# Lectures on the Geometric Anatomy
# of Theoretical Physics

Course delivered in the 2013/14 academic year by

**Dr Frederic P. Schuller**

*Friedrich-Alexander-Universität Erlangen-Nürnberg,*
*Institut für Theoretische Physik III*

*E-mail:* fps@aei.mpg.de

Notes taken by Simon Rea

*E-mail:* s.rea.hw@gmail.com

Last updated on May 22, 2022

**Acknowledgments**

This set of lecture notes accompanies Frederic Schuller's course on the "Geometric anatomy of theoretical physics", taught in the academic year 2013/14 at the Friedrich-Alexander-Universität Erlangen-Nürnberg.

The entire course is hosted on YouTube at the following address:

www.youtube.com/playlist?list=PLPH7f_7ZlzxTi6kS4vCmv4ZKm9u8g5yic

These lecture notes are not endorsed by Dr. Schuller or the University.

While I have tried to correct typos and errors made during the lectures (some helpfully pointed out by YouTube commenters), I have also taken the liberty to add and/or modify some of the material at various points in the notes. Any errors that may result from this are, of course, mine.

If you have any comments regarding these notes, feel free to get in touch. Visit my blog for the most up to date version of these notes

http://mathswithphysics.blogspot.com

My gratitude goes to Dr. Schuller for lecturing this course and for making it available on YouTube.

<div align="right">Simon Rea</div>

# Contents

## Introduction

Theoretical physics is all about casting our concepts about the real world into rigorous mathematical form, for better or worse. But theoretical physical doesn't do that for its own sake. It does so in order to fully explore the implications of what our concepts about the real world are. So, to a certain extent, the spirit of theoretical physics can be cast into the words of Wittgenstein who said: "What we cannot speak about [clearly] we must pass over in silence." Indeed, if we have concepts about the real world and it is not possible to cast them into rigorous mathematical form, that is usually an indicator that some aspects of these concepts have not been well understood.

Theoretical physical aims at casting these concepts into mathematical language. But then, mathematics is just that: it is just a language. If we want to extract physical conclusions from this formulation, we must interpret the language. That is not the purpose or task of mathematics, that is the task of physicists. That is where it gets difficult. But then, again, mathematics is just a language and, going back to Wittgenstein, he said: "The theorems of mathematics all say the same. Namely, nothing." What did he mean by that? Well, obviously, he did not mean that mathematics is useless. He just referred to the fact that if we have a theorem of the type "$A$ if, and only if, $B$", where $A$ and $B$ are propositions, then obviously $B$ says nothing else that $A$ does, and $A$ says nothing else than $B$ does. It is a tautology. However, while from the point of view of logic and mathematics it is a tautology, psychologically, in terms of our understanding of $A$, it may be very useful to have a reformulation of $A$ in terms of $B$.

Thus, with the understanding that mathematics just gives us a language for what we want to do, the idea of this course is to provide proper language for theoretical physics. In particular, we will provide the proper mathematical language for classical mechanics, electromagnetism, quantum mechanics and statistical physics. We are not going to revise all the mathematics that is needed for these four subjects, but rather we will develop the mathematics from a higher point of view assuming some prior knowledge of these subjects.

# 1 Logic of propositions and predicates

## 1.1 Propositional logic

**Definition.** A *proposition* $p$ is a variable[1] 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.** A proposition which is always true is called a *tautology*, while one 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:

| $p$ | $\neg p$ | $\mathrm{id}(p)$ | $\top p$ | $\bot p$ |
|---|---|---|---|---|
| F | T | F | T | F |
| T | F | T | T | F |

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

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:

---

[1] By this we mean a formal expression, with no extra structure assumed.

| $p$ | $q$ | $p \Rightarrow q$ | $p \Leftrightarrow q$ |
|---|---|---|---|
| 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. $\square$

*Remark* 1.2. 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* 1.3. All higher order operators $\heartsuit(p_1, \ldots, p_N)$ 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.** A *predicate* is (informally) a proposition-valued function of some variable or variables. In particular, 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.** 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.** 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 1.4.** *Let $P(x)$ be a predicate. Then:*

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

*Remark* 1.5. 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* 1.6. 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* 1.7. 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* 1.8. 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.** 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 and theory of proofs

**Definition.** An *axiomatic system* is a finite sequence of propositions $a_1, a_2, \ldots, a_N$, which are called the *axioms* of the system.

**Definition.** A *proof* of a proposition $p$ within an axiomatic system $a_1, a_2, \ldots, a_N$ is a finite sequence of propositions $q_1, q_2, \ldots, q_M$ 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* 1.9. If $p$ can be proven within an axiomatic system $a_1, a_2, \ldots, a_N$, we write:

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

and we read "$a_1, a_2, \ldots, a_N$ proves $p$".

*Remark* 1.10. 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* 1.11. 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.** An axiomatic system $a_1, a_2, \ldots, a_N$ is said to be *consistent* if there exists a proposition $q$ which cannot be proven from the axioms. In symbols:

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

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

$$a_1, \ldots, s, \ldots, \neg s, \ldots, a_N.$$

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 1.12.** *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. $\qquad\square$

*Remark* 1.13. 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 1.14.** *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.

## 2  Axioms of set theory

### 2.1  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* 2.1. 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 $=$.

### 2.2  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:*

$$\forall\, x : \forall\, y : (x \in y) \,\underline{\vee}\, \neg(x \in y).$$

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* 2.2 (Russell'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 his 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* 2.3. 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 2.4.** *There is only one empty set, and we denote it by $\varnothing$.*

*Proof.* (Standard textbook style). 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 $x = x'$. $\qquad\square$

This is the proof that is found in standard textbooks. However, we gave a definition of what a proof within the axiomatic system of propositional logic is, and this "proof" does not satisfy our definition of proof.

In order to give a precise proof, we first have to encode our assumptions into a sequence of axioms. These consist of the axioms of propositional logic (the empty sequence) plus:

$$a_1 \Leftrightarrow \forall\, y : y \notin x$$
$$a_2 \Leftrightarrow \forall\, y : y \notin x'$$

i.e. $x$ and $x'$ are both empty sets. We now have to write down a (finite) sequence of propositions:

$$q_1 \Leftrightarrow \ldots$$
$$q_2 \Leftrightarrow \ldots$$
$$\vdots$$
$$q_M \Leftrightarrow x = x'$$

with $M$ to be determined and such that, for each $1 \leq j \leq M$ one of the following is satisfied:

(A) $q_j \Leftrightarrow a_1$ or $q_j \Leftrightarrow a_2$;

(T) $q_j$ is a tautology;

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

These are the three conditions that a sequence of propositions must satisfy in order to be a proof.

*Proof.* (Formal) We begin with a tautology.

$$
\begin{array}{ll}
q_1 \Leftrightarrow y \notin x \Rightarrow \forall\, y : (y \in x \Rightarrow y \in x') & \text{(T)} \\
q_2 \Leftrightarrow \forall\, y : y \notin x & \text{(A) using } a_1 \\
q_3 \Leftrightarrow \forall\, y : (y \in x \Rightarrow y \in x') & \text{(M) using } q_1 \text{ and } q_2
\end{array}
$$

The third step follows since $q_1 \wedge q_2 \Rightarrow q_3$ is of the form:

$$((p \Rightarrow r) \wedge p) \Rightarrow r,$$

where $p \Leftrightarrow y \notin x$ and $r \Leftrightarrow \forall\, y : (y \in x \Rightarrow y \in x')$ and it is easily seen to be true by constructing a truth table. Moreover, by the definition of $\subseteq$, we may rewrite $q_3$ as:

$$q_3 \Leftrightarrow x \subseteq x'.$$

The next three steps are very similar to the first three:

$$q_4 \Leftrightarrow y \notin x' \Rightarrow \forall y : (y \in x' \Rightarrow y \in x) \qquad \text{(T)}$$

$$q_5 \Leftrightarrow \forall y : y \notin x' \qquad \text{(A) using } a_2$$

$$q_6 \Leftrightarrow \forall y : (y \in x' \Rightarrow y \in x) \qquad \text{(M) using } q_4 \text{ and } q_5$$

where again, $q_6$ may be written as:

$$q_6 \Leftrightarrow x' \subseteq x.$$

Finally, we have:

$$q_7 \Leftrightarrow (x \subseteq x') \wedge (x' \subseteq x) \qquad \text{(M) using } q_3 \text{ and } q_6.$$

This follows since since $q_3 \wedge q_6 \Rightarrow q_7$ is of the form $p \Rightarrow p$ which is obviously a tautology. Recalling the definition of $=$, we may rewrite $q_7$ as:

$$q_7 \Leftrightarrow x = x'$$

thereby concluding the proof in seven steps. $\qquad \square$

**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* 2.5. 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* 2.6. 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* 2.7. 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* 2.8. 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.** Let $a_1, a_2, \ldots, a_N$ be sets. We define recursively for all $N \geq 2$:

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

*Remark* 2.9. 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 $\mathrm{im}_R(m)$, 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.** A relation $R$ is said to be *functional* if:

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

**Definition.** 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.

Is 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 2.10.** *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* 2.11. 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* 2.12. 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: $\{y \in m \mid P(y)\} := \varnothing$.

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 $\{y \in m \mid P(y)\} := \mathrm{im}_R(m)$. $\qquad\square$

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* 2.13. 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.** 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 $\bigcap x = \varnothing$, then $a$ and $b$ are said to be *disjoint*.

**Definition.** 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 in 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* 2.14. Let $m = \{a, b\}$. Then $\mathcal{P}(m) = \{\varnothing, \{a\}, \{b\}, \{a, b\}\}$.

*Remark* 2.15. 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:*

$$\exists\, x : \varnothing \in x \wedge \forall\, y : (y \in x \Rightarrow \{y\} \in x).$$

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

$$x = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\}, \ldots\}.$$

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

$$0 := \varnothing, \quad 1 := \{\varnothing\}, \quad 2 := \{\{\varnothing\}\}, \quad 3 := \{\{\{\varnothing\}\}\}, \quad \ldots$$

**Corollary 2.16.** *The "set" $\mathbb{N} := x$ is a set according to axiomatic set theory.*

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

*Remark* 2.17. 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* 2.18. The version of the axiom of infinity tat 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:

$$\mathbb{N} := \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \ldots\}$$

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

*Remark* 2.19. 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:*

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

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.

## 3  Classification of sets

### 3.1  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.** Let $A, B$ be sets. A *map* $\phi \colon A \to B$ 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:

$$\phi \colon A \to B$$
$$a \mapsto \phi(a)$$

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

$$\mathrm{id}_M \colon M \to M$$
$$m \mapsto m.$$

The following is standard terminology for a map $\phi \colon A \to B$:

- the set $A$ is called the *domain* of $\phi$;

- the set $B$ is called the *target* of $\phi$;

- the set $\phi(A) \equiv \mathrm{im}_\phi(A) := \{\phi(a) \mid a \in A\}$ is called the *image* of $A$ under $\phi$.

**Definition.** A *map* $\phi \colon A \to B$ is said to be:

- *injective* if $\ \forall a_1, a_2 \in A : \phi(a_1) = \phi(a_2) \Rightarrow a_1 = a_2$;

- *surjective* if $\mathrm{im}_\phi(A) = B$;

- *bijective* if it is both injective and surjective.

**Definition.** Two sets $A$ and $B$ are called *(set-theoretic) isomorphic* if there exists a bijection $\phi \colon A \to B$. In this case, we write $A \cong_{\mathrm{set}} B$.

*Remark* 3.2. If there is any bijection $A \to B$ 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** (Classification of sets). A set $A$ is:

- *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 $A \cong_{\text{set}} \{1, 2, \ldots, N\}$ for some $N \in \mathbb{N}$ and we say that the *cardinality* of $A$, denoted by $|A|$, is $N$.

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

$$\psi \circ \phi \colon A \to C$$
$$a \mapsto \psi(\phi(a)).$$

This is often represented by drawing the following diagram



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 3.3.** *Composition of maps is associative.*

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

$$\xi \circ (\psi \circ \phi) \colon A \to D$$
$$a \mapsto \xi(\psi(\phi(a)))$$

and:

$$(\xi \circ \psi) \circ \phi \colon A \to D$$
$$a \mapsto \xi(\psi(\phi(a))).$$

Thus $\xi \circ (\psi \circ \phi) = (\xi \circ \psi) \circ \phi$. $\qquad\square$

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

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

$$\phi^{-1} \circ \phi = \mathrm{id}_A$$
$$\phi \circ \phi^{-1} = \mathrm{id}_B\,.$$

Equivalently, we require the following diagram to commute:

$$\mathrm{id}_A \circlearrowright A \underset{\phi^{-1}}{\overset{\phi}{\rightleftarrows}} B \circlearrowleft \mathrm{id}_B$$

The inverse map is only defined for bijections. However, the following notion, which we will often meet in topology, is defined for any map.

**Definition.** Let $\phi\colon A \to B$ be a map and let $V \subseteq B$. Then we define the set:

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

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

**Proposition 3.4.** *Let $\phi\colon A \to B$ be a map, let $U, V \subseteq B$ and $C = \{C_j \mid j \in J\} \subseteq \mathcal{P}(B)$. Then:*

*i)* $\mathrm{preim}_\phi(\varnothing) = \varnothing$ *and* $\mathrm{preim}_\phi(B) = A$*;*

*ii)* $\mathrm{preim}_\phi(U \setminus V) = \mathrm{preim}_\phi(U) \setminus \mathrm{preim}_\phi(V)$*;*

*iii)* $\mathrm{preim}_\phi(\bigcup C) = \bigcup_{j \in J} \mathrm{preim}_\phi(C_j)$ *and* $\mathrm{preim}_\phi(\bigcap C) = \bigcap_{j \in J} \mathrm{preim}_\phi(C_j)$*.*

*Proof.* i) By definition, we have:

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

and:

$$\mathrm{preim}_\phi(\varnothing) = \{a \in A : \phi(a) \in \varnothing\} = \varnothing.$$

ii) We have:

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

iii) We have:

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

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

## 3.2 Equivalence relations

**Definition.** 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: $\forall\, m, n \in M : m \sim n \Leftrightarrow n \sim n$;

   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* 3.5. 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.** 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 3.6.** *Let $\sim$ be an equivalence relation on $M$. Then:*

   *i) $a \in [m] \Rightarrow [a] = [m]$;*

   *ii) either $[m] = [n]$ or $[m] \cap [n] = \varnothing$.*

*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 $[m] \cap [n] \neq \varnothing$. 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 $[m] = [n]$. $\qquad\square$

**Definition.** 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* 3.7. Due to the axiom of choice, there exists a complete system of representatives for $\sim$, i.e. a set $R$ such that $R \cong_{\mathrm{set}} M/\!\sim$.

*Remark* 3.8. 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* 3.9. 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:
$$[0] = [2] = [4] = \cdots = [-2] = [-4] = \cdots$$

and:
$$[1] = [3] = [5] = \cdots = [-1] = [-3] = \cdots$$

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:
$$\oplus \colon \mathbb{Z}/\!\sim \times \; \mathbb{Z}/\!\sim \; \to \mathbb{Z}/\!\sim$$

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* 3.10. 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:

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

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

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

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

where:

$$0 := \varnothing, \quad 1 := \{\varnothing\}, \quad 2 := \{\{\varnothing\}\}, \quad 3 := \{\{\{\varnothing\}\}\}, \quad \ldots$$

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.** The *successor map* $S$ on $\mathbb{N}$ is defined by:

$$S \colon \mathbb{N} \to \mathbb{N}$$
$$n \mapsto \{n\}.$$

*Example* 3.11. Consider $S(2)$. Since $2 := \{\{\varnothing\}\}$, we have $S(2) = \{\{\{\varnothing\}\}\} =: 3$. 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 $\mathbb{N}^* := \mathbb{N} \setminus \{\varnothing\}$.

**Definition.** The *predecessor map* $P$ on $\mathbb{N}^*$ is defined by:

$$P \colon \mathbb{N}^* \to \mathbb{N}$$
$$n \mapsto m \text{ such that } m \in n.$$

*Example* 3.12. We have $P(2) = P(\{\{\varnothing\}\}) = \{\varnothing\} = 1$.

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

$$S^n := S \circ S^{P(n)} \qquad \text{if } n \in \mathbb{N}^*$$
$$S^0 := \mathrm{id}_{\mathbb{N}}.$$

We are now ready to define addition.

**Definition.** The *addition* operation on $\mathbb{N}$ is defined as a map:

$$+ \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$
$$(m, n) \mapsto m + n := S^n(m).$$

*Example* 3.13. 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:
$$m + 0 = S^0(m) = \mathrm{id}_{\mathbb{N}}(m) = m$$

and:
$$0 + m = S^m(0) = S^{P(m)}(1) = S^{P(P(m))}(2) = \cdots = S^0(m) = m.$$

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 $\mathbb{N} \times \mathbb{N}$.

**Definition.** 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 $m + s = r + n$.

**Definition.** 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 $(1, 2), (2, 3), (112, 113), \ldots$

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* 3.14. 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:

$$\iota \colon \mathbb{N} \hookrightarrow \mathbb{Z}$$
$$n \mapsto [(n,0)]$$

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

**Definition.** 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.** We define the *addition of integers* $+_{\mathbb{Z}} \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ 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* 3.15. $2 +_{\mathbb{Z}} (-3) := [(2,0)] +_{\mathbb{Z}} [(0,3)] = [(2,3)] = [(0,1)] =: -1$. 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* 3.16. 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:

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

We also have the *canonical embedding* of $\mathbb{Z}$ into $\mathbb{Q}$:

$$\iota \colon \mathbb{Z} \hookrightarrow \mathbb{Q}$$
$$p \mapsto [(p,1)]$$

**Definition.** We define the *addition of rational numbers* $+_{\mathbb{Q}} \colon \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ by:

$$[(p,q)] +_{\mathbb{Q}} [(r,s)] := [(ps + rq, qs)]$$

and *multiplication of rational numbers* by:

$$[(p, q)] \cdot_{\mathbb{Q}} [(r, s)] := [(pr, qs)],$$

where the operations of addition and multiplication that appear on the right hand sides are the ones defined on $\mathbb{Z}$. It is again necessary (but easy) to check that these operations are both well-defined.

There are many ways to construct the reals from the rationals. One is to define a set $\mathscr{A}$ of *almost homomorphisms* on $\mathbb{Z}$ and hence define:

$$\mathbb{R} := \mathscr{A}/\sim,$$

where $\sim$ is a "suitable" equivalence relation on $\mathscr{A}$.

## 4  Topological spaces: construction and purpose

### 4.1  Topological spaces

We will now discuss topological spaces based on our previous development of 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 mathematician found to be useful.

**Definition.** Let $M$ be a set. A *topology* on $M$ is a set $\mathcal{O} \subseteq \mathcal{P}(M)$ such that:

   i) $\varnothing \in \mathcal{O}$ and $M \in \mathcal{O}$;

   ii) $\{U, V\} \subseteq \mathcal{O} \Rightarrow \bigcap \{U, V\} \in \mathcal{O}$;

   iii) $C \subseteq \mathcal{O} \Rightarrow \bigcup C \in \mathcal{O}$.

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 $M$.

| $|M|$ | Number of topologies |
|:---:|:---:|
| 1 | 1 |
| 2 | 4 |
| 3 | 29 |
| 4 | 355 |
| 5 | 6,942 |
| 6 | 209,527 |
| 7 | 9,535,241 |

*Example* 4.2. Let $M = \{a, b, c\}$. Then $\mathcal{O} = \{\varnothing, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}\}$ is a topology on $M$ since:

   i) $\varnothing \in \mathcal{O}$ and $M \in \mathcal{O}$;

   ii) Clearly, for any $S \in \mathcal{O}$, $\bigcap \{\varnothing, S\} = \varnothing \in \mathcal{O}$ and $\bigcap \{S, M\} = S \in \mathcal{O}$. Moreover, $\{a\} \cap \{b\} = \varnothing \in \mathcal{O}$, $\{a\} \cap \{a, b\} = \{a\} \in \mathcal{O}$, 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 $\bigcup C = \varnothing \in \mathcal{O}$.

*Example* 4.3. Let $M$ be a set. Then $\mathcal{O} = \{\varnothing, M\}$ is a topology on $M$. Indeed, we have:

  i) $\varnothing \in \mathcal{O}$ and $M \in \mathcal{O}$;

  ii) $\bigcap\{\varnothing, \varnothing\} = \varnothing \in \mathcal{O}$, $\bigcap\{\varnothing, M\} = \varnothing \in \mathcal{O}$, and $\bigcap\{M, M\} = M \in \mathcal{O}$;

  iii) If $M \in C$, then $\bigcup C = M \in \mathcal{O}$, otherwise $\bigcup C = \varnothing \in \mathcal{O}$.

This is called the *chaotic topology* and can be defined on any set.

*Example* 4.4. Let $M$ be a set. Then $\mathcal{O} = \mathcal{P}(M)$ is a topology on $M$. Indeed, we have:

  i) $\varnothing \in \mathcal{P}(M)$ 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.** 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.** 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* 4.5. Let $(M, \mathcal{O})$ be a topological space. Then $\varnothing$ is open since $\varnothing \in \mathcal{O}$. However, $\varnothing$ is also closed since $M \setminus \varnothing = M \in \mathcal{O}$. Similarly for $M$.

*Example* 4.6. Let $M = \{a, b, c\}$ and let $\mathcal{O} = \{\varnothing, \{a\}, \{a, b\}, \{a, b, c\}\}$. 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.** 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 $x$:

$$B_r(x) := \big\{y \in \mathbb{R}^d \mid \sqrt{\sum_{i=1}^d (y_i - x_i)^2} < r\big\},$$

where $x := (x_1, x_2, \ldots, x_d)$ and $y := (y_1, y_2, \ldots, y_d)$, with $x_i, y_i \in \mathbb{R}$.

*Remark* 4.7. 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.** 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 4.8.** *The pair* $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$ *is a topological space.*

*Proof.*     i) First, we need to check whether $\varnothing \in \mathcal{O}_{\text{std}}$, i.e. whether:

$$\forall\, p \in \varnothing : \exists\, r \in \mathbb{R}^+ : B_r(p) \subseteq \varnothing$$

is true. This proposition is of the form $\forall\, p \in \varnothing : Q(p)$, which was defined as being equivalent to:
$$\forall\, p : p \in \varnothing \Rightarrow Q(p).$$

However, since $p \in \varnothing$ is false, the implication is true independent of $p$. Hence the initial proposition is true and thus $\varnothing \in \mathcal{O}_{\text{std}}$.

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 $\mathbb{R}^d$. $\qquad\square$

## 4.2 Construction of new topologies from given ones

**Proposition 4.9.** *Let $(M, \mathcal{O})$ be a topological space and let $N \subset M$. Then:*

$$\mathcal{O}|_N := \{U \cap N \mid U \in \mathcal{O}\} \subseteq \mathcal{P}(N)$$

*is a topology on $N$ called the* induced (subset) topology.

*Proof.*    i) Since $\varnothing \in \mathcal{O}$ and $\varnothing = \varnothing \cap N$, we have $\varnothing \in \mathcal{O}|_N$. 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 $N$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

*Example* 4.10. 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.** 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 $q \colon M \to M/\sim$ be the map:

$$q \colon M \to M/\sim$$
$$m \mapsto [m]$$

Then we have:

$$\mathcal{O}_{M/\sim} := \{U \in M/\sim \mid \mathrm{preim}_q(U) \in \mathcal{O}\}.$$

*Example* 4.11. 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.** Let $(A, \mathcal{O}_A)$ and $(B, \mathcal{O}_B)$ be topological spaces. Then the set $\mathcal{O}_{A \times B}$ defined implicitly by:

$$U \in \mathcal{O}_{A \times B} :\Leftrightarrow \forall\, p \in U : \exists\, (S,T) \in \mathcal{O}_A \times \mathcal{O}_B : S \times T \subseteq U$$

is a topology on $A \times B$ called the *product topology*.

*Remark* 4.12. This definition can easily be extended to $n$-fold cartesian products:

$$U \in \mathcal{O}_{A_1 \times \cdots \times A_n} :\Leftrightarrow \forall\, p \in U : \exists\, (S_1, \ldots, S_n) \in \mathcal{O}_{A_1} \times \cdots \times \mathcal{O}_{A_n} : S_1 \times \cdots \times S_n \subseteq U.$$

*Remark* 4.13. Using the previous definition, one can check that the standard topology on $\mathbb{R}^d$ satisfies:

$$\mathcal{O}_{\mathrm{std}} = \mathcal{O}_{\underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{d \text{ times}}}.$$

Therefore, a more minimalistic definition of the standard topology on $\mathbb{R}^d$ would consist in defining $\mathcal{O}_{\mathrm{std}}$ only for $\mathbb{R}$ (i.e. $d = 1$) and then extending it to $\mathbb{R}^d$ by the product topology.

## 4.3 Convergence

**Definition.** Let $M$ be a set. A *sequence* (of points) in $M$ is a function $q \colon \mathbb{N} \to M$.

**Definition.** 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* 4.14. 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* 4.15. Consider the topological space $(M, \{\varnothing, M\})$. 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:

$$\forall U \in \{\varnothing, M\} : a \in U \Rightarrow \exists N \in \mathbb{N} : \forall n > N : q(n) \in U.$$

This proposition is vacuously true for $U = \varnothing$, 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* 4.16. 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* 4.17. Consider the topological space $(\mathbb{R}^d, \mathcal{O}_{\mathrm{std}})$. Then, a sequence $q \colon \mathbb{N} \to \mathbb{R}^d$ 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* 4.18. 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 $(\mathbb{R}, \mathcal{O}_{\mathrm{std}})$.

## 4.4 Continuity

**Definition.** Let $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ be topological spaces and let $\phi \colon M \to N$ be a map. Then, $\phi$ is said to be *continuous* (with respect to the topologies $\mathcal{O}_M$ and $\mathcal{O}_N$) if:

$$\forall S \in \mathcal{O}_N , \ \mathrm{preim}_\phi(S) \in \mathcal{O}_M,$$

where $\mathrm{preim}_\phi(S) := \{m \in M : \phi(m) \in S\}$ 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* 4.19. If $M$ is equipped with the discrete topology, or $N$ with the chaotic topology, then any map $\phi \colon M \to N$ 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:

$$\mathrm{preim}_\phi(S) = \{m \in M : \phi(m) \in S\} \subseteq M \in \mathcal{P}(M) = \mathcal{O}_M.$$

If instead $\mathcal{O}_N = \{\varnothing, N\}$ (and $\mathcal{O}_M$ is any topology), then either $S = \varnothing$ or $S = N$ and thus, we have:

$$\mathrm{preim}_\phi(\varnothing) = \varnothing \in \mathcal{O}_M \quad \text{and} \quad \mathrm{preim}_\phi(N) = M \in \mathcal{O}_M.$$

*Example* 4.20. Let $M = \{a, b, c\}$ and $N = \{1, 2, 3\}$, with respective topologies:

$$\mathcal{O}_M = \{\varnothing, \{b\}, \{a, c\}, \{a, b, c\}\} \quad \text{and} \quad \mathcal{O}_N = \{\varnothing, \{2\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\},$$

and let $\phi\colon M \to N$ by defined by:

$$\phi(a) = 2, \quad \phi(b) = 1, \quad \phi(c) = 2.$$

Then $\phi$ is continuous. Indeed, we have:

$$\text{preim}_\phi(\varnothing) = \varnothing, \qquad \text{preim}_\phi(\{2\}) = \{a, c\}, \qquad \text{preim}_\phi(\{3\}) = \varnothing,$$
$$\text{preim}_\phi(\{1, 3\}) = \{b\}, \qquad \text{preim}_\phi(\{2, 3\}) = \{a, c\}, \qquad \text{preim}_\phi(\{1, 2, 3\}) = \{a, b, c\},$$

and hence $\text{preim}_\phi(S) \in \mathcal{O}_M$ for all $S \in \mathcal{O}_N$.

*Example* 4.21. Consider $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$ and $(\mathbb{R}^s, \mathcal{O}_{\text{std}})$. Then $\phi\colon \mathbb{R}^d \to \mathbb{R}^s$ 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.** Let $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ be topological spaces. A bijection $\phi\colon M \to N$ is called a *homeomorphism* if both $\phi\colon M \to N$ and $\phi^{-1}\colon N \to M$ are continuous.

*Remark* 4.22. Homeo(morphism)s are the structure-preserving maps in topology.

If there exists a homeomorphism $\phi$ between $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$,



then $\phi$ provides a one-to-one pairing of the open sets of $M$ with the open sets of $N$.

**Definition.** 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 $M \cong_{\text{set}} N$.

# 5 Topological spaces: some heavily used invariants

## 5.1 Separation properties

**Definition.** 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.** 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:

$$\forall\, p, q \in M : p \neq q \Rightarrow \exists\, U(p), V(q) \in \mathcal{O} : U(p) \cap V(q) = \varnothing.$$

*Example* 5.1. The topological space $(\mathbb{R}^d, \mathcal{O}_{\text{std}})$ is T2 and hence also T1.

*Example* 5.2. The Zariski topology on an algebraic variety is T1 but not T2.

*Example* 5.3. The topological space $(M, \{\varnothing, M\})$ 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* 5.4. There are many other "T" properties, including a *T2½* property which differs from T2 in that the neighbourhoods are closed.

## 5.2 Compactness and paracompactness

**Definition.** 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 an *open* cover if $C \subseteq \mathcal{O}$.

**Definition.** Let $C$ be a cover. Then any subset $\widetilde{C} \subseteq C$ such that $\widetilde{C}$ 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.** A topological space $(M, \mathcal{O})$ is said to be *compact* if every open cover has a finite subcover.

**Definition.** 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 5.5** (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:

$$\exists\, r \in \mathbb{R}^+ : S \subseteq B_r(0).$$

*Remark* 5.6. It is also possible to generalize this result to arbitrary metric spaces. A *metric space* is a pair $(M, d)$ where $M$ is a set and $d \colon M \times M \to \mathbb{R}$ is a map such that for any $x, y, z \in M$ the following conditions hold:

i) $d(x, y) \geq 0$;

ii) $d(x, y) = 0 \Leftrightarrow x = y$;

iii) $d(x, y) = d(y, x)$;

iv) $d(x, y) \leq d(x, z) + d(y, z)$.

A metric structure on a set $M$ induces a topology $\mathcal{O}_d$ on $M$ by:

$$U \in \mathcal{O}_d :\Leftrightarrow \forall p \in U : \exists r \in \mathbb{R}^+ : B_r(p) \subseteq U,$$

where the open ball in a metric space is defined as:

$$B_r(p) := \{x \in M \mid d(p, x) < r\}.$$

In this setting, one can prove that a subset $S \subseteq M$ of a metric space $(M, d)$ is compact if, and only if, it is complete and totally bounded.

*Example* 5.7. 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* 5.8. 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:

$$C := \{(n, n + 1) \mid n \in \mathbb{Z}\} \cup \{(n + \tfrac{1}{2}, n + \tfrac{3}{2}) \mid n \in \mathbb{Z}\}.$$

This corresponds to the following picture.
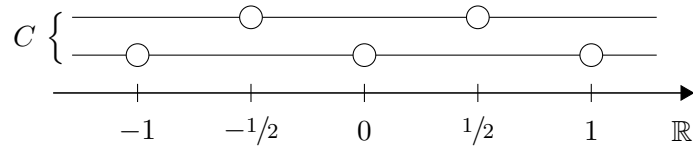


It is clear that removing even one element from $C$ will cause $C$ to fail to be an open cover of $\mathbb{R}$. Therefore, there is no finite subcover of $C$ and hence, $\mathbb{R}$ is not compact.

**Theorem 5.9.** *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.** 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:

$$\{U \in R \mid U \cap U(p) \neq \varnothing\}$$

   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.** A topological space $(M, \mathcal{O})$ is said to be *paracompact* if every open cover has an open refinement that is locally finite.

**Corollary 5.10.** *If a topological space is compact, then it is also paracompact.*

**Definition.** A topological space $(M, \mathcal{O})$ is said to be *metrisable* if there exists a metric $d$ such that the topology induced by $d$ is precisely $\mathcal{O}$, i.e. $\mathcal{O}_d = \mathcal{O}$.

**Theorem 5.11** (Stone). *Every metrisable space is paracompact.*

*Example* 5.12. The space $(\mathbb{R}^d, \mathcal{O}_{\mathrm{std}})$ is metrisable since $\mathcal{O}_{\mathrm{std}} = \mathcal{O}_d$ where $d = \|\cdot\|_2$. Hence it is paracompact by Stone's theorem.

*Remark* 5.13. 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 5.14.** *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 5.15.** *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.** 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 $C$ is an open cover, then $\mathcal{F}$ is said to be *subordinate* to the cover $C$ if:

$$\forall f \in \mathcal{F} : \exists U \in C : f(x) \neq 0 \Rightarrow x \in U.$$

**Theorem 5.16.** *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.*

*Example* 5.17. Let $\mathbb{R}$ be equipped with the standard topology. Then $\mathbb{R}$ is paracompact by Stone's theorem. Hence, every open cover of $\mathbb{R}$ admits a partition of unity subordinate to that cover. As a simple example, consider $\mathcal{F} = \{f, g\}$, where:

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ 1 - x^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

Then $\mathcal{F}$ is a partition of unity of $\mathbb{R}$. Indeed, $f, g \colon \mathbb{R} \to [0, 1]$ are both continuous, condition i) is satisfied since $\mathcal{F}$ itself is finite, and we have $\forall\, x \in \mathbb{R} : f(x) + g(x) = 1$.

Let $C := \{(-\infty, 1), (0, \infty)\}$. Then $C$ is an open cover of $\mathbb{R}$ and since:

$$f(x) \neq 0 \Rightarrow x \in (0, \infty) \quad \text{and} \quad g(x) \neq 0 \Rightarrow x \in (-\infty, 1),$$

the partition of unity $\mathcal{F}$ is subordinate to the open cover $C$.

## 5.3   Connectedness and path-connectedness

**Definition.** 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* 5.18. 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 5.19.** *The interval $[0, 1] \subseteq \mathbb{R}$ equipped with the subset topology is connected.*

**Theorem 5.20.** *A topological space $(M, \mathcal{O})$ is connected if, and only if, the only subsets that are both open and closed are $\varnothing$ and $M$.*

*Proof.* ($\Rightarrow$) Suppose, for the sake of contradiction, that there exists $U \subseteq M$ such that $U$ is both open and closed and $U \notin \{\varnothing, M\}$. Consider the sets $U$ and $M \setminus U$. Clearly, we have $U \cap M \setminus U = \varnothing$. Moreover, $M \setminus U$ is open since $U$ is closed. Therefore, $U$ and $M \setminus U$ are two open, non-empty, non-intersecting sets such that $M = U \cup M \setminus U$, contradicting the connectedness of $(M, \mathcal{O})$.

($\Leftarrow$) Suppose that $(M, \mathcal{O})$ is not connected. Then there exist open, non-empty, non-intersecting subsets $A, B \subseteq M$ such that $M = A \cup B$. Clearly, $A \neq M$, otherwise we would have $B = \varnothing$. Moreover, since $B$ is open, $A = M \setminus B$ is closed. Hence, $A$ is a set which is both open and closed and $A \notin \{\varnothing, M\}$. $\qquad \square$
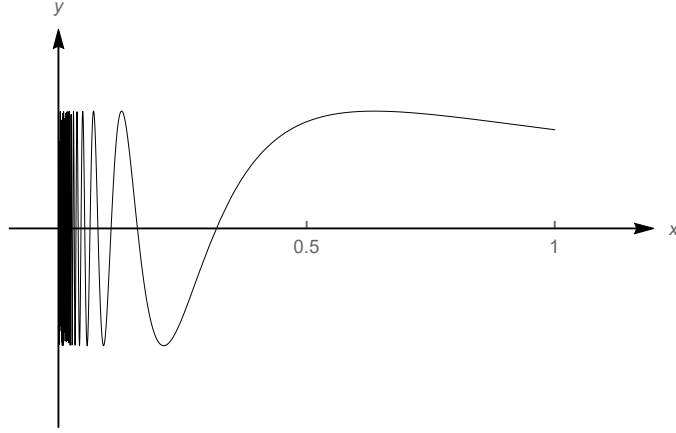
**Definition.** A topological space $(M, \mathcal{O})$ is said to be *path-connected* if for every pair of points $p, q \in M$ there exists a continuous curve $\gamma \colon [0, 1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$.

*Example* 5.21. The space $(\mathbb{R}^d, \mathcal{O}_{\mathrm{std}})$ 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* 5.22. 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 $(S, \mathcal{O}_{\mathrm{std}}|_S)$ is connected but not path-connected.

**Theorem 5.23.** *If a topological space is path-connected, then it is also connected.*

*Proof.* Let $(M, \mathcal{O})$ be path-connected but not connected. Then there exist open, non-empty, non-intersecting subsets $A, B \subset M$ such that $M = A \cup B$. Let $p \in A$ and $q \in B$. Since $(M, \mathcal{O})$ is path-connected, there exists a continuous curve $\gamma \colon [0, 1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$. Then:

$$[0, 1] = \mathrm{preim}_\gamma(M) = \mathrm{preim}_\gamma(A \cup B) = \mathrm{preim}_\gamma(A) \cup \mathrm{preim}_\gamma(B).$$

The sets $\mathrm{preim}_\gamma(A)$ and $\mathrm{preim}_\gamma(B)$ are both open, non-empty and non-intersecting, contradicting the fact that $[0, 1]$ is connected. $\qquad\square$
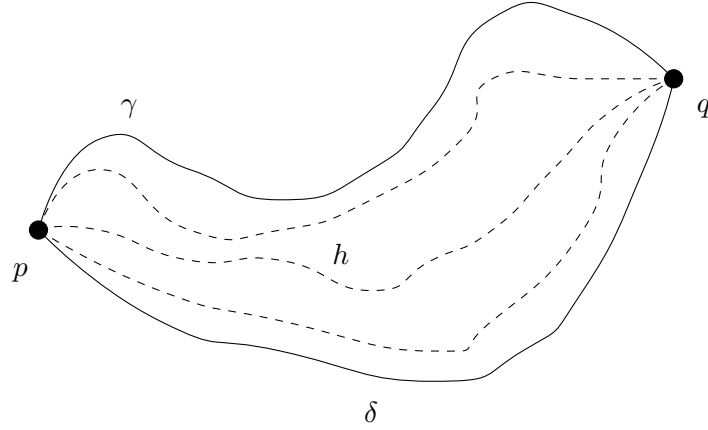
## 5.4 Homotopic curves and the fundamental group

**Definition.** Let $(M, \mathcal{O})$ be a topological space. Two curves $\gamma, \delta \colon [0, 1] \to M$ 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 $h \colon [0, 1] \times [0, 1] \to M$ 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 5.24.** *Let $\gamma \sim \delta :\Leftrightarrow$ "$\gamma$ and $\delta$ are homotopic". Then, $\sim$ is an equivalence relation.*

**Definition.** Let $(M, \mathcal{O})$ be a topological space. Then, for every $p \in M$, we define the *space of loops* at $p$ by:

$$\mathscr{L}_p := \{\gamma \colon [0, 1] \to M \mid \gamma \text{ is continuous and } \gamma(0) = \gamma(1)\}.$$

**Definition.** Let $\mathscr{L}_p$ be the space of loops at $p \in M$. We define the *concatenation* operation $* \colon \mathscr{L}_p \times \mathscr{L}_p \to \mathscr{L}_p$ 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.** 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:

$$\pi_1(p) := \mathscr{L}_p / \sim = \{[\gamma] \mid \gamma \in \mathscr{L}_p\},$$

where $\sim$ is the homotopy equivalence relation, together with the map

$$\bullet \colon \pi_1(p) \times \pi_1(p) \to \pi_1(p)$$
$$(\gamma, \delta) \mapsto [\gamma] \bullet [\delta] := [\gamma * \delta].$$

*Remark* 5.25. Recall that a group is a pair $(G, \bullet)$ where $G$ is a set and $\bullet \colon G \times G \to G$ is a map (also called *binary operation*) such that:

i) $\forall\, a, b, c \in G : (a \bullet b) \bullet c = a \bullet (b \bullet c)$;

ii) $\exists\, e \in G : \forall\, g \in G : g \bullet e = e \bullet g = g$;

iii) $\forall\, g \in G : \exists\, g^{-1} \in G : g \bullet g^{-1} = g^{-1} \bullet g = e$.

A group is called *abelian (or commutative)* if, in addition, $a \bullet b = b \bullet a$ for all $a, b \in G$.

A *group isomorphism* between two groups $(G, \bullet)$ and $(H, \circ)$ is a bijection $\phi \colon G \to H$ such that:

$$\forall\, a, b \in G : \phi(a \bullet b) = \phi(a) \circ \phi(b).$$

If there exists a group isomorphism between $(G, \bullet)$ and $(H, \circ)$, we say that $G$ and $H$ are (group theoretic) isomorphic and we write $G \cong_{\mathrm{grp}} H$.

The operation $\bullet$ is associative (since concatenation is associative); the neutral element of the fundamental group $(\pi_1(p), \bullet)$ is (the equivalence class of) the constant curve $\gamma_e$ defined by:

$$\gamma_e \colon [0, 1] \to M$$
$$\lambda \mapsto \gamma_e(0) = p$$

Finally, for each $[\gamma] \in \pi_1(p)$, the inverse under $\bullet$ is the element $[-\gamma]$, where $-\gamma$ is defined by:

$$-\gamma \colon [0, 1] \to M$$
$$\lambda \mapsto \gamma(1 - \lambda)$$

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.

A property of a topological space is called an *invariant* if any two homeomorphic spaces share the property. 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.

*Example* 5.26. The 2-sphere is defined as the set:

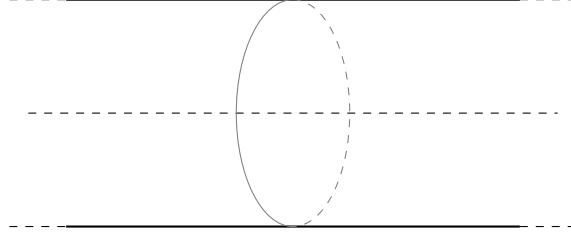$$S^2 := \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

equipped with the subset topology inherited from $\mathbb{R}^3$.
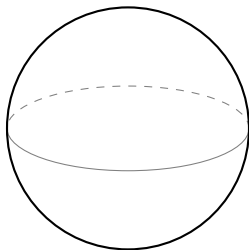


The sphere has the property that all the loops at any point are homotopic, hence the fundamental group (at every point) of the sphere is the trivial group:

$$\forall \, p \in S^2 : \pi_1(p) = 1 := \{[\gamma_e]\}.$$

*Example* 5.27. 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:

$$\forall \, p \in C : (\pi_1(p), \bullet) \cong_{\mathrm{grp}} (\mathbb{Z}, +).$$

*Example* 5.28. The 2-torus is defined as the set $T^2 := S^1 \times S^1$ equipped with the product topology.
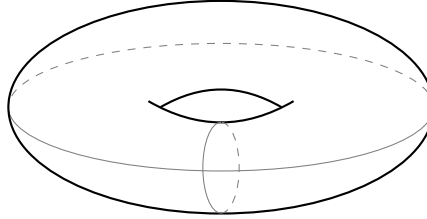


A loop in $T^2$ can intuitively wind around the cylinder-like part of the torus as well as around the hole of the torus. That is, there are two independent winding numbers and hence:

$$\forall \, p \in T^2 : \pi_1(p) \cong_{\mathrm{grp}} \mathbb{Z} \times \mathbb{Z},$$

where $\mathbb{Z} \times \mathbb{Z}$ is understood as a group under pairwise addition.

# 6 Topological manifolds and bundles

## 6.1 Topological manifolds

**Definition.** 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 $x \colon U(p) \to x(U(p)) \subseteq \mathbb{R}^d$. 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 6.1.** *Let $M$ be a d-dimensional manifold and let $U, V \subseteq M$ be open, with $U \cap V \neq \varnothing$. If $x$ and $y$ are two homeomorphisms*

$$x \colon U \to x(U) \subseteq \mathbb{R}^d \qquad and \qquad y \colon V \to y(V) \subseteq \mathbb{R}^{d'},$$

*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* 6.2. 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.** 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* 6.3. 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.** 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* 6.4. 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* 6.5. The cylinder $C = S^1 \times \mathbb{R}$ is a 2-dimensional manifold.

## 6.2  Bundles

Products are very useful. Very often in physics 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*.[2]
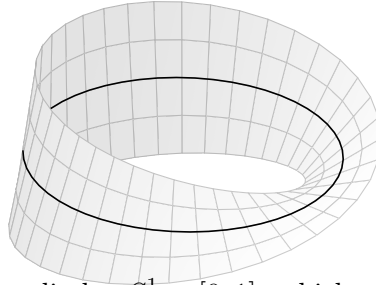


It looks locally like the finite cylinder $S^1 \times [0,1]$, which we can picture as the circle $S^1$ (the thicker line in figure) with the finite interval $[0,1]$ attached to each of its points in a "smooth" way. The Möbius strip has a "twist", which makes it globally different from the cylinder.

**Definition.** A *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 $\pi \colon E \to M$ called the *projection map*.

We will often denote the bundle $(E, \pi, M)$ by $E \xrightarrow{\pi} M$.

**Definition.** Let $E \xrightarrow{\pi} M$ be a bundle and let $p \in M$. Then, $F_p := \mathrm{preim}_\pi(\{p\})$ is called the *fibre* at the point $p$.

Intuitively, the fibre 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 fibre $F_p$ to the point $p$.
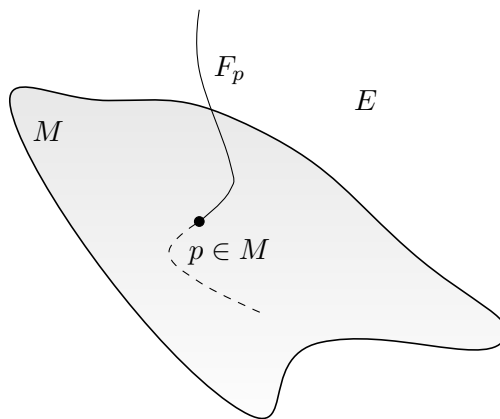
---
[2]The TikZ code for the Möbius strip was written by Jake on TeX.SE.

*Example* 6.6. 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:

$$\pi \colon M \times N \to M$$
$$(p, q) \mapsto p$$

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* 6.7. In a bundle, different points of the base manifold may have (topologically) different fibres. 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.** Let $E \xrightarrow{\pi} M$ be a bundle and let $F$ be a manifold. Then, $E \xrightarrow{\pi} M$ is called a *fibre bundle*, with (typical) fibre $F$, if:

$$\forall\, p \in M : \text{preim}_\pi(\{p\}) \cong_{\text{top}} F.$$

A fibre bundle is often represented diagrammatically as:

$$\begin{array}{ccc} F & \longrightarrow & E \\ & & \downarrow{\scriptstyle \pi} \\ & & M \end{array}$$

*Example* 6.8. The bundle $M \times N \xrightarrow{\pi} M$ is a fibre bundle with fibre $F := N$.

*Example* 6.9. The Möbius strip is a fibre bundle $E \xrightarrow{\pi} S^1$, with fibre $F := [0,1]$, where $E \neq S^1 \times [0,1]$, i.e. the Möbius strip is not a product bundle.

*Example* 6.10. A $\mathbb{C}$-line bundle over $M$ is the fibre bundle $(E, \pi, M)$ with fibre $\mathbb{C}$. 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.
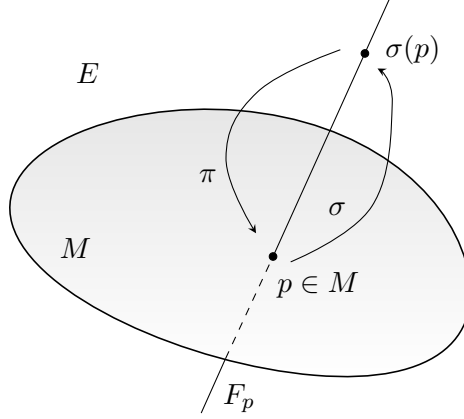
**Definition.** Let $E \xrightarrow{\pi} M$ be a bundle. A map $\sigma \colon M \to E$ is called a *(cross-)section* of the bundle if $\pi \circ \sigma = \text{id}_M$.

Intuitively, a section is a map $\sigma$ which sends each point $p \in M$ to *some* point $\sigma(p)$ in its fibre $F_p$, so that the projection map $\pi$ takes $\sigma(p) \in F_p \subseteq E$ back to the point $p \in M$.

*Example* 6.11. Let $(M \times F, \pi, M)$ be a product bundle. Then, a section of this bundle is a map:

$$\sigma \colon M \to M \times F$$
$$p \mapsto (p, s(p))$$

where $s \colon M \to F$ is any map.

**Definition.** 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.** 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:

$$\pi' := \pi|_{\mathrm{preim}_\pi(N)}$$

**Definition.** Let $E \xrightarrow{\pi} M$ and $E' \xrightarrow{\pi'} M'$ be bundles and let $u \colon E \to E'$ and $v \colon M \to M'$ be maps. Then $(u, v)$ is called a *bundle morphism* if the following diagram commutes:

$$
\begin{array}{ccc}
E & \xrightarrow{\ u\ } & E' \\
{\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi'} \\
M & \xrightarrow{\ v\ } & M'
\end{array}
$$

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.** 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:

$$
\begin{array}{ccc}
E & \underset{u^{-1}}{\overset{u}{\rightleftarrows}} & E' \\
{\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi'} \\
M & \underset{v^{-1}}{\overset{v}{\rightleftarrows}} & M'
\end{array}
$$

Such a $(u, v)$ is called a *bundle isomorphism* and we write $E \xrightarrow{\pi} M \cong_{\mathrm{bdl}} E' \xrightarrow{\pi'} M'$.

Bundle isomorphisms are the structure-preserving maps for bundles.

**Definition.** 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.** 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* 6.12. The cylinder $C$ is trivial as a bundle, and hence also locally trivial.

*Example* 6.13. The Möbious strip is not trivial but it is locally trivial.

From now on, we will mostly consider locally trivial bundles.

*Remark* 6.14. 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.** Let $E \xrightarrow{\pi} M$ be a bundle and let $f \colon M' \to M$ 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:

$$\begin{aligned} u \colon \quad E' &\to E \\ (m', e) &\mapsto e \end{aligned}$$

This corresponds to the diagram:

$$\begin{array}{ccc} E' & \xrightarrow{\;u\;} & E \\ {\scriptstyle \pi'}\big\downarrow & & \big\downarrow {\scriptstyle \pi} \\ M' & \xrightarrow{\;f\;} & M \end{array}$$

*Remark* 6.15. 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 $f$.

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:

$$\pi(\sigma(f(m'))) = (\pi \circ \sigma \circ f)(m') = (\mathrm{id}_M \circ f)(m') = f(m')$$

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:

$$\sigma' \colon M' \to E'$$
$$m' \mapsto (m', (\sigma \circ f)(m'))$$

satisfies $\pi' \circ \sigma' = \mathrm{id}_{M'}$ and it is thus a section on the pull-back bundle $E' \xrightarrow{\pi'} M'$.

### 6.3 Viewing manifolds from atlases

**Definition.** Let $(M, \mathcal{O})$ be a $d$-dimensional manifold. Then, a pair $(U, x)$ where $U \in \mathcal{O}$ and $x \colon U \to x(U) \subseteq \mathbb{R}^d$ is a homeomorphism, is said to be a *chart* of the manifold.

The *component functions (or maps)* of $x \colon U \to x(U) \subseteq \mathbb{R}^d$ are the maps:

$$x^i \colon U \to \mathbb{R}$$
$$p \mapsto \mathrm{proj}_i(x(p))$$

for $1 \le i \le d$, where $\mathrm{proj}_i(x(p))$ 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.** An *atlas* of a manifold $M$ is a collection $\mathscr{A} := \{(U_\alpha, x_\alpha) \mid \alpha \in \mathcal{A}\}$ of charts such that:

$$\bigcup_{\alpha \in \mathcal{A}} U_\alpha = M.$$

**Definition.** Two charts $(U, x)$ and $(V, y)$ are said to be $\mathcal{C}^0$-*compatible* if either $U \cap V = \varnothing$ or the map:

$$y \circ x^{-1} \colon x(U \cap V) \to y(U \cap V)$$

is continuous.

Note that $y \circ x^{-1}$ is a map from a subset of $\mathbb{R}^d$ to a subset of $\mathbb{R}^d$.

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 my 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.

*Remark* 6.16. The map $y \circ x^{-1}$ (and its inverse $x \circ y^{-1}$) is called the *co-ordinate change map* or *chart transition map*.

**Definition.** 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.** A $\mathcal{C}^0$-atlas $\mathscr{A}$ is said to be a *maximal atlas* if for every $(U, x) \in \mathscr{A}$, we have $(V, y) \in \mathscr{A}$ for all $(V, y)$ charts that are $\mathcal{C}^0$-compatible with $(U, x)$.

*Example* 6.17. Not every $\mathcal{C}^0$-atlas is a maximal atlas. Indeed, consider $(\mathbb{R}, \mathcal{O}_{\text{std}})$ and the atlas $\mathscr{A} := (\mathbb{R}, \text{id}_{\mathbb{R}})$. Then $\mathscr{A}$ 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 $((0, 1), \text{id}_{\mathbb{R}}) \notin \mathscr{A}$.

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 $\gamma \colon \mathbb{R} \to M$. 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 $\gamma \colon \mathbb{R} \to M$ 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 $\gamma \colon \text{preim}_\gamma(U) \to U$ directly, we study the map:

$$x \circ \gamma \colon \text{preim}_\gamma(U) \to x(U) \subseteq \mathbb{R}^d,$$

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 $\gamma \colon \text{preim}_\gamma(U) \to U$ (real, as opposed to its co-ordinate representation).
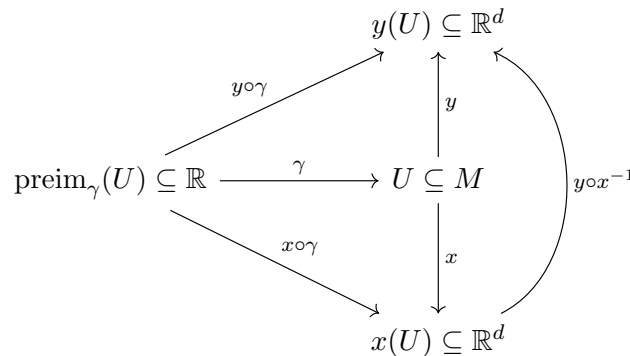


At some point you may wish to use a different "co-ordinate system" to answer a different question. In this case, you would chose a different chart $(U, y)$ and then study the map $y \circ \gamma$

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 $\mathrm{preim}_\gamma(U) \subseteq \mathbb{R}$ and $x(U), y(U) \subseteq \mathbb{R}^d$ together with the maps between them, which is our representation of the real world.

# 7 Differentiable structures: definition and classification

## 7.1 Adding structure by refining the (maximal) atlas

We saw previously that for a topological manifold $(M, \mathcal{O})$, the concept of a $\mathcal{C}^0$-atlas was 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.** An atlas $\mathscr{A}$ for a topological manifold is called a $\maltese$-*atlas* if any two charts $(U, x), (V, y) \in \mathscr{A}$ are $\maltese$-compatible.

In other words, either $U \cap V = \varnothing$ or if $U \cap V \neq \varnothing$, then the transition map $y \circ x^{-1}$ from $x(U \cap V)$ to $y(U \cap V)$ must be $\maltese$.

$$U \cap V \subseteq M$$

$$x(U \cap V) \subseteq \mathbb{R}^{\dim M} \xrightarrow{\quad y \circ x^{-1} \quad} y(U \cap V) \subseteq \mathbb{R}^{\dim M}$$

with arrows labelled $x$ and $y$ from $U \cap V$.

Before you think Dr Schuller finally went nuts, the symbol $\maltese$ is being used as a placeholder for any of the following:

- $\maltese = \mathcal{C}^0$: this just reduces to the previous definition;

- $\maltese = \mathcal{C}^k$: the transition maps are $k$-times continuously differentiable as maps between open subsets of $\mathbb{R}^{\dim M}$;

- $\maltese = \mathcal{C}^\infty$: the transition maps are smooth (infinitely many times differentiable); equivalently, the atlas is $\mathcal{C}^k$ for all $k \geq 0$;

- $\maltese = \mathcal{C}^\omega$: the transition maps are (real) analytic, which is stronger than being smooth;

- $\maltese = $ 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 $\frac{1}{2} \dim M$.

As an aside, if you haven't met the Cauchy-Riemann equations yet, recall than since $\mathbb{R}^2$ and $\mathbb{C}$ are isomorphic as sets, we can identify the function

$$
\begin{aligned}
f \colon \quad \mathbb{R}^2 &\to \mathbb{R}^2 \\
(x, y) &\mapsto (u(x, y), v(x, y))
\end{aligned}
$$

where $u, v \colon \mathbb{R}^2 \to \mathbb{R}$, with the function

$$
\begin{aligned}
f \colon \quad \mathbb{C} &\to \mathbb{C} \\
x + \mathrm{i}y &\mapsto u(x, y) + \mathrm{i}v(x, y).
\end{aligned}
$$

If $u$ and $v$ are real differentiable at $(x_0, y_0)$, then $f = u + iv$ is complex differentiable at $z_0 = x_0 + iy_0$ if, and only if, $u$ and $v$ satisfy

$$\frac{\partial u}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0) \quad \wedge \quad \frac{\partial u}{\partial y}(x_0, y_0) = -\frac{\partial v}{\partial x}(x_0, y_0),$$

which are known as the Cauchy-Riemann equations. Note that differentiability in the complex plane is a much stronger condition than differentiability over the real numbers. If you want to know more, you should take a course in complex analysis or function theory.

We now go back to manifolds.

**Theorem 7.1** (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.** A $\mathcal{C}^k$-*manifold* is a triple $(M, \mathcal{O}, \mathscr{A})$, where $(M, \mathcal{O})$ is a topological manifold and $\mathscr{A}$ is a maximal $\mathcal{C}^k$-atlas.

*Remark* 7.2. 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.** Two ✿-atlases $\mathscr{A}$, $\mathscr{B}$ are *compatible* if their union $\mathscr{A} \cup \mathscr{B}$ is again a ✿-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* 7.3. Let $(M, \mathcal{O}) = (\mathbb{R}, \mathcal{O}_{std})$. Consider the two atlases $\mathscr{A} = \{(\mathbb{R}, \mathrm{id}_\mathbb{R})\}$ and $\mathscr{B} = \{(\mathbb{R}, x)\}$, where $x \colon a \mapsto \sqrt[3]{a}$. 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 $\mathrm{id}_\mathbb{R}$). Hence they are both $\mathcal{C}^\infty$-atlases.

Consider now $\mathscr{A} \cup \mathscr{B}$. The transition map $\mathrm{id}_\mathbb{R} \circ x^{-1}$ is the map $a \mapsto a^3$, which is smooth. However, the other transition map, $x \circ \mathrm{id}_\mathbb{R}^{-1}$, is the map $x$, which is not even differentiable once (the first derivative at 0 does not exist). Consequently, $\mathscr{A}$ and $\mathscr{B}$ 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.

## 7.2 Differentiable manifolds

**Definition.** Let $\phi\colon M \to N$ be a map, where $(M, \mathcal{O}_M, \mathscr{A}_M)$ and $(N, \mathcal{O}_N, \mathscr{A}_N)$ are $\mathcal{C}^k$-manifolds. Then $\phi$ is said to be $(\mathcal{C}^k\text{-})differentiable$ at $p \in M$ if for some charts $(U, x) \in \mathscr{A}_M$ with $p \in U$ and $(V, y) \in \mathscr{A}_N$ 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.

$$
\begin{array}{ccc}
U \subseteq M & \xrightarrow{\quad\phi\quad} & V \subseteq N \\
\downarrow{\scriptstyle x} & & \downarrow{\scriptstyle y} \\
x(U) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{y \circ \phi \circ x^{-1}} & y(V) \subseteq \mathbb{R}^{\dim N}
\end{array}
$$

The above diagram shows a typical theme with manifolds. We have a map $\phi\colon M \to N$ 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 7.4.** *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 $(U, x) \in \mathscr{A}_M$ with $p \in U$ and $(V, y) \in \mathscr{A}_N$ with $\phi(p) \in V$, then $\widetilde{y} \circ \phi \circ \widetilde{x}^{-1}$ is differentiable at $\widetilde{x}(p)$ for all charts $(\widetilde{U}, \widetilde{x}) \in \mathscr{A}_M$ with $p \in \widetilde{U}$ and $(\widetilde{V}, \widetilde{y}) \in \mathscr{A}_N$ with $\phi(p) \in \widetilde{V}$.

$$
\begin{array}{ccc}
\widetilde{x}(U \cap \widetilde{U}) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{\widetilde{y} \circ \phi \circ \widetilde{x}^{-1}} & \widetilde{y}(V \cap \widetilde{V}) \subseteq \mathbb{R}^{\dim N} \\
\uparrow{\scriptstyle \widetilde{x}} & & \uparrow{\scriptstyle \widetilde{y}} \\
U \cap \widetilde{U} \subseteq M & \xrightarrow{\quad\phi\quad} & V \cap \widetilde{V} \subseteq N \\
\downarrow{\scriptstyle x} & & \downarrow{\scriptstyle y} \\
x(U \cap \widetilde{U}) \subseteq \mathbb{R}^{\dim M} & \xrightarrow{y \circ \phi \circ x^{-1}} & y(V \cap \widetilde{V}) \subseteq \mathbb{R}^{\dim N}
\end{array}
$$

with outer arrows $\widetilde{x} \circ x^{-1}$ on the left and $\widetilde{y} \circ y^{-1}$ on the right.

Consider the map $\widetilde{x} \circ x^{-1}$ in the diagram above. Since the charts $(U, x)$ and $(\widetilde{U}, \widetilde{x})$ belong to the same $\mathcal{C}^k$-atlas $\mathscr{A}_M$, by definition the transition map $\widetilde{x} \circ x^{-1}$ is $\mathcal{C}^k$-differentiable as a map between subsets of $\mathbb{R}^{\dim M}$, and similarly for $\widetilde{y} \circ y^{-1}$. We now notice that we can write:

$$
\widetilde{y} \circ \phi \circ \widetilde{x}^{-1} = (\widetilde{y} \circ y^{-1}) \circ (y \circ \phi \circ x^{-1}) \circ (\widetilde{x} \circ x^{-1})^{-1}
$$

and since the composition of $\mathcal{C}^k$ maps is still $\mathcal{C}^k$, we are done. $\qquad\square$

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* 7.5. Consider the smooth manifolds $(\mathbb{R}^d, \mathcal{O}_{\text{std}}, \mathscr{A}_d)$ and $(\mathbb{R}^{d'}, \mathcal{O}_{\text{std}}, \mathscr{A}'_d)$, where $\mathscr{A}_d$ and $\mathscr{A}'_d)$ 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 $f\colon \mathbb{R}^d \to \mathbb{R}^{d'}$ be a map. The diagram defining the differentiability of $f$ with respect to these charts is

$$
\begin{array}{ccc}
\mathbb{R}^d & \xrightarrow{\quad f \quad} & \mathbb{R}^{d'} \\
\downarrow{\scriptstyle \text{id}_{\mathbb{R}^d}} & & \downarrow{\scriptstyle \text{id}_{\mathbb{R}^{d'}}} \\
\mathbb{R}^d & \xrightarrow{\ \text{id}_{\mathbb{R}^{d'}} \circ f \circ (\text{id}_{\mathbb{R}^d})^{-1}\ } & \mathbb{R}^{d'}
\end{array}
$$

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* 7.6. Let $(M, \mathcal{O}, \mathscr{A})$ be a $d$-dimensional smooth manifold and let $(U, x) \in \mathscr{A}$. Then $x\colon U \to x(U) \subseteq \mathbb{R}^d$ is smooth. Indeed, we have

$$
\begin{array}{ccc}
U & \xrightarrow{\quad x \quad} & x(U) \\
\downarrow{\scriptstyle x} & & \downarrow{\scriptstyle \text{id}_{x(U)}} \\
x(U) \subseteq \mathbb{R}^d & \xrightarrow{\ \text{id}_{x(U)} \circ x \circ x^{-1}\ } & x(U) \subseteq \mathbb{R}^d
\end{array}
$$

Hence $x\colon U \to x(U)$ 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 $x^i := \text{proj}_i \circ x\colon U \to \mathbb{R}$ are also smooth. Indeed, consider the diagram

$$
\begin{array}{ccc}
U & \xrightarrow{\quad x^i \quad} & \mathbb{R} \\
\downarrow{\scriptstyle x} & & \downarrow{\scriptstyle \text{id}_{\mathbb{R}}} \\
x(U) \subseteq \mathbb{R}^d & \xrightarrow{\ \text{id}_{\mathbb{R}} \circ x^i \circ x^{-1}\ } & \mathbb{R}
\end{array}
$$

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.

## 7.3 Classification of differentiable structures

**Definition.** Let $\phi\colon M \to N$ 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.** Two manifolds $(M, \mathcal{O}_M, \mathscr{A}_M)$, $(N, \mathcal{O}_N, \mathscr{A}_N)$ are said to be *diffeomorphic* if there exists a diffeomorphism $\phi\colon M \to N$ between them. We write $M \cong_{\mathrm{diff}} N$.

Note that if the differentiable structure is understood (or irrelevant), we typically write $M$ instead of the triple $(M, \mathcal{O}_M, \mathscr{A}_M)$.

*Remark* 7.7. 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 7.8** (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 $(\mathbb{R}, \mathcal{O}_{\mathrm{std}})$ with two incompatible atlases $\mathscr{A}$ and $\mathscr{B}$. Let $\mathscr{A}_{\mathrm{max}}$ and $\mathscr{B}_{\mathrm{max}}$ be their extensions to maximal atlases, and consider the smooth manifolds $(\mathbb{R}, \mathcal{O}_{\mathrm{std}}, \mathscr{A}_{\mathrm{max}})$ and $(\mathbb{R}, \mathcal{O}_{\mathrm{std}}, \mathscr{B}_{\mathrm{max}})$. 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 $(\mathbb{R}^4, \mathcal{O}_{\mathrm{std}})$.

In the compact case there are only partial results. By way of example, here is one such result.

**Proposition 7.9.** *If $(M, \mathcal{O})$, with $\dim M = 4$, has $b_2 > 18$, where $b_2$ is the second Betti number, then there are countably many non-diffeomorphic smooth structures that we can equip $(M, \mathcal{O})$ with.*

Betti numbers are defined in terms of homology groups, but intuitively we have:

- $b_0$ is the number of connected components a space has;

- $b_1$ is the number of circular (1-dimensional) holes a space has;

- $b_2$ is the number of 2-dimensional holes a space has;

and so on. Hence if a manifold has a number of 2-dimensional holes greater than 18, then there only countably many structures that we can choose from, but they are still infinitely many.

## 8 Tensor space theory I: over a field

### 8.1 Vector spaces

We begin with a quick review of vector spaces.

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

- $(K, +)$ is an abelian group, i.e.

    i) $\forall\, a, b, c \in K : (a + b) + c = a + (b + c)$;

    ii) $\exists\, 0 \in K : \forall\, a \in K : a + 0 = 0 + a = a$;

    iii) $\forall\, a \in K : \exists -a \in K : a + (-a) = (-a) + a = 0$;

    iv) $\forall\, a, b \in K : a + b = b + a$;

- $(K^*, \cdot)$, where $K^* := K \setminus \{0\}$, is an abelian group, i.e.

    v) $\forall\, a, b, c \in K^* : (a \cdot b) \cdot c = a \cdot (b \cdot c)$;

    vi) $\exists\, 1 \in K^* : \forall\, a \in K^* : a \cdot 1 = 1 \cdot a = a$;

    vii) $\forall\, a \in K^* : \exists\, a^{-1} \in K^* : a \cdot a^{-1} = a^{-1} \cdot a = 1$;

    viii) $\forall\, a, b \in K^* : a \cdot b = b \cdot a$;

- the maps $+$ and $\cdot$ satisfy the distributive property:

    ix) $\forall\, a, b, c \in K : (a + b) \cdot c = a \cdot c + b \cdot c$.

*Remark* 8.1. In the above definition, we included axiom iv for the sake of clarity, but in fact it can be proven starting from the other axioms.

*Remark* 8.2. A weaker notion that we will encounter later is that of a *ring*. This is also defined as a triple $(R, +, \cdot)$, but we do not require axiom vi, vii and viii to hold. If a ring satisfies axiom vi, it is called a *unital ring*, and if it satisfies axiom viii, it is called a *commutative ring*. We will mostly consider unital rings, a call them just rings.

*Example* 8.3. The triple $(\mathbb{Z}, +, \cdot)$ is a commutative, unital ring. However, it is not a field since 1 and $-1$ are the only two elements which admit an inverse under multiplication.

*Example* 8.4. The sets $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields under the usual $+$ and $\cdot$ operations.

*Example* 8.5. An example of a non-commutative ring is the set of real $m \times n$ matrices $M_{m \times n}(\mathbb{R})$ under the usual operations.

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

$$\oplus \colon V \times V \to V$$
$$\odot \colon K \times V \to V$$

are maps satisfying the following axioms:

- $(V, \oplus)$ is an abelian group;

- the map $\odot$ is an *action* of $K$ on $(V, \oplus)$:

  i) $\forall\, \lambda \in K : \forall\, v, w \in V : \lambda \odot (v \oplus w) = (\lambda \odot v) \oplus (\lambda \odot w)$;

  ii) $\forall\, \lambda, \mu \in K : \forall\, v \in V : (\lambda + \mu) \odot v = (\lambda \odot v) \oplus (\mu \odot v)$;

  iii) $\forall\, \lambda, \mu \in K : \forall\, v \in V : (\lambda \cdot \mu) \odot v = \lambda \odot (\mu \odot v)$;

  iv) $\forall\, v \in V : 1 \odot v = v$.

Vector spaces are also called *linear spaces*. Their elements are called *vectors*, while the elements of $K$ are often called *scalars*, and the map $\odot$ is called *scalar multiplication*. You should already be familiar with the various vector space constructions from your linear algebra course. For example, recall:

**Definition.** Let $(V, \oplus, \odot)$ be a vector space over $K$ and let $U \subseteq V$ be non-empty. Then we say that $(U, \oplus|_{U \times U}, \odot|_{K \times U})$ is a *vector subspace* of $(V, \oplus, \odot)$ if:

i) $\forall\, u_1, u_2 \in U : u_1 \oplus u_2 \in U$;

ii) $\forall\, u \in U : \forall\, \lambda \in K : \lambda \odot u \in U$.

More succinctly, if $\forall\, u_1, u_2 \in U : \forall\, \lambda \in K : (\lambda \odot u_1) \oplus u_2 \in U$.

Also recall that if we have $n$ vector spaces over $K$, we can form the $n$-fold Cartesian product of their underlying sets and make it into a vector space over $K$ by defining the operations $\oplus$ and $\odot$ componentwise.

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

**Definition.** Let $(V, \oplus, \odot)$, $(W, \boxplus, \boxdot)$ be vector spaces over the same field $K$ and let $f \colon V \to W$ be a map. We say that $f$ is a *linear map* if for all $v_1, v_2 \in V$ and all $\lambda \in K$

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

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.** A bijective linear map is called a *linear isomorphism* of vector spaces. Two vector spaces are said to be *isomorphic* is there exists a linear isomorphism between them. We write $V \cong_{\text{vec}} W$.

*Remark* 8.6. Note that, unlike what happens with topological spaces, the inverse of a bijective linear map is automatically linear, hence we do not need to specify this in the definition of linear isomorphism.

**Definition.** Let $V$ and $W$ be vector spaces over the same field $K$. Define the set

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

where the notation $f \colon V \xrightarrow{\sim} W$ stands for "$f$ is a linear map from $V$ to $W$".

The hom-set $\mathrm{Hom}(V, W)$ can itself be made into a vector space over $K$ by defining:

$$\oplus \colon \mathrm{Hom}(V, W) \times \mathrm{Hom}(V, W) \to \mathrm{Hom}(V, W)$$
$$(f, g) \mapsto f \oplus g$$

where

$$f \oplus g \colon V \xrightarrow{\sim} W$$
$$v \mapsto (f \oplus g)(v) := f(v) + g(v),$$

and

$$\diamond \colon K \times \mathrm{Hom}(V, W) \to \mathrm{Hom}(V, W)$$
$$(\lambda, f) \mapsto \lambda \diamond f$$

where

$$\lambda \diamond f \colon V \xrightarrow{\sim} W$$
$$v \mapsto (\lambda \diamond f)(v) := \lambda f(v).$$

It is easy to check that both $f \oplus g$ 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 $\lambda \diamond f \in \mathrm{Hom}(V, W)$. One should also check that $\oplus$ and $\diamond$ satisfy the vector space axioms.

*Remark* 8.7. Notice that in the definition of vector space, none of the axioms require that $K$ necessarily be a field. In fact, just a (unital) ring would suffice. Vector spaces over rings, have a name of their own. They are called *modules* over a ring, and we will meet them later.

For the moment, it is worth pointing out that everything we have done so far applies equally well to modules over a ring, up to and including the definition of $\mathrm{Hom}(V, W)$. However, if we try to make $\mathrm{Hom}(V, W)$ into a module, we run into trouble. Notice that in the derivation above, we used the fact the multiplication in a field is commutative. But this is not the case in general in a ring.

The following are commonly used terminology.

**Definition.** Let $V$ be a vector space. An *endomorphism* of $V$ is a linear map $V \to V$. We write $\mathrm{End}(V) := \mathrm{Hom}(V, V)$.

**Definition.** Let $V$ be a vector space. An *automorphism* of $V$ is a linear isomorphism $V \to V$. We write $\operatorname{Aut}(V) := \{f \in \operatorname{End}(V) \mid f \text{ is an isomorphism}\}$.

*Remark* 8.8. Note that, unlike $\operatorname{End}(V)$, $\operatorname{Aut}(V)$ is *not* a vector space as was claimed in lecture. It however a group under the operation of composition of linear maps.

**Definition.** Let $V$ be a vector space over $K$. The *dual* vector space to $V$ is

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

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 products.

## 8.2 Tensors and tensor spaces

**Definition.** Let $V$, $W$, $Z$ be vector spaces over $K$. A map $f \colon V \times W \to Z$ 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 $V \to Z$ and $W \to Z$, respectively.

*Remark* 8.9. Compare this with the definition of a linear map $f \colon V \times W \xrightarrow{\sim} Z$:

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

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* 8.10. The map $f \colon \mathbb{R}^2 \to \mathbb{R}$ 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.** Let $V$ be a vector space over $K$. A $(p, q)$-*tensor* $T$ on $V$ is a multilinear map

$$T \colon \underbrace{V^* \times \cdots \times V^*}_{p \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q \text{ copies}} \to K.$$

We write

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

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

*Remark* 8.11. By convention, a $(0, 0)$ on $V$ is just an element of $K$, and hence $T_0^0 V = K$.

*Remark* 8.12. 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 $f\colon A \to B$ satisfying some property $p$. Recall that the notation $f\colon A \to B$ 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:

$$\{f \in \mathcal{P}(A \times B) \mid f\colon A \to B \text{ and } p(f)\}.$$

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

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

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

$$\oplus\colon T_q^p V \times T_q^p V \to T_q^p V$$
$$(T, S) \mapsto T \oplus S$$

and

$$\odot\colon K \times T_q^p V \to T_q^p V$$
$$(\lambda, T) \mapsto \lambda \odot T,$$

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

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

**Definition.** 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:

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

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

Some examples are in order.

*Example* 8.13.     a) $T_1^0 V := \{T \mid T\colon V \xrightarrow{\sim} K\} = \mathrm{Hom}(V, K) =: V^*$. Note that here multilinear is the same as linear since the maps only have one argument.

b) $T^1_1 V \equiv V \otimes V^* := \{T \mid T$ is a bilinear map $V^* \times V \to K\}$. 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:

$$\widehat{T}\colon V^* \xrightarrow{\sim} V^*$$
$$\omega \mapsto T(-, \omega)$$

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

$$T(-, \omega)\colon V \xrightarrow{\sim} K$$
$$v \mapsto T(v, \omega).$$

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

$$T\colon V \times V^* \to K$$
$$(v, \omega) \mapsto T(v, \omega) := (\widehat{T}(\omega))(v).$$

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

$$T^1_1 V \cong_{\text{vec}} \text{End}(V^*).$$

Other examples we would like to consider are

c) $T^0_1 V \stackrel{?}{\cong}_{\text{vec}} V$: 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.

The definition of dimension hinges on the notion of a basis. Given a vector space Without any additional structure, the only notion of basis that we can define is a so-called Hamel basis.

**Definition.** 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 $\{b_1, \ldots, b_N\}$ of $\mathcal{B}$ is linearly independent, i.e.

$$\sum_{i=1}^{N} \lambda^i b_i = 0 \;\Rightarrow\; \lambda^1 = \cdots = \lambda^N = 0;$$

- $\mathcal{B}$ is a *generating* or *spanning set* of $V$, i.e.

$$\forall\, v \in V : \exists\, v^1, \ldots, v^M \in K : \exists\, b_1, \ldots, b_M \in \mathcal{B} : v = \sum_{i=1}^{M} v^i b_i.$$

*Remark* 8.14. We can write the second condition more succinctly by defining

$$\operatorname{span}_K(\mathcal{B}) := \left\{ \sum_{i=1}^{n} \lambda^i b_i \ \middle| \ \lambda^i \in K \wedge b_i \in \mathcal{B} \wedge n \geq 1 \right\}$$

and thus writing $V = \operatorname{span}_K(\mathcal{B})$.

*Remark* 8.15. 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 8.16.** *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*

- $\operatorname{span}(S) = V \ \Rightarrow \ |S| \geq |\mathcal{B}|$;

- *$S$ is linearly independent $\ \Rightarrow \ |S| \leq |\mathcal{B}|$.*

**Definition.** 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 8.17.** *If $\dim V < \infty$ and $S \subseteq V$, then we have the following:*

- *if $\operatorname{span}_K(S) = V$ 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 8.18.** *If $\dim V < \infty$, then $(V^*)^* \cong_{\mathrm{vec}} V$.*

*Sketch of proof.* One constructs an explicit isomorphism as follows. Define the *evaluation map* as

$$\operatorname{ev} \colon V \xrightarrow{\sim} (V^*)^*$$
$$v \mapsto \operatorname{ev}_v$$

where

$$\operatorname{ev}_v \colon V^* \xrightarrow{\sim} K$$
$$\omega \mapsto \operatorname{ev}_v(\omega) := \omega(v)$$

The linearity of ev follows immediately from the linearity of the elements of $V^*$, while that of $\operatorname{ev}_v$ from the fact that $V^*$ is a vector space. One then shows that ev is both injective and surjective, and hence an isomorphism. $\qquad \square$

*Remark* 8.19. 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* 8.20. It is not hard to show that $(V^*)^* \cong_{\text{vec}} V$ implies $T_1^0 V \cong_{\text{vec}} V$ and $T_1^1 V \cong_{\text{vec}}$ End$(V)$. So the last two hold in finite dimensions, but they need not hold in infinite dimensions.

*Remark* 8.21. 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. Some people say: *"A gentleman only chooses a basis if he must."*

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.** Let $V$ be a finite-dimensional vector space with basis $\mathcal{B} = \{e_1, \ldots, e_{\dim V}\}$. The *dual basis* to $\mathcal{B}$ is the unique basis $\mathcal{B}' = \{f^1, \ldots, f^{\dim V}\}$ of $V^*$ such that

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

*Remark* 8.22. 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.

The proper treatment of this matter falls within the scope of *category theory*, and the relevant notion is called *natural isomorphism*. See, for instance, the book *Basic Category Theory* by Tom Leinster for an introduction to the subject.

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 $\mathcal{B}$. The notion of coordinates can also be generalised to the case of tensors.

**Definition.** Let $V$ be a finite-dimensional vector space over $K$ with basis $\mathcal{B} = \{e_1, \ldots, e_{\dim V}\}$ and let $T \in T_q^p V$. We define the *components* of $T$ in the basis $\mathcal{B}$ to be the numbers

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

where $1 \leq a_i, b_j \leq \dim V$ and $\{f^1, \ldots, f^{\dim V}\}$ is the dual basis to $\mathcal{B}$.

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

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

where the $e_{a_i}$s are understood as elements of $T_0^1 V \cong_{\text{vec}} V$ and the $f^{b_i}s$ 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 $T_q^p V$.

**Change of basis**

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

$$\widetilde{e}_i = \sum_{j=1}^{d} A^j{}_i \, e_j \qquad \text{for } 1 \leq i \leq d.$$

for some $A^j{}_i \in K$. Similarly, we have

$$e_i = \sum_{j=1}^{d} B^j{}_i \, \widetilde{e}_j \qquad \text{for } 1 \leq i \leq d.$$

for some $B^j{}_i \in K$. It is a standard linear algebra result that the matrices $A$ and $B$, with entries $A^j{}_i$ and $B^j{}_i$ respectively, are invertible and, in fact, $A^{-1} = B$.

## 8.3 Notational conventions

**Einstein's summation convention**

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

$$v = v^i e_i \qquad \text{and} \qquad T = T^{ij}{}_k e_i \otimes e_j \otimes f^k$$

instead of

$$v = \sum_{i=1}^{d} v^i e_i \qquad \text{and} \qquad T = \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} T^{ij}{}_k e_i \otimes e_j \otimes f^k.$$

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. For example

$$v^i e_i = v^k e_k, \qquad T^{ij}{}_k = T^{imj}{}_{km}, \qquad A_{ij} = C_k C^k B_{ij} + C_i C^k B_{kj}$$

are all valid expressions, while

$$c^k v^i e_i = c^k v^k e_k, \qquad T^{ij}{}_j = T^{imj}{}_{km}, \qquad A_{ij} = C_k C^k + C_i C^l B_{lj}$$

are not. 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 upstair indices.

*Example* 8.23. Using the summation convention, we have:

a) $f^a(v) = f^a(v^b e_b) = v^b f^a(e_b) = v^b \delta^a_b = v^a$;

b) $\omega(e_b) = (\omega_a f^a)(e_b) = \omega_a f^a(e_b) = \omega_b$;

c) $\omega(v) = \omega_a f^a(v^b e_b) = \omega_a v^a$;

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

*Remark* 8.24. 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 $\phi\colon V \times W \to Z$, and let $v = v^i e_i \in V$ and $w = w^i \widetilde{e}_i \in W$. Then we have:

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

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.

*Remark* 8.25. 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 $\omega = \omega_i f^i \in V^*$ 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:

$$v = v^i e_i \quad \leftrightsquigarrow \quad v \,\widehat{=}\, \begin{pmatrix} v^1 \\ \vdots \\ v^d \end{pmatrix}$$

and

$$\omega = \omega_i f^i \quad \leftrightsquigarrow \quad \omega \,\widehat{=}\, (\omega_1, \ldots, \omega_d).$$

Given $\phi \in \mathrm{End}(V) \cong_{\mathrm{vec}} T^1_1 V$, recall that we can write $\phi = \phi^i{}_j\, e_i \otimes f^j$, where $\phi^i{}_j := \phi(f^i, e_j)$ 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:

$$\phi = \phi^i{}_j\, e_i \otimes f^j \quad \leftrightsquigarrow \quad \phi \,\widehat{=}\, \begin{pmatrix} \phi^1{}_1 & \phi^1{}_2 & \cdots & \phi^1{}_d \\ \phi^2{}_1 & \phi^2{}_2 & \cdots & \phi^2{}_d \\ \vdots & \vdots & \ddots & \vdots \\ \phi^d{}_1 & \phi^d{}_2 & \cdots & \phi^d{}_d \end{pmatrix}$$

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. Its usefulness stems from the following example.

*Example* 8.26. If $\dim V < \infty$, then we have $\mathrm{End}(V) \cong_{\mathrm{vec}} T_1^1 V$. Explicitly, if $\phi \in \mathrm{End}(V)$, 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 $\phi \in \mathrm{End}(V)$ are $\phi^a{}_b := f^a(\phi(e_b))$.

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

$$
\begin{aligned}
(\phi \circ \psi)^a{}_b &:= (\phi \circ \psi)(f^a, e_b) \\
&:= f^a((\phi \circ \psi)(e_b)) \\
&= f^a((\phi(\psi(e_b)))) \\
&= f^a(\phi(\psi^m{}_b e_m)) \\
&= \psi^m{}_b f^a(\phi(e_m)) \\
&:= \psi^m{}_b \phi^a{}_m.
\end{aligned}
$$

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 $\phi\psi$.

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.

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.

**Change of components under a change of basis**

Recall that if $\{e_a\}$ and $\{\widetilde{e}_a\}$ are basis of $V$, we have

$$\widetilde{e}_a = A^b{}_a e_b \qquad \text{and} \qquad e_a = B^m{}_a \widetilde{e}_m,$$

with $A^{-1} = B$. 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 $v = v^a e_a = \widetilde{v}^a \widetilde{e}_a \in V$. Then:

$$v^a = f^a(v) = f^a(\widetilde{v}^b \widetilde{e}_b) = \widetilde{v}^b f^a(\widetilde{e}_b) = \widetilde{v}^b f^a(A^m{}_b e_m) = A^m{}_b \widetilde{v}^b f^a(e_m) = A^a{}_b \widetilde{v}^b.$$

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

$$\omega_a := \omega(e_a) = \omega(B^m{}_a \widetilde{e}_m) = B^m{}_a \omega(\widetilde{e}_m) = B^m{}_a \widetilde{\omega}_m.$$

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

$$v^a = A^a{}_b \widetilde{v}^b \qquad \omega_a = B^b{}_a \widetilde{\omega}_b$$
$$\widetilde{v}^a = B^a{}_b v^b \qquad \widetilde{\omega}_a = A^b{}_a \omega_b$$

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

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

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

i.e. the upstair indices transform like vector indices, and the downstair indices transform like covector indices.

## 8.4 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.

*Remark* 8.27. Recall that, if $\phi \in T^1_1 V$, then we can arrange the components $\phi^a{}_b$ in matrix form:

$$\phi = \phi^a{}_b\, e_a \otimes f^b \quad \longleftrightarrow \quad \phi \,\hat{=}\, \begin{pmatrix} \phi^1{}_1 & \phi^1{}_2 & \cdots & \phi^1{}_d \\ \phi^2{}_1 & \phi^2{}_2 & \cdots & \phi^2{}_d \\ \vdots & \vdots & \ddots & \vdots \\ \phi^d{}_1 & \phi^d{}_2 & \cdots & \phi^d{}_d \end{pmatrix}$$

Similarly, if we have $g \in T^0_2 V$, its components are $g_{ab} := g(e_a, e_b)$ and we can write

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

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:

$$\phi \to A^{-1}\phi A \qquad \text{and} \qquad g \to A^T g A,$$

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, 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".

We will need some preliminary definitions.

**Definition.** Let $M$ be a set. A *permutation* of $M$ is a bijection $M \to M$.

**Definition.** The *symmetric group* of order $n$, denoted $S_n$, is the set of permutations of $\{1, \ldots, n\}$ under the operation of functional composition.

**Definition.** A *transposition* is a permutation which exchanges two elements, keeping all other elements fixed.

**Proposition 8.28.** *Every permutation $\pi \in S_n$ can be written as a product (composition) of transpositions in $S_n$.*

While this decomposition is not unique, for each given $\pi \in S_n$, the number of transpositions in its decomposition is always either even or odd. Hence, we can define the *sign* (or *signature*) of $\pi \in S_n$ as:

$$\mathrm{sgn}(\pi) := \begin{cases} +1 & \text{if } \pi \text{ is the product of an even number of transpositions} \\ -1 & \text{if } \pi \text{ is the product of an odd number of transpositions.} \end{cases}$$

**Definition.** 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.

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

Note that a 0-form is a scalar, and a 1-form is a covector. A $d$-form is also called a *top form*. There are several equivalent definitions of $n$-form. For instance, we have the following.

**Proposition 8.29.** *A $(0, n)$-tensor $\omega$ is an $n$-form if, and only if, $\omega(v_1, \ldots, v_n) = 0$ whenever $\{v_1, \ldots, v_n\}$ is linearly dependent.*

While $T_n^0 V$ is certainly non-empty when $n > d$, the proposition immediately implies that any $n$-form with $n > d$ must be identically zero. This is because a collection of more than $d$ vectors from a $d$-dimensional vector space is necessarily linearly dependent.

**Proposition 8.30.** *Denote by $\Lambda^n V$ the vector space of n-forms on $V$. Then we have*

$$\dim \Lambda^n V = \begin{cases} \binom{d}{n} & \text{if } 1 \le n \le d \\ 0 & \text{if } n > d, \end{cases}$$

*where $\binom{d}{n} = \frac{d!}{n!(d-n)!}$ is the binomial coefficient, read as "d choose n".*

In particular, $\dim \Lambda^d V = 1$. This means that

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

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

**Definition.** 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.** Let $\dim V = d$ and let $\omega \in \Lambda^d V$ be a volume form on $V$. Given $v_1, \ldots, v_d \in V$, the *volume* spanned by $v_1, \ldots, v_d$ is

$$\mathrm{vol}(v_1, \ldots, v_d) := \omega(v_1, \ldots, v_d).$$

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

*Remark* 8.31. 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.** Let $V$ be a $d$-dimensional vector space and let $\phi \in \mathrm{End}(V) \cong_{\mathrm{vec}} T^1_1 V$. The *determinant* of $\phi$ is

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

for some volume form $\omega \in \Lambda^d V$ and some basis $\{e_1, \ldots, e_d\}$ 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 $\omega, \omega' \in \Lambda^d V$, then there is a $c \in K$ such that $\omega = c\,\omega'$, and hence

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

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

Note that $\phi$ needs to be an endomorphism because we need to apply $\omega$ to $\phi(e_1), \ldots, \phi(e_d)$, and thus $\phi$ needs to output a vector.

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.

*Remark* 8.32. 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 $g \to A^T g A$. 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

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

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

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

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.

## 9 Differential structures: the pivotal concept of tangent vector spaces

### 9.1 Tangent spaces to a manifold

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.** Let $M$ be a manifold. We define the infinite-dimensional vector space over $\mathbb{R}$ with underlying set

$$\mathcal{C}^\infty(M) := \{f\colon M \to \mathbb{R} \mid f \text{ is smooth}\}$$

and operations defined pointwise, i.e. for any $p \in M$,

$$(f+g)(p) := f(p) + g(p)$$
$$(\lambda f)(p) := \lambda f(p).$$

A routine check shows that this is indeed a vector space. We can similarly define $\mathcal{C}^\infty(U)$, with $U$ an open subset of $M$.

**Definition.** A *smooth curve* on $M$ is a smooth map $\gamma\colon \mathbb{R} \to M$, where $\mathbb{R}$ is understood as a 1-dimensional manifold.

This definition also applies to smooth maps $I \to M$ for an open interval $I \subseteq \mathbb{R}$.

**Definition.** Let $\gamma\colon \mathbb{R} \to M$ 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

$$X_{\gamma,p}\colon \mathcal{C}^\infty(M) \xrightarrow{\sim} \mathbb{R}$$
$$f \mapsto (f \circ \gamma)'(0),$$

where $\mathbb{R}$ is understood as a 1-dimensional vector space over the field $\mathbb{R}$.

Note that $f \circ \gamma$ is a map $\mathbb{R} \to \mathbb{R}$, hence we can calculate the usual derivative and evaluate it at 0.

*Remark* 9.1. 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 $\gamma$ 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.** 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

$$T_pM := \{X_{\gamma,p} \mid \gamma \text{ is a smooth curve through } p\},$$

addition

$$\oplus \colon T_p M \times T_p M \to T_p M$$
$$(X_{\gamma,p}, X_{\delta,p}) \mapsto X_{\gamma,p} \oplus X_{\delta,p},$$

and scalar multiplication

$$\odot \colon \mathbb{R} \times T_p M \to T_p M$$
$$(\lambda, X_{\gamma,p}) \mapsto \lambda \odot X_{\gamma,p},$$

both defined pointwise, i.e. for any $f \in \mathcal{C}^\infty(M)$,

$$(X_{\gamma,p} \oplus X_{\delta,p})(f) := X_{\gamma,p}(f) + X_{\delta,p}(f)$$
$$(\lambda \odot X_{\gamma,p})(f) := \lambda X_{\gamma,p}(f).$$

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.

**Proposition 9.2.** *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

$$\begin{aligned}
X_{\sigma,p}(f) &:= (f \circ \sigma)'(0) \\
&= [f \circ x^{-1} \circ ((x \circ \gamma) + (x \circ \delta) - x(p))]'(0)
\end{aligned}$$

where $(f \circ x^{-1}) \colon \mathbb{R}^d \to \mathbb{R}$ and $((x \circ \gamma) + (x \circ \delta) - x(p)) \colon \mathbb{R} \to \mathbb{R}^d$, so by the multivariable chain rule

$$= [\partial_a(f \circ x^{-1})(x(p))]\,((x^a \circ \gamma) + (x^a \circ \delta) - x^a(p))'(0)$$

where $x^a$, with $1 \leq a \leq d$, are the component functions of $x$, and since the derivative is linear, we get

$$
\begin{aligned}
&= \left[\partial_a (f \circ x^{-1})(x(p))\right] \left((x^a \circ \gamma)'(0) + (x^a \circ \delta)'(0)\right) \\
&= (f \circ x^{-1} \circ x \circ \gamma)'(0) + (f \circ x^{-1} \circ x \circ \delta)'(0) \\
&= (f \circ \gamma)'(0) + (f \circ \delta)'(0) \\
&=: (X_{\gamma,p} \oplus X_{\delta,p})(f).
\end{aligned}
$$

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 $\lambda \odot X_{\gamma,p} = X_{\sigma,p} \in T_p M$. $\qquad\square$

*Remark* 9.3. We now give a slightly different (but equivalent) definition of $T_p M$. Consider the set of smooth curves

$$
S = \{\gamma \colon I \to M \mid \text{with } I \subseteq \mathbb{R} \text{ open, } 0 \in I \text{ and } \gamma(0) = p\}
$$

and define the equivalence relation $\sim$ on $S$

$$
\gamma \sim \delta \; :\Leftrightarrow \; (x \circ \gamma)'(0) = (x \circ \delta)'(0)
$$

for some (and hence every) chart $(U, x)$ containing $p$. Then, we can define

$$
T_p M := S/\!\sim .
$$

## 9.2 Algebras and derivations

Before we continue looking at properties of tangent spaces, we will have a short aside on algebras and derivations.

**Definition.** An *algebra* over a field $K$ is a quadruple $(A, +, \cdot, \bullet)$, where $(A, +, \cdot)$ is a $K$-vector space and $\bullet$ is a *product* on $A$, i.e. a ($K$-)bilinear map $\bullet \colon A \times A \to A$.

*Example* 9.4. Define a product on $\mathcal{C}^\infty(M)$ by

$$
\begin{aligned}
\bullet \colon \mathcal{C}^\infty(M) \times \mathcal{C}^\infty(M) &\to \mathcal{C}^\infty(M) \\
(f, g) &\mapsto f \bullet g,
\end{aligned}
$$

where $f \bullet g$ is defined pointwise. Then $(\mathcal{C}^\infty(M), +, \cdot, \bullet)$ is an algebra over $\mathbb{R}$.

The usual qualifiers apply to algebras as well.

**Definition.** 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 $\forall\, v, w \in A : v \bullet w = w \bullet v$.

*Example* 9.5. Clearly, $(\mathcal{C}^\infty(M), +, \cdot, \bullet)$ is an associative, unital, commutative algebra.

An important class of algebras are the so-called Lie algebras, in which the product $v \bullet w$ is usually denoted $[v, w]$.

**Definition.** A *Lie algebra* $A$ is an algebra whose product $[-, -]$, called *Lie bracket*, satisfies

i) antisymmetry: $\forall\, v \in A : [v, v] = 0$;

ii) 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

The antisymmetry condition immediately implies $[v, w] = -[w, v]$ for all $v, w \in A$, hence a (non-trivial) Lie algebra cannot be unital.

*Example* 9.6. Let $V$ be a vector space over $K$. Then $(\mathrm{End}(V), +, \cdot, \circ)$ is an associative, unital, non-commutative algebra over $K$. Define

$$[-, -] \colon \mathrm{End}(V) \times \mathrm{End}(V) \to \mathrm{End}(V)$$
$$(\phi, \psi) \mapsto [\phi, \psi] := \phi \circ \psi - \psi \circ \phi.$$

It is instructive to check that $(\mathrm{End}(V), +, \cdot, [-, -])$ is a Lie algebra over $K$. In this case, the Lie bracket is typically called the *commutator*.

In general, given an associative algebra $(A, +, \cdot, \bullet)$, if we define

$$[v, w] := v \bullet w - w \bullet v,$$

then $(A, +, \cdot, [-, -])$ is a Lie algebra.

**Definition.** Let $A$ be an algebra. A *derivation* on $A$ is a linear map $D \colon A \xrightarrow{\sim} A$ satisfying the Leibniz rule

$$D(v \bullet w) = D(v) \bullet w + v \bullet D(w)$$

for all $v, w \in A$.

*Remark* 9.7. The definition of derivation can be extended to include maps $A \to B$, with suitable structures. The obvious first attempt would be to consider two algebras $(A, +_A, \cdot_A, \bullet_A)$, $(B, +_B, \cdot_B, \bullet_B)$, and require $D \colon A \xrightarrow{\sim} B$ to satisfy

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

However, this is meaningless as it stands since $\bullet_B \colon B \times B \to B$, but on the right hand side $\bullet_B$ acts on elements from $A$ too. In order for this to work, $B$ needs to be a equipped with a product by elements of $A$, both from the left and from the right. The structure we are looking for is called a *bimodule* over $A$, and we will meet this later on.

*Example* 9.8. 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.

*Example* 9.9. Consider again the Lie algebra $(\mathrm{End}(V), +, \cdot, [-, -])$ and fix $\xi \in \mathrm{End}(V)$. If we define

$$D_\xi := [\xi, -] \colon \mathrm{End}(V) \xrightarrow{\sim} \mathrm{End}(V)$$
$$\phi \mapsto [\xi, \phi],$$

then $D_\xi$ is a derivation on $(\mathrm{End}(V), +, \cdot, [-, -])$ since it is linear and

$$
\begin{aligned}
D_\xi([\phi, \psi]) &:= [\xi, [\phi, \psi]] \\
&= -[\psi, [\xi, \phi]] - [\phi, [\psi, \xi]] && \text{(by the Jacobi identity)} \\
&= [[\xi, \phi], \psi] + [\phi, [\xi, \psi]] && \text{(by antisymmetry)} \\
&=: [D_\xi(\phi), \psi] + [\phi, D_\xi(\psi)].
\end{aligned}
$$

This construction works in general Lie algebras as well.

*Example* 9.10. We denote by $\mathrm{Der}_K(A)$ the set of derivations on a $K$-algebra $(A, +, \cdot, \bullet)$. This set can be endowed with a $K$-vector space structure by defining the operations pointwise but, by a previous example, it cannot be made into an algebra under composition of derivations.

However, derivations are maps, so we can still compose them as maps and define

$$[-, -] \colon \mathrm{Der}_K(A) \times \mathrm{Der}_K(A) \to \mathrm{Der}_K(A)$$
$$(D_1, D_2) \mapsto [D_1, D_2] := D_1 \circ D_2 - D_2 \circ D_1.$$

The map $[D_1, D_2]$ is (perhaps surprisingly) a derivation, since it is linear and

$$
\begin{aligned}
[D_1, D_2](v \bullet w) &:= (D_1 \circ D_2 - D_2 \circ D_1)(v \bullet w) \\
&= D_1(D_2(v \bullet w)) - D_2(D_1(v \bullet w)) \\
&= D_1(D_2(v) \bullet w + v \bullet D_2(w)) - D_2(D_1(v) \bullet w + v \bullet D_1(w)) \\
&= D_1(D_2(v) \bullet w) + D_1(v \bullet D_2(w)) - D_2(D_1(v) \bullet w) - D_2(v \bullet D_1(w)) \\
&= D_1(D_2(v)) \bullet w + \cancel{D_2(v) \bullet D_1(w)} + \cancel{D_1(v) \bullet D_2(w)} + v \bullet D_1(D_2(w)) \\
&\quad - D_2(D_1(v)) \bullet w - \cancel{D_1(v) \bullet D_2(w)} - \cancel{D_2(v) \bullet D_1(w)} - v \bullet D_2(D_1(w)) \\
&= (D_1(D_2(v)) - D_2(D_1(v))) \bullet w + v \bullet (D_1(D_2(w)) - D_2(D_1(w))) \\
&= [D_1, D_2](v) \bullet w + v \bullet [D_1, D_2](w)
\end{aligned}
$$

Then $(\mathrm{Der}_K(A), +, \cdot, [-, -])$ is a Lie algebra over $K$.

If we have a manifold, we can define the related notion of derivation on an open subset at a point.

**Definition.** 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 $D \colon \mathcal{C}^\infty(U) \xrightarrow{\sim} \mathbb{R}$ satisfying the Leibniz rule

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

We denote by $\mathrm{Der}_p(U)$ the $\mathbb{R}$-vector space of derivations on $U$ at $p$, with operations defined pointwise.

*Example* 9.11. The tangent vector $X_{\gamma,p}$ is a derivation on $U \subseteq M$ at $p$, where $U$ is any neighbourhood of $p$. In fact, our definition of the tangent space is equivalent to

$$T_p M := \mathrm{Der}_p(U),$$

for some open $U$ containing $p$. One can show that this does not depend on which neighbourhood $U$ of $p$ we pick.

## 9.3 A basis for the tangent space

The following is a crucially important result about tangent spaces.

**Theorem 9.12.** *Let $M$ be a manifold and let $p \in M$. Then*

$$\dim T_p M = \dim M.$$

*Remark* 9.13. 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 $x \colon U \to x(U) \subseteq \mathbb{R}^{\dim M}$, 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 $\gamma_{(a)} \colon \mathbb{R} \to U$ through $p$ by requiring $(x^b \circ \gamma_{(a)})(t) = \delta_a^b t$, i.e.

$$\gamma_{(a)}(0) := p$$
$$\gamma_{(a)}(t) := x^{-1} \circ (0, \ldots, 0, t, 0, \ldots, 0)$$

where the $t$ is in the $a^{\text{th}}$ position, with $1 \le a \le \dim M$. Let us calculate the action of the tangent vector $X_{\gamma_{(a)}, p} \in T_p M$ on an arbitrary function $f \in \mathcal{C}^\infty(U)$:

$$
\begin{aligned}
X_{\gamma_{(a)}, p}(f) &:= (f \circ \gamma_{(a)})'(0) \\
&= (f \circ \mathrm{id}_U \circ \gamma_{(a)})'(0) \\
&= (f \circ x^{-1} \circ x \circ \gamma_{(a)})'(0) \\
&= [\partial_b(f \circ x^{-1})(x(p))] \, (x^b \circ \gamma_{(a)})'(0) \\
&= [\partial_b(f \circ x^{-1})(x(p))] \, (\delta_a^b t)'(0) \\
&= [\partial_b(f \circ x^{-1})(x(p))] \, \delta_a^b \\
&= \partial_a(f \circ x^{-1})(x(p))
\end{aligned}
$$

We introduce a special notation for this tangent vector:

$$\left(\frac{\partial}{\partial x^a}\right)_p := X_{\gamma_{(a)},p},$$

where the $x$ refers to the chart map. We now claim that

$$\mathcal{B} = \left\{\left(\frac{\partial}{\partial x^a}\right)_p \in T_pM \;\middle|\; 1 \le a \le \dim M\right\}$$

is a basis of $T_pM$. First, we show that $\mathcal{B}$ spans $T_pM$.

Let $X \in T_pM$. Then, by definition, there exists some smooth curve $\sigma$ through $p$ such that $X = X_{\sigma,p}$. For any $f \in \mathcal{C}^\infty(U)$, we have

$$
\begin{aligned}
X(f) &= X_{\sigma,p}(f) \\
&:= (f \circ \sigma)'(0) \\
&= (f \circ x^{-1} \circ x \circ \sigma)'(0) \\
&= \left[\partial_b(f \circ x^{-1})(x(p))\right](x^b \circ \sigma)'(0) \\
&= (x^b \circ \sigma)'(0)\left(\frac{\partial}{\partial x^b}\right)_p(f).
\end{aligned}
$$

Since $(x^b \circ \sigma)'(0) =: X^b \in \mathbb{R}$, we have:

$$X = X^b\left(\frac{\partial}{\partial x^b}\right)_p,$$

i.e. any $X \in T_pM$ 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 $0 \in T_pM$.

Recall that, given the chart $(U, x)$, the coordinate maps $x^b\colon U \to \mathbb{R}$ are smooth, i.e. $x^b \in \mathcal{C}^\infty(U)$. Thus, we can feed them into the left hand side to obtain

$$
\begin{aligned}
0 &= \lambda^a\left(\frac{\partial}{\partial x^a}\right)_p(x^b) \\
&= \lambda^a\,\partial_a(x^b \circ x^{-1})(x(p)) \\
&= \lambda^a\,\partial_a(\mathrm{proj}_b)(x(p)) \\
&= \lambda^a\delta^b_a \\
&= \lambda^b
\end{aligned}
$$

i.e. $\lambda^b = 0$ for all $1 \le b \le \dim M$. So $\mathcal{B}$ is indeed a basis of $T_pM$, and since by construction $|\mathcal{B}| = \dim M$, the proof is complete. $\qquad\square$

*Remark* 9.14. 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* 9.15. 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.

*Remark* 9.16. 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 $x^a$, it is notationally consistent with it, in the following sense.

Let $M = \mathbb{R}^d$, $(U, x) = (\mathbb{R}^d, \mathrm{id}_{\mathbb{R}^d})$ 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 $x = x^{-1} = \mathrm{id}_{\mathbb{R}^d}$. Moreover, we have $\mathrm{proj}_a = x^a$. Thus, we can think of $x^1, \ldots, x^d$ as the independent variables of $f$, and we can then write

$$\left(\frac{\partial}{\partial x^a}\right)_p (f) = \frac{\partial f}{\partial x^a}(p).$$

**Definition.** Let $X \in T_p M$ be a tangent vector and let $(U, x)$ be a chart containing $p$. If

$$X = X^a \left(\frac{\partial}{\partial x^a}\right)_p,$$

then the real numbers $X^1, \ldots, X^{\dim M}$ are called the *components* of $X$ with respect to the tangent space basis induced by the chart $(U, x)$. The basis $\{\left(\frac{\partial}{\partial x^a}\right)_p\}$ is also called a *co-ordinate basis*.

**Proposition 9.17.** *Let $X \in T_p M$ and let $(U, x)$ and $(V, y)$ be two charts containing $p$. Then we have*
$$\left(\frac{\partial}{\partial x^a}\right)_p = \partial_a(y^b \circ x^{-1})(x(p)) \left(\frac{\partial}{\partial y^b}\right)_p$$

*Proof.* Assume w.l.o.g. that $U = V$. Since $\left(\frac{\partial}{\partial x^a}\right)_p \in T_p M$ and $\{\left(\frac{\partial}{\partial x^a}\right)_p\}$ forms a basis, we must have
$$\left(\frac{\partial}{\partial x^a}\right)_p = \lambda^b \left(\frac{\partial}{\partial y^b}\right)_p$$

for some $\lambda^b$. Let us determine what the $\lambda^b$ are by applying both sides of the equation to the coordinate maps $y^c$:

$$\left(\frac{\partial}{\partial x^a}\right)_p (y^c) = \partial_a(y^c \circ x^{-1})(x(p));$$

$$\lambda^b \left(\frac{\partial}{\partial y^b}\right)_p (y^c) = \lambda^b \, \partial_b(y^c \circ y^{-1})(y(p))$$

$$= \lambda^b \, \partial_b(\text{proj}_c)(y(p))$$

$$= \lambda^b \, \delta_b^c$$

$$= \lambda^c.$$

Hence

$$\lambda^c = \partial_a(y^c \circ x^{-1})(x(p)).$$

Substituting this expression for $\lambda^c$ gives the result. $\qquad\square$

**Corollary 9.18.** *Let $X \in T_pM$ and let $(U, x)$ and $(V, y)$ be two charts containing $p$. Denote by $X^a$ and $\widetilde{X}^a$ the coordinates of $X$ with respect to the tangent bases induced by the two charts, respectively. Then we have:*

$$\widetilde{X}^a = \partial_b(y^a \circ x^{-1})(x(p)) \, X^b.$$

*Proof.* Applying the previous result,

$$X = X^a \left(\frac{\partial}{\partial x^a}\right)_p = X^a \, \partial_a(y^b \circ x^{-1})(x(p)) \left(\frac{\partial}{\partial y^b}\right)_p.$$

Hence, we read-off $\widetilde{X}^b = \partial_a(y^b \circ x^{-1})(x(p)) \, X^a$. $\qquad\square$

*Remark* 9.19. By abusing notation, we can write the previous equations in a more familiar form. Denote by $y^b$ the maps $y^b \circ x^{-1} \colon x(U) \subseteq \mathbb{R}^{\dim M} \to \mathbb{R}$; these are real functions of $\dim M$ independent real variables. Since here we are only interested in what happens at the point $p \in M$, we can think of the maps $x^1, \ldots, x^{\dim M}$ as the independent variables of each of the $y^b$.

This is a general fact: if $\{*\}$ is a singleton (we let $*$ denote its unique element) and $x \colon \{*\} \to A$, $y \colon A \to B$ are maps, then $y \circ x$ is the same as the map $y$ with independent variable $x$. Intuitively, $x$ just "chooses" an element of $A$.

Hence, we have $y^b = y^b(x^1, \ldots, x^{\dim M})$ and we can write

$$\left(\frac{\partial}{\partial x^a}\right)_p = \frac{\partial y^b}{\partial x^a}(x(p)) \left(\frac{\partial}{\partial y^b}\right)_p \qquad \text{and} \qquad \widetilde{X}^b = \frac{\partial y^b}{\partial x^a}(x(p)) \, X^a,$$

which correspond to our earlier $e_a = A^b{}_a \widetilde{e}_b$ and $\widetilde{v}^b = A^b{}_a v^a$. The function $y = y(x)$ expresses the new co-ordinates in terms of the old ones, and $A^b{}_a$ is the *Jacobian* matrix of this map, evaluated at $x(p)$. The inverse transformation, of course, is given by

$$B^b{}_a = (A^{-1})^b{}_a = \frac{\partial x^b}{\partial y^a}(y(p)).$$

*Remark* 9.20. The formula for change of components of vectors under a change of chart suggests yet another way to define the tangent space to $M$ at $p$.

Let $\mathscr{A}_p : \{(U, x) \in \mathscr{A} \mid p \in U\}$ be the set of charts on $M$ containing $p$. A *tangent vector* $v$ at $p$ is a map

$$v : \mathscr{A}_p \to \mathbb{R}^{\dim M}$$

satisfying

$$v((V, y)) = A\, v((U, x))$$

where $A$ is the Jacobian matrix of $y \circ x^{-1} : \mathbb{R}^{\dim M} \to \mathbb{R}^{\dim M}$ at $x(p)$. In components, we have

$$[v((V, y))]^b = \frac{\partial y^b}{\partial x^a}(x(p))\, [v((U, x))]^a.$$

The tangent space $T_p M$ is then defined to be the set of all tangent vectors at $p$, endowed with the appropriate vector space structure.

What we have given above is the mathematically rigorous version of the definition of vector typically found in physics textbooks, i.e. that a vector is a "set of numbers" $v^a$ which, under a change of coordinates $y = y(x)$, transform as

$$\widetilde{v}^b = \frac{\partial y^b}{\partial x^a}\, v^a.$$

For a comparison of the different definitions of $T_p M$ that we have presented and a proof of their equivalence, refer to Chapter 2 of *Vector Analysis*, by Klaus Jänich.

## 10 Construction of the tangent bundle

### 10.1 Cotangent spaces and the differential

Since the tangent space is a vector space, we can do all the constructions we saw previously in the abstract vector space setting.

**Definition.** 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 $T_p M \cong_{\mathrm{vec}} T_p^* M$. If $\{\left(\frac{\partial}{\partial x^a}\right)_p\}$ is the basis of $T_p M$ induced by some chart $(U, x)$, then the dual basis is denoted as $\{(\mathrm{d}x^a)_p\}$. We have, by definition

$$(\mathrm{d}x^a)_p \left( \left(\frac{\partial}{\partial x^b}\right)_p \right) = \delta_b^a.$$

Once we have the cotangent space, we can define the tensor spaces.

**Definition.** Let $M$ be a manifold and $p \in M$. The *tensor space* $(T_s^r)_p M$ is defined as

$$(T_s^r)_p M := T_s^r(T_p M) = \underbrace{T_p M \otimes \cdots \otimes T_p M}_{r \text{ copies}} \otimes \underbrace{T_p^* M \otimes \cdots \otimes T_p^* M}_{s \text{ copies}}.$$

**Definition.** Let $M$ and $N$ be manifolds and let $\phi \colon M \to N$ be smooth. The *differential* (or *derivative*) *of* $\phi$ *at* $p \in M$ is the linear map

$$\mathrm{d}_p \phi \colon T_p M \xrightarrow{\sim} T_{\phi(p)} N$$
$$X \mapsto \mathrm{d}_p \phi\,(X)$$

where $\mathrm{d}_p \phi\,(X)$ is the tangent vector to $N$ at $\phi(p)$

$$\mathrm{d}_p \phi\,(X) \colon \mathcal{C}^\infty(N) \xrightarrow{\sim} \mathbb{R}$$
$$g \mapsto (\mathrm{d}_p \phi\,(X))(g) := X(g \circ \phi).$$

If this definition looks confusing, it is worth it to pause and think about what it is saying. Intuitively, if $\phi$ takes us from $M$ to $N$, then $\mathrm{d}_p \phi$ takes us from $T_p M$ to $T_{\phi(p)} N$. The way in which it does so, is the following.



Given $X \in T_p M$, we want to construct $\mathrm{d}_p \phi\,(X) \in T_{\phi(p)} N$, i.e. a derivation on $N$ at $f(p)$. Derivations act on functions. So, given $g \colon N \to \mathbb{R}$, we want to construct a real number by using $\phi$ and $X$. There is really only one way to do is. If we precompose $g$ with $\phi$, we obtain $g \circ \phi \colon M \to \mathbb{R}$, which is an element of $\mathcal{C}^\infty(M)$. We can then happily apply $X$ to this function to obtain a real number. You should check that $\mathrm{d}_p \phi\,(X)$ is indeed a tangent vector to $N$.

*Remark* 10.1. Note that, to be careful, we should replace $\mathcal{C}^\infty(M)$ and $\mathcal{C}^\infty(N)$ above with $\mathcal{C}^\infty(U)$ and $\mathcal{C}^\infty(V)$, where $U \subseteq M$ and $V \subseteq N$ are open and contain $p$ and $\phi(p)$, respectively.

*Example* 10.2. If $M = \mathbb{R}^d$ and $N = \mathbb{R}^{d'}$, then the differential of $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ at $p \in \mathbb{R}^d$

$$\mathrm{d}_p f : T_p \mathbb{R}^d \cong_{\text{vec}} \mathbb{R}^d \to T_{f(p)} \mathbb{R}^{d'} \cong_{\text{vec}} \mathbb{R}^{d'}$$

is none other than the Jacobian of $f$ at $p$.

A special case of the differential is the gradient of a function in $\mathcal{C}^\infty(M)$.

**Definition.** Let $M$ be a manifold and let $f : M \to \mathbb{R}$ be smooth. The *gradient of $f$ at* $p \in M$ is the covector

$$\mathrm{d}_p f : T_p M \xrightarrow{\sim} T_{f(p)} \mathbb{R} \cong_{\text{vec}} \mathbb{R}$$
$$X \mapsto \mathrm{d}_p f(X) := X(f).$$

In fact, we can define the *gradient operator at $p \in M$* as the $\mathbb{R}$-linear map

$$\mathrm{d}_p : \mathcal{C}^\infty(U) \xrightarrow{\sim} T_p^* M$$
$$f \mapsto \mathrm{d}_p f,$$

with $p \in U \subseteq M$.

*Remark* 10.3. Note that, by writing $\mathrm{d}_p f(X) := X(f)$, we have committed a slight (but nonetheless real) abuse of notation. Since $\mathrm{d}_p f(X) \in T_{f(p)} \mathbb{R}$, it takes in a function and return a real number, but $X(f)$ is already a real number! This is due to the fact that we have implicitly employed the isomorphism

$$\iota_d : T_p \mathbb{R}^d \to \mathbb{R}^d$$
$$X \mapsto (X(\mathrm{proj}_1), \ldots, X(\mathrm{proj}_d)),$$

which, when $d = 1$, reads

$$\iota_1 : T_p \mathbb{R} \to \mathbb{R}$$
$$X \mapsto X(\mathrm{id}_\mathbb{R}).$$

In our case, we have

$$\mathrm{d}_p f(X) := X(- \circ f) \mapsto X(\mathrm{id}_\mathbb{R} \circ f) = X(f).$$

This notwithstanding, the best way to think of $\mathrm{d}_p f$ is as a covector, i.e. $\mathrm{d}_p f$ takes in a tangent vector $X$ and returns the real number $X(f)$, in a linear fashion.

Recall that if $(U, x)$ is a chart on $M$, then the co-ordinate maps $x^a : U \to x(U) \subseteq \mathbb{R}^{\dim M}$ are smooth functions on $U$. We can thus apply the gradient operator $\mathrm{d}_p$ (with $p \in U$) to each of them to obtain $(\dim M)$-many elements of $T_p^* M$.

**Proposition 10.4.** *Let $(U, x)$ be a chart on $M$, with $p \in U$. The set $\mathcal{B} = \{\mathrm{d}_p x^a \mid 1 \le a \le \dim M\}$ forms a basis of $T_p^* M$.*

*Proof.* We already know that $T_p^* M = \dim M$, since it is the dual space to $T_p M$. As $|\mathcal{B}| = \dim M$ by construction, it suffices to show that it is linearly independent. Suppose that

$$\lambda_a \, \mathrm{d}_p x^a = 0,$$

for some $\lambda_a \in \mathbb{R}$. Applying the left hand side to the basis element $\left(\frac{\partial}{\partial x^b}\right)_p$ yields

$$
\begin{aligned}
\lambda_a \, \mathrm{d}_p x^a \left( \left(\frac{\partial}{\partial x^b}\right)_p \right) &= \lambda_a \left(\frac{\partial}{\partial x^b}\right)_p (x^a) && \text{(definition of } \mathrm{d}_p x^a) \\
&= \lambda_a \, \partial_b (x^a \circ x^{-1})(x(p)) && \text{(definition of } \left(\frac{\partial}{\partial x^b}\right)_p) \\
&= \lambda_a \, \partial_b (\mathrm{proj}_a)(x(p)) \\
&= \lambda_a \, \delta_b^a \\
&= \lambda_b.
\end{aligned}
$$

Therefore, $\mathcal{B}$ is linearly independent and hence a basis of $T_p^* M$. Moreover, since we have shown that

$$\mathrm{d}_p x^a \left( \left(\frac{\partial}{\partial x^b}\right)_p \right) = \delta_b^a,$$

this basis is, in fact, the dual basis to $\{ \left(\frac{\partial}{\partial x^a}\right)_p \}$. $\qquad \square$

*Remark* 10.5. Note a slight subtlety. Given a chart $(U, x)$ and the induced basis $\{ \left(\frac{\partial}{\partial x^a}\right)_p \}$ of $T_p M$, the dual basis to $\{ \left(\frac{\partial}{\partial x^a}\right)_p \}$ exists simply by virtue of $T_p^* M$ being the dual space to $T_p M$. What we have shown above is that the elements of this dual basis are given explicitly by the gradients of the co-ordinate maps of $(U, x)$. In our notation, we have

$$(\mathrm{d} x^a)_p = \mathrm{d}_p x^a, \qquad 1 \leq a \leq \dim M.$$

## 10.2 Push-forward and pull-back

The push-forward of a smooth map $\phi \colon M \to N$ at $p \in M$ is just another name for the differential of $\phi$ at $p$. We give the definition again in order to establish the new notation.

**Definition.** Let $\phi \colon M \to N$ be a smooth map between smooth manifolds. The *push-forward* of $\phi$ at $p \in M$ is the linear map:

$$
\begin{aligned}
(\phi_*)_p \colon T_p M &\xrightarrow{\sim} T_{\phi(p)} N \\
X &\mapsto (\phi_*)_p(X) := X(- \circ \phi).
\end{aligned}
$$

If $\gamma \colon \mathbb{R} \to M$ is a smooth curve on $M$ and $\phi \colon M \to N$ is smooth, then $\phi \circ \gamma \colon \mathbb{R} \to N$ is a smooth curve on $N$.

**Proposition 10.6.** *Let $\phi \colon M \to N$ 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 $(\phi_*)_p(X_{\gamma,p}) = X_{\phi \circ \gamma, \phi(p)}$. $\qquad\square$

Related to the push-forward, there is the notion of pull-back of a smooth map.

**Definition.** Let $\phi\colon M \to N$ be a smooth map between smooth manifolds. The *pull-back* of $\phi$ at $p \in M$ is the linear map:

$$
\begin{aligned}
(\phi^*)_p\colon T^*_{\phi(p)}N &\xrightarrow{\sim} T^*_p M \\
\omega &\mapsto (\phi^*)_p(\omega),
\end{aligned}
$$

where $(\phi^*)_p(\omega)$ is defined as

$$
\begin{aligned}
(\phi^*)_p(\omega)\colon T_p M &\xrightarrow{\sim} \mathbb{R} \\
X &\mapsto \omega((\phi_*)_p(X)),
\end{aligned}
$$

In words, if $\omega$ is a covector on $N$, its pull-back $(\phi^*)_p(\omega)$ 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 $\omega$ to them to produce a real number.

*Remark* 10.7. If you don't see it immediately, then you should spend some time proving that all the maps that we have defined so far and *claimed* to be linear are, in fact, linear.

*Remark* 10.8. We have seen that, given a smooth $\phi\colon M \to N$, we can push a vector $X \in T_p M$ forward to a vector $(\phi_*)_p(X) \in T_{\phi(p)}N$, and pull a covector $\omega \in T^*_{\phi(p)}N$ back to a covector $(\phi^*)_p(\omega) \in T^*_p M$.



However, if $\phi\colon M \to N$ is a diffeomorphism, then we can also pull a vector $Y \in T_{\phi(p)}N$ back to a vector $(\phi^*)_p(Y) \in T_p M$, and push a covector $\eta \in T^*_p M$ forward to a covector $(\phi_*)_p(\eta) \in T^*_{\phi(p)}N$, by using $\phi^{-1}$ as follows:

$$
\begin{aligned}
(\phi^*)_p(Y) &:= ((\phi^{-1})_*)_{\phi(p)}(Y) \\
(\phi_*)_p(\eta) &:= ((\phi^{-1})^*)_{\phi(p)}(\eta).
\end{aligned}
$$

$$\mathcal{C}^\infty(M) \xrightarrow{\ -\circ\phi^{-1}\ } \mathcal{C}^\infty(N) \qquad\qquad T_pM \xleftarrow{\ ((\phi^{-1})_*)_{\phi(p)}\ } T_{\phi(p)}N$$

$$(\phi^*)_p(Y) \searrow \qquad \downarrow Y \qquad\qquad \eta \downarrow \qquad \swarrow (\phi_*)_p(\eta)$$

$$\mathbb{R} \qquad\qquad\qquad \mathbb{R}$$

This is only possible if $\phi$ is a diffeomorphism. In general, you should keep in mind that

<div align="center">

*Vectors are pushed forward,*
*covectors are pulled back.*

</div>

*Remark* 10.9. Given a smooth map $\phi\colon M \to N$, if $f \in \mathcal{C}^\infty(N)$, then $f \circ \phi$ is often called the pull-back of $f$ along $\phi$. Similarly, if $\gamma$ is a curve on $M$, then $\phi \circ \gamma$ is called the push-forward of $\gamma$ along $\phi$. For example, we can say that

- the push-forward of a tangent vector acting on a function is the tangent vector acting on the pull-back of the function;

- the push-forward of a tangent vector to a curve is the tangent vector to the push-forward of the curve.

## 10.3 Immersions and 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.** A smooth map $\phi\colon M \to N$ is said to be an *immersion* of $M$ into $N$ if the derivative

$$\mathrm{d}_p\phi \equiv (\phi_*)_p\colon T_pM \xrightarrow{\ \sim\ } T_{\phi(p)}N$$

is injective, for all $p \in M$. The manifold $M$ is said to be an *immersed submanifold* of $N$.

From the theory of linear algebra, we immediately deduce that, for $\phi\colon M \to N$ 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* 10.10. Consider the map $\phi\colon S^1 \to \mathbb{R}^2$ whose image is reproduced below.
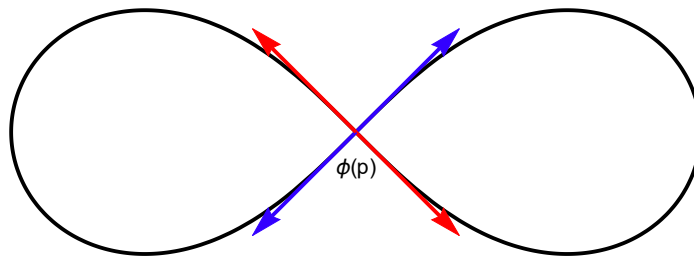
The map $\phi$ is not injective, i.e. there are $p, q \in S^1$, with $p \neq q$ and $\phi(p) = \phi(q)$. Of course, this means that $T_{\phi(p)}\mathbb{R}^2 = T_{\phi(q)}\mathbb{R}^2$. However, the maps $(\phi_*)_p$ and $(\phi_*)_q$ are both injective, with their images being represented by the blue and red arrows, respectively. Hence, the map $\phi$ is immersion.

**Definition.** A smooth map $\phi \colon M \to N$ is said to be a *(smooth) embedding* of $M$ into $N$ if

- $\phi \colon M \to N$ is an immersion;

- $M \cong_{\text{top}} \phi(M) \subseteq N$, where $\phi(M)$ carries the subset topology inherited from $N$.

The manifold $M$ is said to be an *embedded submanifold* of $N$.

*Remark* 10.11. If a continuous map between topological spaces satisfies the second condition above, then it is called a *topological embedding*. Therefore, a smooth embedding is a topological embedding which is also an immersion (as opposed to simply being a smooth 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 10.12** (Whitney). *Any smooth manifold $M$ can be*

- *embedded in $\mathbb{R}^{2\dim M}$;*

- *immersed in $\mathbb{R}^{2\dim M - 1}$.*

*Example* 10.13. The Klein bottle can be embedded in $\mathbb{R}^4$ but not in $\mathbb{R}^3$. It can, however, be immersed in $\mathbb{R}^3$.

What we have presented above is referred to as the *strong* version of Whitney's theorem. There is a weak version as well, but there are also even stronger versions of this result, such as the following.

**Theorem 10.14.** *Any smooth manifold can be immersed in $\mathbb{R}^{2\dim M - a(\dim M)}$, where $a(n)$ is the number of $1$s in a binary expansion of $n \in \mathbb{N}$.*

*Example* 10.15. If $\dim M = 3$, then as

$$3_{10} = (1 \times 2^1 + 1 \times 2^0)_{10} = 11_2,$$

we have $a(\dim M) = 2$, and thus every 3-dimensional manifold can be immersed into $\mathbb{R}^4$. Note that even the strong version of Whitney's theorem only tells us that we can immerse $M$ into $\mathbb{R}^5$.

## 10.4 The tangent bundle

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 $T_pM$. 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.

**Definition.** Given a smooth manifold $M$, the *tangent bundle* of $M$ is the disjoint union of all the tangent spaces to $M$, i.e.

$$TM := \coprod_{p \in M} T_pM,$$

equipped with the canonical projection map

$$\pi \colon TM \to M$$
$$X \mapsto p,$$

where $p$ is the unique $p \in M$ such that $X \in T_pM$.

We now 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 $\mathscr{A}_M$ be a smooth atlas on $M$ and let $(U, x) \in \mathscr{A}_M$. If $X \in \mathrm{preim}_\pi(U) \subseteq TM$, then $X \in T_{\pi(X)}M$, by definition of $\pi$. Moreover, since $\pi(X) \in U$, we can expand $X$ in terms of the basis induced by the chart $(U, x)$:

$$X = X^a \left( \frac{\partial}{\partial x^a} \right)_{\pi(X)},$$

where $X^1, \dots, X^{\dim M} \in \mathbb{R}$. We can then define the map

$$\xi \colon \mathrm{preim}_\pi(U) \to x(U) \times \mathbb{R}^{\dim M} \cong_{\mathrm{set}} \mathbb{R}^{2 \dim M}$$
$$X \mapsto (x(\pi(X)), X^1, \dots, X^{\dim M}).$$

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 $(\mathrm{preim}_\pi(U), \xi)$ is a chart on $TM$ and

$$\mathscr{A}_{TM} := \{ (\mathrm{preim}_\pi(U), \xi) \mid (U, x) \in \mathscr{A}_M \}$$

is a smooth atlas on $TM$. Note that, from its definition, it is clear that $\xi$ is a bijection. We will not show that $(\mathrm{preim}_\pi(U), \xi)$ is a chart here, but we will show that $\mathscr{A}_{TM}$ is a smooth atlas.

**Proposition 10.16.** *Any two charts* $(\mathrm{preim}_\pi(U), \xi), (\mathrm{preim}_\pi(\widetilde{U}), \widetilde{\xi}) \in \mathscr{A}_{TM}$ *are* $\mathcal{C}^\infty$-*compatible.*

*Proof.* Let $(U, x)$ and $(\widetilde{U}, \widetilde{x})$ be the two charts on $M$ giving rise to $(\mathrm{preim}_\pi(U), \xi)$ and $(\mathrm{preim}_\pi(\widetilde{U}), \widetilde{\xi})$, respectively. We need to show that the map

$$\widetilde{\xi} \circ \xi^{-1} \colon x(U \cap \widetilde{U}) \times \mathbb{R}^{\dim M} \to \widetilde{x}(U \cap \widetilde{U}) \times \mathbb{R}^{\dim M}$$

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, $\widetilde{\xi} \circ \xi^{-1}$ acts as

$$\widetilde{x} \circ x^{-1} \colon x(U \cap \widetilde{U}) \to \widetilde{x}(U \cap \widetilde{U})$$
$$x(p) \mapsto \widetilde{x}(p),$$

while on the remaining $\dim M$ components it acts as the change of vector components we met previously, i.e.
$$X^a \mapsto \widetilde{X}^a = \partial_b(y^a \circ x^{-1})(x(p)) \, X^b.$$

Hence, we have

$$\widetilde{\xi} \circ \xi^{-1} \colon \qquad x(U \cap \widetilde{U}) \times \mathbb{R}^{\dim M} \to \widetilde{x}(U \cap \widetilde{U}) \times \mathbb{R}^{\dim M}$$
$$(x(\pi(X)), X^1, \ldots, X^{\dim M}) \mapsto (\widetilde{x}(\pi(X)), \widetilde{X}^1, \ldots, \widetilde{X}^{\dim M}),$$

which is smooth in each component, and hence smooth. $\qquad\square$

The tangent bundle of a smooth manifold $M$ is therefore itself a smooth manifold of dimension $2\dim M$, and the projection $\pi \colon TM \to M$ is smooth with respect to this structure.

Similarly, one can construct the *cotangent bundle* $T^*M$ to $M$ by defining

$$T^*M := \coprod_{p \in M} T^*_p M$$

and going through the above again, using the dual basis $\{(\mathrm{d}x^a)_p\}$ instead of $\{\left(\frac{\partial}{\partial x^a}\right)_p\}$.

## 11 Tensor space theory II: over a ring

### 11.1 Vector fields

Now that we have defined the tangent bundle, we are ready to define vector fields.

**Definition.** Let $M$ be a smooth manifold, and let $TM \xrightarrow{\pi} M$ be its tangent bundle. A *vector field* on $M$ is a smooth section of the tangent bundle, i.e. a smooth map $\sigma \colon M \to TM$ such that $\pi \circ \sigma = \mathrm{id}_M$.

$$
\begin{array}{c}
TM \\
\sigma \Big\uparrow \Big\downarrow \pi \\
M
\end{array}
$$

We denote the set of all vector fields on $M$ by $\Gamma(TM)$, i.e.

$$\Gamma(TM) := \{\sigma \colon M \to TM \mid \sigma \text{ is smooth and } \pi \circ \sigma = \mathrm{id}_M\}.$$

This is, in fact, the standard notation for the set of all sections on a bundle.

*Remark* 11.1. An equivalent definition is that a vector field $\sigma$ on $M$ is a derivation on the algebra $\mathcal{C}^\infty(M)$, i.e. an $\mathbb{R}$-linear map

$$\sigma \colon \mathcal{C}^\infty(M) \xrightarrow{\sim} \mathcal{C}^\infty(M)$$

satisfying the Leibniz rule (with respect to pointwise multiplication on $\mathcal{C}^\infty(M)$)

$$\sigma(fg) = g\,\sigma(f) + f\,\sigma(g).$$

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.

*Example* 11.2. Let $(U, x)$ be a chart on $M$. For each $1 \leq a \leq \dim M$, the map

$$
\begin{aligned}
\sigma \colon U &\to TU \\
p &\mapsto \left(\frac{\partial}{\partial x^a}\right)_p
\end{aligned}
$$

is a vector field on the submanifold $U$. We can also think of this as a linear map

$$
\begin{aligned}
\frac{\partial}{\partial x^a} \colon \mathcal{C}^\infty(U) &\xrightarrow{\sim} \mathcal{C}^\infty(U) \\
f &\mapsto \frac{\partial}{\partial x^a}(f) = \partial_a(f \circ x^{-1}) \circ x.
\end{aligned}
$$

By abuse of notation, one usually denotes the right hand side above simply as $\partial_a f$.

Recall that, given a smooth map $\phi\colon M \to N$, the push-forward $(\phi_*)_p$ is a linear map that takes in a tangent vector in $T_pM$ and outputs a tangent vector in $T_{\phi(p)}N$. Of course, we have one such map for each $p \in M$. We can collect them all into a single smooth map.

**Definition.** Let $\phi\colon M \to N$ be smooth. The *push-forward* $\phi_*$ is defined as

$$\phi_*\colon TM \to TN$$
$$X \mapsto (\phi_*)_{\pi(X)}(X).$$

Any vector $X \in TM$ must belong to $T_pM$ for some $p \in M$, namely $p = \pi(X)$. The map $\phi_*$ simply takes any vector $X \in TM$ and applies the usual push-forward at the "right" point, producing a vector in $TN$. One can similarly define $\phi^*\colon T^*N \to T^*M$.

The ideal next step would be to try to construct a map $\Phi_*\colon \Gamma(TM) \to \Gamma(TN)$ that allows us to push vector fields on $M$ forward to vector fields on $N$. Given $\sigma \in \Gamma(TM)$, we would like to construct a $\Phi_*(\sigma) \in \Gamma(TN)$. This is not trivial, since $\Phi_*(\sigma)$ needs to be, in particular, a smooth map $N \to TN$. Note that the composition $\phi_* \circ \sigma$ is a map $M \to TN$, and hence $\mathrm{im}_{\phi_* \circ \sigma}(M) \subseteq TN$. Thus, we can try to define $\Phi_*(\sigma)$ by mapping each $p \in N$ to some tangent vector in $\mathrm{im}_{\phi_* \circ \sigma}(M)$. Unfortunately, there are at least two ways in which this can go wrong.

1. The map $\phi$ may fail to be injective. Then, there would be two points $p_1, p_2 \in M$ such that $p_1 \neq p_2$ and $\phi(p_1) = \phi(p_2) =: q \in N$. Hence, we would have two tangent vectors on $N$ with base-point $q$, namely $(\phi_* \circ \sigma)(p_1)$ and $(\phi_* \circ \sigma)(p_2)$. These two *need not* be equal, and if they are not then the map $\Phi_*(\sigma)$ is ill-defined at $q$.

2. The map $\phi$ may fail to be surjective. Then, there would be some $q \in N$ such that there is no $X \in \mathrm{im}_{\phi_* \circ \sigma}(M)$ with $\pi(X) = q$ (where $\pi\colon TN \to N$). The map $\Phi_*(\sigma)$ would then be undefined at $q$.

3. Even if the map $\phi$ is bijective, its inverse $\phi^{-1}$ may fail to be smooth. But then $\Phi_*(\sigma)$ would not be guaranteed to be a smooth map.

Of course, everything becomes easier if $\phi\colon M \to N$ is a diffeomorphism.

$$
\begin{array}{ccc}
TM & \xrightarrow{\ \phi_*\ } & TN \\[4pt]
{\scriptstyle\sigma}\uparrow & & \uparrow{\scriptstyle\Phi_*(\sigma)} \\[4pt]
M & \xrightarrow{\ \phi\ } & N
\end{array}
$$

If $\sigma \in \Gamma(TM)$, we can define the *push-forward* $\Phi_*(\sigma) \in \Gamma(TN)$ as

$$\Phi_*(\sigma) := \phi_* \circ \sigma \circ \phi^{-1}.$$

More generally, if $\phi\colon M \to N$ is smooth and $\sigma \in \Gamma(TM)$, $\tau \in \Gamma(TN)$, we can define $\Phi_*(\sigma) = \tau$ if $\sigma$ and $\tau$ are *$\phi$-related*, i.e. if they satisfy

$$\tau \circ \phi = \phi_* \circ \sigma.$$

We can equip the set $\Gamma(TM)$ with the following operations. The first is our, by now familiar, pointwise addition:

$$\oplus \colon \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$$
$$(\sigma, \tau) \mapsto \sigma \oplus \tau,$$

where

$$\sigma \oplus \tau \colon M \to \Gamma(TM)$$
$$p \mapsto (\sigma \oplus \tau)(p) := \sigma(p) + \tau(p).$$

Note that the $+$ on the right hand side above is the addition in $T_pM$. More interestingly, we also define the following multiplication operation:

$$\odot \colon \mathcal{C}^\infty(M) \times \Gamma(TM) \to \Gamma(TM)$$
$$(f, \sigma) \mapsto f \odot \sigma,$$

where

$$f \odot \sigma \colon M \to \Gamma(TM)$$
$$p \mapsto (f \odot \sigma)(p) := f(p)\sigma(p).$$

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 $T_pM$.

If we consider the triple $(\mathcal{C}^\infty(M), +, \bullet)$, where $\bullet$ is pointwise function multiplication as defined in the section on algebras and derivations, then the triple $(\Gamma(TM), \oplus, \odot)$ 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 $T_pM$;

- $\Gamma(TM) \setminus \{0\}$ satisfies

  i) $\forall\, f \in \mathcal{C}^\infty(M) : \forall\, \sigma, \tau \in \Gamma(TM) \setminus \{0\} : f \odot (\sigma \oplus \tau) = (f \odot \sigma) \oplus (f \odot \tau)$;

  ii) $\forall\, f, g \in \mathcal{C}^\infty(M) : \forall\, \sigma \in \Gamma(TM) \setminus \{0\} : (f + g) \odot \sigma = (f \odot \sigma) \oplus (g \odot \sigma)$;

  iii) $\forall\, f, g \in \mathcal{C}^\infty(M) : \sigma \in \Gamma(TM) \setminus \{0\} : (f \bullet g) \odot \sigma = f \odot (g \odot \sigma)$;

  iv) $\forall\, \sigma \in \Gamma(TM) \setminus \{0\} : 1 \odot \sigma = \sigma$,

  where $1 \in \mathcal{C}^\infty(M)$ maps every $p \in M$ to $1 \in \mathbb{R}$.

These are precisely the axioms for a vector space! The only obstacle to saying that $\Gamma(TM)$ is a vector space over $\mathcal{C}^\infty(M)$ is that the triple $(\mathcal{C}^\infty(M), +, \bullet)$ is *not* an algebraic field, but only a ring. We could simply talk about "vector spaces over rings", but vector spaces over ring have wildly different properties than vector spaces over fields, so much so that they have their own name: *modules*.

*Remark* 11.3. 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 $\Gamma(TM)$. However, a basis for this vector space is necessarily uncountably infinite, and hence it does not provide a very useful decomposition for our vector fields.

Instead, the operation $\odot$ that we have defined 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 within the module structure.

## 11.2   Rings and modules over a ring

Unlike mathematicians, most people who apply mathematics tend to consider rings and modules somewhat esoteric objects, but they are not esoteric at all. As we have seen, they arise naturally in the study of manifolds and their unusual properties, at least when compared to fields and vector spaces, are of direct geometric relevance and make us understand the subject better.

For your benefit, we first recall some basic facts about rings.

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

- $(R, +)$ is an abelian group:

    i) $\forall\, a, b, c \in R : (a + b) + c = a + (b + c)$;

    ii) $\exists\, 0 \in R : \forall\, a \in R : a + 0 = 0 + a = a$;

    iii) $\forall\, a \in R : \exists\, {-a} \in R : a + (-a) = (-a) + a = 0$;

    iv) $\forall\, a, b \in R : a + b = b + a$;

- the operation $\cdot$ is associative and distributes over addition:

    v) $\forall\, a, b, c \in R : (a \cdot b) \cdot c = a \cdot (b \cdot c)$;

    vi) $\forall\, a, b, c \in R : (a + b) \cdot c = a \cdot c + b \cdot c$;

    vii) $\forall\, a, b, c \in R : a \cdot (b + c) = a \cdot b + a \cdot c$.

Note that since $\cdot$ is not required to be commutative, axioms vi and vii are both necessary.

**Definition.** 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* 11.4. The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all rings under the usual operations. They are also all fields, except $\mathbb{Z}$.

*Example* 11.5. Let $M$ be a smooth manifold. Then

- $(\mathcal{C}^\infty(M), +, \cdot)$, where $\cdot$ is scalar multiplication (by a real number), is an $\mathbb{R}$-vector space. It is not a ring since $\cdot$ is not a map $\mathcal{C}^\infty(M) \times \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M)$.

- $(\mathcal{C}^\infty(M), +, \bullet)$, where $\bullet$ is pointwise multiplication of maps, is a commutative, unital ring, but not a division ring and hence, not a field.

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

**Definition.** Let $(R, +, \cdot)$ be a unital ring. A triple $(M, \oplus, \odot)$ is called an *R-module* if the maps

$$\oplus \colon M \times M \to M$$
$$\odot \colon R \times M \to M$$

satisfy the vector space axioms, i.e. $(M, \oplus)$ is an abelian group and for all $r, s \in R$ and all $m, n \in M$, we have

i) $r \odot (m \oplus n) = (r \odot m) \oplus (r \odot n)$;

ii) $(r + s) \odot m = (r \odot m) \oplus (s \odot m)$;

iii) $(r \cdot s) \odot m = r \odot (s \odot m)$;

iv) $1 \odot m = m$.

Most definitions we had for vector spaces carry over unaltered to modules, including that of a basis, i.e. a linearly independent spanning set.

*Remark* 11.6. Even though we will not need this, we note as an aside that what we have defined above is a *left* $R$-module, since multiplication has only been defined (and hence only makes sense) on the left. The definition of a *right* $R$-module is completely analogous. Moreover, if $R$ and $S$ are two unital rings, then we can define $M$ to be an *R-S-bimodule* if it is a left $R$-module and a right $S$-module. The bimodule structure is precisely what is needed to generalise the notion of derivation that we have met before.

*Example* 11.7. Any ring $R$ is trivially a module over itself.

*Example* 11.8. The triple $(\Gamma(TM), \oplus, \odot)$ is a $\mathcal{C}^\infty(M)$-module.

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

## 11.3 Bases for 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.

**Theorem 11.9.** *If $D$ is a division ring, then any $D$-module $V$ admits a basis.*

**Corollary 11.10.** *Every vector space has a basis, since any field is also a division ring.*

Before we delve into the proof, let us consider some geometric examples.

*Example* 11.11.    a) Let $M = \mathbb{R}^2$ and consider $v \in \Gamma(T\mathbb{R}^2)$. It is a fact from standard vector analysis that any such $v$ can be written uniquely as

$$v = v^1 e_1 + v^2 e_2$$

for some $v^1, v^2 \in \mathcal{C}^\infty(\mathbb{R}^2)$ and $e_1, e_2 \in \Gamma(T\mathbb{R}^2)$. Hence, even though $\Gamma(T\mathbb{R}^2)$ is a $\mathcal{C}^\infty(\mathbb{R}^2)$-module and $\mathcal{C}^\infty(\mathbb{R}^2)$ is *not* a division ring, it still has a basis. Note that the coefficients in the linear expansion of $v$ are functions.

This example shows that the converse to the above theorem is not true: if $D$ is not a division ring, then a $D$-module may or may not have a basis.

b) Let $M = S^2$. A famous result in algebraic topology, known as the *hairy ball theorem*, states that there is no non-vanishing smooth tangent vector field on even-dimensional $n$-spheres. Hence, we can multiply any smooth vector field $v \in \Gamma(TS^2)$ by a function $f \in \mathcal{C}^\infty(S^2)$ which is zero everywhere except where $v$ is, obtaining $fv = 0$ despite $f \neq 0$ and $v \neq 0$. Therefore, there is no set of linearly independent vector fields on $S^2$, much less a basis.

The proof of the theorem requires the axiom of choice, in the equivalent form known as Zorn's lemma.

**Lemma 11.12** (Zorn)**.** *A partially ordered set $P$ whose every totally ordered subset $T$ has an upper bound in $P$ contains a maximal element.*

Of course, we now need to define the new terms that appear in the statement of Zorn's lemma.

**Definition.** A *partially ordered set* (*poset* for short) is a pair $(P, \leq)$ where $P$ is a set and $\leq$ is a *partial order* on $P$, i.e. a relation on $P$ satisfying

i) *reflexivity:*  $\forall\, a \in P : a \leq a$;

ii) *anti-symmetry:*  $\forall\, a, b \in P : (a \leq b \wedge b \leq a) \Rightarrow a = b$;

iii) *transitivity:*  $\forall\, a, b, c \in P : (a \leq b \wedge b \leq c) \Rightarrow a \leq c$.

In a partially ordered set, while every element is related to itself by reflexivity, two distinct elements need not be related.

**Definition.** A *totally ordered set* is a pair $(P, \leq)$ where $P$ is a set and $\leq$ is a *total order* on $P$, i.e. a relation on $P$ satisfying

   a) *anti-symmetry:* $\forall\, a, b \in P : (a \leq b \wedge b \leq a) \Rightarrow a = b$;

   b) *transitivity:* $\forall\, a, b, c \in P : (a \leq b \wedge b \leq c) \Rightarrow a \leq c$;

   c) *totality:* $\forall\, a, b \in P : a \leq b \vee b \leq a$.

Note that a total order is a special case of a partial order since, by letting $a = b$ in the totality condition, we obtain reflexivity. In a totally ordered set, every pair of elements is related.

**Definition.** Let $(P, \leq)$ be a partially ordered set and let $T \subseteq P$. An element $u \in P$ is said to be an *upper bound* for $T$ if
$$\forall\, t \in T : \ t \leq u.$$

Every single-element subset of a partially ordered set has at least one upper bound by reflexivity. However, in general, a subset may not have any upper bound. For example, if the subset contains an element which is not related to any other element.

**Definition.** Let $(P, \leq)$ be a partially ordered set. A *maximal element* of $P$ is an element $m \in P$ such that
$$\nexists\, a \in P : \ m \leq a$$

or, alternatively,
$$\forall\, a \in P : \ (m \leq a) \Rightarrow (m = a).$$

Note that this is *not* equivalent to
$$\forall\, a \in P : \ a \leq m$$

unless $(P, \leq)$ is a totally ordered set.

We are now ready to prove the theorem.

*Proof of* **??**. We will tackle this one step at a time.

   a) Let $S \subseteq V$ be a generating set of $V$, i.e.
$$\forall\, v \in V : \exists\, e_1, \ldots, e_N \in S : \exists\, d^1, \ldots, d^N \in D : \ v = d^a e_a.$$

      A generating set always exists, as one may take $S = V$.

   b) Define a partially ordered set $(P, \leq)$ by
$$P := \{U \in \mathcal{P}(S) \mid U \text{ is linearly independent}\}$$

      and $\leq := \subseteq$, that is, we partial order by set-theoretic inclusion.

c) Let $T \subseteq P$ be any totally ordered subset of $P$. Then $\bigcup T$ is an upper bound for $T$, and it is linearly independent (by the total ordering assumption). Hence $\bigcup T \in P$.

By Zorn's lemma, $P$ has a maximal element. Le $\mathcal{B} \in P$ be any such element. BY construction, $\mathcal{B}$ is a maximal (with respect to inclusion) linearly independent subset of the generating set $S$.

d) We now claim that $S = \operatorname{span}_D(\mathcal{B})$. Indeed, let $v \in S \setminus \mathcal{B}$. Then $\mathcal{B} \cap \{v\} \in \mathcal{P}(S)$. Since $\mathcal{B}$ is maximal, the set $\mathcal{B} \cap \{v\}$ is *not* linearly independent. Hence

$$\exists\, e_1, \dots, e_N \in \mathcal{B} : \exists\, d, d^1, \dots, d^N \in D : \ d^a e_a + dv = 0,$$

where the coefficients $d, d^1, \dots, d^N$ are not all zero. In particular, $d \neq 0$, for if it was, it would immediately follow that $d^a e_a = 0$ for some $d^1, \dots, d^N$, not all zero, contradicting the linear independence of $\mathcal{B}$.

Since $D$ is a division ring and $d \neq 0$, there exists a multiplicative inverse for $d$. Then we can multiply both sides of the above equation by $d^{-1} \in D$ to obtain

$$v = (d^{-1} \cdot d^a)\, e_a.$$

Hence $S = \operatorname{span}_D(\mathcal{B})$.

e) We therefore have
$$V = \operatorname{span}_D(S) = \operatorname{span}_D(\mathcal{B})$$

and thus $\mathcal{B}$ is a basis of $V$. $\qquad\square$

We stress again that if $D$ is not a division ring, then a $D$-module may, but need not, have a basis.

## 11.4 Module constructions and applications

As for vector spaces, we can perform the usual constructions with modules as well.

**Definition.** 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 componentwise.

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.** An $R$-module $M$ is said to be

- *finitely generated* if it has a finite generating set;

- *free* is 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* 11.13. As we have seen, $\Gamma(T\mathbb{R}^2)$ is free while $\Gamma(TS^2)$ is not.

*Example* 11.14. Clearly, every free module is also projective.

**Definition.** Let $M$ and $N$ be two (left) $R$-modules. A map $f\colon M \to N$ is said to be an *R-linear map*, or an *R-module homomorphism*, 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 $N$.

A bijective module homomorphism is said to be a *module isomorphism*, and we write $M \cong_{\text{mod}} N$ if there exists a module isomorphism between them.

If $M$ and $N$ are right $R$-modules, then the linearity condition is written as

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

**Proposition 11.15.** *If a finitely generated module $R$-module $F$ is free, and $d \in \mathbb{N}$ is the cardinality of a finite basis, then*

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

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 11.16** (Serre, Swan, et al.)**.** *Let $E$ be a vector fibre bundle over a smooth manifold $M$. Then, the set $\Gamma(E)$ of all smooth section of $E$ over $M$ is a finitely generated, projective $\mathcal{C}^\infty(M)$-module.*

A vector fibre bundle is a fibre bundle in which every fibre is a vector space. An example is the tangent bundle to a manifold.

*Remark* 11.17. An immediate consequence of the theorem is that, for any vector fibre bundle $E$ over $M$, there exists a $\mathcal{C}^\infty(M)$-module $Q$ such that the direct sum $\Gamma(E) \oplus Q$ is free. If $Q$ can be chosen to be the trivial module $\{0\}$, then $\Gamma(E)$ is itself free, as it is the case with $\Gamma(T\mathbb{R}^2)$. In a sense, the module $Q$ quantifies the failure of $\Gamma(E)$ to have a basis.

**Theorem 11.18.** *Let $P, Q$ be finitely generated (projective) modules over a commutative ring $R$. Then*

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

*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. For instance

$$\operatorname{Hom}_{\mathcal{C}^\infty(M)}(\Gamma(TM), \mathcal{C}^\infty(M)) =: \Gamma(TM)^*.$$

One can show that $\Gamma(TM)^*$ coincides with the smooth sections over the cotangent bundle $\Gamma(T^*M)$, i.e. the covector fields. Recall that, just like a vector field, a covector field is a smooth section of the cotangent bundle $T^*M$, that is, a smooth map $\omega\colon M \to T^*M$ with $\pi \circ \omega = \text{id}_M$. Unlike what we had with vector fields, we can always define the pull-back of a covector field along any smooth map between manifolds.

**Definition.** Let $\phi\colon M \to N$ be smooth and let $\omega \in \Gamma(T^*N)$. We define the *pull-back* $\Phi^*(\omega) \in \Gamma(T^*M)$ of $\omega$ as

$$\Phi^*(\omega)\colon M \to T^*M$$
$$p \mapsto \Phi^*(\omega)(p),$$

where

$$\Phi^*(\omega)(p)\colon T_pM \xrightarrow{\sim} \mathbb{R}$$
$$X \mapsto \Phi^*(\omega)(p)(X) := \omega(\phi(p))(\phi_*(X)),$$

as in the following diagram

$$
\begin{array}{ccc}
T^*M & \xleftarrow{\quad \phi^* \quad} & T^*N \\
\Big\uparrow{\scriptstyle \Phi^*(\omega)} & & \Big\uparrow{\scriptstyle \omega} \\
M & \xrightarrow{\quad \phi \quad} & N
\end{array}
$$

We can, of course, generalise these ideas by constructing the $(r,s)$ tensor bundle of $M$

$$T^r_s M := \coprod_{p \in M} (T^r_s)_p M$$

and hence define a type $(r,s)$ tensor field on $M$ to be an element of $\Gamma(T^r_s M)$, i.e. a smooth section of the tensor bundle. If $\phi\colon M \to N$ is smooth, we can define the pull-back of contravariant (i.e. type $(0,q)$) tensor fields by generalising the pull-back of covector fields. If $\phi\colon M \to N$ is a diffeomorphism, then we can define the pull-back of any smooth $(p,q)$ tensor field $\tau \in \Gamma(T^r_s N)$ as

$$\Phi^*(\tau)(p)(\omega_1, \ldots, \omega_r, X_1, \ldots, X_s)$$
$$:= \tau(\phi(p))\big((\phi^{-1})^*(\omega_1), \ldots, (\phi^{-1})^*(\omega_r), \phi_*(X_1), \ldots, \phi_*(X_s)\big),$$

with $\omega_i \in T^*_p M$ and $X_i \in T_p M$.

There is, however, an equivalent characterisation of tensor fields as genuine multilinear maps. This is, in fact, the standard textbook definition.

**Definition.** Let $M$ be a smooth manifold. A smooth $(r,s)$ *tensor field* $\tau$ on $M$ is a $\mathcal{C}^\infty(M)$-multilinear map

$$\tau\colon \underbrace{\Gamma(T^*M) \times \cdots \times \Gamma(T^*M)}_{r \text{ copies}} \times \underbrace{\Gamma(TM) \times \cdots \times \Gamma(TM)}_{s \text{ copies}} \to \mathcal{C}^\infty(M).$$

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 $\omega(X) \in \mathcal{C}^\infty(M)$ by

$$(\omega(X))(p) := \omega(p)(X(p)).$$

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 $\omega \colon \Gamma(TM) \xrightarrow{\sim} \mathcal{C}^\infty(M)$ is $\mathcal{C}^\infty(M)$-linear.

Similarly, the set $\Gamma(T^r_s M)$ 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

$$\otimes \colon \Gamma(T^p_q M) \times \Gamma(T^r_s M) \to \Gamma(T^{p+r}_{q+s} M)$$
$$(\tau, \sigma) \mapsto \tau \otimes \sigma$$

analogously to what we had with tensors on a vector space, i.e.

$$(\tau \otimes \sigma)(\omega_1, \ldots, \omega_p, \omega_{p+1}, \ldots, \omega_{p+r}, X_1, \ldots, X_q, X_{q+1}, \ldots, X_{q+s})$$
$$:= \tau(\omega_1, \ldots, \omega_p, X_1, \ldots, X_q) \, \sigma(\omega_{p+1}, \ldots, \omega_{p+r}, X_{q+1}, \ldots, X_{q+s}),$$

with $\omega_i \in \Gamma(T^*M)$ 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 $T_pM$, 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.

## 12 Grassmann algebra and de Rham cohomology

### 12.1 Differential forms

**Definition.** 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.

$$\omega(X_1, \ldots, X_n) = \text{sgn}(\pi)\,\omega(X_{\pi(1)}, \ldots, X_{\pi(n)}),$$

for any $\pi \in S_n$, with $X_i \in \Gamma(TM)$.

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 $T_pM$.

*Example* 12.1.    a) A manifold $M$ is said to be *orientable* if it admits an oriented atlas, i.e. an atlas in which all chart transition maps, which are maps between open subsets of $\mathbb{R}^{\dim M}$, have a positive determinant.

     If $M$ is orientable, then there exists a nowhere vanishing top form ($n = \dim M$) on $M$ providing the volume.

   b) 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.

   c) In classical mechanics, if $Q$ is a smooth manifold describing the possible system configurations, then the phase space is $T^*Q$. There exists a canonically defined 2-form on $T^*Q$ known as a symplectic form, which we will define later.

If $\omega$ is an $n$-form, then $n$ is said to be the *degree* of $\omega$. 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.

*Example* 12.2. Of course, we have $\Omega^0(M) \equiv \mathcal{C}^\infty(M)$ and $\Omega^1(M) \equiv \Gamma(T^0_1 M) \equiv \Gamma(T^*M)$.

Similarly to the case of forms on vector spaces, we have $\Omega^n(M) = \{0\}$ for $n > \dim M$, and otherwise $\dim \Omega^n(M) = \binom{\dim M}{n}$, as a $\mathcal{C}^\infty(M)$-module.

We can specialise the pull-back of tensors to differential forms.

**Definition.** Let $\phi \colon M \to N$ 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

$$\Phi^*(\omega) \colon M \to T^*M$$
$$p \mapsto \Phi^*(\omega)(p),$$

where

$$\Phi^*(\omega)(p)(X_1, \ldots, X_n) := \omega(\phi(p))\big(\phi_*(X_1), \ldots, \phi_*(X_n)\big),$$

for $X_i \in T_pM$.

The map $\Phi^*\colon \Omega^n(N) \to \Omega^n(M)$ is $\mathbb{R}$-linear, and its action on $\Omega^0(M)$ is simply

$$\Phi^*\colon \Omega^0(M) \to \Omega^0(M)$$
$$f \mapsto \Phi^*(f) := f \circ \phi.$$

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. Hence, we define the following.

**Definition.** Let $M$ be a smooth manifold. We define the *wedge* (or *exterior*) *product* of forms as the map

$$\wedge\colon \Omega^n(M) \times \Omega^m(M) \to \Omega^{n+m}(M)$$
$$(\omega, \sigma) \mapsto \omega \wedge \sigma,$$

where

$$(\omega \wedge \sigma)(X_1, \ldots, X_{n+m}) := \frac{1}{n!\,m!} \sum_{\pi \in S_{n+m}} \mathrm{sgn}(\pi)(\omega \otimes \sigma)(X_{\pi(1)}, \ldots, X_{\pi(n+m)})$$

and $X_1, \ldots, X_{n+m} \in \Gamma(TM)$. By convention, for any $f, g \in \Omega^0(M)$ and $\omega \in \Omega^n(M)$, we set

$$f \wedge g := fg \qquad \text{and} \qquad f \wedge \omega = \omega \wedge f = f\omega.$$

*Example* 12.3. 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* 12.4. If $(U, x)$ is a chart on $M$, then every $n$-form $\omega \in \Omega^n(U)$ can be expressed locally on $U$ as

$$\omega = \omega_{a_1 \cdots a_n}\,\mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n},$$

where $\omega_{a_1 \cdots a_n} \in \mathcal{C}^\infty(U)$ and $1 \le a_1 < \cdots < a_n \le \dim M$. The $\mathrm{d}x^{a_i}$ appearing above are the covector fields (1-forms)

$$\mathrm{d}x^{a_i}\colon p \mapsto \mathrm{d}_p x^{a_i}.$$

The pull-back distributes over the wedge product.

**Theorem 12.5.** *Let $\phi\colon M \to N$ 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 $X_1, \ldots, X_{n+m} \in T_pM$. Then we have

$$
\begin{aligned}
\big(\Phi^*(\omega) &\wedge \Phi^*(\sigma)\big)(p)(X_1, \ldots, X_{n+m}) \\
&= \frac{1}{n!\,m!} \sum_{\pi \in S_{n+m}} \mathrm{sgn}(\pi)\big(\Phi^*(\omega) \otimes \Phi^*(\sigma)\big)(p)(X_{\pi(1)}, \ldots, X_{\pi(n+m)}) \\
&= \frac{1}{n!\,m!} \sum_{\pi \in S_{n+m}} \mathrm{sgn}(\pi)\Phi^*(\omega)(p)(X_{\pi(1)}, \ldots, X_{\pi(n)}) \\
&\qquad\qquad\qquad\qquad \Phi^*(\sigma)(p)(X_{\pi(n+1)}, \ldots, X_{\pi(n+m)}) \\
&= \frac{1}{n!\,m!} \sum_{\pi \in S_{n+m}} \mathrm{sgn}(\pi)\omega(\phi(p))\big(\phi_*(X_{\pi(1)}), \ldots, \phi_*(X_{\pi(n)})\big) \\
&\qquad\qquad\qquad\qquad \sigma(\phi(p))\big(\phi_*(X_{\pi(n+1)}), \ldots, \phi_*(X_{\pi(n+m)})\big) \\
&= \frac{1}{n!\,m!} \sum_{\pi \in S_{n+m}} \mathrm{sgn}(\pi)(\omega \otimes \sigma)(\phi(p))\big(\phi_*(X_{\pi(1)}), \ldots, \phi_*(X_{\pi(n+m)})\big) \\
&= (\omega \wedge \sigma)(\phi(p))\big(\phi_*(X_1), \ldots, \phi_*(X_{n+m})\big) \\
&= \Phi^*(\omega \wedge \sigma)(p)(X_1, \ldots, X_{n+m}).
\end{aligned}
$$

Since $p \in M$ was arbitrary, the statement follows. $\qquad\square$

## 12.2 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.** Let $M$ be a smooth manifold. Define the $\mathcal{C}^\infty(M)$-module

$$\mathrm{Gr}(M) \equiv \Omega(M) := \bigoplus_{n=0}^{\dim M} \Omega^n(M).$$

The *Grassmann algebra* on $M$ is the algebra $(\Omega(M), +, \cdot, \wedge)$, where

$$\wedge\colon \Omega(M) \times \Omega(M) \to \Omega(M)$$

is the linear continuation of the previously defined $\wedge\colon \Omega^n(M) \times \Omega^m(M) \to \Omega^{n+m}(M)$.

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* 12.6. 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* 12.7. 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 12.8.** *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 is $\omega \in \Omega^n(M)$ and $\sigma \in \Omega^m(M)$, then locally on a chart $(U, x)$ we can write

$$\omega = \omega_{a_1 \cdots a_n} \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n}$$
$$\sigma = \sigma_{b_1 \cdots b_m} \mathrm{d}x^{b_1} \wedge \cdots \wedge \mathrm{d}x^{b_m}$$

with $1 \leq a_1 < \cdots < a_n \leq \dim M$ and similarly for the $b_i$. The coefficients $\omega_{a_1 \cdots a_n}$ and $\sigma_{b_1 \cdots b_m}$ are smooth functions in $\mathcal{C}^\infty(U)$. Since $\mathrm{d}x^{a_i}, \mathrm{d}x^{b_j} \in \Omega^1(M)$, we have

$$\begin{aligned}
\omega \wedge \sigma &= \omega_{a_1 \cdots a_n} \sigma_{b_1 \cdots b_m} \, \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n} \wedge \mathrm{d}x^{b_1} \wedge \cdots \wedge \mathrm{d}x^{b_m} \\
&= (-1)^n \, \omega_{a_1 \cdots a_n} \sigma_{b_1 \cdots b_m} \, \mathrm{d}x^{b_1} \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n} \wedge \mathrm{d}x^{b_2} \wedge \cdots \wedge \mathrm{d}x^{b_m} \\
&= (-1)^{2n} \, \omega_{a_1 \cdots a_n} \sigma_{b_1 \cdots b_m} \, \mathrm{d}x^{b_1} \wedge \mathrm{d}x^{b_2} \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n} \wedge \mathrm{d}x^{b_3} \wedge \cdots \wedge \mathrm{d}x^{b_m} \\
&\;\;\vdots \\
&= (-1)^{nm} \, \omega_{a_1 \cdots a_n} \sigma_{b_1 \cdots b_m} \, \mathrm{d}x^{b_1} \wedge \cdots \wedge \mathrm{d}x^{b_m} \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n} \\
&= (-1)^{nm} \, \sigma \wedge \omega
\end{aligned}$$

since we have swapped 1-forms $nm$-many times. $\qquad\square$

*Remark* 12.9. 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

$$\varphi \wedge \psi = \cdots \psi \wedge \varphi$$

does not make sense in principle, because the different parts of $\varphi$ and $\psi$ can have different commutation behaviours.

## 12.3   The exterior derivative

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:

$$\mathrm{d} \colon \mathcal{C}^\infty(M) \xrightarrow{\sim} \Gamma(T^*M)$$
$$f \mapsto \mathrm{d}f$$

where, of course, $\mathrm{d}f \colon p \mapsto \mathrm{d}_p f$. Alternatively, we can think of $\mathrm{d}f$ as the $\mathbb{R}$-linear map

$$\mathrm{d}f \colon \Gamma(TM) \xrightarrow{\sim} \mathcal{C}^\infty(M)$$
$$X \mapsto \mathrm{d}f(X) = X(f).$$

*Remark* 12.10. Locally on some chart $(U, x)$ on $M$, the covector field (or 1-form) $\mathrm{d}f$ can be expressed as

$$\mathrm{d}f = \lambda_a \, \mathrm{d}x^a$$

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

$$\mathrm{d}f\left(\frac{\partial}{\partial x^b}\right) = \frac{\partial}{\partial x^b}(f) = \partial_b f$$

and

$$\lambda_a \, \mathrm{d}x^a\left(\frac{\partial}{\partial x^b}\right) = \lambda_a \frac{\partial}{\partial x^b}(x^a) = \lambda_a \, \delta_b^a = \lambda_b.$$

Hence, the local expression of $\mathrm{d}f$ on $(U, x)$ is

$$\mathrm{d}f = \partial_a f \, \mathrm{d}x^a.$$

Note that the operator $\mathrm{d}$ satisfies the Leibniz rule

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

We can also understand this as an operator that takes in 0-forms and outputs 1-forms

$$\mathrm{d} \colon \Omega^0(M) \xrightarrow{\sim} \Omega^1(M).$$

This can then be extended to an operator which acts on any $n$-form. We will need the following definition.

**Definition.** Let $M$ be a smooth manifold and let $X, Y \in \Gamma(TM)$. The commutator (or Lie bracket) of $X$ and $Y$ is defined as

$$[X, Y] \colon \mathcal{C}^\infty(M) \xrightarrow{\sim} \mathcal{C}^\infty(M)$$
$$f \mapsto [X, Y](f) := X(Y(f)) - Y(X(f)),$$

where we are using the definition of vector fields as $\mathbb{R}$-linear maps $\mathcal{C}^\infty(M) \xrightarrow{\sim} \mathcal{C}^\infty(M)$.

**Definition.** The *exterior derivative* on $M$ is the $\mathbb{R}$-linear operator

$$\mathrm{d}\colon \Omega^n(M) \xrightarrow{\sim} \Omega^{n+1}(M)$$
$$\omega \mapsto \mathrm{d}\omega$$

with $\mathrm{d}\omega$ being defined as

$$
\begin{aligned}
\mathrm{d}\omega(X_1,\ldots,X_{n+1}) :=& \sum_{i=1}^{n+1}(-1)^{i+1}\, X_i\big(\omega(X_1,\ldots,\widehat{X_i},\ldots,X_{n+1})\big) \\
&+ \sum_{i<j}(-1)^{i+j}\,\omega\big([X_i,X_j],X_1,\ldots,\widehat{X_i},\ldots,\widehat{X_j},\ldots,X_{n+1}\big),
\end{aligned}
$$

where $X_i \in \Gamma(TM)$ and the hat denotes omissions.

*Remark* 12.11. 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* 12.12. In the case $n = 1$, the form $\mathrm{d}\omega \in \Omega^2(M)$ is given by

$$\mathrm{d}\omega(X,Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y]).$$

Let us check that this is indeed a 2-form, i.e. an antisymmetric, $\mathcal{C}^\infty(M)$-multilinear map

$$\mathrm{d}\omega\colon \Gamma(TM) \times \Gamma(TM) \to \mathcal{C}^\infty(M).$$

By using the antisymmetry of the Lie bracket, we immediately get

$$\mathrm{d}\omega(X,Y) = -\,\mathrm{d}\omega(Y,X).$$

Moreover, thanks to this identity, it suffices to check $\mathcal{C}^\infty(M)$-linearity in the first argument only. Additivity is easily checked

$$
\begin{aligned}
\mathrm{d}\omega(X_1+X_2,Y) &= (X_1+X_2)(\omega(Y)) - Y(\omega(X_1+X_2)) - \omega([X_1+X_2,Y]) \\
&= X_1(\omega(Y)) + X_2(\omega(Y)) - Y(\omega(X_1)+\omega(X_2)) - \omega([X_1,Y]+[X_2,Y]) \\
&= X_1(\omega(Y)) + X_2(\omega(Y)) - Y(\omega(X_1)) - Y(\omega(X_2)) - \omega([X_1,Y]) - \omega([X_2,Y]) \\
&= \mathrm{d}\omega(X_1,Y) + \mathrm{d}\omega(X_2,Y).
\end{aligned}
$$

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

$$\begin{aligned}
\mathrm{d}\omega(fX, Y) &= fX(\omega(Y)) - Y(\omega(fX)) - \omega([fX, Y]) \\
&= fX(\omega(Y)) - Y(f\omega(X)) - \omega(f[X, Y] - Y(f)X) \\
&= fX(\omega(Y)) - fY(\omega(X)) - Y(f)\omega(X) - f\omega([X, Y]) + \omega(Y(f)X) \\
&= fX(\omega(Y)) - fY(\omega(X)) - \cancel{Y(f)\omega(X)} - f\omega([X, Y]) + \cancel{Y(f)\omega(X)} \\
&= f\,\mathrm{d}\omega(X, Y),
\end{aligned}$$

which is what we wanted.

The exterior derivative satisfies a graded version of the Leibniz rule with respect to the wedge product.

**Theorem 12.13.** *Let $\omega \in \Omega^n(M)$ and $\sigma \in \Omega^m(M)$. Then*

$$\mathrm{d}(\omega \wedge \sigma) = \mathrm{d}\omega \wedge \sigma + (-1)^n\,\omega \wedge \mathrm{d}\sigma.$$

*Proof.* We will work in local coordinates. Let $(U, x)$ be a chart on $M$ and write

$$\begin{aligned}
\omega &= \omega_{a_1 \cdots a_n}\mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n} =: \omega_A \mathrm{d}x^A \\
\sigma &= \sigma_{b_1 \cdots b_m}\mathrm{d}x^{b_1} \wedge \cdots \wedge \mathrm{d}x^{b_m} =: \sigma_B \mathrm{d}x^B.
\end{aligned}$$

Locally, the exterior derivative operator d acts as

$$\mathrm{d}\omega = \mathrm{d}\omega_A \wedge \mathrm{d}x^A.$$

Hence

$$\begin{aligned}
\mathrm{d}(\omega \wedge \sigma) &= \mathrm{d}(\omega_A \sigma_B\,\mathrm{d}x^A \wedge \mathrm{d}x^B) \\
&= \mathrm{d}(\omega_A \sigma_B) \wedge \mathrm{d}x^A \wedge \mathrm{d}x^B \\
&= (\sigma_B \mathrm{d}\omega_A + \omega_A \mathrm{d}\sigma_B) \wedge \mathrm{d}x^A \wedge \mathrm{d}x^B \\
&= \sigma_B \mathrm{d}\omega_A \wedge \mathrm{d}x^A \wedge \mathrm{d}x^B + \omega_A \mathrm{d}\sigma_B \wedge \mathrm{d}x^A \wedge \mathrm{d}x^B \\
&= \sigma_B \mathrm{d}\omega_A \wedge \mathrm{d}x^A \wedge \mathrm{d}x^B + (-1)^n \omega_A \mathrm{d}x^A \wedge \mathrm{d}\sigma_B \wedge \mathrm{d}x^B \\
&= \sigma_B \mathrm{d}\omega \wedge \mathrm{d}x^B + (-1)^n \omega_A \mathrm{d}x^A \wedge \mathrm{d}\sigma \\
&= \mathrm{d}\omega \wedge \sigma + (-1)^n\,\omega \wedge \mathrm{d}\sigma
\end{aligned}$$

since we have "anticommuted" the 1-form $\mathrm{d}\sigma_B$ through the $n$-form $\mathrm{d}x^A$, picking up $n$ minus signs in the process. $\qquad\square$

An important property of the exterior derivative is the following.

**Theorem 12.14.** *Let $\phi\colon M \to N$ be smooth. For any $\omega \in \Omega^n(N)$, we have*

$$\Phi^*(\mathrm{d}\omega) = \mathrm{d}(\Phi^*(\omega)).$$

*Proof (sketch).* We first show that this holds for 0-forms (i.e. smooth functions).

Let $f \in \mathcal{C}^{\infty}(N)$, $p \in M$ and $X \in T_p M$. Then

$$
\begin{aligned}
\Phi^*(\mathrm{d}f)(p)(X) &= \mathrm{d}f(\phi(p))(\phi_*(X)) & &\text{(definition of } \Phi^*) \\
&= \phi_*(X)(f) & &\text{(definition of } \mathrm{d}f) \\
&= X(f \circ \phi) & &\text{(definition of } \phi_*) \\
&= \mathrm{d}(f \circ \phi)(p)(X) & &\text{(definition of } \mathrm{d}(f \circ \phi)) \\
&= \mathrm{d}(\Phi^*(f))(p)(X) & &\text{(definition of } \Phi^*),
\end{aligned}
$$

so that we have $\Phi^*(\mathrm{d}f) = \mathrm{d}(\Phi^*(f))$.

The general result follows from the linearity of $\Phi^*$ and the fact that the pull-back distributes over the wedge product. $\qquad\square$

*Remark* 12.15. Informally, we can write this result as $\Phi^*\mathrm{d} = \mathrm{d}\Phi^*$, 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 $\mathrm{d}\colon \Omega^n(N) \to \Omega^{n+1}(N)$, while it is $\mathrm{d}\colon \Omega^n(M) \to \Omega^{n+1}(M)$ on the right hand side.

*Remark* 12.16. Of course, we could also combine the operators d into a single operator acting on the Grassmann algebra on $M$

$$
\mathrm{d}\colon \Omega(M) \to \Omega(M)
$$

by linear continuation.

*Example* 12.17. In the modern formulation of Maxwell's electrodynamics, the electric and magnetic fields $E$ and $B$ are taken to be a 1-form and a 2-form on $\mathbb{R}^3$, respectively:

$$
E := E_x \mathrm{d}x + E_y \mathrm{d}y + E_z \mathrm{d}z
$$
$$
B := B_x \mathrm{d}y \wedge \mathrm{d}z + B_y \mathrm{d}z \wedge \mathrm{d}x + B_z \mathrm{d}x \wedge \mathrm{d}y.
$$

The electromagnetic field strength $F$ is then defined as the 2-form on $\mathbb{R}^4$

$$
F := B + E \wedge \mathrm{d}t.
$$

In components, we can write

$$
F = F_{\mu\nu}\, \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu,
$$

where $(\mathrm{d}x^0, \mathrm{d}x^1, \mathrm{d}x^2, \mathrm{d}x^3) \equiv (\mathrm{d}t, \mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$ and

$$
F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}
$$

The field strength satisfies the equation

$$
\mathrm{d}F = 0.
$$

This is called the homogeneous Maxwell's equation and it is, in fact, equivalent to the two homogeneous Maxwell's (vectorial) equations

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}.$$

In order to cast the remaining Maxwell's equations into the language of differential forms, we need a further operation on forms, called the Hodge star operator.

Recall from the standard theory of electrodynamics that the two equations above imply the existence of the electric and vector potentials $\varphi$ and $\mathbf{A} = (A_x, A_y, A_z)$, satisfying

$$\mathbf{B} = \nabla \times \mathbf{A}$$

$$\mathbf{E} = -\nabla \varphi - \frac{\partial \mathbf{A}}{\partial t}.$$

Similarly, the equation $\mathrm{d}F = 0$ on $\mathbb{R}^4$ implies the existence of an electromagnetic 4-potential (or gauge potential) form $A \in \Omega^1(\mathbb{R}^4)$ such that

$$F = \mathrm{d}A.$$

Indeed, we can take

$$A := -\varphi \, \mathrm{d}t + A_x \mathrm{d}x + A_y \mathrm{d}y + A_z \mathrm{d}z.$$

**Definition.** Let $M$ be a smooth manifold. A 2-form $\omega \in \Omega^2(M)$ is said to be a *symplectic form* on $M$ if $\mathrm{d}\omega = 0$ and if it is non-degenerate, i.e.

$$(\forall \, Y \in \Gamma(TM) : \omega(X, Y) = 0) \Rightarrow X = 0.$$

A manifold equipped with a symplectic form is called a *symplectic manifold.*

*Example* 12.18. In the Hamiltonian formulation of classical mechanics one is especially interested in the cotangent bundle $T^*Q$ of some configuration space $Q$. Similarly to what we did when we introduced the tangent bundle, we can define (at least locally) a system of coordinates on $T^*Q$ by

$$(q^1, \ldots, q^{\dim Q}, p_1, \ldots, p_{\dim Q}),$$

where the $p_i$'s are the generalised momenta on $Q$ and the $q^i$'s are the generalised coordinates on $Q$ (recall that $\dim T^*Q = 2 \dim Q$). We can then define a 1-form $\theta \in \Omega^1(T^*Q)$ by

$$\theta := p_i \, \mathrm{d}q^i$$

called the symplectic potential. If we further define

$$\omega := \mathrm{d}\theta \in \Omega^2(T^*Q),$$

then we can calculate that

$$\mathrm{d}\omega = \mathrm{d}(\mathrm{d}\theta) = \cdots = 0.$$

Moreover, $\omega$ is non-degenerate and hence it is a symplectic form on $T^*Q$.

## 12.4   de Rham cohomology

The last two examples suggest two possible implications. In the electrodynamics example, we saw that

$$(\mathrm{d}F = 0) \Rightarrow (\exists\, A : F = \mathrm{d}A),$$

while in the Hamiltonian mechanics example we saw that

$$(\exists\, \theta : \omega = \mathrm{d}\theta) \Rightarrow (\mathrm{d}\omega = 0).$$

**Definition.** Let $M$ be a smooth manifold and let $\omega \in \Omega^n(M)$. We say that $\omega$ is

- *closed* if $\mathrm{d}\omega = 0$;

- *exact* if $\exists\, \sigma \in \Omega^{n-1}(M) : \omega = \mathrm{d}\sigma$.

The question of whether every closed form is exact and vice versa, i.e. whether the implications

$$(\mathrm{d}\omega = 0) \Leftrightarrow (\exists\, \sigma : \omega = \mathrm{d}\sigma)$$

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 12.19.** *Let $M$ be a smooth manifold. The operator*

$$\mathrm{d}^2 \equiv \mathrm{d} \circ \mathrm{d} \colon \Omega^n(M) \to \Omega^{n+2}(M)$$

*is identically zero, i.e.* $\mathrm{d}^2 = 0$.

For the proof, we will need the following concepts.

**Definition.** Given an object which carries some indices, say $T_{a_1,\ldots,a_n}$, we define the *anti-symmetrization* of $T_{a_1,\ldots,a_n}$ as

$$T_{[a_1\cdots a_n]} := \frac{1}{n!} \sum_{\pi \in S_n} \mathrm{sgn}(\pi)\, T_{\pi(a_1)\cdots\pi(a_n)}.$$

Similarly, the *symmetrization* of $T_{a_1,\ldots,a_n}$ is defined as

$$T_{(a_1\cdots a_n)} := \frac{1}{n!} \sum_{\pi \in S_n} T_{\pi(a_1)\cdots\pi(a_n)}.$$

Some special cases are

$$T_{[ab]} = \frac{1}{2}(T_{ab} - T_{ba}), \qquad T_{(ab)} = \frac{1}{2}(T_{ab} + T_{ba})$$

$$T_{[abc]} = \frac{1}{6}(T_{abc} + T_{bca} + T_{cab} - T_{bac} - T_{cba} - T_{acb})$$

$$T_{(abc)} = \frac{1}{6}(T_{abc} + T_{bca} + T_{cab} + T_{bac} + T_{cba} + T_{acb})$$

Of course, we can (anti)symmetrize only some of the indices

$$T^{ab}{}_{[cd]e} = \frac{1}{2}(T^{ab}{}_{cde} - T^{ab}{}_{dce}).$$

It is easy to check that in a contraction (i.e. a sum), we have

$$T_{a_1\cdots a_n}S^{a_1\cdots[a_i\cdots a_j]\cdots a_n} = T_{a_1\cdots[a_i\cdots a_j]\cdots a_n}S^{a_1\cdots a_n}$$

and

$$T_{a_1\cdots(a_i\cdots a_j)\cdots a_n}S^{a_1\cdots[a_i\cdots a_j]\cdots a_n} = 0.$$

*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

$$\omega = \omega_{a_1\cdots a_n}\mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n}.$$

Then, we have

$$\begin{aligned}
\mathrm{d}\omega &= \mathrm{d}\omega_{a_1\cdots a_n} \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n}\\
&= \partial_b\omega_{a_1\cdots a_n}\mathrm{d}x^b \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n},
\end{aligned}$$

and hence

$$\mathrm{d}^2\omega = \partial_c\partial_b\omega_{a_1\cdots a_n}\mathrm{d}x^c \wedge \mathrm{d}x^b \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n}.$$

Since $\mathrm{d}x^c \wedge \mathrm{d}x^b = -\mathrm{d}x^b \wedge \mathrm{d}x^c$, we have

$$\mathrm{d}x^c \wedge \mathrm{d}x^b = \mathrm{d}x^{[c} \wedge \mathrm{d}x^{b]}.$$

Moreover, by Schwarz's theorem, we have $\partial_c\partial_b\omega_{a_1\cdots a_n} = \partial_b\partial_c\omega_{a_1\cdots a_n}$ and hence

$$\partial_c\partial_b\omega_{a_1\cdots a_n} = \partial_{(c}\partial_{b)}\omega_{a_1\cdots a_n}.$$

Thus

$$\begin{aligned}
\mathrm{d}^2\omega &= \partial_c\partial_b\omega_{a_1\cdots a_n}\mathrm{d}x^c \wedge \mathrm{d}x^b \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n}\\
&= \partial_{(c}\partial_{b)}\omega_{a_1\cdots a_n}\mathrm{d}x^{[c} \wedge \mathrm{d}x^{b]} \wedge \mathrm{d}x^{a_1} \wedge \cdots \wedge \mathrm{d}x^{a_n}\\
&= 0.
\end{aligned}$$

Since this holds for any $\omega$, we have $\mathrm{d}^2 = 0$. $\qquad\square$

**Corollary 12.20.** *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

$$0 \xrightarrow{\mathrm{d}} \Omega^0(M) \xrightarrow{\mathrm{d}} \Omega^1(M) \xrightarrow{\mathrm{d}} \cdots \xrightarrow{\mathrm{d}} \Omega^n(M) \xrightarrow{\mathrm{d}} \Omega^{n+1}(M) \xrightarrow{\mathrm{d}} \cdots \xrightarrow{\mathrm{d}} \Omega^{\dim M}(M) \xrightarrow{\mathrm{d}} 0,$$

where we now think of the spaces $\Omega^n(M)$ as $\mathbb{R}$-vector spaces. Recall from linear algebra that, given a linear map $\phi \colon V \to W$, 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 $W$

$$\operatorname{im}(\phi) := \{\phi(v) \mid v \in V\},$$

called the *image* of $\phi$.

Going back to our chain of maps, the equation $\mathrm{d}^2 = 0$ is equivalent to

$$\operatorname{im}(\mathrm{d} \colon \Omega^n(M) \to \Omega^{n+1}(M)) \subseteq \ker(\mathrm{d} \colon \Omega^{n+1}(M) \to \Omega^{n+2}(M))$$

for all $0 \leq n \leq \dim M - 2$. Moreover, we have

$$\omega \in \Omega^n(M) \text{ is closed} \;\Leftrightarrow\; \omega \in \ker(\mathrm{d} \colon \Omega^n(M) \to \Omega^{n+1}(M))$$
$$\omega \in \Omega^n(M) \text{ is exact} \;\Leftrightarrow\; \omega \in \operatorname{im}(\mathrm{d} \colon \Omega^{n-1}(M) \to \Omega^n(M)).$$

The traditional notation for the spaces on the right hand side above is

$$Z^n := \ker(\mathrm{d} \colon \Omega^n(M) \to \Omega^{n+1}(M)),$$
$$B^n := \operatorname{im}(\mathrm{d} \colon \Omega^{n-1}(M) \to \Omega^n(M)),$$

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 $\mathrm{d}^2 = 0$ 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 12.21** (Poincaré). *Let $M \subseteq \mathbb{R}^d$ be a simply connected domain. Then*

$$Z^n = B^n, \qquad \forall\, n > 0.$$

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.** 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,

$$H^n(M) \cong_{\mathrm{vec}} 0.$$

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* 12.22. 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* 12.23. By Poincaré lemma, we have

$$H^n(M) \cong_{\text{vec}} 0$$

for any simply connected $M \subseteq \mathbb{R}^d$.

# 13 Lie groups and their Lie algebras

Lie theory is a topic of the utmost importance in both physics and differential geometry.

## 13.1 Lie groups

**Definition.** A *Lie group* is a group $(G, \bullet)$, where $G$ is a smooth manifold and the maps

$$\mu \colon G \times G \to G$$
$$(g_1, g_2) \mapsto g_1 \bullet g_2$$

and

$$i \colon G \to G$$
$$g \mapsto g^{-1}$$

are both smooth. Note that $G \times G$ inherits a smooth atlas from the smooth atlas of $G$.

**Definition.** The *dimension* of a Lie group $(G, \bullet)$ is the dimension of $G$ as a manifold.

*Example* 13.1.     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) Let $\mathrm{GL}(n, \mathbb{R}) = \{\phi \colon \mathbb{R}^n \xrightarrow{\sim} \mathbb{R}^n \mid \det \phi \neq 0\}$. 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 $\phi \colon \mathbb{R}^n \xrightarrow{\sim} \mathbb{R}^n$ and $\mathbb{R}^{2n}$. The condition $\det \phi \neq 0$ is a so-called *open condition*, meaning that $\mathrm{GL}(n, \mathbb{R})$ can be identified with an open subset of $\mathbb{R}^{2n}$, from which it then inherits a smooth structure.

Then, $(\mathrm{GL}(n, \mathbb{R}), \circ)$ is a Lie group called the *general linear group*.

d) Let $V$ be an $n$-dimensional $\mathbb{R}$-vector space equipped with a pseudo inner product, i.e. an bilinear map $(-, -) \colon V \times V \to \mathbb{R}$ satisfying

    i) symmetry: $\forall\, v, w \in V : \ (v, w) = (w, v)$;

    ii) non-degeneracy: $(\forall\, w \in V : (v, w) = 0) \Rightarrow v = 0$.

Ordinary inner products satisfy a stronger condition than non-degeneracy, called positive definiteness, which is $\forall\, v \in V : (v, v) \geq 0$ and $(v, v) = 0 \Rightarrow v = 0$.

Given a symmetric bilinear map $(-, -)$ on $V$, there is always a basis $\{e_a\}$ of $V$ such that $(e_a, e_a) = \pm 1$ and zero otherwise. If we get $p$-many 1s and $q$-many $-1$s (with $p + q = n$, of course), then the pair $(p, q)$ is called the *signature* of the map. Positive definiteness is the requirement that the signature be $(n, 0)$, although in relativity we

require the signature to be $(n-1, 1)$. A theorem states that there are (up to isomorphism) only as many pseudo inner products on $V$ as there are different signatures.

We can define the set

$$\mathrm{O}(p, q) := \{\phi \colon V \xrightarrow{\sim} V \mid \forall\, v, w \in V : (\phi(v), \phi(w)) = (v, w)\}.$$

The pair $(\mathrm{O}(p, q), \circ)$ is a Lie group called the *orthogonal group* with respect to the pseudo inner product $(-, -)$. This is, in fact, a Lie subgroup of $\mathrm{GL}(p + q, \mathbb{R})$. Some notable examples are $\mathrm{O}(3, 1)$, which is known as the *Lorentz group* in relativity, and $\mathrm{O}(3, 0)$, which is the 3-dimensional rotation group.

**Definition.** Let $(G, \bullet)$ and $(H, \circ)$ be Lie groups. A map $\phi \colon G \to H$ is *Lie group homomorphism* if it is a group homomorphism and a smooth map.

A *Lie group isomorphism* is a group homomorphism which is also a diffeomorphism.

## 13.2 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.** Let $(G, \bullet)$ be a Lie group and let $g \in G$. The map

$$\ell_g \colon G \to G$$
$$h \mapsto \ell_g(h) := g \bullet h \equiv gh$$

is called the *left translation* by $g$.

If there is no danger of confusion, we usually suppress the $\bullet$ notation.

**Proposition 13.2.** *Let $G$ be a Lie group. For any $g \in G$, the left translation map $\ell_g \colon G \to G$ is a diffeomorphism.*

*Proof.* Let $h, h' \in G$. Then, we have

$$\ell_g(h) = \ell_g(h') \iff gh = gh' \iff h = h'.$$

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 $\mu \colon G \times G \to G$ is smooth by definition, so is $\ell_g$.

The inverse map is $(\ell_g)^{-1} = \ell_{g^{-1}}$, since

$$\ell_{g^{-1}} \circ \ell_g = \ell_g \circ \ell_{g^{-1}} = \mathrm{id}_G.$$

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 a diffeomorphism. $\qquad\square$

Note that, in general, $\ell_g$ is not an isomorphism of groups, i.e.

$$\ell_g(hh') \neq \ell_g(h)\,\ell_g(h')$$

in general. However, as the final part of the previous proof suggests, we do have

$$\ell_g \circ \ell_h = \ell_{gh}$$

for all $g, h \in G$.

Since $\ell_g\colon G \to G$ is a diffeomorphism, we have a well-defined push-forward map

$$(L_g)_*\colon \Gamma(TG) \to \Gamma(TG)$$
$$X \mapsto (L_g)_*(X)$$

where

$$(L_g)_*(X)\colon G \to TG$$
$$h \mapsto (L_g)_*(X)(h) := (\ell_g)_*(X(g^{-1}h)).$$

We can draw the diagram

$$
\begin{array}{ccc}
TG & \xrightarrow{\;(\ell_g)_*\;} & TG \\
\Big\uparrow{\scriptstyle X} & & \Big\uparrow{\scriptstyle (L_g)_*(X)} \\
G & \xrightarrow{\;\ell_g\;} & G
\end{array}
$$

Note that this is exactly the same as our previous

$$\Phi_*(\sigma) := \phi_* \circ \sigma \circ \phi^{-1}.$$

By introducing the notation $X|_h := X(h)$, so that $X|_h \in T_h G$, we can write

$$(L_g)_*(X)|_h := (\ell_g)_*(X|_{g^{-1}h}).$$

Alternatively, recalling that the map $\ell_g$ is a diffeomorphism and relabelling the elements of $G$, we can write this as

$$(L_g)_*(X)|_{gh} := (\ell_g)_*(X|_h).$$

A further reformulation comes from considering the vector field $X \in \Gamma(TG)$ as an $\mathbb{R}$-linear map $X\colon \mathcal{C}^\infty(G) \xrightarrow{\sim} \mathcal{C}^\infty(G)$. Then, for any $f \in \mathcal{C}^\infty(G)$

$$(L_g)_*(X)(f) := X(f \circ \ell_g).$$

**Proposition 13.3.** *Let $G$ be a Lie group. For any $g, h \in G$, we have*

$$(L_g)_* \circ (L_h)_* = (L_{gh})_*.$$

*Proof.* Let $f \in \mathcal{C}^\infty(G)$. Then, we have

$$\begin{aligned}
\big((L_g)_* \circ (L_h)_*\big)(X)(f) &= (L_g)_*\big((L_h)_*(X)\big)(f) \\
&= (L_h)_*(X)(f \circ \ell_g) \\
&= X(f \circ \ell_g \circ \ell_h) \\
&= X(f \circ \ell_{gh}) \\
&=: (L_{gh})_*(X)(f),
\end{aligned}$$

as we wanted to show. $\qquad\square$

The previous identity applies to the pointwise push-forward as well, i.e.

$$\big((\ell_{g_1})_* \circ (\ell_{g2})_*\big)(X|_h) = (\ell_{g_1 g_2})_*(X|_h)$$

for any $g_1, g_2, h \in G$ and $X|_h \in T_h G$.

### 13.3 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.** Let $G$ be a Lie group. A vector field $X \in \Gamma(TG)$ is said to be *left-invariant* if

$$\forall\, g \in G: \ (L_g)_*(X) = X.$$

Equivalently, we can require this to hold pointwise

$$\forall\, g, h \in G: \ (\ell_g)_*(X|_h) = X|_{gh}.$$

By recalling the last reformulation of the push-forward, we have that $X \in \Gamma(TG)$ is left-invariant if, and only if

$$\forall\, f \in \mathcal{C}^\infty(G): \ X(f \circ \ell_g) = X(f) \circ \ell_g.$$

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. One can check that $\mathcal{L}(G)$ is closed under

$$+: \mathcal{L}(G) \times \mathcal{L}(G) \to \mathcal{L}(G)$$
$$\cdot: \mathcal{C}^\infty(G) \times \mathcal{L}(G) \to \mathcal{L}(G),$$

only for the constant functions in $\mathcal{C}^\infty(G)$. Therefore, $\mathcal{L}(G)$ is not a $\mathcal{C}^\infty(G)$-submodule of $\Gamma(TG)$, but it is an $\mathbb{R}$-vector subspace of $\Gamma(TG)$.

Recall that, 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 $\mathbb{R}$.

**Theorem 13.4.** *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 $j \colon T_e G \xrightarrow{\sim} \mathcal{L}(G)$. Define

$$j \colon T_e G \to \Gamma(TG)$$
$$A \mapsto j(A),$$

where

$$j(A) \colon G \to TG$$
$$g \mapsto j(A)|_g := (\ell_g)_*(A).$$

i) First, we show that for any $A \in T_e G$, $j(A)$ is a smooth vector field on $G$. It suffices to check that for any $f \in \mathcal{C}^\infty(G)$, we have $j(A)(f) \in \mathcal{C}^\infty(G)$. Indeed

$$
\begin{aligned}
(j(A)(f))(g) &= j(A)|_g(f) \\
&:= (\ell_g)_*(A)(f) \\
&= A(f \circ \ell_g) \\
&= (f \circ \ell_g \circ \gamma)'(0),
\end{aligned}
$$

where $\gamma$ is a curve through $e \in G$ whose tangent vector at $e$ is $A$. The map

$$\varphi \colon \mathbb{R} \times G \to \mathbb{R}$$
$$
\begin{aligned}
(t, g) \mapsto \varphi(t, g) &:= (f \circ \ell_g \circ \gamma)(t) \\
&= f(g\gamma(t))
\end{aligned}
$$

is a composition of smooth maps, hence it is smooth. Then

$$(j(A)(f))(g) = (\partial_1 \varphi)(0, g)$$

depends smoothly on $g$ and thus $j(A)(f) \in \mathcal{C}^\infty(G)$.

ii) Let $g, h \in G$. Then, for every $A \in T_e G$, we have

$$
\begin{aligned}
(\ell_g)_*(j(A)|_h) &:= (\ell_g)_*((\ell_h)_*(A)) \\
&= (\ell_{gh})_*(A) \\
&= j(A)|_{gh},
\end{aligned}
$$

so $j(A) \in \mathcal{L}(G)$. Hence, the map $j$ is really $j \colon T_e G \to \mathcal{L}(G)$.

iii) Let $A, B \in T_e G$ and $\lambda \in \mathbb{R}$. Then, for any $g \in G$

$$
\begin{aligned}
j(\lambda A + B)|_g &= (\ell_g)_*(\lambda A + B) \\
&= \lambda (\ell_g)_*(A) + (\ell_g)_*(B) \\
&= \lambda j(A)|_g + j(B)|_g,
\end{aligned}
$$

since the push-forward is an $\mathbb{R}$-linear map. Hence, we have $j \colon T_e G \xrightarrow{\sim} \mathcal{L}(G)$.

iv) Let $A, B \in T_e G$. Then

$$
\begin{aligned}
j(A) = j(B) &\Leftrightarrow \forall\, g \in G : j(A)|_g = j(B)|_g \\
&\Rightarrow j(A)|_e = j(B)|_e \\
&\Leftrightarrow (\ell_e)_*(A) = (\ell_e)_*(B) \\
&\Leftrightarrow A = B,
\end{aligned}
$$

since $(\ell_e)_* = \mathrm{id}_{TG}$. Hence, the map $j$ is injective.

v) Let $X \in \mathcal{L}(G)$. Define $A^X := X|_e \in T_e G$. Then, we have

$$
j(A^X)|_g = (\ell_g)_*(A^X) = (\ell_g)_*(X|_e) = X_{ge} = X_g,
$$

since $X$ is left-invariant. Hence $X = j(A^X)$ and thus $j$ is surjective.

Therefore, $j \colon T_e G \xrightarrow{\sim} \mathcal{L}(G)$ is indeed a linear isomorphism. $\qquad\square$

**Corollary 13.5.** *The space $\mathcal{L}(G)$ is finite-dimensional and $\dim \mathcal{L}(G) = \dim G$.*

We will soon see that the identification of $\mathcal{L}(G)$ and $T_e G$ goes beyond the level of linear isomorphism, as they are isomorphic as Lie algebras. Recall 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.

Given $X, Y \in \Gamma(TM)$, we defined their Lie bracket, or commutator, as

$$
[X, Y](f) := X(Y(f)) - Y(X(f))
$$

for any $f \in \mathcal{C}^\infty(M)$. You 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 suppress the $+$ and $\cdot$ when they are clear from the context. In the case of a manifold that is also a Lie group, we have the following.

**Theorem 13.6.** *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

$$
\begin{aligned}
[X, Y](f \circ \ell_g) &:= X(Y(f \circ \ell_g)) - Y(X(f \circ \ell_g)) \\
&= X(Y(f) \circ \ell_g) - Y(X(f) \circ \ell_g) \\
&= X(Y(f)) \circ \ell_g - Y(X(f)) \circ \ell_g \\
&= \big(X(Y(f)) - Y(X(f))\big) \circ \ell_g \\
&= [X, Y](f) \circ \ell_g.
\end{aligned}
$$

Hence, $[X, Y]$ is left-invariant. $\qquad\square$

**Definition.** Let $G$ be a Lie group. The *associated Lie algebra* of $G$ is $\mathcal{L}(G)$.

Note $\mathcal{L}(G)$ is a rather complicated object, since its elements are vector fields, hence we would like to work with $T_e G$ instead, whose elements are tangent vectors. Indeed, we can use the bracket on $L(G)$ to define a bracket on $T_e G$ such that they be isomorphic as Lie algebras. First, let us define the isomorphism of Lie algebras.

**Definition.** Let $(L_1, [-,-]_{L_1})$ and $(L_2, [-,-]_{L_2})$ be Lie algebras over the same field. A linear map $\phi\colon L_1 \xrightarrow{\sim} L_2$ is a *Lie algebra homomorphism* if

$$\forall\, x, y \in L_1 : \ \phi([x,y]_{L_1}) = [\phi(x), \phi(y)]_{L_2}.$$

If $\phi$ is bijective, then it is a *Lie algebra isomorphism* and we write $L_1 \cong_{\text{Lie alg}} L_2$.

By using the bracket $[-,-]_{\mathcal{L}(G)}$ on $\mathcal{L}(G)$ we can define, for any $A, B \in T_e G$

$$[A, B]_{T_e G} := j^{-1}\big([j(A), j(B)]_{\mathcal{L}(G)}\big),$$

where $j^{-1}(X) = X|_e$. Equipped with these brackets, we have

$$\mathcal{L}(G) \cong_{\text{Lie alg}} T_e G.$$

## 14 Classification of Lie algebras and Dynkin diagrams

Given a Lie group, we have seen how we can construct a Lie algebra as the space of left-invariant vector fields or, equivalently, 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.

However, here we will consider a question that is independent of where a Lie algebra comes from, namely that of the classification of Lie algebras.

### 14.1 Lie algebras

While it is possible to classify Lie algebras more generally, we will only consider the classification of finite-dimensional complex Lie algebras, i.e. Lie algebras $(L, [-, -])$ where $L$ is a finite-dimensional $\mathbb{C}$-vector space.

*Example* 14.1. Of course, any complex Lie group $G$ (where $G$ is a complex manifold) gives rise to a complex Lie algebra.

If $A, B$ are Lie subalgebras of a Lie algebra $(L, [-, -])$ over $K$, then

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

is again a Lie subalgebra of $L$.

**Definition.** A Lie algebra $L$ is said to be *abelian* if

$$\forall\, x, y \in L : \ [x, y] = 0.$$

Equivalently, $[L, L] = 0$, 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. Even from the classification point of view, 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.** An *ideal* $I$ of a Lie algebra $L$ is a Lie subalgebra such that $[I, L] \subseteq I$, i.e.

$$\forall\, x \in I : \forall\, y \in L : \ [x, y] \in I.$$

The ideals $0$ and $L$ are called the *trivial ideals* of $L$.

**Definition.** A Lie algebra $L$ is said to be

- *simple* if it is non-abelian and it contains no non-trivial ideals;

- *semi-simple* if it contains no non-trivial abelian ideals.

*Remark* 14.2. 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.** Let $L$ be a Lie algebra. The Lie subalgebra

$$L' := [L, L]$$

is called the *derived subalgebra* of $L$.

We can form a sequence of Lie subalgebras

$$L \supseteq L' \supseteq L'' \supseteq \cdots \supseteq L^{(n)} \supseteq \cdots$$

called the *derived series* of $L$.

**Definition.** 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.** 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.** 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 14.3** (Levi)**.** *Any finite-dimensional complex Lie algebra $L$ can be decomposed as*

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

*where $R$ is a solvable Lie algebra and $L_1, \ldots, L_n$ 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 14.4.** *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.

## 14.2 The adjoint map and the Killing form

**Definition.** 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

$$\text{ad}_x \colon L \xrightarrow{\sim} L$$
$$y \mapsto \text{ad}_x(y) := [x, y].$$

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

$$\text{ad} \colon L \xrightarrow{\sim} \text{End}(L)$$
$$x \mapsto \text{ad}(x) := \text{ad}_x.$$

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

*Proof.* It remains to check that 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 $\text{ad}([x, y]) = [\text{ad}(x), \text{ad}(y)]$. $\qquad\qquad\square$

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

$$\kappa \colon L \times L \to K$$
$$(x, y) \mapsto \kappa(x, y) := \text{tr}(\text{ad}_x \circ \text{ad}_y),$$

where 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 14.6.** *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 ad is a homomorphism.

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

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

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

**Proposition 14.7** (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.** A linear map $\phi \colon V \xrightarrow{\sim} V$ 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 $\operatorname{ad}_z$ is anti-symmetric with respect to $\kappa$, i.e.

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

In order to do computations, it is useful to introduce a basis $\{E_i\}$ on $L$.

**Definition.** Let $L$ be a Lie algebra over $K$ and let $\{E_i\}$ be a basis. Then, we have

$$[E_i, E_j] = C^k{}_{ij} E_k$$

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

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

$$C^k{}_{ij} = -C^k{}_{ji}$$

while the Jacobi identity becomes

$$C^n{}_{im} C^m{}_{jk} + C^n{}_{jm} C^m{}_{ki} + C^n{}_{km} C^m{}_{ij} = 0.$$

We can now express both the adjoint maps and the Killing form in terms of components with respect to a basis.

**Proposition 14.8.** *Let $L$ be a Lie algebra and let $\{E_i\}$ be a basis. Then*

*i)* $(\mathrm{ad}_{E_i})^k{}_j = C^k{}_{ij}$

*ii)* $\kappa_{ij} = C^m{}_{ik} C^k{}_{jm}$

*where $C^k{}_{ij}$ are the structure constants of $L$ with respect to $\{E_i\}$.*

*Proof.*   i) Denote by $\{\varepsilon^i\}$ the dual basis to $\{E_i\}$. Then, we have

$$
\begin{aligned}
(\mathrm{ad}_{E_i})^k{}_j &:= \varepsilon^k(\mathrm{ad}_{E_i}(E_j)) \\
&= \varepsilon^k([E_i, E_j]) \\
&= \varepsilon^k(C^m{}_{ij} E_m) \\
&= C^m{}_{ij} \varepsilon^k(E_m) \\
&= C^k{}_{ij},
\end{aligned}
$$

since $\varepsilon^k(E_m) = \delta^k_m$.

ii) Recall from linear algebra that if $V$ is finite-dimensional, for any $\phi \in \mathrm{End}(V)$ we have $\mathrm{tr}(\phi) = \Phi^k{}_k$, 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, we have

$$
\begin{aligned}
\kappa_{ij} &:= \kappa(E_i, E_j) \\
&= \mathrm{tr}(\mathrm{ad}_{E_i} \circ \mathrm{ad}_{E_j}) \\
&= (\mathrm{ad}_{E_i} \circ \mathrm{ad}_{E_j})^k{}_k \\
&= (\mathrm{ad}_{E_i})^m{}_k (\mathrm{ad}_{E_j})^k{}_m \\
&= C^m{}_{ik} C^k{}_{jm},
\end{aligned}
$$

where we used the same notation for the linear maps and their matrices. $\qquad\square$

## 14.3 The fundamental roots and 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.** 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 $\{h_1, \ldots, h_r\}$ of $H$ which can be extended to a basis $\{h_1, \ldots, h_r, e_1, \ldots, e_{d-r}\}$ of $L$ such that $e_1, \ldots, e_{d-r}$ are eigenvectors of $\mathrm{ad}(h)$ for any $h \in H$, i.e.

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

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

The basis $\{h_1, \ldots, h_r, e_1, \ldots, e_{d-r}\}$ is known as a *Cartan-Weyl basis* of $L$.

**Theorem 14.9.** *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 of $L$ is abelian.*

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

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

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

**Definition.** The maps $\lambda_1, \ldots, \lambda_{d-r} \in H^*$ are called the *roots* of $L$. The collection

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

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 $\mathrm{ad}(h)$ 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.** A set of *fundamental roots* $\Pi := \{\pi_1, \ldots, \pi_f\}$ 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 $n_1, \ldots, n_f \in \mathbb{N}$ and $\varepsilon \in \{+1, -1\}$ such that

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

We can write the last equation more concisely as $\lambda \in \mathrm{span}_{\varepsilon, \mathbb{N}}(\Pi)$. Observe that, for any $\lambda \in \Phi$, the coefficients of $\pi_1, \ldots, \pi_f$ in the expansion above always have the same sign. Indeed, we have $\mathrm{span}_{\varepsilon, \mathbb{N}}(\Pi) \neq \mathrm{span}_{\mathbb{Z}}(\Pi)$.

**Theorem 14.10.** *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 $\mathrm{span}_{\mathbb{C}}(\Pi) = H^*$, that is, $\Pi$ is a basis of $H^*$.*

**Corollary 14.11.** *We have $|\Pi| = r$, where $r$ is the rank of L.*

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

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

$$i \colon V \xrightarrow{\sim} V^*$$
$$v \mapsto i(v) := B(v, -)$$

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

$$B^* \colon V^* \times V^* \to K$$
$$(\phi, \psi) \mapsto B^*(\phi, \psi) := B(i^{-1}(\phi), i^{-1}(\psi)).$$

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 14.12.** *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 $\{h_1, \ldots, h_r, e_{r+1}, \ldots, e_d\}$ 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

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

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

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

which is what we wanted. $\qquad\qquad\square$

We can now define

$$\kappa^* : H^* \times H^* \to \mathbb{C}$$
$$(\mu, \nu) \mapsto \kappa^*(\mu, \nu) := \kappa(i^{-1}(\mu), i^{-1}(\nu)),$$

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

*Remark* 14.13. If $\{h_i\}$ is a basis of $H$, the components of $\kappa^*$ with respect to the dual basis satisfy

$$(\kappa^*)^{ij} \kappa_{jk} = \delta^i_k.$$

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 $H^*_{\mathbb{R}} := \operatorname{span}_{\mathbb{R}}(\Pi)$. Note that we have the following chain of inclusions

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

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

**Theorem 14.14.** *i) For any $\alpha, \beta \in H^*_{\mathbb{R}}$, we have $\kappa^*(\alpha, \beta) \in \mathbb{R}$.*

*ii) $\kappa^* : H^*_{\mathbb{R}} \times H^*_{\mathbb{R}} \to \mathbb{R}$ 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.** 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 $\varphi := \cos^{-1}\left( \dfrac{\kappa^*(\alpha, \beta)}{|\alpha||\beta|} \right)$.

We need one final ingredient for our classification result.

**Definition.** For any $\lambda \in \Phi \subseteq H_{\mathbb{R}}^*$, define the linear map

$$s_\lambda \colon H_{\mathbb{R}}^* \xrightarrow{\sim} H_{\mathbb{R}}^*$$
$$\mu \mapsto s_\lambda(\mu),$$

where

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

The map $s_\lambda$ is called a *Weyl transformation* and the set

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

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

**Theorem 14.15.**    *i) The Weyl group $W$ is generated by the fundamental roots in $\Pi$, in the sense that for some $1 \le n \le r$, with $r = |\Pi|$,*

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

*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,*

$$\forall\, \lambda \in \Phi : \forall\, w \in W : \; w(\lambda) \in \Phi.$$

## 14.4   Dynkin diagrams and the Cartan classification

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

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

Since $s_{\pi_i}(\pi_j) \in \Phi$ and $\Phi \subseteq \mathrm{span}_{\varepsilon, \mathbb{N}}(\Pi)$, for all $1 \le i \ne j \le r$ we must have

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

**Definition.** The *Cartan matrix* of a Lie algebra is the $r \times r$ matrix $C$ with entries

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

where the $C_{ij}$ should not be confused with the structure constants $C^k{}_{ij}$.

**Theorem 14.16.** *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.** Given a Cartan matrix $C$, the $ij$-th *bond number* is

$$n_{ij} := C_{ij} C_{ji} \qquad \text{(no summation implied).}$$

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 $180°$, 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 14.17.** *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.** A *Dynkin diagram* associated to a Cartan matrix is constructed as follows.

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

$$\underset{\pi_i}{\bigcirc}$$

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

| $n_{ij} = 0$ | $n_{ij} = 1$ | $n_{ij} = 2$ | $n_{ij} = 3$ |
|---|---|---|---|
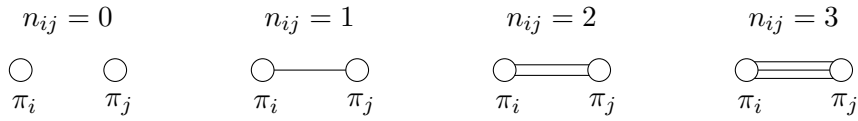


3. If $n_{ij} = 2$ or $3$, draw an arrow on the lines from the longer root to the shorter root.

| $|\pi_i| > |\pi_j|$ | $|\pi_i| < |\pi_j|$ | $|\pi_i| > |\pi_j|$ | $|\pi_i| < |\pi_j|$ |
|---|---|---|---|



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 14.18** (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*
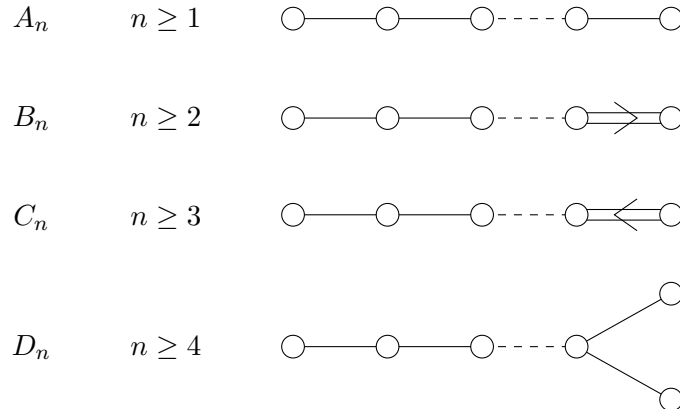
*where the restrictions on n ensure that we don't get repeated diagrams (the diagram $D_2$ is excluded since it is disconnected and does not correspond to a simple Lie algebra)*

*ii) five exceptional cases*



$E_6$

$E_7$

$E_8$

$F_4$

$G_2$
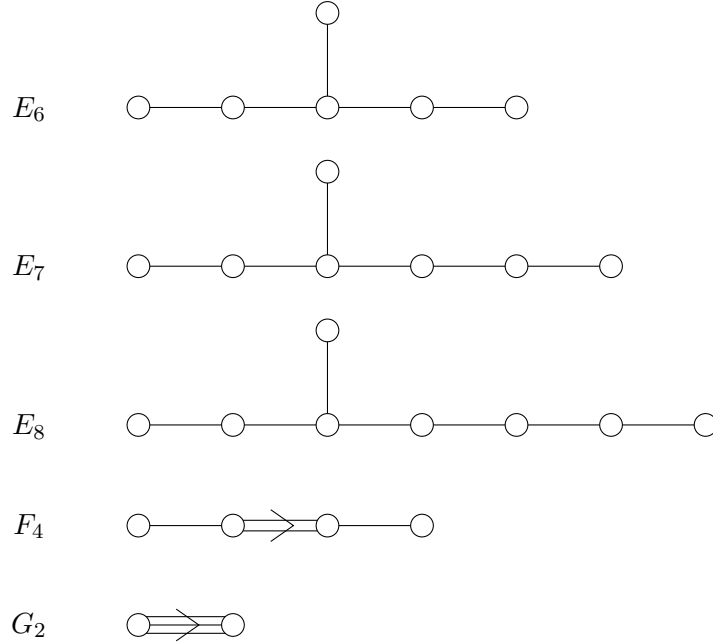
*and no other. These are all the possible (connected) Dynkin diagrams.*

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.

## 15  The Lie group $\mathrm{SL}(2,\mathbb{C})$ and its Lie algebra $\mathfrak{sl}(2,\mathbb{C})$

### 15.1  The structure of $\mathrm{SL}(2,\mathbb{C})$

Recall the the Lie group structure is the combination of a number of simpler structures, which we will now examine in detail for the *special linear group* of degree 2 over $\mathbb{C}$, also known as the *relativistic spin group*.

$\mathrm{SL}(2,\mathbb{C})$ **as a set**

We define the following subset of $\mathbb{C}^4 := \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C}$

$$\mathrm{SL}(2,\mathbb{C}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{C}^4 \ \Big|\ ad - bc = 1 \right\},$$

where the array is just an alternative notation for a quadruple.

$\mathrm{SL}(2,\mathbb{C})$ **as a group**

We define an operation

$$\bullet \colon \mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C}) \to \mathrm{SL}(2,\mathbb{C})$$

$$\left( \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right) \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

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 $\mathrm{SL}(2,\mathbb{C})$, 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 $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$;

iii) admits inverses: for each $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$, we have $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$ and

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

Hence, we have $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Therefore, the pair $(\mathrm{SL}(2,\mathbb{C}), \bullet)$ is a (non-commutative) group.

## $\mathrm{SL}(2, \mathbb{C})$ 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 $z \in \mathbb{C}$.
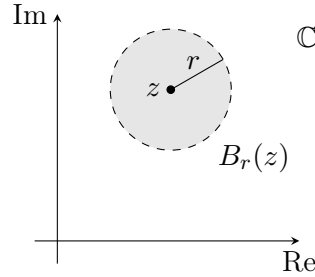


Define $\mathcal{O}_{\mathbb{C}}$ implicitly by

$$U \in \mathcal{O}_{\mathbb{C}} \ :\Leftrightarrow\ \forall\, z \in U : \exists\, r > 0 : B_r(z) \subseteq U.$$

Then, the pair $(\mathbb{C}, \mathcal{O}_{\mathbb{C}})$ is a topological space. In fact, we have

$$(\mathbb{C}, \mathcal{O}_{\mathbb{C}}) \cong_{\mathrm{top}} (\mathbb{R}^2, \mathcal{O}_{\mathrm{std}}).$$

We can then equip $\mathbb{C}^4$ with the product topology so that we can finally define

$$\mathcal{O} := (\mathcal{O}_{\mathbb{C}})|_{\mathrm{SL}(2,\mathbb{C})},$$

so that the pair $(\mathrm{SL}(2, \mathbb{C}), \mathcal{O})$ is a topological space. In fact, it is a connected topological space, and we will need this property later on.

## $\mathrm{SL}(2, \mathbb{C})$ 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 $\mathscr{A}$ 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 $\mathbb{C}^d$.

Let $U$ be the set

$$U := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}) \ \middle|\ a \neq 0 \right\}$$

and define the map

$$\begin{aligned} x \colon \quad U &\to x(U) \subseteq \mathbb{C}^* \times \mathbb{C} \times \mathbb{C} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto (a, b, c), \end{aligned}$$

where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$. With a little work, one can show that $U$ is an open subset of $(\mathrm{SL}(2,\mathbb{C}), \mathcal{O})$ and $x$ is a homeomorphism with inverse

$$x^{-1}: \quad x(U) \to U$$
$$(a, b, c) \mapsto \begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix}.$$

However, since $\mathrm{SL}(2,\mathbb{C})$ contains elements with $a = 0$, the chart $(U, x)$ does not cover the whole space, and hence we need at least one more chart. We thus define the set

$$V := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C}) \ \middle| \ b \neq 0 \right\}$$

and the map

$$y: \quad V \to x(V) \subseteq \mathbb{C} \times \mathbb{C}^* \times \mathbb{C}$$
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (a, b, d).$$

Similarly to the above, $V$ is open and $y$ is a homeomorphism with inverse

$$y^{-1}: \quad x(V) \to V$$
$$(a, b, d) \mapsto \begin{pmatrix} a & b \\ \frac{ad-1}{b} & d \end{pmatrix}.$$

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 $\mathscr{A}_{\mathrm{top}} := \{(U, x), (V, y)\}$ is an atlas, and since every atlas is automatically a $\mathcal{C}^0$-atlas, the triple $(\mathrm{SL}(2,\mathbb{C}), \mathcal{O}, \mathscr{A}_{\mathrm{top}})$ 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 $\mathscr{A}$, we have to check that every transition map between charts in $\mathscr{A}$ is differentiable in the usual sense.

In our case, we have the atlas $\mathscr{A}_{\mathrm{top}} := \{(U, x), (V, y)\}$. We evaluate

$$(y \circ x^{-1})(a, b, c) = y\left( \begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix} \right) = (a, b, \tfrac{1+bc}{a}).$$

Hence we have the transition map

$$y \circ x^{-1}: x(U \cap V) \to y(U \cap V)$$
$$(a, b, c) \mapsto (a, b, \tfrac{1+bc}{a}).$$

Similarly, we have

$$(x \circ y^{-1})(a, b, d) = y\left( \begin{pmatrix} a & b \\ \frac{ad-1}{b} & d \end{pmatrix} \right) = (a, b, \tfrac{ad-1}{b}).$$

Hence, the other transition map is

$$x \circ y^{-1} \colon y(U \cap V) \to x(U \cap V)$$
$$(a, b, c) \mapsto (a, b, \tfrac{ad-1}{b}).$$

Since $a \neq 0$ and $b \neq 0$, the transition maps are complex differentiable (this is a good time to review your complex analysis!).

Therefore, the atlas $\mathscr{A}_{\text{top}}$ is a differentiable atlas. By defining $\mathscr{A}$ to be the maximal differentiable atlas containing $\mathscr{A}_{\text{top}}$, we have that $(\mathrm{SL}(2, \mathbb{C}), \mathcal{O}, \mathscr{A})$ is 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

$$\mu \colon \mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) \to \mathrm{SL}(2, \mathbb{C})$$
$$(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} e & f \\ g & h \end{pmatrix}) \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

and

$$i \colon \mathrm{SL}(2, \mathbb{C}) \to \mathrm{SL}(2, \mathbb{C})$$
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}$$

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 $(U, x), (V, y) \in \mathscr{A}$.

$$\begin{array}{ccc} U \subseteq \mathrm{SL}(2, \mathbb{C}) & \xrightarrow{\quad i \quad} & V \subseteq \mathrm{SL}(2, \mathbb{C}) \\ \downarrow x & & \downarrow y \\ x(U) \subseteq \mathbb{C}^3 & \xrightarrow{\quad y \circ i \circ x^{-1} \quad} & y(V) \subseteq \mathbb{C}^3 \end{array}$$

However, since $\mathrm{SL}(2, \mathbb{C})$ is connected, the differentiability of the transition maps in $\mathscr{A}$ implies that if $y \circ i \circ x^{-1}$ is differentiable for any two given charts, then it is differentiable for all charts in $\mathscr{A}$. 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

$$(y \circ i \circ x^{-1})(a, b, c) = (y \circ i)(\begin{pmatrix} a & b \\ c & \frac{1+bc}{a} \end{pmatrix}) = y(\begin{pmatrix} \frac{1+bc}{a} & -b \\ -c & a \end{pmatrix}) = (\tfrac{1+bc}{a}, -b, a)$$

which is certainly complex differentiable as a map between open subsets of $\mathbb{C}^3$ (recall that $a \neq 0$ on $x(U)$).

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 $((\mathrm{SL}(2, \mathbb{C}), \mathcal{O}, \mathscr{A}), \bullet)$ is a 3-dimensional complex Lie group.

## 15.2 The Lie algebra of $\mathrm{SL}(2, \mathbb{C})$

Recall that to every Lie group $G$, there is an associated Lie algebra $\mathcal{L}(G)$, where

$$\mathcal{L}(G) := \{X \in \Gamma(TG) \mid \forall\, g, h \in G : (\ell_g)_*(X|_h) = X_{gh}\},$$

which we then proved to be isomorphic to the Lie algebra $T_e G$ with Lie bracket

$$[A, B]_{T_e G} := j^{-1}([j(A), j(B)]_{\mathcal{L}(G)})$$

induced by the Lie bracket on $\mathcal{L}(G)$ via the isomorphism $j$

$$j(A)|_g := (\ell_g)_*(A).$$

In the case of $\mathrm{SL}(2, \mathbb{C})$, the left translation map by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is

$$\ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} : \mathrm{SL}(2, \mathbb{C}) \to \mathrm{SL}(2, \mathbb{C})$$

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

By using the standard notation $\mathfrak{sl}(2, \mathbb{C}) \equiv \mathcal{L}(\mathrm{SL}(2, \mathbb{C}))$, we have

$$\mathfrak{sl}(2, \mathbb{C}) \cong_{\text{Lie alg}} T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C}).$$

We would now like to explicitly determine the Lie bracket on $T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C})$, and hence determine its structure constants.

Recall that if $(U, x)$ is a chart on a manifold $M$ and $p \in U$, then the chart $(U, x)$ induces a basis of the tangent space $T_p M$. We shall use our previously defined chart $(U, x)$ on $\mathrm{SL}(2, \mathbb{C})$, where $U := \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{C}) \mid a \neq 0\}$ and

$$\begin{aligned} x : \quad U &\to x(U) \subseteq \mathbb{C}^3 \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto (a, b, c). \end{aligned}$$

Note that the $d$ appearing here is completely redundant, since the membership condition of $\mathrm{SL}(2, \mathbb{C})$ forces $d = \frac{1+bc}{a}$. However, we will keep writing the $d$ to avoid having a fraction in a matrix in a subscript.

The chart $(U, x)$ contains $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and hence we get an induced co-ordinate basis

$$\left\{ \left(\frac{\partial}{\partial x^i}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \in T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C}) \,\Big|\, 1 \leq i \leq 3 \right\}$$

so that any $A \in T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C})$ can be written as

$$A = \alpha \left( \frac{\partial}{\partial x^1} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} + \beta \left( \frac{\partial}{\partial x^2} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} + \gamma \left( \frac{\partial}{\partial x^3} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)},$$

for some $\alpha, \beta, \gamma \in \mathbb{C}$. Since the Lie bracket is bilinear, its action on these basis vectors uniquely extends to the whole of $T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C})$ by linear continuation. Hence, we simply have to determine the action of the Lie bracket of $\mathfrak{sl}(2, \mathbb{C})$ on the images under the isomorphism $j$ of these basis vectors.

Let us now determine the image of these co-ordinate induced basis elements under the isomorphism $j$. The object

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \right) \in \mathfrak{sl}(2, \mathbb{C})$$

is a left-invariant vector field on $\mathrm{SL}(2, \mathbb{C})$. It assigns to each point $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in U \subseteq \mathrm{SL}(2, \mathbb{C})$ the tangent vector

$$j \left( \left( \frac{\partial}{\partial x^i} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \right) \Big|_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} := \left( \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \right)_* \left( \frac{\partial}{\partial x^i} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \in T_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C}).$$

This tangent vector is a $\mathbb{C}$-linear map $\mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C})) \xrightarrow{\sim} \mathbb{C}$, where $\mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C}))$ is the $\mathbb{C}$-vector space (in fact, the $\mathbb{C}$-algebra) of smooth complex-valued functions on $\mathrm{SL}(2, \mathbb{C})$ although, to be precise, since we are working in a chart we should only consider functions defined on $U$. For (the restriction to $U$ of) any $f \in \mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C}))$ we have, explicitly,

$$\left( \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \right)_* \left( \frac{\partial}{\partial x^i} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} (f) = \left( \frac{\partial}{\partial x^i} \right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \left( f \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \right)$$

$$= \partial_i \left( f \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1} \right) (x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)),$$

where the argument of $\partial_i$ in the last line is a map $x(U) \subseteq \mathbb{C}^3 \to \mathbb{C}$, hence $\partial_i$ is simply the operation of complex differentiation with respect to the $i$-th (out of the 3) complex variable of the map $f \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1}$, which is then to be evaluated at $x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \mathbb{C}^3$. By inserting an identity in the composition, we have

$$= \partial_i \left( f \circ \mathrm{id}_U \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1} \right) (x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right))$$

$$= \partial_i \left( f \circ (x^{-1} \circ x) \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1} \right) (x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right))$$

$$= \partial_i \left( (f \circ x^{-1}) \circ (x \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1}) \right) (x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)),$$

where $f \circ x^{-1} \colon x(U) \subseteq \mathbb{C}^3 \to \mathbb{C}$ and $(x \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1}) \colon x(U) \subseteq \mathbb{C}^3 \to x(U) \subseteq \mathbb{C}^3$ and hence, we can use the multi-dimensional chain rule to obtain

$$= \left( \partial_m (f \circ x^{-1}) \left( (x \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1}) (x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)) \right) \right) \left( \partial_i (x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1}) (x \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)) \right),$$

with the summation going from $m = 1$ to $m = 3$. The first factor is simply

$$\partial_m(f \circ x^{-1})\left(\left(x \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}\right)\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)\right) = \partial_m(f \circ x^{-1})\left(x\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right)$$

$$=: \left(\frac{\partial}{\partial x^m}\right)_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}(f).$$

To see what the second factor is, we first consider the map $x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1}$. This map acts on the triple $(e, f, g) \in x(U)$ as

$$(x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1})(e, f, g) = (x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)})\begin{pmatrix} e & f \\ g & \frac{1+fg}{e} \end{pmatrix}$$

$$= x^m\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} e & f \\ g & \frac{1+fg}{e} \end{pmatrix}\right)$$

$$= x^m\left(\begin{pmatrix} ae + bg & af + \frac{b(1+fg)}{e} \\ ce + dg & cf + \frac{d(1+fg)}{e} \end{pmatrix}\right),$$

and since $x^m := \mathrm{proj}_m \circ x$, with $m \in \{1, 2, 3\}$, we have

$$(x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1})(e, f, g) = \mathrm{proj}_m(ae + bg, af + \frac{b(1+fg)}{e}, ce + dg),$$

the map $\mathrm{proj}_m$ simply picks the $m$-th component of the triple. 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 $g$. We can write the result as

$$\partial_i(x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1})(e, f, g) = D(e, f, g)^m{}_i,$$

where $m$ labels the rows and $i$ the columns of the matrix

$$D(e, f, g) = \begin{pmatrix} a & 0 & b \\ -\frac{b(1+fg)}{e^2} & a + \frac{bg}{e} & \frac{bf}{e} \\ c & 0 & d \end{pmatrix}.$$

Finally, by evaluating this at $(e, f, g) = x\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) = (1, 0, 0)$, we obtain

$$\partial_i(x^m \circ \ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \circ x^{-1})\left(x\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)\right) = D^m{}_i,$$

where, by recalling that $d = \frac{1+bc}{a}$,

$$D := D(1, 0, 0) = \begin{pmatrix} a & 0 & b \\ -b & a & 0 \\ c & 0 & \frac{1+bc}{a} \end{pmatrix}.$$

Putting the two factors back together yields

$$\left(\ell_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}\right)_* \left(\frac{\partial}{\partial x^i}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}(f) = D^m{}_i \left(\frac{\partial}{\partial x^m}\right)_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}(f).$$

Since this holds for an arbitrary $f \in \mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C}))$, we have

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)}\right)\bigg|_{\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)} := \left(\ell_{\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)}\right)_* \left(\frac{\partial}{\partial x^i}\right)_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)} = D^m{}_i \left(\frac{\partial}{\partial x^m}\right)_{\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)},$$

and since the point $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in U \subseteq \mathrm{SL}(2, \mathbb{C})$ is also arbitrary, we have

$$j\left(\left(\frac{\partial}{\partial x^i}\right)_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)}\right) = D^m{}_i \frac{\partial}{\partial x^m} \in \mathfrak{sl}(2, \mathbb{C}),$$

where $D$ is now the corresponding matrix of co-ordinate functions

$$D := \begin{pmatrix} x^1 & 0 & x^2 \\ -x^2 & x^1 & 0 \\ x^3 & 0 & \frac{1+x^2 x^3}{x^1} \end{pmatrix}.$$

Note that while the three vector fields

$$\frac{\partial}{\partial x^m} : \mathrm{SL}(2, \mathbb{C}) \to T\,\mathrm{SL}(2, \mathbb{C})$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \left(\frac{\partial}{\partial x^m}\right)_{\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)}$$

are not individually left-invariant, their linear combination with coefficients $D^m{}_i$ is indeed left-invariant. Recall that these vector fields

   i) are $\mathbb{C}$-linear maps

$$\frac{\partial}{\partial x^m} : \mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C})) \xrightarrow{\sim} \mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C}))$$

$$f \mapsto \partial_m(f \circ x^{-1}) \circ x;$$

   ii) satisfy the Leibniz rule

$$\frac{\partial}{\partial x^m}(fg) = f\frac{\partial}{\partial x^m}(g) + g\frac{\partial}{\partial x^m}(f);$$

   iii) act on the coordinate functions $x^i \in \mathcal{C}^\infty(\mathrm{SL}(2, \mathbb{C}))$ as

$$\frac{\partial}{\partial x^m}(x^i) = \partial_m(x^i \circ x^{-1}) \circ x = \partial_m(\mathrm{proj}_i \circ x \circ x^{-1}) \circ x = \delta^i_m \circ x = \delta^i_m,$$

   since the composition of a constant function with any composable function is just the constant function.

Hence, we have an expansion of the images of the basis of $T_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)}\mathrm{SL}(2, \mathbb{C})$ under $j$:

$$j\left(\left(\frac{\partial}{\partial x^1}\right)_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)}\right) = x^1 \frac{\partial}{\partial x^1} - x^2 \frac{\partial}{\partial x^2} + x^3 \frac{\partial}{\partial x^3}$$

$$j\left(\left(\frac{\partial}{\partial x^2}\right)_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)}\right) = x^1 \frac{\partial}{\partial x^2}$$

$$j\left(\left(\frac{\partial}{\partial x^3}\right)_{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)}\right) = x^2 \frac{\partial}{\partial x^1} + \frac{1+x^2 x^3}{x^1} \frac{\partial}{\partial x^3}.$$

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

$$\left[ j\left( \left(\frac{\partial}{\partial x^i}\right)_{\left(\begin{smallmatrix}1 & 0\\0 & 1\end{smallmatrix}\right)}\right), j\left( \left(\frac{\partial}{\partial x^k}\right)_{\left(\begin{smallmatrix}1 & 0\\0 & 1\end{smallmatrix}\right)}\right) \right] = \left[ D^m{}_i \frac{\partial}{\partial x^m}, D^n{}_k \frac{\partial}{\partial x^n} \right].$$

Letting this act on an arbitrary $f \in \mathcal{C}^\infty(\mathrm{SL}(2,\mathbb{C}))$, 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, we have

$$D^n{}_k \frac{\partial}{\partial x^n}\left( D^m{}_i \frac{\partial}{\partial x^m}(f) \right) = D^n{}_k \frac{\partial}{\partial x^n}(D^m{}_i)\,(\partial_m(f \circ x^{-1}) \circ x) + D^n{}_k D^m{}_i \,\partial_n\partial_m(f \circ x^{-1}) \circ x.$$

Hence, recalling that $\partial_m\partial_n = \partial_n\partial_m$ by Schwarz's theorem, we have

$$\begin{aligned}
\left[ D^m{}_i \frac{\partial}{\partial x^m}, D^n{}_k \frac{\partial}{\partial x^n} \right](f) &= D^m{}_i \frac{\partial}{\partial x^m}(D^n{}_k)\,(\partial_n(f \circ x^{-1}) \circ x) + \cancel{D^m{}_i D^n{}_k \,\partial_m\partial_n(f \circ x^{-1}) \circ x}\\
&\quad - D^n{}_k \frac{\partial}{\partial x^n}(D^m{}_i)\,(\partial_m(f \circ x^{-1}) \circ x) - \cancel{D^n{}_k D^m{}_i \,\partial_n\partial_m(f \circ x^{-1}) \circ x}\\
&= \left( D^m{}_i \frac{\partial}{\partial x^m}(D^n{}_k) - D^m{}_k \frac{\partial}{\partial x^m}(D^n{}_i) \right)\partial_n(f \circ x^{-1}) \circ x\\
&= \left( D^m{}_i \frac{\partial}{\partial x^m}(D^n{}_k) - D^m{}_k \frac{\partial}{\partial x^m}(D^n{}_i) \right)\frac{\partial}{\partial x^n}(f),
\end{aligned}$$

where we relabelled some dummy indices. Since the $f \in \mathcal{C}^\infty(\mathrm{SL}(2,\mathbb{C}))$ was arbitrary,

$$\left[ D^m{}_i \frac{\partial}{\partial x^m}, D^n{}_k \frac{\partial}{\partial x^n} \right] = \left( D^m{}_i \frac{\partial}{\partial x^m}(D^n{}_k) - D^m{}_k \frac{\partial}{\partial x^m}(D^n{}_i) \right)\frac{\partial}{\partial x^n}.$$

We can now evaluate this explicitly. 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

$$\left[D^m{}_1 \frac{\partial}{\partial x^m}, D^n{}_3 \frac{\partial}{\partial x^n}\right] = \left(D^m{}_1 \frac{\partial}{\partial x^m}(D^1{}_3) - D^m{}_3 \frac{\partial}{\partial x^m}(D^1{}_1)\right)\frac{\partial}{\partial x^1}$$

$$+ \left(D^m{}_1 \frac{\partial}{\partial x^m}(D^2{}_3) - D^m{}_3 \frac{\partial}{\partial x^m}(D^2{}_1)\right)\frac{\partial}{\partial x^2}$$

$$+ \left(D^m{}_1 \frac{\partial}{\partial x^m}(D^3{}_3) - D^m{}_3 \frac{\partial}{\partial x^m}(D^3{}_1)\right)\frac{\partial}{\partial x^3}$$

$$= -2x^2 \frac{\partial}{\partial x^1} - 2\left(\frac{1+x^2 x^3}{x^1}\right)\frac{\partial}{\partial x^3}$$

and

$$\left[D^m{}_2 \frac{\partial}{\partial x^m}, D^n{}_3 \frac{\partial}{\partial x^n}\right] = \left(D^m{}_2 \frac{\partial}{\partial x^m}(D^1{}_3) - D^m{}_3 \frac{\partial}{\partial x^m}(D^1{}_2)\right)\frac{\partial}{\partial x^1}$$

$$+ \left(D^m{}_2 \frac{\partial}{\partial x^m}(D^2{}_3) - D^m{}_3 \frac{\partial}{\partial x^m}(D^2{}_2)\right)\frac{\partial}{\partial x^2}$$

$$+ \left(D^m{}_2 \frac{\partial}{\partial x^m}(D^3{}_3) - D^m{}_3 \frac{\partial}{\partial x^m}(D^3{}_2)\right)\frac{\partial}{\partial x^3}$$

$$= (D^2{}_1 - D^1{}_3)\frac{\partial}{\partial x^1} + D^2{}_3 \frac{\partial}{\partial x^2} - D^3{}_2 \frac{\partial}{\partial x^3}$$

$$= x^1 \frac{\partial}{\partial x^1} - x^2 \frac{\partial}{\partial x^2} + x^3 \frac{\partial}{\partial x^3},$$

where the differentiation rules that we have used come from the definition of the vector field $\frac{\partial}{\partial x^m}$, the Leibniz rule, and the action on co-ordinate functions.

By applying $j^{-1}$, which is just evaluation at the identity, to these vector fields, we finally see that the induced Lie bracket on $T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C})$ satisfies

$$\left[\left(\frac{\partial}{\partial x^1}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}, \left(\frac{\partial}{\partial x^2}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}\right] = 2\left(\frac{\partial}{\partial x^2}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}$$

$$\left[\left(\frac{\partial}{\partial x^1}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}, \left(\frac{\partial}{\partial x^3}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}\right] = -2\left(\frac{\partial}{\partial x^3}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}$$

$$\left[\left(\frac{\partial}{\partial x^2}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}, \left(\frac{\partial}{\partial x^3}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}\right] = \left(\frac{\partial}{\partial x^1}\right)_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}.$$

Hence, the structure constants of $T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} \mathrm{SL}(2, \mathbb{C})$ with respect to the co-ordinate basis are

$$C^2{}_{12} = 2, \qquad C^3{}_{13} = -2, \qquad C^1{}_{23} = 1,$$

with all other being either zero or related to these by anti-symmetry.

## 16 Dynkin diagrams from Lie algebras, and vice versa

### 16.1 The simplicity of $\mathfrak{sl}(2,\mathbb{C})$

We have seen that the non-zero structure constants of $T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)}\mathrm{SL}(2,\mathbb{C})$ are

$$C^2{}_{12} = 2, \qquad C^3{}_{13} = -2, \qquad C^1{}_{23} = 1,$$

plus those related by anti-symmetry.

**Proposition 16.1.** *Two Lie algebras $A$ and $B$ are isomorphic if, and only if, there exists a basis of $A$ and a basis of $B$ in which the structure constants of $A$ and $B$ are the same.*

Since we have already proved that $T_e G \cong_{\mathrm{Lie\,alg}} \mathcal{L}(G)$ for any Lie group $G$, we can deduce the existence of a basis $\{X_1, X_2, X_2\}$ of $\mathfrak{sl}(2,\mathbb{C})$ with respect to which the structure constants are those listed above. In other words, we have

$$[X_1, X_2] = 2X_2,$$
$$[X_1, X_3] = -2X_3,$$
$$[X_2, X_3] = X_1.$$

In this basis, 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. Explicitly, we have

$$\begin{aligned}
\kappa_{11} &= C^m{}_{1n} C^n{}_{1m} \\
&= C^1{}_{1n} C^n{}_{11} + C^2{}_{1n} C^n{}_{12} + C^3{}_{1n} C^n{}_{13} \\
&= C^2{}_{12} C^2{}_{12} + C^3{}_{13} C^3{}_{13} \\
&= 8.
\end{aligned}$$

Since $\kappa$ is symmetric, we only need to determine $\kappa_{ij}$ for $i \leq j$. By writing the components in a $3 \times 3$ array, we find

$$[\kappa_{ij}] = \begin{pmatrix} 8 & 0 & 0 \\ 0 & -8 & 0 \\ 0 & 0 & 8 \end{pmatrix},$$

which is just shorthand for

$$\kappa(X_1, X_1) = 8, \qquad \kappa(X_2, X_2) = -8, \qquad \kappa(X_3, X_3) = 8,$$

and $\kappa(X_i, X_j) = 0$ whenever $i \neq j$.

**Proposition 16.2.** *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. $\qquad\square$

*Remark* 16.3. 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 $\mathrm{SL}(2, \mathbb{C})$ is not a compact Lie group.

In fact, $\mathfrak{sl}(2, \mathbb{C})$ is more than just semi-simple.

**Proposition 16.4.** *The Lie algebra $\mathfrak{sl}(2, \mathbb{C})$ is simple.*

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

$$\forall\, x \in I : \forall\, y \in L : \ [x, y] \in I.$$

*Proof.* 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

$$
\begin{aligned}
{}[\alpha X_1 + \beta X_2 + \gamma X_3, X_1] &= -2\beta X_1 + 2\gamma X_3, \\
[\alpha X_1 + \beta X_2 + \gamma X_3, X_2] &= 2\alpha X_2 - \gamma X_1, \\
[\alpha X_1 + \beta X_2 + \gamma X_3, X_3] &= -2\alpha X_3 + \beta X_1.
\end{aligned}
$$

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 $I \subseteq \mathrm{span}_{\mathbb{C}}(\{X_2, X_3\})$ and hence we must have $\beta = \gamma = 0$ as well;

ii) if $\beta = 0$, then $I \subseteq \mathrm{span}_{\mathbb{C}}(\{X_1, X_3\})$, hence we must have $\alpha = 0$, so that in fact $I \subseteq \mathrm{span}_{\mathbb{C}}(\{X_3\})$, and hence $\gamma = 0$ as well;

iii) if $\gamma = 0$, then $I \subseteq \mathrm{span}_{\mathbb{C}}(\{X_1, X_2\})$, hence we must have $\alpha = 0$, so that in fact $I \subseteq \mathrm{span}_{\mathbb{C}}(\{X_2\})$, and hence $\beta = 0$ as well.

In all cases, we have $I = 0$. Therefore, there are no non-trivial ideals of $\mathfrak{sl}(2, \mathbb{C})$. $\qquad\square$

## 16.2 The roots and Dynkin diagram of $\mathfrak{sl}(2,\mathbb{C})$

By observing the bracket relations of the basis elements of $\mathfrak{sl}(2,\mathbb{C})$, we can see that

$$H := \operatorname{span}_{\mathbb{C}}(\{X_1\})$$

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

$$\operatorname{ad}(h)X_2 = \xi[X_1, X_2] = 2\xi X_2,$$
$$\operatorname{ad}(h)X_3 = \xi[X_1, X_3] = -2\xi X_3.$$

Recall that in the section on Lie algebras, we re-interpreted these eigenvalue equations in terms of functionals $\lambda_2, \lambda_3 \in H^*$

$$\lambda_2\colon \ H \xrightarrow{\sim} \mathbb{C} \qquad\qquad \lambda_3\colon \ H \xrightarrow{\sim} \mathbb{C}$$
$$\xi X_1 \mapsto 2\xi, \qquad\qquad\qquad \xi X_1 \mapsto -2\xi$$

whereby

$$\operatorname{ad}(h)X_2 = \lambda_2(h)X_2,$$
$$\operatorname{ad}(h)X_3 = \lambda_3(h)X_3.$$

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 $\lambda \in \operatorname{span}_{\epsilon,\mathbb{N}}(\Pi)$.

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

$$s_{\lambda_2}\colon H^*_{\mathbb{R}} \to H^*_{\mathbb{R}}$$
$$\mu \mapsto \mu - 2\frac{\kappa^*(\lambda_2, \mu)}{\kappa^*(\lambda_2, \lambda_2)}\lambda_2.$$

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 $A_1 = \mathfrak{sl}(2,\mathbb{C})$.

## 16.3 Reconstruction of $A_2$ from its Dynkin diagram

We have seen an example of how to construct the Dynkin diagram of a Lie algebra, albeit the simplest of this kind. Let us now consider the opposite direction. We will start from the 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 $\varphi = 60°$ and $\varphi = 120°$.
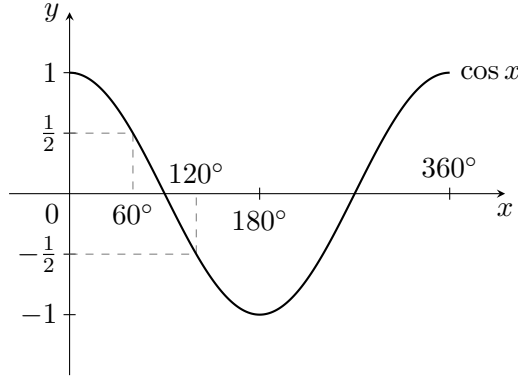


By definition, we have

$$\cos\varphi = \frac{\kappa^*(\pi_1, \pi_2)}{|\pi_1|\,|\pi_2|},$$

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 $\varphi = 120°$. We can thus plot the two fundamental roots in a plane as follows.
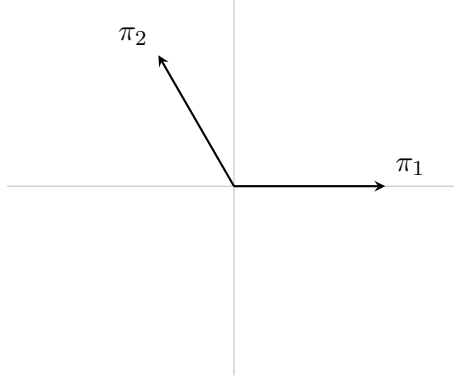
We can determine all the other roots in $\Phi$ by repeated action of the Weyl group. For instance, we easily find that $s_{\pi_1}(\pi_1) = -\pi_1$ and $s_{\pi_2}(\pi_2) = -\pi_2$. We also have

$$s_{\pi_1}(\pi_2) = \pi_2 - 2\frac{\kappa^*(\pi_1, \pi_2)}{\kappa^*(\pi_1, \pi_1)}\pi_1 = \pi_2 - 2(-\tfrac{1}{2})\pi_1 = \pi_1 + \pi_2.$$

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.
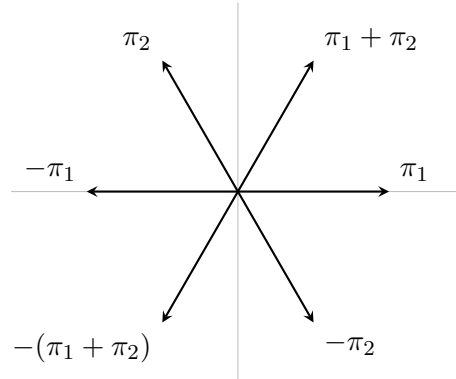


Since $H^* = \mathrm{span}_{\mathbb{C}}(\Pi)$, we have $\dim H^* = 2$, thus the dimension of the Cartan subalgebra is also 2. Since $|\Phi| = 6$, we know that any Cartan-Weyl basis of the Lie algebra $A_2$ must have $2 + 6 = 8$ elements. Hence, the dimension of $A_2$ is 8.

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 $\{h_1, h_2, e_3, \ldots, e_8\}$ a Cartan-Weyl basis of $A_2$, so that $H = \mathrm{span}_{\mathbb{C}}(\{h_1, h_2\})$ and the $e_\alpha$ are eigenvectors of every $h \in H$. Since $A_2$ is simple, $H$ is abelian and hence

$$[h_1, h_2] = 0 \quad \Rightarrow \quad C^k_{\;12} = C^k_{\;21} = 0, \quad \forall\, 1 \le k \le 8.$$

To each $e_\alpha$, for $3 \le \alpha \le 8$, there is an associated $\lambda_\alpha \in \Phi$ such that

$$\forall\, h \in H : \ \mathrm{ad}(h)e_\alpha = \lambda_\alpha(h)e_\alpha.$$

In particular, for the basis elements $h_1, h_2$,

$$[h_1, e_\alpha] = \mathrm{ad}(h_1)e_\alpha = \lambda_\alpha(h_1)e_\alpha,$$
$$[h_2, e_\alpha] = \mathrm{ad}(h_2)e_\alpha = \lambda_\alpha(h_2)e_\alpha,$$

so that we have

$$C^1{}_{1\alpha} = C^2{}_{1\alpha} = 0, \quad C^\alpha{}_{1\alpha} = \lambda_\alpha(h_1), \quad \forall\, 3 \le \alpha \le 8,$$
$$C^1{}_{2\alpha} = C^2{}_{2\alpha} = 0, \quad C^\alpha{}_{2\alpha} = \lambda_\alpha(h_2), \quad \forall\, 3 \le \alpha \le 8.$$

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,

$$\mathrm{ad}(h_i)[e_\alpha, e_\beta] = (\lambda_\alpha(h_i) + \lambda_\beta(h_i))[e_\alpha, e_\beta].$$

If $\lambda_\alpha + \lambda_\beta \in \Phi$, we have $[e_\alpha, e_\beta] = \xi e_\gamma$ for some $3 \le \gamma \le 8$ 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

$$\mathrm{ad}(h)[e_3, e_4] = (\pi_1 + \pi_2)(h)[e_3, e_4],$$

and hence $[e_3, e_4]$ is an eigenvector of $\mathrm{ad}(h)$ 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

$$\big(\forall\, h \in H : [h, x] = 0\big) \Rightarrow x \in H.$$

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 con be used to determine the remaining structure constants of $A_2$.

## 17 Representation theory of Lie groups and Lie algebras

Lie groups and 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. 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.

### 17.1 Representations of Lie algebras

**Definition.** Let $L$ be a Lie algebra. A *representation* of $L$ is a Lie algebra homomorphism

$$\rho \colon L \xrightarrow{\sim} \mathrm{End}(V),$$

where $V$ is some finite-dimensional vector space over the same field as $L$.

Recall that a linear map $\rho \colon L \xrightarrow{\sim} \mathrm{End}(V)$ 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 $\mathrm{End}(V)$.

**Definition.** Let $\rho \colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a representation of $L$.

i) The vector space $V$ is called the *representation space* of $\rho$.

ii) The *dimension* of the representation $\rho$ is $\dim V$.

*Example* 17.1. 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 $\rho \colon \mathfrak{sl}(2, \mathbb{C}) \xrightarrow{\sim} \mathrm{End}(\mathbb{C}^2)$ be the linear map defined by

$$\rho(X_1) := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \rho(X_2) := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \rho(X_3) := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

(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

$$[\rho(X_1), \rho(X_3)] = \rho([X_1, X_3]),$$
$$[\rho(X_2), \rho(X_3)] = \rho([X_2, X_3]).$$

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

$$\mathrm{im}_\rho(\mathfrak{sl}(2, \mathbb{C})) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{End}(\mathbb{C}^2) \,\Big|\, a + d = 0 \right\}$$
$$= \{ \phi \in \mathrm{End}(\mathbb{C}^2) \mid \mathrm{tr}\, \phi = 0 \}.$$

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.

Two representations of a Lie algebra can be related in the following sense.

**Definition.** Let $L$ be a Lie algebra and let

$$\rho_1 \colon L \xrightarrow{\sim} \mathrm{End}(V_1), \qquad \rho_2 \colon L \xrightarrow{\sim} \mathrm{End}(V_2)$$

be representations of $L$. A linear map $f \colon V_1 \xrightarrow{\sim} V_2$ is a *homomorphism of representations* if

$$\forall\, x \in L : \ f \circ \rho_1(x) = \rho_2(x) \circ f.$$

Equivalently, if the following diagram commutes for all $x \in L$.

$$
\begin{array}{ccc}
V_1 & \xrightarrow{\ \ f\ \ } & V_2 \\
{\scriptstyle \rho_1(x)}\big\downarrow & & \big\downarrow{\scriptstyle \rho_2(x)} \\
V_1 & \xrightarrow{\ \ f\ \ } & V_2
\end{array}
$$

If in addition $f \colon V_1 \xrightarrow{\sim} V_2$ is a linear isomorphism, then $f^{-1} \colon V_2 \xrightarrow{\sim} V_1$ is automatically a homomorphism of representations, since

$$f \circ \rho_1(x) = \rho_2(x) \circ f \ \Leftrightarrow\ f^{-1} \circ (f \circ \rho_1(x)) \circ f^{-1} = f^{-1} \circ (\rho_2(x) \circ f) \circ f^{-1}$$
$$\Leftrightarrow\ \rho_1(x) \circ f^{-1} = f^{-1} \circ \rho_2(x).$$

**Definition.** An *isomorphism of representations* of Lie algebras is a bijective homomorphism of representations.

Isomorphic representations necessarily have the same dimension.

*Example* 17.2. Consider $\mathfrak{so}(3, \mathbb{R})$, the Lie algebra of the rotation group $\mathrm{SO}(3, \mathbb{R})$. It is a 3-dimensional Lie algebra over $\mathbb{R}$. It has a basis $\{J_1, J_2, J_3\}$ satisfying

$$[J_i, J_j] = C^k{}_{ij} J_k,$$

where the structure constants $C^k{}_{ij}$ are defined by first "pulling the index $k$ down" using the Killing form $\kappa_{ab} = C^m{}_{an}C^n{}_{bm}$ to obtain $C_{kij} := \kappa_{km}C^m{}_{ij}$, 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

$$[J_1, J_2] = J_3,$$
$$[J_2, J_3] = J_1,$$
$$[J_3, J_1] = J_2.$$

Define a linear map $\rho_{\text{vec}} \colon \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3)$ by

$$\rho_{\text{vec}}(J_1) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \rho_{\text{vec}}(J_2) := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad \rho_{\text{vec}}(J_3) := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

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

$$\rho_{\text{spin}} \colon \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{C}^2),$$

with $\mathbb{C}^2$ understood as a 4-dimensional $\mathbb{R}$-vector space, defined by

$$\rho_{\text{spin}}(J_1) := -\frac{i}{2}\,\sigma_1, \qquad \rho_{\text{spin}}(J_2) := -\frac{i}{2}\,\sigma_2, \qquad \rho_{\text{spin}}(J_3) := -\frac{i}{2}\,\sigma_3,$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

You can again check that this is a representation of $\mathfrak{so}(3, \mathbb{R})$. Since

$$\dim \mathbb{R}^3 = 3 \neq 4 = \dim \mathbb{C}^2,$$

the representations $\rho_{\text{vec}}$ and $\rho_{\text{spin}}$ are not isomorphic.

Any (non-abelian) Lie algebra always has at least two special representations.

**Definition.** Let $L$ be a Lie algebra. A *trivial representation* of $L$ is defined by

$$\rho_{\text{trv}} \colon L \xrightarrow{\sim} \text{End}(V)$$
$$x \mapsto \rho_{\text{trv}}(x) := 0,$$

where $0$ denotes the trivial endomorphism on $V$.

**Definition.** The *adjoint representation* of $L$ is

$$\rho_{\mathrm{adj}} \colon L \xrightarrow{\sim} \mathrm{End}(L)$$
$$x \mapsto \rho_{\mathrm{adj}}(x) := \mathrm{ad}(x).$$

These are indeed representations since we have already shown that ad is a Lie algebra homomorphism, while for the trivial representations we have

$$\forall\, x, y \in L \colon\ \rho_{\mathrm{trv}}([x, y]) = 0 = [\rho_{\mathrm{trv}}(x), \rho_{\mathrm{trv}}(y)].$$

**Definition.** A representation $\rho \colon L \xrightarrow{\sim} \mathrm{End}(V)$ is called *faithful* if $\rho$ is injective, i.e.

$$\dim(\mathrm{im}_{\rho}(L)) = \dim L.$$

*Example* 17.3. 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

$$\mathrm{ad}(x) = \mathrm{ad}(y) \Leftrightarrow \forall\, z \in L : \mathrm{ad}(x)z = \mathrm{ad}(y)z$$
$$\Leftrightarrow \forall\, z \in L : [x, z] = [y, z]$$
$$\Leftrightarrow \forall\, z \in L : [x - y, z] = 0.$$

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 ad is injective.

**Definition.** Given two representations $\rho_1 \colon L \xrightarrow{\sim} \mathrm{End}(V_1)$, $\rho_2 \colon L \xrightarrow{\sim} \mathrm{End}(V_2)$, we can construct new representations called

   i) the *direct sum representation*

$$\rho_1 \oplus \rho_2 \colon L \xrightarrow{\sim} \mathrm{End}(V_1 \oplus V_2)$$
$$x \mapsto (\rho_1 \oplus \rho_2)(x) := \rho_1(x) \oplus \rho_2(x)$$

   ii) the *tensor product representation*

$$\rho_1 \otimes \rho_2 \colon L \xrightarrow{\sim} \mathrm{End}(V_1 \times V_2)$$
$$x \mapsto (\rho_1 \otimes \rho_2)(x) := \rho_1(x) \otimes \mathrm{id}_{V_2} + \mathrm{id}_{V_1} \otimes \rho_2(x).$$

*Example* 17.4. The direct sum representation $\rho_{\mathrm{vec}} \oplus \rho_{\mathrm{spin}} \colon \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \mathrm{End}(\mathbb{R}^3 \oplus \mathbb{C}^2)$ given in block-matrix form by

$$(\rho_{\mathrm{vec}} \oplus \rho_{\mathrm{spin}})(x) = \left( \begin{array}{c|c} \rho_{\mathrm{vec}}(x) & 0 \\ \hline 0 & \rho_{\mathrm{spin}}(x) \end{array} \right)$$

is a 7-dimensional representation of $\mathfrak{so}(3, \mathbb{R})$.

**Definition.** A representation $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ 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 $\rho|_U\colon L \xrightarrow{\sim} \mathrm{End}(U)$.

**Definition.** A representation is *irreducible* if it is not reducible.

*Example* 17.5.     i) The representation $\rho_{\mathrm{vec}} \oplus \rho_{\mathrm{spin}}\colon \mathfrak{so}(3,\mathbb{R}) \xrightarrow{\sim} \mathrm{End}(\mathbb{R}^3 \oplus \mathbb{C}^2)$ is reducible since, for example, we have a subspace $\mathbb{R}^3 \oplus 0$ such that

$$\forall\, x \in \mathfrak{so}(3,\mathbb{R}) : \forall\, u \in \mathbb{R}^3 \oplus 0 :\ (\rho_{\mathrm{vec}} \oplus \rho_{\mathrm{spin}})(x)u \in \mathbb{R}^3 \oplus 0.$$

ii) The representations $\rho_{\mathrm{vec}}$ and $\rho_{\mathrm{spin}}$ are both irreducible.

*Remark* 17.6. 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 con be decomposed as the direct sum of irreducible representations, which can then be classified according to their so-called *highest weights*.

## 17.2   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.** Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a representation of a complex Lie algebra $L$. We define the *$\rho$-Killing form* on $L$ as

$$
\begin{aligned}
\kappa_\rho\colon L \times L &\xrightarrow{\sim} \mathbb{C} \\
(x,y) &\mapsto \kappa_\rho(x,y) := \mathrm{tr}(\rho(x) \circ \rho(y)).
\end{aligned}
$$

Of course, the Killing form we have considered so far is just $\kappa_{\mathrm{ad}}$. Similarly to $\kappa_{\mathrm{ad}}$, every $\kappa_\rho$ is symmetric and associative with respect to the Lie bracket of $L$.

**Proposition 17.7.** *Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ 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 $\{X_1, \ldots, X_{\dim L}\}$ is a basis of $L$, then the dual basis $\{\widetilde{X}^1, \ldots, \widetilde{X}^{\dim L}\}$ of $L^*$ is defined by

$$\widetilde{X}^i(X_j) = \delta^i_j.$$

By using the isomorphism induced by $\kappa_\rho$, we can find some $\xi_1, \ldots, \xi_{\dim L} \in L$ such that we have $\kappa(\xi_i, -) = \widetilde{X}^i$ or, equivalently,

$$\forall\, x \in L : \ \kappa_\rho(x, \xi_i) = \widetilde{X}^i(x).$$

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 17.8.** *Let $\{X_i\}$ and $\{\xi_j\}$ be defined as above. Then*

$$[X_j, \xi_k] = \sum_{m=1}^{\dim L} C^k{}_{mj} \xi_m,$$

*where $C^k{}_{mj}$ are the structure constants with respect to $\{X_i\}$.*

*Proof.* By using the associativity of $\kappa_\rho$, we have

$$\kappa_\rho(X_i, [X_j, \xi_k]) = \kappa_\rho([X_i, X_j], \xi_k) = C^m{}_{ij}\kappa_\rho(X_m, \xi_k) = C^m{}_{ij}\delta_{mk} = C^k{}_{ij}.$$

But we also have

$$\kappa_\rho\!\Big(X_i, \sum_{m=1}^{\dim L} C^k{}_{mj}\xi_m\Big) = \sum_{m=1}^{\dim L} C^k{}_{mj}\kappa_\rho(X_i, \xi_m) = \sum_{m=1}^{\dim L} C^k{}_{mj}\delta_{im} = C^k{}_{ij}.$$

Therefore

$$\forall\, 1 \leq i \leq \dim L : \ \kappa_\rho\!\Big(X_i, [X_j, \xi_k] - \sum_{m=1}^{\dim L} C^k{}_{mj}\xi_m\Big) = 0$$

and hence, the result follows from the non-degeneracy of $\kappa_\rho$. $\qquad\qquad\square$

We are now ready to define the Casimir operator and prove the subsequent theorem.

**Definition.** Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a faithful representation of a complex (compact) Lie algebra $L$ and let $\{X_1, \ldots, X_{\dim L}\}$ be a basis of $L$. The *Casimir operator* associated to the representation $\rho$ is the endomorphism $\Omega_\rho\colon V \xrightarrow{\sim} V$

$$\Omega_\rho := \sum_{i=1}^{\dim L} \rho(X_i) \circ \rho(\xi_i).$$

**Theorem 17.9.** *Let $\Omega_\rho$ the Casimir operator of a representation $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$. Then*

$$\forall\, x \in L : \ [\Omega_\rho, \rho(x)] = 0,$$

*that is, $\Omega_\rho$ commutes with every endomorphism in $\mathrm{im}_\rho(L)$.*

*Proof.* Note that the bracket above is that on $\text{End}(V)$. Let $x = x^k X_k \in L$. Then

$$[\Omega_\rho, \rho(x)] = \left[\sum_{i=1}^{\dim L} \rho(X_i) \circ \rho(\xi_i), \rho(x^k X_k)\right]$$

$$= \sum_{i,k=1}^{\dim L} x^k [\rho(X_i) \circ \rho(\xi_i), \rho(X_k)].$$

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

$$[AB, C] = ABC - CBA$$
$$= ABC - CBA - ACB + ACB$$
$$= A[B, C] + [A, C]B.$$

Hence, by applying this, we obtain

$$\sum_{i,k=1}^{\dim L} x^k [\rho(X_i) \circ \rho(\xi_i), \rho(X_k)] = \sum_{i,k=1}^{\dim L} x^k \big(\rho(X_i) \circ [\rho(\xi_i), \rho(X_k)] + [\rho(X_i), \rho(X_k)] \circ \rho(\xi_i)\big)$$

$$= \sum_{i,k=1}^{\dim L} x^k \big(\rho(X_i) \circ \rho([\xi_i, X_k]) + \rho([X_i, X_k]) \circ \rho(\xi_i)\big)$$

$$= \sum_{i,k,m=1}^{\dim L} x^k \big(\rho(X_i) \circ \rho(-C^i{}_{mk}\xi_m) + \rho(C^m{}_{ik}X_m) \circ \rho(\xi_i)\big)$$

$$= \sum_{i,k,m=1}^{\dim L} x^k \big(-C^i{}_{mk}\rho(X_i) \circ \rho(\xi_m) + C^m{}_{ik}\rho(X_m) \circ \rho(\xi_i)\big)$$

$$= \sum_{i,k,m=1}^{\dim L} x^k \big(-C^i{}_{mk}\rho(X_i) \circ \rho(\xi_m) + C^i{}_{mk}\rho(X_i) \circ \rho(\xi_m)\big)$$

$$= 0,$$

where we have swapped the dummy summation indices $i$ and $m$ in the second term. $\qquad\square$

**Lemma 17.10** (Schur). *If $\rho\colon L \xrightarrow{\sim} \text{End}(V)$ 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 17.11.** *The Casimir operator of $\rho\colon L \xrightarrow{\sim} \text{End}(V)$ is $\Omega_\rho = c_\rho \,\text{id}_V$, where*

$$c_\rho = \frac{\dim L}{\dim V}.$$

*Proof.* We have

$$\mathrm{tr}(\Omega_\rho) = \mathrm{tr}(c_\rho\,\mathrm{id}_V) = c_\rho\dim V$$

and

$$
\begin{aligned}
\mathrm{tr}(\Omega_\rho) &= \mathrm{tr}\left(\sum_{i=1}^{\dim L}\rho(X_i)\circ\rho(\xi_i)\right)\\
&= \sum_{i=1}^{\dim L}\mathrm{tr}(\rho(X_i)\circ\rho(\xi_i))\\
&= \sum_{i=1}^{\dim L}\kappa_\rho(X_i,\xi_i)\\
&= \sum_{i=1}^{\dim L}\delta_{ii}\\
&= \dim L,
\end{aligned}
$$

which is what we wanted. $\square$

*Example* 17.12. 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 $\rho_{\mathrm{vec}}\colon\mathfrak{so}(3,\mathbb{R})\xrightarrow{\sim}\mathrm{End}(\mathbb{R}^3)$ is defined by

$$
\rho_{\mathrm{vec}}(J_1) := \begin{pmatrix} 0 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 0\end{pmatrix},\qquad
\rho_{\mathrm{vec}}(J_2) := \begin{pmatrix} 0 & 0 & 1\\ 0 & 0 & 0\\ -1 & 0 & 0\end{pmatrix},\qquad
\rho_{\mathrm{vec}}(J_3) := \begin{pmatrix} 0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}.
$$

Let us first evaluate the components of $\kappa_{\rho_{\mathrm{vec}}}$. We have

$$
\begin{aligned}
(\kappa_{\rho_{\mathrm{vec}}})_{11} := \kappa_{\rho_{\mathrm{vec}}}(J_1,J_1) &= \mathrm{tr}(\rho_{\mathrm{vec}}(J_1)\circ\rho_{\mathrm{vec}}(J_1))\\
&= \mathrm{tr}((\rho_{\mathrm{vec}}(J_1))^2)\\
&= \mathrm{tr}\begin{pmatrix} 0 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 0\end{pmatrix}^2\\
&= \mathrm{tr}\begin{pmatrix} 0 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\end{pmatrix}\\
&= -2.
\end{aligned}
$$

After calculating the other components similarly, we find

$$
[(\kappa_{\rho_{\mathrm{vec}}})_{ij}] = \begin{pmatrix} -2 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & -2\end{pmatrix}.
$$

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* 17.13. 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

$$
\rho_{\text{spin}}(J_1) := -\frac{i}{2} \sigma_1, \qquad \rho_{\text{spin}}(J_2) := -\frac{i}{2} \sigma_2, \qquad \rho_{\text{spin}}(J_3) := -\frac{i}{2} \sigma_3,
$$

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

$$
\begin{aligned}
(\kappa_{\rho_{\text{spin}}})_{11} := \kappa_{\rho_{\text{spin}}}(J_1, J_1) &= \text{tr}(\rho_{\text{spin}}(J_1) \circ \rho_{\text{spin}}(J_1)) \\
&= \text{tr}((\rho_{\text{spin}}(J_1))^2) \\
&= (-\tfrac{i}{2})^2 \, \text{tr}(\sigma_1^2) \\
&= -\tfrac{1}{4} \, \text{tr}(\text{id}_{\mathbb{C}^2}) \\
&= -1.
\end{aligned}
$$

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

$$
\begin{aligned}
\Omega_{\rho_{\text{spin}}} &:= \sum_{i=1}^{3} \rho_{\text{spin}}(J_i) \circ \rho_{\text{spin}}(\xi_i) \\
&= \sum_{i=1}^{3} \rho_{\text{spin}}(J_i) \circ \rho_{\text{spin}}(-J_i) \\
&= -\sum_{i=1}^{3} (\rho_{\text{spin}}(J_i))^2 \\
&= -\left(-\frac{\mathrm{i}}{2}\right)^2 \sum_{i=1}^{3} \sigma_i^2 \\
&= \frac{1}{4} \sum_{i=1}^{3} \mathrm{id}_{\mathbb{C}^2} \\
&= \frac{3}{4} \mathrm{id}_{\mathbb{C}^2},
\end{aligned}
$$

in accordance with the fact that

$$
\frac{\dim \mathfrak{so}(3, \mathbb{R})}{\dim \mathbb{C}^2} = \frac{3}{4}.
$$

## 17.3 Representations of Lie groups

We now turn to representations of Lie groups. Given a vector space $V$, recall that the subset of $\text{End}(V)$ 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.** A *representation* of a Lie group $(G, \bullet)$ is a Lie group homomorphism

$$
R \colon G \to \text{GL}(V)
$$

for some finite-dimensional vector space $V$.

Recall that $R \colon G \to \text{GL}(V)$ 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

$$
R(e) = \text{id}_V \qquad \text{and} \qquad R(g^{-1}) = R(g)^{-1}.
$$

*Example* 17.14. Consider the Lie group $\mathrm{SO}(2, \mathbb{R})$. As a smooth manifold, $\mathrm{SO}(2, \mathbb{R})$ 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 $\theta \colon U \to [0, 2\pi) \subseteq \mathbb{R}$ on $S^1$ by mapping each point in $U$ to an "angle" in $[0, 2\pi)$.
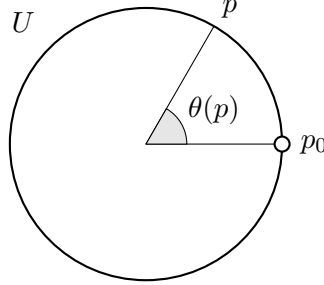


The operation

$$p_1 \bullet p_2 := (\theta(p_1) + \theta(p_2)) \mod 2\pi$$

endows $S^1 \cong_{\mathrm{diff}} \mathrm{SO}(2, \mathbb{R})$ with the structure of a Lie group. Then, a representation of $\mathrm{SO}(2, \mathbb{R})$ is given by

$$R \colon \mathrm{SO}(2, \mathbb{R}) \to \mathrm{GL}(\mathbb{R}^2)$$
$$p \mapsto \begin{pmatrix} \cos \theta(p) & \sin \theta(p) \\ -\sin \theta(p) & \cos \theta(p) \end{pmatrix}.$$

Indeed, the addition formulæfor sine and cosine imply that

$$R(p_1 \bullet p_2) = R(p_1) \circ R(p_2).$$

*Example* 17.15. Let $G$ be a Lie group (we suppress the $\bullet$ in this example). For each $g \in G$, define the Adjoint map

$$\mathrm{Ad}_g \colon G \to G$$
$$h \mapsto ghg^{-1}.$$

Note the capital "A" to distinguish this from the adjoint map on Lie algebras. Since $\mathrm{Ad}_g$ is a composition of the Lie group multiplication and inverse map, it is a smooth map. Moreover, we have

$$\mathrm{Ad}_g(e) = geg^{-1} = gg^{-1} = e.$$

Hence, the push-forward of $\mathrm{Ad}_g$ at the identity is the map

$$(\mathrm{Ad}_{g*})_e \colon T_e G \xrightarrow{\sim} T_{\mathrm{Ad}_g(e)} G = T_e G.$$

Thus, we have $\mathrm{Ad}_g \in \mathrm{End}(T_e G)$. In fact, you can check that

$$(\mathrm{Ad}_{g^{-1}*})_e \circ (\mathrm{Ad}_{g*})_e = (\mathrm{Ad}_{g*})_e \circ (\mathrm{Ad}_{g^{-1}*})_e = \mathrm{id}_{T_e G},$$

and hence we have, in particular, $\mathrm{Ad}_g \in \mathrm{GL}(T_e G) \cong_{\mathrm{Lie\,grp}} \mathrm{GL}(\mathcal{L}(G))$.

We can therefore construct a map

$$\mathrm{Ad}\colon G \to \mathrm{GL}(T_e G)$$

$$g \mapsto \mathrm{Ad}_{g*}$$

which, as you can check, is a representation of $G$ on its Lie algebra.

*Remark 17.16.* 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

$$(R_*)_e \colon T_e G \xrightarrow{\sim} T_{\mathrm{id}_V} \mathrm{GL}(V).$$

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 $\mathrm{GL}(V)$. In fact, in the previous example we have

$$(\mathrm{Ad}_*)_e = \mathrm{ad},$$

where ad is the adjoint representation of $T_e G$.

## 18 Reconstruction of a Lie group from its 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.

### 18.1 Integral curves

**Definition.** Let $M$ be a smooth manifold and let $Y \in \Gamma(TM)$. An *integral curve* of $Y$ is smooth curve $\gamma \colon (-\varepsilon, \varepsilon) \to M$, with $\varepsilon > 0$, such that

$$\forall \lambda \in (-\varepsilon, \varepsilon) : \ X_{\gamma, \gamma(\lambda)} = Y|_{\gamma(\lambda)}.$$

It follows from the local existence and uniqueness of solutions to ordinary differential equations that, given any $Y \in \Gamma(TM)$ and any $p \in M$, there exist $\varepsilon > 0$ and a smooth curve $\gamma \colon (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = p$ which is an integral curve of $Y$.

Moreover, integral curves are locally unique. By this we mean that if $\gamma_1$ and $\gamma_2$ are both integral curves of $Y$ through $p$, i.e. $\gamma_1(0) = \gamma_2(0) = p$, then $\gamma_1 = \gamma_2$ on the intersection of their domains of definition. We can get genuine uniqueness as follows.

**Definition.** The *maximal integral curve* of $Y \in \Gamma(TM)$ through $p \in M$ is the unique integral curve $\gamma \colon I_{\max}^p \to M$ of $Y$ through $p$, where

$$I_{\max}^p := \bigcup \{I \subseteq \mathbb{R} \mid \text{there exists an integral curve } \gamma \colon I \to M \text{ of } Y \text{ through } p\}.$$

For a given vector field, in general, $I_{\max}^p$ will differ from point to point.

**Definition.** A vector field is *complete* if $I_{\max}^p = \mathbb{R}$ for all $p \in M$.

We have the following result.

**Theorem 18.1.** *On a compact manifold, every vector field is complete.*

On a Lie group, even if non-compact, there are always complete vector fields.

**Theorem 18.2.** *Every left-invariant vector field on a Lie group is complete.*

The maximal 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.

### 18.2 The exponential map

Let $G$ be a Lie group. Recall that given any $A \in T_eG$, we can define the uniquely determined left-invariant vector field $X^A := j(A)$ via the isomorphism $j \colon T_eG \xrightarrow{\sim} \mathcal{L}(G)$ as

$$X^A|_g := (\ell_g)_*(A).$$

Then let $\gamma^A \colon \mathbb{R} \to G$ be the maximal integral curve of $X^A$ through $e \in G$.

**Definition.** Let $G$ be a Lie group. The *exponential map* is defined as

$$\exp \colon T_e G \to G$$
$$A \mapsto \exp(A) := \gamma^A(1)$$

**Theorem 18.3.**    *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*

$$\exp|_V \colon V \to \exp(V) \subseteq G$$

*is bijective and both* $\exp|_V$ *and* $(\exp|_V)^{-1}$ *are smooth.*

*ii) If* $G$ *is compact, then* $\exp$ *is surjective.*

Note that the maximal integral curve of $X^0$ is the constant curve $\gamma^0(\lambda) \equiv e$, and hence we have $\exp(0) = e$. Then first part of the theorem then 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 $T_e G \to G$. But as $G$ is compact and $T_e G$ is not, they are not diffeomorphic.

**Proposition 18.4.** *Let* $G$ *be a Lie group. The image of* $\exp \colon T_e G \to G$ *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* 18.5. Let $B \colon V \times V$ be a pseudo inner product on $V$. Then

$$\mathrm{O}(V) := \{\phi \in \mathrm{GL}(V) \mid \forall\, v, w \in V : B(\phi(v), \phi(w)) = B(v, w)\}$$

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 $\phi \in \mathrm{O}(V)$ has determinant $1$ or $-1$. Since det is multiplicative, we have a subgroup

$$\mathrm{SO}(V) := \{\phi \in \mathrm{O}(V) \mid \det \phi = 1\}.$$

These are, in fact, Lie subgroups of $\mathrm{GL}(V)$. The Lie group $\mathrm{SO}(V)$ is connected while

$$\mathrm{O}(V) = \mathrm{SO}(V) \cup \{\phi \in \mathrm{O}(V) \mid \det \phi = -1\}$$

is disconnected. Since $\mathrm{SO}(V)$ contains $\mathrm{id}_V$, we have

$$\mathfrak{so}(V) := T_{\mathrm{id}_V}\mathrm{SO}(V) = T_{\mathrm{id}_V}\mathrm{O}(V) =: \mathfrak{o}(V)$$

and

$$\exp(\mathfrak{so}(V)) = \exp(\mathfrak{o}(V)) = \mathrm{SO}(V).$$

*Example* 18.6. Choosing a basis $A_1, \ldots, A_{\dim G}$ 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 $0 \le \mu, \nu \le 3$ 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 $\mathfrak{o}(3,1)$. For convenience, instead of denoting a basis of $\mathfrak{o}(3,1)$ as $\{M^i \mid i = 1, \ldots, 6\}$, we will denote it as $\{M^{\mu\nu} \mid 0 \le \mu, \nu \le 3\}$ 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 $\{M^{\mu\nu} \mid 0 \le \mu, \nu \le 3\}$, 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 $M^{\mu\nu}$,

$$\lambda = \tfrac{1}{2} \omega_{\mu\nu} M^{\mu\nu}$$

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 have

$$\Lambda = \exp(\lambda) = \exp(\tfrac{1}{2} \omega_{\mu\nu} M^{\mu\nu}) \in \mathrm{O}(3,1).$$

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 $\text{id}_{\mathbb{R}^4} \in \text{SO}^+(3,1)$, we have $\exp(\mathfrak{o}(3,1)) = \text{SO}^+(3,1)$. Then $\{M^{\mu\nu}\}$ provides a nice co-ordinatisation of $\text{SO}^+(3,1)$ 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 $\exp(\frac{1}{2}\omega_{\mu\nu}M^{\mu\nu}) \in \text{SO}^+(3,1)$ 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 $\text{SO}^+(3,1)$ as being generated by $\{M^{\mu\nu}\}$.

A representation $\rho \colon T_{\text{id}_{\mathbb{R}^4}}\text{SO}^+(3,1) \xrightarrow{\sim} \text{End}(\mathbb{R}^4)$ is given by

$$\rho(M^{\mu\nu})^a{}_b := \eta^{\nu a}\delta^\mu_b - \eta^{\mu a}\delta^\nu_b$$

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

$$R \colon \text{SO}^+(3,1) \to \text{GL}(\mathbb{R}^4)$$

via the exponential map by defining

$$R(\Lambda) = \exp(\tfrac{1}{2}\omega_{\mu\nu}\rho(M^{\mu\nu})).$$

Then, the map exp becomes the usual exponential (series) of matrices.

**Definition.** A *one-parameter subgroup* of a Lie group $G$ is a Lie group homomorphism

$$\xi \colon \mathbb{R} \to G,$$

with $\mathbb{R}$ understood as a Lie group under ordinary addition.

*Example* 18.7. Let $M$ be a smooth manifold and let $Y \in \Gamma(TM)$ be a complete vector field. The *flow* of $Y$ is the smooth map

$$\Theta \colon \mathbb{R} \times M \to M$$
$$(\lambda, p) \mapsto \Theta_\lambda(p) := \gamma_p(\lambda),$$

where $\lambda_p$ is the maximal integral curve of $Y$ through $p$. For a fixed $p$, we have

$$\Theta_0 = \text{id}_M, \qquad \Theta_{\lambda_1} \circ \Theta_{\lambda_2} = \Theta_{\lambda_1 + \lambda_2}, \qquad \Theta_{-\lambda} = \Theta_\lambda^{-1}.$$

For each $\lambda \in \mathbb{R}$, the map $\Theta_\lambda$ is a diffeomorphism $M \to M$. Denoting by $\text{Diff}(M)$ the group (under composition) of the diffeomorphisms $M \to M$, we have that the map

$$\xi \colon \mathbb{R} \to \text{Diff}(M)$$
$$\lambda \mapsto \Theta_\lambda$$

is a one-parameter subgroup of $\text{Diff}(M)$.

**Theorem 18.8.** *Let $G$ be a Lie group.*

*i) Let $A \in T_e G$. The map*

$$\xi^A \colon \mathbb{R} \to G$$
$$\lambda \mapsto \xi^A(\lambda) := \exp(\lambda A)$$

*is a one-parameter subgroup.*

*ii) Every one-parameter subgroup of $G$ has the form $\xi^A$ for some $A \in T_e G$.*

Therefore, the Lie algebra allows us to study all the one-parameter subgroups of the Lie group.

**Theorem 18.9.** *Let $G$ and $H$ be Lie groups and let $\phi \colon G \to H$ be a Lie group homomorphism. Then, for all $A \in T_{e_G} G$, we have*

$$\phi(\exp(A)) = \exp((\phi_*)_{e_G} A).$$

*Equivalently, the following diagram commutes.*

$$
\begin{array}{ccc}
T_{e_G} G & \xrightarrow{\;(\phi_*)_{e_G}\;} & T_{e_H} H \\
{\scriptstyle \exp}\big\downarrow & & \big\downarrow{\scriptstyle \exp} \\
G & \xrightarrow{\quad \phi \quad} & H
\end{array}
$$

In particular, for $\phi \equiv \operatorname{Ad}_g \colon G \to G$, we have

$$\operatorname{Ad}_g(\exp(A)) = \exp((\operatorname{Ad}_{g_*})_e A).$$

## 19 Principal fibre bundles

Very roughly speaking, a principal fibre bundle is a bundle whose typical fibre is a Lie group. Principal fibre bundles are so immensely important because they allow us to understand any fibre bundle with fibre $F$ on which a Lie group $G$ acts. These are then called associated fibre bundles, and will be discussed later on.

### 19.1 Lie group actions on a manifold

**Definition.** Let $(G, \bullet)$ be a Lie group and let $M$ be a smooth manifold. A smooth map

$$\rhd \colon G \times M \to M$$
$$(g, p) \mapsto g \rhd p$$

satisfying

i) $\forall\, p \in M : \ e \rhd p = p$;

ii) $\forall\, g_1, g_2 \in G : \forall\, p \in M : \ (g_1 \bullet g_2) \rhd p = g_1 \rhd (g_2 \rhd p)$,

is called a *left Lie group action*, or *left $G$-action*, on $M$.

**Definition.** A manifold equipped with a left $G$-action is called a *left $G$-manifold*.

*Remark* 19.1. Note that in the above definition, the smooth structures on $G$ and $M$ were only used in the requirement that $\rhd$ 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* 19.2. Let $G$ be a Lie group and let $R \colon G \to \mathrm{GL}(V)$ be a representation of $G$ on a vector space $V$. Define a map

$$\rhd \colon G \times V \to V$$
$$(g, v) \mapsto g \rhd v := R(g)v.$$

We easily check that $e \rhd v := R(e)v = \mathrm{id}_V v = v$ and

$$\begin{aligned}
(g_1 \bullet g_2) \rhd v &:= R(g_1 \bullet g_2)v \\
&= (R(g_1) \circ R(g_2))v \\
&= R(g_1)(R(g_2)v) \\
&= g_1 \rhd (g_2 \rhd v),
\end{aligned}$$

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 $\rhd$ 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.** Similarly, a *right G-action* on $M$ is a smooth map

$$\lhd \colon M \times G \to M$$
$$(p, g) \mapsto p \lhd g$$

satisfying

i) $\forall\, p \in M : \; p \lhd g = p$;

ii) $\forall\, g_1, g_2 \in G : \forall\, p \in M : \; p \lhd (g_1 \bullet g_2) = (p \lhd g_1) \lhd g_2$.

**Proposition 19.3.** *Let $\rhd$ be a left G-action on $M$. Then*

$$\lhd \colon M \times G \to M$$
$$(p, g) \mapsto p \lhd g := g^{-1} \rhd p$$

*is a right G-action on $M$.*

*Proof.* First note that $\lhd$ is smooth since it is a composition of $\rhd$ and the inverse map on $G$, which are both smooth. We have $p \lhd e := e \rhd p = p$ and

$$
\begin{aligned}
p \lhd (g_1 \bullet g_2) &:= (g_1 \bullet g_2)^{-1} \rhd p \\
&= (g_2^{-1} \bullet g_1^{-1}) \rhd p \\
&= g_2^{-1} \rhd (g_1^{-1} \rhd p) \\
&= g_2^{-1} \rhd (p \lhd g_1) \\
&= (p \lhd g_1) \lhd g_2,
\end{aligned}
$$

for all $p \in M$ and $g_1, g_2 \in G$, and hence $\lhd$ is a right $G$-action. $\qquad\square$

*Remark* 19.4. 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* 19.5. Recall that if we have a basis $e_1, \ldots, e_{\dim M}$ of $T_pM$ and $X^1, \ldots, X^{\dim M}$ are the components of some $X \in T_pM$ in this basis, then under a change of basis

$$\widetilde{e}_a = A^b{}_a e_b,$$

we have $X = \widetilde{X}^a \widetilde{e}_a$, where

$$\widetilde{X}^a = (A^{-1})^a{}_b X^b.$$

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 $\mathrm{GL}(\dim M, \mathbb{R})$ on the basis and a left action of the same $\mathrm{GL}(\dim M, \mathbb{R})$ on the components.

**Definition.** Let $G, H$ be Lie groups, let $\rho\colon G \to H$ be a Lie group homomorphism and let

$$\rhd\colon G \times M \to M,$$
$$\blacktriangleright\colon H \times N \to N$$

be left actions of $G$ and $H$ on some smooth manifolds $M$ and $N$, respectively. Then, a smooth map $f\colon M \to N$ is said to be *$\rho$-equivariant* if the diagram

$$
\begin{array}{ccc}
G \times M & \xrightarrow{\;\;\rho \times f\;\;} & H \times N \\
{\scriptstyle \rhd}\Big\downarrow & & \Big\downarrow{\scriptstyle \blacktriangleright} \\
M & \xrightarrow{\;\;\;\;f\;\;\;\;} & N
\end{array}
$$

where $(\rho \times f)(g, p) := (\rho(g), f(p)) \in H \times N$, commutes. Equivalently,

$$\forall\, g \in G : \forall\, p \in M : \; f(g \rhd p) = \rho(g) \blacktriangleright f(p).$$

In other words, if $\rho\colon G \to H$ 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* 19.6. Note that by setting $\rho = \mathrm{id}_G$ or $f = \mathrm{id}_M$, 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.** Let $\rhd\colon G \times M \to M$ be a left $G$-action. For each $p \in M$, we define the *orbit* of $p$ as the set

$$G_p := \{q \in M \mid \exists\, g \in G : q = g \rhd p\}.$$

Alternatively, the orbit of $p$ is the image of $G$ under the map $(- \rhd p)$. It consists of all the points in $M$ that can be reached from $p$ by successive applications of the action $\rhd$.

*Example* 19.7. Consider the action induced by representation of $\mathrm{SO}(2, \mathbb{R})$ as rotation matrices in $\mathrm{End}(\mathbb{R}^2)$. The orbit of any $p \in \mathbb{R}^2$ is the circle of radius $|p|$ centred at the origin.
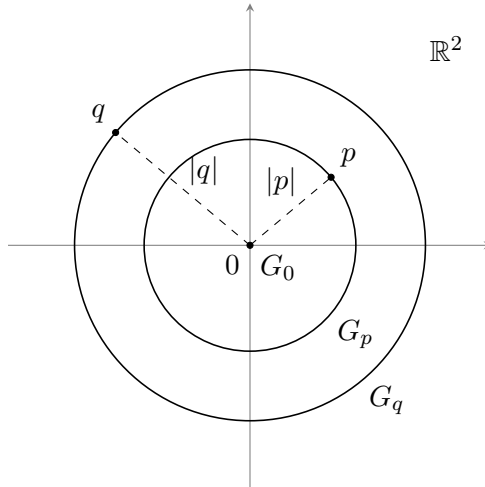
It should be intuitively clear from the definition that the orbits of two points are either disjoint or coincide. In fact, we have the following.

**Proposition 19.8.** *Let $\rhd\colon G \times M \to M$ be an action on $M$. Define a relation on $M$*

$$p \sim q \;:\Leftrightarrow\; \exists\, g \in G : q = g \rhd p.$$

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

*Proof.* Let $p, q, r \in M$. We have

i) $p \sim p$ since $p = e \rhd p$;

ii) $p \sim q \Rightarrow q \sim p$ since, if $q = g \rhd p$, then

$$p = e \rhd p = (g^{-1} \bullet g) \rhd p = g^{-1} \rhd (g \rhd p) = g^{-1} \rhd q;$$

iii) $(p \sim q \text{ and } q \sim r) \Rightarrow p \sim r$ since, if $q = g_1 \rhd p$ and $r = g_2 \rhd q$, then

$$r = g_2 \rhd (g_1 \rhd p) = (g_1 \bullet g_2) \rhd p.$$

Therefore, $\sim$ is an equivalence relation on $M$. $\qquad\square$

The equivalence classes of $\sim$ are, by definition, the orbits.

**Definition.** Let $\rhd\colon G \times M \to M$ be an action on $M$. The *orbit space* of $M$ is

$$M/G := M/\!\sim\, = \{G_p \mid p \in M\}.$$

*Example* 19.9. The orbit space of our previous $\mathrm{SO}(2, \mathbb{R})$-action on $\mathbb{R}^2$ is the partition of $\mathbb{R}^2$ into concentric circles centred at the origin, plus the origin itself.

**Definition.** Let $\rhd\colon G \times M \to M$ be a $G$-action on $M$. The *stabiliser* of $p \in M$ is

$$S_p := \{g \in G \mid g \rhd p = p\}.$$

Note that for each $p \in M$, the stabiliser $S_p$ is a subgroup of $G$.

*Example* 19.10. In our $\mathrm{SO}(2, \mathbb{R})$ example, we have $S_p = \{\mathrm{id}_{\mathbb{R}^2}\}$ for $p \neq 0$ and $S_0 = \mathrm{SO}(2, \mathbb{R})$.

**Definition.** A left $G$-action $\rhd\colon G \times M \to M$ 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 $p = g \rhd p$.

*Example* 19.11. The action $\rhd\colon G \times V \to V$ induced by a representation $R\colon G \to \mathrm{GL}(V)$ is never free since we always have $S_0 = G$.

*Example* 19.12. Consider the action $\rhd\colon \mathrm{T}(n) \times \mathbb{R}^n \to \mathbb{R}^n$ of the $n$-dimensional translation group $\mathrm{T}(n)$ on $\mathbb{R}^n$. We have, rather trivially, $\mathrm{T}(n)_p = \mathbb{R}^n$ for every $p \in \mathbb{R}^n$. It is also easy to show that this action is free and transitive.

**Proposition 19.13.** *Let* $\triangleright\colon G \times M \to M$ *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

$$
\begin{aligned}
g_1 \triangleright p = g_2 \triangleright p \quad &\Leftrightarrow \quad g_2^{-1} \triangleright (g_1 \triangleright p) = g_2^{-1} \triangleright (g_2 \triangleright p) \\
&\Leftrightarrow \quad (g_2^{-1} \bullet g_1) \triangleright p = (g_2^{-1} \bullet g_2) \triangleright p \\
&\Leftrightarrow \quad (g_2^{-1} \bullet g_1) \triangleright p = (e \triangleright p) \\
&\Leftrightarrow \quad (g_2^{-1} \bullet g_1) \triangleright p = p.
\end{aligned}
$$

Hence $g_2^{-1} \bullet g_1 \in S_p$, but since $\triangleright$ is free we have $S_p = \{e\}$, and thus $g_1 = g_2$. $\qquad\square$

**Proposition 19.14.** *If* $\triangleright\colon G \times M \to M$ *is a free action, then*

$$\forall\, p \in G:\ G_p \cong_{\mathrm{diff}} G.$$

*Example* 19.15. Define $\triangleright\colon \mathrm{SO}(2,\mathbb{R}) \times \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}^2 \setminus \{0\}$ to coincide with the action induced by the representation of $\mathrm{SO}(2,\mathbb{R}^2)$ on $\mathbb{R}^2$ for each non-zero point of $\mathbb{R}^2$. Then this action is free, since we have $S_p = \{\mathrm{id}_{\mathbb{R}^2}\}$ for $p \neq 0$, and the previous proposition implies

$$\forall\, p \in \mathbb{R}^2 \setminus \{0\}:\ \mathrm{SO}(2,\mathbb{R})_p \cong_{\mathrm{diff}} \mathrm{SO}(2,\mathbb{R}) \cong_{\mathrm{diff}} S^1.$$

## 19.2 Principal fibre bundles

This is a good time to review our earlier section on bundles. We can specialize our definition of bundle to define a *smooth bundle*, which is just a bundle $(E, \pi, M)$ where $E$ and $M$ are smooth manifolds and the projection $\pi\colon E \to M$ is smooth. Two smooth bundles $(E, \pi, M)$ and $(E', \pi', M')$ are isomorphic if there exist diffeomorphisms $u, f$ such that the following diagram commutes

$$
\begin{array}{ccc}
E & \overset{u}{\longrightarrow} & E' \\
{\scriptstyle \pi}\big\downarrow & & \big\downarrow{\scriptstyle \pi'} \\
M & \underset{f}{\longrightarrow} & M'
\end{array}
$$

**Definition.** Let $G$ be a Lie group. A smooth bundle $(E, \pi, M)$ is called a *principal $G$-bundle* if $E$ is equipped with a free right $G$-action and

$$
\begin{array}{ccc}
E & & E \\
{\scriptstyle \pi}\big\downarrow & \cong_{\mathrm{bdl}} & \big\downarrow{\scriptstyle \rho} \\
M & & E/G
\end{array}
$$

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 free, for each $p \in E$ we have

$$\mathrm{preim}_\rho(G_p) = G_p \cong_{\mathrm{diff}} G.$$

We said at beginning that, roughly speaking, a principal bundle is a bundle whose fibre at each point is a Lie group. Note that the formal definition is that a principal $G$-bundle is a bundle which is isomorphic to a bundle whose fibres are the orbits under the right action of $G$, which are themselves isomorphic to $G$ since the action is free.

*Remark* 19.16. 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 fibre of $E \xrightarrow{\pi} M$. The isomorphism in our definition enforces the fibre-wise transitivity since $G$ acts transitively on each $G_p$ by the definition of orbit.

*Example* 19.17.    a) Let $M$ be a smooth manifold. Consider the space

$$L_p M := \{(e_1, \ldots, e_{\dim M}) \mid e_1, \ldots, e_{\dim M} \text{ is a basis of } T_p M\} \cong_{\mathrm{vec}} \mathrm{GL}(\dim M, \mathbb{R}).$$

We know from linear algebra that the bases of a vector space are related to each other by invertible linear transformations. Hence, we have

$$L_p M \cong_{\mathrm{vec}} \mathrm{GL}(\dim M, \mathbb{R}).$$

We define the frame bundle of $M$ as

$$LM := \coprod_{p \in M} L_p M$$

with the obvious projection map $\pi \colon LM \to M$ sending each basis $(e_1, \ldots, e_{\dim M})$ to the unique point $p \in M$ such that $(e_1, \ldots, e_{\dim M})$ 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

$$\dim LM = \dim M + \dim T_p M = \dim M + (\dim M)^2.$$

b) We would now like to make $LM \xrightarrow{\pi} M$ into a principal $\mathrm{GL}(\dim M, \mathbb{R})$-bundle. We define a right $\mathrm{GL}(\dim M, \mathbb{R})$-action on $LM$ by

$$(e_1, \ldots, e_{\dim M}) \lhd g := (g^a{}_1 e_a, \ldots, g^a{}_{\dim M} e_a),$$

where $g^a{}_b$ are the components of the endomorphism $g \in \mathrm{GL}(\dim M, \mathbb{R})$ with respect to the standard basis on $\mathbb{R}^n$. Note that if $(e_1, \ldots, e_{\dim M}) \in L_p M$, we must also have $(e_1, \ldots, e_{\dim M}) \lhd g \in L_p M$. This action is free since

$$(e_1, \ldots, e_{\dim M}) \lhd g = (e_1, \ldots, e_{\dim M}) \Leftrightarrow (g^a{}_1 e_a, \ldots, g^a{}_{\dim M} e_a) = (e_1, \ldots, e_{\dim M})$$

and hence, by linear independence, $g^a{}_b = \delta^a_b$, so $g = \mathrm{id}_{\mathbb{R}^n}$. Note that since all bases of each $T_p M$ are related by some $g \in \mathrm{GL}(\dim M, \mathbb{R})$, $\lhd$ is also fibre-wise transitive.

c) We now have to show that

$$
\begin{array}{ccc}
LM & & LM \\
\pi \downarrow & \cong_{\text{bdl}} & \downarrow \rho \\
M & & LM \big/ \operatorname{GL}(\dim M, \mathbb{R})
\end{array}
$$

i.e. that there exist smooth maps $u$ and $f$ such that the diagram

$$
\begin{array}{ccc}
LM \underset{u^{-1}}{\overset{u}{\rightleftarrows}} & LM \\
\pi \downarrow & \downarrow \rho \\
M \underset{f^{-1}}{\overset{f}{\rightleftarrows}} & LM \big/ \operatorname{GL}(\dim M, \mathbb{R})
\end{array}
$$

commutes. We can simply choose $u = u^{-1} = \operatorname{id}_{LM}$, while we define $f$ as

$$f \colon M \to LM \big/ \operatorname{GL}(\dim M, \mathbb{R})$$

$$p \mapsto \operatorname{GL}(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})},$$

where $(e_1, \dots, e_{\dim M})$ is some basis of $T_p M$, i.e. $(e_1, \dots, e_{\dim M}) \in \operatorname{preim}_\pi(\{p\})$. Note that $f$ is well-defined since every basis of $T_p M$ gives rise to the same orbit in the orbit space $LM \big/ \operatorname{GL}(\dim M, \mathbb{R})$. Moreover, it is injective since

$$f(p) = f(p') \iff \operatorname{GL}(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} = \operatorname{GL}(\dim M, \mathbb{R})_{(e'_1, \dots, e'_{\dim M})},$$

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 $LM \big/ \operatorname{GL}(\dim M, \mathbb{R})$ 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

$$f^{-1} \colon \quad LM \big/ \operatorname{GL}(\dim M, \mathbb{R}) \to M$$

$$\operatorname{GL}(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} \mapsto \pi((e_1, \dots, e_{\dim M})).$$

Finally, we have

$$(\rho \circ \operatorname{id}_{LM})(e_1, \dots, e_{\dim M}) = \operatorname{GL}(\dim M, \mathbb{R})_{(e_1, \dots, e_{\dim M})} = (f \circ \pi)(e_1, \dots, e_{\dim M})$$

and thus $LM \xrightarrow{\pi} M$ is a principal $G$-bundle, called the *frame bundle* of $M$.

*Remark* 19.18. 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.

## 19.3 Principal bundle 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

$$
\begin{array}{ccc}
E & \xrightarrow{\ \ u\ \ } & E' \\
\downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi'} \\
M & \xrightarrow{\ \ f\ \ } & M'
\end{array}
$$

commutes, that is, $f \circ \pi = \pi' \circ u$.

**Definition.** Let $(P, \pi, M)$ and $(Q, \pi', N)$ be principal $G$-bundles. A *principal bundle morphism* from $(P, \pi, M)$ to $(Q, \pi', N)$ is a pair of smooth maps $(u, f)$ such that the diagram

$$
\begin{array}{ccc}
P & \xrightarrow{\ \ u\ \ } & Q \\
\uparrow{\scriptstyle \lhd G} & & \uparrow{\scriptstyle \blacktriangleleft G} \\
P & \xrightarrow{\ \ u\ \ } & Q \\
\downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi'} \\
M & \xrightarrow{\ \ f\ \ } & N
\end{array}
$$

commutes, that is for all $p \in P$ and $g \in G$, we have

$$
(f \circ \pi)(p) = (\pi' \circ u)(p)
$$
$$
u(p \lhd g) = u(p) \blacktriangleleft g.
$$

Note that $P \xrightarrow{\ \lhd G\ } P$ is a shorthand for the inclusion of $P$ into the product $P \times G$ followed by the right action $\lhd$, i.e.

$$
P \xrightarrow{\ \lhd G\ } P \quad = \quad P \xrightarrow{\ \ i_1\ \ } P \times G \xrightarrow{\ \lhd\ } P
$$

and similarly for $Q \xrightarrow{\ \blacktriangleleft G\ } Q$.

**Definition.** A principal bundle morphism between two principal $G$-bundles is an *isomorphism* (or *diffeomorphism*) *of principal bundles* if it is also a bundle isomorphism.

*Remark* 19.19. Note that the passage from principal bundle morphism to principal bundle isomorphism does not require any extra condition involving the Lie group $G$. We will soon see that this is because the two bundles are both principal $G$-bundles. We can further generalise the notion of principal bundle morphism as follows.

**Definition.** Let $(P, \pi, M)$ be a principal $G$-bundle, let $(Q, \pi', N)$ be a principal $H$-bundle, and let $\rho\colon G \to H$ be a Lie group homomorphism. A *principal bundle morphism* from $(P, \pi, M)$ to $(Q, \pi', N)$ is a pair of smooth maps $(u, f)$ such that the diagram

$$
\begin{array}{ccc}
P & \xrightarrow{\;\;u\;\;} & Q \\
\uparrow{\scriptstyle \triangleleft} & & \uparrow{\scriptstyle \blacktriangleleft} \\
P \times G & \xrightarrow{\;u \times \rho\;} & Q \times H \\
\uparrow{\scriptstyle i_1} & & \uparrow{\scriptstyle i_1} \\
P & \xrightarrow{\;\;u\;\;} & Q \\
\downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi'} \\
M & \xrightarrow{\;\;f\;\;} & N
\end{array}
$$

commutes, that is $f \circ \pi = \pi' \circ u$ and $u$ is a $\rho$-equivariant map

$$\forall\, p \in P : \forall\, g \in G : \; u(p \triangleleft g) = u(p) \blacktriangleleft \rho(g).$$

**Definition.** A principal bundle morphism between principal $G$-bundle and a principal $H$-bundle is an *isomorphism* (or *diffeomorphism*) *of principal bundles* if it is also a bundle isomorphism and $\rho$ is a Lie group isomorphism.

**Lemma 19.20.** *Let $(P, \pi, M)$ and $(Q, \pi', M)$ be principal $G$-bundles over the same base manifold $M$. Then, any $u\colon P \to Q$ such that $(u, \mathrm{id}_M)$ is a principal bundle morphism is necessarily a diffeomorphism.*

$$
\begin{array}{ccc}
P & \xrightarrow{\;\;u\;\;} & Q \\
\uparrow{\scriptstyle \triangleleft G} & & \uparrow{\scriptstyle \blacktriangleleft G} \\
P & \xrightarrow{\;\;u\;\;} & Q \\
 & {\scriptstyle \pi}\searrow \quad \swarrow{\scriptstyle \pi'} & \\
 & M &
\end{array}
$$

*Proof.* We already know that $u$ is smooth since $(u, \mathrm{id}_M)$ 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 fibre. As the action of $G$ on $P$ is fibre-wise transitive, there is a unique $g \in G$ such that $p_1 = p_2 \lhd g$. Then

$$u(p_1) = u(p_2 \lhd g) = u(p_2) \blacktriangleleft g = u(p_1) \blacktriangleleft g,$$

so $g \in S_{u(p_1)}$, but since $\blacktriangleleft$ is free, we have $g = e$ and thus

$$p_1 = p_2 \lhd e = p_2.$$

Therefore $u$ is injective.

ii) Let $q \in Q$. Choose some $p \in \mathrm{preim}_\pi(\pi'(q))$. Then, we have

$$\pi'(u(p)) = \pi(p) = \pi'(q)$$

so that $u(p)$ and $q$ belong to the same fibre. Hence, there is a unique $g \in G$ such that $q = u(p) \blacktriangleleft g$. We thus have

$$q = u(p) \blacktriangleleft g = u(p \lhd g)$$

and since $p \lhd g \in P$, the map $u$ is surjective.

Hence, $u$ is a diffeomorphism. $\qquad\square$

**Definition.** A principal $G$-bundle $(P, \pi, M)$ if it is called *trivial* if it is isomorphic as a principal $G$-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

$$\begin{aligned} \blacktriangleleft : (M \times G) \times G &\to M \times G \\ ((p, g), g') &\mapsto (p, g) \blacktriangleleft g' := (p, g \bullet g'). \end{aligned}$$

By the previous lemma, a principal $G$-bundle $(P, \pi, M)$ is trivial if there exists a smooth map $u : P \to M \times G$ 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 19.21.** *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 $\sigma \colon M \to P$ such that $\pi \circ \sigma = \mathrm{id}_M$.*

*Proof.* ($\Rightarrow$) Suppose that $(P, \pi, M)$ is trivial. Then there exists a diffeomorphism $u \colon P \to M \times G$ which make the following diagram commute



We can define

$$\sigma \colon M \to P$$
$$m \mapsto u^{-1}(m, e),$$

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

$$(\pi \circ \sigma)(pm = \pi(u^{-1}(m, e)) = \pi_1(m, e) = m,$$

for all $m \in M$, hence $\pi \circ \sigma = \mathrm{id}_M$ and thus $\sigma \in \Gamma(P)$.

($\Leftarrow$) Suppose that there exists a smooth section $\sigma \colon M \to P$. Let $p \in P$ and consider the point $\sigma(\pi(p)) \in P$. We have

$$\pi(\sigma(\pi(p))) = \mathrm{id}_M(\pi(p)) = \pi(p),$$

hence $\sigma(\pi(p))$ and $p$ belong to the same fibre, 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

$$\chi_\sigma \colon P \to G$$
$$p \mapsto \chi_\sigma(p)$$

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 fibres 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

$$\sigma(\pi(p)) \lhd \chi_\sigma(p \lhd g) = \sigma(\pi(p)) \lhd (\chi_\sigma(p) \bullet g)$$

and hence

$$\chi_\sigma(p \lhd g) = (\chi_\sigma(p) \bullet g).$$

We can now define the map

$$u_\sigma \colon P \to M \times G$$
$$p \mapsto (\pi(p), \chi_\sigma(p)).$$

By our previous lemma, it suffices to show that $u_\sigma$ is a principal bundle map.



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

$$
\begin{aligned}
u_\sigma(p \lhd g) &= (\pi(p \lhd g), \chi_\sigma(p \lhd g)) \\
&= (\pi(p), \chi_\sigma(p) \bullet g)) \\
&= (\pi(p), \chi_\sigma(p)) \blacktriangleleft g \\
&= u_\sigma(p) \blacktriangleleft g
\end{aligned}
$$

for all $p \in P$ and $g \in G$, so the upper square also commutes and hence $(P, \pi, M)$ is a trivial bundle. $\qquad \square$

*Example* 19.22. The existence of a section on the frame bundle $LM$ can be reduced to the existence of $(\dim M)$ non-everywhere vanishing linearly independent vector fields on $M$. Since no such vector field exists on even-dimensional spheres, $LS^{2n}$ is always non-trivial.

## 20 Associated fibre bundles

An associated fibre bundle is a fibre 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.

### 20.1 Associated fibre bundles

**Definition.** 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

$$\begin{aligned} \pi_F \colon \quad P_F &\to M \\ [p, f] &\mapsto \pi(p), \end{aligned}$$

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 $(P_F, \pi_F, M)$.

*Example* 20.1. Recall that the frame bundle $(LM, \pi, M)$ is a principal $\mathrm{GL}(d, \mathbb{R})$-bundle, where $d = \dim M$, with right $G$-action $\triangleleft \colon LM \times G \to LM$ given by

$$(e_1, \ldots, e_d) \triangleleft g := (g^a{}_1 e_a, \ldots, g^a{}_d e_a).$$

Let $F := \mathbb{R}^d$ (as a smooth manifold) and define a left action

$$\begin{aligned} \triangleright \colon \mathrm{GL}(d, \mathbb{R}) \times \mathbb{R}^d &\to \mathbb{R}^d \\ (g, x) &\mapsto g \triangleright x, \end{aligned}$$

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

$$\begin{array}{ccc} LM_{\mathbb{R}^d} & \xrightarrow{\quad u \quad} & TM \\ {\scriptstyle \pi_{\mathbb{R}^d}}\big\downarrow & & \big\downarrow{\scriptstyle \pi} \\ M & \xrightarrow{\quad \mathrm{id}_M \quad} & M \end{array}$$

where $(TM, \pi, M)$ is the tangent bundle of $M$, and $u$ is defined as

$$u\colon \quad LM_{\mathbb{R}^d} \quad \to TM$$
$$[(e_1, \ldots, e_d), x] \mapsto x^a e_a.$$

The inverse map $u^{-1}\colon TM \to LM_{\mathbb{R}^d}$ works as follows. Given any $X \in TM$, pick any basis $(e_1, \ldots, e_d)$ 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

$$u^{-1}(X) := [(e_1, \ldots, e_d), x].$$

The map $u^{-1}$ is well-defined since, while the pair $((e_1, \ldots, e_d), x) \in LM \times \mathbb{R}^d$ clearly depends on the choice of basis, the equivalence class

$$[(e_1, \ldots, e_d), x] \in LM_{\mathbb{R}^d} := (LM \times \mathbb{R}^d)/\sim_G$$

does not. It includes all pairs $((e_1, \ldots, e_d) \lhd g, g^{-1} \rhd x)$ for every $g \in \mathrm{GL}(d, \mathbb{R})$, i.e. every choice of basis together with the "right" components $x \in \mathbb{R}^d$.

*Remark* 20.2. 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 $\mathrm{GL}(d, \mathbb{R})$, the specific right action $\lhd$ 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 $TM$.

*Example* 20.3. Consider the principal $\mathrm{GL}(d, \mathbb{R})$-bundle $(LM, \pi, M)$ again, with the same right action as before. This time we define

$$F := (\mathbb{R}^d)^{\times p} \times (\mathbb{R}^{d*})^{\times q} := \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{p \text{ times}} \times \underbrace{\mathbb{R}^{d*} \times \cdots \times \mathbb{R}^{d*}}_{q \text{ times}}$$

with left $\mathrm{GL}(d, \mathbb{R})$-action $\rhd\colon \mathrm{GL}(d, \mathbb{R}) \times F \to F$ given by

$$(g \rhd f)^{a_1 \cdots a_p}{}_{b_1 \cdots b_q} := g^{a_1}{}_{\widetilde{a}_1} \cdots g^{a_p}{}_{\widetilde{a}_p} (g^{-1})^{\widetilde{b}_1}{}_{b_1} \cdots (g^{-1})^{\widetilde{b}_q}{}_{b_q} f^{\widetilde{a}_1 \cdots \widetilde{a}_p}{}_{\widetilde{b}_1 \cdots \widetilde{b}_q}.$$

Then, the associated bundle $(LM_F, \pi_F, M)$ thus constructed is isomorphic to $(T^p_q M, \pi, M)$, the $(p, q)$-tensor bundle on $M$.

Now for something new, consider the following.

**Definition.** Let $M$ be a smooth manifold and let $(LM, \pi, M)$ be its frame bundle, with right $\mathrm{GL}(d, \mathbb{R})$-action as above. Let $F := (\mathbb{R}^d)^{\times p} \times (\mathbb{R}^{d*})^{\times q}$ and define a left $\mathrm{GL}(d, \mathbb{R})$-action on $F$ by

$$(g \rhd f)^{a_1 \cdots a_p}{}_{b_1 \cdots b_q} := (\det g^{-1})^\omega \, g^{a_1}{}_{\widetilde{a}_1} \cdots g^{a_p}{}_{\widetilde{a}_p} (g^{-1})^{\widetilde{b}_1}{}_{b_1} \cdots (g^{-1})^{\widetilde{b}_q}{}_{b_q} f^{\widetilde{a}_1 \cdots \widetilde{a}_p}{}_{\widetilde{b}_1 \cdots \widetilde{b}_q},$$

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* 20.4. 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

$$(g \rhd f) = (\det g^{-1})^\omega f,$$

   which is the transformation law for a *scalar density of weight* $\omega$.

   iii) If $\mathrm{GL}(d, \mathbb{R})$ 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.

*Example* 20.5. 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 $g \in \mathrm{GL}(\dim V, K)$, 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.

## 20.2 Associated bundle morphisms

**Definition.** Let $(P_F, \pi_F, M)$ to $(Q_F, \pi'_F, N)$ be the associated bundles (with the same fibre $F$) of two principal $G$-bundles $(P, \pi, M)$ and $(Q, \pi', N)$. An *associated bundle map* between the associated bundles is a bundle map $(\widetilde{u}, v)$ between them such that for some $u$, the pair $(u, v)$ is a principal bundle map between the underlying principal $G$-bundles and

$$\widetilde{u}([p, f]) := [u(p), f].$$

Equivalently, the following diagrams both commute.



**Definition.** An associated bundle map $(\widetilde{u}, v)$ is an *associated bundle isomorphism* if $\widetilde{u}$ and $v$ are invertible and $(\widetilde{u}^{-1}, v^{-1})$ is also an associated bundle map.

*Remark* 20.6. Note that two associated $F$-fibre 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 $F$-fibre bundle $(E, \pi, M)$ is called trivial if there exists a bundle isomorphism

$$
\begin{array}{ccccc}
F & \longrightarrow & E & \xrightarrow{\ \ u\ \ } & M \times F \\
& & \ \searrow{\scriptstyle \pi} & & \swarrow{\scriptstyle \pi_1} \\
& & & M &
\end{array}
$$

while a principal $G$-bundle is called trivial if there exists a principal bundle isomorphism

$$
\begin{array}{ccc}
P & \xrightarrow{\ \ u\ \ } & M \times G \\
{\scriptstyle \lhd G}\uparrow & & \uparrow{\scriptstyle \blacktriangleleft G} \\
P & \xrightarrow{\ \ u\ \ } & M \times G \\
\ \searrow{\scriptstyle \pi} & & \swarrow{\scriptstyle \pi_1} \\
& M &
\end{array}
$$

**Definition.** An associated bundle $(P_F, \pi_F, M)$ is called *trivial* if the underlying principal $G$-bundle $(P, \pi, M)$ is trivial.

**Proposition 20.7.** *A trivial associated bundle is a trivial fibre bundle.*

Note that the converse does not hold. An associated bundle can be trivial as a fibre bundle but not as an associated bundle, i.e. the underlying principal fibre bundle need not be trivial simply because the associated bundle is trivial as a fibre bundle.

**Definition.** Let $H$ be a closed Lie subgroup of $G$. Let $(P, \pi, M)$ be a principal $H$-bundle and $(Q, \pi', M)$ a principal $G$-bundle. If there exists a principal bundle map from $(P, \pi, M)$ to $(Q, \pi', M)$, 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 $(Q, \pi', M)$, while $(Q, \pi', M)$ is called a *$G$-extension* of $(P, \pi, M)$.

**Theorem 20.8.** *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.*

*Example* 20.9.     i) The bundle $(LM/\operatorname{SO}(d), \pi, M)$ always has a section, and since $\operatorname{SO}(d)$ is a closed Lie subgroup of $\operatorname{GL}(d, \mathbb{R})$, the frame bundle can be restricted to a principal $\operatorname{SO}(d)$-bundle. This is related to the fact that any manifold can be equipped with a Riemannian metric.

   ii) The bundle $(LM/\operatorname{SO}(1, d-1), \pi, M)$ may or may not have a section. For example, the bundle $(LS^2/\operatorname{SO}(1,1), \pi, S^2)$ does not admit any section, and hence we cannot restrict $(LS^2/\operatorname{SO}(1,1), \pi, S^2)$ to a principal $\operatorname{SO}(1,1)$-bundle, even though $\operatorname{SO}(1,1)$ is a closed Lie subgroup of $\operatorname{GL}(2, \mathbb{R})$. This is related to the fact that the 2-sphere cannot be equipped with a Lorentzian metric.

## 21 Connections and connection 1-forms

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 fibres 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.

### 21.1 Connections on a principal bundle

Let $(P, \pi, M)$ be a principal $G$-bundle. Recall that every element of $A \in T_e G$ gives rise to a left invariant vector field on $G$ which we denoted by $X^A$. However, we will now reserve this notation for a vector field on $P$ instead. Given $A \in T_e G$, we define $X^A \in \Gamma(TP)$ by

$$X_p^A \colon \mathcal{C}^\infty(P) \xrightarrow{\sim} \mathbb{R}$$
$$f \mapsto [f(p \lhd \exp(tA))]'(0),$$

where the derivative is to be taken with respect to $t$. We also define the maps

$$i_p \colon T_e G \to T_p P$$
$$A \mapsto X_p^A,$$

which can be shown to be a Lie algebra homomorphism.

**Definition.** 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

$$V_p P := \ker((\pi_*)_p)$$
$$= \{X_p \in T_p P \mid (\pi_*)_p(X_p) = 0\}.$$

**Lemma 21.1.** *For all $A \in T_e G$ and $p \in P$, we have $X_p^A \in V_p P$.*

*Proof.* Since the action of $G$ simply permutes the elements within each fibre, we have

$$\pi(p) = \pi(p \lhd \exp(tA)),$$

for any $t$. Let $f \in \mathcal{C}^\infty(M)$ be arbitrary. Then

$$\begin{aligned}
(\pi_*)_p X_p^A(f) &= X_p^A(f \circ \pi) \\
&= [(f \circ \pi)(p \lhd \exp(tA))]'(0) \\
&= [f(\pi(p))]'(0) \\
&= 0,
\end{aligned}$$

since $f(\pi(p))$ is constant. Hence $X_p^A \in V_p P$. Alternatively, one can also argue that $(\pi_*)_p X_p^A$ is the tangent vector to a constant curve on $M$. $\qquad\square$

In particular, the map $i_p \colon T_e G \xrightarrow{\sim} V_p P$ is now a bijection. The idea of a connection is to make a choice of how to "connect" the individual points in "neighbouring" fibres in a principal fibre bundle.

**Definition.** Let $(P, \pi, M)$ be a principal bundle and let $p \in P$. A *horizontal subspace* at $p$ 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

$$X_p = \mathrm{hor}(X_p) + \mathrm{ver}(X_p),$$

with $\mathrm{hor}(X_p) \in H_p P$ and $\mathrm{ver}(X_p) \in V_p P$.

**Definition.** 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

$$(\lhd\, g)_* X_p \in H_{p \lhd g} P,$$

where $(\lhd\, g)_*$ is the push-forward of the map $(- \lhd g) \colon P \to P$ and it is a bijection. We can also write this condition more concisely as

$$(\lhd\, g)_*(H_p P) = H_{p \lhd g} P.$$

ii) For every smooth $X \in \Gamma(TP)$, the two summands in the unique decomposition

$$X|_p = \mathrm{hor}(X|_p) + \mathrm{ver}(X|_p)$$

at each $p \in P$, extend to smooth $\mathrm{hor}(X), \mathrm{ver}(X) \in \Gamma(TM)$.

The definition formalises the idea that the assignment of an $H_p P$ to each $p \in P$ should be "smooth" within each fibre (i) as well as between different fibres (ii).

*Remark* 21.2. For each $X_p \in T_p P$, both $\mathrm{hor}(X_p)$ and $\mathrm{ver}(X_p)$ depend on the choice of $H_p P$.

## 21.2 Connection one-forms

Technically, the choice of a horizontal subspace $H_pP$ at each $p \in P$ providing a connection is conveniently encoded in the thus induced Lie-algebra-valued one-form

$$\omega_p \colon T_pP \xrightarrow{\sim} T_eG$$
$$X_p \mapsto \omega_p(X_p) := i_p^{-1}(\mathrm{ver}(X_p))$$

**Definition.** The map $\omega \colon p \to \omega_p$ sending each $p \in P$ to the $T_eG$-valued one-form $\omega_p$ is called the *connection one-form* with respect to the connection.

*Remark* 21.3. 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

$$H_pP = \ker(\omega_p).$$

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 21.4.** *For all $p \in P$, $g \in G$ and $A \in T_eG$, we have*

$$(\lhd g)_* X_p^A = X_{p \lhd g}^{(\mathrm{Ad}_{g^{-1}})_* A}.$$

*Proof.* Let $f \in \mathcal{C}^\infty(P)$ be arbitrary. We have

$$
\begin{aligned}
(\lhd g)_* X_p^A(f) &= X_p^A(f \circ (- \lhd g)) \\
&= [f(p \lhd \exp(tA) \lhd g)]'(0) \\
&= [f(p \lhd g \lhd g^{-1} \lhd \exp(tA) \lhd g)]'(0) \\
&= [f(p \lhd g \lhd (g^{-1} \bullet \exp(tA) \bullet g))]'(0) \\
&= [f(p \lhd g \lhd \mathrm{Ad}_{g^{-1}}(\exp(tA)))]'(0) \\
&= [f(p \lhd g \lhd \exp(t(\mathrm{Ad}_{g^{-1}})_* A))]'(0) \\
&= X_{p \lhd g}^{(\mathrm{Ad}_{g^{-1}})_* A}(f),
\end{aligned}
$$

which is what we wanted. $\square$

**Theorem 21.5.** *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 $\omega_p \circ i_p = \mathrm{id}_{T_eG}$.*

$$
\begin{array}{ccc}
T_eG & \xrightarrow{\quad i_p \quad} & V_pP \\
& \searrow{\scriptstyle \mathrm{id}_{T_eG}} & \downarrow{\scriptstyle \omega_p|_{V_pP}} \\
& & T_eG
\end{array}
$$

b) $((\lhd\, g)^*\omega)|_p(X_p) = (\mathrm{Ad}_{g^{-1}})_*(\omega_p(X_p))$

$$
\begin{array}{ccc}
T_pP & \xrightarrow{\ \ \omega_p\ \ } & T_eG \\
 & \searrow{\scriptstyle ((\lhd g)^*\omega)|_p} & \downarrow{\scriptstyle (\mathrm{Ad}_{g^{-1}})_*} \\
 & & T_eG
\end{array}
$$

c) $\omega$ is a smooth one-form.

*Proof.*    a) Since $X_p^A \in V_pP$, by definition of $\omega$ we have

$$
\omega_p(X_p^A) := i_p^{-1}(\mathrm{ver}(X_p^A)) = i_p^{-1}(X_p^A) = A.
$$

b) First observe that the left hand side is linear in $X_p$. Consider the two cases

b.1) Suppose that $X_p \in V_pP$. Then $X_p = X_p^A$ for some $A \in T_eG$. Hence

$$
\begin{aligned}
((\lhd\, g)^*\omega)|_p(X_p^A) &= \omega_{p \lhd g}((\lhd\, g)_* X_p^A) \\
&= \omega_{p \lhd g}\left(X_{p \lhd g}^{(\mathrm{Ad}_{g^{-1}})_* A}\right) \\
&= (\mathrm{Ad}_{g^{-1}})_* A \\
&= (\mathrm{Ad}_{g^{-1}})_*(\omega_p(X_p^A))
\end{aligned}
$$

b.2) Suppose now that $X_p \in H_pP = \ker(\omega_p)$. Then

$$
((\lhd\, g)^*\omega)|_p(X_p) = \omega_{p \lhd g}((\lhd\, g)_* X_p) = 0
$$

since $(\lhd\, g)_* X_p \in H_{p \lhd g}P = \ker(\omega_{p \lhd g})$.

Let $X_p \in T_pP$. We have

$$
\begin{aligned}
((\lhd\, g)^*\omega)|_p(X_p) &= ((\lhd\, g)^*\omega)|_p(\mathrm{ver}(X_p) + \mathrm{hor}(X_p)) \\
&= ((\lhd\, g)^*\omega)|_p(\mathrm{ver}(X_p)) + ((\lhd\, g)^*\omega)|_p(\mathrm{hor}(X_p)) \\
&= (\mathrm{Ad}_{g^{-1}})_*(\omega_p(\mathrm{ver}(X_p))) + 0 \\
&= (\mathrm{Ad}_{g^{-1}})_*(\omega_p(\mathrm{ver}(X_p))) + (\mathrm{Ad}_{g^{-1}})_*(\omega_p(\mathrm{hor}(X_p))) \\
&= (\mathrm{Ad}_{g^{-1}})_*(\omega_p(\mathrm{ver}(X_p) + \mathrm{hor}(X_p))) \\
&= (\mathrm{Ad}_{g^{-1}})_*(\omega_p(X_p))
\end{aligned}
$$

c) We have $\omega = i^{-1} \circ \mathrm{ver}$ and both $i^{-1}$ and ver are smooth.    $\square$

## 22 Local representations of a connection on the base manifold: Yang-Mills fields

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.

### 22.1 Yang-Mills fields and local representations

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

$$\omega \colon \Gamma(TP) \xrightarrow{\sim} T_e G$$

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) $((\lhd\, g)^* \omega)(X) = (\mathrm{Ad}_{g^{-1}})_*(\omega(X))$.

If the pair $(u, f)$ is a principal bundle automorphism of $(P, \pi, M)$, i.e. if the diagram

$$
\begin{array}{ccc}
P & \xrightarrow{\ \ u\ \ } & P \\
\Big\uparrow{\scriptstyle \lhd G} & & \Big\uparrow{\scriptstyle \lhd G} \\
P & \xrightarrow{\ \ u\ \ } & P \\
\Big\downarrow{\scriptstyle \pi} & & \Big\downarrow{\scriptstyle \pi} \\
M & \xrightarrow{\ \ f\ \ } & M
\end{array}
$$

commutes, we should be able to pull a connection one-form $\omega$ on $P$ back to another connection one-form $u^*\omega$ on $P$.

$$
\begin{array}{ccc}
\Gamma(TP) & \xrightarrow{\ \ \omega\ \ } & T_e G \\
\Big\uparrow{\scriptstyle u} & \nearrow{\scriptstyle u^*\omega} & \\
\Gamma(TP) & &
\end{array}
$$

Recall that for a one-form $\omega \colon \Gamma(TN) \xrightarrow{\sim} \mathcal{C}^\infty(N)$, we defined

$$\Phi^*(\omega) \colon \Gamma(TM) \xrightarrow{\sim} \mathcal{C}^\infty(M)$$
$$X \mapsto \omega(\Phi_*(X)) \circ \phi$$

for any diffeomorphism $\phi\colon M \to N$. 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 $T_eG$, its domain is still $\Gamma(TP)$ and if $u\colon P \to P$ is a diffeomorphism of $P$, then $u_*X \in \Gamma(TP)$ and so

$$u^*\omega\colon X \mapsto (u^*\omega)(X) := \omega(u_*(X)) \circ u$$

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 $\sigma\colon U \to P$ be a local section of $P$, i.e. $\pi \circ \sigma = \mathrm{id}_U$.

$$
\begin{array}{c}
P \\
\uparrow {\scriptstyle \lhd G} \\
P \\
{\scriptstyle \pi} \downarrow \;\; \uparrow {\scriptstyle \sigma} \\
U
\end{array}
$$

**Definition.** Given a connection one-form $\omega$ on $P$, such a local section $\sigma$ induces

   i) a *Yang-Mills field* $\omega^U\colon \Gamma(TU) \xrightarrow{\sim} T_eG$ given by

$$\omega^U := \sigma^*\omega;$$

   ii) a *local trivialisation* of the principal bundle $P$, i.e. a map

$$
\begin{aligned}
h\colon U \times G &\to P \\
(m, g) &\mapsto \sigma(m) \lhd g;
\end{aligned}
$$

   iii) a *local representation* of $\omega$ on $U$ by

$$h^*\omega\colon \Gamma(T(U \times G)) \xrightarrow{\sim} T_eG.$$

Note that, at each point $(m, g) \in U \times G$, we have

$$T_{(m,g)}(U \times G) \cong_{\mathrm{Lie\,alg}} T_mU \oplus T_gG.$$

*Remark* 22.1. 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 $\omega$.

## 22.2 The Maurer-Cartan form

Th relation between the Yang-Mills field and the local representation is provided by the following result.

**Theorem 22.2.** *For all $v \in T_m U$ and $\gamma \in T_g G$, we have*

$$(h^* \omega)_{(m,g)}(v, \gamma) = (\mathrm{Ad}_{g^{-1}})_*(\omega^U(v)) + \Xi_g(\gamma),$$

*where $\Xi_g$ is the Maurer-Cartan form*

$$\Xi_g \colon T_g G \xrightarrow{\sim} T_e G$$
$$L_g^A \mapsto A.$$

*Remark* 22.3. 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 $\ell_g \colon G \to G$ is a diffeomorphism of $G$. As such, its push-forward at any point is a linear isomorphism. In particular, we have

$$((\ell_g)_*)_e \colon T_e G \xrightarrow{\sim} T_g G,$$

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* 22.4. Any chart $(U, x)$ of a smooth manifold $M$ induces a local section $\sigma \colon U \to LM$ of the frame bundle of $M$ by

$$\sigma(m) := \left( \left( \frac{\partial}{\partial x^1} \right)_m, \ldots, \left( \frac{\partial}{\partial x^{\dim M}} \right)_m \right) \in L_m M.$$

Since $\mathrm{GL}(\dim M, \mathbb{R})$ can be identified with an open subset of $\mathbb{R}^{(\dim M)^2}$, we have

$$T_e \mathrm{GL}(\dim M, \mathbb{R}) \cong_{\mathrm{Lie\,alg}} \mathbb{R}^{(\dim M)^2},$$

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

$$\omega \colon \Gamma(LM) \xrightarrow{\sim} T_e \mathrm{GL}(\dim M, \mathbb{R})$$

can thus be given in terms of $(\dim M)^2$ functions

$$\omega^i{}_j \colon \Gamma(LM) \xrightarrow{\sim} \mathbb{R}, \qquad 1 \le i, j \le \dim M.$$

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

$$(\omega^U)_m = \omega^U_\mu(m) \, (\mathrm{d}x^\mu)_m,$$

where $1 \le \mu \le \dim M$ and

$$\omega^U_\mu(m) := (\omega^U)_m\left(\left(\frac{\partial}{\partial x^\mu}\right)_m\right).$$

Since $(\omega^U)_m \colon T_m U \xrightarrow{\sim} T_e G$, we have $\omega^U_\mu \colon U \to T_e G$. Hence, by employing he same isomorphism as above, we can identify each $\omega^U_\mu(m)$ with a square $\dim M \times \dim M$ matrix and define the symbol

$$\Gamma^i{}_{j\mu}(m) := (\omega^U(m))^i{}_{j\mu} := (\omega^U_\mu(m))^i{}_j,$$

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 $\Gamma^i{}_{j\mu}(m)$ 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 **??** 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* 22.5. 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 $x^i{}_j \colon \mathrm{GL}^+(d, \mathbb{R}) \to \mathbb{R}$ denote the corresponding co-ordinate functions

$$
\begin{array}{ccc}
\mathrm{GL}^+(d, \mathbb{R}) & \xrightarrow{\quad x \quad} & x(\mathrm{GL}^+(d, \mathbb{R})) \subseteq \mathbb{R}^{d^2} \\
& \searrow{\scriptstyle x^i{}_j} & \downarrow{\scriptstyle \mathrm{proj}^i{}_j} \\
& & \mathbb{R}
\end{array}
$$

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

$$L^A \colon \mathcal{C}^\infty(\mathrm{GL}^+(d, \mathbb{R})) \xrightarrow{\sim} \mathcal{C}^\infty(\mathrm{GL}^+(d, \mathbb{R}))$$

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:

$$
\begin{aligned}
(L^A x^i{}_j)(g) &= [x^i{}_j(g \bullet \exp(tA))]'(0) \\
&= [x^i{}_j(g \bullet \mathrm{e}^{tA})]'(0) \\
&= (g^i{}_k(\mathrm{e}^{tA})^k{}_j)'(0) \\
&= g^i{}_k A^k{}_j,
\end{aligned}
$$

where we have used the fact that for a matrix Lie group, the exponential map is just the ordinary exponential

$$
\exp(A) = \mathrm{e}^A := \sum_{n=0}^{\infty} \frac{1}{n!} A^n.
$$

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 $\mathrm{GL}(d, \mathbb{R})$

$$
(\Xi_g)^i{}_j := (g^{-1})^i{}_k (\mathrm{d}x^k{}_j)_g.
$$

Indeed, we can quickly check that

$$
\begin{aligned}
(\Xi_g)^i{}_j(L^A) &= (g^{-1})^i{}_k (\mathrm{d}x^k{}_j)_g \left( g^p{}_r A^r{}_q \left( \frac{\partial}{\partial x^p{}_q} \right)_g \right) \\
&= (g^{-1})^i{}_k g^p{}_r A^r{}_q \delta^k_p \delta^q_j \\
&= (g^{-1})^i{}_p g^p{}_r A^r{}_j \\
&= \delta^i_r A^r{}_j \\
&= A^i{}_j.
\end{aligned}
$$

## 22.3   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.
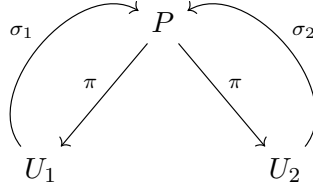


Suppose, for instance, that we have two open subsets $U_1, U_2 \subseteq M$ and consider the Yang-Mills fields associated to two local connections $\sigma_1, \sigma_2$. If $\omega^{U_1}$ and $\omega^{U_2}$ are both local versions of a unique connection one-form, then is $U_1 \cap U_2 \neq \varnothing$, the Yang-Mills fields $\omega^{U_1}$ and $\omega^{U_2}$ should satisfy some compatibility condition on $U_1 \cap U_2$.

**Definition.** Within the above set-up, the *gauge map* is the map

$$\Omega \colon U_1 \cap U_2 \to G$$

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 22.6.** *Under the above assumptions, we have*

$$(\omega^{U_2})_m = (\mathrm{Ad}_{\Omega^{-1}(m)})_*(\omega^{U_1}) + (\Omega^*\Xi_g)_m.$$

*Example* 22.7. 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 $\Xi_g \colon T_g G \to T_e G$ and $\Omega \colon U_1 \cap U_2 \to T_e G$, we have $\Omega^*\Xi_g \colon T(U_1 \cap U_2) \to T_e G$. Let $x$ be a chart map near the point $m \in U_1 \cap U_2$. We have

$$
\begin{aligned}
((\Omega^*\Xi_g)_m)^i{}_j\left(\left(\frac{\partial}{\partial x^\mu}\right)_m\right) &= (\Xi_{\Omega(m)})^i{}_j\left(\Omega_*\left(\frac{\partial}{\partial x^\mu}\right)_m\right) \\
&= (\Omega(m)^{-1})^i{}_k (\mathrm{d}\widetilde{x}^k{}_j)_{\Omega(m)}\left(\Omega_*\left(\frac{\partial}{\partial x^\mu}\right)_m\right) \\
&= (\Omega(m)^{-1})^i{}_k\left(\Omega_*\left(\frac{\partial}{\partial x^\mu}\right)_m\right)(\widetilde{x}^k{}_j) \\
&= (\Omega(m)^{-1})^i{}_k\left(\frac{\partial}{\partial x^\mu}\right)_m(\widetilde{x}^k{}_j \circ \Omega) \\
&= (\Omega(m)^{-1})^i{}_k\left(\frac{\partial}{\partial x^\mu}\right)_m(\Omega(m))^k{}_j.
\end{aligned}
$$

hence, we can write

$$
\begin{aligned}
((\Omega^*\Xi_g)_m)^i{}_j &= (\Omega(m)^{-1})^i{}_k\left(\frac{\partial}{\partial x^\mu}\right)_m(\Omega(m))^k{}_j \mathrm{d}x^\mu \\
&=: (\boldsymbol{\Omega}^{-1}\mathrm{d}\boldsymbol{\Omega})^i{}_j.
\end{aligned}
$$

Let us now compute the other summand. Recall that $\mathrm{Ad}_g$ is the map

$$
\begin{aligned}
\mathrm{Ad}_g \colon G &\to G \\
h &\mapsto g \bullet h \bullet g^{-1}
\end{aligned}
$$

and since $\mathrm{Ad}_g(e) = e$, the push-forward $((\mathrm{Ad}_g)_*)_e \colon T_e G \xrightarrow{\sim} T_e G$ is a linear endomorphism of $T_e G$. Moreover, since here $G = \mathrm{GL}(d, \mathbb{R})$ is a matrix Lie group, we have

$$((\mathrm{Ad}_g)_* A)^i{}_j = g^i{}_k A^k{}_l (g^{-1})^l{}_j =: (\boldsymbol{g}\boldsymbol{A}\boldsymbol{g}^{-1})^i{}_j.$$

Hence, we have

$$(\mathrm{Ad}_{\Omega^{-1}(m)})_*(\omega^{U_1}) = (\Omega(m)^{-1})^i{}_k(\omega^{U_1})_m)^k{}_l(\Omega(m))^l{}_j$$

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

$$\Omega^i{}_j = \frac{\partial y^i}{\partial x^j} := \partial_j (y^i \circ x^{-1}) \circ x$$

$$(\Omega^{-1})^i{}_j = \frac{\partial x^i}{\partial y^j} := \partial_j (x^i \circ y^{-1}) \circ y$$

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.

## 23 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 fibres 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.

### 23.1 Horizontal lifts to the principal bundle

**Definition.** Let $(P, \pi, M)$ be a principal $G$-bundle equipped with a connection and let $\gamma \colon [0, 1] \to M$ be a curve on $M$. The *horizontal lift of $\gamma$ through $p_0 \in P$* is the unique curve

$$\gamma^{\uparrow} \colon [0, 1] \to P$$

with $\gamma^{\uparrow}(0) = p_0 \in \mathrm{preim}_{\pi}(\{\gamma(0)\})$ satisfying

  i) $\pi \circ \gamma^{\uparrow} = \gamma$;

  ii) $\forall \, \lambda \in [0, 1]: \ \mathrm{ver}(X_{\gamma^{\uparrow}, \gamma^{\uparrow}(\lambda)}) = 0$;

  iii) $\forall \, \lambda \in [0, 1]: \ \pi_*(X_{\gamma^{\uparrow}, \gamma^{\uparrow}(\lambda)}) = X_{\gamma, \gamma(\lambda)}$.

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 fibre 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 23.1.* Note that the uniqueness in the above definition only stems from the choice of $p_0 \in \mathrm{preim}_{\pi}(\{\gamma(0)\})$. A curve on $M$ has several horizontal lifts to a curve on $P$, but there is only one such curve going through each point $p_0 \in \mathrm{preim}_{\pi}(\{\gamma(0)\})$. 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 $\gamma \colon [0, 1] \to M$ is to proceed in two steps:

  i) "Generate" the horizontal lift by starting from some arbitrary curve $\delta \colon [0, 1] \to P$ such that $\pi \circ \delta = \gamma$ by action of a suitable curve $g \colon (0, 1) \to G$ 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 $g\colon [0,1] \to P$ by a path-ordered integral over the local Yang-Mills field.

We have the following result characterising the curve $g$ appearing above.

**Theorem 23.2.** *The (first order) ODE satisfied by the curve $g\colon [0,1] \to G$ is*

$$(\mathrm{Ad}_{g(\lambda)^{-1}})_*(\omega_{\delta(\lambda)}(X_{\delta,\delta(\lambda)})) + \Xi_{g(\lambda)}(X_{g,g(\lambda)}) = 0$$

*with the initial condition $g(0) = g_0$.*

**Corollary 23.3.** *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 $\sigma\colon U \to P$ 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

$$
\begin{aligned}
\omega_{\delta(\lambda)}(X_{\delta,\delta(\lambda)}) &= \omega_{\delta(\lambda)}(\sigma_*(X_{\gamma,\gamma(\lambda)})) \\
&= (\sigma^*\omega)_{\gamma(\lambda)}(X_{\gamma,\gamma(\lambda)}) \\
&= (\omega^U)_{\gamma(\lambda)}(X_{\gamma,\gamma(\lambda)}) \\
&= \omega^U_\mu(\gamma(\lambda))(\mathrm{d}x^\mu)_{\gamma(\lambda)}\left(X^\nu_\gamma(\gamma(\lambda))\left(\frac{\partial}{\partial x^\nu}\right)_{\gamma(\lambda)}\right) \\
&= \omega^U_\mu(\gamma(\lambda))X^\nu_\gamma(\gamma(\lambda))(\mathrm{d}x^\mu)_{\gamma(\lambda)}\left(\left(\frac{\partial}{\partial x^\nu}\right)_{\gamma(\lambda)}\right) \\
&= \omega^U_\mu(\gamma(\lambda))X^\nu_\gamma(\gamma(\lambda))\,\delta^\mu_\nu \\
&= \omega^U_\mu(\gamma(\lambda))X^\mu_\gamma(\gamma(\lambda)).
\end{aligned}
$$

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 $\Gamma_\mu := \omega^U_\mu$ and $\dot{\gamma}^\mu(\lambda) := X^\mu_\gamma(\gamma(\lambda))$, together with the initial condition $g(0) = g_0$.

## 23.2 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

$$g(t) = g_0 - \int_0^t d\lambda_1 \, \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \left( g_0 - \int_0^{\lambda_1} d\lambda_2 \, \Gamma_\nu(\gamma(\lambda_2)) \dot{\gamma}^\nu(\lambda_2) \right)$$

$$= g_0 - \int_0^t d\lambda_1 \, \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) g_0 + \int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \, \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \Gamma_\nu(\gamma(\lambda_2)) \dot{\gamma}^\nu(\lambda_2) g(\lambda_2).$$

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

$$g(t) = g_0 - \int_0^t d\lambda_1 \, \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) g_0$$

$$+ \int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \, \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \Gamma_\nu(\gamma(\lambda_2)) \dot{\gamma}^\nu(\lambda_2) g_0$$

$$\vdots$$

$$+ (-1)^n \int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \cdots \int_0^{\lambda_{n-1}} d\lambda_n \, \Gamma_\mu(\gamma(\lambda_1)) \dot{\gamma}^\mu(\lambda_1) \cdots \Gamma_\nu(\gamma(\lambda_n)) \dot{\gamma}^\nu(\lambda_n) \, g(\lambda_n).$$

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
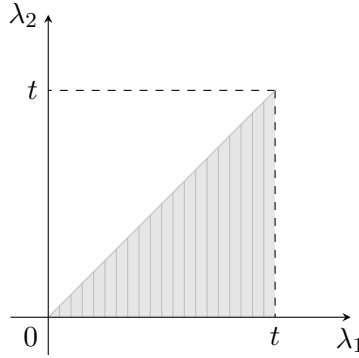


and if the integrand $f(\lambda_1, \lambda_2)$ is invariant under the exchange $\lambda_1 \leftrightarrow \lambda_2$, we have

$$\int_0^t d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \, f(\lambda_1, \lambda_2) = \frac{1}{2} \int_0^t d\lambda_1 \int_0^t d\lambda_2 \, f(\lambda_1, \lambda_2).$$

Generalising to $n$ dimensions, we have

$$\int_0^t d\lambda_1 \cdots \int_0^{\lambda_{n-1}} d\lambda_n \, f(\lambda_1, \ldots, \lambda_n) = \frac{1}{n!} \int_0^t d\lambda_1 \cdots \int_0^t d\lambda_n \, f(\lambda_1, \ldots, \lambda_n)$$

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

$$\int_0^t \mathrm{d}\lambda_1 \cdots \int_0^t \mathrm{d}\lambda_n \, f_1(\lambda_1) \cdots f_n(\lambda_n) = \left(\int_0^t \mathrm{d}\lambda_1 \, f(\lambda_1)\right) \cdots \left(\int_0^t \mathrm{d}\lambda_n \, f(\lambda_n)\right)$$

so that, in our case, we would have

$$g(t) = \left(\sum_{n=0}^\infty \frac{(-1)^n}{n!} \left(\int_0^t \mathrm{d}\lambda \, \Gamma_\mu(\gamma(\lambda))\dot\gamma^\mu(\lambda)\right)^n\right) g_0$$
$$= \exp\left(-\int_0^t \mathrm{d}\lambda \, \Gamma_\mu(\gamma(\lambda))\dot\gamma^\mu(\lambda)\right) g_0.$$

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

$$g(t) = \mathrm{P}\exp\left(-\int_0^t \mathrm{d}\lambda \, \Gamma_\mu(\gamma(\lambda))\dot\gamma^\mu(\lambda)\right) g_0,$$

where the *path-ordered exponential* $\mathrm{P}\exp$ is defined to yield the correct expression for $g(t)$.

Summarising, we have the following.

**Proposition 23.4.** *For a principal $G$-bundle $(P, \pi, M)$, where $G$ is a matrix Lie group, the horizontal lift of a curve $\gamma\colon [0,1] \to U$ through $p_p \in \mathrm{preim}_\pi(\{U\})$, where $(U, x)$ is a chart on $M$, is given in terms of a local section $\sigma\colon U \to P$ by the explicit expression*

$$\gamma^\uparrow(\lambda) = (\sigma \circ \gamma)(\lambda) \lhd \left(\mathrm{P}\exp\left(-\int_0^\lambda \mathrm{d}\widetilde\lambda \, \Gamma_\mu(\gamma(\widetilde\lambda))\dot\gamma^\mu(\widetilde\lambda)\right) g_0\right).$$

**Definition.** Let $\gamma_p^\uparrow\colon [0,1] \to P$ be the horizontal lift through $p \in \mathrm{preim}_\pi(\{\gamma(0)\})$ of the curve $\gamma\colon [0,1] \to M$. The *parallel transport map along $\gamma$* is the map

$$T_\gamma\colon \mathrm{preim}_\pi(\{\gamma(0)\}) \to \mathrm{preim}_\pi(\{\gamma(1)\})$$
$$p \mapsto \gamma_p^\uparrow(1).$$

*Remark* 23.5. The parallel transport is, in fact, a bijection between the fibres $\mathrm{preim}_\pi(\{\gamma(0)\})$ and $\mathrm{preim}_\pi(\{\gamma(1)\})$. It is injective since there is a unique horizontal lift of $\gamma$ through each point $p \in \mathrm{preim}_\pi(\{\gamma(0)\})$, and horizontal lifts through different points do not intersect. It is surjective since for each $q \in \mathrm{preim}_\pi(\{\gamma(1)\})$ we can find a $p$ such that $q = \gamma_p^\uparrow(1)$ as follows. Let $\widetilde p \in \mathrm{preim}_\pi(\{\gamma(0)\})$. Then $\gamma_{\widetilde p}^\uparrow(1)$ belongs to the same fibre as $q$ and hence there exists a unique $g \in G$ such that $q = \gamma_{\widetilde p}^\uparrow(1) \lhd g$. Recall that

$$\gamma_{\widetilde p}^\uparrow(\lambda) = (\sigma \circ \gamma)(\lambda) \lhd (\mathrm{P}\exp(\cdots)g_0)$$

where $g_0$ is the unique $g_0 \in G$ such that $\widetilde p = (\sigma \circ \gamma)(0) \lhd g_0$. Define $p \in \mathrm{preim}_\pi(\{\gamma(0)\})$ by

$$p := \widetilde p \lhd g = (\sigma \circ \gamma)(0) \lhd (g_0 \bullet g).$$

Then we have

$$\gamma_p^\uparrow(1) = (\sigma \circ \gamma)(1) \lhd (\mathrm{P}\exp(\cdots)g_0 \bullet g)$$
$$= (\sigma \circ \gamma)(1) \lhd (\mathrm{P}\exp(\cdots)g_0) \lhd g$$
$$= \gamma_{\tilde{p}}^\uparrow(1) \lhd g$$
$$= q.$$

**Loops and holonomy groups**

Consider the case of loops, i.e. curves $\gamma\colon [0,1] \to M$ for which $\gamma(0) = \gamma(1)$. Fix some $p \in \mathrm{preim}_\pi(\{\gamma(0)\})$. 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 fibre. Hence, there exists a unique $g_\gamma \in G$ such that

$$\gamma_p^\uparrow(1) = \gamma_p^\uparrow(0) \lhd g_\gamma = p \lhd g_\gamma.$$

**Definition.** Let $\omega$ be a connection one-form on the principal $G$-bundle $(P, \pi, M)$. Let $\gamma\colon [0,1] \to M$ be a loop with base-point $a \in M$, i.e. $\gamma(0) = \gamma(1) = a$. The subgroup of $G$

$$\mathrm{Hol}_a(\omega) := \{g_\gamma \mid \gamma_p^\uparrow(1) = p \lhd g_\gamma \text{ for some loop } \gamma\}$$

is called the *holonomy group* of $\omega$ on $P$ at the base-point $a$.

### 23.3 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.** 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 fibre bundle of $P$ on whose typical fibre $F$ the Lie group $G$ acts on the left by $\rhd$. Let $\gamma\colon [0,1] \to M$ be a curve on $M$ and let $\gamma_p^\uparrow$ be its horizontal lift to $P$ through $p \in \mathrm{preim}_\pi(\{\gamma(0)\})$. Then the *horizontal lift* of $\gamma$ to the associated bundle $P_F$ through the point $[p, f] \in P_F$ is the curve

$$\gamma_{[p,f]}^{\overset{P_F}{\uparrow}}\colon [0,1] \to P_F$$
$$\lambda \mapsto [\gamma_p^\uparrow(\lambda), f]$$

For instance, we have the obvious parallel transport map.

**Definition.** The *parallel transport map* on the associated bundle is given by

$$T_\gamma^{P_F}\colon \mathrm{preim}_{\pi_F}(\{\gamma(0)\}) \to \mathrm{preim}_{\pi_F}(\{\gamma(1)\})$$
$$[p, f] \mapsto \gamma_{[p,f]}^{\overset{P_F}{\uparrow}}(1).$$

*Remark 23.6.* If $F$ is a vector space and $\rhd\colon G \times F \to F$ is fibre-wise linear, i.e. for each fixed $g \in G$, the map $(g \rhd -)\colon F \to F$ is linear, then $(P_F, \pi_F, M)$ is called a *vector bundle*. The basic idea of a covariant derivative is as follows. Let $\sigma\colon U \to P_F$ 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 $\gamma \colon (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = m$ such that $X = X_{\gamma,m}$. Then for any $0 \leq t < \varepsilon$, the points $\gamma^{\overset{P_F}{\uparrow}}_{[\sigma(m),f]}(t)$ and $\sigma(\gamma(t))$ lie in the same fibre of $P_F$. But since the fibres are vector spaces, we can write the differential quotient

$$\frac{\sigma(\gamma(t)) - \gamma^{\overset{P_F}{\uparrow}}_{[\sigma(m),f]}(t)}{t},$$

where the minus sign denotes the additive inverse in the vector space $\mathrm{preim}_{\pi_F}(\{\gamma(t)\})$ 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

$$\lim_{t \to 0} \frac{\sigma(\gamma(t)) - \gamma^{\overset{P_F}{\uparrow}}_{[\sigma(m),f]}(t)}{t}$$

(of course, this makes sense as soon as we have a topology on the fibres). We will soon present a more abstract approach.

## 24 Curvature and torsion on principal bundles

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.

### 24.1 Covariant exterior derivative and curvature

**Definition.** 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, $\mathcal{C}^\infty(P)$-multilinear map) with values in some module $V$. Then then *exterior covariant derivative* of $\phi$ is

$$\mathrm{D}\phi\colon \quad \Gamma(TP)^{\times(k+1)} \to V$$
$$(X_1, \ldots, X_{k+1}) \mapsto \mathrm{d}\phi(\mathrm{hor}(X_1), \ldots, \mathrm{hor}(X_{k+1})).$$

**Definition.** 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 $P$

$$\Omega\colon \Gamma(TP) \times \Gamma(TP) \to T_e G$$

defined by

$$\Omega := \mathrm{D}\omega.$$

For calculational purposes, we would like to make this definition a bit more explicit.

**Proposition 24.1.** *Let $\omega$ be a connection one-form and $\Omega$ its curvature. Then*

$$\Omega = \mathrm{d}\omega + \omega \wedge\!\!\!\wedge \omega \qquad\qquad (\star)$$

*with the second term on the right hand side defined as*

$$(\omega \wedge\!\!\!\wedge \omega)(X, Y) := [\![\omega(X), \omega(Y)]\!]$$

*where $X, Y \in \Gamma(TP)$ and the double bracket denotes the Lie bracket on $T_e G$.*

*Remark* 24.2. 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

$$\Omega^i{}_j = \mathrm{d}\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j.$$

*Proof.* Since $\Omega$ is $\mathcal{C}^\infty$-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

$$
\begin{aligned}
\Omega(X^A, X^B) &:= D\omega(X^A, X^B) \\
&= d\omega(\mathrm{hor}(X^A), \mathrm{hor}(X^B)) \\
&= d\omega(0, 0) \\
&= 0
\end{aligned}
$$

while the right hand side is

$$
\begin{aligned}
d\omega(X^A, X^B) + (\omega \wedge \omega)(X^A, X^B) &= X^A(\omega(X^B)) - X^B(\omega(X^A)) \\
&\quad - \omega([X^A, X^B]) + [\![\omega(X^A), \omega(X^B)]\!] \\
&= X^A(B) - X^B(A) \\
&\quad - \omega(X^{[\![A,B]\!]}) + [\![A, B]\!] \\
&= -[\![A, B]\!] + [\![A, B]\!] \\
&= 0.
\end{aligned}
$$

Note that we have used the fact that the map

$$
\begin{aligned}
i\colon T_e G &\to \Gamma(TP) \\
A &\mapsto X^A
\end{aligned}
$$

is a Lie algebra homomorphism, and hence

$$
X^{[\![A,B]\!]} = i([\![A, B]\!]) = [i(A), i(B)] = [X^A, X^B],
$$

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

$$
\Omega(X, Y) := D\omega(X, Y) = d\omega(\mathrm{hor}(X), \mathrm{hor}(Y)) = d\omega(X, Y)
$$

and

$$
(\omega \wedge \omega)(X, Y) = [\![\omega(X), \omega(Y)]\!] = [\![0, 0]\!] = 0.
$$

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

$$
\Omega(X, X^A) := D\omega(X, X^A) = d\omega(\mathrm{hor}(X), \mathrm{hor}(X^A)) = d\omega(\mathrm{hor}(X), 0) = 0.
$$

while the right hand side gives

$$
\begin{aligned}
d\omega(X, X^A) + (\omega \wedge \omega)(X, X^A) &= X(\omega(X^A)) - X^A(\omega(X)) \\
&\quad - \omega([X, X^A]) + [\![\omega(X), \omega(X^A)]\!] \\
&= X(A) - X^A(0) \\
&\quad - \omega(X^{[\![A,B]\!]}) + [\![0, A]\!] \\
&= -\omega([X, X^A]) \\
&= 0,
\end{aligned}
$$

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. $\qquad\square$

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 $T_eG$-valued one-form $\omega$ on $P$. By using the notation $\Omega^1(P) \otimes T_eG$ for the collection (in fact, bundle) of all $T_eG$-valued one-forms, we have $\omega \in \Omega^1(P) \otimes T_eG$. If $\sigma \in \Gamma(TU)$ is a local section on $M$, we defined the Yang-Mills field $\omega^U \in \Omega^1(U) \otimes T_eG$ by pulling $\omega$ back along $\sigma$.

**Definition.** 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

$$\mathrm{Riem} \equiv F := \sigma^*\Omega \in \Omega^2(U) \otimes T_eG$$

is called the *Yang-Mills field strength*.

*Remark* 24.3. Observe that the equation $\Omega = \mathrm{d}\omega + \omega \wedge\!\!\!\wedge \omega$ on $P$ immediately gives

$$\begin{aligned}
\sigma^*\Omega &= \sigma^*(\mathrm{d}\omega + \omega \wedge\!\!\!\wedge \omega) \\
&= \sigma^*(\mathrm{d}\omega) + \sigma^*(\omega \wedge\!\!\!\wedge \omega) \\
&= \mathrm{d}(\sigma^*\omega) + \sigma^*\omega \wedge\!\!\!\wedge \sigma^*\omega.
\end{aligned}$$

Since Riem is a two-form, we can write

$$\mathrm{Riem}_{\mu\nu} = (\mathrm{d}\omega^U)_{\mu\nu} + \omega^U_\mu \wedge\!\!\!\wedge \omega^U_\nu.$$

In the case of a matrix Lie group, by writing $\Gamma^i{}_{j\mu} := (\omega^U)^i{}_{j\mu}$, we can further express this in components as

$$\mathrm{Riem}^i{}_{j\mu\nu} = \partial_\nu \Gamma^i{}_{j\mu} - \partial_\mu \Gamma^i{}_{j\nu} + \Gamma^i{}_{k\mu}\Gamma^k{}_{j\nu} - \Gamma^i{}_{k\nu}\Gamma^k{}_{j\mu}$$

from which we immediately observe that Riem is symmetric in the last two indices, i.e.

$$\mathrm{Riem}^i{}_{j[\mu\nu]} = 0.$$

**Theorem 24.4** (First Bianchi identity)**.** *Let $\Omega$ be the curvature two-form associated to a connection one-form $\omega$ on a principal bundle. Then*

$$\mathrm{D}\Omega = 0.$$

*Remark* 24.5. Note that since $\Omega = \mathrm{D}\omega$, Bianchi's identity can be rewritten as $\mathrm{D}^2\Omega = 0$. However, unlike the exterior derivative $d$, the covariant exterior derivative does *not* satisfy $\mathrm{D}^2 = 0$ in general.

## 24.2 Torsion

**Definition.** 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) $\forall X \in \Gamma(TP) : \ \theta(\mathrm{ver}(X)) = 0$;

(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* 24.6. Consider the frame bundle $(LM, \pi, M)$ and define

$$\theta \colon \Gamma(T(LM)) \to \mathbb{R}^{\dim M}$$
$$X \mapsto (u_{\pi(X)}^{-1} \circ \pi_*)(X)$$

where for each $e := (e_1, \ldots, e_{\dim M}) \in LM$, $u_e$ is defined as

$$u_e \colon \qquad \mathbb{R}^{\dim M} \xrightarrow{\sim} T_{\pi(e)} M$$
$$(x^1, \ldots, x^{\dim M}) \mapsto x^i e_i.$$

To describe the inverse map $u_e^{-1}$ explicitly, note that to every frame $(e_1, \ldots, e_{\dim M}) \in LM$, there exists a co-frame $(f^1, \ldots, f^{\dim M}) \in L^*M$ such that

$$u_e^{-1} \colon T_{\pi(e)} M \xrightarrow{\sim} \mathbb{R}^{\dim M}$$
$$Z \mapsto (f^1(Z), \ldots, f^{\dim M}(Z)).$$

**Definition.** 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

$$\Theta := \mathrm{D}\theta \in \Omega^2(P) \otimes V$$

is the *torsion* of $\omega$ with respect to $\theta$.

*Remark* 24.7. 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 a canonical choice of such a form on any frame bundle 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 $\omega \wedgedouble \theta$ would be meaningless. What we have, instead, is the following

$$\Theta = \mathrm{d}\theta + \omega \wedgehalf \theta,$$

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

$$\Theta^i = \mathrm{d}\theta^i + \omega^i{}_k \wedge \theta^k.$$

**Theorem 24.8** (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*

$$\mathrm{D}\Theta = \Omega \wedge\!\!\!\wedge \theta.$$

*Remark* 24.9. 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.

## 25 Covariant derivatives

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 $\sigma \colon U \to P_F$ 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 will now discuss.

### 25.1 Equivalence of local sections and equivariant functions

**Theorem 25.1.** *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 $\sigma \colon U \to P_F$ are in bijective correspondence with $G$-equivariant functions $\phi \colon \operatorname{preim}_\pi(U) \subseteq P \to F$, where the $G$-equivariance condition is*

$$\forall\, g \in G : \forall\, p \in \operatorname{preim}_\pi(U) : \ \phi(p \lhd g) = g^{-1} \rhd \phi(p).$$

*Proof.* (a) Let $\phi \operatorname{preim}_\pi(U) \to F$ be $G$-equivariant. Define

$$\sigma_\phi \colon U \to P_F$$
$$m \mapsto [p, \phi(p)]$$

where $p$ is any point in $\operatorname{preim}_\pi(\{m\})$. First, we should check that $\sigma_\phi$ is well-defined. Let $p, \widetilde{p} \in \operatorname{preim}_\pi(\{m\})$. Then, there exists a unique $g \in G$ such that $\widetilde{p} = p \lhd g$. Then, by the $G$-equivariance of $\phi$, we have

$$[\widetilde{p}, \phi(\widetilde{p})] = [p \lhd g, \phi(p \lhd g)] = [p \lhd g, g^{-1} \rhd \phi(p)] = [p, \phi(p)]$$

and hence, $\sigma_\phi$ is well-defined. Moreover, since for all $g \in G$

$$\pi_F([p, \phi(p)]) = \pi(p) = \pi(p \lhd g) = \pi_F([p \lhd g, g^{-1} \rhd \phi(p)]),$$

we have $\pi_F \circ \sigma_\phi = \operatorname{id}_U$ and thus, $\sigma_\phi$ is a local section.

(b) Let $\sigma \colon U \to P_F$ be a local section. Define

$$\phi_\sigma \colon \operatorname{preim}_\pi(U) \to F$$
$$p \mapsto i_p^{-1}(\sigma(\pi(p)))$$

where $i_p^{-1}$ is the inverse of the map

$$i_p \colon F \to \operatorname{preim}_{\pi_F}(\{\pi(p)\}) \subseteq P_F$$
$$f \mapsto [p, f].$$

Observe that, for all $g \in G$, we have

$$i_p(f) := [p, f] = [p \lhd g, g^{-1} \rhd f] =: i_{p \lhd g}(g^{-1} \rhd f).$$

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.

$$
\sigma_{\phi_\sigma} = \sigma, \qquad \phi_{\sigma_\phi} = \phi.
$$

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 $p \in \operatorname{preim}_\pi(U)$. 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, $\phi_{\sigma_\phi} = \phi$. $\qquad\square$

## 25.2  Linear actions on associated vector fibre bundles

We now specialise to the case where $F$ is a vector space, and hence we can require the left action $G \triangleright \colon F \xrightarrow{\sim} F$ to be linear.

**Proposition 25.2.** *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 $\phi \colon P \to F$ 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 25.3.** *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*

$$
\mathrm{d}\phi(X^A) + \omega(X^A) \triangleright \phi = 0.
$$

*Proof.* Since $\phi$ is $G$-equivariant, by applying the previous proposition, we have

$$\phi(p \lhd \exp(At)) = \exp(-At) \rhd \phi(p)$$

for any $p \in P$. Hence, differentiating with respect to $t$ yields

$$(\phi(p \lhd \exp(At)))'(0) = (\exp(-At) \rhd \phi(p))'(0)$$
$$d_p\phi(X^A) = -A \rhd \phi(p)$$
$$d_p\phi(X^A) = -\omega(X^A) \rhd \phi(p)$$

for all $p \in P$ and hence, the claim holds. $\square$

### 25.3 Construction of the covariant derivative

We now wish to construct a covariant derivative, i.e. an "operator" $\nabla$ such that for any local section $\sigma \colon U \subseteq M \to P_F$ and any $X \in T_m U$ with $m \in U$, we have that $\nabla_X \sigma$ is again a local section $U \to P_F$ 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 $\sigma, \tau \colon U \to P_F$, any $f \in \mathcal{C}^\infty(U)$, 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 $\phi \colon P \to F$ is

$$D\phi := d\phi \circ \mathrm{hor}\,.$$

We now have the following result.

**Proposition 25.4.** *Let $\phi \colon P \to F$ be $G$-equivariant and let $X \in T_p P$. Then*

$$D\phi(X) = d\phi(X) + \omega(X) \rhd \phi$$

*Proof.* (a) Suppose that $X$ is vertical, that is, $X = X^A$ for some $A \in T_e G$. Then,

$$D\phi(X) = d\phi(\mathrm{hor}(X)) = 0$$

and

$$d\phi(X^A) + \omega(X^A) \rhd \phi = 0$$

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

$$D\phi(X) = d\phi(X) + \omega(X) \rhd \phi. \qquad \square$$

Hence, it is clear from this proposition that $\mathrm{D}\phi(X)$, which we can also write as $\mathrm{D}_X\phi$, 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 $P \to F$ rather than a local section o $(P_F, \pi_F, M)$.

We can obtain a covariant derivative from $\mathrm{D}$ by introducing a local trivialisation on the bundle $(P, \pi, M)$. Indeed, let $s \colon U \subseteq M \to P$ be a local section. Then, we can pull back the following objects

$$\phi \colon P \to F \qquad \rightsquigarrow \qquad s^*\phi := \phi \circ s \colon U \to P_F$$
$$\omega \in \Omega^1(M) \otimes T_e G \qquad \rightsquigarrow \qquad \omega^U := s^*\omega \in \Omega^1(U) \otimes T_e G$$
$$\mathrm{D}\phi \in \Omega^1(M) \otimes F \qquad \rightsquigarrow \qquad s^*(\mathrm{D}\phi) \in \Omega^1(U) \otimes F.$$

It is, in fact, for this last object that we will be able to define the covariant derivative. Let $X \in TU$. Then

$$\begin{aligned}
(s^*\mathrm{D}\phi)(X) &= s^*(\mathrm{d}\phi + \omega \triangleright \phi)(X) \\
&= s^*(\mathrm{d}\phi)(X) + s^*(\omega \triangleright \phi)(X) \\
&= \mathrm{d}(s^*\phi)(X) + s^*(\omega)(X) \triangleright s^*\phi \\
&= \mathrm{d}\sigma(X) + \omega^U(X) \triangleright \sigma
\end{aligned}$$

where we renamed $s^*\phi =: \sigma$. In summary, we can write

$$\nabla_X \sigma = \mathrm{d}\sigma(X) + \omega^U(X) \triangleright \sigma$$

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* 25.5. 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 $F$.

# 26 Application: Quantum mechanics on curved spaces

# 27 Application: Spin structures

# 28 Application: Kinematical and dynamical symmetries

## Further readings

There are several books covering some, most, all, or much more of the course material.

- Baez, Muniain, *Gauge Fields, Knots and Gravity*, World Scientific 1994

- Choquet-Bruhat, DeWitt-Morette, *Analysis, manifolds, and physics, Part I: Basics (Revised edition)*, North-Holland 1982

- Choquet-Bruhat, DeWitt-Morette, *Analysis, manifolds, and physics, Part II (Revised and enlarged edition)*, North-Holland 2000

- Fecko, *Differential geometry and Lie groups for physicists*, Cambridge University Press 2006

- Frankel, *The Geometry of Physics: An Introduction* (Third edition), Cambridge University Press 2011

- Hamilton, *Mathematical Gauge Theory: With Applications to the Standard Model of Particle Physics*, Springer 2017
  (preliminary version http://www.mathematik.uni-muenchen.de/~hamilton/gaugetheory.php)

- Hassani, *Mathematical Physics: A Modern Introduction to Its Foundations (Second edition)*, Springer 2013

- Lam, *Topics in Contemporary Mathematical Physics (Second edition)*, World Scientific 2015

- Naber, *Topology, Geometry and Gauge Fields: Foundations (Second edition)*, Springer 2010

- Nakahara, *Geometry, Topology and Physics (Second edition)*, Institute of Physics Publishing 2003

- Rudolph, Schmidt, *Differential Geometry and Mathematical Physics: Part I. Manifolds, Lie Groups and Hamiltonian Systems*, Springer 2012

- Rudolph, Schmidt, *Differential Geometry and Mathematical Physics: Part II. Fibre Bundles, Topology and Gauge Fields*, Springer 2017

- Schutz, *Geometrical Methods of Mathematical Physics*, Cambridge University Press 1980

- Szekeres, *A Course in Modern Mathematical Physics: Groups, Hilbert Space and Differential Geometry*, Cambridge University Press 2004

- Westenholz, *Differential Forms in Mathematical Physics* (Revised edition), North Holland 1978

Topic-specific references follow.

**Logic**

- Chiswell, Wilfrid, *Mathematical Logic*, Oxford University Press 2007

- Hammack, *Book of Proof*, Virginia Commonwealth University
  http://www.people.vcu.edu/~rhammack/BookOfProof/

- Mendelson, *Introduction to Mathematical Logic*, CRC Press 2015

**Set theory**

- Adamson, *A Set Theory Workbook*, Birkhäuser 1997

- Smullyan, Fitting, *Set theory and the continuum problem*, Oxford University Press 1996

- Takeuti, Zaring, *Introduction to Axiomatic Set Theory*, Springer 1981

**Topology**

- Adamson, *A General Topology Workbook*, Birkhäuser 1995

- Kalajdzievski, *An Illustrated Introduction to Topology and Homotopy*, CRC Press 2015

- Munkres, *Topology* (Second edition), Pearson 2014

**Topological manifolds**

- Lee, *Introduction to Topological Manifolds* (Second edition), Springer 2011

**Linear algebra**

- Jänich, *Linear algebra*, Springer 1994

- Lang, *Linear Algebra* (Third edition), Springer 1987

- Shakarchi, *Solutions Manual for Lang's Linear Algebra*, Springer 1996

**Differentiable manifolds**

- Gadea, Masqué, Mykytyuk, *Analysis and Algebra on Differentiable Manifolds: A Workbook for Students and Teachers (Second Edition)*, Springer 2013

- Jänich, *Vector Analysis*, Springer 2001

- Lang, *Fundamentals of Differential Geometry*, Springer 1999

- Lee, *Introduction to Smooth Manifolds* (Second edition), Springer 2012

- Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer 1983

**Lie groups and Lie algebras**

- Das, Okubo, *Lie Groups and Lie Algebras for Physicists*, World Scientific 2014

- Erdmann, Wildon, *Introduction to Lie Algebras*, Springer 2006

- Hall, *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction (Second edition)*, Springer 2015

- Kirillov, *An Introduction to Lie Groups and Lie Algebras*, Cambridge University Press 2008

- Stillwell, *Naive Lie Theory*, Springer 2008

**Representation theory**

- Fuchs, Schweigert, *Symmetries, Lie Algebras and Representations: A Graduate Course for Physicists*, Cambridge University Press 1997

- Fulton, Harris, *Representation Theory: A First Course*, Springer 2004

- Zee, *Group Theory in a Nutshell for Physicists*, Princeton University Press 2016

**Principal bundles**

- Sontz, *Principal Bundles: The Classical Case*, Springer 2015

- Sontz, *Principal Bundles: The Quantum Case*, Springer 2015

# Alphabetical Index