

# Mathematical Notes

# Contents

<b>I</b>	<b>Basic Mathematics</b>	<b>4</b>
<b>1</b>	<b>Axiomatic Set Theory</b>	<b>5</b>
1.1	Propositional Logic . . . . .	5
1.2	Predicate Logic . . . . .	7
1.3	Axiomatic Systems & Theory Of Proofs . . . . .	8
1.4	The $\in$ -relation . . . . .	9
1.5	Zermelo-Fraenkel Axioms Of Set Theory . . . . .	10
1.6	Maps Between Sets . . . . .	14
1.7	Equivalence Relations . . . . .	17
1.8	Construction Of $\mathbb{N}$ , $\mathbb{Z}$ , $\mathbb{Q}$ & $\mathbb{R}$ . . . . .	19
<b>2</b>	<b>Algebraic Structures</b>	<b>22</b>
2.1	Groups . . . . .	22
2.2	Fields . . . . .	23
2.3	Vector Spaces . . . . .	23
2.3.1	Linear Maps . . . . .	24
2.3.2	Basis Of Vector Spaces . . . . .	27
2.3.3	Change Of Basis . . . . .	31
2.3.4	Determinants . . . . .	32
2.4	Rings . . . . .	34
2.5	Modules . . . . .	35
2.5.1	Basis Of Modules . . . . .	36
2.6	Algebras . . . . .	36
<b>3</b>	<b>Lie Algebras</b>	<b>37</b>
3.1	Basic Definitions . . . . .	37
3.2	The Adjoint Map & The Killing Form . . . . .	40
3.3	The Fundamental Roots & The Weyl Group . . . . .	42
3.4	Dynkin Diagrams & The Cartan Classification . . . . .	46
3.5	Application: Reconstruction Of $A_2$ From Its Dynkin Diagram . . . . .	48
3.6	Representations Of Lie Algebras . . . . .	51
3.6.1	The Casimir Operator . . . . .	54
<b>4</b>	<b>Topology</b>	<b>59</b>
4.1	Topological Spaces . . . . .	59
4.2	Construction Of New Topologies From Given Ones . . . . .	62
4.3	Convergence & Continuity . . . . .	63
4.4	Invariant Topological Properties . . . . .	65
4.4.1	Separation Properties . . . . .	65
4.4.2	Connectedness & Path-Connectedness . . . . .	67
4.4.3	Homotopic Curves & The Fundamental Group . . . . .	68
<b>5</b>	<b>Topological Manifolds</b>	<b>71</b>
5.1	Topological Manifolds . . . . .	71
5.2	Charts & Atlases . . . . .	72
5.3	Differentiable Manifolds . . . . .	75
5.3.1	Classification Of Differentiable Structures . . . . .	76

5.4	Tangent Space . . . . .	77
5.4.1	Co-Ordinate Induced Basis For The Tangent Space . . . . .	80
5.4.2	Change Of Vector Components Under A Change Of Chart . . . . .	82
5.5	Cotangent Space . . . . .	83
5.5.1	Dual Basis For The Cotangent Space . . . . .	83
5.5.2	Change Of Covector Components Under A Change Of Chart . . . . .	84
5.6	Push-Forward & Pull-Back . . . . .	84
5.6.1	Immersion & Embeddings . . . . .	87
5.7	The Tangent Bundle . . . . .	88
5.7.1	Topological Bundles . . . . .	88
5.7.2	Tangent Bundle . . . . .	92
5.7.3	Vector, Covector & Tensor Fields . . . . .	93
5.8	Differential Forms . . . . .	99
5.8.1	The Grassmann Algebra . . . . .	101
5.8.2	The Exterior Derivative . . . . .	103
5.8.3	De Rham Cohomology . . . . .	105
5.9	Application: $SL(2, \mathbb{C})$ - Part 1 . . . . .	107
<b>6</b>	<b>Extra Structures On Topological Manifolds</b>	<b>112</b>
6.1	Connection / Covariant Derivative . . . . .	112
6.1.1	Parallel Transport . . . . .	116
6.1.2	Torsion . . . . .	118
6.1.3	Curvature . . . . .	119
6.2	Metric Manifolds . . . . .	120
6.2.1	Geodesics . . . . .	122
<b>7</b>	<b>Lie Groups</b>	<b>125</b>
7.1	Lie Groups . . . . .	125
7.2	The Left Translation Map . . . . .	126
7.3	The Lie Algebra Of A Lie Group . . . . .	127
7.4	Application: $SL(2, \mathbb{C})$ - Part 2 . . . . .	131
7.5	Representations Of Lie Groups . . . . .	140
7.6	Reconstruction Of A Lie group From Its Lie Algebra . . . . .	142
7.6.1	The Exponential Map . . . . .	143
7.7	Lie Group Actions On A Manifold . . . . .	146
<b>8</b>	<b>Principal Fiber Bundles</b>	<b>150</b>
8.1	Principal Bundles . . . . .	150
8.1.1	Principal Bundle Maps . . . . .	151
8.2	Associated Bundles . . . . .	154
8.2.1	Associated Bundle Maps . . . . .	155
8.3	Application: Principal Frame Bundle . . . . .	156
8.4	Connections On A Principal $G$ -Bundle . . . . .	160
8.4.1	Connection One-Forms . . . . .	162
8.5	Yang-Mills Fields & Local Representations . . . . .	163
8.5.1	The Maurer-Cartan Form . . . . .	165
8.5.2	The Gauge Map . . . . .	167
8.6	Parallel Transport . . . . .	168
8.6.1	Horizontal Lifts To the principal bundle . . . . .	169
8.6.2	Solution of the horizontal lift ODE by a path-ordered exponential . . . . .	170
8.6.3	Horizontal lifts to the associated bundle . . . . .	172
8.7	Covariant exterior derivative and curvature . . . . .	173
8.7.1	Torsion . . . . .	175
8.8	Covariant Derivatives . . . . .	176
8.8.1	Equivalence of local sections and equivariant functions . . . . .	176
8.8.2	Linear actions on associated vector fiber bundles . . . . .	177
8.8.3	Construction of the covariant derivative . . . . .	178

**Part I**

**Basic Mathematics**

# Chapter 1

## Axiomatic Set Theory

Axiomatic set theory is a branch of mathematical logic that studies sets, which informally are collections of objects. Although any type of object can be collected into a set, set theory is applied most often to objects that are relevant to mathematics. The language of set theory can be used to define nearly all mathematical objects.

The modern study of set theory was initiated by Georg Cantor and Richard Dedekind in the 1870s. After the discovery of paradoxes in naive set theory, such as Russell’s paradox, numerous axiom systems were proposed in the early twentieth century, of which the Zermelo–Fraenkel axioms, with or without the axiom of choice, are the best-known.

Set theory is commonly employed as a foundational system for mathematics, particularly in the form of Zermelo–Fraenkel set theory with the axiom of choice. Beyond its foundational role, set theory is a branch of mathematics in its own right, with an active research community. Contemporary research into set theory includes a diverse collection of topics, ranging from the structure of the real number line to the study of the consistency of large cardinals.

### 1.1 Propositional Logic

**Definition 1.1** (Proposition). A **proposition**  $p$  is a variable<sup>1</sup> that can take the values true ( $T$ ) or false ( $F$ ), and no others.

This is what a proposition is from the point of view of propositional logic. In particular, it is not the task of propositional logic to decide whether a complex statement of the form “there is extraterrestrial life” is true or not. Propositional logic already deals with the complete proposition, and it just assumes that is either true or false. It is also not the task of propositional logic to decide whether a statement of the type “in winter is colder than outside” is a proposition or not (i.e. if it has the property of being either true or false). In this particular case, the statement looks rather meaningless.

**Definition 1.2** (Tautology). A proposition which is always true is called a **tautology**.

**Definition 1.3** (Contradiction). A proposition which is always false is called a **contradiction**.

It is possible to build new propositions from given ones using *logical operators*. The simplest kind of logical operators are *unary* operators, which take in one proposition and return another proposition. There are four unary operators in total, and they differ by the truth value of the resulting proposition which, in general, depends on the truth value of  $p$ . We can represent them in a table as follows:

$p$	$\neg p$	$\text{id}(p)$	$\top p$	$\perp p$
F	T	F	T	F
T	F	T	T	F

where  $\neg$  is the *negation* operator,  $\text{id}$  is the *identity* operator,  $\top$  is the *tautology* operator and  $\perp$  is the *contradiction* operator. These clearly exhaust all possibilities for unary operators.

---

<sup>1</sup>By this we mean a formal expression, with no extra structure assumed.

The next step is to consider *binary* operators, i.e. operators that take in two propositions and return a new proposition. There are four combinations of the truth values of two propositions and, since a binary operator assigns one of the two possible truth values to each of those, we have 16 binary operators in total. The operators  $\wedge$ ,  $\vee$  and  $\underline{\vee}$ , called *and*, *or* and *exclusive or* respectively, should already be familiar to you:

$p$	$q$	$p \wedge q$	$p \vee q$	$p \underline{\vee} q$
F	F	F	F	F
F	T	F	T	T
T	F	F	T	T
T	T	T	T	F

There is one binary operator, the *implication* operator  $\Rightarrow$ , which is sometimes a little ill understood, unless you are already very knowledgeable about these things. Its usefulness comes in conjunction with the *equivalence* operator  $\Leftrightarrow$ . We have:

$p$	$q$	$p \Rightarrow q$	$p \Leftrightarrow q$
F	F	T	T
F	T	T	F
T	F	F	F
T	T	T	T

While the fact that the proposition  $p \Rightarrow q$  is true whenever  $p$  is false may be surprising at first, it is just the definition of the implication operator and it is an expression of the principle “Ex falso quod libet”, that is, from a false assumption anything follows. Of course, you may be wondering why on earth we would want to define the implication operator in this way. The answer to this is hidden in the following result.

**Theorem 1.1.** *Let  $p, q$  be propositions. Then  $(p \Rightarrow q) \Leftrightarrow ((\neg q) \Rightarrow (\neg p))$ .*

*Proof.*

We simply construct the truth tables for  $p \Rightarrow q$  and  $(\neg q) \Rightarrow (\neg p)$ :

$p$	$q$	$\neg p$	$\neg q$	$p \Rightarrow q$	$(\neg q) \Rightarrow (\neg p)$
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	F	F	T	T

The columns for  $p \Rightarrow q$  and  $(\neg q) \Rightarrow (\neg p)$  are identical and hence we are done. □

*Remark 1.1.* We agree on decreasing binding strength in the sequence:

$$\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$$

For example,  $(\neg q) \Rightarrow (\neg p)$  may be written unambiguously as  $\neg q \Rightarrow \neg p$ .

*Remark 1.2.* All higher order operators  $\heartsuit(p_1, \dots, p_N)$  can be constructed from a single binary operator defined by:

$p$	$q$	$p \uparrow q$
F	F	T
F	T	T
T	F	T
T	T	F

This is called the *nand* operator and, in fact, we have  $(p \uparrow q) \Leftrightarrow \neg(p \wedge q)$ .

## 1.2 Predicate Logic

**Definition 1.4** (Predicate). A **predicate** is a proposition-valued function of some variable or variables.

**Definition 1.5** (Relation). A predicate of two variables is called a **relation**.

For example,  $P(x)$  is a proposition for each choice of the variable  $x$ , and its truth value depends on  $x$ . Similarly, the predicate  $Q(x, y)$  is, for any choice of  $x$  and  $y$ , a proposition and its truth value depends on  $x$  and  $y$ .

Just like for propositional logic, it is not the task of predicate logic to examine how predicates are built from the variables on which they depend. In order to do that, one would need some further language establishing the rules to combine the variables  $x$  and  $y$  into a predicate. Also, you may want to specify from which “set”  $x$  and  $y$  come from. Instead, we leave it completely open, and simply consider  $x$  and  $y$  formal variables, with no extra conditions imposed.

This may seem a bit weird since from elementary school one is conditioned to always ask where “ $x$ ” comes from upon seeing an expression like  $P(x)$ . However, it is crucial that we refrain from doing this here, since we want to only later define the notion of set, using the language of propositional and predicate logic. As with propositions, we can construct new predicates from given ones by using the operators define in the previous section. For example, we might have:

$$Q(x, y, z) :\Leftrightarrow P(x) \wedge R(y, z)$$

where the symbol  $:\Leftrightarrow$  means “defined as being equivalent to”.

More interestingly, we can construct a new proposition from a given predicate by using *quantifiers*.

**Definition 1.6** (Universal Quantifier). Let  $P(x)$  be a predicate. Then:

$$\forall x : P(x)$$

is a proposition, which we read as “for all  $x$ ,  $P$  of  $x$  (is true)”, and it is defined to be true if  $P(x)$  is true independently of  $x$ , false otherwise. The symbol  $\forall$  is called **universal quantifier**.

**Definition 1.7** (Existential Quantifier). Let  $P(x)$  be a predicate. Then we define:

$$\exists x : P(x) :\Leftrightarrow \neg(\forall x : \neg P(x))$$

The proposition  $\exists x : P(x)$  is read as “there exists (at least one)  $x$  such that  $P$  of  $x$  (is true)” and the symbol  $\exists$  is called **existential quantifier**.

The following result is an immediate consequence of these definitions.

**Corollary 1.1.** Let  $P(x)$  be a predicate. Then:

$$\forall x : P(x) \Leftrightarrow \neg(\exists x : \neg P(x))$$

*Remark 1.3.* It is possible to define quantification of predicates of more than one variable. In order to do so, one proceeds in steps quantifying a predicate of one variable at each step.

*Example 1.1.*

Let  $P(x, y)$  be a predicate. Then, for fixed  $y$ ,  $P(x, y)$  is a predicate of one variable and we define:

$$Q(y) :\Leftrightarrow \forall x : P(x, y)$$

Hence we may have the following:

$$\exists y : \forall x : P(x, y) :\Leftrightarrow \exists y : Q(y)$$

Other combinations of quantifiers are defined analogously.

*Remark 1.4.* The order of quantification matters (if the quantifiers are not all the same). For a given predicate  $P(x, y)$ , the propositions:

$$\exists y : \forall x : P(x, y) \quad \text{and} \quad \forall x : \exists y : P(x, y)$$

are not necessarily equivalent.

*Example 1.2.*

Consider the proposition expressing the existence of additive inverses in the real numbers. We have:

$$\forall x : \exists y : x + y = 0$$

i.e. for each  $x$  there exists an inverse  $y$  such that  $x + y = 0$ . For 1 this is  $-1$ , for 2 it is  $-2$  etc. Consider now the proposition obtained by swapping the quantifiers in the previous proposition:

$$\exists y : \forall x : x + y = 0$$

What this proposition is saying is that there exists a real number  $y$  such that, no matter what  $x$  is, we have  $x + y = 0$ . This is clearly false, since if  $x + y = 0$  for some  $x$  then  $(x + 1) + y \neq 0$ , so the same  $y$  cannot work for both  $x$  and  $x + 1$ , let alone every  $x$ .

Notice that the proposition  $\exists x : P(x)$  means “there exists *at least one*  $x$  such that  $P(x)$  is true”. Often in mathematics we prove that “there exists *a unique*  $x$  such that  $P(x)$  is true”. We therefore have the following definition.

**Definition 1.8** (Unique Existential Quantifier). *Let  $P(x)$  be a predicate. We define the **unique existential quantifier**  $\exists!$  by:*

$$\exists! x : P(x) :\Leftrightarrow (\exists x : P(x)) \wedge \forall y : \forall z : (P(y) \wedge P(z) \Rightarrow y = z)$$

This definition clearly separates the existence condition from the uniqueness condition. An equivalent definition with the advantage of brevity is:

$$\exists! x : P(x) :\Leftrightarrow (\exists x : \forall y : P(y) \Leftrightarrow x = y)$$

### 1.3 Axiomatic Systems & Theory Of Proofs

**Definition 1.9** (Axiomatic System). *An **axiomatic system** is a finite sequence of propositions  $a_1, a_2, \dots, a_N$ , which are called the axioms of the system.*

**Definition 1.10** (Proof). *A **proof** of a proposition  $p$  within an axiomatic system  $a_1, a_2, \dots, a_N$  is a finite sequence of propositions  $q_1, q_2, \dots, q_M$  such that  $q_M = p$  and for any  $1 \leq j \leq M$  one of the following is satisfied:*

(A)  $q_j$  is a proposition from the list of axioms.

(T)  $q_j$  is a tautology.

(M)  $\exists 1 \leq m, n < j : (q_m \wedge q_n \Rightarrow q_j)$  is true.

*Remark 1.5.* If  $p$  can be proven within an axiomatic system  $a_1, a_2, \dots, a_N$ , we write:

$$a_1, a_2, \dots, a_N \vdash p$$

and we read “ $a_1, a_2, \dots, a_N$  proves  $p$ ”.

*Remark 1.6.* This definition of proof allows to easily recognise a proof. A computer could easily check that whether or not the conditions (A), (T) and (M) are satisfied by a sequence of propositions. To actually find a proof of a proposition is a whole different story.

*Remark 1.7.* Obviously, any tautology that appears in the list of axioms of an axiomatic system can be removed from the list without impairing the power of the axiomatic system.

An extreme case of an axiomatic system is propositional logic. The axiomatic system for propositional logic is the empty sequence. This means that all we can prove in propositional logic are tautologies.

**Definition 1.11** (Consistent). *An axiomatic system  $a_1, a_2, \dots, a_N$  is said to be **consistent** if there exists a proposition  $q$  which cannot be proven from the axioms. In symbols:*

$$\exists q : \neg(a_1, a_2, \dots, a_N \vdash q)$$

The idea behind this definition is the following. Consider an axiomatic system which contains contradicting propositions:

$$a_1, \dots, s, \dots, \neg s, \dots, a_N$$

Then, given *any* proposition  $q$ , the following is a proof of  $q$  within this system:

$$s, \neg s, q$$

Indeed,  $s$  and  $\neg s$  are legitimate steps in the proof since they are axioms. Moreover,  $s \wedge \neg s$  is a contradiction and thus  $(s \wedge \neg s) \Rightarrow q$  is a tautology. Therefore,  $q$  follows from condition (M). This shows that any proposition can be proven within a system with contradictory axioms. In other words, the inability to prove every proposition is a property possessed by no contradictory system, and hence we define a consistent system as one with this property.

Having come this far, we can now state (and prove) an impressively sounding theorem.

**Theorem 1.2.** *Propositional logic is consistent.*

*Proof.*

Suffices to show that there exists a proposition that cannot be proven within propositional logic. Propositional logic has the empty sequence as axioms. Therefore, only conditions (T) and (M) are relevant here. The latter allows the insertion of a proposition  $q_j$  such that  $(q_m \wedge q_n) \Rightarrow q_j$  is true, where  $q_m$  and  $q_n$  are propositions that precede  $q_j$  in the proof sequence. However, since (T) only allows the insertion of a tautology anywhere in the proof sequence, the propositions  $q_m$  and  $q_n$  must be tautologies. Consequently, for  $(q_m \wedge q_n) \Rightarrow q_j$  to be true,  $q_j$  must also be a tautology. Hence, the proof sequence consists entirely of tautologies and thus only tautologies can be proven.

Now let  $q$  be any proposition. Then  $q \wedge \neg q$  is a contradiction, hence not a tautology and thus cannot be proven. Therefore, propositional logic is consistent.  $\square$

*Remark 1.8.* While it is perfectly fine and clear how to define consistency, it is perfectly difficult to prove consistency for a given axiomatic system, propositional logic being a big exception.

**Theorem 1.3.** *Any axiomatic system powerful enough to encode elementary arithmetic is either inconsistent or contains an undecidable proposition, i.e. a proposition that can be neither proven nor disproven within the system.*

An example of an undecidable proposition is the Continuum hypothesis within the Zermelo-Fraenkel axiomatic system.

## 1.4 The $\in$ -relation

Set theory is built on the postulate that there is a fundamental relation (i.e. a predicate of two variables) denoted  $\in$  and read as “epsilon”. There will be no definition of what  $\in$  is, or of what a set is. Instead, we will have nine axioms concerning  $\in$  and sets, and it is only in terms of these nine axioms that  $\in$  and sets are defined at all. Here is an overview of the axioms. We will have:

- 2 basic existence axioms, one about the  $\in$  relation and the other about the existence of the empty set.
- 4 construction axioms, which establish rules for building new sets from given ones. They are the pair set axiom, the union set axiom, the replacement axiom and the power set axiom.
- 2 further existence/construction axioms, these are slightly more advanced and newer compared to the others.
- 1 axiom of foundation, excluding some constructions as not being sets.

Using the  $\in$ -relation we can immediately define the following relations:

- $x \notin y \Leftrightarrow \neg(x \in y)$
- $x \subseteq y \Leftrightarrow \forall a : (a \in x \Rightarrow a \in y)$

- $x = y :\Leftrightarrow (x \subseteq y) \wedge (y \subseteq x)$
- $x \subset y :\Leftrightarrow (x \subseteq y) \wedge \neg(x = y)$

*Remark 1.9.* A comment about notation. Since  $\in$  is a predicate of two variables, for consistency of notation we should write  $\in(x, y)$ . However, the notation  $x \in y$  is much more common (as well as intuitive) and hence we simply define:

$$x \in y :\Leftrightarrow \in(x, y)$$

and we read “ $x$  is in (or belongs to)  $y$ ” or “ $x$  is an element (or a member) of  $y$ ”. Similar remarks apply to the other relations  $\notin$ ,  $\subseteq$  and  $=$ .

## 1.5 Zermelo-Fraenkel Axioms Of Set Theory

**Axiom on the  $\in$ -relation.** *The expression  $x \in y$  is a proposition if, and only if, both  $x$  and  $y$  are sets. In symbols:*

$$\forall x : \forall y : (x \in y) \vee \neg(x \in y)$$

We remarked, previously, that it is not the task of predicate logic to inquire about the nature of the variables on which predicates depend. This first axiom clarifies that the variables on which the relation  $\in$  depend are sets. In other words, if  $x \in y$  is not a proposition (i.e. it does not have the property of being either true or false) then  $x$  and  $y$  are not both sets.

This seems so trivial that, for a long time, people thought that this not much of a condition. But, in fact, it is. It tells us when something is not a set.

*Example 1.3.*

This is the so called “Russel’s Paradox”. Suppose that there is some  $u$  which has the following property:

$$\forall x : (x \notin x \Leftrightarrow x \in u)$$

i.e.  $u$  contains all the sets that are not elements of themselves, and no others. We wish to determine whether  $u$  is a set or not. In order to do so, consider the expression  $u \in u$ . If  $u$  is a set then, by the first axiom,  $u \in u$  is a proposition.

However, we will show that this is not the case. Suppose first that  $u \in u$  is true. Then  $\neg(u \notin u)$  is true and thus  $u$  does not satisfy the condition for being an element of  $u$ , and hence is not an element of  $u$ . Thus:

$$u \in u \Rightarrow \neg(u \in u)$$

and this is a contradiction. Therefore,  $u \in u$  cannot be true. Then, if it is a proposition, it must be false. However, if  $u \notin u$ , then  $u$  satisfies the condition for being a member of  $u$  and thus:

$$u \notin u \Rightarrow \neg(u \notin u)$$

which is, again, a contradiction. Therefore,  $u \in u$  does not have the property of being either true or false (it can be neither) and hence it is not a proposition. Thus, our first axiom implies that  $u$  is not a set, for if it were, then  $u \in u$  would be a proposition.

*Remark 1.10.* The fact that  $u$  as defined above is not a set means that expressions like:

$$u \in u, \quad x \in u, \quad u \in x, \quad x \notin u, \quad \text{etc}$$

are not propositions and thus, they are not part of axiomatic set theory.

**Axiom on the existence of an empty set.** *There exists a set that contains no elements. In symbols:*

$$\exists y : \forall x : x \notin y$$

Notice the use of “an” above. In fact, we have all the tools to prove that there is only one empty set. We do not need this to be an axiom.

**Theorem 1.4.** *There is only one empty set, and we denote it by  $\emptyset$ .*

*Proof.*

Suppose that  $x$  and  $x'$  are both empty sets. Then  $y \in x$  is false as  $x$  is the empty set. But then:

$$(y \in x) \Rightarrow (y \in x')$$

is true, and in particular it is true independently of  $y$ . Therefore:

$$\forall y : (y \in x) \Rightarrow (y \in x')$$

and hence  $x \subseteq x'$ .

Conversely, by the same argument, we have:

$$\forall y : (y \in x') \Rightarrow (y \in x)$$

and thus  $x' \subseteq x$ .

Hence  $(x \subseteq x') \wedge (x' \subseteq x)$  and therefore  $x = x'$ .  $\square$

**Axiom on pair sets.** *Let  $x$  and  $y$  be sets. Then there exists a set that contains as its elements precisely  $x$  and  $y$ . In symbols:*

$$\forall x : \forall y : \exists m : \forall u : (u \in m \Leftrightarrow (u = x \vee u = y))$$

The set  $m$  is called the *pair set* of  $x$  and  $y$  and it is denoted by  $\{x, y\}$ .

*Remark 1.11.* We have chosen  $\{x, y\}$  as the notation for the pair set of  $x$  and  $y$ , but what about  $\{y, x\}$ ? The fact that the definition of the pair set remains unchanged if we swap  $x$  and  $y$  suggests that  $\{x, y\}$  and  $\{y, x\}$  are the same set. Indeed, by definition, we have:

$$(a \in \{x, y\} \Rightarrow a \in \{y, x\}) \wedge (a \in \{y, x\} \Rightarrow a \in \{x, y\})$$

independently of  $a$ , hence  $(\{x, y\} \subseteq \{y, x\}) \wedge (\{y, x\} \subseteq \{x, y\})$  and thus  $\{x, y\} = \{y, x\}$ .

The pair set  $\{x, y\}$  is thus an unordered pair. However, using the axiom on pair sets, it is also possible to define an *ordered pair*  $(x, y)$  such that  $(x, y) \neq (y, x)$ . The defining property of an ordered pair is the following:

$$(x, y) = (a, b) \Leftrightarrow x = a \wedge y = b$$

One candidate which satisfies this property is  $(x, y) := \{x, \{x, y\}\}$ , which is a set by the axiom on pair sets.

*Remark 1.12.* The pair set axiom also guarantees the existence of one-element sets, called *singletons*. If  $x$  is a set, then we define  $\{x\} := \{x, x\}$ . Informally, we can say that  $\{x\}$  and  $\{x, x\}$  express the same amount of information, namely that they contain  $x$ .

**Axiom on union sets.** *Let  $x$  be a set. Then there exists a set whose elements are precisely the elements of the elements of  $x$ . In symbols:*

$$\forall x : \exists u : \forall y : (y \in u \Leftrightarrow \exists s : (y \in s \wedge s \in x))$$

The set  $u$  is denoted by  $\bigcup x$ .

*Example 1.4.*

Let  $a, b$  be sets. Then  $\{a\}$  and  $\{b\}$  are sets by the pair set axiom, and hence  $x := \{\{a\}, \{b\}\}$  is a set, again by the pair set axiom. Then the expression:

$$\bigcup x = \{a, b\}$$

is a set by the union axiom.

Notice that, since  $a$  and  $b$  are sets, we could have immediately concluded that  $\{a, b\}$  is a set by the pair set axiom. The union set axiom is really needed to construct sets with more than 2 elements.

*Example 1.5.*

Let  $a, b, c$  be sets. Then  $\{a\}$  and  $\{b, c\}$  are sets by the pair set axiom, and hence  $x := \{\{a\}, \{b, c\}\}$  is a set, again by the pair set axiom. Then the expression:

$$\bigcup x =: \{a, b, c\}$$

is a set by the union set axiom. This time the union set axiom was really necessary to establish that  $\{a, b, c\}$  is a set, i.e. in order to be able to use it meaningfully in conjunction with the  $\in$ -relation.

The previous example easily generalises to a definition.

**Definition 1.12** (Union Of Sets). *Let  $a_1, a_2, \dots, a_N$  be sets. We define recursively for all  $N \geq 2$ :*

$$\{a_1, a_2, \dots, a_{N+1}\} := \bigcup \{\{a_1, a_2, \dots, a_N\}, \{a_{N+1}\}\}$$

*Remark 1.13.* The fact that the  $x$  that appears in  $\bigcup x$  has to be a set is a crucial restriction. Informally, we can say that it is only possible to take unions of as many sets as would fit into a set. The “collection” of all the sets that do not contain themselves is not a set or, we could say, does not fit into a set. Therefore it is not possible to take the union of all the sets that do not contain themselves. This is very subtle, but also very precise.

**Axiom of replacement.** *Let  $R$  be a functional relation and let  $m$  be a set. Then the image of  $m$  under  $R$ , denoted by  $\text{im}_R(m)$ , is again a set.*

Of course, we now need to define the new terms that appear in this axiom. Recall that a relation is simply a predicate of two variables.

**Definition 1.13** (Functional Relation). *A relation  $R$  is said to be **functional** if:*

$$\forall x : \exists! y : R(x, y)$$

**Definition 1.14** (Image Of A Set Under A Relational Functional Relation). *Let  $m$  be a set and let  $R$  be a functional relation. The **image of  $m$  under  $R$**  consists of all those  $y$  for which there is an  $x \in m$  such that  $R(x, y)$ .*

None of the previous axioms imply that the image of a set under a functional relation is again a set. The assumption that it always is, is made explicit by the axiom of replacement.

It is very likely that the reader has come across a weaker form of the axiom of replacement, called the *principle of restricted comprehension*, which says the following.

**Proposition 1.1.** *Let  $P(x)$  be a predicate and let  $m$  be a set. Then the elements  $y \in m$  such that  $P(y)$  is true constitute a set, which we denote by:*

$$\{y \in m \mid P(y)\}$$

*Remark 1.14.* The principle of restricted comprehension is not to be confused with the “principle” of universal comprehension which states that  $\{y \mid P(y)\}$  is a set for any predicate and was shown to be inconsistent by Russell. Observe that the  $y \in m$  condition makes it so that  $\{y \in m \mid P(y)\}$  cannot have more elements than  $m$  itself.

*Remark 1.15.* If  $y$  is a set, we define the following notation:

$$\forall x \in y : P(x) :\Leftrightarrow \forall x : (x \in y \Rightarrow P(x))$$

and:

$$\exists x \in y : P(x) :\Leftrightarrow \neg(\forall x \in y : \neg P(x))$$

Pulling the  $\neg$  through, we can also write:

$$\begin{aligned} \exists x \in y : P(x) &\Leftrightarrow \neg(\forall x \in y : \neg P(x)) \\ &\Leftrightarrow \neg(\forall x : (x \in y \Rightarrow \neg P(x))) \\ &\Leftrightarrow \exists x : \neg(x \in y \Rightarrow \neg P(x)) \\ &\Leftrightarrow \exists x : (x \in y \wedge P(x)) \end{aligned}$$

where we have used the equivalence  $(p \Rightarrow q) \Leftrightarrow \neg(p \wedge \neg q)$ .

The principle of restricted comprehension is a consequence of the axiom of replacement.

*Proof.*

We have two cases:

1. If  $\neg(\exists y \in m : P(y))$ , then we define:  $\{y \in m \mid P(y)\} := \emptyset$ .
2. If  $\exists \hat{y} \in m : P(\hat{y})$ , then let  $R$  be the functional relation:

$$R(x, y) := (P(x) \wedge x = y) \vee (\neg P(x) \wedge \hat{y} = y)$$

and hence define  $\{y \in m \mid P(y)\} := \text{im}_R(m)$ . □

Don't worry if you don't see this immediately. You need to stare at the definitions for a while and then it will become clear.

*Remark 1.16.* We will rarely invoke the axiom of replacement in full. We will only invoke the weaker principle of restricted comprehension, with which we are all familiar with.

We can now define the intersection and the relative complement of sets.

**Definition 1.15** (Intersection). *Let  $x$  be a set. Then we define the **intersection** of  $x$  by:*

$$\bigcap x := \{a \in \bigcup x \mid \forall b \in x : a \in b\}$$

*If  $a, b \in x$  and  $\bigcap x = \emptyset$ , then  $a$  and  $b$  are said to be disjoint.*

**Definition 1.16** (Complement). *Let  $u$  and  $m$  be sets such that  $u \subseteq m$ . Then the **complement** of  $u$  relative to  $m$  is defined as:*

$$m \setminus u := \{x \in m \mid x \notin u\}$$

*These are both sets by the principle of restricted comprehension, which is ultimately due to axiom of replacement.*

**Axiom on the existence of power sets.** *Let  $m$  be a set. Then there exists a set, denoted by  $\mathcal{P}(m)$ , whose elements are precisely the subsets of  $m$ . In symbols:*

$$\forall x : \exists y : \forall a : (a \in y \Leftrightarrow a \subseteq x)$$

Historically, in naïve set theory, the principle of universal comprehension was thought to be needed in order to define the power set of a set. Traditionally, this would have been (inconsistently) defined as:

$$\mathcal{P}(m) := \{y \mid y \subseteq m\}$$

To define power sets in this fashion, we would need to know, a priori, from which “bigger” set the elements of the power set come from. However, this is not possible based only on the previous axioms and, in fact, there is no other choice but to dedicate an additional axiom for the existence of power sets.

*Example 1.6.*

Let  $m = \{a, b\}$ . Then  $\mathcal{P}(m) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$ .

*Remark 1.17.* If one defines  $(a, b) := \{a, \{a, b\}\}$ , then the *cartesian product*  $x \times y$  of two sets  $x$  and  $y$ , which informally is the set of all ordered pairs of elements of  $x$  and  $y$ , satisfies:

$$x \times y \subseteq \mathcal{P}(\mathcal{P}(\bigcup \{x, y\}))$$

Hence, the existence of  $x \times y$  as a set follows from the axioms on unions, pair sets, power sets and the principle of restricted comprehension.

**Axiom of infinity.** *There exists a set that contains the empty set and, together with every other element  $y$ , it also contains the set  $\{y\}$  as an element. In symbols:*

$$\exists x : \emptyset \in x \wedge \forall y : (y \in x \Rightarrow \{y\} \in x)$$

Let us consider one such set  $x$ . Then  $\emptyset \in x$  and hence  $\{\emptyset\} \in x$ . Thus, we also have  $\{\{\emptyset\}\} \in x$  and so on. Therefore:

$$x = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, \dots\}$$

We can introduce the following notation for the elements of  $x$ :

$$0 := \emptyset, \quad 1 := \{\emptyset\}, \quad 2 := \{\{\emptyset\}\}, \quad 3 := \{\{\{\emptyset\}\}\}, \quad \dots$$

**Corollary 1.2.** *The “set”  $\mathbb{N} := x$  is a set according to axiomatic set theory.*

This would not be the case without the axiom of infinity since it is not possible to prove that  $\mathbb{N}$  constitutes a set from the previous axioms.

*Remark 1.18.* At this point, one might suspect that we would need an extra axiom for the existence of the real numbers. But, in fact, we can define  $\mathbb{R} := \mathcal{P}(\mathbb{N})$ , which is a set by the axiom on power sets.

*Remark 1.19.* The version of the axiom of infinity that we stated is the one that was first put forward by Zermelo. A more modern formulation is the following. *There exists a set that contains the empty set and, together with every other element  $y$ , it also contains the set  $y \cup \{y\}$  as an element.* Here we used the notation:

$$x \cup y := \bigcup \{x, y\}$$

With this formulation, the natural numbers look like:

$$\mathbb{N} := \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \dots\}$$

This may appear more complicated than what we had before, but it is much nicer for two reasons. First, the natural number  $n$  is represented by an  $n$ -element set rather than a one-element set. Second, it generalizes much more naturally to the system of transfinite ordinal numbers where the successor operation  $s(x) = x \cup \{x\}$  applies to transfinite ordinals as well as natural numbers. Moreover, the natural numbers have the same defining property as the ordinals: they are transitive sets strictly well-ordered by the  $\in$ -relation.

**Axiom of choice.** *Let  $x$  be a set whose elements are non-empty and mutually disjoint. Then there exists a set  $y$  which contains exactly one element of each element of  $x$ . In symbols:*

$$\forall x : P(x) \Rightarrow \exists y : \forall a \in x : \exists! b \in a : a \in y$$

where  $P(x) \Leftrightarrow (\exists a : a \in x) \wedge (\forall a : \forall b : (a \in x \wedge b \in x) \Rightarrow \bigcap \{a, b\} = \emptyset)$ .

*Remark 1.20.* The axiom of choice is independent of the other 8 axioms, which means that one could have set theory with or without the axiom of choice. However, standard mathematics uses the axiom of choice and hence so will we. There is a number of theorems that can only be proved by using the axiom of choice. Amongst these we have:

- Every vector space has a basis.
- There exists a complete system of representatives of an equivalence relation.

**Axiom of foundation.** *Every non-empty set  $x$  contains an element  $y$  that has none of its elements in common with  $x$ . In symbols:*

$$\forall x : (\exists a : a \in x) \Rightarrow \exists y \in x : \bigcap \{x, y\} = \emptyset$$

An immediate consequence of this axiom is that there is no set that contains itself as an element.

The totality of all these nine axioms are called *ZFC set theory*, which is a shorthand for Zermelo-Fraenkel set theory with the axiom of Choice.

## 1.6 Maps Between Sets

A recurrent theme in mathematics is the classification of *spaces* by means of structure-preserving *maps* between them.

A space is usually meant to be some set equipped with some structure, which is usually some other set. We will define each instance of space precisely when we will need them. In the case of sets considered themselves as spaces, there is no extra structure beyond the set and hence, the structure may be taken to be the empty set.

**Definition 1.17** (Map). *Let  $A, B$  be sets. A **map**  $\phi: A \rightarrow B$  is a relation such that for each  $a \in A$  there exists exactly one  $b \in B$  such that  $\phi(a, b)$ .*

The standard notation for a map is:

$$\begin{aligned}\phi: A &\rightarrow B \\ a &\mapsto \phi(a)\end{aligned}$$

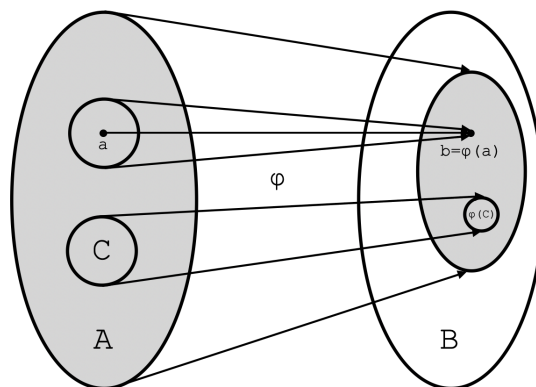
which is technically an abuse of notation since  $\phi$ , being a relation of two variables, should have two arguments and produce a truth value. However, once we agree that for each  $a \in A$  there exists exactly one  $b \in B$  such that  $\phi(a, b)$  is true, then for each  $a$  we can define  $\phi(a)$  to be precisely that unique  $b$ . It is sometimes useful to keep in mind that  $\phi$  is actually a relation.

*Example 1.7.*

Let  $M$  be a set. The simplest example of a map is the *identity map* on  $M$ :

$$\begin{aligned}\text{id}_M: M &\rightarrow M \\ m &\mapsto m\end{aligned}$$

We will now provide some very basic and standard terminology for a map  $\phi: A \rightarrow B$ , that we will be using throughout the notes. It worth spending some time on learning and understanding the terminology.



- The set  $A$  is called the **domain** of  $\phi$ .
- The set  $B$  is called the **codomain** or the **target** of  $\phi$ .
- If  $a$  is an element of  $A$ , then  $\phi(a) = b$  (the value of  $\phi$  when applied to  $a$ ) is called the **image of element** or the **output** of  $a$  under  $\phi$ .
- If  $C$  is a subset of  $A$ , then  $\phi(C)$  (the set of values of  $\phi$  when applied to  $C$ ) is called the **image of subset** of  $C$  under  $\phi$ .
- The set of all elements that the map  $\phi$  can hit in the target  $B$  (grey area in  $B$ ) is called the **image** or the **range** of  $A$  under  $\phi$  (in other words the image of a map is simply the image of its entire domain). Notice that since a map  $\phi$  hits every point of the domain  $A$ , the whole domain  $A$  is covered by  $\phi$  (grey area in  $A$ ). However it is not necessary that the mapping will also cover the whole target  $B$ . This is why the image of a map is not necessarily equal to the whole target.
- The set of all elements of the domain  $A$  that are mapped into a given single element  $b$  of the target  $B$  is called the **fiber** of the element  $b$  under  $\phi$ .
- The subset  $C$  of all elements of the domain  $A$  that are mapped into a subset  $\phi(C)$  of the target  $B$  is called the **preimage** or the **inverse image** of  $\phi(C)$  under  $\phi$ .

- A map  $\phi$  is called **injective** or an **injection** or **one-to-one** if distinct elements of the domain  $A$  map to distinct elements in the target  $B$ , or equivalently if each element of the target  $B$  is mapped to by at most one element of the domain  $A$ :  $\forall a_1, a_2 \in A : \phi(a_1) = \phi(a_2) \Rightarrow a_1 = a_2$ .
- A map  $\phi$  is called **surjective** or a **surjection** or **onto** if its image is equal to the entire domain  $A$ , or equivalently if each element of the target  $B$  is mapped to by at least one element of the domain  $A$ :  $\text{im}_\phi(A) = B$ .
- A map  $\phi$  is called **bijective** or a **bijection**, or **one-to-one and onto** if it is both injective and surjective.

**Definition 1.18** (Isomorphic Sets). *Two sets  $A$  and  $B$  are called **isomorphic** if there exists a bijection  $\phi: A \rightarrow B$ . In this case, we write  $A \cong_{\text{set}} B$ .*

*Remark 1.21.* If there is any bijection  $A \rightarrow B$  then generally there are many.

Bijections are the “structure-preserving” maps for sets. Intuitively, they pair up the elements of  $A$  and  $B$  and a bijection between  $A$  and  $B$  exists only if  $A$  and  $B$  have the same “size”. This is clear for finite sets, but it can also be extended to infinite sets.

**Definition 1.19** (Infinite/Finite Sets). *A set  $A$  is called:*

- Infinite if there exists a proper subset  $B \subset A$  such that  $B \cong_{\text{set}} A$ . In particular, if  $A$  is infinite, we further define  $A$  to be:
  - \* Countably infinite if  $A \cong_{\text{set}} \mathbb{N}$ .
  - \* Uncountably infinite otherwise.
- Finite if it is not infinite. In this case, we have  $A \cong_{\text{set}} \{1, 2, \dots, N\}$  for some  $N \in \mathbb{N}$  and we say that the cardinality of  $A$ , denoted by  $|A|$ , is  $N$ .

Given two maps  $\phi: A \rightarrow B$  and  $\psi: B \rightarrow C$ , we can construct a third map, called the *composition* of  $\phi$  and  $\psi$ , denoted by  $\psi \circ \phi$  (read “psi after phi”), defined by:

$$\begin{aligned} \psi \circ \phi: A &\rightarrow C \\ a &\mapsto \psi(\phi(a)) \end{aligned}$$

This is often represented by drawing the following diagram

$$\begin{array}{ccc} & B & \\ \phi \nearrow & & \searrow \psi \\ A & \xrightarrow{\psi \circ \phi} & C \end{array}$$

and by saying that “the diagram commutes” (although sometimes this is assumed even if it is not explicitly stated). What this means is that every path in the diagram gives the same result. This might seem notational overkill at this point, but later we will encounter situations where we will have many maps, going from many places to many other places and these diagrams greatly simplify the exposition.

**Proposition 1.2.** *Composition of maps is associative.*

*Proof.*

Indeed, let  $\phi: A \rightarrow B$ ,  $\psi: B \rightarrow C$  and  $\xi: C \rightarrow D$  be maps. Then we have:

$$\begin{aligned} \xi \circ (\psi \circ \phi): A &\rightarrow D \\ a &\mapsto \xi(\psi(\phi(a))) \end{aligned}$$

and:

$$\begin{aligned} (\xi \circ \psi) \circ \phi: A &\rightarrow D \\ a &\mapsto \xi(\psi(\phi(a))) \end{aligned}$$

Thus  $\xi \circ (\psi \circ \phi) = (\xi \circ \psi) \circ \phi$ . □

The operation of composition is necessary in order to defined inverses of maps.

**Definition 1.20** (Inverse). *Let  $\phi: A \rightarrow B$  be a bijection. Then the **inverse** of  $\phi$ , denoted  $\phi^{-1}$ , is defined (uniquely) by:*

$$\phi^{-1} \circ \phi = \text{id}_A$$

$$\phi \circ \phi^{-1} = \text{id}_B$$

Equivalently, we require the following diagram to commute:

$$\begin{array}{ccc} & \phi & \\ \text{id}_A \hookrightarrow A & \xrightarrow{\quad} & B \xrightarrow{\quad} \text{id}_B \\ & \phi^{-1} & \end{array}$$

The inverse map is only defined for bijections. However, the notion of the pre-image, which we will often meet in topology, is defined for any map. Given the inverse map we can define the pre-image in a more systematic way as follows.

**Definition 1.21** (Pre-image). *Let  $\phi: A \rightarrow B$  be a map and let  $V \subseteq B$ . Then we define the set:*

$$\text{preim}_\phi(V) := \{a \in A \mid \phi(a) \in V\}$$

*called the **pre-image** of  $V$  under  $\phi$ .*

**Proposition 1.3.** *Let  $\phi: A \rightarrow B$  be a map, let  $U, V \subseteq B$  and  $C = \{C_j \mid j \in J\} \subseteq \mathcal{P}(B)$ . Then:*

- i)  $\text{preim}_\phi(\emptyset) = \emptyset$  and  $\text{preim}_\phi(B) = A$ .
- ii)  $\text{preim}_\phi(U \setminus V) = \text{preim}_\phi(U) \setminus \text{preim}_\phi(V)$ .
- iii)  $\text{preim}_\phi(\bigcup C) = \bigcup_{j \in J} \text{preim}_\phi(C_j)$  and  $\text{preim}_\phi(\bigcap C) = \bigcap_{j \in J} \text{preim}_\phi(C_j)$ .

*Proof.*

i) By definition, we have:

$$\text{preim}_\phi(B) = \{a \in A : \phi(a) \in B\} = A$$

and:

$$\text{preim}_\phi(\emptyset) = \{a \in A : \phi(a) \in \emptyset\} = \emptyset$$

ii) We have:

$$\begin{aligned} a \in \text{preim}_\phi(U \setminus V) &\Leftrightarrow \phi(a) \in U \setminus V \\ &\Leftrightarrow \phi(a) \in U \wedge \phi(a) \notin V \\ &\Leftrightarrow a \in \text{preim}_\phi(U) \wedge a \notin \text{preim}_\phi(V) \\ &\Leftrightarrow a \in \text{preim}_\phi(U) \setminus \text{preim}_\phi(V) \end{aligned}$$

iii) We have:

$$\begin{aligned} a \in \text{preim}_\phi(\bigcup C) &\Leftrightarrow \phi(a) \in \bigcup C \\ &\Leftrightarrow \bigvee_{j \in J} (\phi(a) \in C_j) \\ &\Leftrightarrow \bigvee_{j \in J} (a \in \text{preim}_\phi(C_j)) \\ &\Leftrightarrow a \in \bigcup_{j \in J} \text{preim}_\phi(C_j) \end{aligned}$$

Similarly, we get  $\text{preim}_\phi(\bigcap C) = \bigcap_{j \in J} \text{preim}_\phi(C_j)$ . □

## 1.7 Equivalence Relations

**Definition 1.22** (Equivalence Relation). *Let  $M$  be a set and let  $\sim$  be a relation such that the following conditions are satisfied:*

i) *Reflexivity*:  $\forall m \in M : m \sim m$ .

ii) *Symmetry*:  $\forall m, n \in M : m \sim n \Leftrightarrow n \sim m$ .

iii) *Transitivity*:  $\forall m, n, p \in M : (m \sim n \wedge n \sim p) \Rightarrow m \sim p$ .

Then  $\sim$  is called an **equivalence relation** on  $M$ .

*Example 1.8.*

Consider the following wordy examples.

- a)  $p \sim q :\Leftrightarrow p$  is of the same opinion as  $q$ . This relation is reflexive, symmetric and transitive. Hence, it is an equivalence relation.
- b)  $p \sim q :\Leftrightarrow p$  is a sibling of  $q$ . This relation is symmetric and transitive but not reflexive and hence, it is not an equivalence relation.
- c)  $p \sim q :\Leftrightarrow p$  is taller  $q$ . This relation is transitive, but neither reflexive nor symmetric and hence, it is not an equivalence relation.
- d)  $p \sim q :\Leftrightarrow p$  is in love with  $q$ . This relation is generally not reflexive. People don't like themselves very much. It is certainly not normally symmetric, which is the basis of much drama in literature. It is also not transitive, except in some French films.

**Definition 1.23** (Equivalence Class). *Let  $\sim$  be an equivalence relation on the set  $M$ . Then, for any  $m \in M$ , we define the set:*

$$[m] := \{n \in M \mid m \sim n\}$$

*called the **equivalence class** of  $m$ . Note that the condition  $m \sim n$  is equivalent to  $n \sim m$  since  $\sim$  is symmetric.*

The following are two key properties of equivalence classes.

**Proposition 1.4.** *Let  $\sim$  be an equivalence relation on  $M$ . Then:*

- i)  $a \in [m] \Rightarrow [a] = [m]$ .
- ii) either  $[m] = [n]$  or  $[m] \cap [n] = \emptyset$ .

*Proof.*

- i) Since  $a \in [m]$ , we have  $a \sim m$ . Let  $x \in [a]$ . Then  $x \sim a$  and hence  $x \sim m$  by transitivity. Therefore  $x \in [m]$  and hence  $[a] \subseteq [m]$ . Similarly, we have  $[m] \subseteq [a]$  and hence  $[a] = [m]$ .
- ii) Suppose that  $[m] \cap [n] \neq \emptyset$ . That is:

$$\exists z : z \in [m] \wedge z \in [n]$$

Thus  $z \sim m$  and  $z \sim n$  and hence, by symmetry and transitivity,  $m \sim n$ . This implies that  $m \in [n]$  and hence that  $[m] = [n]$ .  $\square$

**Definition 1.24** (Quotient Set). *Let  $\sim$  be an equivalence relation on  $M$ . Then we define the **quotient set** of  $M$  by  $\sim$  as:*

$$M/\sim := \{[m] \mid m \in M\}$$

*This is indeed a set since  $[m] \subseteq \mathcal{P}(M)$  and hence we can write more precisely:*

$$M/\sim := \{[m] \in \mathcal{P}(M) \mid m \in M\}$$

*Then clearly  $M/\sim$  is a set by the power set axiom and the principle of restricted comprehension.*

*Remark 1.22.* Due to the axiom of choice, there exists a complete system of representatives for  $\sim$ , i.e. a set  $R$  such that  $R \cong_{\text{set}} M/\sim$ .

*Remark 1.23.* Care must be taken when defining maps whose domain is a quotient set if one uses representatives to define the map. In order for the map to be *well-defined* one needs to show that the map is independent of the choice of representatives.

*Example 1.9.*

Let  $M = \mathbb{Z}$  and define  $\sim$  by:

$$m \sim n :\Leftrightarrow n - m \in 2\mathbb{Z}$$

It is easy to check that  $\sim$  is indeed an equivalence relation. Moreover, we have:

$$[0] = [2] = [4] = \dots = [-2] = [-4] = \dots$$

and:

$$[1] = [3] = [5] = \dots = [-1] = [-3] = \dots$$

Thus we have:  $\mathbb{Z}/\sim = \{[0], [1]\}$ . We wish to define an addition  $\oplus$  on  $\mathbb{Z}/\sim$  by inheriting the usual addition on  $\mathbb{Z}$ . As a tentative definition we could have:

$$\oplus: \mathbb{Z}/\sim \times \mathbb{Z}/\sim \rightarrow \mathbb{Z}/\sim$$

being given by:

$$[a] \oplus [b] := [a + b]$$

However, we need to check that our definition does not depend on the choice of class representatives, i.e. if  $[a] = [a']$  and  $[b] = [b']$ , then we should have:

$$[a] \oplus [b] = [a'] \oplus [b'].$$

Indeed,  $[a] = [a']$  and  $[b] = [b']$  means  $a - a' \in 2\mathbb{Z}$  and  $b - b' \in 2\mathbb{Z}$ , i.e.  $a - a' = 2m$  and  $b - b' = 2n$  for some  $m, n \in \mathbb{Z}$ . We thus have:

$$\begin{aligned} [a' + b'] &= [a - 2m + b - 2n] \\ &= [(a + b) - 2(m + n)] \\ &= [a + b] \end{aligned}$$

where the last equality follows since:

$$(a + b) - 2(m + n) - (a + b) = -2(m + n) \in 2\mathbb{Z}$$

Therefore  $[a'] \oplus [b'] = [a] \oplus [b]$  and hence the operation  $\oplus$  is well-defined.

*Example 1.10.*

As a counterexample, with the same set-up as in the previous example, let us define an operation  $\star$  by:

$$[a] \star [b] := \frac{a}{b}$$

This is easily seen to be *ill-defined* since  $[1] = [3]$  and  $[2] = [4]$  but:

$$[1] \star [2] = \frac{1}{2} \neq \frac{3}{4} = [3] \star [4]$$

## 1.8 Construction of $\mathbb{N}$ , $\mathbb{Z}$ , $\mathbb{Q}$ and $\mathbb{R}$

Recall that, invoking the axiom of infinity, we defined the natural numbers:

$$\mathbb{N} := \{0, 1, 2, 3, \dots\}$$

where:

$$0 := \emptyset, \quad 1 := \{\emptyset\}, \quad 2 := \{\{\emptyset\}\}, \quad 3 := \{\{\{\emptyset\}\}\}, \quad \dots$$

We would now like to define an addition operation on  $\mathbb{N}$  by using the axioms of set theory. We will need some preliminary definitions.

**Definition 1.25** (Successor Map). *The **successor map**  $S$  on  $\mathbb{N}$  is defined by:*

$$\begin{aligned} S: \mathbb{N} &\rightarrow \mathbb{N} \\ n &\mapsto \{n\} \end{aligned}$$

*Example 1.11.*

Consider  $S(2)$ . Since  $2 := \{\{\emptyset\}\}$ , we have  $S(2) = \{\{\{\emptyset\}\}\} =: 3$ . Therefore, we have  $S(2) = 3$  as we would have expected.

To make progress, we also need to define the predecessor map, which is only defined on the set  $\mathbb{N}^* := \mathbb{N} \setminus \{\emptyset\}$ .

**Definition 1.26** (Predecessor Map). *The **predecessor map**  $P$  on  $\mathbb{N}^*$  is defined by:*

$$\begin{aligned} P: \mathbb{N}^* &\rightarrow \mathbb{N} \\ n &\mapsto m \text{ such that } m \in n \end{aligned}$$

*Example 1.12.*

We have  $P(2) = P(\{\{\emptyset\}\}) = \{\emptyset\} = 1$ .

**Definition 1.27** ( $n$ -th Power). *Let  $n \in \mathbb{N}$ . The  **$n$ -th power** of  $S$ , denoted  $S^n$ , is defined recursively by:*

$$\begin{aligned} S^n &:= S \circ S^{P(n)} && \text{if } n \in \mathbb{N}^* \\ S^0 &:= \text{id}_{\mathbb{N}} \end{aligned}$$

We are now ready to define addition.

**Definition 1.28** (Addition Of Natural Numbers). *The **addition** operation on  $\mathbb{N}$  is defined as a map:*

$$\begin{aligned} +: \mathbb{N} \times \mathbb{N} &\rightarrow \mathbb{N} \\ (m, n) &\mapsto m + n := S^n(m) \end{aligned}$$

*Example 1.13.*

We have:

$$2 + 1 = S^1(2) = S(2) = 3$$

and:

$$1 + 2 = S^2(1) = S(S^1(1)) = S(S(1)) = S(2) = 3$$

Using this definition, it is possible to show that  $+$  is commutative and associative. The *neutral element* of  $+$  is 0 since:

$$m + 0 = S^0(m) = \text{id}_{\mathbb{N}}(m) = m$$

and:

$$0 + m = S^m(0) = S^{P(m)}(1) = S^{P(P(m))}(2) = \dots = S^0(m) = m$$

Clearly, there exist no inverses for  $+$  in  $\mathbb{N}$ , i.e. given  $m \in \mathbb{N}$  (non-zero), there exist no  $n \in \mathbb{N}$  such that  $m + n = 0$ . This motivates the extension of the natural numbers to the integers. In order to rigorously define  $\mathbb{Z}$ , we need to define the following relation on  $\mathbb{N} \times \mathbb{N}$ .

Let  $\sim$  be the relation on  $\mathbb{N} \times \mathbb{N}$  defined by:

$$(m, n) \sim (p, q) :\Leftrightarrow m + q = p + n$$

It is easy to check that this is an equivalence relation as:

- i)  $(m, n) \sim (m, n)$  since  $m + n = m + n$ .
- ii)  $(m, n) \sim (p, q) \Rightarrow (p, q) \sim (m, n)$  since  $m + q = p + n \Leftrightarrow p + n = m + q$ .

iii)  $((m, n) \sim (p, q) \wedge (p, q) \sim (r, s)) \Rightarrow (m, n) \sim (r, s)$  since we have:

$$m + q = p + n \wedge p + s = r + q$$

hence  $m + q + p + s = p + n + r + q$ , and thus  $m + s = r + n$ .

By equipping this relation we can define the set of integers in the following way.

**Definition 1.29** (Integers). *We define the set of **integers** by:*

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N}) / \sim$$

The intuition behind this definition is that the pair  $(m, n)$  stands for “ $m - n$ ”. In other words, we represent each integer by a pair of natural numbers whose (yet to be defined) difference is precisely that integer. There are, of course, many ways to represent the same integer with a pair of natural numbers in this way. For instance, the integer  $-1$  could be represented by  $(1, 2)$ ,  $(2, 3)$ ,  $(112, 113)$ ,  $\dots$

Notice however that  $(1, 2) \sim (2, 3)$ ,  $(1, 2) \sim (112, 113)$ , etc. and indeed, taking the quotient by  $\sim$  takes care of this “redundancy”. Notice also that this definition relies entirely on previously defined entities.

*Remark 1.24.* In a first introduction to set theory it is not unlikely to find the claim that the natural numbers are part of the integers, i.e.  $\mathbb{N} \subseteq \mathbb{Z}$ . However, according to our definition, this is obviously nonsense since  $\mathbb{N}$  and  $\mathbb{Z} := (\mathbb{N} \times \mathbb{N}) / \sim$  contain entirely different elements. What is true is that  $\mathbb{N}$  can be *embedded* into  $\mathbb{Z}$ , i.e. there exists an *inclusion map*  $\iota$ , given by:

$$\begin{aligned} \iota: \mathbb{N} &\hookrightarrow \mathbb{Z} \\ n &\mapsto [(n, 0)] \end{aligned}$$

and it is in this sense that  $\mathbb{N}$  is included in  $\mathbb{Z}$ .

**Definition 1.30** (Inverse Of Integer). *Let  $n := [(n, 0)] \in \mathbb{Z}$ . Then we define the **inverse** of  $n$  to be  $-n := [(0, n)]$ .*

We would now like to inherit the  $+$  operation from  $\mathbb{N}$ .

**Definition 1.31** (Addition Of Integers). *We define the **addition of integers**  $+_{\mathbb{Z}}: \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$  by:*

$$[(m, n)] +_{\mathbb{Z}} [(p, q)] := [(m + p, n + q)]$$

Since we used representatives to define  $+_{\mathbb{Z}}$ , we would need to check that  $+_{\mathbb{Z}}$  is well-defined. It is an easy exercise.

*Example 1.14.*

$$2 +_{\mathbb{Z}} (-3) := [(2, 0)] +_{\mathbb{Z}} [(0, 3)] = [(2, 3)] = [(0, 1)] =: -1. \text{ Hallelujah!}$$

In a similar fashion, we define the set of *rational numbers* by:

$$\mathbb{Q} := (\mathbb{Z} \times \mathbb{Z}^*) / \sim$$

where  $\mathbb{Z}^* := \mathbb{Z} \setminus \{0\}$  and  $\sim$  is a relation on  $\mathbb{Z} \times \mathbb{Z}^*$  given by:

$$(p, q) \sim (r, s) :\Leftrightarrow ps = qr$$

assuming that a *multiplication* operation on the integers has already been defined.

*Example 1.15.*

We have  $(2, 3) \sim (4, 6)$  since  $2 \times 6 = 12 = 3 \times 4$ .

Similarly to what we did for the integers, here we are representing each rational number by the collection of pairs of integers (the second one in each pair being non-zero) such that their (yet to be defined) ratio is precisely that rational number. Thus, for example, we have:

$$\frac{2}{3} := [(2, 3)] = [(4, 6)] = \dots$$

There are many ways to construct the reals from the rationals however we will skip them for now.

## Chapter 2

# Algebraic Structures

**Definition 2.1** (Algebraic Structures). *A set  $A$  (called the underlying set, carrier set or domain), together with a collection of maps (called operations) on  $A$  of finite arity (typically binary operations), and a finite set of identities, known as axioms, that these operations must satisfy, is called an **algebraic structure**. Some algebraic structures also involve another set (called the scalar set).*

Examples of algebraic structures with a single underlying set include groups, fields and rings. Examples of algebraic structures with two underlying sets include vector spaces, modules, and algebras. In this section we will review the most important algebraic structures for our purposes.

One has to be careful with the terminology since it changes depending on the area of mathematics. For example, in the context of universal algebra, the set  $A$  with this structure is called an algebra, while, in other contexts, it is (somewhat ambiguously) called an algebraic structure, the term algebra being reserved for specific algebraic structures that are vector spaces over a field or modules over a commutative ring.

The properties of specific algebraic structures are studied in abstract algebra. The general theory of algebraic structures has been formalized in universal algebra. The language of category theory is used to express and study relationships between different classes of algebraic and non-algebraic objects. This is because it is sometimes possible to find strong connections between some classes of objects, sometimes of different kinds. For example, Galois theory establishes a connection between certain fields and groups: two algebraic structures of different kinds.

In this chapter we will introduce the basic algebraic structures by giving their definitions and some of their key properties. In later chapter we get into depth in various topics of algebraic structures.

## 2.1 Groups

**Definition 2.2** (Group). *A **group** is a tuple  $(G, \cdot)$ , where  $G$  is a set (called the underlying set of the group) and  $\cdot$  is a map (called operation)  $G \times G \rightarrow G$  satisfying the following four group axioms:*

- *Closure:*  $\forall a, b \in G : a \cdot b \in G$ .
- *Associativity:*  $\forall a, b, c \in G : (a \cdot b) \cdot c = a \cdot (b \cdot c)$ .
- *Neutral Element:*  $\exists e \in G : \forall a \in G : a \cdot e = e \cdot a = a$ .
- *Inverse Element:*  $\forall a \in G : \exists a^{-1} \in G : a \cdot a^{-1} = a^{-1} \cdot a = e$ .

The identity element  $e$  of a group  $G$  is often written as 1 a notation inherited from the multiplicative identity. If a group is abelian, then one may choose to denote the group operation by  $+$  and the identity element by 0.

The result of the group operation may depend on the order of the operands. In other words, the result of combining element  $a$  with element  $b$  need not yield the same result as combining element  $b$  with element  $a$ , so the equation  $a \cdot b = b \cdot a$  may not be true for every two elements  $a$  and  $b$ .

**Definition 2.3** (Abelian Group). *A group  $G$  is called **Abelian** if on top of the four group axioms it also satisfies the axiom of commutativity:*

- *Commutativity:*  $\forall a, b \in G : a \cdot b = b \cdot a$ .

Commutativity always holds in the group of integers under addition, because  $a + b = b + a$  for any two integers (commutativity of addition). The symmetry group is an example of a group that is not abelian.

## 2.2 Fields

**Definition 2.4** (Field). An **(algebraic) field** is a triple  $(K, +, \cdot)$ , where  $K$  is a set and  $+, \cdot$  are maps  $K \times K \rightarrow K$  satisfying the following axioms:

- $(K, +)$  is an abelian group, i.e.:
  - i) *Closure:*  $\forall a, b \in K : a + b \in K$ .
  - ii) *Associativity:*  $\forall a, b, c \in K : (a + b) + c = a + (b + c)$ .
  - iii) *Neutral Element:*  $\exists 0 \in K : \forall a \in K : a + 0 = 0 + a = a$ .
  - iv) *Inverse Element:*  $\forall a \in K : \exists -a \in K : a + (-a) = (-a) + a = 0$ .
  - v) *Commutativity:*  $\forall a, b \in K : a + b = b + a$ .
- $(K^*, \cdot)$ , where  $K^* := K \setminus \{0\}$ , is an abelian group, i.e.:
  - vi) *Closure:*  $\forall a, b \in K^* : a \cdot b \in K^*$ .
  - vii) *Associativity:*  $\forall a, b, c \in K^* : (a \cdot b) \cdot c = a \cdot (b \cdot c)$ .
  - viii) *Neutral Element:*  $\exists 1 \in K^* : \forall a \in K^* : a \cdot 1 = 1 \cdot a = a$ .
  - ix) *Inverse Element:*  $\forall a \in K^* : \exists a^{-1} \in K^* : a \cdot a^{-1} = a^{-1} \cdot a = 1$ .
  - x) *Commutativity:*  $\forall a, b \in K^* : a \cdot b = b \cdot a$ .
- The maps  $+$  and  $\cdot$  satisfy the distributive property:
  - xi)  $\forall a, b, c \in K : (a + b) \cdot c = a \cdot c + b \cdot c$ .

*Remark 2.1.* In the above definition, we included axiom v for the sake of clarity, but in fact it can be proven starting from the other axioms.

## 2.3 Vector Spaces

**Definition 2.5** (K-Vector Space). Let  $(K, +, \cdot)$  be a field. A **K-vector space**, or **vector space over K** or **linear space over K** is a triple  $(V, \oplus, \odot)$ , where  $V$  is a set and:

$$\begin{aligned}\oplus : V \times V &\rightarrow V \\ \odot : K \times V &\rightarrow V\end{aligned}$$

are maps satisfying the following axioms:

- $(V, \oplus)$  is an abelian group i.e.
  - i) *Closure:*  $\forall v, w \in V : v \oplus w \in V$ .
  - ii) *Associativity:*  $\forall v, w, z \in V : (v \oplus w) \oplus z = v \oplus (w \oplus z)$ .
  - iii) *Neutral Element:*  $\exists 0 \in V : \forall v \in V : v \oplus 0 = 0 \oplus v = v$ .
  - iv) *Inverse Element:*  $\forall v \in V : \exists -v \in V : v \oplus (-v) = (-v) \oplus v = 0$ .
  - v) *Commutativity:*  $\forall v, w \in V : v \oplus w = w \oplus v$ .
- The map  $\odot$  is an action of  $K$  on  $(V, \oplus)$ :
  - vi) *Distributivity Of Scalar Multiplication - Vector Addition:*  $\forall \lambda \in K : \forall v, w \in V : \lambda \odot (v \oplus w) = (\lambda \odot v) \oplus (\lambda \odot w)$ .
  - vii) *Distributivity Of Scalar Multiplication - Field Addition:*  $\forall \lambda, \mu \in K : \forall v \in V : (\lambda + \mu) \odot v = (\lambda \odot v) \oplus (\mu \odot v)$ .

viii) *Compatibility Of Scalar Multiplication - Field Multiplication:*  $\forall \lambda, \mu \in K : \forall v \in V : (\lambda \cdot \mu) \odot v = \lambda \odot (\mu \odot v)$ .

ix) *Neutral Element Of Scalar Multiplication:*  $\forall v \in V : 1 \odot v = v$ .

The elements of a vector space are called *vectors*, while the elements of  $K$  are often called *scalars*, and the map  $\odot$  is called *scalar multiplication*.

### 2.3.1 Linear Maps

As usual by now, we will look at the structure-preserving maps between vector spaces.

**Definition 2.6** (Linear Maps). *Let  $(V, \oplus, \odot)$ ,  $(W, \boxplus, \boxtimes)$  be vector spaces over the same field  $K$  and let  $f: V \rightarrow W$  be a map. We say that  $f$  is a **linear map**, or a **homomorphism** and we denote it as  $f: V \xrightarrow{\sim} W$ , if for all  $v_1, v_2 \in V$  and all  $\lambda \in K$ :*

$$f((\lambda \odot v_1) \oplus v_2) = (\lambda \boxtimes f(v_1)) \boxplus f(v_2)$$

From now on, we will drop the special notation for the vector space operations and suppress the dot for scalar multiplication. For instance, we will write the equation above as  $f(\lambda v_1 + v_2) = \lambda f(v_1) + f(v_2)$ , hoping that this will not cause any confusion.

**Definition 2.7** (Linear Isomorphism). *A bijective linear map (or a bijective homomorphism) is called a **linear isomorphism** of vector spaces.*

**Definition 2.8** (Isomorphic Vector Spaces). *Two vector spaces are said to be **isomorphic** if there exists a linear isomorphism between them. We write  $V \cong_{\text{vec}} W$ .*

Based on the linear maps (a.k.a homomorphisms) and bijective linear maps (a.k.a isomorphisms) we can define three important notions of vector spaces:  $\text{Hom}(V, W)$ ,  $\text{End}(V)$  and  $\text{Aut}(V)$ .

**Definition 2.9** ( $\text{Hom}(V, W)$ ). *Let  $V$  and  $W$  be vector spaces over the same field  $K$ . We define the set  $\text{Hom}(V, W)$  as the set of all linear maps (a.k.a all homomorphisms) between  $V$  and  $W$ :*

$$\text{Hom}(V, W) := \{f \mid f: V \xrightarrow{\sim} W\}$$

$\text{Hom}(V, W)$  can itself be made into a vector space over  $K$  by defining:

$$\begin{aligned} \diamond: \text{Hom}(V, W) \times \text{Hom}(V, W) &\rightarrow \text{Hom}(V, W) \\ (f, g) &\mapsto f \diamond g \end{aligned}$$

where:

$$\begin{aligned} f \diamond g: V &\xrightarrow{\sim} W \\ v &\mapsto (f \diamond g)(v) := f(v) + g(v) \end{aligned}$$

and:

$$\begin{aligned} \diamond: K \times \text{Hom}(V, W) &\rightarrow \text{Hom}(V, W) \\ (\lambda, f) &\mapsto \lambda \diamond f \end{aligned}$$

where:

$$\begin{aligned} \lambda \diamond f: V &\xrightarrow{\sim} W \\ v &\mapsto (\lambda \diamond f)(v) := \lambda f(v) \end{aligned}$$

It is easy to check that both  $f \diamond g$  and  $\lambda \diamond f$  are indeed linear maps from  $V$  to  $W$ . For instance, we

have:

$$\begin{aligned}
(\lambda \diamond f)(\mu v_1 + v_2) &= \lambda f(\mu v_1 + v_2) && \text{(by definition)} \\
&= \lambda(\mu f(v_1) + f(v_2)) && \text{(since } f \text{ is linear)} \\
&= \lambda \mu f(v_1) + \lambda f(v_2) && \text{(by axioms i and iii)} \\
&= \mu \lambda f(v_1) + \lambda f(v_2) && \text{(since } K \text{ is a field)} \\
&= \mu(\lambda \diamond f)(v_1) + (\lambda \diamond f)(v_2)
\end{aligned}$$

so that  $\lambda \diamond f \in \text{Hom}(V, W)$ . One should also check that  $\oplus$  and  $\diamond$  satisfy the vector space axioms.

**Definition 2.10** (Endomorphisms). *Let  $V$  be a vector space. An **endomorphism** of  $V$  is a linear map  $V \rightarrow V$ . In other words an endomorphism is a homomorphism whose domain equals the target.*

**Definition 2.11** ( $\text{End}(V)$ ). *Let  $V$  be a vector space. We define the set  $\text{End}(V)$  as the set of all endomorphisms of  $V$ :*

$$\text{End}(V) := \text{Hom}(V, V)$$

It is easy to show that  $\text{End}(V)$  can again itself be made into a vector space over  $K$ .

**Definition 2.12** (Automorphism). *Let  $V$  be a vector space. An **automorphism** of  $V$  is a linear isomorphism  $V \rightarrow V$ . In other words an automorphism is an endomorphism that is also an isomorphism.*

**Definition 2.13** ( $\text{Aut}(V)$ ). *Let  $V$  be a vector space. We define the set  $\text{Aut}(V)$  as the set of all automorphisms of  $V$ :*

$$\text{Aut}(V) := \{f \in \text{End}(V) \mid f \text{ is an isomorphism}\}$$

*Remark 2.2.* Note that  $\text{Aut}(V)$  **cannot** be made into a vector space. It is however a group under the operation of composition of linear maps.

**Definition 2.14** (Dual Vector Space). *Let  $V$  be a vector space over  $K$ . The **dual** vector space to  $V$  is:*

$$V^* := \text{Hom}(V, K)$$

where  $K$  is considered as a vector space over itself.

The dual vector space to  $V$  is the vector space of linear maps from  $V$  to the underlying field  $K$ , which are variously called *linear functionals*, *covectors*, or *one-forms* on  $V$ . The dual plays a very important role, in that from a vector space and its dual, we will construct the tensor space.

**Definition 2.15** (Bilinear Maps). *Let  $V, W, Z$  be vector spaces over  $K$ . A map  $f: V \times W \rightarrow Z$  is said to be **bilinear** if:*

- $\forall w \in W : \forall v_1, v_2 \in V : \forall \lambda \in K : f(\lambda v_1 + v_2, w) = \lambda f(v_1, w) + f(v_2, w).$
- $\forall v \in V : \forall w_1, w_2 \in W : \forall \lambda \in K : f(v, \lambda w_1 + w_2) = \lambda f(v, w_1) + f(v, w_2).$

*i.e. if the maps  $v \mapsto f(v, w)$ , for any fixed  $w$ , and  $w \mapsto f(v, w)$ , for any fixed  $v$ , are both linear as maps  $V \rightarrow Z$  and  $W \rightarrow Z$ , respectively.*

*Remark 2.3.* Compare this with the definition of a linear map  $f: V \times W \xrightarrow{\sim} Z$ :

$$\forall x, y \in V \times W : \forall \lambda \in K : f(\lambda x + y) = \lambda f(x) + f(y).$$

More explicitly, if  $x = (v_1, w_1)$  and  $y = (v_2, w_2)$ , then:

$$f(\lambda(v_1, w_1) + (v_2, w_2)) = \lambda f((v_1, w_1)) + f((v_2, w_2))$$

A bilinear map out of  $V \times W$  is *not* the same as a linear map out of  $V \times W$ . In fact, bilinearity is just a special kind of non-linearity.

*Example 2.1.*

The map  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  given by  $(x, y) \mapsto x + y$  is linear but not bilinear, while the map  $(x, y) \mapsto xy$  is bilinear but not linear.

We can immediately generalise the above to define *multilinear* maps out of a Cartesian product of vector spaces.

**Definition 2.16** (Tensors). *Let  $V$  be a vector space over  $K$ . A  $(p, q)$ -**tensor**  $T$  on  $V$  is a multilinear map:*

$$T: \underbrace{V^* \times \cdots \times V^*}_{p \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q \text{ copies}} \rightarrow K$$

*Remark 2.4.* By convention, a  $(0, 0)$  on  $V$  is just an element of  $K$ , and hence  $T_0^0 V = K$ .

**Definition 2.17** (Covariant / Contravariant Tensor). *A type  $(p, 0)$  tensor is called a **covariant  $p$ -tensor**, while a tensor of type  $(0, q)$  is called a **contravariant  $q$ -tensor**.*

**Definition 2.18** ( $T_q^p V$ ). *We define the set of all  $(p, q)$ -tensors  $T$  as:*

$$T_q^p V := \underbrace{V \otimes \cdots \otimes V}_{p \text{ copies}} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q \text{ copies}} := \{T \mid T \text{ is a } (p, q)\text{-tensor on } V\}$$

*Remark 2.5.* Note that to define  $T_q^p V$  as a set, we should be careful and invoke the principle of restricted comprehension, i.e. we should say where the  $T$ s are coming from. In general, say we want to build a set of maps  $f: A \rightarrow B$  satisfying some property  $p$ . Recall that the notation  $f: A \rightarrow B$  is hiding the fact that is a relation (indeed, a functional relation), and a relation between  $A$  and  $B$  is a subset of  $A \times B$ . Therefore, we ought to write:

$$\{f \in \mathcal{P}(A \times B) \mid f: A \rightarrow B \text{ and } p(f)\}$$

In the case of  $T_q^p V$  we have:

$$T_q^p V := \{T \in \mathcal{P}(\underbrace{V^* \times \cdots \times V^*}_{p \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q \text{ copies}} \times K) \mid T \text{ is a } (p, q)\text{-tensor on } V\}$$

although we will not write this down every time.

The set  $T_q^p V$  can be equipped with a  $K$ -vector space structure by defining:

$$\begin{aligned} \oplus: T_q^p V \times T_q^p V &\rightarrow T_q^p V \\ (T, S) &\mapsto T \oplus S \end{aligned}$$

and:

$$\begin{aligned} \odot: K \times T_q^p V &\rightarrow T_q^p V \\ (\lambda, T) &\mapsto \lambda \odot T \end{aligned}$$

where  $T \oplus S$  and  $\lambda \odot T$  are defined pointwise, as we did with  $\text{Hom}(V, W)$ .

We now define an important way of obtaining a new tensor from two given ones.

**Definition 2.19** (Tensor Product). *Let  $T \in T_q^p V$  and  $S \in T_s^r V$ . The **tensor product** of  $T$  and  $S$  is the tensor  $T \otimes S \in T_{q+s}^{p+r} V$  defined by:*

$$\begin{aligned} (T \otimes S)(\omega_1, \dots, \omega_p, \omega_{p+1}, \dots, \omega_{p+r}, v_1, \dots, v_q, v_{q+1}, \dots, v_{q+s}) \\ := T(\omega_1, \dots, \omega_p, v_1, \dots, v_q) S(\omega_{p+1}, \dots, \omega_{p+r}, v_{q+1}, \dots, v_{q+s}) \end{aligned}$$

with  $\omega_i \in V^*$  and  $v_i \in V$ .

Some examples are in order.

*Example 2.2.*

- a)  $T_1^0 V := \{T \mid T: V \xrightarrow{\sim} K\} = \text{Hom}(V, K) =: V^*$ . Note that here multilinear is the same as linear since the maps only have one argument.

- b)  $T_1^1 V \equiv V \otimes V^* := \{T \mid T \text{ is a bilinear map } V^* \times V \rightarrow K\}$ . We claim that this is the same as  $\text{End}(V^*)$ . Indeed, given  $T \in V \otimes V^*$ , we can construct  $\widehat{T} \in \text{End}(V^*)$  as follows:

$$\begin{aligned}\widehat{T}: V^* &\xrightarrow{\sim} V^* \\ \omega &\mapsto T(-, \omega)\end{aligned}$$

where, for any fixed  $\omega$ , we have:

$$\begin{aligned}T(-, \omega): V &\xrightarrow{\sim} K \\ v &\mapsto T(v, \omega)\end{aligned}$$

The linearity of both  $\widehat{T}$  and  $T(-, \omega)$  follows immediately from the bilinearity of  $T$ . Hence  $T(-, \omega) \in V^*$  for all  $\omega$ , and  $\widehat{T} \in \text{End}(V^*)$ . This correspondence is invertible, since can reconstruct  $T$  from  $\widehat{T}$  by defining:

$$\begin{aligned}T: V \times V^* &\rightarrow K \\ (v, \omega) &\mapsto T(v, \omega) := (\widehat{T}(\omega))(v)\end{aligned}$$

The correspondence is in fact linear, hence an isomorphism, and thus:

$$T_1^1 V \cong_{\text{vec}} \text{End}(V^*)$$

- c)  $T_1^0 V \stackrel{?}{\cong}_{\text{vec}} V$ : while you will find this stated as true in some physics textbooks, it is in fact *not true* in general.
- d)  $T_1^1 V \stackrel{?}{\cong}_{\text{vec}} \text{End}(V)$ : This is also not true in general.
- e)  $(V^*)^* \stackrel{?}{\cong}_{\text{vec}} V$ : This only holds if  $V$  is finite-dimensional (we will define the dimensions of a vector space in the next section).

### 2.3.2 Basis Of Vector Spaces

Given a vector space without any additional structure, the only notion of basis that we can define is a so-called Hamel basis. In order to do so, we first need to define the notion of “span”.

**Definition 2.20** (Span). *Given a vector space  $V$  over a field  $K$ , the span of a set  $S$  of vectors  $\{s_1, \dots, s_N\}$  of  $V$  is defined to be the set of all finite linear combinations of elements (vectors) of  $S$ :*

$$\text{span}_K(S) := \left\{ \sum_{i=1}^n \lambda^i s_i \mid \lambda^i \in K, s_i \in S, n \geq 1 \right\}$$

Now we are ready to define the so called “Hamel basis”.

**Definition 2.21** (Hamel Basis). *Let  $(V, +, \cdot)$  be a vector space over  $K$ . A subset  $\mathcal{B} \subseteq V$  is called a **Hamel basis** for  $V$  if:*

- Every finite subset  $\{b_1, \dots, b_N\}$  of  $\mathcal{B}$  is linearly independent, i.e:

$$\sum_{i=1}^N \lambda^i b_i = 0 \implies \lambda^1 = \dots = \lambda^N = 0$$

- The span of  $\mathcal{B}$  can recreate the whole  $V$ , i.e:

$$V = \text{span}_K(\mathcal{B}) \implies \forall v \in V : \exists v^1, \dots, v^M \in K : \exists b_1, \dots, b_M \in \mathcal{B} : v = \sum_{i=1}^M v^i b_i$$

*Remark 2.6.* Once we have a basis  $\mathcal{B}$ , the expansion of  $v \in V$  in terms of elements of  $\mathcal{B}$  is, in fact, unique. Hence we can meaningfully speak of the *components* of  $v$  in the basis  $\mathcal{B}$ .

*Remark 2.7.* Note that we have been using superscripts for the elements of  $K$ , and these should not be confused with exponents.

The following characterisation of a Hamel basis is often useful.

**Proposition 2.1.** *Let  $V$  be a vector space and  $\mathcal{B}$  a Hamel basis of  $V$ . Then  $\mathcal{B}$  is a minimal spanning and maximal independent subset of  $V$ , i.e., if  $S \subseteq V$ , then:*

- $\text{span}(S) = V \Rightarrow |S| \geq |\mathcal{B}|$ .
- $S$  is linearly independent  $\Rightarrow |S| \leq |\mathcal{B}|$ .

**Definition 2.22** (Dimension Of Vector Space). *Let  $V$  be a vector space. The **dimension** of  $V$  is  $\dim V := |\mathcal{B}|$ , where  $\mathcal{B}$  is a Hamel basis for  $V$ .*

Even though we will not prove it, it is the case that every Hamel basis for a given vector space has the same cardinality, and hence the notion of dimension is well-defined.

**Proposition 2.2.** *If  $\dim V < \infty$  and  $S \subseteq V$ , then we have the following:*

- If  $\text{span}_K(S) = V$  and  $|S| = \dim V$ , then  $S$  is a Hamel basis of  $V$ .
- If  $S$  is linearly independent and  $|S| = \dim V$ , then  $S$  is a Hamel basis of  $V$ .

**Theorem 2.1.** *If  $\dim V < \infty$ , then  $(V^*)^* \cong_{\text{vec}} V$ .*

*Remark 2.8.* Note that while we need the concept of basis to state this result (since we require  $\dim V < \infty$ ), the isomorphism that we have constructed is independent of any choice of basis.

*Remark 2.9.* While a choice of basis often simplifies things, when defining new objects it is important to do so without making reference to a basis. If we do define something in terms of a basis (e.g. the dimension of a vector space), then we have to check that the thing is well-defined, i.e. it does not depend on which basis we choose.

If  $V$  is finite-dimensional, then  $V^*$  is also finite-dimensional and  $V \cong_{\text{vec}} V^*$ . Moreover, given a basis  $\mathcal{B}$  of  $V$ , there is a spacial basis of  $V^*$  associated to  $\mathcal{B}$ .

**Definition 2.23** (Dual Basis). *Let  $V$  be a finite-dimensional vector space with basis  $\mathcal{B} = \{e_1, \dots, e_{\dim V}\}$ . The **dual basis** to  $\mathcal{B}$  is the unique basis  $\mathcal{B}' = \{\epsilon^1, \dots, \epsilon^{\dim V}\}$  of  $V^*$  such that:*

$$\forall 1 \leq i, j \leq \dim V : \quad \epsilon^i(e_j) = \delta_j^i := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

*Remark 2.10.* If  $V$  is finite-dimensional, then  $V$  is isomorphic to both  $V^*$  and  $(V^*)^*$ . In the case of  $V^*$ , an isomorphism is given by sending each element of a basis  $\mathcal{B}$  of  $V$  to a different element of the dual basis  $\mathcal{B}'$ , and then extending linearly to  $V$ . You will (and probably already have) read that a vector space is *canonically* isomorphic to its double dual, but *not* canonically isomorphic to its dual, because an arbitrary choice of basis on  $V$  is necessary in order to provide an isomorphism.

Finally by using a basis (and its dual) we can define the components of a tensor as follows.

**Definition 2.24** (Components Of A Tensor). *Let  $V$  be a finite-dimensional vector space over  $K$  with basis  $\mathcal{B} = \{e_1, \dots, e_{\dim V}\}$  and dual basis  $\{\epsilon^1, \dots, \epsilon^{\dim V}\}$  and let  $T \in T_q^p V$ . We define the **components** of  $T$  in the basis  $\mathcal{B}$  to be the numbers:*

$$T^{a_1 \dots a_p}_{b_1 \dots b_q} := T(\epsilon^{a_1}, \dots, \epsilon^{a_p}, e_{b_1}, \dots, e_{b_q}) \in K$$

where  $1 \leq a_i, b_j \leq \dim V$ .

Just as with vectors, the components completely determine the tensor. Indeed, we can reconstruct the tensor from its components by using the basis:

$$T = \underbrace{\sum_{a_1=1}^{\dim V} \dots \sum_{b_q=1}^{\dim V}}_{p+q \text{ sums}} T^{a_1 \dots a_p}_{b_1 \dots b_q} e_{a_1} \otimes \dots \otimes e_{a_p} \otimes \epsilon^{b_1} \otimes \dots \otimes \epsilon^{b_q}$$

where the  $e_{a_i}$ s are understood as elements of  $T_0^1 V \cong_{\text{vec}} V$  and the  $\epsilon^{b_i}$ s as elements of  $T_1^0 V \cong_{\text{vec}} V^*$ . Note that each summand is a  $(p, q)$ -tensor and the (implicit) multiplication between the components and the tensor product is the scalar multiplication in  $T_q^p V$ .

## Notational Conventions

From now on, we will employ the Einstein's summation convention, which consists in suppressing the summation sign when the indices to be summed over each appear once as a subscript and once as a superscript in the same term. For example, we write:

$$v = v^a e_a, \quad \omega = \omega_a \epsilon^a \quad \text{and} \quad T = T^{ab}_c e_a \otimes e_b \otimes \epsilon^c$$

instead of:

$$v = \sum_{a=1}^d v^a e_a, \quad \omega = \sum_{a=1}^d \omega_a \epsilon^a \quad \text{and} \quad T = \sum_{a=1}^d \sum_{b=1}^d \sum_{c=1}^d T^{ab}_c e_a \otimes e_b \otimes \epsilon^c$$

Indices that are summed over are called *dummy indices*. they always appear in pairs and clearly it doesn't matter which particular letter we choose to denote them, provided it doesn't already appear in the expression. Indices that are not summed over are called *free indices*. expressions containing free indices represent multiple expressions, one for each value of the free indices. free indices must match on both sides of an equation. The ranges over which the indices run are usually understood and not written out.

The convention on which indices go upstairs and which downstairs (which we have already been using) is that:

- The basis vectors of  $V$  carry downstairs indices.
- The basis vectors of  $V^*$  carry upstairs indices.
- All other placements are enforced by the Einstein's summation convention.

For example, since the components of a vector must multiply the basis vectors and be summed over, the Einstein's summation convention requires that they carry upstairs indices.

*Example 2.3.*

Using the summation convention, we have:

- $\epsilon^a(v) = \epsilon^a(v^b e_b) = v^b \epsilon^a(e_b) = v^b \delta_b^a = v^a.$
- $\omega(e_b) = (\omega_a \epsilon^a)(e_b) = \omega_a \epsilon^a(e_b) = \omega_b.$
- $\omega(v) = \omega_a \epsilon^a(v^b e_b) = \omega_a v^a.$

where  $v \in V$ ,  $\omega \in V^*$ ,  $\{e_i\}$  is a basis of  $V$  and  $\{\epsilon^j\}$  is the dual basis to  $\{e_i\}$ .

*Remark 2.11.* The Einstein's summation convention should only be used when dealing with linear spaces and multilinear maps. The reason for this is the following. Consider a map  $\phi: V \times W \rightarrow Z$ , and let  $v = v^i e_i \in V$  and  $w = w^j \tilde{e}_j \in W$ . Then we have:

$$\phi(v, w) = \phi \left( \sum_{i=1}^d v^i e_i, \sum_{j=1}^{\tilde{d}} w^j \tilde{e}_j \right) = \sum_{i=1}^d \sum_{j=1}^{\tilde{d}} \phi(v^i e_i, w^j \tilde{e}_j) = \sum_{i=1}^d \sum_{j=1}^{\tilde{d}} v^i w^j \phi(e_i, \tilde{e}_j)$$

Note that by suppressing the greyed out summation signs, the second and third term above are indistinguishable. But this is only true if  $\phi$  is bilinear! Hence the summation convention should not be used (at least, not without extra care) in other areas of mathematics.

## Matrix Representation

Having chosen a basis for  $V$  and the dual basis for  $V^*$ , it is very tempting to think of  $v = v^i e_i \in V$  and  $\omega = \omega_i \epsilon^i \in V^*$  as  $d$ -tuples of numbers. In order to distinguish them, one may choose to write vectors as *columns* of numbers and covectors as *rows* of numbers:

$$v = v^i e_i \quad \longleftrightarrow \quad v \doteq \begin{pmatrix} v^1 \\ \vdots \\ v^d \end{pmatrix}$$

and:

$$\omega = \omega_i \epsilon^i \quad \rightsquigarrow \quad \omega \hat{=} (\omega_1, \dots, \omega_d)$$

Given  $\phi \in \text{End}(V) \cong_{\text{vec}} T_1^1 V$ , recall that we can write  $\phi = \phi^i_j e_i \otimes \epsilon^j$ , where  $\phi^i_j := \phi(\epsilon^i, e_j)$  are the components of  $\phi$  with respect to the chosen basis. It is then also very tempting to think of  $\phi$  as a square array of numbers:

$$\phi = \phi^i_j e_i \otimes \epsilon^j \quad \rightsquigarrow \quad \phi \hat{=} \begin{pmatrix} \phi^1_1 & \phi^1_2 & \cdots & \phi^1_d \\ \phi^2_1 & \phi^2_2 & \cdots & \phi^2_d \\ \vdots & \vdots & \ddots & \vdots \\ \phi^d_1 & \phi^d_2 & \cdots & \phi^d_d \end{pmatrix}$$

The convention here is to think of the  $i$  index on  $\phi^i_j$  as a *row index*, and of  $j$  as a *column index* (we cannot stress enough that this is pure convention). Hence, once we start using the “matrix representation” (although technically it shouldn’t be called representation since as we will see the word “representation” means something else), we can then express all the linear maps  $\phi$  of  $V$  (a.k.a all the elements of  $\text{End}(V)$ ) as  $n \times n$  matrices. This coincides with the usual picture we have in physics, where all the vectors are represented by a column vector of size  $n$  and all the linear transformations are represented by  $n \times n$  matrices that act on  $v$  and produce another vector  $w$  (hence the  $\text{End}(V)$ ).

Going one step further, notice that not all matrices have an inverse. This coincides with the fact that not all linear maps have an inverse. Since  $\text{End}(V)$  contains all linear maps, it also contains maps that are not linear isomorphisms (a.k.a maps that are not bijections, a.k.a maps that do not have an inverse). However, if we restrict ourselves more, from linear maps to linear isomorphisms then we move from  $\text{End}(V)$  to  $\text{Aut}(V) := \{\phi \in \text{End}(V) \mid \phi \text{ is an isomorphism}\}$ . And if we switch again to the “matrix representation”, now we are dealing with matrices that do have an inverse. We call the set of all these matrices “General Linear Group” and we denote by  $GL(V)$  (we can indeed equip this set with matrix multiplication and show that it is closed under the operation, hence the “group” in the name). From there we can then restrict our transformations even more and then we can get for example the “Special Linear Group” denoted by  $SL(V)$  etc...

For the sake of completeness, let us make a final note that uses the concept of the determinant that we introduce in the next section. Another way to say that a map is an isomorphism is to say that the determinant of the map  $\det \phi \neq 0$ . This condition is a so-called *open condition*, meaning that  $GL(V)$  can be identified with an open subset of  $V$ , from which it then inherits a smooth structure and hence the inverse. By using this we can write  $GL(V) = \text{Aut}(V) = \{\phi \in \text{End}(V) \mid \det \phi \neq 0\}$ . Or in other words since automorphisms are linear isomorphisms between a space and itself,  $GL(V) = \{\phi : V \xrightarrow{\sim} V \mid \det \phi \neq 0\}$ , which coincides with the “matrix representation” of  $GL(V)$  where a matrix has an inverse only when its determinant is non vanishing.

*Example 2.4.*

If  $\dim V < \infty$ , then we have  $\text{End}(V) \cong_{\text{vec}} T_1^1 V$ . Explicitly, if  $\phi \in \text{End}(V)$ , we can think of  $\phi \in T_1^1 V$ , using the same symbol, as:

$$\phi(\omega, v) := \omega(\phi(v))$$

Hence the components of  $\phi \in \text{End}(V)$  are  $\phi^a_b := \epsilon^a(\phi(e_b))$ .

Similarly,  $\omega(v) = \omega_m v^m$  can be thought of as the *dot product*  $\omega \cdot v \equiv \omega^T v$ , and:

$$\phi(v, w) = w_a \phi^a_b v^b \quad \rightsquigarrow \quad \omega^T \phi v$$

The last expression is could mislead you into thinking that the transpose is a “good” notion, but in fact it is not. It is very bad notation. It almost pretends to be basis independent, but it is not at all.

Now consider  $\phi, \psi \in \text{End}(V)$ . Let us determine the components of  $\phi \circ \psi$ . We have:

$$\begin{aligned}
(\phi \circ \psi)^a_b &:= (\phi \circ \psi)(\epsilon^a, e_b) \\
&:= \epsilon^a((\phi \circ \psi)(e_b)) \\
&= \epsilon^a((\phi(\psi(e_b)))) \\
&= \epsilon^a(\phi(\psi^m_b e_m)) \\
&= \psi^m_b \epsilon^a(\phi(e_m)) \\
&:= \psi^m_b \phi^a_m
\end{aligned}$$

The multiplication in the last line is the multiplication in the field  $K$ , and since that's commutative, we have  $\psi^m_b \phi^a_m = \phi^a_m \psi^m_b$ . However, in light of the convention introduced in the previous remark, the latter is preferable. Indeed, if we think of the superscripts as row indices and of the subscripts as column indices, then  $\phi^a_m \psi^m_b$  is the entry in row  $a$ , column  $b$ , of the matrix product  $\phi\psi$ .

The moral of the story is that you should try your best *not* to think of vectors, covectors and tensors as arrays of numbers. Instead, always try to understand them from the abstract, intrinsic, component-free point of view.

### 2.3.3 Change Of Basis

Let  $V$  be a vector space over  $K$  with  $d = \dim V < \infty$  and let  $\{e_1, \dots, e_d\}$  be a basis of  $V$ . Consider a new basis  $\{\tilde{e}_1, \dots, \tilde{e}_d\}$ . Since the elements of the new basis are also elements of  $V$ , we can expand them in terms of the old basis. We have:

$$\tilde{e}_a = \sum_{b=1}^d A^b_a e_b = A^b_a e_b$$

for some  $A^b_a \in K$ . Similarly, we have:

$$e_a = \sum_{m=1}^d B^m_a \tilde{e}_m = B^m_a \tilde{e}_m$$

for some  $B^m_a \in K$ . It is a standard linear algebra result that the matrices  $A$  and  $B$ , with entries  $A^b_a$  and  $B^m_a$  respectively, are invertible and, in fact,  $A^{-1} = B$ . Note that in index notation, the equation  $AB = I$  reads  $A^a_m B^m_b = \delta^a_b$ .

We now investigate how the components of vectors and covectors change under a change of basis.

a) Let  $v = v^a e_a = \tilde{v}^a \tilde{e}_a \in V$ . Then:

$$v^a = \epsilon^a(v) = \epsilon^a(\tilde{v}^b \tilde{e}_b) = \tilde{v}^b \epsilon^a(\tilde{e}_b) = \tilde{v}^b \epsilon^a(A^m_b e_m) = A^m_b \tilde{v}^b \epsilon^a(e_m) = A^a_b \tilde{v}^b$$

b) Let  $\omega = \omega_a \epsilon^a = \tilde{\omega}_a \tilde{\epsilon}^a \in V^*$ . Then:

$$\omega_a := \omega(e_a) = \omega(B^m_a \tilde{e}_m) = B^m_a \omega(\tilde{e}_m) = B^m_a \tilde{\omega}_m$$

Summarising, for  $v \in V$ ,  $\omega \in V^*$  and  $\tilde{e}_a = A^b_a e_b$ , we have:

$$\begin{aligned}
v^a &= A^a_b \tilde{v}^b & \omega_a &= B^b_a \tilde{\omega}_b \\
\tilde{v}^a &= B^a_b v^b & \tilde{\omega}_a &= A^b_a \omega_b
\end{aligned}$$

The result for tensors is a combination of the above, depending on the type of tensor.

c) Let  $T \in T^p_p V$ . Then:

$$T^{a_1 \dots a_p}_{b_1 \dots b_q} = A^{a_1}_{m_1} \dots A^{a_p}_{m_p} B^{n_1}_{b_1} \dots B^{n_q}_{b_q} \tilde{T}^{m_1 \dots m_p}_{n_1 \dots n_q}$$

i.e. the upstairs indices transform like vector indices, and the downstairs indices transform like covector indices.

Coming back (once again) to the “matrix representation”, let’s see now one of the biggest misunderstandings that might come up when we want to perform a change of basis for tensors.

Recall that, if  $\phi \in T_1^1 V$ , then we can arrange the components  $\phi^a_b$  in matrix form:

$$\phi = \phi^a_b e_a \otimes \epsilon^b \quad \longleftrightarrow \quad \phi \hat{=} \begin{pmatrix} \phi^1_1 & \phi^1_2 & \cdots & \phi^1_d \\ \phi^2_1 & \phi^2_2 & \cdots & \phi^2_d \\ \vdots & \vdots & \ddots & \vdots \\ \phi^d_1 & \phi^d_2 & \cdots & \phi^d_d \end{pmatrix}$$

Similarly, if we have  $g \in T_2^0 V$ , its components are  $g_{ab} := g(e_a, e_b)$  and we can write:

$$g = g_{ab} \epsilon^a \otimes \epsilon^b \quad \longleftrightarrow \quad g \hat{=} \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1d} \\ g_{21} & g_{22} & \cdots & g_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ g_{d1} & g_{d2} & \cdots & g_{dd} \end{pmatrix}$$

Needless to say that these two objects could not be more different if they tried. Indeed:

- $\phi$  is an endomorphism of  $V$ . the first index in  $\phi^a_b$  transforms like a vector index, while the second index transforms like a covector index.
- $g$  is a *bilinear form* on  $V$ . both indices in  $g_{ab}$  transform like covector indices.

In linear algebra, you may have seen the two different transformation laws for these objects:

$$\phi \rightarrow A^{-1} \phi A \quad \text{and} \quad g \rightarrow A^T g A$$

where  $A$  is the change of basis matrix. However, once we fix a basis, the matrix representations of these two objects are indistinguishable. It is then very tempting to think that what we can do with a matrix, we can just as easily do with another matrix.

For instance, the eigenvector equation reads:

$$A_m^a v^m = \lambda v^a$$

It is clear from the indices that  $A_m^a$  is the components of an endomorphism. If we attempt to write a similar equation for a bilinear form we will simply fail:

$$g_{mn} v^m = \lambda \underbrace{v_n}_{(?)}$$

Indeed, the eigenvectors equation holds only for endomorphisms, while bilinear forms carry the so called “signature”.

Another example: if we have a rule to calculate the determinant of a square matrix, we should be able to apply it to both of the above matrices. However, the notion of determinant is *only* defined for endomorphisms. The only way to see this is to give a basis-independent definition, i.e. a definition that does not involve the “components of a matrix”. Let’s do that!

### 2.3.4 Determinants

In your previous course on linear algebra, you may have met the determinant of a square matrix as a number calculated by applying a mysterious rule. Using the mysterious rule, you may have shown, with a lot of work, that for example, if we exchange two rows or two columns, the determinant changes sign. But, as we have seen, matrices are the result of pure convention. Hence, one more polemic remark is in order.

We will need some preliminary definitions (we will define  $n$ -forms in a more proper way in next chapters, so for now do not spend a lot of time on them).

**Definition 2.25** (*n-form*). Let  $V$  be a  $d$ -dimensional vector space. An  $n$ -**form** on  $V$  is a  $(0, n)$ -tensor  $\omega$  that is totally antisymmetric, i.e:

$$\forall \pi \in S_n : \omega(v_1, v_2, \dots, v_n) = \text{sgn}(\pi) \omega(v_{\pi(1)}, v_{\pi(2)}, \dots, v_{\pi(n)})$$

Note that a 0-form is a scalar, and a 1-form is a covector. A  $d$ -form is also called a *top form*, and one can show that for two top forms  $\omega$  and  $\omega'$  the following holds:

$$\forall \omega, \omega' \in \Lambda^d V : \exists c \in K : \omega = c \omega'$$

i.e. there is essentially only one top form on  $V$ , up to a scalar factor.

**Definition 2.26** (Choice Of Volume). A choice of top form on  $V$  is called a choice of **volume form** on  $V$ . A vector space with a chosen volume form is then called a vector space with volume.

This terminology is due to the next definition.

**Definition 2.27** (Volume). Let  $\dim V = d$  be the dimension of vector space  $V$  and let  $v_1, \dots, v_d \in V$ , be  $d$  vectors in  $V$ . Then the **volume** spanned by  $v_1, \dots, v_d$  is:

$$\text{vol}(v_1, \dots, v_d) := \omega(v_1, \dots, v_d)$$

where  $\omega$  is the (chosen) top form.

Intuitively, the antisymmetry condition on  $\omega$  makes sure that  $\text{vol}(v_1, \dots, v_d)$  is zero whenever the set  $\{v_1, \dots, v_d\}$  is not linearly independent. Indeed, in that case  $v_1, \dots, v_d$  could only span a  $(d - 1)$ -dimensional hypersurface in  $V$  at most, which should have 0 volume.

*Remark 2.12.* You may have rightfully thought that the notion of volume would require some extra structure on  $V$ , such as a notion of length or angles, and hence an inner product. But instead, we only need a top form.

We are finally ready to define the determinant.

**Definition 2.28** (Determinant). Let  $V$  be a  $d$ -dimensional vector space and let  $\phi \in \text{End}(V) \cong_{\text{vec}} T_1^1 V$ . The determinant of  $\phi$  is:

$$\det \phi := \frac{\omega(\phi(e_1), \dots, \phi(e_d))}{\omega(e_1, \dots, e_d)}$$

for some top form  $\omega$  and some basis  $\{e_1, \dots, e_d\}$  of  $V$ .

The first thing we need to do is to check that this is well-defined. That  $\det \phi$  is independent of the choice of  $\omega$  is clear, since if  $\omega, \omega'$  are top forms, then there is a  $c \in K$  such that  $\omega = c \omega'$ , and hence:

$$\frac{\omega(\phi(e_1), \dots, \phi(e_d))}{\omega(e_1, \dots, e_d)} = \frac{c \omega'(\phi(e_1), \dots, \phi(e_d))}{c \omega'(e_1, \dots, e_d)}.$$

The independence from the choice of basis is more cumbersome to show, but it does hold, and thus  $\det \phi$  is well-defined.

It is very important to notice that  $\phi$  needs to be an endomorphism because we need to apply  $\omega$  to  $\phi(e_1), \dots, \phi(e_d)$ , and thus  $\phi$  needs to output a vector. Which means that the determinant can only be defined for endomorphisms.

Of course, under the identification of  $\phi$  as a matrix, this definition coincides with the usual definition of determinant, and all your favourite results about determinants can be derived from it. However once we switch to “matrix representation” as we said one is not able to distinguish between an endomorphism  $\phi \in T_1^1 V$  and the so called “bilinear form”  $g \in T_2^0 V$ , hence one might think that they can calculate the determinant of the second guy. Let’s see why such a determinant is not well defined.

In your linear algebra course, you may have shown the the determinant is basis-independent as follows: if  $A$  denotes the change of basis matrix, then:

$$\det(A^{-1} \phi A) = \det(A^{-1}) \det(\phi) \det(A) = \det(A^{-1} A) \det(\phi) = \det(\phi)$$

since scalars commute, and  $\det(A^{-1}A) = \det(I) = 1$ .

Recall that the transformation rule for a bilinear form  $g$  under a change of basis is  $g \rightarrow A^T g A$ . The determinant of  $g$  then transforms as:

$$\det(A^T g A) = \det(A^T) \det(g) \det(A) = (\det A)^2 \det(g)$$

i.e. it not invariant under a change of basis. It is not a well-defined object, and thus we should not use it.

We will later meet quantities  $X$  that transform as:

$$X \rightarrow \frac{1}{(\det A)^2} X$$

under a change of basis, and hence they are also not well-defined. However, we obviously have:

$$\det(g)X \rightarrow \frac{(\det A)^2}{(\det A)^2} \det(g)X = \det(g)X$$

so that the product  $\det(g)X$  is a well-defined object. It seems that two wrongs make a right!

In order to make this mathematically precise, we will have to introduce *principal fibre bundles*. Using them, we will be able to give a bundle definition of tensor and of *tensor densities* which are, loosely speaking, quantities that transform with powers of  $\det A$  under a change of basis. We will see all of that in later chapters.

## 2.4 Rings

**Definition 2.29** (Ring). A **ring** is a triple  $(R, +, \cdot)$ , where  $R$  is a set and  $+, \cdot : R \times R \rightarrow R$  are maps satisfying the following axioms:

- $(R, +)$  is an abelian group:
  - i) Closure:  $\forall a, b \in R : a + b \in R$ .
  - ii) Associativity:  $\forall a, b, c \in R : (a + b) + c = a + (b + c)$ .
  - iii) Neutral Element:  $\exists 0 \in R : \forall a \in R : a + 0 = 0 + a = a$ .
  - iv) Inverse Element:  $\forall a \in R : \exists -a \in R : a + (-a) = (-a) + a = 0$ .
  - v) Commutativity:  $\forall a, b \in R : a + b = b + a$ .
- The operation  $\cdot$  is closed and associative in  $R^* := R \setminus \{0\}$ :
  - vi) Closure:  $\forall a, b \in R^* : a \cdot b \in R^*$ .
  - vii) Associativity:  $\forall a, b, c \in R^* : (a \cdot b) \cdot c = a \cdot (b \cdot c)$ .
- The maps  $+$  and  $\cdot$  satisfy the distributive properties:
  - viii)  $\forall a, b, c \in R : (a + b) \cdot c = a \cdot c + b \cdot c$ .
  - ix)  $\forall a, b, c \in R : a \cdot (b + c) = a \cdot b + a \cdot c$ .

Note that since  $\cdot$  is not required to be commutative, axioms viii and ix are both necessary. In the case of fields where  $\cdot$  was commutative, ix followed from viii and commutativity of  $\cdot$ .

**Definition 2.30** (Commutative / Unital / Division Rings). A ring  $(R, +, \cdot)$  is said to be:

- **Commutative** if  $\forall a, b \in R : a \cdot b = b \cdot a$ .
- **Unital** if  $\exists 1 \in R : \forall a \in R : 1 \cdot a = a \cdot 1 = a$ .
- A **division** (or **skew**) ring if it is unital and:

$$\forall a \in R \setminus \{0\} : \exists a^{-1} \in R \setminus \{0\} : a \cdot a^{-1} = a^{-1} \cdot a = 1$$

In a unital ring, an element for which there exists a multiplicative inverse is said to be a *unit*. The set of units of a ring  $R$  is denoted by  $R^*$  (not to be confused with the vector space dual) and forms a group under multiplication. Then,  $R$  is a division ring iff  $R^* = R \setminus \{0\}$ .

*Example 2.5.*

The sets  $\mathbb{Z}$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$ , and  $\mathbb{C}$  are all rings under the usual operations. They are also all fields, except  $\mathbb{Z}$ .

In general, if  $(A, +, \cdot, \bullet)$  is an algebra, then  $(A, +, \bullet)$  is a ring.

## 2.5 Modules

**Definition 2.31** (*R-Module*). Let  $(R, +, \cdot)$  be a unital ring. An ***R-module*** is a triple  $(M, \oplus, \odot)$  where  $M$  is a set and:

$$\oplus: M \times M \rightarrow M$$

$$\odot: R \times M \rightarrow M$$

are maps satisfying the following axioms:

- $(M, \oplus)$  is an abelian group i.e.:
  - i) *Closure*:  $\forall m, n \in M : m \oplus n \in M$ .
  - ii) *Associativity*:  $\forall m, n, s \in M : (m \oplus n) \oplus s = m \oplus (n \oplus s)$ .
  - iii) *Neutral Element*:  $\exists 0 \in M : \forall m \in M : m \oplus 0 = 0 \oplus m = m$ .
  - iv) *Inverse Element*:  $\forall m \in M : \exists -m \in M : m \oplus (-m) = (-m) \oplus m = 0$ .
  - v) *Commutativity*:  $\forall m, n \in M : m \oplus n = n \oplus m$ .
- The map  $\odot$  is an action of  $R$  on  $(M, \oplus)$ :
  - vi) *Distributivity Of Scalar Multiplication - Vector Addition*:  $\forall r \in R : \forall m, n \in M : r \odot (m \oplus n) = (r \odot m) \oplus (r \odot n)$ .
  - vii) *Distributivity Of Scalar Multiplication - Field Addition*:  $\forall r, s \in K : \forall m \in V : (r + s) \odot m = (r \odot m) \oplus (s \odot m)$ .
  - viii) *Compatibility Of Scalar Multiplication - Field Multiplication*:  $\forall r, s \in R : \forall m \in M : (r \cdot s) \odot m = r \odot (s \odot m)$ .
  - ix) *Neutral Element Of Scalar Multiplication*:  $\forall m \in M : 1 \odot m = m$ .

So, modules are simply vector spaces over rings instead of fields. For this reason, most definitions we had for vector spaces carry over unaltered to modules.

*Example 2.6.*

Any ring  $R$  is trivially a module over itself, in the sense that every field  $K$  is a vector space over itself.

In the following, we will usually denote  $\oplus$  by  $+$  and suppress the  $\odot$ , as we did with vector spaces.

**Definition 2.32** (*Direct Sum Of Modules*). The ***direct sum*** of two  $R$ -modules  $M$  and  $N$  is the  $R$ -module  $M \oplus N$ , which has  $M \times N$  as its underlying set and operations (inherited from  $M$  and  $N$ ) defined component-wise.

Note that while we have been using  $\oplus$  to temporarily distinguish two “plus-like” operations in different spaces, the symbol  $\oplus$  is the standard notation for the direct sum.

**Definition 2.33** (*Finitely Generated / Free / Projective Modules*). An  $R$ -module  $M$  is said to be:

- ***Finitely generated*** if it has a finite generating set.
- ***Free*** if it has a basis.
- ***Projective*** if it is a direct summand of a free  $R$ -module  $F$ , i.e.:

$$M \oplus Q = F$$

for some  $R$ -module  $Q$ .

*Example 2.7.*

Clearly, every free module is also projective.

**Definition 2.34** (R-Linear Maps). *Let  $M$  and  $N$  be two  $R$ -modules. A map  $f: M \rightarrow N$  is said to be an  **$R$ -linear map** if:*

$$\forall r \in R : \forall m_1, m_2 \in M : f(rm_1 + m_2) = rf(m_1) + f(m_2)$$

where it should be clear which operations are in  $M$  and which in  $N$ .

**Definition 2.35** (Module Isomorphisms). *A bijective  $R$ -linear map is said to be a **module isomorphism**.*

**Definition 2.36** (Isomorphic Modules). *Two modules are said to be **isomorphic** if there exists a module isomorphism between them. We write  $M \cong_{\text{mod}} N$ .*

**Proposition 2.3.** *If a finitely generated module  $R$ -module  $F$  is free, and  $d \in \mathbb{N}$  is the cardinality of a finite basis, then:*

$$F \cong_{\text{mod}} \underbrace{R \oplus \cdots \oplus R}_{d \text{ copies}} =: R^d$$

One can show that if  $R^d \cong_{\text{mod}} R^{d'}$ , then  $d = d'$  and hence, the concept of dimension is well-defined for finitely generated, free modules.

**Theorem 2.2.** *Let  $P, Q$  be finitely generated (projective) modules over a commutative ring  $R$ . Then:*

$$\text{Hom}_R(P, Q) := \{\phi: P \xrightarrow{\sim} Q \mid \phi \text{ is } R\text{-linear}\}$$

*is again a finitely generated (projective)  $R$ -module, with operations defined pointwise.*

The proof is exactly the same as with vector spaces. As an example, we can use this to define the dual of a module.

### 2.5.1 Basis Of Modules

The key fact that sets modules apart from vector spaces is that, unlike a vector space, an  $R$ -module need not have a basis, unless  $R$  is a division ring. This is actually a well-known theorem that we will state but not prove.

**Theorem 2.3.** *If  $D$  is a division ring, then any  $D$ -module  $V$  admits a basis.*

**Corollary 2.1.** *Every vector space has a basis, since any field is also a division ring.*

## 2.6 Algebras

**Definition 2.37** (Algebra). *Let  $K$  be a field, and let  $A$  be a vector space over  $K$  equipped with an additional bilinear map (called binary operation or product)  $\bullet: A \times A \rightarrow A$ . The quadruple  $(A, +, \cdot, \bullet)$  is called an **algebra** over a field  $K$ .*

**Definition 2.38** (Associative / Unital / Commutative Algebra). *An algebra  $(A, +, \cdot, \bullet)$  is said to be:*

i) **Associative** if  $\forall v, w, z \in A : v \bullet (w \bullet z) = (v \bullet w) \bullet z$ .

ii) **Unital** if  $\exists \mathbf{1} \in A : \forall v \in V : \mathbf{1} \bullet v = v \bullet \mathbf{1} = v$ .

iii) **Commutative** or **abelian** if  $\forall v, w \in A : v \bullet w = w \bullet v$ .

**Definition 2.39** (Derivation). *Let  $A$  and  $B$  be algebras. A **derivation** on  $A$  is a linear map  $D: A \xrightarrow{\sim} B$  satisfying the Leibniz rule:*

$$D(v \bullet_A w) = D(v) \bullet_B w +_B v \bullet_B D(w)$$

for all  $v, w \in A$ .

# Chapter 3

## Lie Algebras

We already defined in the previous chapter that an algebra is a vector space  $A$  with an additional bilinear map (called binary operation or product)  $\bullet: A \times A \rightarrow A$ . A very important class of algebras, that we will also see later, are the so-called Lie algebras, in which the product  $v \bullet w$  is called “Lie bracket” and denoted as  $[v, w]$ . In general Lie algebras are just a very specific class of algebras, hence we might have them introduced in the previous chapter under “algebras”. However, since they are so important, and lengthy, we will introduce them separately in their own chapter.

Lie algebras are closely related to Lie groups, which are groups that are also smooth manifolds: any Lie group gives rise to a Lie algebra, which is its tangent space at the identity. Conversely, to any finite-dimensional Lie algebra over real or complex numbers, there is a corresponding connected Lie group unique up to finite coverings. This correspondence allows one to study the structure and classification of Lie groups in terms of Lie algebras (we will see all of that as we proceed in the notes).

In physics, Lie groups appear as symmetry groups of physical systems, and their Lie algebras (tangent vectors near the identity) may be thought of as infinitesimal symmetry motions. Thus Lie algebras and their representations are used extensively in physics, notably in quantum mechanics and particle physics.

### 3.1 Basic Definitions

**Definition 3.1** (Lie Algebra). A **Lie algebra**  $A$  over a field  $K$  is an algebra whose product  $[-, -]$ , called Lie bracket, satisfies:

- i) *Bilinearity*:  $A \times A \rightarrow A$ :  $[av + w, z] = a[v, w] + [v, z]$ .
- ii) *Antisymmetry*:  $\forall v \in A$ :  $[v, v] = 0$ .
- iii) *The Jacobi identity*:  $\forall v, w, z \in A$ :  $[v, [w, z]] + [w, [z, v]] + [z, [v, w]] = 0$ .

Note that the zeros above represent the additive identity element in  $A$ , not the zero scalar.

Some remarks are in order.

*Remark 3.1.* The antisymmetry condition immediately implies  $[v, w] = -[w, v]$  for all  $v, w \in A$  since:

$$[v + w, v + w] = [v, v] + [v, w] + [w, v] + [w, w] = [v, w] + [w, v] = 0 \implies [v, w] = -[w, v]$$

*Remark 3.2.* Notice that the Lie bracket is not defined as the usual commutator  $[v, w] = vw - wv$ , but is defined very abstractly by the 3 conditions. In other words, anything that satisfies these 3 conditions can be defined as a Lie bracket. Of course one example is the commutator (you can check it yourself)

*Remark 3.3.* Notice that we specifically defined the Lie algebra on top of a field  $K$ . One can construct an algebra over a ring, by imposing all the axioms on a module instead of a vector space. However, in this notes we will stick with Lie algebras on top of a vector space, and more specifically on top of a complex vector space (i.e where the  $K$  field is the complex and to the real numbers), since they are more related to our purposes. In general, same definitions apply for an algebra over a ring with the appropriate changes when needed.

Now let's give some examples of Lie algebras.

*Example 3.1.*

The usual cross product between vectors  $u \times w$  in  $\mathbb{R}^3$  can be shown that satisfies all the requirements for a Lie bracket, hence the vector space  $\mathbb{R}^3$  equipped with the cross product  $[u, w] = u \times w$  is actually a Lie algebra.

*Example 3.2.*

Let  $V$  be a vector space. Recall that we defined the set  $\text{End}(V)$  as the set of all endomorphisms of  $V$ , i.e the set of all linear maps that send  $V$  back to itself. Now we define the following Lie bracket:

$$\begin{aligned} [-, -]: \text{End}(V) \times \text{End}(V) &\rightarrow \text{End}(V) \\ (\phi, \psi) &\mapsto [\phi, \psi] := \phi \circ \psi - \psi \circ \phi \end{aligned}$$

It is instructive to check that this is actually a Lie bracket. Hence,  $(\text{End}(V), +, \cdot, [-, -])$  is a Lie algebra. In this case, the Lie bracket is typically called the *commutator*. (Remember that after having chosen a basis then we can "represent" the elements of  $\text{End}(V)$  as  $n \times n$  matrices over a field  $K$ , with their commutator  $[v, w] = vw - wv$  where here the composition is the usual matrix multiplication).

As usual we can define the concept of homomorphism and isomorphism in the level of Lie algebras.

**Definition 3.2** (Lie Algebra Homomorphism). *A map  $\phi$  between two Lie algebras that preserves both the vector space structure and the bracket structure is called a **Lie algebra homomorphism**.*

**Definition 3.3** (Homomorphic Lie Algebras). *Two Lie algebras over the same field  $K$  are said to be **homomorphic** if there exists a lie algebra homomorphism between them.*

**Definition 3.4** (Lie Algebra Isomorphism). *A bijective Lie algebra homomorphism is called a **Lie algebra isomorphism**.*

**Definition 3.5** (Isomorphic Lie Algebras). *Two Lie algebras over the same field  $K$  are said to be **isomorphic** if there exists a lie algebra isomorphism.*

In what follows we will make heavy use of the following notation that we will give in a form of definition.

**Definition 3.6** (Bracket). *Given two subsets  $A, B$  of a Lie algebra  $L$  we define the **bracket** of these two subsets  $[A, B]$  as the subset defined by the span of all commutators  $[x, y]$  where  $x \in A$  and  $y \in B$ , i.e:*

$$[A, B] := \text{span}_K(\{[x, y] \in L \mid x \in A \text{ and } y \in B\})$$

In other words is just the set of all commutators  $[x, y]$  where  $x \in A$  and  $y \in B$ .

Now let's give some very basic definitions of Lie algebras.

**Definition 3.7** (Abelian Lie Algebra). *A Lie algebra  $L$  is said to be **abelian** if  $\forall x, y \in L : [x, y] = 0$  or equivalently in bracket notation  $[L, L] = 0$ , where  $0$  denotes the trivial Lie algebra  $\{0\}$ .*

Abelian Lie algebras are highly non-interesting as Lie algebras: since the bracket is identically zero, it may as well not be there. On top of that, the vanishing of the bracket implies that, given any two abelian Lie algebras, every linear isomorphism between their underlying vector spaces is automatically a Lie algebra isomorphism. Therefore, for each  $n \in \mathbb{N}$ , there is (up to isomorphism) only one abelian  $n$ -dimensional Lie algebra.

**Definition 3.8** (Subalgebra). *We say  $L'$  is a **subalgebra** of  $L$  if  $L'$  is a vector subspace of  $L$  and  $\forall x, y \in L' : [x, y] \in L'$  or equivalently in bracket notation  $[L', L'] \in L'$ .*

One can prove that if  $A, B$  are Lie subalgebras of a Lie algebra  $L$  over  $K$ , then the bracket  $[A, B]$  is again a Lie subalgebra of  $L$ .

**Definition 3.9** (Ideal). *An **ideal**  $I$  of a Lie algebra  $L$  is a Lie subalgebra such that  $\forall x \in I : \forall y \in L : [x, y] \in I$  or equivalently in bracket notation  $[I, L] \subseteq I$ .*

Note that no matter the Lie algebra, we can show that:  $[0, L] = 0 \subseteq 0$  and  $[L, L] \subseteq L$  hence both  $0$  and  $L$  are always ideals of any Lie algebra.

*Remark 3.4.* Recall from the definition of an algebra (any algebra) that the operation (or product) of the algebra  $\bullet: A \times A \rightarrow A$  is a bilinear map with no need to be surjective. This means that applying the operation to every possible element of the algebra does not guarantee that will give us back the whole algebra (but it does guarantee to give us back a subalgebra). In other words,  $[L, L] \subseteq L$  and not  $[L, L] = L$ .

**Definition 3.10** (Trivial Ideals). *The ideals  $0$  and  $L$  are called the **trivial ideals** of  $L$ .*

**Definition 3.11** (Simple Lie Algebra). *A Lie algebra  $L$  is said to be **simple** if it is non-abelian and it contains no non-trivial ideals.*

**Definition 3.12** (Semi-Simple Lie Algebra). *A Lie algebra  $L$  is said to be **semi-simple** if it contains no non-trivial abelian ideals.*

*Remark 3.5.* Note that any simple Lie algebra is also semi-simple. The requirement that a simple Lie algebra be non-abelian is due to the 1-dimensional abelian Lie algebra, which would otherwise be the only simple Lie algebra which is not semi-simple.

**Definition 3.13** (Derived Subalgebra). *Let  $L$  be a Lie algebra. The Lie subalgebra  $L' := [L, L]$  is called the **derived subalgebra** of  $L$ .*

Hence, once we have a Lie algebra we can compute the derived subalgebra  $L' := [L, L]$ . However since  $L'$  is by itself an algebra we can compute its own derived subalgebra  $L'' := [L', L']$  (which is the derived subalgebra of the derived subalgebra of  $L$ ). And of course we can go on forever.

**Definition 3.14** (Derived Series). *The sequence  $L \supseteq L' \supseteq L'' \supseteq \dots \supseteq L^{(n)} \supseteq \dots$  of Lie subalgebras is called the **derived series** of  $L$  usually denoted by  $L^{(n)}$ .*

**Definition 3.15** (Solvable Lie Algebra). *A Lie algebra  $L$  is **solvable** if there exists  $k \in \mathbb{N}$  such that  $L^{(k)} = 0$ .*

Recall that the direct sum of vector spaces  $V \oplus W$  has  $V \times W$  as its underlying set and operations defined componentwise.

**Definition 3.16** (Direct Sum Of Lie Algebras). *Let  $L_1$  and  $L_2$  be Lie algebras. The **direct sum**  $L_1 \oplus_{\text{Lie}} L_2$  has  $L_1 \oplus L_2$  as its underlying vector space and Lie bracket defined as:*

$$[x_1 + x_2, y_1 + y_2]_{L_1 \oplus_{\text{Lie}} L_2} := [x_1, y_1]_{L_1} + [x_2, y_2]_{L_2}$$

*for all  $x_1, y_1 \in L_1$  and  $x_2, y_2 \in L_2$ . Alternatively, by identifying  $L_1$  and  $L_2$  with the subspaces  $L_1 \oplus 0$  and  $0 \oplus L_2$  of  $L_1 \oplus L_2$  respectively, we require:*

$$[L_1, L_2]_{L_1 \oplus_{\text{Lie}} L_2} = 0$$

*In the following, we will drop the “Lie” subscript and understand  $\oplus$  to mean  $\oplus_{\text{Lie}}$  whenever the summands are Lie algebras.*

There is a weaker notion than the direct sum, defined only for Lie algebras.

**Definition 3.17** (Semi-Direct Sum Of Lie Algebras). *Let  $R$  and  $L$  be Lie algebras. The **semi-direct sum**  $R \oplus_s L$  has  $R \oplus L$  as its underlying vector space and Lie bracket satisfying:*

$$[R, L]_{R \oplus_s L} \subseteq R$$

*i.e.  $R$  is an ideal of  $R \oplus_s L$ .*

We are now ready to state Levi’s decomposition theorem.

**Theorem 3.1** (Levi). *Any finite-dimensional complex Lie algebra  $L$  can be decomposed as:*

$$L = R \oplus_s (L_1 \oplus \dots \oplus L_n)$$

*where  $R$  is a solvable Lie algebra and  $L_1, \dots, L_n$  are simple Lie algebras.*

As of today, no general classification of solvable Lie algebras is known, except for some special cases (e.g. in low dimensions). In contrast, the finite dimensional, simple, complex Lie algebras have been classified completely.

**Proposition 3.1.** *A Lie algebra is semi-simple if, and only if, it can be expressed as a direct sum of simple Lie algebras.*

Hence, the simple Lie algebras are the basic building blocks from which one can build any semi-simple Lie algebra. Then, by Levi's theorem, the classification of simple Lie algebras easily extends to a classification of all semi-simple Lie algebras.

In order to do computations, it is useful to introduce a basis  $\{e_i\}$  on  $L$ . Recall that an algebra is nothing else but a vector space with an extra operation. Hence, we can simply pick a basis  $\{e_i\}$  on the vector space, and then examine how the Lie bracket behaves when we plug in, not any random element of algebra (i.e of the vector space) but specifically the elements of the basis.

**Definition 3.18** (Structure Constants). *Let  $L$  be a Lie algebra over  $K$  and let  $\{e_i\}$  be a basis of the underlying vector space. Then, we have:*

$$[e_i, e_j] = C_{ij}^k e_k$$

for some  $C_{ij}^k \in K$ . The numbers  $C_{ij}^k$  are called the **structure constants** of  $L$  with respect to the basis  $\{e_i\}$ .

*Remark 3.6.* Since the operation of the algebra  $\bullet: A \times A \rightarrow A$ , sends two elements of the algebra to an element of the algebra, this can be translated as sending two elements of the vector space to an element of the vector space, i.e  $[e_i, e_j] = v \in V$  for some fixed  $i$  and  $j$ . However since the final result  $v$  is again an element of the vector space it can also be expressed as a linear combination of the basis  $v = v^k e_k$ . This  $v^k$  is actually the structure constants (again for some fixed  $i$  and  $j$ , if we do not fix them we have to include them on the  $v^k$  hence we obtain  $v^k \rightarrow C_{ij}^k$ ). This is why it is guaranteed that the structure constants  $C_{ij}^k \in K$  exist.

In terms of the structure constants, the anti-symmetry of the Lie bracket reads:

$$[e_i, e_j] = -[e_j, e_i] \implies C_{ij}^k e_k = -C_{ji}^k e_k \implies C_{ij}^k = -C_{ji}^k$$

while after some trivial calculations one can show that the Jacobi identity becomes:

$$C_{im}^n C_{jk}^m + C_{jm}^n C_{ki}^m + C_{km}^n C_{ij}^m = 0$$

## 3.2 The Adjoint Map & The Killing Form

**Definition 3.19** (Adjoint Map). *Let  $L$  be a Lie algebra over  $K$  and let  $x \in L$ . The **adjoint map** with respect to  $x$  is the  $K$ -linear map:*

$$\begin{aligned} \text{ad}_x: L &\xrightarrow{\sim} L \\ y &\mapsto \text{ad}_x(y) := [x, y] \end{aligned}$$

The linearity of  $\text{ad}_x$  follows from the linearity of the bracket in the second argument, while the linearity in the first argument of the bracket implies that the map:

$$\begin{aligned} \text{ad}: L &\xrightarrow{\sim} \text{End}(L) \\ x &\mapsto \text{ad}(x) := \text{ad}_x \end{aligned}$$

itself is also linear. In fact, more is true. Recall that  $\text{End}(L)$  is a Lie algebra with bracket:

$$[\phi, \psi] := \phi \circ \psi - \psi \circ \phi$$

Then, we have the following.

**Proposition 3.2.** *The map  $\text{ad}: L \xrightarrow{\sim} \text{End}(L)$  is a Lie algebra homomorphism.*

*Proof.*

It remains to check that  $\text{ad}$  preserves the brackets. Let  $x, y, z \in L$ . Then:

$$\begin{aligned}
\text{ad}_{[x,y]}(z) &:= [[x, y], z] && \text{(definition of ad)} \\
&= -[[y, z], x] - [[z, x], y] && \text{(Jacobi's identity)} \\
&= [x, [y, z]] - [y, [x, z]] && \text{(anti-symmetry)} \\
&= \text{ad}_x(\text{ad}_y(z)) - \text{ad}_y(\text{ad}_x(z)) \\
&= (\text{ad}_x \circ \text{ad}_y - \text{ad}_y \circ \text{ad}_x)(z) \\
&= [\text{ad}_x, \text{ad}_y](z)
\end{aligned}$$

Hence, we have  $\text{ad}([x, y]) = [\text{ad}(x), \text{ad}(y)]$ . □

By choosing a basis for the vector space, we can express the adjoint map in terms of components with respect to the basis as follows. Start by noting that:

$$\begin{aligned}
\text{ad}: L &\xrightarrow{\sim} \text{End}(L) \\
x &\mapsto \text{ad}(x) := \text{ad}_x
\end{aligned}$$

which means that  $\text{ad}_x$  is an element of  $\text{End}(L)$  hence an endomorphism of  $L$ . Recall that for any vector space  $V$ :  $\text{End}(V) \cong_{\text{vec}} T_1^1 V$  which means that if  $\phi \in \text{End}(V)$ , we can think of  $\phi \in T_1^1 V$ , using the same symbol, as  $\phi(\omega, v) := \omega(\phi(v))$  hence the components of  $\phi \in \text{End}(V)$  are  $\phi_b^a := \epsilon^a(\phi(e_b))$ .

So, in our case, let  $\{e_i\}$  and  $\{\epsilon^i\}$  be a basis and its dual basis of the underlying vector space of a Lie algebra  $L$ . Then:

$$\begin{aligned}
(\text{ad}_{e_i})^k_j &:= \epsilon^k(\text{ad}_{e_i}(e_j)) \\
&= \epsilon^k([e_i, e_j]) \\
&= \epsilon^k(C^m_{ij} e_m) \\
&= C^m_{ij} \epsilon^k(e_m) \\
&= C^k_{ij}
\end{aligned}$$

In other words, the adjoint map represents the structure constants without the need of choosing a basis.

**Definition 3.20** (Killing Form). *Let  $L$  be a Lie algebra over  $K$ . The **Killing form** on  $L$  is the  $K$ -bilinear map:*

$$\begin{aligned}
\kappa: L \times L &\rightarrow K \\
(x, y) &\mapsto \kappa(x, y) := \text{tr}(\text{ad}_x \circ \text{ad}_y)
\end{aligned}$$

where  $\text{tr}$  is the usual trace on the vector space  $\text{End}(L)$ .

Note that the Killing form is not a “form” in the sense that we defined previously. In fact, since  $L$  is finite-dimensional, the trace is cyclic and thus  $\kappa$  is symmetric, i.e:

$$\forall x, y \in L : \kappa(x, y) = \kappa(y, x)$$

An important property of  $\kappa$  is its associativity with respect to the bracket.

**Proposition 3.3.** *Let  $L$  be a Lie algebra. For any  $x, y, z \in L$ , we have:*

$$\kappa([x, y], z) = \kappa(x, [y, z])$$

*Proof.*

This follows easily from the fact that  $\text{ad}$  is a homomorphism.

$$\begin{aligned}
\kappa([x, y], z) &:= \text{tr}(\text{ad}_{[x, y]} \circ \text{ad}_z) \\
&= \text{tr}([\text{ad}_x, \text{ad}_y] \circ \text{ad}_z) \\
&= \text{tr}((\text{ad}_x \circ \text{ad}_y - \text{ad}_y \circ \text{ad}_x) \circ \text{ad}_z) \\
&= \text{tr}(\text{ad}_x \circ \text{ad}_y \circ \text{ad}_z) - \text{tr}(\text{ad}_y \circ \text{ad}_x \circ \text{ad}_z) \\
&= \text{tr}(\text{ad}_x \circ \text{ad}_y \circ \text{ad}_z) - \text{tr}(\text{ad}_x \circ \text{ad}_z \circ \text{ad}_y) \\
&= \text{tr}(\text{ad}_x \circ (\text{ad}_y \circ \text{ad}_z - \text{ad}_z \circ \text{ad}_y)) \\
&= \text{tr}(\text{ad}_x \circ [\text{ad}_y, \text{ad}_z]) \\
&= \text{tr}(\text{ad}_x \circ \text{ad}_{[y, z]}) \\
&=: \kappa(x, [y, z])
\end{aligned}$$

where we used the cyclicity of the trace.  $\square$

As we did for the adjoint map we can also express the Killing form in terms of components with respect to a basis.

Recall from linear algebra that if  $V$  is finite-dimensional, for any  $\phi \in \text{End}(V)$  we have  $\text{tr}(\phi) = \Phi^k_k$ , where  $\Phi$  is the matrix representing the linear map in any basis. Also, recall that the matrix representing  $\phi \circ \psi$  is the product  $\Phi\Psi$ . Using these, by letting  $\{e_i\}$  and  $\{\varepsilon^i\}$  be a basis and its dual basis of the underlying vector space of a Lie algebra  $L$  we have:

$$\begin{aligned}
\kappa_{ij} &:= \kappa(e_i, e_j) \\
&= \text{tr}(\text{ad}_{e_i} \circ \text{ad}_{e_j}) \\
&= (\text{ad}_{e_i} \circ \text{ad}_{e_j})^k_k \\
&= (\text{ad}_{e_i})^m_k (\text{ad}_{e_j})^k_m \\
&= C^m_{ik} C^k_{jm}
\end{aligned}$$

where we used the same notation for the linear maps and their matrices.

We can use  $\kappa$  to give a further equivalent characterisation of semi-simplicity.

**Proposition 3.4** (Cartan's criterion). *A Lie algebra  $L$  is semi-simple if, and only if, the Killing form  $\kappa$  is non-degenerate, i.e:*

$$(\forall y \in L : \kappa(x, y) = 0) \Rightarrow x = 0$$

Hence, if  $L$  is semi-simple, then  $\kappa$  is a pseudo inner product on  $L$ . Recall the following definition from linear algebra.

**Definition 3.21.** *A linear map  $\phi: V \xrightarrow{\sim} V$  is said to be symmetric with respect to the pseudo inner product  $B(-, -)$  on  $V$  if:*

$$\forall v, w \in V : B(\phi(v), w) = B(v, \phi(w))$$

*If, instead, we have:*

$$\forall v, w \in V : B(\phi(v), w) = -B(v, \phi(w))$$

*then  $\phi$  is said to be anti-symmetric with respect to  $B$ .*

The associativity property of  $\kappa$  with respect to the bracket can be restated by saying that, for any  $z \in L$ , the linear map  $\text{ad}_z$  is anti-symmetric with respect to  $\kappa$ , i.e:

$$\forall x, y \in L : \kappa(\text{ad}_z(x), y) = -\kappa(x, \text{ad}_z(y))$$

### 3.3 The Fundamental Roots & The Weyl Group

We will now focus on finite-dimensional semi-simple complex Lie algebras, whose classification hinges on the existence of a special type of subalgebra.

**Definition 3.22** (Cartan Subalgebra). Let  $L$  be a  $d$ -dimensional Lie algebra. A **Cartan subalgebra**  $H$  of  $L$  is a maximal Lie subalgebra of  $L$  with the following property: there exists a basis  $\{h_1, \dots, h_r\}$  of  $H$  which can be extended to a basis  $\{h_1, \dots, h_r, e_1, \dots, e_{d-r}\}$  of  $L$  such that  $e_1, \dots, e_{d-r}$  are eigenvectors of  $\text{ad}(h)$  for any  $h \in H$ , i.e:

$$\forall h \in H : \exists \lambda_\alpha(h) \in \mathbb{C} : \text{ad}(h)e_\alpha = \lambda_\alpha(h)e_\alpha$$

for each  $1 \leq \alpha \leq d - r$ .

The basis  $\{h_1, \dots, h_r, e_1, \dots, e_{d-r}\}$  is known as a *Cartan-Weyl basis* of  $L$ . Of course, we would like to know when we can find such a subalgebra.

**Theorem 3.2.** Let  $L$  be a finite-dimensional semi-simple complex Lie algebra. Then:

- i)  $L$  possesses a Cartan subalgebra.
- ii) All Cartan subalgebras of  $L$  have the same dimension, called the rank of  $L$ .
- iii) Any of Cartan subalgebra  $H$  of  $L$  is abelian, i.e  $[H, H] = 0$ .

Note that we can think of the  $\lambda_\alpha$  appearing above as a map  $\lambda_\alpha : H \rightarrow \mathbb{C}$ . Moreover, for any  $z \in \mathbb{C}$  and  $h, h' \in H$ , we have:

$$\begin{aligned} \lambda_\alpha(zh + h')e_\alpha &= \text{ad}(zh + h')e_\alpha \\ &= [zh + h', e_\alpha] \\ &= z[h, e_\alpha] + [h', e_\alpha] \\ &= z\lambda_\alpha(h)e_\alpha + \lambda_\alpha(h')e_\alpha \\ &= (z\lambda_\alpha(h) + \lambda_\alpha(h'))e_\alpha \end{aligned}$$

Hence  $\lambda_\alpha$  is a  $\mathbb{C}$ -linear map  $\lambda_\alpha : H \xrightarrow{\sim} \mathbb{C}$ , and thus  $\lambda_\alpha \in H^*$ .

**Definition 3.23** (Roots). The maps  $\lambda_1, \dots, \lambda_{d-r} \in H^*$  are called the **roots** of  $L$ .

**Definition 3.24** (Root Set). The collection of the roots of an algebra:

$$\Phi := \{\lambda_\alpha \mid 1 \leq \alpha \leq d - r\} \subseteq H^*$$

is called the **root set** of  $L$ .

One can show that if  $\lambda_\alpha$  were the zero map, then we would have  $e_\alpha \in H$ . Thus, we must have  $0 \notin \Phi$ . Note that a consequence of the anti-symmetry of each  $\text{ad}(h)$  with respect to the Killing form  $\kappa$  is that:

$$\lambda \in \Phi \Rightarrow -\lambda \in \Phi$$

Hence  $\Phi$  is not a linearly independent subset of  $H^*$ .

**Definition 3.25** (Fundamental Roots). A set of **fundamental roots**  $\Pi := \{\pi_1, \dots, \pi_f\}$  is a subset  $\Pi \subseteq \Phi$  such that :

- a)  $\Pi$  is a linearly independent subset of  $H^*$ .
- b) For each  $\lambda \in \Phi$ , there exist  $n_1, \dots, n_f \in \mathbb{N}$  and  $\varepsilon \in \{+1, -1\}$  such that:

$$\lambda = \varepsilon \sum_{i=1}^f n_i \pi_i$$

Since  $n_i \in \mathbb{N}$  this means that they are all positive numbers (as they should be by the definition of a basis). By also picking an  $\varepsilon \in \{+1, -1\}$  to be either  $+1$  or  $-1$ , we are able, no matter the choice of fundamental roots, to obtain the opposite signed ones. That way, observe that, for any  $\lambda \in \Phi$ , the coefficients of  $\pi_1, \dots, \pi_f$  in the expansion above always have the same sign. We can write the last equation more concisely as  $\lambda \in \text{span}_{\varepsilon, \mathbb{N}}(\Pi)$  where in general  $\text{span}_{\varepsilon, \mathbb{N}}(\Pi) \neq \text{span}_{\mathbb{Z}}(\Pi)$ .

**Theorem 3.3.** Let  $L$  be a finite-dimensional semi-simple complex Lie algebra. Then:

i) A set  $\Pi \subseteq \Phi$  of fundamental roots always exists.

ii) We have  $\text{span}_{\mathbb{C}}(\Pi) = H^*$ , that is,  $\Pi$  is a basis of  $H^*$ .

**Corollary 3.1.** We have  $|\Pi| = r$ , where  $r$  is the rank of  $L$ .

*Proof.*

Since  $\Pi$  is a basis,  $|\Pi| = \dim H^* = \dim H = r$ . □

We would now like to use  $\kappa$  to define a pseudo inner product on  $H^*$ . We know from linear algebra that a pseudo inner product  $B(-, -)$  on a finite-dimensional vector space  $V$  over  $K$  induces a linear isomorphism:

$$\begin{aligned} i: V &\xrightarrow{\sim} V^* \\ v &\mapsto i(v) := B(v, -) \end{aligned}$$

which can be used to define a pseudo inner product  $B^*(-, -)$  on  $V^*$  as:

$$\begin{aligned} B^*: V^* \times V^* &\rightarrow K \\ (\phi, \psi) &\mapsto B^*(\phi, \psi) := B(i^{-1}(\phi), i^{-1}(\psi)) \end{aligned}$$

We would like to apply this to the restriction of  $\kappa$  to the Cartan subalgebra. However, a pseudo inner product on a vector space is not necessarily a pseudo inner product on a subspace, since the non-degeneracy condition may fail when considered on a subspace.

**Proposition 3.5.** The restriction of  $\kappa$  to  $H$  is a pseudo inner product on  $H$ .

*Proof.*

Bilinearity and symmetry are automatically satisfied. It remains to show that  $\kappa$  is non-degenerate on  $H$ .

i) Let  $\{h_1, \dots, h_r, e_{r+1}, \dots, e_d\}$  be a Cartan-Weyl basis of  $L$  and let  $\lambda_\alpha \in \Phi$ . Then:

$$\begin{aligned} \lambda_\alpha(h_j)\kappa(h_i, e_\alpha) &= \kappa(h_i, \lambda_\alpha(h_j)e_\alpha) \\ &= \kappa(h_i, [h_j, e_\alpha]) \\ &= \kappa([h_i, h_j], e_\alpha) \\ &= \kappa(0, e_\alpha) \\ &= 0 \end{aligned}$$

Since  $\lambda_\alpha \neq 0$ , there is some  $h_j$  such that  $\lambda_\alpha(h_j) \neq 0$  and hence:

$$\kappa(h_i, e_\alpha) = 0$$

By linearity, we have  $\kappa(h, e_\alpha) = 0$  for any  $h \in H$  and any  $e_\alpha$ .

ii) Let  $h \in H \subseteq L$ . Since  $\kappa$  is non-degenerate on  $L$ , we have:

$$(\forall x \in L : \kappa(h, x) = 0) \Rightarrow h = 0$$

Expand  $x \in L$  in the Cartan-Weyl basis as:

$$x = h' + e$$

where  $h' := x^i h_i$  and  $e := x^\alpha e_\alpha$ . Then, we have:

$$\kappa(h, x) = \kappa(h, h') + x^\alpha \kappa(h, e_\alpha) = \kappa(h, h')$$

Thus, the non-degeneracy condition reads:

$$(\forall h' \in H : \kappa(h, h') = 0) \Rightarrow h = 0$$

which is what we wanted. □

We can now define:

$$\begin{aligned}\kappa^*: H^* \times H^* &\rightarrow \mathbb{C} \\ (\mu, \nu) &\mapsto \kappa^*(\mu, \nu) := \kappa(i^{-1}(\mu), i^{-1}(\nu))\end{aligned}$$

where  $i: H \xrightarrow{\sim} H^*$  is the linear isomorphism induced by  $\kappa$ .

*Remark 3.7.* If  $\{h_i\}$  is a basis of  $H$ , the components of  $\kappa^*$  with respect to the dual basis satisfy :

$$(\kappa^*)^{ij} \kappa_{jk} = \delta_k^i$$

Hence, we can write:

$$\kappa^*(\mu, \nu) = (\kappa^*)^{ij} \mu_i \nu_j$$

where  $\mu_i := \mu(h_i)$ .

We now turn our attention to the real subalgebra  $H_{\mathbb{R}}^* := \text{span}_{\mathbb{R}}(\Pi)$ . Note that we have the following chain of inclusions:

$$\Pi \subseteq \Phi \subseteq \text{span}_{\varepsilon, \mathbb{N}}(\Pi) \subseteq \underbrace{\text{span}_{\mathbb{R}}(\Pi)}_{H_{\mathbb{R}}^*} \subseteq \underbrace{\text{span}_{\mathbb{C}}(\Pi)}_{H^*}$$

The restriction of  $\kappa^*$  to  $H_{\mathbb{R}}^*$  leads to a surprising result.

**Theorem 3.4.** *i) For any  $\alpha, \beta \in H_{\mathbb{R}}^*$ , we have  $\kappa^*(\alpha, \beta) \in \mathbb{R}$ .*

*ii)  $\kappa^*: H_{\mathbb{R}}^* \times H_{\mathbb{R}}^* \rightarrow \mathbb{R}$  is an inner product on  $H_{\mathbb{R}}^*$ .*

This is indeed a surprise! Upon restriction to  $H_{\mathbb{R}}^*$ , instead of being weakened, the non-degeneracy of  $\kappa^*$  gets strengthened to positive definiteness. Now that we have a proper real inner product, we can define some familiar notions from basic linear algebra, such as lengths and angles.

**Definition 3.26** (Length & Angle). *Let  $\alpha, \beta \in H_{\mathbb{R}}^*$ . Then, we define:*

*i) The **length** of  $\alpha$  as  $|\alpha| := \sqrt{\kappa^*(\alpha, \alpha)}$ .*

*ii) The **angle** between  $\alpha$  and  $\beta$  as  $\varphi := \cos^{-1}\left(\frac{\kappa^*(\alpha, \beta)}{|\alpha||\beta|}\right)$ .*

We need one final ingredient for our classification result.

**Definition 3.27** (Weyl Transformation). *For any  $\lambda \in \Phi \subseteq H_{\mathbb{R}}^*$ , define the linear map  $s_{\lambda}$  called a **Weyl transformation**:*

$$\begin{aligned}s_{\lambda}: H_{\mathbb{R}}^* &\xrightarrow{\sim} H_{\mathbb{R}}^* \\ \mu &\mapsto s_{\lambda}(\mu)\end{aligned}$$

where:

$$s_{\lambda}(\mu) := \mu - 2 \frac{\kappa^*(\lambda, \mu)}{\kappa^*(\lambda, \lambda)} \lambda$$

**Definition 3.28** (Weyl Group). *The set:*

$$W := \{s_{\lambda} \mid \lambda \in \Phi\}$$

*is a group under composition of maps, and it is called the **Weyl group**.*

**Theorem 3.5.** *i) The Weyl group  $W$  is generated by the fundamental roots in  $\Pi$ , in the sense that for some  $1 \leq n \leq r$ , with  $r = |\Pi|$ :*

$$\forall w \in W : \exists \pi_1, \dots, \pi_n \in \Pi : w = s_{\pi_1} \circ s_{\pi_2} \circ \dots \circ s_{\pi_n}$$

*ii) Every root can be produced from a fundamental root by the action of  $W$ , i.e:*

$$\forall \lambda \in \Phi : \exists \pi \in \Pi : \exists w \in W : \lambda = w(\pi)$$

iii) The Weyl group permutes the roots, that is:

$$\forall \lambda \in \Phi : \forall w \in W : w(\lambda) \in \Phi$$

### 3.4 Dynkin Diagrams & The Cartan Classification

Consider, for any  $\pi_i, \pi_j \in \Pi$ , the action of the Weyl transformation:

$$s_{\pi_i}(\pi_j) := \pi_j - 2 \frac{\kappa^*(\pi_i, \pi_j)}{\kappa^*(\pi_i, \pi_i)} \pi_i$$

However, since  $s_{\pi_i}(\pi_j) \in \Phi$  and  $\Phi \subseteq \text{span}_{\varepsilon, \mathbb{N}}(\Pi)$  this means that it must be written in terms of the basis as:

$$s_{\pi_i}(\pi_j) \in \Phi = \left( \varepsilon \sum_{i=1}^f n_i \pi_i \right) = C_1 \pi_j + C_2 \pi_i$$

But it is already written in such form since:

$$s_{\pi_i}(\pi_j) = \pi_j - 2 \frac{\kappa^*(\pi_i, \pi_j)}{\kappa^*(\pi_i, \pi_i)} \pi_i = 1 \pi_j + \left( -2 \frac{\kappa^*(\pi_i, \pi_j)}{\kappa^*(\pi_i, \pi_i)} \right) \pi_i$$

and from the first coefficient (a.k.a the number 1) which is positive, we conclude that for all  $1 \leq i \neq j \leq r$  we must have:

$$-2 \frac{\kappa^*(\pi_i, \pi_j)}{\kappa^*(\pi_i, \pi_i)} \in \mathbb{N}$$

**Definition 3.29** (Cartan Matrix). *The **Cartan matrix** of a Lie algebra is the  $r \times r$  matrix  $C$  with entries:*

$$C_{ij} := 2 \frac{\kappa^*(\pi_i, \pi_j)}{\kappa^*(\pi_i, \pi_i)}$$

*Remark 3.8.* The  $C_{ij}$  should not be confused with the structure constants  $C_{ij}^k$ .

**Theorem 3.6.** *To every simple finite-dimensional complex Lie algebra there corresponds a unique Cartan matrix and vice versa (up to relabelling of the basis elements).*

Of course, not every matrix can be a Cartan matrix. For instance, since  $C_{ii} = 2$  (no summation implied), the diagonal entries of  $C$  are all equal to 2, while the off-diagonal entries are either zero or negative. In general,  $C_{ij} \neq C_{ji}$ , so the Cartan matrix is not symmetric, but if  $C_{ij} = 0$ , then necessarily  $C_{ji} = 0$ .

We have thus reduced the problem of classifying the simple finite-dimensional complex Lie algebras to that of finding all the Cartan matrices. This can, in turn, be reduced to the problem of determining all the inequivalent Dynkin diagrams.

**Definition 3.30** (Bond Number). *Given a Cartan matrix  $C$ , the  $ij$ -th **bond number** is:*

$$n_{ij} := C_{ij} C_{ji}$$

Note that we have:

$$\begin{aligned} n_{ij} &= 4 \frac{\kappa^*(\pi_i, \pi_j)}{\kappa^*(\pi_i, \pi_i)} \frac{\kappa^*(\pi_j, \pi_i)}{\kappa^*(\pi_j, \pi_j)} \\ &= 4 \left( \frac{\kappa^*(\pi_i, \pi_j)}{|\pi_i| |\pi_j|} \right)^2 \\ &= 4 \cos^2 \varphi \end{aligned}$$

where  $\varphi$  is the angle between  $\pi_i$  and  $\pi_j$ .

For  $i \neq j$ , the angle  $\varphi$  is neither zero nor  $180^\circ$ , hence  $0 \leq \cos^2 \varphi < 1$ , and therefore:

$$n_{ij} \in \{0, 1, 2, 3\}$$

Since  $C_{ij} \leq 0$  for  $i \neq j$ , the only possibilities are:

$C_{ij}$	$C_{ji}$	$n_{ij}$
0	0	0
-1	-1	1
-1	-2	2
-1	-3	3

Note that while the Cartan matrices are not symmetric, swapping any pair of  $C_{ij}$  and  $C_{ji}$  gives a Cartan matrix which represents the same Lie algebra as the original matrix, with two elements from the Cartan-Weyl basis swapped. This is why we have not included  $(-2, -1)$  and  $(-3, -1)$  in the table above. If  $n_{ij} = 2$  or  $3$ , then the corresponding fundamental roots have different lengths, i.e. either  $|\pi_i| < |\pi_j|$  or  $|\pi_i| > |\pi_j|$ . We also have the following result.

**Proposition 3.6.** *The roots of a simple Lie algebra have, at most, two distinct lengths.*

The redundancy of the Cartan matrices highlighted above is nicely taken care of by considering Dynkin diagrams.

**Definition 3.31** (Dynkin Diagram). *A **Dynkin diagram** associated to a Cartan matrix is constructed as follows:*

1. Draw a circle for every fundamental root in  $\pi_i \in \Pi$ :



2. Draw  $n_{ij}$  lines between the circles representing the roots  $\pi_i$  and  $\pi_j$ :



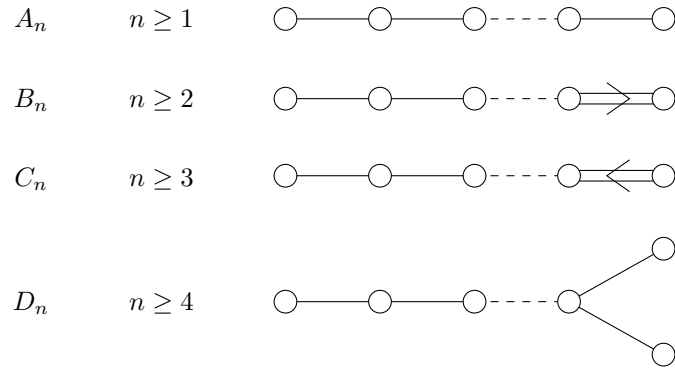
3. If  $n_{ij} = 2$  or  $3$ , draw an arrow on the lines from the longer root to the shorter root:



Dynkin diagrams completely characterise any set of fundamental roots, from which we can reconstruct the entire root set by using the Weyl transformations. The root set can then be used to produce a Cartan-Weyl basis. We are now finally ready to state the much awaited classification theorem.

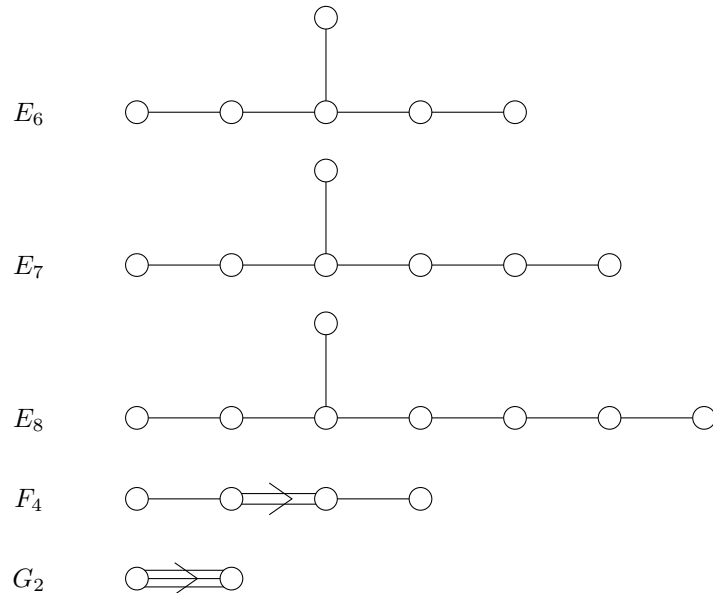
**Theorem 3.7** (Killing, Cartan). *Any simple finite-dimensional complex Lie algebra can be reconstructed from its set of fundamental roots  $\Pi$ , which only come in the following forms:*

- i) There are 4 infinite families:



where the restrictions on  $n$  ensure that we don't get repeated diagrams (the diagram  $D_2$  is excluded since it is disconnected and does not correspond to a simple Lie algebra)

ii) Five exceptional cases:



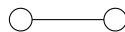
and no other. These are all the possible (connected) Dynkin diagrams.

At last, we have achieved a classification of all simple finite-dimensional complex Lie algebras. The finite-dimensional semi-simple complex Lie algebras are direct sums of simple Lie algebras, and correspond to disconnected Dynkin diagrams whose connected components are the ones listed above.

### 3.5 Application: Reconstruction Of $A_2$ From Its Dynkin Diagram

We have seen how to construct the Dynkin diagram of a Lie algebra. Let us now consider the opposite direction, where we want to retrieve the Lie algebra given a Dynkin diagram. There is no general theory for that, we simply have to follow the opposite procedure of the theory we developed in the previous section, hence we will provide a specific example.

We will start from the  $A_2$  Dynkin diagram:



We immediately see that we have two fundamental roots, i.e.  $\Pi = \{\pi_1, \pi_2\}$ , since there are two circles in the diagram. The bond number is  $n_{12} = 1$ , so the two fundamental roots have the same length. Moreover, by definition:

$$1 = n_{12} = C_{12}C_{21}$$

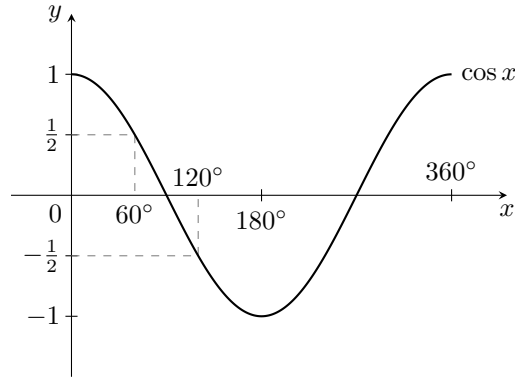
and since the off-diagonal entries of the Cartan matrix are non-positive integers, the only possibility is  $C_{12} = C_{21} = -1$ , so that we have:

$$C = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

To determine the angle  $\varphi$  between  $\pi_1$  and  $\pi_2$ , recall that:

$$1 = n_{12} = 4 \cos^2 \varphi$$

and hence  $|\cos \varphi| = \frac{1}{2}$ . There are two solutions, namely  $\varphi = 60^\circ$  and  $\varphi = 120^\circ$ .



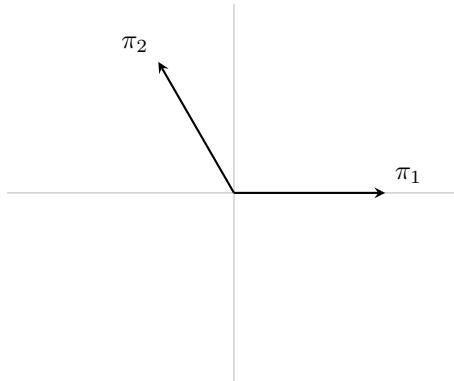
By definition, we have:

$$\cos \varphi = \frac{\kappa^*(\pi_1, \pi_2)}{|\pi_1| |\pi_2|}$$

and therefore:

$$0 > C_{12} = 2 \frac{\kappa^*(\pi_1, \pi_2)}{\kappa^*(\pi_1, \pi_1)} = 2 \frac{|\pi_1| |\pi_2| \cos \varphi}{\kappa^*(\pi_1, \pi_1)} = 2 \frac{|\pi_2|}{|\pi_1|} \cos \varphi$$

It follows that  $\cos \varphi < 0$ , and hence  $\varphi = 120^\circ$ . We can thus plot the two fundamental roots in a plane as follows:



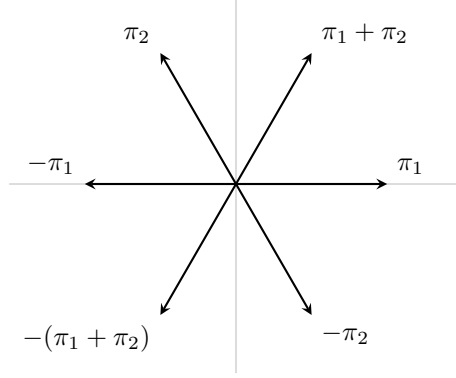
We can determine all the other roots in  $\Phi$  by repeated action of the Weyl group. For instance, we easily find that  $s_{\pi_1}(\pi_1) = -\pi_1$  and  $s_{\pi_2}(\pi_2) = -\pi_2$ . We also have:

$$s_{\pi_1}(\pi_2) = \pi_2 - 2 \frac{\kappa^*(\pi_1, \pi_2)}{\kappa^*(\pi_1, \pi_1)} \pi_1 = \pi_2 - 2(-\frac{1}{2})\pi_1 = \pi_1 + \pi_2$$

Finally, we have  $s_{\pi_1+\pi_2}(\pi_1 + \pi_2) = -(\pi_1 + \pi_2)$ . Any further action by Weyl transformations simply permutes these roots. Hence, we have:

$$\Phi = \{\pi_1, -\pi_1, \pi_2, -\pi_2, \pi_1 + \pi_2, -(\pi_1 + \pi_2)\}$$

and these are all the roots.



Since  $H^* = \text{span}_{\mathbb{C}}(\Pi)$ , we have  $\dim H^* = 2$ , thus the dimension of the Cartan subalgebra is also 2. Since  $|\Phi| = 6$ , we know that any Cartan-Weyl basis of the Lie algebra  $A_2$  must have  $2 + 6 = 8$  elements. Hence, the dimension of  $A_2$  is 8.

To complete our reconstruction of  $A_2$ , we would now like to understand how its bracket behaves. This amounts to finding its structure constants. Note that since  $\dim A_2 = 8$ , the structure constants  $C^k_{ij}$  consist of  $8^3 = 512$  complex numbers (not all unrelated, of course).

Denote by  $\{h_1, h_2, e_3, \dots, e_8\}$  a Cartan-Weyl basis of  $A_2$ , so that  $H = \text{span}_{\mathbb{C}}(\{h_1, h_2\})$  and the  $e_\alpha$  are eigenvectors of every  $h \in H$ . Since  $A_2$  is simple,  $H$  is abelian and hence:

$$[h_1, h_2] = 0 \quad \Rightarrow \quad C^k_{12} = C^k_{21} = 0, \quad \forall 1 \leq k \leq 8$$

To each  $e_\alpha$ , for  $3 \leq \alpha \leq 8$ , there is an associated  $\lambda_\alpha \in \Phi$  such that:

$$\forall h \in H : \text{ad}(h)e_\alpha = \lambda_\alpha(h)e_\alpha$$

In particular, for the basis elements  $h_1, h_2$ :

$$\begin{aligned} [h_1, e_\alpha] &= \text{ad}(h_1)e_\alpha = \lambda_\alpha(h_1)e_\alpha \\ [h_2, e_\alpha] &= \text{ad}(h_2)e_\alpha = \lambda_\alpha(h_2)e_\alpha \end{aligned}$$

so that we have:

$$\begin{aligned} C^1_{1\alpha} &= C^2_{1\alpha} = 0, & C^\alpha_{1\alpha} &= \lambda_\alpha(h_1), & \forall 3 \leq \alpha \leq 8 \\ C^1_{2\alpha} &= C^2_{2\alpha} = 0, & C^\alpha_{2\alpha} &= \lambda_\alpha(h_2), & \forall 3 \leq \alpha \leq 8 \end{aligned}$$

Finally, we need to determine  $[e_\alpha, e_\beta]$ . By using the Jacobi identity, we have:

$$\begin{aligned} [h_i, [e_\alpha, e_\beta]] &= -[e_\alpha, [e_\beta, h_i]] - [e_\beta, [h_i, e_\alpha]] \\ &= -[e_\alpha, -\lambda_\beta(h_i)e_\beta] - [e_\beta, \lambda_\alpha(h_i)e_\alpha] \\ &= \lambda_\beta(h_i)[e_\alpha, e_\beta] + \lambda_\alpha(h_i)[e_\alpha, e_\beta] \\ &= (\lambda_\alpha(h_i) + \lambda_\beta(h_i))[e_\alpha, e_\beta] \end{aligned}$$

that is:

$$\text{ad}(h_i)[e_\alpha, e_\beta] = (\lambda_\alpha(h_i) + \lambda_\beta(h_i))[e_\alpha, e_\beta]$$

If  $\lambda_\alpha + \lambda_\beta \in \Phi$ , we have  $[e_\alpha, e_\beta] = \xi e_\gamma$  for some  $3 \leq \gamma \leq 8$  and  $\xi \in \mathbb{C}$ . Let us label the roots in our previous plot as:

$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\lambda_8$
$\pi_1$	$\pi_2$	$\pi_1 + \pi_2$	$-\pi_1$	$-\pi_2$	$-(\pi_1 + \pi_2)$

Then, for example:

$$\text{ad}(h)[e_3, e_4] = (\pi_1 + \pi_2)(h)[e_3, e_4]$$

and hence  $[e_3, e_4]$  is an eigenvector of  $\text{ad}(h)$  with eigenvalues  $(\pi_1 + \pi_2)(h)$ . But so is  $e_5$ ! Hence, we must have  $[e_3, e_4] = \xi e_5$  for some  $\xi \in \mathbb{C}$ . Similarly,  $[e_5, e_7] = \xi e_3$ , and so on.

If  $\lambda_\alpha + \lambda_\beta \notin \Phi$ , then in order for the equation above to hold, we must have either  $[e_\alpha, e_\beta] = 0$  (so both sides are zero), or  $\lambda_\alpha(h) + \lambda_\beta(h) = 0$  for all  $h$ , i.e.  $\lambda_\alpha + \lambda_\beta = 0$  as a functional. In the latter case, we must have  $[e_\alpha, e_\beta] \in H$ . This follows from a stronger version of the maximality property of the Cartan subalgebra  $H$  of a simple Lie algebra  $L$ , namely that:

$$(\forall h \in H : [h, x] = 0) \Rightarrow x \in H$$

Summarising, we have:

$$[e_\alpha, e_\beta] = \begin{cases} \xi e_\gamma & \text{if } \lambda_\alpha + \lambda_\beta \in \Phi \\ \in H & \text{if } \lambda_\alpha + \lambda_\beta = 0 \\ 0 & \text{otherwise} \end{cases}$$

and these relations can be used to determine the remaining structure constants of  $A_2$ .

### 3.6 Representations Of Lie Algebras

**Definition 3.32** (Representations Of Lie Algebra). *Let  $L$  be a Lie algebra. A **representation** of  $L$  is a Lie algebra homomorphism:*

$$\rho: L \xrightarrow{\sim} \text{End}(V)$$

where  $V$  is some finite-dimensional vector space over the same field as  $L$ .

Recall that a linear map  $\rho: L \xrightarrow{\sim} \text{End}(V)$  is a Lie algebra homomorphism if:

$$\forall x, y \in L : \rho([x, y]) = [\rho(x), \rho(y)] := \rho(x) \circ \rho(y) - \rho(y) \circ \rho(x)$$

where the right hand side is the natural Lie bracket on  $\text{End}(V)$ .

**Definition 3.33** (Representation Space). *Let  $\rho: L \xrightarrow{\sim} \text{End}(V)$  be a representation of  $L$ . The vector space  $V$  is called the **representation space** of  $\rho$ .*

**Definition 3.34** (Dimension Of Representation). *Let  $\rho: L \xrightarrow{\sim} \text{End}(V)$  be a representation of  $L$ . The **dimension** of the representation  $\rho$  is  $\dim V$ .*

*Example 3.3.*

Consider the Lie algebra  $\mathfrak{sl}(2, \mathbb{C})$ . We constructed a basis  $\{X_1, X_2, X_3\}$  satisfying the relations:

$$\begin{aligned} [X_1, X_2] &= 2X_2 \\ [X_1, X_3] &= -2X_3 \\ [X_2, X_3] &= X_1 \end{aligned}$$

Let  $\rho: \mathfrak{sl}(2, \mathbb{C}) \xrightarrow{\sim} \text{End}(\mathbb{C}^2)$  be the linear map defined by:

$$\rho(X_1) := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \rho(X_2) := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \rho(X_3) := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Recall that a linear map is completely determined by its action on a basis, by linear continuation. To

check that  $\rho$  is a representation of  $\mathfrak{sl}(2, \mathbb{C})$ , we calculate:

$$\begin{aligned} [\rho(X_1), \rho(X_2)] &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \\ &= \rho(2X_2) \\ &= \rho([X_1, X_2]) \end{aligned}$$

Similarly, we find:

$$\begin{aligned} [\rho(X_1), \rho(X_3)] &= \rho([X_1, X_3]) \\ [\rho(X_2), \rho(X_3)] &= \rho([X_2, X_3]) \end{aligned}$$

By linear continuation,  $\rho([x, y]) = [\rho(x), \rho(y)]$  for any  $x, y \in \mathfrak{sl}(2, \mathbb{C})$  and hence,  $\rho$  is a 2-dimensional representation of  $\mathfrak{sl}(2, \mathbb{C})$  with representation space  $\mathbb{C}^2$ . Note that we have:

$$\begin{aligned} \text{im}_\rho(\mathfrak{sl}(2, \mathbb{C})) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{End}(\mathbb{C}^2) \mid a + d = 0 \right\} \\ &= \{ \phi \in \text{End}(\mathbb{C}^2) \mid \text{tr } \phi = 0 \} \end{aligned}$$

This is how  $\mathfrak{sl}(2, \mathbb{C})$  is often defined in physics courses, i.e. as the algebra of  $2 \times 2$  complex traceless matrices.

*Example 3.4.*

Consider  $\mathfrak{so}(3, \mathbb{R})$ , the Lie algebra of the rotation group  $\text{SO}(3, \mathbb{R})$ . It is a 3-dimensional Lie algebra over  $\mathbb{R}$ . It has a basis  $\{J_1, J_2, J_3\}$  satisfying:

$$[J_i, J_j] = C_{ij}^k J_k$$

where the structure constants  $C_{ij}^k$  are defined by first “pulling the index  $k$  down” using the Killing form  $\kappa_{ab} = C_{an}^m C_{bm}^n$  to obtain  $C_{kij} := \kappa_{km} C_{ij}^m$ , and then setting:

$$C_{kij} := \varepsilon_{ijk} := \begin{cases} 1 & \text{if } (i j k) \text{ is an even permutation of } (1 2 3) \\ -1 & \text{if } (i j k) \text{ is an odd permutation of } (1 2 3) \\ 0 & \text{otherwise.} \end{cases}$$

By evaluating these, we find:

$$\begin{aligned} [J_1, J_2] &= J_3 \\ [J_2, J_3] &= J_1 \\ [J_3, J_1] &= J_2 \end{aligned}$$

Define a linear map  $\rho_{\text{vec}}: \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3)$  by:

$$\rho_{\text{vec}}(J_1) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \rho_{\text{vec}}(J_2) := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \rho_{\text{vec}}(J_3) := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

You can easily check that this is a representation of  $\mathfrak{so}(3, \mathbb{R})$ . However, as you may be aware from quantum mechanics, there is another representation of  $\mathfrak{so}(3, \mathbb{R})$ , namely:

$$\rho_{\text{spin}}: \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{C}^2)$$

with  $\mathbb{C}^2$  understood as a 4-dimensional  $\mathbb{R}$ -vector space, defined by:

$$\rho_{\text{spin}}(J_1) := -\frac{i}{2} \sigma_1, \quad \rho_{\text{spin}}(J_2) := -\frac{i}{2} \sigma_2, \quad \rho_{\text{spin}}(J_3) := -\frac{i}{2} \sigma_3$$

where  $\sigma_1, \sigma_2, \sigma_3$  are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

You can again check that this is a representation of  $\mathfrak{so}(3, \mathbb{R})$ .

Notice that the two representations have different dimensions:

$$\dim \mathbb{R}^3 = 3 \neq 4 = \dim \mathbb{C}^2$$

Any (non-abelian) Lie algebra always has at least two special representations.

**Definition 3.35** (Trivial Representation). *Let  $L$  be a Lie algebra. A **trivial representation** of  $L$  is defined by:*

$$\begin{aligned} \rho_{\text{triv}}: L &\xrightarrow{\sim} \text{End}(V) \\ x &\mapsto \rho_{\text{triv}}(x) := 0 \end{aligned}$$

where  $0$  denotes the trivial endomorphism on  $V$ .

This is indeed a representation for any  $L$  since:

$$\forall x, y \in L: \rho_{\text{triv}}([x, y]) = 0 = [\rho_{\text{triv}}(x), \rho_{\text{triv}}(y)]$$

**Definition 3.36** (Adjoint Representation). *The **adjoint representation** of  $L$  is:*

$$\begin{aligned} \rho_{\text{adj}}: L &\xrightarrow{\sim} \text{End}(L) \\ x &\mapsto \rho_{\text{adj}}(x) := \text{ad}(x) \end{aligned}$$

Hence the adjoint map we have been using is actually a representation (we have already shown that  $\text{ad}$  is a Lie algebra homomorphism):

**Definition 3.37** (Faithful Representation). *A representation  $\rho: L \xrightarrow{\sim} \text{End}(V)$  is called **faithful** if  $\rho$  is injective, i.e.:*

$$\dim(\text{im}_\rho(L)) = \dim L$$

*Example 3.5.*

All representations considered so far are faithful, except for the trivial representations whenever the Lie algebra  $L$  is not itself trivial. Consider, for instance, the adjoint representation. We have:

$$\begin{aligned} \text{ad}(x) = \text{ad}(y) &\Leftrightarrow \forall z \in L: \text{ad}(x)z = \text{ad}(y)z \\ &\Leftrightarrow \forall z \in L: [x, z] = [y, z] \\ &\Leftrightarrow \forall z \in L: [x - y, z] = 0 \end{aligned}$$

If  $L$  is trivial, then any representation is faithful. Otherwise, there is some non-zero  $z \in L$ , hence we must have  $x - y = 0$ , so  $x = y$ , and thus  $\text{ad}$  is injective.

**Definition 3.38** (Direct Sum / Tensor Product Representations). *Given two representations  $\rho_1: L \xrightarrow{\sim} \text{End}(V_1)$ ,  $\rho_2: L \xrightarrow{\sim} \text{End}(V_2)$ , we can construct new representations called:*

i) *The **direct sum representation**:*

$$\begin{aligned} \rho_1 \oplus \rho_2: L &\xrightarrow{\sim} \text{End}(V_1 \oplus V_2) \\ x &\mapsto (\rho_1 \oplus \rho_2)(x) := \rho_1(x) \oplus \rho_2(x) \end{aligned}$$

ii) *The **tensor product representation**:*

$$\begin{aligned} \rho_1 \otimes \rho_2: L &\xrightarrow{\sim} \text{End}(V_1 \times V_2) \\ x &\mapsto (\rho_1 \otimes \rho_2)(x) := \rho_1(x) \otimes \text{id}_{V_2} + \text{id}_{V_1} \otimes \rho_2(x) \end{aligned}$$

*Example 3.6.*

The direct sum representation  $\rho_{\text{vec}} \oplus \rho_{\text{spin}} : \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3 \oplus \mathbb{C}^2)$  given in block-matrix form by:

$$(\rho_{\text{vec}} \oplus \rho_{\text{spin}})(x) = \left( \begin{array}{c|c} \rho_{\text{vec}}(x) & 0 \\ \hline 0 & \rho_{\text{spin}}(x) \end{array} \right)$$

is a 7-dimensional representation of  $\mathfrak{so}(3, \mathbb{R})$ .

**Definition 3.39** (Reducible Representation). *A representation  $\rho : L \xrightarrow{\sim} \text{End}(V)$  is called **reducible** if there exists a non-trivial vector subspace  $U \subseteq V$  which is invariant under the action of  $\rho$ , i.e.:*

$$\forall x \in L : \forall u \in U : \rho(x)u \in U$$

*In other words,  $\rho$  restricts to a representation  $\rho|_U : L \xrightarrow{\sim} \text{End}(U)$ .*

**Definition 3.40** (Irreducible Representation). *A representation is **irreducible** if it is not reducible.*

*Example 3.7.*

- i) The representation  $\rho_{\text{vec}} \oplus \rho_{\text{spin}} : \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3 \oplus \mathbb{C}^2)$  is reducible since, for example, we have a subspace  $\mathbb{R}^3 \oplus 0$  such that:

$$\forall x \in \mathfrak{so}(3, \mathbb{R}) : \forall u \in \mathbb{R}^3 \oplus 0 : (\rho_{\text{vec}} \oplus \rho_{\text{spin}})(x)u \in \mathbb{R}^3 \oplus 0$$

- ii) The representations  $\rho_{\text{vec}}$  and  $\rho_{\text{spin}}$  are both irreducible.

*Remark 3.9.* Just like the simple Lie algebras are the building blocks of all semi-simple Lie algebras, the irreducible representations of a semi-simple Lie algebra are the building blocks of all finite-dimensional representations of the Lie algebra. Any such representation can be decomposed as the direct sum of irreducible representations, which can then be classified according to their so-called *highest weights*.

### 3.6.1 The Casimir Operator

To every representation  $\rho$  of a compact Lie algebra (i.e. the Lie algebra of a compact Lie group) there is associated an operator  $\Omega_\rho$ , called the Casimir operator. We will need some preparation in order to define it.

**Definition 3.41** ( $\rho$ -Killing Form). *Let  $\rho : L \xrightarrow{\sim} \text{End}(V)$  be a representation of a complex Lie algebra  $L$ . We define the  $\rho$ -**Killing form** on  $L$  as:*

$$\begin{aligned} \kappa_\rho : L \times L &\xrightarrow{\sim} \mathbb{C} \\ (x, y) &\mapsto \kappa_\rho(x, y) := \text{tr}(\rho(x) \circ \rho(y)) \end{aligned}$$

Of course, the Killing form we have considered so far is just  $\kappa_{\text{ad}}$ . Similarly to  $\kappa_{\text{ad}}$ , every  $\kappa_\rho$  is symmetric and associative with respect to the Lie bracket of  $L$ .

**Proposition 3.7.** *Let  $\rho : L \xrightarrow{\sim} \text{End}(V)$  be a faithful representation of a complex semi-simple Lie algebra  $L$ . Then,  $\kappa_\rho$  is non-degenerate.*

Hence,  $\kappa_\rho$  induces an isomorphism  $L \xrightarrow{\sim} L^*$  via:

$$L \ni x \mapsto \kappa_\rho(x, -) \in L^*$$

Recall that if  $\{X_1, \dots, X_{\dim L}\}$  is a basis of  $L$ , then the dual basis  $\{\tilde{X}^1, \dots, \tilde{X}^{\dim L}\}$  of  $L^*$  is defined by:

$$\tilde{X}^i(X_j) = \delta_j^i$$

By using the isomorphism induced by  $\kappa_\rho$ , we can find some  $\xi_1, \dots, \xi_{\dim L} \in L$  such that we have  $\kappa(\xi_i, -) = \tilde{X}^i$  or, equivalently:

$$\forall x \in L : \kappa_\rho(x, \xi_i) = \tilde{X}^i(x)$$

We thus have:

$$\kappa_\rho(X_i, \xi_j) = \delta_{ij} := \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

**Proposition 3.8.** *Let  $\{X_i\}$  and  $\{\xi_j\}$  be defined as above. Then:*

$$[X_j, \xi_k] = \sum_{m=1}^{\dim L} C_{mj}^k \xi_m$$

where  $C_{mj}^k$  are the structure constants with respect to  $\{X_i\}$ .

*Proof.*

By using the associativity of  $\kappa_\rho$ , we have:

$$\kappa_\rho(X_i, [X_j, \xi_k]) = \kappa_\rho([X_i, X_j], \xi_k) = C_{ij}^m \kappa_\rho(X_m, \xi_k) = C_{ij}^m \delta_{mk} = C_{ij}^k$$

But we also have:

$$\kappa_\rho\left(X_i, \sum_{m=1}^{\dim L} C_{mj}^k \xi_m\right) = \sum_{m=1}^{\dim L} C_{mj}^k \kappa_\rho(X_i, \xi_m) = \sum_{m=1}^{\dim L} C_{mj}^k \delta_{im} = C_{ij}^k$$

Therefore:

$$\forall 1 \leq i \leq \dim L : \kappa_\rho\left(X_i, [X_j, \xi_k] - \sum_{m=1}^{\dim L} C_{mj}^k \xi_m\right) = 0$$

and hence, the result follows from the non-degeneracy of  $\kappa_\rho$ .  $\square$

We are now ready to define the Casimir operator and prove the subsequent theorem.

**Definition 3.42.** *Let  $\rho: L \xrightarrow{\sim} \text{End}(V)$  be a faithful representation of a complex (compact) Lie algebra  $L$  and let  $\{X_1, \dots, X_{\dim L}\}$  be a basis of  $L$ . The **Casimir operator** associated to the representation  $\rho$  is the endomorphism  $\Omega_\rho: V \xrightarrow{\sim} V$ :*

$$\Omega_\rho := \sum_{i=1}^{\dim L} \rho(X_i) \circ \rho(\xi_i)$$

**Theorem 3.8.** *Let  $\Omega_\rho$  the Casimir operator of a representation  $\rho: L \xrightarrow{\sim} \text{End}(V)$ . Then:*

$$\forall x \in L : [\Omega_\rho, \rho(x)] = 0$$

that is,  $\Omega_\rho$  commutes with every endomorphism in  $\text{im}_\rho(L)$ .

*Proof.*

Note that the bracket above is that on  $\text{End}(V)$ . Let  $x = x^k X_k \in L$ . Then:

$$\begin{aligned} [\Omega_\rho, \rho(x)] &= \left[ \sum_{i=1}^{\dim L} \rho(X_i) \circ \rho(\xi_i), \rho(x^k X_k) \right] \\ &= \sum_{i,k=1}^{\dim L} x^k [\rho(X_i) \circ \rho(\xi_i), \rho(X_k)] \end{aligned}$$

Observe that if the Lie bracket as the commutator with respect to an associative product, as is the case for  $\text{End}(V)$ , we have:

$$\begin{aligned} [AB, C] &= ABC - CBA \\ &= ABC - CBA - ACB + ACB \\ &= A[B, C] + [A, C]B \end{aligned}$$

Hence, by applying this, we obtain:

$$\begin{aligned}
\sum_{i,k=1}^{\dim L} x^k [\rho(X_i) \circ \rho(\xi_i), \rho(X_k)] &= \sum_{i,k=1}^{\dim L} x^k (\rho(X_i) \circ [\rho(\xi_i), \rho(X_k)] + [\rho(X_i), \rho(X_k)] \circ \rho(\xi_i)) \\
&= \sum_{i,k=1}^{\dim L} x^k (\rho(X_i) \circ \rho([\xi_i, X_k]) + \rho([X_i, X_k]) \circ \rho(\xi_i)) \\
&= \sum_{i,k,m=1}^{\dim L} x^k (\rho(X_i) \circ \rho(-C_{mk}^i \xi_m) + \rho(C_{ik}^m X_m) \circ \rho(\xi_i)) \\
&= \sum_{i,k,m=1}^{\dim L} x^k (-C_{mk}^i \rho(X_i) \circ \rho(\xi_m) + C_{ik}^m \rho(X_m) \circ \rho(\xi_i)) \\
&= \sum_{i,k,m=1}^{\dim L} x^k (-C_{mk}^i \rho(X_i) \circ \rho(\xi_m) + C_{mk}^i \rho(X_i) \circ \rho(\xi_m)) \\
&= 0
\end{aligned}$$

where we have swapped the dummy summation indices  $i$  and  $m$  in the second term.  $\square$

**Lemma 3.1** (Schur). *If  $\rho: L \xrightarrow{\sim} \text{End}(V)$  is irreducible, then any operator  $S$  which commutes with every endomorphism in  $\text{im}_\rho(L)$  has the form:*

$$S = c_\rho \text{id}_V$$

for some constant  $c_\rho \in \mathbb{C}$  (or  $\mathbb{R}$ , if  $L$  is a real Lie algebra).

It follows immediately that  $\Omega_\rho = c_\rho \text{id}_V$  for some  $c_\rho$  but, in fact, we can say more.

**Proposition 3.9.** *The Casimir operator of  $\rho: L \xrightarrow{\sim} \text{End}(V)$  is  $\Omega_\rho = c_\rho \text{id}_V$ , where:*

$$c_\rho = \frac{\dim L}{\dim V}$$

*Proof.*

We have:

$$\text{tr}(\Omega_\rho) = \text{tr}(c_\rho \text{id}_V) = c_\rho \dim V$$

and:

$$\begin{aligned}
\text{tr}(\Omega_\rho) &= \text{tr}\left(\sum_{i=1}^{\dim L} \rho(X_i) \circ \rho(\xi_i)\right) \\
&= \sum_{i=1}^{\dim L} \text{tr}(\rho(X_i) \circ \rho(\xi_i)) \\
&= \sum_{i=1}^{\dim L} \kappa_\rho(X_i, \xi_i) \\
&= \sum_{i=1}^{\dim L} \delta_{ii} \\
&= \dim L
\end{aligned}$$

which is what we wanted.  $\square$

*Example 3.8.*

Consider the Lie algebra  $\mathfrak{so}(3, \mathbb{R})$  with basis  $\{J_1, J_2, J_3\}$  satisfying:

$$[J_i, J_j] = \varepsilon_{ijk} J_k$$

where we assume the summation convention on the lower index  $k$ . Recall that the representation  $\rho_{\text{vec}}: \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3)$  is defined by:

$$\rho_{\text{vec}}(J_1) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \rho_{\text{vec}}(J_2) := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \rho_{\text{vec}}(J_3) := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Let us first evaluate the components of  $\kappa_{\rho_{\text{vec}}}$ . We have:

$$\begin{aligned} (\kappa_{\rho_{\text{vec}}})_{11} &:= \kappa_{\rho_{\text{vec}}}(J_1, J_1) = \text{tr}(\rho_{\text{vec}}(J_1) \circ \rho_{\text{vec}}(J_1)) \\ &= \text{tr}((\rho_{\text{vec}}(J_1))^2) \\ &= \text{tr} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}^2 \\ &= \text{tr} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \\ &= -2 \end{aligned}$$

After calculating the other components similarly, we find:

$$[(\kappa_{\rho_{\text{vec}}})_{ij}] = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Thus,  $\kappa_{\rho_{\text{vec}}}(J_i, \xi_j) = \delta_{ij}$  requires that we define  $\xi_i := -\frac{1}{2}J_i$ . Then, we have:

$$\begin{aligned} \Omega_{\rho_{\text{vec}}} &:= \sum_{i=1}^3 \rho_{\text{vec}}(J_i) \circ \rho_{\text{vec}}(\xi_i) \\ &= \sum_{i=1}^3 \rho_{\text{vec}}(J_i) \circ \rho_{\text{vec}}(-\tfrac{1}{2}J_i) \\ &= -\frac{1}{2} \sum_{i=1}^3 (\rho_{\text{vec}}(J_i))^2 \\ &= -\frac{1}{2} \left( \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}^2 + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}^2 + \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}^2 \right) \\ &= -\frac{1}{2} \left( \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Hence  $\Omega_{\rho_{\text{vec}}} = c_{\rho_{\text{vec}}} \text{id}_{\mathbb{R}^3}$  with  $c_{\rho_{\text{vec}}} = 1$ , which agrees with our previous theorem since:

$$\frac{\dim \mathfrak{so}(3, \mathbb{R})}{\dim \mathbb{R}^3} = \frac{3}{3} = 1$$

*Example 3.9.*

Let us consider the Lie algebra  $\mathfrak{so}(3, \mathbb{R})$  again, but this time with representation  $\rho_{\text{spin}}$ . Recall that this

is given by:

$$\rho_{\text{spin}}(J_1) := -\frac{i}{2} \sigma_1, \quad \rho_{\text{spin}}(J_2) := -\frac{i}{2} \sigma_2, \quad \rho_{\text{spin}}(J_3) := -\frac{i}{2} \sigma_3$$

where  $\sigma_1, \sigma_2, \sigma_3$  are the Pauli matrices. Recalling that  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \text{id}_{\mathbb{C}^2}$ , we calculate:

$$\begin{aligned} (\kappa_{\rho_{\text{spin}}})_{11} &:= \kappa_{\rho_{\text{spin}}}(J_1, J_1) = \text{tr}(\rho_{\text{spin}}(J_1) \circ \rho_{\text{spin}}(J_1)) \\ &= \text{tr}((\rho_{\text{spin}}(J_1))^2) \\ &= \left(-\frac{i}{2}\right)^2 \text{tr}(\sigma_1^2) \\ &= -\frac{1}{4} \text{tr}(\text{id}_{\mathbb{C}^2}) \\ &= -1 \end{aligned}$$

Note that  $\text{tr}(\text{id}_{\mathbb{C}^2}) = 4$ , since  $\text{tr}(\text{id}_V) = \dim V$  and here  $\mathbb{C}^2$  is considered as a 4-dimensional vector space over  $\mathbb{R}$ . Proceeding similarly, we find that the components of  $\kappa_{\rho_{\text{spin}}}$  are:

$$[(\kappa_{\rho_{\text{spin}}})_{ij}] = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Hence, we define  $\xi_i := -J_i$ . Then, we have:

$$\begin{aligned} \Omega_{\rho_{\text{spin}}} &:= \sum_{i=1}^3 \rho_{\text{spin}}(J_i) \circ \rho_{\text{spin}}(\xi_i) \\ &= \sum_{i=1}^3 \rho_{\text{spin}}(J_i) \circ \rho_{\text{spin}}(-J_i) \\ &= -\sum_{i=1}^3 (\rho_{\text{spin}}(J_i))^2 \\ &= -\left(-\frac{i}{2}\right)^2 \sum_{i=1}^3 \sigma_i^2 \\ &= \frac{1}{4} \sum_{i=1}^3 \text{id}_{\mathbb{C}^2} \\ &= \frac{3}{4} \text{id}_{\mathbb{C}^2} \end{aligned}$$

in accordance with the fact that:

$$\frac{\dim \mathfrak{so}(3, \mathbb{R})}{\dim \mathbb{C}^2} = \frac{3}{4}$$

# Chapter 4

## Topology

Topology is concerned with the properties of a geometric object that are preserved under continuous deformations, such as stretching, twisting, crumpling and bending, but not tearing or gluing.

A topological space is a set endowed with a structure, called a topology, which allows defining continuous deformation of subspaces, and, more generally, all kinds of continuity. Euclidean spaces, and, more generally, metric spaces are examples of a topological space, as any distance or metric defines a topology. The deformations that are considered in topology are homeomorphisms and homotopies. A property that is invariant under such deformations is a topological property. Basic examples of topological properties are: the dimension, which allows distinguishing between a line and a surface; compactness, which allows distinguishing between a line and a circle; connectedness, which allows distinguishing a circle from two non-intersecting circles.

The ideas underlying topology go back to Gottfried Leibniz, who in the 17th century envisioned the *geometria situs* and *analysis situs*. Leonhard Euler's Seven Bridges of Königsberg problem and polyhedron formula are arguably the field's first theorems. The term topology was introduced by Johann Benedict Listing in the 19th century, although it was not until the first decades of the 20th century that the idea of a topological space was developed.

### 4.1 Topological Spaces

We will now discuss topological spaces based on our previous development of axiomatic set theory. As we will see, a topology on a set provides the weakest structure in order to define the two very important notions of convergence of sequences to points in a set, and of continuity of maps between two sets. The definition of topology on a set is, at first sight, rather abstract. But on the upside it is also extremely simple. This definition is the result of a historical development, it is the simplest definition of topology that mathematicians found to be useful.

**Definition 4.1** (Topology). *Let  $M$  be a set. A **topology** on  $M$  is a set  $\mathcal{O} \subseteq \mathcal{P}(M)$  such that:*

- i)  $\emptyset \in \mathcal{O}$  and  $M \in \mathcal{O}$ .*
- ii)  $\{U, V\} \subseteq \mathcal{O} \Rightarrow \bigcap \{U, V\} \in \mathcal{O}$ .*
- iii)  $C \subseteq \mathcal{O} \Rightarrow \bigcup C \in \mathcal{O}$ .*

**Definition 4.2** (Topological Space). *Let  $M$  be a set and  $\mathcal{O}$  a topology on the set  $M$ . The pair  $(M, \mathcal{O})$  is called a **topological space**. If we write “let  $M$  be a topological space” then some topology  $\mathcal{O}$  on  $M$  is assumed.*

*Remark 4.1.* Unless  $|M| = 1$ , there are (usually many) different topologies  $\mathcal{O}$  that one can choose on the set  $M$ . The following table summarizes the number of possible topologies (some of them might not satisfy the axioms so they fail to be topologies) given the cardinality of a set.

$ M $	Number of topologies
1	1
2	4
3	29
4	355
5	6,942
6	209,527
7	9,535,241

*Example 4.1.*

Let  $M = \{a, b, c\}$ . Then  $\mathcal{O} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}\}$  is a topology on  $M$  since:

- i)  $\emptyset \in \mathcal{O}$  and  $M \in \mathcal{O}$ .
- ii) Clearly, for any  $S \in \mathcal{O}$ ,  $\bigcap \{\emptyset, S\} = \emptyset \in \mathcal{O}$  and  $\bigcap \{S, M\} = S \in \mathcal{O}$ . Moreover,  $\{a\} \cap \{b\} = \emptyset \in \mathcal{O}$ ,  $\{a\} \cap \{a, b\} = \{a\} \in \mathcal{O}$ , and  $\{b\} \cap \{a, b\} = \{b\} \in \mathcal{O}$ .
- iii) Let  $C \subseteq \mathcal{O}$ . If  $M \in C$ , then  $\bigcup C = M \in \mathcal{O}$ . If  $\{a, b\} \in C$  (or  $\{a\}, \{b\} \in C$ ) but  $M \notin C$ , then  $\bigcup C = \{a, b\} \in \mathcal{O}$ . If either  $\{a\} \in C$  or  $\{b\} \in C$ , but  $\{a, b\} \notin C$  and  $M \notin C$ , then  $\bigcup C = \{a\} \in \mathcal{O}$  or  $\bigcup C = \{b\} \in \mathcal{O}$ , respectively. Finally, if none of the above hold, then  $\bigcup C = \emptyset \in \mathcal{O}$ .

*Example 4.2.*

Let  $M$  be a set. Then  $\mathcal{O} = \{\emptyset, M\}$  is a topology on  $M$ . Indeed, we have:

- i)  $\emptyset \in \mathcal{O}$  and  $M \in \mathcal{O}$ .
- ii)  $\bigcap \{\emptyset, \emptyset\} = \emptyset \in \mathcal{O}$ ,  $\bigcap \{\emptyset, M\} = \emptyset \in \mathcal{O}$ , and  $\bigcap \{M, M\} = M \in \mathcal{O}$ .
- iii) If  $M \in C$ , then  $\bigcup C = M \in \mathcal{O}$ , otherwise  $\bigcup C = \emptyset \in \mathcal{O}$ .

This is called the *chaotic topology* and can be defined on any set.

*Example 4.3.*

Let  $M$  be a set. Then  $\mathcal{O} = \mathcal{P}(M)$  is a topology on  $M$ . Indeed, we have:

- i)  $\emptyset \in \mathcal{P}(M)$  and  $M \in \mathcal{P}(M)$ .
- ii) If  $U, V \in \mathcal{P}(M)$ , then  $\bigcap \{U, V\} \subseteq M$  and hence  $\bigcap \{U, V\} \in \mathcal{P}(M)$ .
- iii) If  $C \subseteq \mathcal{P}(M)$ , then  $\bigcup C \subseteq M$ , and hence  $\bigcup C \in \mathcal{P}(M)$ .

This is called the *discrete topology* and can be defined on any set.

We now give some common terminology regarding topologies.

**Definition 4.3** (Coarser / Finer Topology). *Let  $\mathcal{O}_1$  and  $\mathcal{O}_2$  be two topologies on a set  $M$ . If  $\mathcal{O}_1 \subset \mathcal{O}_2$ , then we say that  $\mathcal{O}_1$  is a **coarser** (or weaker) topology than  $\mathcal{O}_2$ . Equivalently, we say that  $\mathcal{O}_2$  is a **finer** (or stronger) topology than  $\mathcal{O}_1$ .*

Clearly, the chaotic topology is the coarsest topology on any given set, while the discrete topology is the finest.

**Definition 4.4** (Open / Closed Subsets). *Let  $(M, \mathcal{O})$  be a topological space. A subset  $S$  of  $M$  is said to be **open** (with respect to  $\mathcal{O}$ ) if  $S \in \mathcal{O}$  and **closed** (with respect to  $\mathcal{O}$ ) if  $M \setminus S \in \mathcal{O}$ .*

Notice that the notions of open and closed sets, as defined, are not mutually exclusive. A set could be both or neither, or one and not the other.

*Example 4.4.*

Let  $(M, \mathcal{O})$  be a topological space. Then  $\emptyset$  is open since  $\emptyset \in \mathcal{O}$ . However,  $\emptyset$  is also closed since  $M \setminus \emptyset = M \in \mathcal{O}$ . Similarly for  $M$ .

*Example 4.5.*

Let  $M = \{a, b, c\}$  and let  $\mathcal{O} = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}\}$ . Then  $\{a\}$  is open but not closed,  $\{b, c\}$  is closed but not open, and  $\{b\}$  is neither open nor closed.

We will now define what is called the standard topology on  $\mathbb{R}^d$ , where:

$$\mathbb{R}^d := \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{d \text{ times}}$$

We will need the following auxiliary definition.

**Definition 4.5** (Open Balls). *For any  $x \in \mathbb{R}^d$  and any  $r \in \mathbb{R}^+ := \{s \in \mathbb{R} \mid s > 0\}$ , we define the **open ball** of radius  $r$  around the point  $x$ :*

$$B_r(x) := \{y \in \mathbb{R}^d \mid \sqrt{\sum_{i=1}^d (y_i - x_i)^2} < r\}$$

where  $x := (x_1, x_2, \dots, x_d)$  and  $y := (y_1, y_2, \dots, y_d)$ , with  $x_i, y_i \in \mathbb{R}$ .

*Remark 4.2.* The quantity  $\sqrt{\sum_{i=1}^d (y_i - x_i)^2}$  is usually denoted by  $\|y - x\|_2$ , where  $\|\cdot\|_2$  is the 2-norm on  $\mathbb{R}^d$ . However, the definition of a norm on a set requires the set to be equipped with a vector space structure (which we haven't defined yet), while our construction does not. Moreover, our construction can be proven to be independent of the particular norm used to define it, i.e. any other norm will induce the same topological structure.

**Definition 4.6** (Standard Topology). *The **standard topology** on  $\mathbb{R}^d$ , denoted  $\mathcal{O}_{\text{std}}$ , is defined by:*

$$U \in \mathcal{O}_{\text{std}} :\Leftrightarrow \forall p \in U : \exists r \in \mathbb{R}^+ : B_r(p) \subseteq U$$

Of course, simply calling something a topology, does not automatically make it into a topology. We have to prove that  $\mathcal{O}_{\text{std}}$  as we defined it, does constitute a topology.

**Proposition 4.1.** *The pair  $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$  is a topological space.*

*Proof.*

i) First, we need to check whether  $\emptyset \in \mathcal{O}_{\text{std}}$ , i.e. whether is true:

$$\forall p \in \emptyset : \exists r \in \mathbb{R}^+ : B_r(p) \subseteq \emptyset$$

This proposition is of the form  $\forall p \in \emptyset : Q(p)$ , which was defined as being equivalent to:

$$\forall p : p \in \emptyset \Rightarrow Q(p)$$

However, since  $p \in \emptyset$  is false, the implication is true independent of  $p$ . Hence the initial proposition is true and thus  $\emptyset \in \mathcal{O}_{\text{std}}$ .

Second, by definition, we have  $B_r(x) \subseteq \mathbb{R}^d$  independent of  $x$  and  $r$ , hence:

$$\forall p \in \mathbb{R}^d : \exists r \in \mathbb{R}^+ : B_r(p) \subseteq \mathbb{R}^d$$

is true and thus  $\mathbb{R}^d \in \mathcal{O}_{\text{std}}$ .

ii) Let  $U, V \in \mathcal{O}_{\text{std}}$  and let  $p \in U \cap V$ . Then:

$$p \in U \cap V :\Leftrightarrow p \in U \wedge p \in V$$

and hence, since  $U, V \in \mathcal{O}_{\text{std}}$ , we have:

$$\exists r_1 \in \mathbb{R}^+ : B_{r_1}(p) \subseteq U \quad \wedge \quad \exists r_2 \in \mathbb{R}^+ : B_{r_2}(p) \subseteq V$$

Let  $r = \min\{r_1, r_2\}$ . Then:

$$B_r(p) \subseteq B_{r_1}(p) \subseteq U \quad \wedge \quad B_r(p) \subseteq B_{r_2}(p) \subseteq V$$

and hence  $B_r(p) \subseteq U \cap V$ . Therefore  $U \cap V \in \mathcal{O}_{\text{std}}$ .

iii) Let  $C \subseteq \mathcal{O}_{\text{std}}$  and let  $p \in \bigcup C$ . Then,  $p \in U$  for some  $U \in C$  and, since  $U \in \mathcal{O}_{\text{std}}$ , we have:

$$\exists r \in \mathbb{R}^+ : B_r(p) \subseteq U \subseteq \bigcup C$$

Therefore,  $\mathcal{O}_{\text{std}}$  is indeed a topology on  $\mathbb{R}^d$ . □

## 4.2 Construction Of New Topologies From Given Ones

**Definition 4.7** (Induced Topology). *Let  $(M, \mathcal{O})$  be a topological space and let  $N \subset M$ . Then we call the **induced topology** on  $N$  the topology:*

$$\mathcal{O}|_N := \{U \cap N \mid U \in \mathcal{O}\} \subseteq \mathcal{P}(N)$$

Of course we need to prove that this is indeed a topology.

*Proof.*

i) Since  $\emptyset \in \mathcal{O}$  and  $\emptyset = \emptyset \cap N$ , we have  $\emptyset \in \mathcal{O}|_N$ . Similarly, we have  $M \in \mathcal{O}$  and  $N = M \cap N$ , and thus  $N \in \mathcal{O}|_N$ .

ii) Let  $U, V \in \mathcal{O}|_N$ . Then, by definition:

$$\exists S \in \mathcal{O} : U = S \cap N \quad \wedge \quad \exists T \in \mathcal{O} : V = T \cap N$$

We thus have:

$$U \cap V = (S \cap N) \cap (T \cap N) = (S \cap T) \cap N$$

Since  $S, T \in \mathcal{O}$  and  $\mathcal{O}$  is a topology, we have  $S \cap T \in \mathcal{O}$  and hence  $U \cap V \in \mathcal{O}|_N$ .

iii) Let  $C := \{S_\alpha \mid \alpha \in \mathcal{A}\} \subseteq \mathcal{O}|_N$ . By definition, we have:

$$\forall \alpha \in \mathcal{A} : \exists U_\alpha \in \mathcal{O} : S_\alpha = U_\alpha \cap N$$

Then, using the notation:

$$\bigcup_{\alpha \in \mathcal{A}} S_\alpha := \bigcup C = \bigcup \{S_\alpha \mid \alpha \in \mathcal{A}\}$$

and De Morgan's law, we have:

$$\bigcup_{\alpha \in \mathcal{A}} S_\alpha = \bigcup_{\alpha \in \mathcal{A}} (U_\alpha \cap N) = \left( \bigcup_{\alpha \in \mathcal{A}} U_\alpha \right) \cap N$$

Since  $\mathcal{O}$  is a topology, we have  $\bigcup_{\alpha \in \mathcal{A}} U_\alpha \in \mathcal{O}$  and hence  $\bigcup C \in \mathcal{O}|_N$ .

Thus  $\mathcal{O}|_N$  is a topology on  $N$ . □

*Example 4.6.*

Consider  $(\mathbb{R}, \mathcal{O}_{\text{std}})$  and let:

$$N = [-1, 1] := \{x \in \mathbb{R} \mid -1 \leq x \leq 1\}$$

Then  $(N, \mathcal{O}_{\text{std}}|_N)$  is a topological space. The set  $(0, 1]$  is clearly not open in  $(\mathbb{R}, \mathcal{O}_{\text{std}})$  since  $(0, 1] \notin \mathcal{O}_{\text{std}}$ . However, we have:

$$(0, 1] = (0, 2) \cap [-1, 1]$$

where  $(0, 2) \in \mathcal{O}_{\text{std}}$  and hence  $(0, 1] \in \mathcal{O}_{\text{std}}|_N$ , i.e. the set  $(0, 1]$  is open in  $(N, \mathcal{O}_{\text{std}}|_N)$ .

**Definition 4.8** (Quotient Topology). *Let  $(M, \mathcal{O})$  be a topological space and let  $\sim$  be an equivalence relation on  $M$ . Then, the quotient set:*

$$M/\sim = \{[m] \in \mathcal{P}(M) \mid m \in M\}$$

*can be equipped with the **quotient topology**  $\mathcal{O}_{M/\sim}$  defined by:*

$$\mathcal{O}_{M/\sim} := \{U \in M/\sim \mid \bigcup U = \bigcup_{[a] \in U} [a] \in \mathcal{O}\}$$

An equivalent definition of the quotient topology is as follows. Let  $q: M \rightarrow M/\sim$  be the map:

$$\begin{aligned} q: M &\rightarrow M/\sim \\ m &\mapsto [m] \end{aligned}$$

Then we have:

$$\mathcal{O}_{M/\sim} := \{U \in M/\sim \mid \text{preim}_q(U) \in \mathcal{O}\}$$

*Example 4.7.*

The *circle* (or 1-sphere) is defined as the set  $S^1 := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$  equipped with the subset topology inherited from  $\mathbb{R}^2$ . The open sets of the circle are (unions of) open arcs, i.e. arcs without the endpoints. Individual points on the circle are clearly not open since there is no open set of  $\mathbb{R}^2$  whose intersection with the circle is a single point. However, an individual point on the circle is a closed set since its complement is an open arc.

An alternative definition of the circle is the following. Let  $\sim$  be the equivalence relation on  $\mathbb{R}$  defined by:

$$x \sim y :\Leftrightarrow \exists n \in \mathbb{Z} : x = y + 2\pi n$$

Then the circle can be defined as the set  $S^1 := \mathbb{R}/\sim$  equipped with the quotient topology.

**Definition 4.9** (Product Topology). *Let  $(A, \mathcal{O}_A)$  and  $(B, \mathcal{O}_B)$  be topological spaces. Then a topology on  $A \times B$  is defined by the set  $\mathcal{O}_{A \times B}$  called the **product topology** as:*

$$U \in \mathcal{O}_{A \times B} :\Leftrightarrow \forall p \in U : \exists (S, T) \in \mathcal{O}_A \times \mathcal{O}_B : S \times T \subseteq U$$

*Remark 4.3.* This definition can easily be extended to  $n$ -fold cartesian products:

$$U \in \mathcal{O}_{A_1 \times \dots \times A_n} :\Leftrightarrow \forall p \in U : \exists (S_1, \dots, S_n) \in \mathcal{O}_{A_1} \times \dots \times \mathcal{O}_{A_n} : S_1 \times \dots \times S_n \subseteq U$$

*Remark 4.4.* Using the previous definition, one can check that the standard topology on  $\mathbb{R}^d$  satisfies:

$$\mathcal{O}_{\text{std}} = \underbrace{\mathcal{O}_{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}}_{d \text{ times}}$$

Therefore, a more minimalistic definition of the standard topology on  $\mathbb{R}^d$  would consist in defining  $\mathcal{O}_{\text{std}}$  only for  $\mathbb{R}$  (i.e.  $d = 1$ ) and then extending it to  $\mathbb{R}^d$  by the product topology.

## 4.3 Convergence & Continuity

**Definition 4.10** (Sequence). *Let  $M$  be a set. A **sequence** (of points) in  $M$  is a function  $q: \mathbb{N} \rightarrow M$ .*

**Definition 4.11** (Convergence). *Let  $(M, \mathcal{O})$  be a topological space. A sequence  $q$  in  $M$  is said to **converge** against a limit point  $a \in M$  if:*

$$\forall U \in \mathcal{O} : a \in U \Rightarrow \exists N \in \mathbb{N} : \forall n > N : q(n) \in U$$

*Remark 4.5.* An open set  $U$  of  $M$  such that  $a \in U$  is called an *open neighbourhood* of  $a$ . If we denote

this by  $U(a)$ , then the previous definition of convergence can be rewritten as:

$$\forall U(a) : \exists N \in \mathbb{N} : \forall n > N : q(n) \in U$$

*Example 4.8.*

Consider the topological space  $(M, \{\emptyset, M\})$ . Then every sequence in  $M$  converges to every point in  $M$ . Indeed, let  $q$  be any sequence and let  $a \in M$ . Then,  $q$  converges against  $a$  if:

$$\forall U \in \{\emptyset, M\} : a \in U \Rightarrow \exists N \in \mathbb{N} : \forall n > N : q(n) \in U$$

This proposition is vacuously true for  $U = \emptyset$ , while for  $U = M$  we have  $q(n) \in M$  independent of  $n$ . Therefore, the (arbitrary) sequence  $q$  converges to the (arbitrary) point  $a \in M$ .

*Example 4.9.*

Consider the topological space  $(M, \mathcal{P}(M))$ . Then only definitely constant sequences converge, where a sequence is *definitely constant* with value  $c \in M$  if:

$$\exists N \in \mathbb{N} : \forall n > N : q(n) = c$$

This is immediate from the definition of convergence since in the discrete topology all singleton sets (i.e. one-element sets) are open.

*Example 4.10.*

Consider the topological space  $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$ . Then, a sequence  $q : \mathbb{N} \rightarrow \mathbb{R}^d$  converges against  $a \in \mathbb{R}^d$  if:

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall n > N : \|q(n) - a\|_2 < \varepsilon$$

*Example 4.11.*

Let  $M = \mathbb{R}$  and let  $q = 1 - \frac{1}{n+1}$ . Then, since  $q$  is not definitely constant, it is not convergent in  $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ , but it is convergent in  $(\mathbb{R}, \mathcal{O}_{\text{std}})$ .

**Definition 4.12** (Continuity). *Let  $(M, \mathcal{O}_M)$  and  $(N, \mathcal{O}_N)$  be topological spaces and let  $\phi : M \rightarrow N$  be a map. Then,  $\phi$  is said to be **continuous** (with respect to the topologies  $\mathcal{O}_M$  and  $\mathcal{O}_N$ ) if:*

$$\forall S \in \mathcal{O}_N, \text{ preim}_\phi(S) \in \mathcal{O}_M$$

where  $\text{preim}_\phi(S) := \{m \in M : \phi(m) \in S\}$  is the pre-image of  $S$  under the map  $\phi$ .

Informally, one says that  $\phi$  is continuous if the pre-images of open sets are open.

*Example 4.12.*

If  $M$  is equipped with the discrete topology, or  $N$  with the chaotic topology, then any map  $\phi : M \rightarrow N$  is continuous. Indeed, let  $S \in \mathcal{O}_N$ . If  $\mathcal{O}_M = \mathcal{P}(M)$  (and  $\mathcal{O}_N$  is any topology), then we have:

$$\text{preim}_\phi(S) = \{m \in M : \phi(m) \in S\} \subseteq M \in \mathcal{P}(M) = \mathcal{O}_M$$

If instead  $\mathcal{O}_N = \{\emptyset, N\}$  (and  $\mathcal{O}_M$  is any topology), then either  $S = \emptyset$  or  $S = N$  and thus, we have:

$$\text{preim}_\phi(\emptyset) = \emptyset \in \mathcal{O}_M \quad \text{and} \quad \text{preim}_\phi(N) = M \in \mathcal{O}_M$$

*Example 4.13.*

Let  $M = \{a, b, c\}$  and  $N = \{1, 2, 3\}$ , with respective topologies:

$$\mathcal{O}_M = \{\emptyset, \{b\}, \{a, c\}, \{a, b, c\}\} \quad \text{and} \quad \mathcal{O}_N = \{\emptyset, \{2\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

and let  $\phi : M \rightarrow N$  by defined by:

$$\phi(a) = 2, \quad \phi(b) = 1, \quad \phi(c) = 2$$

Then  $\phi$  is continuous. Indeed, we have:

$$\begin{array}{lll} \text{preim}_\phi(\emptyset) = \emptyset, & \text{preim}_\phi(\{2\}) = \{a, c\}, & \text{preim}_\phi(\{3\}) = \emptyset \\ \text{preim}_\phi(\{1, 3\}) = \{b\}, & \text{preim}_\phi(\{2, 3\}) = \{a, c\}, & \text{preim}_\phi(\{1, 2, 3\}) = \{a, b, c\} \end{array}$$

and hence  $\text{preim}_\phi(S) \in \mathcal{O}_M$  for all  $S \in \mathcal{O}_N$ .

*Example 4.14.*

Consider  $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$  and  $(\mathbb{R}^s, \mathcal{O}_{\text{std}})$ . Then  $\phi: \mathbb{R}^d \rightarrow \mathbb{R}^s$  is continuous with respect to the standard topologies if it satisfies the usual  $\varepsilon$ - $\delta$  definition of continuity:

$$\forall a \in \mathbb{R}^d : \forall \varepsilon > 0 : \exists \delta > 0 : \forall 0 < \|x - a\|_2 < \delta : \|\phi(x) - \phi(a)\|_2 < \varepsilon$$

**Definition 4.13** (Homeomorphism). *Let  $(M, \mathcal{O}_M)$  and  $(N, \mathcal{O}_N)$  be topological spaces. A bijection  $\phi: M \rightarrow N$  is called a **homeomorphism** if both  $\phi: M \rightarrow N$  and  $\phi^{-1}: N \rightarrow M$  are continuous.*

*Remark 4.6.* Homeo(morphism)s are the structure-preserving maps in topology.

If there exists a homeomorphism  $\phi$  between  $(M, \mathcal{O}_M)$  and  $(N, \mathcal{O}_N)$ :

$$\begin{array}{ccc} & \phi & \\ M & \xrightarrow{\quad} & N \\ & \xleftarrow{\quad \phi^{-1}} & \end{array}$$

then  $\phi$  provides a one-to-one pairing of the open sets of  $M$  with the open sets of  $N$ .

**Definition 4.14** (Isomorphic Topological Spaces). *If there exists a homeomorphism between two topological spaces  $(M, \mathcal{O}_M)$  and  $(N, \mathcal{O}_N)$ , we say that the two spaces are **homeomorphic** or **topologically isomorphic** and we write  $(M, \mathcal{O}_M) \cong_{\text{top}} (N, \mathcal{O}_N)$ .*

Clearly, if  $(M, \mathcal{O}_M) \cong_{\text{top}} (N, \mathcal{O}_N)$ , then  $M \cong_{\text{set}} N$ .

## 4.4 Invariant Topological Properties

**Definition 4.15** (Invariant Topological Properties). *A property of a topological space is called an **invariant** if any two homeomorphic topological spaces share the property.*

In this section we will mention some of the (almost uncountable) invariant topological properties of topological spaces. A *classification* of topological spaces would be a list of topological invariants such that any two spaces which share these invariants are homeomorphic. As of now, no such list is known!

### 4.4.1 Separation Properties

**Definition 4.16** (T1 Topological Space). *A topological space  $(M, \mathcal{O})$  is said to be **T1** if for any two distinct points  $p, q \in M$ ,  $p \neq q$ :*

$$\exists U(p) \in \mathcal{O} : q \notin U(p)$$

**Definition 4.17** (T2 or Hausdorff Topological Space). *A topological space  $(M, \mathcal{O})$  is said to be **T2** or **Hausdorff** if, for any two distinct points, there exist non-intersecting open neighbourhoods of these two points:*

$$\forall p, q \in M : p \neq q \Rightarrow \exists U(p), V(q) \in \mathcal{O} : U(p) \cap V(q) = \emptyset$$

*Example 4.15.*

The topological space  $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$  is T2 and hence also T1.

*Example 4.16.*

The Zariski topology on an algebraic variety is T1 but not T2.

*Example 4.17.*

The topological space  $(M, \{\emptyset, M\})$  does not have the T1 property since for any  $p \in M$ , the only open neighbourhood of  $p$  is  $M$  and for any other  $q \neq p$  we have  $q \in M$ . Moreover, since this space is not T1, it cannot be T2 either.

*Remark 4.7.* There are many other “T” properties, including a  $T2^{1/2}$  property which differs from T2 in that the neighbourhoods are closed.

**Definition 4.18** (Cover). Let  $(M, \mathcal{O})$  be a topological space. A set  $C \subseteq \mathcal{P}(M)$  is called a **cover** (of  $M$ ) if:

$$\bigcup C = M$$

Additionally, it is said to be an open cover if  $C \subseteq \mathcal{O}$ .

**Definition 4.19** (Open Cover). Let  $(M, \mathcal{O})$  be a topological space. A cover  $C \subseteq \mathcal{P}(M)$  is said to be an **open cover** if  $C \subseteq \mathcal{O}$ .

**Definition 4.20** (Subcover). Let  $C$  be a cover. Then any subset  $\tilde{C} \subseteq C$  such that  $\tilde{C}$  is still a cover, is called a **subcover**. Additionally, it is said to be a finite subcover if it is finite as a set.

**Definition 4.21** (Compact Topological Space). A topological space  $(M, \mathcal{O})$  is said to be **compact** if every open cover has a finite subcover.

**Definition 4.22** (Compact Subset). Let  $(M, \mathcal{O})$  be a topological space. A subset  $N \subseteq M$  is called **compact** if the topological space  $(N, \mathcal{O}|_N)$  is compact.

Determining whether a set is compact or not is not an easy task. Fortunately though, for  $\mathbb{R}^d$  equipped with the standard topology  $\mathcal{O}_{\text{std}}$ , the following theorem greatly simplifies matters.

**Theorem 4.1** (Heine-Borel). Let  $\mathbb{R}^d$  be equipped with the standard topology  $\mathcal{O}_{\text{std}}$ . Then, a subset of  $\mathbb{R}^d$  is compact if, and only if, it is closed and bounded.

A subset  $S$  of  $\mathbb{R}^d$  is said to be *bounded* if:

$$\exists r \in \mathbb{R}^+ : S \subseteq B_r(0)$$

*Example 4.18.*

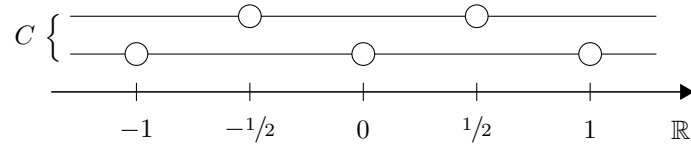
The interval  $[0, 1]$  is compact in  $(\mathbb{R}, \mathcal{O}_{\text{std}})$ . The one-element set containing  $(-1, 2)$  is a cover of  $[0, 1]$ , but it is also a finite subcover and hence  $[0, 1]$  is compact from the definition. Alternatively,  $[0, 1]$  is clearly closed and bounded, and hence it is compact by the Heine-Borel theorem.

*Example 4.19.*

The set  $\mathbb{R}$  is not compact in  $(\mathbb{R}, \mathcal{O}_{\text{std}})$ . To prove this, it suffices to show that there exists a cover of  $\mathbb{R}$  that does not have a finite subcover. To this end, let:

$$C := \{(n, n+1) \mid n \in \mathbb{Z}\} \cup \{(n + \frac{1}{2}, n + \frac{3}{2}) \mid n \in \mathbb{Z}\}$$

This corresponds to the following picture.



It is clear that removing even one element from  $C$  will cause  $C$  to fail to be an open cover of  $\mathbb{R}$ . Therefore, there is no finite subcover of  $C$  and hence,  $\mathbb{R}$  is not compact.

**Theorem 4.2.** Let  $(M, \mathcal{O}_M)$  and  $(N, \mathcal{O}_N)$  be compact topological spaces. Then  $(M \times N, \mathcal{O}_{M \times N})$  is a compact topological space.

The above theorem easily extends to finite cartesian products.

**Definition 4.23** (Refinement). Let  $(M, \mathcal{O})$  be a topological space and let  $C$  be a cover. A **refinement** of  $C$  is a cover  $R$  such that:

$$\forall U \in R : \exists V \in C : U \subseteq V$$

Any subcover of a cover is a refinement of that cover, but the converse is not true in general. A refinement  $R$  is said to be:

- *Open* if  $R \subseteq \mathcal{O}$ .

- *Locally finite* if for any  $p \in M$  there exists a neighbourhood  $U(p)$  such that the set:

$$\{U \in \mathcal{R} \mid U \cap U(p) \neq \emptyset\}$$

is finite as a set.

Compactness is a very strong property. Hence often times it does not hold, but a weaker and still useful property, called paracompactness, may still hold.

**Definition 4.24** (Paracompact Topological Space). *A topological space  $(M, \mathcal{O})$  is said to be **paracompact** if every open cover has an open refinement that is locally finite.*

**Corollary 4.1.** *If a topological space is compact, then it is also paracompact.*

*Remark 4.8.* Paracompactness is, informally, a rather natural property since every example of a non-paracompact space looks artificial. One such example is the *long line* (or *Alexandroff line*). To construct it, we first observe that we could “build”  $\mathbb{R}$  by taking the interval  $[0, 1)$  and stacking countably many copies of it one after the other. Hence, in a sense,  $\mathbb{R}$  is equivalent to  $\mathbb{Z} \times [0, 1)$ . The long line  $L$  is defined analogously as  $L : \omega_1 \times [0, 1)$ , where  $\omega_1$  is an uncountably infinite set. The resulting space  $L$  is not paracompact.

**Theorem 4.3.** *Let  $(M, \mathcal{O}_M)$  be a paracompact space and let  $(N, \mathcal{O}_N)$  be a compact space. Then  $M \times N$  (equipped with the product topology) is paracompact.*

**Corollary 4.2.** *Let  $(M, \mathcal{O}_M)$  be a paracompact space and let  $(N_i, \mathcal{O}_{N_i})$  be compact spaces for every  $1 \leq i \leq n$ . Then  $M \times N_1 \times \cdots \times N_n$  is paracompact.*

**Definition 4.25** (Partition Of Unity). *Let  $(M, \mathcal{O}_M)$  be a topological space. A **partition of unity** of  $M$  is a set  $\mathcal{F}$  of continuous maps from  $M$  to the interval  $[0, 1]$  such that for each  $p \in M$  the following conditions hold:*

- i) *There exists  $U(p)$  such that the set  $\{f \in \mathcal{F} \mid \forall x \in U(p) : f(x) \neq 0\}$  is finite.*
- ii)  $\sum_{f \in \mathcal{F}} f(p) = 1$ .

*If  $\mathcal{C}$  is an open cover, then  $\mathcal{F}$  is said to be subordinate to the cover  $\mathcal{C}$  if:*

$$\forall f \in \mathcal{F} : \exists U \in \mathcal{C} : f(x) \neq 0 \Rightarrow x \in U$$

**Theorem 4.4.** *Let  $(M, \mathcal{O}_M)$  be a Hausdorff topological space. Then  $(M, \mathcal{O}_M)$  is paracompact if, and only if, every open cover admits a partition of unity subordinate to that cover.*

#### 4.4.2 Connectedness & Path-Connectedness

**Definition 4.26** (Connected Topological Space). *A topological space  $(M, \mathcal{O})$  is said to be **connected** unless there exist two non-empty, non-intersecting open sets  $A$  and  $B$  such that  $M = A \cup B$ .*

*Example 4.20.*

Consider  $(\mathbb{R} \setminus \{0\}, \mathcal{O}_{\text{std}}|_{\mathbb{R} \setminus \{0\}})$ , i.e.  $\mathbb{R} \setminus \{0\}$  equipped with the subset topology inherited from  $\mathbb{R}$ . This topological space is not connected since  $(-\infty, 0)$  and  $(0, \infty)$  are open, non-empty, non-intersecting sets such that  $\mathbb{R} \setminus \{0\} = (-\infty, 0) \cup (0, \infty)$ .

**Theorem 4.5.** *The interval  $[0, 1] \subseteq \mathbb{R}$  equipped with the subset topology is connected.*

**Theorem 4.6.** *A topological space  $(M, \mathcal{O})$  is connected if, and only if, the only subsets that are both open and closed are  $\emptyset$  and  $M$ .*

**Definition 4.27** (Path - Connected Topological Space). *A topological space  $(M, \mathcal{O})$  is said to be **path-connected** if for every pair of points  $p, q \in M$  there exists a continuous curve  $\gamma : [0, 1] \rightarrow M$  such that  $\gamma(0) = p$  and  $\gamma(1) = q$ .*

*Example 4.21.*

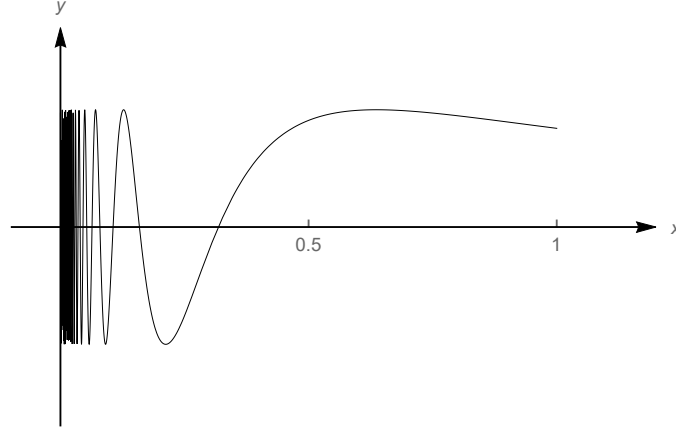
The space  $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$  is path-connected. Indeed, let  $p, q \in \mathbb{R}^d$  and let:

$$\gamma(\lambda) := p + \lambda(q - p)$$

Then  $\gamma$  is continuous and satisfies  $\gamma(0) = p$  and  $\gamma(1) = q$ .

*Example 4.22.*

Let  $S := \{(x, \sin(\frac{1}{x})) \mid x \in (0, 1]\} \cup \{(0, 0)\}$  be equipped with the subset topology inherited from  $\mathbb{R}^2$ .



The space  $(S, \mathcal{O}_{\text{std}}|_S)$  is connected but not path-connected.

**Theorem 4.7.** *If a topological space is path-connected, then it is also connected.*

#### 4.4.3 Homotopic Curves & The Fundamental Group

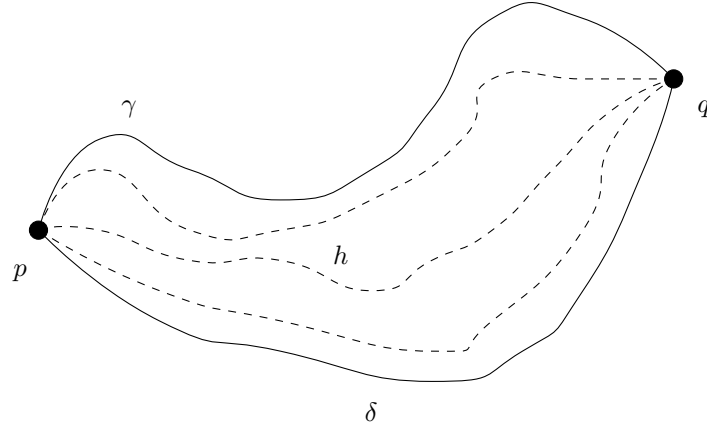
**Definition 4.28** (Homotopic Curves). *Let  $(M, \mathcal{O})$  be a topological space. Two curves  $\gamma, \delta: [0, 1] \rightarrow M$  such that:*

$$\gamma(0) = \delta(0) \quad \text{and} \quad \gamma(1) = \delta(1)$$

*are said to be **homotopic** if there exists a continuous map  $h: [0, 1] \times [0, 1] \rightarrow M$  such that for all  $\lambda \in [0, 1]$ :*

$$h(0, \lambda) = \gamma(\lambda) \quad \text{and} \quad h(1, \lambda) = \delta(\lambda)$$

Pictorially, two curves are homotopic if they can be continuously deformed into one another.



**Proposition 4.2.** *Let  $\gamma \sim \delta \Leftrightarrow$  “ $\gamma$  and  $\delta$  are homotopic”. Then,  $\sim$  is an equivalence relation.*

**Definition 4.29** (Space Of Loops). *Let  $(M, \mathcal{O})$  be a topological space. Then, for every  $p \in M$ , we define the **space of loops** at  $p$  by:*

$$\mathcal{L}_p := \{\gamma: [0, 1] \rightarrow M \mid \gamma \text{ is continuous and } \gamma(0) = \gamma(1)\}$$

**Definition 4.30** (Concatenation). *Let  $\mathcal{L}_p$  be the space of loops at  $p \in M$ . We define the **concatenation operation**  $*$ :  $\mathcal{L}_p \times \mathcal{L}_p \rightarrow \mathcal{L}_p$  by:*

$$(\gamma * \delta)(\lambda) := \begin{cases} \gamma(2\lambda) & \text{if } 0 \leq \lambda \leq \frac{1}{2} \\ \delta(2\lambda - 1) & \text{if } \frac{1}{2} \leq \lambda \leq 1 \end{cases}$$

**Definition 4.31** (Fundamental Group). *Let  $(M, \mathcal{O})$  be a topological space. The **fundamental group**  $\pi_1(p)$  of  $(M, \mathcal{O})$  at  $p \in M$  is the set:*

$$\pi_1(p) := \mathcal{L}_p / \sim = \{[\gamma] \mid \gamma \in \mathcal{L}_p\}$$

where  $\sim$  is the homotopy equivalence relation, together with the map :

$$\begin{aligned} \bullet: \pi_1(p) \times \pi_1(p) &\rightarrow \pi_1(p) \\ (\gamma, \delta) &\mapsto [\gamma] \bullet [\delta] := [\gamma * \delta] \end{aligned}$$

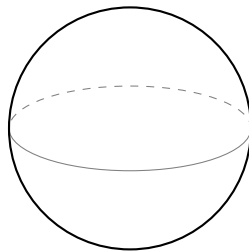
Observe that while all the previously discussed topological properties are “boolean-valued”, i.e. a topological space is either Hausdorff or not Hausdorff, either connected or not connected, and so on, the fundamental group is a “group-valued” property, i.e. the value of the property is not “either yes or no”, but a group.

*Example 4.23.*

The 2-sphere is defined as the set:

$$S^2 := \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

equipped with the subset topology inherited from  $\mathbb{R}^3$ .

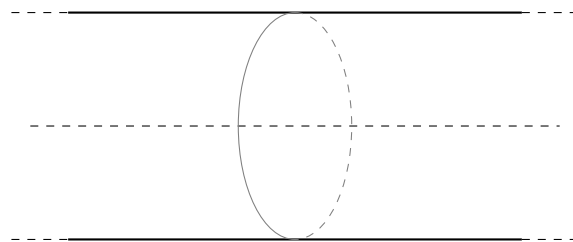


The sphere has the property that all the loops at any point are homotopic, hence the fundamental group (at every point) of the sphere is the trivial group:

$$\forall p \in S^2 : \pi_1(p) = 1 := \{[\gamma_e]\}$$

*Example 4.24.*

The cylinder is defined as  $C := \mathbb{R} \times S^1$  equipped with the product topology.



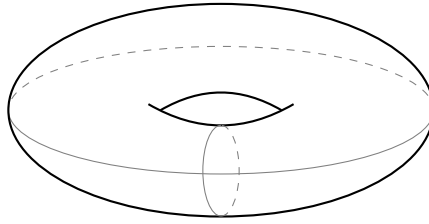
A loop in  $C$  can either go around the cylinder (i.e. around its central axis) or not. If it does not, then it can be continuously deformed to a point (the identity loop). If it does, then it cannot be deformed to the identity loop (intuitively because the cylinder is infinitely long) and hence it is a homotopically different loop. The number of times a loop winds around the cylinder is called the *winding number*. Loops with different winding numbers are not homotopic.

Moreover, loops with different *orientations* are also not homotopic and hence we have:

$$\forall p \in C : (\pi_1(p), \bullet) \cong_{\text{grp}} (\mathbb{Z}, +)$$

*Example 4.25.*

The 2-torus is defined as the set  $T^2 := S^1 \times S^1$  equipped with the product topology.



A loop in  $T^2$  can intuitively wind around the cylinder-like part of the torus as well as around the hole of the torus. That is, there are two independent winding numbers and hence:

$$\forall p \in T^2 : \pi_1(p) \cong_{\text{grp}} \mathbb{Z} \times \mathbb{Z}$$

where  $\mathbb{Z} \times \mathbb{Z}$  is understood as a group under pairwise addition.

## Chapter 5

# Topological Manifolds

A topological manifold is a topological space that locally resembles Euclidean space near each point. More precisely, each point of an  $n$ -dimensional manifold has a neighborhood that is homeomorphic to the Euclidean space of dimension  $n$ . In this more precise terminology, a manifold is referred to as an  $n$ -manifold.

One-dimensional manifolds include lines and circles. Two-dimensional manifolds (also called surfaces) include the plane, the sphere, and the torus, which can all be embedded (formed without self-intersections) in three dimensional real space, but also the Klein bottle and real projective plane, which will always self-intersect when immersed in three-dimensional real space.

Although a manifold locally resembles Euclidean space, meaning that every point has a neighbourhood homeomorphic to an open subset of Euclidean space, globally it may be not homeomorphic to Euclidean space. For example, the surface of the sphere is not homeomorphic to the Euclidean plane, because (among other properties) it has the global topological property of compactness that Euclidean space lacks, but in a region it can be charted by means of map projections of the region into the Euclidean plane (in the context of manifolds they are called charts). When a region appears in two neighbouring charts, the two representations do not coincide exactly and a transformation is needed to pass from one to the other, called a transition map.

The concept of a manifold is central to many parts of geometry and modern mathematical physics because it allows complicated structures to be described and understood in terms of the simpler local topological properties of Euclidean space. Manifolds naturally arise as solution sets of systems of equations and as graphs of functions.

### 5.1 Topological Manifolds

**Definition 5.1** (Topological Manifold). *A paracompact, Hausdorff, topological space  $(M, \mathcal{O})$  is called a  $d$ -dimensional topological manifold if for every point  $p \in M$  there exist a neighbourhood  $U(p)$  and a homeomorphism  $x: U(p) \rightarrow x(U(p)) \subseteq \mathbb{R}^d$ . We also write  $\dim M = d$ .*

Intuitively, a  $d$ -dimensional manifold is a topological space which locally (i.e. around each point) looks like  $\mathbb{R}^d$ . Note that, strictly speaking, what we have just defined are *real* topological manifolds. We could define *complex* topological manifolds as well, simply by requiring that the map  $x$  be a homeomorphism onto an open subset of  $\mathbb{C}^d$ .

**Proposition 5.1.** *Let  $M$  be a  $d$ -dimensional manifold and let  $U, V \subseteq M$  be open, with  $U \cap V \neq \emptyset$ . If  $x$  and  $y$  are two homeomorphisms:*

$$x: U \rightarrow x(U) \subseteq \mathbb{R}^d \quad \text{and} \quad y: V \rightarrow y(V) \subseteq \mathbb{R}^{d'}$$

*then  $d = d'$ .*

This ensures that the concept of dimension is indeed well-defined, i.e. it is the same at every point, at least on each connected component of the manifold.

*Example 5.1.*

Trivially,  $\mathbb{R}^d$  is a  $d$ -dimensional manifold for any  $d \geq 1$ . The space  $S^1$  is a 1-dimensional manifold while the spaces  $S^2$ ,  $C$  and  $T^2$  are 2-dimensional manifolds.

**Definition 5.2** (Topological Submanifold). Let  $(M, \mathcal{O})$  be a topological manifold and let  $N \subseteq M$ . Then  $(N, \mathcal{O}|_N)$  is called a **submanifold** of  $(M, \mathcal{O})$  if it is a manifold in its own right.

*Example 5.2.*

The space  $S^1$  is a submanifold of  $\mathbb{R}^2$  while the spaces  $S^2$ ,  $C$  and  $T^2$  are submanifolds of  $\mathbb{R}^3$ .

**Definition 5.3** (Product Manifold). Let  $(M, \mathcal{O}_M)$  and  $(N, \mathcal{O}_N)$  be topological manifolds of dimension  $m$  and  $n$ , respectively. Then,  $(M \times N, \mathcal{O}_{M \times N})$  is a topological manifold of dimension  $m + n$  called the **product manifold**.

*Example 5.3.*

We have  $T^2 = S^1 \times S^1$  not just as topological spaces, but as topological manifolds as well. This is a special case of the  $n$ -torus:

$$T^n := \underbrace{S^1 \times S^1 \times \cdots \times S^1}_{n \text{ times}}$$

which is an  $n$ -dimensional manifold.

*Example 5.4.*

The cylinder  $C = S^1 \times \mathbb{R}$  is a 2-dimensional manifold.

## 5.2 Charts & Atlases

**Definition 5.4** (Chart). Let  $(M, \mathcal{O})$  be a  $d$ -dimensional manifold. Then, a pair  $(U, x)$  where  $U \in \mathcal{O}$  and  $x: U \rightarrow x(U) \subseteq \mathbb{R}^d$  is a homeomorphism, is said to be a **chart** of the manifold.

**Definition 5.5** (Components / Co-Ordinates Of A Chart). The **component functions (or maps)** of  $x: U \rightarrow x(U) \subseteq \mathbb{R}^d$  are the maps:

$$\begin{aligned} x^i: U &\rightarrow \mathbb{R} \\ p &\mapsto \text{proj}_i(x(p)) \end{aligned}$$

for  $1 \leq i \leq d$ , where  $\text{proj}_i(x(p))$  is the  $i$ -th component of  $x(p) \in \mathbb{R}^d$ . The  $x^i(p)$  are called the **co-ordinates** of the point  $p \in U$  with respect to the chart  $(U, x)$ .

**Definition 5.6** (Atlas). An **atlas** of a manifold  $M$  is a collection  $\mathcal{A} := \{(U_\alpha, x_\alpha) \mid \alpha \in \mathcal{A}\}$  of charts such that:

$$\bigcup_{\alpha \in \mathcal{A}} U_\alpha = M$$

**Definition 5.7** ( $\mathcal{C}^0$ -Compatible Charts). Two charts  $(U, x)$  and  $(V, y)$  are said to be  **$\mathcal{C}^0$ -compatible** if either  $U \cap V = \emptyset$  or the map:

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

is continuous.

Note that  $y \circ x^{-1}$  is a map from a subset of  $\mathbb{R}^d$  to a subset of  $\mathbb{R}^d$ .

$$\begin{array}{ccc} & U \cap V \subseteq M & \\ x \swarrow & & \searrow y \\ x(U \cap V) \subseteq \mathbb{R}^d & \xrightarrow{y \circ x^{-1}} & y(U \cap V) \subseteq \mathbb{R}^d \end{array}$$

Since the maps  $x$  and  $y$  are homeomorphisms, the composition map  $y \circ x^{-1}$  is also a homeomorphism and hence continuous. Therefore, any two charts on a topological manifold are  $\mathcal{C}^0$ -compatible. This definition may thus seem redundant since it applies to every pair of charts. However, it is just a “warm up” since we will later refine this definition and define the *differentiability* of maps on a manifold in terms of  $\mathcal{C}^k$ -compatibility of charts.

**Definition 5.8** (Chart Transition Map). The map  $y \circ x^{-1}$  (and its inverse  $x \circ y^{-1}$ ) is called the **chart transition map**.

**Definition 5.9** ( $\mathcal{C}^0$ -Atlas). A  $\mathcal{C}^0$ -**atlas** of a manifold is an atlas of pairwise  $\mathcal{C}^0$ -compatible charts.

Note that any atlas is also a  $\mathcal{C}^0$ -atlas.

**Definition 5.10** (Maximal Atlas). A  $\mathcal{C}^0$ -atlas  $\mathcal{A}$  is said to be a **maximal atlas** if for every  $(U, x) \in \mathcal{A}$ , we have  $(V, y) \in \mathcal{A}$  for all  $(V, y)$  charts that are  $\mathcal{C}^0$ -compatible with  $(U, x)$ .

*Example 5.5.*

Not every  $\mathcal{C}^0$ -atlas is a maximal atlas. Indeed, consider  $(\mathbb{R}, \mathcal{O}_{\text{std}})$  and the atlas  $\mathcal{A} := (\mathbb{R}, \text{id}_{\mathbb{R}})$ . Then  $\mathcal{A}$  is not maximal since  $((0, 1), \text{id}_{\mathbb{R}})$  is a chart which is  $\mathcal{C}^0$ -compatible with  $(\mathbb{R}, \text{id}_{\mathbb{R}})$  but  $((0, 1), \text{id}_{\mathbb{R}}) \notin \mathcal{A}$ .

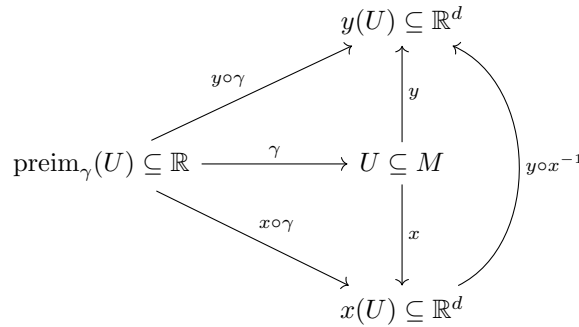
We can now look at “objects on” topological manifolds from two points of view. For instance, consider a curve on a  $d$ -dimensional manifold  $M$ , i.e. a map  $\gamma: \mathbb{R} \rightarrow M$ . We now ask whether this curve is continuous, as it should be if models the trajectory of a particle on the “physical space”  $M$ .

A first answer is that  $\gamma: \mathbb{R} \rightarrow M$  is continuous if it is continuous as a map between the topological spaces  $\mathbb{R}$  and  $M$ .

However, the answer that may be more familiar to you from undergraduate physics is the following. We consider only a portion (open subset  $U$ ) of the physical space  $M$  and, instead of studying the map  $\gamma: \text{preim}_{\gamma}(U) \rightarrow U$  directly, we study the map:

$$x \circ \gamma: \text{preim}_{\gamma}(U) \rightarrow x(U) \subseteq \mathbb{R}^d$$

where  $(U, x)$  is a chart of  $M$ . More likely, you would be checking the continuity of the co-ordinate maps  $x^i \circ \gamma$ , which would then imply the continuity of the “real” curve  $\gamma: \text{preim}_{\gamma}(U) \rightarrow U$  (real, as opposed to its co-ordinate representation).



At some point you may wish to use a different “co-ordinate system” to answer a different question. In this case, you would chose a different chart  $(U, y)$  and then study the map  $y \circ \gamma$  or its co-ordinate maps. Notice however that some results (e.g. the continuity of  $\gamma$ ) obtained in the previous chart  $(U, x)$  can be immediately “transported” to the new chart  $(U, y)$  via the chart transition map  $y \circ x^{-1}$ . Moreover, the map  $y \circ x^{-1}$  allows us to, intuitively speaking, forget about the inner structure (i.e.  $U$  and the maps  $\gamma$ ,  $x$  and  $x$ ) which, in a sense, is the real world, and only consider  $\text{preim}_{\gamma}(U) \subseteq \mathbb{R}$  and  $x(U), y(U) \subseteq \mathbb{R}^d$  together with the maps between them, which is our representation of the real world.

As we already said, for a topological manifold  $(M, \mathcal{O})$ , the concept of a  $\mathcal{C}^0$ -atlas is fully redundant since every atlas is also a  $\mathcal{C}^0$ -atlas. We will now generalise the notion of a  $\mathcal{C}^0$ -atlas, or more precisely, the notion of  $\mathcal{C}^0$ -compatibility of charts, to something which is non-trivial and non-redundant.

**Definition 5.11** ( $\clubsuit$ -Atlas). An atlas  $\mathcal{A}$  for a topological manifold is called a  **$\clubsuit$ -atlas** if any two charts  $(U, x), (V, y) \in \mathcal{A}$  are  $\clubsuit$ -compatible, where the symbol  $\clubsuit$  is being used as a placeholder for any of the following:

- $\clubsuit = \mathcal{C}^0$ : this just reduces to the previous definition.
- $\clubsuit = \mathcal{C}^k$ : the transition maps are  $k$ -times continuously differentiable as maps between open subsets of  $\mathbb{R}^{\dim M}$ .

- $\mathfrak{C} = \mathcal{C}^\infty$ : the transition maps are smooth (infinitely many times differentiable); equivalently, the atlas is  $\mathcal{C}^k$  for all  $k \geq 0$ .
- $\mathfrak{C} = \mathcal{C}^\omega$ : the transition maps are (real) analytic, which is stronger than being smooth.
- $\mathfrak{C} = \text{complex}$ : if  $\dim M$  is even,  $M$  is a complex manifold if the transition maps are continuous and satisfy the Cauchy-Riemann equations; its complex dimension is  $\frac{1}{2} \dim M$ .

In other words, either  $U \cap V = \emptyset$  or if  $U \cap V \neq \emptyset$ , then the transition map  $y \circ x^{-1}$  from  $x(U \cap V)$  to  $y(U \cap V)$  must be  $\mathfrak{C}$ .

$$\begin{array}{ccc}
 & U \cap V \subseteq M & \\
 x \swarrow & & \searrow y \\
 x(U \cap V) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{y \circ x^{-1}} & y(U \cap V) \subseteq \mathbb{R}^{\dim M}
 \end{array}$$

**Theorem 5.1** (Whitney). *Any maximal  $\mathcal{C}^k$ -atlas, with  $k \geq 1$ , contains a  $\mathcal{C}^\infty$ -atlas. Moreover, any two maximal  $\mathcal{C}^k$ -atlases that contain the same  $\mathcal{C}^\infty$ -atlas are identical.*

An immediate implication is that if we can find a  $\mathcal{C}^1$ -atlas for a manifold, then we can also assume the existence of a  $\mathcal{C}^\infty$ -atlas for that manifold. This is not the case for topological manifolds in general: a space with a  $\mathcal{C}^0$ -atlas may not admit any  $\mathcal{C}^1$ -atlas. But if we have at least a  $\mathcal{C}^1$ -atlas, then we can obtain a  $\mathcal{C}^\infty$ -atlas simply by removing charts, keeping only the ones which are  $\mathcal{C}^\infty$ -compatible.

Hence, for the purposes of this course, we will not distinguish between  $\mathcal{C}^k$  ( $k \geq 1$ ) and  $\mathcal{C}^\infty$ -manifolds in the above sense.

We now give the explicit definition of a  $\mathcal{C}^k$ -manifold.

**Definition 5.12** ( $\mathcal{C}^k$ -Manifold). *A  $\mathcal{C}^k$ -manifold is a triple  $(M, \mathcal{O}, \mathcal{A})$ , where  $(M, \mathcal{O})$  is a topological manifold and  $\mathcal{A}$  is a maximal  $\mathcal{C}^k$ -atlas.*

**Definition 5.13** (Smooth Manifold). *A  $\mathcal{C}^\infty$ -manifold is called a **smooth manifold**.*

*Remark 5.1.* A given topological manifold can carry different incompatible atlases.

Note that while we only defined compatibility of charts, it should be clear what it means for two atlases of the same type to be compatible.

**Definition 5.14** (Compatible / Incompatible Atlases). *Two  $\mathfrak{C}$ -atlases  $\mathcal{A}, \mathcal{B}$  are **compatible** if their union  $\mathcal{A} \cup \mathcal{B}$  is again a  $\mathfrak{C}$ -atlas, and are **incompatible** otherwise.*

Alternatively, we can define the compatibility of two atlases in terms of the compatibility of any pair of charts, one from each atlas.

*Example 5.6.*

Let  $(M, \mathcal{O}) = (\mathbb{R}, \mathcal{O}_{\text{std}})$ . Consider the two atlases  $\mathcal{A} = \{(\mathbb{R}, \text{id}_{\mathbb{R}})\}$  and  $\mathcal{B} = \{(\mathbb{R}, x)\}$ , where  $x: a \mapsto \sqrt[3]{a}$ . Since they both contain a single chart, the compatibility condition on the transition maps is easily seen to hold (in both cases, the only transition map is  $\text{id}_{\mathbb{R}}$ ). Hence they are both  $\mathcal{C}^\infty$ -atlases.

Consider now  $\mathcal{A} \cup \mathcal{B}$ . The transition map  $\text{id}_{\mathbb{R}} \circ x^{-1}$  is the map  $a \mapsto a^3$ , which is smooth. However, the other transition map,  $x \circ \text{id}_{\mathbb{R}}^{-1}$ , is the map  $x$ , which is not even differentiable once (the first derivative at 0 does not exist). Consequently,  $\mathcal{A}$  and  $\mathcal{B}$  are not even  $\mathcal{C}^1$ -compatible.

The previous example shows that we can equip the real line with (at least) two different incompatible  $\mathcal{C}^\infty$ -structures. This looks like a disaster as it implies that there is an arbitrary choice to be made about which differentiable structure to use. Fortunately, the situation is not as bad as it looks, as we will see in the next sections.

### 5.3 Differentiable Manifolds

**Definition 5.15** (Differentiable Map). *Let  $\phi: M \rightarrow N$  be a map, where  $(M, \mathcal{O}_M, \mathcal{A}_M)$  and  $(N, \mathcal{O}_N, \mathcal{A}_N)$  are  $\mathcal{C}^k$ -manifolds. Then  $\phi$  is said to be  $(\mathcal{C}^k\text{-})$ **differentiable at**  $p \in M$  if for some charts  $(U, x) \in \mathcal{A}_M$  with  $p \in U$  and  $(V, y) \in \mathcal{A}_N$  with  $\phi(p) \in V$ , the map  $y \circ \phi \circ x^{-1}$  is  $k$ -times continuously differentiable at  $x(p) \in x(U) \subseteq \mathbb{R}^{\dim M}$  in the usual sense.*

$$\begin{array}{ccc} U \subseteq M & \xrightarrow{\phi} & V \subseteq N \\ \downarrow x & & \downarrow y \\ x(U) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{y \circ \phi \circ x^{-1}} & y(V) \subseteq \mathbb{R}^{\dim N} \end{array}$$

The above diagram shows a typical theme with manifolds. We have a map  $\phi: M \rightarrow N$  and we want to define some property of  $\phi$  at  $p \in M$  analogous to some property of maps between subsets of  $\mathbb{R}^d$ . What we typically do is consider some charts  $(U, x)$  and  $(V, y)$  as above and define the desired property of  $\phi$  at  $p \in U$  in terms of the corresponding property of the downstairs map  $y \circ \phi \circ x^{-1}$  at the point  $x(p) \in \mathbb{R}^d$ .

Notice that in the previous definition we only require that *some* charts from the two atlases satisfy the stated property. So we should worry about whether this definition depends on which charts we pick. In fact, this “lifting” of the notion of differentiability from the chart representation of  $\phi$  to the manifold level is well-defined.

**Proposition 5.2.** *The definition of differentiability is well-defined.*

*Proof.*

We want to show that if  $y \circ \phi \circ x^{-1}$  is differentiable at  $x(p)$  for some  $(U, x) \in \mathcal{A}_M$  with  $p \in U$  and  $(V, y) \in \mathcal{A}_N$  with  $\phi(p) \in V$ , then  $\tilde{y} \circ \phi \circ \tilde{x}^{-1}$  is differentiable at  $\tilde{x}(p)$  for all charts  $(\tilde{U}, \tilde{x}) \in \mathcal{A}_M$  with  $p \in \tilde{U}$  and  $(\tilde{V}, \tilde{y}) \in \mathcal{A}_N$  with  $\phi(p) \in \tilde{V}$ .

$$\begin{array}{ccc} \tilde{x}(U \cap \tilde{U}) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{\tilde{y} \circ \phi \circ \tilde{x}^{-1}} & \tilde{y}(V \cap \tilde{V}) \subseteq \mathbb{R}^{\dim N} \\ \uparrow \tilde{x} & & \uparrow \tilde{y} \\ U \cap \tilde{U} \subseteq M & \xrightarrow{\phi} & V \cap \tilde{V} \subseteq N \\ \downarrow x & & \downarrow y \\ x(U \cap \tilde{U}) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{y \circ \phi \circ x^{-1}} & y(V \cap \tilde{V}) \subseteq \mathbb{R}^{\dim N} \end{array}$$

$\tilde{x} \circ x^{-1}$  (left curved arrow)       $\tilde{y} \circ y^{-1}$  (right curved arrow)

Consider the map  $\tilde{x} \circ x^{-1}$  in the diagram above. Since the charts  $(U, x)$  and  $(\tilde{U}, \tilde{x})$  belong to the same  $\mathcal{C}^k$ -atlas  $\mathcal{A}_M$ , by definition the transition map  $\tilde{x} \circ x^{-1}$  is  $\mathcal{C}^k$ -differentiable as a map between subsets of  $\mathbb{R}^{\dim M}$ , and similarly for  $\tilde{y} \circ y^{-1}$ . We now notice that we can write:

$$\tilde{y} \circ \phi \circ \tilde{x}^{-1} = (\tilde{y} \circ y^{-1}) \circ (y \circ \phi \circ x^{-1}) \circ (\tilde{x} \circ x^{-1})^{-1}$$

and since the composition of  $\mathcal{C}^k$  maps is still  $\mathcal{C}^k$ , we are done.  $\square$

This proof shows the significance of restricting to  $\mathcal{C}^k$ -atlases. Such atlases only contain charts for which the transition maps are  $\mathcal{C}^k$ , which is what makes our definition of differentiability of maps between manifolds well-defined.

The same definition and proof work for smooth ( $\mathcal{C}^\infty$ ) manifolds, in which case we talk about *smooth maps*. As we said before, this is the case we will be most interested in.

*Example 5.7.*

Consider the smooth manifolds  $(\mathbb{R}^d, \mathcal{O}_{\text{std}}, \mathcal{A}_d)$  and  $(\mathbb{R}^{d'}, \mathcal{O}_{\text{std}}, \mathcal{A}_{d'})$ , where  $\mathcal{A}_d$  and  $\mathcal{A}_{d'}$  are the maximal atlases containing the charts  $(\mathbb{R}^d, \text{id}_{\mathbb{R}^d})$  and  $(\mathbb{R}^{d'}, \text{id}_{\mathbb{R}^{d'}})$  respectively, and let  $f: \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$  be a map. The diagram defining the differentiability of  $f$  with respect to these charts is:

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{f} & \mathbb{R}^{d'} \\ \downarrow \text{id}_{\mathbb{R}^d} & & \downarrow \text{id}_{\mathbb{R}^{d'}} \\ \mathbb{R}^d & \xrightarrow{\text{id}_{\mathbb{R}^{d'}} \circ f \circ (\text{id}_{\mathbb{R}^d})^{-1}} & \mathbb{R}^{d'} \end{array}$$

and, by definition, the map  $f$  is smooth as a map between manifolds if, and only if, the map  $\text{id}_{\mathbb{R}^{d'}} \circ f \circ (\text{id}_{\mathbb{R}^d})^{-1} = f$  is smooth in the usual sense.

*Example 5.8.*

Let  $(M, \mathcal{O}, \mathcal{A})$  be a  $d$ -dimensional smooth manifold and let  $(U, x) \in \mathcal{A}$ . Then  $x: U \rightarrow x(U) \subseteq \mathbb{R}^d$  is smooth. Indeed, we have:

$$\begin{array}{ccc} U & \xrightarrow{x} & x(U) \\ \downarrow x & & \downarrow \text{id}_{x(U)} \\ x(U) \subseteq \mathbb{R}^d & \xrightarrow{\text{id}_{x(U)} \circ x \circ x^{-1}} & x(U) \subseteq \mathbb{R}^d \end{array}$$

Hence  $x: U \rightarrow x(U)$  is smooth if, and only if, the map  $\text{id}_{x(U)} \circ x \circ x^{-1} = \text{id}_{x(U)}$  is smooth in the usual sense, which it certainly is.

The coordinate maps  $x^i := \text{proj}_i \circ x: U \rightarrow \mathbb{R}$  are also smooth. Indeed, consider the diagram:

$$\begin{array}{ccc} U & \xrightarrow{x^i} & \mathbb{R} \\ \downarrow x & & \downarrow \text{id}_{\mathbb{R}} \\ x(U) \subseteq \mathbb{R}^d & \xrightarrow{\text{id}_{\mathbb{R}} \circ x^i \circ x^{-1}} & \mathbb{R} \end{array}$$

Then,  $x^i$  is smooth if, and only if, the map:

$$\text{id}_{\mathbb{R}} \circ x^i \circ x^{-1} = x^i \circ x^{-1} = \text{proj}_i$$

is smooth in the usual sense, which it certainly is.

### 5.3.1 Classification Of Differentiable Structures

**Definition 5.16** (Diffeomorphism). *Let  $\phi: M \rightarrow N$  be a bijective map between smooth manifolds. If both  $\phi$  and  $\phi^{-1}$  are smooth, then  $\phi$  is said to be a **diffeomorphism**.*

Diffeomorphisms are the structure preserving maps between smooth manifolds.

**Definition 5.17** (Diffeomorphic Manifolds). *Two manifolds  $(M, \mathcal{O}_M, \mathcal{A}_M)$ ,  $(N, \mathcal{O}_N, \mathcal{A}_N)$  are said to be **diffeomorphic** if there exists a diffeomorphism  $\phi: M \rightarrow N$  between them. We write  $M \cong_{\text{diff}} N$ .*

Note that if the differentiable structure is understood (or irrelevant), we typically write  $M$  instead of the triple  $(M, \mathcal{O}_M, \mathcal{A}_M)$ .

*Remark 5.2.* Being diffeomorphic is an equivalence relation. In fact, it is customary to consider diffeomorphic manifolds to be *the same* from the point of view of differential geometry. This is similar to the situation with topological spaces, where we consider homeomorphic spaces to be the same from the point of view of topology. This is typical of all structure preserving maps.

Armed with the notion of diffeomorphism, we can now ask the following question: how many smooth structures on a given topological space are there, up to diffeomorphism? The answer is quite surprising: it depends on the dimension of the manifold!

**Theorem 5.2** (Radon-Moise). *Let  $M$  be a manifold with  $\dim M = 1, 2$ , or  $3$ . Then there is a unique smooth structure on  $M$  up to diffeomorphism.*

Recall that in a previous example, we showed that we can equip  $(\mathbb{R}, \mathcal{O}_{\text{std}})$  with two incompatible atlases  $\mathcal{A}$  and  $\mathcal{B}$ . Let  $\mathcal{A}_{\text{max}}$  and  $\mathcal{B}_{\text{max}}$  be their extensions to maximal atlases, and consider the smooth manifolds  $(\mathbb{R}, \mathcal{O}_{\text{std}}, \mathcal{A}_{\text{max}})$  and  $(\mathbb{R}, \mathcal{O}_{\text{std}}, \mathcal{B}_{\text{max}})$ . Clearly, these are different manifolds, because the atlases are different, but since  $\dim \mathbb{R} = 1$ , they must be diffeomorphic.

The answer to the case  $\dim M > 4$  (we emphasize  $\dim M \neq 4$ ) is provided by *surgery theory*. This is a collection of tools and techniques in topology with which one obtains a new manifold from given ones by performing surgery on them, i.e. by cutting, replacing and gluing parts in such a way as to control topological invariants like the fundamental group. The idea is to understand all manifolds in dimensions higher than 4 by performing surgery systematically. In particular, using surgery theory, it has been shown that there are only finitely many smooth manifolds (up to diffeomorphism) one can make from a topological manifold.

This is not as neat as the previous case, but since there are only finitely many structures, we can still enumerate them, i.e. we can write an exhaustive list.

While finding all the differentiable structures may be difficult for any given manifold, this theorem has an immediate impact on a physical theory that models spacetime as a manifold. For instance, some physicists believe that spacetime should be modelled as a 10-dimensional manifold (we are neither proposing nor condemning this view). If that is indeed the case, we need to worry about which differentiable structure we equip our 10-dimensional manifold with, as each different choice will likely lead to different predictions. But since there are only finitely many such structures, physicists can, at least in principle, devise and perform finitely many experiments to distinguish between them and determine which is the right one, if any.

We now turn to the special case  $\dim M = 4$ . The result is that if  $M$  is a non-compact topological manifold, then there are uncountably many non-diffeomorphic smooth structures that we can equip  $M$  with. In particular, this applies to  $(\mathbb{R}^4, \mathcal{O}_{\text{std}})$ .

## 5.4 Tangent Space

In this section, whenever we say “manifold”, we mean a (real)  $d$ -dimensional differentiable manifold, unless we explicitly say otherwise. We will also suppress the differentiable structure in the notation.

**Definition 5.18** ( $\mathcal{C}^\infty(M)$  Vector Space). *Let  $M$  be a manifold. We define the infinite-dimensional vector space over  $\mathbb{R}$  of all smooth functions on  $M$  with underlying set:*

$$\mathcal{C}^\infty(M) := \{f: M \rightarrow \mathbb{R} \mid f \text{ is smooth}\}$$

*and operations defined pointwise, i.e. for any  $p \in M$ :*

$$\begin{aligned} (f + g)(p) &:= f(p) + g(p) \\ (\lambda f)(p) &:= \lambda f(p) \end{aligned}$$

A routine check shows that this is indeed a vector space.

**Definition 5.19** (Smooth Curve). *A **smooth curve** on  $M$  is a smooth map  $\gamma: \mathbb{R} \rightarrow M$ , where  $\mathbb{R}$  is understood as a 1-dimensional manifold.*

**Definition 5.20** (Directional Derivative Operator). *Let  $\gamma: \mathbb{R} \rightarrow M$  be a smooth curve through  $p \in M$ ; w.l.o.g. let  $\gamma(0) = p$ . The **directional derivative operator** at  $p$  along  $\gamma$  is the linear map:*

$$\begin{aligned} X_{\gamma, p}: \mathcal{C}^\infty(M) &\xrightarrow{\sim} \mathbb{R} \\ f &\mapsto (f \circ \gamma)'(0) \end{aligned}$$

*where  $\mathbb{R}$  is understood as a 1-dimensional vector space over the field  $\mathbb{R}$ .*

Note that  $f \circ \gamma$  is a map  $\mathbb{R} \rightarrow \mathbb{R}$ , hence we can calculate the usual derivative and evaluate it at 0.

**Remark 5.3.** In differential geometry,  $X_{\gamma,p}$  is called the *tangent vector* to the curve  $\gamma$  at the point  $p \in M$ . Intuitively,  $X_{\gamma,p}$  is the velocity  $\dot{\gamma}$  at  $p$ . Consider the curve  $\delta(t) := \gamma(2t)$ , which is the same curve parametrised twice as fast. We have, for any  $f \in \mathcal{C}^\infty(M)$ :

$$X_{\delta,p}(f) = (f \circ \delta)'(0) = 2(f \circ \gamma)'(0) = 2X_{\gamma,p}(f)$$

by using the chain rule. Hence  $X_{\gamma,p}$  scales like a velocity should.

**Definition 5.21** (Tangent Space). *Let  $M$  be a manifold and  $p \in M$ . The **tangent space** to  $M$  at  $p$  is the vector space over  $\mathbb{R}$  with underlying set:*

$$T_p M := \{X_{\gamma,p} \mid \gamma \text{ is a smooth curve through } p\}$$

*addition:*

$$\begin{aligned} \oplus: T_p M \times T_p M &\rightarrow T_p M \\ (X_{\gamma,p}, X_{\delta,p}) &\mapsto X_{\gamma,p} \oplus X_{\delta,p} \end{aligned}$$

*and scalar multiplication:*

$$\begin{aligned} \odot: \mathbb{R} \times T_p M &\rightarrow T_p M \\ (\lambda, X_{\gamma,p}) &\mapsto \lambda \odot X_{\gamma,p} \end{aligned}$$

*both defined pointwise, i.e. for any  $f \in \mathcal{C}^\infty(M)$ :*

$$\begin{aligned} (X_{\gamma,p} \oplus X_{\delta,p})(f) &:= X_{\gamma,p}(f) + X_{\delta,p}(f) \\ (\lambda \odot X_{\gamma,p})(f) &:= \lambda X_{\gamma,p}(f) \end{aligned}$$

Note that the outputs of these operations do not look like elements in  $T_p M$ , because they are not of the form  $X_{\sigma,p}$  for some curve  $\sigma$ . Hence, we need to show that the above operations are, in fact, well-defined.

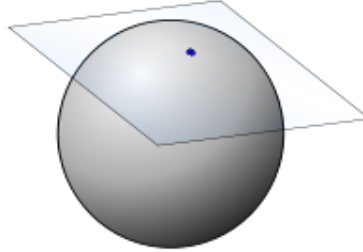


Figure 5.1: A pictorial representation of the tangent space  $T_p M$  of a single point,  $p$ , on a manifold  $M$ . A vector in this  $T_p M$  can represent a possible velocity at  $p$ . After moving in that direction to a nearby point, one's velocity would then be given by a vector in the tangent space of that nearby point, i.e a different tangent space, not shown in here.

**Proposition 5.3.** *Let  $X_{\gamma,p}, X_{\delta,p} \in T_p M$  and  $\lambda \in \mathbb{R}$ . Then, we have  $X_{\gamma,p} \oplus X_{\delta,p} \in T_p M$  and  $\lambda \odot X_{\gamma,p} \in T_p M$ .*

Since the derivative is a local concept, it is only the behaviour of curves near  $p$  that matters. In particular, if two curves  $\gamma$  and  $\delta$  agree on a neighbourhood of  $p$ , then  $X_{\gamma,p}$  and  $X_{\delta,p}$  are the same element of  $T_p M$ . Hence, we can work *locally* by using a chart on  $M$ .

*Proof.*

Let  $(U, x)$  be a chart on  $M$ , with  $U$  a neighbourhood of  $p$ .

i) Define the curve:

$$\sigma(t) := x^{-1}((x \circ \gamma)(t) + (x \circ \delta)(t) - x(p))$$

Note that  $\sigma$  is smooth since it is constructed via addition and composition of smooth maps and, moreover:

$$\begin{aligned}\sigma(0) &= x^{-1}(x(\gamma(0)) + x(\delta(0)) - x(p)) \\ &= x^{-1}(x(p)) + x(p) - x(p) \\ &= x^{-1}(x(p)) \\ &= p\end{aligned}$$

Thus  $\sigma$  is a smooth curve through  $p$ . Let  $f \in \mathcal{C}^\infty(U)$  be arbitrary. Then we have:

$$\begin{aligned}X_{\sigma,p}(f) &:= (f \circ \sigma)'(0) \\ &= [f \circ x^{-1} \circ ((x \circ \gamma) + (x \circ \delta) - x(p))]'(0) \\ &= [\partial_a(f \circ x^{-1})(x(p))]((x^a \circ \gamma) + (x^a \circ \delta) - x^a(p))'(0) \\ &= [\partial_a(f \circ x^{-1})(x(p))]((x^a \circ \gamma)'(0) + (x^a \circ \delta)'(0)) \\ &= (f \circ x^{-1} \circ x \circ \gamma)'(0) + (f \circ x^{-1} \circ x \circ \delta)'(0) \\ &= (f \circ \gamma)'(0) + (f \circ \delta)'(0) \\ &=: (X_{\gamma,p} \oplus X_{\delta,p})(f)\end{aligned}$$

Therefore  $X_{\gamma,p} \oplus X_{\delta,p} = X_{\sigma,p} \in T_p M$ .

- ii) The second part is straightforward. Define  $\sigma(t) := \gamma(\lambda t)$ . This is again a smooth curve through  $p$  and we have:

$$\begin{aligned}X_{\sigma,p}(f) &:= (f \circ \sigma)'(0) \\ &= f'(\sigma(0)) \sigma'(0) \\ &= \lambda f'(\gamma(0)) \gamma'(0) \\ &= \lambda(f \circ \gamma)'(0) \\ &:= (\lambda \odot X_{\gamma,p})(f)\end{aligned}$$

for any  $f \in \mathcal{C}^\infty(U)$ . Hence  $\lambda \odot X_{\gamma,p} = X_{\sigma,p} \in T_p M$ . Hence indeed  $T_p M$  is a vector space.  $\square$

The question is, what exactly  $X_{\gamma,p}$  is mathematically speaking? Since it's a map of the form:

$$X_{\gamma,p}: \mathcal{C}^\infty(M) \xrightarrow{\sim} \mathbb{R}$$

it's clear that it's an element of  $\text{Hom}(\mathcal{C}^\infty(M), \mathbb{R})$ , i.e an element of the dual vector space of  $\mathcal{C}^\infty(M)$ . Which subsequently makes  $T_p M$  a sub-vector space of the dual vector space of  $\mathcal{C}^\infty(M)$ . ( $X_{\gamma,p}$  is a particular choice of a linear map, more specifically the derivative with respect to the parameter, and not **all** possible linear maps. This is why  $T_p M$  is not the whole dual vector space of  $\mathcal{C}^\infty(M)$ )

However, if we take the extra step and turn the  $\mathcal{C}^\infty(M)$  from a vector space to an algebra (by defining an appropriate operation) then we can show that  $X_{\gamma,p}$  is actually a derivation of the algebra.

More specifically we will define a product on  $\mathcal{C}^\infty(M)$  by:

$$\begin{aligned}\bullet: \mathcal{C}^\infty(M) \times \mathcal{C}^\infty(M) &\rightarrow \mathcal{C}^\infty(M) \\ (f, g) &\mapsto f \bullet g\end{aligned}$$

where  $f \bullet g$  is defined pointwise. Then  $(\mathcal{C}^\infty(M), +, \cdot, \bullet)$  is an associative, unital and commutative algebra over  $\mathbb{R}$ .

Now that we have an algebra, let us remind ourselves what a derivation is and also try to combine the definition with our case.

**Definition 5.22** (Derivation (On A Manifold)). *Let  $M$  be a manifold and let  $p \in U \subseteq M$ , where  $U$  is open. A derivation on  $U$  at  $p$  is an  $\mathbb{R}$ -linear map  $D: \mathcal{C}^\infty(U) \xrightarrow{\sim} \mathbb{R}$  satisfying the Leibniz rule:*

$$D(fg) = D(f)g(p) + f(p)D(g)$$

The usual derivative operator is a derivation on  $\mathcal{C}^\infty(\mathbb{R})$ , the algebra of smooth real functions, since it is linear and satisfies the Leibniz rule. (The second derivative operator, however, is not a derivation on  $\mathcal{C}^\infty(\mathbb{R})$ , since it does not satisfy the Leibniz rule. This shows that the composition of derivations need not be a derivation.) Hence, we managed to show that indeed  $X_{\gamma,p}$  is actually a derivation of the algebra of smooth real functions on  $M$ .

### 5.4.1 Co-Ordinate Induced Basis For The Tangent Space

The following is a crucially important result about tangent spaces.

**Theorem 5.3.** *Let  $M$  be a manifold and let  $p \in M$ . Then:*

$$\dim T_p M = \dim M$$

*Remark 5.4.* Note carefully that, despite us using the same symbol, the two “dimensions” appearing in the statement of the theorem are, at least on the surface, entirely unrelated. Indeed, recall that  $\dim M$  is defined in terms of charts  $(U, x)$ , with  $x: U \rightarrow x(U) \subseteq \mathbb{R}^{\dim M}$ , while  $\dim T_p M = |\mathcal{B}|$ , where  $\mathcal{B}$  is a Hamel basis for the vector space  $T_p M$ . The idea behind the proof is to construct a basis of  $T_p M$  from a chart on  $M$ .

*Proof.*

W.l.o.g., let  $(U, x)$  be a chart *centred* at  $p$ , i.e.  $x(p) = 0 \in \mathbb{R}^{\dim M}$ . Define  $(\dim M)$ -many curves  $\gamma_{(a)}: \mathbb{R} \rightarrow U$  through  $p$  by requiring  $(x^b \circ \gamma_{(a)})(t) = \delta_a^b t$ , i.e:

$$\begin{aligned}\gamma_{(a)}(0) &:= p \\ \gamma_{(a)}(t) &:= x^{-1} \circ (0, \dots, 0, t, 0, \dots, 0)\end{aligned}$$

where the  $t$  is in the  $a^{\text{th}}$  position, with  $1 \leq a \leq \dim M$ . Let us calculate the action of the tangent vector  $X_{\gamma_{(a)},p} \in T_p M$  on an arbitrary function  $f \in \mathcal{C}^\infty(U)$ :

$$\begin{aligned}X_{\gamma_{(a)},p}(f) &:= (f \circ \gamma_{(a)})'(0) \\ &= (f \circ \text{id}_U \circ \gamma_{(a)})'(0) \\ &= (f \circ x^{-1} \circ x \circ \gamma_{(a)})'(0) \\ &= [\partial_b (f \circ x^{-1})(x(p))] (x^b \circ \gamma_{(a)})'(0) \\ &= [\partial_b (f \circ x^{-1})(x(p))] (\delta_a^b t)'(0) \\ &= [\partial_b (f \circ x^{-1})(x(p))] \delta_a^b \\ &= \partial_a (f \circ x^{-1})(x(p))\end{aligned}$$

We introduce a special notation for this last line, namely:

$$\partial_a (f \circ x^{-1})(x(p)) := \left( \frac{\partial}{\partial x^a} \right)_p (f)$$

*Remark 5.5.* While the symbol  $\left( \frac{\partial}{\partial x^a} \right)_p$  has nothing to do with the idea of partial differentiation with respect to the variable  $x^a$  (since  $x$  refers to the chart map and no differentiation has been defined there), it is notationally consistent with it, in the following sense.

Let  $M = \mathbb{R}^d$ ,  $(U, x) = (\mathbb{R}^d, \text{id}_{\mathbb{R}^d})$  and let  $\left( \frac{\partial}{\partial x^a} \right)_p \in T_p \mathbb{R}^d$ . If  $f \in \mathcal{C}^\infty(\mathbb{R}^d)$ , then:

$$\left( \frac{\partial}{\partial x^a} \right)_p (f) = \partial_a (f \circ x^{-1})(x(p)) = \partial_a f(p)$$

since  $x = x^{-1} = \text{id}_{\mathbb{R}^d}$ . Moreover, we have  $\text{proj}_a = x^a$ . Thus, we can think of  $x^1, \dots, x^d$  as the independent variables of  $f$ , and we can then write:

$$\left( \frac{\partial}{\partial x^a} \right)_p (f) = \frac{\partial f}{\partial x^a}(p)$$

Hence, up to this point we showed that:

$$X_{\gamma(a),p}(f) = \left( \frac{\partial}{\partial x^a} \right)_p (f)$$

Or by removing the action on the function, simply:

$$X_{\gamma(a),p} = \left( \frac{\partial}{\partial x^a} \right)_p$$

We now claim that:

$$\mathcal{B} = \left\{ \left( \frac{\partial}{\partial x^a} \right)_p \in T_p M \mid 1 \leq a \leq \dim M \right\}$$

is a basis of  $T_p M$ . First, we show that  $\mathcal{B}$  spans  $T_p M$ .

Let  $X_{\gamma,p} \in T_p M$ . For any  $f \in \mathcal{C}^\infty(U)$ , we have:

$$\begin{aligned} X_{\gamma,p}(f) &:= (f \circ \sigma)'(0) \\ &= (f \circ x^{-1} \circ x \circ \gamma)'(0) \\ &= [\partial_b(f \circ x^{-1})(x(p))] (x^b \circ \gamma)'(0) \\ &= (x^b \circ \gamma)'(0) \left( \frac{\partial}{\partial x^b} \right)_p (f) \end{aligned}$$

Since  $(x^b \circ \gamma)'(0) =: X^b \in \mathbb{R}$ , we have:

$$X_{\gamma,p} = X^b \left( \frac{\partial}{\partial x^b} \right)_p$$

i.e. any  $X_{\gamma,p} \in T_p M$  is a linear combination of elements from  $\mathcal{B}$ .

To show linear independence, suppose that:

$$\lambda^a \left( \frac{\partial}{\partial x^a} \right)_p = 0$$

for some scalars  $\lambda^a$ . Note that this is an operator equation, and the zero on the right hand side is the zero operator  $0 \in T_p M$ .

Recall that, given the chart  $(U, x)$ , the coordinate maps  $x^b: U \rightarrow \mathbb{R}$  are smooth, i.e.  $x^b \in \mathcal{C}^\infty(U)$ . Thus, we can feed them into the left hand side to obtain:

$$\begin{aligned} 0 &= \lambda^a \left( \frac{\partial}{\partial x^a} \right)_p (x^b) \\ &= \lambda^a \partial_a (x^b \circ x^{-1})(x(p)) \\ &= \lambda^a \partial_a (\text{proj}_b)(x(p)) \\ &= \lambda^a \delta_a^b \\ &= \lambda^b \end{aligned}$$

i.e.  $\lambda^b = 0$  for all  $1 \leq b \leq \dim M$ . So  $\mathcal{B}$  is indeed a basis of  $T_p M$ , and since by construction  $|\mathcal{B}| = \dim M$ ,

the proof is complete.  $\square$

*Remark 5.6.* While it is possible to define infinite-dimensional manifolds, in this course we will only consider finite-dimensional ones. Hence  $\dim T_p M = \dim M$  will always be finite in this course.

*Remark 5.7.* Note that the basis that we have constructed in the proof is *not* chart-independent. Indeed, each different chart will induce a different tangent space basis, and we distinguish between them by keeping the chart map in the notation for the basis elements.

This is not a cause of concern for our proof however, since every basis of a vector space must have the same cardinality, and hence it suffices to find one basis to determine the dimension.

**Definition 5.23** (Co-Ordinate Induced Basis). *Let  $X_{\gamma,p} \in T_p M$  be a tangent vector and let  $(U, x)$  be a chart containing  $p$ . Then the basis  $\{(\frac{\partial}{\partial x^a})_p\}$  created by the usage of the chart is called a **co-ordinate induced basis**. In this basis an element  $X_{\gamma,p} \in T_p M$  can be expressed as:*

$$X_{\gamma,p} = X^a \left( \frac{\partial}{\partial x^a} \right)_p$$

where the real numbers  $X^1, \dots, X^{\dim M}$  are called the **vector components** of  $X_{\gamma,p}$  with respect to the co-ordinate induced basis by the chart  $(U, x)$ .

### 5.4.2 Change Of Vector Components Under A Change Of Chart

One of the most heavily used concepts is the transformation of the components of a vector under different co-ordinate systems (i.e under a chart transition map that subsequently changes the co-ordinate induced basis). Let's find out the rule.

Let  $X_{\gamma,p} \in T_p M$  and let  $(U, x)$  and  $(V, y)$  be two charts containing  $p$ . Then  $X_{\gamma,p}$  can be expressed in any of the two charts as:

$$X^a_{(y)} \left( \frac{\partial}{\partial y^a} \right)_p = X_{\gamma,p} = X^a_{(x)} \left( \frac{\partial}{\partial x^a} \right)_p$$

Let us act with  $X_{\gamma,p}$  on some smooth function  $f$  of  $\mathcal{C}^\infty(M)$  by using first the components of  $(U, x)$  chart:

$$\begin{aligned} X_{\gamma,p}(f) &= X^a_{(x)} \left( \frac{\partial}{\partial x^a} \right)_p (f) \\ &= X^a_{(x)} \partial_a (f \circ x^{-1})(x(p)) \\ &= X^a_{(x)} \partial_a (f \circ y^{-1} \circ y \circ x^{-1})(x(p)) \\ &= X^a_{(x)} \partial_a (y^b \circ x^{-1})(x(p)) \partial_b (f \circ y^{-1})(y(p)) \\ &= X^a_{(x)} \frac{\partial y^b}{\partial x^a} \left( \frac{\partial}{\partial y^b} \right)_p (f) \end{aligned}$$

Similarly, let us now act with  $X_{\gamma,p}$  on the smooth function  $f$  of  $\mathcal{C}^\infty(M)$  by using the components of  $(V, y)$  chart:

$$X_{\gamma,p}(f) = X^a_{(y)} \left( \frac{\partial}{\partial y^a} \right)_p (f)$$

These expressions are, of course, equal to each other so by suppressing now the action on the function  $f$ ,

we obtain:

$$\begin{aligned} X^a_{(x)} \frac{\partial y^b}{\partial x^a} \left( \frac{\partial}{\partial y^b} \right)_p &= X^b_{(y)} \left( \frac{\partial}{\partial y^b} \right)_p \\ X^a_{(x)} \frac{\partial y^b}{\partial x^a} \left( \frac{\partial}{\partial y^b} \right)_p - X^b_{(y)} \left( \frac{\partial}{\partial y^b} \right)_p &= 0 \\ \left( X^a_{(x)} \frac{\partial y^b}{\partial x^a} - X^b_{(y)} \right) \left( \frac{\partial}{\partial y^b} \right)_p &= 0 \end{aligned}$$

Finally, since the base vectors of  $\left\{ \left( \frac{\partial}{\partial y^a} \right)_p \right\}$  are linearly independent the only way for this equation to be zero is for the coefficients to be zero hence:

$$X^a_{(x)} \frac{\partial y^b}{\partial x^a} - X^b_{(y)} = 0$$

Of finally by solving w.r.t  $X^b_{(y)}$  and renaming the indices:

$$X^a_{(y)} = \frac{\partial y^a}{\partial x^b} X^b_{(x)}$$

This equation shows as how the components of a vector transform under a chart transition map, i.e under the change of charts, i.e from one co-ordinate induced basis to another. Of course the formula agrees completely with the transformations of vector components under the change of basis that we showed in previous chapter:  $\tilde{v}^b = A^b_a v^a$ .

The function  $y^a = y^a(x^1, \dots, x^{\dim M})$  expresses the new co-ordinates in terms of the old ones, and  $A^b_a$  is the *Jacobian* matrix of this map, evaluated at  $x(p)$ . Note that no matter how non-linear the transformations of the co-ordinates are, the vectors always transform in a linear fashion. In a way, “vectors do not care about the non-linearity of co-ordinate transformations”.

## 5.5 Cotangent Space

Since the tangent space is a vector space, we can do all the constructions we saw previously in the abstract vector space setting.

**Definition 5.24** (Cotangent Space). *Let  $M$  be a manifold and  $p \in M$ . The **cotangent space** to  $M$  at  $p$  is:*

$$T_p^* M := (T_p M)^*$$

Since  $\dim T_p M$  is finite, we have  $T_p M \cong_{\text{vec}} T_p^* M$ .

And of course, once we have the cotangent space, we can define the tensor spaces.

**Definition 5.25** (Tensor Space). *Let  $M$  be a manifold and  $p \in M$ . The **tensor space**  $(T_s^r)_p M$  is defined as:*

$$(T_s^r)_p M := T_s^r(T_p M) = \underbrace{T_p M \otimes \dots \otimes T_p M}_{r \text{ copies}} \otimes \underbrace{T_p^* M \otimes \dots \otimes T_p^* M}_{s \text{ copies}}$$

### 5.5.1 Dual Basis For The Cotangent Space

Now let's give a very important definition that will help us to formalize elements, and subsequently a basis, for the cotangent space.

**Definition 5.26** (Gradient). *Let  $M$  be a manifold and let  $f: M \rightarrow \mathbb{R}$  be smooth. The **gradient of  $f$**  at  $p \in M$  is the  $\mathbb{R}$ -linear map:*

$$\begin{aligned} d_p: \mathcal{C}^\infty(M) &\xrightarrow{\sim} T_p^* M \\ f &\mapsto d_p f \end{aligned}$$

with  $p \in U \subseteq M$ , defined by:

$$d_p f(X_{\gamma,p}) := X_{\gamma,p}(f)$$

*Remark 5.8.* Note that since  $d_p$  is a map from  $\mathcal{C}^\infty(M) \xrightarrow{\sim} T_p^*M$  that means that when it acts on a function of  $\mathcal{C}^\infty(M)$  the final result  $d_p f$  is an element of  $T_p^*M$  hence a covector. By its turn, as an element of the dual space of  $T_p M$  it maps elements of  $T_p M$  to the real numbers (that's the definition of the dual space of a vector space). Hence the expression  $d_p f(X)$  must end up to a real number, which is indeed what  $X_{\gamma,p}(f)$  is. By writing  $d_p f(X) := X_{\gamma,p}(f)$ , we have committed a slight (but nonetheless real) abuse of notation, since  $d_p f(X) \in T_{f(p)}\mathbb{R}$  takes in a function and return a real number, but  $X_{\gamma,p}(f)$  is already a real number! However by doing so we can now talk about  $d_p f$  without providing the vector that it acts on. In other words we can talk about covectors without the need of their actions on vectors.

*Remark 5.9.* The gradient of a function is a covector and **not** a vector.

Recall that if  $(U, x)$  is a chart on  $M$ , then the co-ordinate maps  $x^a: U \rightarrow x(U) \subseteq \mathbb{R}^{\dim M}$  are smooth functions on  $U$  hence they belong to  $\mathcal{C}^\infty(M)$ . We can thus apply the gradient operator  $d_p$  (with  $p \in U$ ) to each of them to obtain  $(\dim M)$ -many elements of  $T_p^*M$ .

**Proposition 5.4.** *Let  $(U, x)$  be a chart on  $M$ , with  $p \in U$ . The set  $\mathcal{B} = \{d_p x^a \mid 1 \leq a \leq \dim M\}$  forms the dual basis of  $T_p^*M$ .*

*Proof.*

By simply acting on  $(\frac{\partial}{\partial x^a})_p$  with  $d_p x^a$  (in our notation, we have  $(dx^a)_p = d_p x^a$ ) we obtain:

$$\begin{aligned} d_p x^a \left( \left( \frac{\partial}{\partial x^b} \right)_p \right) &= \left( \frac{\partial}{\partial x^b} \right)_p (x^a) && \text{(definition of } d_p x^a) \\ &= \partial_b (x^a \circ x^{-1})(x(p)) && \text{(definition of } (\frac{\partial}{\partial x^b})_p) \\ &= \partial_b (\text{proj}_a)(x(p)) \\ &= \delta_b^a \end{aligned}$$

Therefore,  $\mathcal{B}$  is, in fact, the dual basis to  $\{(\frac{\partial}{\partial x^a})_p\}$ . □

### 5.5.2 Change Of Covector Components Under A Change Of Chart

Once again, as we did in the vector case with the vector components, one needs to find the transformation of the components of a covector under different co-ordinate systems. We will follow exactly the same procedure.

Let  $\omega_p \in T_p^*M$  and let  $(U, x)$  and  $(V, y)$  be two charts containing  $p$ . Then  $\omega_p$  can be expressed in any of the two charts by using the dual basis as:

$$\omega_{(y)a}(dy^a)_p = \omega_p = \omega_{(x)a}(dx^a)_p$$

By repeating the same process as we did for the vectors it is very easy to show that covectors components transform as:

$$\omega_{(y)a} = \left( \frac{\partial x^b}{\partial y^a} \right)_p \omega_{(x)b}$$

## 5.6 Push-Forward & Pull-Back

**Definition 5.27** (Push-Forward). *Let  $\phi: M \rightarrow N$  be a smooth map between smooth manifolds. The **push-forward** (or **derivative**) of  $\phi$  at  $p \in M$  is the linear map  $(\phi_*)_p$ :*

$$\begin{aligned} (\phi_*)_p: T_p M &\xrightarrow{\sim} T_{\phi(p)} N \\ X_{\gamma,p} &\mapsto (\phi_*)_p(X_{\gamma,p}) \end{aligned}$$

where  $(\phi_*)_p(X_{\gamma,p})$  is defined as:

$$\begin{aligned} (\phi_*)_p(X_{\gamma,p}) : \mathcal{C}^\infty(N) &\xrightarrow{\sim} \mathbb{R} \\ f &\mapsto (\phi_*)_p(X_{\gamma,p})f := X_{\gamma,p}(f \circ \phi) \end{aligned}$$

In other words, since  $(\phi_*)_p$  is a map from one tangent space to another this means that it acts on a tangent vector and produces another one, hence  $(\phi_*)_p(X_{\gamma,p})$  is again a tangent vector (but on  $N$ ). As a tangent vector it can act on a smooth function (again on  $N$ ) and produce a real number, hence the action of a push-forward on a function is simply the one we wrote above.

$$\begin{array}{ccccc} T_p M & \xrightarrow{(\phi_*)_p} & T_{\phi(p)} N & & \\ \downarrow & & \downarrow & & \\ M & \xrightarrow{\phi} & N & \xrightarrow{f} & \mathbb{R} \\ & \searrow f \circ \phi & & \nearrow & \end{array}$$

Figure 5.2:  $(\phi_*)_p$  takes a vector  $X_p \in T_p M$  in the tangent space at the point  $p \in M$  to the vector  $(\phi_*)_p(X_{\gamma,p}) \in T_{\phi(p)} N$  in the tangent space at the point  $\phi(p) \in N$ , such that the action of  $(\phi_*)_p(X_{\gamma,p})$  on any smooth function  $f \in \mathcal{C}^\infty(N)$  results in the same value as the action of  $X_p$  on the function  $(f \circ \phi)$ .

Note that one has to define a push-forward  $(\phi_*)_p$  for every point  $p$  of  $M$ . Although we have only one map  $\phi$  we have many push-forward maps  $(\phi_*)_p$ .

One can compute the components of a push-forward  $(\phi_*)_p$  w.r.t charts  $(U, x) \in \mathcal{A}_M$  and  $(V, y) \in \mathcal{A}_N$ . Let  $p \in U$  and  $\phi(p) \in V$ . Since  $(\frac{\partial}{\partial x^i})_p$  is a vector in  $M$ , we have  $(\phi_*)_p((\frac{\partial}{\partial x^i})_p)$  as a vector in  $N$ . Then we can select a component of this vector by using  $dy^a$  as follows:

$$\begin{aligned} ((\phi_*)_p)_i^a &= dy^a \left( (\phi_*)_p \left( \left( \frac{\partial}{\partial x^i} \right)_p \right) \right) \\ &= (\phi_*)_p \left( \left( \frac{\partial}{\partial x^i} \right)_p \right) y^a \\ &= \left( \frac{\partial}{\partial x^i} \right)_p (y^a \circ \phi) \\ &= \left( \frac{\partial}{\partial x^i} \right)_p (y \circ \phi)^a \\ &= \left( \frac{\partial \hat{\phi}^a}{\partial x^i} \right)_p \end{aligned}$$

Diagrammatically:

$$\begin{array}{ccc} M \supseteq U & \xrightarrow{\phi} & V \subseteq N \\ \downarrow x & \searrow y \circ \phi =: \hat{\phi} & \downarrow y \\ \underbrace{x(U)}_{\subseteq \mathbb{R}^{\dim M}} & \xrightarrow{y \circ \phi \circ x^{-1}} & \underbrace{y(V)}_{\subseteq \mathbb{R}^{\dim N}} \end{array}$$

**Proposition 5.5.** Let  $\phi: M \rightarrow N$  be smooth. The tangent vector  $X_{\gamma,p} \in T_p M$  is pushed forward to the

tangent vector  $X_{\phi \circ \gamma, \phi(p)} \in T_{\phi(p)}N$ , i.e.:

$$(\phi_*)_p(X_{\gamma,p}) = X_{\phi \circ \gamma, \phi(p)}$$

*Proof.*

Let  $f \in \mathcal{C}^\infty(V)$ , with  $(V, x)$  a chart on  $N$  and  $\phi(p) \in V$ . By applying the definitions, we have:

$$\begin{aligned} (\phi_*)_p(X_{\gamma,p})(f) &= (X_{\gamma,p})(f \circ \phi) && \text{(definition of } (\phi_*)_p) \\ &= ((f \circ \phi) \circ \gamma)'(0) && \text{(definition of } X_{\gamma,p}) \\ &= (f \circ (\phi \circ \gamma))'(0) && \text{(associativity of } \circ) \\ &= X_{\phi \circ \gamma, \phi(p)}(f) && \text{(definition of } X_{\phi \circ \gamma, \phi(p)}) \end{aligned}$$

Since  $f$  was arbitrary, we have  $(\phi_*)_p(X_{\gamma,p}) = X_{\phi \circ \gamma, \phi(p)}$ . □

Related to the push-forward, there is the notion of pull-back of a smooth map.

**Definition 5.28** (Pull-Back). *Let  $\phi: M \rightarrow N$  be a smooth map between smooth manifolds. The **pull-back** of  $\phi$  at  $p \in M$  is the linear map:*

$$\begin{aligned} (\phi^*)_p: T_{\phi(p)}^*N &\xrightarrow{\sim} T_p^*M \\ \omega_{\phi(p)} &\mapsto (\phi^*)_p(\omega_{\phi(p)}) \end{aligned}$$

where  $(\phi^*)_p(\omega_{\phi(p)})$  is defined as:

$$\begin{aligned} (\phi^*)_p(\omega_{\phi(p)}): T_p M &\xrightarrow{\sim} \mathbb{R} \\ X_{\gamma,p} &\mapsto (\phi^*)_p(\omega_{\phi(p)})(X_{\gamma,p}) := \omega_{\phi(p)}((\phi_*)_p(X_{\gamma,p})) \end{aligned}$$

In words, if  $\omega_{\phi(p)}$  is a covector on  $N$ , its pull-back  $(\phi^*)_p(\omega_{\phi(p)})$  is a covector on  $M$ . It acts on tangent vectors on  $M$  by first pushing them forward to tangent vectors on  $N$ , and then applying  $\omega_{\phi(p)}$  to them to produce a real number.

As before, one can compute the components of a pull-back  $(\phi^*)_p$  w.r.t charts  $(U, x) \in \mathcal{A}_M$  and  $(V, y) \in \mathcal{A}_N$ . Let  $p \in U$  and  $\phi(p) \in V$ . Since  $(dy^a)_{\phi(p)}$  is a covector in  $N$ , we have  $(\phi^*)_p(dy^a_{\phi(p)})$  as a covector in  $M$ . Then we can select a component of this covector by using  $(\frac{\partial}{\partial x^i})_p$  as follows:

$$\begin{aligned} ((\phi^*)_p)_a &= (\phi^*)_p \left( (dy^a)_{\phi(p)} \right) \left( \left( \frac{\partial}{\partial x^i} \right)_p \right) \\ &= (dy^a)_{\phi(p)} (\phi_*)_p \left( \left( \frac{\partial}{\partial x^i} \right)_p \right) \\ &= \left( \frac{\partial \hat{\phi}^a}{\partial x^i} \right)_p \\ &= ((\phi_*)_p)_i^a \end{aligned}$$

Thus, the components of the push-forward and pull-back maps are exactly the same:

$$\begin{aligned} ((\phi_*)_p(X))^a &= ((\phi_*)_p)_i^a X^i \\ ((\phi^*)_p(\omega))_i &= ((\phi_*)_p)_i^a \omega_a \end{aligned}$$

Diagrammatically, what we've defined so far is the following:

$$\begin{array}{ccc}
\mathcal{C}^\infty(M) & \xleftarrow{-\circ\phi} & \mathcal{C}^\infty(N) \\
\downarrow X_{\gamma,p} & \searrow (\phi_*)_p(X_{\gamma,p}) & \\
\mathbb{R} & & 
\end{array}
\qquad
\begin{array}{ccc}
T_p M & \xrightarrow{(\phi_*)_p} & T_{\phi(p)} N \\
\searrow (\phi^*)_p(\omega_{\phi(p)}) & & \downarrow \omega_{\phi(p)} \\
& & \mathbb{R}
\end{array}$$

*Remark 5.10.* It is quite easy to show that everything we have defined in this section is, in fact, linear.

*Remark 5.11.* We have seen that, given a smooth  $\phi: M \rightarrow N$ , we can push a vector  $X_{\gamma,p} \in T_p M$  forward to a vector  $(\phi_*)_p(X_{\gamma,p}) \in T_{\phi(p)} N$ , and pull a covector  $\omega_{\phi(p)} \in T_{\phi(p)}^* N$  back to a covector  $(\phi^*)_p(\omega_{\phi(p)}) \in T_p^* M$ . In other words both push-forward and pull-back work only in the direction of their definition. However, if  $\phi: M \rightarrow N$  is a diffeomorphism (and only then), we can also pull a vector  $Y_{\gamma,\phi(p)} \in T_{\phi(p)} N$  back to a vector  $(\phi^*)_p(Y_{\gamma,\phi(p)}) \in T_p M$ , and push a covector  $\eta_p \in T_p^* M$  forward to a covector  $(\phi_*)_p(\eta_p) \in T_{\phi(p)}^* N$ , by using  $\phi^{-1}$  as follows:

$$\begin{aligned}
(\phi^*)_p(Y_{\gamma,\phi(p)}) &:= ((\phi^{-1})^*)_{\phi(p)}(Y_{\gamma,\phi(p)}) \\
(\phi_*)_p(\eta_p) &:= ((\phi^{-1})^*)_{\phi(p)}(\eta_p)
\end{aligned}$$

In general, we should keep in mind that:

*Vectors are pushed forward,  
covectors are pulled back.*

### 5.6.1 Immersions & Embeddings

We will now consider the question of under which circumstances a smooth manifold can “sit” in  $\mathbb{R}^d$ , for some  $d \in \mathbb{N}$ . There are, in fact, two notions of sitting inside another manifold, called immersion and embedding.

**Definition 5.29** (Immersion). *A smooth map  $\phi: M \rightarrow N$  is said to be an **immersion** of  $M$  into  $N$  if the push-forward:*

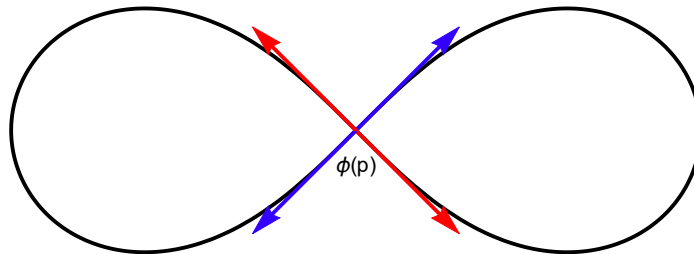
$$(\phi_*)_p: T_p M \xrightarrow{\sim} T_{\phi(p)} N$$

*is injective, for all  $p \in M$ . In that case, the manifold  $M$  is said to be an immersed submanifold of  $N$ .*

From the theory of linear algebra, we immediately deduce that, for  $\phi: M \rightarrow N$  to be an immersion, we must have  $\dim M \leq \dim N$ . A closely related notion is that of a *submersion*, where we require each  $(\phi_*)_p$  to be surjective, and thus we must have  $\dim M \geq \dim N$ . However, we will not need this here.

*Example 5.9.*

Consider the map  $\phi: S^1 \rightarrow \mathbb{R}^2$  whose image is reproduced below.



The map  $\phi$  is not injective, i.e. there are  $p, q \in S^1$ , with  $p \neq q$  and  $\phi(p) = \phi(q)$ . Of course, this means that  $T_{\phi(p)} \mathbb{R}^2 = T_{\phi(q)} \mathbb{R}^2$ . However, the maps  $(\phi_*)_p$  and  $(\phi_*)_q$  are both injective, with their images being represented by the blue and red arrows, respectively. Hence, the map  $\phi$  is immersion.

**Definition 5.30** (Embedding). *A smooth map  $\phi: M \rightarrow N$  is said to be a **embedding** of  $M$  into  $N$  if:*

- $\phi: M \rightarrow N$  is an immersion.
- $M \cong_{\text{top}} \phi(M) \subseteq N$ , where  $\phi(M)$  carries the subset topology inherited from  $N$ .

In that case the manifold  $M$  is said to be an embedded submanifold of  $N$ .

*Remark 5.12.* If a continuous map between topological spaces satisfies the second condition above, then it is called a *topological embedding*. Therefore, an embedding is a topological embedding which is also an immersion (as opposed to simply being a topological embedding).

In the early days of differential geometry there were two approaches to study manifolds. One was the extrinsic view, within which manifolds are defined as special subsets of  $\mathbb{R}^d$ , and the other was the intrinsic view, which is the view that we have adopted here.

Whitney's theorem, which we will state without proof, states that these two approaches are, in fact, equivalent.

**Theorem 5.4** (Whitney). *Any smooth manifold  $M$  can be:*

- Embedded in  $\mathbb{R}^{2 \dim M}$ .
- Immersed in  $\mathbb{R}^{2 \dim M - 1}$ .

*Example 5.10.*

The Klein bottle can be embedded in  $\mathbb{R}^4$  but not in  $\mathbb{R}^3$ . It can, however, be immersed in  $\mathbb{R}^3$ .

## 5.7 The Tangent Bundle

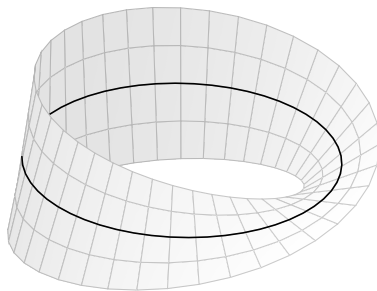
Up to this point, we have defined everything on the level of a point on the manifold. However, since we are interested in describing quantities as a whole in the entire manifold, we would like to define a vector field on a manifold  $M$  as a “smooth” map that assigns to each  $p \in M$  a tangent vector in  $T_p M$ . However, since this would then be a “map” to a different space at each point, it is unclear how to define its smoothness.

The simplest solution is to merge all the tangent spaces into a unique set and equip it with a smooth structure, so that we can then define a vector field as a smooth map between smooth manifolds.

In order to do so, we will need something called “topological bundles”. We will start by defining them more generally and then we will use them for topological manifolds.

### 5.7.1 Topological Bundles

While topological products are very useful, very often one intuitively thinks of the product of two manifolds as attaching a copy of the second manifold to each point of the first. However, not all interesting manifolds can be understood as products of manifolds. A classic example of this is the *Möbius strip*.



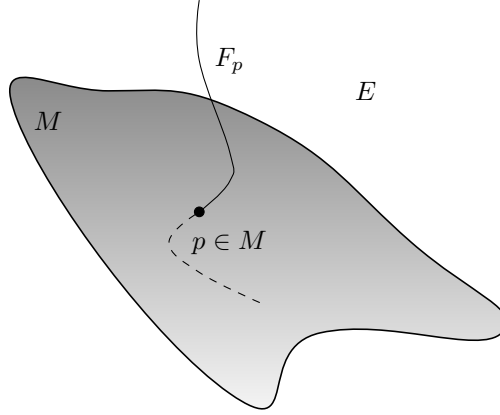
It looks locally like the finite cylinder  $S^1 \times [0, 1]$ , which we can picture as the circle  $S^1$  (the thicker line in figure) with the finite interval  $[0, 1]$  attached to each of its points in a “smooth” way. The Möbius strip has a “twist”, which makes it globally different from the cylinder.

**Definition 5.31** (Topological Bundles). A topological **bundle** (of topological manifolds) is a triple  $(E, \pi, M)$  where  $E$  and  $M$  are topological manifolds called the total space and the base space respectively, and  $\pi$  is a continuous, surjective map  $\pi: E \rightarrow M$  called the projection map.

We will often denote the bundle  $(E, \pi, M)$  by  $E \xrightarrow{\pi} M$ .

**Definition 5.32** (Fiber). Let  $E \xrightarrow{\pi} M$  be a bundle and let  $p \in M$ . Then,  $F_p := \text{preim}_{\pi}(\{p\})$  is called the **fiber** at the point  $p$ .

Intuitively, the fiber at the point  $p \in M$  is a set of points in  $E$  (represented below as a line) attached to the point  $p$ . The projection map sends all the points in the fiber  $F_p$  to the point  $p$ .



*Example 5.11.*

A trivial example of a bundle is the *product bundle*. Let  $M$  and  $N$  be manifolds. Then, the triple  $(M \times N, \pi, M)$ , where:

$$\begin{aligned} \pi: M \times N &\rightarrow M \\ (p, q) &\mapsto p \end{aligned}$$

is a bundle since (one can easily check)  $\pi$  is a continuous and surjective map. Similarly,  $(M \times N, \pi, N)$  with the appropriate  $\pi$ , is also a bundle.

*Example 5.12.*

In a bundle, different points of the base manifold may have (topologically) different fibers. For example, consider the bundle  $E \xrightarrow{\pi} \mathbb{R}$  where:

$$F_p := \text{preim}_{\pi}(\{p\}) \cong_{\text{top}} \begin{cases} S^1 & \text{if } p < 0 \\ \{p\} & \text{if } p = 0 \\ [0, 1] & \text{if } p > 0 \end{cases}$$

**Definition 5.33** (Fiber Bundle). Let  $E \xrightarrow{\pi} M$  be a bundle and let  $F$  be a manifold. Then,  $E \xrightarrow{\pi} M$  is called a **fiber bundle**, with (typical) fiber  $F$ , if:

$$\forall p \in M : \text{preim}_{\pi}(\{p\}) \cong_{\text{top}} F$$

A fiber bundle is often represented diagrammatically as:

$$\begin{array}{ccc} F & \longrightarrow & E \\ & & \downarrow \pi \\ & & M \end{array}$$

*Example 5.13.*

The bundle  $M \times N \xrightarrow{\pi} M$  is a fiber bundle with fiber  $F := N$ .

*Example 5.14.*

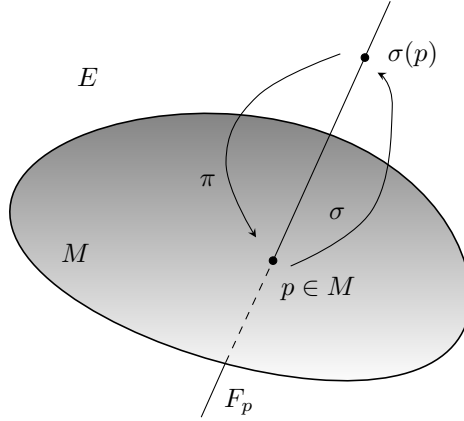
The Möbius strip is a fiber bundle  $E \xrightarrow{\pi} S^1$ , with fiber  $F := [0, 1]$ , where  $E \neq S^1 \times [0, 1]$ , i.e. the Möbius strip is not a product bundle.

*Example 5.15.*

A  $\mathbb{C}$ -line bundle over  $M$  is the fiber bundle  $(E, \pi, M)$  with fiber  $\mathbb{C}$ . Note that the product bundle  $(M \times \mathbb{C}, \pi, M)$  is a  $\mathbb{C}$ -line bundle over  $M$ , but a  $\mathbb{C}$ -line bundle over  $M$  need not be a product bundle.

**Definition 5.34** (Section). *Let  $E \xrightarrow{\pi} M$  be a bundle. A map  $\sigma: M \rightarrow E$  is called a **section** of the bundle if  $\pi \circ \sigma = \text{id}_M$ .*

Intuitively, a section is a map  $\sigma$  which sends each point  $p \in M$  to *some* point  $\sigma(p)$  in its fiber  $F_p$ , so that the projection map  $\pi$  takes  $\sigma(p) \in F_p \subseteq E$  back to the point  $p \in M$ .



*Example 5.16.*

Let  $(M \times F, \pi, M)$  be a product bundle. Then, a section of this bundle is a map:

$$\begin{aligned} \sigma: M &\rightarrow M \times F \\ p &\mapsto (p, s(p)) \end{aligned}$$

where  $s: M \rightarrow F$  is any map.

**Definition 5.35** (Sub-Bundle). *A **sub-bundle** of a bundle  $(E, \pi, M)$  is a triple  $(E', \pi', M')$  where  $E' \subseteq E$  and  $M' \subseteq M$  are submanifolds and  $\pi' := \pi|_{E'}$ .*

**Definition 5.36** (Restricted Bundle). *Let  $(E, \pi, M)$  be a bundle and let  $N \subseteq M$  be a submanifold. The **restricted bundle** (to  $N$ ) is the triple  $(E, \pi', N)$  where:*

$$\pi' := \pi|_{\text{preim}_\pi(N)}$$

**Definition 5.37** (Bundle Morphism). *Let  $E \xrightarrow{\pi} M$  and  $E' \xrightarrow{\pi'} M'$  be bundles and let  $u: E \rightarrow E'$  and  $v: M \rightarrow M'$  be maps. Then  $(u, v)$  is called a **bundle morphism** if the following diagram commutes:*

$$\begin{array}{ccc} E & \xrightarrow{u} & E' \\ \downarrow \pi & & \downarrow \pi' \\ M & \xrightarrow{v} & M' \end{array}$$

*i.e. if  $\pi' \circ u = v \circ \pi$ .*

If  $(u, v)$  and  $(u, v')$  are both bundle morphisms, then  $v = v'$ . That is, given  $u$ , if there exists  $v$  such that  $(u, v)$  is a bundle morphism, then  $v$  is unique.

**Definition 5.38** (Isomorphic Bundles). Two bundles  $E \xrightarrow{\pi} M$  and  $E' \xrightarrow{\pi'} M'$  are said to be **isomorphic (as bundles)** if there exist bundle morphisms  $(u, v)$  and  $(u^{-1}, v^{-1})$  satisfying:

$$\begin{array}{ccc} E & \xrightleftharpoons[u^{-1}]{u} & E' \\ \pi \downarrow & & \downarrow \pi' \\ M & \xrightleftharpoons[v^{-1}]{v} & M' \end{array}$$

Such a  $(u, v)$  is called a bundle isomorphism and we write  $E \xrightarrow{\pi} M \cong_{\text{bdl}} E' \xrightarrow{\pi'} M'$ .

Bundle isomorphisms are the structure-preserving maps for bundles.

**Definition 5.39** (Locally Isomorphic Bundles). A bundle  $E \xrightarrow{\pi} M$  is said to be **locally isomorphic (as a bundle)** to a bundle  $E' \xrightarrow{\pi'} M'$  if for all  $p \in M$  there exists a neighbourhood  $U(p)$  such that the restricted bundle:

$$\text{preim}_{\pi}(U(p)) \xrightarrow{\pi|_{\text{preim}_{\pi}(U(p))}} U(p)$$

is isomorphic to the bundle  $E' \xrightarrow{\pi'} M'$ .

**Definition 5.40** (Trivial / Locally Trivial Bundle). A bundle  $E \xrightarrow{\pi} M$  is said to be:

- i) **Trivial** if it is isomorphic to a product bundle.
- ii) **Locally trivial** if it is locally isomorphic to a product bundle.

*Example 5.17.*

The cylinder  $C$  is trivial as a bundle, and hence also locally trivial.

*Example 5.18.*

The Möbius strip is not trivial but it is locally trivial.

From now on, we will mostly consider locally trivial bundles.

**Remark 5.13.** In quantum mechanics, what is usually called a “wave function” is not a function at all, but rather a section of a  $\mathbb{C}$ -line bundle over physical space. However, if we assume that the  $\mathbb{C}$ -line bundle under consideration is locally trivial, then each section of the bundle can be represented (locally) by a map from the base space to the total space and hence it is appropriate to use the term “wave function”.

**Definition 5.41** (Pull-Back Bundle). Let  $E \xrightarrow{\pi} M$  be a bundle and let  $f: M' \rightarrow M$  be a map from some manifold  $M'$ . The **pull-back bundle of  $E \xrightarrow{\pi} M$**  induced by  $f$  is defined as  $E' \xrightarrow{\pi'} M'$ , where:

$$E' := \{(m', e) \in M' \times E \mid f(m') = \pi(e)\}$$

and  $\pi'(m', e) := m'$ .

If  $E' \xrightarrow{\pi'} M'$  is the pull-back bundle of  $E \xrightarrow{\pi} M$  induced by  $f$ , then one can easily construct a bundle morphism by defining:

$$\begin{aligned} u: E' &\rightarrow E \\ (m', e) &\mapsto e \end{aligned}$$

This corresponds to the diagram:

$$\begin{array}{ccc} E' & \xrightarrow{u} & E \\ \pi' \downarrow & & \downarrow \pi \\ M' & \xrightarrow{f} & M \end{array}$$

*Remark 5.14.* Sections on a bundle pull back to the pull-back bundle. Indeed, let  $E' \xrightarrow{\pi'} M'$  be the pull-back bundle of  $E \xrightarrow{\pi} M$  induced by  $f$ .

$$\begin{array}{ccc} E' & & E \\ \uparrow \sigma' & \nearrow \sigma \circ f & \uparrow \sigma \\ \downarrow \pi' & & \downarrow \pi \\ M' & \xrightarrow{f} & M \end{array}$$

If  $\sigma$  is a section of  $E \xrightarrow{\pi} M$ , then  $\sigma \circ f$  determines a map from  $M'$  to  $E$  which sends each  $m' \in M'$  to  $\sigma(f(m')) \in E$ . However, since  $\sigma$  is a section, we have:

$$\pi(\sigma(f(m'))) = (\pi \circ \sigma \circ f)(m') = (\text{id}_M \circ f)(m') = f(m')$$

and hence  $(m', (\sigma \circ f)(m')) \in E'$  by definition of  $E'$ . Moreover:

$$\pi'(m', (\sigma \circ f)(m')) = m'$$

and hence the map:

$$\begin{aligned} \sigma' : M' &\rightarrow E' \\ m' &\mapsto (m', (\sigma \circ f)(m')) \end{aligned}$$

satisfies  $\pi' \circ \sigma' = \text{id}_{M'}$  and it is thus a section on the pull-back bundle  $E' \xrightarrow{\pi'} M'$ .

### 5.7.2 Tangent Bundle

The reason of introducing the concept of a topological bundle, is because we need it in order to construct the so called “tangent bundle”.

As we said before, we want to define a vector field on a manifold  $M$  as a “smooth” map that assigns to each  $p \in M$  a tangent vector in  $T_p M$ . The idea was to merge all the tangent spaces into a unique set and equip it with a smooth structure, so that we can then define a vector field as a smooth map between smooth manifolds. We can do that through the tangent bundle.

**Definition 5.42** (Tangent Bundle). *Given a smooth manifold  $M$ , the **tangent bundle** of  $M$  is the disjoint union of all the tangent spaces to  $M$ , i.e:*

$$TM := \dot{\bigcup}_{p \in M} T_p M$$

equipped with the canonical projection map:

$$\begin{aligned} \pi : TM &\rightarrow M \\ X_{\gamma,p} &\mapsto p \end{aligned}$$

where  $p$  is the unique  $p \in M$  such that  $X_{\gamma,p} \in T_p M$ .

In other word the projection map, takes a vector from  $TM$  and spits out the point that it belongs  $\pi(X_{\gamma,p}) = p$ .

Since  $TM$  is simply a set (and not a smooth manifold), up to here what we have is a set bundle. In order for this set bundle to turn to a topological bundle as we defined it previously, we need to equip  $TM$  with the structure of a smooth manifold. We can achieve this by constructing a smooth atlas for  $TM$  from a smooth atlas on  $M$ , as follows.

Let  $\mathcal{A}_M$  be a smooth atlas on  $M$  and let  $(U, x) \in \mathcal{A}_M$ . If  $X_{\gamma,p} \in \text{preim}_\pi(U) \subseteq TM$ , then  $X_{\gamma,p} \in T_{\pi(X_{\gamma,p})} M$ , by definition of  $\pi$ . Moreover, since  $\pi(X_{\gamma,p}) = p \in U$ , we can expand  $X_{\gamma,p}$  in terms of the

basis induced by the chart  $(U, x)$ :

$$X_{\gamma,p} = X^a \left( \frac{\partial}{\partial x^a} \right)_p = X^a \left( \frac{\partial}{\partial x^a} \right)_{\pi(X_{\gamma,p})}$$

where  $X^1, \dots, X^{\dim M} \in \mathbb{R}$ . We can then define the map:

$$\begin{aligned} \xi: \text{preim}_\pi(U) &\rightarrow x(U) \times \mathbb{R}^{\dim M} \cong_{\text{set}} \mathbb{R}^{2 \dim M} \\ X_{\gamma,p} &\mapsto (x(\pi(X_{\gamma,p})), X^1, \dots, X^{\dim M}) \end{aligned}$$

Assuming that  $TM$  is equipped with a suitable topology, for instance the initial topology (i.e. the coarsest topology on  $TM$  that makes  $\pi$  continuous), we claim that the pair  $(\text{preim}_\pi(U), \xi)$  is a chart on  $TM$  and:

$$\mathcal{A}_{TM} := \{(\text{preim}_\pi(U), \xi) \mid (U, x) \in \mathcal{A}_M\}$$

is a smooth atlas on  $TM$ . Note that, from its definition, it is clear that  $\xi$  is a bijection. We will not show that  $(\text{preim}_\pi(U), \xi)$  is a chart here, but we will show that  $\mathcal{A}_{TM}$  is a smooth atlas.

**Proposition 5.6.** *Any two charts  $(\text{preim}_\pi(U), \xi), (\text{preim}_\pi(\tilde{U}), \tilde{\xi}) \in \mathcal{A}_{TM}$  are  $C^\infty$ -compatible.*

*Proof.*

Let  $(U, x)$  and  $(\tilde{U}, \tilde{x})$  be the two charts on  $M$  giving rise to  $(\text{preim}_\pi(U), \xi)$  and  $(\text{preim}_\pi(\tilde{U}), \tilde{\xi})$ , respectively. We need to show that the map:

$$\tilde{\xi} \circ \xi^{-1}: x(U \cap \tilde{U}) \times \mathbb{R}^{\dim M} \rightarrow \tilde{x}(U \cap \tilde{U}) \times \mathbb{R}^{\dim M}$$

is smooth, as a map between open subsets of  $\mathbb{R}^{2 \dim M}$ . Recall that such a map is smooth if, and only if, it is smooth componentwise. On the first  $\dim M$  components,  $\tilde{\xi} \circ \xi^{-1}$  acts as:

$$\begin{aligned} \tilde{x} \circ x^{-1}: x(U \cap \tilde{U}) &\rightarrow \tilde{x}(U \cap \tilde{U}) \\ x(p) &\mapsto \tilde{x}(p) \end{aligned}$$

while on the remaining  $\dim M$  components it acts as the change of vector components we met previously, i.e:

$$X^a \mapsto \tilde{X}^a = \partial_b(y^a \circ x^{-1})(x(p)) X^b$$

Hence, we have:

$$\begin{aligned} \tilde{\xi} \circ \xi^{-1}: \quad &x(U \cap \tilde{U}) \times \mathbb{R}^{\dim M} \rightarrow \tilde{x}(U \cap \tilde{U}) \times \mathbb{R}^{\dim M} \\ &(x(\pi(X_{\gamma,p})), X^1, \dots, X^{\dim M}) \mapsto (\tilde{x}(\pi(X_{\gamma,p})), \tilde{X}^1, \dots, \tilde{X}^{\dim M}) \end{aligned}$$

which is smooth in each component, and hence smooth.  $\square$

The tangent bundle of a smooth manifold  $M$  is therefore itself a smooth manifold of dimension  $2 \dim M$ , and the projection  $\pi: TM \rightarrow M$  is smooth with respect to this structure.

Now by using the smooth manifold  $M$  as the base space, the smooth manifold  $TM$  as the total space, and the smooth projection  $\pi$  we can define the topological tangent bundle as the triple:

$$TM \xrightarrow{\pi} M$$

Similarly, one can construct the *cotangent bundle*  $T^*M$  to  $M$  by defining:

$$T^*M := \dot{\bigcup}_{p \in M} T_p^*M$$

and going through the above again, using the dual basis  $\{(dx^a)_p\}$  instead of  $\{(\frac{\partial}{\partial x^a})_p\}$ .

### 5.7.3 Vector, Covector & Tensor Fields

Now that we have defined the tangent and cotangent bundles, we are ready to define fields.

**Definition 5.43** (Vector Field). *Let  $M$  be a smooth manifold, and let  $TM \xrightarrow{\pi} M$  be its tangent bundle. A **vector field**  $X$  on  $M$  is a smooth section of the tangent bundle, i.e. a smooth map  $X: M \rightarrow TM$  such that  $\pi \circ X = \text{id}_M$ .*

$$\begin{array}{c} TM \\ \uparrow X \quad \downarrow \pi \\ M \end{array}$$

This might seem as an abuse of notation since we already use the symbol “ $X$ ” to denote the tangent vectors  $X_{\gamma,p}$ . However, there is a good reason for that since the two are very closely related, as we will see right now.

We have already defined a smooth curve on a manifold  $M$  as a smooth map  $\gamma: \mathbb{R} \rightarrow M$ , where  $\mathbb{R}$  is understood as a 1-dimensional manifold. We will now introduce the concept of the integral curve.

**Definition 5.44** (Integral Curve). *Let  $M$  be a smooth manifold and let  $X$  be a vector field on  $M$ . An **integral curve** of  $X$  is smooth curve  $\gamma: (-\varepsilon, \varepsilon) \rightarrow M$ , with  $\varepsilon > 0$ , such that:*

$$\forall \lambda \in (-\varepsilon, \varepsilon) : X_{\gamma, \gamma(\lambda)} = X(\gamma(\lambda))$$

It follows from the local existence and uniqueness of solutions to ordinary differential equations that, given any vector field  $X$  and any point  $p \in M$ , there exist  $\varepsilon > 0$  and a smooth curve  $\gamma: (-\varepsilon, \varepsilon) \rightarrow M$  with  $\gamma(0) = p$  which is an integral curve of  $X$ .

Moreover, integral curves are locally unique. By this we mean that if  $\gamma_1$  and  $\gamma_2$  are both integral curves of  $X$  through  $p$ , i.e.  $\gamma_1(0) = \gamma_2(0) = p$ , then  $\gamma_1 = \gamma_2$  on the intersection of their domains of definition.

**Definition 5.45** (Complete Vector Field). *A vector field is called **complete** if given an integral curve  $\gamma: (-\varepsilon, \varepsilon) \rightarrow M$ , the domain of the integral curve  $(-\varepsilon, \varepsilon)$  can be extended to  $\mathbb{R}$  and still have:*

$$\forall \lambda \in \mathbb{R} : X_{\gamma, \gamma(\lambda)} = X(\gamma(\lambda))$$

We have the following result.

**Theorem 5.5.** *On a compact manifold, every vector field is complete.*

Since we are only dealing with compact manifolds, we can draw the following conclusion. Every (complete) vector field  $X$  can act on a point  $p$  of a manifold  $M$  and produce the corresponding vector  $X_{\gamma,p}$  in the sense that:

$$X(p) = X_{\gamma,p}$$

and this is the reason why “ $X$ ” and  $X_{\gamma,p}$  are closely related. From now on, we will drop the curve  $\gamma$  from the notation of the vectors  $X_{\gamma,p}$  and we will simply write a vector as  $X_p$ . We do that for two reason: first in order to reduce the complexity of the equations (although one has to remember that there is a curve  $\gamma$  associated to every  $X_p$ ) and second in order to have a consistency with the way we denote covectors (i.e  $\omega_p$ ) where we do not denote the curve  $\gamma$ . Hence from now on the notation that we will be using is:

$$X(p) = X_p \implies X(p)f = X_p f$$

where  $X$  is a vector field and when acted on a point  $p$  produces the corresponding vector  $X_p$ . (From time to time, when we will need to apply the definition of a vector, we will remind ourselves that  $X_p$  actually means  $X_{\gamma,p}$  so we are not confused from the  $\gamma$  in the definition).

As a final note, we also showed that a vector  $X_p$  can be expressed in the co-ordinate induced basis as:

$$X_p = X^a \left( \frac{\partial}{\partial x^a} \right)_p$$

Similarly for a vector field  $X$  we will be writing:

$$X = X^a \frac{\partial}{\partial x^a}$$

where now  $X^a$  are functions with  $X^a(p)$  equal to  $X^a$  from the vector equation i.e:

$$X(p) = X^a(p) \left( \frac{\partial}{\partial x^a} \right)_p = X^a \left( \frac{\partial}{\partial x^a} \right)_p = X_p$$

In other words, this is more of a notation and not really a basis of vector fields  $X$ , since  $\frac{\partial}{\partial x^a}$  must be understood locally and not globally (more on that later).

**Definition 5.46** ( $\Gamma(TM)$ ). We denote the set of all vector fields on  $M$  by  $\Gamma(TM)$ , i.e:

$$\Gamma(TM) := \{X: M \rightarrow TM \mid X \text{ is smooth and } \pi \circ X = \text{id}_M\}$$

This is, in fact, the standard notation for the set of all sections on a bundle.

*Remark 5.15.* An equivalent definition is that a vector field  $X$  on  $M$  is a derivation on the algebra  $\mathcal{C}^\infty(M)$ , i.e. an  $\mathbb{R}$ -linear map:

$$X: \mathcal{C}^\infty(M) \xrightarrow{\sim} \mathcal{C}^\infty(M)$$

satisfying the Leibniz rule (with respect to pointwise multiplication on  $\mathcal{C}^\infty(M)$ ):

$$X(fg) = g X(f) + f X(g)$$

This definition is better suited for some purposes, and later on we will switch from one to the other without making any notational distinction between them. It also implies that a field  $X$  instead of acting on a point  $p$  of the manifold, it can act directly to a function  $f$  (and then evaluate at point  $p$ ):

$$(Xf)(p) = X(p)f = X_p f$$

We can equip the set  $\Gamma(TM)$  with the following operations. The first is our, by now familiar, pointwise addition:

$$\begin{aligned} \oplus: \Gamma(TM) \times \Gamma(TM) &\rightarrow \Gamma(TM) \\ (X, Y) &\mapsto X \oplus Y \end{aligned}$$

where:

$$\begin{aligned} X \oplus Y: M &\rightarrow \Gamma(TM) \\ p &\mapsto (X \oplus Y)(p) := X(p) + Y(p) = X_p + Y_p \end{aligned}$$

Note that the  $+$  on the right hand side above is the addition in  $T_p M$ .

More interestingly, we can define a multiplication operation not by a simple number (i.e an element of  $\mathbb{R}$ ) but with a whole function (i.e an element of  $\mathcal{C}^\infty(M)$ ) as follows:

$$\begin{aligned} \odot: \mathcal{C}^\infty(M) \times \Gamma(TM) &\rightarrow \Gamma(TM) \\ (f, X) &\mapsto f \odot X \end{aligned}$$

where:

$$\begin{aligned} f \odot X: M &\rightarrow \Gamma(TM) \\ p &\mapsto (f \odot X)(p) := f(p)X(p) = f(p)X_p \end{aligned}$$

Note that since  $f \in \mathcal{C}^\infty(M)$ , we have  $f(p) \in \mathbb{R}$  and hence the multiplication above is the scalar multiplication on  $T_p M$ .

*Remark 5.16.* Of course, we could have defined  $\odot$  simply as pointwise *global* scaling, using the reals  $\mathbb{R}$  instead of the real functions  $\mathcal{C}^\infty(M)$ . Then, since  $(\mathbb{R}, +, \cdot)$  is an algebraic field, we would then have the obvious  $\mathbb{R}$ -vector space structure on  $\Gamma(TM)$ , that we will also be using when needed. Let us make two comment regarding this  $\mathbb{R}$ -vector space:

- In the case of a function instead of a number, a function  $f$  acts on the whole manifold  $M$  (it assigns a value  $f(p)$  on every point  $p$  of the manifold) and thus we are able to assign different values to

different points. In the case of  $\mathbb{R}$ -vector space we can only assign the same value to every point (i.e. having a constant vector field).

- A basis for this corresponding  $\mathbb{R}$ -vector space is necessarily uncountably infinite, and hence it does not provide a very useful decomposition for our vector fields. On the other hand the operation  $\odot$  allows for *local* scaling, i.e. we can scale a vector field by a different value at each point, and a much more useful decomposition of vector fields.

The question now is, mathematically speaking, what exactly the triple  $(\Gamma(TM), \oplus, \odot)$  is. Its nature of course depends on what the triple  $(\mathcal{C}^\infty(M), +, \cdot)$  is. Let's recall that the triple  $(\mathcal{C}^\infty(M), +, \cdot)$  can be viewed in two different ways:

- $(\mathcal{C}^\infty(M), +, \cdot)$ , where  $\cdot$  is scalar multiplication (by a real number), is an  $\mathbb{R}$ -vector space.
- $(\mathcal{C}^\infty(M), +, \bullet)$ , where  $\bullet$  is pointwise multiplication of maps, is a commutative, unital ring, but not a division ring since not every function has an inverse at every point (i.e. at all points that a function is zero, we cannot define an inverse since we would divide by zero).

The first view is of no use since if the triple is seen as a vector space over the real numbers, there is nothing else we can do. However, if we consider the second view. i.e. the triple  $(\mathcal{C}^\infty(M), +, \bullet)$ , where  $\bullet$  is pointwise function multiplication as a ring, then the triple  $(\Gamma(TM), \oplus, \odot)$  built on top of this ring satisfies:

- $(\Gamma(TM), \oplus)$  is an abelian group, with  $0 \in \Gamma(TM)$  being the section that maps each  $p \in M$  to the zero tangent vector in  $T_p M$ .
- $\Gamma(TM) \setminus \{0\}$  satisfies:
  - i)  $\forall f \in \mathcal{C}^\infty(M) : \forall X, Y \in \Gamma(TM) \setminus \{0\} : f \odot (X \oplus Y) = (f \odot X) \oplus (f \odot Y)$ .
  - ii)  $\forall f, g \in \mathcal{C}^\infty(M) : \forall X \in \Gamma(TM) \setminus \{0\} : (f + g) \odot X = (f \odot X) \oplus (g \odot X)$ .
  - iii)  $\forall f, g \in \mathcal{C}^\infty(M) : X \in \Gamma(TM) \setminus \{0\} : (f \bullet g) \odot X = f \odot (g \odot X)$ .
  - iv)  $\forall X \in \Gamma(TM) \setminus \{0\} : 1 \odot X = X$ , where  $1 \in \mathcal{C}^\infty(M)$  maps every  $p \in M$  to  $1 \in \mathbb{R}$ .

which are precisely the axioms for a vector space! Hence given that the triple  $(\mathcal{C}^\infty(M), +, \bullet)$  is a ring, that turns the triple  $(\Gamma(TM), \oplus, \odot)$  to a  $\mathcal{C}^\infty(M)$ -module.

And this of course is of crucial importance since as we showed in previous chapters, if a ring  $R$  is not a division ring, then a  $R$ -module does not need to have a basis. And since as we already said  $(\mathcal{C}^\infty(M), +, \bullet)$  is not a division ring, the vector fields as  $\mathcal{C}^\infty(M)$ -modules do not need to have a basis! And this is a shame, since if they would have a basis (let's say  $E_i$ ) we would be able to write a vector field  $X$  as:

$$X = X^i E_i$$

where  $X^i$  would be functions acting as components of the vector field!

Let us mention one final thing that will be useful later. Recall from the Lie algebra chapter in the notes, that a Lie algebra over an algebraic field  $K$  is a vector space over  $K$  equipped with a Lie bracket  $[-, -]$ , i.e. a  $K$ -bilinear, antisymmetric map which satisfies the Jacobi identity.

Considering  $\Gamma(TM)$  as an infinite-dimensional  $R$ -vector space, for two vector fields  $X, Y \in \Gamma(TM)$ , we can define their Lie bracket to be the commutator of the fields:

$$[X, Y](f) := X(Y(f)) - Y(X(f))$$

for any  $f \in \mathcal{C}^\infty(M)$ . Now we can check that indeed  $[X, Y] \in \Gamma(TM)$ , and that the bracket is  $\mathbb{R}$ -bilinear, antisymmetric and satisfies the Jacobi identity. Thus,  $(\Gamma(TM), +, \cdot, [-, -])$  is an infinite-dimensional Lie algebra over  $\mathbb{R}$ . (We usually suppress the  $+$  and  $\cdot$  when they are clear from the context).

In a similar manner one can construct a covector field through the use of the cotangent bundle, and from there to define the set of all covector fields  $\Gamma(T^*M)$  and subsequently a triple  $(\Gamma(T^*M), \oplus, \odot)$ .

Finally using  $\Gamma(TM)$  and  $\Gamma(T^*M)$  we can define a tensor field.

**Definition 5.47** (Tensor Field). *Let  $M$  be a smooth manifold. A smooth  $(r, s)$  **tensor field**  $\tau$  on  $M$  is a  $\mathcal{C}^\infty(M)$ -multilinear map:*

$$\tau: \underbrace{\Gamma(T^*M) \times \cdots \times \Gamma(T^*M)}_{r \text{ copies}} \times \underbrace{\Gamma(TM) \times \cdots \times \Gamma(TM)}_{s \text{ copies}} \rightarrow \mathcal{C}^\infty(M)$$

The equivalence of this to the bundle definition is due to the pointwise nature of tensors. For instance, a covector field  $\omega \in \Gamma(T^*M)$  can act on a vector field  $X \in \Gamma(TM)$  to yield a smooth function  $\omega(X) \in \mathcal{C}^\infty(M)$  which if it is evaluated at a specific point  $p$  will yield a specific number:

$$(\omega(X))(p) := \omega(p)(X(p)) = \omega_p(X_p)$$

Then, we see that for any  $f \in \mathcal{C}^\infty(M)$ , we have:

$$(\omega(fX))(p) = \omega(p)(f(p)X(p)) = f(p)\omega(p)(X(p)) =: (f\omega(X))(p)$$

and hence, the map  $\omega: \Gamma(TM) \xrightarrow{\sim} \mathcal{C}^\infty(M)$  is  $\mathcal{C}^\infty(M)$ -linear.

Similarly, the set  $\Gamma(T_s^r M)$  of all  $(r, s)$  smooth tensor fields on  $M$  can be made into a  $\mathcal{C}^\infty(M)$ -module, with module operations defined pointwise.

We can also define the tensor product of tensor fields:

$$\begin{aligned} \otimes: \Gamma(T_q^p M) \times \Gamma(T_s^r M) &\rightarrow \Gamma(T_{q+s}^{p+r} M) \\ (\tau, \sigma) &\mapsto \tau \otimes \sigma \end{aligned}$$

analogously to what we had with tensors on a vector space, i.e:

$$\begin{aligned} (\tau \otimes \sigma)(\omega^{(1)}, \dots, \omega^{(p)}, \omega^{(p+1)}, \dots, \omega^{(p+r)}, X^{(1)}, \dots, X^{(q)}, X^{(q+1)}, \dots, X^{(q+s)}) \\ := \tau(\omega^{(1)}, \dots, \omega^{(p)}, X^{(1)}, \dots, X^{(q)}) \sigma(\omega^{(p+1)}, \dots, \omega^{(p+r)}, X^{(q+1)}, \dots, X^{(q+s)}) \end{aligned}$$

where the upper indices with the parenthesis count different vector and covector fields i.e  $\omega^{(i)} \in \Gamma(T^*M)$  and  $X^{(i)} \in \Gamma(TM)$ .

Therefore, we can think of tensor fields on  $M$  either as sections of some tensor bundle on  $M$ , that is, as maps assigning to each  $p \in M$  a tensor ( $\mathbb{R}$ -multilinear map) on the vector space  $T_p M$ , or as a  $\mathcal{C}^\infty(M)$ -multilinear map as above. We will always try to pick the most useful or easier to understand, based on the context.

To summarize, fields are the generalization of the definitions of vectors, covectors and tensors at a specific point  $p$  of  $M$ , to every possible point  $p$  of manifold  $M$ , hence to the whole manifold  $M$ . In a similar way we can generalize the concept of the gradient of  $f$  at  $p \in M$  in the gradient of  $f$  at  $M$ .

### Generalize (Gradient At A Point) To (Gradient Of The Manifold)

Recall the definition of the gradient operator at a point  $p \in M$ . We can extend that definition to define the ( $\mathbb{R}$ -linear) operator:

$$\begin{aligned} d: \mathcal{C}^\infty(M) &\xrightarrow{\sim} \Gamma(T^*M) \\ f &\mapsto df \end{aligned}$$

where, of course,  $df: p \mapsto d_p f$ . Alternatively, we can think of  $df$  as the  $\mathbb{R}$ -linear map:

$$\begin{aligned} df: \Gamma(TM) &\xrightarrow{\sim} \mathcal{C}^\infty(M) \\ X &\mapsto df(X) = X(f) \end{aligned}$$

*Remark 5.17.* Locally on some chart  $(U, x)$  on  $M$ , the covector field  $df$  can be expressed as:

$$df = \lambda_a dx^a$$

for some smooth functions  $\lambda_i \in \mathcal{C}^\infty(U)$ . To determine what they are, we simply apply both sides to the

vector fields induced by the chart. We have:

$$df\left(\frac{\partial}{\partial x^b}\right) = \frac{\partial}{\partial x^b}(f) = \partial_b f$$

and:

$$\lambda_a dx^a \left(\frac{\partial}{\partial x^b}\right) = \lambda_a \frac{\partial}{\partial x^b}(x^a) = \lambda_a \delta_b^a = \lambda_b$$

Hence, the local expression of  $df$  on  $(U, x)$  is:

$$df = \partial_a f dx^a$$

Note that the operator  $d$  satisfies the Leibniz rule:

$$d(fg) = g df + f dg$$

### Generalize (Push-Forward & Pull-Back At A Point) To (Push-Forward & Pull-Back Of A Manifold)

Finally, we want to generalize the concepts of push-forward and pull-back from a point to the whole manifold (a.k.a from a vector/covector to a vector/covector field). For a good reason we will first start with the pull-back.

Recall that given a map  $\phi: M \rightarrow N$  between smooth manifolds we defined the pull-back of  $\phi$  at  $p \in M$  as the linear map:

$$\begin{aligned} (\phi^*)_p: T_{\phi(p)}^* N &\xrightarrow{\sim} T_p^* M \\ \omega_{\phi(p)} &\mapsto (\phi^*)_p(\omega_{\phi(p)}) \end{aligned}$$

where  $(\phi^*)_p(\omega_{\phi(p)})$  is defined as:

$$\begin{aligned} (\phi^*)_p(\omega_{\phi(p)}): T_p M &\xrightarrow{\sim} \mathbb{R} \\ X_{\gamma,p} &\mapsto (\phi^*)_p(\omega_{\phi(p)})(X_{\gamma,p}) := \omega_{\phi(p)}((\phi_*)_p(X_{\gamma,p})) \end{aligned}$$

Now we can simply extend the definition of a pull-back for a covector at point  $p$  denoted  $(\phi^*)_p$ , to this of a pull-back for a covector field on a manifold  $M$  denoted  $\phi^*$ , by simply acting with  $(\phi^*)_p$  at every point  $p$  of the manifold  $M$ :

$$\begin{aligned} \phi^*: \Gamma(T^* N) &\rightarrow \Gamma(T^* M) \\ \omega &\mapsto \phi^*(\omega) \end{aligned}$$

where  $\phi^*(\omega)$  evaluated at a point  $p$  of  $M$  is:

$$\phi^*(\omega)(p) := (\phi^*)_p(\omega_{\phi(p)})$$

Hence, the pull-back of a covector field evaluated at point  $p$  is equal to the pull-back of the covector  $\omega_{\phi(p)}$  generated by the covector field  $\omega$  at point  $\omega_{\phi(p)}$ .

While the pull-back can be extended from covectors to covector fields without problems, the push-forward of a vector cannot be generalized to the push-forward of a vector field unless the underlying map  $\phi$  is a diffeomorphism between the manifold  $M$  and  $N$ . Let's see why.

Let's start again with the pull-back that we have already defined. Observe that the pull back of a covector field includes the notion of the pull-back of a covector at a point  $p$ . Now, the map  $\phi$ , as a map, maps every single point of its domain  $M$  to a single point of its target  $N$ . Hence the whole target  $M$  is hit by the map, but the whole target  $N$  is not (recall that the part of  $N$  that is hit by the map is called the image of  $\phi$ ). This means that in the case of a pull-back (after we have defined the tangent vectors in both  $M$  and  $N$ ) every single point of the image of  $\phi$  on  $N$  will have a corresponding point back on  $M$  hence the definition of the pull-back of a covector field will be well-defined.

On the other hand, in the case of a push forward we get two problems coming from the fact that, in

general, the map  $\phi$  may not be neither surjective nor injective. First of all if the map  $\phi$  is not surjective that means that the image of  $\phi$  is not equal to the entire domain  $M$  ( $\text{im}_\phi(M) \neq N$ ), hence from a vector field defined on  $M$  we will never be able to define a vector field on  $N$  that lies outside the image of  $\phi$ . Moreover, if  $\phi$  is not injective, that means that distinct elements of the domain  $M$  are mapped to the same element in the target  $N$  hence it might be the case that the push-forward will create many different vectors for one given point  $p$  on  $N$  which will make it ill-defined.

Of course, if the map  $\phi$  is both surjective and injective, hence bijective, hence has an inverse, then none of this problems exist any more, since then the case is similar to the case of pull-backs (both directions of the map behave similarly). But recall that a bijection between topological spaces is called a “homeomorphism”, and moreover if the map is smooth (which in the case of smooth manifolds by definition always is) then the smooth “homeomorphism” is called a “diffeomorphism”.

So we ended up to our initial conclusion that the push-forward of a vector can be generalized to the push-forward of a vector field only if the underlying map  $\phi$  is a diffeomorphism between the manifold  $M$  and  $N$ . Then we can simply follow the same procedure as with the pull-back and define the push-forward of a vector field by simply acting with the push-forward at every point  $\phi(p)$  on the manifold  $N$ .

More specifically, recall that given a map  $\phi: M \rightarrow N$  between smooth manifolds we defined the push-forward of  $\phi$  at  $p \in M$  as the linear map:

$$\begin{aligned} (\phi_*)_p: T_p M &\xrightarrow{\sim} T_{\phi(p)} N \\ X_{\gamma,p} &\mapsto (\phi_*)_p(X_{\gamma,p}) \end{aligned}$$

where  $(\phi_*)_p(X_{\gamma,p})$  is defined as:

$$\begin{aligned} (\phi_*)_p(X_{\gamma,p}): \mathcal{C}^\infty(N) &\xrightarrow{\sim} \mathbb{R} \\ f &\mapsto (\phi_*)_p(X_{\gamma,p})f := X_{\gamma,p}(f \circ \phi) \end{aligned}$$

Now we can simply extend the definition of a push-forward for a vector at point  $p$  denoted  $(\phi_*)_p$ , to this of a push-forward for a vector field on a manifold  $M$  denoted  $\phi_*$ , by simply acting with the push-forward at every point  $p$  of the manifold  $M$ :

$$\begin{aligned} \phi_*: \Gamma(TM) &\rightarrow \Gamma(TN) \\ X &\mapsto \phi_*(X) \end{aligned}$$

where  $\phi_*(X)$  evaluated at a point  $\phi(p)$  of  $N$  is:

$$\phi_*(X)(\phi(p)) := (\phi_*)_p(X_p)$$

Hence the push-forward of a vector field evaluated at point  $\phi(p)$  is equal to the push-forward of the vector  $X|_p$  generated by the vector field  $X$  at point  $p$ .

## 5.8 Differential Forms

**Definition 5.48** (Differential Form). *Let  $M$  be a smooth manifold. A **(differential)  $n$ -form** on  $M$  is a  $(0, n)$  smooth tensor field  $\omega$  which is totally antisymmetric, i.e:*

$$\omega(X_1, \dots, X_n) = \text{sgn}(\pi) \omega(X_{\pi(1)}, \dots, X_{\pi(n)})$$

for any  $\pi \in S_n$ , with  $X_i \in \Gamma(TM)$ . We call  $n$  the degree of the form.

Alternatively, we can define a differential form as a smooth section of the appropriate bundle on  $M$ , i.e. as a map assigning to each  $p \in M$  an  $n$ -form on the vector space  $T_p M$ .

Of course, by definition, differential forms are nothing more but a very specific kind of tensors, hence it's a subset of the tensor space.

*Example 5.19.*

The electromagnetic field strength  $F$  is a differential 2-form built from the electric and magnetic fields, which are also taken to be forms. We will define these later in some detail.

**Definition 5.49** ( $\Omega^n(M)$ ). We denote by  $\Omega^n(M)$  the set of all differential  $n$ -forms on  $M$ , which then becomes a  $\mathcal{C}^\infty(M)$ -module by defining the addition and multiplication operations pointwise.

We have  $\Omega^0(M) \equiv \mathcal{C}^\infty(M)$  since they are  $(0,0)$  tensors a.k.a functions and  $\Omega^1(M) \equiv \Gamma(T_1^0 M) \equiv \Gamma(T^*M)$  since they are  $(0,1)$  tensors a.k.a covectors.

We can specialise the pull-back of tensors to differential forms.

**Definition 5.50** (Pull-Back On Differential Forms). Let  $\phi: M \rightarrow N$  be a smooth map and let  $\omega \in \Omega^n(N)$ . Then we define the **pull-back**  $\Phi^*(\omega) \in \Omega^n(M)$  of  $\omega$  as:

$$\begin{aligned}\Phi^*: \Gamma(T^*N) &\rightarrow \Gamma(T^*M) \\ \omega &\mapsto \Phi^*(\omega)\end{aligned}$$

or equivalently:

$$\begin{aligned}\Phi^*: \Omega^1(N) &\rightarrow \Omega^1(M) \\ \omega &\mapsto \Phi^*(\omega)\end{aligned}$$

where:

$$\Phi^*(\omega)(X^{(1)} \dots, X^{(n)}) := \omega(\phi_*(X^{(1)}), \dots, \phi_*(X^{(n)}))$$

for vector fields  $X^{(i)} \in \Gamma(TM)$  where  $\phi_*(X^{(i)})$  is the push-forward of each vector field.

The map  $\Phi^*: \Omega^n(N) \rightarrow \Omega^n(M)$  is  $\mathbb{R}$ -linear, and its action on  $\Omega^0(M)$  is simply:

$$\begin{aligned}\Phi^*: \Omega^0(M) &\rightarrow \Omega^0(M) \\ f &\mapsto \Phi^*(f) := f \circ \phi\end{aligned}$$

This works for any smooth map  $\phi$ , and it leads to a slight modification of our mantra:

*Vectors are pushed forward,  
forms are pulled back.*

The tensor product  $\otimes$  does not interact well with forms, since the tensor product of two forms is not necessarily a form (it might be, for example, a symmetric  $(0,n)$  tensor which, by definition, is not a form). Hence, we define the following.

**Definition 5.51** (Wedge Product). Let  $M$  be a smooth manifold. We define the **wedge** (or exterior) product of forms as the map:

$$\begin{aligned}\wedge: \Omega^n(M) \times \Omega^m(M) &\rightarrow \Omega^{n+m}(M) \\ (\omega, \sigma) &\mapsto \omega \wedge \sigma\end{aligned}$$

where:

$$(\omega \wedge \sigma)(X_1, \dots, X_{n+m}) := \frac{1}{n!m!} \sum_{\pi \in S_{n+m}} \text{sgn}(\pi) (\omega \otimes \sigma)(X_{\pi(1)}, \dots, X_{\pi(n+m)})$$

and  $X_1, \dots, X_{n+m} \in \Gamma(TM)$ . By convention, for any  $f, g \in \Omega^0(M)$  and  $\omega \in \Omega^n(M)$ , we set:

$$f \wedge g := fg \quad \text{and} \quad f \wedge \omega = \omega \wedge f = f\omega$$

*Example 5.20.*

Suppose that  $\omega, \sigma \in \Omega^1(M)$ . Then, for any  $X, Y \in \Gamma(TM)$ :

$$\begin{aligned}(\omega \wedge \sigma)(X, Y) &= (\omega \otimes \sigma)(X, Y) - (\omega \otimes \sigma)(Y, X) \\ &= (\omega \otimes \sigma)(X, Y) - \omega(Y)\sigma(X) \\ &= (\omega \otimes \sigma)(X, Y) - (\sigma \otimes \omega)(X, Y) \\ &= (\omega \otimes \sigma - \sigma \otimes \omega)(X, Y)\end{aligned}$$

Hence:

$$\omega \wedge \sigma = \omega \otimes \sigma - \sigma \otimes \omega.$$

The wedge product is bilinear over  $\mathcal{C}^\infty(M)$ , that is:

$$(f\omega_1 + \omega_2) \wedge \sigma = f\omega_1 \wedge \sigma + \omega_2 \wedge \sigma$$

for all  $f \in \mathcal{C}^\infty(M)$ ,  $\omega_1, \omega_2 \in \Omega^n(M)$  and  $\sigma \in \Omega^m(M)$ , and similarly for the second argument.

*Remark 5.18.* If  $(U, x)$  is a chart on  $M$ , then every  $n$ -form  $\omega \in \Omega^n(U)$  can be expressed locally on  $U$  as:

$$\omega = \omega_{a_1 \dots a_n} dx^{a_1} \wedge \dots \wedge dx^{a_n}$$

where  $\omega_{a_1 \dots a_n} \in \mathcal{C}^\infty(U)$  and  $1 \leq a_1 < \dots < a_n \leq \dim M$ . The  $dx^{a_i}$  appearing above are the covector fields (1-forms):

$$dx^{a_i}: p \mapsto d_p x^{a_i}$$

The pull-back distributes over the wedge product.

**Theorem 5.6.** *Let  $\phi: M \rightarrow N$  be smooth,  $\omega \in \Omega^n(N)$  and  $\sigma \in \Omega^m(N)$ . Then, we have:*

$$\Phi^*(\omega \wedge \sigma) = \Phi^*(\omega) \wedge \Phi^*(\sigma)$$

*Proof.*

Let  $p \in M$  and  $X_1, \dots, X_{n+m} \in T_p M$ . Then we have:

$$\begin{aligned} & (\Phi^*(\omega) \wedge \Phi^*(\sigma))(p)(X_1, \dots, X_{n+m}) \\ &= \frac{1}{n! m!} \sum_{\pi \in S_{n+m}} \text{sgn}(\pi) (\Phi^*(\omega) \otimes \Phi^*(\sigma))(p)(X_{\pi(1)}, \dots, X_{\pi(n+m)}) \\ &= \frac{1}{n! m!} \sum_{\pi \in S_{n+m}} \text{sgn}(\pi) \Phi^*(\omega)(p)(X_{\pi(1)}, \dots, X_{\pi(n)}) \\ & \quad \Phi^*(\sigma)(p)(X_{\pi(n+1)}, \dots, X_{\pi(n+m)}) \\ &= \frac{1}{n! m!} \sum_{\pi \in S_{n+m}} \text{sgn}(\pi) \omega(\phi(p))(\phi_*(X_{\pi(1)}), \dots, \phi_*(X_{\pi(n)})) \\ & \quad \sigma(\phi(p))(\phi_*(X_{\pi(n+1)}), \dots, \phi_*(X_{\pi(n+m)})) \\ &= \frac{1}{n! m!} \sum_{\pi \in S_{n+m}} \text{sgn}(\pi) (\omega \otimes \sigma)(\phi(p))(\phi_*(X_{\pi(1)}), \dots, \phi_*(X_{\pi(n+m)})) \\ &= (\omega \wedge \sigma)(\phi(p))(\phi_*(X_1), \dots, \phi_*(X_{n+m})) \\ &= \Phi^*(\omega \wedge \sigma)(p)(X_1, \dots, X_{n+m}) \end{aligned}$$

Since  $p \in M$  was arbitrary, the statement follows.  $\square$

### 5.8.1 The Grassmann Algebra

Note that the wedge product takes two differential forms and produces a differential form of a different type. It would be much nicer to have a space which is closed under the action of  $\wedge$ . In fact, such a space exists and it is called the Grassmann algebra of  $M$ .

**Definition 5.52** (Grassmann Algebra). *Let  $M$  be a smooth manifold. Define the  $\mathcal{C}^\infty(M)$ -module:*

$$\text{Gr}(M) \equiv \Omega(M) := \bigoplus_{n=0}^{\dim M} \Omega^n(M)$$

The **Grassmann algebra** on  $M$  is the algebra  $(\Omega(M), +, \cdot, \wedge)$ , where:

$$\wedge: \Omega(M) \times \Omega(M) \rightarrow \Omega(M)$$

is the linear continuation of the previously defined  $\wedge: \Omega^n(M) \times \Omega^m(M) \rightarrow \Omega^{n+m}(M)$ .

Recall that the direct sum of modules has the Cartesian product of the modules as underlying set and module operations defined componentwise. Also, note that by “algebra” here we really mean “algebra over a module”.

*Example 5.21.*

Let  $\psi = \omega + \sigma$ , where  $\omega \in \Omega^1(M)$  and  $\sigma \in \Omega^3(M)$ . Of course, this “+” is neither the addition on  $\Omega^1(M)$  nor the one on  $\Omega^3(M)$ , but rather that on  $\Omega(M)$  and, in fact,  $\psi \in \Omega(M)$ .

Let  $\varphi \in \Omega^n(M)$ , for some  $n$ . Then:

$$\varphi \wedge \psi = \varphi \wedge (\omega + \sigma) = \varphi \wedge \omega + \varphi \wedge \sigma$$

where  $\varphi \wedge \omega \in \Omega^{n+1}(M)$ ,  $\varphi \wedge \sigma \in \Omega^{n+3}(M)$ , and  $\varphi \wedge \psi \in \Omega(M)$ .

*Example 5.22.*

There is a lot of talk about *Grassmann numbers*, particularly in supersymmetry. One often hears that these are “numbers that do not commute, but anticommute”. Of course, objects cannot be commutative or anticommutative by themselves. These qualifiers only apply to operations on the objects. In fact, the Grassmann numbers are just the elements of a Grassmann algebra.

The following result is about the anticommutative behaviour of  $\wedge$ .

**Theorem 5.7.** *Let  $\omega \in \Omega^n(M)$  and  $\sigma \in \Omega^m(M)$ . Then:*

$$\omega \wedge \sigma = (-1)^{nm} \sigma \wedge \omega$$

We say that  $\wedge$  is *graded commutative*, that is, it satisfies a version of anticommutativity which depends on the degrees of the forms.

*Proof.*

First note that if  $\omega, \sigma \in \Omega^1(M)$ , then:

$$\omega \wedge \sigma = \omega \otimes \sigma - \sigma \otimes \omega = -\sigma \wedge \omega$$

Recall that if  $\omega \in \Omega^n(M)$  and  $\sigma \in \Omega^m(M)$ , then locally on a chart  $(U, x)$  we can write:

$$\begin{aligned} \omega &= \omega_{a_1 \dots a_n} dx^{a_1} \wedge \dots \wedge dx^{a_n} \\ \sigma &= \sigma_{b_1 \dots b_m} dx^{b_1} \wedge \dots \wedge dx^{b_m} \end{aligned}$$

with  $1 \leq a_1 < \dots < a_n \leq \dim M$  and similarly for the  $b_i$ . The coefficients  $\omega_{a_1 \dots a_n}$  and  $\sigma_{b_1 \dots b_m}$  are smooth functions in  $\mathcal{C}^\infty(U)$ . Since  $dx^{a_i}, dx^{b_j} \in \Omega^1(M)$ , we have:

$$\begin{aligned} \omega \wedge \sigma &= \omega_{a_1 \dots a_n} \sigma_{b_1 \dots b_m} dx^{a_1} \wedge \dots \wedge dx^{a_n} \wedge dx^{b_1} \wedge \dots \wedge dx^{b_m} \\ &= (-1)^n \omega_{a_1 \dots a_n} \sigma_{b_1 \dots b_m} dx^{b_1} \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} \wedge dx^{b_2} \wedge \dots \wedge dx^{b_m} \\ &= (-1)^{2n} \omega_{a_1 \dots a_n} \sigma_{b_1 \dots b_m} dx^{b_1} \wedge dx^{b_2} \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} \wedge dx^{b_3} \wedge \dots \wedge dx^{b_m} \\ &\vdots \\ &= (-1)^{nm} \omega_{a_1 \dots a_n} \sigma_{b_1 \dots b_m} dx^{b_1} \wedge \dots \wedge dx^{b_m} \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} \\ &= (-1)^{nm} \sigma \wedge \omega \end{aligned}$$

since we have swapped 1-forms  $nm$ -many times. □

*Remark 5.19.* We should stress that this is only true when  $\omega$  and  $\sigma$  are pure degree forms, rather than linear combinations of forms of different degrees. Indeed, if  $\varphi, \psi \in \Omega(M)$ , a formula like:

$$\varphi \wedge \psi = \dots \psi \wedge \varphi$$

does not make sense in principle, because the different parts of  $\varphi$  and  $\psi$  can have different commutation behaviours.

### 5.8.2 The Exterior Derivative

Recall the “extended” definition of the gradient operator of a function  $d$  on the whole manifold  $M$ :

$$\begin{aligned} d: \mathcal{C}^\infty(M) &\xrightarrow{\sim} \Gamma(T^*M) \\ f &\mapsto df \end{aligned}$$

Since  $\Omega^0(M) \equiv \mathcal{C}^\infty(M)$  and  $\Omega^1(M) \equiv \Gamma(T_1^0 M) \equiv \Gamma(T^*M)$ , we can also understand this as an operator that takes in 0-forms and outputs 1-forms:

$$d: \Omega^0(M) \xrightarrow{\sim} \Omega^1(M)$$

This can then be extended to an operator which acts on any  $n$ -form. For this definition, we need to remind ourselves of the definition of commutator we gave in the algebra section of the notes. More precisely, if  $M$  is a smooth manifold and  $X, Y \in \Gamma(TM)$  then the commutator (or Lie bracket) of  $X$  and  $Y$  is defined as:

$$\begin{aligned} [X, Y]: \mathcal{C}^\infty(M) &\xrightarrow{\sim} \mathcal{C}^\infty(M) \\ f &\mapsto [X, Y](f) := X(Y(f)) - Y(X(f)) \end{aligned}$$

where we are using the definition of vector fields as  $\mathbb{R}$ -linear maps  $\mathcal{C}^\infty(M) \xrightarrow{\sim} \mathcal{C}^\infty(M)$ .

Using the commutator we can now extend the gradient as follows.

**Definition 5.53** (Exterior Derivative). *The **exterior derivative** on  $M$  is the  $\mathbb{R}$ -linear operator:*

$$\begin{aligned} d: \Omega^n(M) &\xrightarrow{\sim} \Omega^{n+1}(M) \\ \omega &\mapsto d\omega \end{aligned}$$

with  $d\omega$  being defined as:

$$\begin{aligned} d\omega(X^{(1)}, \dots, X^{(n+1)}) &:= \sum_{i=1}^{n+1} (-1)^{i+1} X^{(i)}(\omega(X^{(1)}, \dots, \widehat{X^{(i)}}, \dots, X^{(n+1)})) \\ &\quad + \sum_{i < j} (-1)^{i+j} \omega([X^{(i)}, X^{(j)}], X^{(1)}, \dots, \widehat{X^{(i)}}, \dots, \widehat{X^{(j)}}, \dots, X^{(n+1)}) \end{aligned}$$

where  $X^{(i)} \in \Gamma(TM)$  and the hat denotes omissions.

*Remark 5.20.* Note that the operator  $d$  is only well-defined when it acts on forms. In order to define a derivative operator on general tensors we will need to add extra structure to our differentiable manifold.

*Example 5.23.*

In the case  $n = 1$ , the form  $d\omega \in \Omega^2(M)$  is given by:

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$$

Let us check that this is indeed a 2-form, i.e. an antisymmetric,  $\mathcal{C}^\infty(M)$ -multilinear map:

$$d\omega: \Gamma(TM) \times \Gamma(TM) \rightarrow \mathcal{C}^\infty(M)$$

By using the antisymmetry of the Lie bracket, we immediately get:

$$d\omega(X, Y) = -d\omega(Y, X)$$

Moreover, thanks to this identity, it suffices to check  $\mathcal{C}^\infty(M)$ -linearity in the first argument only. Addi-

tivity is easily checked:

$$\begin{aligned}
d\omega(X+Y, Z) &= (X+Y)(\omega(Z)) - Z(\omega(X+Y)) - \omega([X+Y, Z]) \\
&= X(\omega(Z)) + Y(\omega(Z)) - Z(\omega(X) + \omega(Y)) - \omega([X, Z] + [Y, Z]) \\
&= X(\omega(Z)) + Y(\omega(Z)) - Z(\omega(X)) - Z(\omega(Y)) - \omega([X, Z]) - \omega([Y, Z]) \\
&= d\omega(X, Z) + d\omega(Y, Z)
\end{aligned}$$

For  $\mathcal{C}^\infty(M)$ -scaling, first we calculate  $[fX, Y]$ . Let  $g \in \mathcal{C}^\infty(M)$ . Then:

$$\begin{aligned}
[fX, Y](g) &= fX(Y(g)) - Y(fX(g)) \\
&= fX(Y(g)) - fY(X(g)) - Y(f)X(g) \\
&= f(X(Y(g)) - Y(X(g))) - Y(f)X(g) \\
&= f[X, Y](g) - Y(f)X(g) \\
&= (f[X, Y] - Y(f)X)(g)
\end{aligned}$$

Therefore:

$$[fX, Y] = f[X, Y] - Y(f)X$$

Hence, we can calculate:

$$\begin{aligned}
d\omega(fX, Y) &= fX(\omega(Y)) - Y(\omega(fX)) - \omega([fX, Y]) \\
&= fX(\omega(Y)) - Y(f\omega(X)) - \omega(f[X, Y] - Y(f)X) \\
&= fX(\omega(Y)) - fY(\omega(X)) - Y(f)\omega(X) - f\omega([X, Y]) + \omega(Y(f)X) \\
&= fX(\omega(Y)) - fY(\omega(X)) - \textcolor{gray}{f\omega([X, Y])} + \textcolor{gray}{\omega(Y(f)X)} \\
&= f d\omega(X, Y)
\end{aligned}$$

which is what we wanted.

The exterior derivative satisfies a graded version of the Leibniz rule with respect to the wedge product.

**Theorem 5.8.** *Let  $\omega \in \Omega^n(M)$  and  $\sigma \in \Omega^m(M)$ . Then:*

$$d(\omega \wedge \sigma) = d\omega \wedge \sigma + (-1)^n \omega \wedge d\sigma$$

*Proof.*

We will work in local coordinates. Let  $(U, x)$  be a chart on  $M$  and write:

$$\begin{aligned}
\omega &= \omega_{a_1 \dots a_n} dx^{a_1} \wedge \dots \wedge dx^{a_n} =: \omega_A dx^A \\
\sigma &= \sigma_{b_1 \dots b_m} dx^{b_1} \wedge \dots \wedge dx^{b_m} =: \sigma_B dx^B
\end{aligned}$$

Locally, the exterior derivative operator  $d$  acts as:

$$d\omega = d\omega_A \wedge dx^A$$

Hence:

$$\begin{aligned}
d(\omega \wedge \sigma) &= d(\omega_A \sigma_B dx^A \wedge dx^B) \\
&= d(\omega_A \sigma_B) \wedge dx^A \wedge dx^B \\
&= (\sigma_B d\omega_A + \omega_A d\sigma_B) \wedge dx^A \wedge dx^B \\
&= \sigma_B d\omega_A \wedge dx^A \wedge dx^B + \omega_A d\sigma_B \wedge dx^A \wedge dx^B \\
&= \sigma_B d\omega_A \wedge dx^A \wedge dx^B + (-1)^n \omega_A dx^A \wedge d\sigma_B \wedge dx^B \\
&= \sigma_B d\omega \wedge dx^B + (-1)^n \omega_A dx^A \wedge d\sigma \\
&= d\omega \wedge \sigma + (-1)^n \omega \wedge d\sigma
\end{aligned}$$

since we have “anticommutated” the 1-form  $d\sigma_B$  through the  $n$ -form  $dx^A$ , picking up  $n$  minus signs in the process.  $\square$

An important property of the exterior derivative is the following.

**Theorem 5.9.** *Let  $\phi: M \rightarrow N$  be smooth. For any  $\omega \in \Omega^n(N)$ , we have:*

$$\Phi^*(d\omega) = d(\Phi^*(\omega))$$

*Remark 5.21.* Informally, we can write this result as  $\Phi^*d = d\Phi^*$ , and say that the exterior derivative “commutes” with the pull-back.

However, you should bear in mind that the two  $d$ ’s appearing in the statement are two different operators. On the left hand side, it is  $d: \Omega^n(N) \rightarrow \Omega^{n+1}(N)$ , while it is  $d: \Omega^n(M) \rightarrow \Omega^{n+1}(M)$  on the right hand side.

*Remark 5.22.* Of course, we could also combine the operators  $d$  into a single operator acting on the Grassmann algebra on  $M$ :

$$d: \Omega(M) \rightarrow \Omega(M)$$

by linear continuation.

### 5.8.3 De Rham Cohomology

**Definition 5.54** (Closed / Exact Forms). *Let  $M$  be a smooth manifold and let  $\omega \in \Omega^n(M)$ . We say that  $\omega$  is:*

- **Closed** if  $d\omega = 0$ .
- **Exact** if  $\exists \sigma \in \Omega^{n-1}(M) : \omega = d\sigma$ .

The question of whether every closed form is exact and vice versa, i.e. whether the implications:

$$(d\omega = 0) \Leftrightarrow (\exists \sigma : \omega = d\sigma)$$

hold in general, belongs to the branch of mathematics called cohomology theory, to which we will now provide an introduction.

The answer for the  $\Leftarrow$  direction is affirmative thanks to the following result.

**Theorem 5.10.** *Let  $M$  be a smooth manifold. The operator:*

$$d^2 \equiv d \circ d: \Omega^n(M) \rightarrow \Omega^{n+2}(M)$$

*is identically zero, i.e.  $d^2 = 0$ .*

*Proof.*

This can be shown directly using the definition of  $d$ . Here, we will instead show it by working in local coordinates.

Recall that, locally on a chart  $(U, x)$ , we can write any form  $\omega \in \Omega^n(M)$  as:

$$\omega = \omega_{a_1 \dots a_n} dx^{a_1} \wedge \dots \wedge dx^{a_n}$$

Then, we have:

$$\begin{aligned} d\omega &= d\omega_{a_1 \dots a_n} \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} \\ &= \partial_b \omega_{a_1 \dots a_n} dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} \end{aligned}$$

and hence :

$$d^2\omega = \partial_c \partial_b \omega_{a_1 \dots a_n} dx^c \wedge dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n}.$$

We can perform a little “trick” in the last equation and write it as twice the half expression:

$$d^2\omega = \frac{1}{2} \partial_c \partial_b \omega_{a_1 \dots a_n} dx^c \wedge dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} + \frac{1}{2} \partial_b \partial_c \omega_{a_1 \dots a_n} dx^c \wedge dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n}$$

Now we can inter-switch the  $c$  and  $b$  dummy indices in the second half part (we can do it since they are just dummy indices) and we get:

$$d^2\omega = \frac{1}{2} \partial_c \partial_b \omega_{a_1 \dots a_n} dx^c \wedge dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} + \frac{1}{2} \partial_b \partial_c \omega_{a_1 \dots a_n} dx^b \wedge dx^c \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n}$$

Since  $dx^b \wedge dx^c = -dx^c \wedge dx^b$ , and moreover, by Schwarz’s theorem, we have  $\partial_c \partial_b \omega_{a_1 \dots a_n} = \partial_b \partial_c \omega_{a_1 \dots a_n}$  we get:

$$d^2\omega = \frac{1}{2} \partial_c \partial_b \omega_{a_1 \dots a_n} dx^c \wedge dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n} - \frac{1}{2} \partial_c \partial_b \omega_{a_1 \dots a_n} dx^c \wedge dx^b \wedge dx^{a_1} \wedge \dots \wedge dx^{a_n}$$

Hence:

$$d^2\omega = 0$$

Since this holds for any  $\omega$ , we have  $d^2 = 0$ . □

**Corollary 5.1.** *Every exact form is closed.*

We can extend the action of  $d$  to the zero vector space  $0 := \{0\}$  by mapping the zero in  $0$  to the zero function in  $\Omega^0(M)$ . In this way, we obtain the chain of  $\mathbb{R}$ -linear maps:

$$0 \xrightarrow{d} \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \dots \xrightarrow{d} \Omega^n(M) \xrightarrow{d} \Omega^{n+1}(M) \xrightarrow{d} \dots \xrightarrow{d} \Omega^{\dim M}(M) \xrightarrow{d} 0$$

where we now think of the spaces  $\Omega^n(M)$  as  $\mathbb{R}$ -vector spaces.

Recall from linear algebra section in the notes that, given a linear map  $\phi: V \rightarrow W$ , one can define the subspace of  $V$ :

$$\ker(\phi) := \{v \in V \mid \phi(v) = 0\}$$

called the *kernel* of  $\phi$ , and the subspace of  $W$ :

$$\text{im}(\phi) := \{\phi(v) \mid v \in V\}$$

called the *image* of  $\phi$ .

Going back to our chain of maps, the equation  $d^2 = 0$  is equivalent to:

$$\text{im}(d: \Omega^n(M) \rightarrow \Omega^{n+1}(M)) \subseteq \ker(d: \Omega^{n+1}(M) \rightarrow \Omega^{n+2}(M))$$

for all  $0 \leq n \leq \dim M - 2$ . Moreover, we have:

$$\begin{aligned} \omega \in \Omega^n(M) \text{ is closed} &\Leftrightarrow \omega \in \ker(d: \Omega^n(M) \rightarrow \Omega^{n+1}(M)) \\ \omega \in \Omega^n(M) \text{ is exact} &\Leftrightarrow \omega \in \text{im}(d: \Omega^{n-1}(M) \rightarrow \Omega^n(M)) \end{aligned}$$

The traditional notation for the spaces on the right hand side above is:

$$\begin{aligned} Z^n &:= \ker(d: \Omega^n(M) \rightarrow \Omega^{n+1}(M)) \\ B^n &:= \operatorname{im}(d: \Omega^{n-1}(M) \rightarrow \Omega^n(M)) \end{aligned}$$

so that  $Z^n$  is the space of closed  $n$ -forms and  $B^n$  is the space of exact  $n$ -forms.

Our original question can be restated as: does  $Z^n = B^n$  for all  $n$ ? We have already seen that  $d^2 = 0$  implies that  $B^n \subseteq Z^n$  for all  $n$  ( $B^n$  is, in fact, a vector subspace of  $Z^n$ ). Unfortunately the equality does not hold in general, but we do have the following result.

**Lemma 5.1** (Poincaré). *Let  $M \subseteq \mathbb{R}^d$  be a simply connected domain. Then:*

$$Z^n = B^n, \quad \forall n > 0$$

In the cases where  $Z^n \neq B^n$ , we would like to quantify by how much the closed  $n$ -forms fail to be exact. The answer is provided by the cohomology group.

**Definition 5.55** (de Rham Cohomology Group). *Let  $M$  be a smooth manifold. The  $n$ -th **de Rham cohomology group** on  $M$  is the quotient  $\mathbb{R}$ -vector space:*

$$H^n(M) := Z^n / B^n$$

You can think of the above quotient as  $Z^n / \sim$ , where  $\sim$  is the equivalence relation:

$$\omega \sim \sigma \Leftrightarrow \omega - \sigma \in B^n$$

The answer to our question as it is addressed in cohomology theory is: every exact  $n$ -form on  $M$  is also closed and vice versa if, only if:

$$H^n(M) \cong_{\text{vec}} 0$$

Of course, rather than an actual answer, this is yet another restatement of the question. However, if we are able to determine the spaces  $H^n(M)$ , then we do get an answer.

A crucial theorem by de Rham states (in more technical terms) that  $H^n(M)$  only depends on the global topology of  $M$ . In other words, the cohomology groups are topological invariants. This is remarkable because  $H^n(M)$  is defined in terms of exterior derivatives, which have everything to do with the local differentiable structure of  $M$ , and a given topological space can be equipped with several inequivalent differentiable structures.

*Example 5.24.*

Let  $M$  be any smooth manifold. We have:

$$H^0(M) \cong_{\text{vec}} \mathbb{R}^{(\# \text{ of connected components of } M)}$$

since the closed 0-forms are just the locally constant smooth functions on  $M$ . As an immediate consequence, we have:

$$H^0(\mathbb{R}) \cong_{\text{vec}} H^0(S^1) \cong_{\text{vec}} \mathbb{R}$$

*Example 5.25.*

By Poincaré lemma, we have:

$$H^n(M) \cong_{\text{vec}} 0$$

for any simply connected  $M \subseteq \mathbb{R}^d$ .

## 5.9 Application: $\text{SL}(2, \mathbb{C})$ - Part 1

In this chapter we will go through an application containing (almost) everything we have mentioned so far. More specifically, we will examine in detail the special linear group of degree 2 over  $\mathbb{C}$ , also known as the relativistic spin group.

### SL(2, ℂ) As A Set

We define the following subset of  $\mathbb{C}^4 := \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C}$ :

$$\text{SL}(2, \mathbb{C}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{C}^4 \mid ad - bc = 1 \right\} \subseteq \mathbb{C}^4$$

where the array is just an alternative notation for a quadruple  $(a, b, c, d)$ . It's this extra constraint  $ad - bc = 1$  that removes one degree of freedom and makes it a subset and not the whole  $\mathbb{C}^4$ .

### SL(2, ℂ) As A Group

We define an operation:

$$\bullet: \text{SL}(2, \mathbb{C}) \times \text{SL}(2, \mathbb{C}) \rightarrow \text{SL}(2, \mathbb{C})$$

$$\left( \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right) \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

where:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & h \end{pmatrix} := \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

Formally, this operation is the same as matrix multiplication. We can check directly that the result of applying  $\bullet$  lands back in  $\text{SL}(2, \mathbb{C})$ , or simply recall that the determinant of a product is the product of the determinants. Moreover, the operation  $\bullet$ :

- i) Is associative (straightforward but tedious to check).
- ii) Has an identity element, namely  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \text{SL}(2, \mathbb{C})$ .
- iii) Admits inverses since for each  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{C})$ , we have  $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \in \text{SL}(2, \mathbb{C})$  and:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \bullet \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\text{hence, we have } \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Therefore, the pair  $(\text{SL}(2, \mathbb{C}), \bullet)$  is a (non-commutative) group.

### SL(2, ℂ) As A Topological Space

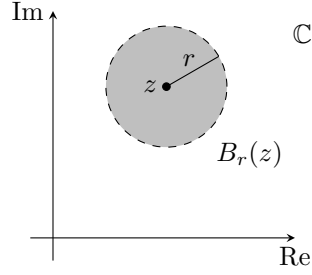
Recall that if  $N$  is a subset of  $M$  and  $\mathcal{O}$  is a topology on  $M$ , then we can equip  $N$  with the subset topology inherited from  $M$ :

$$\mathcal{O}|_N := \{U \cap N \mid U \in \mathcal{O}\}$$

We begin by establishing a topology on  $\mathbb{C}$  as follows. Let:

$$B_r(z) := \{y \in \mathbb{C} \mid |z - y| < r\}$$

be the open ball of radius  $r > 0$  and centre  $z \in \mathbb{C}$ .



Define  $\mathcal{O}_{\mathbb{C}}$  implicitly by:

$$U \in \mathcal{O}_{\mathbb{C}} \iff \forall z \in U : \exists r > 0 : B_r(z) \subseteq U$$

Then, the pair  $(\mathbb{C}, \mathcal{O}_{\mathbb{C}})$  is a topological space. In fact, we have:

$$(\mathbb{C}, \mathcal{O}_{\mathbb{C}}) \cong_{\text{top}} (\mathbb{R}^2, \mathcal{O}_{\text{std}})$$

We can then equip  $\mathbb{C}^4$  with the product topology so that we can finally define:

$$\mathcal{O} := (\mathcal{O}_{\mathbb{C}})|_{\text{SL}(2, \mathbb{C})}$$

so that the pair  $(\text{SL}(2, \mathbb{C}), \mathcal{O})$  is a topological space. In fact, it is a connected topological space, and we will need this property later on.

### SL(2, $\mathbb{C}$ ) As A Topological Manifold

Recall that a topological space  $(M, \mathcal{O})$  is a complex topological manifold if each point  $p \in M$  has an open neighbourhood  $U(p)$  which is homeomorphic to an open subset of  $\mathbb{C}^d$ . Equivalently, there must exist a  $\mathcal{C}^0$ -atlas, i.e. a collection  $\mathcal{A}$  of charts  $(U_\alpha, x_\alpha)$ , where the  $U_\alpha$  are open and cover  $M$  and each  $x$  is a homeomorphism onto a subset of  $\mathbb{C}^d$ . In our case of  $\text{SL}(2, \mathbb{C})$  we will map it locally to  $\mathbb{C}^3$  because as we said in the beginning of this section,  $\text{SL}(2, \mathbb{C})$  (as a set) is a subset of  $\mathbb{C}^4 := \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C}$  due to the constraint  $ad - bc = 1$  that removes one degree of freedom. This is why  $\mathbb{C}^3$  suffices and we do not use  $\mathbb{C}^4$ .

In general (but not in our case of  $\text{SL}(2, \mathbb{C})$ ) nothing stops us from using just one chart  $(U, x)$  and cover the whole topological space. However, as we will see, here we need two charts to cover the whole  $\text{SL}(2, \mathbb{C})$  (in the same manner that a sphere cannot be covered with just one chart).

Let  $U$  be the set:

$$U := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{C}) \mid a \neq 0 \right\}$$

and define the map:

$$\begin{aligned} x: \quad U &\rightarrow x(U) \subseteq \mathbb{C}^* \times \mathbb{C} \times \mathbb{C} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto (a, b, c) \end{aligned}$$

where  $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ .

Notice that given the mapping  $(a, b, c)$  one can reconstruct  $d$  from the constraint on the degree of freedom  $d = \frac{1+bc}{a}$ , hence the mapping to  $\mathbb{C}^3$ .

With a little more work on this direction, one can show that  $U$  is an open subset of  $(\text{SL}(2, \mathbb{C}), \mathcal{O})$  and  $x$  is a homeomorphism with inverse:

$$\begin{aligned} x^{-1}: \quad x(U) &\rightarrow U \\ (a, b, c) &\mapsto \begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix} \end{aligned}$$

This is the reason why we excluded the case  $a = 0$  when we defined the set  $U$  of the chart  $(U, x)$ , since if we hadn't, we wouldn't be able to divide with  $a$  and the map  $x$  wouldn't have an inverse.

However, this makes the chart  $(U, x)$  to not cover the whole  $\mathrm{SL}(2, \mathbb{C})$  since  $U$  as a set takes care only the elements of  $\mathrm{SL}(2, \mathbb{C})$  with  $a \neq 0$ . Hence we need at least one more chart. We thus define the set:

$$V := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}) \mid b \neq 0 \right\}$$

and the map:

$$\begin{aligned} y: \quad V &\rightarrow x(V) \subseteq \mathbb{C} \times \mathbb{C}^* \times \mathbb{C} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto (a, b, d) \end{aligned}$$

Similarly to the above,  $V$  is open and  $y$  is a homeomorphism with inverse:

$$\begin{aligned} y^{-1}: \quad x(V) &\rightarrow V \\ (a, b, d) &\mapsto \begin{pmatrix} a & b \\ \frac{ad-1}{b} & d \end{pmatrix} \end{aligned}$$

An element of  $\mathrm{SL}(2, \mathbb{C})$  cannot have both  $a$  and  $b$  equal to zero, for otherwise  $ad - bc = 0 \neq 1$ . Hence  $\mathcal{A}_{\mathrm{top}} := \{(U, x), (V, y)\}$  is an atlas, and we showed that  $x$  and  $y$  are homeomorphisms hence continuous and invertible. Since every atlas is automatically a  $\mathcal{C}^0$ -atlas, the triple  $(\mathrm{SL}(2, \mathbb{C}), \mathcal{O}, \mathcal{A}_{\mathrm{top}})$  is a 3-dimensional, complex, topological manifold.

### **$\mathrm{SL}(2, \mathbb{C})$ As A Complex Differentiable Manifold**

Recall that to obtain a  $\mathcal{C}^1$ -differentiable manifold from a topological manifold with atlas  $\mathcal{A}$ , we have to check that every transition map between charts in  $\mathcal{A}$  is differentiable in the usual sense. Remember this picture:

$$\begin{array}{ccc} & U \cap V \subseteq \mathrm{SL}(2, \mathbb{C}) & \\ x \swarrow & & \searrow y \\ x(U \cap V) \subseteq \mathbb{C}^3 & \xrightarrow{y \circ x^{-1}} & y(U \cap V) \subseteq \mathbb{C}^3 \end{array}$$

In our case, we have the atlas  $\mathcal{A}_{\mathrm{top}} := \{(U, x), (V, y)\}$ . We evaluate:

$$(y \circ x^{-1})(a, b, c) = y\left(\begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix}\right) = (a, b, \frac{1+bc}{a})$$

Hence we have the transition map:

$$\begin{aligned} y \circ x^{-1}: x(U \cap V) &\rightarrow y(U \cap V) \\ (a, b, c) &\mapsto (a, b, \frac{1+bc}{a}) \end{aligned}$$

Similarly, we have :

$$(x \circ y^{-1})(a, b, d) = y\left(\begin{pmatrix} a & b \\ \frac{ad-1}{b} & d \end{pmatrix}\right) = (a, b, \frac{ad-1}{b})$$

Hence, the other transition map is :

$$\begin{aligned} x \circ y^{-1}: y(U \cap V) &\rightarrow x(U \cap V) \\ (a, b, d) &\mapsto (a, b, \frac{ad-1}{b}) \end{aligned}$$

Since  $a \neq 0$  and  $b \neq 0$ , the transition maps are complex differentiable.

Therefore, the atlas  $\mathcal{A}_{\text{top}}$  is a differentiable atlas. By defining  $\mathcal{A}$  to be the maximal differentiable atlas containing  $\mathcal{A}_{\text{top}}$ , we have that  $(\text{SL}(2, \mathbb{C}), \mathcal{O}, \mathcal{A})$  is a 3-dimensional, complex differentiable manifold.

## Chapter 6

# Extra Structures On Topological Manifolds

In this chapter we will introduce a very important concept called “connection” (or “covariant derivative”) on a differentiable topological manifold. However, we need to make something very clear. “Connection” and “covariant derivative” are not actually the same thing. In elementary courses on differential geometry or general relativity, the notions of connection and covariant derivative are often confused with one another and, sometimes, the terms are even used as synonyms. This happens because one is not in a position to properly define these concepts before introducing the so-called “principal fiber bundles” that we will do in a later chapter.

So for now, in this section, we will make a kind of “naive” introduction to the concept, which however, most of the times, this is how it is introduced in many courses and on top of that it will give us a very first understanding of the topic. Later in the notes, once we introduce “principal fiber bundles” we will revisit these terms in a more systematic way and we will gain a better understanding of what they are exactly and how they generalize.

Having said that, for now we will introduce “connection” and “covariant derivative” as identical concepts which, again, even though they are different, for the purposes of this section we shall not distinguish the two, and we will also introduce “parallel transport” within this framework (later we will also revisit parallel transport).

### 6.1 Connection / Covariant Derivative

So far, we saw that a vector field  $X$  can be used to provide a directional derivative of a function  $f \in C^\infty(M)$  in the direction  $X$  as  $Xf(p) = X_p f$ . Recall that once we introduced the  $(p, q)$  tensor field, we saw that a function  $f \in C^\infty(M)$  is nothing else but a  $(0, 0)$  tensor field. Hence the equation  $Xf$  can be seen as the directional derivative acting on a  $(0, 0)$  tensor field, which leads to the natural question: can we apply the directional derivative to any  $(p, q)$  tensor field?

The short answer is no, and the reason is that we do not know how the field  $X$  can act on anything else other than a function (this is how we defined it). However, one could try to generalize the action of  $X$  to any  $(p, q)$  tensor field by defining a new object  $\nabla_X$ , called “connection” or “covariant derivative”, able to act on any  $(p, q)$  tensor field (for this section we will stick with the term “connection”).

$$\begin{array}{ccc}
 X : C^\infty(M) & \longrightarrow & C^\infty(M) \\
 \vdots \downarrow \text{wavy} & & \vdots \downarrow \text{wavy} \\
 \nabla_X : (p, q)\text{-tensor field} & \longrightarrow & (p, q)\text{-tensor field}
 \end{array}$$

The question is how to define such an object? Since it is closely related to  $X$  it must be consistent with it, so we will define  $\nabla_X$  through some “consistency” conditions that it must satisfy.

**Definition 6.1** (Connection / Covariant Derivative). A **connection** (or **covariant derivative**)  $\nabla_X$  on a smooth manifold  $M$  is a map that takes a pair consisting of a vector field  $X$  and a  $(p, q)$ -tensor field  $T$  and sends them to a  $(p, q)$ -tensor field  $\nabla_X T$  satisfying:

1.  $\nabla_X f = Xf, \forall f \in C^\infty M.$

(in order to be consistent with the already defined action of  $Xf$ )

2.  $\nabla_X(T + S) = \nabla_X T + \nabla_X S.$

(due to linearity of  $X$ )

3. Since it's a notion of derivative, for any  $(p, q)$  tensor field  $T$  the Leibnitz rule must be satisfied:

$$\begin{aligned} \nabla_X T(\omega_1, \dots, \omega_p, Y_1, \dots, Y_q) &= (\nabla_X T)(\omega_1, \dots, \omega_p, Y_1, \dots, Y_q) \\ &\quad + T(\nabla_X \omega_1, \dots, \omega_p, Y_1, \dots, Y_q) + \dots + T(\omega_1, \dots, \nabla_X \omega_p, Y_1, \dots, Y_q) \\ &\quad + T(\omega_1, \dots, \omega_p, \nabla_X Y_1, \dots, Y_q) + \dots + T(\omega_1, \dots, \omega_p, Y_1, \dots, \nabla_X Y_q) \end{aligned}$$

4.  $\nabla_{fX+gZ} T = f\nabla_X T + g\nabla_Z T, \forall f, g \in C^\infty(M).$

(This is  $C^\infty$ -linearity, which means that no matter how the function  $f$  scales the vectors at different points of the manifold, the effect of the scaling at any point is independent of scaling in the neighbourhood and depends only on how the scaling happens at that point)

After formulating this “wish list” of properties which  $\nabla_X$  acting on a tensor field should have, in general there may be many structures that satisfy these conditions. Any remaining freedom in choosing such a  $\nabla_X$ , will need to be provided as additional structure beyond the structure we already have, as we did for example with the topology and the atlases.

**Definition 6.2** (Topological Manifold With A Connection). A **topological manifold with a connection**  $\nabla_X$  is a quadruple  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$ , where  $M$  is a set,  $\mathcal{O}$  is a chosen topology on the set,  $\mathcal{A}$  is a chosen smooth atlas and  $\nabla_X$  is a chosen connection.

The question now is, how much freedom do we have in choosing a connection? In order to answer that we will let it act in “all” possible  $(p, q)$  tensor fields and see what we get.

Starting with the simplest case of a  $(0, 0)$  tensor field, i.e a function  $f$ , by the first condition we simply have:

$$\nabla_X f = Xf$$

So far so good. Let us consider now the next simplest case of an  $(1, 0)$  tensor field, i.e a vector field  $Y$ :

$$\begin{aligned} \nabla_X Y &= \nabla_{(X^i \frac{\partial}{\partial x^i})} \left( Y^m \frac{\partial}{\partial x^m} \right) && \text{(expressed in chart (U, x))} \\ &= X^i \cdot \nabla_{(\frac{\partial}{\partial x^i})} \left( Y^m \frac{\partial}{\partial x^m} \right) && \text{(fourth condition)} \\ &= X^i \left( \nabla_{(\frac{\partial}{\partial x^i})} Y^m \right) \frac{\partial}{\partial x^m} + X^i \cdot Y^m \cdot \left( \nabla_{(\frac{\partial}{\partial x^i})} \frac{\partial}{\partial x^m} \right) && \text{(third condition)} \\ &= X^i \frac{\partial}{\partial x^i} Y^m \frac{\partial}{\partial x^m} + X^i \cdot Y^m \cdot \left( \nabla_{(\frac{\partial}{\partial x^i})} \frac{\partial}{\partial x^m} \right) && \text{(first condition)} \end{aligned}$$

Regarding the last term  $\nabla_{(\frac{\partial}{\partial x^i})} \frac{\partial}{\partial x^m}$ , note that it simply is of the form  $\nabla_X X$  where  $X$  is the co-ordinate induced basis vector. There isn't something more we can do on that, however since  $\frac{\partial}{\partial x^m}$  is a basis vector, i.e a  $(1, 0)$  tensor field, the result of the action of  $\nabla_{(\frac{\partial}{\partial x^i})}$  on it must by definition, yield again a  $(1, 0)$  tensor field i.e a vector that we can express in the co-ordinate induced basis also as:

$$\nabla_{(\frac{\partial}{\partial x^i})} \frac{\partial}{\partial x^m} = \Gamma_{im}^q \frac{\partial}{\partial x^q}$$

Hence:

$$\nabla_X Y = X^i \left( \frac{\partial}{\partial x^i} Y^m \right) \frac{\partial}{\partial x^m} + X^i \cdot Y^m \cdot \Gamma_{im}^q \frac{\partial}{\partial x^q}$$

Thus, by rearranging the indices and the terms and by discarding the basis vectors:

$$(\nabla_X Y)^i = X^m \left( \frac{\partial}{\partial x^m} Y^i \right) + \Gamma_{mn}^i \cdot X^m \cdot Y^n = X(Y^i) + \Gamma_{mn}^i \cdot X^m \cdot Y^n$$

This last equation tells us the freedom we have left with on choosing a structure, after imposing the conditions on  $\nabla_X$  acting on an  $(1,0)$  tensor field, i.e a vector field. Namely, we need to define all the components of  $\Gamma_{nm}^i$  i.e  $(\dim M)^3$ -many functions in order to define a directional derivative of a vector field.

Now we move on to consider the action of  $\nabla_X$  in the case of a  $(0,1)$  tensor field, i.e a covector field  $\omega$ .

$$\begin{aligned} \nabla_X \omega &= \nabla_{(X^i \frac{\partial}{\partial x^i})} (\omega_m dx^m) && \text{(expressed in chart (U,x))} \\ &= X^i \cdot \nabla_{(\frac{\partial}{\partial x^i})} (\omega_m dx^m) && \text{(fourth condition)} \\ &= X^i \left( \nabla_{(\frac{\partial}{\partial x^i})} \omega_m \right) dx^m + X^i \cdot \omega_m \cdot \left( \nabla_{(\frac{\partial}{\partial x^i})} dx^m \right) && \text{(third condition)} \\ &= X^i \frac{\partial}{\partial x^i} \omega_m dx^m + X^i \cdot \omega_m \cdot \left( \nabla_{(\frac{\partial}{\partial x^i})} dx^m \right) && \text{(first condition)} \end{aligned}$$

Once again we have left with the term  $\nabla_{\frac{\partial}{\partial x^i}} (dx^m)$  which with similar reasoning as before, it must be again a covector field that can be expressed in term of the dual co-ordinate induced basis i.e:

$$\nabla_{\frac{\partial}{\partial x^i}} (dx^m) = \Sigma_{ij}^m dx^j$$

The question now is, are these  $\Sigma$ 's independent of  $\Gamma$ 's? We just showed that, on a chart domain  $U$ , the choice of the  $(\dim M)^3$ -many functions  $\Gamma$ 's suffices to fix the action of  $\nabla_X$  on a vector field. Do we need another  $(\dim M)^3$ -many functions  $\Sigma$ 's to fix the action on a covector field? And if it so, will we have to provide more and more coefficients for all individual cases of  $(p,q)$  tensor field?

Fortunately the answer is no! The same  $(\dim M)^3$ -many functions  $\Gamma$ 's fix the action of  $\nabla$  on any tensor field. Let's first show that the  $\Sigma$ 's are actually related to  $\Gamma$ 's.

Consider the following:

$$\nabla_{\frac{\partial}{\partial x^i}} \left( dx^m \left( \frac{\partial}{\partial x^j} \right) \right) dx^j = \left( \nabla_{\frac{\partial}{\partial x^i}} \delta_j^m \right) dx^j = \left( \frac{\partial}{\partial x^i} (\delta_j^m) \right) dx^j = 0$$

However it also is:

$$\begin{aligned} \nabla_{\frac{\partial}{\partial x^i}} \left( dx^m \left( \frac{\partial}{\partial x^j} \right) \right) dx^j &= \left( \nabla_{\frac{\partial}{\partial x^i}} dx^m \right) \left( \frac{\partial}{\partial x^j} \right) dx^j + dx^m \left( \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} \right) dx^j \\ &= \left( \nabla_{\frac{\partial}{\partial x^i}} dx^m \right) \left( \frac{\partial}{\partial x^j} \right) dx^j + dx^m \left( \Gamma_{ij}^q \frac{\partial}{\partial x^q} \right) dx^j \\ &= \left( \nabla_{\frac{\partial}{\partial x^i}} dx^m \right) \left( \frac{\partial}{\partial x^j} dx^j \right) + \Gamma_{ij}^q \left( dx^m \frac{\partial}{\partial x^q} \right) dx^j \\ &= \left( \nabla_{\frac{\partial}{\partial x^i}} dx^m \right) \delta_j^j + \Gamma_{ij}^q \cdot \delta_q^m \cdot dx^j \\ &= \left( \nabla_{\frac{\partial}{\partial x^i}} dx^m \right) + \Gamma_{ij}^m \cdot dx^j \\ &= \Sigma_{ij}^m dx^j + \Gamma_{ij}^m \cdot dx^j \\ &= (\Sigma_{ij}^m + \Gamma_{ij}^m) dx^j \end{aligned}$$

But since, as we showed, this is equal to zero we end up with:

$$(\Sigma_{ij}^m + \Gamma_{ij}^m) = 0 \implies \Sigma_{ij}^m + \Gamma_{ij}^m = 0 \implies \Sigma_{ij}^m = -\Gamma_{ij}^m$$

Hence:

$$\nabla_x \omega = X^i \left( \frac{\partial}{\partial x^i} \omega_m \right) dx^m - X^i \cdot \omega_m \cdot \Gamma_{iq}^m \cdot dx^q$$

Thus, by rearranging the indices and the terms and by discarding the basis vectors:

$$(\nabla_x \omega)_i = X^m \left( \frac{\partial}{\partial x^m} \omega_i \right) - \Gamma_{mi}^n \cdot X^m \cdot \omega_n = X(\omega_i) - \Gamma_{mi}^n \cdot X^m \cdot \omega_n$$

To sum up:

$$\begin{aligned} (\nabla_X Y)^i &= X(Y^i) + \Gamma_{mn}^i \cdot X^m \cdot Y^n \\ (\nabla_x \omega)_i &= X(\omega_i) - \Gamma_{mi}^n \cdot X^m \cdot \omega_n \end{aligned}$$

On top of that, by making use of the Leibnitz rule we can find the action of  $\nabla_X$  on any  $(p, q)$ -tensor field. For example for a  $(1, 2)$ -tensor field  $T$ :

$$(\nabla_X T)_{jk}^i = X(T_{jk}^i) + \Gamma_{ms}^i T_{jk}^s X^m - \Gamma_{mj}^s T_{sk}^i X^m - \Gamma_{mk}^s T_{js}^i X^m$$

Hence, we showed that these  $(\dim M)^3$ -many functions  $\Gamma$ 's suffices to fix the action of  $\nabla_X$  on any  $(p, q)$ -vector field and thus this is all we need to chose in order to completely define the structure of the connection.

**Definition 6.3** (Connection Coefficient Functions). *Given a manifold with a connection  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$  and a chart  $(U, x) \in \mathcal{A}$ , then the **connection coefficient functions**  $\Gamma$ 's on  $M$  of  $\nabla_X$  w.r.t  $(U, x)$  are  $(\dim M)^3$ -many functions given by:*

$$\begin{aligned} \Gamma_i^{jk} : U &\rightarrow \mathbb{R} \\ p &\mapsto \Gamma_{jk}^i(p) := \left( dx^i \left( \nabla_{\left( \frac{\partial}{\partial x^j} \right)} \frac{\partial}{\partial x^k} \right) \right) (p) \end{aligned}$$

One important aspect of the connection coefficient functions is their transformation law under a change of coordinates. Namely, let  $(U, x), (V, y) \in \mathcal{A}$  and  $U \cap V \neq \emptyset$ . Then for the connection coefficient functions in  $(V, y)$  holds:

$$\begin{aligned} \Gamma_{jk(y)}^i &:= dy^i \left( \nabla_{\frac{\partial}{\partial y^j}} \frac{\partial}{\partial y^k} \right) \\ &= \frac{\partial y^i}{\partial x^q} dx^q \left( \nabla_{\frac{\partial x^p}{\partial y^j}} \frac{\partial x^s}{\partial y^k} \frac{\partial}{\partial x^s} \right) \\ &= \frac{\partial y^i}{\partial x^q} dx^q \left( \frac{\partial x^p}{\partial y^j} \left[ \left( \nabla_{\frac{\partial}{\partial x^p}} \frac{\partial x^s}{\partial y^k} \right) \frac{\partial}{\partial x^s} + \frac{\partial x^s}{\partial y^k} \left( \nabla_{\frac{\partial}{\partial x^p}} \frac{\partial}{\partial x^s} \right) \right] \right) \\ &= \frac{\partial y^i}{\partial x^q} \frac{\partial x^p}{\partial y^j} \frac{\partial}{\partial x^p} \frac{\partial x^s}{\partial y^k} \delta_s^q + \frac{\partial y^i}{\partial x^q} \frac{\partial x^p}{\partial y^j} \frac{\partial x^s}{\partial y^k} \Gamma_{sp(x)}^q \\ &= \frac{\partial y^i}{\partial x^q} \frac{\partial x^p}{\partial y^j} \frac{\partial x^s}{\partial y^k} \Gamma_{sp(x)}^q + \frac{\partial y^i}{\partial x^q} \frac{\partial}{\partial y^j} \frac{\partial x^s}{\partial y^k} \\ &= \frac{\partial y^i}{\partial x^q} \frac{\partial x^p}{\partial y^j} \frac{\partial x^s}{\partial y^k} \Gamma_{sp(x)}^q + \frac{\partial y^i}{\partial x^q} \frac{\partial^2 x^q}{\partial y^j \partial y^k} \end{aligned}$$

Notice that the change of connection coefficient function under the change of chart  $(U \cap V, x) \rightarrow (U \cap V, y)$  do not follow the usual transformation law of a tensor due to the second term of the equation. If that part was missing then  $\Gamma$ 's would be the components of a tensor. Notice also that this extra term carries a second derivative hence for linear transformation between coordinates in two charts, the term  $\frac{\partial^2 x^q}{\partial y^j \partial y^k}$

always vanishes and then the transformation law resembles the one of a tensor (without meaning that  $\Gamma$ 's became the components of a tensor suddenly).

$\Gamma$ 's transformation law has another great impact. Since the second term, the one destroying the tensor components transformation law, does not carry and  $\Gamma$ 's, that means that even if  $\Gamma$ 's are zero in one chart, they might be non zero in the other charts (unless the chart transition maps are all linear, however, there is no reason not to select a coordinate which is not a linear transformation of another one). In other words just the change of coordinate system can suddenly give non-vanishing  $\Gamma$ 's even if we didn't have them in the first coordinate system.

That leads us to the following definition/theorem that we will not prove.

**Definition 6.4** (Normal Coordinates). *Let  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$  be a manifold with a connection. Having chosen a point  $p$ , one can construct a chart  $(U, x)$  with  $p \in U$  such that the symmetric part of  $\Gamma$ 's vanish at the point  $p$  (not necessarily in any neighbourhood). That is:*

$$\forall p \in M, \exists (U, x) \in \mathcal{A} : p \in U : \Gamma_{(jk)(x)}^i(p) = 0$$

*Such  $(U, x)$  is called a **normal coordinate chart** of  $\nabla_X$  at  $p \in M$ .*

### 6.1.1 Parallel Transport

Parallel transport is a way of transporting geometrical data along smooth curves in a manifold. If the manifold is equipped with a connection (a covariant derivative or connection on the tangent bundle), then this connection allows one to transport vectors of the manifold along curves so that they stay parallel with respect to the connection.

The parallel transport for a connection thus supplies a way of, in some sense, moving the local geometry of a manifold along a curve: that is, of connecting the geometries of nearby points. There may be many notions of parallel transport available, but a specification of one, one way of connecting up the geometries of points on a curve, is tantamount to providing a connection. In fact, the usual notion of connection is the infinitesimal analogue of parallel transport. Or, vice versa, parallel transport is the local realization of a connection.

As parallel transport supplies a local realization of the connection, it also supplies a local realization of the curvature that we will see in a while.

**Definition 6.5** (Parallel Transport). *Let  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$  be a smooth manifold with connection. A vector field  $Y$  on  $M$  is said to be **parallelly transported** along a smooth curve  $\gamma : \mathbb{R} \rightarrow M$  if :*

$$\nabla_X Y = 0$$

*where  $X$  is a vector field along this curve  $\gamma$ , i.e  $X(p) = X_{\gamma, p}$ ,  $\forall p \in M$ .*

*Remark 6.1.* Even though parallelly transported sounds like an action, it is really a property.

A slightly weaker condition is the notion of “parallel”.

**Definition 6.6** (Parallel). *Let  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$  be a smooth manifold with connection. A vector field  $Y$  on  $M$  is said to be **parallel** along a smooth curve  $\gamma : \mathbb{R} \rightarrow M$  if :*

$$\nabla_X Y = \mu Y$$

*where  $\mu$  is a function  $\mu : \mathbb{R} \rightarrow \mathbb{R}$ .*

Finally we also introduce the crucial concept of autoparallel transport.

**Definition 6.7** (Autoparallel Transport). *Let  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$  be a smooth manifold with connection. A vector field  $X$  on  $M$  is said to be **autoparallelly transported** along a smooth curve  $\gamma : \mathbb{R} \rightarrow M$  if :*

$$\nabla_X X = 0$$

We can manipulate the autoparallel transport equation by invoking a co-ordinate induced basis and get one of the most heavily used equation in physics as follows:

$$\begin{aligned} (\nabla_X X)^i &= X(X^i) + \Gamma_{mn}^i \cdot X^m \cdot X^n \\ &= X^m \left( \frac{\partial}{\partial x^m} X^i \right) + \Gamma_{mn}^i \cdot X^m \cdot X^n \\ &= \ddot{\gamma}^i + \Gamma_{mn}^i \cdot \dot{\gamma}^m \cdot \dot{\gamma}^n \end{aligned}$$

where in the last line we used a different notation for the directional derivative, namely:

$$\begin{aligned} X^m: \mathcal{C}^\infty(M) &\xrightarrow{\sim} \mathcal{C}^\infty(M) \\ f &\mapsto ((f \circ \gamma)'(0))^m := \dot{\gamma}^m \end{aligned}$$

and we also used a fact that we can show (but we won't):

$$X^m \left( \frac{\partial}{\partial x^m} X^i \right) = \ddot{\gamma}^i$$

Hence the autoparallel transport equations reads:

$$\ddot{\gamma}^i + \Gamma_{mn}^i \cdot \dot{\gamma}^m \cdot \dot{\gamma}^n = 0$$

which is the chart expression of the condition that  $\gamma$  be autoparallely transported.

*Example 6.1.*

In Euclidean plane having a chart:

$$(U = \mathbb{R}^2, x = id_{\mathbb{R}^2}, \Gamma_{jk}^i(x) = 0)$$

leads to the following autoparallely transported equation:

$$\ddot{\gamma}_{(x)}^m = 0 \implies \gamma_{(x)}^m(\lambda) = a^m \lambda + b^m$$

where  $a, b \in \mathbb{R}^d$ , which is the uniform motion.

*Example 6.2.*

Consider the round sphere  $(S^2, \mathcal{O}, \mathcal{A}, \nabla_{round})$ . Consider the chart  $x(p) = (\theta, \phi)$  where  $\theta \in (0, \pi)$  and  $\phi \in (0, 2\pi)$ . In this chart  $\nabla_{round}$  is given by:

$$\begin{aligned} \Gamma_{22}^1(x) (x^{-1}(\theta, \phi)) &:= -\sin \theta \cos \theta \\ \Gamma_{12}^2(x) (x^{-1}(\theta, \phi)) &= \Gamma_{21}^2(x) (x^{-1}(\theta, \phi)) := \cot \theta \end{aligned}$$

and all other  $\Gamma$ 's vanish.

Then, by using the sloppy notation (familiar to us from classical mechanics) i.e  $x^1(p) = \theta(p)$  and  $x^2(p) = \phi(p)$ , the autoparallel transport equation reads:

$$\begin{aligned} \ddot{\theta} + \Gamma_{22}^1 \dot{\phi} \dot{\phi} &= 0 \\ \ddot{\phi} + 2\Gamma_{12}^2 \dot{\theta} \dot{\phi} &= 0 \end{aligned}$$

or:

$$\begin{aligned} \ddot{\theta} - \sin \theta \cos \theta \dot{\phi} \dot{\phi} &= 0 \\ \ddot{\phi} + 2 \cot \theta \dot{\theta} \dot{\phi} &= 0 \end{aligned}$$

which is exactly what one would obtain by the Lagrange equations. It can be seen that the above equations are satisfied at the equator where  $\theta(\lambda) = \pi/2$ , and  $\phi(\lambda) = \omega\lambda + \phi_0$  (running around the equator at constant speed  $\omega$ ). Thus, this curve is autoparallel. However,  $\phi(\lambda) = \omega\lambda^2 + \phi_0$  wouldn't be autoparallel.

### 6.1.2 Torsion

We say that the connection coefficients are not a tensor. The question now is, is any tensorial information in the connectio? In other words can we use  $\nabla_X$  to define tensors on  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$ ?

**Definition 6.8** (Torsion). *The **torsion** of a connection  $\nabla_X$  is the  $(1, 2)$ -tensor field:*

$$T(\omega, X, Y) := \omega(\nabla_X Y - \nabla_Y X - [X, Y])$$

where  $[X, Y]$ , called the commutator of  $X$  and  $Y$  is a vector field defined by  $[X, Y]f := X(Yf) - Y(Xf)$ .

*Proof.*

The definition of torsion is in a form of a theorem since we need to show that what we defined is actually a tensor. We shall check that  $T$  is  $C^\infty$ -linear in each entry:

- $T(f\omega, X, Y) = f\omega(\nabla_X Y - \nabla_Y X - [X, Y]) = fT(\omega, X, Y)$ .
- $T(\omega + \psi, X, Y) = (\omega + \psi)(\nabla_X Y - \nabla_Y X - [X, Y]) = T(\omega, X, Y) + T(\psi, X, Y)$ .
- It is:

$$\begin{aligned} T(\omega, fX, Y) &= \omega(\nabla_{fX} Y - \nabla_Y(fX) - [fX, Y]) \\ &= \omega(f\nabla_X Y - (\nabla_Y(f))X - f(\nabla_Y X) - [fX, Y]) \\ &= \omega(f\nabla_X Y - (Yf)X - f(\nabla_Y X) - [fX, Y]) \\ &= \omega(f\nabla_X Y - (Yf)X - f(\nabla_Y X) - fX(Y) + Y(fX)) \\ &= \omega(f\nabla_X Y - (Yf)X - f(\nabla_Y X) - fX(Y) + Y(f)X + fY(X)) \\ &= \omega(f\nabla_X Y - (Yf)X - f(\nabla_Y X) - (fX(Y) - fY(X)) + Y(f)X) \\ &= \omega(f\nabla_X Y - (Yf)X - f(\nabla_Y X) - f[X, Y] + (Yf)X) \\ &= \omega(f\nabla_X Y - f(\nabla_Y X) - f[X, Y]) \\ &= f\omega(\nabla_X Y - (\nabla_Y X) - [X, Y]) \\ &= fT(\omega, X, Y) \end{aligned}$$

Since  $T(\omega, X, Y) = -T(\omega, Y, X)$ , scaling in the last factor need not be checked separately.

- Finally it is:

$$\begin{aligned} T(\omega, X + Z, Y) &= \omega(\nabla_{X+Z} Y - \nabla_Y(X + Z) - [(X + Z), Y]) \\ &= \omega(\nabla_X Y + \nabla_Z Y - \nabla_Y X - \nabla_Y Z - (X + Z)Y + Y(X + Z)) \\ &= \omega(\nabla_X Y + \nabla_Z Y - \nabla_Y X + \nabla_Y Z - X(Y) - Z(Y) + Y(X) + Y(Z)) \\ &= \omega(\nabla_X Y - \nabla_Y X - X(Y) + Y(X)) + \omega(\nabla_Z Y - \nabla_Y Z - Z(Y) + Y(Z)) \\ &= \omega(\nabla_X Y - \nabla_Y X - [X, Y]) + \omega(\nabla_Z Y - \nabla_Y Z - [Z, Y]) \\ &= T(\omega, X, Y) + T(\omega, Z, Y) \end{aligned}$$

Since  $T(\omega, X, Y) = -T(\omega, Y, X)$ , additivity in the last factor need not be checked separately.

□

Now the question is what is exactly the information that torsion provides about the connection. The answer for that can be seen by invoking a chart, since the tensor components of torsion (or coefficient

function if you like) are:

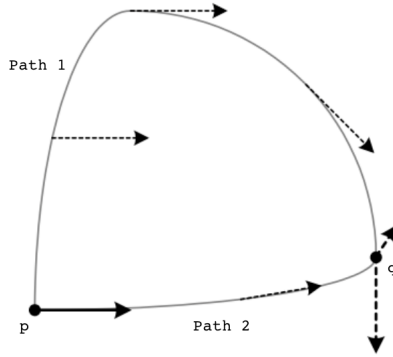
$$\begin{aligned}
T_{ab}^i &:= T\left(dx^i, \frac{\partial}{\partial x^a}, \frac{\partial}{\partial x^b}\right) \\
&= dx^i \left( \nabla_{\frac{\partial}{\partial x^a}} \left( \frac{\partial}{\partial x^b} \right) - \nabla_{\frac{\partial}{\partial x^b}} \left( \frac{\partial}{\partial x^a} \right) - \left[ \frac{\partial}{\partial x^a}, \frac{\partial}{\partial x^b} \right] \right) \\
&= dx^i \left( \Gamma_{ab}^c \frac{\partial}{\partial x^c} - \Gamma_{ba}^c \frac{\partial}{\partial x^c} \right) \\
&= \Gamma_{ab}^c dx^i \left( \frac{\partial}{\partial x^c} \right) - \Gamma_{ba}^c dx^i \left( \frac{\partial}{\partial x^c} \right) \\
&= \Gamma_{ab}^c \delta_c^i - \Gamma_{ba}^c \delta_c^i = \Gamma_{ab}^i - \Gamma_{ba}^i = 2\Gamma_{[ab]}^i
\end{aligned}$$

Hence the torsion is simply the antisymmetric part of connection coefficients which is also the part that one cannot get rid of with coordinate transformation, which of course make sense since if it is not zero in one chart it will not be zero everywhere.

**Definition 6.9** (Torsion-Free). *A manifold with connection  $(M, \mathcal{O}, \mathcal{A}, \nabla_X)$  is called **torsion-free** if the torsion of its connection is zero. That is,  $T = 0$ .*

### 6.1.3 Curvature

In this section we will show how the connection is related to the curvature of the underlying manifold. Let's start with an intuitive picture. In general, if we parallel transport a vector from a point  $p$  to a point  $q$  along two different paths (path 1) and (path 2), the resulting vectors at point  $q$  are different.



We said, in general, because, for example, if we parallel transport a vector in a Euclidean space, where the parallel transport is defined in our usual sense, the resulting vector does not depend on the path along which it has been parallel transported. This non-integrability of parallel transport characterizes the intrinsic notion of curvature, which does not depend on the special coordinates chosen.

Since the notion of parallel transport contains the connection, one expects that the connection has some information regarding the curvature of the manifold. And indeed this is the case. (Recall that the connection is just a generalization of the directional derivative). In order to formalize this mathematically, we ask what is the difference in applying two connections (here maybe the term covariant derivatives is more intuitive) in different order, i.e what is  $\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z$  (where this can be seen as the difference between path 1 and path 2). In other words we ask by how much the covariant derivatives fail to commute. The so called “Riemann curvature” is a quantity that measures exactly this failure.

**Definition 6.10** (Riemann Curvature). *The **Riemann curvature** of a connection  $\nabla_X$  is the  $(1,3)$ -tensor field:*

$$Riem(\omega, Z, X, Y) := \omega(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z)$$

where it can be shown that it is  $C^\infty$ -linear in each slot.

So indeed, as we can see:

$$\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z = \text{Riem}(\cdot, Z, X, Y) + \nabla_{[X, Y]} Z$$

In other word the difference we are interested in is the Riemann tensor plus a correction term. For this correction term note that by invoking a chart  $(U, x)$ , and by using the co-ordinate induced basis:

$$\left( \nabla_{\frac{\partial}{\partial x^a}} \nabla_{\frac{\partial}{\partial x^b}} Z \right)^m - \left( \nabla_{\frac{\partial}{\partial x^b}} \nabla_{\frac{\partial}{\partial x^a}} Z \right)^m = R_{nab}^m Z^n + \nabla_{[\frac{\partial}{\partial x^a}, \frac{\partial}{\partial x^b}]} Z$$

However basis vectors always commute so the last term vanishes, and now we can see how the *Riem* tensor components  $R_{nab}^m$  contain all the information about how the  $\nabla_{\frac{\partial}{\partial x^a}}$  and  $\nabla_{\frac{\partial}{\partial x^b}}$  fail to commute if they act on a vector field. Similarly if they act on a tensor field, there are several terms on the right hand side of the equation like the one term above and if they act on a function, of course they commute. Observe that, being a tensor, *Riem* vanishes in all coordinate systems if it vanishes in one coordinate system.

Now let's try to calculate the components of Riemann tensor.

$$\begin{aligned} R_{\sigma\mu\nu}^\rho &= \text{Riem}(dx^\rho, \frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}, \frac{\partial}{\partial x^\sigma}) \\ &= dx^\rho \left( \nabla_{\frac{\partial}{\partial x^\mu}} \nabla_{\frac{\partial}{\partial x^\nu}} \left( \frac{\partial}{\partial x^\sigma} \right) - \nabla_{\frac{\partial}{\partial x^\nu}} \nabla_{\frac{\partial}{\partial x^\mu}} \left( \frac{\partial}{\partial x^\sigma} \right) - \nabla_{[\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}]} \left( \frac{\partial}{\partial x^\sigma} \right) \right) \\ &= dx^\rho \left( \nabla_{\frac{\partial}{\partial x^\mu}} \left( \Gamma_{\nu\sigma}^\tau \frac{\partial}{\partial x^\tau} \right) - \nabla_{\frac{\partial}{\partial x^\nu}} \left( \Gamma_{\mu\sigma}^\tau \frac{\partial}{\partial x^\tau} \right) \right) \\ &= dx^\rho \left( \nabla_{\frac{\partial}{\partial x^\mu}} (\Gamma_{\nu\sigma}^\tau) \cdot \frac{\partial}{\partial x^\tau} + \Gamma_{\nu\sigma}^\tau \nabla_{\frac{\partial}{\partial x^\mu}} \left( \frac{\partial}{\partial x^\tau} \right) - \nabla_{\frac{\partial}{\partial x^\nu}} (\Gamma_{\mu\sigma}^\tau) \cdot \frac{\partial}{\partial x^\tau} - \Gamma_{\mu\sigma}^\tau \nabla_{\frac{\partial}{\partial x^\nu}} \left( \frac{\partial}{\partial x^\tau} \right) \right) \\ &= dx^\rho \left( \frac{\partial}{\partial x^\mu} (\Gamma_{\nu\sigma}^\tau) \cdot \frac{\partial}{\partial x^\tau} + \Gamma_{\nu\sigma}^\tau \Gamma_{\mu\tau}^\lambda \frac{\partial}{\partial x^\lambda} - \frac{\partial}{\partial x^\nu} (\Gamma_{\mu\sigma}^\tau) \cdot \frac{\partial}{\partial x^\tau} - \Gamma_{\mu\sigma}^\tau \Gamma_{\nu\tau}^\lambda \frac{\partial}{\partial x^\lambda} \right) \\ &= \frac{\partial}{\partial x^\mu} (\Gamma_{\nu\sigma}^\tau) dx^\rho \left( \frac{\partial}{\partial x^\tau} \right) + \Gamma_{\nu\sigma}^\tau \Gamma_{\mu\tau}^\lambda dx^\rho \left( \frac{\partial}{\partial x^\lambda} \right) - \frac{\partial}{\partial x^\nu} (\Gamma_{\mu\sigma}^\tau) dx^\rho \left( \frac{\partial}{\partial x^\tau} \right) - \Gamma_{\mu\sigma}^\tau \Gamma_{\nu\tau}^\lambda dx^\rho \left( \frac{\partial}{\partial x^\lambda} \right) \\ &= \frac{\partial}{\partial x^\mu} (\Gamma_{\nu\sigma}^\tau) \delta_\tau^\rho + \Gamma_{\nu\sigma}^\tau \Gamma_{\mu\tau}^\lambda \delta_\lambda^\rho - \frac{\partial}{\partial x^\nu} (\Gamma_{\mu\sigma}^\tau) \delta_\tau^\rho - \Gamma_{\mu\sigma}^\tau \Gamma_{\nu\tau}^\lambda \delta_\lambda^\rho \\ &= \frac{\partial}{\partial x^\mu} (\Gamma_{\nu\sigma}^\rho) + \Gamma_{\nu\sigma}^\tau \Gamma_{\mu\tau}^\rho - \frac{\partial}{\partial x^\nu} (\Gamma_{\mu\sigma}^\rho) - \Gamma_{\mu\sigma}^\tau \Gamma_{\nu\tau}^\rho \end{aligned}$$

Hence:

$$\frac{\partial}{\partial x^\mu} (\Gamma_{\nu\sigma}^\rho) - \frac{\partial}{\partial x^\nu} (\Gamma_{\mu\sigma}^\rho) + \Gamma_{\mu\tau}^\rho \Gamma_{\nu\sigma}^\tau - \Gamma_{\nu\tau}^\rho \Gamma_{\mu\sigma}^\tau$$

Coming back to the Euclidean space given that the connection is 0, it is straightforward to see that  $\text{Riem} = 0$  hence there is no curvature and this is why parallel transport does not depend on the path along which is performed.

## 6.2 Metric Manifolds

In this chapter we will establish yet another structure on a smooth manifold that allows one to assign vectors in each tangent space a length (and an angle between vectors in the same tangent space). From this structure, one can then define a notion of length of a curve. Then we can look at shortest curves (which will be called geodesics).

Requiring then that the shortest curves coincide with the straight curves (w.r.t.  $\nabla_X$ ) will result in  $\nabla_X$  being determined by the metric structure  $g$ .  $\nabla_X$ , in turn determines the curvature given by *Riem*.

**Definition 6.11** (Metric Tensor). *A metric  $g$  on a smooth manifold  $M$  is a  $(0, 2)$ -tensor field:*

$$g : \Gamma(TM) \times \Gamma(TM) \rightarrow C^\infty(M)$$

$$(X, Y) \mapsto g(X, Y)$$

*satisfying the following properties:*

- *Symmetry:*  $g(X, Y) = g(Y, X) \quad \forall X, Y \in \Gamma(TM)$ .
- *Non-degeneracy:* The so called “musical map”  $\flat$  defined as:

$$\flat : \Gamma(TM) \rightarrow \Gamma(T^*M)$$

$$X \mapsto \flat(X)$$

*where:*

$$\flat(X)(Y) := g(X, Y)$$

*is a  $C^\infty$ -isomorphism in other words, it is invertible.*

**Definition 6.12** (Inverse Metric Tensor). *The  $(2, 0)$ -tensor field  $g^{-1}$  with respect to a metric  $g$  is the symmetric map:*

$$g^{-1} : \Gamma(T^*M) \times \Gamma(T^*M) \rightarrow C^\infty(M)$$

$$(\omega, \sigma) \mapsto g^{-1}(\omega, \sigma) := \omega(\flat^{-1}(\sigma))$$

*with:*

$$\flat^{-1} : \Gamma(T^*M) \rightarrow \Gamma(TM)$$

$$\omega \mapsto \flat^{-1}(\omega)$$

*where:*

$$\flat^{-1}(\omega)(\sigma) := g^{-1}(\omega, \sigma)$$

*Remark 6.2.* In components:  $g_{\mu\nu} = g_{\nu\mu}$  and  $(g^{-1})^{\mu\nu} g_{\nu\rho} = \delta_\rho^\mu$ .

*Remark 6.3.* The musical map turns a vector  $X$  to a covector  $(\flat(X))$ . Given that  $(\flat(X))$  is a covector it carries, by the predefined convention, a “lower index”  $(\flat(X))_\mu$ . Hence in tensor components:

$$(\flat(X))_\mu := g_{\mu\nu} X^\nu$$

Similarly the the inverse musical map turns a covector  $\omega$  to a vector  $(\flat^{-1}(\omega))$ . Given that  $(\flat^{-1}(\omega))$  is a vector it carries, by the predefined convention, an “upper index”  $(\flat^{-1}(\omega))^\mu$ . Hence in tensor components:

$$(\flat^{-1}(\omega))^\mu := g^{\mu\nu} \omega_\nu$$

Most of the time people drop the  $\flat$  and  $\flat^{-1}$  symbol and simply write  $(\flat(X))_\mu \rightarrow X_\mu$  and  $(\flat^{-1}(\omega))^\mu \rightarrow \omega^\mu$ , hence the equations read:

$$X_\mu := g_{\mu\nu} X^\nu \quad \text{and} \quad \omega^\mu := g^{\mu\nu} \omega_\nu$$

which is the familiar “lower the index” and “raise the index” formula. However the problem with these formulae is that when we see  $X_\mu$  (or  $\omega^\mu$ ) we don’t really know if it is actually a vector (or a covector) or if it’s the musical map (or inverse musical map) applied to a covector (or a vector). For this reason we will stick to the musical map (and the inverse musical map) notation.

*Example 6.3.*

Consider the sphere  $(S^2, \mathcal{O}, \mathcal{A})$  and the chart  $(U, x)$ :

$$x^1 = \varphi \in (0, 2\pi), \quad x^2 = \theta \in (0, \pi)$$

Define the metric of the round sphere of radius  $R$ :

$$g_{ij}(x^{-1}(\theta, \varphi)) = \begin{bmatrix} R^2 & 0 \\ 0 & R^2 \sin^2 \theta \end{bmatrix}$$

where the matrix is just a way of collecting this information.

As we mentioned in the “vector space” chapter, bilinear forms carry a signature. Subsequently the metric tensor as a bilinear form carries a signature. The signature  $(v, p, r)$  of a metric tensor  $g$  is the number (counted with multiplicity) of positive, negative and zero eigenvalues of the real symmetric matrix  $g_{\mu\nu}$  of the metric tensor with respect to a basis. Alternatively, it can be defined as the dimensions of a maximal positive and null subspace.

The signature completely classifies the metric up to a choice of basis. The signature is often denoted by a pair of integers  $(v, p)$  (implying  $r = 0$ ), or as an explicit list of signs of eigenvalues such as  $(+, -, -, -)$  or  $(-, +, +, +)$  for the signatures  $(1, 3, 0)$  and  $(3, 1, 0)$ , respectively.

**Definition 6.13** (Riemannian Metric). *A metric is called **Riemannian** if it is a metric with a positive definite signature  $(v, 0)$ , i.e.:  $(+, +, \dots, +)$ .*

**Definition 6.14** (Lorentzian Metric). *A metric is called **Lorentzian** if it is a metric with signature  $(v, 1)$  or  $(1, p)$ , i.e.:  $(+, -, \dots, -)$ .*

### 6.2.1 Geodesics

**Definition 6.15** (Speed Of A Curve). *On a Riemannian metric manifold  $(M, \mathcal{O}, \mathcal{A}, g)$ , the **speed** of a curve at the point  $p = \gamma(\lambda)$  is the number:*

$$s(\lambda) = \sqrt{g(X_{\gamma, \gamma(\lambda)}, X_{\gamma, \gamma(\lambda)})}$$

**Definition 6.16** (Length Of A Curve). *Let  $\gamma : (0, 1) \rightarrow M$  be a smooth curve. Then the **length** of  $\gamma$  denoted by  $L[\gamma] \in \mathbb{R}$  is the number:*

$$L[\gamma] := \int_0^1 d\lambda s(\lambda) = \int_0^1 d\lambda \sqrt{g(X_{\gamma, \gamma(\lambda)}, X_{\gamma, \gamma(\lambda)})}$$

Note that in contrast with the usual way of teaching, velocity is more fundamental than speed and speed is more fundamental than length!

*Example 6.4.*

Reconsider the round sphere of radius  $R$  with metric:

$$g_{ij} = \begin{bmatrix} R^2 & 0 \\ 0 & R^2 \sin^2 \theta \end{bmatrix}$$

Consider its equator:

$$\begin{aligned} \theta(\lambda) &:= (x^1 \circ \gamma)(\lambda) = \frac{\pi}{2} \implies \theta'(\lambda) = 0 \\ \varphi(\lambda) &:= (x^2 \circ \gamma)(\lambda) = 2\pi\lambda^3 \implies \varphi'(\lambda) = 6\pi\lambda^2 \end{aligned}$$

In this chart the length is:

$$\begin{aligned} L[\gamma] &= \int_0^1 d\lambda \sqrt{g_{ij}(x^{-1}(\theta(\lambda), \varphi(\lambda)))(x^i \circ \gamma)'(\lambda)(x^j \circ \gamma)'(\lambda)} \\ &= \int_0^1 d\lambda \sqrt{R^2 \cdot 0 + R^2 \sin^2(\theta(\lambda)) 36\pi^2 \lambda^4} \\ &= 6\pi R \int_0^1 d\lambda \lambda^2 = 6\pi R \left[ \frac{1}{3} \lambda^3 \right]_0^1 = 2\pi R \end{aligned}$$

Observe that for any reparametrization of  $\gamma$ , the factors from the reparametrization will cancel with the corresponding factors coming from the metric and the result will be independent of the reparametrization (as it should be since the length of the curve should not depend on the parametrization). This actually is a theorem that can be easily proved (we skip the proof).

**Theorem 6.1.** *For any curve  $\gamma : (0, 1) \rightarrow M$  and any smooth, bijective and increasing reparametrization  $\sigma : (0, 1) \rightarrow (0, 1)$  of  $\gamma$ :*

$$L[\gamma] = L[\gamma \circ \sigma]$$

**Definition 6.17** (Geodesic). *A curve  $\gamma : (0, 1) \rightarrow M$  on a Riemannian manifold  $(M, \mathcal{O}, \mathcal{A}, g)$  is called a **geodesic** if it is a stationary curve with respect to a length functional  $L$ .*

As before let's right the directional derivative using the "curve notation":

$$\begin{aligned} X^m : \mathcal{C}^\infty(M) &\xrightarrow{\sim} \mathcal{C}^\infty(M) \\ f &\mapsto ((f \circ \gamma)'(0))^m := \dot{\gamma}^m \end{aligned}$$

Thus the length reads:

$$\begin{aligned} L[\gamma] &:= \int_0^1 d\lambda s(\lambda) = \int_0^1 d\lambda \sqrt{g(\dot{\gamma}, \dot{\gamma})} = \int_0^1 d\lambda \sqrt{g_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu} \\ 0 = \delta L[\gamma] &= \int_0^1 d\lambda \cdot \delta \left( \sqrt{g_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu} \right) = \int_0^1 d\lambda \cdot \frac{\delta (g_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu)}{2\sqrt{g_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu}} \end{aligned}$$

Using the product rule we get:

$$0 = \int_0^1 d\lambda \cdot \left( \dot{\gamma}^\mu \dot{\gamma}^\nu \delta g_{\mu\nu} + g_{\mu\nu} \delta \dot{\gamma}^\mu \dot{\gamma}^\nu + g_{\mu\nu} \dot{\gamma}^\mu \delta \dot{\gamma}^\nu \right) = \int_0^1 d\lambda \cdot \left( \dot{\gamma}^\mu \dot{\gamma}^\nu \partial_\alpha g_{\mu\nu} \delta x^\alpha + 2g_{\mu\nu} \delta \dot{\gamma}^\mu \dot{\gamma}^\nu \right)$$

Integrating by-parts the last term and dropping the total derivative (which equals to zero at the boundaries) we get that:

$$\begin{aligned} 0 &= \int_0^1 d\lambda \cdot \left( \dot{\gamma}^\mu \dot{\gamma}^\nu \partial_\alpha g_{\mu\nu} \delta x^\alpha - 2\delta x^\mu \frac{d}{d\tau} (g_{\mu\nu} \dot{\gamma}^\nu) \right) \\ &= \int_0^1 d\lambda \cdot \left( \dot{\gamma}^\mu \dot{\gamma}^\nu \partial_\alpha g_{\mu\nu} \delta x^\alpha - 2\delta x^\mu \partial_\alpha g_{\mu\nu} \dot{\gamma}^\alpha \dot{\gamma}^\nu - 2\delta x^\mu g_{\mu\nu} \ddot{\gamma}^\nu \right) \\ &= \int_0^1 d\lambda \cdot \delta x^\mu \cdot \left( -2g_{\mu\nu} \ddot{\gamma}^\nu + \dot{\gamma}^\alpha \dot{\gamma}^\nu \partial_\mu g_{\alpha\nu} - 2\partial_\alpha g_{\mu\nu} \dot{\gamma}^\alpha \dot{\gamma}^\nu \right) \\ &= \int_0^1 d\lambda \cdot \delta x^\mu \cdot \left( -2g_{\mu\nu} \ddot{\gamma}^\nu + \dot{\gamma}^\alpha \dot{\gamma}^\nu \partial_\mu g_{\alpha\nu} - \partial_\alpha g_{\mu\nu} \dot{\gamma}^\alpha \dot{\gamma}^\nu - \partial_\nu g_{\mu\alpha} \dot{\gamma}^\nu \dot{\gamma}^\alpha \right) \\ &= \int_0^1 d\lambda \cdot \delta x^\mu \cdot \left( g_{\mu\nu} \ddot{\gamma}^\nu + \frac{1}{2} \dot{\gamma}^\alpha \dot{\gamma}^\nu (\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}) \right) \end{aligned}$$

Hence by Hamilton's principle we find that the Euler-Lagrange equation is:

$$g_{\mu\nu} \ddot{\gamma}^\nu + \frac{1}{2} \dot{\gamma}^\alpha \dot{\gamma}^\nu (\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}) = 0$$

Multiplying by the inverse metric tensor  $g^{\mu\beta}$  we get that:

$$\ddot{\gamma}^\beta + \frac{1}{2} g^{\mu\beta} (\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}) \dot{\gamma}^\alpha \dot{\gamma}^\nu = 0$$

Thus we get the geodesic equation:

$$\ddot{\gamma}^\beta + \Gamma_{\alpha\nu}^\beta \dot{\gamma}^\alpha \dot{\gamma}^\nu = 0$$

where  $\Gamma_{\alpha\nu}^{\beta}$  are called “Christoffel symbols” defined in terms of the metric tensor as:

$$\Gamma_{\alpha\nu}^{\beta} = \frac{1}{2}g^{\mu\beta}(\partial_{\mu}g_{\alpha\nu} + \partial_{\nu}g_{\mu\alpha} - \partial_{\alpha}g_{\mu\nu})$$

# Chapter 7

## Lie Groups

A Lie group is a group whose elements are organized continuously and smoothly, as opposed to discrete groups, where the elements are separated, thus this makes Lie groups differentiable manifolds and as such can be studied using differential calculus, in contrast with the case of more general topological groups. Lie groups are named after Norwegian mathematician Sophus Lie, who laid the foundations of the theory of continuous transformation groups. One of the key ideas in the theory of Lie groups is to replace the global object, the group, with its local or linearised version, which Lie himself called its “infinitesimal group” and which has since become known as its Lie algebra.

In rough terms, a Lie group is a continuous group: it is a group whose elements are described by several real parameters. As such, Lie groups provide a natural model for the concept of continuous symmetry, such as rotational symmetry in three dimensions. Lie groups are widely used in many parts of modern mathematics and physics. Lie’s original motivation for introducing Lie groups was to model the continuous symmetries of differential equations, in much the same way that finite groups are used in Galois theory to model the discrete symmetries of algebraic equations.

Lie groups (and their associated Lie algebras) play a major role in modern physics, with the Lie group typically playing the role of a symmetry of a physical system. Here, the representations of the Lie group (or of its Lie algebra) are especially important. Representation theory is used extensively in particle physics. Groups whose representations are of particular importance include the rotation group  $SO(3)$  (or its double cover  $SU(2)$ ), the special unitary group  $SU(3)$  and the Poincare group.

### 7.1 Lie Groups

**Definition 7.1** (Lie Group). A **Lie group** is a group  $(G, \bullet)$ , where  $G$  is a smooth manifold and the maps:

$$\begin{aligned}\mu: G \times G &\rightarrow G \\ (g_1, g_2) &\mapsto g_1 \bullet g_2\end{aligned}$$

and:

$$\begin{aligned}i: G &\rightarrow G \\ g &\mapsto g^{-1}\end{aligned}$$

are both smooth. Note that  $G \times G$  inherits a smooth atlas from the smooth atlas of  $G$ .

**Definition 7.2** (Dimension Of Lie Group). The **dimension** of a Lie group  $(G, \bullet)$  is the dimension of  $G$  as a manifold.

*Example 7.1.*

- a) Consider  $(\mathbb{R}^n, +)$ , where  $\mathbb{R}^n$  is understood as a smooth  $n$ -dimensional manifold. This is a commutative (or abelian) Lie group (since  $\bullet$  is commutative), often called the  $n$ -dimensional translation group.

- b) Let  $S^1 := \{z \in \mathbb{C} \mid |z| = 1\}$  and let  $\cdot$  be the usual multiplication of complex numbers. Then  $(S^1, \cdot)$  is a commutative Lie group usually denoted  $U(1)$ .
- c) As we discussed in the chapter of vector spaces,  $\text{Aut}(V) := \{\phi: V \xrightarrow{\sim} V \mid \det \phi \neq 0\}$  is the set of all of linear isomorphisms on  $V$  that we denoted as  $GL(V)$ . To make it more concrete, we can take as the vector space  $V = \mathbb{R}^n$  hence  $GL(n, \mathbb{R}) = \{\phi: \mathbb{R}^n \xrightarrow{\sim} \mathbb{R}^n \mid \det \phi \neq 0\}$ . This set can be endowed with the structure of a smooth  $n^2$ -dimensional manifold, by noting that there is a bijection between linear maps  $\phi: \mathbb{R}^n \xrightarrow{\sim} \mathbb{R}^n$  and  $\mathbb{R}^{n^2}$ . Then,  $(GL(n, \mathbb{R}), \circ)$  is a Lie group called the *general linear group*.

**Definition 7.3** (Lie Group Homomorphism). *Let  $(G, \bullet)$  and  $(H, \circ)$  be Lie groups. A map  $\phi: G \rightarrow H$  is a **Lie group homomorphism** if it is a group homomorphism and a smooth map.*

**Definition 7.4** (Lie Group Isomorphism). *A **Lie group isomorphism** is a Lie group homomorphism which is also a diffeomorphism of the underlying manifolds.*

## 7.2 The Left Translation Map

To every element of a Lie group there is associated a special map. Note that everything we will do here can be done equivalently by using right translation maps.

**Definition 7.5** (Left Translation). *Let  $(G, \bullet)$  be a Lie group and let  $g \in G$ . The map:*

$$\begin{aligned} \ell_g: G &\rightarrow G \\ h &\mapsto \ell_g(h) := g \bullet h \equiv gh \end{aligned}$$

*is called the **left translation** by  $g$ .*

One might think that this is an overkill of notation since we already had the operation between two element from the group structure. However the left translation map is different, since we first have to fix an element of the group  $g$  (hence the index in  $\ell_g$ ) and then apply this element to the whole group (a.k.a to each element of the group).

If there is no danger of confusion, we usually suppress the  $\bullet$  notation.

**Proposition 7.1.** *Let  $G$  be a Lie group. For any  $g \in G$ , the left translation map  $\ell_g: G \rightarrow G$  is a isomorphism.*

*Proof.*

Let  $h, h' \in G$ . Then, we have:

$$\ell_g(h) = \ell_g(h') \Leftrightarrow gh = gh' \Leftrightarrow h = h'$$

Moreover, for any  $h \in G$ , we have  $g^{-1}h \in G$  and:

$$\ell_g(g^{-1}h) = gg^{-1}h = h$$

Therefore,  $\ell_g$  is a bijection on  $G$ .

Note that:

$$\ell_g = \mu(g, -)$$

and since  $\mu: G \times G \rightarrow G$  is smooth by definition, so is  $\ell_g$ .

The inverse map is  $(\ell_g)^{-1} = \ell_{g^{-1}}$ , since:

$$\ell_{g^{-1}} \circ \ell_g = \ell_g \circ \ell_{g^{-1}} = \text{id}_G$$

Then, for the same reason as above with  $g$  replaced by  $g^{-1}$ , the inverse map  $(\ell_g)^{-1}$  is also smooth. Hence, the map  $\ell_g$  is indeed an isomorphism.  $\square$

Note that,  $\ell_g$  is not an isomorphism of groups, i.e.:

$$\ell_g(hh') \neq \ell_g(h) \ell_g(h')$$

in general. However that does not stop  $\ell_g$  from being an isomorphism, which actually means that is a diffeomorphism of the underlying manifolds.

Since a lie group is a topological manifold, on top of being a group, this means that at any point  $g$  we can define the tangent space, and by following the analysis we did in the previous chapter to define fields on  $G$ . Recall from the previous chapter that once we have a diffeomorphism  $\phi$  between two manifolds  $M$  and  $N$ , we can define the push-forward of a vector field  $X$  as:

$$\phi_*(X)(\phi(p)) := (\phi_*)_p(X_p)$$

where  $X_p$  is the vector created by the field  $X$  at point  $p$ .

Coming in our case, we just showed that the map  $\ell_g: G \rightarrow G$  is a diffeomorphism so we can push-forward any vector field  $X$  on  $G$  to another vector field (again on  $G$  since the maps is between the same manifold). So in our case  $\phi_*(X) = (\ell_g)_*(X)$  and for any point  $h$  in  $G$ :  $\ell_g(h) = gh$  so the push-forward equation reads:

$$\ell_{g*}(X)(gh) := (\ell_{g*})_h(X_h)$$

### 7.3 The Lie Algebra Of A Lie Group

In Lie theory, we are typically not interested in general vector fields, but rather on special class of vector fields which are invariant under the induced push-forward of the left translation maps  $\ell_g$ .

**Definition 7.6** (Left Invariant Vector Fields). *Let  $G$  be a Lie group. A vector field  $X \in \Gamma(TG)$  is said to be **left-invariant** if:*

$$\forall g \in G: \ell_{g*}(X) = X$$

*Equivalently, we can require this to hold pointwise:*

$$\forall g, h \in G: (\ell_{g*})_h(X_h) = X_{gh}$$

*where basically what this means is that if we take the vector  $X_h$  at  $h$  and the vector  $X_{gh}$  at  $gh$ , then by pushing forward  $X_h$  with  $(\ell_{g*})_h$  (hence bringing it to the point  $gh$ ), the final result should be equal to the vector  $X_{gh}$ .*

We can manipulate a bit the pointwise formulation to yield another reformulation. Since both sides are vectors we can let them act on a function  $f$ :

$$(\ell_{g*})_h(X_h)f = X_{gh}f$$

By using the definition of a push-forward of a vector  $(\phi_*)_p(X_p)f := X_p(f \circ \phi)$  the left part of the equation reads:

$$(\ell_{g*})_h(X_h)f = X_h(f \circ \ell_g) = X(h)(f \circ \ell_g) = X(f \circ \ell_g)(h)$$

The right part can be manipulated as follows:

$$X_{gh}f = X(gh)f = (Xf)(gh) = (Xf)(\ell_g(h)) = (Xf) \circ \ell_g(h)$$

By substituting both final expressions back to the original one and discarding the point  $h$  since they must be true for any  $h$  we obtain the last reformulation of the push-forward:

$$X(f \circ \ell_g) = X(f) \circ \ell_g$$

Once again, let's emphasize that this equations holds only for left invariant vector fields and not all vector fields  $X$  of  $\Gamma(TG)$ .

**Definition 7.7** ( $\mathcal{L}(G)$  (As A Set)). *We denote the set of all left-invariant vector fields on  $G$  as  $\mathcal{L}(G)$ .*

Of course:

$$\mathcal{L}(G) \subseteq \Gamma(TG)$$

but, in fact, more is true. Recall that we equipped  $(\Gamma(TG))$  with two operations and we showed that  $(\Gamma(TG), +, \cdot)$  is in fact a  $\mathcal{C}^\infty(G)$ -submodule. One can check that  $\mathcal{L}(G)$  is closed under:

$$\begin{aligned} +: \mathcal{L}(G) \times \mathcal{L}(G) &\rightarrow \mathcal{L}(G) \\ \cdot: \mathcal{C}^\infty(G) \times \mathcal{L}(G) &\rightarrow \mathcal{L}(G), \end{aligned}$$

therefore,  $\mathcal{L}(G)$  is a  $\mathcal{C}^\infty(G)$ -submodule of  $\Gamma(TG)$ .

On top of that we said that  $(\Gamma(TG), +, \cdot)$  can also be seen as an  $\mathbb{R}$ -vector space. Up to now, we have refrained from thinking of  $\Gamma(TG)$  as an  $\mathbb{R}$ -vector space since it is infinite-dimensional and, even worse, a basis is in general uncountable. A priori, this could be true for  $\mathcal{L}(G)$  as well, but we will see that the situation is, in fact, much nicer as  $\mathcal{L}(G)$  will turn out to be a finite-dimensional vector space over  $\mathbb{R}$  (as an  $\mathbb{R}$ -vector subspace of  $\Gamma(TG)$ ). Let's see why, since the reason why it is so it's of crucial importance.

**Theorem 7.1.** *Let  $G$  be a Lie group with identity element  $e \in G$ . Then  $\mathcal{L}(G) \cong_{\text{vec}} T_e G$ .*

*Proof.*

We will construct a linear isomorphism  $j: T_e G \xrightarrow{\sim} \mathcal{L}(G)$ . Define:

$$\begin{aligned} j: T_e G &\rightarrow \Gamma(TG) \\ X_{\gamma,e} &\mapsto j(X_{\gamma,e}) \end{aligned}$$

where  $j(X_{\gamma,e})$  is define as:

$$\begin{aligned} j(X_{\gamma,e}): G &\rightarrow TG \\ g &\mapsto j(X_{\gamma,e})(g) := (\ell_{g*})_e(X_{\gamma,e}) \quad \forall g \in G \end{aligned}$$

*Remark 7.1.* Recall that we dropped the curve  $\gamma$  in the notation of a vector  $X_p$  to save space on the notation. However since we are going to need the definition of a vector, that contains the curve, we will write it again as  $X_{\gamma,p}$  for now.

Now we have to prove that this is actually a linear isomorphism, and we will do it in steps.

- i) First we need to check that the definition we provided is consistent. Focusing on the left side of the equation,  $j(X_{\gamma,e})$  as a field associates to every point  $g$  of the manifold  $G$  a tangent vector, hence the first part of the definition  $j(X_{\gamma,e})(g)$  is a tangent vector of the tangent space  $T_g G$ .

Now we need to see if the second part of the equation is again a vector on  $T_g G$ . Recall that under a push-forward a vector of  $T_p M$  goes to a vector of  $T_{\phi(p)} N$ , hence in our case the vector  $X_{\gamma,e}$  of  $T_e G$  goes to the vector  $(\ell_{g*})_e X_{\gamma,e}$  which lies on  $T_{\ell_g(e)} G = T_{ge} G = T_g G$ .

Hence there is consistency in our definition.

- ii) Second, we show that for any  $X_{\gamma,e} \in T_e G$ ,  $j(X_{\gamma,e})$  is a smooth vector field on  $G$ . It suffices to check that for any  $f \in \mathcal{C}^\infty(G)$ , we have  $j(X_{\gamma,e})(f) \in \mathcal{C}^\infty(G)$ . Indeed:

$$\begin{aligned} j(X_{\gamma,e})(g) &:= (\ell_{g*})_e(X_{\gamma,e}) && \text{(definition of } j) \\ j(X_{\gamma,e})(g)(f) &= (\ell_{g*})_e(X_{\gamma,e})(f) && \text{(act on a function)} \\ &= X_{\gamma,e}(f \circ \ell_g) && \text{(definition of push-forward)} \\ &= (f \circ \ell_g \circ \gamma)'(0) && \text{(definition of tangent vector)} \end{aligned}$$

The map:

$$\begin{aligned} \varphi: \mathbb{R} \times G &\rightarrow \mathbb{R} \\ (t, g) &\mapsto \varphi(t, g) := (f \circ \ell_g \circ \gamma)(t) \\ &= f(g\gamma(t)) \end{aligned}$$

is a composition of smooth maps, hence it is smooth. Then:

$$j(X_{\gamma,e})(g)(f) = (\partial_1 \varphi)(0, g)$$

depends smoothly on  $g$  and thus  $j(X_{\gamma,e})(f) \in \mathcal{C}^\infty(G)$ .

- iii) We proved that  $j(X_{\gamma,e})$  is indeed a smooth vector field, however now we need to prove that it is a left invariant vector field since it is an element of  $\Gamma(TG)$ . To prove that we need to show that it's invariant under left translations. Let  $g, h \in G$ . Then, for every  $X_{\gamma,e} \in T_e G$ , we have:

$$\begin{aligned} j(X_{\gamma,e})(g) &:= (\ell_{g*})_e(X_{\gamma,e}) && \text{(definition of } j) \\ (\ell_{h*})_e(j(X_{\gamma,e})(g)) &= (\ell_{h*})_e((\ell_{g*})_e(X_{\gamma,e})) && \text{(acting with } (\ell_{h*})_e) \\ &= (\ell_{gh*})_e(X_{\gamma,e}) && \text{(one can show that)} \\ &= j(X_{\gamma,e})(gh) && \text{(definition of } j) \end{aligned}$$

so  $j(X_{\gamma,e}) \in \mathcal{L}(G)$ . Hence, the map  $j$  is really  $j: T_e G \rightarrow \mathcal{L}(G)$ .

- iv) We also need to check the linearity. Let  $X_{\gamma,e}, Y_{\gamma,e} \in T_e G$  and  $\lambda \in \mathbb{R}$ . Then, for any  $g \in G$ :

$$\begin{aligned} j(\lambda X_{\gamma,e} + Y_{\gamma,e})(g) &= (\ell_{g*})_e(\lambda X_{\gamma,e} + Y_{\gamma,e}) && \text{(definition of } j) \\ &= \lambda(\ell_{h*})_e(X_{\gamma,e}) + (\ell_{h*})_e(Y_{\gamma,e}) \\ &= \lambda j(X_{\gamma,e})(g) + j(Y_{\gamma,e})(g) \end{aligned}$$

since the push-forward is an  $\mathbb{R}$ -linear map. Hence, we have  $j: T_e G \xrightarrow{\sim} \mathcal{L}(G)$ .

- v) We also need to check that the map is injective. Let  $X_{\gamma,e}, Y_{\gamma,e} \in T_e G$ . Then:

$$\begin{aligned} j(X_{\gamma,e}) = j(Y_{\gamma,e}) &\Leftrightarrow \forall g \in G : j(X_{\gamma,e})(g) = j(Y_{\gamma,e})(g) \\ &\Rightarrow j(X_{\gamma,e})(e) = j(Y_{\gamma,e})(e) \\ &\Leftrightarrow (\ell_{e*})_e(X_{\gamma,e}) = (\ell_{e*})_e(Y_{\gamma,e}) && \text{(definition of } j) \\ &\Leftrightarrow X_{\gamma,e} = Y_{\gamma,e} && ((\ell_{e*})_e = \text{id}_{T_e G}) \end{aligned}$$

Hence, the map  $j$  is injective.

- vi) Finally we need to check that the map is surjective. Let  $X \in \mathcal{L}(G)$ . Then, we have:

$$\begin{aligned} j(X_{\gamma,e})(g) &= (\ell_{g*})_e(X_{\gamma,e}) && \text{(definition of } j) \\ &= (\ell_{g*})_e(X(e)) && (X(e) = X_{\gamma,e}) \\ &= X(ge) && (X \text{ is left-invariant}) \\ &= X(g) \end{aligned}$$

Hence  $X = j(X_{\gamma,e})$  and thus  $j$  is surjective.

Therefore,  $j: T_e G \xrightarrow{\sim} \mathcal{L}(G)$  is indeed a linear isomorphism. □

**Corollary 7.1.** *It is  $\dim \mathcal{L}(G) = \dim G$ , hence  $\mathcal{L}(G)$  turns out to be a finite-dimensional vector space over  $\mathbb{R}$  (as an  $\mathbb{R}$ -vector subspace of  $\Gamma(TG)$ ).*

So we proved that indeed  $\mathcal{L}(G)$  is a finite-dimensional vector space, but as we said there is something more important here. Namely, since  $j: T_e G \xrightarrow{\sim} \mathcal{L}(G)$  is a linear isomorphism this means that the spaces  $\mathcal{L}(G)$  and  $T_e G$  are isomorphic which with its turn it means that the map  $j$ , first of all has an inverse, and most importantly maps each element of  $\mathcal{L}(G)$  to exactly one element of  $T_e G$ . In other words there are as many left invariant fields as there are tangent vectors to the Lie group at the identity, so instead of studying the (quite complicated) left invariant vector fields we can simply study the (quite simpler) tangent vectors at the identity.

However, we can go one step further now. Recall from previous chapter that by considering  $\Gamma(TM)$  as an infinite-dimensional  $R$ -vector space, for two vector fields  $X, Y \in \Gamma(TM)$ , we can define their Lie bracket

to be the commutator of the fields:

$$[X, Y](f) := X(Y(f)) - Y(X(f))$$

for any  $f \in \mathcal{C}^\infty(M)$ . Then  $(\Gamma(TG), +, \cdot, [-, -])$  is a Lie algebra. Coming to our case, since  $\mathcal{L}(G)$  is a subvector space (and a submodule) of  $\Gamma(TM)$  we can inherit the commutator to  $\mathcal{L}(G)$  and ask if  $\mathcal{L}(G)$  is closed under the commutator so  $(\mathcal{L}(G), +, \cdot, [-, -])$  is a subalgebra of  $(\Gamma(TG), +, \cdot, [-, -])$ . Indeed, this is the case.

**Theorem 7.2.** *Let  $G$  be a Lie group. Then  $\mathcal{L}(G)$  is a Lie subalgebra of  $\Gamma(TG)$ .*

*Proof.*

A Lie subalgebra of a Lie algebra is simply a vector subspace which is closed under the action of the Lie bracket. Therefore, we only need to check that:

$$\forall X, Y \in \mathcal{L}(G) : [X, Y] \in \mathcal{L}(G)$$

Let  $X, Y \in \mathcal{L}(G)$ . For any  $g \in G$  and  $f \in \mathcal{C}^\infty(G)$ , we have:

$$\begin{aligned} [X, Y](f \circ \ell_g) &:= X(Y(f \circ \ell_g)) - Y(X(f \circ \ell_g)) \\ &= X(Y(f) \circ \ell_g) - Y(X(f) \circ \ell_g) \\ &= X(Y(f)) \circ \ell_g - Y(X(f)) \circ \ell_g \\ &= (X(Y(f)) - Y(X(f))) \circ \ell_g \\ &= [X, Y](f) \circ \ell_g \end{aligned}$$

Hence,  $[X, Y]$  is left-invariant. □

To summarise, we began with  $\mathcal{L}(G)$  as a set of all left invariant vector fields of  $G$ , which is a subset of  $\Gamma(TG)$ , then we inherited the  $+$  and  $\cdot$  of  $\Gamma(TG)$  to  $\mathcal{L}(G)$  and we showed that it is also a submodule and a subvector space of  $\Gamma(TG)$ , and finally we inherited the Lie bracket from  $\Gamma(TG)$  and we showed that it is also a subalgebra of  $\Gamma(TG)$ . From now on when we will be referring to  $\mathcal{L}(G)$ , we will mean its algebra structure.

**Definition 7.8** (The Lie Algebra Of A Lie Group). *Let  $G$  be a Lie group. We call the Lie algebra  $\mathcal{L}(G)$  of all left invariant vector fields of  $G$  the **Lie algebra of the Lie group  $G$** .*

We already showed that the underlying vector space of the Lie algebra of a Lie group  $\mathcal{L}(G)$  is isomorphic to the tangent vector of the Lie group  $G$  at the identity  $T_e G$ . We will now see that the identification of  $\mathcal{L}(G)$  and  $T_e G$  goes beyond the level of linear isomorphism as vector spaces, as they are isomorphic as algebras. Of course we cannot simply inherit the commutator from  $\mathcal{L}(G)$  to  $T_e G$  because the commutator is defined for fields and it works because one field can act on another. In the case of  $T_e G$  we cannot act with a vector on another vector. However, we can use the commutator from  $\mathcal{L}(G)$  to construct an appropriate Lie bracket on  $T_e G$  such that they be isomorphic as algebras.

Recall from the Lie algebras chapter in the notes that two algebras are called isomorphic if there exists an isomorphism between them, a.k.a a bijective map  $\phi$  such that:

$$\forall x, y \in L_1 : \phi([x, y]_{L_1}) = [\phi(x), \phi(y)]_{L_2}$$

Well, we already have a bijective map  $j$  so by using the commutator  $[-, -]_{\mathcal{L}(G)}$  on  $\mathcal{L}(G)$  we can define a Lie bracket for any  $X_e, Y_e \in T_e G$  as:

$$[X_e, Y_e]_{T_e G} := j^{-1}([j(X_e), j(Y_e)]_{\mathcal{L}(G)})$$

where  $j^{-1}$  works the opposite way, i.e takes a vector field and produces a vector.

Equipped with these Lie bracket, by definition, we have:

$$\mathcal{L}(G) \cong_{\text{Lie alg}} T_e G$$

Hence, given a Lie group we have seen how we can construct its corresponding Lie algebra as the space of left-invariant vector fields and we also showed that this algebra is isomorphic to the algebra of tangent vectors at the identity. We will later explore the opposite direction, i.e. given a Lie algebra, we will see how to construct a Lie group whose associated Lie algebra is the one we started from.

## 7.4 Application: $\mathrm{SL}(2, \mathbb{C})$ - Part 2

In the first part of the application in the previous chapter, we defined the set  $\mathrm{SL}(2, \mathbb{C})$  as a subset of  $\mathbb{C}^4 := \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C}$ . Then we showed that:

- $\mathrm{SL}(2, \mathbb{C})$  can be made into a group.
- $\mathrm{SL}(2, \mathbb{C})$  can be made into a topological space.
- $\mathrm{SL}(2, \mathbb{C})$  can be made into a topological manifold.
- $\mathrm{SL}(2, \mathbb{C})$  can be made into a complex differentiable manifold.

Hence we have left with  $\mathrm{SL}(2, \mathbb{C})$  as a 3-dimensional, complex differentiable manifold.

### $\mathrm{SL}(2, \mathbb{C})$ As A Lie Group

We equipped  $\mathrm{SL}(2, \mathbb{C})$  with both a group and a manifold structure. In order to obtain a Lie group structure, we have to check that these two structures are compatible, that is, we have to show that the two maps:

$$\begin{aligned} \mu: \mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) &\rightarrow \mathrm{SL}(2, \mathbb{C}) \\ \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right) &\mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & h \end{pmatrix} \end{aligned}$$

and :

$$\begin{aligned} i: \mathrm{SL}(2, \mathbb{C}) &\rightarrow \mathrm{SL}(2, \mathbb{C}) \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \end{aligned}$$

are differentiable with respect to the differentiable structure on  $\mathrm{SL}(2, \mathbb{C})$ .

For instance, for the inverse map  $i$ , we have to show that the map  $y \circ i \circ x^{-1}$  is differentiable in the usual for any pair of charts  $(U, x), (V, y) \in \mathcal{A}$ .

$$\begin{array}{ccc} U \subseteq \mathrm{SL}(2, \mathbb{C}) & \xrightarrow{i} & V \subseteq \mathrm{SL}(2, \mathbb{C}) \\ \downarrow x & & \downarrow y \\ x(U) \subseteq \mathbb{C}^3 & \xrightarrow{y \circ i \circ x^{-1}} & y(V) \subseteq \mathbb{C}^3 \end{array}$$

where we remind that:

$$\begin{aligned} x^{-1}: x(U) &\rightarrow U \\ (a, b, c) &\mapsto \begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix} \end{aligned}$$

and:

$$\begin{aligned} y: V &\rightarrow x(V) \subseteq \mathbb{C} \times \mathbb{C}^* \times \mathbb{C} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto (a, b, d) \end{aligned}$$

Since  $\mathrm{SL}(2, \mathbb{C})$  is connected, the differentiability of the transition maps in  $\mathcal{A}$  implies that if  $y \circ i \circ x^{-1}$  is differentiable for any two given charts, then it is differentiable for all charts in  $\mathcal{A}$ . Hence, we can simply let  $(U, x)$  and  $(V, y)$  be the two charts on  $\mathrm{SL}(2, \mathbb{C})$  defined above. Then, we have:

$$(y \circ i \circ x^{-1})(a, b, c) = (y \circ i)\left(\begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix}\right) = y\left(\begin{pmatrix} \frac{1+bc}{a} & -b \\ -c & a \end{pmatrix}\right) = \left(\frac{1+bc}{a}, -b, a\right)$$

which is certainly complex differentiable as a map between open subsets of  $\mathbb{C}^3$  (recall that  $a \neq 0$  on  $x(U)$ ). We have to do the whole process again for  $x \circ i \circ y^{-1}$  and show that is complex differentiable (which it is), hence we conclude that indeed the map  $i$  is complex differentiable.

Checking that  $\mu$  is complex differentiable is slightly more involved, since we first have to equip  $\mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C})$  with a suitable “product differentiable structure” and then proceed as above. Once that is done, we can finally conclude that  $((\mathrm{SL}(2, \mathbb{C}), \mathcal{O}, \mathcal{A}), \bullet)$  is a 3-dimensional complex Lie group.

### The Lie Algebra $\mathfrak{sl}(2, \mathbb{C})$ Of The Lie Group $\mathrm{SL}(2, \mathbb{C})$

Recall that to every Lie group  $G$ , there is an associated Lie algebra  $\mathcal{L}(G)$  of all left invariant vector fields of  $G$  i.e:

$$\mathcal{L}(G) := \{X \in \Gamma(TG) \mid \forall g \in G : \ell_{g*}(X) = X\}$$

where the left translation map  $\ell_g$  was given by:

$$\ell_g(h) := g \bullet h \equiv gh$$

Coming to our case we have that  $G = \mathrm{SL}(2, \mathbb{C})$  hence the corresponding Lie algebra of  $\mathrm{SL}(2, \mathbb{C})$  usually denoted by small letters  $\mathfrak{sl}(2, \mathbb{C})$  is:

$$\mathfrak{sl}(2, \mathbb{C}) := \mathcal{L}(\mathrm{SL}(2, \mathbb{C})) := \{X \in \Gamma(T\mathrm{SL}(2, \mathbb{C})) \mid \forall g \in G : \ell_{g*}(X) = X\}$$

where the left translation map at a point  $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  of  $\mathrm{SL}(2, \mathbb{C})$  is:

$$\ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

Now, our goal is to classify  $\mathfrak{sl}(2, \mathbb{C})$  as a Lie algebra. The first step we need to do it to find its structure constants. This can be done by considering two vector fields  $X, Y \in \Gamma(T\mathrm{SL}(2, \mathbb{C}))$  and computing the commutator:

$$[X, Y] := X(Y) - Y(X)$$

However, as we proved earlier, we can always use the fact that the corresponding Lie algebra  $\mathcal{L}(G)$  of a Lie group  $G$  is isomorphic to the Lie algebra  $T_e G$  with Lie bracket:

$$[X_e, Y_e]_{T_e G} := j^{-1}([j(X_e), j(Y_e)]_{\mathcal{L}(G)})$$

with:

$$j(X_e)(g) := (\ell_{g*})_e(X_e), \quad \forall g \in G$$

Hence, in our case we can study the lie algebra  $\mathfrak{sl}(2, \mathbb{C})$  by focusing on the tangent space  $T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \mathrm{SL}(2, \mathbb{C})$  of  $\mathrm{SL}(2, \mathbb{C})$ , and compute the Lie bracket there as:

$$[X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, Y_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}]_{T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \mathrm{SL}(2, \mathbb{C})} := j^{-1}([j(X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}), j(Y_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}})]_{\mathfrak{sl}(2, \mathbb{C})})$$

with:

$$j(X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}})\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \left(\ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}*}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\left(X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right)$$

So we need to compute this bracket. First thing first, for any vector  $X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \in T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \mathrm{SL}(2, \mathbb{C})$  we have

to compute  $j(X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}})$ .

Recall that if  $(U, x)$  is a chart on a manifold  $M$  and  $p \in U$ , then the chart  $(U, x)$  induces the co-ordinate induced basis which is a basis of the tangent space  $T_p M$ . In our case we can use the chart  $(U, x)$  that contains the identity  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  (we must include the identity since we are interested in the tangent space at the identity  $T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \text{SL}(2, \mathbb{C})$ ) and hence we get an induced co-ordinate basis:

$$\left\{ \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \in T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \text{SL}(2, \mathbb{C}) \mid 1 \leq i \leq 3 \right\}$$

so that any  $X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \in T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \text{SL}(2, \mathbb{C})$  can be written as:

$$X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} = X^i \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \quad X^i \in \mathbb{C}$$

Hence it suffices to compute the bracket only for the basis, since every other vector can be reconstructed by using the basis.

(In what follows. note that the  $d$  appearing is completely redundant, since the membership condition of  $\text{SL}(2, \mathbb{C})$  forces  $d = \frac{1+bc}{a}$ . However, we will keep writing the  $d$  to avoid having a fraction in a matrix in a subscript).

Let us now determine the image of these co-ordinate induced basis elements under the isomorphism  $j$ . The object:

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right) \in \mathfrak{sl}(2, \mathbb{C})$$

is a left-invariant vector field on  $\text{SL}(2, \mathbb{C})$ . It assigns to each point  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U \subseteq \text{SL}(2, \mathbb{C})$  the tangent vector:

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right) \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) := \left( \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}*} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}$$

By acting on a function  $f$ :

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right) \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) (f) := \left( \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}*} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} (f)$$

by using the definition of the push-forward:

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right) \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) (f) = \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} (f \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}})$$

and by expanding the notation of  $\left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}$  to its actual form we obtain:

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right) \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) (f) = \partial_i (f \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1}) (x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})$$

We can now manipulate a bit the left part of the equation by inserting an identity in the composition and we obtain:

$$\begin{aligned} j \left( \left( \frac{\partial}{\partial x^i} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right) \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) (f) &= \partial_i (f \circ \text{id}_U \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1}) (x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) \\ &= \partial_i (f \circ (x^{-1} \circ x) \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1}) (x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) \\ &= \partial_i ((f \circ x^{-1}) \circ (x \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})) (x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) \end{aligned}$$

where  $f \circ x^{-1}: x(U) \subseteq \mathbb{C}^3 \rightarrow \mathbb{C}$  and  $(x \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1}): x(U) \subseteq \mathbb{C}^3 \rightarrow x(U) \subseteq \mathbb{C}^3$  and hence, we can use the multi-dimensional chain rule to obtain:

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right)\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)(f) = \left(\partial_m(f \circ x^{-1})((x \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})(x(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})))\right)\left(\partial_i(x^m \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})(x(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}))\right)$$

with the summation going from  $m = 1$  to  $m = 3$ .

The first factor of the right part of the equation is simply:

$$\begin{aligned} (\partial_m(f \circ x^{-1})((x \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})(x(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})))) &= (\partial_m(f \circ x^{-1})((x \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}})x^{-1}(x(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})))) \\ &= \partial_m(f \circ x^{-1})((x \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}})(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})) \\ &= \partial_m(f \circ x^{-1})(x(\begin{pmatrix} a & b \\ c & d \end{pmatrix})) \\ &= \left(\frac{\partial}{\partial x^m}\right)_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}(f) \end{aligned}$$

To see what the second factor of the right part of the equation is, we first consider the map  $x^m \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1}$ . This map acts on the triple  $(e, f, g) \in x(U)$  as:

$$\begin{aligned} (x^m \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})(e, f, g) &= (x^m \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}})\begin{pmatrix} e & f \\ g & \frac{1+fg}{e} \end{pmatrix} \\ &= x^m\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & \frac{1+fg}{e} \end{pmatrix}\right) \\ &= x^m\left(\begin{pmatrix} ae + bg & af + \frac{b(1+fg)}{e} \\ ce + dg & cf + \frac{d(1+fg)}{e} \end{pmatrix}\right) \\ &= (ae + bg, af + \frac{b(1+fg)}{e}, ce + dg) \end{aligned}$$

We now have to apply  $\partial_i$  to this map, with  $i \in \{1, 2, 3\}$ , i.e. we have to differentiate with respect to each of the three complex variables  $e$ ,  $f$ , and  $g$ . Of course depending to which derivative we take, and to which slot we apply it we end up with  $3 \times 3$  different results, that we will collect in one matrix just for notation purposes. More specifically we can write the result as:

$$\partial_i(x^m \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})(e, f, g) = D(e, f, g)^m_i$$

where  $m$  labels the rows and  $i$  the columns of the matrix. (So for example the first row of the first column means the derivative of the first component (a.k.a  $ae + bg$ ) with respect to the first variable (a.k.a  $e$ ) hence the result (a.k.a  $a$ ). If we do that for all components and all derivatives we end up with:

$$D(e, f, g) = \begin{pmatrix} a & 0 & b \\ -\frac{b(1+fg)}{e^2} & a + \frac{bg}{e} & \frac{bf}{e} \\ c & 0 & d \end{pmatrix}$$

Finally, by evaluating this at  $(e, f, g) = x(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) = (1, 0, 0)$ , we obtain:

$$\partial_i(x^m \circ \ell_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \circ x^{-1})(x(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})) = D(1, 0, 0)^m_i$$

where, by recalling that  $d = \frac{1+bc}{a}$  we simply get:

$$D(1, 0, 0) = \begin{pmatrix} a & 0 & b \\ -b & a & 0 \\ c & 0 & \frac{1+bc}{a} \end{pmatrix} := D^m_i$$

Hence we computed the two parts of the equation above, so we can combine them and we get:

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right)\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)(f) = D^m_i \left(\frac{\partial}{\partial x^m}\right)_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}(f)$$

Since this holds for an arbitrary  $f \in \mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C}))$ , we have:

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right)\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = D^m_i \left(\frac{\partial}{\partial x^m}\right)_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}$$

and since the point  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U \subseteq \mathrm{SL}(2, \mathbb{C})$  is also arbitrary, we have:

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right) = D^m_i \frac{\partial}{\partial x^m} \in \mathfrak{sl}(2, \mathbb{C})$$

where  $D$  is now the corresponding matrix of co-ordinate functions:

$$D := \begin{pmatrix} x^1 & 0 & x^2 \\ -x^2 & x^1 & 0 \\ x^3 & 0 & \frac{1+x^2x^3}{x^1} \end{pmatrix}$$

Hence, we have an expansion of the images of the basis of  $T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \mathrm{SL}(2, \mathbb{C})$  under  $j$ :

$$\begin{aligned} j\left(\left(\frac{\partial}{\partial x^1}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right) &= x^1 \frac{\partial}{\partial x^1} - x^2 \frac{\partial}{\partial x^2} + x^3 \frac{\partial}{\partial x^3} \\ j\left(\left(\frac{\partial}{\partial x^2}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right) &= x^1 \frac{\partial}{\partial x^2} \\ j\left(\left(\frac{\partial}{\partial x^3}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right) &= x^2 \frac{\partial}{\partial x^1} + \frac{1+x^2x^3}{x^1} \frac{\partial}{\partial x^3} \end{aligned}$$

Note that while the three vector fields:

$$\begin{aligned} \frac{\partial}{\partial x^m} : \mathrm{SL}(2, \mathbb{C}) &\rightarrow T \mathrm{SL}(2, \mathbb{C}) \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto \left(\frac{\partial}{\partial x^m}\right)_{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \end{aligned}$$

are not individually left-invariant, their linear combination with coefficients  $D^m_i$  is indeed left-invariant.

Hence we found :

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right) \in \mathfrak{sl}(2, \mathbb{C})$$

We now have to calculate the bracket (in  $\mathfrak{sl}(2, \mathbb{C})$ ) of every pair of these. We can also do them all at once, which is a good exercise in index gymnastics. We have:

$$\left[ j\left(\left(\frac{\partial}{\partial x^i}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right), j\left(\left(\frac{\partial}{\partial x^k}\right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\right) \right] = \left[ D^m_i \frac{\partial}{\partial x^m}, D^n_k \frac{\partial}{\partial x^n} \right]$$

Letting this act on an arbitrary  $f \in \mathcal{C}^\infty(\text{SL}(2, \mathbb{C}))$ , by definition:

$$\left[ D^m_i \frac{\partial}{\partial x^m}, D^n_k \frac{\partial}{\partial x^n} \right] (f) := D^m_i \frac{\partial}{\partial x^m} \left( D^n_k \frac{\partial}{\partial x^n} (f) \right) - D^n_k \frac{\partial}{\partial x^n} \left( D^m_i \frac{\partial}{\partial x^m} (f) \right)$$

The first term gives:

$$\begin{aligned} D^m_i \frac{\partial}{\partial x^m} \left( D^n_k \frac{\partial}{\partial x^n} (f) \right) &= D^m_i \frac{\partial}{\partial x^m} (D^n_k \partial_n (f \circ x^{-1}) \circ x) \\ &= D^m_i \frac{\partial}{\partial x^m} (D^n_k) (\partial_n (f \circ x^{-1}) \circ x) + D^m_i D^n_k \frac{\partial}{\partial x^m} (\partial_n (f \circ x^{-1}) \circ x) \\ &= D^m_i \frac{\partial}{\partial x^m} (D^n_k) (\partial_n (f \circ x^{-1}) \circ x) + D^m_i D^n_k \partial_m (\partial_n (f \circ x^{-1}) \circ x \circ x^{-1}) \circ x \\ &= D^m_i \frac{\partial}{\partial x^m} (D^n_k) (\partial_n (f \circ x^{-1}) \circ x) + D^m_i D^n_k \partial_m \partial_n (f \circ x^{-1}) \circ x \end{aligned}$$

Similarly, for the second term:

$$D^n_k \frac{\partial}{\partial x^n} \left( D^m_i \frac{\partial}{\partial x^m} (f) \right) = D^n_k \frac{\partial}{\partial x^n} (D^m_i) (\partial_m (f \circ x^{-1}) \circ x) + D^n_k D^m_i \partial_n \partial_m (f \circ x^{-1}) \circ x$$

Hence, recalling that  $\partial_m \partial_n = \partial_n \partial_m$  by Schwarz's theorem, we have:

$$\begin{aligned} \left[ D^m_i \frac{\partial}{\partial x^m}, D^n_k \frac{\partial}{\partial x^n} \right] (f) &= D^m_i \frac{\partial}{\partial x^m} (D^n_k) (\partial_n (f \circ x^{-1}) \circ x) + \cancel{D^m_i D^n_k \partial_m \partial_n (f \circ x^{-1}) \circ x} \\ &\quad - D^n_k \frac{\partial}{\partial x^n} (D^m_i) (\partial_m (f \circ x^{-1}) \circ x) - \cancel{D^n_k D^m_i \partial_n \partial_m (f \circ x^{-1}) \circ x} \\ &= \left( D^m_i \frac{\partial}{\partial x^m} (D^n_k) - D^n_k \frac{\partial}{\partial x^m} (D^m_i) \right) \partial_n (f \circ x^{-1}) \circ x \\ &= \left( D^m_i \frac{\partial}{\partial x^m} (D^n_k) - D^n_k \frac{\partial}{\partial x^m} (D^m_i) \right) \frac{\partial}{\partial x^n} (f) \end{aligned}$$

where we relabelled some dummy indices.

Since the  $f \in \mathcal{C}^\infty(\text{SL}(2, \mathbb{C}))$  was arbitrary:

$$\left[ D^m_i \frac{\partial}{\partial x^m}, D^n_k \frac{\partial}{\partial x^n} \right] = \left( D^m_i \frac{\partial}{\partial x^m} (D^n_k) - D^n_k \frac{\partial}{\partial x^m} (D^m_i) \right) \frac{\partial}{\partial x^n}$$

We can now evaluate this explicitly for each basis vector.

For  $i = 1$  and  $k = 2$ , we have:

$$\begin{aligned} \left[ D^m_1 \frac{\partial}{\partial x^m}, D^n_2 \frac{\partial}{\partial x^n} \right] &= \left( \cancel{D^m_1 \frac{\partial}{\partial x^m} (D^1_2)} - D^m_2 \frac{\partial}{\partial x^m} (D^1_1) \right) \frac{\partial}{\partial x^1} \\ &\quad + \left( D^m_1 \frac{\partial}{\partial x^m} (D^2_2) - D^m_2 \frac{\partial}{\partial x^m} (D^2_1) \right) \frac{\partial}{\partial x^2} \\ &\quad + \left( \cancel{D^m_1 \frac{\partial}{\partial x^m} (D^3_2)} - D^m_2 \frac{\partial}{\partial x^m} (D^3_1) \right) \frac{\partial}{\partial x^3} \\ &= -D^1_2 \frac{\partial}{\partial x^1} + (D^1_1 + D^2_2) \frac{\partial}{\partial x^2} - D^3_2 \frac{\partial}{\partial x^3} \\ &= 2x^1 \frac{\partial}{\partial x^2} \end{aligned}$$

Similarly, we compute:

$$\begin{aligned}
\left[ D^m_1 \frac{\partial}{\partial x^m}, D^n_3 \frac{\partial}{\partial x^n} \right] &= \left( D^m_1 \frac{\partial}{\partial x^m} (D^1_3) - D^m_3 \frac{\partial}{\partial x^m} (D^1_1) \right) \frac{\partial}{\partial x^1} \\
&\quad + \left( \cancel{D^m_1 \frac{\partial}{\partial x^m} (D^2_3)} - D^m_3 \frac{\partial}{\partial x^m} (D^2_1) \right) \frac{\partial}{\partial x^2} \\
&\quad + \left( D^m_1 \frac{\partial}{\partial x^m} (D^3_3) - D^m_3 \frac{\partial}{\partial x^m} (D^3_1) \right) \frac{\partial}{\partial x^3} \\
&= -2x^2 \frac{\partial}{\partial x^1} - 2\left(\frac{1+x^2x^3}{x^1}\right) \frac{\partial}{\partial x^3}
\end{aligned}$$

and:

$$\begin{aligned}
\left[ D^m_2 \frac{\partial}{\partial x^m}, D^n_3 \frac{\partial}{\partial x^n} \right] &= \left( D^m_2 \frac{\partial}{\partial x^m} (D^1_3) - \cancel{D^m_3 \frac{\partial}{\partial x^m} (D^1_2)} \right) \frac{\partial}{\partial x^1} \\
&\quad + \left( \cancel{D^m_2 \frac{\partial}{\partial x^m} (D^2_3)} - D^m_3 \frac{\partial}{\partial x^m} (D^2_2) \right) \frac{\partial}{\partial x^2} \\
&\quad + \left( D^m_2 \frac{\partial}{\partial x^m} (D^3_3) - \cancel{D^m_3 \frac{\partial}{\partial x^m} (D^3_2)} \right) \frac{\partial}{\partial x^3} \\
&= (D^2_1 - D^1_3) \frac{\partial}{\partial x^1} + D^2_3 \frac{\partial}{\partial x^2} - D^3_2 \frac{\partial}{\partial x^3} \\
&= x^1 \frac{\partial}{\partial x^1} - x^2 \frac{\partial}{\partial x^2} + x^3 \frac{\partial}{\partial x^3}
\end{aligned}$$

where the deleted parts come from the zeros in  $D_i^m$ .

So now we know the commutator in  $\mathfrak{sl}(2, \mathbb{C})$ :

$$[j(X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}), j(Y_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}})]_{\mathfrak{sl}(2, \mathbb{C})}$$

The only thing that remains in order to fully obtain the commutator back in  $T_e G$  is simply to apply  $j^{-1}$ , since remember that:

$$[X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, Y_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}]_{T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \mathrm{SL}(2, \mathbb{C})} := j^{-1}([j(X_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}), j(Y_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}})]_{\mathfrak{sl}(2, \mathbb{C})})$$

Hence, by applying  $j^{-1}$ , which is just evaluation at the identity, to these vector fields, we finally see that the induced Lie bracket on  $T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \mathrm{SL}(2, \mathbb{C})$  satisfies:

$$\begin{aligned}
\left[ \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right] &= 2 \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \\
\left[ \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right] &= -2 \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \\
\left[ \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right] &= \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}
\end{aligned}$$

which can be rewritten as:

$$\begin{aligned} \left[ \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right] &= 0 \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} + 2 \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} + 0 \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \\ \left[ \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right] &= 0 \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} + 0 \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} + (-2) \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \\ \left[ \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}, \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \right] &= 1 \left( \frac{\partial}{\partial x^1} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} + 0 \left( \frac{\partial}{\partial x^2} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} + 0 \left( \frac{\partial}{\partial x^3} \right)_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \end{aligned}$$

Hence, the structure constants of  $T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \text{SL}(2, \mathbb{C})$  with respect to the co-ordinate basis are:

$$C^2_{12} = 2, \quad C^3_{13} = -2, \quad C^1_{23} = 1$$

with all other being either zero or related to these by anti-symmetry.

However remember from the Lie algebra part of the notes that two Lie algebras  $L$  and  $L'$  are isomorphic if, and only if, there exists a basis of  $L$  and a basis of  $L'$  in which the structure constants of  $L$  and  $L'$  are the same.

Since we have already proved that  $T_e G \cong_{\text{Lie alg}} \mathcal{L}(G)$  for any Lie group  $G$ , we can deduce the existence of a basis  $\{e_1, e_2, e_3\}$  of  $\mathfrak{sl}(2, \mathbb{C})$  with respect to which the structure constants are those listed above.

*Remark 7.2.* From now on we will be denoting the basis of  $T_e G$  as  $\{X_1, X_2, X_3\}$  since this is the common notation. However, do not be confused by thinking that  $\{X_1, X_2, X_3\}$  are tangent vectors at points 1, 2 and 3.

Hence, finally we induce from the basis and the structure constants of  $T_{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \text{SL}(2, \mathbb{C})$  a basis and structure constants of  $\mathfrak{sl}(2, \mathbb{C})$  as:

$$\begin{aligned} [X_1, X_2] &= 2X_2 \\ [X_1, X_3] &= -2X_3 \\ [X_2, X_3] &= X_1 \end{aligned}$$

*Remark 7.3.* This is quite familiar since it coincides with our familiar commutation relations that the matrices of  $\text{SL}(2, \mathbb{C})$  satisfy. However, we have to keep in mind that nothing has been mentioned regarding matrices.  $\{X_1, X_2, X_3\}$  are abstract vector fields and not matrices. However, in the next chapter we will see the connection between those two when we will talk about representation theory.

Now that we found the structure constants of  $\mathfrak{sl}(2, \mathbb{C})$  we can use them in order to classify it. We will follow the steps we explained in the corresponding Lie algebra chapter of the notes.

In the basis  $\{X_1, X_2, X_3\}$ , the Killing form of  $\mathfrak{sl}(2, \mathbb{C})$  has components:

$$\kappa_{ij} = C^m_{in} C^n_{jm}$$

with all indices ranging from 1 to 3.

As an example, for  $i = j = 1$  we have:

$$\begin{aligned} \kappa_{11} &= C^m_{1n} C^n_{1m} \\ &= \cancel{C^1_{1n} C^n_{11}} + C^2_{1n} C^n_{12} + C^3_{1n} C^n_{13} \\ &= C^2_{12} C^2_{12} + C^3_{13} C^3_{13} \\ &= 8 \end{aligned}$$

Similarly we can compute all the other components. (Since  $\kappa$  is symmetric, we only need to determine  $\kappa_{ij}$  for  $i \leq j$ ).

After computing all the components we can write them in a  $3 \times 3$  array, i.e:

$$[\kappa_{ij}] = \begin{pmatrix} 8 & 0 & 0 \\ 0 & -8 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

which is just shorthand for :

$$\kappa(X_1, X_1) = 8, \quad \kappa(X_2, X_2) = -8, \quad \kappa(X_3, X_3) = 8, \quad \kappa(X_i, X_j) = 0, \quad \forall i \neq j$$

**Proposition 7.2.** *The Lie algebra  $\mathfrak{sl}(2, \mathbb{C})$  is semi-simple.*

*Proof.*

Since the diagonal entries of  $\kappa$  are all non-zero, the Killing form is non-degenerate. By Cartan's criterion, this implies that  $\mathfrak{sl}(2, \mathbb{C})$  is semi-simple.  $\square$

*Remark 7.4.* There is one more thing that can be read off from the components of  $\kappa$ , namely, that it is an *indefinite* form, i.e. the sign of  $\kappa(X, X)$  can be positive or negative depending on which  $X \in \mathfrak{sl}(2, \mathbb{C})$  we pick.

A result from Lie theory states that the Killing form on the Lie algebra of a compact Lie group is always negative semi-definite, i.e.  $\kappa(X, X)$  is always negative or zero, for all  $X$  in the Lie algebra. Hence, we can conclude that  $SL(2, \mathbb{C})$  is not a compact Lie group.

**Proposition 7.3.** *The Lie algebra  $\mathfrak{sl}(2, \mathbb{C})$  is simple.*

*Proof.*

Recall that a Lie algebra is said to be simple if it contains no non-trivial ideals, and that an ideal  $I$  of a Lie algebra  $L$  is a Lie subalgebra of  $L$  such that:

$$\forall x \in I : \forall y \in L : [x, y] \in I$$

Consider the ideal of  $\mathfrak{sl}(2, \mathbb{C})$ :

$$I := \{\alpha X_1 + \beta X_2 + \gamma X_3 \mid \alpha, \beta, \gamma \text{ restricted so that } I \text{ is an ideal}\}$$

Since the bracket is bilinear, it suffices to check the result of bracketing an arbitrary element of  $I$  with each of the basis vectors of  $\mathfrak{sl}(2, \mathbb{C})$ . We find:

$$[\alpha X_1 + \beta X_2 + \gamma X_3, X_1] = -2\beta X_1 + 2\gamma X_3$$

$$[\alpha X_1 + \beta X_2 + \gamma X_3, X_2] = 2\alpha X_2 - \gamma X_1$$

$$[\alpha X_1 + \beta X_2 + \gamma X_3, X_3] = -2\alpha X_3 + \beta X_1$$

We need to choose  $\alpha, \beta, \gamma$  so that the results always land back in  $I$ . Of course, we can choose  $\alpha, \beta, \gamma \in \mathbb{C}$  and  $\alpha = \beta = \gamma = 0$ , which correspond respectively to the trivial ideals  $\mathfrak{sl}(2, \mathbb{C})$  and  $0$ . If none of  $\alpha, \beta, \gamma$  is zero, then you can check that the right hand sides above are linearly independent, so that  $I$  contains three linearly independent vectors. Since the only  $n$ -dimensional subspace of an  $n$ -dimensional vector space is the vector space itself, we have  $I = L$ . Thus, we are left with the following cases:

- i) If  $\alpha = 0$ , then  $I \subseteq \text{span}_{\mathbb{C}}(\{X_2, X_3\})$  and hence we must have  $\beta = \gamma = 0$  as well.
- ii) If  $\beta = 0$ , then  $I \subseteq \text{span}_{\mathbb{C}}(\{X_1, X_3\})$ , hence we must have  $\alpha = 0$ , so that in fact  $I \subseteq \text{span}_{\mathbb{C}}(\{X_3\})$ , and hence  $\gamma = 0$  as well.
- iii) If  $\gamma = 0$ , then  $I \subseteq \text{span}_{\mathbb{C}}(\{X_1, X_2\})$ , hence we must have  $\alpha = 0$ , so that in fact  $I \subseteq \text{span}_{\mathbb{C}}(\{X_2\})$ , and hence  $\beta = 0$  as well.

In all cases, we have  $I = 0$ . Therefore, there are no non-trivial ideals of  $\mathfrak{sl}(2, \mathbb{C})$ .  $\square$

By observing the bracket relations of the basis elements of  $\mathfrak{sl}(2, \mathbb{C})$ , we can see that:

$$H := \text{span}_{\mathbb{C}}(\{X_1\})$$

is a Cartan subalgebra of  $\mathfrak{sl}(2, \mathbb{C})$ . Indeed, for any  $h \in H$ , there exists a  $\xi \in \mathbb{C}$  such that  $h = \xi X_1$ , and hence we have:

$$\begin{aligned}\mathrm{ad}(h)X_2 &= \xi[X_1, X_2] = 2\xi X_2 \\ \mathrm{ad}(h)X_3 &= \xi[X_1, X_3] = -2\xi X_3\end{aligned}$$

Recall that in the section on Lie algebras, we re-interpreted these eigenvalue equations in terms of functionals  $\lambda_2, \lambda_3 \in H^*$ :

$$\begin{array}{ll}\lambda_2: & H \xrightarrow{\sim} \mathbb{C} \\ & \xi X_1 \mapsto 2\xi\end{array} \qquad \begin{array}{ll}\lambda_3: & H \xrightarrow{\sim} \mathbb{C} \\ & \xi X_1 \mapsto -2\xi\end{array}$$

whereby:

$$\begin{aligned}\mathrm{ad}(h)X_2 &= \lambda_2(h)X_2 \\ \mathrm{ad}(h)X_3 &= \lambda_3(h)X_3\end{aligned}$$

Then,  $\lambda_2$  and  $\lambda_3$  are called the roots of  $\mathfrak{sl}(2, \mathbb{C})$ , so that the root set is  $\Phi = \{\lambda_2, \lambda_3\}$ .

Of course, we are mainly interested in a subset  $\Pi \subset \Phi$  of fundamental roots, which satisfies:

- i)  $\Pi$  is a linearly independent subset of  $H^*$ .
- ii) For any  $\lambda \in \Phi$ , we have  $\lambda \in \mathrm{span}_{\mathbb{N}}(\Pi)$ .

We can choose  $\Pi := \{\lambda_2\}$ , even though  $\Pi := \{\lambda_3\}$  would work just as well.

Since  $|\Pi| = 1$ , the Weyl group is generated by the single Weyl transformation:

$$\begin{aligned}s_{\lambda_2}: H_{\mathbb{R}}^* &\rightarrow H_{\mathbb{R}}^* \\ \mu &\mapsto \mu - 2 \frac{\kappa^*(\lambda_2, \mu)}{\kappa^*(\lambda_2, \lambda_2)} \lambda_2\end{aligned}$$

Recall that we can recover the entire root set  $\Phi$  by acting on the fundamental roots with Weyl transformations. Indeed, we have:

$$s_{\lambda_2}(\lambda_2) = \lambda_2 - 2 \frac{\kappa^*(\lambda_2, \lambda_2)}{\kappa^*(\lambda_2, \lambda_2)} \lambda_2 = \lambda_2 - 2\lambda_2 = -\lambda_2 = \lambda_3$$

as expected.

Since there is only one fundamental root, the Cartan matrix is actually just a  $1 \times 1$  matrix. Its only entry is a diagonal entry, and since  $\mathfrak{sl}(2, \mathbb{C})$  is simple, we have:

$$C = (2).$$

The Dynkin diagram of  $\mathfrak{sl}(2, \mathbb{C})$  is simply:

$$\bigcirc$$

Hence, with reference to the Cartan classification, we have  $\mathfrak{sl}(2, \mathbb{C}) = A_1$ .

## 7.5 Representations Of Lie Groups

Lie groups (and the associated Lie algebras) are used in physics mostly in terms of what are called representations. Very often they are even defined in terms of their concrete representations. In this chapter we took a more abstract approach by defining a Lie group as a smooth manifold with a compatible group structure, and its associated Lie algebra as the space of left-invariant vector fields, which we then showed to be isomorphic to the tangent space at the identity. However, as we defined the notion of representations for Lie algebras in the chapter of Lie algebras, we can define a similar notion of representations of Lie groups.

Given a vector space  $V$ , recall that the subset of  $\text{Aut}(V)$  consisting of the invertible endomorphisms and denoted:

$$\text{GL}(V) \equiv \text{Aut}(V) := \{\phi \in \text{End}(V) \mid \det \phi \neq 0\}$$

forms a group under composition, called the automorphism group (or general linear group) of  $V$ . Moreover, if  $V$  is a finite-dimensional  $K$ -vector space, then  $V \cong_{\text{vec}} K^{\dim V}$  and hence the group  $\text{GL}(V)$  can be given the structure of a Lie group via:

$$\text{GL}(V) \cong_{\text{Lie grp}} \text{GL}(K^{\dim V}) := \text{GL}(\dim V, K)$$

This is, of course, if we have established a topology and a differentiable structure on  $K^d$ , as is the case for  $\mathbb{R}^d$  and  $\mathbb{C}^d$ .

**Definition 7.9** (Representation Of Lie Group). *A **representation** of a Lie group  $(G, \bullet)$  is a Lie group homomorphism:*

$$R: G \rightarrow \text{GL}(V)$$

for some finite-dimensional vector space  $V$ .

Recall that  $R: G \rightarrow \text{GL}(V)$  is a Lie group homomorphism if it is smooth and:

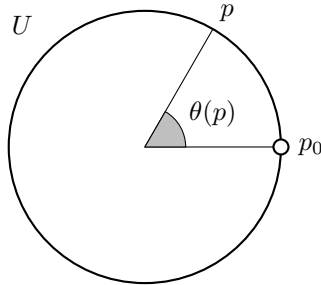
$$\forall g_1, g_2 \in G : R(g_1 \bullet g_2) = R(g_1) \circ R(g_2)$$

Note that, as is the case with any group homomorphism, we have:

$$R(e) = \text{id}_V \quad \text{and} \quad R(g^{-1}) = R(g)^{-1}$$

*Example 7.2.*

Consider the Lie group  $\text{SO}(2, \mathbb{R})$ . As a smooth manifold,  $\text{SO}(2, \mathbb{R})$  is isomorphic to the circle  $S^1$ . Let  $U = S^1 \setminus \{p_0\}$ , where  $p$  is any point of  $S^1$ , so that we can define a chart  $\theta: U \rightarrow [0, 2\pi) \subseteq \mathbb{R}$  on  $S^1$  by mapping each point in  $U$  to an “angle” in  $[0, 2\pi)$ .



The operation:

$$p_1 \bullet p_2 := (\theta(p_1) + \theta(p_2)) \mod 2\pi$$

endows  $S^1 \cong_{\text{diff}} \text{SO}(2, \mathbb{R})$  with the structure of a Lie group. Then, a representation of  $\text{SO}(2, \mathbb{R})$  is given by:

$$R: \text{SO}(2, \mathbb{R}) \rightarrow \text{GL}(\mathbb{R}^2)$$

$$p \mapsto \begin{pmatrix} \cos \theta(p) & \sin \theta(p) \\ -\sin \theta(p) & \cos \theta(p) \end{pmatrix}$$

Indeed, the addition formula for sine and cosine imply that:

$$R(p_1 \bullet p_2) = R(p_1) \circ R(p_2)$$

*Example 7.3.*

Let  $G$  be a Lie group (we suppress the  $\bullet$  in this example). For each  $g \in G$ , define the Adjoint map:

$$\begin{aligned} \text{Ad}_g : G &\rightarrow G \\ h &\mapsto ghg^{-1} \end{aligned}$$

Note the capital “A” to distinguish this from the adjoint map on Lie algebras. Since  $\text{Ad}_g$  is a composition of the Lie group multiplication and inverse map, it is a smooth map. Moreover, we have:

$$\text{Ad}_g(e) = geg^{-1} = gg^{-1} = e$$

Hence, the push-forward of  $\text{Ad}_g$  at the identity is the map:

$$(\text{Ad}_{g*})_e : T_e G \xrightarrow{\sim} T_{\text{Ad}_g(e)} G = T_e G$$

Thus, we have  $\text{Ad}_g \in \text{End}(T_e G)$ . In fact, you can check that:

$$(\text{Ad}_{g^{-1}*})_e \circ (\text{Ad}_{g*})_e = (\text{Ad}_{g*})_e \circ (\text{Ad}_{g^{-1}*})_e = \text{id}_{T_e G}$$

and hence we have, in particular,  $\text{Ad}_g \in \text{GL}(T_e G) \cong_{\text{Lie grp}} \text{GL}(\mathcal{L}(G))$

We can therefore construct a map:

$$\begin{aligned} \text{Ad} : G &\rightarrow \text{GL}(T_e G) \\ g &\mapsto \text{Ad}_{g*} \end{aligned}$$

which, as you can check, is a representation of  $G$  on its Lie algebra.

*Remark 7.5.* Since a representation  $R$  of a Lie group  $G$  is required to be smooth, we can always consider its differential or push-forward at the identity:

$$(R_*)_e : T_e G \xrightarrow{\sim} T_{\text{id}_V} \text{GL}(V)$$

Since for any  $A, B \in T_e G$  we have:

$$(R_*)_e[A, B] = [(R_*)_e A, (R_*)_e B]$$

the map  $(R_*)_e$  is a representation of the Lie algebra of  $G$  on the vector space  $\text{GL}(V)$ . In fact, in the previous example we have:

$$(\text{Ad}_*)_e = \text{ad}$$

where  $\text{ad}$  is the adjoint representation of  $T_e G$ .

## 7.6 Reconstruction Of A Lie group From Its Lie Algebra

We have seen in detail how to construct a Lie algebra from a given Lie group. We would now like to consider the inverse question, i.e. whether, given a Lie algebra, we can construct a Lie group whose associated Lie algebra is the given one and, if this is the case, whether this correspondence is bijective.

We will find that the answer to the first question is affirmative. Given a Lie algebra, we will construct a Lie group via something called the exponential map. However, we can already answer in the negative the second question, since there are examples of Lie groups which are not isomorphic but give rise to the same Lie algebra. Hence, the correspondence between Lie groups and Lie algebras cannot be bijective.

Recall that in the chapter of topological manifolds we defined the integral curve  $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$  on a smooth manifold  $M$  as:

$$\forall \lambda \in (-\varepsilon, \varepsilon) : X_{\gamma, \gamma(\lambda)} = X(\gamma(\lambda))$$

and based on that we defined a complete vector field as the vector field where the integral  $(-\varepsilon, \varepsilon)$  of  $\gamma$  can be extended to the whole  $\mathbb{R}$  which from there followed that for any  $p$  on  $M$  it is  $X(p) = X_p$ .

Finally we gave a theorem that stated that on a compact manifold, every vector field is complete, and

given that we only work with compact manifolds we have been using the fact that  $X(p) = X_p$ .

However, on a Lie group, even if non-compact, there are always complete vector fields.

**Theorem 7.3.** *Every left-invariant vector field on a Lie group is complete.*

The integral curves of left-invariant vector fields are crucial in the construction of the map that allows us to go from a Lie algebra to a Lie group.

### 7.6.1 The Exponential Map

Let  $G$  be a Lie group. Recall that given any  $X_e \in T_e G$ , we can define the uniquely determined left-invariant vector field  $X := j(A)$  via the isomorphism  $j: T_e G \xrightarrow{\sim} \mathcal{L}(G)$  as:

$$X(g) = j(X_e)(g) := (\ell_{g*})_e(X_e)$$

Then let  $\gamma: \mathbb{R} \rightarrow G$  be the maximal integral curve of  $X$  with  $\gamma(0) = e \in G$ .

**Definition 7.10** (Exponential Map). *Let  $G$  be a Lie group. The **exponential map** is defined as:*

$$\begin{aligned} \exp: T_e G &\rightarrow G \\ X_e &\mapsto \exp(X_e) := \gamma(1) \end{aligned}$$

**Theorem 7.4.** *i) The map  $\exp$  is smooth and a local diffeomorphism around  $0 \in T_e G$ , i.e. there exists an open  $V \subseteq T_e G$  containing 0 such that the restriction:*

$$\exp|_V: V \rightarrow \exp(V) \subseteq G$$

*is bijective and both  $\exp|_V$  and  $(\exp|_V)^{-1}$  are smooth.*

*ii) If  $G$  is compact, then  $\exp$  is surjective.*

The first part of the theorem says that we can recover a neighbourhood of the identity of  $G$  from a neighbourhood of the identity of  $T_e G$ .

Since  $T_e G$  is a vector space, it is non-compact (intuitively, it extends infinitely far away in every direction) and hence, if  $G$  is compact,  $\exp$  cannot be injective. This is because, by the second part of the theorem, it would then be a diffeomorphism  $T_e G \rightarrow G$ . But as  $G$  is compact and  $T_e G$  is not, they are not diffeomorphic.

**Proposition 7.4.** *Let  $G$  be a Lie group. The image of  $\exp: T_e G \rightarrow G$  is the connected component of  $G$  containing the identity.*

Therefore, if  $G$  itself is connected, then  $\exp$  is again surjective. Note that, in general, there is no relation between connected and compact topological spaces, i.e. a topological space can be either, both, or neither.

*Example 7.4.*

Let  $B: V \times V$  be a pseudo inner product on  $V$ . Then:

$$\mathcal{O}(V) := \{\phi \in \text{GL}(V) \mid \forall v, w \in V : B(\phi(v), \phi(w)) = B(v, w)\}$$

is called the *orthogonal group* of  $V$  with respect to  $B$ . Of course, if  $B$  or the base field of  $V$  need to be emphasised, they can be included in the notation. Every  $\phi \in \mathcal{O}(V)$  has determinant 1 or  $-1$ . Since the determinant is multiplicative, we have a subgroup:

$$\text{SO}(V) := \{\phi \in \mathcal{O}(V) \mid \det \phi = 1\}$$

These are, in fact, Lie subgroups of  $\text{GL}(V)$ . The Lie group  $\text{SO}(V)$  is connected while:

$$\mathcal{O}(V) = \text{SO}(V) \cup \{\phi \in \mathcal{O}(V) \mid \det \phi = -1\}$$

is disconnected. Since  $\text{SO}(V)$  contains  $\text{id}_V$ , we have:

$$\mathfrak{so}(V) := T_{\text{id}_V} \text{SO}(V) = T_{\text{id}_V} \mathcal{O}(V) =: \mathfrak{o}(V)$$

and:

$$\exp(\mathfrak{so}(V)) = \exp(\mathfrak{o}(V)) = \mathrm{SO}(V)$$

*Example 7.5.*

Choosing a basis  $X_1, \dots, X_{\dim G}$  of  $T_e G$  often provides a convenient co-ordinatisation of  $G$  near  $e$ . Consider, for example, the Lorentz group:

$$\mathrm{O}(3, 1) \equiv \mathrm{O}(\mathbb{R}^4) = \{\Lambda \in \mathrm{GL}(\mathbb{R}^4) \mid \forall x, y \in \mathbb{R}^4 : B(\Lambda(x), \Lambda(y)) = B(x, y)\},$$

where  $B(x, y) := \eta_{\mu\nu} x^\mu y^\nu$ , with  $0 \leq \mu, \nu \leq 3$  and:

$$[\eta^{\mu\nu}] = [\eta_{\mu\nu}] := \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The Lorentz group  $\mathrm{O}(3, 1)$  is 6-dimensional, hence so is the Lorentz algebra  $\mathfrak{o}(3, 1)$ . We could simply denote the basis of  $\mathfrak{o}(3, 1)$  as  $\{X_i \mid i = 1, \dots, 6\}$ , but it is used to denote it as  $\{M^{\mu\nu} \mid 0 \leq \mu, \nu \leq 3\}$  and require that the indices  $\mu, \nu$  be anti-symmetric, i.e:

$$M^{\mu\nu} = -M^{\nu\mu}$$

Then  $M^{\mu\nu} = 0$  when  $\rho = \sigma$ , and the set  $\{M^{\mu\nu} \mid 0 \leq \mu, \nu \leq 3\}$ , while technically not linearly independent, contains the 6 independent elements that we want to consider as a basis. These basis elements satisfy the following bracket relation:

$$[M^{\mu\nu}, M^{\rho\sigma}] = \eta^{\nu\sigma} M^{\mu\rho} + \eta^{\mu\rho} M^{\nu\sigma} - \eta^{\nu\rho} M^{\mu\sigma} - \eta^{\mu\sigma} M^{\nu\rho}$$

Any element  $\lambda \in \mathfrak{o}(3, 1)$  can be expressed as linear combination of the  $M^{\mu\nu}$ :

$$\lambda = \frac{1}{2} \omega_{\mu\nu} M^{\mu\nu}$$

where the indices on the coefficients  $\omega_{\mu\nu}$  are also anti-symmetric, and the factor of  $\frac{1}{2}$  ensures that the sum over all  $\mu, \nu$  counts each anti-symmetric pair only once. Then, we can recover part of the Lie Group  $\mathrm{O}(3, 1)$  by the use of the exponential map:

$$\Lambda = \exp(\lambda) = \exp(\frac{1}{2} \omega_{\mu\nu} M^{\mu\nu}) \in \mathrm{O}(3, 1)$$

The subgroup of  $\mathrm{O}(3, 1)$  consisting of the the space-orientation preserving Lorentz transformations, or *proper* Lorentz transformations, is denoted by  $\mathrm{SO}(3, 1)$ . The subgroup consisting of the time-orientation preserving, or *orthochronous*, Lorentz transformations is denoted by  $\mathrm{O}^+(3, 1)$ . The Lie group  $\mathrm{O}(3, 1)$  is disconnected. Its four connected components are:

- i)  $\mathrm{SO}^+(3, 1) := \mathrm{SO}(3, 1) \cap \mathrm{O}^+(3, 1)$ , also called the *restricted Lorentz group*, consisting of the proper orthochronous Lorentz transformations.
- ii)  $\mathrm{SO}(3, 1) \setminus \mathrm{O}^+(3, 1)$ , the proper non-orthochronous transformations.
- iii)  $\mathrm{O}^+(3, 1) \setminus \mathrm{SO}(3, 1)$ , the improper orthochronous transformations.
- iv)  $\mathrm{O}(3, 1) \setminus (\mathrm{SO}(3, 1) \cup \mathrm{O}^+(3, 1))$ , the improper non-orthochronous transformations.

Since  $\mathrm{id}_{\mathbb{R}^4} \in \mathrm{SO}^+(3, 1)$ , we have  $\exp(\mathfrak{o}(3, 1)) = \mathrm{SO}^+(3, 1)$ . Then  $\{M^{\mu\nu}\}$  provides a nice co-ordinatisation of  $\mathrm{SO}^+(3, 1)$  since, if we choose:

$$[\omega_{\mu\nu}] = \begin{pmatrix} 0 & \psi_1 & \psi_2 & \psi_3 \\ -\psi_1 & 0 & \varphi_3 & -\varphi_2 \\ -\psi_2 & -\varphi_3 & 0 & \varphi_1 \\ -\psi_3 & \varphi_2 & -\varphi_1 & 0 \end{pmatrix}$$

then the Lorentz transformation  $\exp(\frac{1}{2} \omega_{\mu\nu} M^{\mu\nu}) \in \mathrm{SO}^+(3, 1)$  corresponds to a boost in the  $(\psi_1, \psi_2, \psi_3)$  direction and a space rotation by  $(\varphi_1, \varphi_2, \varphi_3)$ . Indeed, in physics one often thinks of the Lie group  $\mathrm{SO}^+(3, 1)$  as being generated by  $\{M^{\mu\nu}\}$ .

A representation  $\rho: T_{\text{id}_{\mathbb{R}^4}}\text{SO}^+(3,1) \xrightarrow{\sim} \text{End}(\mathbb{R}^4)$  is given by:

$$\rho(M^{\mu\nu})^a{}_b := \eta^{\nu a} \delta_b^\mu - \eta^{\mu a} \delta_b^\nu$$

which is probably how you have seen the  $M^{\mu\nu}$  themselves defined in some previous course on relativity theory. Using this representation, we get a corresponding representation:

$$R: \text{SO}^+(3,1) \rightarrow \text{GL}(\mathbb{R}^4)$$

via the exponential map by defining:

$$R(\Lambda) = \exp(\frac{1}{2}\omega_{\mu\nu}\rho(M^{\mu\nu}))$$

Then, the map  $\exp$  becomes the usual exponential (series) of matrices.

**Definition 7.11** (One-Parameter Subgroup). A **one-parameter subgroup** of a Lie group  $G$  is a Lie group homomorphism:

$$\xi: \mathbb{R} \rightarrow G$$

with  $\mathbb{R}$  understood as a Lie group under ordinary addition.

*Example 7.6.*

Let  $M$  be a smooth manifold and let  $X \in \Gamma(TM)$  be a complete vector field. The *flow* of  $X$  is the smooth map:

$$\begin{aligned} \Theta: \mathbb{R} \times M &\rightarrow M \\ (\lambda, p) &\mapsto \Theta_\lambda(p) := \gamma_p(\lambda) \end{aligned}$$

where  $\lambda_p$  is the maximal integral curve of  $X$  through  $p$ . For a fixed  $p$ , we have:

$$\Theta_0 = \text{id}_M, \quad \Theta_{\lambda_1} \circ \Theta_{\lambda_2} = \Theta_{\lambda_1 + \lambda_2}, \quad \Theta_{-\lambda} = \Theta_\lambda^{-1}$$

For each  $\lambda \in \mathbb{R}$ , the map  $\Theta_\lambda$  is a diffeomorphism  $M \rightarrow M$ . Denoting by  $\text{Diff}(M)$  the group (under composition) of the diffeomorphisms  $M \rightarrow M$ , we have that the map:

$$\begin{aligned} \xi: \mathbb{R} &\rightarrow \text{Diff}(M) \\ \lambda &\mapsto \Theta_\lambda \end{aligned}$$

is a one-parameter subgroup of  $\text{Diff}(M)$ .

**Theorem 7.5.** Let  $G$  be a Lie group.

i) Let  $X_e \in T_e G$ . The map:

$$\begin{aligned} \xi^{X_e}: \mathbb{R} &\rightarrow G \\ \lambda &\mapsto \xi^{X_e}(\lambda) := \exp(\lambda X_e) \end{aligned}$$

is a one-parameter subgroup.

ii) Every one-parameter subgroup of  $G$  has the form  $\xi^{X_e}$  for some  $X_e \in T_e G$ .

Therefore, the Lie algebra allows us to study all the one-parameter subgroups of the Lie group.

**Theorem 7.6.** Let  $G$  and  $H$  be Lie groups and let  $\phi: G \rightarrow H$  be a Lie group homomorphism. Then, for all  $X_e \in T_e G$ , we have:

$$\phi(\exp(X_e)) = \exp((\phi_*)_e X_e)$$

Equivalently, the following diagram commutes:

$$\begin{array}{ccc} T_{e_G} G & \xrightarrow{(\phi_*)_e} & T_{e_H} H \\ \exp \downarrow & & \downarrow \exp \\ G & \xrightarrow{\phi} & H \end{array}$$

In particular, for  $\phi \equiv \text{Ad}_g: G \rightarrow G$ , we have:

$$\text{Ad}_g(\exp(X_e)) = \exp((\text{Ad}_{g*})_e X_e)$$

## 7.7 Lie Group Actions On A Manifold

**Definition 7.12** (Left Lie Group Action). *Let  $(G, \bullet)$  be a Lie group and let  $M$  be a smooth manifold. A smooth map:*

$$\begin{aligned} \triangleright: G \times M &\rightarrow M \\ (g, p) &\mapsto g \triangleright p \end{aligned}$$

*satisfying:*

$$i) \quad \forall p \in M : e \triangleright p = p.$$

$$ii) \quad \forall g_1, g_2 \in G : \forall p \in M : (g_1 \bullet g_2) \triangleright p = g_1 \triangleright (g_2 \triangleright p).$$

*is called a **left Lie group action**, or **left  $G$ -action**, on  $M$ .*

**Definition 7.13** (Left  $G$ -Manifold). *A manifold equipped with a left  $G$ -action is called a **left  $G$ -manifold**.*

*Remark 7.6.* Note that in the above definition, the smooth structures on  $G$  and  $M$  were only used in the requirement that  $\triangleright$  be smooth. By dropping this condition, we obtain the usual definition of a group action on a set. Some of the definitions that we will soon give for Lie groups and smooth manifolds, such as those of orbits and stabilisers, also have clear analogues to the case of bare groups and sets.

*Example 7.7.*

Let  $G$  be a Lie group and let  $R: G \rightarrow \text{GL}(V)$  be a representation of  $G$  on a vector space  $V$ . Define a map:

$$\begin{aligned} \triangleright: G \times V &\rightarrow V \\ (g, v) &\mapsto g \triangleright v := R(g)v \end{aligned}$$

We easily check that  $e \triangleright v := R(e)v = \text{id}_V v = v$  and:

$$\begin{aligned} (g_1 \bullet g_2) \triangleright v &:= R(g_1 \bullet g_2)v \\ &= (R(g_1) \circ R(g_2))v \\ &= R(g_1)(R(g_2)v) \\ &= g_1 \triangleright (g_2 \triangleright v) \end{aligned}$$

for any  $v \in V$  and any  $g_1, g_2 \in G$ . Moreover, if we equip  $V$  with the usual smooth structure, the map  $\triangleright$  is smooth and hence a Lie group action on  $V$ . It follows that representations of Lie groups are just a special case of left Lie group actions. We can therefore think of left  $G$ -actions as generalised representations of  $G$  on some manifold.

**Definition 7.14** (Right Lie Group Action). *Similarly, a **right Lie group action** or **right  $G$ -action** on  $M$  is a smooth map:*

$$\begin{aligned} \triangleleft: M \times G &\rightarrow M \\ (p, g) &\mapsto p \triangleleft g \end{aligned}$$

*satisfying:*

$$i) \quad \forall p \in M : p \triangleleft g = p.$$

$$ii) \quad \forall g_1, g_2 \in G : \forall p \in M : p \triangleleft (g_1 \bullet g_2) = (p \triangleleft g_1) \triangleleft g_2.$$

**Proposition 7.5.** *Let  $\triangleright$  be a left  $G$ -action on  $M$ . Then:*

$$\begin{aligned} \triangleleft: M \times G &\rightarrow M \\ (p, g) &\mapsto p \triangleleft g := g^{-1} \triangleright p \end{aligned}$$

is a right  $G$ -action on  $M$ .

*Proof.*

First note that  $\triangleleft$  is smooth since it is a composition of  $\triangleright$  and the inverse map on  $G$ , which are both smooth. We have  $p \triangleleft e := e \triangleright p = p$  and:

$$\begin{aligned} p \triangleleft (g_1 \bullet g_2) &:= (g_1 \bullet g_2)^{-1} \triangleright p \\ &= (g_2^{-1} \bullet g_1^{-1}) \triangleright p \\ &= g_2^{-1} \triangleright (g_1^{-1} \triangleright p) \\ &= g_2^{-1} \triangleright (p \triangleleft g_1) \\ &= (p \triangleleft g_1) \triangleleft g_2 \end{aligned}$$

for all  $p \in M$  and  $g_1, g_2 \in G$ , and hence  $\triangleleft$  is a right  $G$ -action.  $\square$

*Remark 7.7.* Since for each  $g \in G$  we also have  $g^{-1} \in G$ , if we need *some* action of  $G$  on  $M$ , then a left action is just as good as a right action. Only later, within the context of principal and associated fibre bundles, we will attach separate “meanings” to left and right actions. Some of the next definitions and results will only be given in terms of left actions, but they obviously apply to right actions as well.

*Remark 7.8.* Recall that if we have a basis  $\{e_1, \dots, e_{\dim M}\}$  of  $T_p M$  and  $X^1, \dots, X^{\dim M}$  are the components of some  $X_p \in T_p M$  in this basis, then under a change of basis:

$$\tilde{e}_a = A^b_a e_b$$

we have  $X = \tilde{X}^a \tilde{e}_a$ , where:

$$\tilde{X}^a = (A^{-1})^a_b X^b$$

Once expressed in terms of principal and associated fibre bundles, we will see that the “recipe” of labelling the basis by lower indices and the vector components by upper indices, as well as their transformation law, can be understood as a right action of  $\text{GL}(\dim M, \mathbb{R})$  on the basis and a left action of the same  $\text{GL}(\dim M, \mathbb{R})$  on the components.

**Definition 7.15** ( $\rho$ -Equivariant Map). *Let  $G, H$  be Lie groups, let  $\rho: G \rightarrow H$  be a Lie group homomorphism and let:*

$$\begin{aligned} \triangleright: G \times M &\rightarrow M \\ \blacktriangleright: H \times N &\rightarrow N \end{aligned}$$

*be left actions of  $G$  and  $H$  on some smooth manifolds  $M$  and  $N$ , respectively. Then, a smooth map  $f: M \rightarrow N$  is said to be  $\rho$ -equivariant if the diagram:*

$$\begin{array}{ccc} G \times M & \xrightarrow{\rho \times f} & H \times N \\ \triangleright \downarrow & & \downarrow \blacktriangleright \\ M & \xrightarrow{f} & N \end{array}$$

*where  $(\rho \times f)(g, p) := (\rho(g), f(p)) \in H \times N$ , commutes. Equivalently:*

$$\forall g \in G : \forall p \in M : f(g \triangleright p) = \rho(g) \blacktriangleright f(p)$$

In other words, if  $\rho: G \rightarrow H$  is a Lie group homomorphism, then the  $\rho$ -equivariant maps are the “action-preserving” maps between the  $G$ -manifold  $M$  and the  $H$ -manifold  $N$ .

*Remark 7.9.* Note that by setting  $\rho = \text{id}_G$  or  $f = \text{id}_M$ , the notion of  $f$  being  $\rho$ -equivariant reduces to what we might call a homomorphism of  $G$ -manifolds in the former case, and a homomorphism of left actions on  $M$  in the latter.

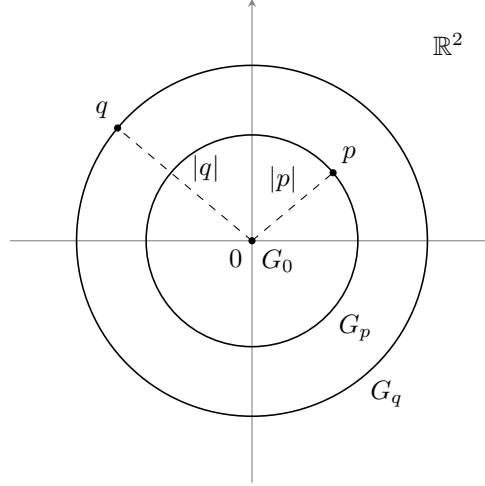
**Definition 7.16** (Orbit). Let  $\triangleright : G \times M \rightarrow M$  be a left  $G$ -action. For each  $p \in M$ , we define the **orbit** of  $p$  as the set:

$$G_p := \{q \in M \mid \exists g \in G : q = g \triangleright p\}$$

Alternatively, the orbit of  $p$  is the image of  $G$  under the map  $(- \triangleright p)$ . It consists of all the points in  $M$  that can be reached from  $p$  by successive applications of the action  $\triangleright$ .

*Example 7.8.*

Consider the action induced by representation of  $\text{SO}(2, \mathbb{R})$  as rotation matrices in  $\text{End}(\mathbb{R}^2)$ . The orbit of any  $p \in \mathbb{R}^2$  is the circle of radius  $|p|$  centred at the origin.



It should be intuitively clear from the definition that the orbits of two points are either disjoint or coincide. In fact, we have the following.

**Proposition 7.6.** Let  $\triangleright : G \times M \rightarrow M$  be an action on  $M$ . Define a relation on  $M$ :

$$p \sim q :\Leftrightarrow \exists g \in G : q = g \triangleright p$$

Then  $\sim$  is an equivalence relation on  $M$ .

*Proof.*

Let  $p, q, r \in M$ . We have:

i)  $p \sim p$  since  $p = e \triangleright p$ .

ii)  $p \sim q \Rightarrow q \sim p$  since, if  $q = g \triangleright p$ , then:

$$p = e \triangleright p = (g^{-1} \bullet g) \triangleright p = g^{-1} \triangleright (g \triangleright p) = g^{-1} \triangleright q$$

iii)  $(p \sim q \text{ and } q \sim r) \Rightarrow p \sim r$  since, if  $q = g_1 \triangleright p$  and  $r = g_2 \triangleright q$ , then:

$$r = g_2 \triangleright (g_1 \triangleright p) = (g_1 \bullet g_2) \triangleright p$$

Therefore,  $\sim$  is an equivalence relation on  $M$ . □

The equivalence classes of  $\sim$  are, by definition, the orbits.

**Definition 7.17** (Orbit Space). Let  $\triangleright : G \times M \rightarrow M$  be an action on  $M$ . The **orbit space** of  $M$  is:

$$M/G := M/\sim = \{G_p \mid p \in M\}$$

*Example 7.9.*

The orbit space of our previous  $\text{SO}(2, \mathbb{R})$ -action on  $\mathbb{R}^2$  is the partition of  $\mathbb{R}^2$  into concentric circles centred at the origin, plus the origin itself.

**Definition 7.18** (Stabiliser). Let  $\triangleright: G \times M \rightarrow M$  be a  $G$ -action on  $M$ . The **stabiliser** of  $p \in M$  is:

$$S_p := \{g \in G \mid g \triangleright p = p\}$$

Note that for each  $p \in M$ , the stabiliser  $S_p$  is a subgroup of  $G$ .

*Example 7.10.*

In our  $\text{SO}(2, \mathbb{R})$  example, we have  $S_p = \{\text{id}_{\mathbb{R}^2}\}$  for  $p \neq 0$  and  $S_0 = \text{SO}(2, \mathbb{R})$ .

**Definition 7.19** (Free / Transitive Actions). A left  $G$ -action  $\triangleright: G \times M \rightarrow M$  is said to be:

- i) **Free** if for all  $p \in M$ , we have  $S_p = \{e\}$ .
- ii) **Transitive** if for all  $p, q \in M$ , there exists  $g \in G$  such that  $p = g \triangleright q$ .

*Example 7.11.*

The action  $\triangleright: G \times V \rightarrow V$  induced by a representation  $R: G \rightarrow \text{GL}(V)$  is never free since  $\text{GL}(V)$  contains all the invertible linear transformation which, by definition, map 0 to 0 hence we always have  $S_0 = G$ .

*Example 7.12.*

Consider the action  $\triangleright: T(n) \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  of the  $n$ -dimensional translation group  $T(n)$  on  $\mathbb{R}^n$ . We have, rather trivially,  $T(n)_p = \mathbb{R}^n$  for every  $p \in \mathbb{R}^n$ . It is also easy to show that this action is free and transitive.

**Proposition 7.7.** Let  $\triangleright: G \times M \rightarrow M$  be a free action. Then:

$$g_1 \triangleright p = g_2 \triangleright p \quad \Leftrightarrow \quad g_1 = g_2$$

*Proof.*

The ( $\Leftarrow$ ) direction is obvious. Suppose that there exist  $p \in M$  and  $g_1, g_2 \in G$  such that  $g_1 \triangleright p = g_2 \triangleright p$ . Then:

$$\begin{aligned} g_1 \triangleright p = g_2 \triangleright p &\Leftrightarrow g_2^{-1} \triangleright (g_1 \triangleright p) = g_2^{-1} \triangleright (g_2 \triangleright p) \\ &\Leftrightarrow (g_2^{-1} \bullet g_1) \triangleright p = (g_2^{-1} \bullet g_2) \triangleright p \\ &\Leftrightarrow (g_2^{-1} \bullet g_1) \triangleright p = (e \triangleright p) \\ &\Leftrightarrow (g_2^{-1} \bullet g_1) \triangleright p = p \end{aligned}$$

Hence  $g_2^{-1} \bullet g_1 \in S_p$ , but since  $\triangleright$  is free we have  $S_p = \{e\}$ , and thus  $g_1 = g_2$ . □

**Proposition 7.8.** If  $\triangleright: G \times M \rightarrow M$  is a free action, then:

$$\forall p \in G: G_p \cong_{\text{diff}} G$$

*Example 7.13.*

Define  $\triangleright: \text{SO}(2, \mathbb{R}) \times \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^2 \setminus \{0\}$  to coincide with the action induced by the representation of  $\text{SO}(2, \mathbb{R}^2)$  on  $\mathbb{R}^2$  for each non-zero point of  $\mathbb{R}^2$ . Then this action is free, since we have  $S_p = \{\text{id}_{\mathbb{R}^2}\}$  for  $p \neq 0$ , and the previous proposition implies:

$$\forall p \in \mathbb{R}^2 \setminus \{0\}: \text{SO}(2, \mathbb{R})_p \cong_{\text{diff}} \text{SO}(2, \mathbb{R}) \cong_{\text{diff}} S^1$$

## Chapter 8

# Principal Fiber Bundles

A principal fiber bundle (often called simply principal bundle) is a mathematical object that formalizes some of the essential features of the Cartesian product  $X \times G$  of a space  $X$  with a group  $G$ . In other words, very roughly speaking, a principal bundle is a fiber bundle whose typical fiber is a Lie group.

A common example of a principal bundle is the frame bundle of a vector bundle, which consists of all ordered bases of the vector space attached to each point. The group  $G$  in this case is the general linear group, which acts on the right in the usual way: by changes of basis. Since there is no natural way to choose an ordered basis of a vector space, a frame bundle lacks a canonical choice of identity cross section.

Principal bundles have important applications in topology and differential geometry. They have also found application in physics where they form part of the foundational framework of gauge theories. Principal bundles are so immensely important because they allow us to understand any fiber bundle with fiber  $F$  on which a Lie group  $G$  acts. These are then called associated fiber bundles (usually called associated bundles), and will be discussed later on.

### 8.1 Principal Bundles

**Definition 8.1** (Principal Bundle). *Let  $G$  be a Lie group. A smooth bundle  $(E, \pi, M)$  is called a **principal bundle** (or sometimes **principal  $G$ -bundle**) if  $E$  is equipped with a free right  $G$ -action and:*

$$\begin{array}{ccc} E & & E \\ \pi \downarrow & \cong_{\text{bdl}} & \downarrow \rho \\ M & & E/G \end{array}$$

where  $\rho$  is the quotient map, defined by sending each  $p \in E$  to its equivalence class (i.e. orbit) in the orbit space  $E/G$ .

Observe that since the right action of  $G$  on  $E$  is, by definition, free, for each  $p \in E$  we have that the fiber is the whole group:

$$\text{preim}_\rho(G_p) = G_p \cong_{\text{diff}} G$$

We said at beginning that, roughly speaking, a principal bundle is a bundle whose fiber at each point is a Lie group. Note that the formal definition is that a principal bundle is a bundle which is isomorphic to a bundle whose fibers are the orbits under the right action of  $G$ , which are themselves isomorphic to  $G$  since the action is free.

*Remark 8.1.* A slight generalisation would be to consider smooth bundles  $E \xrightarrow{\pi} M$  where  $E$  is equipped with a right  $G$ -action which is free and transitive on each fiber of  $E \xrightarrow{\pi} M$ . The isomorphism in our definition enforces the fiber-wise transitivity since  $G$  acts transitively on each  $G_p$  by the definition of orbit.

### 8.1.1 Principal Bundle Maps

As usual once we have defined a new object we want to define maps between these objects and study their morphisms.

Recall that a bundle morphism (also called simply a bundle map) between two bundles  $(E, \pi, M)$  and  $(E', \pi', M')$  is a pair of maps  $(u, f)$  such that the diagram:

$$\begin{array}{ccc} E & \xrightarrow{u} & E' \\ \pi \downarrow & & \downarrow \pi' \\ M & \xrightarrow{f} & M' \end{array}$$

commutes, that is,  $f \circ \pi = \pi' \circ u$ . As usual, two smooth bundles  $(E, \pi, M)$  and  $(E', \pi', M')$  are isomorphic if  $u, f$  are diffeomorphisms.

We want now to extend these concepts from bundles to principal bundles. The extension is not automatic since we have to take care of the extra structure of the underlying group.

**Definition 8.2** (Principal Bundle Map). *Let  $(P, \pi, M)$  be a principal  $G$ -bundle, let  $(P', \pi', M')$  be a principal  $G'$ -bundle, and let  $\rho: G \rightarrow G'$  be a Lie group homomorphism. A **principal bundle morphism** from  $(P, \pi, M)$  to  $(P', \pi', M')$  is a pair of smooth maps  $(u, f)$  such that the diagram:*

$$\begin{array}{ccc} P & \xrightarrow{u} & P' \\ \wr G \uparrow & & \uparrow \wr' G' \\ P & \xrightarrow{u} & P' \\ \pi \downarrow & & \downarrow \pi' \\ M & \xrightarrow{f} & M' \end{array}$$

commutes, that is:

$$\begin{aligned} \forall p \in P : (f \circ \pi)(p) &= (\pi' \circ u)(p) \\ \forall p \in P : \forall g \in G : u(p \wr g) &= u(p) \wr' \rho(g) \end{aligned}$$

*Remark 8.2.* Note that one can restrict this general definition to the case where the same group  $G$  acts in both principal bundles (i.e  $G = G'$ ). In this case there is no Lie group homomorphism  $\rho: G \rightarrow G'$ , and the elements of the group  $G$  can act in both total spaces  $P$  and  $P'$  as they are. In this case the commutativity of the diagram translates to:

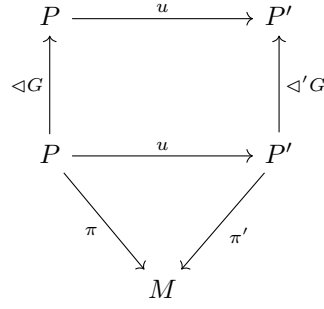
$$\begin{aligned} \forall p \in P : (f \circ \pi)(p) &= (\pi' \circ u)(p) \\ \forall p \in P : \forall g \in G : u(p \wr g) &= u(p) \wr g \end{aligned}$$

where we keep the symbol  $\wr'$  because we may define a different action (of the same group) in the two Principal bundles.

**Definition 8.3** (Principal Bundle Isomorphism/Diffeomorphism). *A principal bundle morphism between a principal  $G$ -bundle and a principal  $G'$ -bundle is an **isomorphism** (or **diffeomorphism**) of **principal bundles** if it is also a bundle isomorphism and  $\rho$  is a Lie group isomorphism.*

*Remark 8.3.* Similarly in the case where  $G = G'$ , a principal bundle morphism between two principal  $G$ -bundles is an **isomorphism** (or **diffeomorphism**) of **principal bundles** if it is also a bundle isomorphism.

**Lemma 8.1.** *Let  $(P, \pi, M)$  and  $(P', \pi', M)$  be principal bundles over the same base manifold  $M$ . Then, any  $u: P \rightarrow P'$  such that  $(u, \text{id}_M)$  is a principal bundle morphism is necessarily a diffeomorphism.*



*Proof.*

We already know that  $u$  is smooth since  $(u, \text{id}_M)$  is assumed to be a principal bundle morphism. It remains to check that  $u$  is bijective and its inverse is also smooth.

i) Let  $p_1, p_2 \in P$  be such that  $u(p_1) = u(p_2)$ . Then:

$$\pi(p_1) = \pi'(u(p_1)) = \pi'(u(p_2)) = \pi(p_2)$$

that is,  $p_1$  and  $p_2$  belong to the same fiber. As the action of  $G$  on  $P$  is fiber-wise transitive, there is a unique  $g \in G$  such that  $p_1 = p_2 \triangleleft g$ . Then:

$$u(p_1) = u(p_2 \triangleleft g) = u(p_2) \triangleleft' g = u(p_1) \triangleleft' g$$

so  $g \in S_{u(p_1)}$ , but since  $\triangleleft'$  is free, we have  $g = e$  and thus:

$$p_1 = p_2 \triangleleft e = p_2$$

Therefore  $u$  is injective.

ii) Let  $p' \in P'$ . Choose some  $p \in \text{preim}_\pi(\pi'(p'))$ . Then, we have:

$$\pi'(u(p)) = \pi(p) = \pi'(p')$$

so that  $u(p)$  and  $p'$  belong to the same fiber. Hence, there is a unique  $g \in G$  such that  $p' = u(p) \triangleleft' g$ . We thus have:

$$p' = u(p) \triangleleft' g = u(p \triangleleft g)$$

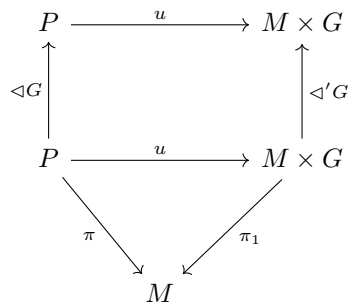
and since  $p \triangleleft g \in P$ , the map  $u$  is surjective.

Hence,  $u$  is a diffeomorphism.  $\square$

**Definition 8.4** (Trivial Principal Bundle). A principal  $G$ -bundle  $(P, \pi, M)$  is called **trivial** if it is isomorphic as a principal bundle to the principal  $G$ -bundle  $(M \times G, \pi_1, M)$  where  $\pi_1$  is the projection onto the first component and the action is defined as:

$$\begin{aligned}
\triangleleft' : (M \times G) \times G &\rightarrow M \times G \\
((p, g), g') &\mapsto (p, g) \triangleleft' g' := (p, g \bullet g')
\end{aligned}$$

By the previous lemma, a principal  $G$ -bundle  $(P, \pi, M)$  is trivial if there exists a smooth map  $u: P \rightarrow M \times G$  such that the following diagram commutes:



The following result provides a necessary and sufficient criterion for when a principal bundle is trivial. Note that while we have used the lower case letter  $p$  almost exclusively to denote points of the base manifold  $M$ , in the next proof we will use it to denote points of the total space  $P$  instead.

**Theorem 8.1.** *A principal  $G$ -bundle  $(P, \pi, M)$  is trivial if, and only if, there exists a smooth section  $\sigma \in \Gamma(P)$ , that is, a smooth  $\sigma: M \rightarrow P$  such that  $\pi \circ \sigma = \text{id}_M$ .*

*Proof.*

( $\Rightarrow$ ) Suppose that  $(P, \pi, M)$  is trivial. Then there exists a diffeomorphism  $u: P \rightarrow M \times G$  which make the following diagram commute:

$$\begin{array}{ccc}
 & P & \\
 \lhd G \uparrow & & \\
 P & \xleftarrow{u^{-1}} & M \times G \\
 \pi \searrow & & \swarrow \pi_1 \\
 & M &
 \end{array}$$

We can define:

$$\begin{aligned}
 \sigma: M &\rightarrow P \\
 m &\mapsto u^{-1}(m, e)
 \end{aligned}$$

where  $e$  is the identity of  $G$ . Then  $\sigma$  is smooth since it is the composition of  $u^{-1}$  with the map  $p \mapsto (p, e)$ , which are both smooth. We also have:

$$(\pi \circ \sigma)(m) = \pi(u^{-1}(m, e)) = \pi_1(m, e) = m$$

for all  $m \in M$ , hence  $\pi \circ \sigma = \text{id}_M$  and thus  $\sigma \in \Gamma(P)$ .

( $\Leftarrow$ ) Suppose that there exists a smooth section  $\sigma: M \rightarrow P$ . Let  $p \in P$  and consider the point  $\sigma(\pi(p)) \in P$ . We have:

$$\pi(\sigma(\pi(p))) = \text{id}_M(\pi(p)) = \pi(p)$$

hence  $\sigma(\pi(p))$  and  $p$  belong to the same fiber, and thus there exists a unique group element in  $G$  which links the two points via  $\lhd$ . Since this element depends on both  $\sigma$  and  $p$ , let us denote it by  $\chi_\sigma(p)$ . Then,  $\chi_\sigma$  defines a function:

$$\begin{aligned}
 \chi_\sigma: P &\rightarrow G \\
 p &\mapsto \chi_\sigma(p)
 \end{aligned}$$

and we can write:

$$\forall p \in P: p = \sigma(\pi(p)) \lhd \chi_\sigma(p)$$

In particular, for any other  $g \in G$  we have  $p \lhd g \in P$  and thus:

$$p \lhd g = \sigma(\pi(p \lhd g)) \lhd \chi_\sigma(p \lhd g) = \sigma(\pi(p)) \lhd \chi_\sigma(p \lhd g)$$

where the second equality follows from the fact that the fibers of  $P$  are precisely the orbits under the action of  $G$ .

On the other hand, we can act on the right with an arbitrary  $g \in G$  directly to obtain :

$$p \lhd g = (\sigma(\pi(p)) \lhd \chi_\sigma(p)) \lhd g = \sigma(\pi(p)) \lhd (\chi_\sigma(p) \bullet g)$$

Combining the last two equations yields:

$$\sigma(\pi(p)) \triangleleft \chi_\sigma(p \triangleleft g) = \sigma(\pi(p)) \triangleleft (\chi_\sigma(p) \bullet g)$$

and hence:

$$\chi_\sigma(p \triangleleft g) = (\chi_\sigma(p) \bullet g)$$

We can now define the map:

$$\begin{aligned} u_\sigma: P &\rightarrow M \times G \\ p &\mapsto (\pi(p), \chi_\sigma(p)) \end{aligned}$$

By our previous lemma, it suffices to show that  $u_\sigma$  is a principal bundle map.

$$\begin{array}{ccc} P & \xrightarrow{u_\sigma} & M \times G \\ \uparrow \triangleleft G & & \uparrow \triangleleft' G \\ P & \xrightarrow{u_\sigma} & M \times G \\ \searrow \pi & & \swarrow \pi_1 \\ & M & \end{array}$$

By definition, we have:

$$(\pi_1 \circ u_\sigma)(p) = \pi_1(\pi(p), \chi_\sigma(p)) = \pi(p)$$

for all  $p \in P$ , so the lower triangle commutes. Moreover, we have:

$$\begin{aligned} u_\sigma(p \triangleleft g) &= (\pi(p \triangleleft g), \chi_\sigma(p \triangleleft g)) \\ &= (\pi(p), \chi_\sigma(p) \bullet g) \\ &= (\pi(p), \chi_\sigma(p)) \triangleleft' g \\ &= u_\sigma(p) \triangleleft' g \end{aligned}$$

for all  $p \in P$  and  $g \in G$ , so the upper square also commutes and hence  $(P, \pi, M)$  is a trivial bundle.  $\square$

## 8.2 Associated Bundles

An associated fiber bundle (often called simply associated bundle) is a fiber bundle which is associated (in a precise sense) to a principal  $G$ -bundle. Associated bundles are related to their underlying principal bundles in a way that models the transformation law for components under a change of basis.

**Definition 8.5** (Associated Bundle). *Let  $(P, \pi, M)$  be a principal  $G$ -bundle and let  $F$  be a smooth manifold equipped with a left  $G$ -action  $\triangleright$ . We define:*

i)  $P_F := (P \times F)/\sim_G$ , where  $\sim_G$  is the equivalence relation:

$$(p, f) \sim_G (p', f') \quad :\Leftrightarrow \quad \exists g \in G : \begin{cases} p' &= p \triangleleft g \\ f' &= g^{-1} \triangleright f \end{cases}$$

We denote the points of  $P_F$  as  $[p, f]$ .

ii) The map:

$$\begin{aligned} \pi_F: P_F &\rightarrow M \\ [p, f] &\mapsto \pi(p) \end{aligned}$$

which is well-defined since, if  $[p', f'] = [p, f]$ , then for some  $g \in G$ :

$$\pi_F([p', f']) = \pi_F([p \triangleleft g, g^{-1} \triangleright f]) := \pi(p \triangleleft g) = \pi(p) =: \pi_F([p, f])$$

The **associated bundle** (to  $(P, \pi, M)$ ,  $F$  and  $\triangleright$ ) is the bundle  $(P_F, \pi_F, M)$ .

### 8.2.1 Associated Bundle Maps

Once again, once we have defined associated bundles, we want to define maps between them and study their morphisms.

**Definition 8.6** (Associated Bundle Map). Let  $(P_F, \pi_F, M)$  to  $(Q_F, \pi'_F, N)$  be the associated bundles (with the same fiber  $F$ ) of two principal  $G$ -bundles  $(P, \pi, M)$  and  $(P', \pi', M')$ . An **associated bundle map** between the associated bundles is a bundle map  $(\tilde{u}, \tilde{h})$  between them such that for some  $u$ , the pair  $(u, h)$  is a principal bundle map between the underlying principal  $G$ -bundles, i.e:

$$\forall p \in P : (\tilde{h} \circ \pi)(p) = (\pi' \circ u)(p)$$

$$\forall p \in P : \forall g \in G : u(p \triangleleft g) = u(p) \triangleleft' g$$

and on top of that:

$$\forall [p, f] \in P_F : (h \circ \pi_F)([p, f]) = (\pi'_F \circ \tilde{u})([p, f])$$

where :

$$\forall [p, f] \in P_F : \tilde{u}([p, f]) := [u(p), f]$$

$$\forall m \in M : \tilde{h}(m) := h(m)$$

Equivalently, the following **two** diagrams both commute:

$$\begin{array}{ccc} P_F & \xrightarrow{\tilde{u}} & P'_F \\ \pi_F \downarrow & & \downarrow \pi'_F \\ M & \xrightarrow{\tilde{h}} & M' \end{array} \quad \begin{array}{ccc} P & \xrightarrow{u} & P' \\ \triangleleft G \uparrow & & \uparrow \triangleleft' G \\ P & \xrightarrow{u} & P' \\ \pi \downarrow & & \downarrow \pi' \\ M & \xrightarrow{h} & M' \end{array}$$

*Remark 8.4.* Since  $\forall m \in M : \tilde{h}(m) := h(m)$  this simply implies that  $\tilde{h} = h$ . From now on we will simply denote  $(\tilde{u}, \tilde{h})$  a  $(\tilde{u}, h)$ .

**Definition 8.7** (Associated Bundle Isomorphism). An associated bundle map  $(\tilde{u}, h)$  is an **associated bundle isomorphism** if  $\tilde{u}$  and  $h$  are invertible and  $(\tilde{u}^{-1}, h^{-1})$  is also an associated bundle map.

*Remark 8.5.* Note that two associated  $F$ -fiber bundles may be isomorphic as bundles but not as associated bundles. In other words, there may exist a bundle isomorphism between them, but there may not exist any bundle isomorphism between them which can be written as in the definition for some principal bundle isomorphism between the underlying principal bundles.

Recall that an  $F$ -fiber bundle  $(E, \pi, M)$  is called trivial if there exists a bundle isomorphism:

$$\begin{array}{ccccc} F & \longrightarrow & E & \xrightarrow{u} & M \times F \\ & & \searrow \pi & & \swarrow \pi_1 \\ & & & M & \end{array}$$

while a principal  $G$ -bundle is called trivial if there exists a principal bundle isomorphism:

$$\begin{array}{ccc}
 P & \xrightarrow{u} & M \times G \\
 \uparrow \triangleleft G & & \uparrow \triangleleft' G \\
 P & \xrightarrow{u} & M \times G \\
 \searrow \pi & & \swarrow \pi_1 \\
 & M &
 \end{array}$$

**Definition 8.8** (Trivial Associated Bundle). *An associated bundle  $(P_F, \pi_F, M)$  is called **trivial** if the underlying principal  $G$ -bundle  $(P, \pi, M)$  is trivial.*

**Proposition 8.1.** *A trivial associated bundle is a trivial fiber bundle.*

Note that the converse does not hold. An associated bundle can be trivial as a fiber bundle but not as an associated bundle, i.e. the underlying principal fiber bundle need not be trivial simply because the associated bundle is trivial as a fiber bundle.

**Definition 8.9** ( $H$ -Restriction /  $G$ -Extension). *Let  $H$  be a closed Lie subgroup of  $G$ . Let  $(P, \pi, M)$  be a principal  $H$ -bundle and  $(P', \pi', M)$  a principal  $G$ -bundle. If there exists a principal bundle map from  $(P, \pi, M)$  to  $(P', \pi', M)$ , i.e. a smooth bundle map which is equivariant with respect to the inclusion of  $H$  into  $G$ , then  $(P, \pi, M)$  is called an  $H$ -restriction of  $(P', \pi', M)$ , while  $(P', \pi', M)$  is called a  $G$ -extension of  $(P, \pi, M)$ .*

**Theorem 8.2.** *Let  $H$  be a closed Lie subgroup of  $G$ .*

- i) *Any principal  $H$ -bundle can be extended to a principal  $G$ -bundle.*
- ii) *A principal  $G$ -bundle  $(P, \pi, M)$  can be restricted to a principal  $H$ -bundle if, and only if, the bundle  $(P/H, \pi', M)$  has a section.*

### 8.3 Application: Principal Frame Bundle

One of the most important examples of a principal  $G$ -bundle is the so called “principal frame bundle”. Given its importance we will study it as a separate application. We will begin simply by defining it as a principal  $G$ -bundle. We will do that in steps so it’s clear what we’re doing.

#### Frame Bundle

Let  $M$  be a smooth manifold. Consider the space:

$$L_p M := \{(e_1, \dots, e_{\dim M}) \mid e_1, \dots, e_{\dim M} \text{ is a basis of } T_p M\}$$

In other words,  $L_p M$  is simply the collection (as a tuple) of the vectors of a basis of  $T_p M$ . We know from linear algebra that the bases of a vector space are related to each other by invertible linear transformations. Hence, we have:

$$L_p M \cong_{\text{vec}} \text{GL}(\dim M, \mathbb{R})$$

We define the frame bundle of  $M$  in the similar way as the tangent bundle:

$$LM := \bigcup_{p \in M} L_p M$$

with the obvious projection map  $\pi: LM \rightarrow M$  sending each basis  $(e_1, \dots, e_{\dim M})$  to the unique point  $p \in M$  such that  $(e_1, \dots, e_{\dim M})$  is a basis of  $T_p M$ .

By proceeding similarly to the case of the tangent bundle, we can equip  $LM$  with a smooth structure inherited from that of  $M$ . We then find:

$$\dim LM = \dim M + \dim T_p M = \dim M + (\dim M)^2$$

## Principal Frame Bundle

We would now like to make  $LM \xrightarrow{\pi} M$  into a principal  $G$ -bundle. Let us remind ourselves some things we mentioned earlier.

Recall that we defined the set of all invertible maps  $\text{Aut}(V) := \{\phi \in \text{End}(V) \mid \phi \text{ is an isomorphism}\}$ , and we further said that we can equip this set with matrix multiplication and show that it is closed under the operation, hence we turn it into a group that we call  $GL(V)$ . On top of that, we can show that  $GL(V)$  is actually a topological group, hence a Lie group, so it can be used as the group of a principal  $G$ -bundle.

Recall also that the elements  $g$  of  $GL(V)$  are actually “one-one” tensors (i.e.  $g \in T_1^1 V$ ) and that once we pick a basis and we switch to the matrix notation then we can denote them as  $g_b^a$ . Then we usually denote the group as  $GL(n, V)$  where  $n$  indicates the number of basis vectors (which is of course the dimension of the manifold) and  $V$  the same underlying space as before.

Hence in our case, we use as a group  $G$  the group  $GL(\dim M, \mathbb{R})$ . We define a right  $GL(\dim M, \mathbb{R})$ -action on  $LM$  by:

$$(e_1, \dots, e_{\dim M}) \triangleleft g := (g_1^a e_a, \dots, g_{\dim M}^a e_a)$$

where  $g_b^a$  are the components of the automorphism (usually called endomorphism but the invertibility is implied)  $g \in GL(\dim M, \mathbb{R})$  with respect to the standard basis on  $\mathbb{R}^n$ . Note that if  $(e_1, \dots, e_{\dim M}) \in L_p M$ , we must also have  $(e_1, \dots, e_{\dim M}) \triangleleft g \in L_p M$ , hence we just defined a mathematically proper way to talk about a change of basis (linear basis transformations).

This action is free since:

$$(e_1, \dots, e_{\dim M}) \triangleleft g = (e_1, \dots, e_{\dim M}) \Leftrightarrow (g_1^a e_a, \dots, g_{\dim M}^a e_a) = (e_1, \dots, e_{\dim M})$$

and hence, by linear independence,  $g_b^a = \delta_b^a$ , so  $g = \text{id}_{\mathbb{R}^n}$ . Note that since all bases of each  $T_p M$  are related by some  $g \in GL(\dim M, \mathbb{R})$ ,  $\triangleleft$  is also fiber-wise transitive.

We now have to show that:

$$\begin{array}{ccc} LM & & LM \\ \pi \downarrow & \cong_{\text{bdl}} & \downarrow \rho \\ M & & LM / GL(\dim M, \mathbb{R}) \end{array}$$

i.e. that there exist smooth maps  $u$  and  $f$  such that the diagram:

$$\begin{array}{ccc} LM & \xrightleftharpoons[u^{-1}]{u} & LM \\ \pi \downarrow & & \downarrow \rho \\ M & \xrightleftharpoons[f^{-1}]{f} & LM / GL(\dim M, \mathbb{R}) \end{array}$$

commutes.

We can simply choose  $u = u^{-1} = \text{id}_{LM}$ , while we define  $f$  as:

$$\begin{aligned} f: M &\rightarrow LM / GL(\dim M, \mathbb{R}) \\ p &\mapsto GL(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} \end{aligned}$$

where  $(e_1, \dots, e_{\dim M})$  is some basis of  $T_p M$ , i.e.  $(e_1, \dots, e_{\dim M}) \in \text{preim}_\pi(\{p\})$ . Note that  $f$  is well-defined since every basis of  $T_p M$  gives rise to the same orbit in the orbit space  $LM / GL(\dim M, \mathbb{R})$ . Moreover, it is injective since:

$$f(p) = f(p') \Leftrightarrow GL(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} = GL(\dim M, \mathbb{R})_{(e'_1, \dots, e'_{\dim M})}$$

which is true only if  $(e_1, \dots, e_{\dim M})$  and  $(e'_1, \dots, e'_{\dim M})$  are basis of the same tangent space, so  $p = p'$ .

It is clearly surjective since every orbit in  $LM/\mathrm{GL}(\dim M, \mathbb{R})$  is the orbit of some basis of some tangent space  $T_p M$  at some point  $p \in M$ . The inverse map is given explicitly by:

$$\begin{aligned} f^{-1}: \quad LM/\mathrm{GL}(\dim M, \mathbb{R}) &\rightarrow M \\ \mathrm{GL}(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} &\mapsto \pi((e_1, \dots, e_{\dim M})) \end{aligned}$$

Finally, we have:

$$(\rho \circ \mathrm{id}_{LM})(e_1, \dots, e_{\dim M}) = \mathrm{GL}(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} = (f \circ \pi)(e_1, \dots, e_{\dim M})$$

and thus  $LM \xrightarrow{\pi} M$  is a principal  $G$ -bundle, called the *principal frame bundle* of  $M$  or simply frame bundle.

*Remark 8.6.* A note to the careful reader. As we have just done in the previous example, in the following we will sometimes simply assume that certain maps are smooth, instead of rigorously proving it.

### Associated Frame Bundle

Up to that point we showed that the frame bundle  $(LM, \pi, M)$  is a principal  $\mathrm{GL}(d, \mathbb{R})$ -bundle, where  $d = \dim M$ , with right  $G$ -action  $\triangleleft: LM \times G \rightarrow LM$  given by:

$$(e_1, \dots, e_d) \triangleleft g := (g^a_1 e_a, \dots, g^a_d e_a)$$

Let  $F := \mathbb{R}^d$  (as a smooth manifold) and define a left action:

$$\begin{aligned} \triangleright: \mathrm{GL}(d, \mathbb{R}) \times \mathbb{R}^d &\rightarrow \mathbb{R}^d \\ (g, x) &\mapsto g \triangleright x \end{aligned}$$

where:

$$(g \triangleright x)^a := g^a_b x^b$$

Then  $(LM_{\mathbb{R}^d}, \pi_{\mathbb{R}^d}, \mathbb{R}^d)$  is the associated bundle. In fact, we have a bundle isomorphism:

$$\begin{array}{ccc} LM_{\mathbb{R}^d} & \xrightarrow{u} & TM \\ \pi_{\mathbb{R}^d} \downarrow & & \downarrow \pi \\ M & \xrightarrow{\mathrm{id}_M} & M \end{array}$$

where  $(TM, \pi, M)$  is the tangent bundle of  $M$ , and  $u$  is defined as:

$$\begin{aligned} u: \quad LM_{\mathbb{R}^d} &\rightarrow TM \\ [(e_1, \dots, e_d), x] &\mapsto x^a e_a \end{aligned}$$

The inverse map  $u^{-1}: TM \rightarrow LM_{\mathbb{R}^d}$  works as follows. Given any  $X \in TM$ , pick any basis  $(e_1, \dots, e_d)$  of the tangent space at the point  $\pi(X) \in M$ , i.e. any element of  $L_{\pi(X)} M$ . Decompose  $X$  as  $x^a e_a$ , with each  $x^a \in \mathbb{R}$ , and define:

$$u^{-1}(X) := [(e_1, \dots, e_d), x]$$

The map  $u^{-1}$  is well-defined since, while the pair  $((e_1, \dots, e_d), x) \in LM \times \mathbb{R}^d$  clearly depends on the choice of basis, the equivalence class :

$$[(e_1, \dots, e_d), x] \in LM_{\mathbb{R}^d} := (LM \times \mathbb{R}^d)/\sim_G$$

does not. It includes all pairs  $((e_1, \dots, e_d) \triangleleft g, g^{-1} \triangleright x)$  for every  $g \in \mathrm{GL}(d, \mathbb{R})$ , i.e. every choice of basis together with the “right” components  $x \in \mathbb{R}^d$ .

Even though the associated bundle  $(LM_{\mathbb{R}^d}, \pi_{\mathbb{R}^d}, \mathbb{R}^d)$  is isomorphic to the tangent bundle  $(TM, \pi, M)$ , note a subtle difference between the two. On the tangent bundle, the transformation law for a change of basis

and the related transformation law for components are *deduced* from the definitions by undergraduate linear algebra.

On the other hand, the transformation laws on  $LM_{\mathbb{R}^d}$  were *chosen* by us in its definition. We chose the Lie group  $GL(d, \mathbb{R})$ , the specific right action  $\triangleleft$  on  $LM$ , the space  $\mathbb{R}^d$ , and the specific left action on  $\mathbb{R}^d$ . It just happens that, with these choices, the resulting associated bundle is isomorphic to the tangent bundle. Of course, we have the freedom to make different choices and construct bundles which behave very differently from  $TM$ .

For example, consider now the principal  $GL(d, \mathbb{R})$ -bundle  $(LM, \pi, M)$  again, with the same right action as before, however this time we define a different fiber:

$$F := (\mathbb{R}^d)^{\times p} \times (\mathbb{R}^{d*})^{\times q} := \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{p \text{ times}} \times \underbrace{\mathbb{R}^{d*} \times \cdots \times \mathbb{R}^{d*}}_{q \text{ times}}$$

with left  $GL(d, \mathbb{R})$ -action  $\triangleright: GL(d, \mathbb{R}) \times F \rightarrow F$  given by:

$$(g \triangleright f)^{a_1 \cdots a_p}_{b_1 \cdots b_q} := g^{a_1}_{\tilde{a}_1} \cdots g^{a_p}_{\tilde{a}_p} (g^{-1})^{\tilde{b}_1}_{b_1} \cdots (g^{-1})^{\tilde{b}_q}_{b_q} f^{\tilde{a}_1 \cdots \tilde{a}_p}_{\tilde{b}_1 \cdots \tilde{b}_q}$$

Then, the associated bundle  $(LM_F, \pi_F, M)$  thus constructed is isomorphic to  $(T_q^p M, \pi, M)$ , the  $(p, q)$ -tensor bundle on  $M$ .

Now for something new, consider the following.

**Definition 8.10** (Tensor Density). *Let  $M$  be a smooth manifold and let  $(LM, \pi, M)$  be its frame bundle, with right  $GL(d, \mathbb{R})$ -action as above. Let  $F := (\mathbb{R}^d)^{\times p} \times (\mathbb{R}^{d*})^{\times q}$  and define a left  $GL(d, \mathbb{R})$ -action on  $F$  by:*

$$(g \triangleright f)^{a_1 \cdots a_p}_{b_1 \cdots b_q} := (\det g^{-1})^\omega g^{a_1}_{\tilde{a}_1} \cdots g^{a_p}_{\tilde{a}_p} (g^{-1})^{\tilde{b}_1}_{b_1} \cdots (g^{-1})^{\tilde{b}_q}_{b_q} f^{\tilde{a}_1 \cdots \tilde{a}_p}_{\tilde{b}_1 \cdots \tilde{b}_q}$$

where  $\omega \in \mathbb{Z}$ . Then the associated bundle  $(LM_F, \pi_F, M)$  is called the  $(p, q)$ -**tensor  $\omega$ -density bundle** on  $M$ , and its sections are called  $(p, q)$ -**tensor densities of weight  $\omega$** .

*Remark 8.7.* Some special cases include the following:

- i) If  $\omega = 0$ , we recover the  $(p, q)$ -tensor bundle on  $M$ .
- ii) If  $F = \mathbb{R}$  (i.e.  $p = q = 0$ ), the left action reduces to:

$$(g \triangleright f) = (\det g^{-1})^\omega f$$

which is the transformation law for a *scalar density of weight  $\omega$* .

- iii) If  $GL(d, \mathbb{R})$  is restricted in such a way that we always have  $(\det g^{-1}) = 1$ , then tensor densities are indistinguishable from ordinary tensor fields. This is why you probably haven't met tensor densities in your special relativity course.

*Remark 8.8.* Recall that if  $B$  is a bilinear form on a  $K$ -vector space  $V$ , the determinant of  $B$  is not independent from the choice of basis. Indeed, if  $\{e_a\}$  and  $\{e'_b := g^a_b e_a\}$  are both basis of  $V$ , where  $g \in GL(\dim V, K)$ , then:

$$(\det B)' = (\det g^{-1})^2 \det B$$

Once recast in the principal and associated bundle formalism, we find that the determinant of a bilinear form is a scalar density of weight 2.

*Remark 8.9.* As a final note, remember the theorem we stated in the previous section, that for a closed Lie subgroup  $H$  of  $G$ :

- i) Any principal  $H$ -bundle can be extended to a principal  $G$ -bundle.
- ii) A principal  $G$ -bundle  $(P, \pi, M)$  can be restricted to a principal  $H$ -bundle if, and only if, the bundle  $(P/H, \pi', M)$  has a section.

Coming to our case:

- i) The bundle  $(LM/\mathrm{SO}(d), \pi, M)$  always has a section, and since  $\mathrm{SO}(d)$  is a closed Lie subgroup of  $\mathrm{GL}(d, \mathbb{R})$ , the frame bundle can be restricted to a principal  $\mathrm{SO}(d)$ -bundle. This is related to the fact that any manifold can be equipped with a Riemannian metric.
- ii) The bundle  $(LM/\mathrm{SO}(1, d-1), \pi, M)$  may or may not have a section. For example, the bundle  $(LS^2/\mathrm{SO}(1, 1), \pi, S^2)$  does not admit any section, and hence we cannot restrict  $(LS^2/\mathrm{SO}(1, 1), \pi, S^2)$  to a principal  $\mathrm{SO}(1, 1)$ -bundle, even though  $\mathrm{SO}(1, 1)$  is a closed Lie subgroup of  $\mathrm{GL}(2, \mathbb{R})$ . This is related to the fact that the 2-sphere cannot be equipped with a Lorentzian metric.

This concludes our application of principle fibre bundles to the very useful case of principle frame bundles. Now we switch to the concept of connections on principle fibre bundles.

## 8.4 Connections On A Principal $G$ -Bundle

In elementary courses on differential geometry or general relativity, the notions of connection, parallel transport and covariant derivative are often confused with one another. Sometimes, the terms are even used as synonyms. If you have seen any of that before, it is probably best to forget about it for the time being.

What a connection really is, is just additional structure on a principal bundle consisting is a “smooth” assignment of a particular vector space at each point of the base manifold compatible with the right action of the Lie group on the principal bundle. Such an assignment is, in fact, equivalent to a certain Lie-algebra-valued one-form on the principal bundle, as we will discuss below. Later, we will see that a connection on a principal bundle induces a parallel transport map on the principal bundle, which in turn induces a parallel transport map on any of its associated bundles. If the fibers of the associated bundle carry a vector space structure, then the parallel transport can be used to define a covariant derivative on the associated bundle.

Hence the conceptual sequence “connection, parallel transport covariant derivative” is in decreasing order of generality, and it should be clear that treating these terms as synonyms will inevitably lead to confusion. We will now discuss the first of these in some detail.

Let  $(P, \pi, M)$  be a principal  $G$ -bundle. Recall that every element of  $X_e \in T_e G$  gives rise to a left invariant vector field on  $G$  which we denoted by  $X = j(X_e) = (\ell_{g*})_e(X_e)$ . At this point we will change our notation a bit since we will invoke vectors from different space. Namely, we will denote the element  $X_e$  as  $A_e$  hence  $A_e \in T_e G$ .

We now define for every point  $p \in P$  a map  $i_p$  as follows:

$$\begin{aligned} i_p: T_e G &\rightarrow T_p P \\ A_e &\mapsto X_p^{A_e} \end{aligned}$$

where:

$$\begin{aligned} X_p^{A_e}: \mathcal{C}^\infty(P) &\xrightarrow{\sim} \mathbb{R} \\ f &\mapsto X_p^{A_e}(f) := [f(p \triangleleft \exp(tA_e))]'(0) \end{aligned}$$

where the derivative is to be taken with respect to  $t$ . It can be shown that  $i_p$  is actually a Lie algebra homomorphism.

**Definition 8.11** (Vertical Subspace). *Let  $(P, \pi, M)$  be a principal bundle and let  $p \in P$ . The **vertical subspace** at  $p$  is the vector subspace of  $T_p P$  given by:*

$$\begin{aligned} V_p P &:= \ker((\pi_*)_p) \\ &= \{X_p \in T_p P \mid (\pi_*)_p(X_p) = 0\} \end{aligned}$$

**Lemma 8.2.** *For all  $A_e \in T_e G$  and  $p \in P$ , we have  $X_p^{A_e} \in V_p P$ .*

*Proof.*

Since the action of  $G$  simply permutes the elements within each fiber, we have:

$$\pi(p) = \pi(p \triangleleft \exp(tA_e))$$

for any  $t$ . Let  $f \in C^\infty(M)$  be arbitrary. Then:

$$\begin{aligned} (\pi_*)_p X_p^{A_e}(f) &= X_p^{A_e}(f \circ \pi) \\ &= [(f \circ \pi)(p \triangleleft \exp(tA_e))]'(0) \\ &= [f(\pi(p))]'(0) \\ &= 0 \end{aligned}$$

since  $f(\pi(p))$  is constant. Hence  $X_p^{A_e} \in V_p P$ . Alternatively, one can also argue that  $(\pi_*)_p X_p^{A_e}$  is the tangent vector to a constant curve on  $M$ .  $\square$

In particular, the map  $i_p: T_e G \xrightarrow{\sim} V_p P$  is now a bijection. The idea of a connection is to make a choice of how to “connect” the individual points in “neighbouring” fibers in a principal fiber bundle.

**Definition 8.12** (Horizontal Subspace). *Let  $(P, \pi, M)$  be a principal bundle and let  $p \in P$ . A **horizontal subspace** at  $p$  is a vector subspace  $H_p P$  of  $T_p P$  which is complementary to  $V_p P$ , i.e.:*

$$T_p P = H_p P \oplus V_p P$$

The choice of horizontal space at  $p \in P$  is not unique. However, once a choice is made, there is a unique decomposition of each  $X_p \in T_p P$  as:

$$X_p = \text{hor}(X_p) + \text{ver}(X_p)$$

with  $\text{hor}(X_p) \in H_p P$  and  $\text{ver}(X_p) \in V_p P$ .

**Definition 8.13** (Connection). *A **connection** on a principal  $G$ -bundle  $(P, \pi, M)$  is a choice of horizontal space at each  $p \in P$  such that:*

i) *For all  $g \in G$ ,  $p \in P$  and  $X_p \in H_p P$ , we have:*

$$(\triangleleft g)_* X_p \in H_{p \triangleleft g} P$$

*where  $(\triangleleft g)_*$  is the push-forward of the map  $(-\triangleleft g): P \rightarrow P$  and it is a bijection.*

ii) *For every smooth vector field  $X \in \Gamma(TP)$ , the two summands in the unique decomposition:*

$$X(p) = \text{hor}(X(p)) + \text{ver}(X(p))$$

*at each  $p \in P$ , extend to smooth  $\text{hor}(X), \text{ver}(X) \in \Gamma(TM)$ .*

The definition formalises the idea that the assignment of an  $H_p P$  to each  $p \in P$  should be “smooth” within each fiber (i) as well as between different fibers (ii).

*Remark 8.10.* For each  $X_p \in T_p P$ , both  $\text{hor}(X_p)$  and  $\text{ver}(X_p)$  depend on the choice of  $H_p P$ .

### 8.4.1 Connection One-Forms

Technically, the choice of a horizontal subspace  $H_p P$  at each  $p \in P$  providing a connection is conveniently encoded in the thus induced Lie-algebra-valued one-form:

$$\begin{aligned} \omega_p: T_p P &\xrightarrow{\sim} T_e G \\ X_p &\mapsto \omega_p(X_p) := i_p^{-1}(\text{ver}(X_p)) \end{aligned}$$

**Definition 8.14** (Connection One-Form). *The map  $\omega: p \rightarrow \omega_p$  sending each  $p \in P$  to the  $T_e G$ -valued one-form  $\omega_p$  is called the **connection one-form** with respect to the connection.*

*Remark 8.11.* We have seen how to produce a one-form from a choice of horizontal spaces (i.e. a connection). The choice of horizontal spaces can be recovered from  $\omega$  by:

$$H_p P = \ker(\omega_p)$$

Of course, not every (Lie-algebra-valued) one-form on  $P$  is such that  $\ker(\omega_p)$  gives a connection on the principal bundle. What we would now like to do is to study some crucial properties of  $\omega$ . We will then elevate these properties to a definition of connection one-form absent a connection, so that we may re-define the notion of connection in terms of a connection one-form.

**Lemma 8.3.** *For all  $p \in P$ ,  $g \in G$  and  $A_e \in T_e G$ , we have:*

$$(\triangleleft g)_p X_p^{A_e} = X_{p \triangleleft g}^{(\text{Ad}_{g^{-1}})_e A_e}$$

*Proof.*

Let  $f \in \mathcal{C}^\infty(P)$  be arbitrary. We have:

$$\begin{aligned} (\triangleleft g)_p X_p^{A_e}(f) &= X_p^{A_e}(f \circ (- \triangleleft g)) \\ &= [f(p \triangleleft \exp(tA_e) \triangleleft g)]'(0) \\ &= [f(p \triangleleft g \triangleleft g^{-1} \triangleleft \exp(tA_e) \triangleleft g)]'(0) \\ &= [f(p \triangleleft g \triangleleft (g^{-1} \bullet \exp(tA_e) \bullet g))](0) \\ &= [f(p \triangleleft g \triangleleft \text{Ad}_{g^{-1}}(\exp(tA_e)))]'(0) \\ &= [f(p \triangleleft g \triangleleft \exp(t(\text{Ad}_{g^{-1}})_e A_e))](0) \\ &= X_{p \triangleleft g}^{(\text{Ad}_{g^{-1}})_e A_e}(f) \end{aligned}$$

which is what we wanted. □

**Theorem 8.3.** *A connection one-form  $\omega$  with respect to a connection satisfies:*

a) *For all  $p \in P$ , we have  $\omega_p(X_p^A) = A$ , that is  $\omega_p \circ i_p = \text{id}_{T_e G}$ .*

$$\begin{array}{ccc} T_e G & \xrightarrow{i_p} & V_p P \\ & \searrow \text{id}_{T_e G} & \downarrow \omega_p|_{V_p P} \\ & & T_e G \end{array}$$

b)  *$((\triangleleft g)^* \omega)|_p(X_p) = (\text{Ad}_{g^{-1}})_*(\omega_p(X_p))$*

$$\begin{array}{ccc} T_p P & \xrightarrow{\omega_p} & T_e G \\ & \searrow ((\triangleleft g)^* \omega)|_p & \downarrow (\text{Ad}_{g^{-1}})_* \\ & & T_e G \end{array}$$

c)  *$\omega$  is a smooth one-form.*

*Proof.*

a) Since  $X_p^A \in V_p P$ , by definition of  $\omega$  we have

$$\omega_p(X_p^A) := i_p^{-1}(\text{ver}(X_p^A)) = i_p^{-1}(X_p^A) = A.$$

b) First observe that the left hand side is linear in  $X_p$ . Consider the two cases

b.1) Suppose that  $X_p \in V_p P$ . Then  $X_p = X_p^A$  for some  $A \in T_e G$ . Hence

$$\begin{aligned} ((\triangleleft g)^* \omega)|_p(X_p^A) &= \omega_{p \triangleleft g}((\triangleleft g)_* X_p^A) \\ &= \omega_{p \triangleleft g}(X_{p \triangleleft g}^{(\text{Ad}_{g^{-1}})^* A}) \\ &= (\text{Ad}_{g^{-1}})_* A \\ &= (\text{Ad}_{g^{-1}})_*(\omega_p(X_p^A)) \end{aligned}$$

b.2) Suppose now that  $X_p \in H_p P = \ker(\omega_p)$ . Then

$$((\triangleleft g)^* \omega)|_p(X_p) = \omega_{p \triangleleft g}((\triangleleft g)_* X_p) = 0$$

since  $(\triangleleft g)_* X_p \in H_{p \triangleleft g} P = \ker(\omega_{p \triangleleft g})$ .

Let  $X_p \in T_p P$ . We have

$$\begin{aligned} ((\triangleleft g)^* \omega)|_p(X_p) &= ((\triangleleft g)^* \omega)|_p(\text{ver}(X_p) + \text{hor}(X_p)) \\ &= ((\triangleleft g)^* \omega)|_p(\text{ver}(X_p)) + ((\triangleleft g)^* \omega)|_p(\text{hor}(X_p)) \\ &= (\text{Ad}_{g^{-1}})_*(\omega_p(\text{ver}(X_p))) + 0 \\ &= (\text{Ad}_{g^{-1}})_*(\omega_p(\text{ver}(X_p))) + (\text{Ad}_{g^{-1}})_*(\omega_p(\text{hor}(X_p))) \\ &= (\text{Ad}_{g^{-1}})_*(\omega_p(\text{ver}(X_p) + \text{hor}(X_p))) \\ &= (\text{Ad}_{g^{-1}})_*(\omega_p(X_p)) \end{aligned}$$

c) We have  $\omega = i^{-1} \circ \text{ver}$  and both  $i^{-1}$  and  $\text{ver}$  are smooth. □

## 8.5 Yang-Mills Fields & Local Representations

We have seen how to associate a connection one-form to a connection, i.e. a certain Lie-algebra-valued one-form to a smooth choice of horizontal spaces on the principal bundle. We will now study how we can express this connection one-form locally on the base manifold of the principal bundle.

Recall that a connection one-form on a principal bundle  $(P, \pi, M)$  is a smooth Lie-algebra-valued one-form, i.e. a smooth map

$$\omega: \Gamma(TP) \xrightarrow{\sim} T_e G$$

which “behaves like a one-form”, in the sense that it is  $\mathbb{R}$ -linear and satisfies the Leibniz rule, and such that, in addition, for all  $A \in T_e G$ ,  $g \in G$  and  $X \in \Gamma(TP)$ , we have

- i)  $\omega(X^A) = A$ ;
- ii)  $((\triangleleft g)^* \omega)(X) = (\text{Ad}_{g^{-1}})_*(\omega(X))$ .

If the pair  $(u, f)$  is a principal bundle automorphism of  $(P, \pi, M)$ , i.e. if the diagram

$$\begin{array}{ccc} P & \xrightarrow{u} & P \\ \triangleleft G \uparrow & & \uparrow \triangleleft G \\ P & \xrightarrow{u} & P \\ \pi \downarrow & & \downarrow \pi \\ M & \xrightarrow{f} & M \end{array}$$

commutes, we should be able to pull a connection one-form  $\omega$  on  $P$  back to another connection one-form

$u^*\omega$  on  $P$ .

$$\begin{array}{ccc} \Gamma(TP) & \xrightarrow{\omega} & T_e G \\ \uparrow u & \nearrow u^*\omega & \\ \Gamma(TP) & & \end{array}$$

Recall that for a one-form  $\omega: \Gamma(TN) \xrightarrow{\sim} \mathcal{C}^\infty(N)$ , we defined

$$\begin{aligned} \Phi^*(\omega): \Gamma(TM) &\xrightarrow{\sim} \mathcal{C}^\infty(M) \\ X &\mapsto \omega(\Phi_*(X)) \circ \phi \end{aligned}$$

for any diffeomorphism  $\phi: M \rightarrow N$ . One might be worried about whether this and similar definitions apply to Lie-algebra-valued one-forms but, in fact, they do. In our case, even though  $\omega$  lands in  $T_e G$ , its domain is still  $\Gamma(TP)$  and if  $u: P \rightarrow P$  is a diffeomorphism of  $P$ , then  $u_*X \in \Gamma(TP)$  and so

$$u^*\omega: X \mapsto (u^*\omega)(X) := \omega(u_*(X)) \circ u$$

is again a Lie-algebra-valued one-form. Note that we will no longer distinguish notationally between the push-forward of tangent vectors and of vector fields.

In practice, e.g. for calculational purposes, one may wish to restrict attention to some open subset  $U \subseteq M$ . Let  $\sigma: U \rightarrow P$  be a local section of  $P$ , i.e.  $\pi \circ \sigma = \text{id}_U$ .

$$\begin{array}{c} P \\ \uparrow \triangleleft G \\ P \\ \downarrow \pi \quad \curvearrowright \sigma \\ U \end{array}$$

**Definition 8.15.** *Given a connection one-form  $\omega$  on  $P$ , such a local section  $\sigma$  induces*

i) a Yang-Mills field  $\omega^U: \Gamma(TU) \xrightarrow{\sim} T_e G$  given by

$$\omega^U := \sigma^*\omega;$$

ii) a local trivialisation of the principal bundle  $P$ , i.e. a map

$$\begin{aligned} h: U \times G &\rightarrow P \\ (m, g) &\mapsto \sigma(m) \triangleleft g; \end{aligned}$$

iii) a local representation of  $\omega$  on  $U$  by

$$h^*\omega: \Gamma(T(U \times G)) \xrightarrow{\sim} T_e G.$$

Note that, at each point  $(m, g) \in U \times G$ , we have

$$T_{(m,g)}(U \times G) \cong_{\text{Lie alg}} T_m U \oplus T_g G.$$

*Remark 8.12.* Both the Yang-Mills field  $\omega^U$  and the local representation  $h^*\omega$  encode the information carried by  $\omega$  locally on  $U$ . Since  $h^*\omega$  involves  $U \times G$  while  $\omega^U$  doesn't, one might guess that  $h^*\omega$  gives a more “accurate” picture of  $\omega$  on  $U$  than the Yang-Mills field. But in fact, this is not the case. They both contain the same amount of local information about the connection one-form  $\omega$ .

### 8.5.1 The Maurer-Cartan Form

The relation between the Yang-Mills field and the local representation is provided by the following result.

**Theorem 8.4.** *For all  $v \in T_m U$  and  $\gamma \in T_g G$ , we have*

$$(h^* \omega)_{(m,g)}(v, \gamma) = (\text{Ad}_{g^{-1}})_*(\omega^U(v)) + \Xi_g(\gamma),$$

where  $\Xi_g$  is the Maurer-Cartan form

$$\begin{aligned} \Xi_g: T_g G &\xrightarrow{\sim} T_e G \\ L_g^A &\mapsto A. \end{aligned}$$

*Remark 8.13.* Note that we have represented a generic element of  $T_g G$  as  $L_g^A$ . This is due to the following. Recall that the left translation map  $\ell_g: G \rightarrow G$  is a diffeomorphism of  $G$ . As such, its push-forward at any point is a linear isomorphism. In particular, we have

$$((\ell_g)_*)_e: T_e G \xrightarrow{\sim} T_g G,$$

that is, the tangent space at any point  $g \in G$  can be canonically identified with the tangent space at the identity. Hence, we can write any element of  $T_g G$  as

$$L_g^A := ((\ell_g)_*)_e(A)$$

for some  $A \in T_e G$ .

Let us consider some specific examples.

*Example 8.1.*

Any chart  $(U, x)$  of a smooth manifold  $M$  induces a local section  $\sigma: U \rightarrow LM$  of the frame bundle of  $M$  by

$$\sigma(m) := \left( \left( \frac{\partial}{\partial x^1} \right)_m, \dots, \left( \frac{\partial}{\partial x^{\dim M}} \right)_m \right) \in L_m M.$$

Since  $\text{GL}(\dim M, \mathbb{R})$  can be identified with an open subset of  $\mathbb{R}^{(\dim M)^2}$ , we have

$$T_e \text{GL}(\dim M, \mathbb{R}) \cong_{\text{Lie alg}} \mathbb{R}^{(\dim M)^2},$$

where  $\mathbb{R}^{(\dim M)^2}$  is understood as the algebra of  $\dim M \times \dim M$  square matrices, with bracket induced by matrix multiplication. In fact, this holds for any open subset of a vector space, when considered as a smooth manifold. A connection one-form

$$\omega: \Gamma(LM) \xrightarrow{\sim} T_e \text{GL}(\dim M, \mathbb{R})$$

can thus be given in terms of  $(\dim M)^2$  functions

$$\omega^i_j: \Gamma(LM) \xrightarrow{\sim} \mathbb{R}, \quad 1 \leq i, j \leq \dim M.$$

The associated Yang-Mills field  $\omega^U := \sigma^* \omega$  is, at each point  $m \in U$ , a Lie-algebra-valued one-form on the vector space  $T_m U$ . By using the co-ordinate induced basis and its dual basis, we can express  $(\omega^U)_m$  in terms of components as

$$(\omega^U)_m = \omega_\mu^U(m) (dx^\mu)_m,$$

where  $1 \leq \mu \leq \dim M$  and

$$\omega_\mu^U(m) := (\omega^U)_m \left( \left( \frac{\partial}{\partial x^\mu} \right)_m \right).$$

Since  $(\omega^U)_m: T_m U \xrightarrow{\sim} T_e G$ , we have  $\omega_\mu^U: U \rightarrow T_e G$ . Hence, by employing the same isomorphism as above, we can identify each  $\omega_\mu^U(m)$  with a square  $\dim M \times \dim M$  matrix and define the symbol

$$\Gamma^i_{j\mu}(m) := (\omega^U(m))^i_{j\mu} := (\omega_\mu^U(m))^i_j,$$

usually referred to as the *Christoffel symbol*. The middle term is just an alternative notation for the

right-most side. Note that, even though all three indices  $i, j, \mu$  run from 1 to  $\dim M$ , the numbers  $\Gamma_{j\mu}^i(m)$  do not constitute the components of a  $(1,2)$ -tensor on  $U$ . Only the  $\mu$  index transforms as a one-form component index, i.e.

$$((g \triangleright \omega^U(m))^i_j)_\mu = (g^{-1})^\nu (\omega^U(m))^i_{j\nu}$$

for  $g \in \mathrm{GL}(\dim M, \mathbb{R})$ , while the  $i, j$  indices simply label different one-forms,  $(\dim M)^2$  in total.

Note that the Maurer-Cartan form appearing in Theorem 8.4 only depends on the Lie group (and its Lie algebra), not on the principal bundle  $P$  or the restriction  $U \subseteq M$ . In the following example, we will go through the explicit calculation of the Maurer-Cartan form of the Lie group  $\mathrm{GL}(d, \mathbb{R})$ .

*Example 8.2.*

Let  $(\mathrm{GL}^+(d, \mathbb{R}), x)$  be a chart on  $\mathrm{GL}(d, \mathbb{R})$ , where  $\mathrm{GL}^+(d, \mathbb{R})$  denotes an open subset of  $\mathrm{GL}(d, \mathbb{R})$  containing the identity  $\mathrm{id}_{\mathbb{R}^d}$ , and let  $x^i_j: \mathrm{GL}^+(d, \mathbb{R}) \rightarrow \mathbb{R}$  denote the corresponding co-ordinate functions

$$\begin{array}{ccc} \mathrm{GL}^+(d, \mathbb{R}) & \xrightarrow{x} & x(\mathrm{GL}^+(d, \mathbb{R})) \subseteq \mathbb{R}^{d^2} \\ & \searrow x^i_j & \downarrow \mathrm{proj}^i_j \\ & & \mathbb{R} \end{array}$$

so that  $x^i_j(g) := g^i_j$ . Recall that the co-ordinate functions are smooth maps on the chart domain, i.e. we have  $x^i_j \in \mathcal{C}^\infty(\mathrm{GL}^+(d, \mathbb{R}))$ . Also recall that to each  $A \in T_{\mathrm{id}_{\mathbb{R}^d}} \mathrm{GL}(d, \mathbb{R})$  there is associated a left-invariant vector field

$$L^A: \mathcal{C}^\infty(\mathrm{GL}^+(d, \mathbb{R})) \xrightarrow{\sim} \mathcal{C}^\infty(\mathrm{GL}^+(d, \mathbb{R}))$$

which, at each point  $g \in \mathrm{GL}(d, \mathbb{R})$ , is the tangent vector to the curve

$$\gamma^A(t) := g \bullet \exp(tA).$$

Consider the action of  $L^A$  on the co-ordinate functions:

$$\begin{aligned} (L^A x^i_j)(g) &= [x^i_j(g \bullet \exp(tA))]'(0) \\ &= [x^i_j(g \bullet e^{tA})]'(0) \\ &= (g^i_k (e^{tA})^k_j)'(0) \\ &= g^i_k A^k_j, \end{aligned}$$

where we have used the fact that for a matrix Lie group, the exponential map is just the ordinary exponential

$$\exp(A) = e^A := \sum_{n=0}^{\infty} \frac{1}{n!} A^n.$$

Hence, we can write

$$L^A|_g = g^i_k A^k_j \left( \frac{\partial}{\partial x^i_j} \right)_g$$

from which we can read-off the Maurer-Cartan form of  $\mathrm{GL}(d, \mathbb{R})$

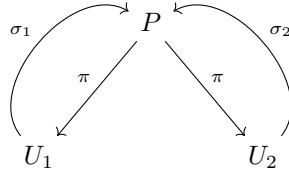
$$(\Xi_g)^i_j := (g^{-1})^i_k (dx^k_j)_g.$$

Indeed, we can quickly check that

$$\begin{aligned}
(\Xi_g)^i_j(L^A) &= (g^{-1})^i_k(dx^k_j)_g \left( g^p_r A^r_q \left( \frac{\partial}{\partial x^p_q} \right)_g \right) \\
&= (g^{-1})^i_k g^p_r A^r_q \delta^k_p \delta^q_j \\
&= (g^{-1})^i_p g^p_r A^r_j \\
&= \delta^i_r A^r_j \\
&= A^i_j.
\end{aligned}$$

### 8.5.2 The Gauge Map

In physics, we are often prompted to write down a Yang-Mills field because we have local information about a connection. We can then try to reconstruct the global connection by glueing the Yang-Mills fields on several open subsets of our manifold.



Suppose, for instance, that we have two open subsets  $U_1, U_2 \subseteq M$  and consider the Yang-Mills fields associated to two local connections  $\sigma_1, \sigma_2$ . If  $\omega^{U_1}$  and  $\omega^{U_2}$  are both local versions of a unique connection one-form, then is  $U_1 \cap U_2 \neq \emptyset$ , the Yang-Mills fields  $\omega^{U_1}$  and  $\omega^{U_2}$  should satisfy some compatibility condition on  $U_1 \cap U_2$ .

**Definition 8.16.** *Within the above set-up, the gauge map is the map*

$$\Omega: U_1 \cap U_2 \rightarrow G$$

where, for each  $m \in U_1 \cap U_2$ , the Lie group element  $\Omega(m) \in G$  satisfies

$$\sigma_2(m) = \sigma_1(m) \triangleleft \Omega(m).$$

Note that since the  $G$ -action  $\triangleleft$  on  $P$  is free, for each  $m$  there exists a unique  $\Omega(m)$  satisfying the above condition, and hence the gauge map  $\Omega$  is well-defined.

**Theorem 8.5.** *Under the above assumptions, we have*

$$(\omega^{U_2})_m = (\text{Ad}_{\Omega^{-1}(m)})_*(\omega^{U_1}) + (\Omega^* \Xi_g)_m.$$

*Example 8.3.*

Consider again the frame bundle  $LM$  of some manifold  $M$ . Let us evaluate explicitly the pull-back along  $\Omega$  of the Maurer-Cartan form. Since  $\Xi_g: T_g G \rightarrow T_e G$  and  $\Omega: U_1 \cap U_2 \rightarrow T_e G$ , we have  $\Omega^* \Xi_g: T(U_1 \cap U_2) \rightarrow T_e G$ . Let  $x$  be a chart map near the point  $m \in U_1 \cap U_2$ . We have

$$\begin{aligned}
((\Omega^* \Xi_g)_m)^i_j \left( \left( \frac{\partial}{\partial x^\mu} \right)_m \right) &= (\Xi_{\Omega(m)})^i_j \left( \Omega_* \left( \frac{\partial}{\partial x^\mu} \right)_m \right) \\
&= (\Omega(m)^{-1})^i_k (d\tilde{x}^k_j)_{\Omega(m)} \left( \Omega_* \left( \frac{\partial}{\partial x^\mu} \right)_m \right) \\
&= (\Omega(m)^{-1})^i_k \left( \Omega_* \left( \frac{\partial}{\partial x^\mu} \right)_m \right) (\tilde{x}^k_j) \\
&= (\Omega(m)^{-1})^i_k \left( \frac{\partial}{\partial x^\mu} \right)_m (\tilde{x}^k_j \circ \Omega) \\
&= (\Omega(m)^{-1})^i_k \left( \frac{\partial}{\partial x^\mu} \right)_m (\Omega(m))^k_j.
\end{aligned}$$

hence, we can write

$$\begin{aligned} ((\Omega^* \Xi_g)_m)^i_j &= (\Omega(m)^{-1})^i_k \left( \frac{\partial}{\partial x^\mu} \right)_m (\Omega(m))^k_j dx^\mu \\ &=: (\Omega^{-1} d\Omega)^i_j. \end{aligned}$$

Let us now compute the other summand. Recall that  $\text{Ad}_g$  is the map

$$\begin{aligned} \text{Ad}_g: G &\rightarrow G \\ h &\mapsto g \bullet h \bullet g^{-1} \end{aligned}$$

and since  $\text{Ad}_g(e) = e$ , the push-forward  $((\text{Ad}_g)_*)_e: T_e G \xrightarrow{\sim} T_e G$  is a linear endomorphism of  $T_e G$ . Moreover, since here  $G = \text{GL}(d, \mathbb{R})$  is a matrix Lie group, we have

$$((\text{Ad}_g)_* A)^i_j = g^i_k A^k_l (g^{-1})^l_j =: (g A g^{-1})^i_j.$$

Hence, we have

$$(\text{Ad}_{\Omega^{-1}(m)})_*(\omega^{U_1}) = (\Omega(m)^{-1})^i_k (\omega^{U_1})^k_l (\Omega(m))^l_j$$

Altogether, we find that the transition rule for the Yang-Mills fields on the intersection of  $U_1$  and  $U_2$  is given by

$$(\omega^{U_2})^i_{j\mu} = (\Omega^{-1})^i_k (\omega^{U_1})^k_{l\mu} \Omega^l_j + (\Omega^{-1})^i_k \partial_\mu (\Omega^{-1})^k_j.$$

As an application, consider the spacial case in which the sections  $\sigma_1, \sigma_2$  are induced by co-ordinate charts  $(U_1, x)$  and  $(U_2, y)$ . Then we have

$$\begin{aligned} \Omega^i_j &= \frac{\partial y^i}{\partial x^j} := \partial_j (y^i \circ x^{-1}) \circ x \\ (\Omega^{-1})^i_j &= \frac{\partial x^i}{\partial y^j} := \partial_j (x^i \circ y^{-1}) \circ y \end{aligned}$$

and hence

$$(\omega^{U_2})^i_{j\nu} = \frac{\partial y^\mu}{\partial x^\nu} \left( \frac{\partial x^i}{\partial y^k} (\omega^{U_1})^k_{l\mu} \frac{\partial y^l}{\partial x^j} + \frac{\partial x^i}{\partial y^k} \frac{\partial^2 y^k}{\partial x^\mu \partial x^j} \right).$$

You may recognise this as the transformation law for the Christoffel symbols from general relativity.

## 8.6 Parallel Transport

We now come to the second term in the sequence “connection, parallel transport, covariant derivative”. The idea of parallel transport on a principal bundle hinges on that of horizontal lift of a curve on the base manifold, which is a lifting to a curve on the principal bundle in the sense that the projection to the base manifold of this curve gives the curve we started with. In particular, if the principal bundle is equipped with a connection, we would like to impose some extra conditions on this lifting, so that it “connects” nearby fibers in a nice way. We will then consider the same idea on an associated bundle and see how we can induce a derivative operator if the associated bundle is a vector bundle.

### 8.6.1 Horizontal Lifts To the principal bundle

**Definition 8.17.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle equipped with a connection and let  $\gamma: [0, 1] \rightarrow M$  be a curve on  $M$ . The horizontal lift of  $\gamma$  through  $p_0 \in P$  is the unique curve

$$\gamma^\uparrow: [0, 1] \rightarrow P$$

with  $\gamma^\uparrow(0) = p_0 \in \text{preim}_\pi(\{\gamma(0)\})$  satisfying

$$i) \quad \pi \circ \gamma^\uparrow = \gamma;$$

$$ii) \quad \forall \lambda \in [0, 1] : \text{ver}(X_{\gamma^\uparrow, \gamma^\uparrow(\lambda)}) = 0;$$

$$iii) \quad \forall \lambda \in [0, 1] : \pi_*(X_{\gamma^\uparrow, \gamma^\uparrow(\lambda)}) = X_{\gamma, \gamma(\lambda)}.$$

Intuitively, a horizontal lift of a curve  $\gamma$  on  $M$  is a curve  $\gamma^\uparrow$  on  $P$  such that each point  $\gamma^\uparrow(\lambda) \in P$  belongs to the fiber of  $\gamma(\lambda)$  (condition i), the tangent vectors to the curve  $\gamma^\uparrow$  have no vertical component (condition ii), i.e. they lie entirely in the horizontal spaces at each point, and finally the projection of the tangent vector to  $\gamma^\uparrow$  at  $\gamma^\uparrow(\lambda)$  coincides with the tangent vector to the curve  $\gamma$  at  $\pi(\gamma^\uparrow(\lambda)) = \gamma(\lambda)$ .

*Remark 8.14.* Note that the uniqueness in the above definition only stems from the choice of  $p_0 \in \text{preim}_\pi(\{\gamma(0)\})$ . A curve on  $M$  has several horizontal lifts to a curve on  $P$ , but there is only one such curve going through each point  $p_0 \in \text{preim}_\pi(\{\gamma(0)\})$ . Clearly, different horizontal lifts cannot intersect each other.

Our strategy to write down an explicit expression for the horizontal lift through  $p_0 \in P$  of a curve  $\gamma: [0, 1] \rightarrow M$  is to proceed in two steps:

- i) “Generate” the horizontal lift by starting from some arbitrary curve  $\delta: [0, 1] \rightarrow P$  such that  $\pi \circ \delta = \gamma$  by action of a suitable curve  $g: (0, 1) \rightarrow G$  so that

$$\gamma^\uparrow(\lambda) = \delta(\lambda) \triangleleft g(\lambda).$$

The suitable curve  $g$  will be the solution to an ordinary differential equation with initial condition  $g(0) = g_0$ , where  $g_0$  is the unique element in  $G$  such that

$$\delta(0) \triangleleft g_0 = p_0 \in P.$$

- ii) We will explicitly solve (locally) this differential equation for  $g: [0, 1] \rightarrow P$  by a path-ordered integral over the local Yang-Mills field.

We have the following result characterising the curve  $g$  appearing above.

**Theorem 8.6.** *The (first order) ODE satisfied by the curve  $g: [0, 1] \rightarrow G$  is*

$$(\text{Ad}_{g(\lambda)}^{-1})_*(\omega_{\delta(\lambda)}(X_{\delta, \delta(\lambda)})) + \Xi_{g(\lambda)}(X_{g, g(\lambda)}) = 0$$

with the initial condition  $g(0) = g_0$ .

**Corollary 8.1.** *If  $G$  is a matrix group, then the above ODE takes the form*

$$g(\lambda)^{-1}(\omega_{\delta(\lambda)}(X_{\delta, \delta(\lambda)}))g(\lambda) + g(\lambda)^{-1}\dot{g}(\lambda) = 0$$

where we denoted matrix multiplication by juxtaposition and  $\dot{g}(\lambda)$  denotes the derivative with respect to  $\lambda$  of the matrix entries of  $g$ . Equivalently, by multiplying both sides on the left by  $g(\lambda)$ ,

$$\dot{g}(\lambda) = -(\omega_{\delta(\lambda)}(X_{\delta, \delta(\lambda)}))g(\lambda).$$

In order to further massage this ODE, let us consider a chart  $(U, x)$  on the base manifold  $M$ , such that the image of  $\gamma$  is entirely contained in  $U$ . A local section  $\sigma: U \rightarrow P$  induces

- i) a Yang-Mills field  $\omega^U$ ;
- ii) a curve on  $P$  by  $\delta := \sigma \circ \gamma$ .

In fact, since the only condition imposed on  $\delta$  is that  $\pi \circ \delta = \gamma$ , choosing a such a curve  $\delta$  is equivalent to choosing a local section  $\sigma$ . Note that we have

$$\sigma_*(X_{\gamma, \gamma(\lambda)}) = X_{\delta, \delta(\lambda)},$$

and hence

$$\begin{aligned}
\omega_{\delta(\lambda)}(X_{\delta,\delta(\lambda)}) &= \omega_{\delta(\lambda)}(\sigma_*(X_{\gamma,\gamma(\lambda)})) \\
&= (\sigma^*\omega)_{\gamma(\lambda)}(X_{\gamma,\gamma(\lambda)}) \\
&= (\omega^U)_{\gamma(\lambda)}(X_{\gamma,\gamma(\lambda)}) \\
&= \omega_\mu^U(\gamma(\lambda))(dx^\mu)_{\gamma(\lambda)} \left( X_\gamma^\nu(\gamma(\lambda)) \left( \frac{\partial}{\partial x^\nu} \right)_{\gamma(\lambda)} \right) \\
&= \omega_\mu^U(\gamma(\lambda)) X_\gamma^\nu(\gamma(\lambda)) (dx^\mu)_{\gamma(\lambda)} \left( \left( \frac{\partial}{\partial x^\nu} \right)_{\gamma(\lambda)} \right) \\
&= \omega_\mu^U(\gamma(\lambda)) X_\gamma^\nu(\gamma(\lambda)) \delta_\nu^\mu \\
&= \omega_\mu^U(\gamma(\lambda)) X_\gamma^\mu(\gamma(\lambda)).
\end{aligned}$$

Thus, in the special case of a matrix Lie group, the ODE reads

$$\dot{g}(\lambda) = -\Gamma_\mu(\gamma(\lambda)) \dot{\gamma}^\mu(\lambda),$$

where  $\Gamma_\mu := \omega_\mu^U$  and  $\dot{\gamma}^\mu(\lambda) := X_\gamma^\mu(\gamma(\lambda))$ , together with the initial condition  $g(0) = g_0$ .

### 8.6.2 Solution of the horizontal lift ODE by a path-ordered exponential

As a first step towards the solution of our ODE, consider

$$g(t) := g_0 - \int_0^t d\lambda \Gamma_\mu(\gamma(\lambda)) \dot{\gamma}^\mu(\lambda) g(\lambda).$$

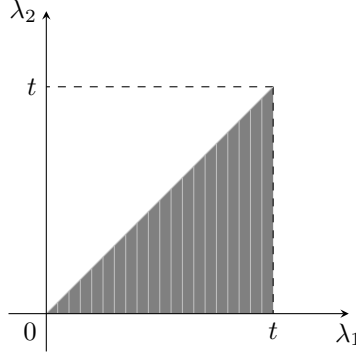
This doesn't seem to have brought us far since the function  $g$  that we would like to determine appears again on the right hand side. However, we can now iterate this definition to obtain

$$\begin{aligned}
g(t) &= g_0 - \int_0^t d\lambda_1 \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \left( g_0 - \int_0^{\lambda_1} d\lambda_2 \Gamma_\nu(\gamma(\lambda_2)) \dot{\gamma}^\nu(\lambda_2) \right) \\
&= g_0 - \int_0^t d\lambda_1 \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) g_0 + \int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \Gamma_\nu(\gamma(\lambda_2)) \dot{\gamma}^\nu(\lambda_2) g(\lambda_2).
\end{aligned}$$

Matters seem to only get worse, until one realises that the first integral no longer contains the unknown function  $g$ . Hence, the above expression provides a “first-order” approximation to  $g$ . It is clear that we can get higher-order approximations by iterating this process

$$\begin{aligned}
g(t) &= g_0 - \int_0^t d\lambda_1 \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) g_0 \\
&\quad + \int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \Gamma_\nu(\gamma(\lambda_2)) \dot{\gamma}^\nu(\lambda_2) g_0 \\
&\quad \vdots \\
&\quad + (-1)^n \int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \cdots \int_0^{\lambda_{n-1}} d\lambda_n \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \cdots \Gamma_\nu(\gamma(\lambda_n)) \dot{\gamma}^\nu(\lambda_n) g(\lambda_n).
\end{aligned}$$

Note how the range of each integral depends on the integration variable of the previous integral. It would much nicer if we could have the same range in each integral. In fact, there is a standard trick to achieve this. The region of integration in the double integral is



and if the integrand  $f(\lambda_1, \lambda_2)$  is invariant under the exchange  $\lambda_1 \leftrightarrow \lambda_2$ , we have

$$\int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 f(\lambda_1, \lambda_2) = \frac{1}{2} \int_0^t d\lambda_1 \int_0^t d\lambda_2 f(\lambda_1, \lambda_2).$$

Generalising to  $n$  dimensions, we have

$$\int_0^t d\lambda_1 \cdots \int_0^{\lambda_{n-1}} d\lambda_n f(\lambda_1, \dots, \lambda_n) = \frac{1}{n!} \int_0^t d\lambda_1 \cdots \int_0^t d\lambda_n f(\lambda_1, \dots, \lambda_n)$$

if  $f$  is invariant under any permutation of its arguments. Moreover, since each term in our integrands only depends on one integration variable at a time, we can use

$$\int_0^t d\lambda_1 \cdots \int_0^t d\lambda_n f_1(\lambda_1) \cdots f_n(\lambda_n) = \left( \int_0^t d\lambda_1 f_1(\lambda_1) \right) \cdots \left( \int_0^t d\lambda_n f_n(\lambda_n) \right)$$

so that, in our case, we would have

$$\begin{aligned} g(t) &= \left( \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left( \int_0^t d\lambda \Gamma_{\mu}(\gamma(\lambda)) \dot{\gamma}^{\mu}(\lambda) \right)^n \right) g_0 \\ &= \exp \left( - \int_0^t d\lambda \Gamma_{\mu}(\gamma(\lambda)) \dot{\gamma}^{\mu}(\lambda) \right) g_0. \end{aligned}$$

However, our integrands are Lie-algebra-valued (that is, matrix valued), and since the factors therein need not commute, they are not invariant under permutations of the independent variables. Hence, the above formula doesn't work. Instead, we write

$$g(t) = \text{P exp} \left( - \int_0^t d\lambda \Gamma_{\mu}(\gamma(\lambda)) \dot{\gamma}^{\mu}(\lambda) \right) g_0,$$

where the *path-ordered exponential* P exp is defined to yield the correct expression for  $g(t)$ .

Summarising, we have the following.

**Proposition 8.2.** *For a principal  $G$ -bundle  $(P, \pi, M)$ , where  $G$  is a matrix Lie group, the horizontal lift of a curve  $\gamma: [0, 1] \rightarrow U$  through  $p_p \in \text{preim}_{\pi}(\{U\})$ , where  $(U, x)$  is a chart on  $M$ , is given in terms of a local section  $\sigma: U \rightarrow P$  by the explicit expression*

$$\gamma^{\uparrow}(\lambda) = (\sigma \circ \gamma)(\lambda) \triangleleft \left( \text{P exp} \left( - \int_0^{\lambda} d\tilde{\lambda} \Gamma_{\mu}(\gamma(\tilde{\lambda})) \dot{\gamma}^{\mu}(\tilde{\lambda}) \right) g_0 \right).$$

**Definition 8.18.** *Let  $\gamma_p^{\uparrow}: [0, 1] \rightarrow P$  be the horizontal lift through  $p \in \text{preim}_{\pi}(\{\gamma(0)\})$  of the curve  $\gamma: [0, 1] \rightarrow M$ . The parallel transport map along  $\gamma$  is the map*

$$\begin{aligned} T_{\gamma}: \text{preim}_{\pi}(\{\gamma(0)\}) &\rightarrow \text{preim}_{\pi}(\{\gamma(1)\}) \\ p &\mapsto \gamma_p^{\uparrow}(1). \end{aligned}$$

*Remark 8.15.* The parallel transport is, in fact, a bijection between the fibers  $\text{preim}_{\pi}(\{\gamma(0)\})$  and  $\text{preim}_{\pi}(\{\gamma(1)\})$ . It is injective since there is a unique horizontal lift of  $\gamma$  through each point  $p \in$

$\text{preim}_\pi(\{\gamma(0)\})$ , and horizontal lifts through different points do not intersect. It is surjective since for each  $q \in \text{preim}_\pi(\{\gamma(1)\})$  we can find a  $p$  such that  $q = \gamma_p^\uparrow(1)$  as follows. Let  $\tilde{p} \in \text{preim}_\pi(\{\gamma(0)\})$ . Then  $\gamma_{\tilde{p}}^\uparrow(1)$  belongs to the same fiber as  $q$  and hence there exists a unique  $g \in G$  such that  $q = \gamma_{\tilde{p}}^\uparrow(1) \triangleleft g$ . Recall that

$$\gamma_{\tilde{p}}^\uparrow(\lambda) = (\sigma \circ \gamma)(\lambda) \triangleleft (\text{P exp}(\cdots)g_0)$$

where  $g_0$  is the unique  $g_0 \in G$  such that  $\tilde{p} = (\sigma \circ \gamma)(0) \triangleleft g_0$ . Define  $p \in \text{preim}_\pi(\{\gamma(0)\})$  by

$$p := \tilde{p} \triangleleft g = (\sigma \circ \gamma)(0) \triangleleft (g_0 \bullet g).$$

Then we have

$$\begin{aligned} \gamma_p^\uparrow(1) &= (\sigma \circ \gamma)(1) \triangleleft (\text{P exp}(\cdots)g_0 \bullet g) \\ &= (\sigma \circ \gamma)(1) \triangleleft (\text{P exp}(\cdots)g_0) \triangleleft g \\ &= \gamma_{\tilde{p}}^\uparrow(1) \triangleleft g \\ &= q. \end{aligned}$$

### Loops and holonomy groups

Consider the case of loops, i.e. curves  $\gamma: [0, 1] \rightarrow M$  for which  $\gamma(0) = \gamma(1)$ . Fix some  $p \in \text{preim}_\pi(\{\gamma(0)\})$ . The condition that  $\pi \circ \gamma_p^\uparrow = \gamma$  then implies that  $\gamma_p^\uparrow(0)$  and  $\gamma_p^\uparrow(1)$  belong to the same fiber. Hence, there exists a unique  $g_\gamma \in G$  such that

$$\gamma_p^\uparrow(1) = \gamma_p^\uparrow(0) \triangleleft g_\gamma = p \triangleleft g_\gamma.$$

**Definition 8.19.** Let  $\omega$  be a connection one-form on the principal  $G$ -bundle  $(P, \pi, M)$ . Let  $\gamma: [0, 1] \rightarrow M$  be a loop with base-point  $a \in M$ , i.e.  $\gamma(0) = \gamma(1) = a$ . The subgroup of  $G$

$$\text{Hol}_a(\omega) := \{g_\gamma \mid \gamma_p^\uparrow(1) = p \triangleleft g_\gamma \text{ for some loop } \gamma\}$$

is called the holonomy group of  $\omega$  on  $P$  at the base-point  $a$ .

### 8.6.3 Horizontal lifts to the associated bundle

Almost everything that we have done so far transfers with ease to an associated bundle via the following definition.

**Definition 8.20.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle and  $\omega$  a connection one-form on  $P$ . Let  $(P_F, \pi_F, M)$  be an associated fiber bundle of  $P$  on whose typical fiber  $F$  the Lie group  $G$  acts on the left by  $\triangleright$ . Let  $\gamma: [0, 1] \rightarrow M$  be a curve on  $M$  and let  $\gamma_p^\uparrow$  be its horizontal lift to  $P$  through  $p \in \text{preim}_\pi(\{\gamma(0)\})$ . Then the horizontal lift of  $\gamma$  to the associated bundle  $P_F$  through the point  $[p, f] \in P_F$  is the curve

$$\begin{aligned} \gamma_{[p, f]}^\uparrow: [0, 1] &\rightarrow P_F \\ \lambda &\mapsto [\gamma_p^\uparrow(\lambda), f] \end{aligned}$$

For instance, we have the obvious parallel transport map.

**Definition 8.21.** The parallel transport map on the associated bundle is given by

$$\begin{aligned} T_\gamma^{P_F}: \text{preim}_{\pi_F}(\{\gamma(0)\}) &\rightarrow \text{preim}_{\pi_F}(\{\gamma(1)\}) \\ [p, f] &\mapsto \gamma_{[p, f]}^\uparrow(1). \end{aligned}$$

*Remark 8.16.* If  $F$  is a vector space and  $\triangleright: G \times F \rightarrow F$  is fiber-wise linear, i.e. for each fixed  $g \in G$ , the map  $(g \triangleright -): F \rightarrow F$  is linear, then  $(P_F, \pi_F, M)$  is called a *vector bundle*. The basic idea of a covariant derivative is as follows. Let  $\sigma: U \rightarrow P_F$  be a local section of the associated bundle. We would like to define the derivative of  $\sigma$  at the point  $m \in U \subseteq M$  in the direction  $X \in T_m M$ . By definition, there exists a curve  $\gamma: (-\varepsilon, \varepsilon) \rightarrow M$  with  $\gamma(0) = m$  such that  $X = X_{\gamma, m}$ . Then for any  $0 \leq t < \varepsilon$ , the points  $\gamma_{[\sigma(m), f]}^\uparrow(t)$  and  $\sigma(\gamma(t))$  lie in the same fiber of  $P_F$ . But since the fibers are vector spaces, we can write

the differential quotient

$$\frac{\sigma(\gamma(t)) - \gamma_{[\sigma(m), f]}^{\uparrow P_F}(t)}{t},$$

where the minus sign denotes the additive inverse in the vector space  $\text{preim}_{\pi_F}(\{\gamma(t)\})$  and hence define the derivative of  $\sigma$  at the point  $m$  in the direction  $X$ , or the derivative of  $\sigma$  along  $\gamma$  at  $\gamma(0) = m$ , by

$$\lim_{t \rightarrow 0} \frac{\sigma(\gamma(t)) - \gamma_{[\sigma(m), f]}^{\uparrow P_F}(t)}{t}$$

(of course, this makes sense as soon as we have a topology on the fibers). We will soon present a more abstract approach.

## 8.7 Covariant exterior derivative and curvature

Usually, in more elementary treatments of differential geometry or general relativity, curvature and torsion are mentioned together as properties of a covariant derivative over the tangent or the frame bundle. Since we will soon define the notion of curvature on a general principal bundle equipped with a connection, one might expect that there be a general definition of torsion on a principal bundle with a connection. However, this is not the case. Torsion requires additional structure beyond that induced by a connection. The reason why curvature and torsion are sometimes presented together is that frame bundles are already equipped, in a canonical way, with the extra structure required to define torsion.

**Definition 8.22.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle with connection one-form  $\omega$ . Let  $\phi$  be a  $k$ -form (i.e. an anti-symmetric,  $C^\infty(P)$ -multilinear map) with values in some module  $V$ . Then the exterior covariant derivative of  $\phi$  is

$$\begin{aligned} D\phi: \Gamma(TP)^{\times(k+1)} &\rightarrow V \\ (X_1, \dots, X_{k+1}) &\mapsto d\phi(\text{hor}(X_1), \dots, \text{hor}(X_{k+1})). \end{aligned}$$

**Definition 8.23.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle with connection one-form  $\omega$ . The curvature of the connection one-form  $\omega$  is the Lie-algebra-valued 2-form on  $P$

$$\Omega: \Gamma(TP) \times \Gamma(TP) \rightarrow T_e G$$

defined by

$$\Omega := D\omega.$$

For calculational purposes, we would like to make this definition a bit more explicit.

**Proposition 8.3.** Let  $\omega$  be a connection one-form and  $\Omega$  its curvature. Then

$$\Omega = d\omega + \omega \mathbin{\frown} \omega \tag{*}$$

with the second term on the right hand side defined as

$$(\omega \mathbin{\frown} \omega)(X, Y) := \llbracket \omega(X), \omega(Y) \rrbracket$$

where  $X, Y \in \Gamma(TP)$  and the double bracket denotes the Lie bracket on  $T_e G$ .

*Remark 8.17.* If  $G$  is a matrix Lie group, and hence  $T_e G$  is an algebra of matrices of the same size as those of  $G$ , then we can write

$$\Omega_j^i = d\omega_j^i + \omega_k^i \wedge \omega_j^k.$$

*Proof.*

Since  $\Omega$  is  $C^\infty$ -bilinear, it suffices to consider the following three cases.

- a) Suppose that  $X, Y \in \Gamma(TP)$  are both vertical, that is, there exist  $A, B \in T_e G$  such that  $X = X^A$  and  $Y = X^B$ . Then the left hand side of our equation reads

$$\begin{aligned}\Omega(X^A, X^B) &:= D\omega(X^A, X^B) \\ &= d\omega(\text{hor}(X^A), \text{hor}(X^B)) \\ &= d\omega(0, 0) \\ &= 0\end{aligned}$$

while the right hand side is

$$\begin{aligned}d\omega(X^A, X^B) + (\omega \mathbin{\mathbb{A}} \omega)(X^A, X^B) &= X^A(\omega(X^B)) - X^B(\omega(X^A)) \\ &\quad - \omega([X^A, X^B]) + \llbracket \omega(X^A), \omega(X^B) \rrbracket \\ &= X^A(B) - X^B(A) \\ &\quad - \omega(X^{[A, B]}) + \llbracket A, B \rrbracket \\ &= -\llbracket A, B \rrbracket + \llbracket A, B \rrbracket \\ &= 0.\end{aligned}$$

Note that we have used the fact that the map

$$\begin{aligned}i: T_e G &\rightarrow \Gamma(TP) \\ A &\mapsto X^A\end{aligned}$$

is a Lie algebra homomorphism, and hence

$$X^{[A, B]} = i(\llbracket A, B \rrbracket) = [i(A), i(B)] = [X^A, X^B],$$

where the single square brackets denote the Lie bracket on  $\Gamma(TP)$ .

- b) Suppose that  $X, Y \in \Gamma(TP)$  are both horizontal. Then we have

$$\Omega(X, Y) := D\omega(X, Y) = d\omega(\text{hor}(X), \text{hor}(Y)) = d\omega(X, Y)$$

and

$$(\omega \mathbin{\mathbb{A}} \omega)(X, Y) = \llbracket \omega(X), \omega(Y) \rrbracket = \llbracket 0, 0 \rrbracket = 0.$$

Hence the equation holds in this case.

- c) W.l.o.g suppose that  $X \in \Gamma(TP)$  is horizontal while  $Y = X^A \in \Gamma(TP)$  is vertical. Then the left hand side is

$$\Omega(X, X^A) := D\omega(X, X^A) = d\omega(\text{hor}(X), \text{hor}(X^A)) = d\omega(\text{hor}(X), 0) = 0.$$

while the right hand side gives

$$\begin{aligned}d\omega(X, X^A) + (\omega \mathbin{\mathbb{A}} \omega)(X, X^A) &= X(\omega(X^A)) - X^A(\omega(X)) \\ &\quad - \omega([X, X^A]) + \llbracket \omega(X), \omega(X^A) \rrbracket \\ &= X(A) - X^A(0) \\ &\quad - \omega(X^{[A, B]}) + \llbracket 0, A \rrbracket \\ &= -\omega([X, X^A]) \\ &= 0,\end{aligned}$$

where the only non-trivial step, which is left as an exercise, is to show that if  $X$  is horizontal and  $Y$  is vertical, then  $[X, Y]$  is again horizontal.  $\square$

We would now like to relate the curvature on a principal bundle to (local) objects on the base manifold, just like we have done for the connection one-form. Recall that a connection one-form on a principal  $G$ -bundle  $(P, \pi, M)$  is a  $T_e G$ -valued one-form  $\omega$  on  $P$ . By using the notation  $\Omega^1(P) \otimes T_e G$  for the collection (in fact, bundle) of all  $T_e G$ -valued one-forms, we have  $\omega \in \Omega^1(P) \otimes T_e G$ . If  $\sigma \in \Gamma(TU)$  is a local section on  $M$ , we defined the Yang-Mills field  $\omega^U \in \Omega^1(U) \otimes T_e G$  by pulling  $\omega$  back along  $\sigma$ .

**Definition 8.24.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle and let  $\Omega$  be the curvature associated to a connection one-form on  $P$ . Let  $\sigma \in \Gamma(TU)$  be a local section on  $M$ . Then, the two-form

$$\text{Riem} \equiv F := \sigma^* \Omega \in \Omega^2(U) \otimes T_e G$$

is called the Yang-Mills field strength.

*Remark 8.18.* Observe that the equation  $\Omega = d\omega + \omega \mathbb{A} \omega$  on  $P$  immediately gives

$$\begin{aligned} \sigma^* \Omega &= \sigma^*(d\omega + \omega \mathbb{A} \omega) \\ &= \sigma^*(d\omega) + \sigma^*(\omega \mathbb{A} \omega) \\ &= d(\sigma^* \omega) + \sigma^* \omega \mathbb{A} \sigma^* \omega. \end{aligned}$$

Since  $\text{Riem}$  is a two-form, we can write

$$\text{Riem}_{\mu\nu} = (d\omega^U)_{\mu\nu} + \omega_\mu^U \mathbb{A} \omega_\nu^U.$$

In the case of a matrix Lie group, by writing  $\Gamma_{j\mu}^i := (\omega^U)^i_{j\mu}$ , we can further express this in components as

$$\text{Riem}_{j\mu\nu}^i = \partial_\nu \Gamma_{j\mu}^i - \partial_\mu \Gamma_{j\nu}^i + \Gamma_{k\mu}^i \Gamma_{j\nu}^k - \Gamma_{k\nu}^i \Gamma_{j\mu}^k$$

from which we immediately observe that  $\text{Riem}$  is symmetric in the last two indices, i.e.

$$\text{Riem}_{j[\mu\nu]}^i = 0.$$

**Theorem 8.7** (First Bianchi identity). Let  $\Omega$  be the curvature two-form associated to a connection one-form  $\omega$  on a principal bundle. Then

$$D\Omega = 0.$$

*Remark 8.19.* Note that since  $\Omega = D\omega$ , Bianchi's identity can be rewritten as  $D^2\Omega = 0$ . However, unlike the exterior derivative  $d$ , the covariant exterior derivative does *not* satisfy  $D^2 = 0$  in general.

### 8.7.1 Torsion

**Definition 8.25.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle and let  $V$  be the representation space of a linear  $(\dim M)$ -dimensional representation of the Lie group  $G$ . A solder(ing) form on  $P$  is a one-form  $\theta \in \Omega^1(P) \otimes V$  such that

- (i)  $\forall X \in \Gamma(TP) : \theta(\text{ver}(X)) = 0$ ;
- (ii)  $\forall g \in G : g \triangleright ((\triangleleft g)^* \theta) = \theta$ ;
- (iii)  $TM$  and  $P_V$  are isomorphic as associated bundles.

A solder form provides an identification of  $V$  with each tangent space of  $M$ .

*Example 8.4.*

Consider the frame bundle  $(LM, \pi, M)$  and define

$$\begin{aligned} \theta : \Gamma(T(LM)) &\rightarrow \mathbb{R}^{\dim M} \\ X &\mapsto (u_{\pi(X)}^{-1} \circ \pi_*)(X) \end{aligned}$$

where for each  $e := (e_1, \dots, e_{\dim M}) \in LM$ ,  $u_e$  is defined as

$$\begin{aligned} u_e : \mathbb{R}^{\dim M} &\xrightarrow{\sim} T_{\pi(e)} M \\ (x^1, \dots, x^{\dim M}) &\mapsto x^i e_i. \end{aligned}$$

To describe the inverse map  $u_e^{-1}$  explicitly, note that to every frame  $(e_1, \dots, e_{\dim M}) \in LM$ , there exists a co-frame  $(f^1, \dots, f^{\dim M}) \in L^*M$  such that

$$\begin{aligned} u_e^{-1} : T_{\pi(e)} M &\xrightarrow{\sim} \mathbb{R}^{\dim M} \\ Z &\mapsto (f^1(Z), \dots, f^{\dim M}(Z)). \end{aligned}$$

**Definition 8.26.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle with connection one-form  $\omega$  and let  $\theta \in \Omega^1(P) \otimes V$  be a solder form on  $P$ . Then

$$\Theta := D\theta \in \Omega^2(P) \otimes V$$

is the torsion of  $\omega$  with respect to  $\theta$ .

*Remark 8.20.* You can now see that the “extra structure” required to define the torsion is a choice of solder form. The previous example shows that there is a canonical choice of such a form on any frame bundle.

We would like to have a similar formula for  $\Theta$  as we had for  $\Omega$ . However, since  $\Theta$  and  $\theta$  are both  $V$ -valued but  $\omega$  is  $T_e G$ -valued, the term  $\omega \mathbin{\lrcorner} \theta$  would be meaningless. What we have, instead, is the following

$$\Theta = d\theta + \omega \mathbin{\lrcorner} \theta,$$

where the half-double wedge symbol intuitively indicates that we let  $\omega$  act on  $\theta$ . More precisely, in the case of a matrix Lie group, recalling that  $\dim G = \dim T_e G = \dim V$ , we have

$$\Theta^i = d\theta^i + \omega^i_k \mathbin{\lrcorner} \theta^k.$$

**Theorem 8.8** (Second Bianchi identity). Let  $\Theta$  be the torsion of a connection one-form  $\omega$  with respect to a solder form  $\theta$  on a principal bundle. Then

$$D\Theta = \Omega \mathbin{\lrcorner} \theta.$$

*Remark 8.21.* Like connection one-forms and curvatures two-forms, a torsion two-form  $\Theta$  can also be pulled back to the base manifold along a local section  $\sigma$  as  $T := \sigma^* \Theta$ . In fact, *this* is the torsion that one typically meets in general relativity.

## 8.8 Covariant Derivatives

### 8.8.1 Equivalence of local sections and equivariant functions

Recall that if  $F$  is a vector space and  $(P, \pi, M)$  a principal  $G$ -bundle equipped with a connection, we can use the parallel transport on the associated bundle  $(P_F, \pi_F, M)$  and the vector space structure of  $F$  to define the differential quotient of a local section  $\sigma: U \rightarrow P_F$  along an integral curve of some tangent vector  $X \in TU$ . This then allowed us to define the covariant derivative of  $\sigma$  at the point  $\pi(X) \in U$  in the direction of  $X \in TU$ .

This approach to the concept of covariant derivative is very intuitive and geometric, but it is a disaster from a technical point of view as it is quite difficult to implement. There is, in fact, a neater approach to covariant differentiation, which we will now discuss.

**Theorem 8.9.** Let  $(P, \pi, M)$  be a principal  $G$ -bundle and  $(P_F, \pi_F, M)$  be an associated bundle. Let  $(U, x)$  be a chart on  $M$ . The local sections  $\sigma: U \rightarrow P_F$  are in bijective correspondence with  $G$ -equivariant functions  $\phi: \text{preim}_\pi(U) \subseteq P \rightarrow F$ , where the  $G$ -equivariance condition is

$$\forall g \in G : \forall p \in \text{preim}_\pi(U) : \phi(g \triangleleft p) = g^{-1} \triangleright \phi(p).$$

*Proof.*

(a) Let  $\phi: \text{preim}_\pi(U) \rightarrow F$  be  $G$ -equivariant. Define

$$\begin{aligned} \sigma_\phi: U &\rightarrow P_F \\ m &\mapsto [p, \phi(p)] \end{aligned}$$

where  $p$  is any point in  $\text{preim}_\pi(\{m\})$ . First, we should check that  $\sigma_\phi$  is well-defined. Let  $p, \tilde{p} \in \text{preim}_\pi(\{m\})$ . Then, there exists a unique  $g \in G$  such that  $\tilde{p} = p \triangleleft g$ . Then, by the  $G$ -equivariance of  $\phi$ , we have

$$[\tilde{p}, \phi(\tilde{p})] = [p \triangleleft g, \phi(p \triangleleft g)] = [p \triangleleft g, g^{-1} \triangleright \phi(p)] = [p, \phi(p)]$$

and hence,  $\sigma_\phi$  is well-defined. Moreover, since for all  $g \in G$

$$\pi_F([p, \phi(p)]) = \pi(p) = \pi(p \triangleleft g) = \pi_F([p \triangleleft g, g^{-1} \triangleright \phi(p)]),$$

we have  $\pi_F \circ \sigma_\phi = \text{id}_U$  and thus,  $\sigma_\phi$  is a local section.

(b) Let  $\sigma: U \rightarrow P_F$  be a local section. Define

$$\begin{aligned} \phi_\sigma: \text{preim}_\pi(U) &\rightarrow F \\ p &\mapsto i_p^{-1}(\sigma(\pi(p))) \end{aligned}$$

where  $i_p^{-1}$  is the inverse of the map

$$\begin{aligned} i_p: F &\rightarrow \text{preim}_{\pi_F}(\{\pi(p)\}) \subseteq P_F \\ f &\mapsto [p, f]. \end{aligned}$$

Observe that, for all  $g \in G$ , we have

$$i_p(f) := [p, f] = [p \triangleleft g, g^{-1} \triangleright f] =: i_{p \triangleleft g}(g^{-1} \triangleright f).$$

Let us now show that  $\phi_\sigma$  is  $G$ -equivariant. We have

$$\begin{aligned} \phi_\sigma(p \triangleleft g) &= i_{p \triangleleft g}^{-1}(\sigma(\pi(p \triangleleft g))) \\ &= i_{p \triangleleft g}^{-1}(\sigma(\pi(p))) \\ &= i_{p \triangleleft g}^{-1}(i_p(\phi_\sigma(p))) \\ &= i_{p \triangleleft g}^{-1}(i_{p \triangleleft g}(g^{-1} \triangleright \phi_\sigma(p))) \\ &= g^{-1} \triangleright \phi_\sigma(p), \end{aligned}$$

which is what we wanted.

(c) We now show that these constructions are the inverses of each other, i.e.

$$\sigma_{\phi_\sigma} = \sigma, \quad \phi_{\sigma_\phi} = \phi.$$

Let  $m \in U$ . Then, we have

$$\begin{aligned} \sigma_{\phi_\sigma}(m) &= [p, \phi_\sigma(p)] \\ &= [p, i_p^{-1}(\sigma(\pi(p)))] \\ &= i_p(i_p^{-1}(\sigma(\pi(p)))) \\ &= \sigma(\pi(p)) \\ &= \sigma(m) \end{aligned}$$

and hence  $\sigma_{\phi_\sigma} = \sigma$ . Now let  $p \in \text{preim}_\pi(U)$ . Then, we have

$$\begin{aligned} \phi_{\sigma_\phi}(p) &= i_p^{-1}(\sigma_\phi(\pi(p))) \\ &= i_p^{-1}([p, \phi(p)]) \\ &= i_p^{-1}(i_p(\phi(p))) \\ &= \phi(p) \end{aligned}$$

and hence,  $\phi_{\sigma_\phi} = \phi$ . □

## 8.8.2 Linear actions on associated vector fiber bundles

We now specialise to the case where  $F$  is a vector space, and hence we can require the left action  $G \triangleright: F \xrightarrow{\sim} F$  to be linear.

**Proposition 8.4.** *Let  $(P, \pi, M)$  be a principal  $G$ -bundle, and let  $(P_F, \pi_F, M)$  be an associated bundle, where  $G$  is a matrix Lie group,  $F$  is a vector space, and the left  $G$ -action on  $F$  is linear. Let  $\phi: P \rightarrow F$*

be  $G$ -equivariant. Then

$$\phi(p \triangleleft \exp(At)) = \exp(-At) \triangleright \phi(p),$$

where  $p \in P$  and  $A \in T_e G$ .

**Corollary 8.2.** *With the same assumptions as above, let  $A \in T_e G$  and let  $\omega$  be a connection one-form on  $(P, \pi, M)$ . Then*

$$d\phi(X^A) + \omega(X^A) \triangleright \phi = 0.$$

*Proof.*

Since  $\phi$  is  $G$ -equivariant, by applying the previous proposition, we have

$$\phi(p \triangleleft \exp(At)) = \exp(-At) \triangleright \phi(p)$$

for any  $p \in P$ . Hence, differentiating with respect to  $t$  yields

$$\begin{aligned} (\phi(p \triangleleft \exp(At)))'(0) &= (\exp(-At) \triangleright \phi(p))'(0) \\ d_p \phi(X^A) &= -A \triangleright \phi(p) \\ d_p \phi(X^A) &= -\omega(X^A) \triangleright \phi(p) \end{aligned}$$

for all  $p \in P$  and hence, the claim holds.  $\square$

### 8.8.3 Construction of the covariant derivative

We now wish to construct a covariant derivative, i.e. an “operator”  $\nabla$  such that for any local section  $\sigma: U \subseteq M \rightarrow P_F$  and any  $X \in T_m U$  with  $m \in U$ , we have that  $\nabla_X \sigma$  is again a local section  $U \rightarrow P_F$  and

- i)  $\nabla_{fX+Y} \sigma = f \nabla_X \sigma + \nabla_Y \sigma$
- ii)  $\nabla_X (\sigma + \tau) = \nabla_X \sigma + \nabla_X \tau$
- iii)  $\nabla_X f \sigma = X(f) \sigma + f \nabla_X \sigma$

for any sections  $\sigma, \tau: U \rightarrow P_F$ , any  $f \in C^\infty(U)$ , and any  $X, Y \in T_m U$ .

These (together with  $\nabla_X f := X(f)$ ) are usually presented as the defining properties of the covariant derivative in more elementary treatments.

Recall that functions are a special case of forms, namely the 0-forms, and hence the exterior covariant derivative a function  $\phi: P \rightarrow F$  is

$$D\phi := d\phi \circ \text{hor}.$$

We now have the following result.

**Proposition 8.5.** *Let  $\phi: P \rightarrow F$  be  $G$ -equivariant and let  $X \in T_p P$ . Then*

$$D\phi(X) = d\phi(X) + \omega(X) \triangleright \phi$$

*Proof.*

- (a) Suppose that  $X$  is vertical, that is,  $X = X^A$  for some  $A \in T_e G$ . Then,

$$D\phi(X) = d\phi(\text{hor}(X)) = 0$$

and

$$d\phi(X^A) + \omega(X^A) \triangleright \phi = 0$$

by the previous corollary.

- (b) Suppose that  $X$  is horizontal. Then,

$$D\phi(X) = d\phi(X)$$

and  $\omega(X) = 0$ , so that we have

$$D\phi(X) = d\phi(X) + \omega(X) \triangleright \phi.$$

$\square$

Hence, it is clear from this proposition that  $D\phi(X)$ , which we can also write as  $D_X\phi$ , is  $\mathcal{C}^\infty(P)$ -linear in the  $X$ -slot, additive in the  $\phi$ -slot and satisfies property iii) above. However, it also clearly *not* a covariant derivative since  $X \in TP$  rather than  $X \in TM$  and  $\phi$  is a  $G$ -equivariant function  $P \rightarrow F$  rather than a local section of  $(P_F, \pi_F, M)$ .

We can obtain a covariant derivative from  $D$  by introducing a local trivialisation on the bundle  $(P, \pi, M)$ . Indeed, let  $s: U \subseteq M \rightarrow P$  be a local section. Then, we can pull back the following objects

$$\begin{array}{lll} \phi: P \rightarrow F & \rightsquigarrow & s^*\phi := \phi \circ s: U \rightarrow P_F \\ \omega \in \Omega^1(M) \otimes T_e G & \rightsquigarrow & \omega^U := s^*\omega \in \Omega^1(U) \otimes T_e G \\ D\phi \in \Omega^1(M) \otimes F & \rightsquigarrow & s^*(D\phi) \in \Omega^1(U) \otimes F. \end{array}$$

It is, in fact, for this last object that we will be able to define the covariant derivative. Let  $X \in TU$ . Then

$$\begin{aligned} (s^*D\phi)(X) &= s^*(d\phi + \omega \triangleright \phi)(X) \\ &= s^*(d\phi)(X) + s^*(\omega \triangleright \phi)(X) \\ &= d(s^*\phi)(X) + s^*(\omega)(X) \triangleright s^*\phi \\ &= d\sigma(X) + \omega^U(X) \triangleright \sigma \end{aligned}$$

where we renamed  $s^*\phi =: \sigma$ . In summary, we can write

$$\nabla_X \sigma = d\sigma(X) + \omega^U(X) \triangleright \sigma$$

One can check that this satisfies all the properties that we wanted a covariant derivative to satisfy. Of course, we should note that this is a local definition.

*Remark 8.22.* Observe that the definition of covariant derivative depends on two choices which can be made quite independently of each other, namely, the choice of connection one-form  $\omega$  (which determines  $\omega^U$ ) and the choice of linear left action  $\triangleright$  on  $F$ .