , and  $(y, V)$  around  $q$ , such that for each  $(t_1, \dots, t_N) \in x(U)$ ,

$$(y \circ f \circ x^{-1})(t_1, \dots, t_n) = (t_1, \dots, t_k, *, \dots, *),$$

where

$$N \cap V = \{q \in V : y_1(q), \dots, y_k(q) = 0\}.$$

For each  $p \in U$ , we can set

$$V_p = \text{span} \left( \left. \frac{\partial}{\partial x^1} \right|_p, \dots, \left. \frac{\partial}{\partial x^k} \right|_p \right)$$

and we then find

$$W_q = \text{span} \left( \left. \frac{\partial}{\partial y^1} \right|_q, \dots, \left. \frac{\partial}{\partial y^k} \right|_q \right).$$

Now for each  $q \in V$ , we have

$$T_q N = \text{span} \left( \left. \frac{\partial}{\partial y^{k+1}} \right|_q, \dots, \left. \frac{\partial}{\partial y^K} \right|_q \right)$$

and

$$T_p S = \text{span} \left( \left. \frac{\partial}{\partial x^{k+1}} \right|_p, \dots, \left. \frac{\partial}{\partial x^n} \right|_p \right).$$

One checks that in our construction,  $W_q$  and thus  $V_p$ , inherit the orientation given by the bases in the space above. But this means that an oriented basis for  $T_p S$  is given by the span above. Thus the trivialization is orientation preserving on  $U \cap f^{-1}(V)$ . But such sets are easily seen to cover  $S$ , so the orientation we have given on  $S$  is consistent.

**Lemma 16.15.** *Let  $N^m$  be a closed submanifold of  $E^d$ ,  $M^n$  a compact manifold with boundary, and  $f : M \rightarrow E$  a smooth map with  $f$  and  $\partial f$  transverse to  $N$ , where  $M$ ,  $N$ , and  $E$  are all oriented. Then the induced orientation on  $\partial[f^{-1}(N)]$  is (the same?) orientation as that induces on  $(\partial f)^{-1}(N)$ .*

*Proof.* Fix  $p \in \partial f^{-1}(N)$  and let  $f(p) = q$ . Consider the following vector spaces

- Let  $V_p = T_p(\partial f)^{-1}(N)$ .
- Let  $O_p$  be a complementary subspace to  $V$  in  $T_p(f^{-1}(N))$ , oriented outward.
- Let  $W_p$  be  $T_p(f^{-1}(N))$ .
- Let  $U_p$  be a complementary subspace to  $T_p(f^{-1}(N))$  in  $T_p(M)$ .

The orientation on  $V_p$  and  $U_p$  relative to the construction  $\partial[f^{-1}(N)]$  is given by the equations

$$O_p \oplus V_p = W_p,$$

$$U_p \oplus W_p = T_p(M),$$

and

$$f_*(U_p) \oplus T_p(N) = T_p(E).$$

In short,

$$U_p \oplus O_p \oplus V_p = T_p(M)$$

and

$$f_*(U_p) \oplus T_p(N) = T_p(E).$$

In the construction  $(\partial f)^{-1}(N)$ , the orientation on  $V_p$  and  $U_p$  is given by the equations

$$f_*(U_p) \oplus T_p(N) = T_p(E),$$

$$U_p \oplus V_p = T_p(\partial M)$$

and

$$O_p \oplus T_p(\partial M) = T_p(M).$$

In short,

$$O_p \oplus U_p \oplus V_p = T_p(M),$$

and

$$f_*(U_p) \oplus T_p(N) = T_p(E).$$

Comparing the two pairs of equations shows that these orientations differ by  $(-1)^{n+m-d}$ .  $\square$

*Remark.* In particular, if  $n + m = d$ , the two orientations are the same.

Now, given oriented manifolds  $M$ ,  $N$ , and  $E$ , where  $M$  is compact,  $N$  is a closed submanifold of  $N$ ,  $\dim(M) + \dim(N) = \dim(E)$ , and  $f : M \rightarrow E$  is a smooth map, then  $f^{-1}(N)$  is a finite, oriented set of points, and we define

$$I(f, N) = \sum \{ \text{sign}(p) : p \in f^{-1}(N) \}.$$

Here  $\text{sign}(p)$  equals  $+1$  if  $f_*(T_p M) \oplus T_p N$  has the same orientation as  $T_p E$ , and equals  $-1$  if  $f_*(T_p M) \oplus T_p N$  has the opposite orientation to  $T_p E$ . If  $f_0$  and  $f_1$  are homotopic, and both

transverse to  $N$ , we can find a homotopy  $H : [0, 1] \times M \rightarrow E$  with  $H$  and  $\partial H$  transverse to  $N$ . Since  $[0, 1] \times M$  has a natural orientation,  $H^{-1}(N)$  is an oriented, compact, one dimensional manifold with boundary. The boundary has an induced orientation, either as  $(\partial H^{-1})(N)$ , or as  $\partial[H^{-1}(N)]$ , with both orientations being the same. But viewing the orientation via the construction  $\partial[H^{-1}(N)]$ , we see

$$\sum_{p \in \partial H^{-1}(N)} \text{sgn}(p) = 0.$$

The viewpoint as  $(\partial H^{-1}(N))$ , combined with this equation, shows that

$$\sum_{p \in f_0^{-1}(N)} \text{sgn}(p) = \sum_{p \in f_1^{-1}(N)} \text{sgn}(p).$$

This gives the invariance under homotopy. Thus  $I(f, N)$  is defined for any smooth map  $f$  and closed, oriented submanifold, and is a homotopy invariant of  $f$ . Just as with the modulo two case, the quantity is trivial among maps along boundaries, for the same reason.

**Theorem 16.16.** *Let  $M$  be the boundary of a compact manifold with boundary  $W$ , and let  $f : M \rightarrow E$ . If  $f$  extends to a map  $F : W \rightarrow E$ , then  $I(f, N) = 0$  for any closed submanifold of  $W$  with dimensions complementary to  $M$ .*

Just as with the modulo two intersection numbers, if  $M$  and  $N$  are compact oriented manifolds with complementary dimension, and  $f : M \rightarrow E$  and  $g : N \rightarrow E$  are smooth maps, where  $E$  is an oriented manifold, then we can define  $I(f, g)$  as  $(-1)^{\dim E} I(f \times g, \Delta_E)$ . More precisely, if  $f$  and  $g$  are transverse, and  $(p, q) \in M \times N$  with  $f(p) = g(q) = r$ , then we assign  $(p, q)$  a value  $+1$  if  $f_*(T_p M) \oplus g_*(T_q N)$  has the same orientation as  $T_r E$ , and a value  $-1$  if it has the orientation number. Then we find

$$I(f, g) = \sum_{f(p)=g(q)} \text{sgn}((p, q)).$$

This is clearly a homotopy invariant of both  $f$  and  $g$ . But unlike in the case modulo two, it is not necessarily true that  $I(f, g) = I(g, f)$ . Indeed, for a given point  $(p, q) \in M \times N$  with  $f(p) = g(q)$ , we find

$$f_*(T_p M) \oplus g_*(T_q N) = (-1)^{(\dim M)(\dim N)} [g_*(T_q N) \oplus f_*(T_p M)].$$

Thus we conclude  $I(f, g) = (-1)^{(\dim M)(\dim N)} I(g, f)$ .

If  $M$  is an oriented submanifold of an oriented manifold  $N$ , with

$$\dim(N) = 2 \dim(M),$$

and  $\dim(M)$  is odd, then the last calculation gives that  $I(M, M) = -I(M, M)$ , so  $I(M, M) = 0$ . In particular, we conclude  $I_2(M, M)$  is even, so if  $M$  and  $N$  are manifolds such that  $I_2(M, M)$  is odd, then  $N$  cannot be orientable. In particular, we therefore showed in the last section that the Möbius strip is not orientable.

## 16.3 Degree and Winding Numbers

Intersection numbers can be used via certain geometric construction to give further invariants to smooth maps on manifolds. In particular, two powerful invariants are the *degree* of a map, and the *winding number* of a map around a point. The most powerful of these invariants occur on oriented manifolds, but the invariants can also be considered modulo two even on nonorientable manifolds.

Let us begin by considering a map  $f : M \rightarrow N$ , where  $M$  and  $N$  are compact oriented manifolds,  $N$  is connected, and  $\dim(M) = \dim(N)$ . Then the quantity  $I(f, q)$  is independent of the value  $q \in N$  chosen, and we define this quantity to be the *degree* of  $f$ . We know from this definition that if  $q \in N$  is a regular value of  $f$ , then  $f^{-1}(q)$  consists of finitely many points  $\{p_1, \dots, p_N\}$ . If we let  $\text{sgn}(p_i) = +1$  if  $f_*|_{p_i}$  is orientation preserving, and  $\text{sgn}(p_i) = -1$  if  $f_*|_{p_i}$  is orientation reversing, then

$$\deg(f) = \sum_{i=1}^N \text{sgn}(p_i).$$

If we remove the assumption of orientability, then  $\deg(f)$  is no longer definable, but we can define the quantity  $\deg_2(f)$ , which is equal to  $I_2(f, q)$ . If  $q$  is a regular value, then  $\deg_2(f)$  is equal to the parity of  $\#f^{-1}(q)$ . These quantities are homotopy invariants of the degrees of the maps involved. Moreover, in both the orientable and not oriented case, if  $M = \partial W$  for some oriented / non-oriented manifold with boundary  $W$ , then  $f$  extends to a smooth map on  $W$  if and only if  $\deg(f) = 0$  or  $\deg_2(f) = 0$ .

**Example.** *The only intersection number on the spheres  $S^k$  are the degree maps. If  $f : M \rightarrow S^k$  is a smooth map, where  $0 < \dim(M) < k$ , then for any closed submanifold  $N$  of  $S^k$  with dimension complementary to  $M$ , there must exist some point  $p \in S^k - f(M) - N$ . Since  $S^k - \{p\}$  is contractible,  $f$  is homotopic to the constant map, so  $I_2(f, N) = 0$ . A simple corollary of this is that  $S^2$  is not diffeomorphic to the torus  $\mathbb{T}^2$ , because there exists two loops on  $\mathbb{T}^2$  which intersect an odd number of times.*

In a very special case, we can consider a further invariant. If  $x \in \mathbb{R}^{n+1}$ ,  $M^n$  is an oriented, compact manifold, and  $f : M \rightarrow \mathbb{R}^{n+1} - \{x\}$ , then we define the *winding number* of  $f$  around  $x$ , denoted  $W(f, x)$ , as the degree of the map  $\tilde{f}_x : M \rightarrow S^n$  given by setting

$$\tilde{f}_x(p) = \frac{f(p) - x}{|f(p) - x|}.$$

Then  $W(f, x)$  is a homotopy invariant of maps from  $M$  to  $\mathbb{R}^{n+1} - \{x\}$ . In the non-orientable case, we can consider the winding number modulo two, defined via the degree modulo two. One of the main uses of the winding number is to obtain a generalization of the argument principle; if  $M = \partial W$  for some compact manifold with boundary  $W$ , and there exists a map  $F : W \rightarrow \mathbb{R}^{n+1} - \{x\}$  extending  $f$ , then  $W(f, x) = 0$  in the orientable case, and  $W_2(f, x) = 0$  in the nonorientable case.



Let us now use these invariants to prove some powerful theorems characterizing the behaviour of smooth maps. We will sometimes use the modulo two invariants, but often we will be forced to use the general invariants, thus requiring orientability of our maps. Thus we will be able to measure how much information the modulo two invariants throw away as compared to the general invariants. Let us begin with the fundamental theorem of algebra.

**Theorem 16.17.** *Every polynomial has a root over  $\mathbb{C}$ .*

*Proof.* Let  $f : \mathbb{C} \rightarrow \mathbb{C}$  be given by a map  $f(z) = z^n + g(z)$ , where  $\deg(g) < \deg(f)$ . If  $R$  is sufficiently large, depending on  $n$  and the coefficients of  $g$ , then  $f$  has no zeroes on the boundary of the ball  $B$  of radius  $R$  centered at the origin. If we let  $\tilde{f} : \partial B \rightarrow \mathbb{C} - \{0\}$  denote the restriction of  $f$  to  $\partial B$ , then it is also true that  $\tilde{f}$  is homotopic to the map  $g : \partial B \rightarrow \mathbb{C} - \{0\}$  given by  $g(z) = z^n$ . But then it is easy to see that  $W(\tilde{f}, 0) = W(g, 0) = n$ . If  $f$  has no zeroes on  $B$ , then we would conclude that  $W(\tilde{f}, 0) = 0$ , so we conclude that  $f$  must actually have a zero on the interior of  $B$ .  $\square$

*Remark.* Using winding numbers modulo two enables us to prove the fundamental theorem of algebra for odd degree polynomials, but not for even degree polynomials.

Now let us move onto an important theorem of algebraic topology; the Jordan-Brouwer separation theorem.

**Lemma 16.18.** *Suppose  $M$  is the boundary of a compact manifold  $W^n$ , and let  $\tilde{f} : W \rightarrow \mathbb{R}^n$  be a smooth map extending some smooth map  $f : M \rightarrow \mathbb{R}^n$ . Suppose  $x$  is a regular value of  $\tilde{f}$  that does not belong to the image of  $f$ . Then  $\tilde{f}^{-1}(x)$  is a finite set  $\{x_1, \dots, x_N\}$ , and  $W_2(f, x)$  has the same parity as  $N$ .*

*Proof.* If  $\tilde{f}^{-1}(x)$  is empty, the boundary theorem implies that  $W_2(f, x)$  is even, so the result is trivial in this case. If  $\tilde{f}^{-1}(x) = \{x_1, \dots, x_N\}$ , since  $\tilde{f}$  is a diffeomorphism in a neighbourhood of each  $x_i$ , there is  $\varepsilon > 0$  such that if  $B$  is the closed ball of radius  $\varepsilon$  around  $x$ , and  $B_i = \tilde{f}^{-1}(B)$ , then  $\tilde{f}$  is a diffeomorphism between  $B_i$  and  $B$  for each  $i$ . Let  $f_i$  be the restriction of  $\tilde{f}$  to  $\partial B_i$ . The boundary theorem implies that

$$W_2(f, x) = W_2(f_1, x) + \dots + W_2(f_N, x),$$

But in this case it is obvious that  $W_2(f_i, x)$  is odd for each  $i$ , which completes the proof.  $\square$

Now we can prove the separation theorem. We note that if  $M$  is a compact, connected hypersurface in  $\mathbb{R}^n$  of codimension one, and if  $M = \partial N$  for some  $N$ , then we would conclude from the theorem above that  $W_2(M, x)$  is odd for  $x \in N$ , and  $W_2(M, x)$  is even for  $x \notin N$ . To prove the Jordan separation theorem, it now suffices to reverse this construction.

**Lemma 16.19.** *Suppose  $M$  is a compact connected hypersurface in  $\mathbb{R}^n$ . If  $x \in M$ , and  $U$  is an open neighbourhood of  $x$ , then for any  $y \in \mathbb{R}^n - M$ , there exists a curve connecting  $y$  to a point in  $U$ , not passing through any point in  $M$ .*

*Proof.* Fix  $y \in \mathbb{R}^n - M$ , and let  $U$  be the connected component of  $\mathbb{R}^n - M$  containing  $y$ . Then every point in  $U$  is connected to  $y$  via a path not containing any points in  $M$ , and it suffices to show that  $\overline{U}$  contains  $M$ . Since  $U$  is closed in  $\mathbb{R}^n - M$ ,  $\overline{U} - U \subset M$ . Since  $U$  cannot be closed in  $\mathbb{R}^n - M$ , since  $\mathbb{R}^n$  is connected,  $\overline{U} - U$  is non-empty. It is certainly closed in  $M$ , so it suffices to show that  $\overline{U} - U$  is open. But this is easy to see working locally in coordinates.  $\square$

A simple corollary of this lemma is that  $\mathbb{R}^n - M$  has at most two components, since locally  $M$  looks like a hyperplane, which cuts the plane into two parts.

**Theorem 16.20.** *Let  $M$  be a compact, connected hypersurface in  $\mathbb{R}^n$ . Then there exists a connected, compact manifold with boundary  $N$  with  $\partial N = M$ .*

*Proof.* We note that  $W_2(M, x_1) = W_2(M, x_2)$  for any points  $x_1, x_2$  in the same connected components of  $\mathbb{R}^n - M$ . Thus the set

$$N^\circ = \{x : W_2(M, x) = 1\}$$

is a union of connected components of  $\mathbb{R}^n - M$ . Since  $W_2(M, x) = 0$  for suitably large  $x$ , the only possible case is that  $N^\circ = \emptyset$ , or  $N^\circ$  is a connected subset of  $\mathbb{R}^n - M$ . If  $l$  is a closed line segment between two points  $x_1$  and  $x_2$ , with  $l$  transversal to  $M$ , then it is easy to see that  $W_2(M, x_1) = W_2(M, x_2) + \#(l \cap M)$ . In particular, this enables us to show  $N^\circ$  is nonempty. Now since  $N^\circ$  is bounded, if we set  $N = N^\circ \cup M$ , then  $N$  is compact. Working locally shows that  $N$  is a compact manifold with boundary, which completes the proof.  $\square$

An interesting corollary of the Jordan Brouwer separation theorem is that every hypersurface of  $\mathbb{R}^n$  of codimension one is orientable, since every submanifold of dimension  $n$  is orientable. Next, we move onto the Borsuk-Ulam theorem, which we state in the following form. We begin with a special case of the Hopf degree theorem, which we prove later.

**Lemma 16.21.** *The homotopy classes of maps between  $S^1$  and itself are determined by the degree maps.*

*Proof.* Let  $f : S^1 \rightarrow S^1$  be a smooth map. If we let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be the covering space map of  $f$ , i.e. the map which for each  $t \in \mathbb{R}$ ,

$$f(\cos t, \sin t) = (\cos(g(t)), \sin(g(t))).$$

Then there is  $n \in \mathbb{Z}$  such that for each  $t \in \mathbb{R}$  such that  $g(t + 2\pi) = g(t) + 2\pi n$ . There is a homotopy  $\tilde{H} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$  obtained by setting

$$\tilde{H}(t_1, t_2) = t_1 g(t_2) + (1 - t_1) n t_2$$

Clearly  $\tilde{H}$  descends to a homotopy  $H : [0, 1] \times S^1 \rightarrow S^1$  between  $f$  and the map  $f_n : S^1 \rightarrow S^1$  given by setting  $f_n(z) = nz$ . From this statement, it is obvious that  $n = \deg(f)$ , and thus this quantity uniquely characterizes the homotopy class of  $f_n$ .  $\square$

**Theorem 16.22.** Let  $f : S^n \rightarrow \mathbb{R}^{n+1} - \{0\}$  be a smooth map satisfying the symmetry condition  $f(-x) = -f(x)$ . Then  $W_2(f, 0) = 1$ .

*Proof.* By homotopy invariance, we may assume  $f : S^n \rightarrow S^n$ , and we thus need to show  $\deg_2(f)$  is odd. We prove this result by induction on  $n$ . For the base case  $n = 1$ , we work in coordinates. The symmetry condition guarantees the existence of a smooth map  $g : \mathbb{R} \rightarrow \mathbb{R}$  and an odd integer  $m$  such that

$$f(\cos t, \sin t) = (\cos(g(t)), \sin(g(t))),$$

and  $g(t + \pi) = g(t) + \pi m$  for each  $t \in \mathbb{R}$ . But this means that  $g(t + 2\pi) = g(t) + 2\pi m$ , and thus  $\deg_2(f)$  is congruent to  $m$ , which is odd, completing the proof. Now we consider the case  $n > 1$ . Consider  $S^{n-1}$  as the equator of  $S^n$ , and let  $f_0 : S^{n-1} \rightarrow \mathbb{R}^{n+1} - \{0\}$  denote the restriction of  $f$  to the equator. Choose a unit vector  $a \in S^n$  such that  $a$  and  $-a$  are both regular values for

$$\frac{f}{|f|} : S^n \rightarrow S^n \quad \text{and} \quad \frac{f_0}{|f_0|} : S^{n-1} \rightarrow S^n.$$

The second condition implies that  $f_0(S^{n-1})$  is disjoint from the line  $l$  generated by  $a$ , and the first condition implies that  $f$  intersects  $l$  transversally. By definition,  $W_2(f, 0)$  has the same parity as the number of intersection points of  $f/|f|$  with  $a$ . By symmetry,  $f/|f|$  has the same number of intersection points with  $a$  as with  $-a$ . Thus  $W_2(f, 0)$  has the same parity as half the number of intersection points of  $f$  with the line  $l$ . If  $U_+$  denotes the upper hemisphere of  $S^n$ , and  $f_+ : U_+ \rightarrow \mathbb{R}^{n+1} - \{0\}$  denotes the restriction of  $f$  to  $U_+$ , then symmetry again implies that  $W_2(f, 0)$  has the same parity as the number of intersection points of  $f_+$  with  $l$ . We would like to use the boundary theorem and the inductive hypothesis, but the dimensions don't quite add up. But if we let  $V$  be the orthogonal complement of  $l$ , and let  $\pi : \mathbb{R}^{n+1} - \{0\} \rightarrow \mathbb{R}^n$  denote the orthogonal projection map. By the induction hypothesis, we can consider  $\pi \circ f_0 : S^{n-1} \rightarrow \mathbb{R}^n - \{0\}$ , and so  $W_2(\pi \circ f_0, 0)$  is odd. But we know that 0 is a regular value for  $\pi \circ f_+$ , so  $W_2(\pi \circ f_0, 0)$  has the same parity as  $(\pi \circ f_+)^{-1}(0) = f_+^{-1}(l)$ , which completes the proof.  $\square$

*Remark.* We remark that if  $f : S^n \rightarrow \mathbb{R}^{n+1} - \{0\}$  satisfies  $f(-x) = -f(x)$ , then  $f$  intersects every line through the origin. For if  $l$  does not intersect  $f$ , then any  $a \in S^n \cap l$  is a regular value of  $f/|f|$ , so  $W_2(f, 0)$  has the same parity as the number of point on  $f^{-1}(l)$ , which is odd, giving a contradiction.

Here are other interesting results which follow from the Borsuk-Ulam theorem.

**Theorem 16.23.** If  $f_1, \dots, f_n : S^n \rightarrow \mathbb{R}$  are smooth functions satisfying  $f_i(-x) = -f_i(x)$ , then there must be some  $x \in S^n$  such that  $f_i(x) = 0$  for all  $i$ .

*Proof.* For otherwise we have a map  $f = (f_1, \dots, f_n, 0) : S^n \rightarrow \mathbb{R}^{n+1} - \{0\}$ . Then the Borsuk-Ulam theorem implies that  $W(f, 0) = 0$ , but  $f$  does not intersect the  $x_{n+1}$  axis.  $\square$

*Remark.* More generally, if  $g_1, \dots, g_n : S^n \rightarrow \mathbb{R}$  are smooth functions, then there exists  $x \in S^n$  such that  $g_i(-x) = g_i(x)$  for all  $i \in \{1, \dots, n\}$ .

Thus we know that there exists a position on the surface of the earth for which the temperature and the altitude are exactly the same on opposite sides.

## 16.4 Euler Characteristic

## 16.5 Lefschetz Fixed Point Theory

Let  $f : M \rightarrow M$  be a smooth map on a compact oriented manifold. We want to understand the fixed points of  $f$  topologically. We define

$$\Gamma(f) = \{(x, f(x)) \in M \times M : x \in M\}.$$

The *Lefschetz number*, denoted  $L(f)$ , is the intersection number  $I(\Delta_M, \Gamma(f))$ . If  $M$  is non-orientable, we can consider the Lefschetz number modulo two, denoted  $L_2(f)$ , but for simplicity we deal with oriented Lefschetz number.

**Lemma 16.24.** *If  $L(f)$  is nonzero, then  $f$  has a fixed point.*

*Proof.* If  $f$  has no fixed points, then  $\Gamma(f)$  is transverse to  $\Delta_M$ , so  $L(f) = 0$ , since  $\Gamma(f)$  is disjoint to  $\Delta_M$ .  $\square$

The most powerful results are obtained for *Lefschetz maps*  $f : M \rightarrow M$ , such that  $\Gamma(f)$  is transverse to  $\Delta_M$ . Then  $L(f)$  gives precisely the oriented number of fixed points for  $f$ . A simple condition can be used to check whether a map  $\Gamma(f)$  is transverse to  $\Delta_M$  at a point  $(p, p)$ ; it suffices to check that  $f_*|_p - I_p$  is invertible at each fixed point  $p$  of  $f$ , or equivalently, 1 is not an eigenvalue of  $f_*|_p$ . If, for a function  $f : M \rightarrow M$ , we say a fixed point  $p \in M$  is a *Lefschetz fixed point* if  $f_*|_p$  does not have 1 as an eigenvalue, then a map is Lefschetz precisely when all its fixed points are Lefschetz. Of course, Lefschetz maps are generic.

**Lemma 16.25.** *Every map  $f : M \rightarrow M$  is homotopic to a Lefschetz map.*

*Proof.* We consider a ball  $B$  in  $\mathbb{R}^K$ , and a map  $F : B \times M \rightarrow M$  such that  $F(0, p) = f(p)$  for all  $p \in M$ , and for each  $p \in M$ , the map  $x \mapsto F(x, p)$  is a submersion. Thus the map  $G : B \times M \rightarrow M \times M$  defined by setting  $G(x, p) = (p, F(x, p))$  is also a submersion. By the transversality theorem, for almost all  $x \in B$ , the map  $g_x(p) = (p, F(x, p))$  is transverse to  $\Delta_M$ . But then  $f_x(p) = F(x, p)$  is a Lefschetz map.  $\square$

Now consider a Lefschetz map  $f : M^n \rightarrow M^n$ , and a fixed point  $p \in M$ . To determine the orientation number assigned to  $(p, p) \in \Gamma(f) \cap \Delta_M$ , we refer to the definition. Let  $q = (p, p)$ . If  $T_p(M)$  has an oriented basis  $\{e_1, \dots, e_n\}$ , then  $T_q(\Gamma(f))$  has an oriented basis  $\{e_1 \oplus f_*(e_1), \dots, e_n \oplus f_*(e_n)\}$ , and  $\Delta_M$  has an oriented basis  $\{e_1 \oplus e_1, \dots, e_n \oplus e_n\}$ . This means that the orientation of  $\Delta_M \oplus T_q(\Gamma(f))$  is given by the basis

$$\{e_1 \oplus e_1, \dots, e_n \oplus e_n, e_1 \oplus f_*(e_1), \dots, e_n \oplus f_*(e_n)\}.$$

But subtracting vectors in a basis does not change orientation, so this basis has the same orientation as the basis

$$\{e_1 \oplus e_1, \dots, e_n \oplus e_n, 0 \oplus (f_*(e_1) - e_1), \dots, 0 \oplus (f_*(e_n) - e_n)\}.$$

But since  $f$  is Lefschetz,  $f_*(e_1) - e_1, \dots, f_*(e_n) - e_n$ , we may subtract a linear combination of the latter vectors to the former vectors, so this basis has the same orientation as the basis

$$\{e_1 \oplus 0, \dots, e_n \oplus 0, 0 \oplus (f_*(e_1) - e_1), \dots, 0 \oplus (f_*(e_n) - e_n)\}.$$

Clearly, the orientation of this basis relative to the standard basis

$$\{e_1 \oplus 0, \dots, e_n \oplus 0, 0 \oplus e_1, \dots, 0 \oplus e_n\}$$

on  $T_q(M \times M)$  is given by the sign of the determinant  $\det(f_*|_p - I_p)$ . Thus we conclude that the sign of the determinant gives the orientation number of the point  $(p, p)$ . Thus, for a Lefschetz fixed point  $p$  of a map  $f : M \rightarrow M$ , we let  $L_p(f)$  denote the determinant of  $\det(f_*|_p - I_p)$

**Example.** Let us consider the two dimensional case as an example. Let the function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a function with a Lefschetz fixed point at the origin. Then  $f(x) = Ax + \varepsilon(x)$ , where  $\varepsilon(x) = f(x) - Ax$ , and  $\varepsilon(x)/|x| \rightarrow 0$  as  $x \rightarrow 0$ . Then, under a suitably choice of coordinates around the origin, we may assume

$$A = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}.$$

Then  $L_0(f) = \text{sgn}[(\alpha_1 - 1)(\alpha_2 - 1)]$ . If  $\alpha_1, \alpha_2 > 1$ , then  $L_0(f) = +1$ , and  $f$  is an ‘expanding map’ about the origin. If  $\alpha_1, \alpha_2 < 1$ , then  $L_0(f) = +1$  again, and  $f$  is a ‘shrinking map’ contracting around the origin. If  $\alpha_1 < 1$  and  $\alpha_2 > 1$ , then  $L_0(f) = -1$ , and  $f$  is a saddle point about the origin, attracting some points and repelling others.

Since the Lefschetz number are defined in terms of intersection numbers, it is easy to see that the Lefschetz number is a homotopy invariant of the map  $f$ . If  $i : M \rightarrow M$  is the identity map, then

$$L(i) = I(\Delta, \Delta) = \chi(M).$$

Thus if  $f$  is homotopic to the identity, then  $L(f) = \chi(M)$ . Recall that if  $X$  is a smooth vector field on the manifold  $M$ , then it induces a flow map  $\phi_t : M \rightarrow M$  for each  $t \in \mathbb{R}$ . As  $t$  varies, we get a homotopy, so we conclude that the Lefschetz number of  $\phi_t$  is equal to the Euler characteristic of  $M$  for each  $t \in \mathbb{R}$ .

**Example.** Consider the sphere  $S^2$  in  $\mathbb{R}^3$ . If  $X$  is the vector field on  $\mathbb{R}^3$  with  $X = (0, 0, -1)$ , then we can obtain a vector field  $Y$  on  $S^2$  by letting  $Y_p$  be the orthogonal projection of  $X_p$  onto  $T_p(S^2)$ . For small  $t$ , it is easy to see the induced flow  $\phi_t : S^2 \rightarrow S^2$  has two fixed points, at the top and bottom of the sphere. Around the top point,  $\phi_t$  acts like an expanding map, and around the bottom point,  $\phi_t$  acts like a shrinking map. Thus we conclude that  $L(\phi_t) = 2$ , and thus, the  $\chi(S^2) = 2$ . Similar techniques constructing flows along a compact surface  $S$  of genus  $g$ , which has one expanding point, one shrinking point, and  $2g$  saddle points, show that  $\chi(S) = 2 - 2g$ .

Even if being Lefschetz is generic, we want some techniques to geometrically understand the Lefschetz number of non Lefschetz maps. For simplicity, we consider maps  $f : M \rightarrow M$  with only finitely many fixed points.

**Theorem 16.26.** *Let  $f : M \rightarrow M$  be a map, and let  $U$  be an open set of  $M$  containing only a single fixed point  $p$  of  $f$ . Then there exists a map  $g : M \rightarrow M$  homotopic to  $f$ , such that all fixed points for  $g$  in  $U$  are Lefschetz, and  $g(p) = f(p)$  for  $p \notin U$ .*

*Proof.* Assume without loss of generality that  $M = \mathbb{R}^d$  and that  $f$  has only a single fixed point globally, from which the general case reduces by working locally. Let  $x_0 \in M$  be the fixed point of  $f$ . Let  $\rho : M \rightarrow [0, 1]$  be a smooth function such that  $\rho(x) = 1$  in a neighbourhood  $V$  of  $x_0$ , and  $\rho$  vanishes outside a compact set  $K \subset U$ . For  $a \in \mathbb{R}^d$  and  $x \in M$ , define

$$f_a(x) = f(x) + \rho(x)a.$$

Since  $f$  has no fixed points in the compact set  $K - V$ , there exists  $c > 0$  such that  $|f(x) - x| \geq c$  for all  $x \in K - V$ . In particular, if  $|a| < c/2$ , then for  $x \in K - V$ ,  $|f_a(x) - x| \geq c/2$ . Thus  $f_a$  has no fixed points outside of  $V$ . Using Sard's theorem, we may find  $|a| < c/2$  such that  $a$  is a regular value for the map  $x \mapsto f(x) - x$ . This means precisely that for each fixed point  $p$  of the map  $f_a$ , the linear map  $f_*|_p - I_p$  is invertible. But since every fixed point  $p$  of  $f_a$  occurs in  $V$ ,  $(f_a)_*|_p = f_*|_p$ . Thus  $f_a$  is a Lefschetz map satisfying the requirements of the Lemma.  $\square$

Thus if we perturb an isolated fixed point of a map  $f : M \rightarrow M$  locally, it breaks up into finitely many fixed points. The perturbation might introduce these fixed points 'out of nowhere', but not in a way which changes  $Lf$ . Thus the Lefschetz fixed points created must have equal numbers of positive and negative orientation numbers.

Fortunately, there is a way of calculating this 'local' orientation number without having to perturb the fixed point, as long as we work in coordinates, using a very similar method to the winding number method we considering in the last chapter. Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a map with a single fixed point at the origin. Let  $U$  be an open set around the origin with smooth boundary  $N = \partial U$ . Let  $g : N \rightarrow S^{n-1}$  be defined by setting

$$g(x) = \frac{f(x) - x}{|f(x) - x|}$$

We claim the degree of  $g$  gives the Lefschetz number of  $f$  at the origin, provided that  $f$  is Lefschetz.

**Theorem 16.27.** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a Lefschetz map, and 0 is a fixed point, then for the map  $g : N \rightarrow S^{n-1}$  given by*

$$g(x) = \frac{f(x) - x}{|f(x) - x|},$$

*we have  $\deg(g) = L_0(f)$ .*

*Proof.* If  $B$  is a small ball around the origin in  $U$ , and we let  $g' : \partial B \rightarrow S^{n-1}$  denote the map

$$g'(x) = \frac{f(x) - x}{|f(x) - x|},$$

then clearly  $\deg(g) = \deg(g')$ , because both maps extend to a map on  $\overline{U} - B$ . Thus we may assume  $N$  is a ball around the origin with arbitrarily small radius. Since  $f(0) = 0$ , we can write  $f(x) = Ax + \varepsilon(x)$ , where  $\varepsilon(x)/|x| \rightarrow 0$  as  $x \rightarrow 0$ , and  $\det(A - I) \neq 0$ . Set  $f_i(x) = Ax + t\varepsilon(x)$ . Then for  $x \in \partial B$ , we have

$$|f_i(x) - x| = |(A - I)x + t\varepsilon(x)| \geq \left( \|(A - I)^{-1}\| - \frac{t\varepsilon(x)}{|x|} \right) \cdot |x|.$$

If the ball  $B$  has small enough radius, then we conclude that  $|f_i(x) - x| > 0$  for all  $x \in \partial B$ , so that  $g$  is homotopic to the map

$$\tilde{g}(x) = \frac{(A - I)x}{|(A - I)x|}.$$

It thus suffices to prove that  $\deg(\tilde{g}) = +1$  if  $\det(A - I) > 0$ , and  $\deg(\tilde{g}) = -1$  if  $\det(A - I) < 0$ . But this follows because  $SL(n)$  is path connected, so that if  $\det(A - I) > 0$ , then  $\tilde{g}$  is homotopic to the map

$$x \mapsto \frac{x}{|x|},$$

which is clearly oriented. On the other hand, if  $\det(A - I) < 0$ , then  $\tilde{g}$  is homotopic to

$$(x_1, \dots, x_n) \mapsto \frac{(-x_1, \dots, x_n)}{|x|}$$

which is orientation reversing. □

Note that for any map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  with finitely many fixed points  $\{p_1, \dots, p_m\}$ , we can define the quantities  $L_{p_i}(f)$  for each  $i$ . Fix  $\varepsilon > 0$  such that the balls  $B_\varepsilon(p_1), \dots, B_\varepsilon(p_m)$  are disjoint. We can thus define a new function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , homotopic to  $f$ , such that  $g$  is Lefschetz, and agrees with  $f$  outside of  $B_{\varepsilon/2}(p_1) \cup \dots \cup B_{\varepsilon/2}(p_m)$ . In particular, the degree of the maps  $\tilde{f}_i : \partial B_i \rightarrow S^{n-1}$  given by setting

$$\tilde{f}_i(x) = \frac{f(x) - x}{|f(x) - x|}.$$

is the same as the degree of the maps  $\tilde{g}_i : \partial B_i \rightarrow S^{n-1}$

$$\tilde{g}_i(x) = \frac{g(x) - x}{|g(x) - x|},$$

since both maps are equal. But if, for each  $i \in \{1, \dots, m\}$ , we consider the fixed points  $\{q_{i1}, \dots, q_{ik_i}\}$  for  $g$  in  $B_i$ , and considering small balls  $q_{ij} \in B_{ij} \subset B_i$ , we can define  $\tilde{g}_{ij} : \partial B_{ij} \rightarrow S^{n-1}$

$$\tilde{g}_{ij}(x) = \frac{g(x) - x}{|g(x) - x|}.$$

The boundary theorem implies that

$$\deg(g_i) = \sum \deg(g_{ij}) = \sum L_{q_{ij}}(g) = L(g)$$

then considering small balls an application of the boundary theorem and the previous lemma shows that

$$L(f) = L(g) = \sum_{i,j} \deg(g_{ij}) = \sum_i \deg(g_i) = \sum_i \deg(f_i) = \sum_i L_{p_i}(f).$$

Thus we have found a way to measure the Lefschetz number of a map without having to perturb the map into a Lefschetz map.



# **Part IV**

## **Differential Geometry**

Our goal is now to apply the foundational tools we have developed to analyze Riemannian manifolds in order to study the classical methods of differential geometry, most importantly, the properties of *curvature*, as initially developed alongside the calculus by Newton, Leibnitz, and Huygens, and then later revolutioned by Gauss in the late 19th century.

# Chapter 17

## Curves

### 17.1 Planar Curves

We begin our study by understanding the *curvature* of planar curves. By a planar curve, we mean an immersed, connected submanifold of  $\mathbb{R}^2$ , and we will assume this curve is  $C^2$ . We begin with the local theory, so we assume the existence of a local parametrization  $\mathbf{c} : [a, b] \rightarrow \mathbb{R}^2$ , such that  $\mathbf{c}'(t) \neq 0$  for all  $t \in [a, b]$ . In particular, this means that the arclength function

$$s(t) = \int_a^t |\mathbf{c}'(u)| \, du$$

is  $C^2$  diffeomorphism from  $[a, b]$  onto  $[0, L]$ , where  $L = \int_a^b |\mathbf{c}'(u)| \, du$  is the length of the portion of the curve parametrized by  $\mathbf{c}$ . We now find that

$$\frac{d\mathbf{c}}{ds} = \frac{d\mathbf{c}}{dt} \bigg/ \frac{ds}{dt} = \frac{\mathbf{c}'(t)}{|\mathbf{c}'(t)|}.$$

We write  $\mathbb{T} : [a, b] \rightarrow \mathbb{R}^2$  to be the function

$$\mathbb{T}(t) = \frac{\mathbf{c}'(t)}{|\mathbf{c}'(t)|}.$$

Thus, when  $\mathbf{c}$  is an arclength parameterization,  $\mathbf{c}'(s) = \mathbb{T}(s)$ . We also define  $\mathbf{n} : [a, b] \rightarrow \mathbb{R}^2$  to be the unit vector orthogonal to  $\mathbb{T}$ , such that  $[\mathbb{T}, \mathbf{n}]$  is an oriented basis of  $\mathbb{R}^2$  (i.e.  $\mathbf{n}$  is ‘anticlockwise’ to  $\mathbb{T}$ ).

The intuitive concept of curvature is vague. Nonetheless, we expect a straight line to have no curvature, and a circle with smaller radius should be curving more than a circle with lower radius. The definition of curvature we give here will follow this intuition, assigning a circle of radius  $R$  a constant curvature of  $1/R$  at each point, and assigning the line a curvature of zero at each point. To extend this concept to arbitrary curves, we try and apply the same intuitions that brought us the tangent line, i.e. approximating a curve locally by a circle, the *osculating circle* of the curve at a point. If the osculating circle exists at a point  $\mathbf{c}(t_0)$ , and

has radius  $R$ , then we define the curvature at  $\mathbf{c}(t_0)$  to be  $1/R$ . To find this circle, we recall that in the plane, any three non-colinear points lie on a unique circle. Thus to find the osculating circle at a point  $\mathbf{c}(t_0)$ , we might take  $t_1, t_2, t_3$  close to  $t_0$ , hope that  $\mathbf{c}(t_1), \mathbf{c}(t_2)$ , and  $\mathbf{c}(t_3)$  are not colinear, consider the circle  $C(t_1, t_2, t_3)$  defined by these three points, and then define the osculating circle  $C(t_0)$  to be the ‘limit’ of the circles  $C(t_1, t_2, t_3)$  and  $t_1, t_2, t_3 \rightarrow t_0$ . This limit will exist provided that we assume  $\mathbf{c}''(t_0) \neq 0$ .

**Theorem 17.1.** *Suppose  $\mathbf{c} : [a, b] \rightarrow \mathbb{R}^2$  is an arclength parameterization of a curve. If  $\mathbf{c}''(s_0) \neq 0$ , then for  $s_1, s_2, s_3$  close to  $s_0$ ,  $\mathbf{c}(s_1), \mathbf{c}(s_2)$ , and  $\mathbf{c}(s_3)$  are not colinear, and the osculating circle*

$$C(s_0) = \lim_{s_1, s_2, s_3 \rightarrow s_0} C(s_1, s_2, s_3)$$

*is well defined, and it is necessary for  $\mathbf{c}''(t_0)$  to be nonzero in order for this circle to exist. The radius of this circle is  $1/|\mathbf{c}''(s_0)|$ .*

*Proof.* Suppose  $\mathbf{c}(s_1), \mathbf{c}(s_2)$ , and  $\mathbf{c}(s_3)$  are colinear for  $s_1 < s_2 < s_3$ . Then the mean-value theorem implies that there exists  $\xi_1 \in (s_1, s_2)$  and  $\xi_2 \in (s_2, s_3)$  such that  $\mathbf{c}'(\xi_1) = \mathbf{c}'(\xi_2)$ . If this is true for a family of tuples  $(s_1, s_2, s_3)$  that contain  $(s_0, s_0, s_0)$  in their closure, then the resulting  $(\xi_1, \xi_2)$  also converge to  $(s_0, s_0)$ , and we see (again Cauchy Mean-Value Theorem) that  $\mathbf{c}''(s_0) = 0$ , which gives a contradiction.

Thus under the assumption that  $\mathbf{c}''(s_0) \neq 0$ , the circle  $C(s_1, s_2, s_3)$  is well defined for  $s_1 < s_2 < s_3$  sufficiently close to  $s_0$ . Let us assume these circles converge to some  $C(s_0)$ , and determine what this circle must be. Since the circle clearly must contain the point  $\mathbf{c}(s_0)$ , it suffices to determine the centre  $x_0$  of  $C(s_0)$  to uniquely determine the circle. To find  $x_0$ , let  $x = x(s_1, s_2, s_3)$  denote the center of the circle  $C(s_1, s_2, s_3)$ , and let  $r = r(s_1, s_2, s_3)$  denote its radius. Then the function

$$f(t) = |\mathbf{c}(t) - x|^2 - r^2$$

vanishes at the three points  $s_1, s_2$ , and  $s_3$ . It follows that  $f'(t) = 0$  for a point in  $(s_1, s_2)$  and a point in  $(s_2, s_3)$ , and thus that  $f''(t) = 0$  for a point in  $(s_1, s_3)$ . Taking  $s_1, s_2, s_3 \rightarrow t$ , and assuming  $x \rightarrow x_0$  and  $r \rightarrow r_0$ , we find that the first and second derivatives of the function

$$f_0(t) = |\mathbf{c}(t) - x_0|^2 - r_0^2$$

vanish at  $s_0$ . But this means that

$$\mathbf{c}'(s_0) \cdot (\mathbf{c}(s_0) - x_0) = 0$$

and

$$\mathbf{c}''(s_0) \cdot (\mathbf{c}(s_0) - x_0) + |\mathbf{c}'(s_0)|^2 = 0.$$

The existence of these conditions means that  $\mathbf{c}''(s_0)$  cannot possibly be a scalar multiple of  $\mathbf{c}'(s_0)$ , so we see this condition is necessary for an osculating circle to exist.

Without loss of generality, we may assume that  $\mathbf{c}$  is an arclength parameterization. Then  $\mathbf{c}''(s_0)$  is orthogonal to  $\mathbf{c}'(s_0) = \mathbb{T}(s_0)$ , and so the first condition implies we must have  $\mathbf{c}(s_0) - x_0 = (1/\kappa) \cdot \mathbf{n}(s_0)$  for some constant  $\kappa$ . The second equation now reads

$$(1/\kappa) (\mathbf{c}''(s_0) \cdot \mathbf{n}(s_0)) + 1 = 0$$

Now  $\mathbf{c}''(s_0)$  is orthogonal to  $\mathbf{c}'(s_0)$ , and thus is aligned with  $\mathbf{n}(s_0)$ , so  $\mathbf{c}''(s_0) \cdot \mathbf{n}(s_0) = \pm |\mathbf{c}''(s_0)|$ , and so  $\kappa = \mp |\mathbf{c}''(s_0)|$ . Thus the radius of the osculating circle is  $1/|\mathbf{c}''(s_0)|$ .  $\square$

This result motivates us to define the (*unsigned*) *curvature* of a curve at the point  $\mathbf{c}(s_0)$  to be  $|\mathbf{c}''(s_0)|$ . The *signed curvature* of the curve is the quantity  $\kappa$  such that  $\mathbf{c}''(s_0) = \kappa \mathbf{n}(s_0)$ . Note that the signed curvature is not an invariance of the curve by itself, but depends on the curves orientation. If one reverses the orientation of a curve, the curvature flips sign.

It is often useful to have a formula for  $\kappa$  that does not require we work in an arclength parameterization. We note first that, in an arclength parameterization,

$$\kappa = \det(\mathbf{c}'(s_0), \mathbf{c}''(s_0)).$$

In a general parameterization, we have

$$\frac{d\mathbf{c}}{ds} = \frac{d\mathbf{c}}{dt} \frac{dt}{ds} = \left( \frac{d\mathbf{c}}{dt} \right) \left( \frac{ds}{dt} \right)^{-1}$$

and

$$\frac{d^2\mathbf{c}}{ds^2} = \frac{d^2\mathbf{c}}{dt^2} \left( \frac{dt}{ds} \right)^2 + \frac{d\mathbf{c}}{dt} \frac{d^2t}{ds^2} = \left( \frac{ds}{dt} \right)^{-2} \left( \frac{d^2\mathbf{c}}{dt^2} - \frac{d^2s}{dt^2} \left( \frac{ds}{dt} \right)^{-1} \frac{d\mathbf{c}}{dt} \right).$$

Thus we get that

$$\begin{aligned} \kappa &= \left( \frac{ds}{dt} \right)^{-3} \det \left( \frac{d\mathbf{c}}{dt}, \frac{d^2\mathbf{c}}{dt^2} - \frac{d^2s}{dt^2} \left( \frac{ds}{dt} \right)^{-1} \frac{d\mathbf{c}}{dt} \right) \\ &= \left( \frac{ds}{dt} \right)^{-3} \det \left( \frac{d\mathbf{c}}{dt}, \frac{d^2\mathbf{c}}{dt^2} \right). \end{aligned}$$

In coordinates,

$$\kappa = \frac{\mathbf{c}'_1 \mathbf{c}''_2 - \mathbf{c}'_2 \mathbf{c}''_1}{((\mathbf{c}'_1)^2 + (\mathbf{c}'_2)^2)^{3/2}}.$$

Thus we have a way of calculating curvature without having to rely on an arclength parameterization every time.

**Theorem 17.2.** *The only planar curves with constant curvature are lines and circles.*

*Proof.* Suppose that  $\mathbf{c}$  is an arclength parameterization of a curve with constant curvature  $\kappa$ . Then

$$(\mathbb{T}'_1)^2 + (\mathbb{T}'_2)^2 = \kappa^2.$$

Differentiating, we find that

$$\mathbb{T}' \cdot \mathbb{T}'' = 0.$$

But this means that  $\mathbb{T}''$  is a scalar multiple of  $\mathbb{T}$ , i.e. we can write  $\mathbb{T}''(s) = \mu(s) \mathbb{T}(s)$ . Differentiating the equation  $\mathbb{T} \cdot \mathbb{T}' = 0$  gives that  $\kappa^2 + \mu(s) = 0$ , so that  $\mu(s) = -\kappa^2$  for all  $s$ . Thus we have  $\mathbb{T}'' = -\kappa^2 \mathbb{T}$ . The only solutions to this differential equation are

$$\mathbb{T} = (-\sin(\kappa s + \phi), \cos(\kappa s + \phi)).$$

But this means that  $\mathbf{c}(s) = (a + \cos(\kappa s + \phi), b + \sin(\kappa s + \phi))$ , which traces out the required circle.  $\square$

More generally,  $\kappa$  determines any curve, up to a Euclidean symmetry.

# Chapter 18

## Moving Frames

Let's start by considering frames in  $\mathbb{R}^3$ . Let  $e_1, e_2, e_3$  be a right handed smooth orthonormal frame. Since we are working in Euclidean space, we can think of the  $e_n$  as vector fields. Because  $e_1, e_2, e_3$  are a basis, we can write

$$e'_i = \omega_{i1}e_1 + \omega_{i2}e_2 + \omega_{i3}e_3$$

Since  $e_i \cdot e_j = \delta_{ij}$ , we can differentiate on both sides to conclude that  $e'_i \cdot e_j + e_i \cdot e'_j = 0$ . This implies

$$\omega_{i1} = e'_i \cdot e_1 = -e'_1 \cdot e_i = -\omega_{1i}$$

Thus the matrix  $\Omega = (\omega_{ij})$  is skew symmetric. In particular,  $e'_i \cdot e_i = 0$ . If we consider the one forms  $\sigma_k = dx(e_k)dx + dy(e_k)dy + dz(e_k)dz$ , then

# Chapter 19

## Differential Topology Notes

A *connection*  $d_A$  on a bundle  $E$  is an  $\mathbb{R}$  linear map  $d_A : \Omega^0(M, E) \rightarrow \Omega^1(M, E)$  satisfies  $d_A(fs) = df \wedge s + fd_As$  for all  $f \in C^\infty(M)$ ,  $s \in \Gamma(E)$ . Note that the support of  $d_As$  must be contained in the support of  $s$ . The operator  $d_A$  is local, and for local frames  $s^1, \dots, s^n$ , we have  $d_As^i = \sum a_j^i s^j$ . This is the connection matrix, or matrix valued one forms, or gauge fields.

### 19.1 November 2nd

Recall that the space of endomorphisms on a bundle  $E$  is isomorphic naturally to  $E \otimes E^*$ . Given a connection  $A$  on  $E$ , we get an induced connection on  $E \otimes E^*$ , also denoted by  $A$  (though really we should denote it by  $A \otimes A^*$ ). If  $g$  is a section of  $E \otimes E^*$ , then  $g$  acts on a section  $s$  of  $E$  to give another section  $gs$ . We let  $(d_Ag)(s) = d_A(gs) - g(d_As) = [d_A, g]s$ . We can check that  $(d_Ag)(fs) = f(d_Ag)(s)$  so  $d_Ag$  is a one form on  $E \otimes E^*$ , and also  $d_A(fg) = (df)g + f(d_Ag)$ , so  $d_A$  is a connection on  $E \otimes E$ . More generally,  $d_A(g \circ h) = (d_Ag) \circ h + g \circ (d_Ah)$ , which can be verified by a quick calculation. If  $\theta$  is a  $p$  form on  $\text{End}(E)$ , we let  $(d_A\theta)(s) = d_A(\theta s) - (-1)^p \theta(d_As)$ .

**Theorem 19.1** (Bianchi Identity). *If  $A$  is a connection on  $E$ , inducing a curvature  $F_A$ , then  $d_AF_A = 0$ .*

*Proof.*

$$d_AF = [d_A, F_A] = d_A \circ F_A - F_A \circ d_A = d_A^3 - d_A^3 = 0$$

□

### 19.2 The Variation of Curvature Formula

The space  $A(E)$  is affine moddled on  $\Omega^1(M, \text{End}(E))$ . Given  $A_0 \in A(E)$ , every connection is of the form  $A_0 + a$ , for some one form  $a$ . Then  $d_A = d_{A_0} + a$ , and

$$F_{A_0+a} = (d_{A_0} + a)^2 = d_{A_0}^2 + d_{A_0}a + ad_{A_0} + a^2 = F_{A_0} + (d_{A_0}a) + a^2$$



Given any element  $a$  of  $\Omega^1(M, \text{End}(E))$ , the trace  $\text{tr}$  gives us a one form  $\text{tr}(a)$  on  $E$ .

**Lemma 19.2.** *For any  $\eta \in \Omega^p(M, \text{End}(E))$ , and  $\theta \in \Omega^q(M, \text{End}(E))$ ,*

$$\text{Tr}(\theta\eta) = \text{Tr}((-1)^{pq}\eta\theta)$$

*and for any connection  $A \in A(E)$ ,*

$$d\text{Tr}(\theta) = \text{Tr}(d_A\theta)$$

*Proof.* The first statement is immediate when written in any local frame. For the second property,

$$d_{A+a}\theta = [d_A + a, \theta] = [d_A, \theta] + [a, \theta] = (d_A\theta) + [a, \theta]$$

Taking traces on both sides, we conclude

$$\text{Tr}(d_{A+a}\theta) = \text{Tr}(d_A\theta) + \text{tr}(a\theta - (-1)^q a\theta)$$

But the right-most term is zero, which gives that the trace is independant of  $A$ . In particular, if we take a local frame and a trivial connection, we get the required formula.  $\square$

**Theorem 19.3.** *If  $E$  is a complex vector bundle, and  $A$  is a connection on  $E$ , then  $i\text{Tr}(F_A)/2\pi$  is a closed two form, and the class  $c_1(E) = [\text{itr}(F_A)/2\pi]$  is invariant of  $A$ .*

*Proof.* For any  $A$ ,  $d(\text{Tr}(F_A)) = \text{Tr}(d_A F_A) = 0$ , which gives closure. Now

$$\text{Tr}(F_{A+a}) - \text{Tr}(F_A) = \text{Tr}(F_A + d_A a + a^2 - F_A) = \text{Tr}(d_A a) + \text{Tr}(a^2) = d\text{Tr}(a)$$

because  $\text{Tr}(a^2) = 0$ .  $\square$

# Chapter 20

## Homology of Vector Bundles

The theory of characteristic classes enables us to assign invariants to smooth bundles.

Let's use this construction to obtain an interesting  $K$  vector bundle on projective space. A point in projective space  $\mathbf{RP}^n$  can be viewed as a line through the origin in  $\mathbb{R}^{n+1}$ . We thus assign a vector bundle  $L$  known as the *Hopf line bundle* to  $\mathbf{RP}^n$  by taking the subbundle of  $\varepsilon^{n+1}(\mathbf{RP}^n)$  consisting of all vectors  $v_l$ , for  $v \in \mathbb{R}^{n+1}$  and  $l \in \mathbf{RP}^n$ , such that  $v \in l$ . To see that this is truly a line bundle, we consider the continuous sections

$$l \mapsto \frac{l_i e_j - l_j e_i}{l_k}$$

which are continuous where  $l_k \neq 0$ , and nonzero where  $l_i$  and  $l_j$  are not both vanishing. We can also consider complex projective space  $\mathbf{CP}^n$ , and the resultant complex Hopf line bundle, which we shall also denote by  $L$ .

One interesting application of the Hopf line bundle is to understand the tangent bundle to projective space. If we consider  $L$  as a subbundle of  $\varepsilon^{n+1}(\mathbb{R}^{n+1})$ , with the trivial Riemannian metric, then we can consider the normal bundle  $L^\perp$ . The immersion  $\pi : S^n \rightarrow \mathbf{RP}^n$  induces  $\pi_* : TS^n \rightarrow T\mathbf{RP}^n$ , and for a given  $v \in T\mathbf{RP}^n$ ,  $\pi_*^{-1}(v)$  consists of two vectors  $w_x$  and  $-w_{-x}$ , where  $x, -x \in l$ . Thus  $v$  corresponds to the linear map  $f_v : l \rightarrow l^\perp$  mapping  $x$  to  $w$  and  $-x$  to  $-w$ . Conversely if  $f : l \rightarrow l^\perp$  is given, if we take  $x, -x \in S^n \cap l$  and consider their images  $f(x) = w$ ,  $f(-x) = -w$ , then  $\pi_*(w_x) = \pi_*(-w_{-x})$  is some vector  $v$ , and  $f = f_v$ . It is easy to see this action is smooth, and so we obtain a bundle equivalence between  $T\mathbf{RP}^n$  and  $\text{Hom}(L, L^\perp)$ . The same is true of  $T\mathbf{CP}^n$ , where the homomorphisms are complex linear.

A simple corollary of this is that  $T\mathbf{RP}^n \oplus \varepsilon^1(\mathbf{RP}^n)$  is isomorphic  $L^{\oplus(n+1)}$  to the  $n+1$  fold sum  $L \oplus \cdots \oplus L$ . First, we notice that for any line bundle  $\xi$ ,  $\text{Hom}(\xi, \xi)$  is trivial (the identity homomorphism constitutes a global section). Thus we find

$$\begin{aligned} T\mathbf{RP}^n \oplus \varepsilon^1(\mathbf{RP}^n) &\cong \text{Hom}(L, L^\perp) \oplus \text{Hom}(L, L) \cong \text{Hom}(L, L^\perp \oplus L) \\ &\cong \text{Hom}(L, \varepsilon^{n+1}(\mathbf{RP}^n)) \cong (L^*)^{\oplus(n+1)} \end{aligned}$$

The same argument justifies that  $T\mathbf{CP}^n$  is isomorphic to the complex dual  $(L^*)^{\oplus(n+1)}$ . In the case of a real vector bundle, by taking a Riemannian metric on a real dual bundle we obtain

an isomorphism between it's dual and itself, so  $L^*$  is isomorphic to  $L$ . This is no longer true in the complex space. But we do find that  $T^*\mathbf{CP}^n \oplus \varepsilon^1(\mathbf{CP}^n)$  is isomorphic to  $L^{\oplus(n+1)}$ .

## 20.1 Vector Valued Differential Forms

We have been considering the differential forms  $\omega \in \Omega^k(TM)$ , which assign to each point  $p$  an alternating map  $\omega(p)$  in  $k$  variables into  $\mathbb{R}$ . If we replace  $\mathbb{R}$  by any other vector space  $V$ , then we obtain a *vector valued differential form*. We can view these forms as sections of  $\text{Hom}\left(\bigwedge^k(TM), V\right)$ . In view of the natural isomorphisms

$$\text{Hom}\left(\bigwedge^k(W), V\right) \cong \left(\bigwedge^k(W)\right)^* \otimes V \cong \bigwedge^k(W^*) \otimes V$$

The space of vector valued differential forms is viewed as sections of  $\bigwedge^k(TM^*) \otimes V$ . We denote the space of smooth  $V$  valued  $k$  forms as  $\Omega^k(M, V)$ .

If  $V$  has a basis  $e_1, \dots, e_n$ , then for any  $\omega \in \Omega^k(M, V)$ , we have coefficients  $a_i$  such that  $\omega(p)(v_1, \dots, v_n) = \sum a^i(v_1, \dots, v_n)e_i$ . The functions  $a_i$  are alternating, and therefore they are normal  $k$  forms in  $\Omega^k(M)$ . But this means that  $\omega = \sum a^i \otimes e_i$ . In particular, we see that  $\Omega^k(M, V)$  is isomorphic to the direct sum of  $n$  copies of  $\Omega^k(M)$ .

If we allow  $V$  to vary from point to point, we obtain the notion of a differential form taking values in a vector bundle. If  $E$  is a vector bundle, then an  $E$  valued  $k$  forms assigns to each point  $p$  an alternating map from  $M_p^k$  to  $E_p$ . Arguing as before, such forms are sections of the bundle  $\bigwedge^k T^*M \otimes E$ . We denote the space of sections as  $\Omega^k(M, E)$ . We can introduce coefficients just as in  $\Omega^k(M, V)$ , but only in coordinate systems which trivialize  $E$ .

## 20.2 Algebraic Understand of Connections

Recall that a connection on a smooth bundle  $(\xi, E)$  over a manifold  $M$  is a map  $\nabla : \Gamma(M) \times \Gamma(E) \rightarrow \Gamma(E)$ , whose image for  $X \in \Gamma(M)$  and  $s \in \Gamma(E)$  is denoted  $\nabla_X(s)$ , which is  $C^\infty(M)$  linear in  $X$  and satisfies the Leibnitz rule  $\nabla_X(fs) = X(f)s + f\nabla_X(s)$ . Since a connection is surely bilinear in  $X$  and  $s$ , we can consider  $\nabla$  as a map from  $\Gamma(E)$  to  $\Omega^1(M, E)$ . The Leibnitz rule  $\nabla$  then takes the form  $\nabla(fs) = df \otimes s + f + \nabla(s)$ . A section  $s$  of a vector bundle is called *flat* if  $\nabla(s) = 0$ .

**Theorem 20.1.** *If  $\nabla, \nabla' : \Gamma(E) \rightarrow A^1(E)$  are connections, then  $\nabla - \nabla'$  is  $C^\infty(M)$  linear, and so can be identified as an element of  $A^1(\text{End}(E))$ . Conversely, if  $\nabla$  is a connection, and  $a \in A^1(\text{End}(E))$ , then  $\nabla + a$  is a connection.*

*Proof.* We calculate that

$$(\nabla - \nabla')(fs) = (df \otimes s + f\nabla(s)) - (df \otimes s + f\nabla'(s)) = f(\nabla - \nabla')(s)$$

which gives the  $C^\infty(M)$  linearity. Thus for each  $X$ ,  $(\nabla - \nabla')_X$  is a map from  $\Gamma(E)$  to itself, which by  $C^\infty$  linearity localizes to an element of  $\text{End}(E)$ . Conversely, if  $a \in A^1(\text{End}(E))$ , then we have

$$(\nabla + a)(fs) = f\nabla s + df \otimes s + fa(s) = f(\nabla + a)(s) + df \otimes s$$

so the map satisfies the Leibnitz rule.  $\square$

*Remark.* The family of all connections on a vector bundle is therefore an affine space over  $A^1(\text{End}(E))$ , and once we have found a single connection  $\nabla$ , all other connections can be written as  $\nabla + a$ , where  $a : M \rightarrow M_n(\mathbf{C})$  is a *matrix valued* one form, at least locally in coordinates.

Since  $\Gamma(E)$  is equal to  $A^0(E)$ ,  $\nabla$  maps  $A^0(E)$  to  $A^1(E)$ . We can actually extend  $\nabla$  so it maps  $A^n(E)$  to  $A^{n+1}(E)$  for all  $n$ . To do this, for  $\alpha \in \Omega^n(M)$ , and  $s \in \Gamma(E)$ , we write  $\nabla(\alpha \otimes s) = d\alpha \otimes s + (-1)^n \alpha \wedge \nabla s$ . This is a well defined bilinear operation by the Leibnitz theorem in the base situation, so  $\nabla(\alpha \otimes (fs)) = \nabla(f(\alpha \otimes s))$ . And now we have the more generalized Leibnitz rule, which for every  $\alpha \in \Omega^k(M)$  and  $\beta \in A^l(E)$  we have

$$\nabla(\alpha \wedge \beta) = (d\alpha \wedge \beta) + (-1)^k (\beta \wedge \nabla \alpha)$$

In particular, this enables us to define the *curvature*  $F_\nabla = \nabla \circ \nabla$ , mapping  $\Gamma(E)$  to  $A^2(E)$ . It is really an element of  $A^2(\text{End}(E))$ , because it is  $C^\infty(M)$  linear.

**Theorem 20.2.** *The curvature map  $F_\nabla$  is  $C^\infty(M)$  linear, so we can consider the curvature as an element of  $A^2(\text{End}(E))$ .*

*Proof.* We compute

$$\begin{aligned} \nabla(\nabla(fs)) &= \nabla(df \otimes s + f\nabla s) \\ &= (d^2f \otimes s - df \wedge \nabla s) + df \otimes \nabla(s) + fF_\nabla(s) \\ &= fF_\nabla(s) \end{aligned}$$

Thus for two vector fields  $X$  and  $Y$  we can consider  $F_\nabla(X, Y)$  as an endomorphism on  $E$  of each fibre.  $\square$

**Example.** On  $\mathbb{R}^n$ , with the canonical tangent bundle, we have the trivial connection  $\nabla_X Y = X(Y)$ . This has vanishing curvature  $F_\nabla = 0$ . Any other connection is of the form  $\nabla + A$ , where  $A$  is a matrix of one forms. We obtain that

$$\begin{aligned} F_{\nabla+A}(X) &= (\nabla + A)(\nabla + A)(X) = (\nabla + A)(\nabla X + A(X)) \\ &= A(\nabla X) + \nabla(A(X)) + A^2(X) = \nabla(A)(X) + (A \wedge A)(X) \end{aligned}$$

# Bibliography

- [1] Michael Spivak, *A Concise Introduction to Differential Geometry: Vol. One*
- [2] James Lee, *An Introduction to Smooth Manifolds*
- [3] Paul Halmos, *Naive Set Theory*
- [4] Wikipedia, *Lie Groups*