

Geometrical methods for Physics

Iberê Kuntz¹ and Christopher Fritz²

*Physics & Astronomy, University of Sussex
Falmer, Brighton, BN1 9QH, United Kingdom*

¹E-mail: ibere.kuntz@sussex.ac.uk

²E-mail: c.fritz@sussex.ac.uk

Contents

I	Geometry and Topology	3
1	Preliminaries	4
1.1	Maps and Sets	4
1.2	Morphisms	7
1.3	Topological spaces and manifolds	8
1.4	Tensors	10
2	Homotopy groups	14
2.1	Fundamental groups	14
2.2	Properties of the fundamental group	16
2.3	Examples of fundamental groups	17
3	Differential Topology	20
3.1	Charts and Coordinates	20
3.2	Functions and Curves	24
3.3	Vectors and One-forms	26
3.4	Induced maps and Flows	28
3.5	Lie Derivatives and Brackets	32
3.6	Differential Forms and Tensors	35
4	Differential Geometry	41
4.1	Metric	41
4.2	Connections	45
4.3	Geodesics	50
4.4	Torsion and Curvature	51

II	Applications	53
5	General Relativity	54
5.1	Postulates	54
5.2	Tetrad formalism	55
5.3	Hodge Map	59
5.4	Curvature and Torsion	63
5.5	Electromagnetism and Einstein's Equations	64
6	Topological defects on condensed matter	67
6.1	Topological charges	72

Part I

Geometry and Topology

Chapter 1

Preliminaries

In the present chapter we are going to introduce a few basic mathematical concepts that will be necessary in this course. We will discuss only briefly each new concept and will stick mainly to the basic definitions, without proving anything in detail. The reason for this is that it would take a whole course to go through the theory underlying each of them. The good news is that we do not need the whole theory to start doing geometry. The only prerequisites for this chapter (and for the rest of course) are calculus and linear algebra.

1.1 Maps and Sets

From the point of view of a physicist, perhaps the most daunting aspect of differential geometry is its comparative rigor and accompanying notation. So before diving headlong into the subject, it might be worth spending the space of a chapter just elucidating some of the more basic concepts and notations that can make texts on differential geometry seem a bit alien. Hopefully it will become clear that this is simply another way of looking at objects (like functions) which are already familiar, but allow for the concept to be generalized.

If that seems a bit incoherent, let's start by looking at an example. Take the function

$$f(x) = x^2 + 5. \tag{1.1}$$

Now what is this actually saying? Well if we insert some value x into our function, it is instructing us to square that value and then add five. The result is another number, so we could take $x = 1$ and get 6 or $x = 5$ and get 30. So what is actually happening here is that the function f is assigning a number to each value of x (also a number). Since

clearly, both belong to the real numbers \mathbb{R} , we can think of f as a **map** from the space of real numbers to the space of real numbers

$$f : \mathbb{R} \longrightarrow \mathbb{R}. \tag{1.2}$$

We could also take a function in several variables, i.e.

$$f(x, y, z) = y^{-3} - e^{3z} + x - 2, \tag{1.3}$$

which is a map that takes three numbers into a single value or

$$f : \mathbb{R}^3 \longrightarrow \mathbb{R}. \tag{1.4}$$

The symbol \mathbb{R}^3 simply means that we take three copies of the real line (one for each variable) and map it into a single number. More generally we have a function in an arbitrary number of variables which we now denote

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}. \tag{1.5}$$

We could talk further about this example, but the point here is that we have changed the way in which we look at functions; rather than an expression in some variables, we can see its “underlying property” as a map. This doesn’t just apply to functions, but can be generalized to maps between any kind of (mathematical) object.

Now, let X and Y be two sets and define a map as a procedure through which we assign to each $x \in X$ some $y \in Y$

$$f : X \longrightarrow Y. \tag{1.6}$$

The set X is referred to as the **domain** and Y as the **range** of f . But, the map f needn’t *use up* all of Y . So we have the **image**, which is all points in Y assigned to points in X through f

$$f(X) = \{y \in Y \mid y = f(x), x \in X\} \tag{1.7}$$

This is another bit of notation we will have to get used to. It denotes the set $f(X) \subset Y$ and can be read as: The set of all elements y in Y where each such element is given by mapping some element x in X to Y . It may be a bit of a mouthful, but it is simply all elements of Y that satisfy a certain condition on the right hand side of \mid . Returning to our example above, if we wanted a map which assigns to each real number (domain) a positive real (range) number we would write

$$f : \mathbb{R} \longrightarrow \mathbb{R}_{>0} \tag{1.8}$$

and define the range as the set

$$\mathbb{R}_{>0} = \{x \in \mathbb{R} \mid x > 0\} \tag{1.9}$$

whose elements are strictly greater than zero. We will look at more examples shortly, but first let us consider some properties of maps. Firstly a map $f : X \longrightarrow Y$ can have certain properties (see Figure 1.1):

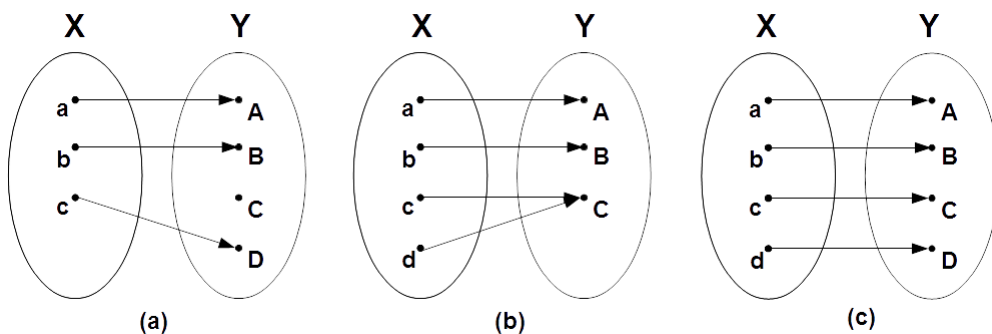


Figure 1.1: a) Injective function, b) surjective function and c) bijective function.

- **Injective** (or one-to-one): Each element in the domain maps to *one and only one* element of the range i.e. $\forall x, x' \in X$ if $x \neq x'$ then $f(x) \neq f(x')$.
- **Surjective** (or onto): Each element in the domain maps to *at least one* element of the range i.e. $\forall y \in Y$ there exists at least one $x \in X$ such that $y = f(x)$.
- **Bijective** (one-to-one correspondence): Each element in the domain maps to *one and only one* element of the *entire* range i.e. a map f is both injective and surjective. A map f has an inverse if and only if it is bijective.

Although a map needn't be any of these. So, to take an example, if we let $f : \mathbb{R} \rightarrow \mathbb{R}$ and $f(x) = \cos(x)$, then this function is neither injective nor surjective: Since $\cos(0) = \cos(2\pi) = \cos(4\pi) = \dots = 1$, injectivity fails, and since $\cos(x)$ only takes values in the range from -1 to 1 so does surjectivity. However, we can define the sets $X = \{x \in \mathbb{R} \mid 0 \leq x \leq 2\pi\}$ and $Y = \{x \in \mathbb{R} \mid -1 \leq x \leq 1\}$ and take these to be the domain and range respectively. Now the map $f : X \longrightarrow Y$ is bijective, so both surjective and injective.

Another important property is that maps can be **composed**: If we have $f : X \rightarrow Z$ and $g : Z \rightarrow Y$ then we also have a map $h : X \rightarrow Y$ given by the composition

$$h = g \circ f. \tag{1.10}$$

Which is instructing us to first apply f followed by g . Note that this only works if the domain of g corresponds to the range of f . For example, consider $f, g : \mathbb{R} \rightarrow \mathbb{R}$ (which may be a bit reductive, but ought to illustrate the point) where this is clearly satisfied and fix f, g to be the functions $f(x) = \cos(x)$ and $g(x) = x^2 - 1$. Now we can find the composite function $h = g \circ f$ of some number x by $h(x) = g \circ f(x) = g(f(x)) = \cos^2(x) - 1$.

We also note the existence of inverse maps; that is for some $f : X \rightarrow Y$ there may also (but not necessarily) exist an inverse $f^{-1} : Y \rightarrow X$. They can be composed $f^{-1} \circ f = f \circ f^{-1} = \text{id}$ where id is the **identity operator** and simply maps elements of a set back to themselves. So once again we take $f : \mathbb{R} \rightarrow \mathbb{R}$ as $f(x) = \cos(x)$ which also gives us $f^{-1} : \mathbb{R} \rightarrow \mathbb{R}$ as $f^{-1}(x) = \cos^{-1}(x)$ and $f^{-1} \circ f(x) = \cos^{-1}(\cos(x)) = x$.

1.2 Morphisms

The modern way to deal with mathematics is through structures. A mathematical structure is to be viewed as some sort of relations imposed over a set of objects (this could be numbers, or matrices, or bananas, etc). A set A together with some structure Ω forms what we will call space and which will be denoted as an ordered pair (A, Ω) . Essentially there are two types of structures: the algebraic structures and the geometrical ones. How to define them and consequently classify all of the mathematical structures in these two categories is not a trivial task, but this will not bother us here. In fact, in this course we will only assume that a structure is algebraic whenever there is a product \circ between the elements of the set A , denoted (A, \circ) .

In this sense, the trend in mathematics is to isolate some structure from the rest and try to understand everything about it, exhausting all of the possible definitions, theorems and so on. The way this usually happens is through the generalization of intuitive concepts. The idea of distance, for example, is very intuitive to all of us, but the mathematicians have gone beyond that and have introduced a new structure called metric space, turning the results we know from Euclidian space into axioms. This way we can get a deep understanding of what distance is without getting confused due to other concepts that are also present in Euclidian space.

In order to better understand some structure, we have to know when two spaces are equivalent with respect to this structure. To this notion of equivalence we give the name of morphism. More precisely, if A and B are two sets together with some structure, then a morphism between A and B is a map $\phi : A \rightarrow B$ which preserves the underlying structure. In the special case that ϕ has an inverse $\phi^{-1} : B \rightarrow A$ that is also a morphism, we call ϕ an isomorphism. If there is an isomorphism between A and B , then A and B are said to be isomorphic. When the structure is algebraic, a morphism $\phi : (A, \circ) \rightarrow (B, \cdot)$ means that

$$\phi(a \circ b) = \phi(a) \cdot \phi(b),$$

for every $a, b \in A$. In this case the word homomorphism is often used instead of morphism.

Example 1.2.1.

1. If A and B are sets with no structure, then any function $f : A \rightarrow B$ is a morphism. If f is invertible, then it is an isomorphism.
2. If V and W are vector spaces, then any linear transformation $T : V \rightarrow W$ is a homomorphism.



As we go ahead in this course we will find different structures and different types of morphisms with respect to it. Our job here is to study these structures and to identify them in physics. By the way, this is essentially the role of mathematical physics.

1.3 Topological spaces and manifolds

One of the most important concepts in mathematics and that also appears everywhere in physics is the idea of the *continuum*. We assume almost everywhere in physics that the functions are continuous, even though this is not always explicitly mentioned. The reader might be wondering right now how we could talk about continuity if the only thing we have is a set of numbers. The answer is in another structure called topology¹.

Definition 1. *Let X be a set. A topology on X is a collection τ of subsets of X , called open sets, satisfying:*

- a) *The entire set X and the empty set \emptyset are open;*

¹The word topology might refer to either the structure or the field/subject itself.

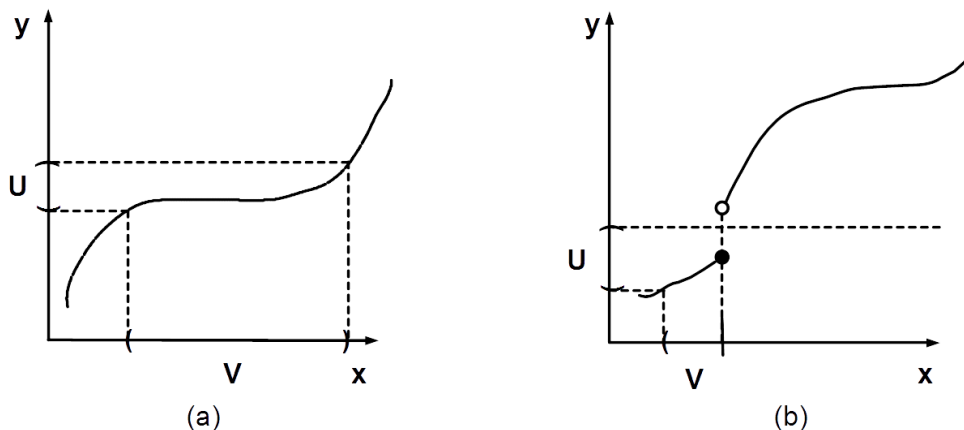


Figure 1.2: A map $f : V \rightarrow U$. (a) f is continuous and the inverse image of U is open. (b) f is discontinuous and the inverse image of U is not open; V includes its end points.

b) *The union of any family of open sets is open;*

c) *The intersection of any finite family of open subsets is open.*

The pair (X, τ) is called a topological space. Whenever the topology τ is clear from the context, we will write X for the topological space.

Note that the notion of open sets was not defined previously and here they appear as mere objects belonging to a topological space. In topology, open sets are not defined indeed, but they are characterized by the rules of the above definition. This happens quite often in mathematics and for this reason we give these non-defined objects a name: primitive concepts. In summary we do not know what they are but we know how to operate them. Likewise, vector is a primitive concept, so if you cannot accept open sets because they are not defined, always remember that you have been operating vectors since your childhood.

Before we jump to continuity, here is an useful definition: a neighborhood of $p \in X$ is an open set containing p . Now we are able to talk about continuity. Let X and Y be topological spaces. A function $F : X \rightarrow Y$ is continuous if for every open set $U \subset Y$ the preimage $F^{-1}(U)$ is open in X . In other words, a continuous function is simply a function which maps open sets into open sets in the reverse direction. This is the essence of continuity (see the illustration in 1.2). When $X = Y = \mathbb{R}$, it can be proved that this definition reduces to that one you probably have seen in your calculus course.

Further, a map (in the sense of a function) has associated to it a so-called **differentiability class** denoted by C^n . This relates to its properties under differentiation so that, for example, a continuous map whose first order derivative is also continuous, but where the second order isn't, is in C^1 . All continuous maps are automatically in C^0 . All functions whose k^{th} order derivatives exist and are continuous, are in C^k . E.g. $f(x) = |x|$, which is obviously in C^0 . However, its derivative $f'(x) = 1$ if $x > 0$ and $f'(x) = -1$ if $x < 0$ is not continuous and hence is not an element of C^1 . Meanwhile, the function $f(x) = |x|^2$

has first order derivative $f'(x) = |x|$ if $x > 0$ and $f'(x) = -|x|$ if $x < 0$ which is continuous, but its second order derivative is not. Hence $|x|^2$ is in C^1 . In general we have $|x|^n \in C^{n-1}$. On the other hand, if a function is infinitely many times differentiable with all derivatives being continuous then it is in the special class C^∞ and is called **smooth**. Examples include the ordinary exponential and trigonometric functions.

Since a continuous function maps open sets into open sets, it preserves the topology structure. Therefore, continuous functions are the morphisms of topological spaces. The isomorphisms in this context are called homeomorphisms (be careful, this is not the same as the homomorphisms).

As you may have realized, topological spaces are very general spaces and this allows us to obtain very general results as well. Unfortunately, on the other hand, they are so general that they include some pathological cases that are not of our interest. As an example, consider the topology $\tau = \{X, \emptyset\}$ over X , formed by the entire set itself and the empty set; this is called the trivial topology. In this case it can be proved that every sequence converges to every point of X and that every function into X is continuous.

To avoid these cases, which appears when X does not have enough subspaces, the following additional axiom is often assumed. For every pair of different points $p, q \in X$, there exist disjoint open subsets $U, V \subset X$ (i.e., $U \cap V = \emptyset$) such that $p \in U$ and $q \in V$. Topological spaces that satisfy this axiom are called Hausdorff topological space or simply Hausdorff space.

A topological manifold M of dimension n is a Hausdorff topological space M which for every point $p \in M$ there exists a neighborhood of p homeomorphic to an open set of \mathbb{R}^n . Intuitively speaking, this means that every neighborhood of M looks like the usual real space.

1.4 Tensors

It is impossible to overstress the usefulness of tensors for us. In mathematics they play a very important role in a broad range of fields, specially in geometry. In physics they appear everywhere as well, from moment of inertia to quantum field theory. Sadly tensors are still poorly introduced in physics through its components and as objects defined by some transformation. This leads to a lot of misconceptions and confusions. Here we will introduce them in an abstract way, without using coordinates, and then show that when we pick up some coordinate frame we get the usual formulas used in the physics literature.

Before anything, it is important to be clear concerning notations. We are going to use subscripts in vectors and superscript in coordinates of vectors whenever necessary. On the other hand, dual vectors will receive superscripts and its coordinates subscripts. In

addition, we will use Einstein notation when summing over indices, that is, we will sum over repeated indices and will omit the sum sign. So for example, in a basis $\{e_i\}$ a vector will be expressed as $v = v^i e_i$, while in a basis $\{\epsilon^j\}$ a dual vector will be expressed as $\omega = \omega^j \epsilon_j$.

From linear algebra we know that a linear functional is a map $T : V \rightarrow \mathbb{R}$ from a vector space V to the real numbers which satisfies linearity:

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v), \quad \alpha, \beta \in \mathbb{R} \quad u, v \in V.$$

Now let V_1, \dots, V_n be vector spaces. A multilinear map is a map $T : V_1 \times \dots \times V_n \rightarrow \mathbb{R}$ which satisfies linearity in each of its components. Multilinear maps can also be added and multiplied by scalars:

$$(aT + bS)(v_1, \dots, v_n) = aT(v_1, \dots, v_n) + bS(v_1, \dots, v_n).$$

Therefore, the set of all multilinear maps also forms a vector space itself, denoted

$$V_1^* \otimes \dots \otimes V_n^*$$

and called tensor product of dual spaces. Now let's consider s copies of a vector space V and r copies of its dual V^* , that is,

$$T : \underbrace{V^* \otimes \dots \otimes V^*}_{r \text{ times}} \otimes \underbrace{V \otimes \dots \otimes V}_{s \text{ times}} \rightarrow \mathbb{R}.$$

These multilinear maps are called tensors of type (r, s) on V . Since the dual of the dual space V^{**} is isomorphic to V (prove it), the vector space of tensors can be written as

$$V^{(r,s)} \equiv \underbrace{V \otimes \dots \otimes V}_{r \text{ times}} \otimes \underbrace{V^* \otimes \dots \otimes V^*}_{s \text{ times}}.$$

We usually set $V^{(0,0)}$ to be the real line \mathbb{R} and hence the tensors in this case are scalars. For $V^{(0,1)} = V^*$ we have the usual linear functionals already found in linear algebra (also called covectors) and for $V^{(1,0)} = V$ we have the ordinary vector space.

Let T and S be tensors of type (r, s) and (p, q) respectively. We define the tensor product between them as the tensor of type $(r + p, s + q)$ given by

$$\begin{aligned} T \otimes S(\omega_1, \dots, \omega_r, \rho_1, \dots, \rho_p, u_1, \dots, u_s, v_1, \dots, v_q) \\ = T(\omega_1, \dots, \omega_r, u_1, \dots, u_s) S(\rho_1, \dots, \rho_p, v_1, \dots, v_q). \end{aligned}$$

Exercise 1.4.1. Show that the tensor product is indeed a tensor by showing it is linear in all of its arguments.

Up to now we have not introduced any coordinates. In physics language, this means that everything at this point does not depend on the observer. Now we are going to introduce such coordinates, but always bear in mind that this is just a way to express a tensor in *that* special coordinate system. Naturally the tensor coordinates will change from one coordinate system to another, but this does not mean that the tensor itself is changing. It is redundant to prove this (and you will not find any serious book proving it) because the tensor is not defined using coordinates. So let's say $\{e_i\}$ is a basis of V and $\{\epsilon^j\}$ a basis of V^* . The set of vectors

$$\{e_{i_1} \otimes \dots \otimes e_{i_r} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_s}\}$$

forms a basis of the vector space $V^{(r,s)}$. Therefore, any tensor T of type (r,s) can be expressed in this basis as

$$T = T_{j_1 \dots j_s}^{i_1 \dots i_r} e_{i_1} \otimes \dots \otimes e_{i_r} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_s}, \quad (1.11)$$

where the coefficients are defined by $T_{j_1 \dots j_s}^{i_1 \dots i_r} = T(\epsilon^{i_1}, \dots, \epsilon^{i_r}, e_{j_1}, \dots, e_{j_s})$. These coefficients are what physicists usually call tensor, but as you can see, they are actually obtained when the actual tensors are expressed in a coordinate basis.

After introducing coordinates, the next natural question is: how the coefficients of T expressed in different basis are related to each other? Or physically, how do we compare the physical quantity T measured by different observers? Let $\{e_i\}$ and $\{e'_j\}$ be two different bases of V . They are related by the changing basis matrix $A = [A_i^j]$:

$$e_i = A_i^k e'_k. \quad (1.12)$$

The dual vectors transform as²

$$\epsilon^j = A_m^j \epsilon^m, \quad (1.13)$$

where A_m^j are the components of the inverse matrix A^{-1} . Using Equations (1.12) and (1.13) in (1.11) we get

$$\begin{aligned} T &= T_{j_1 \dots j_s}^{i_1 \dots i_r} e_{i_1} \otimes \dots \otimes e_{i_r} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_s} \\ &= T_{j_1 \dots j_s}^{i_1 \dots i_r} A_{i_1}^{k_1} \dots A_{i_r}^{k_r} A_{m_1}^{j_1} \dots A_{m_s}^{j_s} e'_{k_1} \otimes \dots \otimes e'_{k_r} \otimes \epsilon^{m_1} \otimes \dots \otimes \epsilon^{m_s} \\ &= T_{m_1 \dots m_s}^{k_1 \dots k_r} e'_{k_1} \otimes \dots \otimes e'_{k_r} \otimes \epsilon^{m_1} \otimes \dots \otimes \epsilon^{m_s}, \end{aligned}$$

²The fact that the dual vectors transform with the inverse of the changing basis matrix is a result from Linear Algebra. If you do not remember that I encourage you to try to prove it.

where the unprimed coefficients are the components of T with respect to the basis $\{e_i\}$ and the primed ones are the components of T with respect to $\{e'_j\}$. Therefore, the components of the tensor T transform as

$$T'^{k_1 \dots k_r}_{m_1 \dots m_s} = T^{i_1 \dots i_r}_{j_1 \dots j_s} A^{k_1}_{i_1} \dots A^{k_r}_{i_r} A'^{j_1}_{m_1} \dots A'^{j_s}_{m_s}. \quad (1.14)$$

And this is the “definition” of tensors one encounters in physics.

Example 1.4.1. We will see on Chapter 4 that a metric g is a tensor of type $(0, 2)$. So if $\{e^i\}$ is a basis for V^* we can write

$$g = g_{ij} e^i \otimes e^j.$$

In this case, Equation 1.14 becomes

$$g'_{ij} = g_{mn} A_i^m A_j^n.$$

■

Chapter 2

Homotopy groups

Mathematically, the classification of topological defects is done with the use of homotopy groups, which are part of a huge mathematics field known as Algebraic Topology, whose aim is to use algebra to study topological problems.

Then naturally, before learning about the physics of topological defects, we have to introduce some tools from the homotopy theory.

2.1 Fundamental groups

Let $\alpha : I = [0, 1] \rightarrow X$ be a curve in X such that $\alpha(0) = \alpha(1) = x_0$. Such curves are called loops based at x_0 . It is possible to endow the set of all of these loops with an algebraic structure as follows.

Definition 2. Let $\alpha, \beta : I \rightarrow X$ be loops such that $\alpha(1) = \beta(0)$. The product $\alpha\beta$ is defined as:

$$\alpha\beta(s) = \begin{cases} \alpha(2s), & 0 \leq s \leq \frac{1}{2}, \\ \beta(2s - 1), & \frac{1}{2} \leq s \leq 1. \end{cases}$$

Geometrically, this definition corresponds to the curve obtained by traversing the image of α and then the image of β . It is straightforward to show that the loop $\alpha\beta$ is continuous since $\alpha(1) = \beta(0)$.

To proceed with the construction of the algebraic structure, we also have to define the inverse and the identity elements. With the last definition in mind it is natural to define $\alpha^{-1}(s) = \alpha(1 - s)$, for each $s \in I$, which corresponds to traversing the loop α in the other

way around. Finally, the identity element is defined by $c_x(s) = x$, i.e., the image $c_x(I)$ is a single point.

These definitions could suggest that $\alpha^{-1}\alpha = \alpha\alpha^{-1} = c_x$, however this is not true as can be easily verified. To manage this problem we need the concept of homotopy.

Definition 3. A homotopy in x_0 is a continuous function $h : I \times I \rightarrow X$ such that:

$$h_0(s) = \alpha(s), \quad h_1(s) = \beta(s), \quad \forall s \in I, \quad (2.1)$$

$$h_t(0) = h_t(1) = x_0, \quad \forall t \in I, \quad (2.2)$$

where α and β are loops based at x_0 . Then we say that there exists a homotopy between α and β based at x_0 , written $\alpha \sim \beta$. In this conditions, α and β are said to be homotopic.

In other words, two loops are homotopic if they can be continuously deformed to each other.

Remarkably, the relation $\alpha \sim \beta$ is an equivalence relation and its equivalence class $[\alpha] = \{\beta | \alpha \sim \beta\}$ is called the homotopic class of α . In fact, the relation \sim satisfies the properties:

- *Reflexivity*: $\alpha \sim \alpha$. The homotopy can be given by $h_t(s) = \alpha(s)$ for any $t \in I$.
- *Symmetry*: Let $\alpha \sim \beta$ with the homotopy $h_t(s)$ such that $h_0(s) = \alpha(s)$ and $h_1(s) = \beta(s)$. Then $\beta \sim \alpha$, where the homotopy is given by $h_{1-t}(s)$.
- *Transitivity*: Let $\alpha \sim \beta$ and $\beta \sim \gamma$. Then $\alpha \sim \gamma$. In fact, if $f_t(s)$ is a homotopy between α and β and $g_t(s)$ is a homotopy between β and γ , a homotopy between α and γ can be given by

$$h_t(s) = \begin{cases} f_{2t}(s), & 0 \leq t \leq \frac{1}{2} \\ g_{2t-1}(s), & \frac{1}{2} < t \leq 1. \end{cases}$$

Equipped with the set of homotopy classes at x , denoted by $\pi_1(X, x)$, and defining the operation $[\alpha][\beta] = [\alpha\beta]$, we get the so called fundamental group. The fundamental group is a group indeed, since it satisfies all of the axioms of a group:

- *Associativity*: A homotopy $f_t(s)$ between $(\alpha\beta)\gamma$ and $\alpha(\beta\gamma)$ can be given by

$$f_t(s) = \begin{cases} \alpha\left(\frac{4s}{1+t}\right), & 0 \leq s \leq \frac{1+t}{4}, \\ \beta(4s-t-1), & \frac{1+t}{4} < s \leq \frac{2+t}{4}, \\ \gamma\left(\frac{4s-t-2}{2-t}\right), & \frac{2+t}{4} < s \leq 1. \end{cases}$$

Therefore, we can simply write $[\alpha\beta\gamma]$ to denote $[(\alpha\beta)\gamma]$ ou $[\alpha(\beta\gamma)]$.

- *Identity element*: Define a homotopy as

$$f_t(s) = \begin{cases} \alpha\left(\frac{2s}{1+t}\right), & 0 \leq s \leq \frac{1+t}{2}, \\ x, & \frac{1+t}{2} < s \leq 1. \end{cases}$$

Clearly this is a homotopy between αc_x and α . Analogously, a homotopy between $c_x\alpha$ and α is given by

$$f_t(s) = \begin{cases} x, & 0 \leq s \leq \frac{1-t}{2}, \\ \alpha\left(\frac{2s-1+t}{1+t}\right), & \frac{1-t}{2} < s \leq 1. \end{cases}$$

This show that $[\alpha][c_x] = [c_x][\alpha] = [\alpha]$.

- *Inverse*: Define a homotopy $f_t(s)$ by

$$h_t(s) = \begin{cases} \alpha(2s(1-t)), & 0 \leq s \leq \frac{1}{2}, \\ \alpha(2(1-s)(1-t)), & \frac{1}{2} < s \leq 1. \end{cases}$$

Naturally, $f_0(s) = \alpha\alpha^{-1}$ and $f_1(s) = c_x$ and, consequently,

$$[\alpha\alpha^{-1}] = [\alpha][\alpha^{-1}] = [c_x].$$

This shows that $[\alpha^{-1}] = [\alpha]^{-1}$.

Briefly, the fundamental group $\pi_1(X, x_0)$ is the group formed by all of the homotopy classes at x_0 , where $[c_x]$ is the identity element and, given $[\alpha]$, $[\alpha]^{-1} = [\alpha^{-1}]$ is its inverse element.

2.2 Properties of the fundamental group

In this section some of the most important properties of the fundamental group will be presented. It is worth noting that the fundamental group was built considering very broad conditions. Nonetheless, if we restrict our attention to some more specific cases (but still general enough to have an enormous applicability) we could be able to prove interesting properties.

In this sense, consider a path connected topological space X (i.e., a space which any two points $x_0, x_1 \in X$ can be connected by a continuous curve α such that $\alpha(0) = x_0$ and $\alpha(1) = x_1$). It is possible to show that there exists an isomorphism between $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$. The proof is out of the scope of this course and can be found in any introductory book about homotopy. With this result, we can omit the base point of the homotopy in path connected spaces and simply write $\pi_1(X)$.

The homotopy of loops can be easily generalized to continuous functions. Let X, Y be topological spaces and $f, g : X \rightarrow Y$ be continuous functions. If there exists another continuous function $F : X \times I \rightarrow Y$ such that $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$, then we say that f is homotopic to g and we denote $f \sim g$. The function F is called the homotopy between f and g .

Two topological spaces X and Y are of the same homotopy type, denoted $X \simeq Y$, if there are continuous functions $f : X \rightarrow Y$ and $g : Y \rightarrow X$ such that $f \circ g \sim id_Y$ and $g \circ f \sim id_X$. The function f is called homotopy equivalence and g is its inverse.

Now it is possible to state one of the most important properties of this section: if $f : X \rightarrow Y$ is a homotopy equivalence between X and Y , $\pi_1(X, x_0)$ is isomorphic to $\pi_1(Y, f(x_0))$. As a corollary we have that the fundamental group is invariant under homeomorphisms, i.e., it is a topological invariant.

In this sense, fundamental groups classify topological spaces in a less restrict way than homeomorphisms do. Nevertheless, it is necessary to emphasize that fundamental groups are used in physics to classify maps and field configurations instead of topological spaces.

2.3 Examples of fundamental groups

Although there are no systematic procedure to calculate the fundamental group, it is possible to find it in some specific cases through some simple considerations.

In the case of a circumference S^1 the fundamental group is isomorphic to the integers; $\pi_1(S^1) \cong \mathbb{Z}$. Although the proof of this claim is not so obvious (see [B9]), its outcome can be easily understood. Let's suppose that we encircle a cylinder with an elastic band. If the elastic band encircles the cylinder n times, then this configuration cannot be continuously deformed in another which encircles the cylinder $m \neq n$ times. Moreover, if the elastic band encircles the cylinder n times and then m times, it encircles the cylinder $n + m$ times in total.

Another interesting topological space is the real projective line $\mathbb{R}P^1$ which is the topological space made through the identification of the points of a circumference to its respective antipodes. This space can be thought of as the semicircle with identified ends. For this reason, the real projective line $\mathbb{R}P^1$ is topologically equivalent to the circumference

S^1 . Therefore, its fundamental group is also isomorphic to the integers:

$$\pi_1(\mathbb{R}P^1) \cong \mathbb{Z}.$$

Besides $\mathbb{R}P^1$, there are other spaces defined analogously but with higher dimensions. To these spaces we give the name of real projective space $\mathbb{R}P^n$. In the same way as before, $\mathbb{R}P^n$ is obtained by the identification of p and $-p$ for each $p \in S^n$. However, in this case $\mathbb{R}P^n$ is not topologically equivalent to S^n . In fact, its fundamental group is

$$\pi_1(\mathbb{R}P^n) \cong \mathbb{Z}_2, \quad n > 1,$$

where $\mathbb{Z}_2 = \{\bar{0}, \bar{1}\}$ represents the quotient space of \mathbb{Z} by the equivalence relation $x \equiv y \pmod{2}$, i.e., \mathbb{Z}_2 is the set of the equivalent classes

$$\bar{a} = [a] = \{x | x \equiv a \pmod{2}\}.$$

As a result, the operation $x \equiv y \pmod{2}$ divide the integers in two classes: the odd and even numbers.

One class of important topological spaces is the one obtained by the cartesian product of other spaces. In these cases, we have that if X and Y are topological spaces, then

$$\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \oplus \pi_1(Y, y_0), \quad (2.3)$$

where \oplus represents the direct sum. To prove this statement, let's define the projections $p_1 : X \times Y \rightarrow X$ and $p_2 : X \times Y \rightarrow Y$. If α is a loop in $X \times Y$ at the point (x_0, y_0) , then $\alpha_1 = p_1(\alpha)$ is a loop in X at x_0 and $\alpha_2 = p_2(\alpha)$ is a loop in Y at y_0 . Reciprocally, any pair of loops α_1 of X at x_0 and α_2 of Y at y_0 determine an unique loop $\alpha = (\alpha_1, \alpha_2)$ of $X \times Y$ at (x_0, y_0) . Define a homomorphism

$$\begin{aligned} \phi : \pi_1(X \times Y, (x_0, y_0)) &\rightarrow \pi_1(X, x_0) \oplus \pi_1(Y, y_0) \\ \phi([\alpha]) &\mapsto \phi([\alpha]) = ([\alpha_1], [\alpha_2]). \end{aligned}$$

By construction ϕ has inverse, hence ϕ is the isomorphism we were looking for.

As an application of this result, we have that the fundamental group of the torus $T^2 = S^1 \times S^1$ is:

$$\pi_1(T^2) \cong \pi_1(S^1) \oplus \pi_1(S^1) \cong \mathbb{Z} \oplus \mathbb{Z}.$$

Another example is the cylinder $X = S^1 \times \mathbb{R}$ which has the fundamental group

$$\pi_1(X) \cong \mathbb{Z} \oplus \{e\} \cong \mathbb{Z}.$$

Up to this point we have seen the basics about homotopy to start doing physics. We could go further and generalize all of this to higher dimensions. To do so, instead of considering the fundamental group $\pi_1(X)$, we would define the n th homotopy group $\pi_n(X)$ as the group that classifies n -loops $\alpha : I^n \rightarrow X$, where I^n denotes the n -cube $I \times \cdots \times I$. For $n = 2$, for example, $\pi_2(X)$ would classify the homotopy classes of spheres in X . As our goal here is not to explain the whole theory of homotopy in detail, we leave some good references for the interested reader [A1, B9].

Chapter 3

Differential Topology

3.1 Charts and Coordinates

The notion of a topological space ought to be clear and is the starting point to define smooth manifolds. If what we want to do is study physics, which we take to mean the motions of particles, simply specifying a topology doesn't get us very far. In order to introduce and generalize such things as velocity and acceleration, we need some sort of differential structure, that is to say we need to figure out how to translate our usual flat space calculus to some arbitrary curved space. A manifold is a type of topology that in the neighborhood of each point looks like Euclidean (or in physics Minkowski) space. This is an extremely important property as it will allow us to endow our space a so-called **coordinate chart** which gives us access to the differential structure and apply our normal calculus developed in flat space. Why is this important? Conventionally, we are used to being able to visualize (so essentially to draw pictures of) the physical process we are interested in. What this equates to is being able to frame quantities like momentum and position relative to some coordinate axis (so we can specify everything in terms of x, y and z (and possibly t) coordinates). And indeed, we know that at a given scale, we can describe everything in terms of this simple 3D Euclidean (or 4D Minkowski) coordinate system. This may sound familiar as the equivalence principle in GR. What this means is that any space can be considered to be flat at small enough scales, or to put it differently, every space looks locally Euclidean (or Minkowski). Another important aspect, is that this gives us a notions of calculus on a manifold since it is, at these sufficiently small scales, simply the same calculus we are familiar with (that is, the calculus of \mathbb{R}^n).

However, a manifold¹ isn't simply isomorphic to \mathbb{R}^n . Rather, we use something called an **Atlas** which is a set of pairs $\{(U_i, \psi_i)\}$. We denote by \mathcal{M} a topological space which is called a manifold if it has

- U_i an open subset $U_i \subseteq \mathcal{M}$ which satisfies $\bigcup_i U_i = \mathcal{M}$. So the union of all subsets cover the entire manifold
- ψ_i a homeomorphism into an open subset of \mathbb{R}^n so that $\psi_i : U_i \rightarrow \psi(U_i) \subseteq \mathbb{R}^n$
- Given a nonempty intersection $U_i \cap U_j \neq \emptyset$, there exists a **transition function** $\phi_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ or, more precisely $\phi_{ij} : \psi_i(U_i \cap U_j) \rightarrow \psi_j(U_i \cap U_j)$, given by $\phi_{ij} = \psi_j \circ \psi_i^{-1}$. ϕ_{ij} which is also a homeomorphism.

See the illustration in (3.1). So, if we take a point on an m -dimensional manifold $p \in U_i \subseteq \mathcal{M}$, a map ψ_i assigns m coordinate values $\{x^\mu\}$ where $1 \leq \mu \leq m$. So think of labeling each point with coordinates $(x, y, z \dots)$. Meanwhile if $p \in U_i \cap U_j$, ψ_i assigns $\{x^\mu\}$ and ψ_j assigns $\{y^\nu\}$ where $1 \leq \mu, \nu \leq m$, the transition functions relating x to y are functions in m variables of the form $x^\mu = x^\mu(y)$. Moreover, for physics there are particular classes of manifolds that are interesting

- **Differential Manifold:** A manifold whose Atlases are equipped with transition functions that are differentiable i.e. $\phi \in C^k$
- **Smooth Manifold:** Same as a differential manifold, but the transition functions are smooth: $\phi \in C^\infty$

We will be concentrating on the latter since all examples we will encounter here are going to be of smooth manifolds. Let's illustrate this by looking at a more heuristic example: the two sphere - S^2 .

We take S^2 with unit radius embedded in \mathbb{R}^3 where points are labeled by (x, y, z) . A sphere is denoted as $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ so all points equidistant from the origin $(0, 0, 0)$. Now, since the sphere is a two dimensional surface, we want a map into \mathbb{R}^2 , where points are labeled as (X, Y) , so that we need $\phi_i : S^2 \rightarrow \mathbb{R}^2$, a function taking points on the sphere to the plane. There are many different ways of doing this;

¹Along the rest of this chapter manifolds will mean differentiable/smooth manifolds. Do not get confused with the more general notion of a topological manifold introduced in Chapter 1.

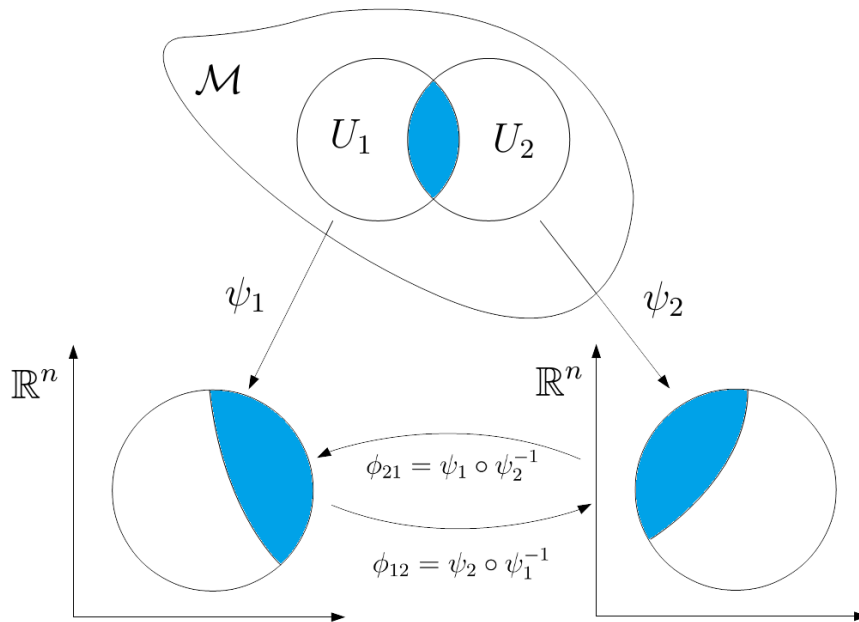


Figure 3.1: Illustration of an Atlas: The charts ψ_1 and ψ_2 map the open subsets U_1 and U_2 into \mathbb{R}^m and \mathbb{R}^n . In the nonempty intersection $U_1 \cap U_2$ one has transition functions $\phi_{12} = \phi_{21}^{-1}$ between charts.

one is so-called *stereographic projection* (see 3.2). We take the plane to lie at the height of the sphere's equator perpendicular to the z -axis and project from the north pole $(0, 0, 1)$, denoting this map ψ_N which relates points as

$$X = \frac{x}{1-z}, \quad Y = \frac{y}{1-z}. \quad (3.1)$$

Notice however, ψ_N is not well defined at the north pole where $z = 1$. That means our open set U_N cannot be all of S^2 , but rather must be a subset excluding that point. So we have a set $U_N = \{S^2 - \text{north pole}\}$. To cover the entire sphere, we need at least one more open set² which includes the north pole. A simple choice is the south pole at $(0, 0, -1)$, though we could choose any other as long as it results in a chart which includes $(0, 0, 1)$. Similar to before this gives a set $U_S = \{S^2 - \text{south pole}\}$ with chart

$$X' = \frac{x}{1+z}, \quad Y' = \frac{y}{1+z}, \quad (3.2)$$

covering all of the sphere except the south pole. In summary we have two projective charts on the sphere mapping into the subsets

²In fact, a two sphere can never be coordinated by less than two charts. If it could, it would be homeomorphic to (i.e. the same as) the plane.

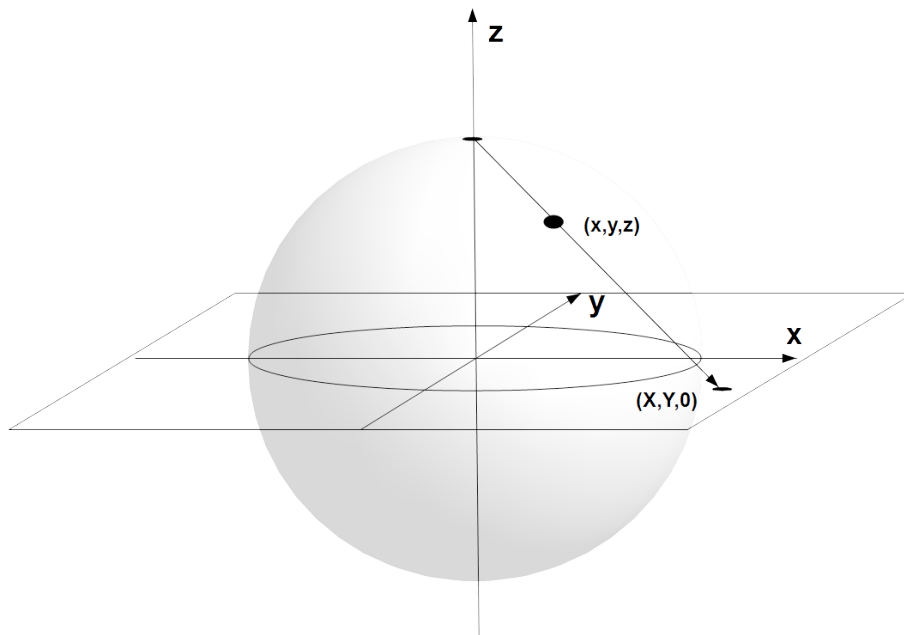


Figure 3.2: Stereographic projection of a two-Sphere: the line originating at the north pole intersects the sphere at a point and maps it on to the plane

$$\begin{aligned} \psi_N(U_N) &= \left\{ \left(\frac{x}{1-z}, \frac{y}{1-z} \right) \mid (x, y, z) \in \mathbb{R}^3 - (0, 0, 1) \right\}, \\ \psi_N(U_S) &= \left\{ \left(\frac{x}{1+z}, \frac{y}{1+z} \right) \mid (x, y, z) \in \mathbb{R}^3 - (0, 0, -1) \right\}. \end{aligned} \quad (3.3)$$

The transition function ϕ_{NS} between these two charts can be found from the definition of ϕ_{ij} by composing the above charts appropriately. They are given by

$$X = \frac{X'}{X'^2 + Y'^2}, \quad Y = \frac{Y'}{X'^2 + Y'^2}. \quad (3.4)$$

Now, as mentioned before, there is no one way of doing this and we could choose different coordinates entirely. A more familiar example are spherical coordinates: we take \mathbb{R}^2 to be parametrized by angles (α, β) and can find

$$\alpha = \cos^{-1}(z), \quad \beta = \tan^{-1}\left(\frac{y}{x}\right), \quad (3.5)$$

or conversely

$$x = \sin(\alpha) \cos(\beta), \quad y = \sin(\alpha) \sin(\beta), \quad z = \cos(\alpha). \quad (3.6)$$

The parameters (α, β) are defined over the ranges $0 \leq \alpha \leq \pi$ and $0 \leq \beta < 2\pi$, so, as with the projective chart, these coordinates don't cover the entire sphere. Our covering excludes all points along the z -axis at $(0, 0, z)$ since we get that $\beta = \tan^{-1}(0) = \infty$. Also since $\sin(0) = \sin(2\pi)$ and $\cos(0) = \cos(2\pi)$ and we want the covering to be an open

set, we exclude $\beta = 2\pi$. The end result is a single chart which covers the sphere except for the poles and the great circle section connecting them. One could think of taking an inflatable ball, cutting it open along one side and flattening it out. More coverings of the entire sphere using this parametrization could be found by taking $(\alpha + \delta, \beta + \gamma)$. We can also find transition functions between our polar coordinates and stereographic ones. These are given by (for ψ_N)

$$X = \cot\left(\frac{\alpha}{2}\right) \cos(\beta), \quad Y = \cot\left(\frac{\alpha}{2}\right) \sin(\beta). \quad (3.7)$$

So we have at least two completely different coordinate systems for the sphere, though there are many more. An everyday example of this, which also neatly demonstrates the difficulties with choosing an appropriate mapping, can be found in every Atlas (the book in this case). Earth, being (almost) a sphere, needs to be mapped into plane in order to produce a map that can be drawn on a flat piece of paper. The most familiar one is the so-called Mercator projection. This has the drawback of hugely distorting the relative proportions of landmasses the closer they are to the poles. For example, Greenland looks to be larger than Africa, though in actual fact the latter is about 14 times larger. Alternatives include the Gall-Peters and Dymaxion maps, the latter being a projection first onto a polygon (Icosahedron) and then into the plane.

You have hopefully realized that everything we have been doing here is simply an example of coordinate transformations. Different charts are simply different coordinate systems and the statement that they all parametrize the same topology, albeit in different ranges, is an idea of central importance in general relativity. What we have done is to hopefully give some different intuition for something that is already quite familiar.

3.2 Functions and Curves

Now we have established what a manifold is, the next thing is to begin to look at the structures it can possess e.g. functions, vector fields, tensors etc. Or to put it another way, how can one realize objects that are useful in physics within this formalism. To begin with, we briefly return to maps, but now specifically maps between manifolds.

So, let's take $f : M \rightarrow N$, where M and N are manifolds of dimension m and n respectively and a point $p \in M$ is mapped to a point $f(p) \in N$. With a coordinate chart (U, ψ) on M and (V, θ) on N such that $p \in U$ and $f(p) \in V$, f is assigned a coordinate representation by the composition

$$\theta \circ f \circ \psi^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^n. \quad (3.8)$$

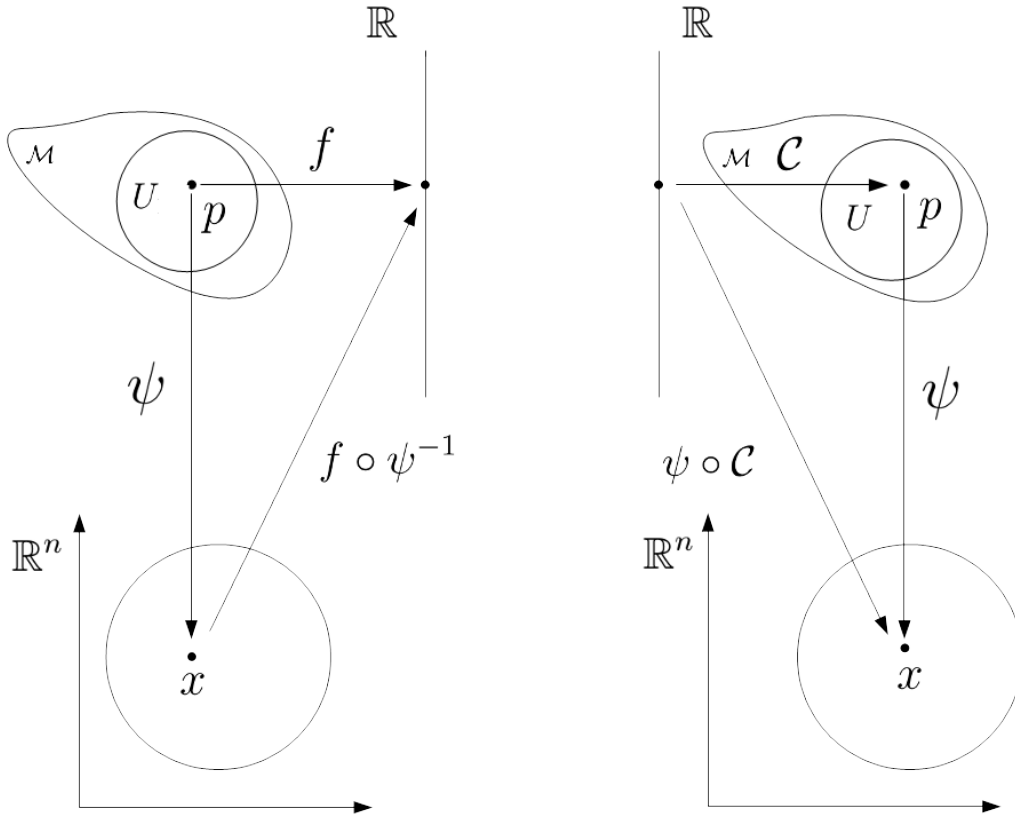


Figure 3.3: Functions and curves

Taking $\psi(p) = \{x^\mu\}$ and $\theta(f(p)) = \{y^\nu\}$ where $1 \leq \mu, \leq m$ and $1 \leq \nu \leq n$, this is simply the function $y = \theta \circ f \circ \psi^{-1}(x)$ which can be more compactly (if not entirely correctly) be expressed as $y = f(x)$. Note that if f is a diffeomorphism (as previously defined) then M and N are said to be diffeomorphic and $M \equiv N$ (that is, they are the same manifold) and obviously $\dim(M) = \dim(N)$.

With all of this in mind, we now move on to two special cases of mappings: **Functions** and **Curves**. Both of these concern a mapping between a manifold M and the real number line \mathbb{R} . A function (that being a function on a manifold as opposed to the more general treatment of the term in previous sections) is a map $f : M \rightarrow \mathbb{R}$. It has coordinate representation $f \circ \psi^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}$ and the set of smooth functions on M is denoted by $\mathcal{F}(M)$.

Conversely, a curve is a map $\mathcal{C} : \mathbb{R} \rightarrow M$ with coordinate representation $\psi \circ \mathcal{C} : \mathbb{R} \rightarrow \mathbb{R}^m$. If a curve is closed (e.g. if we identify $\mathcal{C}(0) = \mathcal{C}(1)$) it is regarded as the map $\mathcal{C} : S^1 \rightarrow M$ (S^1 being a circle). This is illustrated in 3.4.

3.3 Vectors and One-forms

We are now in a position to look at some more interesting objects. One may question the purpose of the previous section, until the difficulties in defining something like a vector become apparent. We are used to the notion of a vector as a straight arrow, but how is this concept translated to a manifold? Where is the origin? What does it even mean for something to be 'straight'? So, we define a vector using both a function and a curve - it will be a **tangent vector**. You will hopefully remember that the derivative of a function is always tangent to that function. Let's take two curves $\mathcal{C}_1, \mathcal{C}_2 : \mathbb{R} \rightarrow M$ and $\lambda \in \mathbb{R}$ a parameter along the curves so that, for a point $p \in M$, we have $p = \mathcal{C}_1(0) = \mathcal{C}_2(0)$. Then, \mathcal{C}_1 and \mathcal{C}_2 are said to be tangent if and only if

$$\left. \frac{dx^\mu(\mathcal{C}_1(\lambda))}{d\lambda} \right|_{\lambda=0} = \left. \frac{dx^\nu(\mathcal{C}_2(\lambda))}{d\lambda} \right|_{\lambda=0}. \quad (3.9)$$

Note that while this does use a coordinate chart, if two curves are not tangent in a particular chart, they won't be in any intersecting chart. A tangent vector at a point p is then the collection of all curves for which the above is satisfied. Formally referred to as an equivalence class, as a set it is expressed as

$$[\mathcal{C}] = \left\{ \mathcal{C}(\lambda) \mid \mathcal{C}_i(0) = \mathcal{C}_j(0) \text{ and } \left. \frac{dx^\mu(\mathcal{C}_i(\lambda))}{d\lambda} \right|_{\lambda=0} = \left. \frac{dx^\nu(\mathcal{C}_j(\lambda))}{d\lambda} \right|_{\lambda=0}, \forall i, j \right\}. \quad (3.10)$$

So let's be more explicit. Take a curve $\mathcal{C} : \mathbb{R} \rightarrow M$ parametrized once again by λ and a function $f : M \rightarrow \mathbb{R}$. Then find the derivative of the function along the curve and express this in local coordinates

$$\left. \frac{df(\mathcal{C}(\lambda))}{d\lambda} \right|_{\lambda=0} = \left. \frac{df}{dx^\mu} \frac{dx^\mu(\mathcal{C}(\lambda))}{d\lambda} \right|_{\lambda=0}. \quad (3.11)$$

That is, we can find the derivative of a function $f(\mathcal{C}(\lambda))$ at a point $\lambda = 0$ by applying an operator

$$X = X^\mu \frac{d}{dx^\mu}. \quad (3.12)$$

In other words

$$X[f] = X^\mu \frac{df}{dx^\mu} = \left. \frac{df(\mathcal{C}(\lambda))}{d\lambda} \right|_{\lambda=0}. \quad (3.13)$$

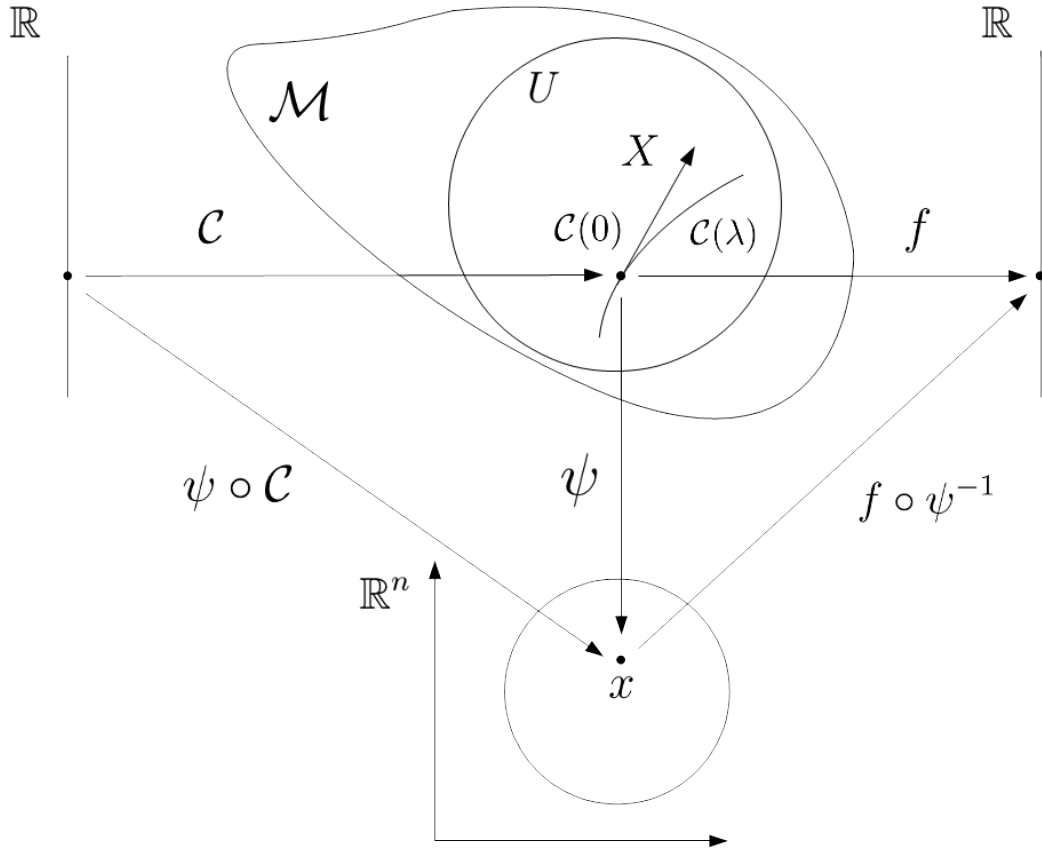


Figure 3.4: Tangent vectors are defined in terms a function f and curve \mathcal{C}

It is now this X which serves as our tangent vector, the $\frac{d}{dx^\mu}$ can be thought of as an arrow pointing along the direction of the coordinate x^μ and forms a vector basis. Furthermore, it is clearly identifiable with our basis e_μ from the preliminaries section. The X^μ 's are then the vector components. The set of all X at a point p (the equivalence class of all curves tangent to each other) is referred to as the **tangent space** and denoted $T_p M$. Their action is $X : T_p M \rightarrow \mathbb{R}^m$. The basis is regarded as $\left\{ \frac{d}{dx^\mu} \right\} \in T_p M$

Conversely we also have the notion of a dual tangent space called **cotangent space**, denoted by $T_p^* M$. In this case, an element $\eta : T_p^* M \rightarrow \mathbb{R}^m$ is referred to as a **cotangent** or alternatively a **one-form**. Some notation that will become important later is $\Omega^1(M) = T_p^* M$. As with a general vector space, the inner product exists as the natural pairing between $T_p M$ and $T_p^* M$ so that, by denoting the basis of the latter by $\{dx^\mu\}$, we have

$$\left\langle dx^\mu, \frac{d}{dx^\nu} \right\rangle = \delta^\mu_\nu. \quad (3.14)$$

A general element $\eta \in T_p^* M$ can be expressed as

$$\eta = \eta_\mu dx^\mu. \quad (3.15)$$

Lastly, let's take $p \in U_i \cap U_j$ with charts assigning $\{x^\mu\}$ and $\{y^\nu\}$ respectively. Then a vector $V \in T_p M$ can be expanded as both

$$V = V^\mu \frac{\partial}{\partial x^\mu} = \tilde{V}^\nu \frac{\partial}{\partial y^\nu}. \quad (3.16)$$

So that they are related as

$$V^\mu = \tilde{V}^\nu \frac{\partial x^\mu}{\partial y^\nu}. \quad (3.17)$$

The vector transforms so that V itself is invariant. Similarly for $\eta \in T_p^* M$

$$\eta = \eta_\mu dx^\mu = \tilde{\eta}_\nu dy^\nu, \quad (3.18)$$

so the relation is

$$\eta_\mu = \tilde{\eta}_\nu \frac{\partial y^\nu}{\partial x^\mu}. \quad (3.19)$$

A quick note on notation: Reading the preceding sections carefully ought to make expressions like $\frac{\partial f}{\partial x^\mu}$ look somewhat strange. That is because this is strictly speaking, an abuse of notation and we ought to be using

$$\frac{\partial}{\partial x^\mu} (f \circ \psi^{-1}(x)) \quad (3.20)$$

instead (since the coordinate representation of a function depends on ψ). However, for the sake of clarity, we will be sticking to the 'abuse of notation' used in this chapter as it is the more familiar and compact.

3.4 Induced maps and Flows

Now that we have seen some of the structures interesting to physics emerging, we might perhaps take some time to examine how they are manipulated. But first, we return, once again, to maps. Having introduced tangent spaces, we now need to examine their properties under mapping also. This is important since, by definition, $T_p M$ exists only at a point and is distinct from $T_{p'} M$, the tangent space at any other point. So, it stands to reason that we somehow need to relate them. Let's take another look at a

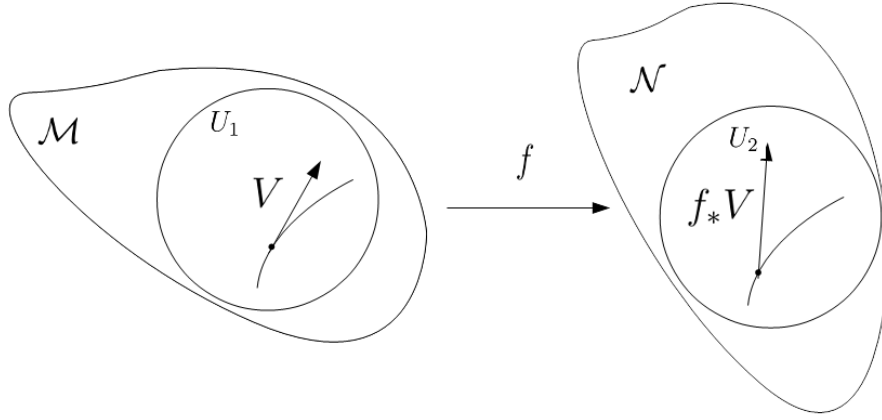


Figure 3.5: The map $f : \mathcal{M} \rightarrow \mathcal{N}$ induces the map $f_* : T_p \mathcal{M} \rightarrow T_p \mathcal{N}$ on the tangent space

smooth map $f : M \rightarrow N$ between two manifolds, recalling that it maps points $p \in M$ to $f(p) \in N$. What this does, is to give us an **induced map** which we will denote

$$f_* : T_p M \rightarrow T_{f(p)} M, \quad (3.21)$$

acting on the tangent space. That is to mean, if we have some smooth map acting on points on the manifold, it induces a map acting on the vectors that live at those points. Explicitly, taking a vector $V \in T_p M$ and denoting its action on some function by $V[g]$, we get

$$f_* V[g \circ \psi^{-1}(y)] = V[f \circ g \circ \psi^{-1}(x)] \quad (3.22)$$

in two coordinate charts on M and N . Ok, let's try to be even more explicit and take two vectors $V \in T_p M$ and $W \in T_{f(p)} N$ which are related by $f_*(V) = W$. Both can be expanded in a coordinate basis as $V = V^\mu \frac{\partial}{\partial x^\mu}$ and $W = W^\mu \frac{\partial}{\partial y^\mu}$. Then, taking $g = x^\mu$ (a coordinate function) and $f : \{x^\mu\} \rightarrow \{y^\mu\}$, we get

$$V[x^\mu] = V^\alpha \frac{\partial x^\mu}{\partial x^\alpha} = V^\mu \quad (3.23)$$

and define our mapping so that

$$(f_* V)[x^\mu] = W[x^\mu(y)] = W^\alpha \frac{\partial x^\mu(y)}{\partial y^\alpha}. \quad (3.24)$$

So that we can relate the components as

$$V^\mu = W^\alpha \frac{\partial x^\mu(y)}{\partial y^\alpha}, \quad (3.25)$$

which hopefully looks familiar as a coordinate transformation. This juxtaposes before, where we showed that a coordinate transformation was to be thought of changing charts.

That, in fact, is what is called a **passive coordinate transformation** whereas what we just derived is an **active coordinate transformation**. To illustrate this let's take our example from before - the sphere S^2 - and apply this. We look at our projective coordinate chart for the north pole and realize it as a map from \mathbb{R}^3 onto our sphere in \mathbb{R}^2 , so $f : \mathbb{R}^3 \rightarrow S^2 \subset \mathbb{R}^2$ so that

$$f : (x, y, z) \rightarrow (X, Y) = \left(\frac{x}{1-z}, \frac{y}{1-z} \right). \quad (3.26)$$

Now take some vector $V \in T_p\mathbb{R}^3$ as $V = a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} + c \frac{\partial}{\partial z}$ at some point $p = (x, y, z)$ and map it to $f_*V \in T_{f(p)}\mathbb{R}^2$. Then, applying this map

$$\begin{aligned} f_*V &= V^\nu \frac{\partial y^\mu}{\partial x^\nu} \frac{\partial}{\partial y^\mu} \\ &= a \left(\frac{\partial X}{\partial x} + \frac{\partial X}{\partial y} + \frac{\partial X}{\partial z} \right) \frac{\partial}{\partial X} + b \left(\frac{\partial Y}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Y}{\partial z} \right) \frac{\partial}{\partial Y} \\ &= a \frac{1-z-x}{(1-z)^2} \frac{\partial}{\partial X} + b \frac{1-z-y}{(1-z)^2} \frac{\partial}{\partial Y} \end{aligned} \quad (3.27)$$

Alternatively, we could also take

$$f : (x, y, z) \rightarrow (\alpha, \beta) = \left(\cos^{-1}(z), \tan^{-1} \left(\frac{y}{x} \right) \right), \quad (3.28)$$

giving

$$\begin{aligned} f_*V &= V^\nu \frac{\partial y^\mu}{\partial x^\nu} \frac{\partial}{\partial y^\mu} \\ &= a \left(\frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial y} + \frac{\partial \alpha}{\partial z} \right) \frac{\partial}{\partial \alpha} + b \left(\frac{\partial \beta}{\partial x} + \frac{\partial \beta}{\partial y} + \frac{\partial \beta}{\partial z} \right) \frac{\partial}{\partial \beta} \\ &= -a \frac{1}{\sqrt{1-z^2}} \frac{\partial}{\partial \alpha} + b \frac{x-y}{x^2+y^2} \frac{\partial}{\partial \beta}. \end{aligned} \quad (3.29)$$

To compliment the above, we also have a second map acting on the dual tangent space $f^* : T_{f(p)}^*M \rightarrow T_p^*M$. The action of this map, taking one forms $\eta : T_p^*M$ and $\epsilon : T_{f(p)}^*M$ and identifying as before $f_*(\eta) = \epsilon$, is, in components

$$\epsilon_\mu = \frac{\partial x^\mu}{\partial y^\nu} \eta^\nu. \quad (3.30)$$

Note that it acts in the opposite direction of f_* . This directionality does in fact, lend names to the maps: f_* is called **pushforward** and f^* **pullback**.

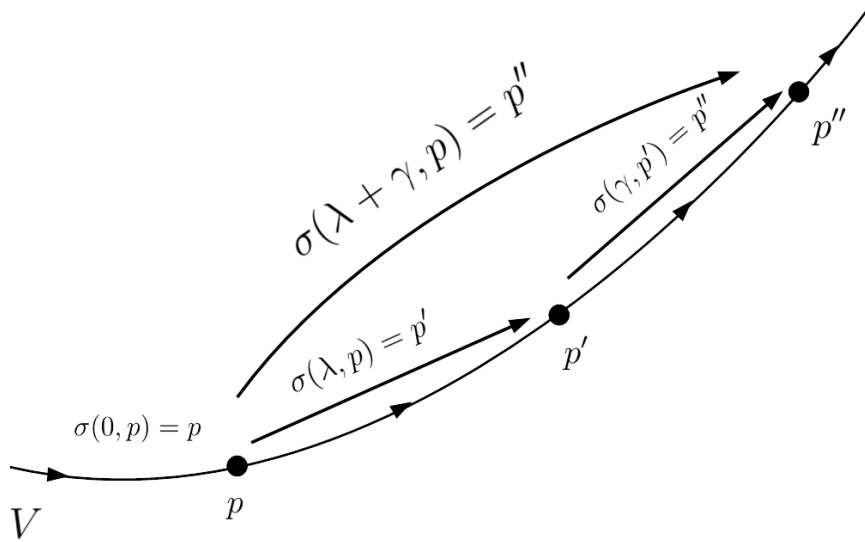


Figure 3.6: Some properties of the flow σ induced by a vector field V

The next concept we want to tackle is that of flow, since it will lead us into the subject of the next chapter. Take a curve $\mathcal{C} : \mathbb{R} \rightarrow M$ in some coordinate chart and a vector $V \in T_x M$ (for simplicity we will denote both the point on the manifold and its coordinate representation as x). Now write

$$\frac{d}{d\lambda} = \frac{dx^\mu}{d\lambda} \frac{\partial}{\partial x^\mu}, \quad (3.31)$$

so we are dealing with the tangent to that curve in a coordinate chart. For the vector write

$$V[x^\nu(\lambda)] = V^\mu \frac{\partial}{\partial x^\mu} x^\nu(\lambda). \quad (3.32)$$

Now, an **integral curve**, is a curve $x(\lambda)$ whose tangent vector at a *point* $x^\nu(\lambda)$ is $V|_x$. So using the above equations we can write

$$V^\mu(x(\lambda)) = \frac{d}{d\lambda} x^\mu(\lambda). \quad (3.33)$$

The solutions to this equation are denoted by $\sigma(\lambda, x_0)$ where x_0 corresponds to the coordinates of the point along the integral curve where $\lambda = 0$. This translates to the initial condition

$$\sigma^\mu(0, x_0) = x_0, \quad (3.34)$$

and the above is written

$$V^\mu(\sigma^\mu(\lambda, x_0)) = \frac{d}{d\lambda} \sigma^\mu(\lambda, x_0). \quad (3.35)$$

We also have some of its properties illustrated in (3.6). This map is referred to as the **flow** generated by a vector field V .

3.5 Lie Derivatives and Brackets

Now we are ready to look at some more interesting structures. It seems to be fact that many people struggle with the intuition behind Lie Derivatives. Hopefully, this section will show that as a concept these aren't so nebulous as many suppose them to be and that there is in fact, very simple intuition behind them.

To start with, let us take two vector fields V and W and their flows $\tau(\lambda, x)$ and $\sigma(\lambda, x)$ so that we have

$$V^\mu = \frac{d}{d\lambda}\tau^\mu(\lambda, x), \quad W^\mu = \frac{d}{d\lambda}\sigma^\mu(\lambda, x). \quad (3.36)$$

Now let's ask (and pay attention here), if we took a vector V at a point p somewhere on the integral curve generated by W (so $\sigma^\mu(\lambda, p)$) and then 'pushed' it along the curve some distance ϵ (so to $\sigma^\mu(\lambda + \epsilon, p)$), how would V change? This is exactly what we are trying to answer using the **Lie Derivative** and is illustrated in (3.8). For simplicity we will speak of a single vector V , though this should of course always be vector field also we will suppress most of the information in $\sigma^\mu(\lambda + \epsilon, p')$ so that this becomes σ_ϵ to save on notation (and since we are only interested in the displacement). But we must bear in mind that vectors at different points on the manifold also live in different tangent spaces. So we use what we learned about induced maps in the previous section: To compare vectors that reside at different points, we must compare one to the pullback of the other. In this case, the map we are using is the flow and as you might imagine, the flow from p to $\sigma_\epsilon(p)$ induces a pull back on the tangent space so $\sigma_{*-\epsilon} : T_{\sigma_\epsilon(p)} \rightarrow T_p$. The Lie derivative of V along W is thus defined by

$$\mathcal{L}_W V = \lim_{\epsilon \rightarrow 0} \frac{\sigma_{*-\epsilon} V |_{\sigma_\epsilon(p)} - V |_p}{\epsilon}. \quad (3.37)$$

To obtain an explicit expression for this, we take a chart with $\{x^\mu\}$ and relate coordinates at a point p to those at $\sigma_\epsilon(p)$ by

$$\begin{aligned} x'^\mu(\sigma_\epsilon(p)) &\approx x'^\mu(\sigma_0(p)) + \epsilon \frac{d}{d\lambda} x'^\mu(\sigma_\epsilon(p)) |_{\epsilon=0} + \mathcal{O}(\epsilon^2) \\ &\approx x^\mu(p) + \epsilon W^\mu(x(p)) + \mathcal{O}(\epsilon^2). \end{aligned} \quad (3.38)$$

This means we can write

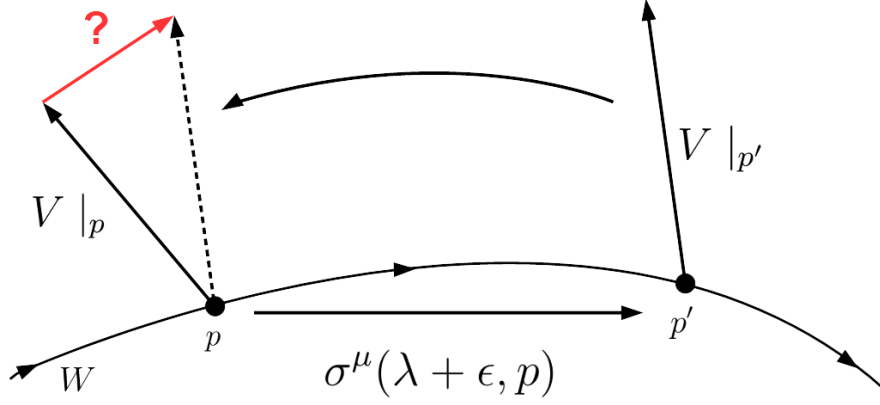


Figure 3.7: What is the Lie derivative trying to determine? How do we compare vectors that 'flow' along another vector field to each other?

$$\begin{aligned} V|_{\sigma_\epsilon(p)} &= V^\mu(x^\nu + \epsilon W^\nu(x)) \frac{\partial}{\partial x^\mu} \\ &\approx \left(V^\mu(x) + \epsilon W^\nu(x) \frac{\partial}{\partial x^\nu} V^\mu(x) \right) \frac{\partial}{\partial x^\mu}. \end{aligned} \quad (3.39)$$

Then the pullback can be calculated as

$$\begin{aligned} \sigma_{*- \epsilon} V|_{\sigma_\epsilon(p)} &= V^\nu(x(p')) \frac{\partial x^\mu(p')}{\partial x^\nu(p)} \frac{\partial}{\partial x^\nu(p)} \\ &= \left(V^\nu(x) + \epsilon W^\alpha(x) \frac{\partial}{\partial x^\alpha} V^\nu(x) \right) \left(\delta_\nu^\mu + \epsilon \frac{\partial}{\partial x^\nu} W^\mu(x) \right) \frac{\partial}{\partial x^\nu} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (3.40)$$

Substituting this in (3.37), we obtain an expression for the Lie derivative of two vector fields

$$\mathcal{L}_W V = \left(W^\mu \frac{\partial}{\partial x^\mu} V^\nu - V^\mu \frac{\partial}{\partial x^\mu} W^\nu \right) \frac{\partial}{\partial x^\nu}. \quad (3.41)$$

Now, interestingly we can look at another object that, on the surface, wouldn't seem related to this one. The **Lie bracket** is a map

$$[\ , \] : T_p M \times T_p M \rightarrow T_p M. \quad (3.42)$$

Though it should be said that it can act on any kind of vector space, presently we are dealing only with tangent vectors. Its action is taking two vectors $V, W \in T_p M$ and returning a single one. So, by acting on a function $f \in \mathcal{F}$, we have

$$[V, W][f] = V[W[f]] - W[V[f]]. \quad (3.43)$$

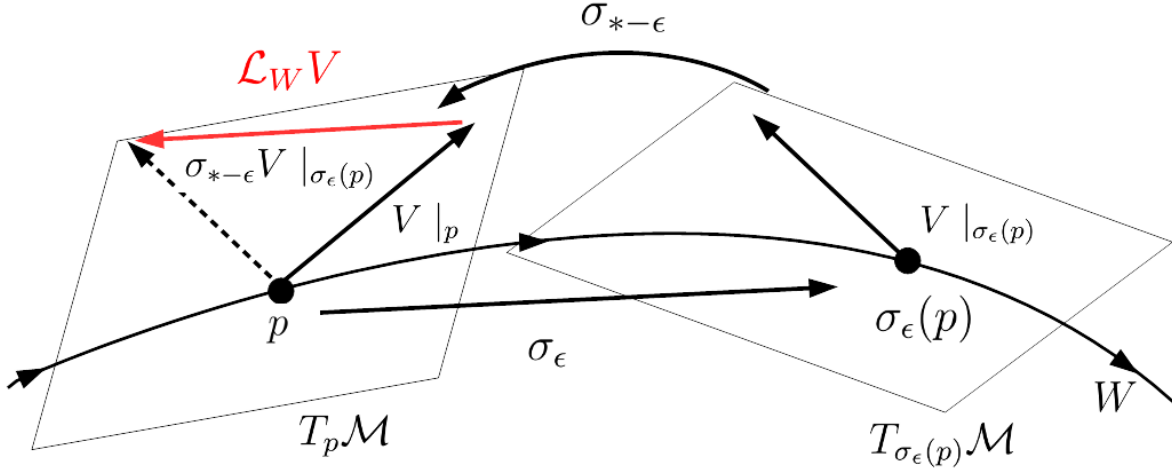


Figure 3.8: The Lie derivative determines the rate of change of a vector field along a flow generated by another vector field

Remembering from before how vectors act on functions: $V[f] = V^\mu \frac{\partial f}{\partial x^\mu}$ and that this is itself a function, this, in a coordinate chart, can be expanded as

$$[V, W][f] = \left(V^\mu \frac{\partial}{\partial x^\mu} W^\nu - W^\mu \frac{\partial}{\partial x^\mu} V^\nu \right) \frac{\partial f}{\partial x^\nu}. \quad (3.44)$$

From here we see that

$$\mathcal{L}_W V = [V, W]. \quad (3.45)$$

What this is telling us, is whether the two flows associated with the vector fields V and W commute with each other. Some properties of the Lie derivative:

- Action on functions: for a function f the Lie derivative is $\mathcal{L}_V f = V[f]$
- Linearity:

$$\mathcal{L}_{fW} V = f \mathcal{L}_V W - W[f]V$$

$$\mathcal{L}_W fV = f \mathcal{L}_V W + V[f]W$$

Meanwhile, for the Lie bracket we have the usual properties

- Skew-Symmetry:

$$[V, W] = -[W, V]$$

- Linearity: for some constant $k_1, k_2 \in \mathbb{R}$

$$[V, k_1 W + k_2 X] = k_1 [V, W] + k_2 [V, X]$$

$$[k_1 V + k_2 X, W] = k_1 [V, W] + k_2 [X, W]$$

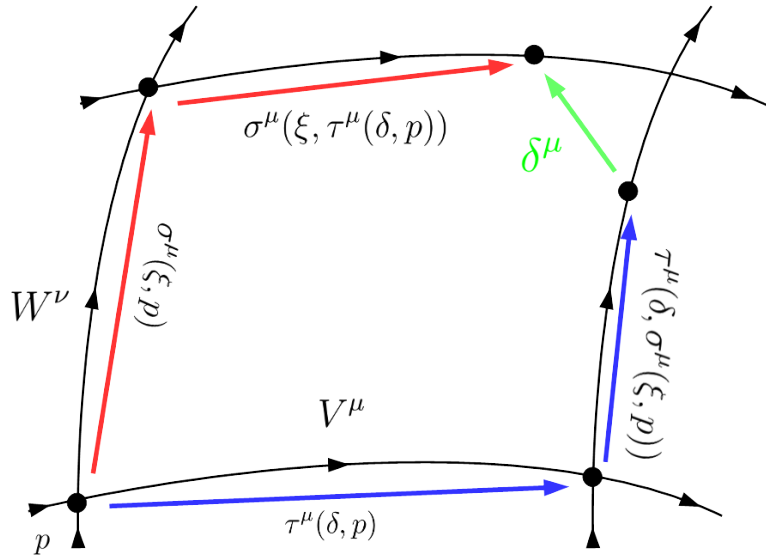


Figure 3.9: The Lie bracket can be thought of measuring the failure of two flows σ and τ - here generated by V^μ and W^ν respectively - to commute: $[V^\mu, W^\nu] = \delta^\mu$

- Jacobi identity

$$[X, [V, W]] + [W, [X, V]] + [V, [W, X]] = 0$$

It ought to be obvious that through (3.45) the properties of either one of these carry over to the other.

3.6 Differential Forms and Tensors

Before moving on to some physical applications of everything we have learned so far, we are going to relate everything we have done so far to the concept of tensors encountered in the introduction. There, we saw them as multilinear maps on a vector space with a composition called the tensor product (more generally and formally a type of monoid). We may also use this structure to construct tensors on our tangent space by identifying $\{e_\mu\} = \{\frac{\partial}{\partial x^\mu}\}$ and the dual $\{\epsilon^\mu\} = \{dx^\mu\}$.

A quick side note here on something that will become important in a later chapter: So far we have denoted the basis of $T_p M$ by using partial derivatives. This is useful when thinking about vector fields and their actions on functions, but it isn't the only choice. We could take some basis $\{\hat{e}_\mu\}$ which is related to $\{\hat{e}_\mu\} = \{A^\nu_\mu \frac{\partial}{\partial x^\nu}\}$ where $A^\nu_\mu \in GL(d)$. That is to say, A is a general linear (square) matrix of dimension d corresponding to the dimension of the basis and is a general coordinate transformation ((3.25) is a particular example). When the basis of a tangent space is referred to without writing it out in

partial derivatives, it is known as a **non-coordinate basis**.

For now, we will be sticking to our normal tangent space basis. The space of tensors of valence (q, r) at a point is denoted by $\mathcal{T}_{r,p}^q M$. A tensor $T \in \mathcal{T}_{r,p}^q M$ can then be expanded as

$$T = T^{\mu_1 \mu_2 \dots \mu_q}{}_{\nu_1 \nu_2 \dots \nu_r} \frac{\partial}{\partial x^{\mu_1}} \otimes \frac{\partial}{\partial x^{\mu_2}} \otimes \dots \otimes \frac{\partial}{\partial x^{\mu_q}} \otimes dx^{\nu_1} \otimes dx^{\nu_2} \otimes \dots \otimes dx^{\nu_r} \quad (3.46)$$

Taking $f : M \rightarrow N$ a diffeomorphism, the pullback naturally extends to $(q, 0)$ tensors as a map $f_* : \mathcal{T}_{0,p}^q \rightarrow \mathcal{T}_{0,f(p)}^q$. While for $(0, r)$ tensors the push forward is $f^* : \mathcal{T}_{r,f(p)}^0 \rightarrow \mathcal{T}_{r,p}^0$. As an example, we take a $(1, 1)$ tensor and function so that in some coordinate chart $f : \{x^\mu\} \rightarrow \{y^\mu\}$, and consider

$$f_* T = f_* \left(T^\mu{}_\nu \frac{\partial}{\partial x^\mu} \otimes dx^\nu \right) = T^\mu{}_\nu f_* \left(\frac{\partial}{\partial x^\mu} \right) \otimes f_*(dx^\nu) = T^\mu{}_\nu \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial x^\nu}{\partial y^\beta} \frac{\partial}{\partial y^\alpha} \otimes dy^\beta \quad (3.47)$$

Now, this leads us into our next topic. We note that the composition \otimes does not say anything about the symmetry properties of the resultant tensor. We know that from physics that tensors can possess symmetric and/or antisymmetric indices, however the tensor product by itself does not allow for permutations of the tensor factors and hence the indices. What we look at then is a particularly important type of antisymmetric $(0, r)$ tensor called a **differential form**. Recall from earlier that we made use of the notation $\Omega^1 M$ for the cotangent space calling into mind the question as to the purpose of ¹. Now we are going to define a totally antisymmetric $(0, r)$ tensor using the **wedge product**, or **exterior product**, which, at the simplest level, is a map

$$\wedge : \Omega^1 M \otimes \Omega^1 M \rightarrow \Omega^2 M \quad (3.48)$$

Where $\Omega^2 M$ is the space of totally antisymmetric **two forms**. For example

$$\wedge (dx^\mu \otimes dx^\nu) = dx^\mu \wedge dx^\nu = \frac{1}{2} (dx^\mu \otimes dx^\nu - dx^\nu \otimes dx^\mu) \quad (3.49)$$

so that it is simply the antisymmetrisation of the tensor product which has the property

$$dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu \quad (3.50)$$

from where we see that $dx^\mu \wedge dx^\mu = 0$. More generally we can have a differential form

$$dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_r} \quad (3.51)$$

which is referred to as an **r -form** and is an element of $\Omega^r M$ the integer r is called the **degree**. Meanwhile the wedge product extends to higher degree forms as

$$\wedge : \Omega^m M \otimes \Omega^n M \rightarrow \Omega^{m+n} M \quad (3.52)$$

From (3.50) , we note the important property

$$dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_r} = 0 \quad (3.53)$$

if any index is repeated. A general r -form $\omega \in \Omega^r M$ is expressed as

$$\omega = \frac{1}{r!} \omega_{\mu_1 \mu_2 \cdots \mu_r} dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_r} \quad (3.54)$$

Now, something interesting to note: If we have a manifold of dimension m , then the highest degree r of any form, will be $r = m$. This is because if we have m basis one-forms corresponding to the number of coordinates, we will not be able to build a form with degree higher than m without repeating any element (according to (3.53)). More generally, due to the antisymmetry of the wedge product, the maximum number of elements in $\Omega^r M$ for a manifold with dimension m is

$$\binom{m}{r} = \frac{m!}{(m-r)!r!} \quad (3.55)$$

Generally, for any forms of arbitrary degree $\omega \in \Omega^r M$ and $\xi \in \Omega^p M$ we have

$$\begin{aligned} \omega \wedge \xi &= -\xi \wedge \omega \\ \omega \wedge \omega &= 0 \end{aligned} \quad (3.56)$$

For completeness, we note that $\Omega^0 M = \mathcal{F}(M)$, the space of functions. Let's look at an example: Take a manifold with $\dim(M) = 4$ and a chart with coordinates $\{x^\mu\}$ where $1 \leq \mu \leq 4$ and one form basis $\{dx^\mu\}$. Now what sort of forms can we build out of this? See 3.10.

The highest degree form is called a **top-form**. Next, we have the **exterior derivative**, a map

$$d : \Omega^r M \rightarrow \Omega^{r+1} M \quad (3.57)$$

$\Omega^0 M$	$\Omega^1 M$	$\Omega^2 M$	$\Omega^3 M$	$\Omega^4 M$
$\mathcal{F}(M)$	dx^1	$dx^1 \wedge dx^2$	$dx^1 \wedge dx^2 \wedge dx^3$	$dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4$
	dx^2	$dx^1 \wedge dx^3$	$dx^1 \wedge dx^2 \wedge dx^4$	
	dx^3	$dx^1 \wedge dx^4$	$dx^1 \wedge dx^3 \wedge dx^4$	
	dx^4	$dx^2 \wedge dx^3$	$dx^2 \wedge dx^3 \wedge dx^4$	
		$dx^2 \wedge dx^4$		
		$dx^3 \wedge dx^4$		

Figure 3.10: All possible combination of forms in 4 dimensions

Explicitly, acting on some r -form ω , this is

$$d\omega = \frac{1}{r!} \left(\frac{\partial}{\partial x^\nu} \omega_{\mu_1 \mu_2 \dots \mu_r} \right) dx^\nu \wedge dx^{\mu_1} \wedge dx^{\mu_2} \wedge \dots \wedge dx^{\mu_r} \quad (3.58)$$

For example, if we take \mathbb{R}^3 with coordinate chart (x, y, z) a 0-form $\omega_0 = \omega_0(x, y, z)$, then clearly

$$d\omega_0 = \frac{\partial \omega_0}{\partial x} dx + \frac{\partial \omega_0}{\partial y} dy + \frac{\partial \omega_0}{\partial z} dz \quad (3.59)$$

By extension, if $\omega_1 = \omega_x(x, y, z)dx + \omega_y(x, y, z)dy + \omega_z(x, y, z)dz$ then

$$d\omega_1 = \left(\frac{\partial \omega_y}{\partial x} - \frac{\partial \omega_x}{\partial y} \right) dx \wedge dy + \left(\frac{\partial \omega_x}{\partial z} - \frac{\partial \omega_z}{\partial x} \right) dz \wedge dx + \left(\frac{\partial \omega_x}{\partial z} - \frac{\partial \omega_z}{\partial x} \right) dy \wedge dz \quad (3.60)$$

and so on until one arrives at the top form $\omega_3 = \omega_3(x, y, z)dx \wedge dy \wedge dz$ where clearly

$$d\omega_3 = 0 \quad (3.61)$$

We also note an important property of the exterior derivative, for any r -form $\omega \in \Omega^r M$

$$d^2\omega = 0 \quad (3.62)$$

it is a nil-potent operator. You may have noticed the section on Lie derivatives defined their action only on a vector basis, but not on the dual one forms. For this extension we need one further map. The **interior contraction** (or **interior product**), is a map

$$i : TM \times \Omega^r M \rightarrow \Omega^{r-1} M \quad (3.63)$$

and denotes the contraction of a vector with an r -form. Note that it is *not* the inverse of the exterior derivative. For $V \in TM$, $W \in TM$, $\omega \in \Omega M$, $\eta \in \Omega^r M$ and $\gamma \in \Omega^p M$ we have

- $i_V \omega = \omega(V) = (\omega_\mu V^\nu) dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \omega_\mu V^\mu$
- $i_V(\eta \wedge \gamma) = (i_V \eta) \wedge \gamma + (-1)^{\deg(\eta)} \eta \wedge (i_V \gamma)$
- $i_V i_W \gamma = -i_W i_V \gamma$
- $i_V i_V \gamma = 0$

Where the second property allows us to extend i to forms of any degree. As an example, let's look again at \mathbb{R}^2 and take $\omega_1 = \omega_x dx + \omega_y dy + \omega_z dz$ with a vector $V_1 = V \frac{\partial}{\partial x}$. The we have the contraction

$$i_{V_1} \omega_1 = i_{V_1}(\omega_x dx) + i_{V_1}(\omega_y dy) + i_{V_1}(\omega_z dz) = V \omega_x \quad (3.64)$$

Meanwhile, taking $\omega_2 = \omega_x dx \wedge dy + \omega_y dy \wedge dz + \omega_z dz \wedge dx$ we can calculate the contraction

$$\begin{aligned} i_{V_1} \omega_2 &= i_{V_1}(\omega_x dx \wedge dy) + i_{V_1}(\omega_y dy \wedge dz) + i_{V_1}(\omega_z dz \wedge dx) \\ &= V \omega_x dy - V \omega_z dz \end{aligned} \quad (3.65)$$

Now we can finally write the Lie derivative for differential forms as

$$\mathcal{L}_V \omega = di_V \omega + i_V d\omega \quad (3.66)$$

An interesting exercise at this point might be to try and derive this expression for one forms based on the derivation of the Lie derivative for vector fields. Our last point is the action of the Lie derivative on tensors (since we now know how it acts on the dual basis). Generally, if we take the tensor product of an arbitrary vector and r -form and act with the Lie derivative we get

$$\mathcal{L}_V(W \otimes \omega) = \mathcal{L}_V(W) \otimes \omega + W \otimes \mathcal{L}_V(\omega) \quad (3.67)$$

which is simply a version of the usual Leibniz rule for differentiation. So now, if we take a $(1, 1)$ tensor $T \in \mathcal{T}^1_1$ the above means that we have

$$\begin{aligned}
\mathcal{L}_V T &= \mathcal{L}_V \left(T^\mu{}_\nu \frac{\partial}{\partial x^\mu} \otimes dx^\nu \right) \\
&= V[T^\mu{}_\nu] \frac{\partial}{\partial x^\mu} \otimes dx^\nu + T^\mu{}_\nu \mathcal{L}_V \left(\frac{\partial}{\partial x^\mu} \right) \otimes dx^\nu + T^\mu{}_\nu \frac{\partial}{\partial x^\mu} \otimes \mathcal{L}_V(dx^\nu)
\end{aligned} \tag{3.68}$$

remembering that we also need to act on the scalar component $T^\mu{}_\nu$.

Chapter 4

Differential Geometry

In Chapter 3 we have studied manifolds in the most abstract way. It is very remarkable how far we can get by studying those geometrical structures without even defining the concept of distance. In this chapter we take a step further and we introduce the concept of metric. This is a fundamental concept in geometry since it enables us to define every single geometrical quantity, including distances, angles and volumes, in terms of it. As a matter of fact, one could argue that the metric is the borderline between topology and geometry, even though this subject is far more subtle than it seems.

The metric also appears everywhere in physics, specially in General Relativity where it plays the role of a fundamental tensor field in nature, interacting with matter accordingly to the Einstein's field equation. We will see more on this in the second part of this course. For the time being, let's see how it beautifully appears in pure geometry.

4.1 Metric

We know from Euclidean geometry that the dot product satisfies the following properties for $u, v \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$:

1. Symmetry: $u \cdot v = v \cdot u$
2. Distributivity: $u \cdot (av + bw) = au \cdot v + bu \cdot w$
3. If $u \cdot v = 0$ for all $v \in \mathbb{R}^n$, then $u = 0$.

It turns out that these properties are the most important properties that characterize such kind of product. In order to generalize this notion of product to smooth manifolds,

we have to keep these properties in mind and take them as axioms. Then we get the following definition.

Definition 4. *Let V be a real vector space of finite dimension. A real inner product on V is a map $V \times V \rightarrow \mathbb{R}$ that assigns a real number $\langle u, v \rangle \in \mathbb{R}$ to every pair of vectors $u, v \in V$ satisfying the following three conditions:*

1. *Symmetry: $\langle u, v \rangle = \langle v, u \rangle$*
2. *Distributivity: $\langle u, av + bw \rangle = a\langle u, v \rangle + b\langle u, w \rangle$*
3. *If $\langle u, v \rangle = 0$ for all $v \in V$, then $u = 0$.*

Now we are able to introduce the object that gives name to this section. A metric tensor (or simply metric) on a smooth manifold M is a $(0, 2)$ -tensor field¹ satisfying the following axioms:

1. Symmetry: $g(X, Y) = g(Y, X)$
2. Positive definiteness: $g(X, X) > 0$ if $X \neq 0$.

This means that the metric assigns an inner product structure to every tangent space of a manifold M , that is, for every $p \in M$ the map $g_p : T_p M \times T_p M \rightarrow \mathbb{R}$ is an inner product. Sometimes we will also use the notation $\langle X, Y \rangle_p \doteq g_p(X, Y)$ to emphasize this last fact.

See that we are assuming the existence of a new object, the metric is not a consequence of the previous theory of smooth manifolds. That is the why we say it is an additional structure and we call the pair (M, g) , formed by a smooth manifold M together with a metric g , by Riemannian manifold.

We may go even further relaxing the positive-definite condition of the metric and instead assuming that it is just non-degenerate $g(X, Y) \neq 0$. A manifold together with such a metric is called pseudo-Riemannian manifold or semi-Riemannian manifold.

Example 4.1.1.

¹A tensor field on a manifold M is a bit more general than a tensor (not a field) in the following sense. Let $\mathcal{F}(M)$ be the set of all smooth real-valued functions on M . Then a (r, s) -tensor field is defined to be \mathcal{F} -linear (instead of \mathbb{R} -linear as is the case for tensors) in each of its arguments. For example, a $(0, 2)$ -tensor field T satisfies

$$T_p(fX + gY) = f(p)T_p(X) + g(p)T_p(Y), \quad p \in M \quad \text{and} \quad f, g \in \mathcal{F}(M).$$

1. The real space \mathbb{R}^n together with the dot product is a Riemannian manifold called Euclidian space and denoted E^n
2. Spacetime together with the metric given as a solution of Einstein's equations is a pseudo-Riemannian manifold
3. Any surface immersed in E^n is a Riemannian manifold
4. The configuration space of a Lagrangian mechanical system is a Riemannian manifold

■

In any coordinate system $(U; x^i)$ we can write

$$g = g_{ij} dx^i \otimes dx^j, \quad (4.1)$$

where $g_{ij} = g(\partial_i, \partial_j)$ are the coefficients of the metric tensor. In the physics literature g_{ij} is what is called metric tensor, but keep in mind that this is an abuse of nomenclature. If we introduce the symmetric product of two 1-forms ω and η , denoted by juxtaposition, the notation used in Equation 4.1 can be shortened:

$$\omega\eta = \frac{1}{2}(\omega \otimes \eta + \eta \otimes \omega).$$

Due to the symmetry of g_{ij} , Equation 4.1 can be written in the following way:

$$g = g_{ij} dx^i dx^j. \quad (4.2)$$

As it was stressed before, all of the geometrical concepts can be written in terms of the metric. The canonical way of defining distance in Euclidean space, for example, is taking the square root of the inner product. Since the metric assigns an inner product to each point of a manifold, one would conclude that

$$\|\cdot\| = \sqrt{g(\cdot, \cdot)}.$$

Moreover, the length of a curve α on a manifold is defined to be

$$s = \int \|\alpha'(t)\| dt = \int \sqrt{g(\alpha'(t), \alpha'(t))} dt = \int \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt.$$

Then, from the fundamental theorem of calculus, we get the formal expression (in the poor sense²) usually taken by physicists as the “definition” of the metric

$$ds^2 = g_{ij}dx^i dx^j. \quad (4.3)$$

It is important to note that Equation 4.3 is very formal and objects like $dx^i dx^j$ ³ and ds^2 have no mathematical meaning here. But fortunately, the notation used in differential and integral calculus is very good and physicists do not need to know about differential forms to proceed with their formal calculations.

Another example is the angle between two tangent vectors X and Y on a manifold is given by

$$\cos \theta = \frac{g(X, Y)}{\|X\| \|Y\|} = \frac{g(X, Y)}{\sqrt{g(X, X)g(Y, Y)}}$$

and as we can see it is written solely in terms of the vectors and of the metric.

We will close this section by showing that the metric establishes an isomorphism between vectors and covectors. Let $u = u^i \partial_i$ be a vector on a manifold M equipped with a metric g . We can define a covector T_u by

$$T_u(v) = \sum_i g \otimes u(\partial_i, v, dx^i).$$

Exercise 4.1.1. Convince yourself that T_u is indeed a covector.

Exercise 4.1.2. Show that $T_u(v) = g(u, v)$.

Its coordinates are given by

$$u_i \equiv T_u(\partial_i) = g_{ij}u^j. \quad (4.4)$$

This process is called lowering the index. Conversely, if $\omega = \omega_i dx^i$ is a covector, then the vector T_ω defined by

$$T_\omega(\rho) = \sum_i g^{-1} \otimes \omega(dx^i, \rho, \partial_i),$$

where $g^{-1} = g^{ij} \partial_i \otimes \partial_j$ is the inverse of the metric g , has components

$$\omega^i = g^{ij} \omega_j. \quad (4.5)$$

²When one speaks about formality in mathematics, he means the management of objects concerning their form and not the mathematical concept behind them. In physics formality means exactly the opposite.

³Observe that this is not the symmetric product defined above. Instead, this is the result of “canceling” dt 's from both sides of the equation.

This process is called raising the index. This shows that for every vector there is a corresponding covector and vice versa. Componentwise, they are simply related through the metric as in Equations (4.4) and (4.5).

Exercise 4.1.3. Show that raising and lowering index in succession produces no effect. This is the reason we can keep the same kernel letter u for the vector components u^i and the covector components u_i .

4.2 Connections

A curve in Euclidean space is straight if and only if its acceleration vanishes identically. As we know, this is a very important feature of Euclidean geometry and one of the most fundamental concepts in mechanics as well. Hence, it would be great if we could keep this idea on general smooth manifolds and then look for the “straight lines” over them. It turns out it is completely possible to do so if we generalize the notion of acceleration and, in turn, the concept of derivative of vectors. The latter can be easily generalized since we always expect that any notion of first derivative should satisfy the linear and Leibnitz rules. As such, we are going to introduce the covariant derivative as a primitive concept satisfying these properties. If $\mathcal{T}(M)$ is the set of all smooth vector fields on a manifold M , then an affine connection (or simply connection) ∇ is a map

$$\nabla : \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{T}(M) \quad (4.6)$$

$$(X, Y) \quad \mapsto \nabla_X Y \quad (4.7)$$

which satisfies the axioms

$$\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z, \quad (4.8a)$$

$$\nabla_{(X+Y)}Z = \nabla_X Z + \nabla_Y Z, \quad (4.8b)$$

$$\nabla_{(fX)}Y = f\nabla_X Y, \quad (4.8c)$$

$$\nabla_X(fY) = X[f]Y + f\nabla_X Y, \quad (4.8d)$$

where $f : M \rightarrow \mathbb{R}$ is any differentiable function, $X, Y, Z \in \mathcal{T}(M)$ and $X[f]$ represents the directional derivative of f in the direction of X at p , defined by

$$X[f] = \left. \frac{d}{dt} f(p + tX) \right|_{t=0} = \frac{\partial f(p)}{\partial x^i} X^i. \quad (4.9)$$

The last equality of 4.9 was resulted through the application of the chain rule.

The connection produces a derivative $\nabla_X Y$ with the above properties, called covariant derivative. Intuitively, the value of $\nabla_X Y$ at a point $p \in M$ is the rate of change of Y in the direction of the unit vector $X(p)$.

The covariant derivative of a vector field Y along a curve $\gamma : \mathbb{R} \rightarrow M$ is defined to be

$$\frac{DY}{dt} = D_{\dot{\gamma}}Y,$$

where $\dot{\gamma}$ is the tangent vector to the curve. On a manifold there are infinite many different ways to define a connection that could satisfy Equations (4.8). Despite this, we can restrict this range of options if we demand that such object obeys some other properties. In this sense, a Levi-Civita connection is a connection such that, for any $X, Y, Z \in \mathcal{T}(M)$:

$$\nabla_X [g(Y, Z)] = g(\nabla_X Y, Z) + g(Y, \nabla_X Z), \quad (4.10)$$

$$[X, Y] = \nabla_X Y - \nabla_Y X, \quad (4.11)$$

where $[X, Y]$ is the Lie bracket of the vector fields X and Y .

The property (4.10), known as compatibility with the metric, says the metric is preserved along any curve on M and, hence, angles and volumes are also kept constant. On the other hand, the relation (4.11) implies the symmetry of the connection and of the Christoffel symbols, as we will see later on, and is associated to torsionless connections. Unless stated otherwise, we are going to consider only Levi-Civita connections.

In what follows we will illustrate some of the concepts introduced above with concrete examples. The simplest case is the one of vector fields in the Euclidean space \mathbb{R}^3 . In this case the covariant derivative

$$\bar{\nabla} : \mathcal{T}(E^3) \times \mathcal{T}(E^3) \rightarrow \mathcal{T}(E^3)$$

of $W = W^i \partial_i$ with respect to the vector X at p is defined by

$$\bar{\nabla}_X W = \frac{d}{dt} W(p + tX)(0), \quad (4.12)$$

$$= \frac{d}{dt} W^i(p + tX)(0) \partial_i, \quad (4.13)$$

$$= X[W^i] \partial_i, \quad (4.14)$$

i.e., the covariant derivative is the directional derivative of the vector field W in the direction of v , which is the derivative of the Equation (4.9) calculated for each component of W . To show that (4.13) is in fact a covariant derivative, we just need to verify it

satisfies the properties (4.8). To do this we just have to verify them for each component of the Equation (4.14). Therefore, from the linearity of the usual derivative, we get (4.8a):

$$\begin{aligned}\bar{\nabla}_X(Y^i + Z^i) &= X[Y^i + Z^i] = \left. \frac{d}{dt}(Y^i + Z^i)(p + tX) \right|_{t=0} \\ &= \left. \frac{d}{dt} [Y^i(p + tX) + Z^i(p + tX)] \right|_{t=0} \\ &= X[Y^i] + X[Z^i] = \bar{\nabla}_X Y^i + \bar{\nabla}_X Z^i.\end{aligned}$$

Using the chain rule we get (4.8b):

$$\begin{aligned}\bar{\nabla}_{X+Y} Z^i &= \left. \frac{d}{dt} Z^i(p + t(X + Y)) \right|_{t=0} \\ &= \frac{\partial Z^i}{\partial x^j} (X^j + Y^j) \\ &= \frac{\partial Z^i}{\partial x^j} X^j + \frac{\partial Z^i}{\partial x^j} Y^j \\ &= X[Z^i] + Y[Z^i] = \bar{\nabla}_X Z^i + \bar{\nabla}_Y Z^i.\end{aligned}$$

The relation (4.8c) is obtained through the direct application of the chain rule and of the Leibnitz rule:

$$\begin{aligned}\bar{\nabla}_{fX} Y^i &= (fX)[Y^i] = \left. \frac{d}{dt} Y^i(p + tfX) \right|_{t=0} \\ &= \left. \frac{\partial Y^i}{\partial x^j} \frac{d}{dt} (p^j + tfX^j) \right|_{t=0} \\ &= \left. \frac{\partial Y^i}{\partial x^j} \left[f + t \left(\frac{\partial f}{\partial x^k} \dot{x}^k \right) \right] \right|_{t=0} X^j \\ &= \frac{\partial Y^i}{\partial x^j} f X^j \\ &= fX[Y^i] = f\bar{\nabla}_X Y^i.\end{aligned}$$

Lastly, property (4.8d) is resulted from the application of the Leibnitz rule:

$$\begin{aligned}\bar{\nabla}_X fY^i &= X[fY^i] = \left. \frac{d}{dt} (fY^i)(p + tX) \right|_{t=0} \\ &= \left. \frac{d}{dt} f(p + tX) \right|_{t=0} Y^i + f \left. \frac{d}{dt} Y^i(p + tX) \right|_{t=0} \\ &= X[f]Y^i + fX[Y^i] = X[f]Y^i + f\bar{\nabla}_X Y^i.\end{aligned}$$

Exercise 4.2.1. The covariant derivative of vector fields on a surface S immersed in \mathbb{R}^3 is defined by the projection of (4.13) on the surface S :

$$\nabla : \mathcal{T}(S) \times \mathcal{T}(S) \rightarrow \mathcal{T}(S) \quad (4.15)$$

$$(X, Y) \mapsto \nabla_X Y = \bar{\nabla}_X Y - \langle \bar{\nabla}_X Y, N \rangle N, \quad (4.16)$$

where

$$N = \frac{\partial_u \times \partial_v}{\|\partial_u \times \partial_v\|}$$

is the normal vector field to the surface, also known as Gauss map, and $\bar{\nabla}_X Y$ represents the Euclidean covariant derivative given by Equation (4.14). Prove that Equation (4.16) is indeed a covariant derivative.

Geometrically, the definition (4.16) measures the rate of variation of a vector field Y , defined over a surface S , in the direction of the unit vector X , also defined over S . If we imagine $\gamma(t)$ as being the curve which describes the position of a particle constraint on a surface S , then the covariant derivative $\nabla_{\gamma'(t)} \gamma'(t)$ has the meaning of the particle's acceleration vector over this surface.

It is interesting to note that so far the covariant derivative has been used abstractly, that is, without the need of introducing a reference frame. However, it is useful to express it as coordinates of such a reference system to make calculations easier. So let $\{\partial_i(p)\}$ be a basis for $T_p M$. The covariant derivative of a vector from this basis with respect to another vector from the same basis is again another vector of this basis. Therefore, we can write

$$\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k \partial_k, \quad (4.17)$$

where the coefficients Γ_{ij}^k are called Christoffel symbols. Moreover, from Equation (4.11) we find that:

$$\begin{aligned} [\partial_i, \partial_j] &= \nabla_{\partial_i} \partial_j - \nabla_{\partial_j} \partial_i \\ &= (\Gamma_{ij}^k - \Gamma_{ji}^k) \partial_k. \end{aligned}$$

Since $\partial_i \partial_j f = \partial_j \partial_i f$, it follows that:

$$\Gamma_{ij}^k = \Gamma_{ji}^k, \quad (4.18)$$

i.e., the Christoffel symbols are symmetric in the lower indices.

Hence, in order to specify the covariant derivative of arbitrary vector fields is enough to specify it for each basis vector ∂_j with respect to ∂_i . Then, using Equation (4.17) together with the properties (4.8), we find that for generic vectors $X = X^i \partial_i$ and $Y = Y^j \partial_j$ that:

$$\nabla_X Y = \nabla_{X^i \partial_i} (Y^j \partial_j) \quad (4.19)$$

$$= X^i \nabla (Y^j \partial_j) \quad (4.20)$$

$$= X^i Y^j \nabla_{\partial_i} \partial_j + X^i \partial_j \nabla_{\partial_i} Y^j \quad (4.21)$$

$$= X^i Y^j \Gamma_{ij}^k \partial_k + X^i \frac{\partial Y^j}{\partial x^i} \partial_j. \quad (4.22)$$

Since the index j in the last term is dummy, we can change it for k and we get

$$\nabla_X Y = \left(\frac{\partial Y^k}{\partial x^i} + Y^j \Gamma_{ij}^k \right) X^i \partial_k. \quad (4.23)$$

If we repeat the above argument for vector fields Y defined along curves γ , we would get

$$\frac{DY}{dt} = \left(\frac{dY^k}{dt} + Y^j \Gamma_{ij}^k \frac{dx^i}{dt} \right) \partial_k. \quad (4.24)$$

Exercise 4.2.2. Prove Equation (4.24).

It is possible to find the Christoffel coefficients Γ_{ij}^k only in terms of the metric and its derivatives from the metric compatibility condition (see Equation (4.10)):

$$0 = \frac{\partial g_{ik}}{\partial x^l} - g_{mk} \Gamma_{kl}^m - g_{im} \Gamma_{kl}^m. \quad (4.25)$$

Using that the Christoffel symbols are symmetric in the lower indices, Equation (4.25) can be solved explicitly for Γ_{ij}^k :

$$\Gamma_{ij}^k = \frac{g^{kl}}{2} \left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l} \right). \quad (4.26)$$

Be aware that the Equation (4.26) only holds for the Levi-Civita connection. If torsion is included, as happens in Einstein-Cartan gravity for example, the Christoffel symbols cannot be expressed in terms of the metric and its derivatives. In fact, connections are, in general, completely independent of the metric since they form different mathematical structures. Fortunately, General Relativity has the Levi-Civita connection and Equation (4.26) is all we need to specify it.

4.3 Geodesics

A vector field Y is said to be parallel along a curve γ if its covariant derivative along γ vanishes for all t in the curve's parameter range, that is,

$$\frac{DY}{dt} = 0, \quad t \in I \subset \mathbb{R}. \quad (4.27)$$

A curve whose tangent vector is parallel along itself is called a geodesic:

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0, \quad t \in I. \quad (4.28)$$

Therefore, the geodesics are precisely the “straight lines” on manifolds that we were looking for, simply because their accelerations vanish.

According to Equation (4.24), in a coordinate system $(U; x^i)$ Equation (5.1) becomes

$$\frac{d^2 x^k}{dt^2} + \Gamma_{ij}^k \frac{dx^j}{dt} \frac{dx^i}{dt} = 0. \quad (4.29)$$

This is the equation which describes the geodesics in a neighborhood U of a manifold. It must be observed that Equation (4.29) was determined without specifying any specific connection, hence whether Christoffel symbols in this equation can be determined through the metric, as in Equation (4.26), or not depends upon the chosen connection.

The parameter t used to parametrize γ such that it satisfies Equation (4.29) is called affine parameter. Under a reparametrization of γ , say $t' = f(t)$, the tangent vector becomes

$$\dot{\gamma}' = \frac{dx^i}{dt'} \partial_i = \frac{1}{f'(t)} \frac{dx^i}{dt} \partial_i,$$

where $f'(t) = df/dt$ and, using Equation (4.29), we have

$$\begin{aligned} \frac{d^2 x^k}{dt'^2} + \Gamma_{ij}^k \frac{dx^j}{dt'} \frac{dx^i}{dt'} &= \frac{1}{f'(t)} \frac{d}{dt} \left(\frac{1}{f'(t)} \right) \frac{dx^i}{dt} \\ &= -\frac{f''(t)}{f'(t)^2} \frac{dx^i}{dt}. \end{aligned}$$

Hence the new parameter t' is an affine parameter if, and only if, $f''(t) = 0$ or $t' = at + b$ for a, b real constants. This shows that the fact of being a geodesic depends on the parametrization. Curves that can be transformed into geodesics through a reparametrization are called pre-geodesics.

4.4 Torsion and Curvature

In the transformation law of Γ_{jk}^i , the term involving second derivatives is symmetric in the indices jk . It follows that the antisymmetrized quantity

$$T_{jk}^i = \Gamma_{jk}^i - \Gamma_{kj}^i \quad (4.30)$$

does transform as a tensor, due to the cancellation of the non-tensorial parts. To express this idea in a non-coordinate way, we define the torsion map $\tau : \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{T}(M)$ by

$$\tau(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

Exercise 4.4.1. Show that τ is antisymmetric, i.e., $\tau(X, Y) = -\tau(Y, X)$ and \mathcal{F} -linear (see the footnote in the beginning of this chapter).

Hence τ gives rise to a tensor field T of type $(1, 2)$ defined as

$$T(\omega, X, Y) = \langle \omega, \tau(X, Y) \rangle.$$

It is called torsion tensor of the connection ∇ .

Exercise 4.4.2. By applying T to basis vectors, prove that its coefficients in a local coordinate chart $(U; x^i)$ is given by Equation (4.30).

We call a connection torsion-free or symmetric if its torsion tensor vanishes identically, $T = 0$. In this case, Equation (4.30) shows that its components are symmetric with respect to any coordinates.

A similar problem occurs when commuting repeated covariant derivatives on a vector or tensor field. The map $P : \mathcal{T}^*(M) \times \mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{F}(M)$ defined by

$$P(\omega, X, Y, Z) = \langle \omega, \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z \rangle$$

fails to be a tensor of type $(1, 3)$ since it is not \mathcal{F} -linear in the three vector field arguments. To remedy this we proceed in the same way as before for creating the torsion tensor.

For any vector fields X, Y define the operator $\rho_{X,Y} : \mathcal{T}(M) \rightarrow \mathcal{T}(M)$ by

$$\rho_{X,Y} Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

Exercise 4.4.3. Prove that $\rho_{X,Y} Z$ is \mathcal{F} -linear in all of its components.

Therefore, we can define a tensor field of type (1, 3) by setting

$$R(\omega, X, Y, Z) = \langle \omega, \rho_{X,Y}Z \rangle,$$

known as Riemann curvature tensor of the connection ∇ .

In a local coordinate chart $(U; x^i)$, the components of the curvature tensor is given by

$$\begin{aligned} R_{jkl}^i &= R(dx^i, \partial_j, \partial_k, \partial_l) \\ &= \langle dx^i, \nabla_k \nabla_l \partial_j - \nabla_l \nabla_k \partial_j - \nabla_{[\partial_k, \partial_l]} \partial_j \rangle \\ &= \langle dx^i, \nabla_k (\Gamma_{jl}^m \partial_m) - \nabla_l (\Gamma_{jk}^m \partial_m) \rangle \\ &= \langle dx^i, \Gamma_{jl,k}^m \partial_m + \Gamma_{jl}^m \Gamma_{mk}^p \partial_p - \Gamma_{jk,l}^m \partial_m - \Gamma_{jk}^m \Gamma_{ml}^p \partial_p \rangle, \end{aligned}$$

where $\Gamma_{jl,k}^m = \frac{\partial \Gamma_{jl}^m}{\partial x^k}$. Therefore

$$R_{jkl}^i = \Gamma_{jl,k}^i + \Gamma_{jl}^m \Gamma_{mk}^i - \Gamma_{jk,l}^i - \Gamma_{jk}^m \Gamma_{ml}^i. \quad (4.31)$$

The Ricci (curvature) tensor is then defined by

$$Ric(X, Z) = \sum_i R(dx^i, X, \partial_i, Z),$$

whose coordinates are

$$R_{ij} \equiv Ric(\partial_i, \partial_j) = R_{ikj}^k = g^{lk} R_{likj}. \quad (4.32)$$

This process in which two indices are set equal to each other and summed over is called contraction. Contracting again we get the Ricci scalar

$$R = R_i^i = g^{ij} R_{ij}. \quad (4.33)$$

If we are using the Levi-Civita connection, then the Riemann tensor, the Ricci tensor and the Ricci scalar can be calculated directly from the metric. That is why the metric is chosen to be the fundamental field in General Relativity; it carries all of the geometrical and physical information we need.

Part II

Applications

Chapter 5

General Relativity

Now we have seen the basics of differential geometry, it is about time we moved on to some applications. One may wonder as to the purpose of the previous sections (in particular why they seem so long). The answer to this, is that we are trying to obtain new intuition for something we are already quite familiar with. For this reason, it is important to spend some time familiarizing oneself with the basics, in this case the notions of topological spaces, manifolds etc. The actual application should then be almost only a detail. What we are trying to do, is not to learn new physics, but rather to think about the physics we already know in a different way. The following sections are not intended as - could never be - a re-derivation of all of physics based on differential geometry, but rather a list of specific examples. Hopefully familiar ones, they are intended to show how we can take the contents of the previous section and apply to things we already know and hopefully, gain some insight by doing so. At the end of the day, a formalism (like tensor calculus or path integrals) is only a tool and it is important to have the knowledge to choose the right tool for the job. In this section we will be looking specifically at General Relativity (GR). What follows assumes that the reader is already familiar with GR at the level of an undergraduate course. We are going to look some (hopefully) familiar examples and then cast them in different ways based on the previous chapters.

5.1 Postulates

The theory of GR is just a match of geometrical concepts with physical ones. Precisely, this is done by setting up the postulates of GR as follows. The universe is a four-dimensional Pseudo-Riemannian manifold (M, g) of index $+2$ (i.e., with a metric whose signature has 3 plus signs corresponding to spatial coordinates and 1 minus sign

corresponding to time coordinate; the index is defined to be the sum of these signs) called spacetime. Its points are called events. Test particles “free fall” along spacetime, since they are subjected to no external forces, following its geodesics:

$$\frac{d^2x^\mu}{ds^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{ds} \frac{dx^\rho}{ds} = 0. \quad (5.1)$$

The geodesics followed by massive particles are assumed to be time-like, whereas massless particles, e.g. photons, move along null-like geodesics. Since the geodesic equation (5.1) does not depend on the mass of the particle, all particles have the same geodesic. This is the equivalence principle: all particles undergo the same acceleration in a gravitational field.

The Einstein field equations are

$$Ric - \frac{1}{2}Rg = 8\pi T, \quad (5.2)$$

or in a coordinate chart

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu},$$

where $T = T_{\mu\nu}dx^\mu \otimes dx^\nu$ is the stress-energy tensor of the matter fields and we are using units such that Newton constant and the speed of light are $G = c = 1$. Keep in mind that Equation (5.2) cannot be proved, it was first obtained by brute force (just like Newton’s second law and Schrödinger equation) in an attempt of getting a relation between curvature and energy. That being said, we can adopt a variational approach in which the field equations (5.2) can be deduced. This is accomplished by using the Hilbert-Einstein action

$$S = \int \sqrt{-\det(g)} R. \quad (5.3)$$

Varying this action with respect to the metric field and setting the result to zero leads us to (5.2). However, by adopting the above action we are just shifting our ignorance because now what cannot be proved is Equation (5.3).

5.2 Tetrad formalism

To begin with, we are going to take a closer look at something that was already touched on in previous sections, a **non-coordinate basis** (also called a **tetrad**). In the section on manifolds, we formulated all our quantities in terms of the tangent space basis $\{\partial_\mu\}$. It has the drawback of explicitly depending on some choice coordinates x^μ and as you

may recall, no one coordinate systems is unique and there is generally no canonical choice (that is, there is no preferred system of coordinates). An example of where this becomes an issue, is the Schwarzschild spacetime which, if you recall, describes black holes. Since it is by definition spherically symmetric, the easiest choice of coordinate chart is spherical coordinates. But, as you hopefully recall, this leaves us with an apparent singularity at the black holes event horizon which can be dealt with by a change in coordinates (e.g. Eddington-Finkelstein or Kruskal-Szekeres). What we see however, is a coordinate system which makes calculations easy but suffers from a significant pathology while trying to eliminate it, can make the analysis all the more difficult. Mathematically speaking, neither of these scenarios is preferred and both are equally valid in the regions their coverings overlap. However, we still need some notion of coordinates if we are to do physics.

We are going to avail ourselves of a tetrad basis $\{e_\mu\}$ which is related to our tangent space basis by some coordinate transformation

$$e_\alpha = \frac{\partial y^\nu}{\partial x^\alpha} \frac{\partial}{\partial y^\nu} \quad (5.4)$$

If you recall, any coordinate bases can be related to any other by a transformation of this kind. Now however, we are going to denote the transformation matrix by

$$e^\nu{}_\alpha = \frac{\partial y^\nu}{\partial x^\alpha} \quad (5.5)$$

and take a look at some of its properties. Firstly, and most obviously, we can recast (5.4) as

$$e_\alpha = e^\nu{}_\alpha \frac{\partial}{\partial x^\nu} \quad (5.6)$$

While, conversely for the dual basis $\{dx^\mu\}$ on T^*M , we have

$$e^\alpha = e^\alpha{}_\nu dx^\nu \quad (5.7)$$

We then require that the basis $\{e_\alpha\}$ be **orthonormal**, that is, the inner product satisfies

$$\langle e_\alpha, e_\beta \rangle = \delta_{\alpha\beta} \quad (5.8)$$

This means we have a metric in this basis given by

$$g = g_{\mu\nu} dx^\mu \otimes dx^\nu = g_{\mu\nu} e_\alpha{}^\nu e_\beta{}^\mu e^\alpha \otimes e^\beta = \eta_{\mu\nu} e^\alpha \otimes e^\beta = \eta \quad (5.9)$$

where η is the flat space Minkowski metric

$$\eta = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5.10)$$

This is nothing more than a coordinate transformation into a frame where the metric is flat. The matrices $\{e_\alpha{}^\nu\}$ are called **vierbein**¹ and correspond to whichever coordinate transformation maps one to this frame. But hold on, what about curvature, surely you can't just do a coordinate transformation and get rid of it? Indeed one can't. Usually, our tangent space basis $\{\partial_\mu\}$ is rather trivial, so that one has, for example, $[\partial_\mu, \partial_\nu]f = 0$ for any function f . This is obvious since partial derivatives commute with each other. However, the basis $\{e_\alpha\}$ isn't as trivial, since it will generally contain some functional dependence and then $[e_\mu, e_\nu]f \neq 0$ as for the general Lie bracket of vector fields. What this special coordinate transformation amounts to is taking all the parametric dependence out of the metric and putting it into the basis vectors in such a way that they are orthonormal. The most astounding thing about this, is that it is *always* possible to find such a transformation. Also, it is worth noting that a basis given by (5.4) will have a Lie bracket of the form

$$[e_\mu, e_\nu] \big|_p = c_{\mu\nu}{}^\alpha(p) e_\alpha \big|_p \quad (5.11)$$

Further, we note that $e^\alpha{}_\mu e^\nu{}_\alpha = \delta_\mu{}^\nu$ and $e^\alpha{}_\mu e^\mu{}_\beta = \delta^\alpha{}_\beta$. It is important to distinguish between the indices on this matrix when raising and lowering, since this is done in different frames with different metrics. For example, the transformation in (5.5) one has $e_{\mu\alpha} = g_{\mu\nu} e^\nu{}_\alpha$ and $e^{\nu\beta} = \eta^{\alpha\beta} e^\nu{}_\alpha$. Also note $e^\nu{}_\alpha e_{\nu\beta} = \eta_{\alpha\beta}$ and $e^\nu{}_\alpha e^{\mu\alpha} = g^{\mu\nu}$.

Let's have look at some examples. First consider the standard spherical metric on S^2

$$ds^2 = d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi \quad (5.12)$$

This can easily be converted into the form (5.9) if we take a basis

$$e^1 = d\theta \quad e^2 = \sin(\theta) d\phi \quad (5.13)$$

¹Vierbein is German and translates as 'four-leg' which is specific to four dimensions. In arbitrary dimensions the term **vielbein** is used (many-leg). In a specific number of dimensions one can also have, for example **zweibein**, **dreibein** etc. (receptively two-leg and three-leg)

The tetrad is then read of as

$$e^\alpha{}_\mu = \begin{pmatrix} 1 & 0 \\ 0 & \sin(\theta) \end{pmatrix} \quad (5.14)$$

Now, consider the Schwarzschild metric in spherical coordinates:

$$ds^2 = - \left(1 - \frac{2GM}{r}\right) dt \otimes dt + \left(1 - \frac{2GM}{r}\right)^{-1} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2(\theta) d\phi \otimes d\phi \quad (5.15)$$

Here we see similarly

$$\begin{aligned} e^0 &= \left(1 - \frac{2GM}{r}\right)^{1/2} dt & e^1 &= \left(1 - \frac{2GM}{r}\right)^{-1/2} dr \\ e^2 &= r d\theta & e^3 &= r \sin(\theta) d\phi \end{aligned} \quad (5.16)$$

with

$$e^\alpha{}_\mu = \begin{pmatrix} \left(1 - \frac{2GM}{r}\right)^{1/2} & 0 & 0 & 0 \\ 0 & \left(1 - \frac{2GM}{r}\right)^{-1/2} & 0 & 0 \\ 0 & 0 & r & 0 \\ 0 & 0 & 0 & r \sin(\theta) \end{pmatrix} \quad (5.17)$$

A more interesting example is perhaps a non-diagonal metric (in two dimensions for simplicity)

$$ds^2 = du \otimes dv + dv \otimes du \quad (5.18)$$

Here we can take

$$e^1 = \frac{1}{\sqrt{2}}(du - dv) \quad e^2 = \frac{1}{\sqrt{2}}(du + dv) \quad (5.19)$$

which you can verify:

$$ds^2 = -e^1 \otimes e^1 + e^2 \otimes e^2 \quad (5.20)$$

Does it look familiar? This corresponds to

$$e^\alpha{}_\mu = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (5.21)$$

What all of this hopefully makes clear is that we can always choose to work in a non-coordinate basis. We also note that this basis can be composed the same way as the coordinate basis using the wedge product i.e. $e^\mu \wedge e^\nu$. The top form in this basis is special and referred to as a **volume form**

$$\Omega_{\mathcal{M}} = e^1 \wedge e^2 \wedge \cdots \wedge e^m = \sqrt{|g|} dx^1 \wedge dx^2 \wedge \cdots \wedge dx^m \quad (5.22)$$

where g is the determinant of the metric. This is nothing more than the volume element from GR (think of integral measures $\int \sqrt{|g|} d^m x$). Lastly, we mention for convenience sake, a coordinate free version of raising and lowering indices. The metric can be viewed as an isomorphism $g : T_p M \rightarrow T_p^* M$ and $g^{-1} : T_p^* M \rightarrow T_p M$. Explicitly, for a vector $V \in T_p M$ we will denote its **metric dual** one-form as $\tilde{V} \in T_p^* M$ where these are related by

$$\tilde{V} = g(V, -) = g(V^\alpha e_\alpha, e_\beta) e^\beta = \eta_{\alpha\beta} V^\alpha e^\beta = \tilde{V}_\beta e^\beta \quad (5.23)$$

For a one form $\omega \in T_p^* M$ its corresponding metric dual is

$$\tilde{\omega} = g^{-1}(\omega, -) = g^{-1}(\omega_\alpha e^\alpha, e^\beta) e_\beta = \eta^{\alpha\beta} \omega_\alpha e_\beta = \tilde{\omega}^\beta e_\beta \quad (5.24)$$

It should be evident that we have $\tilde{\tilde{\omega}} = \omega$ and $\tilde{\tilde{V}} = V$. A note on conventions, in the literature you might find this method a raising and lowering indices referred to as a *musical isomorphism*. So, lowering the index of a vector is, independently of any basis, denoted by V^\flat and likewise raising the index of a one form by ω^\sharp . If you are familiar with musical notation, you might see the logic behind this, although we will not be using this particular notation here.

5.3 Hodge Map

I we are to examine an alternative formulation of GR, we are going to need one more, rather useful map. Let's take another look at the Table 5.1, but now repeat the same exercise in a non coordinate basis of dimension four.

$\Omega^0 M$	$\Omega^1 M$	$\Omega^2 M$	$\Omega^3 M$	$\Omega^4 M$
1	e^1	$e^1 \wedge e^2$	$e^1 \wedge e^2 \wedge e^3$	$e^1 \wedge e^2 \wedge e^3 \wedge e^4$
	e^2	$e^1 \wedge e^3$	$e^1 \wedge e^2 \wedge e^4$	
	e^3	$e^1 \wedge e^4$	$e^1 \wedge e^3 \wedge e^4$	
	e^4	$e^2 \wedge e^3$	$e^2 \wedge e^3 \wedge e^4$	
		$e^2 \wedge e^4$		
		$e^3 \wedge e^4$		

Table 5.1: All possible combination of basis forms in 4 dimensions, but with a non-coordinate basis.

Looking at this table, we notice that opposite columns have the same number of elements. So, as vector spaces, $\dim(\Omega^0 M) = \dim(\Omega^4 M)$ and $\dim(\Omega^1 M) = \dim(\Omega^3 M)$. This is indicative of a duality that exists between the vector spaces: If we have an m dimensional manifold with exterior algebra $\Omega^n M$, the spaces of $\dim(n)$ and $\dim(m - n)$ are related by **Hodge duality**. As a map, we have the **Hodge star** so that

$$\star : \Omega^n M \longrightarrow \Omega^{n-m} M \quad (5.25)$$

For $\omega \in \Omega^r M$, $V \in TM$ and $f \in \mathcal{F}(M)$ it satisfies the identities

$$\begin{aligned} \star(f\omega) &= f \star \omega \\ \star(\omega \wedge \tilde{V}) &= i_V \star \omega \end{aligned} \quad (5.26)$$

We also have the special identity for the volume form

$$\star \mathbf{1} = e^1 \wedge e^2 \wedge \cdots \wedge e^m \quad (5.27)$$

So what about actual evaluation? One method is to use the second identity in (5.26). So if we take $\gamma, \epsilon \in \Omega^1 M$ this can be applied to give

$$\begin{aligned} \star(\omega \wedge \gamma) &= i_{\tilde{\gamma}} \star \omega \\ \star(\omega \wedge \gamma \wedge \epsilon) &= i_{\tilde{\epsilon}} \star(\omega \wedge \gamma) = i_{\tilde{\epsilon}}(i_{\tilde{\gamma}} \star \omega) \end{aligned} \quad (5.28)$$

This successive unpacking allows us to construct the Hodge star for any degree form by considering its expansion in a non-coordinate basis. Taking now $\omega_1 \in \Omega^1 M$ we can obtain

$$\star \omega_1 = \star(\omega_{1\mu} e^\mu) = i_{e_\mu} \star \omega_{1\mu} = i_{e^\mu} \omega_{1\mu} \star \mathbf{1} \quad (5.29)$$

Where the Hodge star of a scalar quantity is simply that scalar multiplied by the volume form and $\tilde{e}^\mu = e_\mu$. To check this, remember that the action of \star is to map an r -form into an $(m-r)$ -form. Here, we are starting with a one-form and applying \star means that we contract the metric dual with the volume form to give an $(m-1)$ -form. So if we were in four dimensions, the above would be a three form (see 5.1). Now for $\omega_2 \in \Omega^2 M$, we can similarly calculate

$$\star \omega_2 = \star(\omega_{2\mu\nu} e^\mu \wedge e^\nu) = i_{e_\mu} (\star \omega_{2\mu\nu} e^\nu) = i_{e_\mu} (i_{e_\nu} (\omega_{2\mu\nu} \star \mathbf{1})) \quad (5.30)$$

Generally for $\omega \in \Omega^r M$ we can iterate this to give

$$\star \omega = \star(\omega_{\mu_1 \dots \mu_r} e^{\mu_1} \wedge \dots \wedge e^{\mu_r}) = \omega_{\mu_1 \dots \mu_r} (i_{e_{\mu_1}} \dots i_{e_{\mu_r}} \star \mathbf{1}) \quad (5.31)$$

Applying the Hodge star twice returns the original form up to some sign dependent on the conventions chosen so that

$$\star \star \omega = (-1)^{r(m-r)} \omega \quad (5.32)$$

if the metric is Riemannian and

$$\star \star \omega = (-1)^{1+r(m-r)} \omega \quad (5.33)$$

if it is Lorentzian. Lastly, we note that we also have an inverse \star^{-1} . By demanding $\star \star^{-1} = \star^{-1} \star = \text{id}$, we get that

$$\begin{aligned} \star^{-1} \omega &= (-1)^{r(m-r)} \star \omega \\ \star^{-1} \omega &= (-1)^{1+r(m-r)} \star \omega \end{aligned} \quad (5.34)$$

if the metric is Riemannian or Lorentzian respectively. We can use these operations for an alternative definition of the inner product between two forms

$$\langle \omega, \gamma \rangle = g^{\alpha\beta} \omega_\alpha \gamma_\beta = \star^{-1} (\omega \wedge \star \gamma) \quad (5.35)$$

There is a rather interesting example of this specific to three dimensions. Let's take \mathbb{R}^3 with the basis $\{dx, dy, dz\}$ (obviously, in this case $e^\mu = dx^\mu$). The volume form is $\star \mathbf{1} = dx \wedge dy \wedge dz$ while for the metric we have

$$ds^2 = dx \otimes dx + dy \otimes dy + dz \otimes dz \quad (5.36)$$

Now consider

$$\begin{aligned}
\star(dx \wedge dy) &= i_{\partial_y} i_{\partial_x} \star \mathbf{1} \\
&= i_{\partial_y} i_{\partial_x} (dx \wedge dy \wedge dz) \\
&= i_{\partial_y} (dy \wedge dz) \\
&= dz
\end{aligned} \tag{5.37}$$

We are in three dimensions applying the Hodge star to a two form, so the result should be a one form. Repeating, we also get

$$\star(dx \wedge dz) = -dy, \quad \star(dy \wedge dz) = dx \tag{5.38}$$

This ought to look familiar:

$$\mathbf{e}_x \times \mathbf{e}_y = \mathbf{e}_z, \quad \mathbf{e}_x \times \mathbf{e}_z = -\mathbf{e}_y, \quad \mathbf{e}_y \times \mathbf{e}_z = \mathbf{e}_x \tag{5.39}$$

What we have just obtained is, in a sense, a generalization of the cross product, but one that can be used in an arbitrary number of dimensions. The cross product in \mathbb{R}^3 is special since it is only in this number of dimensions that we can take the product of two vectors (co-vectors) and obtain another vector. In four dimensions, repeating the above would give a two form, in five dimensions a three form and so on. For completeness we also show

$$\begin{aligned}
\star \star (dx \wedge dy) &= \star dz \\
&= i_{\partial_z} \star (dx \wedge dy \wedge dz) \\
&= dx \wedge dy
\end{aligned} \tag{5.40}$$

which gives the same as applying (5.32) directly.

Lastly, we note that there exists a definition of some familiar operators in this formalism that allows us to apply them in any number of dimensions. For some function $f \in \mathcal{F}(M)$ and vector $V \in T_p M$

$$\begin{aligned}
\text{grad}(f) &= \widetilde{df} \\
\text{div}(V) &= \star d \widetilde{V} \\
\text{curl}(V) &= \star^{-1} d \star \widetilde{V}
\end{aligned} \tag{5.41}$$

As an exercise, try to relate these to the usual divergence, gradient and curl operators in three dimensions:

$$\begin{aligned}
\text{grad}(f) &= \nabla f \\
\text{div}(V) &= \nabla \cdot V \\
\text{curl}(V) &= \nabla \times V
\end{aligned} \tag{5.42}$$

5.4 Curvature and Torsion

We were introduced to the ideas of connections, curvature and torsion in a previous chapter. In light of everything we have just covered in this chapter, one may well ask how it all applies to those ideas. To how this changes let's first consider a connection ∇ and its action on a non-coordinate basis $\{e_\mu\}$ (and its corresponding dual)

$$\nabla_{e_\nu} e_\mu = \Gamma^\gamma_{\nu\mu} e_\gamma = \omega^\gamma_\mu(e_\nu) e_\gamma, \tag{5.43}$$

where we have introduced a set of m^2 one forms $\{\omega^\gamma_\mu\}$ called **connection one-forms** related to the Christoffel symbols by either $\Gamma^\gamma_{\nu\mu} e_\gamma = \omega^\gamma_\mu(e_\nu) e_\gamma = i_{e_\nu}(\omega^\gamma_\mu) e_\gamma$ or $\Gamma^\gamma_{\nu\mu} e^\nu = \omega^\gamma_\mu$. We can define the **curvature two-form**

$$R^\alpha_\beta = \frac{1}{2} R^\alpha_{\beta\gamma\delta} e^\gamma \wedge e^\delta \tag{5.44}$$

using the Riemann curvature tensor and the **torsion two form**

$$T^\alpha = \frac{1}{2} T^\alpha_{\gamma\delta} e^\gamma \wedge e^\delta \tag{5.45}$$

from the torsion tensor. These satisfy **Cartan's structure equations**

$$\begin{aligned}
de^\alpha + \omega^\alpha_\beta \wedge e^\beta &= T^\alpha \\
d\omega^\alpha_\beta + \omega^\alpha_\gamma \wedge \omega^\gamma_\beta &= R^\alpha_\beta
\end{aligned} \tag{5.46}$$

Further, if our connection is Levi-Civita (that is torsion free and $\nabla g = 0$), we can find the connection one-forms from our basis using

$$\omega_{\mu\nu} = \frac{1}{2} (i_{e_\nu} de_\mu - i_{e_\mu} de_\nu + e_\alpha i_{e_\mu} i_{e_\nu} de^\alpha) \tag{5.47}$$

Once again, let's take an example: We look at S^2 with the metric

$$ds^2 = d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi. \tag{5.48}$$

We already demonstrated in a previous example that the basis is given by

$$e^1 = d\theta \quad e^2 = \sin(\theta)d\phi \quad (5.49)$$

with the dual basis

$$e_1 = \partial_\theta \quad e_2 = \frac{1}{\sin(\theta)}\partial_\phi. \quad (5.50)$$

So now we can calculate the nonzero connection one-forms as

$$\omega_{12} = -\omega_{21} = -\cot(\theta)e^2 = \cos(\theta)d\phi. \quad (5.51)$$

Since we are working with the Levi-Civita connection to obtain this expression, we know that the right hand side of the first of Cartan's structure equations must be zero. Substituting the above confirms the same for the left hand side:

$$\begin{aligned} de^1 + \omega^1_2 \wedge e^2 &= 0, \\ de^2 + \omega^2_1 \wedge e^1 &= \cot(\theta)e^1 \wedge e^2 + \cot(\theta)e^2 \wedge e^1 = 0. \end{aligned} \quad (5.52)$$

Meanwhile, the second structure equation can be solved for the curvature two form to give

$$R_{12} = -R_{21} = -e^1 \wedge e^2 \quad (5.53)$$

There are two things to note about this result. Firstly, it is frame independent. Since the basis e^μ can be identified with the polar basis in any way (so the great circle section connecting the poles can be chosen to lie anywhere) the above answer is valid in all frames. Secondly, if you have ever gone through the rigmarole of calculating the curvature by using the Christoffel symbols, it is hopefully clear that this way is considerably shorter (Christoffel symbols have $\frac{1}{2}(m^2 + m)$ components while the connection one-forms have $\frac{1}{2}(m^2 - m)$). This is a point we are going to illustrate further in the next section.

5.5 Electromagnetism and Einstein's Equations

An extremely useful application lies in electromagnetism. Expressed in conventional vector calculus, Maxwells equations (in differential form) are

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \rho, \\
\nabla \cdot \mathbf{B} &= 0, \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \\
\nabla \times \mathbf{B} &= \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t},
\end{aligned} \tag{5.54}$$

where \mathbf{E} and \mathbf{B} are the electric and magnetic fields respectively, ρ is the charge density and \mathbf{J} the current density. These are related to the vector potential \mathbf{A} by

$$\begin{aligned}
\mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t}, \\
\mathbf{B} &= \nabla \times \mathbf{A}.
\end{aligned} \tag{5.55}$$

As you might recall, the operators used here are not suited to general relativity owing to them being defined only in three dimensions. Using the contents of this section however, we are in a position to recast these in a general and coordinate independent form. Now, we treat \mathbf{A} as a one-form. The **Maxwell two-form** is obtained by

$$\mathbf{F} = d\mathbf{A}. \tag{5.56}$$

Maxwell's equations are now cast in a much more compact form ²

$$\begin{aligned}
d\mathbf{F} &= 0, \\
d\star \mathbf{F} &= j,
\end{aligned} \tag{5.57}$$

where j is the current three form. As an exercise, take this $\mathbb{R}^{1,3}$ (flat Minkowski space) and by expanding the Maxwell two-form as

$$\mathbf{F} = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu = E_x dx \wedge dt + E_y dy \wedge dt + E_z dz \wedge dt + B_z dx \wedge dy - B_y dx \wedge dz + B_x dy \wedge dz \tag{5.58}$$

and the current three form

$$\star j = \rho dt + j_x dx + j_y dy + j_z dz \tag{5.59}$$

try to obtain the Maxwell equation in (5.55). We now turn to Einstein's equations, how are they to be recast in this form? Take first the contraction

²It has been said that the progress of physics can be tracked through the ever increasing compactification of Maxwell's equations.

$$P_\mu = R^\alpha{}_\mu(e_{\alpha}) = i_{e_\alpha} R^\alpha{}_\mu \quad (5.60)$$

called the **Ricci one-form**. This is related to the Ricci tensor by

$$\text{Ric} = P_\mu \otimes e^\mu. \quad (5.61)$$

We then further consider the **Einstein three-form**

$$G_\mu = R_{\alpha\beta} \wedge \star(e^\alpha \wedge e^\beta \wedge e_\mu), \quad (5.62)$$

which we relate back to the traditional Einstein tensor by

$$\text{Ein} = -\frac{1}{2} \star G_\mu \otimes e^\mu. \quad (5.63)$$

If we then define the **curvature scalar** by

$$\mathcal{P} = i_{e^\alpha} i_{e_\beta} R^\alpha{}_\beta, \quad (5.64)$$

we can recast this in the more familiar form (try as an exercise)

$$\text{Ein} = \text{Ric} - \frac{1}{2} \mathcal{P} g. \quad (5.65)$$

We are, however, not obliged to work in this frame and can in fact write Einstein's equations as

$$G_\mu = 8\pi\tau_\mu, \quad (5.66)$$

where τ_μ is the **stress-energy-momentum three-form** (SEM three-form). A common example, drawing on what we covered above, is the SEM three-form for the electromagnetic field

$$\tau_\mu^{EM} = \frac{1}{2} (i_{e^\mu} F \wedge \star F - F \wedge i_{e^\mu} \star F). \quad (5.67)$$

Or, for a scalar field satisfying the Klein-Gordon equation

$$d \star d\phi - m^2 \phi \star \mathbf{1} = 0, \quad (5.68)$$

it is

$$\tau_\mu^S = \frac{1}{2} (i_{e^\mu} d\phi \wedge \star d\phi - d\phi \wedge i_{e^\mu} \star d\phi) - \frac{1}{2} m^2 \phi^2 \star e_\mu. \quad (5.69)$$

Chapter 6

Topological defects on condensed matter

Differently from particle physics, which tries to understand the matter looking for its fundamental constituent, condensed matter physics tries to explain it through the emergent behavior resulted from the interaction of zillions of electrons, atoms and molecules. To accomplish this, it is necessary to look for the variables and degrees of freedom most important. These variables are often brought together in the so called order parameter, which is a (scalar, vector or tensor) field $\psi : X \rightarrow M$ that maps points from the physical space X into the abstract space M , known as order parameter space. The space M is defined as the space in which the parameter ψ is varied without changing the energy of the system. It is for this reason that this space is also called degeneracy space, since all of the points in this space have the same energy.

Let's see some examples to better illustrate this idea. The order parameter of a bidimensional ferromagnetic material, for instance, is given by the vector field possessing constant magnitude that describes the spins direction in the material. The degeneracy space is then a circle whose points correspond to a direction of the vector field. If \hat{u} and \hat{v} are orthonormal vectors on the plane, the order parameter can be given by

$$f(\vec{r}) = \hat{u} \cos \phi(\vec{r}) + \hat{v} \sin \phi(\vec{r}).$$

The same order parameter space describes the superfluid helium-4, in which the order parameter is the complex scalar field with constant magnitude ψ_0 but with arbitrary phase $\phi(\vec{r})$

$$\psi(\vec{r}) = \psi_0 \exp(i\phi(\vec{r})).$$

In this case the degeneracy space can be thought as a unit circle in the complex plane. Another interesting example is the nematic liquid crystal. It is one of the phases of a

liquid crystal whose molecules have directional symmetry. This means the molecules are aligned to a vector field \mathbf{N} , called director. Considering bidimensional nematics, the order parameter is the director vector \mathbf{N} itself, which takes values on the space S^1 . However, the director vector also has inversion symmetry; this means that \mathbf{N} and $-\mathbf{N}$ represent the same physical state. Therefore, the topological space which describes the nematics phase states is the projective line $\mathbb{R}P^1$, that is, the space obtained by identifying the points $p \in S^1$ with their respect antipodes (Figure 6.1).

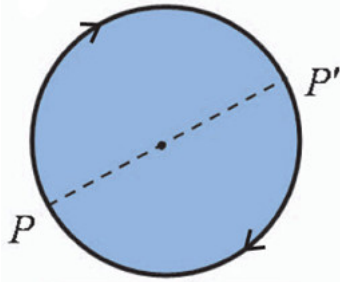


Figure 6.1: Representation of the manifold $\mathbb{R}P^1$ such that the points P and P' are identified. Figure taken from http://ieeexplore.ieee.org/ieee_pilot/articles/06/ttg2009061457/article.html.

When the space M is degenerated, i.e., formed by only one point, we say ψ (and therefore the material itself) is homogeneous. Nonetheless, the field $\psi(\mathbf{r})$ in general varies point to point and there are even regions in the material where ψ is not defined at all. These singular regions are the defects. If this defect can be removed by a continuous transformation of the order parameter ψ , such that the field configuration be transformed into a configuration locally homogeneous, we say the defect is topologically unstable. On the other hand, if the defect cannot be eliminated by continuous transformations on the order parameter, we say the defect is topologically stable, or simply that it is a topological defect. The topological defects can be points (no dimension), lines (1 dimension) or surfaces (2 dimensions). From now on, whenever we say defect is to be understood topological defect, unless we explicitly state otherwise.

Let's suppose that the topological space X is filled with a condensed matter system to be studied. Suppose also that exists a topological defect in this media. By simplicity, let's say that this defect is a point defect. Draw a loop α around the defect in X (Figure 6.2). Through the order parameter ψ , the loop α is mapped into a loop $\beta = \psi \circ \alpha$ in M . It is from the loops in M , more precisely from the fundamental group of M , that we investigate the topology of M and, consequently, classify the topological defects in X . If M has holes or offer restrictions on the deformation of an arbitrary loop β , X has

defects. In other words, the space X might have topological defects whenever there exists a non-trivial homotopy class in M . On the other hand, if every loop is homeotopic to a constant loop c_x , then the physical system in X can be continuously deformed into a defectless configuration. Hence, for each non-trivial homotopy class of M we associate a topological defect. Consequently, the fundamental group gives us enough information for a topological classification of the defects in condensed matter systems.

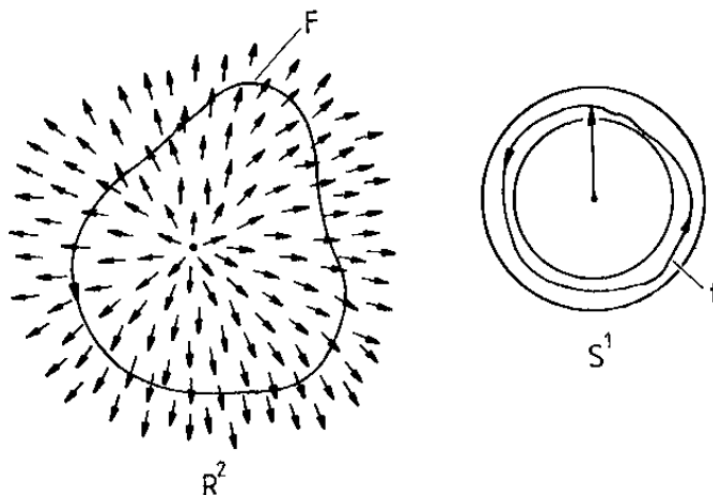


Figure 6.2: Loop around a point defect in an ordered media (left) and the respective loop mapped into the degeneracy space (right). Figure taken from [C14].

Consider as an example point defects in a bidimensional nematic liquid crystal. Since its degeneracy space is $M = \mathbb{R}P^1$, its fundamental group is $\pi_1(\mathbb{R}P^1) \cong \mathbb{Z}$ (see Section 2.3). Therefore we can label the topological defects of a nematic liquid crystal by integer numbers.

On the other hand, if we consider line defects in a three-dimensional nematic liquid crystal, this situation changes drastically. The degeneracy space becomes $\mathbb{R}P^2$, which is the analogous of $\mathbb{R}P^1$ in three dimensions (Figure 6.3). Its fundamental group is $\pi_1(\mathbb{R}P^2) \cong \mathbb{Z}_2$ and, hence, there is only one configuration with topological defect because the other one can be continuously deformed into a free-defect configuration.

This shows that the topological classification of defects is dependent on the dimension of the underlying physical system. In fact, to alter the dimension of a physical system also alters the dimension of the degeneracy space M , which can make a very big difference in its topology. This can be easily seen considering the topologies of S^1 and S^2 as an example. Make a loop in S^1 and try to shrink it into a point (Figure 6.4). Repeat this

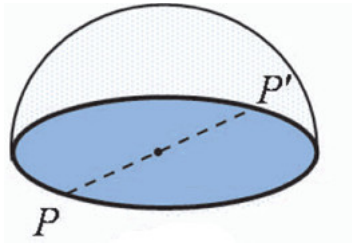


Figure 6.3: Representation of the manifold $\mathbb{R}P^2$ whose points P and P' are identified. Figure taken from http://ieeexplore.ieee.org/ieee_pilot/articles/06/ttg2009061457/article.html.

procedure for S^2 . The extra dimension in S^2 enable us to shrink every loop into a single point.

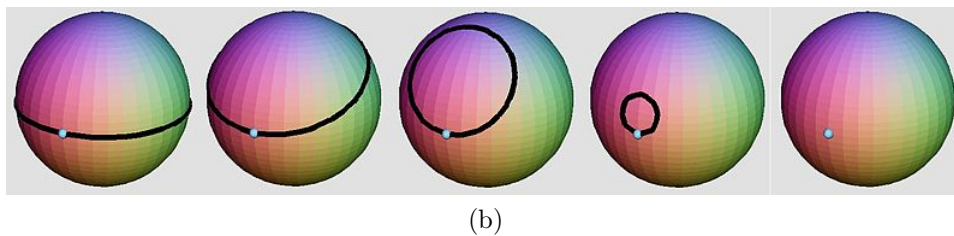
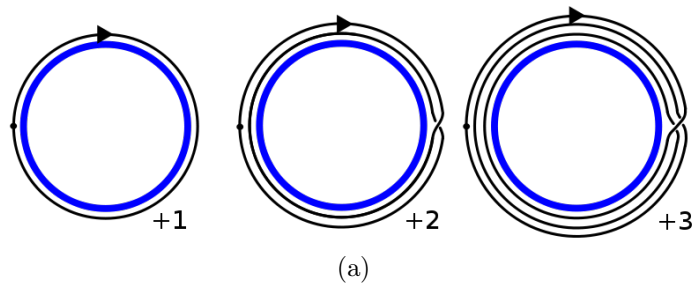


Figure 6.4: Difference in the topology of (a) S^1 and (b) S^2 . In S^1 it is impossible to shrink loops into a point. Figures taken from http://en.wikipedia.org/wiki/Homotopy_groups_of_spheres.

Obtaining the degeneracy space in general is a complicate and delicate task. This involves the concept of symmetric spaces that we have not introduced and, therefore, we will not describe this procedure in detail; the interested reader may consult [C15,C16] for more information. Although the technical proof is quite difficult, the idea is very simple and is closely related to the one used to build Lagrangians in particle physics from known

symmetries. The general rule just says that the degeneracy space of an ordered media is given by the factor group between the total symmetry group G of the energy functional (group of disordered states) by the subgroup H (group of ordered states), whose order parameter is invariant under the subgroup actions. Therefore, in symbols, the degeneracy space is simply given by G/H .

As an example, let's take a look at the bidimensional smectics liquid crystal. The smectics molecules have the same symmetries as the nematics, with the addition of a position symmetry, which forces the molecules to organize in layers (Figure 6.5). In this case, the

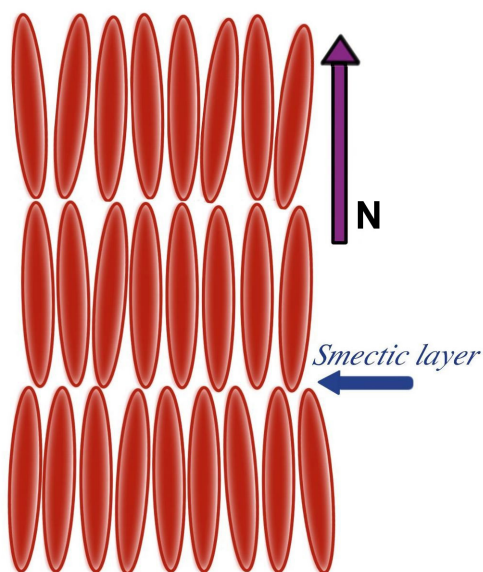


Figure 6.5: Illustration of a smectic liquid crystal. The red rods represent the molecules.

group G can be chosen as the Euclidian group E , that is, the group of all translations and rotations in the plane. Observe in Figure 6.5 that the smectics have continuous symmetry along the layers and discrete symmetry in the direction of \mathbf{N} . Moreover, due to the symmetry of the molecules ($\mathbf{N} \equiv -\mathbf{N}$), smectics also have rotation symmetry of order 2 (group C_2), i.e., they are invariant under rotations of π around the axis normal to the plane which the molecules belong. Hence, the symmetry group of the ordered states is

$$H = (\mathbb{R} \times \mathbb{Z}) \rtimes C_2,$$

where \times and \rtimes are the direct and semi-direct product¹, respectively. It is easy to see (but not so easy to prove) that the factor group E/H is the Klein bottle (Figure 6.6).

¹See [B10] for the definitions and for more details of groups in physics.

In fact, to perform a rotation of π followed by a translation of δ is equivalent to perform a translation of $-\delta$ (convince yourself of this by looking at Figure 6.5). This makes the degeneracy space to take a twisted form and the only way for this to be true is a Klein bottle degeneracy space, whose fundamental group is

$$\pi_1(E/H) \cong \mathbb{Z} \rtimes \mathbb{Z}. \quad (6.1)$$

Therefore, topological defects in smectics liquid crystal are described by a ordered pair of integer numbers (b, k) , where b characterizes defects in the molecules direction (called disclination) and k in the molecules position (called dislocation).

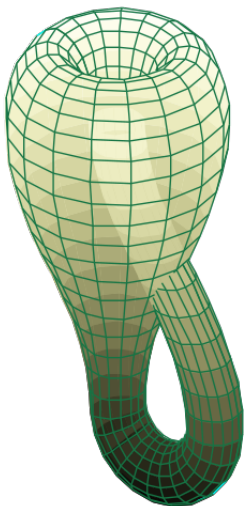


Figure 6.6: Klein bottle immersed in \mathbb{R}^3 .

This example of smectics shows that the application of the homotopy theory for the classification of topological defects in materials with translational order has limitations. Part of the problem is that this theory predicts the existence of defects that are not present in the physical system. In fact, it was shown in [C17] that smectics cannot have disclinations with charge b greater than $+1$ (see next section), but charges $b > +1$ are still predicted by the theory.

6.1 Topological charges

As we have seen in the last section, there are some cases where the defects can be totally described by an integer number. This is always the case when the fundamental

group is isomorphic to the additive group of integers. Defects with charges include vortex in superconductors, dislocations in crystals and point defects in isotropic ferromagnetic, nematics and smectics materials. Therefore, analogously to the electric charge, some defects have topological charges. Mathematically, the topological charge of a defect is the degree of a map or the Poincaré-Hopf index, which is the number of turns that the order parameter $\psi(r)$ performs when r runs a closed loop around the defect. This way the charges can be expressed as functionals of the order parameter and, for this reason, are called analytical topological invariants.

In the nematics, for example, we have seen that the order parameter is described by the vector field \mathbf{N} . The topological charge is then defined by [A2,A3]

$$m = \frac{1}{2\pi} \angle(\mathbf{N}, \mathbf{V}) = \frac{1}{2\pi} \oint_{\gamma} \frac{d\phi}{ds} ds = \frac{\phi(b) - \phi(a)}{2\pi},$$

where \mathbf{V} is any unit vector field, ϕ is the angle between \mathbf{N} and \mathbf{V} and γ is a curve defined on the range $a \leq s \leq b$ such that $\gamma(a) = \gamma(b)$. Since the nematics molecules are symmetric by the inversion $\mathbf{N} \mapsto -\mathbf{N}$, the possible values for this topological charge are semi-integers (Figure 6.7):

$$m \in \frac{\mathbb{Z}}{2} = \{0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \dots\}.$$

However, the set \mathbb{Z} is isomorphic to the semi-integers $\frac{\mathbb{Z}}{2}$, because for each $n \in \mathbb{Z}$ we can assign $f(n) = \frac{n}{2}$ whose inverse is $f^{-1}(n) = 2n$. Consequently, an isomorphism also exists between the bidimensional nematics fundamental group $\pi_1(\mathbb{R}P^1)$, which classifies the defect, and the semi-integers $\frac{\mathbb{Z}}{2}$ that label the topological charge. Symbolically we write

$$\pi_1(\mathbb{R}P^1) \cong \frac{\mathbb{Z}}{2}.$$

This shows that we can assign to each topological defect (that is, to each homotopy class) a topological charge and, hence, we can classify them through these charges. Therefore, defects of different topological charges cannot be transformed into each other continuously. In addition, since the group $\pi_1(\mathbb{R}P^1)$ is isomorphic to the additive group of semi-integers, two defects of charges m_1 and m_2 can be summed, producing a single defect of charge $m = m_1 + m_2$. Consequently, defects of opposite sign and same modulus can annihilate themselves producing a configuration with no defects.

This also occurs partly in the smectics because they inherit the director vector from the nematics phase. However, in other materials like the ferromagnetic ones which do not have that inversion symmetry (spins in opposite directions represent different physical states), only integer values for the topological charge m are allowed, since configurations

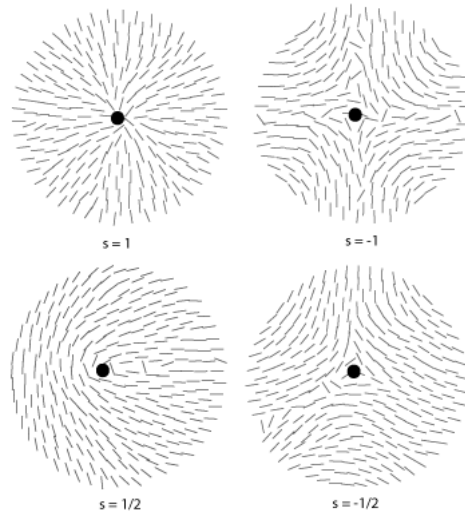


Figure 6.7: Configuration of the nematics molecules in the presence of topological defects of different charges s . Figure taken from http://www.doitpoms.ac.uk/tlplib/liquid_crystals/printall.php.

with semi-integer charges would cost infinite energy due to the discontinuity in the field of spins (see Figure 6.8).

This chapter was intended to give a brief introduction to the subject and to try to show what has been done currently considering applications of topology, or more precisely of homotopy theory, in physics. It was not intended to give a deep description of the physics of condensed matter systems though and the reader that feels uncomfortable with such physical concepts is encourage to take a look in the rich literature. Some good references on condensed matter include [B11,C18–C20]. Topological defects in these systems can be found in [C14,C21,C22]. Clearly there are a lot of other places besides topological defects where topology could be useful and others where it has already proved its usefulness, including geometrical phases, quantum anomalies and topological insulators. Outside condensed matter, topology has also been used to study quantum field theories (specially topological quantum field theories) and topological questions in gravity. As one can imagine, the literature is vast and we hope the reader has enjoyed the present course and could use it as a bridge to others more profound and advanced.

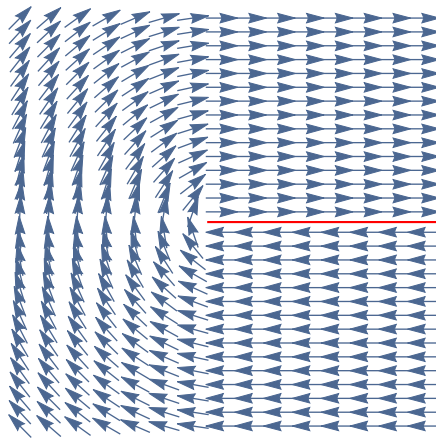


Figure 6.8: Spins configuration (represented in blue) if there were a point defect of charge $+\frac{1}{2}$ in the origin. The discontinuity (represented in red) extends along the whole positive x axis. In order to this configuration to exist, would be necessary an infinite amount of energy.

Bibliography

Geometry and Topology

- [A1] A. Hatcher, *Algebraic Topology*. Cambridge University Press, 2009.
- [A2] B. O’neill, *Elementary differential geometry*. Elsevier, 2 ed., 2006.
- [A3] A. Pressley, *Elementary Differential Geometry*. Springer, 2010.
- [A4] M. Carmo, *Riemmanian Geometry*. Birkhäuser, 1992.
- [A5] M. D. Carmo, *Differential geometry of curves and surfaces*. Pearson, 1976.
- [A6] J. Lee, *Introduction to Topological Manifolds*. Springer, 2010.
- [A7] J. Lee, *Introduction to Smooth Manifolds*. Springer, 2012.
- [A8] J. Lee, *Riemannian Manifolds: An Introduction to Curvature*. Springer, 1997.

Applications in Physics

- [B9] M. Nakahara, *Geometry, topology and physics*. CRC Press, 2003.
- [B10] J. Cornwell, *Group Theory in Physics: An Introduction*. Academic Press, 1997.
- [B11] R. Kamien, “The geometry of soft materials: a primer,” *Reviews of Modern Physics*, vol. 74, no. October, 2002.

- [B12] P. Szekeres, *A Course in Modern Mathematical Physics: Groups, Hilbert Space and Differential Geometry*. Cambridge University Press, 2004.
- [B13] J. Baez and J. Muniain, *Gauge Fields, Knots and Gravity*. World Scientific Publishing Company, 1994.

Condensed Matter and Topological Defects

- [C14] H. Trebin, “The topology of non-uniform media in condensed matter physics,” *Advances in Physics*, 1982.
- [C15] M. Kleman and J. Friedel, “Disclinations, dislocations and continuous defects: a reappraisal,” *Reviews of Modern Physics*, p. 72, apr 2007.
- [C16] D. Ruppel and E. Sackmann, “On defects in different phases of two-dimensional lipid bilayers,” *Journal de Physique*, 1983.
- [C17] V. Poénaru, “Some aspects of the theory of defects of ordered media and gauge fields related to foliations,” *Communications in mathematical physics*, vol. 136, pp. 127–136, 1981.
- [C18] P. Chaikin and T. Lubensky, *Principles of condensed matter physics*. Cambridge University Press, 2000.
- [C19] P. D. Gennes and J. Prost, *The physics of liquid crystals*. Oxford University Press, 1993.
- [C20] M. Kleman and O. D. Lavrentovich, *Soft Matter Physics: An Introduction*. Partially Ordered Systems, New York, NY: Springer New York, 2004.
- [C21] N. Mermin, “The topological theory of defects in ordered media,” *Reviews of Modern Physics*, no. 3, 1979.
- [C22] M. Kurik and O. Lavrentovich, “Defects in liquid crystals: homotopy theory and experimental studies,” *Soviet Physics Uspekhi*, vol. 196, 1988.