Mathematical Notes

Contents

Part I

Basic Mathematics

Chapter 1

Axiomatic Set Theory

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

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

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

1.1 Propositional Logic

Definition 1.1 (Proposition)**.** *A* ***proposition*** *p is a variable*[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 1.2 (Tautology)**.** *A proposition which is always true is called a* ***tautology****.*

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

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

p	$\neg p$	$\mathrm{id}(p)$	$\top p$	$\bot p$
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.

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

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

p	q	$p \wedge q$	$p \vee q$	$p \veebar q$
F	F	F	F	F
F	T	F	T	T
T	F	F	T	T
T	T	T	T	F

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

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

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

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

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

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

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

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

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

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

Remark 1.2. All higher order operators $\heartsuit(p_1, \ldots, p_N)$ 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 1.4 (Predicate)**.** *A **predicate** is a proposition-valued function of some variable or variables.*

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

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

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

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

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

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

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

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

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

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

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

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

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

The following result is an immediate consequence of these definitions.

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

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

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

> *Example* 1.1.
> Let $P(x, y)$ be a predicate. Then, for fixed y, $P(x, y)$ is a predicate of one variable and we define:
>
> $$Q(y) :\Leftrightarrow \forall\, x : P(x, y)$$
>
> Hence we may have the following:
>
> $$\exists\, y : \forall\, x : P(x, y) :\Leftrightarrow \exists\, y : Q(y)$$
>
> Other combinations of quantifiers are defined analogously.

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

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

are not necessarily equivalent.

Example 1.2.
Consider the proposition expressing the existence of additive inverses in the real numbers. We have:

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

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

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

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

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

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

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

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

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

1.3 Axiomatic Systems & Theory Of Proofs

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

Definition 1.10 (Proof)**.** *A* ***proof*** *of a proposition p within an axiomatic system $a_1, a_2, \ldots, a_N$ 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 1.5. 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 1.6. This definition of proof allows to easily recognise a proof. A computer could easily check that whether or not the conditions (A), (T) and (M) are satisfied by a sequence of propositions. To actually find a proof of a proposition is a whole different story.

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

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

Definition 1.11 (Consistent)**.** *An axiomatic system $a_1, a_2, \ldots, a_N$ 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 1.2. *Propositional logic is consistent.*

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

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

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

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

1.4 The $\in$-relation

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

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

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

- $x \notin y :\Leftrightarrow \neg(x \in y)$

- $x \subseteq y :\Leftrightarrow \forall\, a : (a \in x \Rightarrow a \in y)$
- $x = y :\Leftrightarrow (x \subseteq y) \wedge (y \subseteq x)$
- $x \subset y :\Leftrightarrow (x \subseteq y) \wedge \neg(x = y)$

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

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

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

1.5 Zermelo-Fraenkel Axioms Of Set Theory

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

$$\forall\, x : \forall\, y : (x \in y) \veebar \neg(x \in y)$$

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

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

Example 1.3.
This is the so called "Russel's Paraox". 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 1.10. The fact that u as defined above is not a set means that expressions like:

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

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

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

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

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

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

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

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

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

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

and hence $x \subseteq x'$.

Conversely, by the same argument, we have:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 1.4.
Let a, b be sets. Then $\{a\}$ and $\{b\}$ are sets by the pair set axiom, and hence $x := \{\{a\}, \{b\}\}$ is a set, again by the pair set axiom. Then the expression:

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

is a set by the union axiom.

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

Example 1.5.
Let a, b, c be sets. Then $\{a\}$ and $\{b, c\}$ are sets by the pair set axiom, and hence $x := \{\{a\}, \{b, c\}\}$ is a set, again by the pair set axiom. Then the expression:

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

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

The previous example easily generalises to a definition.

Definition 1.12 (Union Of Sets)**.** *Let $a_1, a_2, \ldots, a_N$ 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 1.13. The fact that the x that appears in $\bigcup x$ has to be a set is a crucial restriction. Informally, we can say that it is only possible to take unions of as many sets as would fit into a set. The "collection" of all the sets that do not contain themselves is not a set or, we could say, does not fit into a set. Therefore it is not possible to take the union of all the sets that do not contain themselves. This is very subtle, but also very precise.

Axiom of replacement. *Let R be a functional relation and let m be a set. Then the image of m under R, denoted by $\mathrm{im}_R(m)$, 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 1.13 (Functional Relation)**.** *A relation R is said to be **functional** if:*

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

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

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

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

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

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

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

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

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

and:

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

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

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

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

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

Proof.
We have two cases:

1. If $\neg(\exists\, y \in m : P(y))$, then we define: $\{y \in m \mid P(y)\} := \varnothing$.
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 $\{y \in m \mid P(y)\} := \mathrm{im}_R(m)$. □

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

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

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

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

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

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

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

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

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

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

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

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

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

To define power sets in this fashion, we would need to know, a priori, from which "bigger" set the elements of the power set come from. However, this 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 1.6.
Let $m = \{a, b\}$. Then $\mathcal{P}(m) = \{\varnothing, \{a\}, \{b\}, \{a, b\}\}$.

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

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

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

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

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

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 1.2. *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 1.18. At this point, one might suspect that we would need an extra axiom for the existence of the real numbers. But, in fact, we can define $\mathbb{R} := \mathcal{P}(\mathbb{N})$, which is a set by the axiom on power sets.

Remark 1.19. The version of the axiom of infinity 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 1.20. The axiom of choice is independent of the other 8 axioms, which means that one could have set theory with or without the axiom of choice. However, standard mathematics uses the axiom of choice and hence so will we. There is a number of theorems that can only be proved by using the axiom of choice. Amongst these we have:

- every vector space has a basis;
- there exists a complete system of representatives of an equivalence relation.

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

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

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.

1.6 Maps Between Sets

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

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

Definition 1.17 (Map)**.** *Let A, B be sets. A **map** $\phi\colon A \to B$ 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:

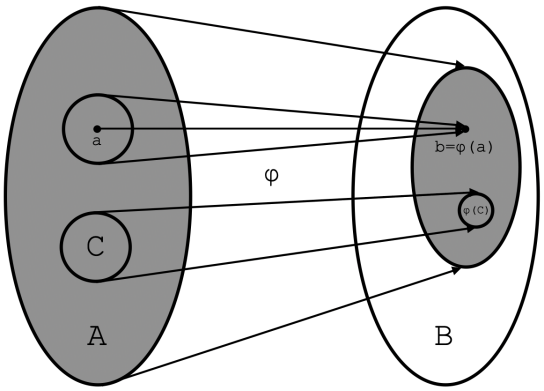
$$\begin{aligned} \phi\colon A &\to B \\ a &\mapsto \phi(a) \end{aligned}$$

which is technically an abuse of notation since ϕ, 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 ϕ is actually a relation.

Example 1.7.
Let M be a set. The simplest example of a map is the *identity map* on M:

$$\begin{aligned} \mathrm{id}_M\colon M &\to M \\ m &\mapsto m \end{aligned}$$

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



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

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

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

Remark 1.21. 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.

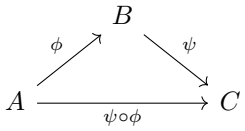
Definition 1.19 (Infinite/Finite Sets)**.** *A set A is called:*

- Infinite *if there exists a proper subset $B \subset A$ such that $B \cong_{\text{set}} A$. In particular, if A is infinite, we further define A to be:*
  - Countably *infinite if $A \cong_{\text{set}} \mathbb{N}$.*
  - Uncountably *infinite otherwise.*
- Finite *if it is not infinite. In this case, we have $A \cong_{\text{set}} \{1, 2, \ldots, N\}$ 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 ϕ and ψ, denoted by $\psi \circ \phi$ (read "psi after phi"), defined by:

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

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 1.2. *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:

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

and:

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

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

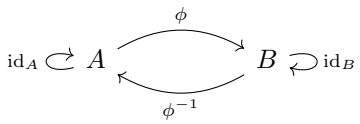
The operation of composition is necessary in order to defined inverses of maps.

Definition 1.20 (Inverse)**.** *Let $\phi\colon A \to B$ be a bijection. Then the **inverse** of ϕ, denoted ϕ^{-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:



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

Definition 1.21 (Pre-image)**.** *Let $\phi\colon A \to B$ 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 ϕ.*

Proposition 1.3. *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 $\mathrm{preim}_\phi(\bigcap C) = \bigcap_{j\in J} \mathrm{preim}_\phi(C_j)$. □

1.7 Equivalence Relations

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

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

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

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 1.8.
Consider the following wordy examples.

a) $p \sim q :\Leftrightarrow p$ is of the same opinion as q. This relation is reflexive, symmetric and transitive. Hence, it is an equivalence relation.

b) $p \sim q :\Leftrightarrow p$ is a sibling of q. This relation is symmetric and transitive but not reflexive and hence, it is not an equivalence relation.

c) $p \sim q :\Leftrightarrow p$ is taller q. This relation is transitive, but neither reflexive nor symmetric and hence, it is not an equivalence relation.

d) $p \sim q :\Leftrightarrow p$ is in love with q. This relation is generally not reflexive. People don't like themselves very much. It is certainly not normally symmetric, which is the basis of much drama in literature. It is also not transitive, except in some French films.

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

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

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

The following are two key properties of equivalence classes.

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

i) $a \in [m] \Rightarrow [a] = [m]$.

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

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 $[m] = [n]$. □

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

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

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

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

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

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

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

Example 1.9.
Let $M = \mathbb{Z}$ and define $\sim$ by:

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

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

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

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 1.10.
As a counterexample, with the same set-up as in the previous example, let us define an operation $\star$ by:

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

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

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

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

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

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

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

$$\begin{aligned} S\colon \mathbb{N} &\to \mathbb{N} \\ n &\mapsto \{n\} \end{aligned}$$

Example 1.11.
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 1.26 (Predecessor Map)**.** *The* ***predecessor map*** P *on* $\mathbb{N}^*$ *is defined by:*

$$\begin{aligned} P\colon \mathbb{N}^* &\to \mathbb{N} \\ n &\mapsto m \quad \textit{such that} \quad m \in n \end{aligned}$$

Example 1.12.
We have $P(2) = P(\{\{\varnothing\}\}) = \{\varnothing\} = 1$.

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

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

We are now ready to define addition.

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

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

Example 1.13.
We have:

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

and:

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

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

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

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

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

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

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

i) $(m, n) \sim (m, n)$ since $m + n = m + n$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Definition 1.31 (Addition Of Integers)**.** *We define the* addition of integers $+_{\mathbb{Z}} \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ *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 1.14.
$2 +_{\mathbb{Z}} (-3) := [(2,0)] +_{\mathbb{Z}} [(0,3)] = [(2,3)] = [(0,1)] =: -1$. 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 1.15.
We have $(2,3) \sim (4,6)$ since $2 \times 6 = 12 = 3 \times 4$.

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

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

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

Chapter 2

Algebraic Structures

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

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

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

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

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

2.1 Groups

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

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

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

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

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

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

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

2.2 Fields

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

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

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

2.3 Vector Spaces

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

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

are maps satisfying the following axioms:

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

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

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

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

2.3.1 Linear Maps

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

Definition 2.6 (Linear Maps)**.** *Let $(V, \oplus, \odot)$, $(W, \boxplus, \boxdot)$ 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**, or a **homomorphism** and we denote it as $f\colon V \xrightarrow{\sim} W$, if for all $v_1, v_2 \in V$ and all $\lambda \in K$:*

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

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

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

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

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

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

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

$$\begin{aligned} \oplus\colon \text{Hom}(V, W) \times \text{Hom}(V, W) &\to \text{Hom}(V, W) \\ (f, g) &\mapsto f \oplus g \end{aligned}$$

where:

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

and:

$$\begin{aligned} \odot\colon K \times \text{Hom}(V, W) &\to \text{Hom}(V, W) \\ (\lambda, f) &\mapsto \lambda \odot f \end{aligned}$$

where:

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

It is easy to check that both $f \oplus g$ and $\lambda \odot f$ 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.

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

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

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

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

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

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

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

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

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

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

where K is considered as a vector space over itself.

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

Definition 2.15 (Bilinear Maps)**.** *Let V, W, Z be vector spaces over K. A map $f\colon V \times W \to Z$ 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 2.3. 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* 2.1.
> 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 2.16 (Tensors)**.** *Let V be a vector space over K. A (p,q)-**tensor** T on V is a multilinear map:*

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

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

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

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

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

Remark 2.5. Note that to define $T_q^p V$ as a set, we should be careful and invoke the principle of restricted comprehension, i.e. we should say where the Ts 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:

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

and

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

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

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

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

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

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

Some examples are in order.

Example 2.2.

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

b) $T^1_1 V \equiv V \otimes V^* := \{T \mid T \text{ is a bilinear map } V^* \times V \to K\}$. We claim that this is the same as $\mathrm{End}(V^*)$. Indeed, given $T \in V \otimes V^*$, we can construct $\widehat{T} \in \mathrm{End}(V^*)$ as follows:

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

where, for any fixed ω, we have:

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

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

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

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

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

c) $T^0_1 V \overset{?}{\cong}_{\mathrm{vec}} V$: while you will find this stated as true in some physics textbooks, it is in fact *not true* in general.

d) $T^1_1 V \overset{?}{\cong}_{\mathrm{vec}} \mathrm{End}(V)$: This is also not true in general.

e) $(V^*)^* \overset{?}{\cong}_{\mathrm{vec}} V$: This only holds if V is finite-dimensional (we will define the dimensions of a vector space in the next section).

2.3.2 Basis Of Vector Spaces

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

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

$$\mathrm{span}_K(\mathcal{S}) := \left\{ \sum_{i=1}^{n} \lambda^i s_i \;\middle|\; \lambda^i \in K, s_i \in S, n \geq 1 \right\}$$

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

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

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

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

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

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

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

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

The following characterisation of a Hamel basis is often useful.

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

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

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

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

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

- *if $\operatorname{span}_K(S) = V$ 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 2.1. *If $\dim V < \infty$, then $(V^*)^* \cong_{\text{vec}} V$.*

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

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

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

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

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

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

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

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

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

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

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

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

where the e_{a_i}s are understood as elements of $T^1_0 V \cong_{\text{vec}} V$ and the $\epsilon^{b_i}s$ as elements of $T^0_1 V \cong_{\text{vec}} V^*$. 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 $T^p_q V$.

Notational Conventions

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

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

instead of:

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

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

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

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

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

Example 2.3.
Using the summation convention, we have:

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

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

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

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

Remark 2.11. The Einstein's summation convention should only be used when dealing with linear spaces and multilinear maps. The reason for this is the following. Consider a map $\phi\colon V \times W \to Z$, 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 ϕ is bilinear! Hence the summation convention should not be used (at least, not without extra care) in other areas of mathematics.

Matrix Representation

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

$$v = v^i e_i \quad \rightsquigarrow \quad v \mathrel{\hat{=}} \begin{pmatrix} v^1 \\ \vdots \\ v^d \end{pmatrix}$$

and

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

Given $\phi \in \mathrm{End}(V) \cong_{\mathrm{vec}} T^1_1 V$, recall that we can write $\phi = \phi^i{}_j \, e_i \otimes \epsilon^j$, where $\phi^i{}_j := \phi(\epsilon^i, e_j)$ are the components of ϕ with respect to the chosen basis. It is then also very tempting to think of ϕ as a square array of numbers:

$$\phi = \phi^i{}_j \, e_i \otimes \epsilon^j \quad \rightsquigarrow \quad \phi \mathrel{\hat{=}} \begin{pmatrix} \phi^1{}_1 & \phi^1{}_2 & \cdots & \phi^1{}_d \\ \phi^2{}_1 & \phi^2{}_2 & \cdots & \phi^2{}_d \\ \vdots & \vdots & \ddots & \vdots \\ \phi^d{}_1 & \phi^d{}_2 & \cdots & \phi^d{}_d \end{pmatrix}$$

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

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

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

Example 2.4.
If $\dim V < \infty$, then we have $\mathrm{End}(V) \cong_{\mathrm{vec}} T^1_1 V$. 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 $\phi^a{}_b := \epsilon^a(\phi(e_b))$.

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

$$\phi(v, w) = w_a \, \phi^a{}_b \, v^b \quad \rightsquigarrow \quad \omega^T \phi v$$

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

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

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

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

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

2.3.3 Change Of Basis

Let V be a vector space over K with $d = \dim V < \infty$ and let $\{e_1, \ldots, e_d\}$ 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:

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

for some $A^b{}_a \in K$. Similarly, we have:

$$e_a = \sum_{m=1}^{d} B^m{}_a \widetilde{e}_m = B^m{}_a \widetilde{e}_m$$

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

- ϕ 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". Let's do that!

2.3.4 Determinants

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

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

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

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

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

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

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

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

This terminology is due to the next definition.

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

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

where ω *is the (chosen) top form.*

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

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

We are finally ready to define the determinant.

Definition 2.28 (Determinant)**.** *Let* V *be a* d*-dimensional vector space and let* $\phi \in \mathrm{End}(V) \cong_{\mathrm{vec}} T^1_1 V$*. The* determinant *of* ϕ *is:*

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

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

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

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

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

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

Of course, under the identification of ϕ as a matrix, this definition coincides with the usual definition of determinant, and all your favourite results about determinants can be derived from it. However once we switch to "matrix representation" as we said one is not able to distinguish between an endomorphism $\phi \in T^1_1 V$ and the so called "bilinear form" $g \in T^0_2 V$, hence one might think that they can calculate the determinant of the second guy. Let's see why such a determinant is not well defined.

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

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

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

Recall that the transformation rule for a bilinear form g under a change of basis is $g \to A^T g A$. 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. We will see all of that in later chapters.

2.4 Rings

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

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

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

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

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

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

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

> *Example* 2.5.
> The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all rings under the usual operations. They are also all fields, except $\mathbb{Z}$.

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

2.5 Modules

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

$$\oplus \colon M \times M \to M$$
$$\odot \colon R \times M \to M$$

are maps satisfying the following axioms:

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

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

Example 2.6.
Any ring R is trivially a module over itself, in the sense that every field K is a vector space over itself.

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

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

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

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

- ***Finitely generated** if it has a finite generating set.*
- ***Free** 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 2.7.
Clearly, every free module is also projective.

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

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

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

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

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

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

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

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

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

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

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.

2.5.1 Basis Of Modules

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

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

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

2.6 Algebras

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

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

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

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

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

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

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

for all $v, w \in A$.

Chapter 3

Lie Algebras

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

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

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

3.1 Basic Definitions

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

i) Bilinearity: $A \times A \to A$*:* $[av + w, z] = a[v, w] + [v, z]$

ii) Antisymmetry: $\forall\, v \in A : [v, v] = 0$

iii) The Jacobi identity: $\forall\, v, w, z \in A : [v, [w, z]] + [w, [z, v]] + [z, [v, w]] = 0$

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

Some remarks are in order.

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

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

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

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

Now let's give some examples of Lie algebras.

Example 3.1.
The usual cross product between vectors $u \times w$ in $\mathbb{R}^3$ can be shown that satisfies all the requirements for a Lie bracket, hence the vector space $\mathbb{R}^3$ equipped with the cross product $[u, w] = u \times w$ is actually a Lie algebra.

Example 3.2.
Let V be a vector space. Recall that we defined the set $\mathrm{End}(V)$ as the set of all endomorphisms of V, i.e the set of all linear maps that send V back to itself. Now we define the following Lie bracket:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.2 The Adjoint Map & The Killing Form

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

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

The linearity of ad_x follows from the linearity of the bracket in the second argument, while the linearity in the first argument of the bracket implies that the map:

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

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

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

Then, we have the following.

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

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

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

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

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

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

which means that ad_x is an element of $\mathrm{End}(L)$ hence an endomorphism of L. Recall that for any vector space V: $\mathrm{End}(V) \cong_{\mathrm{vec}} T^1_1 V$ which means that 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 $\phi^a{}_b := \epsilon^a(\phi(e_b))$.

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

$$\begin{aligned}
(\mathrm{ad}_{e_i})^k{}_j &:= \varepsilon^k(\mathrm{ad}_{e_i}(e_j))\\
&= \varepsilon^k([e_i, e_j])\\
&= \varepsilon^k(C^m{}_{ij} e_m)\\
&= C^m{}_{ij} \varepsilon^k(e_m)\\
&= C^k{}_{ij}
\end{aligned}$$

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

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

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

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

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 κ is symmetric, i.e:

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

An important property of κ is its associativity with respect to the bracket.

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

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

Proof.
This follows easily from the fact that ad is a homomorphism.

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

where we used the cyclicity of the trace. □

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

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

$$\begin{aligned}
\kappa_{ij} &:= \kappa(e_i, e_j) \\
&= \mathrm{tr}(\mathrm{ad}_{e_i} \circ \mathrm{ad}_{e_j}) \\
&= (\mathrm{ad}_{e_i} \circ \mathrm{ad}_{e_j})^k{}_k \\
&= (\mathrm{ad}_{e_i})^m{}_k (\mathrm{ad}_{e_j})^k{}_m \\
&= C^m{}_{ik} C^k{}_{jm}
\end{aligned}$$

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

We can use κ to give a further equivalent characterisation of semi-simplicity.

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

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

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

Definition 3.21. *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 ϕ is said to be anti-symmetric *with respect to B.*

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

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

3.3 The Fundamental Roots & The Weyl Group

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

Definition 3.22 (Cartan Subalgebra)**.** *Let L be a d-dimensional Lie algebra. A **Cartan subalgebra** H of L is a maximal Lie subalgebra of L with the following property: there exists a basis $\{h_1, \dots, h_r\}$ of H which can be extended to a basis $\{h_1, \dots, h_r, e_1, \dots, e_{d-r}\}$ of L such that $e_1, \dots, e_{d-r}$ are eigenvectors of* $\mathrm{ad}(h)$ *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 $\{h_1, \dots, h_r, e_1, \dots, e_{d-r}\}$ is known as a *Cartan-Weyl basis* of L. Of course, we would like to know when we can find such a subalgebra.

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

i) L possesses a Cartan subalgebra.

ii) all Cartan subalgebras of L have the same dimension, called the rank *of L.*

iii) any of Cartan subalgebra H of L is abelian, i.e $[H, H] = 0$.

Note that we can think of the λ_α 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 λ_α is a $\mathbb{C}$-linear map $\lambda_\alpha \colon H \xrightarrow{\sim} \mathbb{C}$, and thus $\lambda_\alpha \in H^*$.

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

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

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

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

One can show that if λ_α 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 κ is that:

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

Hence Φ is not a linearly independent subset of H^*.

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

a) Π is a linearly independent subset of H^*.

b) For each $\lambda \in \Phi$, there exist $n_1, \dots, n_f \in \mathbb{N}$ and $\varepsilon \in \{+1, -1\}$ such that:

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

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

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

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

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

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

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

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

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

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

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

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

Proposition 3.5. *The restriction of κ to H is a pseudo inner product on H.*

Proof.
Bilinearity and symmetry are automatically satisfied. It remains to show that κ 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_α.

ii) Let $h \in H \subseteq L$. Since κ 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. □

We can now define:

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

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

Remark 3.7. If $\{h_i\}$ is a basis of H, the components of κ^* 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 $H^*_{\mathbb{R}} := \mathrm{span}_{\mathbb{R}}(\Pi)$. Note that we have the following chain of inclusions:

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

The restriction of κ^* to $H^*_{\mathbb{R}}$ leads to a surprising result.

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

ii) $\kappa^*\colon H^*_{\mathbb{R}} \times H^*_{\mathbb{R}} \to \mathbb{R}$ 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 κ^* gets strengthened to positive definiteness. Now that we have a proper real inner product, we can define some familiar notions from basic linear algebra, such as lengths and angles.

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

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

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

We need one final ingredient for our classification result.

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

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

where

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

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

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

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

Theorem 3.5. *i) The Weyl group W is generated by the fundamental roots in Π, 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:

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

3.4 Dynkin Diagrams & The Cartan Classification

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

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

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

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

But it is already written in such from since:

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

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

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

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

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

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

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

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

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

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

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

Note that we have:

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

where φ is the angle between π_i and π_j.

For $i \neq j$, the angle φ 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 3.6. *The roots of a simple Lie algebra have, at most, two distinct lengths.*

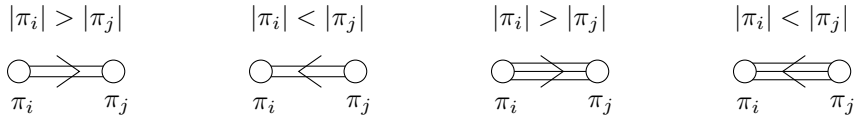
The redundancy of the Cartan matrices highlighted above is nicely taken care of by considering Dynkin diagrams.

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

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

2. *Draw n_{ij} lines between the circles representing the roots π_i and π_j;*

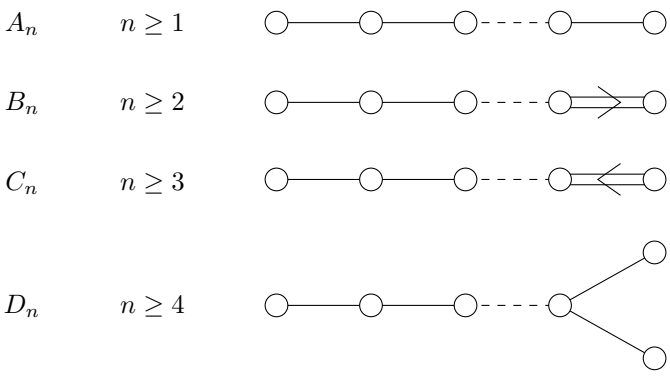
3. *If $n_{ij} = 2$ or 3, draw an arrow on the lines from the longer root to the shorter root.*



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

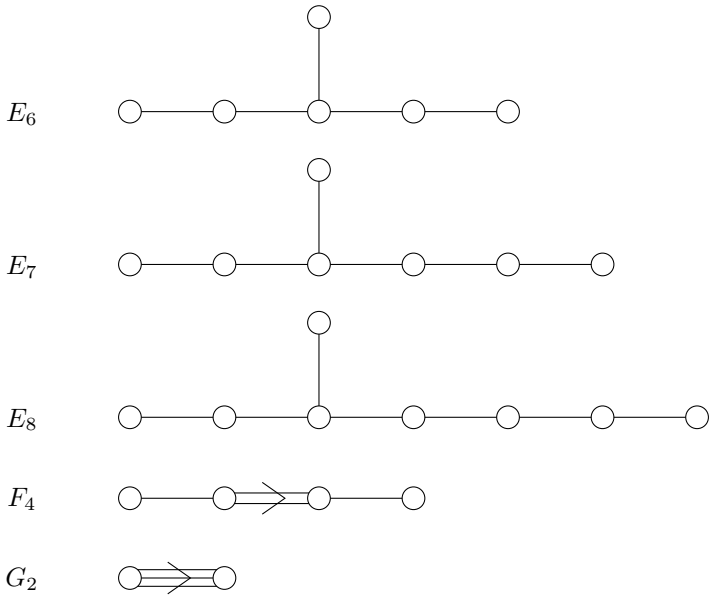
Theorem 3.7 (Killing, Cartan)**.** *Any simple finite-dimensional complex Lie algebra can be reconstructed from its set of fundamental roots Π, which only come in the following forms:*

i) There are 4 infinite families:



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

ii) Five exceptional cases:



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

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

3.5 Application: Reconstruction Of A_2 From Its Dynkin Diagram

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

We will start from the A_2 Dynkin diagram:

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

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

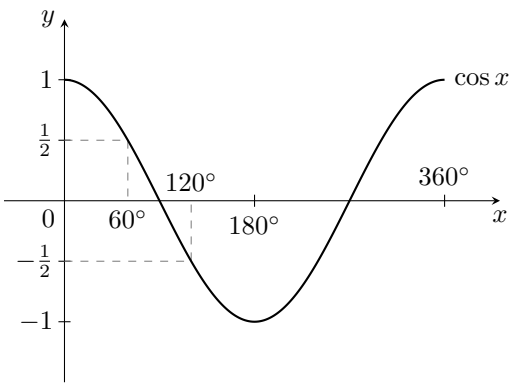
and since the off-diagonal entries of the Cartan matrix are non-positive integers, the only possibility is $C_{12} = C_{21} = -1$, so that we have:

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

To determine the angle φ between π_1 and π_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 $\varphi = 120°$.



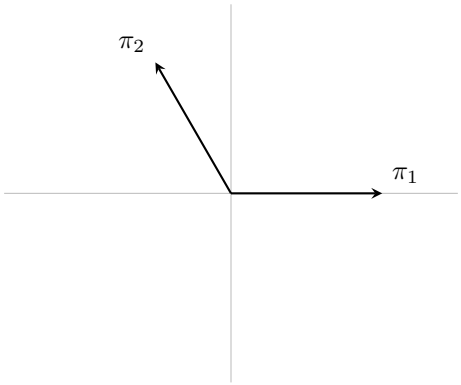
By definition, we have:

$$\cos\varphi = \frac{\kappa^*(\pi_1, \pi_2)}{|\pi_1|\,|\pi_2|}$$

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:



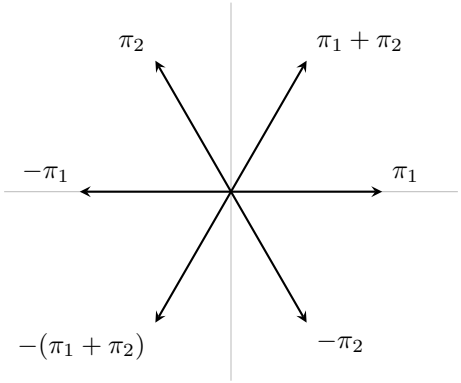
We can determine all the other roots in Φ by repeated action of the Weyl group. For instance, we easily find that $s_{\pi_1}(\pi_1) = -\pi_1$ and $s_{\pi_2}(\pi_2) = -\pi_2$. We also have:

$$s_{\pi_1}(\pi_2) = \pi_2 - 2\frac{\kappa^*(\pi_1, \pi_2)}{\kappa^*(\pi_1, \pi_1)}\pi_1 = \pi_2 - 2(-\tfrac{1}{2})\pi_1 = \pi_1 + \pi_2$$

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 $H^* = \text{span}_{\mathbb{C}}(\Pi)$, we have $\dim H^* = 2$, thus the dimension of the Cartan subalgebra is also 2. Since $|\Phi| = 6$, we know that any Cartan-Weyl basis of the Lie algebra A_2 must have $2+6 = 8$ elements. Hence, the dimension of A_2 is 8.

To complete our reconstruction of A_2, we would now like to understand how its bracket behaves. This amounts to finding its structure constants. Note that since $\dim A_2 = 8$, the structure constants $C^k{}_{ij}$ consist of $8^3 = 512$ complex numbers (not all unrelated, of course).

Denote by $\{h_1, h_2, e_3, \ldots, e_8\}$ a Cartan-Weyl basis of A_2, so that $H = \text{span}_{\mathbb{C}}(\{h_1, h_2\})$ and the e_α are eigenvectors of every $h \in H$. Since A_2 is simple, H is abelian and hence:

$$[h_1, h_2] = 0 \quad \Rightarrow \quad C^k{}_{12} = C^k{}_{21} = 0, \quad \forall\, 1 \leq k \leq 8$$

To each e_α, for $3 \leq \alpha \leq 8$, there is an associated $\lambda_\alpha \in \Phi$ such that:

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

In particular, for the basis elements h_1, h_2:

$$[h_1, e_\alpha] = \text{ad}(h_1)e_\alpha = \lambda_\alpha(h_1)e_\alpha$$
$$[h_2, e_\alpha] = \text{ad}(h_2)e_\alpha = \lambda_\alpha(h_2)e_\alpha$$

so that we have:

$$C^1{}_{1\alpha} = C^2{}_{1\alpha} = 0, \quad C^\alpha{}_{1\alpha} = \lambda_\alpha(h_1), \quad \forall\, 3 \leq \alpha \leq 8$$
$$C^1{}_{2\alpha} = C^2{}_{2\alpha} = 0, \quad C^\alpha{}_{2\alpha} = \lambda_\alpha(h_2), \quad \forall\, 3 \leq \alpha \leq 8$$

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:

$$\text{ad}(h_i)[e_\alpha, e_\beta] = (\lambda_\alpha(h_i) + \lambda_\beta(h_i))[e_\alpha, e_\beta]$$

If $\lambda_\alpha + \lambda_\beta \in \Phi$, we have $[e_\alpha, e_\beta] = \xi e_\gamma$ for some $3 \leq \gamma \leq 8$ and $\xi \in \mathbb{C}$. Let us label the roots in our previous plot as:

λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
π_1	π_2	$\pi_1 + \pi_2$	$-\pi_1$	$-\pi_2$	$-(\pi_1 + \pi_2)$

Then, for example:

$$\text{ad}(h)[e_3, e_4] = (\pi_1 + \pi_2)(h)[e_3, e_4]$$

and hence $[e_3, e_4]$ is an eigenvector of $\mathrm{ad}(h)$ 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.

3.6 Representations Of Lie Algebras

Definition 3.32 (Representations Of Lie Algebra)**.** *Let L be a Lie algebra. A **representation** of L is a Lie algebra homomorphism:*

$$\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$$

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 3.33 (Representation Space)**.** *Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a representation of L. The vector space V is called the **representation space** of ρ.*

Definition 3.34 (Dimension Of Representation)**.** *Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a representation of L. The **dimension** of the representation ρ is $\dim V$.*

Example 3.3.
Consider the Lie algebra $\mathfrak{sl}(2, \mathbb{C})$. We constructed a basis $\{X_1, X_2, X_3\}$ satisfying the relations:

$$\begin{aligned} [X_1, X_2] &= 2X_2 \\ [X_1, X_3] &= -2X_3 \\ [X_2, X_3] &= X_1 \end{aligned}$$

Let $\rho\colon \mathfrak{sl}(2, \mathbb{C}) \xrightarrow{\sim} \mathrm{End}(\mathbb{C}^2)$ 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 ρ 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, ρ is a 2-dimensional representation of $\mathfrak{sl}(2, \mathbb{C})$ with representation space $\mathbb{C}^2$. Note that we have:

$$\begin{aligned}\mathrm{im}_\rho(\mathfrak{sl}(2, \mathbb{C})) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{End}(\mathbb{C}^2) \;\middle|\; a + d = 0 \right\} \\ &= \{\phi \in \mathrm{End}(\mathbb{C}^2) \mid \mathrm{tr}\, \phi = 0\}\end{aligned}$$

This is how $\mathfrak{sl}(2, \mathbb{C})$ is often defined in physics courses, i.e. as the algebra of 2×2 complex traceless matrices.

Example 3.4.
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 $\rho_{\mathrm{vec}} \colon \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \mathrm{End}(\mathbb{R}^3)$ by:

$$\rho_{\mathrm{vec}}(J_1) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \rho_{\mathrm{vec}}(J_2) := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad \rho_{\mathrm{vec}}(J_3) := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

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:

$$\rho_{\mathrm{spin}} \colon \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\sim} \mathrm{End}(\mathbb{C}^2)$$

with $\mathbb{C}^2$ understood as a 4-dimensional $\mathbb{R}$-vector space, defined by:

$$\rho_{\mathrm{spin}}(J_1) := -\frac{\mathrm{i}}{2}\sigma_1, \qquad \rho_{\mathrm{spin}}(J_2) := -\frac{\mathrm{i}}{2}\sigma_2, \qquad \rho_{\mathrm{spin}}(J_3) := -\frac{\mathrm{i}}{2}\sigma_3$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

You can again check that this is a representation of $\mathfrak{so}(3, \mathbb{R})$.

Notice that the two representations have different dimensions:

$$\dim \mathbb{R}^3 = 3 \neq 4 = \dim \mathbb{C}^2$$

Any (non-abelian) Lie algebra always has at least two special representations.

Definition 3.35 (Trivial Representation)**.** *Let L be a Lie algebra. A **trivial representation** of L is defined by:*

$$\begin{aligned}\rho_{\text{trv}} \colon L &\xrightarrow{\sim} \text{End}(V) \\ x &\mapsto \rho_{\text{trv}}(x) := 0\end{aligned}$$

where 0 denotes the trivial endomorphism on V.

This is indeed a representation for any L since:

$$\forall\, x, y \in L : \ \rho_{\text{trv}}([x,y]) = 0 = [\rho_{\text{trv}}(x), \rho_{\text{trv}}(y)]$$

Definition 3.36 (Ajdoint Representation)**.** *The **adjoint representation** of L is:*

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

Hence the adjoint map we have been using is actually a representation (we have already shown that ad is a Lie algebra homomorphism):

Definition 3.37 (Faithful Representation)**.** *A representation $\rho \colon L \xrightarrow{\sim} \text{End}(V)$ is called **faithful** if ρ is injective, i.e:*

$$\dim(\text{im}_\rho(L)) = \dim L$$

Example 3.5.
All representations considered so far are faithful, except for the trivial representations whenever the Lie algebra L is not itself trivial. Consider, for instance, the adjoint representation. We have:

$$\begin{aligned}\text{ad}(x) = \text{ad}(y) &\Leftrightarrow \forall\, z \in L : \text{ad}(x)z = \text{ad}(y)z \\ &\Leftrightarrow \forall\, z \in L : [x,z] = [y,z] \\ &\Leftrightarrow \forall\, z \in L : [x-y,z] = 0\end{aligned}$$

If L is trivial, then any representation is faithful. Otherwise, there is some non-zero $z \in L$, hence we must have $x - y = 0$, so $x = y$, and thus ad is injective.

Definition 3.38 (Direct Sum / Tensor Product Representations)**.** *Given two representations $\rho_1 \colon L \xrightarrow{\sim} \text{End}(V_1)$, $\rho_2 \colon L \xrightarrow{\sim} \text{End}(V_2)$, we can construct new representations called:*

*i) The **direct sum representation**:*

$$\begin{aligned}\rho_1 \oplus \rho_2 \colon L &\xrightarrow{\sim} \text{End}(V_1 \oplus V_2) \\ x &\mapsto (\rho_1 \oplus \rho_2)(x) := \rho_1(x) \oplus \rho_2(x)\end{aligned}$$

*ii) The **tensor product representation**:*

$$\begin{aligned}\rho_1 \otimes \rho_2 \colon L &\xrightarrow{\sim} \text{End}(V_1 \times V_2) \\ x &\mapsto (\rho_1 \otimes \rho_2)(x) := \rho_1(x) \otimes \text{id}_{V_2} + \text{id}_{V_1} \otimes \rho_2(x)\end{aligned}$$

Example 3.6.
The direct sum representation $\rho_{\text{vec}} \oplus \rho_{\text{spin}} \colon \mathfrak{so}(3,\mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3 \oplus \mathbb{C}^2)$ given in block-matrix form by:

$$(\rho_{\text{vec}} \oplus \rho_{\text{spin}})(x) = \left(\begin{array}{c|c} \rho_{\text{vec}}(x) & 0 \\ \hline 0 & \rho_{\text{spin}}(x) \end{array}\right)$$

is a 7-dimensional representation of $\mathfrak{so}(3,\mathbb{R})$.

Definition 3.39 (Reducible Representation)**.** *A representation $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ is called* ***reducible*** *if there exists a non-trivial vector subspace $U \subseteq V$ which is invariant under the action of ρ, i.e:*

$$\forall\, x \in L : \forall\, u \in U :\ \rho(x)u \in U$$

In other words, ρ restricts to a representation $\rho|_U\colon L \xrightarrow{\sim} \mathrm{End}(U)$.

Definition 3.40 (Irreducible Representation)**.** *A representation is* ***irreducible*** *if it is not reducible.*

Example 3.7.

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 ρ_{vec} and ρ_{spin} are both irreducible.

Remark 3.9. Just like the simple Lie algebras are the building blocks of all semi-simple Lie algebras, the irreducible representations of a semi-simple Lie algebra are the building blocks of all finite-dimensional representations of the Lie algebra. Any such representation con be decomposed as the direct sum of irreducible representations, which can then be classified according to their so-called *highest weights*.

3.6.1 The Casimir Operator

To every representation ρ of a compact Lie algebra (i.e. the Lie algebra of a compact Lie group) there is associated an operator Ω_ρ, called the Casimir operator. We will need some preparation in order to define it.

Definition 3.41 (ρ-Killing Form)**.** *Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a representation of a complex Lie algebra L. We define the* ***ρ-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 κ_{ad}. Similarly to κ_{ad}, every κ_ρ is symmetric and associative with respect to the Lie bracket of L.

Proposition 3.7. *Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a faithful representation of a complex semi-simple Lie algebra L. Then, κ_ρ is non-degenerate.*

Hence, κ_ρ induces an isomorphism $L \xrightarrow{\sim} L^*$ via

$$L \ni x \mapsto \kappa_\rho(x,-) \in L^*.$$

Recall that if $\{X_1,\dots,X_{\dim L}\}$ is a basis of L, then the dual basis $\{\widetilde{X}^1,\dots,\widetilde{X}^{\dim L}\}$ of L^* is defined by

$$\widetilde{X}^i(X_j) = \delta^i_j.$$

By using the isomorphism induced by κ_ρ, we can find some $\xi_1,\dots,\xi_{\dim L} \in L$ 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 3.8. *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 κ_ρ, 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

$$\forall\, 1 \le i \le \dim L:\ \kappa_\rho\Big(X_i,[X_j,\xi_k] - \sum_{m=1}^{\dim L} C^k{}_{mj}\xi_m\Big) = 0$$

and hence, the result follows from the non-degeneracy of κ_ρ. □

We are now ready to define the Casimir operator and prove the subsequent theorem.

Definition 3.42. *Let $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ be a faithful representation of a complex (compact) Lie algebra L and let $\{X_1,\dots,X_{\dim L}\}$ be a basis of L. The* Casimir operator *associated to the representation ρ 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 3.8. *Let Ω_ρ 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, Ω_ρ commutes with every endomorphism in $\mathrm{im}_\rho(L)$.

Proof.
Note that the bracket above is that on $\mathrm{End}(V)$. Let $x = x^k X_k \in L$. Then

$$\begin{aligned}[\Omega_\rho,\rho(x)] &= \Big[\sum_{i=1}^{\dim L}\rho(X_i)\circ\rho(\xi_i),\rho(x^kX_k)\Big]\\ &= \sum_{i,k=1}^{\dim L} x^k[\rho(X_i)\circ\rho(\xi_i),\rho(X_k)].\end{aligned}$$

Observe that if the Lie bracket as the commutator with respect to an associative product, as is the case for $\mathrm{End}(V)$, we have

$$\begin{aligned}[AB,C] &= ABC - CBA\\ &= ABC - CBA - ACB + ACB\\ &= A[B,C] + [A,C]B.\end{aligned}$$

Hence, by applying this, we obtain

$$\begin{aligned}
\sum_{i,k=1}^{\dim L} x^k[\rho(X_i)\circ\rho(\xi_i),\rho(X_k)] &= \sum_{i,k=1}^{\dim L} x^k\big(\rho(X_i)\circ[\rho(\xi_i),\rho(X_k)]+[\rho(X_i),\rho(X_k)]\circ\rho(\xi_i)\big)\\
&= \sum_{i,k=1}^{\dim L} x^k\big(\rho(X_i)\circ\rho([\xi_i,X_k])+\rho([X_i,X_k])\circ\rho(\xi_i)\big)\\
&= \sum_{i,k,m=1}^{\dim L} x^k\big(\rho(X_i)\circ\rho(-C^i{}_{mk}\xi_m)+\rho(C^m{}_{ik}X_m)\circ\rho(\xi_i)\big)\\
&= \sum_{i,k,m=1}^{\dim L} x^k\big(-C^i{}_{mk}\rho(X_i)\circ\rho(\xi_m)+C^m{}_{ik}\rho(X_m)\circ\rho(\xi_i)\big)\\
&= \sum_{i,k,m=1}^{\dim L} x^k\big(-C^i{}_{mk}\rho(X_i)\circ\rho(\xi_m)+C^i{}_{mk}\rho(X_i)\circ\rho(\xi_m)\big)\\
&= 0,
\end{aligned}$$

where we have swapped the dummy summation indices i and m in the second term. □

Lemma 3.1 (Schur)**.** *If $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ is irreducible, then any operator S which commutes with every endomorphism in $\mathrm{im}_\rho(L)$ has the form*

$$S = c_\rho \,\mathrm{id}_V$$

for some constant $c_\rho \in \mathbb{C}$ (or $\mathbb{R}$, if L is a real Lie algebra).

It follows immediately that $\Omega_\rho = c_\rho \,\mathrm{id}_V$ for some c_ρ but, in fact, we can say more.

Proposition 3.9. *The Casimir operator of $\rho\colon L \xrightarrow{\sim} \mathrm{End}(V)$ is $\Omega_\rho = c_\rho \,\mathrm{id}_V$, 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

$$\begin{aligned}
\mathrm{tr}(\Omega_\rho) &= \mathrm{tr}\Big(\sum_{i=1}^{\dim L} \rho(X_i)\circ\rho(\xi_i)\Big)\\
&= \sum_{i=1}^{\dim L} \mathrm{tr}(\rho(X_i)\circ\rho(\xi_i))\\
&= \sum_{i=1}^{\dim L} \kappa_\rho(X_i,\xi_i)\\
&= \sum_{i=1}^{\dim L} \delta_{ii}\\
&= \dim L,
\end{aligned}$$

which is what we wanted. □

Example 3.8.
Consider the Lie algebra $\mathfrak{so}(3,\mathbb{R})$ with basis $\{J_1, J_2, J_3\}$ satisfying

$$[J_i, J_j] = \varepsilon_{ijk} J_k,$$

where we assume the summation convention on the lower index k. Recall that the representation $\rho_{\text{vec}}\colon \mathfrak{so}(3,\mathbb{R}) \xrightarrow{\sim} \text{End}(\mathbb{R}^3)$ is defined by

$$\rho_{\text{vec}}(J_1) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \rho_{\text{vec}}(J_2) := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad \rho_{\text{vec}}(J_3) := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Let us first evaluate the components of $\kappa_{\rho_{\text{vec}}}$. We have

$$\begin{aligned}
(\kappa_{\rho_{\text{vec}}})_{11} := \kappa_{\rho_{\text{vec}}}(J_1, J_1) &= \text{tr}(\rho_{\text{vec}}(J_1) \circ \rho_{\text{vec}}(J_1)) \\
&= \text{tr}((\rho_{\text{vec}}(J_1))^2) \\
&= \text{tr} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}^2 \\
&= \text{tr} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \\
&= -2.
\end{aligned}$$

After calculating the other components similarly, we find

$$[(\kappa_{\rho_{\text{vec}}})_{ij}] = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Thus, $\kappa_{\rho_{\text{vec}}}(J_i, \xi_j) = \delta_{ij}$ requires that we define $\xi_i := -\frac{1}{2} J_i$. Then, we have

$$\begin{aligned}
\Omega_{\rho_{\text{vec}}} &:= \sum_{i=1}^{3} \rho_{\text{vec}}(J_i) \circ \rho_{\text{vec}}(\xi_i) \\
&= \sum_{i=1}^{3} \rho_{\text{vec}}(J_i) \circ \rho_{\text{vec}}(-\tfrac{1}{2} J_i) \\
&= -\frac{1}{2} \sum_{i=1}^{3} (\rho_{\text{vec}}(J_i))^2 \\
&= -\frac{1}{2} \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}^2 + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}^2 + \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}^2 \right) \\
&= -\frac{1}{2} \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

Hence $\Omega_{\rho_{\text{vec}}} = c_{\rho_{\text{vec}}} \, \text{id}_{\mathbb{R}^3}$ with $c_{\rho_{\text{vec}}} = 1$, which agrees with our previous theorem since

$$\frac{\dim \mathfrak{so}(3,\mathbb{R})}{\dim \mathbb{R}^3} = \frac{3}{3} = 1.$$

Example 3.9.
Let us consider the Lie algebra $\mathfrak{so}(3,\mathbb{R})$ again, but this time with representation ρ_{spin}. Recall that this is given by

$$\rho_{\text{spin}}(J_1) := -\frac{\mathrm{i}}{2} \sigma_1, \qquad \rho_{\text{spin}}(J_2) := -\frac{\mathrm{i}}{2} \sigma_2, \qquad \rho_{\text{spin}}(J_3) := -\frac{\mathrm{i}}{2} \sigma_3,$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. Recalling that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \mathrm{id}_{\mathbb{C}^2}$, we calculate

$$\begin{aligned}(\kappa_{\rho_{\mathrm{spin}}})_{11} := \kappa_{\rho_{\mathrm{spin}}}(J_1, J_1) &= \mathrm{tr}(\rho_{\mathrm{spin}}(J_1) \circ \rho_{\mathrm{spin}}(J_1)) \\ &= \mathrm{tr}((\rho_{\mathrm{spin}}(J_1))^2) \\ &= (-\tfrac{\mathrm{i}}{2})^2 \,\mathrm{tr}(\sigma_1^2) \\ &= -\tfrac{1}{4}\,\mathrm{tr}(\mathrm{id}_{\mathbb{C}^2}) \\ &= -1.\end{aligned}$$

Note that $\mathrm{tr}(\mathrm{id}_{\mathbb{C}^2}) = 4$, since $\mathrm{tr}(\mathrm{id}_V) = \dim V$ and here $\mathbb{C}^2$ is considered as a 4-dimensional vector space over $\mathbb{R}$. Proceeding similarly, we find that the components of $\kappa_{\rho_{\mathrm{spin}}}$ are

$$[(\kappa_{\rho_{\mathrm{spin}}})_{ij}] = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Hence, we define $\xi_i := -J_i$. Then, we have

$$\begin{aligned}\Omega_{\rho_{\mathrm{spin}}} &:= \sum_{i=1}^{3} \rho_{\mathrm{spin}}(J_i) \circ \rho_{\mathrm{spin}}(\xi_i) \\ &= \sum_{i=1}^{3} \rho_{\mathrm{spin}}(J_i) \circ \rho_{\mathrm{spin}}(-J_i) \\ &= -\sum_{i=1}^{3} (\rho_{\mathrm{spin}}(J_i))^2 \\ &= -\Big(-\frac{\mathrm{i}}{2}\Big)^2 \sum_{i=1}^{3} \sigma_i^2 \\ &= \frac{1}{4} \sum_{i=1}^{3} \mathrm{id}_{\mathbb{C}^2} \\ &= \frac{3}{4}\,\mathrm{id}_{\mathbb{C}^2},\end{aligned}$$

in accordance with the fact that

$$\frac{\dim \mathfrak{so}(3, \mathbb{R})}{\dim \mathbb{C}^2} = \frac{3}{4}.$$