MA103
Introduction to Abstract Mathematics
Lent Term 2021/22 – Weeks 6-10

Lecture Notes

These lecture notes are based on notes originally written by Amol Sasane with revisions by Konrad Swanepoel and Emilio Pierro

Contents

1 Groups

- Biggs, N. L. *Discrete Mathematics*. Chapter 20.
- Liebeck, M. *A Concise Introduction to Pure Mathematics*. Chapters 25–26.

1.1 Introduction

Groups are the most fundamental objects in abstract algebra. Many of the algebraic objects with which you are already familiar are already groups – $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ for example. However, in order to appreciate what a group is more generally (or rather: more abstractly!) we shall see many different examples and flavours of them. In mathematics it is important to see a wide variety of examples, so that we do not fall into the trap of thinking that the properties of the above groups are somehow characteristic of all groups. Indeed, the groups mentioned are quite "boring" and rather unilluminating in some sense. They are far from typical.

An important way to think about groups in general is to consider them as collections of "symmetries" of a particular object. We shall be deliberately vague about what we mean by a "symmetry", since in practice, as a group theorist, we are able to decide that for ourselves. Going further, one can consider a group as being a "measure" of how symmetric an object is.

For example, we might consider an equilateral triangle and our "symmetries" are going to be "ways I can move the triangle around so that it looks the same afterwards" and hopefully this coincides with what we intuitively feel a symmetry of the triangle (or a shape in general) should be. Under this definition, then, an equilateral triangle has six "symmetries" – the 3 reflections in axes passing through the centre and one of its corners; and the 3 rotations (say clockwise) around the centre (including the most important symmetry of all: the symmetry that leaves our triangle unchanged). As we consider regular n-gons with more and more sides, we see that they generally have twice as many symmetries as they do sides.

The extremal example of this would be the circle whose symmetry group has the same cardinality as $\mathbb{R}$ and this corresponds to our intuition that the circle is "highly symmetric". This particular family of groups is so important it has its own name: the dihedral groups.

Before giving the formal definition of a group we highlight some of the important features of what we have just considered. We would really like it to be the case that if we were to take two different symmetries of our object, perform one after the other (which we call their "composition" – similar to the composition of bijective functions, for example), then this is still a symmetry. This is where the "algebra" comes in – we are really "multiplying" symmetries together. Going a bit further we might have a (finite) sequence of many symmetries of our object that we would like to perform one after the other. If we were to label the points of our object and keep track of them as we perform these symmetries, a different ordering of the symmetries might send the

labels to different places. A priori this is perfectly allowed, since, as in a multiplication table, we only specify what the product of two symmetries is. For example, we declare that $1 + 2 = 3$, and that $2 + 3 = 5$, but we need slightly more to assert that $(1 + 2) + 3 = 1 + (2 + 3)$. In other words it shouldn't matter how we "group" – no pun intended – the pairs of adjacent symmetries together. To reiterate – this is important because we *only* define how to multiply two symmetries at a time. Of course we also believe that leaving something unchanged should be considered a symmetry, and finally, we should be able to undo or "invert" any symmetry. Going back to our equilateral triangle, each of our reflections we could "undo" by performing the reflection again; each of our rotations we could "undo" by rotating back in the opposite direction.

In general, as we consider more and more abstract groups, the "object" for which our group is the collection of symmetries of might become more and more abstract as well. Here it was easy to picture a triangle or a square. With a little imagination you might be able to convince yourself that the elements of $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$ (with addition) could be thought of as certain symmetries (translations) of the real line. Many groups out there unfortunately do not have such easily tangible objects on which they act. It is necessary therefore to be able to study a group abstractly and so we now turn to the abstract definition of a group.

1.2 Definition of a Group

Ultimately, a group will consist of two things: a non-empty set S and a rule $*$ for combining any two elements in S. We can think of our rule as a function f with domain $S \times S$ and eventually we require that the codomain of f will be S. This is really just an abstract way of saying, e.g. if I take two integers (i.e. an element of $\mathbb{Z} \times \mathbb{Z}$) and add them together, I'd like their sum to belong to $\mathbb{Z}$ too. We take this opportunity to point out a bit of standard mathematical notation. To refer to the function, we use the arrow "$\to$" in between the domain and codomain, e.g. $f : S \times S \to S$. When we describe what the function actually does, we use the arrow "$\mapsto$" as follows: $(a, b) \mapsto a + b$. This is read as "the element $(a, b) \in S \times S$ is mapped to the element $a + b \in S$".

The elements of S can in principle be **anything** – numbers, functions, rigid motions, permutations, shuffles of a deck of cards, anything at all. Our rule can also in principle be **anything** – addition, multiplication, composition, performing one shuffle followed by another shuffle – anything. Generally when speaking about groups we will tend to use the word "multiplication" to mean whatever our rule $*$ is. For our group to have any hope of making sense, though, we would like our multiplication to be a *binary operation*:

Definition 1.1. A **binary operation on a set** S is a function $f\colon S \times S \to S$.

This captures the idea that the composition of two symmetries should again be a symmetry. The functional notation $c = f(a, b)$ is not very convenient for what is going to follow, and so instead, the element obtained by applying the binary operation to a pair (a, b) is usually denoted using a notation resembling that used for addition or multiplication:

$$c = a * b, \quad \text{or} \quad ab, \quad \text{or} \quad a \circ b, \quad \text{or} \quad a + b \quad \text{and so on,}$$

with a fixed choice being made for the particular operation in question. In this notation we then have that $a * b$ is a binary operation on S iff $a * b \in S$ for any $a, b \in S$. Let us see some examples and non-examples.

Examples 1.2.

1. Addition of integers is a binary operation on $\mathbb{Z}$. Indeed, the sum of two integers is yet another integer, and addition is the function $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ that assigns $a + b$ to the pair (a, b), denoted by $(a, b) \mapsto a + b$.

2. Multiplication of real numbers is a binary operation on $\mathbb{R}$.

3. If a, b are rational numbers, then let $a * b = a + b - ab$. The function from $\mathbb{Q} \times \mathbb{Q}$ to $\mathbb{Q}$ given by $(a, b) \mapsto a * b$ is a binary operation on $\mathbb{Q}$.

4. If a, b are real numbers, then define $a * b = \sqrt{a^2 + b^2}$. Then $(a, b) \mapsto a * b$ is not a binary operation on $\mathbb{Q}$, since $1 \in \mathbb{Q}$, but $1 * 1 = \sqrt{2} \notin \mathbb{Q}$. However, $(a, b) \mapsto a * b$ is a binary operation on $\mathbb{R}$.

5. Let $n \in \mathbb{N}$, and let S denote the set of $n \times n$ matrices with real entries. Then matrix multiplication is a binary operation on S.

6. Let $n \in \mathbb{N}$, and let $GL(n, \mathbb{R})$ denote the set of all invertible matrices of size $n \times n$ with real entries. Then matrix multiplication is a binary operation on $GL(n, \mathbb{R})$. Indeed, if $A, B \in GL(n, \mathbb{R})$, then the matrix AB is again a matrix of size $n \times n$ with real entries, and moreover, since A and B are invertible, it follows that AB is also invertible.

7. Let $C(\mathbb{R})$ denote the set of all continuous functions on $\mathbb{R}$. Let addition of functions be defined as follows: if f, g belong to $C(\mathbb{R})$, then

$$(f + g)(x) = f(x) + g(x), \quad x \in \mathbb{R}.$$

   Then addition of functions is a binary operation on $C(\mathbb{R})$, since the sum of continuous functions is again continuous as proved in Michaelmas Term.

8. If $a, b \in \mathbb{N}$, then let

$$a * b = \frac{a}{b}.$$

   $*$ is not a binary operation on $\mathbb{N}$, since $1 * 2 = \frac{1}{2} \notin \mathbb{N}$. The function $(a, b) \mapsto \frac{a}{b}$ is also not a binary operation on $\mathbb{Q}$, since $(1, 0)$ is not mapped to any rational number by the function.

9. Let N be a set of size $n \in \mathbb{N}$ and let S be the set of all bijections from N to itself. Let $f, g \in S$. Our binary operation on S will be composition of functions and we denote it by $\circ$ as usual. Recall in Michaelmas Term we proved that a composition of bijections is again a bijection, hence $\circ : S \times S \to S$ given by $(f, g) \mapsto g \circ f$ is indeed a binary operation.

Now that we have abstracted the definition of "multiplication" we are ready to formally define the additional desirable properties that our set and multiplication should have.

Definition 1.3. A **group** is a pair $(G, *)$, where G is a non-empty set, that satisfies the following four **axioms**:

Closure: For all $a, b \in G$, $a * b \in G$. In other words, that $*$ is a binary operation on G.

Associativity: For all $a, b, c \in G$, the product $a * (b * c) = (a * b) * c$.

Identity: There exists an element $e \in G$ such that $e * g = g = g * e$ for all $g \in G$. Such an element e is referred to as an **identity element**.

Inverses: For every $g \in G$, there exists an element $g^{-1} \in G$ such that $g * g^{-1} = e = g^{-1} * g$. Such an element g^{-1} is called an **inverse** of the element g in the group G.

Where the binary operation is clear, we shall simply write G for $(G, *)$.

Warning 1.4. We may write 1_G for the identity element of G (or 0_G when we think of the operation as being some sort of "addition"), like in the very next example. Take care because **we may use this notation even for groups whose elements are not numbers!**

Let us now see a few examples of groups.

Examples 1.5.

1. $(\mathbb{Z}, +)$ is a group, as may be seen by checking each of the axioms.

   Closure: For any $a, b \in \mathbb{Z}$, their sum $a + b \in \mathbb{Z}$.

   Associativity: Similarly, for all $a, b, c \in \mathbb{Z}$, $(a + b) + c = a + (b + c)$.

   Identity: The element $0 \in \mathbb{Z}$ is an **identity** element since for all $a \in \mathbb{Z}$, $a + 0 = a = 0 + a$.

   Inverses: Finally, if $a \in \mathbb{Z}$, then $-a \in \mathbb{Z}$ and $a + (-a) = 0 = -a + a$.

2. $(\mathbb{R}, .)$ is not a group, although it does satisfy three of the four axioms.

   Closure: For all $a, b \in \mathbb{R}$, $ab \in \mathbb{R}$.

   Associativity: For all $a, b, c \in \mathbb{R}$, $(ab)c = a(bc)$.

   Identity: If e is an identity element, then we must have $ae = a = ea$ for all $a \in \mathbb{R}$, and in particular, with $a = 1$, we should have $1e = 1$, and so, $e = 1$. And so if e is an identity element, then it must necessarily be equal to 1. We then check that 1 serves as an identity element: for all $a \in \mathbb{R}$, $a1 = a = 1a$.

   ~~**Inverses**~~: We do not have the inverses axiom, however, since $0 \in \mathbb{R}$, but for any potential inverse $b \in \mathbb{R}$ we find that $0b = b0 = 0 \neq 1 = e$, a contradiction.

3. However, the set $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ of non-zero real numbers with multiplication forms a group.

   Closure: For closure we notice that for all $a, b \in \mathbb{R}$, **if in addition** $a, b \neq 0$, then $ab \in \mathbb{R}$ **as well as** $ab \neq 0$.

   Associativity: The associative axiom passes to $\mathbb{R}^*$ and so is satisfied.

   Identity: Also, since $1 \in \mathbb{R}^*$, the **identity** axiom is still satisfied.

   Inverses: Finally, let $a \in \mathbb{R}^*$. To find out if a has an inverse x, we have to solve the equations $ax = xa = 1$. It is clear that $x = \frac{1}{a}$ (which exists, since $a \neq 0$). Hence the inverses axiom is satisfied. Thus we have for any $a \in \mathbb{R}^*$ that $a^{-1} = \frac{1}{a}$ serves as an inverse, since $a \cdot \frac{1}{a} = \frac{1}{a} \cdot a = 1$.

4. Similarly, $(\mathbb{R}^+, \times)$ the set of positive real numbers with ordinary multiplication is also a group.

5. For $n \in \mathbb{N}$, let $GL(n, \mathbb{R})$ denote the set of all invertible $n \times n$ matrices with real entries. Then $GL(n, \mathbb{R})$ is group with matrix multiplication.

   Closure: There are two ways that we can see closure:

   i. For all $A, B \in GL(n, \mathbb{R})$, since A and B are invertible $n \times n$ matrices, $\det A \neq 0$ and $\det B \neq 0$. Therefore, $\det(AB) = (\det A)(\det B) \neq 0$, which gives that AB is also an invertible $n \times n$ matrix. That is, $AB \in GL(n, \mathbb{R})$, satisfying the **closure** axiom.

   ii. Alternatively, given A and B as before, is their product AB an invertible matrix? Yes: it's inverse is the product $B^{-1}A^{-1}$, hence $AB \in GL(n, \mathbb{R})$.

   Associativity: For all $A, B, C \in GL(n, \mathbb{R})$, $(AB)C = A(BC)$, since matrix multiplication is associative.

Identity: The identity matrix

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

serves as an **identity** element. I_n is an invertible matrix of size $n \times n$ with real entries, and so it belongs to $GL(n, \mathbb{R})$, and moreover, for all $A \in GL(n, \mathbb{R})$, $AI_n = A = I_n A$.

Inverses: If $A \in GL(n, \mathbb{R})$, then A is an invertible matrix, and so there exists a matrix A^{-1} such that $AA^{-1} = I_n = A^{-1}A$. The matrix A^{-1} is thus in $GL(n, \mathbb{R})$, and serves as an inverse of A satisfying the **inverses** axiom.

This group is called the **general linear group**.

6. Let $m, n \in \mathbb{N}$ and let $M_{m,n}(\mathbb{R})$ be the set of matrices of size $m \times n$ with entries in $\mathbb{R}$. This set, along with usual matrix addition is a group. Let $A, B \in M_{m,n}(\mathbb{R})$ where A and B are as follows:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}$$

where all of the a_{ij}, $b_{ij} \in \mathbb{R}$.

Closure: It is clear that $A + B$ is again an element of $M_{m,n}(\mathbb{R})$.

Associativity: In each entry of the matrix we are performing addition in $\mathbb{R}$, and so $M_{m,n}(\mathbb{R})$ inherits **associativity** from $(\mathbb{R}, +)$.

Identity: The zero matrix is an identity for $M_{m,n}(\mathbb{R})$.

Inverses: The inverses of a matrix A is the matrix $-A$, which is obtained by negating every entry in A. Hence **inverses** is also satisfied.

7. Let $C(\mathbb{R})$ be as before. This is a group with addition of functions:

Closure: For all $f, g \in C(\mathbb{R})$, $f + g \in C(\mathbb{R})$ since the sum of continuous functions is continuous.

Associativity: For all $f, g, h \in C(\mathbb{R})$, and any $x \in \mathbb{R}$ we have

$$\begin{aligned} (f + (g + h))(x) &= f(x) + (g + h)(x) \\ &= f(x) + (g(x) + h(x)) \\ &= (f(x) + g(x)) + h(x) \\ &\qquad \text{(since addition is associative in } \mathbb{R}\text{!)} \\ &= (f + g)(x) + h(x) \\ &= ((f + g) + h)(x). \end{aligned}$$

Hence $f + (g + h) = (f + g) + h$.

Identity: The constant function $\mathbf{0}$ defined by $\mathbf{0}(x) = 0$ for all $x \in \mathbb{R}$, serves as an identity element. $\mathbf{0}$ is a continuous function on $\mathbb{R}$ and so $\mathbf{0} \in C(\mathbb{R})$. Moreover, for all $f \in C(\mathbb{R})$, we have for all $x \in \mathbb{R}$:

$$\begin{aligned}(f + \mathbf{0})(x) &= f(x) + \mathbf{0}(x) \\ &= f(x) + 0 \\ &= f(x) \text{ (0 is an identity element for addition in } \mathbb{R}) \\ &= 0 + f(x) \\ &= \mathbf{0}(x) + f(x) \\ &= (\mathbf{0} + f)(x).\end{aligned}$$

Hence $f + \mathbf{0} = f = \mathbf{0} + f$.

Inverses: If $f \in C(\mathbb{R})$, then define $-f$ by $(-f)(x) = -f(x)$, for $x \in \mathbb{R}$. Given $f \in C(\mathbb{R})$, we have for all $x \in \mathbb{R}$:

$$\begin{aligned}(f + (-f))(x) &= f(x) + (-f)(x) \\ &= f(x) + (-f(x)) \\ &= 0 = \mathbf{0}(x) = 0 \\ &= -f(x) + f(x) \\ &= (-f)(x) + f(x) \\ &= (-f + f)(x).\end{aligned}$$

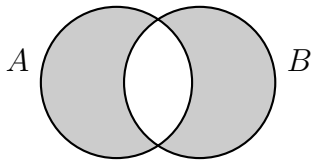
Hence $f + (-f) = \mathbf{0} = -f + f$.

8. For any set S, its **power set** $\mathcal{P}(S)$ is defined to be the set consisting of all subsets of S:

$$\mathcal{P}(S) = \{A \mid A \subseteq S\}.$$

Define the following binary operation on $\mathcal{P}(A)$, called the **symmetric difference operation**:

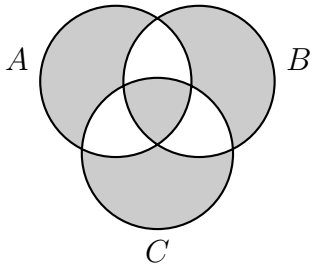
$$A \triangle B = (A \setminus B) \cup (B \setminus A), \quad A, B \in \mathcal{P}(S).$$



Then $\mathcal{P}(S)$ with $\triangle$ is a group:

Closure: For any $A, B \in \mathcal{P}(S)$ we have $A, B \subseteq S$, and therefore, $A \setminus B, B \setminus A \subseteq S$ and $A \triangle B = (A \setminus B) \cup (B \setminus A) \subseteq S$. Thus $A \triangle B \in \mathcal{P}(S)$.

Associativity: For any $A, B, C \in \mathcal{P}(S)$ we have to check that $A \triangle (B \triangle C) = (A \triangle B) \triangle C$. It is easily checked that both expressions correspond to the grey part in the following Venn diagram:



Identity: Since $A \triangle \varnothing = A = \varnothing \triangle A$ for all $A \in \mathcal{P}(S)$, the empty set $\varnothing$ serves as an identity element.

Inverses: Since $A \triangle A = \varnothing$, each element is its own inverse: $A^{-1} = A$ for all $A \in \mathcal{P}(S)$.

9. Given $n \in \mathbb{N}$, let S_n be the set of all bijections from the set $\{1, 2, \ldots, n\}$ to itself, and let $\circ$ denote the composition of functions, so $\alpha \circ \beta$ means the function with $(\alpha \circ \beta)(x) = \alpha(\beta(x))$. Then $(S_n, \circ)$ is a group, called the **symmetric group on n symbols**.

   Closure: You have seen that the composition of two bijections is a bijection, so indeed if α and β are in S_n then $\alpha \circ \beta$ is also in S_n.

   Associativity: If α, β, γ are three elements of S_n, and $x \in \{1, \ldots, n\}$, then

   $$\Big((\alpha \circ \beta) \circ \gamma\Big)(x) = (\alpha \circ \beta)(\gamma(x)) = \alpha(\beta(\gamma(x))) = \alpha((\beta \circ \gamma)(x)) = \Big(\alpha \circ (\beta \circ \gamma)\Big)(x),$$

   and so $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$ for all $\alpha, \beta, \gamma \in S_n$.

   Identity: The identity function ι, with $\iota(x) = x$ for each x, is a bijection, and evidently $\iota \circ \alpha = \alpha = \alpha \circ \iota$ for each $\alpha \in S_n$.

   Inverses: If $\alpha \in S_n$ is a bijection, then the inverse function α^{-1} is also a bijection, and serves as an inverse to α in $(S_n, \circ)$.

Activity 1.6. *Determine which group axioms are satisfied by the following sets and rules. Hence determine which are groups and which are not.*

1. *The set $G = \mathbb{Z}$ and the rule $a * b = \frac{a+b}{ab}$ for $a, b \in G$.*
2. *The set $G = \mathbb{Q}$ and the rule $a * b = -ab$ for $a, b \in G$.*
3. *The set $G = \mathbb{Q} \setminus \{0\}$ and the rule $a * b = -ab$ for $a, b \in G$.*

Definition 1.7. A group $(G, *)$ is said to be a **finite group** if the set G has finite cardinality. The **order of a finite group** $(G, *)$ is the cardinality of G. A group is said to be an **infinite group** if it is not finite.

Examples 1.8.

1. The set $\{1, -1, i, -i\}$ with multiplication is a finite group of order 4.
2. The set $\mathbb{Z}$ with addition is an infinite group.
3. The set $\mathcal{P}(S)$ with the symmetric difference operation $\triangle$ is a finite group if S is a finite set, and an infinite group if S is an infinite set.
4. The set S_n of bijections from a finite set of order n to itself is a finite group. Its order is $n!$.

A finite group can be completely described by writing its *group table*.

Definition 1.9. The **group table** of a finite group is a table that displays the law of composition as follows: the elements of the group are listed in the first row and the first column. Conventionally, the two lists have the group elements in the same order, with the identity element first. Given $a, b \in G$, the element $a * b$ is entered in the row corresponding to a and the column corresponding to b, as shown below.

$*$	$\cdots$	b	$\cdots$
$\vdots$			
a		$a * b$	
$\vdots$			

We clarify this further by considering a few examples.

Examples 1.10.

1. The finite group $\{-1, 1\}$ with multiplication can be described by the group table given below.

$\cdot$	1	-1
1	1	-1
-1	-1	1

   The table completely describes the binary operation: $1 \cdot 1 = 1$, $1 \cdot (-1) = -1$, $-1 \cdot 1 = -1$ and $-1 \cdot (-1) = 1$.

2. Recall the group S_n we described in Example 4.5.9. Let us consider the case $n = 3$, since this is small enough to work with explicitly. The group S_3 has $3! = 6$ elements, listed as follows:

x	1	2	3
$\iota(x)$	1	2	3

x	1	2	3
$\alpha(x)$	1	3	2

x	1	2	3
$\beta(x)$	3	2	1

x	1	2	3
$\gamma(x)$	2	1	3

x	1	2	3
$\delta(x)$	2	3	1

x	1	2	3
$\varepsilon(x)$	3	1	2

   We can now form the group table for $(S_3, \circ)$. For example, since

$$\begin{aligned}\alpha \circ \beta(1) &= \alpha(\beta(1)) = \alpha(3) = 2,\\ \alpha \circ \beta(2) &= \alpha(\beta(2)) = \alpha(2) = 3,\\ \alpha \circ \beta(3) &= \alpha(\beta(3)) = \alpha(1) = 1,\end{aligned}$$

   we obtain that $\alpha \circ \beta = \delta$. In a similar way, the whole group table may be calculated:

$\circ$	ι	α	β	γ	δ	ε
ι	ι	α	β	γ	δ	ε
α	α	ι	δ	ε	β	γ
β	β	ε	ι	δ	γ	α
γ	γ	δ	ε	ι	α	β
δ	δ	γ	α	β	ε	ι
ε	ε	β	γ	α	ι	δ

   Given that the group S_n more generally has $n!$ elements, you would not wish to write the group table out when n is much bigger than 4. Moreover, you would not like to use a different symbol for every different element of G. Think about what we do for integers – we only really use 10 different symbols. In the next section we shall consider how to use fewer symbols to write down all the elements of groups in general.

3. The finite group $\mathcal{P}(\{1,2\})$ with the symmetric difference operation $\triangle$ has four elements:

$$\mathcal{P}(\{1,2\}) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$$

and has the following group table:

$\triangle$	$\varnothing$	$\{1\}$	$\{2\}$	$\{1,2\}$
$\varnothing$	$\varnothing$	$\{1\}$	$\{2\}$	$\{1,2\}$
$\{1\}$	$\{1\}$	$\varnothing$	$\{1,2\}$	$\{2\}$
$\{2\}$	$\{2\}$	$\{1,2\}$	$\varnothing$	$\{1\}$
$\{1,2\}$	$\{1,2\}$	$\{2\}$	$\{1\}$	$\varnothing$

Again the table completely describes the binary operation. For example,

$$\{1,2\} \triangle \{2\} = (\{1,2\} \setminus \{2\}) \cup (\{2\} \setminus \{1,2\}) = \{1\} \cup \varnothing = \{1\}.$$

More generally the group $\mathcal{P}(S)$ for an arbitrary set S has order $2^{|S|}$ so you **really** do not want to be writing down group tables with different symbols for each of its elements.

1.3 Proving theorems, laws of exponents, generators

We now prove a few elementary theorems concerning groups.

Theorem 1.11. *There is a unique identity element in a group.*

Proof. Let e and e' be identity elements in $(G, *)$. Since $e \in G$ and e' is an identity, we obtain

$$e = e * e'.$$

Moreover, since $e' \in G$ and e is an identity, we also have

$$e * e' = e'.$$

Consequently, $e = e'$. □

Theorem 1.12. *Let $(G, *)$ be a group and let $a \in G$. Then a has a unique inverse.*

Proof. Let the group have the identity e. If a_1 and a_2 are inverses of a, then we have

$$\begin{aligned} a_1 &= a_1 * e \text{ (since } a_1 \in G \text{ and } e \text{ is the identity)} \\ &= a_1 * (a * a_2) \text{ (since } a_2 \text{ is an inverse of } a) \\ &= (a_1 * a) * a_2 \text{ (associativity)} \\ &= e * a_2 \text{ (since } a_1 \text{ is an inverse of } a) \\ &= a_2 \text{ (since } a_2 \in G \text{ and } e \text{ is the identity).} \end{aligned}$$

□

Activity 1.13.

Let $(G, *)$ be a group and let $a, b, c \in G$. Prove that if $a * b = a * c$, then $b = c$.

Example 1.14. *Use* **associativity** *to show that $(a*b)*(c*d) = (a*(b*c))*d$ for any elements a, b, c, d of a group $(G, *)$.*

Solution. Note in each step how the associativity axiom is applied:

$$\begin{aligned}(a*b)*(c*d) &= a*(b*(c*d))\\ &= a*((b*c)*d)\\ &= (a*(b*c))*d.\end{aligned}$$

□

In the previous example we saw that by **associativity**, two ways of using parentheses on $a*b*c*d$ resulted in equal expressions. It can be shown that in general, for any elements $a_1, a_2, \ldots, a_n$ of a group $(G, *)$, any two ways of parenthesising $a_1 * a_2 * \cdots * a_n$ will result in the same element of the group. We therefore in general do not show the parentheses. Instead we write $a_1 * a_2 * \cdots * a_n$, or even just or $a_1 a_2 \ldots a_n$, when it is clear which binary operation is meant. In particular, powers of elements in a group behave as we would expect them to.

Definition 1.15. Let G be a group and let $a \in G$. We define

$$a^0 = e \text{ and } a^n = a^{n-1} * a \text{ for } n \in \mathbb{N}.$$

Moreover, if $n \in \mathbb{N}$, define

$$a^{-n} = (a^n)^{-1}.$$

It can be shown that the usual **laws of exponents** hold: for all $m, n \in \mathbb{Z}$,

$$a^m * a^n = a^{m+n} \text{ and } (a^m)^n = a^{mn}.$$

Warning 1.16. It is **not necessarily true** that $(a*b)^n = a^n * b^n$.

Following this line of thought one can begin to study groups in a purely abstract fashion, by simply giving a list of symbols and relations that they obey. We make this more formal with the following definition, but we shall not pursue this more abstract development much further after this section.

Definition 1.17. Let G be a group and let S be a subset of elements of G (note that S can equal G). A **word in** S is a string of elements $s_1^{k_1} s_2^{k_2} \ldots s_n^{k_n}$ where each of the $s_i \in S$. We say that S is a **generating set** for G if every element of G can be written as a word using only the elements of S (and their inverses). By $\langle S \rangle$ we mean the group consisting of all words that can be written using the elements of S and their inverses. We say that a group is **finitely generated** if there exists a finite set S such that $G = \langle S \rangle$.

Remark 1.18. We may also write $\langle s_1, \ldots, s_n \rangle$ instead of $\langle S \rangle$ where the s_i are elements of the group.

We now use this to rethink the group S_3 as we saw it in Example 4.10.2.

Theorem 1.19. *Every element of the group S_3 can be generated by $\alpha(x)$ and $\varepsilon(x)$.*

Proof. Our aim is to show that the remaining bijections: $\iota(x)$, $\beta(x)$, $\gamma(x)$ and $\delta(x)$ can all be written as words using only $\alpha(x)$, $\varepsilon(x)$ and their inverses. It can easily be checked that $\alpha(x) = \alpha^{-1}(x)$ and so composing these two together gives us the identity bijection $\iota(x)$. The bijection $\varepsilon(x)$ is not equal to its own inverse, in fact $\varepsilon^{-1} = \delta(x)$, which can easily be checked. It remains to find expressions for $\beta(x)$ and $\gamma(x)$. Let us try the composition $\alpha \circ \varepsilon(x) = \alpha(\varepsilon(x))$. This maps $1 \to 2$, $2 \to 1$ and $3 \to 3$ and so it is equal to $\gamma(x)$. We leave it as an exercise for the reader to show that $\varepsilon \circ \alpha(x) = \beta(x)$. □

The group table for S_3 could now be more easily expressed as follows. For brevity, we write e.g. $\alpha\varepsilon$ for $\alpha \circ \varepsilon(x)$.

$\circ$	ι	α	$\varepsilon\alpha$	$\alpha\varepsilon$	ε^{-1}	ε
ι	ι	α	$\varepsilon\alpha$	$\alpha\varepsilon$	ε^{-1}	ε
α	α	ι	ε^{-1}	ε	$\varepsilon\alpha$	$\alpha\varepsilon$
$\varepsilon\alpha$	$\varepsilon\alpha$	ε	ι	ε^{-1}	$\alpha\varepsilon$	α
$\alpha\varepsilon$	$\alpha\varepsilon$	ε^{-1}	ε	ι	α	$\varepsilon\alpha$
ε^{-1}	ε^{-1}	$\alpha\varepsilon$	α	$\varepsilon\alpha$	ε	ι
ε	ε	$\varepsilon\alpha$	$\alpha\varepsilon$	α	ι	ε^{-1}

This is more useful, but again, you still would not wish to do this for larger groups. Here are some more examples of finitely and non-finitely generated groups.

Examples 1.20.

1. Any finite group is of course finitely generated.
2. Let $G = \mathbb{Z}$. This group is generated by 1 and so $G = \langle 1 \rangle$ and so again G is finitely generated.
3. Let $G = (\mathbb{Q}, +)$. This group does not possess a finite generating set, proving so is left as an exercise.

1.4 Subgroups

Definition 1.21. Let $(G, *)$ be a group and let H be a subset of G. We say that H is a **subgroup** of G if H satisfies the group axioms with the operation $*$ restricted to H.

Activity 1.22. *Let $(G, *)$ be a group with identity 1_G and let $(H, *)$ be a subgroup of G with identity 1_H. Prove that $1_H = 1_G$.*

Examples 1.23.

1. The subset of even integers $\{2m \mid m \in \mathbb{Z}\}$ is a subgroup of the group of integers $\mathbb{Z}$ with addition. Indeed, the sum of even numbers is even (and so **closure** holds), 0 is even (and so **identity** holds), and finally, given the even number $2m$, $-2m = 2(-m)$ is even as well (and so **inverses** holds).
2. The group of integers with addition $(\mathbb{Z}, +)$ is a subgroup of the group of rational numbers with addition $(\mathbb{Q}, +)$, which in turn is a subgroup of the group of real numbers with addition $(\mathbb{R}, +)$.
3. If G is a group with identity e, then $\{e\}$ and G are both subgroups of G.
4. Let n be a natural number and let $G = GL(n, \mathbb{R})$ with the usual matrix multiplication. An important subgroup of G is the set of matrices with determinant 1, which we denote $SL(n, \mathbb{R})$. If $A, B \in SL(n, \mathbb{R})$, then $\det(A) = \det(B) = 1$, hence $\det(AB) = 1$, and so this set satisfies **closure**. The existence of **inverses** is proved similarly. **Associativity** is immediate, since the binary operation is matrix multiplication. Finally, the **identity** matrix (of $GL(n, \mathbb{R})$) has determinant 1 and so this is indeed the identity for $SL(n, \mathbb{R})$.
5. The subset of 2×2 **symmetric matrices** with real entries, namely

$$\left\{ \begin{pmatrix} a & b \\ b & d \end{pmatrix} \;\middle|\; a, b, d \in \mathbb{R} \right\}$$

is a subgroup of the set of all 2×2 matrices with real entries together with matrix addition. Indeed, given any two symmetric matrices

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} \text{ and } \begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix},$$

their sum

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} = \begin{pmatrix} a+a' & b+b' \\ b+b' & d+d' \end{pmatrix}$$

is also symmetric, and so **closure** holds. Clearly the identity element

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

is symmetric, and so **identity** also holds. Finally, the inverse (with respect to matrix addition) of any symmetric matrix

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix},$$

is the element

$$\begin{pmatrix} -a & -b \\ -b & -d \end{pmatrix}$$

which is also symmetric, and so **inverses** holds.

6. The subset of *upper triangular* invertible matrices,

$$UT(2,\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \;\middle|\; a,b,d \in \mathbb{R} \text{ and } ad \neq 0 \right\}$$

is a subgroup of the group $GL(2,\mathbb{R})$ with matrix multiplication. Indeed, if

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix} \in UT(2,\mathbb{R}),$$

then

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix} = \begin{pmatrix} aa' & ab'+bd' \\ 0 & dd' \end{pmatrix} \in UT(2,\mathbb{R}),$$

and so **closure** holds. Clearly

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in UT(2,\mathbb{R}),$$

and so **identity** also holds. Finally,

$$\text{if } \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in UT(2,\mathbb{R}), \text{ then } \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{a} & -\frac{b}{ad} \\ 0 & \frac{1}{d} \end{pmatrix} \in UT(2,\mathbb{R}).$$

For a given group G checking all of the group axioms for a subgroup H can be quite tedious, we would like to find a much simpler test. Indeed, there is one.

Theorem 1.24 (The Subgroup Criterion)**.** *Let G be a group and let H be a subset of G. Then, H is a subgroup of G if and only if H is non-empty and for all $g,h \in H$, $gh^{-1} \in H$.*

Proof. There are two directions to prove here. One direction is immediate, and so we prove the reverse implication.

Assume then, that H is a non-empty subset of G, and that for all $g, h \in H$, that $gh^{-1} \in H$. Call this property $(*)$. Associativity is immediate, since this is inherited from G, and it remains to prove that H satisfies Closure, Identity and Inverses.

Let $g \in H$, which exists since H is non-empty. By property $(*)$, $gg^{-1} = 1_G \in H$, and so H satisfies the Identity axiom. Since $1_G \in H$, for all $g \in H$ we have $1_G g^{-1} = g^{-1} \in H$, and so H satisfies Inverses. Finally, let $g, h \in H$. Then $h^{-1} \in H$, since H satisfies Inverses, and so we can form the product $g(h^{-1})^{-1} = gh$ which, by Property $(*)$, belongs to H and so H also satisfies Closure. Hence, H is a subset of G and a group, and so it is a subgroup of G. □

1.5 The order of an element of a group

Definition 1.25. Let G be a group and suppose that $a \in G$.

1. If there exists an $m \in \mathbb{N}$ such that $a^m = e$, then a is said to have **finite order**.
2. If a has finite order, then the **order of** a, denoted by $\text{ord}(a)$, is
$$\text{ord}(a) = \min \{m \in \mathbb{N} \mid a^m = e\}.$$
3. If a does not have finite order, then a is said to have **infinite order**.

Warning 1.26. Take care not to confuse the **order of an element** with the **order of the group**. Go back and check these definitions if you are unsure.

Examples 1.27.

1. The element -1 has order 2 in the group of non-zero real numbers with multiplication.
2. The element 2 has infinite order in the group of integers with addition.
3. The element $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ is an element of order 3 in the group $GL(3, \mathbb{R})$.

Activity 1.28. *Determine the order of* $\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$ *in* $GL(2, \mathbb{R})$.

Activity 1.29. *Let $(G, *)$ be any group and let $a, b \in G$. Prove that the order of $a * b$ is the same as the order of $b * a$. Note that we are not assuming that G is abelian.*

We now prove that given any element from a group, the set of its powers forms a subgroup of the group.

Theorem 1.30. *Suppose that $(G, *)$ is a group and $a \in G$. Let $\langle a \rangle = \{a^n \mid n \in \mathbb{Z}\}$. Then:*

1. *$\langle a \rangle$ is a subgroup of G. Moreover, $\langle a \rangle$ is the* smallest *subgroup of G containing a.*
2. *If a is an element with finite order m, then $\langle a \rangle = \{e, a, a^2, a^3, \ldots, a^{m-1}\}$.*

Proof.

1. We use the Subgroup Criterion to prove this. Let $g, h \in \langle a \rangle$. Then, there exist k, l such that $g = a^k$ and $h = a^l$. The product $gh^{-1} = a^k a^{-l} = a^{k-l}$ is then clearly an element of $\langle a \rangle$. So $\langle a \rangle$ is a subgroup of G.

   Conversely, if H is any subgroup of G containing the element a, then since H is closed under multiplication, H contains the set $\{a^k \mid k \in \mathbb{Z}\} = \langle a \rangle$. In other words $\langle a \rangle \subseteq H$.

2. Clearly $\{e, a, a^2, a^3, \ldots, a^{m-1}\} \subseteq \langle a \rangle$. Conversely, if $n \in \mathbb{Z}$, then there exist integers q and r, such that $0 \leq r \leq m-1$, and $n = q \cdot m + r$. So

$$\begin{aligned} a^n &= a^{q \cdot m + r} = a^{q \cdot m} * a^r = (a^m)^q * a^r = (e)^q * a^r \\ &= e * a^r = a^r \in \{e, a, a^2, a^3, \ldots, a^{m-1}\}. \end{aligned}$$

□

Example 1.31. The element $[1]_5$ in the group $(\mathbb{Z}_5, \oplus)$ (addition modulo 5) has finite order, since

$$[1]_5 \oplus [1]_5 \oplus [1]_5 \oplus [1]_5 \oplus [1]_5 = [0]_5,$$

and the subgroup $\langle [1]_5 \rangle = \{[0]_5, [1]_5, [2]_5, [3]_5, [4]_5\}$ is in fact the whole group. □

The above example motivates the following definition.

Definition 1.32. A group G is said to be a **cyclic group** if there exists an element $a \in G$ such that $G = \langle a \rangle$. That is, if such a group can be generated by one element.

1.6 Abelian groups

Cyclic groups are a particular type of groups more generally known as abelian groups which we now discuss. The groups in Examples 4.5.1, 4.5.2, 4.5.7 and 4.5.8 also satisfy the following axiom:

Commutativity: For all $a, b \in G$, $a * b = b * a$.

On the other hand, the group in Example 4.5.5 does not satisfy **commutativity** for any $n \geq 2$. For example, when $n = 2$:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

This gives rise to the following natural definition.

Definition 1.33. A group $(G, *)$ is said to be an **abelian group**[1] if its binary operation is **commutative**, that is, for all $a, b \in G$, $a * b = b * a$.

Examples 1.34.

1. The group $(\mathbb{Z}, +)$ of integers with addition is an abelian group.
2. The group $(\mathbb{R}^*, \times)$ of non-zero real numbers with multiplication is an abelian group.
3. Let $m, n \in \mathbb{N}$. The group $(M_{m,n}(\mathbb{R}), +)$ is an abelian group.
4. The group $(C(\mathbb{R}), +)$ is an abelian group.
5. For any S, the group $(\mathcal{P}(S), \triangle)$ is an abelian group.

[1] after the Norwegian mathematician Niels Henrik Abel (1802–1829).

6. The symmetric group S_3 is not an abelian group. For instance, referring to the revised group table in Example 4.10.2, we see that $\alpha\varepsilon \neq \varepsilon\alpha$.

Activity 1.35. *Let $n \in \mathbb{N}$ and let $G = GL(n, \mathbb{Q})$. Prove that G is abelian if and only if $n = 1$.*

We often use **additive notation** when working with abelian groups. We then

use a symbol such as '$+$', '$\oplus$', '$\boxplus$', etc., to denote the binary operation,

use a symbol such as 0, $\boldsymbol{o}$, etc., to denote the identity element,

and denote the inverse of an element x by $-x$.

Furthermore, instead of writing a^m for $a + a + \cdots + a$, we use the more natural-looking notation $m \cdot a$. As before, we extend it also to $m = 0$ by writing $0 \cdot a = 0$ (where the 0 on the right-hand side is the identity element of the group) and to $m < 0$, where we still write $m \cdot a$ instead of a^m. In particular, in this notation we then have $-x = (-1) \cdot x$.

This notation is nothing more than a new notation for a^m in the case of abelian groups where additive notation is used. The laws of exponents look as follows in additive notation:

$$m \cdot a + n \cdot a = (m + n) \cdot a \quad \text{and} \quad m \cdot (n \cdot a) = (mn) \cdot a.$$

1.7 Homomorphisms and isomorphisms

The study of algebra is not just about *structures* with algebraic operations on them, but also about functions between such structures, that in an obvious sense "respect" the algebraic operations. The following definition illustrates this idea in the case of groups.

Definition 1.36. Let $(G, *)$ and $(H, \circ)$ be groups. A **homomorphism** $\varphi\colon G \to H$ is a function such that

$$\text{for all } a, b \in G, \quad \varphi(a * b) = \varphi(a) \circ \varphi(b).$$

What this is saying in words is the following. We have a multiplication in G, a multiplication in H and a function φ from G to H. Starting off with two elements in G, if we first multiply them in G, and then apply φ to their product, we should obtain the same element in H if we had first mapped the elements into H using φ and then multiplied their images in H.

Examples 1.37.

1. Let $G = (\mathbb{R}, +)$ and $H = (\mathbb{R}^+, \times)$. The exponential function from G to H, given by $f(x) = 2^x$, is a homomorphism. To see this, let $x, y \in G$. Then

$$f(x + y) = 2^{x+y} = 2^x \times 2^y = f(x) \times f(y)$$

2. Let G be the group $GL(2, \mathbb{R})$ with matrix multiplication, and H be the group of non-zero real numbers with multiplication. Then the determinant function $\det\colon G \to H$ is a homomorphism, since for all 2×2 real matrices A, B we have $\det(AB) = \det(A)\det(B)$. This function is well-defined, since invertible functions have non-zero determinant.

3. Let G be the group $C(\mathbb{R})$ of continuous functions on $\mathbb{R}$ with addition, and H be the group $\mathbb{R}$ with addition. Then the function $f \mapsto f(1)$ is a homomorphism. Indeed, if f, g are continuous functions on $\mathbb{R}$, then $(f + g)(1) = f(1) + g(1)$, by the definition of $f + g$.

4. If G is a group, then the identity map $\iota\colon G \to G$ defined by $\iota(a) = a$ for all $a \in G$, and the trivial map $\zeta\colon G \to G$ defined by $\zeta(a) = e$ for all $a \in G$ are homomorphisms.

Thus a homomorphism is a function between two groups that respects the binary operation. In the next theorem we show that it preserves the identity element and the inverses of elements as well.

Theorem 1.38. *Let G be a group with identity 1_G and H be a group with identity 1_H. If $\varphi\colon G \to H$ is a homomorphism, then:*

1. $\varphi(1_G) = 1_H$.
2. *If $a \in G$, then $(\varphi(a))^{-1} = \varphi(a^{-1})$.*

Proof. We have

$$1_H \circ \varphi(e) = \varphi(e) = \varphi(e * e) = \varphi(e) \circ \varphi(e),$$

and so cancelling $\varphi(e)$ on both sides, we obtain $1_H = \varphi(e)$. Next,

$$\varphi(a^{-1}) *' \varphi(a) = \varphi(a^{-1} * a) = \varphi(e) = 1_H$$

and similarly,

$$1_H = \varphi(e) = \varphi(a * a^{-1}) = \varphi(a) \circ \varphi(a^{-1}).$$

Thus $\varphi(a^{-1}) \circ \varphi(a) = 1_H = \varphi(a) \circ \varphi(a^{-1})$, and by the uniqueness of the inverse of $\varphi(a)$ in H, we obtain $(\varphi(a))^{-1} = \varphi(a^{-1})$. □

Warning 1.39. When working with a homomorphism say $\varphi : G \to H$, pay careful attention to **where** any multiplication is taking place. The multiplication $g * h$ is taking place in G whereas the multiplication $\varphi(g) \circ \varphi(h)$ is taking place in H. For this reason we try to use different symbols for the group multiplication.

Every group homomorphism φ determines two important subsets, which we now define.

Definition 1.40. Let G, H be groups and let $\varphi\colon (G, *) \to (H, \circ)$ be a group homomorphism.

1. The **kernel** of φ is the set $\ker(\varphi) = \{g \in G \mid \varphi(g) = 1_H\}$.
2. The **image** of φ is the set $\mathrm{im}(\varphi) = \{h \in H \mid \exists g \in G \text{ such that } \varphi(g) = h\}$.

Activity 1.41. *Let $G = (\mathbb{Z}, +)$ and let $n \in \mathbb{N}$. Let $\varphi : G \to G$ be the function given by $\varphi(x) = xn$. Prove that φ is a homomorphism. What is the kernel of φ? What is the image of φ?*

It is no accident that in the examples we have given that the kernel of a homomorphism is a subgroup of the original group. Using Theorem 1.38, we now prove the following more general result.

Theorem 1.42. *Let $(G, *)$, $(H, \circ)$ be groups and let $\varphi\colon G \to H$ be a group homomorphism. Then:*

1. $\ker(\varphi)$ *is a subgroup of G.*
2. $\mathrm{im}(\varphi)$ *is a subgroup of H.*

Proof. 1. We shall use the Subgroup Criterion to prove that $\ker(\varphi)$ is a subgroup. First, since 1_G itself belongs to $\ker(\varphi)$, this set is non-empty. Let $a, b \in \ker(\varphi)$ be arbitrary. Our aim is to show that $a * b^{-1} \in \ker(\varphi)$, that is, that $\varphi(a * b^{-1}) = 1_H$. Since φ is a homomorphism, $\varphi(a * b^{-1}) = \varphi(a) \circ \varphi(b^{-1})$. Also, as we saw earlier, $\varphi(b^{-1}) = \varphi(b)^{-1}$. Hence

$$\varphi(a * b^{-1}) = \varphi(a) \circ \varphi(b^{-1}) = \varphi(a) \circ \varphi(b)^{-1} = 1_H * 1_H^{-1} = 1_H$$

2. We now also check that $\mathrm{im}(\varphi)$ is a subgroup of H, again using the Subgroup Criterion. Consider the subset $\mathrm{im}(\varphi)$ of H. This is non-empty, since $\varphi(1_G) = 1_H \in H$. Let h_1, h_1 belong to $\mathrm{im}(\varphi)$, then there exist elements g_1, g_2 in G such that $\varphi(g_1) = h_1$ and $\varphi(g_2) = h_2$. Then, $\varphi(g_2^{-1}) = \varphi(g_2)^{-1} = h_2^{-1}$. Consequently, $\varphi(g_1 * g_2^{-1}) = \varphi(g_1) \circ \varphi(g_2^{-1}) = h_1 h_2^{-1}$, and so there exists an element in G, namely $g_1 * g_2^{-1}$, such that $\varphi(g_1 * g_2^{-1}) = h_1 \circ h_2^{-1}$. In other words, $h_1 \circ h_2^{-1} \in \mathrm{im}(\varphi)$. Thus $\mathrm{im}(\varphi)$ is a subgroup of H.

□

Examples 1.43.

1. Let $G = (\mathbb{R}, +)$ and let $H = (\mathbb{R}^+, \times)$. The exponential function from G to H, given by $x \mapsto 2^x$, is a homomorphism with kernel the trivial subgroup of G comprising the element 0, and image the whole group of positive reals with multiplication. To see this, it can be shown that given any $y > 0$, there exists a unique real number (called the *logarithm of y to the base* 2, denoted by $\log_2 y$) such that $y = 2^{\log_2 y}$.

2. Let G be the group $GL(2, \mathbb{R})$ with matrix multiplication, and $H = (\mathbb{R}^*, \times)$. Then the determinant function $\det\colon G \to H$ is a homomorphism. Its kernel is the set of all invertible matrices with determinant equal to 1. This is the group $SL(2, \mathbb{R})$ which we defined previously. The image of this homomorphism is the whole group of non-zero reals: indeed, given any real number a not equal to zero, we have that

$$A := \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \in GL(2, \mathbb{R}),$$

and $\det(A) = a$.

3. Let G be the group $C(\mathbb{R})$ of continuous functions with addition on $\mathbb{R}$, as before. Let H be the group $\mathbb{R}$ with addition. Then the function $f \mapsto f(1)$ is a homomorphism, and its kernel is the set of all continuous functions on $\mathbb{R}$ that have a zero at $x = 1$ (for instance $f(x) = x - 1$ belongs to the kernel). The image is the set of all real numbers, since given any $a \in \mathbb{R}$, the constant function $f(x) = a$ for all $x \in \mathbb{R}$ is continuous, and $f(1) = a$.

In Example 1 above, the homomorphism between the two groups was also bijective. We give a special name to such homomorphisms.

Definition 1.44. Let G, H be groups. A homomorphism $\varphi\colon G \to H$ is said to be an **isomorphism** if it is bijective.

Examples 1.45.

1. Let G be $\mathbb{R}$ with addition, and H be the set of positive reals with multiplication. The exponential function from G to H, given by $x \mapsto 2^x$, is an isomorphism.

2. If G is a group, then the identity function $\iota\colon G \to G$ defined by $\iota(a) = a$ for all $a \in G$ is an isomorphism.

3. Let G be the subgroup of $GL(2, \mathbb{R})$ comprising all matrices of the form $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$, where $x \in \mathbb{R}$. Let H be the group $\mathbb{R}$ with addition. Then the function from G to H, given by $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \mapsto x$, is an isomorphism.

Activity 1.46. *Let $G = (\mathbb{R}, *)$ where $*$ is defined as $a * b = a + b + 1$ for all $a, b \in G$. Show that G and $(\mathbb{R}, +)$ are isomorphic. (Hint: in order to determine what the homomorphism should be, try to determine how to map the identity of G to the identity of $\mathbb{R}$).*

Isomorphisms are important because their existence between two groups means that the two groups are essentially the "same", in the sense that as far as algebraic properties go, there is no real difference between them. It might also be the case that given several isomorphic versions of a group, one might be much easier to work with than the rest. Which of the two groups in Example 4.45.3 would you rather work with?

Definition 1.47. Two groups G and H are called **isomorphic** if there exists an isomorphism $\varphi\colon G \to H$.

1.8 Cosets and Lagrange's theorem

The main goal of this section is to prove a fundamental and very useful result in Group Theory due to Lagrange[2]. In order to get there, we introduce the concept of cosets. These turn out to be something that we have seen before, without us realising it. We begin with the definition:

Definition 1.48. Let $(G, *)$ be a group and let H be a subgroup of G. The **left cosets** of H are the sets

$$g * H = \{g * h \mid h \in H\}$$

as g runs through all elements in G. Similarly, the **right cosets** of H are the sets

$$H * g = \{h * g \mid h \in H\}$$

as g runs through all elements in G.

In general, properties of the right cosets are more or less the same as those of the left cosets (as you might expect) hence we shall mostly refer only to the left cosets from now on. We shall eventually prove that the left (respectively right) cosets of a subgroup H of a group G form a partition of the group G. In the definition given above, the element g is in general not unique. It is entirely possible (as we shall see) that there may be two different elements of G, say g and g' such that $gH = g'H$. Let us see a concrete example for a small group for which we can write down all of the elements quite easily.

Example 1.49. Consider the symmetric group S_3. We shall consider the left cosets of the subgroups $H = \langle\alpha\rangle$ and $K = \langle\varepsilon\rangle$. These can easily be read off from the group table. Recall first that as a set:

$$S_3 = \{\iota, \alpha, \varepsilon, \varepsilon^2, \alpha\varepsilon, \varepsilon\alpha\}$$

The left cosets of H are the **sets**:

$$H = \{\iota, \alpha\}, \quad \varepsilon H = \{\varepsilon\iota, \varepsilon\alpha\} = \{\varepsilon, \varepsilon\alpha\}, \quad \varepsilon^2 H = \{\varepsilon^2\iota, \varepsilon^2\alpha\} = \{\varepsilon^2, \alpha\varepsilon\}$$

whereas the right cosets of H are the **sets**:

$$H = \{\iota, \alpha\}, \quad H\varepsilon = \{\iota\varepsilon, \alpha\varepsilon\} = \{\varepsilon, \alpha\varepsilon\}, \quad H\varepsilon^2 = \{\iota\varepsilon^2, \alpha\varepsilon^2\} = \{\varepsilon^2, \varepsilon\alpha\}.$$

In both cases we see that a partition of S_3 is obtained, however $\varepsilon H \neq H\varepsilon$ and $\varepsilon^2 H \neq H\varepsilon^2$.

Similarly, the left cosets of K are:

$$K = \{\iota, \varepsilon, \varepsilon^2\}, \quad \alpha K = \{\alpha\iota, \varepsilon\alpha, \varepsilon^2\alpha\} = \{\alpha, \varepsilon\alpha, \alpha\varepsilon\}.$$

Note that in this case the right cosets are identical to the left cosets, and this is always true in general for subgroups whose order is half the order of the group (why?).

[2] Joseph Louis Lagrange (1736–1813)

Another important example is one we have seen before: the construction of $\mathbb{Z}_n$ from $\mathbb{Z}$. Let us recall how that construction went using our new context.

Example 1.50. Let $G = (\mathbb{Z}, +)$ and let $H = (4\mathbb{Z}, +)$ where $4\mathbb{Z}$ is the set of integer multiples of 4. Make sure that you are happy that $(4\mathbb{Z}, +)$ really is a subgroup before we continue. What are the left cosets of H? Of course we have the trivial coset H itself:

$$H = \{0 + h \mid h \in H\}.$$

But, as in the definition, there is nothing special about 0. We could have taken any other element in H, such as 16:

$$16 + H = \{16 + h \mid h \in H\} = H.$$

What about the coset $1 + H$?

$$1 + H = \{1 + h \mid h \in H\} = \{\ldots, 1 + (-8), 1 + (-4), 1 + 0, 1 + 4, 1 + 8, \ldots\}.$$

This is exactly the same as the set we denoted $[1]_4$ before. What are the left cosets of H, then?

$$H, \quad 1 + H, \quad 2 + H, \quad 3 + H.$$

What about the right cosets? Since addition in G is commutative, it should come as no surprise that the right cosets are identical to the left cosets

$$H, \quad H + 1 = 1 + H, \quad H + 2 = 2 + H, \quad H + 3 = 3 + H.$$

This is true in general for subgroups of abelian groups. It is also true for certain subgroups of non-abelian groups.

In the above example, there was nothing special about 4, indeed, the same construction would have worked with any other natural number. We shall usually omit writing the binary operation using cosets hereafter.

Activity 1.51. *In the previous example we asked that $n \geq 2$. What happens if $n = 1$? $n = 0$?*

We now prove Lagrange's Theorem by showing that the left cosets of a group partition the group.

Theorem 1.52 (Lagrange's Theorem)**.** *Let H be a subgroup of a group G and let a, $b \in G$. If aH and bH are two left cosets of H, then either $aH = bH$ or $aH \cap bH = \varnothing$. In other words, the left cosets of H partition the group G. In particular, the order of H divides the order of G.*

Proof. Let G be a group, let H be a subgroup of G and let $a, b \in G$. Suppose that $aH \cap bH \neq \varnothing$ so that there exists $g \in aH \cap bH$. By definition, there exists $h_1, h_2 \in H$ such that $g = ah_1 = bh_2$. Our aim is to show that $aH \subseteq bH$ and $bH \subseteq aH$. Let $ah \in aH$ for some element $h \in H$. Since $a = bh_2h_1^{-1}$ we can write $ah = bh_2h_1^{-1}h$ and so $ah \in bH$, since $h_2h_1^{-1}h \in H$. This shows that $aH \subseteq bH$. The reverse inclusion is similar. □

Remark 1.53. The statement and proof in the case of right cosets is similar.

It is useful to refer to the number of cosets of a given subgroup H in G.

Definition 1.54. Let G be a group and let H be a subgroup of G. The number of left (or right) cosets of H in G is called the **index of H in G** and is written $|G : H|$. It follows from Lagrange's Theorem that $|G| = |H||G : H|$.

We conclude with a number of corollaries of Lagrange's Theorem.

Corollary 1.55. *Let G be a finite group and let $a \in G$. Then $a^{|G|} = e$. In particular, the order of a divides the order of G.*

Proof. Let a have order m. Recall that $\langle a \rangle$ is a subgroup of G, so Theorem 1.30 implies that $|\langle a \rangle| = m$ divides $|G|$, and so $|G| = m \cdot k$ for some $k \in \mathbb{N}$. Thus also

$$a^{|G|} = a^{m \cdot k} = (a^m)^k = e^k = e.$$

□

The following theorem characterises all groups whose order is a prime number.

Corollary 1.56. *If G is a group with prime order p, then G is cyclic, and $G = \langle a \rangle$ for every $a \in G \setminus \{e\}$.*

Proof. If $a \neq e$, then a has order $m \neq 1$. Since m divides p, and p is prime, it follows that $m = p$ and $\langle a \rangle$ is a subgroup of G of order p. As G itself has order p, it now follows that $G = \langle a \rangle$, and so G is cyclic. □

Activity 1.57. *Show that S_4 does not have an element of order 13. Does S_4 have an element of order 24? What about order 12?*

1.9 Coset action and Normal Subgroups – non-examinable

This section is non-examinable and purely to tie a few threads together for personal interest (should you have it...).

1.9.1 Coset Actions

In the beginning we said that ideally we would like a nice geometric object for our group to act on, and this would hopefully give some intuition for the multiplication. Failing that we would have to try to understand the group abstractly. It turns out that a group can always "act on itself" in a certain sense which we briefly describe now.

With our geometric example, we could encode the action of the group by thinking about where each group element sent the vertices of our regular polygon. More generally, as you will find in MA211, finite (and some infinite) groups can act on a collection of n points with the elements of the group described by where they send each point. In this way, all finite groups can naturally be seen as a subgroup of the set of all permutations on n points - the symmetric group S_n. Indeed, this is a result known as Cayley's Theorem[3], one of the pioneers of group theory.

The most natural collection of things that a group can act on, in some sense, is the set of left (or right) cosets of one of its subgroups. Going back to our example of S_3, if we were to take the three left cosets corresponding to the subgroup of order 2, these would behave in exactly the same way under multiplication as the points 1, 2 and 3 do in our geometric picture. There is a catch though, left cosets must always be multiplied by elements on the left and right cosets must always be multiplied by elements on the right. What happens when the left and right cosets are equal, though?

[3] Arthur Cayley (1821–1895).

1.9.2 Normal Subgroups

Earlier we said that for an abelian group G and a subgroup H of G, the right and left cosets were equal and that for non-abelian groups this happened for certain subgroups. We give a name to subgroups whose left and right cosets are equal.

Definition 1.58. Let G be a group and let H be a subgroup of G. If the left cosets of H are equal to the right cosets of H, we say that H is a **normal subgroup**.

We can immediately say the following:

Theorem 1.59. *Let G be a group. If H is an index 2 subgroup of G, then H is a normal subgroup. If G is an abelian group, then every subgroup of G is a normal subgroup.*

In the case of a group G acting on the cosets of a normal subgroup N, we are able to multiply the cosets on the left or the right and the behaviour of the set of cosets will be identical. This allows us to define a new group from the subgroup N, the **quotient of G by N** and we write this as G/N. What does this quotient group look like?

Theorem 1.60 (The Homomorphism Theorem)**.** *Let G and H be groups and let φ be a homomorphism from G to H. The kernel of φ is a normal subgroup of G. Writing $N = ker(\varphi)$, the image of φ is isomorphic to $G/ker(\varphi)$.*

We finish off this chapter on group theory with an example of a quotient group with which we are already very familiar with.

Example 1.61. Let $G = (\mathbb{Z}, +)$. Any subgroup of G is a normal subgroup, since G is abelian, and can be generated by a natural number $n \in \mathbb{Z}$. Letting $N = n\mathbb{Z}$, the quotient group $(\mathbb{Z}/n\mathbb{Z}, \oplus)$ is precisely the group $\mathbb{Z}_n$.

1.10 Comments on selected Activities

Comment on Activity 1.6. In the case $G = \mathbb{Z}$ the given operation is not a binary operation; indeed, when $a = 0$ its output is not even defined.

In the cases where $G = \mathbb{Q}$ it is clear that closure is satisfied and can easily be checked that associativity holds. An identity element is the element -1; but then 0 does not have an inverse. For an element $a \in \mathbb{Q}$, it can easily be determined that $a^{-1} = \frac{1}{a}$. Hence the third example is the only example which satisfies all the group axioms.

Comment on Activity 1.51. If we take $n = 1$, the subgroup $H = \mathbb{Z}$ and so we are left with only one coset: $\mathbb{Z}$ itself. If we take $n = 0$, the subgroup $H = \{0\}$ and we have one coset for each elements of $\mathbb{Z}$. These two cases are somehow trivial in that that they either leave us where we started, or collapse everything into the identity, hence we tend to ignore them.

1.11 Solutions to selected Activities

Solution to Exercise 1.22. Since 1_H is the identity element of H, it follows that $1_H * 1_H = 1_H$. Since $1_H \in G$, it follows that $1_H * 1_G = 1_H$, and that $1_H^{-1} = 1_H$. Then $1_H * 1_H = 1_H * 1_G$. Multiplying on the left by 1_H^{-1}, we achieve $1_H = 1_G$.

Solution to Exercise 1.28. Recall that the identity element in $GL(2, \mathbb{R})$ is the identity matrix I. Let

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}.$$

We then compute:

$$A^2 = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix},$$

$$A^3 = A^2A = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I.$$

Notice then that A^6 equals the identity matrix and so $4 \leq \text{ord}(A) \leq 6$. We then check that A^4 and A^5 are not the identity matrix, but this is even simpler:

$$A^4 = A^3A = -A, \quad A^5 = A^3A^2 = -A^2$$

neither of which is the identity. Hence A has order 6.

Solution to Exercise 1.29. We shall suppress the binary operation for brevity. There are two cases to consider, either ab has finite order, or it has infinite order. Suppose that ab has finite order $k \in \mathbb{N}$ and consider the order of ba. What is $(ba)^k$? This is:

$$baba \dots ba = b(ab)^{k-1}a = b(ab)^{k-1}a(bb^{-1}) = b(ab)^kb^{-1} = bb^{-1} = 1_G.$$

The order of ba is at most k in that case. Now, for a contradiction, suppose that there exists $l < k$ such that $(ba)^l = 1_G$. Then

$$(ab)^l = a(ba)^la^{-1} = 1_G,$$

contradicting the fact that ab had order greater than l.

If ab did not have finite order, but, for a contradiction, ba had finite order k, then the above argument shows that ab also has order k, a contradiction, and so ba also does not have finite order.

Solution to Exercise 1.35. We have two directions to prove here. If $n = 1$, then $GL(1, \mathbb{Q})$ is the set of (multiplicatively) invertible rational numbers, that is: $\mathbb{Q} \setminus \{0\}$. Hence $GL(1, \mathbb{Q}) = (\mathbb{Q} \setminus \{0\}, \times)$ which is clearly abelian. If $n \geq 2$, from what we know about matrix multiplication, it is sufficient to prove this for $n = 2$, since we can naturally embed a 2×2 matrix into an $n \times n$ in the top left corner and let there be 1s along the rest of the diagonal and 0s elsewhere. Almost any pair of matrices in $GL(2, \mathbb{Q})$ will do, e.g.:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Solution to Exercise 1.41. Let $\varphi : G \to G$ be the map $\varphi(x) = xn$ as given. Then

$$\varphi(x + y) = (x + y)n = xn + yn = \varphi(x) + \varphi(y)$$

hence φ is indeed a homomorphism. The kernel of φ is everything that φ maps to the identity of G, that is 0. The only way that $\varphi(x) = xn = 0$, then, is if $x = 0$. The image of φ is the set $\{xn \mid x \in \mathbb{Z}\}$.

Solution to Exercise 1.46. We let $R = (\mathbb{R}, +)$ for brevity. As the hint suggests, let's see what the identity of G is. We want $1_G * a = a * 1_G = a + 1_G + 1 = a$ and hence $1_G = -1$. Now, let $a, b \in G$. We want $f : G \to R$ such that:

$$f(a) + f(b) = f(a * b) = f(a + b + 1) = a + b + 1 + 1.$$

This suggests that we could let $f(a) = a + 1$. We see that this would indeed be a homomorphism and it is visibly a bijection (of sets) from G to R. Finally, we check that $f(1_G)$ is the identity of R: $f(-1) = -1 + 1 = 0$. Hence f is our desired isomorphism.

Solution to Exercise 1.57. The order of S_4 is $|S_4| = 4! = 24$. Suppose that S_4 had an element g of order 13, then $H = \langle g \rangle$ is a subgroup of S_4 of order 13. By Langrage's Theorem, the order of H should divide the order of S_4, but we see that this is not true. Hence, S_4 does not have an element of order 13.

Could S_4 have an element of order 24? We have seen that $|S_4| = 24$, so if it did, that would imply that S_4 is cyclic, which it is not. Perhaps the easiest way to see that S_4 is not cyclic is to see that it is not even abelian, as we proved in the homework exercises.

What about an element of order 12? This requires some thought and there are a few different methods. Perhaps the easiest is to brute force check the order of every element of S_4, but let's do this in a sophisticated way. I would recommend drawing pictures as you go through this solution. An element $g \in S_4$ can in (principle) fix 0, 1, 2, 3 or 4 elements of $\{1, 2, 3, 4\}$. If g fixes all 4 elements, it is the identity. If g fixes 3 elements...where on earth is it going to send the fourth? Hence for a non-trivial element g, it can really only fix 0, 1 or 2 elements. If g fixes 2 elements, there isn't a lot it can do to the other two apart from swap them and so g has order 2. If g fixes 1 element, it can be thought of as an element of S_3, and we know the orders of elements there: 2 or 3. Finally, if g fixes 0 elements, one of two things can happen. It could behave as follows:

$$g(1) = 2, \quad g(2) = 1, \quad g(3) = 4, \quad g(4) = 3$$

where hopefully it isn't too hard to see that applying g twice gives the identity element. The other thing that could happen is as follows. Suppose without loss of generality that $g(1) = 2$ and $g(2) \neq 1$. Again, without loss of generality, let's say that $g(2) = 3$. If $g(3) = 1$, this would fix 4 and give an element of 3, so it must be that $g(3) = 4$. Finally, all that is left is $g(4) = 1$ and it is not too hard to see that this element has order 4.

This is of course a rather long-winded approach to take, you will be glad to know that if you take MA211, a much simpler method for working with elements in symmetric groups will appear – the cycle notation which we have seen as non-examinable so far.

2

Abstract vector spaces

- Anthony, M. and Harvey, M. *Linear Algebra: Concepts and Methods*. Cambridge University Press 2012. Chapters 5, 6 and Sections 7.1 and 7.2 of Chapter 7.

2.1 Introduction

In this chapter we introduce our second important example of an algebraic structure, called a vector space. Roughly speaking, a vector space is a set of elements called "vectors", where any two vectors can be "added", resulting in a new vector, and any vector can be multiplied by an element from $\mathbb{R}$ so as to give a new vector. This is a generalisation of the usual idea of the ordinary n-dimensional vector space $\mathbb{R}^n$, because many other sets also have a natural addition and multiplication with reals, for example sets of matrices, sets of functions, and sets of sequences. We start off with a precise definition, which will allow all these examples to be vector spaces.

2.2 The definition of a vector space

Definition 2.1. A **vector space** V is a set together with two functions, $+ \colon V \times V \to V$, called **vector addition**, and $\cdot \colon \mathbb{R} \times V \to V$, called **scalar multiplication**, such that $(V, +)$ is an abelian group and the following three properties hold:

Multiplicative identity: For all $\boldsymbol{v} \in V$, $1 \cdot \boldsymbol{v} = \boldsymbol{v}$.

Associativity of scalar multiplication: For all $\alpha, \beta \in \mathbb{R}$ and all $\boldsymbol{v} \in V$, $\alpha \cdot (\beta \cdot \boldsymbol{v}) = (\alpha\beta) \cdot \boldsymbol{v}$.

Distributivity: For all $\alpha, \beta \in \mathbb{R}$ and all $\boldsymbol{v} \in V$, $(\alpha + \beta) \cdot \boldsymbol{v} = \alpha \cdot \boldsymbol{v} + \beta \cdot \boldsymbol{v}$,
and for all $\alpha \in \mathbb{R}$ and all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\alpha \cdot (\boldsymbol{v}_1 + \boldsymbol{v}_2) = \alpha \cdot \boldsymbol{v}_1 + \alpha \cdot \boldsymbol{v}_2$.

The elements of a vector space are called **vectors**.

Since the above definition requires $(V, +)$ to be an abelian group, a vector space also has the following five properties built into its definition: (note the use of additive notation)

Closure: For all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\boldsymbol{v}_1 + \boldsymbol{v}_2 \in V$.

Associativity: For all $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in V$, $\boldsymbol{v}_1 + (\boldsymbol{v}_2 + \boldsymbol{v}_3) = (\boldsymbol{v}_1 + \boldsymbol{v}_2) + \boldsymbol{v}_3$.

Identity: There exists an element $\boldsymbol{o} \in V$ (called the[1] **zero vector**) such that for all $\boldsymbol{v} \in V$, $\boldsymbol{v} + \boldsymbol{o} = \boldsymbol{v} = \boldsymbol{o} + \boldsymbol{v}$.

[1] Since there is a unique identity element in a group, the zero vector is unique.

Inverses: For every $\boldsymbol{v} \in V$, there exists a unique[2] element in V, denoted by $-\boldsymbol{v}$, such that $\boldsymbol{v} + (-\boldsymbol{v}) = \boldsymbol{o} = -\boldsymbol{v} + \boldsymbol{v}$.

Commutativity: For all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{v}_2 + \boldsymbol{v}_1$.

Examples 2.2.

1. Let $m, n \in \mathbb{N}$, and consider the set $\mathbb{R}^{m\times n}$ of $m \times n$ matrices having real entries. In the previous Chapter, we showed that this is a group, to see that it is abelian immediately follows from the commutativity of addition in $\mathbb{R}$.

   We next define scalar multiplication as follows: if $\alpha \in \mathbb{R}$ and

   $$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n},$$

   then

   $$\alpha \cdot A = \begin{pmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{pmatrix}. \tag{2.1}$$

   Thus $\alpha \cdot A \in \mathbb{R}^{m\times n}$, and moreover the vector space axioms are satisfied:

   Multiplicative identity: If $A \in \mathbb{R}^{m\times n}$, then by (2.1)

   $$\begin{aligned} 1 \cdot A = 1 \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} &= \begin{pmatrix} 1a_{11} & \dots & 1a_{1n} \\ \vdots & & \vdots \\ 1a_{m1} & \dots & 1a_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = A. \end{aligned}$$

   Associativity of scalar multiplication: For all $\alpha, \beta \in \mathbb{R}$ and all $A \in \mathbb{R}^{m\times n}$, using (2.1) three times, as well as the associativity of multiplication of real numbers,

   $$\begin{aligned} \alpha \cdot (\beta \cdot A) = \alpha \cdot \left(\beta \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \right) &= \alpha \cdot \begin{pmatrix} \beta a_{11} & \dots & \beta a_{1n} \\ \vdots & & \vdots \\ \beta a_{m1} & \dots & \beta a_{mn} \end{pmatrix} \\ = \begin{pmatrix} \alpha(\beta a_{11}) & \dots & \alpha(\beta a_{1n}) \\ \vdots & & \vdots \\ \alpha(\beta a_{m1}) & \dots & \alpha(\beta a_{mn}) \end{pmatrix} &= \begin{pmatrix} (\alpha\beta) a_{11} & \dots & (\alpha\beta) a_{1n} \\ \vdots & & \vdots \\ (\alpha\beta) a_{m1} & \dots & (\alpha\beta) a_{mn} \end{pmatrix} \\ = (\alpha\beta) \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} &= (\alpha\beta) \cdot A. \end{aligned}$$

   Distributivity: There are two identities to check, and for both we repeatedly use (2.1), as well as the distributive laws for $\mathbb{R}$.

[2] Recall that in a group, every element has a unique inverse.

First, for all $\alpha, \beta \in \mathbb{R}$ and all $A \in \mathbb{R}^{m\times n}$,

$$\begin{aligned}
(\alpha+\beta)\cdot A &= (\alpha+\beta)\cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} (\alpha+\beta)a_{11} & \dots & (\alpha+\beta)a_{1n} \\ \vdots & & \vdots \\ (\alpha+\beta)a_{m1} & \dots & (\alpha+\beta)a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} \alpha a_{11}+\beta a_{11} & \dots & \alpha a_{1n}+\beta a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1}+\beta a_{m1} & \dots & \alpha a_{mn}+\beta a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{pmatrix} + \begin{pmatrix} \beta a_{11} & \dots & \beta a_{1n} \\ \vdots & & \vdots \\ \beta a_{m1} & \dots & \beta a_{mn} \end{pmatrix} \\
&= \alpha\cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \beta\cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \\
&= \alpha\cdot A + \beta\cdot A.
\end{aligned}$$

Also, for all $\alpha \in \mathbb{R}$ and all $A, B \in \mathbb{R}^{m\times n}$,

$$\begin{aligned}
\alpha\cdot(A+B) &= \alpha\cdot\left(\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix}\right) \\
&= \alpha\cdot \begin{pmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \dots & a_{mn}+b_{mn} \end{pmatrix} \\
&= \begin{pmatrix} \alpha(a_{11}+b_{11}) & \dots & \alpha(a_{1n}+b_{1n}) \\ \vdots & & \vdots \\ \alpha(a_{m1}+b_{m1}) & \dots & \alpha(a_{mn}+b_{mn}) \end{pmatrix} \\
&= \begin{pmatrix} \alpha a_{11}+\alpha b_{11} & \dots & \alpha a_{1n}+\alpha b_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1}+\alpha b_{m1} & \dots & \alpha a_{mn}+\alpha b_{mn} \end{pmatrix} \\
&= \begin{pmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{pmatrix} + \begin{pmatrix} \alpha b_{11} & \dots & \alpha b_{1n} \\ \vdots & & \vdots \\ \alpha b_{m1} & \dots & \alpha b_{mn} \end{pmatrix} \\
&= \alpha\cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \alpha\cdot \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} \\
&= \alpha\cdot A + \alpha\cdot B.
\end{aligned}$$

Hence $\mathbb{R}^{m\times n}$ is a vector space with matrix addition and with scalar multiplication defined by (2.1). If $n = 1$, then we denote the vector space of column vectors $\mathbb{R}^{m\times 1}$ by $\mathbb{R}^m$.

2. We already know that the set $C(\mathbb{R})$ of continuous functions $\mathbb{R} \to \mathbb{R}$ with addition of functions is an abelian group (Example 4.34.4). Let scalar multiplication be defined as

follows:

$$\text{if } \alpha \in \mathbb{R} \text{ and } f \in C(\mathbb{R}), \text{ then } (\alpha \cdot f)(x) = \alpha f(x), \quad x \in \mathbb{R}. \tag{2.2}$$

(We say that scalar multiplication, and addition, are defined "pointwise".) Then $\alpha \cdot f \in C(\mathbb{R})$, and moreover, the vector space axioms are satisfied:

Multiplicative identity: Let $f \in C(\mathbb{R})$. For all $x \in \mathbb{R}$, we have

$$(1 \cdot f)(x) = 1f(x) = f(x),$$

and so $1 \cdot f = f$.

Associativity of scalar multiplication: Let $\alpha, \beta \in \mathbb{R}$ and $f \in C(\mathbb{R})$. For all $x \in \mathbb{R}$, we have

$$\begin{aligned}(\alpha \cdot (\beta \cdot f))(x) &= \alpha(\beta \cdot f)(x) \\ &= \alpha(\beta f(x)) \\ &= (\alpha\beta)f(x) \\ &= ((\alpha\beta) \cdot f)(x),\end{aligned}$$

and so $(\alpha \cdot (\beta \cdot f)) = (\alpha\beta) \cdot f$.

Distributivity: There are two identities to check. First, let $\alpha, \beta \in \mathbb{R}$ and $f \in C(\mathbb{R})$. For all $x \in \mathbb{R}$, we have

$$\begin{aligned}((\alpha + \beta) \cdot f)(x) &= (\alpha + \beta)f(x) \\ &= \alpha f(x) + \beta f(x) \\ &= (\alpha \cdot f)(x) + (\beta \cdot f)(x) \\ &= (\alpha \cdot f + \beta \cdot f)(x),\end{aligned}$$

and so $(\alpha + \beta) \cdot f = \alpha \cdot f + \beta \cdot f$.

Second, let $\alpha \in \mathbb{R}$ and $f, g \in C(\mathbb{R})$. For all $x \in \mathbb{R}$, we have

$$\begin{aligned}(\alpha \cdot (f + g))(x) &= \alpha(f + g)(x) \\ &= \alpha(f(x) + g(x)) \\ &= \alpha f(x) + \alpha g(x) \\ &= (\alpha \cdot f)(x) + (\alpha \cdot g)(x) \\ &= (\alpha \cdot f + \alpha \cdot g)(x),\end{aligned}$$

and so $\alpha \cdot (f + g) = \alpha \cdot f + \alpha \cdot g$.

Hence $C(\mathbb{R})$ with addition and scalar multiplication is a vector space, called the **vector space of continuous functions** on $\mathbb{R}$.

Activity 2.3. *Show that the empty set is not a vector space, despite (vacuously) satisfying all three of the vector space properties: multiplicative identity, associativity of scalar multiplication and distributivity.*

We now prove a few elementary properties of vector spaces.

Theorem 2.4. *Let V be a vector space. Then the following hold:*

1. *For all $\boldsymbol{v} \in V$, $0 \cdot \boldsymbol{v} = \boldsymbol{o}$.*

2. *For all* $\alpha \in \mathbb{R}$, $\alpha \cdot \boldsymbol{o} = \boldsymbol{o}$.

3. *If* $\boldsymbol{v} \in V$, *then* $(-1) \cdot \boldsymbol{v} = -\boldsymbol{v}$.

Proof.

1. To see this, we use the distributive law to write

$$0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} = (0 + 0) \cdot \boldsymbol{v} = 0 \cdot \boldsymbol{v}.$$

   Now add $-(0 \cdot \boldsymbol{v})$ on both sides, to obtain $0 \cdot \boldsymbol{v} = \boldsymbol{o}$.

2. Similarly, $\alpha \cdot \boldsymbol{o} + \alpha \cdot \boldsymbol{o} = \alpha \cdot (\boldsymbol{o} + \boldsymbol{o}) = \alpha \cdot \boldsymbol{o}$, and hence $\alpha \cdot \boldsymbol{o} = \boldsymbol{o}$.

3. Finally, we have

$$\begin{aligned} \boldsymbol{v} + (-1) \cdot \boldsymbol{v} &= 1 \cdot \boldsymbol{v} + (-1) \cdot \boldsymbol{v} \text{ (since } 1 \cdot \boldsymbol{v} = \boldsymbol{v}) \\ &= (1 + (-1)) \cdot \boldsymbol{v} \text{ (distributive law)} \\ &= 0 \cdot \boldsymbol{v} \text{ (since } 1 + (-1) = 0) \\ &= \boldsymbol{o} \text{ (since } 0 \cdot \boldsymbol{v} = \boldsymbol{o}), \end{aligned}$$

   and so $\boldsymbol{v} + (-1) \cdot \boldsymbol{v} = \boldsymbol{o} = (-1) \cdot \boldsymbol{v} + \boldsymbol{v}$. Hence $(-1) \cdot \boldsymbol{v}$ is an inverse of $\boldsymbol{v}$. But the inverse of an element from a group is unique, and so $(-1) \cdot \boldsymbol{v} = -\boldsymbol{v}$. □

Activity 2.5.

Let U *and* V *be vector spaces, and define*

$$U \times V = \{(u, v) \mid \boldsymbol{u} \in U, \boldsymbol{v} \in V\}.$$

Show that $U \times V$, *with addition defined by* $(\boldsymbol{u}_1, \boldsymbol{v}_1) + (\boldsymbol{u}_2, \boldsymbol{v}_2) = (\boldsymbol{u}_1 + \boldsymbol{u}_2, \boldsymbol{v}_1 + \boldsymbol{v}_2)$ *and scalar multiplication by* $\alpha \cdot (\boldsymbol{u}, \boldsymbol{v}) = (\alpha \cdot \boldsymbol{u}, \alpha \cdot \boldsymbol{v})$, *is a vector space.*

2.3 Subspaces

Definition 2.6. Let V be a vector space. A subset U is called a *subspace of* V if U is a vector space with the vector addition and scalar multiplication of V restricted to U.

It is very cumbersome to check all the axioms to verify that a subset of V is a subspace. Fortunately, we have an analogous Criterion to the Subgroup Criterion in the previous Chapter. The following theorem gives a shortcut for checking whether a subset of V is a subspace.

Theorem 2.7 (The Subspace Criterion)**.** *Let* V *be a vector space and let* U *be a subset of* V. *Then* U *is a subspace of* V *if and only if* U *is a non-empty subset of* V *and for all* $\alpha, \beta \in \mathbb{R}$ *and* $\boldsymbol{u}, \boldsymbol{v} \in U$, $\alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v} \in U$.

Proof. Suppose first that U is a subspace of V. Given $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{v} \in U$, then clearly $\alpha \cdot \boldsymbol{u}$ and $\beta \cdot \boldsymbol{v}$ are in U, because U is a vector space, and for the same reason their sum $\alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v} \in U$.

Conversely, suppose that U is a non-empty subset of V and for all $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{v} \in U$, $\alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v} \in U$. We then have to check that U is an abelian group, as well as satisfying the three additional properties in the definition of a vector space.

Abelian group: To show that U is a group, we use the Subgroup Criterion (Theorem 1.24). Let $\boldsymbol{u}, \boldsymbol{v} \in U$, which exists since U is non-empty. By Theorem 1.24, we have to show that $\boldsymbol{u} - \boldsymbol{v} \in U$. However, this follows from the assumption, since $\boldsymbol{u} - \boldsymbol{v} = 1 \cdot \boldsymbol{u} + (-1) \cdot \boldsymbol{v}$, by the

Multiplicative identity in V and Theorem 2.4. Since $(V,+)$ is an abelian group, it follows that the subgroup U is also abelian.

Multiplicative identity: Let $\boldsymbol{u} \in U$. Since then $\boldsymbol{u} \in V$, it follows from the Multiplicative identity in V that $1 \cdot \boldsymbol{u} = \boldsymbol{u}$.

Associativity of scalar multiplication: Since this holds in V, we have that $\alpha \cdot (\beta \cdot \boldsymbol{v}) = (\alpha\beta) \cdot \boldsymbol{v}$ for all $\alpha, \beta \in \mathbb{R}$ and all $\boldsymbol{v} \in V$. It then obviously holds for all $\boldsymbol{v}$ in the subset U.

Distributivity. Similarly, since we already have that Distributivity holds in V, it will also hold for all vectors in the subset U. □

Two important subspaces of a vector space V are the subspaces $\{\boldsymbol{o}\}$ and V itself. This leads to the following definition.

Definition 2.8. Let V be a vector space. If a subspace U of V is neither $\{\boldsymbol{o}\}$ nor V, then it is called a **proper subspace** of V.

Examples 2.9.

1. Consider the vector space $\mathbb{R}^{2\times 2}$ with matrix addition and scalar multiplication defined by (2.1). Then the set of **upper triangular matrices**

$$U_1 = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \,\middle|\, a, b, d \in \mathbb{R} \right\}$$

   is a subspace of $\mathbb{R}^{2\times 2}$.

2. Consider the set of all polynomial functions

$$P(\mathbb{R}) = \left\{ p\colon \mathbb{R} \to \mathbb{R} \,\middle|\, \begin{array}{l} \exists n \in \mathbb{N} \cup \{0\} \text{ and} \\ a_0, a_1, a_2, \ldots, a_n \in \mathbb{R} \text{ such that} \\ p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \\ \text{for all } x \in \mathbb{R} \end{array} \right\}.$$

   Then $U = P(\mathbb{R})$ is a subspace of the vector space $C(\mathbb{R})$ with addition of functions and scalar multiplication defined by (2.2).

3. Let $n \in \mathbb{N} \cup \{0\}$. Consider the set of all polynomial functions of degree at most n:

$$P_n(\mathbb{R}) = \left\{ p\colon \mathbb{R} \to \mathbb{R} \,\middle|\, \begin{array}{l} \exists a_0, a_1, a_2, \ldots, a_n \in \mathbb{R} \text{ such that} \\ p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \\ \text{for all } x \in \mathbb{R} \end{array} \right\}.$$

   Then $U = P_n(\mathbb{R})$ is a subspace of the vector space $P(\mathbb{R})$, as well as being a subspace of the vector space of all continuous functions $C(\mathbb{R})$.

Activity 2.10. *Prove that the set of* ***symmetric matrices*** *with usual matrix addition and multiplication*

$$U_2 = \left\{ \begin{pmatrix} a & b \\ b & d \end{pmatrix} \,\middle|\, a, b, d \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^{2\times 2}$.

2.4 Linear combinations

Definitions 2.11. Let V be a vector space.

1. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are vectors in V and $\alpha_1, \ldots, \alpha_n$ belong to $\mathbb{R}$, then the vector $\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n$ is called a **linear combination of the vectors** $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.

2. Let S be a non-empty subset of a vector space V. The **span** of S, denoted by $\operatorname{lin}(S)$, is defined as the set of all possible linear combinations[3] of vectors from S:

$$\operatorname{lin}(S) = \{\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n \mid n \in \mathbb{N},\ \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in S,\ \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}\,.$$

   We also define the **span of the empty set** as $\operatorname{lin}(\varnothing) = \{\boldsymbol{o}\}$.

The term **linear span** of S is also used instead of span.

Examples 2.12.

1. Let $m \in \mathbb{N}$. Any vector in the vector space $\mathbb{R}^m$ is a linear combination of the vectors

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \boldsymbol{e}_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

   Hence $\operatorname{lin}(\{\boldsymbol{e}_1, \boldsymbol{e}_2 \ldots, \boldsymbol{e}_m\}) = \mathbb{R}^m$.

2. Let $f_0(x) = 1$ for all $x \in \mathbb{R}$, and for each $k \in \mathbb{N}$, let $f_k(x) = x^k$ for all $x \in \mathbb{R}$. Then any polynomial of degree at most n is a linear combination of $f_0, f_1, \ldots, f_n$. Hence $P_n(\mathbb{R}) = \operatorname{lin}(\{f_0, f_1, \ldots, f_n\})$.

   Furthermore, any polynomial in $P(\mathbb{R})$ is a linear combination of polynomials in the infinite set $\{f_k \mid k = 0, 1, 2, \ldots\}$, so it follows that $P(\mathbb{R}) = \operatorname{lin}(\{f_k \mid k = 0, 1, 2, \ldots\})$.

The span of a set S of vectors from the vector space V turns out to be a special subspace of the vector space V.

Theorem 2.13. *Let V be a vector space and S be a subset of V. Then $\operatorname{lin}(S)$ is a subspace of V containing S. Moreover, $\operatorname{lin}(S)$ is the smallest subspace of V that contains S in the following sense: For any subspace U of V, if $S \subseteq U$, then $\operatorname{lin}(S) \subseteq U$.*

Proof. The result is true in the case where $S = \varnothing$, as $\operatorname{lin}(\varnothing) = \{\boldsymbol{o}\}$ contains the empty set and is the smallest subspace of V.

Suppose that S is non-empty. Then we have for any $\boldsymbol{v} \in S$ that $\boldsymbol{v} = 1 \cdot \boldsymbol{v}$, so is a linear combination of elements from S, hence $\boldsymbol{v} \in \operatorname{lin}(S)$. This shows that $S \subseteq \operatorname{lin}(S)$.

We next show that $\operatorname{lin}(S)$ is a subspace of V by using the Subspace Criterion Theorem 2.7. Let $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{v} \in \operatorname{lin}(S)$. Then we know that

$$\boldsymbol{u} = \alpha_1 \cdot \boldsymbol{u}_1 + \cdots + \alpha_n \cdot \boldsymbol{u}_n \text{ and } \boldsymbol{v} = \beta_1 \cdot \boldsymbol{v}_1 + \cdots + \beta_m \cdot \boldsymbol{v}_m$$

for some vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in S$ and scalars $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m \in \mathbb{R}$. Consequently,

$$\begin{aligned} \alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v} &= \alpha \cdot (\alpha_1 \cdot \boldsymbol{u}_1 + \cdots + \alpha_n \cdot \boldsymbol{u}_n) + \beta \cdot (\beta_1 \cdot \boldsymbol{v}_1 + \cdots + \beta_m \cdot \boldsymbol{v}_m) \\ &= \alpha \cdot (\alpha_1 \cdot \boldsymbol{u}_1) + \cdots + \alpha \cdot (\alpha_n \cdot \boldsymbol{u}_n) + \beta \cdot (\beta_1 \cdot \boldsymbol{v}_1) + \cdots + \beta \cdot (\beta_m \cdot \boldsymbol{v}_m) \\ &= (\alpha\alpha_1) \cdot \boldsymbol{u}_1 + \cdots + (\alpha\alpha_n) \cdot \boldsymbol{u}_n + (\beta\beta_1) \cdot \boldsymbol{v}_1 + \cdots + (\beta\beta_m) \cdot \boldsymbol{v}_m \in \operatorname{lin}(S), \end{aligned}$$

[3] Note that although S might be infinite, a linear combination, by definition, is always a linear combination of a *finite* set of vectors from S.

where we have used Distributivity and Associativity of scalar multiplication. By the Subspace Criterion it follows that $\text{lin}(S)$ is a subspace of V.

Finally, if U is a subspace of V that contains S, then for any linear combination $\alpha_1 \cdot \boldsymbol{v}_1 + \alpha_2 \cdot \boldsymbol{v}_2 + \cdots + \alpha_n \cdot \boldsymbol{v}_n$ of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ from S, since $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are contained in U and U is a vector space, the linear combination $\alpha_1 \cdot \boldsymbol{v}_1 + \alpha_2 \cdot \boldsymbol{v}_2 + \cdots + \alpha_n \cdot \boldsymbol{v}_n$ also belongs to U, and so $\text{lin}(S) \subseteq U$.

Hence $\text{lin}(S)$ is the smallest subspace of V containing S. □

Definition 2.14. Let V be a vector space with subset S. If $V = \text{lin}(S)$, then we say that V is **spanned by** the set S, and also that the set S is a **generating set** of V.

Examples 2.15.

1. Theorem 2.13 gives an alternative way of showing that $P_n(\mathbb{R})$ is a subspace of $P(\mathbb{R})$ and $P(\mathbb{R})$ is a subspace of $C(\mathbb{R})$. Indeed, $P_n(\mathbb{R})$ is spanned by the polynomials $1, x, x^2, \ldots, x^n$, and $P(\mathbb{R})$ has generating set $\{1, x, x^2, \ldots\}$.

2. For any $\boldsymbol{v} \in V$, the set $U = \{\alpha \cdot v \mid \alpha \in \mathbb{R}\}$ is a subspace of V, since clearly $U = \text{lin}(\{\boldsymbol{v}\})$.

2.5 Linear dependence

Definitions 2.16. Let V be a vector space and let $n \geq 1$.

1. The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ are called **linearly dependent** if there exist $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, not all 0,[4] such that $\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n = \boldsymbol{o}$.

2. The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are called **linearly independent** if they are not linearly dependent, that is,

$$\forall \alpha_1, \ldots, \alpha_n \in \mathbb{R}, \text{ if } \alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n = \boldsymbol{o}, \text{ then } \alpha_1 = \cdots = \alpha_n = 0.$$

3. An arbitrary subset S of vectors from V is said to be a **linearly dependent set** if for some $n \geq 1$ there exist distinct vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in S$ that are linearly dependent. An arbitrary subset S of vectors from V is said to be a **linearly independent set** if it is not a linearly dependent set.

Remarks 2.17. These definitions are somewhat subtle, but they are crucial for understanding vector spaces and linear algebra. Here are some important observations.

1. In the definition of linear independence of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$, we allow repetitions. However, whenever there is a repetition in this list of vectors, they are immediately linearly dependent. For example, if $\boldsymbol{v}_1 = \boldsymbol{v}_2$, then $(-1) \cdot \boldsymbol{v}_1 + 1 \cdot \boldsymbol{v}_2 + 0 \cdot \boldsymbol{v}_3 + \cdots + 0 \cdot \boldsymbol{v}_n = \boldsymbol{o}$, so $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are linearly dependent.

   On the other hand, when we consider the linear independence of sets of vectors, there are no repetitions because there are no repetitions in sets.

2. A subset S is linearly independent if and only if for all $n \geq 1$ and all choices of distinct $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in S$, $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are linearly independent.

Examples 2.18.

[4] Note that $\alpha_1, \ldots, \alpha_n$ are not all 0 iff at least one of them is non-zero. Another way of writing this is that $(\alpha_1, \alpha_2, \ldots, \alpha_n) \neq (0, 0, \ldots, 0)$.

1. Let V be a vector space. Then any finite set of vectors from V containing the zero vector is linearly dependent. Indeed if $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ and $\boldsymbol{v}_k = \boldsymbol{o}$, then

$$0 \cdot \boldsymbol{v}_1 + \cdots + 0 \cdot \boldsymbol{v}_{k-1} + 1 \cdot \boldsymbol{v}_k + 0 \cdot \boldsymbol{v}_{k+1} + \cdots + 0 \cdot \boldsymbol{v}_n = \boldsymbol{o}.$$

2. The vectors

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \boldsymbol{e}_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

   are linearly independent in $\mathbb{R}^m$. Indeed if $\alpha_1 \cdot \boldsymbol{e}_1 + \boldsymbol{e}_2 + \cdots + \alpha_m \cdot \boldsymbol{e}_m = \boldsymbol{o}$, then

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} = \alpha_1 \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \alpha_2 \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \alpha_m \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

   and so $\alpha_1 = \cdots = \alpha_m = 0$.

3. Let S_1 and S_2 be subsets of a vector space V with $S_1 \subseteq S_2$. If S_1 is linearly dependent, then S_2 is also linearly dependent. Equivalently, if S_2 is linearly independent, then its subset S_1 is also linearly independent.

4. Consider the polynomials $f_0(x) = 1$ and for $k = 1, \ldots, n$, $f_k(x) = x^k$, $x \in \mathbb{R}$. Then the $n+1$ polynomials $f_0, f_1, \ldots, f_n$ are linearly independent in the vector space $P_n(\mathbb{R})$.

   Indeed, suppose that

$$\alpha_0 \cdot f_0 + \alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \ldots \alpha_n \cdot f_n = \boldsymbol{o},$$

   where $\boldsymbol{o}$ is the zero polynomial $\boldsymbol{o}(x) = 0 \ \forall x \in \mathbb{R}$. This means that for all $x \in \mathbb{R}$ we have

$$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots \alpha_n x^n = 0.$$

   From this it can be shown that $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$.

5. Similarly, the infinite set $\{f_0, f_1, f_2, \ldots\}$ is a linearly independent set in the vector space $P(\mathbb{R})$.

   If it were not, then it would contain a finite subset $\{f_{i_1}, f_{i_2}, \ldots, f_{i_k}\}$ that is linearly dependent (where $1 \leq i_1 < i_2 < \cdots < i_k$). If we include all the missing f_i as well, we obtain that $\{f_0, f_1, f_2, \ldots, f_n\}$ is still linearly dependent, where $n = i_k$. This contradicts the previous example.

6. It follows from the definition that the empty set is not linearly dependent, because in the definition of a linearly dependent set we require the existence of at least one vector in the set. Thus the empty set is a linearly independent subset of any vector space.

It might seem obvious that subspaces of finite-dimensional vector spaces are also finite-dimensional, but proving this for a general, abstract vector space is slightly tricky. (Note that we haven't even defined what it means for a vector space to be finite-dimensional.) A crucial ingredient is the following lemma saying that any linearly dependent set of vectors has a redundant element.

Lemma 2.19 (Redundancy)**.** *Let V be a vector space and let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in V$ be linearly dependent. Then there exists at least one $j \in \{1, 2, \ldots, n\}$ such that*

(a) *$\boldsymbol{v}_j$ is in the span of the previous vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}$, that is, $\boldsymbol{v}_j \in \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1}\})$,*

(b) *and $\boldsymbol{v}_j$ can be removed without changing the span:*

$$\operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}) = \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \setminus \{\boldsymbol{v}_j\}).$$

Proof. Since $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are linearly dependent, there exist $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$, not all zero, such that

$$\alpha_1 \cdot \boldsymbol{v}_1 + \alpha_2 \cdot \boldsymbol{v}_2 + \cdots + \alpha_n \cdot \boldsymbol{v}_n = \boldsymbol{o}.$$

Let j be the last index such that $\alpha_j \neq 0$. Then

$$\alpha_1 \cdot \boldsymbol{v}_1 + \alpha_2 \cdot \boldsymbol{v}_2 + \cdots + \alpha_{j-1} \cdot \boldsymbol{v}_{j-1} + \alpha_j \cdot \boldsymbol{v}_j = \boldsymbol{o},$$

hence

$$\boldsymbol{v}_j = -\frac{\alpha_1}{\alpha_j}\boldsymbol{v}_1 - \cdots - \frac{\alpha_{j-1}}{\alpha_j}\boldsymbol{v}_{j-1}, \tag{2.3}$$

so $\boldsymbol{v}_j \in \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1}\})$, which gives (a). (Note that if $j = 1$, this just means that $\boldsymbol{v}_j = \boldsymbol{o} \in \operatorname{lin}(\varnothing)$.)

To show (b), we first show the inclusion $\operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}) \subseteq \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \cancel{\boldsymbol{v}_j}, \ldots \boldsymbol{v}_n\})$. Let $\boldsymbol{v} \in \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\})$. Then we can write $\boldsymbol{v}$ as a linear combination

$$\boldsymbol{v} = \beta_1 \cdot \boldsymbol{v}_1 + \cdots + \beta_j \cdot \boldsymbol{v}_j + \cdots + \beta_n \cdot \boldsymbol{v}_n$$

for some $\beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R}$. We can now replace the $\boldsymbol{v}_j$ that occurs in this linear combination by the right-hand side of (2.3), to obtain a linear combination that does not use $\boldsymbol{v}_j$. This shows that $\boldsymbol{v} \in \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \cancel{\boldsymbol{v}_j}, \ldots, \boldsymbol{v}_n\})$. (Again note that when $j = 1$, we can simply remove $\beta_j \cdot \boldsymbol{v}_j$ from the linear combination, since then $\boldsymbol{v}_j = \boldsymbol{o}$.)

The opposite inclusion $\operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \cancel{\boldsymbol{v}_j}, \ldots, \boldsymbol{v}_n\}) \subseteq \operatorname{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\})$ is clear, since any linear combination of vectors from $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ that does not use $\boldsymbol{v}_j$ can be changed into one that uses $\boldsymbol{v}_j$ as well, just by adding $0 \cdot \boldsymbol{v}_j$ to the linear combination. □

Theorem 2.20. *Let V be a vector space, with linearly independent $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m \in V$, and the set $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n\}$ spanning V. Then $m \leq n$.*

Proof. We start with the list of spanning vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$. We will prove that $m \leq n$ in m steps, where in each step we add a $\boldsymbol{v}_i$ to this list and remove a $\boldsymbol{w}_j$, while keeping the property that the vectors span the whole V, which is needed for passing to the next step.

In the first step we add $\boldsymbol{v}_1$ to the list to obtain $\boldsymbol{v}_1, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$. These vectors are linearly dependent, since $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ span V, so $\boldsymbol{v}_1$ is a linear combination of them. By the Redundancy Lemma 2.19, one of the vectors $\boldsymbol{v}_1, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ is a linear combination of previous vectors, and can be removed without changing the span. This element must be some $\boldsymbol{w}_j$, because $\boldsymbol{v}_1$ is linearly independent, hence $\boldsymbol{v}_1 \neq \boldsymbol{o}$, so is not a linear combination of the empty set. Thus $\boldsymbol{v}_1, \boldsymbol{w}_1, \ldots, \cancel{\boldsymbol{w}_j}, \ldots, \boldsymbol{w}_n$ still span V.

In the 2nd step, we add $\boldsymbol{v}_2$ to this list, after $\boldsymbol{v}_1$. Again, because $\boldsymbol{v}_1, \boldsymbol{w}_1, \ldots, \cancel{\boldsymbol{w}_j}, \ldots, \boldsymbol{w}_n$ span V, $\boldsymbol{v}_2$ is a linear combination of them, so the resulting vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{w}_1, \ldots, \cancel{\boldsymbol{w}_j}, \ldots, \boldsymbol{w}_n$ are linearly dependent. Again by the Redundancy Lemma 2.19, one of the vectors is a linear combination of previous ones, and can be removed without changing the span. This element must again be some $\boldsymbol{w}_k$, because $\boldsymbol{v}_1, \boldsymbol{v}_2$ are linearly independent. Thus $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{w}_1, \ldots, \cancel{\boldsymbol{w}_j}, \ldots, \cancel{\boldsymbol{w}_k}, \ldots, \boldsymbol{w}_n$ still span V.

In Step i, we already have a list of n vectors starting with $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{i-1}, \ldots$ that spans V, so $\boldsymbol{v}_i$ is a linear combination of them. We insert $\boldsymbol{v}_i$ into this sequence right after $\boldsymbol{v}_{i-1}$, before the $\boldsymbol{w}$'s to obtain a linearly dependent sequence of vectors. Again by the Redundancy Lemma 2.19, one of the vectors is a linear combination of previous ones, and as before, this vector must be some $\boldsymbol{w}_k$, since $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i$ are linearly independent. We remove this vector to obtain n vectors starting off with $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i$ that still span V.

Since we have m linearly independent vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$, we can continue up to Step m. Then, after step m, we have a list of n vectors starting off with $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$. It follows that $m \leq n$. □

Examples 2.21.

1. The vectors $\begin{bmatrix}-10\\11\\8\end{bmatrix}, \begin{bmatrix}4\\20\\3\end{bmatrix}, \begin{bmatrix}17\\12\\-2\end{bmatrix}, \begin{bmatrix}-5\\-5\\1\end{bmatrix}$ are linearly dependent in $\mathbb{R}^3$.

   Indeed, since the vectors $\boldsymbol{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \boldsymbol{e}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$ span $\mathbb{R}^3$, it follows by Theorem 2.20 that 4 vectors of $\mathbb{R}^3$ cannot be linearly independent.

2. The vectors $\begin{bmatrix}1\\-10\\11\\8\end{bmatrix}, \begin{bmatrix}5\\4\\20\\3\end{bmatrix}, \begin{bmatrix}-6\\17\\12\\-2\end{bmatrix}$ do not span $\mathbb{R}^4$.

   Indeed, since the vectors $\boldsymbol{e}_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \boldsymbol{e}_3 = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \boldsymbol{e}_4 = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ are linearly independent in $\mathbb{R}^4$, it follows by Theorem 2.20 that no 3 vectors can span $\mathbb{R}^4$.

Activity 2.22. *Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ be linearly independent vectors in a vector space V. Let $V_1 = \text{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\})$ and $V_2 = \text{lin}(\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\})$. Find $V_1 \cap V_2$, and justify your answer in detail.*

2.6 Finite-dimensional and infinite-dimensional

We are now in a position to distinguish between finite-dimensional and infinite-dimensional vector spaces (before we define the dimension of a vector space).

Definition 2.23. A vector space V is called **finite-dimensional** if V has a finite generating set; in other words, if there exists a finite subset S of vectors from V such that $V = \text{lin}(S)$. A vector space V is called **infinite-dimensional** if it is not finite-dimensional.

Examples 2.24.

1. The vector space $\mathbb{R}^{m\times n}$ is finite-dimensional.

   For a set that spans $\mathbb{R}^{m\times n}$ we can take all mn matrices that have a 1 in a single position and zeroes elsewhere. For instance, the $2 \times 3 = 6$ matrices

$$\begin{bmatrix}1&0&0\\0&0&0\end{bmatrix}, \quad \begin{bmatrix}0&1&0\\0&0&0\end{bmatrix}, \quad \begin{bmatrix}0&0&1\\0&0&0\end{bmatrix},$$

$$\begin{bmatrix}0&0&0\\1&0&0\end{bmatrix}, \quad \begin{bmatrix}0&0&0\\0&1&0\end{bmatrix}, \quad \begin{bmatrix}0&0&0\\0&0&1\end{bmatrix}$$

span $\mathbb{R}^{2\times 3}$.

2. The vector space $P_n(\mathbb{R})$ of all polynomial functions of degree at most n is finite-dimensional.

   The set of polynomials $\{1, x, x^2, \ldots, x^n\}$ spans $P_n(\mathbb{R})$.

3. The vector space $P(\mathbb{R})$ of all polynomial functions from $\mathbb{R}$ to $\mathbb{R}$ is infinite-dimensional.

   We can show this by contradiction. Suppose that $P(\mathbb{R})$ is finite-dimensional. Then there is a finite set $\{p_1, p_2, \ldots, p_n\}$ of polynomials that spans $P(\mathbb{R})$. Let N be the largest degree of the polynomials $p_1, \ldots, p_n$. Then any linear combination of these polynomials will have degree at most N. In particular, the polynomial x^{N+1} will not be in $\text{lin}(\{p_1, p_2, \ldots, p_n\})$, a contradiction.

We would also like to show that the vector space $C(\mathbb{R})$ of all continuous functions is infinite-dimensional. This seems clear, as it contains the infinite-dimensional vector space $P(\mathbb{R})$ as a subspace. This begs the question: Why are the subspaces of a finite-dimensional vector space also finite-dimensional? We establish this in the following result. Its proof uses the Redundancy Lemma 2.19 and Theorem 2.20.

Theorem 2.25. *Every subspace of a finite-dimensional vector space is finite-dimensional.*

Proof. Let V be a finite-dimensional vector space and let U be a subspace of V. Suppose that V is the span of m vectors. We will show in at most m steps that U is the span of at most m vectors.

Step 1: If $U = \{\boldsymbol{o}\}$, then U is the span of the empty set, so is finite-dimensional, and the proof is finished. Otherwise U contains a non-zero vector $\boldsymbol{v}_1$, which on its own is linearly independent, and we can go on to Step 2.

Step j: We have already chosen $j-1$ linearly independent vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1} \in U$. If $U = \text{lin}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}\})$, then U is finite-dimensional, and the proof is finished. Otherwise, we can prepare for the next step: There exists a vector $\boldsymbol{v}_j \in U \setminus \text{lin}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}\})$. Since $\boldsymbol{v}_j$ is not in the span of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}$, it follows by Lemma 2.19 that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}, \boldsymbol{v}_j$ are linearly independent in U. We can now go on to Step $j+1$.

After each Step j, we have j linearly independent vectors, and by Theorem 2.20 it follows that $j \leq m$. Thus at some stage, before m steps, we will not be able to continue to the next step, which means that U is the span of a finite set of vectors. We conclude that U is finite-dimensional. □

Example 2.26.

1. The vector space $C(\mathbb{R})$ of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is infinite-dimensional.

   Indeed, by the previous theorem, since $C(\mathbb{R})$ contains an infinite-dimensional subspace $P(\mathbb{R})$, it must itself be infinite-dimensional.

2.7 Basis and dimension

Definition 2.27. Let V be a vector space. Then a set B of vectors is called a **basis** of V if it is linearly independent and spans V.

Examples 2.28.

1. The vectors
$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \boldsymbol{e}_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$
   form a basis of $\mathbb{R}^m$.

2. Similarly, $\mathbb{R}^{m\times n}$ has a basis of mn elements.

3. The empty set $\varnothing$ is a basis of the subspace $\{\boldsymbol{o}\}$ of any vector space.

4. The set $\{1, x, x^2, \dots, x^n\}$ is a basis of the vector space $P_n(\mathbb{R})$ of polynomials of degree at most n.

5. The infinite set $\{1, x, x^2, x^3, \dots\}$ is a basis of the vector space $P_n(\mathbb{R})$ of polynomials.

Theorem 2.29 (Criterion for a basis)**.** *Let V be a finite-dimensional vector space. A subset S of V is a basis of V if and only if every $\boldsymbol{v} \in V$ can be written as a linear combination of vectors from S in a unique way.*

Proof. First assume that S is a basis of V. Because S is linearly independent, and V is the span of some finite set, it follows that S is a finite set by Theorem 2.20. Thus we can write $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n\}$ for some $n \in \mathbb{N}$. Let $\boldsymbol{v} \in V$. Because S is a basis, it spans V, hence $\boldsymbol{v}$ is a linear combination of elements from S:

$$\boldsymbol{v} = \alpha_1 \cdot \boldsymbol{v}_1 + \alpha_2 \cdot \boldsymbol{v}_2 + \cdots + \alpha_n \cdot \boldsymbol{v}_n$$

for some $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. To show uniqueness, suppose that there is also another representation

$$\boldsymbol{v} = \beta_1 \cdot \boldsymbol{v}_1 + \beta_2 \cdot \boldsymbol{v}_2 + \cdots + \alpha_m \cdot \boldsymbol{v}_m$$

for some $\beta_1, \dots, \beta_n \in \mathbb{R}$. If we subtract the second representation of $\boldsymbol{v}$ from the first, we obtain

$$\boldsymbol{o} = (\alpha_1 - \beta_1) \cdot \boldsymbol{v}_1 + (\alpha_2 - \beta_2) \cdot \boldsymbol{v}_2 + \cdots + (\alpha_n - \beta_n) \cdot \boldsymbol{v}_n,$$

and because $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ are linearly independent, we conclude that $\alpha_1 - \beta_1 = 0, \alpha_2 - \beta_2 = 0, \dots, \alpha_n - \beta_n = 0$. We conclude that $\alpha_i = \beta_i$ for each $i = 1, \dots, n$, which shows that the two linear combinations are the same. Therefore, $\boldsymbol{v}$ has a unique representation as a linear combination of vectors from S.

For the opposite direction, we assume that every $\boldsymbol{v} \in V$ can be written as a linear combination of vectors from S in a unique way. In particular, this shows that S spans V.

To show that S is linearly independent, write

$$\boldsymbol{o} = \alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n.$$

We can also write $\boldsymbol{o}$ as $\boldsymbol{o} = 0 \cdot \boldsymbol{v}_1 + \cdots + 0 \cdot \boldsymbol{v}_n$. Because we assumed that any vector in V can be written as a linear combination of vectors from S in a unique way, we obtain that $\alpha_1 = 0, \alpha_2 = 0, \dots, \alpha_n = 0$. This shows linear independence of S.

Therefore, S is a basis. □

From now on, we only consider finite-dimensional vector spaces. We will show that any finite-dimensional vector space has a basis, and that any two bases have the same cardinality. (This is also true for infinite-dimensional vector spaces, but much harder to show.)

Theorem 2.30. *Let V be a finite-dimensional vector space, and let $S \subseteq V$ be a spanning set of V. Then S contains a basis of V.*

Proof. We use several steps to choose linearly independent vectors from S until V is spanned by these vectors.

Step 1: Choose any non-zero $\boldsymbol{v}_1 \in S$. (If S is empty or $S = \{\boldsymbol{o}\}$, then $V = \{\boldsymbol{o}\}$, and has the empty set as basis, so there is nothing to prove.) If $V = \operatorname{lin}(\{\boldsymbol{v}_1\})$, then we are done. Otherwise, we go to Step 2.

Step j: We have already chosen linearly independent $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1} \in S$ in Step $j-1$ such that $V \neq \mathrm{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1}\})$. Let $\boldsymbol{v}_j$ be any vector in $S \setminus \mathrm{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1}\})$. Then $\boldsymbol{v}_j$ is not a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1}$, hence by Lemma 2.19, $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_j$ are linearly independent. If $V = \mathrm{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_j\})$, then we are done. Otherwise, we go to step $j+1$.

Because V is finite-dimensional, it is spanned by some finite set of m vectors, say. By Theorem 2.20, in each step we have $j \leq m$. Thus, by Step m or earlier, we will have to stop with V being the span of the linearly independent subset $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_j\}$ of S. □

Corollary 2.31. *Every finite-dimensional vector space has a basis.*

Proof. In Theorem 2.30, let S be the whole vector space V, which surely spans V. □

Although a finite-dimensional vector space can of course have many different bases, the next result says that two bases cannot have different cardinalities.

Corollary 2.32. *If a vector space V has a basis with n elements, then every basis of V has n elements.*

Proof. Let B be a basis of V with n elements, and let B' be another basis of V.

Suppose first that B' has more than n elements. Take any $n+1$ distinct vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{v}_{n+1}$ from B'. Since $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{v}_{n+1}$ are linearly independent, and B spans V, it follows from Theorem 2.20 that $n+1 \leq |B| = n$, a contradiction. Therefore, B' is finite and $|B'| \leq n$.

Since B is a linearly independent set and B' spans V, a second application of Theorem 2.20 gives that $|B| \leq |B'|$. It follows that $|B'| = n$. □

We now know by Corollary 2.31 that every finite-dimensional vector space has a basis, and by Corollary 2.32 that any two bases have the same cardinality. This motivates the following natural definition.

Definition 2.33. Let V be a finite-dimensional vector space. Then the **dimension** of V, denoted by $\dim(V)$, is defined to be the cardinality of a basis of V.

Activity 2.34. *Prove the following:*

1. $\dim(\mathbb{R}^m) = m$,
2. $\dim(\mathbb{R}^{m \times n}) = mn$,
3. $\dim(P_n(\mathbb{R})) = n+1$,
4. *the dimension of the zero subspace of any vector space is* $\dim(\{\boldsymbol{o}\}) = 0$.

We have already seen that any subspace of a finite-dimensional vector space is also finite-dimensional, so has a dimension too. The next theorem shows that the dimension of the subspace cannot exceed the dimension of the whole space.

Theorem 2.35. *Let V be a finite-dimensional vector space with a subspace U. Then* $\dim(U) \leq \dim(V)$.

Proof. By Theorem 2.25, U is finite-dimensional. By Corollary 2.31, U has a basis B_1 and V has a basis B_2. Since B_1 is a linearly independent subset of V and B_2 a spanning subset of V, we obtain by Theorem 2.20 that $|B_1| \leq |B_2|$. □

The next activity asks you to prove that any linearly independent set of n vectors in an n-dimensional vector space is already a basis; there is no need to check that they span the vector space.

Activity 2.36. *Let V be a vector space (of any dimension) and suppose that there exist n linearly independent vectors in V for some $n \in \mathbb{N}$. If S is a spanning subset of V of size n, then S is also a basis of V.*

We have the following dual results whose proof we also leave as an activity. It is needed in the proof of the Rank-Nullity theorem (Theorem 2.43).

Activity 2.37. *Let V be a finite-dimensional vector space of dimension $n \in \mathbb{N}$, and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in V$ be linearly independent. Then, there exist $n-k$ vectors, $\{\boldsymbol{w}_{k+1}, \ldots, \boldsymbol{w}_n\}$ such that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k, \boldsymbol{w}_{k+1}, \ldots, \boldsymbol{w}_n\}$ is a basis of V.*

2.8 Linear transformations

Definition 2.38. Let U, V be two vector spaces. A function $T\colon U \to V$ is called a **linear transformation** if it satisfies the following two properties:

Additivity: For all $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$, $T(\boldsymbol{u}_1 + \boldsymbol{u}_2) = T(\boldsymbol{u}_1) + T(\boldsymbol{u}_2)$.

Homogeneity: For all $\boldsymbol{u} \in U$, and all $\alpha \in \mathbb{R}$, $T(\alpha \cdot \boldsymbol{u}) = \alpha \cdot T(\boldsymbol{u})$.

Just as group homomorphisms are functions that respect group operations, linear transformations are functions that respect vector space operations. In fact, a linear transformation is also a homomorphism: Recall that U and V are abelian groups under vector addition. The Additivity property of a linear transformation says exactly that a linear transformation is a homomorphism from $(U, +)$ to $(V, +)$. Thus we can also say that a linear transformation $T\colon U \to V$ is a homorphism $(U, +) \to (V, +)$ that satisfies the Homogeneity property.

Examples 2.39.

1. For any vector spaces V and W, define the **zero map** $\boldsymbol{o}\colon V \to W$ by $\boldsymbol{o}(\boldsymbol{x}) = \boldsymbol{o}_W$ for all $\boldsymbol{x} \in V$, where $\boldsymbol{o}_W$ is the zero vector in W. Then $\boldsymbol{o}$ is a linear transformation.

2. For any vector space V, the **identity map** $I\colon V \to V$ defined by $I(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in V$, is a linear transformation.

3. Let $m, n \in \mathbb{N}$. Let $A \in \mathbb{R}^{m \times n}$. Then the function $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T_A(\boldsymbol{x}) = A\boldsymbol{x} \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^n,$$

is a linear transformation. To show this, we require the Additivity property: For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$T_A(\boldsymbol{x} + \boldsymbol{y}) = A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y} = T_A(\boldsymbol{x}) + T_A(\boldsymbol{y}),$$

and the Homogeneity property: For any $\alpha \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$,

$$T(\alpha \cdot \boldsymbol{x}) = A(\alpha \cdot \boldsymbol{x}) = \alpha \cdot (A\boldsymbol{x}) = \alpha T_A(\boldsymbol{x}).$$

The only non-trivial parts of this proof are to show that $A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$ and $A(\alpha \cdot \boldsymbol{x}) = \alpha \cdot (A\boldsymbol{x})$. Each of these can be checked by writing out the matrix operations in detail. First note that $A\boldsymbol{x}$ is the product of the matrix A with the column vector $\boldsymbol{x}$. If we write

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \quad \text{and} \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

then

$$A\boldsymbol{x} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{pmatrix}$$

To show that $A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$ for all

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n,$$

we start with the left-hand side:

$$A(\boldsymbol{x} + \boldsymbol{y}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n} a_{1k}(x_k + y_k) \\ \vdots \\ \sum_{k=1}^{n} a_{mk}(x_k + y_k) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} (a_{1k}x_k + a_{1k}y_k) \\ \vdots \\ \sum_{k=1}^{n} (a_{mk}x_k + a_{mk}y_k) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n} a_{1k}x_k + \sum_{k=1}^{n} a_{1k}y_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k + \sum_{k=1}^{n} a_{mk}y_k \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{pmatrix} + \begin{pmatrix} \sum_{k=1}^{n} a_{1k}y_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}y_k \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= A\boldsymbol{x} + A\boldsymbol{y},$$

thus ending with the right-hand side. Similarly, to show that $A(\alpha \cdot \boldsymbol{x}) = \alpha \cdot (A\boldsymbol{x})$ for all $\alpha \in \mathbb{R}$ and all vectors

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n,$$

we calculate as follows:

$$A(\alpha \cdot \boldsymbol{x}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \left(\alpha \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n} a_{1k}(\alpha x_k) \\ \vdots \\ \sum_{k=1}^{n} a_{mk}(\alpha x_k) \end{pmatrix} = \begin{pmatrix} \alpha \sum_{k=1}^{n} a_{1k} x_k \\ \vdots \\ \alpha \sum_{k=1}^{n} a_{mk} x_k \end{pmatrix}$$

$$= \alpha \cdot \begin{pmatrix} \sum_{k=1}^{n} a_{1k} x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk} x_k \end{pmatrix} = \alpha \cdot \left(\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right)$$

$$= \alpha \cdot (A\boldsymbol{x}).$$

4. The function $T\colon C(\mathbb{R}) \to \mathbb{R}$ given by

$$Tf = f(1) \text{ for all } f \in C(\mathbb{R}),$$

is a linear transformation from the vector space $C(\mathbb{R})$ to the vector space $\mathbb{R}$. Indeed,

$$T(f+g) = (f+g)(1) = f(1) + g(1) = T(f) + T(g)$$

for all $f, g \in C(\mathbb{R})$, which gives the Additivity property. Furthermore,

$$T(\alpha \cdot f) = (\alpha \cdot f)(1) = \alpha f(1) = \alpha T(f)$$

for all $\alpha \in \mathbb{R}$ and all $f \in C(\mathbb{R})$, and so the Homogeneity property holds too. Thus T is a linear transformation.

5. Define $D\colon P(\mathbb{R}) \to P(\mathbb{R})$ by $D(p) = p'$, the derivative of the polynomial $p \in P(\mathbb{R})$. Then D is a linear transformation.

Additivity: For all $p, q \in P(\mathbb{R})$, $D(p+q) = (p+q)' = p' + q' = D(p) + D(q)$.

Homogeneity: For all $\alpha \in \mathbb{R}$ and $p \in P(\mathbb{R})$, $T(\alpha \cdot p) = (\alpha \cdot p)' = \alpha \cdot p' = \alpha \cdot T(p)$.

Thus the Additivity property for D just means that the derivative of a sum is the sum of the derivatives, and the Homogeneity property means that the derivative of a constant times a function equals the constant times the derivative.

6. Define $T\colon C(\mathbb{R}) \to \mathbb{R}$ by $T(f) = \int_0^1 f(x)dx$. Then T is a linear transformation.

   Additivity: For any $f, g \in C(\mathbb{R})$,

   $$T(f+g) = \int_0^1 (f+g)(x)dx = \int_0^1 (f(x)+g(x))dx = \int_0^1 f(x)dx + \int_0^1 g(x)dx = T(f)+T(g).$$

   Homogeneity: For any $\alpha \in \mathbb{R}$ and $f \in C(\mathbb{R})$,

   $$T(\alpha \cdot f) = \int_0^1 (\alpha \cdot f)(x)dx = \int_0^1 \alpha f(x)dx = \alpha \int_0^1 f(x)dx = \alpha T(f).$$

In Definitions 1.40 in the previous chapter, we defined the kernel and image of a homomorphism. Since, as we already observed, a linear transformation is a homomorphism, we can also consider its kernel and image. Indeed, given a linear transformation $T\colon U \to V$, the kernel of T is the subset

$$\ker(T) = \{\boldsymbol{u} \in U \mid T(\boldsymbol{u}) = \boldsymbol{o}\}$$

of U, and the image of T is the subset

$$\mathrm{im}(T) = \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U \text{ such that } T(\boldsymbol{u}) = \boldsymbol{v}\}$$

of V. These are not new definitions—we have just rewritten the definitions of kernel and image of a homomorphism. The kernel of a linear transformation T is also sometimes called its **null space**, because it is exactly the subset of U where T has the value $\boldsymbol{o} \in V$.

Examples 2.40.

1. Let $A \in \mathbb{R}^{m\times n}$, and let $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation defined by $T_A(\boldsymbol{x}) = A\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{R}^n$. Then the kernel of the linear transformation is the set of all vectors

   $$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

   such that the system of linear equations

   $$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

   is simultaneously satisfied. Thus, $\ker(T_A)$ is the set of all solutions to this homogeneous linear system.

   The range of T_A is the set of all vectors

   $$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$$

such that there exists a vector

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

such that

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= y_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= y_m. \end{aligned}$$

Thus, $\operatorname{im}(T_A)$ is the union of $\{\boldsymbol{o}\}$ with the set of all $\boldsymbol{y} \in \mathbb{R}^m$ for which this non-homogeneous system has a solution.

2. The function $T\colon C(\mathbb{R}) \to \mathbb{R}$ given by $Tf = f(1)$ for all $f \in C(\mathbb{R})$, is a linear transformation from the vector space $C(\mathbb{R})$ to the vector space $\mathbb{R}$, with kernel equal to the set of continuous functions that vanish at the point 1. The range of T is the whole vector space $\mathbb{R}$.

By Theorem 1.42, we know that $\ker(T)$ and $\operatorname{im}(T)$ are subgroups of the abelian groups $(U, +)$ and $(V, +)$, respectively. The following result shows, not surprisingly, that both sets are in fact vector subspaces.

Theorem 2.41. *Let U, V be vector spaces and $T\colon U \to V$ a linear transformation. Then*

1. $\ker(T)$ *is a subspace of* U.
2. $\operatorname{im}(T)$ *is a subspace of* V.

Proof. We will use the Subspace Criterion (Theorem 2.7) in each case. Let $\boldsymbol{o}_U, \boldsymbol{o}_V$ denote the zero vectors in the vector spaces U, V, respectively. Since $T(\boldsymbol{o}_U) = \boldsymbol{o}_V$ both $\ker(T)$ and $\operatorname{im}(T)$ are non-empty.

Let us first check that $\ker(T)$ is a subspace of U. Let $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{v} \in \ker(T)$. Thus $T(\boldsymbol{u}) = T(\boldsymbol{v}) = \boldsymbol{o}_V$. We have to show that $\alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v} \in \ker(T)$, and to do this, we calculate

$$\begin{aligned} T(\alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v}) &= T(\alpha \cdot \boldsymbol{u}) + T(\beta \cdot \boldsymbol{v}) \\ &= \alpha \cdot T(\boldsymbol{u}) + \beta \cdot T(\boldsymbol{v}) \\ &= \alpha \cdot \boldsymbol{o}_V + \beta \cdot \boldsymbol{o}_V \\ &= \boldsymbol{o}_V + \boldsymbol{o}_V = \boldsymbol{o}_V. \end{aligned}$$

We conclude that $\alpha \cdot \boldsymbol{u} + \beta \cdot \boldsymbol{v} \in \ker(T)$. By Theorem 2.7, $\ker(T)$ is a subspace of U.

We next check that $\operatorname{im}(T)$ is a subspace of V. Let $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \operatorname{im}(T)$. Then there exist $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$ such that $T(\boldsymbol{u}_1) = \boldsymbol{v}_1$ and $T(\boldsymbol{u}_2) = \boldsymbol{v}_2$. We have to show that $\alpha \cdot \boldsymbol{v}_1 + \beta \cdot \boldsymbol{v}_2 \in \operatorname{im}(T)$. This means that we have to show the existence of a $\boldsymbol{u} \in U$ such that $T(\boldsymbol{u}) = \alpha \cdot \boldsymbol{v}_1 + \beta \cdot \boldsymbol{v}_2$. It is not hard to guess that a candidate for $\boldsymbol{u}$ should be $\alpha \cdot \boldsymbol{u}_1 + \beta \cdot \boldsymbol{u}_2$. Let us check this: First, note that $\alpha \cdot \boldsymbol{u}_1 + \beta \cdot \boldsymbol{u}_2 \in U$ because $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$ and U is a vector space. Second,

$$\begin{aligned} T(\alpha \cdot \boldsymbol{u}_1 + \beta \cdot \boldsymbol{u}_2) &= T(\alpha \cdot \boldsymbol{u}_1) + T(\beta \cdot \boldsymbol{u}_2) \\ &= \alpha \cdot T(\boldsymbol{u}_1) + \beta \cdot T(\boldsymbol{u}_2) \\ &= \alpha \cdot \boldsymbol{v}_1 + \beta \cdot \boldsymbol{v}_2, \end{aligned}$$

so we conclude that $\alpha \cdot \boldsymbol{v}_1 + \beta \cdot \boldsymbol{v}_2 \in \operatorname{im}(T)$. By Theorem 2.7, $\operatorname{im}(T)$ is a subspace of V. □

Now that we know that $\ker(T)$ and $\operatorname{im}(T)$ are vector spaces, we can consider their dimensions if they are finite-dimensional.

Definition 2.42. Let U, V be vector spaces with U finite-dimensional and $T\colon U \to V$ a linear transformation. Then

1. the **nullity** of T is defined to be the dimension of the kernel of T:
$$\operatorname{nullity}(T) = \dim(\ker(T)),$$
2. and the **rank** of T is defined to be the dimension of the image of T:
$$\operatorname{rank}(T) = \dim(\operatorname{im}(T)).$$

Note that it is irrelevant whether V is finite-dimensional or not in the above definition. Our final result shows a connection between the nullity and the rank of a linear transformation. It also shows that the above definition makes sense by showing that $\ker(T)$ and $\operatorname{im}(T)$ are finite-dimensional if U is finite-dimensional.

Theorem 2.43 (Rank-Nullity Theorem)**.** *Let U be a finite-dimensional vector space, V be a vector space, and let $T\colon U \to V$ be a linear transformation. Then* $\ker(T)$ *and* $\operatorname{im}(T)$ *are finite-dimensional, and*
$$\operatorname{nullity}(T) + \operatorname{rank}(T) = \dim(U).$$

Proof. By Theorem 2.41, $\ker(T)$ is a subspace of the finite-dimensional U, so is also finite-dimensional by Theorem 2.25 and has a basis by Corollary 2.31. Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ be a basis of $\ker(T)$. By Activity 2.37, since this set is linearly independent, it can be extended to a basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ of U.

So far, we have $\dim(\ker(T)) = m$ and $\dim(U) = m+n$. What remains to prove is that $\operatorname{im}(T)$ is finite-dimensional and has dimension n, since then we will have
$$\operatorname{nullity}(T) + \operatorname{rank}(T) = \dim(\ker(T)) + \dim(\operatorname{im}(T)) = m + n = \dim(U),$$
which will finish the proof of the theorem.

To show that $\operatorname{im}(T)$ has dimension n, we will show that $\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$ is a basis. First, we show that $\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$ spans $\operatorname{im}(T)$, and then we show that the set is linearly independent.

Since $\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\} \subseteq \operatorname{im}(T)$, we also have
$$\operatorname{lin}(\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}) \subseteq \operatorname{im}(T)$$
by Theorem 2.13. To show the opposite inclusion, let $\boldsymbol{v} \in \operatorname{im}(T)$. Then $\boldsymbol{v} = T(\boldsymbol{u})$ for some $\boldsymbol{u} \in U$. We can write $\boldsymbol{u}$ as a linear combination of vectors from the basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ of U:
$$\boldsymbol{u} = \alpha_1 \cdot \boldsymbol{u}_1 + \alpha_2 \cdot \boldsymbol{u}_2 + \cdots + \alpha_m \cdot \boldsymbol{u}_m + \beta_1 \cdot \boldsymbol{v}_1 + \beta_2 \cdot \boldsymbol{v}_2 + \cdots + \beta_n \cdot \boldsymbol{v}_n$$
for some $\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R}$. Then

$$\begin{aligned}
\boldsymbol{v} &= T(\boldsymbol{u}) \\
&= T(\alpha_1 \cdot \boldsymbol{u}_1 + \alpha_2 \cdot \boldsymbol{u}_2 + \cdots + \alpha_m \cdot \boldsymbol{u}_m + \beta_1 \cdot \boldsymbol{v}_1 + \beta_2 \cdot \boldsymbol{v}_2 + \cdots + \beta_n \cdot \boldsymbol{v}_n) \\
&= \alpha_1 \cdot T(\boldsymbol{u}_1) + \alpha_2 \cdot T(\boldsymbol{u}_2) + \cdots + \alpha_m \cdot T(\boldsymbol{u}_m) + \beta_1 \cdot T(\boldsymbol{v}_1) + \beta_2 \cdot T(\boldsymbol{v}_2) + \cdots + \beta_n \cdot T(\boldsymbol{v}_n) \\
&= \alpha_1 \cdot \boldsymbol{o}_V + \alpha_2 \cdot \boldsymbol{o}_V + \cdots + \alpha_m \cdot \boldsymbol{o}_V + \beta_1 \cdot T(\boldsymbol{v}_1) + \beta_2 \cdot T(\boldsymbol{v}_2) + \cdots + \beta_n \cdot T(\boldsymbol{v}_n) \\
&= \beta_1 \cdot T(\boldsymbol{v}_1) + \beta_2 \cdot T(\boldsymbol{v}_2) + \cdots + \beta_n \cdot T(\boldsymbol{v}_n) \\
&\in \operatorname{lin}(\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}),
\end{aligned}$$

where we have used the properties of a linear transformation, as well as the fact that $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \ker(T)$. It follows that $\text{im}(T) \subseteq \text{lin}(\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\})$.

So far we have shown that $\text{im}(T) = \text{lin}(\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\})$, which means in particular, that $\text{im}(T)$ is finite-dimensional. We next show that $\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$ is linearly independent. Suppose that

$$\boldsymbol{o}_V = \gamma_1 \cdot T(\boldsymbol{v}_1) + \gamma_2 \cdot T(\boldsymbol{v}_2) + \cdots + \gamma_n \cdot T(\boldsymbol{v}_n)$$

for some $\gamma_1, \gamma_2, \ldots, \gamma_n \in \mathbb{R}$. Then

$$\begin{aligned} \boldsymbol{o}_V &= T(\gamma_1 \cdot \boldsymbol{v}_1) + T(\gamma_2 \cdot \boldsymbol{v}_2) + \cdots + T(\gamma_n \cdot \boldsymbol{v}_n) \\ &= T(\gamma_1 \cdot \boldsymbol{v}_1 + \gamma_2 \cdot \boldsymbol{v}_2 + \cdots + \gamma_n \cdot \boldsymbol{v}_n). \end{aligned}$$

Therefore, $\gamma_1 \cdot \boldsymbol{v}_1 + \gamma_2 \cdot \boldsymbol{v}_2 + \cdots + \gamma_n \cdot \boldsymbol{v}_n \in \ker(T)$, so we can write this vector as a linear combination of the basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ of $\ker(T)$:

$$\gamma_1 \cdot \boldsymbol{v}_1 + \gamma_2 \cdot \boldsymbol{v}_2 + \cdots + \gamma_n \cdot \boldsymbol{v}_n = \delta_1 \cdot \boldsymbol{u}_1 + \delta_2 \cdot \boldsymbol{u}_2 + \cdots + \delta_m \cdot \boldsymbol{u}_m$$

for some $\delta_1, \delta_2, \ldots, \delta_m \in \mathbb{R}$. Thus

$$\gamma_1 \cdot \boldsymbol{v}_1 + \gamma_2 \cdot \boldsymbol{v}_2 + \cdots + \gamma_n \cdot \boldsymbol{v}_n - \delta_1 \cdot \boldsymbol{u}_1 - \delta_2 \cdot \boldsymbol{u}_2 - \cdots - \delta_m \cdot \boldsymbol{u}_m = \boldsymbol{o}_U,$$

and since $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are linearly independent, it follows that $\gamma_1 = \gamma_2 = \cdots = \gamma_n = 0$.[5] Therefore, $\{T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_n)\}$ is linearly independent, hence a basis of $\text{im}(T)$. This shows that $\dim(\text{im}(T)) = n$, which finishes the proof. □

2.9 Solutions to selected Activities

Solution to Exercise 2.3. The empty set is not an abelian group.

Solution to Exercise 2.22. It is clear that $\text{lin}(\{\mathbf{v}_2, \mathbf{v}_3\}) \subseteq V_1 \cap V_2$. We claim that this is in fact an equality, and so it remains to show the reverse inclusion.

The elements of V_1 are of the form $\alpha_1\mathbf{v_1} + \alpha_2\mathbf{v_2} + \alpha_3\mathbf{v_3}$ and elements of V_2 are of the form $\beta_2\mathbf{v_2} + \beta_3\mathbf{v_3} + \beta_4\mathbf{v_4}$. If $v \in V_1 \cap V_2$, then there must exist appropriate α_i and β_j such that

$$\alpha_1\mathbf{v_1} + \alpha_2\mathbf{v_2} + \alpha_3\mathbf{v_3} = \beta_2\mathbf{v_2} + \beta_3\mathbf{v_3} + \beta_4\mathbf{v_4}.$$

The zero vector of course belongs to $V_1 \cap V_2$ and so we must be able to write

$$\mathbf{0} = \alpha_1\mathbf{v_1} + (\alpha_2 - \beta_2)\mathbf{v_2} + (\alpha_3 - \beta_3)\mathbf{v_3} - \beta_4\mathbf{v_4}.$$

Since $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ are linearly independent, it follows that $\alpha_1 = \beta_4 = 0$, $\alpha_2 = \beta_2$ and $\alpha_3 = \beta_3$. Hence $\mathbf{v} \in \text{lin}(\{\mathbf{v_2}, \mathbf{v_3}\})$.

Solution to Exercise 2.36. This follows from Lemma 2.19 and Theorem 2.20.

Solution to Exercise 2.37. Beginning with $S = \{\mathbf{v_1}, \ldots, \mathbf{v_k}\}$, keep adding $\mathbf{w_i}$ not in the span of S, until it is no longer possible (similar to the proof of Theorem 2.20. This process will stop at or before you end up with n vectors in total (due to Theorem 2.20). Since the claim is that you end up with n vectors in total, Corollary 2.32 states that the resulting set is a basis.

[5] The δ_i are also all equal to 0, but we don't need this.