# Groups and Rings

David Pierce

January 7, 2014, 11:41 a.m.

Matematik Bölümü
Mimar Sinan Güzel Sanatlar Üniversitesi
dpierce@msgsu.edu.tr
http://mat.msgsu.edu.tr/~dpierce/

*Groups and Rings*



Mathematics Department
Mimar Sinan Fine Arts University
Istanbul, Turkey
`http://mat.msgsu.edu.tr/~dpierce/`
`dpierce@msgsu.edu.tr`

# Preface

There have been several versions of the present text.

1. The first draft was my record of the first semester of the graduate course in algebra given at Middle East Technical University in Ankara in 2008–9. I had taught the same course also in 2003–4. The main reference for the course was Hungerford's *Algebra* [19].

2. I revised my notes when teaching algebra a third time, in 2009–10. Here I started making some attempt to indicate how theorems were going to be used later. What is now §1.4 (the development of the natural numbers from the Peano Axioms) was originally prepared for a course called Non-Standard Analysis, given at the Nesin Mathematics Village, Şirince, in the summer of 2009. I built up the foundational Chapter 1 around this section.

3. Another revision, but only partial, came in preparation for a course at Mimar Sinan Fine Arts University in Istanbul in 2013–4. I expanded Chapter 1, out of a desire to give some indication of how mathematics, and especially algebra, could be built up from some simple axioms about the relation of membership—that is, from set theory. This building up, however, is not part of the course proper.

4. The present version of the notes represents a more thorough-going revision, made during and after the course at Mimar Sinan. I try to make more use of examples, introducing them as early as possible. The number theory that has always been in the background has been integrated more explicitly into the text (see page 38). I have tried to distinguish more clearly between what is essential to the course and what is not; the starred sections comprise most of what is not essential.

All along, I have treated groups, not merely as structures satisfying certain axioms, but as structures isomorphic to groups of symmetries of sets.

The equivalence of the two points of view has been established in the theorem named for Cayley (in §2.1, on page 57). Now it is pointed out (in that section) that standard structures like $(\mathbb{Q}^+, 1, {}^{-1}, \cdot)$ and $(\mathbb{Q}, 0, -, +)$, are also groups, even though they are not obviously symmetry groups. Several of these structures are constructed in Chapter 1. (In earlier editions they were constructed later.)

Symmetry groups as such are investigated more thoroughly now, in §§2.2 and 2.3, *before* the group axioms are simplified in §2.4.

Rings are defined in Part I, on groups, so that their groups of units are available as examples of groups, especially in §5.1 on semidirect products (page 144). Also rings are needed to produce rings of matrices and their groups of units, as in §3.1 (page 83).

I give many page-number references, first of all for my own convenience in the composition of the text at the computer. Thus the capabilities of Leslie Lamport's LaTeX program in automating such references are invaluable. Writing the text could hardly have been contemplated in the first place without Donald Knuth's original TeX program. I now use the `scrbook` document class of KOMA-Script, "developed by Markus Kohm and based on earlier work by Frank Neukam" [28, p. 236].

Ideally every theorem would have an historical reference. This is a distant goal, but I have made some moves in this direction.

The only exercises in the text are the theorems whose proofs are not already supplied. Ideally more exercises would be supplied, perhaps in the same manner.

# Contents

# List of Figures

# Introduction

Published around 300 B.C.E., the *Elements* of Euclid is a model of mathematical exposition. Each of is thirteen books consists mainly of statements followed by proofs. The statements are usually called **Propositions** today [7, 8], although they have no particular title in the original text [6]. By their content, they can be understood as *theorems* or *problems.* Writing six hundred years after Euclid, Pappus of Alexandria explains the difference [34, p. 566]:

> Those who wish to make more skilful distinctions in geometry find it worthwhile to call
>
> - a **problem** (πρόβλημα), that in which it is *proposed* (προβάλλεται) to do or construct something;
> - a **theorem** (θεώρημα), that in which the consequences and necessary implications of certain hypotheses *are investigated* (θεωρεῖται).

For example, Euclid's first proposition is the the problem of constructing an equilateral triangle. His fifth proposition is the theorem that the base angles of an isosceles triangle are equal to one another.

Each proposition of the present notes has one of four titles: **Lemma, Theorem, Corollary,** or **Porism.** Each proposition may be followed by an explicitly labelled proof, which is terminated with a box □. *If there is no proof, the reader is expected to supply her or his own proof, as an exercise.* No propositions are to be accepted on faith.

Nonetheless, for an algebra course, some propositions are more important than others. The full development of the foundational Chapter 1 below would take a course in itself, but is not required for algebra as such.

In these notes, a proposition may be called a lemma if it will be used to prove a theorem, but then never used again. Lemmas in these notes are numbered sequentially. Theorems are also numbered sequentially, independently from the lemmas. A statement that can be proved easily from a theorem is called a corollary and is numbered with the theorem. So for example Theorem 14 on page 33 is followed by Corollary 14.1.

Some propositionss can be obtained easily, not from a preceding theorem itself, but from its proof. Such propositions are called *porisms* and, like corollaries, are numbered with the theorems from whose proofs they are derived. So for example Porism 121.1 on page 122 follows Theorem 121.

The word *porism* and its meaning are explained, in the 5th century C.E., by Proclus in his commentary on the first book of Euclid's *Elements* [30, p. 212]:

> "Porism" is a term applied to a certain kind of problem, such as those in the *Porisms* of Euclid. But it is used in its special sense when as a result of what is demonstrated some other theorem comes to light without our propounding it. Such a theorem is therefore called a "porism," as being a kind of incidental gain resulting from the scientific demonstration.

The translator explains that the word *porism* comes from the verb πορίζω, meaning to furnish or provide.

The original source for much of the material of these notes is Hungerford's *Algebra* [19], or sometimes Lang's *Algebra* [23], but there are various rearrangements and additions. The back cover of Hungerford's book quotes a review:

> Hungerford's exposition is clear enough that an average graduate student can read the text on his own and understand most of it.

I myself aim for logical clarity; but I do not intend for these notes to be a replacement for lectures in a classroom. Such lectures may amplify some parts, while glossing over others. As a graduate student myself, I understood a course to consist of the teacher's lectures, and the most useful reference was not a printed book, but the notes that I took in my own hand. I still occasionally refer to those notes today.

Hungerford is inspired by category theory, of which his teacher Saunders Mac Lane was one of the creators. Categories are defined in the present text in §4.5 (page 131). The spirit of category theory is seen at the beginning of Hungerford's Chapter I, "Groups":

> There is a basic truth that applies not only to groups but also to many other algebraic objects (for example, rings, modules, vector spaces, fields): in order to study effectively an object with a given algebraic

> structure, it is necessary to study as well the functions that preserve the given algebraic structure (such functions are called homomorphisms).

Hungerford's term *object* here reflects the usage of category theory. Taking inspiration from model theory, the present notes will often use the term *structure* instead. Structures are defined in §1.6 (page 40). The examples of objects named by Hungerford are all structures in the sense of model theory, although not every object in a category is a structure in this sense.

When a word is printed in **boldface** in these notes, the word is a technical term whose meaning can be inferred from the surrounding text.

# 1. Mathematical foundations

As suggested in the Introduction, the full details of this chapter are not strictly part of an algebra course, but are logically presupposed by the course.

One purpose of the chapter is to establish the notation whereby

$$\mathbb{N} = \{1, 2, 3, \dots\}, \qquad\qquad \upomega = \{0, 1, 2, \dots\}.$$

The elements of $\upomega$ are the von-Neumann natural numbers,[1] so that if $n \in \upomega$, then

$$n = \{0, \dots, n-1\}.$$

In particular, $n$ is itself a set with $n$ elements. When $n = 0$, this means $n$ is the empty set. A cartesian power $A^n$ can be understood as the set of functions from $n$ to $A$. Then a typical element of $A^n$ can be written as $(a_0, \dots, a_{n-1})$. Most people write $(a_1, \dots, a_n)$ instead; and when they want an $n$-element set, they use $\{1, \dots, n\}$. This is a needless complication, since it leaves us with no simple abbreviation for an $n$-element set.

Another purpose of this chapter is to review the construction, not only of the sets $\mathbb{N}$ and $\upomega$, but the sets $\mathbb{Q}^+$, $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}^+$, and $\mathbb{R}$ derived from them. We ultimately have certain *structures,* namely:

- the *semigroup* $(\mathbb{N}, +)$;
- the *monoids* $(\upomega, 0, +)$ and $(\mathbb{N}, 1, \cdot)$;

[1] The letter $\upomega$ is not the minuscule English letter called *double u,* but the minuscule Greek *omega,* which is probably in origin a double o. Obtained with the control sequence `\upomega` from the `upgreek` package for LaTeX, the $\upomega$ used here is upright, unlike the standard slanted $\omega$ (obtained with `\omega`). The latter $\omega$ might be used as a variable (as for example on page 187). We shall similarly distinguish between the constant $\uppi$ (used for the ratio of the circumference to the diameter of a circle, as well as for the *canonical projection* defined on page 113 and the *coordinate projections* defined on pages 120 and 134) and the variable $\pi$ (pages 64 and 186).

- the *groups* $(\mathbb{Q}^+, 1, {}^{-1}, \cdot)$, $(\mathbb{Q}, 0, -, +)$, $(\mathbb{Z}, 0, -, +)$, $(\mathbb{R}^+, 1, {}^{-1}, \cdot)$, and $(\mathbb{R}, 0, -, +)$;
- the *rings* $(\mathbb{Z}, 0, -, +, 1, \cdot)$, $(\mathbb{Q}, 0, -, +, 1, \cdot)$, and $(\mathbb{R}, 0, -, +, 1, \cdot)$.

## 1.1. Sets and geometry

Most objects of mathematical study can be understood as *sets.* A set is a special kind of *collection.* A **collection** is many things, considered as one. Those many things are the **members** or **elements** of the collection. The members **compose** the collection, and the collection **comprises** them.[2] Each member **belongs** to the collection and is **in** the collection, and the collection **contains** the member.

Designating certain collections as sets, we shall identify some properties of them that will allow us to do the mathematics that we want. These properties will be expressed by *axioms.* We shall use versions of the so-called Zermelo–Fraenkel Axioms with the Axiom of Choice. The collection of these axioms is denoted by ZFC. Most of these axioms were described by Zermelo in 1908 [39].

We study study sets axiomatically, because a naïve approach can lead to contradictions. For example, one might think naïvely that there was a collection of all collections. But there can be no such collection, because if there were, then there would be a collection of all collections that did not contain themselves, and *this* collection would contain itself if and only if it did not. This result is the **Russell Paradox,** described in a letter [31] from Russell to Frege in 1902.

The propositions of Euclid's *Elements* concern points and lines in a plane and in space. Some of these lines are *straight* lines, and some are circles. Two straight lines that meet at a point make an *angle.* Unless otherwise stated, straight lines have endpoints. It is possible to compare two straight lines, or two angles: if they can be made to coincide, they are *equal* to one another. This is one of Euclid's so-called *common notions.* If a straight line has an endpoint on another straight line, two angles are

[2] Thus the relations named by the verbs *compose* and *comprise* are converses of one another; but native English speakers often confuse these two verbs.

created. If they are equal to one another, then they are called *right angles.* One of Euclid's *postulates* is that all right angles are equal to one another. The other postulates tell us things that we can do: Given a center and radius, we can draw a circle. From any given point to another, we can draw a straight line, and we can extend an existing straight line beyond its endpoints; indeed, given *two* straight lines, with another straight line cutting them so as to make the interior angles on the same side together less than two right angles, we can extend the first two straight lines so far that they will intersect one another.

Using the common notions and the postulates, Euclid proves propositions: the problems and theorems discussed in the Introduction above. The common notions and the postulates do not *create* the plane or the space in which the propositions are set. The plane or the space exists already. The Greek word γεωμετρία has the original meaning of *earth measurement,* that is, surveying. People knew how to measure the earth long before Euclid's *Elements* was written.

Similarly, people were doing mathematics long before set theory was developed. Accordingly, the set theory presented here will assume that sets already exist. Where Euclid has postulates, we shall have axioms. Where Euclid has definitions and common notions and certain unstated assumptions, we shall have definitions and certain logical principles.

It is said of the *Elements,*

> A critical study of Euclid, with, of course, the advantage of present insights, shows that he uses dozens of assumptions that he never states and undoubtedly did not recognize. [20, p. 87]

One of these assumptions is that two circles will intersect if each of them passes through the center of the other. (This assumption is used to construct an equilateral triangle.) But it is impossible to state *all* of one's assumptions. We shall assume, for example, that if a formal sentence $\forall x\ \varphi(x)$ is true, what this means is that $\varphi(a)$ is true for arbitrary $a$. *This* means $\varphi(b)$ is true, and $\varphi(c)$ is true, and so on. However, there is nothing at the moment called $a$ or $b$ or $c$ or whatever. For that matter, we have no actual formula called $\varphi$. There is nothing called $x$, and moreover there will never be anything called $x$ in the way that there might be something called $a$. Nonetheless, we assume that everything we have said about $\varphi$, $x$, $a$, $b$, and $c$ makes sense.

The elements of every set will be sets themselves. By definition, two sets will *equal* if they have the same elements. There will be an *empty set,* denoted by

$$\varnothing;$$

this will have no elements. If $a$ is a set, then there will be a set denoted by

$$\{a\},$$

with the unique element $a$. If $b$ is also a set, then there will be a set denoted by

$$a \cup b,$$

whose members are precisely the members of $a$ and the members of $b$. Thus there will be sets $a \cup \{b\}$ and $\{a\} \cup \{b\}$; the latter is usually written as

$$\{a, b\}.$$

If $c$ is another set, we can form the set $\{a, b\} \cup \{c\}$, which we write as

$$\{a, b, c\},$$

and so forth. This will allow us to build up the following infinite sequence:

$$\varnothing, \qquad \{\varnothing\}, \qquad \{\varnothing, \{\varnothing\}\}, \qquad \Big\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\Big\}, \qquad \dots$$

By definition, these sets will be the natural numbers 0, 1, 2, 3, ... To be more precise, they are the **von Neumann natural numbers** [38].

## 1.2. Set theory

### 1.2.1. Notation

Our formal axioms for set theory will be written in a certain logic, whose symbols are:

1) variables, as $x$, $y$, and $z$;
2) the symbol $\in$ denoting the membership relation;
3) the Boolean connectives of propositional logic:
   a) the singulary connective $\neg$ ("not"), and
   b) the binary connectives $\vee$ ("or"), $\wedge$ ("and"), $\Rightarrow$ ("implies"), and $\Leftrightarrow$ ("if and only if");
4) parentheses;
5) quantification symbols $\exists$ ("there exists") and $\forall$ ("for all").

We may also introduce constants, as $a$, $b$, and $c$, or $A$, $B$, and $C$, to stand for particular sets. A variable or a constant is called a *term.* If $t$ and $u$ are terms, then the expression

$$t \in u$$

is called an **atomic formula.** It means $t$ is a member of $u$. From atomic formulas, other formulas are built up *recursively* by use of the symbols above, according to certain rules, namely,

1) if $\varphi$ is a formula, then so is $\neg\varphi$;
2) if $\varphi$ and $\psi$ are formulas, then so is $(\varphi * \psi)$, where $*$ is one of the binary Boolean connectives;
3) if $\varphi$ is a formula and $x$ is variable, then $\exists x\ \varphi$ and $\forall x\ \varphi$ are formulas.

The formula $\neg\ t \in u$ says $t$ is *not* a member of $u$. We usually abbreviate the formula by

$$t \notin u.$$

The expression $\forall z\ (z \in x \Rightarrow z \in y)$ is the formula saying that every element of $x$ is an element of $y$. Another way to say this is that $x$ is a **subset** of $y$, or that $y$ **includes** $x$. We abbreviate this formula by[3]

$$x \subseteq y.$$

[3]The relation $\subseteq$ of being included is completely different from the relation $\in$ of being contained. However, many mathematicians confuse these relations in words, using the word *contained* to describe both.

The formula $x \subseteq y \land y \subseteq x$ says that $x$ and $y$ have the same members, so that they are **equal** by the definition foretold above (page 16); in this case we use the abbreviation

$$x = y.$$

All occurrences of $x$ in the formulas $\exists x\ \varphi$ and $\forall x\ \varphi$ are **bound,**[4] and they remain bound when other formulas are built up from these formulas. Occurrences of a variable that are not bound are **free.**

### 1.2.2. Classes and equality

A **singulary**[5] formula is a formula in which only one variable occurs freely. If $\varphi$ is a singulary formula with free variable $x$, we may write $\varphi$ as

$$\varphi(x).$$

If $a$ is a set, then by replacing every free occurrence of $x$ in $\varphi$ with $a$, we obtain the formula

$$\varphi(a),$$

which is called a **sentence** because it has no free variables. This sentence is true or false, depending on which set $a$ is. If the sentence is true, then $a$ can be said to **satisfy** the formula $\varphi$. There is a collection of all sets that satisfy $\varphi$: we denote this collection by

$$\{x \colon \varphi(x)\}.$$

Such a collection is called a **class.** In particular, it is the class **defined** by the formula $\varphi$. If we give this class the name $\boldsymbol{C}$, then the expression

$$x \in \boldsymbol{C}$$

means just $\varphi(x)$.

---

[4]The word *bound* here is the past participle of the verb *to bind.* The etymologically unrelated verb *to bound* is also used in mathematics, but its past participle is *bounded.*

[5]The word **unary** is more common, but less etymologically correct.

A formula in which only two variables occur freely is **binary.** If $\psi$ is such a formula, with free variables $x$ and $y$, then we may write $\psi$ as

$$\psi(x, y).$$

We shall want this notation for proving Theorem 1 below. If needed, we can talk about ternary formulas $\chi(x, y, z)$, and so on.

The definition of equality of sets can be expressed by the sentences

$$\forall x\ \forall y\ \bigl(x = y \Rightarrow (a \in x \Leftrightarrow a \in y)\bigr), \tag{1.1}$$

$$\forall x\ \forall y\ \bigl((a \in x \Leftrightarrow a \in y) \Rightarrow x = y\bigr), \tag{1.2}$$

where $a$ is an arbitrary set. The **Equality Axiom** is that equal sets belong to the same sets:

$$\forall x\ \forall y\ \bigl(x = y \Rightarrow (x \in a \Leftrightarrow y \in a)\bigr). \tag{1.3}$$

The meaning of the sentences (1.1) and (1.3) is that equal sets satisfy the same atomic formulas.

**Theorem 1.** *Equal sets satisfy the same formulas:*

$$\forall x\ \forall y\ \Bigl(x = y \Rightarrow \bigl(\varphi(x) \Leftrightarrow \varphi(y)\bigr)\Bigr). \tag{1.4}$$

*Proof.* Suppose $a$ and $b$ are equal sets. By symmetry, it is enough to show

$$\varphi(a) \Rightarrow \varphi(b) \tag{1.5}$$

for all singulary formulas $\varphi(x)$. As noted, we have (1.5) whenever $\varphi(x)$ is an atomic formula $x \in c$ or $c \in x$. If we have (1.5) when $\varphi$ is $\psi$, then we have it when $\varphi$ is $\neg\psi$. If we have (1.5) when $\varphi$ is $\psi$ or $\chi$, then we have it when $\varphi$ is $(\psi * \chi)$, where $*$ is one of the binary connectives. If we have (1.5) when $\varphi(x)$ is of the form $\psi(x, c)$, then we have it when $\varphi(x)$ is $\forall y\ \psi(x, y)$ or $\exists y\ \psi(x, y)$. Therefore we do have (1.5) in all cases. □

The foregoing is a proof by **induction.** Such a proof is possible because formulas are defined recursively. See §1.4 below (page 31). Actually we have glossed over some details. This may cause confusion; but then the details themselves could cause confusion. What we are really proving is all of the sentences of one of the infinitely many forms

$$\left.\begin{array}{c} \forall x\ \forall y\ \Big(x = y \Rightarrow \big(\varphi(x) \Leftrightarrow \varphi(y)\big)\Big), \\ \forall x\ \forall y\ \forall z\ \Big(x = y \Rightarrow \big(\varphi(x, z) \Leftrightarrow \varphi(y, z)\big)\Big), \\ \forall x\ \forall y\ \forall z\ \forall z'\ \Big(x = y \Rightarrow \big(\varphi(x, z, z') \Leftrightarrow \varphi(y, z, z')\big)\Big), \\ \dotfill, \end{array}\right\} \tag{1.6}$$

where no constant occurs in any of the formulas $\varphi$. Assuming $a = b$, it is enough to prove every sentence of one of the forms

$$\begin{gathered} \varphi(a) = \varphi(b), \\ \varphi(a, c) = \varphi(b, c), \\ \varphi(a, c, c') = \varphi(b, c, c'), \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{gathered}$$

We have tried to avoid writing all of this out, by allowing constants to occur implicitly in formulas, and by understanding $\forall x\ \varphi(x)$ to mean $\varphi(a)$ for arbitrary $a$, as suggested above (page 15). We could abbreviate the sentences in (1.6) as

$$\forall x\ \forall y\ \forall z_1\ \ldots \forall z_n\ \Big(x = y \Rightarrow \\ \big(\varphi(x, z_1, \ldots, z_n) \Leftrightarrow \varphi(y, z_1, \ldots, z_n)\big)\Big). \tag{1.7}$$

However, we would have to explain what $n$ was and what the dots of ellipsis meant. The expression in (1.7) means one of the formulas in the infinite list suggested in (1.6), and there does not seem to be a better way to say it than that.

The sentence (1.4) is usually taken as a logical axiom, like one of Euclid's common notions. Then (1.1) and (1.3) are special cases of this axiom, but (1.2) is no longer true, either by definition or by proof. So this too must be taken as an axiom, which is called the **Extension Axiom.**

In any case, all of the sentences (1.1), (1.2), (1.3), and (1.4) end up being true. They tell us that equal sets are precisely those sets that are logically indistinguishable. We customarily treat equality as *identity.* We consider equal sets to be the *same* set. If $a = b$, we may say simply that $a$ is $b$.

Similarly, in ordinary mathematics, since $1/2 = 2/4$, we consider $1/2$ and $2/4$ to be the same. In ordinary *life* they are distinct: $1/2$ is one thing, namely one half, while $2/4$ is two things, namely two quarters. In mathematics, we ignore this distinction.

As with sets, so with classes, we define them to be **equal** if they have the same members. Thus whenever

$$\forall x\ \bigl(\varphi(x) \Leftrightarrow \psi(x)\bigr),$$

the formulas $\varphi$ and $\psi$ define equal classes. Again we consider equality as identity.

Finally, a set and a class can be considered as **equal** if they have the same members. Thus if $\boldsymbol{C}$ is the class defined by $\varphi(x)$, then the expression

$$a = \boldsymbol{C}$$

means $\forall x\ \bigl(x \in a \Leftrightarrow \varphi(x)\bigr)$.

**Theorem 2.** *Every set is a class.*

*Proof.* The set $a$ is the class $\{x \colon x \in a\}$. □

However, there is no reason to expect the converse to be true.

**Theorem 3.** *Not every class is a set.*

*Proof.* There are formulas $\varphi(x)$ such that

$$\forall y\ \neg\forall x\ \bigl(x \in y \Leftrightarrow \varphi(x)\bigr).$$

Indeed, let $\varphi(x)$ be the formula $x \notin x$. Then

$$\forall y\ \neg\bigl(y \in y \Leftrightarrow \varphi(y)\bigr).$$ □

More informally, the argument is that the class $\{x \colon x \notin x\}$ is not a set, because if it were a set $a$, then $a \in a \Leftrightarrow a \notin a$, which is a contradiction. This is what was given above as the Russell Paradox (page 14). Another example of a class that is not a set is given by the *Burali-Forti Paradox* on page 50 below.

### 1.2.3. Construction of sets

We have established what it means for sets to be equal. We have established that sets are examples, but not the only examples, of the collections called classes. However, we have not officially exhibited any sets. We do this now. The **Empty Set Axiom** is

$$\exists x\ \forall y\ y \notin x.$$

As noted above (page 16), the set whose existence is asserted by this axiom is denoted by $\varnothing$. This set is the class $\{x\colon x \neq x\}$.

We now obtain the sequence $0$, $1$, $2$, $\dots$, described above (page 16). We use the Empty Set Axiom to start the sequence. We continue by means of the **Adjunction Axiom:** if $a$ and $b$ are sets, then the set denoted by $a \cup \{b\}$ exists. Formally, the axiom is

$$\forall x\ \forall y\ \exists z\ \forall w\ (w \in z \Leftrightarrow w \in x \lor w = y).$$

In writing this sentence, we follow the convention whereby the connectives $\lor$ and $\land$ are more binding than $\Rightarrow$ and $\Leftrightarrow$, so the expression $(w \in z \Leftrightarrow w \in x \lor w = y)$ means the formula $\big(w \in z \Leftrightarrow (w \in x \lor w = y)\big)$.

We can understand the Adjunction Axiom as saying that, for all sets $a$ and $b$, the class $\{x\colon x \in a \lor x = b\}$ is actually a set. Adjunction is not one of Zermelo's original axioms of 1908; but the following is Zermelo's **Pairing Axiom:**

**Theorem 4.** *For any two sets $a$ and $b$, the set $\{a, b\}$ exists:*

$$\forall x\ \forall y\ \exists z\ \forall w\ (w \in z \Leftrightarrow w = x \lor w = y).$$

*Proof.* By Empty Set and Adjunction, $\varnothing \cup \{a\}$ exists, but this is just $\{a\}$. Then $\{a\} \cup \{b\}$ exists by Adjunction again. □

The theorem is that the class $\{x\colon x = a \lor x = b\}$ is always a set. Actually Zermelo does not have a Pairing Axiom as such, but he has an **Elementary Sets Axiom,** which consists of what we have called the Empty Set Axiom and the Pairing Axiom.[6]

[6] Zermelo also requires that for every set $a$ there be a set $\{a\}$; but this can be understood as a special case of pairing.

Every class $\boldsymbol{C}$ has a **union,** which is the class $\{x\colon \exists y\ (x \in y \wedge y \in \boldsymbol{C})\}$. This class is denoted by

$$\bigcup \boldsymbol{C}.$$

This notation is related as follows with the notation for the classes involved in the Adjunction Axiom:

**Theorem 5.** *For all sets $a$ and $b$, $a \cup \{b\} = \bigcup\{a, \{b\}\}$.*

We can now use the more general notation

$$a \cup b = \bigcup\{a, b\}.$$

The **Union Axiom** is that the union of a *set* is always a set:

$$\forall x\ \exists y\ y = \bigcup x.$$

The Adjunction Axiom is a consequence of the Empty-Set, Pairing, and Union Axioms. This why Zermelo did not need Adjunction as an axiom. We state it as an axiom, because we can do a lot of mathematics with it that does not require the full force of the Union Axiom. We shall however use the Union Axiom when considering unions of chains of structures (as on page 79 below).

Suppose $A$ is a set and $\boldsymbol{C}$ is the class $\{x\colon \varphi(x)\}$. Then we can form the class

$$A \cap \boldsymbol{C},$$

which is defined by the formula $x \in A \wedge \varphi(x)$. The **Separation Axiom** is that this class is a set. Standard notation for this set is

$$\{x \in A\colon \varphi(x)\}. \tag{1.8}$$

However, this notation is unfortunate. Normally the formula $x \in A$ is read as a sentence of ordinary language, namely "$x$ belongs to $A$" or "$x$ is in $A$." However, the expression in (1.8) is read as "the set of $x$ in $A$ such that $\varphi$ holds of $x$"; in particular, $x \in A$ here is read as the noun phrase

"$x$ in $A$" (or "$x$ belonging to $A$," or "$x$ that are in $A$," or something like that).[7]

Actually Separation is a *scheme* of axioms, one for each singulary formula $\varphi$:

$$\forall x\ \exists y\ \forall z\ \bigl(z \in y \Leftrightarrow z \in x \land \varphi(z)\bigr).$$

In most of mathematics, and in particular in the other sections of these notes, one need not worry too much about the distinction between sets and classes. But it is logically important. It turns out that the objects of interest in mathematics can be understood as sets. Indeed, we have already defined natural numbers as sets. We can talk about sets by means of formulas. Formulas define classes of sets, as we have said. Some of these classes turn out to be sets themselves; but again, there is no reason to expect all of them to be sets, and indeed by Theorem 3 some of them are not sets. *Sub-classes* of sets are sets, by the Separation Axiom; but some classes are too big to be sets. The class $\{x\colon x = x\}$ of all sets is not a set, since if it were, then the sub-class $\{x\colon x \notin x\}$ would be a set, and it is not.

Every set $a$ has a *power class,* namely the class $\{x\colon x \subseteq a\}$ of all subsets of $a$. This class is denoted by

$$\mathscr{P}(a).$$

The **Power Set Axiom** is that this class is a set:

$$\forall x\ \exists y\ y = \mathscr{P}(x).$$

Then $\mathscr{P}(a)$ can be called the **power set** of $a$. In the main text, after this chapter, we shall not explicitly mention power sets until page 184. However, the Power Set Axiom is of fundamental importance for allowing us to prove Theorem 9 on page 27 below.

We want the **Axiom of Infinity** to be that the collection $\{0, 1, 2, \dots\}$ of natural numbers as defined on page 16 is a set. It is not obvious

[7]Ambiguity of expressions like $x \in A$ (is it a noun or a sentence?) is common in mathematical writing, as for example in the abbreviation of $\forall\varepsilon\ (\varepsilon > 0 \Rightarrow \varphi)$ as $(\forall\varepsilon > 0\ )\varphi$. Such ambiguity is avoided in these notes. However, certain ambiguities are tolerated: letters like $a$ and $A$ stand sometimes for sets, sometimes for *names* for sets.

how to formulate this as a sentence of our logic. However, the indicated collection contains 0, which by definition is the empty set; also, for each of its elements $n$, the collection contains also $n \cup \{n\}$. Let $\mathbf{I}$ be the class of all *sets* with these properties: that is,

$$\mathbf{I} = \big\{x\colon 0 \in x \wedge \forall y\ (y \in x \Rightarrow y \cup \{y\} \in x)\big\}.$$

Thus, if it exists, the set of natural numbers will belong to $\mathbf{I}$. Furthermore, the set of natural numbers will be the *smallest* element of $\mathbf{I}$. But we still must make this precise. For an arbitrary class $\boldsymbol{C}$, we define

$$\bigcap \boldsymbol{C} = \{x\colon \forall y\ (y \in \boldsymbol{C} \Rightarrow x \in y)\}.$$

This class is the **intersection** of $\boldsymbol{C}$.

**Theorem 6.** *If $a$ and $b$ are two sets, then*

$$a \cap b = \bigcap \{a, b\}.$$

*If $a \in \boldsymbol{C}$, then*

$$\bigcap \boldsymbol{C} \subseteq a,$$

*so in particular $\bigcap \boldsymbol{C}$ is a set. However, $\bigcap \varnothing$ is the class of all sets, which is not a set.*

We can now define[8]

$$\omega = \bigcap \mathbf{I}. \tag{1.9}$$

**Theorem 7.** *The following conditions are equivalent.*

1. $\mathbf{I} \neq \varnothing$.

2. $\omega$ *is a set.*

3. $\omega \in \mathbf{I}$.

[8] Some writers define $\bigcap \boldsymbol{C}$ only when $\boldsymbol{C}$ is a nonempty set. This would make our definition of $\omega$ invalid without the Axiom of Infinity.

Any of the equivalent conditions in the theorem can be taken as the Axiom of Infinity. This does not by itself establish that $\omega$ has the properties we expect of the natural numbers; we still have to do some work. We shall do this in §1.5 (p. 38).

The **Axiom of Choice** can be stated in any of several equivalent versions. One of these versions is that every set can be **well-ordered:** that is, the set can be given a linear ordering (as defined on page 37 below) so that every nonempty subset has a least element (as in Theorem 23 on page 38). However, we have not yet got a way to understand an ordering as a set. An ordering is a kind of binary relation, and a binary formula can be understood to define a binary relation. But we cannot yet use our logical symbolism to say that such a relation *exists.* We shall be able to do so in the next section. We shall use the Axiom of Choice:

- to establish that every set has a *cardinality* (page 51);
- to prove Theorem 204, that every PID is a UFD (page 195);
- to prove Zorn's Lemma (page 202;
- hence to prove Stone's theorem on representations of Boolean rings (page 204).

The Axiom can also used to show:

- that direct sums are not always the same as direct products (page 126);
- that nonprincipal ultraproducts of fields exist (page 210).

For the record, we have now named all of the axioms given by Zermelo in 1908: (I) Extension, (II) Elementary Sets, (III) Separation, (IV) Power Set, (V) Union, (VI) Choice, and (VII) Infinity. Zermelo assumes that equality is identity: but his assumption is our Theorem 1. In fact Zermelo does not use logical formalism as we have. We prefer to define equality with (1.1) and (1.2) and then use the Axioms of (i) the Empty Set, (ii) Equality, (iii) Adjunction, (iv) Separation, (v) Union, (vi) Power Set, (vii) Infinity, and (viii) Choice. But these two collections of definitions and axioms are logically equivalent.

Apparently Zermelo overlooked on axiom, the *Replacement Axiom,* which was supplied in 1922 by Skolem [32] and by Fraenkel.[9] We shall give this

[9] I have not been able to consult Fraenkel's original papers. However, according to van Heijenoort [36, p. 291], Lennes also suggested something like the Replacement

axiom in the next section.

An axiom never needed in ordinary mathematics is the *Foundation Axiom.* Stated originally by von Neumann [37], it ensures that certain pathological situations, like a set containing itself, are impossible. It does this by declaring that every nonempty set has an element that is disjoint from it: $\forall x\ \exists y\ (x \neq \varnothing \Rightarrow y \in x \land x \cap y = \varnothing)$. We shall never use this.

The collection called ZFC is Zermelo's axioms, along with Replacement and Foundation. If we leave out Choice, we have what is called ZF.

## 1.3. Functions and relations

Given two sets $a$ and $b$, we define

$$(a, b) = \big\{\{a\}, \{a, b\}\big\}.$$

This set is the **ordered pair** whose first entry is $a$ and whose second entry is $b$. The purpose of the definition is to make the following theorem true.

**Theorem 8.** *Two ordered pairs are equal if and only if their first entries are equal and their second entries are equal:*

$$(a, b) = (x, y) \Leftrightarrow a = x \land b = y.$$

If $A$ and $B$ are sets, then we define

$$A \times B = \{z \colon \exists x\ \exists y\ (z = (x, y) \land x \in A \land y \in B)\}.$$

This is the **cartesian product** of $A$ and $B$.

**Theorem 9.** *The cartesian product of two sets is a set.*

---

Axiom at around the same time (1922) as Skolem and Fraenkel; but Cantor had suggested such an axiom in 1899.

*Proof.* If $a \in A$ and $b \in B$, then $\{a\}$ and $\{a, b\}$ are elements of $\mathscr{P}(A \cup B)$, so $(a, b) \in \mathscr{P}(\mathscr{P}(A \cup B))$, and therefore

$$A \times B \subseteq \mathscr{P}(\mathscr{P}(A \cup B)). \qquad \square$$

An **ordered triple** $(x, y, z)$ can be defined as $\big((x, y), z\big)$, and so forth.

A **function** or **map** from $A$ to $B$ is a subset $f$ of $A \times B$ such that, for each $a$ in $A$, there is exactly one $b$ in $B$ such that $(a, b) \in f$. Then instead of $(a, b) \in f$, we write

$$f(a) = b. \tag{1.10}$$

We have then

$$A = \{x \colon \exists y\ f(x) = y\},$$

that is, $A = \{x \colon \exists y\ (x, y) \in f\}$. The set $A$ is called the **domain** of $f$. A function is sometimes said to be a function **on** its domain. For example, the function $f$ here is a function on $A$. The **range** of $f$ is the subset

$$\{y \colon \exists x\ f(x) = y\}$$

of $B$. If this range is actually equal to $B$, then we say that $f$ is **surjective onto** $B$, or simply that $f$ is **onto** $B$. Strictly speaking, it would not make sense to say $f$ was surjective or onto, simply.

A function $f$ is **injective** or **one-to-one,** if

$$\forall x\ \forall z\ (f(x) = f(z) \Rightarrow x = z).$$

The expression $f(x) = f(z)$ is an abbreviation of $\exists y\ (f(x) = y \wedge f(z) = y)$, which is another way of writing $\exists y\ \big((x, y) \in f \wedge (z, y) \in f\big)$. An injective function from $A$ *onto* $B$ is a **bijection** from $A$ to $B$.

If it is not convenient to name a function with a single letter like $f$, we may write the function as

$$x \mapsto f(x),$$

where the expression $f(x)$ would be replaced by some particular expression involving $x$. As an abbreviation of the statement that $f$ is a function from $A$ to $B$, we may write

$$f \colon A \to B. \tag{1.11}$$

Thus, while the symbol $f$ can be understood as a *noun,* the expression $f\colon A \to B$ is a complete *sentence.* If we say, “Let $f\colon A \to B$,” we mean let $f$ be a function from $A$ to $B$.

If $f\colon A \to B$ and $D \subseteq A$, then the subset $\{y\colon \exists x\ (x \in D \land y = f(x)\}$ of $B$ can be written as one of[10]

$$\{f(x)\colon x \in D\}, \qquad\qquad f[D].$$

This set is the **image** of $D$ under $f$. Similarly, we can write

$$A \times B = \{(x, y)\colon x \in A \land y \in B\}.$$

Then variations on this notation are possible. For example, if $f\colon A \to B$ and $D \subseteq A$, we can define

$$f \restriction D = \{(x, y) \in f\colon x \in D\}.$$

**Theorem 10.** *If $f\colon A \to B$ and $D \subseteq A$, then*

$$f \restriction D\colon D \to B$$

*and, for all $x$ in $D$, $f \restriction D(x) = f(x)$.*

If $f\colon A \to B$ and $g\colon B \to C$, then we can define

$$g \circ f = \{(x, z)\colon \exists y\ (f(x) = y \land g(y) = z)\};$$

this is called the **composite** of $(g, f)$.

**Theorem 11.** *If $f\colon A \to B$ and $g\colon B \to C$, then*

$$g \circ f\colon A \to C.$$

*If also $h\colon C \to D$, then*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

[10]The notation $f(D)$ is also used, but the ambiguity is dangerous, at least in set theory as such.

We define

$$\mathrm{id}_A = \{(x, x) \colon x \in A\};$$

this is the **identity** on $A$.

**Theorem 12.** $\mathrm{id}_A$ *is a bijection from* $A$ *to itself. If* $f \colon A \to B$, *then*

$$f \circ \mathrm{id}_A = f, \qquad \mathrm{id}_B \circ f = f.$$

If $f$ is a bijection from $A$ to $B$, we define

$$f^{-1} = \{(y, x) \colon f(x) = y\};$$

this is the **inverse** of $f$.

**Theorem 13.**

1. *The inverse of a bijection from* $A$ *to* $B$ *is a bijection from* $B$ *to* $A$.
2. *Suppose* $f : A \to B$ *and* $g : B \to A$. *Then* $f$ *is a bijection from* $A$ *to* $B$ *whose inverse is* $g$ *if and only if*

$$g \circ f = \mathrm{id}_A, \qquad f \circ g = \mathrm{id}_B .$$

In the definition of the cartesian product $A \times B$ and of a functions from $A$ to $B$, we may replace the sets $A$ and $B$ with classes. For example, we may speak of the function $x \mapsto \{x\}$ on the class of all sets. If $\boldsymbol{F}$ is a function on some class $\boldsymbol{C}$, and $A$ is a *subset* of $\boldsymbol{C}$, then by the **Replacement Axiom,** the image $\boldsymbol{F}[A]$ is also a set. For example, if we are given a function $n \mapsto G_n$ on $\omega$, then by Replacement the class $\{G_n \colon n \in \omega\}$ is a set. Then the union of this class is a set, which we denote by

$$\bigcup_{n \in \omega} G_n.$$

A **singulary operation** on $A$ is a function from $A$ to itself; a **binary** on $A$ is a function from $A \times A$ to $A$. A **binary relation** on $A$ is a subset of $A \times A$; if $R$ is such, and $(a, b) \in R$, we often write

$$a \mathrel{R} b.$$

A singulary operation on $A$ is a particular kind of binary relation on $A$; for such a relation, we already have the special notation in (1.10). The reader will be familiar with other kinds of binary relations, such as *orderings.* We are going to define a particular binary relation on page 36 below and prove that it is an ordering.

## 1.4. An axiomatic development of the natural numbers

In the preceding sections, we sketched an axiomatic approach to set theory. Now we start over with an axiomatic approach to the natural numbers alone. In the section after this, we shall show that the set $\omega$ does actually provide a *model* of the axioms for natural numbers developed in the present section.

For the moment though, we forget the definition of $\omega$. We forget about starting the natural numbers with 0. Children learn to count starting with 1, not 0. Let us understand the natural numbers to compose *some* set called $\mathbb{N}$. This set has a distinguished **initial element,** which we call **one** and denote by

$$1.$$

On the set $\mathbb{N}$ there is also a distinguished singulary operation of **succession,** namely the operation

$$n \mapsto n+1,$$

where $n+1$ is called the **successor** of $n$. Note that some other expression like $S(n)$ might be used for this successor. For the moment, we have no binary operation called $+$ on $\mathbb{N}$.

I propose to refer to the ordered triple $(\mathbb{N}, 1, n \mapsto n+1)$ as an *iterative structure.* In general, by an **iterative structure,** I mean any set that has a distinuished element and a distinguished singulary operation. Here the underlying set can be called the **universe** of the structure. For a simple notational distinction between a structure and its universe, if the universe is $A$, the structure itself might be denoted by a fancier version of this letter, such as the Fraktur version $\mathfrak{A}$. See Appendix A (p. 222) for Fraktur versions, and their handwritten forms, for all of the Latin letters.

The iterative structure $(\mathbb{N}, 1, n \mapsto n+1)$ is distinguished among iterative structures by satisfying the following axioms.

1. 1 is not a successor: $1 \neq n+1$.
2. Succession is injective: if $m+1 = n+1$, then $m = n$.
3. The structure admits **proof by induction,** in the following sense. Every subset $A$ of the universe must be the whole universe, provided $A$ has the following two closure properties:

   a) $1 \in A$, and
   b) for all $n$, if $n \in A$, then $n+1 \in A$.

These axioms seem to have been discovered originally by Dedekind [5, II, VI (71), p. 67]; but they were written down also by Peano [29], and they are often known as the **Peano axioms.**

Suppose $(A, b, f)$ is an iterative structure. If we successively compute $b$, $f(b)$, $f(f(b))$, $f(f(f(b)))$, and so on, either we always get a new element of $A$, or we reach an element that we have already seen. In the latter case, if the first repeated element is $b$, then the first Peano axiom fails. If it is not $b$, then the second Peano axiom fails. The last Peano axiom, the Induction Axiom, would ensure that every element of $A$ was reached by our computations. None of the three axioms implies the others, although the Induction Axiom implies that exactly one of the other two axioms holds [16].

The following theorem will allow us to define all of the usual operations on $\mathbb{N}$. The theorem is difficult to prove. Not the least difficulty is seeing that the theorem *needs* to be proved.[11]

*Homomorphisms* will be defined generally on page 41, but meanwhile we need a special case. A **homomorphism** from $(\mathbb{N}, 1, n \mapsto n+1)$ to an iterative structure $(A, b, f)$ is a function $h$ from $\mathbb{N}$ to $A$ such that

1) $h(1) = b$, and
2) $h(n+1) = f(h(n))$ for all $n$ in $\mathbb{N}$.

[11] Peano did not see this need, but Dedekind did. Landau discusses the matter [22, pp. ix–x].

**Theorem 14** (Recursion). *For every iterative structure, there is exactly one homomorphism from* $(\mathbb{N}, 1, n \mapsto n+1)$ *to this structure.*

*Proof.* Given an iterative structure $(A, b, f)$, we seek a homomorphism $h$ from $(\mathbb{N}, 1, x \mapsto n+1)$ to $(A, b, f)$. Then $h$ will be a particular subset of $\mathbb{N} \times A$. Let $B$ be the set whose elements are the subsets $C$ of $\mathbb{N} \times A$ such that, if $(n, y) \in C$, then either

1) $(n, y) = (1, b)$ or else
2) $C$ has an element $(m, x)$ such that $(n, y) = (m+1, f(x))$.

In particular, $\{(1, b)\} \in B$. Also, if $C \in B$ and $(m, x) \in C$, then

$$C \cup \{(m+1, f(x))\} \in B.$$

Let $R = \bigcup B$; so $R$ is a subset of $\mathbb{N} \times A$. We may say $R$ is a *relation from* $\mathbb{N}$ *to* $A$. If $(n, y) \in R$, then (as suggested on page 30 above) we may write also

$$n \; R \; y.$$

Since $\{(1, b)\} \in B$, we have $1 \; R \; b$. Also, if $m \; R \; x$, then $(m, x) \in C$ for some $C$ in $B$, so $C \cup \{(m+1, f(x))\} \in B$, and therefore $(m+1) \; R \; f(x)$. Thus $R$ is the desired function $h$, provided $R$ is actually a *function* from $\mathbb{N}$ to $A$. Proving that $R$ is a function from $\mathbb{N}$ to $R$ has two stages.

1. Let $D$ be the set of all $n$ in $\mathbb{N}$ for which there is $y$ in $A$ such that $n \; R \; y$. Then we have just seen that $1 \in D$, and if $n \in D$, then $n+1 \in D$. By induction, $D = \mathbb{N}$. Thus if $R$ is a function, its domain is $\mathbb{N}$.

2. Let $E$ be the set of all $n$ in $\mathbb{N}$ such that, for all $y$ in $A$, if $n \; R \; y$ and $n \; R \; z$, then $y = z$. Suppose $1 \; R \; y$. Then $(1, y) \in C$ for some $C$ in $B$. Since $1$ is not a successor, we must have $y = b$, by definition of $B$. Therefore $1 \in E$. Suppose $n \in E$, and $(n+1) \; R \; y$. Then $(n+1, y) \in C$ for some $C$ in $B$. Again since $1$ is not a successor, we must have

$$(n+1, y) = (m+1, f(x))$$

for some $(m, x)$ in $C$. Since succession is injective, we must have $m = n$. Thus, $y = f(x)$ for some $x$ in $A$ such that $n \; R \; x$. Since $n \in E$, we know $x$ is *unique* such that $n \; R \; x$. Therefore $y$ is unique such that $(n+1) \; R \; y$. Thus $n+1 \in E$. By induction, $E = \mathbb{N}$.

So $R$ is the desired function $h$. Finally, $h$ is unique by induction. □

Note well that the proof uses all three of the Peano Axioms. The Recursion Theorem is often used in the following form.

**Corollary 14.1.** *For every set $A$ with a distinguished element $b$, and for every function $F$ from $\mathbb{N} \times B$ to $B$, there is a unique function $H$ from $\mathbb{N}$ to $A$ such that*

1) *$H(1) = b$, and*
2) *$H(n+1) = F(n, H(n))$ for all $n$ in $\mathbb{N}$.*

*Proof.* Let $h$ be the unique homomorphism from $(\mathbb{N}, 1, n \mapsto n+1)$ to $(\mathbb{N} \times A, (1, b), f)$, where $f$ is the operation $(n, x) \mapsto (n+1, F(n, x)))$. In particular, $h(n)$ is always an ordered pair. By induction, the first entry of $h(n)$ is always $n$; so there is a function $H$ from $\mathbb{N}$ to $A$ such that $h(n) = (n, H(n))$. Then $H$ is as desired. By induction, $H$ is unique. □

We can now use recursion to define, on $\mathbb{N}$, the binary operation

$$(x, y) \mapsto x + y$$

of **addition,** and the binary operation

$$(x, y) \mapsto x \cdot y$$

of **multiplication.** More precisely, for each $n$ in $\mathbb{N}$, we recursively define the operations $x \mapsto n + x$ and $x \mapsto n \cdot x$. The definitions are:

$$\begin{aligned} n + 1 &= n + 1, & n + (m+1) &= (n+m) + 1, \\ n \cdot 1 &= n, & n \cdot (m+1) &= n \cdot m + n. \end{aligned} \tag{1.12}$$

The definition of addition might also be written as $n + 1 = S(n)$ and $n + S(m) = S(n + m)$. In place of $x \cdot y$, we often write $xy$.

**Lemma 1.** *For all $n$ and $m$ in $\mathbb{N}$,*

$$1 + n = n + 1, \qquad (m+1) + n = (m+n) + 1.$$

*Proof.* Induction. □

**Theorem 15.** *Addition on* $\mathbb{N}$ *is*

1) ***commutative:*** $n + m = m + n$*; and*
2) ***associative:*** $n + (m + k) = (n + m) + k$*.*

*Proof.* Induction and the lemma. □

**Theorem 16.** *Addition on* $\mathbb{N}$ *allows* ***cancellation:*** *if* $n + x = n + y$*, then* $x = y$*.*

*Proof.* Induction, and injectivity of succession. □

The analogous proposition for multiplication is Corollary 22.1 below.

**Lemma 2.** *For all* $n$ *and* $m$ *in* $\mathbb{N}$*,*

$$1 \cdot n = n, \qquad (m + 1) \cdot n = m \cdot n + n.$$

*Proof.* Induction. □

**Theorem 17.** *Multiplication on* $\mathbb{N}$ *is*

1) *commutative:* $nm = mn$*;*
2) ***distributive*** *over addition:* $n(m + k) = nm + nk$*; and*
3) *associative:* $n(mk) = (nm)k$*.*

*Proof.* Induction and the lemma. □

Landau [22] proves *using induction alone* that $+$ and $\cdot$ exist as given by the recursive definitions above. However, Theorem 16 needs more than induction. So does the existence of the **factorial** function defined by

$$1! = 1, \qquad (n + 1)! = n! \cdot (n + 1).$$

So does **exponentiation,** defined by

$$n^1 = n, \qquad n^{m+1} = n^m \cdot n.$$

The usual ordering $<$ of $\mathbb{N}$ is defined recursively as follows. First note that $m \leqslant n$ means simply $m < n$ or $m = n$. Then the definition of $<$ is:

1) $m \nless 1$ (that is, $\neg\, m < 1$);
2) $m < n + 1$ if and only if $m \leqslant n$.

In particular, $n < n + 1$. Really, it is the sets $\{x \in \mathbb{N} \colon x < n\}$ that are defined by recursion:

$$\{x \in \mathbb{N} \colon x < 1\} = \varnothing,$$
$$\{x \in \mathbb{N} \colon x < n + 1\} = \{x \in \mathbb{N} \colon x < n\} \cup \{n\} = \{x \in \mathbb{N} \colon x \leqslant n\}.$$

We now have $<$ as a binary relation on $\mathbb{N}$; we must *prove* that it is an ordering.

**Theorem 18.** *The relation $<$ is **transitive** on $\mathbb{N}$, that is, if $k < m$ and $m < n$, then $k < n$.*

*Proof.* Induction on $n$. □

**Theorem 19.** *The relation $<$ is **irreflexive** on $\mathbb{N}$: $m \nless m$.*

*Proof.* Since every element $k$ of $\mathbb{N}$ is less than some other element (namely $k + 1$), it is enough to prove

$$k < n \Rightarrow k \nless k.$$

We do this by induction on $n$. The claim is vacuously true when $n = 1$. Suppose it is true when $n = m$. If $k < m + 1$, then $k < m$ or $k = m$. If $k < m$, then by inductive hypothesis $k \nless k$. If $k = m$, but $k < k$, then $k < m$, so again $k \nless k$. Thus the claim holds when $n = m + 1$. By induction, it holds for all $n$. □

**Lemma 3.** $1 \leqslant m$.

*Proof.* Induction. □

**Lemma 4.** *If $k < m$, then $k + 1 \leqslant m$.*

*Proof.* The claim is vacuously true when $m = 1$. Suppose it is true when $m = n$. Say $k < n + 1$. Then $k \leqslant n$. If $k = n$, then $k + 1 = n + 1$, so $k + 1 \leqslant n + 1$. If $k < n$, then $k + 1 \leqslant n$ by inductive hypothesis, so $k + 1 < n + 1$ by transitivity (Theorem 18), and therefore $k + 1 \leqslant n + 1$. Thus the claim holds when $m = n + 1$. By induction, the claim holds for all $m$. □

**Theorem 20.** *The relation $<$ is* ***total*** *on $\mathbb{N}$: either $k \leqslant m$ or $m < k$.*

*Proof.* By Lemma 3, the claim is true when $k = 1$. Suppose it is true when $k = \ell$. If $m \not< \ell + 1$, then $m \nleqslant \ell$. In this case, we have both $m \neq \ell$ and $m \not< \ell$. Also, by the inductive hypothesis, $\ell \leqslant m$, so $\ell < m$, and hence $\ell + 1 \leqslant m$ by Lemma 4. □

Because of Theorems 18, 19, and 20, the relation $<$ is a **linear ordering** of $\mathbb{N}$, and $\mathbb{N}$ is **linearly ordered** by $<$.

**Theorem 21.** *For all $m$ and $n$ in $\mathbb{N}$, we have $m < n$ if and only if the equation*

$$m + x = n \tag{1.13}$$

*is soluble in $\mathbb{N}$.*

*Proof.* By induction on $k$, if $m + k = n$, then $m < n$. We prove the converse by induction on $n$. We never have $m < 1$. Suppose for some $r$ that, for all $m$, if $m < r$, then the equation $m + x = r$ is soluble. Suppose also $m < r + 1$. Then $m < r$ or $m = r$. In the former case, by inductive hypothesis, the equation $m + x = r$ has a solution $k$, and therefore $m + (k + 1) = r + 1$. If $m = r$, then $m + 1 = r + 1$. Thus the equation $m + x = r + 1$ is soluble whenever $m < r + 1$. By induction, for all $n$ in $\mathbb{N}$, if $m < n$, then (1.13) is soluble in $\mathbb{N}$. □

**Theorem 22.** *If $k < \ell$, then*

$$k + m < \ell + m, \qquad km < \ell m.$$

Here the first conclusion is a refinement of Theorem 16; the second yields the following analogue of Theorem 16 for multiplication.

**Corollary 22.1.** *If $km = \ell m$, then $k = \ell$.*

**Theorem 23.** $\mathbb{N}$ *is well-ordered by* $<$*: every nonempty set of natural numbers has a least element.*

*Proof.* Suppose $A$ is a set of natural numbers with no least element. Let $B$ be the set of natural numbers $n$ such that, if $m \leqslant n$, then $m \notin A$. Then $1 \in B$, since otherwise 1 would be the least element of $A$. Suppose $m \in B$. Then $m + 1 \in B$, since otherwise $m + 1$ would be the least element of $A$. By induction, $B = \mathbb{N}$, so $A = \varnothing$. □

The members of $\mathbb{N}$ are the **positive integers;** the full set $\mathbb{Z}$ of *integers* will be defined formally in §1.7 below, on page 45. As presented in Books VII–IX of Euclid's *Elements,* number theory is a study of the positive integers; but a consideration of all integers is useful in this study, and the integers that will constitute a motivating example, first of a group (page 55), and then of a ring (page 80). Fundamental topics of number theory developed in the main text are:

- greatest common divisors, the Euclidean algorithm, and numbers prime to one another (sub-§3.2.4, page 100);
- prime numbers, Fermat's Theorem, and Euler's generalization of this (§3.5, page 108);
- Chinese Remainder Theorem, primitive roots (§4.7, page 142);
- Euclid's Lemma (§7.2, page 182);
- the Fundamental Theorem of Arithmetic (§7.4, page 192).

## 1.5. A construction of the natural numbers

For an arbitrary set $a$, let

$$a' = a \cup \{a\}.$$

If $A$ belongs to the class $\mathbf{I}$ defined in (1.9) on page 25, then $0 \in A$, and $A$ is closed under the operation $x \mapsto x'$, and so $(A, 0, ')$ is an iterative structure. In particular, by the Axiom of Infinity, $\omega$ is a set, so $(\omega, 0, ')$ is an iterative structure.

**Theorem 24.** *The structure* $(\omega, 0, ')$ *satisfies the Peano Axioms.*

*Proof.* There are three things to prove.

1. In $(\omega, 0, ')$, the initial element 0 is not a successor, because for all sets $a$, the set $a'$ contains $a$, so it is nonempty.

2. $(\omega, 0, ')$ admits induction, because, if $A \subseteq \omega$, and $A$ contains 0 and is closed under $x \mapsto x'$, then $A \in \mathbf{I}$, so $\bigcap \mathbf{I} \subseteq A$, that is, $\omega \subseteq A$.

3. It remains to establish that $x \mapsto x'$ is injective on $\omega$. On page 36, we used recursion to define a relation $<$ on $\mathbb{N}$ so that

$$m \nless 1, \qquad m < n + 1 \Leftrightarrow m < n \lor m = n. \tag{1.14}$$

Everything that we proved about this relation required only these properties, and induction. On $\omega$, we do not know whether we have recursion, but we have (1.14) when $<$ is $\in$ and 1 is 0: that is, we have

$$m \notin 0, \qquad m \in n' \Leftrightarrow m \in n \lor m = n.$$

Therefore $\in$ must be a linear ordering of $\omega$, by the proofs in the previous section. We also have Lemma 4 for $\in$, that is, if $n$ in $\omega$, and $m \in n$, then either $m' = n$ or $m' \in n$. In either case, $m' \in n'$. Thus, if $m \neq n$, then either $m \in n$ or $n \in m$, and so $m' \in n'$ or $n' \in m'$, and therefore $m' \neq n'$. □

Given sets $A$ and $B$, we define

$$A \smallsetminus B = \{x \in A \colon x \notin B\}.$$

As a corollary of the foregoing theorem, we have that the iterative structure $(\omega \smallsetminus \{0\}, 1, ')$ also satisfies the Peano Axioms. We may henceforth assume that $(\mathbb{N}, 1, x \mapsto x + 1)$ is this structure. In particular,

$$\mathbb{N} = \omega \smallsetminus \{0\}.$$

Thus we no longer need the Peano Axioms as axioms; they are theorems about $(\mathbb{N}, 1, x \mapsto x + 1)$ and $(\omega, 0, ')$.

We extend the definitions of addition and multiplication on $\mathbb{N}$ to allow their arguments to be 0:

$$n + 0 = n = 0 + n, \qquad n \cdot 0 = 0 = 0 \cdot n.$$

**Theorem 25.** *Addition and multiplication are commutative and associative on $\omega$, and multiplication distributes over addition.*

In particular, the equations (1.12) making up the recursive definitions of addition and multiplication on $\mathbb{N}$ are still valid on $\omega$. The same goes for factorials and exponentiation when we define

$$0! = 1, \qquad\qquad n^0 = 1.$$

## 1.6. Structures

For us, the point of using the von-Neumann definition of the natural numbers is that, under this definition, a natural number $n$ is a particular set, namely $\{0, \dots, n-1\}$, with $n$ elements. We denote the set of functions from a set $B$ to a set $A$ by

$$A^B.$$

In particular then, $A^n$ is the set of functions from $\{0, \dots, n-1\}$ into $A$. We can denote such a function by one of

$$(x_0, \dots, x_{n-1}), \qquad\qquad (x_i \colon i < n),$$

so that

$$A^n = \{(x_0, \dots, x_{n-1}) \colon x_i \in A\}.$$

Thus, $A^2$ can be identified with $A \times A$, and $A^1$ with $A$ itself. There is exactly one function from 0 to $A$, namely 0; so

$$A^0 = \{0\} = 1.$$

An $n$-ary **relation** on $A$ is a subset of $A^n$; an $n$**-ary operation** on $A$ is a function from $A^n$ to $A$. Relations and operations that are 2-ary, 1-ary, or 0-ary can be called **binary, singulary,** or **nullary,** respectively; after the appropriate identifications, this agrees with the terminology used in §1.3. A nullary operation on $A$ can be identified with an element of $A$.

Generalizing the terminology used at the beginning of §1.4, we define a **structure** as a set together with some distinguished relations and operations on the set; as before, the set is the **universe** of the structure.

Again, if the universe is $A$, then the whole structure might be denoted by $\mathfrak{A}$; if $B$, then $\mathfrak{B}$.

The **signature** of a structure comprises a symbol for each distinguished relation and operation of the structure. For example, we have so far obtained $\mathbb{N}$ as a structure in the signature $\{1, +, \cdot, <\}$. We may then write out this structure as

$$(\mathbb{N}, 1, +, \cdot, <).$$

In this way of writing the structure, an expression like $+$ stands not for the *symbol* of addition, but for the actual operation on $\mathbb{N}$. In general, if $s$ is a symbol of the signature of $\mathfrak{A}$, then the corresponding relation or operation on $A$ can, for precision, be denoted by $s^{\mathfrak{A}}$, in case there is another structure around with the same signature. We use this notation in writing the next definition, and later on page 91.

A **homomorphism** from a structure $\mathfrak{A}$ to a structure $\mathfrak{B}$ of the same signature is a function $h$ from $A$ to $B$ that *preserves* the distinguished relations and operations: this means

$$\begin{gathered} h(f^{\mathfrak{A}}(x_0, \dots, x_{n-1})) = f^{\mathfrak{B}}(h(x_0), \dots, h(x_{n-1})), \\ (x_0, \dots, x_{n-1}) \in R^{\mathfrak{A}} \Rightarrow (h(x_0), \dots, h(x_{n-1})) \in R^{\mathfrak{B}}, \end{gathered} \tag{1.15}$$

for all $n$-ary operation-symbols $f$ and relation-symbols $R$ of the signature, for all $n$ in $\omega$. To indicate that $h$ is a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$, we may write

$$h \colon \mathfrak{A} \to \mathfrak{B}$$

(rather than simply $h \colon A \to B$). We have already seen a special case of a homomorphism in the Recursion Theorem (Theorem 14 on page 33 above).

**Theorem 26.** *If $h \colon \mathfrak{A} \to \mathfrak{B}$ and $g \colon \mathfrak{B} \to \mathfrak{C}$, then*

$$g \circ h \colon \mathfrak{A} \to \mathfrak{C}.$$

A homomorphism is an **embedding** if it is injective and if the converse of (1.15) also holds. A surjective embedding is an **isomorphism.**

**Theorem 27.** *The function* $\mathrm{id}_A$ *is an isomorphism from* $\mathfrak{A}$ *to itself. The following are equivalent conditions on a bijective homomorphism* $h$ *from* $\mathfrak{A}$ *to* $\mathfrak{B}$*:*

1) $\mathfrak{B}$ *is an isomorphism from* $\mathfrak{A}$ *to* $\mathfrak{B}$*,*
2) $h^{-1}$ *is a homomorphism from* $\mathfrak{B}$ *to* $\mathfrak{A}$*,*
3) $h^{-1}$ *is an isomorphism from* $\mathfrak{B}$ *to* $\mathfrak{A}$*.*

If there is an isomorphism from a structure $\mathfrak{A}$ to a structure $\mathfrak{B}$, then these two structures are said to be **isomorphic** to one another, and we may write

$$\mathfrak{A} \cong \mathfrak{B}.$$

In this case $\mathfrak{A}$ and $\mathfrak{B}$ are indistinguishable as structures, and so (out of laziness perhaps) we may *identify* them, treating them as the *same* structure. We have already done this, in a sense, with $(\mathbb{N}, 1, x \mapsto x+1)$ and $(\omega \smallsetminus \{0\}, 1, ')$. However, we never actually had a set called $\mathbb{N}$, until we identified it with $\omega \smallsetminus \{0\}$.

A **substructure** of a structure $\mathfrak{B}$ is a structure $\mathfrak{A}$ of the same signature such that $A \subseteq B$ and the **inclusion** $x \mapsto x$ of $A$ in $B$ is an embedding of $\mathfrak{A}$ in $\mathfrak{B}$.

**Model theory** studies structures as such. **Universal algebra** studies **algebras,** which are sets with distinguished operations, but no distinguished relations (except for equality). In other words, an algebra is a structure in a signature with no symbols for relations (except equality).

We shall study mainly the algebras called *groups* and the algebras called *rings.* Meanwhile, we have the algebra $(\mathbb{N}, 1, +, \cdot)$, and we shall have more examples in the next section.

A **reduct** of a structure is obtained by ignoring some of its operations and relations, while the universe remains the same. The original structure is then an **expansion** of the reduct. For example, $(\mathbb{N}, +)$ is a reduct of $(\mathbb{N}, +, \cdot, <)$, and the latter is an expansion of the former.

## 1.7. Constructions of the integers and rationals

The following theorem is an example of something like *localization,* which will be the topic of §7.5 (p. 198). One learns the theorem implicitly in school, when one learns about fractions (as on page 21 above). Perhaps fractions are our first encounter with nontrivial equivalence-classes.

Let $\approx$ be the binary relation on $\mathbb{N} \times \mathbb{N}$ given by[12]

$$(a, b) \approx (x, y) \Leftrightarrow ay = bx. \tag{1.16}$$

**Lemma 5.** *The relation $\approx$ on $\mathbb{N} \times \mathbb{N}$ is an equivalence-relation.*

If $(a, b) \in \mathbb{N} \times \mathbb{N}$, let its equivalence-class with respect to $\approx$ be denoted by $a/b$ or

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

Let the set of all such equivalence-classes be denoted by

$$\mathbb{Q}^+.$$

This set comprises the **positive rational numbers.**

**Theorem 28.** *There are well-defined operations $+$, $^{-1}$, and $\cdot$ on $\mathbb{Q}^+$ given by the rules*

$$\frac{a}{b} + \frac{x}{y} = \frac{ay + bx}{by},$$
$$\left(\frac{x}{y}\right)^{-1} = \frac{y}{x},$$
$$\frac{a}{b} \cdot \frac{x}{y} = \frac{ax}{by}.$$

*There is a linear ordering $<$ of $\mathbb{Q}^+$ given by*

$$\frac{a}{b} < \frac{x}{y} \Leftrightarrow ay < bx.$$

[12] As a binary relation on $\mathbb{N} \times \mathbb{N}$, the relation $\approx$ is a subset of $(\mathbb{N} \times \mathbb{N})^2$, which we identify with $\mathbb{N}^4$.

*The structure* $(\mathbb{N},+,\cdot,<)$ *embeds in* $(\mathbb{Q}^+,+,\cdot,<)$ *under the map* $x \mapsto x/1$. *Addition and multiplication are commutative and associative on* $\mathbb{Q}^+$, *and multiplication distributes over addition. Moreover,*

$$\frac{1}{1}\cdot\frac{x}{y}=\frac{x}{y}, \qquad \left(\frac{x}{y}\right)^{-1}\cdot\frac{x}{y}=\frac{1}{1}, \tag{1.17}$$

*Finally,*

$$\frac{1}{1}<\frac{a}{b}\wedge\frac{1}{1}<\frac{x}{y}\Rightarrow\frac{1}{1}<\frac{a}{b}\cdot\frac{x}{y}. \tag{1.18}$$

The operations on $\mathbb{Q}^+$ in the theorem are said to be *well defined* because it is not immediately obvious that they exist at all. It is possible that $a/b = c/d$ although $(a,b) \neq (c,d)$. In this case one must check that (for example) $(ay+bx)/(by) = (cy+dx)/(dy)$. See page 93 below.

Because multiplication is commutative and associative on $\mathbb{Q}^+$, and (1.17) holds, the structure $(\mathbb{Q}^+, 1/1, {}^{-1}, \cdot)$ is an **commutative group.** Because in addition $\mathbb{Q}^+$ is linearly ordered by $<$, and (1.18) holds, the structure $(\mathbb{Q}^+, 1/1, {}^{-1}, \cdot, <)$ is an **ordered group.**

In the theorem, the natural number $n$ is *not* a rational number, but $n/1$ is a rational number. However, we henceforth *identify* $n$ and $n/1$: we treat them as the same thing. Then we have $\mathbb{N} \subseteq \mathbb{Q}^+$.

In the definition (1.16) of $\approx$, if we replace multiplication with addition, then instead of the positive rational numbers, we obtain the *integers.* Probably this construction of the integers is not learned in school. If it were, possibly students would never think that $-x$ is automatically a negative number. In any case, by applying this construction of the integers to the positive rational numbers, we obtain all of the rational numbers as follows. Let $\sim$ be the binary relation on $\mathbb{Q}^+ \times \mathbb{Q}^+$ given by

$$(a,b) \sim (x,y) \Leftrightarrow a+y=b+x. \tag{1.19}$$

**Lemma 6.** *The relation* $\sim$ *on* $\mathbb{Q}^+ \times \mathbb{Q}^+$ *is an equivalence-relation.*

If $(a,b) \in \mathbb{Q}^+ \times \mathbb{Q}^+$, let its equivalence-class with respect to $\sim$ be denoted by

$$a-b.$$

Let the set of such equivalence-classes be denoted by

$$\mathbb{Q}.$$

**Theorem 29.** *There are well-defined operations $-$, $+$, and $\cdot$ on $\mathbb{Q}$ given by the rules*

$$\begin{aligned} -(x-y) &= y-x, \\ (a-b)+(x-y) &= (a+x)-(b+y), \\ (a-b)\cdot(x-y) &= (ax+by)-(ay+bx). \end{aligned}$$

*There is a dense linear ordering $<$ of $\mathbb{Q}$ given by*

$$a-b<x-y \Leftrightarrow a+y<b+x.$$

*The structure $(\mathbb{Q}^+,+,\cdot,<)$ embeds in $(\mathbb{Q},+,\cdot,<)$ under the map $x \mapsto (x+1)-1$. The structure $(\mathbb{Q},1-1,-,+,<)$ is an ordered group. Moreover, multiplication is also commutative and associative on $\mathbb{Q}$, and it distributes over addition.*

We identify $\mathbb{Q}^+$ with its image in $\mathbb{Q}$. Now we can refer to the elements of $\mathbb{Q}$ as the **rational numbers.** We denote $1-1$ by

$$0.$$

Then $\mathbb{Q}^+=\{x\in\mathbb{Q}\colon 0<x\}$. We can now define

$$\mathbb{Z}=\{x-y\colon (x,y)\in\mathbb{N}\times\mathbb{N}\};$$

this is the set of **integers.**

**Theorem 30.** *The structure $(\mathbb{Z},0,-,+,1,\cdot,<)$ is a well-defined substructure of $(\mathbb{Q},0,-,+,1,\cdot,<)$. The structure $(\mathbb{Z},0,-,+,<)$ is an ordered group.*

We can also think of $\mathbb{Q}$ as arising from $\mathbb{Z}$ by the same construction that gives us $\mathbb{Q}^+$ from $\mathbb{N}$. This gives us the following.

**Theorem 31.** *There is a surjective function $(x, y) \mapsto x/y$ from the product $\mathbb{Z} \times (\mathbb{Z} \smallsetminus \{0\})$ to $\mathbb{Q}$ such that*

$$\frac{a}{b} + \frac{x}{y} = \frac{ay + bx}{by},$$
$$1 = \frac{1}{1},$$
$$\frac{a}{b} \cdot \frac{x}{y} = \frac{ax}{by}.$$

*Then*

$$\frac{a}{b} < \frac{x}{y} \Leftrightarrow ay < bx.$$

*There is an operation $x \mapsto x^{-1}$ on $\mathbb{Q} \smallsetminus \{0\}$ given by*

$$\left(\frac{x}{y}\right)^{-1} = \frac{y}{x}.$$

*Then $(\mathbb{Q} \smallsetminus \{0\}, 1, {}^{-1}, \cdot)$ is a commutative group. Finally,*

$$0 < x \land 0 < y \Rightarrow 0 < x \cdot y. \tag{1.20}$$

Because $(\mathbb{Q}, 0, -, 1, <)$ is an ordered group, and $(\mathbb{Q} \smallsetminus \{0\}, 1, {}^{-1}, \cdot)$ is a commutative group, and multiplication distributes over addition in $\mathbb{Q}$, and (1.20) holds, the structure $(\mathbb{Q}, 0, -, +, 1, \cdot, <)$ is an **ordered field.** However, the ordering of $\mathbb{Q}$ is not **complete,** that is, there are subsets with upper bounds, but no *suprema* (least upper bounds). An example is the set $\{x \in \mathbb{Q} \colon 0 < x \land x^2 < 2\}$.

## 1.8. A construction of the reals

There is a technique due to Dedekind for completing $(\mathbb{Q}, <)$ to obtain the completely ordered set $(\mathbb{R}, <)$. As Dedekind says explicitly [5, pp. 39–40], the original inspiration for the technique is the definition of *proportion* found in Book V of Euclid's *Elements.*

In the geometry of Euclid, let us refer to the collection of straight lines that are equal to a given straight line (in the sense of page 14 above) as the *length* of that straight line. Two lengths of straight lines can be *added* together by taking two particular lines with those lengths and setting them end to end. Then lengths of straight lines compose the set of positive elements of an ordered group. Therefore individual lengths can be *multiplied,* that is, taken several times. Indeed, if $A$ is a length, and $n \in \mathbb{N}$, we can define the multiple $nA$ of $x$ recursively:

$$1A = A, \qquad\qquad (n+1)A = nA + A.$$

It is assumed that, for any two lengths $A$ and $B$, some multiple of $A$ is greater than $B$: this is the **archimedean property.** If $C$ and $D$ are two more lengths, then $A$ has to $B$ the *same ratio* that $C$ has to $D$, provided that, for all $k$ and $m$ in $\mathbb{N}$,

$$kA > mB \Leftrightarrow kC > mD.$$

In this case, the four lengths $A$, $B$, $C$, and $D$ are *proportional,* and we may write

$$A : B : : C : D.$$

We can write the condition for this proportionality as

$$\left\{\frac{x}{y} \in \mathbb{Q}^+ : xB < yA\right\} = \left\{\frac{x}{y} \in \mathbb{Q}^+ : xD < yC\right\}$$

Dedekind's observation is that such sets can be defined independently of all geometrical considerations. Indeed, we may define a **positive real number** as a nonempty, proper subset $C$ of $\mathbb{Q}^+$ such that

1) if $a \in C$ and $b \in \mathbb{Q}^+$ and $b < a$, then $b \in C$, and
2) if $C$ has a supremum in $\mathbb{Q}^+$, this supremum does not belong to $C$.

Let the set of all positive real numbers be denoted by

$$\mathbb{R}^+.$$

**Theorem 32.** *The set $\mathbb{R}^+$ is completely ordered by proper inclusion. There are well-defined operations $+$, $^{-1}$, and $\cdot$ on $\mathbb{Q}^+$ given by the rules*

$$C + D = \{x + y \colon x \in C \wedge y \in D\},$$
$$C^{-1} = \{x^{-1} \colon x \in \mathbb{Q}^+ \wedge \exists y\ (y \in \mathbb{Q}^+ \smallsetminus C \wedge y < x)\},$$
$$C \cdot D = \{x \cdot y \colon x \in C \wedge y \in D\}.$$

*Then $(\mathbb{Q}^+, +, ^{-1}, \cdot)$ embeds in $(\mathbb{R}^+, +, ^{-1}, \cdot)$ under $y \mapsto \{x \in \mathbb{Q}^+ : x < y\}$.*

Let us identify $\mathbb{Q}^+$ with its image in $\mathbb{R}^+$. We may also write $\subset$ on $\mathbb{R}^+$ as $<$.

For every $n$ in $\omega$, an $n$-ary operation $f$ on $\mathbb{R}^+$ is **continuous** if, for every $(A_i \colon i < n)$ in $(\mathbb{R}^+)^n$, for every $\varepsilon$ in $\mathbb{Q}^+$, there is $(\delta_i \colon i < n)$ in $(\mathbb{Q}^+)^n$ such that, for all $(X_i \colon i < n)$ in $(\mathbb{R}^+)^n$, if

$$\bigwedge_{i<n} A_i - \delta_i < X_i < A_i + \delta_i,$$

then

$$f(A_i \colon i < n) - \varepsilon < f(X_i \colon i < n) < f(A_i \colon i < n) + \varepsilon.$$

**Theorem 33.** *The operations* $+$, $^{-1}$, *and* $\cdot$ *on* $\mathbb{R}^+$ *are continuous. Every composite of continuous functions on* $\mathbb{R}^+$ *is continuous.*

**Lemma 7.** *The only continuous singulary operation on* $\mathbb{R}^+$ *that is* $1$ *on* $\mathbb{Q}$ *is* $1$ *everywhere.*

**Theorem 34.** *The structure* $(\mathbb{R}^+, 1, {}^{-1}, \cdot, <)$ *is an ordered group, and addition is commutative and associative on* $\mathbb{R}^+$*, and multiplication distributes over addition on* $\mathbb{R}^+$*.*

Now define $\sim$ on $\mathbb{R}^+ \times \mathbb{R}^+$ as in (1.19). Just as before, this is an equivalence relation. The set of its equivalence-classes is denoted by

$$\mathbb{R}.$$

Just as before, we obtain the ordered field $(\mathbb{R}, 0, -, +, {}^{-1}, \cdot, <)$. But now, the ordering is complete. We identify $\mathbb{R}^+$ with its image in $\mathbb{R}$. The elements of $\mathbb{R}$ are the **real numbers.**

**Lemma 8.** *For every* $n$ *in* $\mathbb{N}$*, for every element* $A$ *of a completely and densely ordered group, the equation*

$$nX = A$$

*is soluble in the group.*

**Theorem 35.** *Suppose* $(G, 0, -, +, <)$ *is a completely and densely ordered group, and* $u$ *is a positive element of* $G$*, and* $b$ *is an element of* $\mathbb{R}^+$ *such that* $1 < b$*. Then there is an isomorphism from* $(G, 0, -, +, <)$ *to* $(\mathbb{R}^+, 1, {}^{-1}, \cdot, <)$ *taking* $u$ *to* $b$*.*

By the previous theorem, the completely ordered groups $(\mathbb{R}, 0, -, +, <)$ and $(\mathbb{R}^+, 1, {}^{-1}, \cdot, <)$ are isomorphic, and indeed for every $b$ in $\mathbb{R}^+$ such that $b > 1$, there is an isomorphism taking 1 to $b$. This isomorphism is denoted by

$$x \mapsto b^x,$$

and its inverse is

$$x \mapsto \log_b x.$$

**Theorem 36** (Intermediate Value Theorem). *If $f$ is a continuous singulary operation on $\mathbb{R}$, and $f(a) \cdot f(b) < 0$, then $f$ has a zero between $a$ and $b$.*

Hence for example the function $x \mapsto x^2 - 2$ must have a zero in $\mathbb{R}$ between 1 and 2. More generally, if $A \subseteq \mathbb{R}$, then the set of *polynomial functions over $A$* is obtained from the set of constant functions taking values in $A$, along with $-$, $+$, $\cdot$, and the projections $(x_0, \dots, x_{n-1}) \mapsto x_i$, by closing under taking composites. Then all polynomial functions over $\mathbb{R}$ are continuous, and so the Intermediate Value Theorem applies to the singulary polynomial functions. Therefore the ordered field $\mathbb{R}$ is said to be **real-closed.** However, there are smaller real-closed ordered fields: we establish this in the next section.

## 1.9. Countability

A set is **countable** if it embeds in $\omega$; otherwise the set is **uncountable.**

**Theorem 37.** *The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are all countable.*

**Theorem 38.** *$\mathscr{P}(\omega)$ is uncountable.*

*Proof.* Suppose $f$ is an injection from $\omega$ to $\mathscr{P}(\omega)$. Then the subset $\{x \colon x \notin f(x)\}$ of $\omega$ is not in the range of $f$, by a variant of the Russell Paradox: if $\{x \colon x \notin f(x)\} = f(a)$, then $a \in f(a) \Leftrightarrow a \notin f(a)$. $\square$

**Theorem 39.** *The set $\mathbb{R}$ is uncountable.*

*Proof.* We shall use the notation whose properties will be established in sub-§2.3.3 (p. 71). For every subset $A$ of $\omega$, let $g(A)$ be the set of rational numbers $x$ such that, for some $n$ in $\omega$,

$$x < \sum_{k \in A \cap n} \frac{2}{3^k}.$$

Then $g(A)$ is a real number by the original definition. The function $A \mapsto g(A)$ from $\mathscr{P}(\omega)$ to $\mathbb{R}$ is injective. $\square$

However, suppose we let $A^{\mathrm{rc}}$ be the smallest field that contains all zeros from $\mathbb{R}$ of singulary polynomial functions over $A$. If we define $A_0 = \mathbb{Q}$ and $A_{n+1} = {A_n}^{\mathrm{rc}}$, then $\bigcup_{n\in\omega} A_n$ will contain all zeros from $\mathbb{R}$ of singulary polynomial functions over itself. In fact $\bigcup_{n\in\omega} A_n$ will be $\mathbb{Q}^{\mathrm{rc}}$. But this field is countable.

We can say more about a set than whether it is countable or uncountable. The main reason for doing this here is that it provides a good example of a *classification*: see §3.7 on page 116 below. A class is **transitive** if it properly includes all of its elements. A transitive *set* is an **ordinal** if it is well-ordered by the relation of membership. Then all of the elements of $\omega$ are ordinals, and so is $\omega$ itself. The class of all ordinals can be denoted by

$$\mathbf{ON}.$$

**Theorem 40.** *The class* **ON** *is transitive and well-ordered by membership.*

In particular, **ON** cannot contain itself; so it is not a set. This result is the **Burali-Forti Paradox** [1].

**Theorem 41.** *Every well-ordered set $(A, <)$ is isomorphic to a unique ordinal. The isomorphism is a certain function $f$ on $A$, and this function is determined by the rule*

$$f(b) = \{f(x) \colon x < b\}.$$

There are three classes of ordinals.

1. A **successor** is an ordinal $\alpha'$ for some ordinal $\alpha$.

2. The least ordinal, 0, is in a class by itself.
3. A **limit** is an ordinal that is neither 0 nor a successor.

Then $\omega$ is the least limit ordinal.

Two sets are **equipollent** if there is a bijection between them. An ordinal is a **cardinal** if it is the least ordinal that is equipollent with it.

**Theorem 42.** *Every element of* $\omega$ *is a cardinal. So is* $\omega$ *itself.*

The class of cardinals can be denoted by

$$\mathbf{CN}.$$

By the Axiom of Choice, every set is equipollent with some unique cardinal. This is the **cardinality** or **size** of that set. The cardinality of an arbitrary set $A$ is denoted by

$$|A|.$$

A countable set has cardinality $\omega$ or less; uncountable sets have cardinality greater than $\omega$. The **finite** sets are those whose cardinalities are less then $\omega$; other sets are **infinite.**

**Theorem 43.** *A set is infinite if and only if it is in bijection with a proper subset of itself.*

**Theorem 44.** *There is a bijection from* $\mathbf{ON}$ *to* $\mathbf{CN} \smallsetminus \omega$ *(the class of infinite cardinals).*

The bijection of the theorem is denoted by

$$\alpha \mapsto \aleph_\alpha.$$

Thus $\omega = \aleph_0$, and $|\mathbb{R}| = \aleph_\alpha$ for some ordinal $\alpha$ that is greater than 0. The *Continuum Hypothesis* is that $|\mathbb{R}| = \aleph_1$, but we shall make no use of this.

# Part I.

# Groups

# 2. Basic properties of groups and rings

We define both groups and rings in this chapter. We define rings (in §2.5, page 79), because at the beginning of the next chapter (§3.1, page 83) we shall define certain groups—namely *general linear groups*—in terms of rings.

## 2.1. Groups

Given a set $A$, we may refer to a bijection from $A$ to itself as a **symmetry** or **permutation** of $A$. Let us denote the set of these symmetries by

$$\mathrm{Sym}(A).$$

This set can be equipped with:

1) the element $\mathrm{id}_A$, which is the **identity** on $A$;
2) the singulary operation $f \mapsto f^{-1}$, which is **inversion;**
3) the binary operation $(f, g) \mapsto f \circ g$, which is **composition.**

(The functions $\mathrm{id}_A$, $f^{-1}$, and $f \circ g$ are defined in §1.3, page 27). The structure or algebra denoted by

$$(\mathrm{Sym}(A), \mathrm{id}_A, {}^{-1}, \circ)$$

is the **complete group of symmetries** of $A$. A substructure of this can be called simply a **group of symmetries** of $A$. (Structures, substructures, and algebras are defined in §1.6, page 40.)

We may use the expression $\mathrm{Sym}(A)$ to denote the whole structure $(\mathrm{Sym}(A), \mathrm{id}_A, {}^{-1}, \circ)$. Then, when we speak of a **subgroup** of $\mathrm{Sym}(A)$, we mean a subset that contains the identity and is closed under inversion and composition.

**Theorem 45.** *For all sets $A$, for all elements $f$, $g$, and $h$ of a group of symmetries of $A$,*

$$\begin{aligned} f \circ \mathrm{id}_A &= f, \\ \mathrm{id}_A \circ f &= f, \\ f \circ f^{-1} &= \mathrm{id}_A, \\ f^{-1} \circ f &= \mathrm{id}_A, \\ (f \circ g) \circ h &= f \circ (g \circ h). \end{aligned}$$

*Proof.* Theorems 12, 13, and 11 in §1.3 (page 27). □

A **group** is a structure with the properties of a group of symmetries given by the last theorem, Theorem 45. That is, a group is a structure $(G, \mathrm{e}, {}^{-1}, \cdot)$ in which the following equations are *identities* (are true for all values of the variables):

$$\begin{aligned} x \cdot \mathrm{e} &= x, \\ \mathrm{e} \cdot x &= x, \\ x \cdot x^{-1} &= \mathrm{e}, \\ x^{-1} \cdot x &= \mathrm{e}, \\ (x \cdot y) \cdot z &= x \cdot (y \cdot z). \end{aligned}$$

We may say also that these equations are the *axioms* of groups: this means that their *generalizations* ($\forall x\ x \cdot \mathrm{e} = x$ and so forth) are true in every group, by definition. According to these axioms, in every group $(G, \mathrm{e}, {}^{-1}, \cdot)$,

1) the binary operation $\cdot$ is associative,
2) the element e is an identity with respect to $\cdot$,
3) the singulary operation ${}^{-1}$ is inversion with respect to $\cdot$ and e.

The identity and the inversion will turn out to be uniquely determined by the binary operation, by Theorem 68 on page 76.

A group is called **abelian** if its binary operation is commutative. If $A$ has at least three elements, then $\mathrm{Sym}(A)$ is not abelian. However, every one-element set $\{a\}$ becomes an abelian group when we define

$$\mathrm{e} = a, \qquad a^{-1} = a, \qquad a \cdot a = a.$$

This group is a **trivial group.** All trivial groups are isomorphic to one another. Therefore, as suggested on page 42, we tend to identify them with one another, referring to each of them as *the* trivial group.

Besides symmetry groups and the trivial group, we have four examples of groups from §1.7 (page 43), namely

$$(\mathbb{Q}^+, 1, {}^{-1}, \cdot), \quad (\mathbb{Q}, 0, -, +), \quad (\mathbb{Z}, 0, -, +), \quad (\mathbb{Q} \smallsetminus \{0\}, 1, {}^{-1}, \cdot),$$

and three examples from §1.8 (page 46):

$$(\mathbb{R}^+, 1, {}^{-1}, \cdot), \qquad (\mathbb{R}, 0, -, +), \qquad (\mathbb{R} \smallsetminus \{0\}, 1, {}^{-1}, \cdot).$$

These seven examples are all abelian. Four of them are the origin of a terminological convention. In an arbitrary group $(G, \mathrm{e}, {}^{-1}, \cdot)$, the operation $\cdot$ is usually called **multiplication.** We usually write $g \cdot h$ as $gh$. The element $g^{-1}$ is the **inverse** of $g$. The element e is the **identity,** and it is sometimes denoted by 1 rather than e.

Evidently the groups of rational numbers, of integers, and of real numbers use different notation. These groups are said to be written **additively.** Additive notation is often used for abelian groups, but almost never for other groups. It will be useful to have one more example of an abelian group. Actually there will be one example for each positive integer. If $a$ and $b$ are arbitrary integers for which the equation

$$ax = b$$

has a solution in $\mathbb{Z}$, then we say that $a$ **divides** $b$, or $a$ is a **divisor** or **factor** of $b$, or $b$ is a **multiple** of $a$, and we may write

$$a \mid b.$$

Using the notation due to Gauss [9, p. 1], for a positive integer $n$ and arbitrary integers $a$ and $b$ we write

$$a \equiv b \pmod{n}$$

if $n \mid a - b$. In this case we say $a$ and $b$ are **congruent** with respect to the **modulus** $n$. This manner of speaking is abbreviated by putting the Latin word *modulus* into the ablative case: $a$ and $b$ are congruent ***modulo*** $n$.[1] Still following Gauss, we may say too that $a$ is a **residue** of $b$ with respect to the modulus $n$.

**Theorem 46.** *Let $n \in \mathbb{N}$.*

1. *Congruence* modulo $n$ *is an equivalence-relation on $\mathbb{Z}$.*
2. *If $a \equiv x$ and $b \equiv y \pmod{n}$, then*

$$-a \equiv -x \ \ \& \ \ a + b \equiv x + y \ \ \& \ \ ab \equiv xy \pmod{n}.$$

Thus congruence *modulo* $n$ is an example of a *congruence* in the sense to be defined on page 94. The set of congruence-classes of integers *modulo* $n$ can be denoted by

$$\mathbb{Z}_n.$$

If $a$ is some integer, we can denote its congruence-class *modulo* $n$ by something like $[a]$ or $\bar{a}$, or more precisely by

$$a + n\mathbb{Z}.$$

(This is a *coset* in the sense to be defined in §3.4, page 106.)

**Theorem 47.** *For every positive integer $n$, the function*

$$x \mapsto x + n\mathbb{Z}$$

*from $\{0, \ldots, n-1\}$ to $\mathbb{Z}_n$ is a bijection.*

*Proof.* If $0 \leqslant i < j < n$, then $1 \leqslant j - i < n$, and so $nx > j - i$ for all $x$ in $\mathbb{N}$. By Theorem 22 (page 37),

$$i \not\equiv j \pmod{n}.$$

[1] The ablative case of Latin corresponds roughly to the *-den hali* of Turkish. Gauss writes in Latin; however, instead of *modulo n*, he says *secundum modulum n*, “according to the modulus $n$” [10, p. 2].

Thus the given map is injective. If $k \in Z$, let $a$ be its least nonnegative residue (which exists by Theorem 23). Then $a < n$ (since otherwise $0 \leqslant a - n < a$, and $a - n$ is also a residue of $k$). Thus

$$a + n\mathbb{Z} = k + n\mathbb{Z}.$$

So the given map is surjective. □

Again given a positive integer $n$, we may treat an arbitary integer as a name for its own congruence-class *modulo* $n$. In particular, by the last theorem, we may consider $\mathbb{Z}_n$ as being the set $\{0, \ldots, n-1\}$, where these $n$ elements are understood to be distinct. By Theorem 46, we have a well-defined structure $(\mathbb{Z}_n, 0, -, +, 1, \cdot)$, where 0 and 1 stand for their respective congruence-classes $n\mathbb{Z}$ and $1 + n\mathbb{Z}$. The following theorem is then easy to prove. In fact the formal verification will be made even easier by Theorem 84 on page 94.

**Theorem 48.** *For each $n$ in $\mathbb{N}$, the structure $(\mathbb{Z}_n, 0, -, +)$ is an abelian group.*

The (multiplicative) groups of positive rational numbers, of nonzero rational numbers, of positive real numbers, and of nonzero real numbers, and the (additive) groups of integers, rational numbers, real numbers, and integers with respect to some modulus, are not obviously symmetry groups. But they can be *embedded* in symmetry groups, in the sense of §1.6 (page 40). Indeed, every element $g$ of a group $G$ (written multiplicatively) determines a singulary operation $\lambda_g$ on $G$, given by

$$\lambda_g(x) = gx.$$

Then we have the following.

**Theorem 49** (Cayley)**.** *For every group $(G, \mathrm{e}, {}^{-1}, \cdot)$, the function*

$$x \mapsto \lambda_x$$

*embeds $(G, \mathrm{e}, {}^{-1}, \cdot)$ in the group $(\mathrm{Sym}(G), \mathrm{id}_G, {}^{-1}, \circ)$ of symmetries.*

*Proof.* We first observe that

$$\lambda_{\mathrm{e}} = \mathrm{id}_G, \qquad \lambda_{g \cdot h} = \lambda_g \circ \lambda_h,$$

because

$$\begin{gathered}\lambda_{\mathrm{e}}(x) = \mathrm{e} \cdot x = x = \mathrm{id}_G(x),\\ \lambda_{g \cdot h}(x) = (g \cdot h) \cdot x = g \cdot (h \cdot x) = \lambda_g(\lambda_h(x)) = (\lambda_g \circ \lambda_h)(x).\end{gathered}$$

Consequently, by Theorem 13 (page 30), each $\lambda_g$ has an inverse, and

$$(\lambda_g)^{-1} = \lambda_{g^{-1}}.$$

This establishes $x \mapsto \lambda_x \colon G \to \mathrm{Sym}(G)$ and in fact

$$x \mapsto \lambda_x \colon (G, \mathrm{e}, {}^{-1}, \cdot) \to (\mathrm{Sym}(G), \mathrm{id}_G, {}^{-1}, \circ)$$

—that is, by the notational convention established on page 41, $x \mapsto \lambda_x$ is a *homomorphism* from the one group to the other. It is an embedding, since if $\lambda_g = \lambda_h$, then in particular

$$g = g\,\mathrm{e} = \lambda_g(\mathrm{e}) = \lambda_h(\mathrm{e}) = h\,\mathrm{e} = h. \qquad \square$$

By Cayley's Theorem, every group can be considered as a symmetry group.

## 2.2. Symmetry groups

In case $n \in \omega$, then in place of $\mathrm{Sym}(n)$ the notation

$$\mathrm{S}_n$$

is also used. However, most people probably understand $\mathrm{S}_n$ as the complete group of symmetries of the set $\{1, \dots, n\}$. It does not really matter whether $\{0, \dots, n-1\}$ or $\{1, \dots, n\}$ is used; we just need a set with $n$ elements, and we are using $\{0, \dots, n-1\}$, which is $n$, as this set.

In the following, the *factorial* of a natural number was defined on pages 35 and 40, and the *cardinality* of a set was defined on page 51.

**Theorem 50.** *For each $n$ in $\omega$,*

$$|\mathrm{Sym}(n)| = n!$$

The group $\mathrm{Sym}(0)$ has a unique element, $\mathrm{id}_0$, which is itself $0$, that is, $\varnothing$. The group $\mathrm{Sym}(1)$ has the unique element $\mathrm{id}_1$, which is $\{(0,0)\}$. Thus

$$\mathrm{Sym}(0) = 1, \qquad\qquad \mathrm{Sym}(1) = \big\{\{(0,0)\}\big\}.$$

As groups, they are both trivial. We can think of the next symmetry groups—$\mathrm{Sym}(2)$, $\mathrm{Sym}(3)$, and so on—in terms of the following notion.

### 2.2.1. Automorphism groups

An **automorphism** of a structure is an isomorphism from the structure to itself. The set of automorphisms of a structure $\mathfrak{A}$ can be denoted by

$$\mathrm{Aut}(\mathfrak{A}).$$

We have $\mathrm{Aut}(\mathfrak{A}) \subseteq \mathrm{Sym}(A)$, where as usual $A$ is the universe of $\mathfrak{A}$; and we have more:

**Theorem 51.** *For every structure $\mathfrak{A}$, the set $\mathrm{Aut}(\mathfrak{A})$ is the universe of a substructure of the group of symmetries of $A$.*

*Proof.* $\mathrm{Aut}(\mathfrak{A})$ contains $\mathrm{id}_A$ and is closed under inversion and composition. □

Thus we may speak of $\mathrm{Aut}(\mathfrak{A})$ as the **automorphism group** of $\mathfrak{A}$.

### 2.2.2. Automorphism groups of graphs

It will be especially useful to consider automorphism groups of *graphs*. As a structure, a **graph on** a set $A$ is an ordered pair $(A, E)$, where $E$ is an antisymmetric, reflexive binary relation on $A$. This means

$$\neg\, x \mathrel{E} x, \qquad\qquad x \mathrel{E} y \Leftrightarrow y \mathrel{E} x.$$

The elements of $A$ are called **vertices** of the graph. If $b \mathrel{E} c$, then the set $\{b, c\}$ is called an **edge** of the graph. An edge is an example of an **(unordered) pair,** that is, a set with exactly two elements. The set of unordered pairs of elements of a set $A$ can be denoted by

$$[A]^2.$$

Every graph on a given set is determined by its edges, and moreover every subset of $[A]^2$ determines a graph on $A$. This result can be stated as follows.

**Theorem 52.** *For every set $A$, there is a bijection*

$$E \mapsto \bigl\{\{x, y\} \colon (x, y) \in E\bigr\}$$

*from the set of antisymmetric, reflexive binary relations on $A$ to $\mathscr{P}([A]^2)$.*

For our purposes, the **triangle** is the graph on 3 and edge set $[3]^2$. In a word, it is the **complete graph on** 3. Therefore every permutation of 3 is an automorphism of the triangle. The vertices of this triangle can be envisioned as the points $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ in the space $\mathbb{R}^3$. An automorphism of this triangle then induces a permutation of the coordinate axes of $\mathbb{R}^3$.

Similarly, the **tetrahedron** is the complete graph on 4, and so each permutation of 4 is an automorphism of the tetrahedron. The tetrahedron can be envisioned as having vertices $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$ in $\mathbb{R}^4$.

In general, $\mathrm{Sym}(n)$ can be understood as comprising the permutations of the coordinate axes of $\mathbb{R}^n$. In this way, an element $\sigma$ of $\mathrm{Sym}(n)$ determines the permutation

$$(x_i \colon i < n) \mapsto (x_{\sigma^{-1}(i)} \colon i < n)$$

of $\mathbb{R}^n$. The reason why we use $\sigma^{-1}$ in this rule is the following. Suppose we denote by $f_\sigma$ the permutation of $\mathbb{R}^n$ given by this rule. Then

$$\begin{aligned}
f_\tau(f_\sigma(x_i \colon i < n)) &= f_\tau(x_{\sigma^{-1}(i)} \colon i < n) \\
&= (x_{\sigma^{-1}(\tau^{-1}(i))} \colon i < n) \\
&= (x_{(\tau\sigma)^{-1}(i)} \colon i < n) \\
&= f_{\tau\sigma}(x_i \colon i < n).
\end{aligned}$$

Thus $\sigma \mapsto f_\sigma$ is a homomorphism from $\operatorname{Sym}(n)$ to $\operatorname{Sym}(\mathbb{R}^n)$. Another way to see this is to recall that an element $(x_i \colon i < n)$ of $\mathbb{R}^n$ is just a function $i \mapsto x_i$ from $n$ to $\mathbb{R}$. Denoting this function simply by $x$, we have

$$f_\sigma(x) = x \circ \sigma^{-1},$$
$$f_\tau(f_\sigma(x)) = x \circ \sigma^{-1} \circ \tau^{-1} = x \circ (\tau \circ \sigma)^{-1} = f_{\tau\sigma}(x).$$

This idea will come back in §3.1 (p. 83). Meanwhile, we are going to develop a way to distinguish the *orientation-preserving* permutations of the axes, namely the permutations that can be achieved by rotation without reflection.

If $n \geqslant 3$, we may consider the $n$**-gon** to be the graph on $n$ with the $n$ vertices

$$\{0,1\}, \qquad \{1,2\}, \qquad \{2,3\}, \qquad \dots, \qquad \{n-2,n-1\}, \qquad \{n-1,0\}$$

Considering $n$ as $\mathbb{Z}_n$, we can also write these edges more symmetrically as

$$\{i, i+1\},$$

where $i \in \mathbb{Z}_n$. The 3-gon is the triangle. The **square** is the 4-gon. The $n$**th dihedral group,** denoted by one of

$$\operatorname{Dih}(n), \qquad\qquad \mathrm{D}_n,$$

is the automorphism group of the $n$-gon; it is a subgroup of $\operatorname{Sym}(n)$.

**Theorem 53.** *If $n \geqslant 3$, then every element $\sigma$ of $\operatorname{Dih}(n)$ is determined by $(\sigma(0), \sigma(1))$. Moreover, $\sigma(0)$ can have any value in $n$, and then $\sigma(1)$ can and must be $\sigma(0) \pm 1$. Thus*

$$|\operatorname{Dih}(n)| = 2n.$$

Theorem 95 on page 102 will build on this theorem.

### 2.2.3. A homomorphism

Every permutation of 4 is an automorphism of the tetrahedron. It can also be understood as a permutation of a certain set of three elements as follows.

**Theorem 54.** *There is a surjective homomorphism from* $\mathrm{Sym}(4)$ *onto* $\mathrm{Sym}(3)$.

*Proof.* Let $A$ be the set consisting of the three partitions

$$\{\{0,1\},\{2,3\}\}, \qquad \{\{0,2\},\{1,3\}\}, \qquad \{\{0,3\},\{1,2\}\}$$

of 4 into two pairs. If $\sigma \in \mathrm{Sym}(4)$, there is an element $\tilde{\sigma}$ in $\mathrm{Sym}(A)$ given by

$$\tilde{\sigma}\Big(\{\{i,j\},\{k,\ell\}\}\Big) = \Big(\{\{\sigma(i),\sigma(j)\},\{\sigma(k),\sigma(\ell)\}\}\Big).$$

Then $\sigma \mapsto \tilde{\sigma}$ is a surjective homomorphism from $\mathrm{Sym}(4)$ to $\mathrm{Sym}(A)$. □

This homomorphism will be of use later: in an example on page 113, and then in the proof of Theorem 114 on page 115, which will be used on page 117.

### 2.2.4. Cycles

We now consider symmetry groups of arbitrary sets. We shall be interested in the results mainly for finite sets; but obtaining the results for infinite sets also will take no more work. For any set $A$, for any $\sigma$ in $\mathrm{Sym}(A)$, we make the recursive definition

$$\sigma^0 = \mathrm{id}_A, \qquad \sigma^{n+1} = \sigma \circ \sigma^n.$$

If $n \in \mathbb{N}$, we also define

$$\sigma^{-n} = (\sigma^n)^{-1}.$$

Thus we have a function $n \mapsto \sigma^n$ from $\mathbb{Z}$ to $\mathrm{Sym}(A)$.

**Theorem 55.** *For every set* $A$*, for every* $\sigma$ *in* $\mathrm{Sym}(A)$*, the function* $n \mapsto \sigma^n$ *from* $\mathbb{Z}$ *to* $\mathrm{Sym}(A)$ *is a homomorphism of groups.*

*Proof.* Since $\sigma^0 = \mathrm{id}_A$ and $\sigma^{-n} = (\sigma^n)^{-1}$ for *all* $n$ in $\mathbb{Z}$, it remains to show

$$\sigma^{n+m} = \sigma^n \circ \sigma^m \tag{2.1}$$

for all $m$ and $n$ in $\mathbb{Z}$. We start with the the case where $m$ and $n$ are in $\omega$. Here we use induction on $n$. The claim holds easily if $n = 0$. Suppose it holds when $n = k$. Then

$$\begin{aligned} \sigma^{(k+1)+m} &= \sigma^{(k+m)+1} \\ &= \sigma \circ \sigma^{k+m} \\ &= \sigma \circ (\sigma^k \circ \sigma^m) \\ &= (\sigma \circ \sigma^k) \circ \sigma^m \\ &= \sigma^{k+1} \circ \sigma^m, \end{aligned}$$

and so (2.1) holds when $n = k + 1$. By induction, it holds for all $n$ in $\omega$, for all $m$ in $\omega$. Hence in this case also we have

$$\sigma^{-n-m} = (\sigma^{m+n})^{-1} = (\sigma^m \circ \sigma^n)^{-1} = \sigma^{-n} \circ \sigma^{-m}.$$

Finally, if also $m \leqslant n$, then we have $\sigma^{n-m} \circ \sigma^m = \sigma^n$, so

$$\begin{aligned} \sigma^{n-m} &= \sigma^n \circ (\sigma^m)^{-1} = \sigma^n \circ \sigma^{-m}, \\ \sigma^{m-n} = (\sigma^{n-m})^{-1} &= (\sigma^n \circ \sigma^{-m})^{-1} = \sigma^m \circ \sigma^{-n}. \end{aligned}$$

This completes all cases of (2.1). □

If $b \in A$ and $\sigma \in \mathrm{Sym}(A)$, then the set $\{\sigma^n(b) \colon n \in \mathbb{Z}\}$ is called the **orbit of $b$ under $\sigma$**. A subset of $A$ is an **orbit under** $\sigma$ if it is the orbit under $\sigma$ of some element of $A$. So for example if we think of the tetrahedron as a pyramid with an equilateral triangular base, and we let $\sigma$ be the automorphism that rotates the base clockwise by 120°, then the orbit under $\sigma$ of any vertex of the base is the set of vertices of the base.

An orbit is **trivial** if it has size 1; if it is larger, it is **nontrivial.** Then a permutation is a **cycle** if, under it, there is exactly one nontrivial orbit. Cycles are like prime numbers, by Theorem 58 below. Under the identity, there are no nontrivial cycles. As we do not consider 1 to be a prime number, so we do not consider the identity to be a cycle.

If the nontrivial orbits under some cycles are disjoint from one another, then the cycles themselves are said to be **disjoint** from one another. If $\sigma$ and $\tau$ are disjoint cycles, then $\sigma\tau = \tau\sigma$, and so on for larger numbers of disjoint cycles: the order of multiplying them makes no difference to the product. It even makes sense to talk about the product of an infinite set of disjoint cycles:

**Theorem 56.** *Suppose* $\Sigma$ *is a set of disjoint cycles in* $\operatorname{Sym}(A)$, *where the nontrivial orbit under each* $\sigma$ *in* $\Sigma$ *is* $A_\sigma$. *Then there is a unique element* $\pi$ *of* $\operatorname{Sym}(A)$ *given by*

$$\pi(x) = \begin{cases} \sigma(x), & \text{if } x \in A_\sigma, \\ x, & \text{if } x \in A \smallsetminus \bigcup_{\sigma\in\Sigma} A_\sigma. \end{cases}$$

*Proof.* The rule gives us *at least* one value of $\pi(x)$ for each $x$ in $A$; and this value is itself in $A$. But there is *at most* one value, because the sets $A_\sigma$ are known to be disjoint from one another, so that if $x \in A_\sigma$, and $\sigma \neq \tau$, then $x \notin A_\tau$. Thus $\pi$ is unique. Also $\pi\colon A \to A$. Moreover, each $\sigma$ in $\Sigma$, restricted to $A_\sigma$, is a permutation of $A_\sigma$. Thus, replacing each $\sigma$ with $\sigma^{-1}$, we obtain $\pi^{-1}$ by the given rule. Therefore $\pi \in \operatorname{Sym}(A)$. □

The permutation $\pi$ found in the theorem is the **product** of the cycles in $\Sigma$. We may denote this product by

$$\prod \Sigma.$$

In the notation of the theorem, if $i \mapsto \sigma_i$ is a bijection from some set $I$ to $\Sigma$, then we can write

$$\prod_{i\in I} \sigma_i = \prod \Sigma.$$

This function $i \mapsto \sigma_i$ can be called an **indexing** of $\Sigma$ by $I$. The product given by the theorem is independent of any indexing. If $j \mapsto \tau_j$ is an indexing of $\Sigma$ by some set $J$, then there must be a bijection $f$ from $I$ to $J$ such that $\tau_{f(i)} = \sigma_i$ for each $i$ in $I$, and so by the theorem,

$$\prod_{j\in J} \tau_j = \prod_{i\in I} \sigma_i = \prod_{i\in I} \tau_{f(i)}.$$

Next, instead of disjoint cycles, we consider disjoint orbits under some one permutation.

**Theorem 57.** *Any two distinct orbits under the same permutation are disjoint. In particular, if $a$ belongs to an orbit under $\sigma$, then that orbit is $\{\sigma^k(a)\colon k \in \mathbb{Z}\}$. If this orbit has size $n$ for some $n$ in $\mathbb{N}$, then the orbit is $\{\sigma^k(a)\colon k \in n\}$.*

*Proof.* We prove the contrapositive of the first claim. Suppose $a$ and $b$ have intersecting orbits under $\sigma$. Then for some $m$ and $n$ in $\mathbb{Z}$ we have $\sigma^m(a) = \sigma^n(b)$. In this case, for all $k$ in $\omega$,

$$\sigma^k(a) = \sigma^{n+k-m}(b).$$

Thus the orbit of $a$ is included in the orbit of $b$. By symmetry, the two orbits are the same.

For the final claim, suppose the orbit of $a$ is finite. Then for some $i$ in $\mathbb{Z}$ and $n$ in $\mathbb{N}$, we must have

$$\sigma^i(a) = \sigma^{i+n}(a). \tag{2.2}$$

Then $a = \sigma^{\pm n}(a)$, and so, by induction, for all $k$ in $\mathbb{Z}$ we have $a = \sigma^{kn}(a)$, and more generally

$$i \equiv j \Rightarrow \sigma^i(a) = \sigma^j(a) \pmod{n}.$$

Therefore, by Theorem 47, the orbit of $a$ is $\{\sigma^i\colon i \in n\}$. If $n$ is minimal such that, for some $i$, (2.2), then $n$ the size of the orbit of $a$. □

**Theorem 58.** *For every set $A$, every element of* $\operatorname{Sym}(A)$ *is uniquely the product of disjoint cycles.*

*Proof.* Supposing $\sigma \in A$, let $I$ be the set of nontrivial orbits under $\sigma$. These are all disjoint from one another, by Theorem 57. For each $i$ in $I$, we can define a unique cycle $\sigma_i$ that agrees with $\sigma$ on $i$, but otherwise is the identity. Then $\sigma = \prod_{i\in I}\sigma_i$. Suppose $\sigma = \prod\Sigma$ for some set $\Sigma$ of disjoint cycles. Then for each $i$ in $I$, we must have $\sigma_i \in \Sigma$. Moreover, $i \mapsto \sigma_i$ must be a bijection from $I$ to $\Sigma$. □

The cardinality of the unique nontrivial orbit under a cycle is the **order** of the cycle. We may say that the identity has order 1. Then orders come from the set $\mathbb{N} \cup \{\aleph_0\}$, which is $\omega' \smallsetminus \{0\}$.

### 2.2.5. Notation

Suppose $\sigma \in \operatorname{Sym}(n)$ for some $n$. Then

$$\sigma = \big\{\big(0, \sigma(0)\big), \dots, \big(n-1, \sigma(n-1)\big)\big\}.$$

We might write this equation a bit more simply in the form

$$\sigma = \left\{\begin{matrix} 0 & \dots & n-1 \\ \sigma(0) & \dots & \sigma(n-1) \end{matrix}\right\}. \tag{2.3}$$

This is a set with $n$ elements, and each of those elements is an ordered pair, here written vertically. The braces in (2.3) might be replaced with parentheses, as in

$$\begin{pmatrix} 0 & \cdots & n-1 \\ \sigma(0) & \cdots & \sigma(n-1) \end{pmatrix}.$$

However, this notation is potentially misleading, because it does not stand for a *matrix* such as we shall define in §3.1 (p. 83). In a matrix, the order of the columns (as well as the rows) matters; but in (2.3), the order of the columns does not matter. The order of the rows *does* matter. Indeed, we have

$$\left\{\begin{matrix} \sigma(0) & \dots & \sigma(n-1) \\ 0 & \dots & n-1 \end{matrix}\right\} = \sigma^{-1}.$$

Suppose $\sigma$ is a cycle, and $k$ belongs to the nontrivial orbit under it. Then we may use for $\sigma$ the notation

$$\begin{pmatrix} k & \sigma(k) & \cdots & \sigma^{m-1}(k) \end{pmatrix}, \tag{2.4}$$

where $m$ is the order of $\sigma$. By Theorem 57, we can replace $k$ with any member of the same cycle. So the expression in (2.4) should be understood, not as a matrix, but rather as a ring or a circle,[2] as in Figure 2.1

$$k$$

$$\sigma^5(k) \qquad \qquad \sigma(k)$$

$$\sigma^4(k) \qquad \qquad \sigma^2(k)$$

$$\sigma^3(k)$$

Figure 2.1. A cycle.

where $m = 6$. In general, the circle can be broken and written in one line in $m$ different ways, as

$$\begin{pmatrix}\sigma^i(k) & \cdots & \sigma^{m-1}(k) & k & \sigma(k) & \cdots & \sigma^{i-1}(k)\end{pmatrix}$$

for any $i$ in $m$. The identity $\mathrm{id}_n$ might be denoted by $(0)$, or even by $(i)$ for any $i$ in $n$.

When $n$ is small, we can just list the elements of $\mathrm{Sym}(n)$, according to their factorizations into disjoint cycles. For example, $\mathrm{Sym}(3)$ consists of

$$\begin{gathered}
(0),\\
(0\ 1),\ (0\ 2),\ (1\ 2),\\
(0\ 1\ 2),\ (0\ 2\ 1),
\end{gathered}$$

where no nontrivial factorizations are possible,, while $\mathrm{Sym}(4)$ consists of

$$\begin{gathered}
(0),\\
(0\ 1),\ (0\ 2),\ (0\ 3),\ (1\ 2),\ (1\ 3),\ (2\ 3),\\
(0\ 1\ 2),\ (0\ 1\ 3),\ (0\ 2\ 3),\ (1\ 2\ 3),\\
(0\ 1)(2\ 3),\ (0\ 2)(1\ 3),\ (0\ 3)(1\ 2),\\
(0\ 1\ 2\ 3),\ (0\ 1\ 3\ 2),\ (0\ 2\ 1\ 3),\ (0\ 2\ 3\ 1),\ (0\ 3\ 1\ 2),\ (0\ 3\ 2\ 1).
\end{gathered}$$

[2]The English word "circle" comes from the Latin *circulus* (which is a diminutive form of *circus*); "cycle" comes ultimately from the Greek κύκλος. Both *circulus* and κύκλος mean something round; and κύκλος is cognate with "wheel."

For larger $n$, one might like to have some additional principle of organization. But then the whole study of groups might be understood as a search for such principles (for organizing the elements of a group, or organizing all groups).

If $m < n$, the map $\sigma \mapsto \sigma \cup \mathrm{id}_{n \smallsetminus m}$ is an embedding of the group $\mathrm{Sym}(m)$ in $\mathrm{Sym}(n)$. Similarly each $\mathrm{Sym}(n)$ embeds in $\mathrm{Sym}(\omega)$; but the latter has many elements that are not in the image of any $\mathrm{Sym}(n)$. Indeed, we have the following, which can be obtained as a corollary of Theorem 38.

**Theorem 59.** $\mathrm{Sym}(\omega)$ *is uncountable.*

### 2.2.6. Even and odd permutations

An element of $\mathrm{Sym}(n)$ is said to be **even** if, in its factorization as a product of disjoint cycles, there is an even number of cycles of even order. Otherwise the permutation is **odd.** Thus cycles of even order are odd; cycles of odd order are even. The reason for this peculiar situation is suggested by Theorem 60 below.

Meanwhile, if $m < n$, then, under the embedding $\sigma \mapsto \sigma \cup \mathrm{id}_{n \smallsetminus m}$ just discussed of $\mathrm{Sym}(m)$ in $\mathrm{Sym}(n)$, evenness and oddness are preserved. That is, $\sigma$ in $\mathrm{Sym}(m)$ is even if and only if $\sigma \cup \mathrm{id}_{n \smallsetminus m}$ is even.

We define the **signum** function sgn from $\mathrm{Sym}(n)$ to $\{\pm 1\}$ by

$$\mathrm{sgn}(\sigma) = \begin{cases} 1, & \text{if } \sigma \text{ is even,} \\ -1, & \text{if } \sigma \text{ is odd.} \end{cases}$$

Theorem 67 on page 74 below is that this function is a homomorphism.

A cycle of order $n$ can be called an $n$**-cycle.** It is consistent with this terminology to consider the identity as a 1-cycle. A 2-cycle is also called a **transposition.**

**Theorem 60.** *Every finite permutation is a product of transpositions. A cycle of order $m$ is a product of $m-1$ transpositions.*

*Proof.* $\begin{pmatrix}0 & 1 & \cdots & m-1\end{pmatrix} = \begin{pmatrix}0 & m-1\end{pmatrix} \cdots \begin{pmatrix}0 & 2\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix}$. □

Thus an even permutation is the product of an even number of transpositions, and an odd permutation is the product of an odd number of permutations. If the converse is true, then the signum function must be a homomorphism.

However, proving that converse is not especially easy. The neatest approach might seem to be as follows. A **tournament** on set $A$ is an irreflexive, antisymmetric, total binary relation on $A$. This means, if $i$ and $j$ are distinct elements of $A$, then exactly one of $(i, j)$ and $(j, i)$ belongs to a given tournament on $A$, but $(i, i)$ never belongs. If $(i, j)$ belongs to a given tournament, we can think of $i$ as the winner of a match between $i$ and $j$; this is the reason for the name *tournament.* If $T$ is a tournament on $n$, and $\sigma \in \operatorname{Sym}(n)$, we can define

$$\tilde{\sigma}(T) = \{(\sigma(i), \sigma(j))\colon (i, j) \in T\}.$$

This is another (or possibly the same) tournament on $n$. Fixing a particular tournament $U$ on $n$, such as $\{(i, j)\colon i < j < n\}$, we let

$$A = \{\tilde{\sigma}(U)\colon \sigma \in \operatorname{Sym}(n)\}.$$

Then every $\tilde{\sigma}$, restricted to $A$, is a permutation of $A$, and indeed the map $\sigma \mapsto \tilde{\sigma} \restriction A$ is a homomorphism from $\operatorname{Sym}(n)$ to $\operatorname{Sym}(A)$. Let

$$A_0 = \{T \in A\colon |T \smallsetminus U| \text{ is even}\}, \qquad A_1 = A \smallsetminus A_0.$$

We should like to show that, for every $\sigma$ in $\operatorname{Sym}(n)$, for each $i$ in 2, the set $\{\tilde{\sigma}(T)\colon T \in A_i\}$ is $A_i$ again, if $\sigma$ is even, and $A_{1-i}$ if $\sigma$ is odd. Thus we should obtain a homomorphism from $\operatorname{Sym}(n)$ to $\operatorname{Sym}(\{A_0, A_1\})$, and the signum function would be a homomorphism. However, proving all of these things seems to be no easier than just proving directly Theorem 67 on page 74 below.

## 2.3. Monoids and semigroups

### 2.3.1. Definitions

The structure $(\mathbb{N}, 1, \cdot)$ cannot *expand* to a group, that is, it cannot be given an operation of inversion so that the structure becomes a group.

(See page 42.) The structure is however a *monoid.* A **monoid** is a structure $(M, \mathrm{e}, \cdot)$ satisfying the axioms

$$x\,\mathrm{e} = x$$
$$\mathrm{e}\,x = x,$$
$$(xy)z = x(yz).$$

In particular, if $(G, \mathrm{e}, {}^{-1}, \cdot)$ is a group, then the *reduct* $(G, \mathrm{e}, \cdot)$ is a monoid.

Not every monoid is the reduct of a group: the example of $(\mathbb{N}, 1, \cdot)$ shows this. So does the example of a set $M$ with an element e and at least one other element, if we define $xy$ to be e for all $x$ and $y$ in $M$.

For another example, given an arbitrary set $A$, we have the monoid $(A^A, \mathrm{id}_A, \circ)$. (See page 40.) However, if $A$ has at least two elements, then $A^A$ has elements (for example, constant functions) that are not injective and are therefore not invertible.

If $(M, \mathrm{e}, \cdot)$ is a monoid, then by the proof of Cayley's Theorem on page 57, the map $x \mapsto \lambda_x$ is a homomorphism from $(M, \mathrm{e}, \cdot)$ to $(M^M, \mathrm{id}_M, \circ)$. However, this homomorphism might not be an embedding.

Even though the monoid $(\mathbb{N}, 1, \cdot)$ does not expand to a group, it embeds in the monoid $(\mathbb{Q}^+, 1, \cdot)$, which expands to the group $(\mathbb{Q}^+, 1, {}^{-1}, \cdot)$, by the method of fractions learned in school and reviewed as Theorem 28 on page 43 above. There is no such embedding if we replace the monoid $(\mathbb{N}, 1, \cdot)$ with the monoid $(A^A, \mathrm{id}_A, \circ)$ for a set $A$ with at least two elements. For, in this case, Lemma 5 on page 43 is false, because multiplication on $A^A$ does not allow cancellation in the sense of Theorem 16 on page 35.

However, Theorem 28 does not actually require the identity 1 in the monoid $(\mathbb{N}, 1, \cdot)$. After appropriate modifications, the method of the theorem allows us to obtain the group $(\mathbb{Q}, 0, -, +)$ such that $(\mathbb{Q}^+, +)$ embeds in the reduct $(\mathbb{Q}, +)$. This is shown in Theorem 29 on page 45. The proof goes through, even though $(\mathbb{Q}^+, +)$ does not expand to a monoid. By the same method, $(\mathbb{Z}, 0, -, +)$ can be obtained directly from $(\mathbb{N}, +)$.

The structures $(\mathbb{N}, +)$ and $(\mathbb{Q}^+, +)$ are *semigroups.* In general, a **semigroup** is a structure $(S, \cdot)$ satisfying the identity

$$(xy)z = x(yz).$$

If $(M, \mathrm{e}, \cdot)$ is a monoid, then the reduct $(M, \cdot)$ is a semigroup. But not every semigroup is the reduct of a monoid: for example $(\mathbb{N}, +)$ and $(\mathbb{Q}^+, +)$ are not reducts of monoids. Or let $O$ be the set of all operations $f$ on $\omega^\omega$ such that, for all $n$ in $\omega$, $f(n) > n$: then $O$ is closed under composition, so $(O, \circ)$ is a semigroup; but it has no identity.

The structure $(\mathbb{Q}, 0, -, +, 1, \cdot)$ is an example of a *ring* (or more precisely associative ring); in fact it is a *field,* and it embeds in the field $(\mathbb{R}, 0, -, +, 1, \cdot)$ of real numbers, as follows from Theorem 32 on page 47. Rings and fields as such will be defined formally in §2.5, beginning on page 79.

### 2.3.2. Some homomorphisms

We defined powers of symmetries on page 62. By the same definition, we obtain at least the *positive* powers of elements of semigroups:

$$a^1 = a, \qquad\qquad a^{n+1} = a \cdot a^n.$$

**Theorem 61.** *Suppose $(S, \cdot)$ is a semigroup, and $m$ and $n$ range over $\mathbb{N}$.*

1. *For all $a$ in $S$,*
$$a^{m+n} = a^m a^n.$$
*That is, if $a \in S$, then*
$$n \mapsto a^n \colon (\mathbb{N}, +) \to (S, \cdot).$$

2. *For all $a$ in $S$,*
$$a^{mn} = (a^m)^n. \tag{2.5}$$
*That is,*
$$n \mapsto (a \mapsto a^n) \colon (\mathbb{N}, 1, \cdot) \to (S^S, \mathrm{id}_S, \circ). \tag{2.6}$$

*Proof.* We use induction. The first part is proved like Theorem 55. For the second part, we have $a^{n \cdot 1} = a^n = (a^n)^1$, and if $a^{nm} = (a^n)^m$, then

$$a^{n(m+1)} = a^{nm+n} = a^{nm} a^n = (a^n)^m a^n = (a^n)^{m+1}.$$

This establishes (2.5). If we write $f_x(y)$ for $y^x$, then (2.5) becomes

$$f_{mn} = f_n \circ f_m.$$

Since $mn = nm$, we get (2.6). □

In a monoid, we define

$$a^0 = \mathrm{e}\,.$$

**Theorem 62.** *Suppose $(M, \mathrm{e}, \cdot)$ is a monoid.*

1. *If $a \in M$, then $x \mapsto a^x \colon (\omega, 0, +) \to (M, \mathrm{e}, \cdot)$.*
2. *$x \mapsto (y \mapsto y^x) \colon (\omega, 1, \cdot) \to (M^M, \mathrm{id}_A, \circ)$.*

In a group, we define

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

**Theorem 63.** *Suppose $(G, \mathrm{e}, {}^{-1}, \cdot)$ is a group.*

1. *If $a \in G$, then $x \mapsto a^x \colon (\mathbb{Z}, 0, -, +) \to (G, \mathrm{e}, {}^{-1}, \cdot)$.*
2. *$x \mapsto (y \mapsto y^x) \colon (\mathbb{Z}, 1, \cdot) \to (G^G, \mathrm{id}_G, \circ)$.*

We shall use the following in Theorem 160 on page 156.

**Theorem 64.** *If $x^2 = \mathrm{e}$ for all $x$ in some group, then that group is abelian.*

### 2.3.3. Pi and Sigma notation

We can generalize the taking of powers in a semigroup as follows. Given elements $a_i$ of a semigroup, where $i$ ranges over $\omega$, we define certain **iterated products** recursively by

$$\prod_{i<0} a_i = 1, \qquad \prod_{i<n+1} a_i = \Bigl(\prod_{i<n} a_i\Bigr) \cdot a_n.$$

We may also write $\prod_{i<n} a_i$ as

$$a_0 \cdots a_{n-1}.$$

This product depends not just on the set $\{a_i\colon i<n\}$, but on the function $i\mapsto a_i$ on $n$. As on page 40, we may denote this function by one of

$$(a_0,\dots,a_{n-1}),\qquad\qquad (a_i\colon i<n).$$

Then the product $\prod_{i<n}a_i$ could also be written as

$$\prod(a_i\colon i<n).$$

By associativity of multiplication in semigroups, we obtain the following.

**Theorem 65.** *In a semigroup,*

$$\prod_{i<n+m}a_i=\prod_{i<n}a_i\cdot\prod_{j<m}a_{n+j}.$$

If the operation on a semigroup is commutative, we usually write it additively, and then we may define

$$\sum_{i<0}a_i=0,\qquad\qquad \sum_{i<n+1}a_i=\sum_{i<n}a_i+a_n.$$

We may also write $\sum_{i<n}a_i$ as

$$a_0+\dots+a_{n-1}.$$

However, we use multiplicative notation for the following.

**Theorem 66.** *In a commutative semigroup, for all $n$ in $\mathbb{N}$, for all $\sigma$ in* $\operatorname{Sym}(n)$*,*

$$\prod_{i<n}a_{\sigma(i)}=\prod_{i<n}a_i.$$

*Proof.* Suppose first that $\sigma$ is the transposition $(k\ \ell)$, where $k<\ell$. Let

$$b=\prod_{i<k}a_i,\qquad c=\prod_{i<\ell-k-1}a_{k+i+1},\qquad d=\prod_{i<n-\ell-1}a_{\ell+i+1}.$$

By Theorem 65 and commutativity,

$$\begin{aligned}\prod_{i<n} a_{\sigma(i)} &= b \cdot a_\ell \cdot c \cdot a_k \cdot d \\ &= b \cdot a_\ell \cdot a_k \cdot c \cdot d \\ &= b \cdot a_k \cdot a_\ell \cdot c \cdot d \\ &= b \cdot a_k \cdot c \cdot a_\ell \cdot d = \prod_{i<n} a_i.\end{aligned}$$

So the claim holds when $\sigma$ is a transposition. In this case we have

$$\prod_{i<n} a_{\tau\sigma(i)} = \prod_{i<n} a_{\tau(i)}$$

for all $\tau$ in $\mathrm{Sym}(n)$. Since every finite permutation is a product of transpositions by Theorem 60, we obtain the claim in general. □

By this theorem, if we have a function $i \mapsto a_i$ from some finite set $I$ into a commutative semigroup, then the notation

$$\prod_{i \in I} a_i$$

makes sense. We use such notation in the next theorem, Theorem 67. We may denote the function $i \mapsto a_i$ on $I$ by

$$(a_i \colon i \in I),$$

and we may refer to it as an **indexed set,** specifically as an indexed subset of the commutative semigroup in question. The set $I$ is the **index set** for this indexed set.

### 2.3.4. Alternating groups

**Theorem 67.** *The function* sgn *is a homomorphism from* $\mathrm{Sym}(n)$ *to* $\{\pm 1\}$.

*Proof.* If $\sigma \in \mathrm{Sym}(n)$, then there is a well-defined function $X \mapsto q_\sigma(X)$ from $[n]^2$ to $\{\pm 1\}$ given by

$$q_\sigma(\{i,j\}) = \frac{\sigma(i) - \sigma(j)}{i - j}.$$

Since multiplication in $\{\pm 1\}$ is commutative, we can define

$$f(\sigma) = \prod_{X \in [n]^2} q_\sigma(X).$$

If $\sigma = \begin{pmatrix} k & \ell \end{pmatrix}$, then

$$\begin{aligned} f(\sigma) &= q_\sigma(\{k,\ell\}) \cdot \prod_{i \in n \smallsetminus \{k,\ell\}} \big(q_\sigma(\{i,\ell\}) \cdot q_\sigma(\{k,i\})\big) \\ &= \frac{\ell - k}{k - \ell} \cdot \prod_{i \in n \smallsetminus \{k,\ell\}} \Big(\frac{i-k}{i-\ell} \cdot \frac{\ell - i}{k - i}\Big) \\ &= -1. \end{aligned}$$

If $\tau \in \mathrm{Sym}(n)$, we can define an element $\hat{\tau}$ of $\mathrm{Sym}([n]^2)$ by

$$\hat{\tau}(\{i,j\}) = \{\tau(i), \tau(j)\}.$$

By Theorem 66,

$$f(\sigma) = \prod_{X \in [n]^2} q_\sigma(\hat{\tau}(X)),$$

so

$$\begin{aligned} f(\sigma\tau) &= \prod_{\{i,j\} \in [n]^2} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{i - j} \\ &= \prod_{\{i,j\} \in [n]^2} \left(\frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)} \cdot \frac{\tau(i) - \tau(j)}{i - j}\right) \\ &= \prod_{X \in [n]^2} \big(q_\sigma(\hat{\tau}(X)) \cdot q_\tau(X)\big) \\ &= \prod_{X \in [n]^2} q_\sigma(\hat{\tau}(X)) \cdot \prod_{X \in [n]^2} q_\tau(X) \\ &= f(\sigma) \cdot f(\tau). \end{aligned}$$

Thus $f(\tau) = 1$ if and only if $\tau$ is the product of an even number of transpositions, and otherwise $f(\tau) = -1$. Therefore $f$ must agree with $\sigma$ on $\mathrm{Sym}(n)$, and so sgn must be a homomorphism. □

We have as a corollary that the even permutations of $n$ compose a subgroup of $\mathrm{Sym}(n)$. This subgroup is the **alternating group** of degree $n$ and is denoted by

$$\mathrm{Alt}(n).$$

If $n > 1$, there is a permutation $\sigma \mapsto \sigma \circ (0\ 1)$ of $\mathrm{Sym}(n)$ itself that takes even elements to odd. In this case, $\mathrm{Alt}(n)$ is half the size of $\mathrm{Sym}(n)$. However, $\mathrm{Alt}(1) = \mathrm{Sym}(1)$. For this reason, one may wish to say that that $\mathrm{Alt}(n)$ is defined only when $n \geqslant 2$. This makes Theorem 120 (page 119 below) simpler to state.

## 2.4. Simplifications

If a semigroup $(G, \cdot)$ expands to a group $(G, \mathrm{e}, {}^{-1}, \cdot)$, then the semigroup $(G, \cdot)$ itself is often called a group. But this usage must be justified.

**Theorem 68.** *A semigroup can expand to a group in only one way.*

*Proof.* Let $(G, \mathrm{e}, {}^{-1}, \cdot)$ be a group. If $\mathrm{e}'$ were a second identity, then

$$\mathrm{e}'\,x = \mathrm{e}\,x, \qquad \mathrm{e}'\,xx^{-1} = \mathrm{e}\,xx^{-1}, \qquad \mathrm{e}' = \mathrm{e}\,.$$

If $a'$ were a second inverse of $a$, then

$$a'a = a^{-1}a, \qquad a'aa^{-1} = a^{-1}aa^{-1}, \qquad a' = a^{-1}.$$ □

Establishing that a particular structure is a group is made easier by the following.

**Theorem 69.** *Any structure satisfying the identities*

$$\mathrm{e}x = x,$$
$$x^{-1}x = \mathrm{e},$$
$$x(yz) = (xy)z$$

*is a group. In other words, any semigroup with a left-identity and with left-inverses is a group.*

*Proof.* We need to show $x\,\mathrm{e} = x$ and $xx^{-1} = \mathrm{e}$. To establish the latter, using the given identies we have

$$(xx^{-1})(xx^{-1}) = x(x^{-1}x)x^{-1} = x\mathrm{e}x^{-1} = xx^{-1},$$

and so

$$xx^{-1} = \mathrm{e}xx^{-1} = (xx^{-1})^{-1}(xx^{-1})(xx^{-1}) = (xx^{-1})^{-1}(xx^{-1}) = \mathrm{e}.$$

Hence also

$$x\mathrm{e} = x(x^{-1}x) = (xx^{-1})x = \mathrm{e}x = x. \qquad \square$$

The theorem has an obvious "dual" involving right-identities and right-inverses. By the theorem, the semigroups that expand to groups are precisely the semigroups that satisfy the axiom

$$\exists z\ (\forall x\ zx = x \wedge \forall x\ \exists y\ yx = z),$$

which is logically equivalent to

$$\exists z\ \forall x\ \forall y\ \exists u\ (zx = x \wedge uy = z). \tag{2.7}$$

We shall show that this sentence is more complex than need be.

Thanks to Theorem 68, if a semigroup $(G, \cdot)$ does expand to a group, then we may unambiguously refer to $(G, \cdot)$ itself as a group. Furthermore, we may refer to $G$ as a group: this is commonly done, although, theoretically, it may lead to ambiguity.

**Theorem 70.** *Let $G$ be a nonempty semigroup. The following are equivalent.*

1. *$G$ expands to a group.*
2. *Each equation $ax = b$ and $ya = b$ with parameters from $G$ has a solution in $G$.*
3. *Each equation $ax = b$ and $ya = b$ with parameters from $G$ has a unique solution in $G$.*

*Proof.* Immediately (3)⇒(2). Almost as easily, (1)⇒(3). For, if $a$ and $b$ belong to some semigroup that expands to a group, we have $ax = b \Leftrightarrow x = a^{-1}b$; and we know by Theorem 68 that $a^{-1}$ is uniquely determined. Likewise for $ya = b$.

Finally we show (2)⇒(1). Suppose $G$ is a nonempty semigroup in which all equations $ax = b$ and $ya = b$ have solutions. If $c \in G$, let $\mathrm{e}$ be a solution to $yc = c$. If $b \in G$, let $d$ be a solution to $cx = b$. Then

$$\mathrm{e}b = \mathrm{e}(cd) = (\mathrm{e}c)d = cd = b.$$

Since $b$ was chosen arbitrarily, $\mathrm{e}$ is a left identity. Since the equation $yc = \mathrm{e}$ has a solution, $c$ has a left inverse. But $c$ is an arbitrary element of $G$. By Theorem 69, we are done. □

Now we have that the semigroups that expand to groups are just the semigroups that satisfy the axiom

$$\forall x \; \forall y \; \exists z \; \exists w \; (xz = y \wedge wx = y).$$

This may not look simpler than (2.7), but it is. It should be understood as

$$\forall x \; \forall y \; \exists z \; \exists w \; (xz = y \wedge wx = y),$$

which is a sentence of the general form $\forall\exists$; whereas (2.7) is of the form $\exists\forall\exists$).

**Theorem 71.** *A map $f$ from one group to another is a homomorphism, provided it is a homomorphism of semigroups, that is, $f(xy) = f(x)f(y)$.*

*Proof.* In a group, if $a$ is an element, then the identity is the unique solution of $xa = a$, and $a^{-1}$ is the unique solution of $yaa = a$. A semigroup homomorphism $f$ takes solutions of these equations to solutions of $xb = b$ and $ybb = b$, where $b = f(a)$. □

*Inclusion* of a substructure in a larger structure is a homomorphism. In particular, if $(G, \mathrm{e}, {}^{-1}, \cdot)$ and $(H, \mathrm{e}, {}^{-1}, \cdot)$ are groups, we have

$$(G, \cdot) \subseteq (H, \cdot) \implies (G, \mathrm{e}, {}^{-1}, \cdot) \subseteq (H, \mathrm{e}, {}^{-1}, \cdot).$$

If an arbitrary class of structures is axiomatized by $\forall\exists$ sentences, then the class is "closed under unions of chains" in the sense that, if $\mathfrak{A}_0 \subseteq \mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \cdots$, where each $\mathfrak{A}_k$ belongs to the class, then the union of all of these structures also belongs to the class. In fact the converse is also true, by the so-called Chang–Łoś–Suszko Theorem [3, 25]. With this theorem, and with Theorem 71 in place of 70, we can still conclude that the theory of groups in the signature $\{\cdot\}$ has $\forall\exists$ axioms, although we may not know what they are.

Theorem 71 fails with monoids in place of groups. For example, $(\mathbb{Z}, 1, \cdot)$ and $(\mathbb{Z} \times \mathbb{Z}, (1, 1), \cdot)$ are monoids (the latter being the product of the former with itself as defined in §3.2), and $x \mapsto (x, 0)$ is an embedding of the semigroup $(\mathbb{Z}, \cdot)$ in $(\mathbb{Z} \times \mathbb{Z}, \cdot)$, but it is not an embedding of the monoids.

## 2.5. Associative rings

A homomorphism from a structure to itself is an **endomorphism.** Recall from page 54 that a group in which the multiplication is commutative is said to be an **abelian group,** and (page 55) its operation is usually written additively. The set of endomorphisms of an abelian group can be made into an abelian group in which:

1) the identity is the constant function $x \mapsto \mathrm{e}$;
2) additive inversion converts $f$ to $x \mapsto -f(x)$;
3) addition converts $(f, g)$ to $x \mapsto f(x) + g(x)$.

If $E$ is an abelian group, let the abelian group of its endomorphisms be denoted by

$$\mathrm{End}(E).$$

The set of endomorphisms of $E$ can also be made into a monoid in which the identity is the identity function $\mathrm{id}_E$, and multiplication is functional composition. This multiplication distributes in both senses over addition:

$$f \circ (g + h) = f \circ g + f \circ h, \qquad (f + g) \circ h = f \circ h + g \circ h.$$

We may denote the two combined structures—abelian group and monoid together—by

$$(\mathrm{End}(E), \mathrm{id}_E, \circ);$$

this is the **complete ring of endomorphisms of** $E$. A substructure of $(\mathrm{End}(E), \mathrm{id}_E, \circ)$ can be called simply a **ring of endomorphisms** $E$.

An **associative ring** is a structure $(R, 0, -, +, 1, \cdot)$ such that

1) $(R, 0, -, +)$ is an abelian group,
2) $(R, 1, \cdot)$ is a monoid,
3) the multiplication distributes in both senses over addition.

Then rings of endomorphisms are associative rings.[3] It may be convenient to write an associative ring as $(R, 1, \cdot)$, where $R$ is implicitly an abelian group. We might even say simply that $R$ is an associative ring.

An associative ring is usually just called a ring; however, we shall consider some rings that are not associative rings in §6.2 (page 174). Some authors might not require an associative ring to have a multiplicative identity.[4] We require it, so that the next theorem holds. As with a group, so with an associative ring, an element $a$ determines a singulary operation $\lambda_a$ on the structure, the operation being given by

$$\lambda_a(x) = ax.$$

Then we have an analogue of Cayley's Theorem (page 57):

**Theorem 72.** *For every associative ring* $(R, 1, \cdot)$, *the function*

$$x \mapsto \lambda_x$$

*embeds* $(R, 1, \cdot)$ *in* $(\mathrm{End}(R), \mathrm{id}_R, \circ)$.

In an associative ring, if the multiplication commutes, then the ring is a **commutative ring.** For example, $(\mathbb{Z}, 0, -, +, 1, \cdot)$ and $(\mathbb{Q}, 0, -, +, 1, \cdot)$ are commutative rings. The following is easy to check, but can be seen as a consequence of Theorem 85 on page 94 below, which is itself easy to prove, especially given Theorem 84.

[3]See note 1 on page 188 for the origin of the term *ring.*

[4]For Lang [23, ch. II, §1, p. 83], a ring is what we have defined as an associative ring. For Hungerford [19, ch. III, §1, p. 115], what we call an associative ring is a *ring with identity.*

**Theorem 73.** $(\mathbb{Z}_n, 0, -, +, 1, \cdot)$ *is a commutative ring.*

In an associative ring, an element with both a left and a right multiplicative inverse can be called simply **invertible;** it is also called a **unit.**

**Theorem 74.** *In an associative ring, the units compose a group with respect to multiplication. In particular, a unit has a unique left inverse, which is also a right inverse.*

The group of units of an associative ring $R$ is denoted by

$$R^\times.$$

For example, $\mathbb{Z}^\times = \{1, -1\}$. Evidently all two-element groups are isomorphic to this one.

By the theorem, if an element of an associative ring has both a left inverse and a right inverse, then they are equal. However, possibly an element can have a right inverse, but not a left inverse. We can construct an example by means of the following.

**Theorem 75.** *If $I$ is a set and $G$ is a group, then the set $G^I$ of functions from $I$ to $G$ is a group with multiplication given by*

$$(x_i \colon i \in I) \cdot (y_i \colon i \in I) = (x_i \cdot y_i \colon i \in I).$$

Now let $G$ be any nontrivial group. An arbitrary element $(x_n \colon n \in \omega)$ of $G^\omega$ can be written also as

$$(x_0, x_1, \dots).$$

Then $\mathrm{End}(G^\omega)$ contains elements $f$ and $g$ given by

$$\begin{aligned} f(x_0, x_1, \dots) &= (x_1, x_2, x_3, x_4, \dots), \\ g(x_0, x_1, \dots) &= (x_0, x_0, x_1, x_2, \dots), \end{aligned}$$

so that

$$\begin{aligned} fg(x_0, x_1, \dots) &= (x_0, x_1, x_2, \dots), \\ gf(x_0, x_1, \dots) &= (x_1, x_1, x_2, \dots). \end{aligned}$$

In particular, $g$ is a right inverse of $f$, but not a left inverse. The construction in Theorem 75 will be generalized on page 121.

If $R$ is a commutative ring, and $R^\times = R \setminus \{0\}$, then $R$ is called a **field.** For example, $\mathbb{Q}$ and $\mathbb{R}$ are fields. The field $\mathbb{C}$ can be defined as $\mathbb{R} \times \mathbb{R}$ with the appropriate operations: see page 97.

The trivial group $\{0\}$ becomes the trivial associative ring when we define $1 = 0$ and $0 \cdot 0 = 0$. This ring is not a field, because its only element 0 is a unit.

# 3. Groups

## 3.1. *General linear groups

The purpose of this section is to define some families of examples of groups, besides the finite symmetry groups $\operatorname{Sym}(n)$.

By Cayley's Theorem, page 57, we know that every finite group embeds, for some $n$ in $\omega$, in $\operatorname{Sym}(n)$. We know in turn (from page 68) that each $\operatorname{Sym}(n)$ embeds in $\operatorname{Sym}(\omega)$, which however is uncountable by Theorem 59. For every commutative ring $R$, for every $n$ in $\omega$, we shall define the group $\mathrm{GL}_n(R)$ of *invertible $n \times n$ matrices over $R$.* Both $\operatorname{Sym}(n)$ and $R^\times$ embed in $\mathrm{GL}_n(R)$. If $R$ is countable, then so is $\mathrm{GL}_n(R)$. If $R$ is finite, then so is $\mathrm{GL}_n(R)$. In any case, $\mathrm{GL}_n(R)$ can be understood as the automorphism group of $R^n$, when this is considered as an *$R$-module.*

We shall use invertible matrices over $\mathbb{Z}$ in classifying the *finitely generated* abelian groups, in §4.7 (page 139).

### 3.1.1. Additive groups of matrices

For any commutative ring $R$, for any two elements $m$ and $n$ of $\omega$, a function $(i,j) \mapsto a^i_j$ from $m \times n$ to $R$ can be called an $m \times n$ **matrix over** $R$ and denoted by the expression

$$\begin{pmatrix} a^0_0 & \cdots & a^0_{n-1} \\ \vdots & \ddots & \vdots \\ a^{m-1}_0 & \cdots & a^{m-1}_{n-1} \end{pmatrix},$$

which has $m$ rows and $n$ columns. We may abbreviate this matrix to

$$(a^i_j)^{i<m}_{j<n},$$

or simply

$$(a^i_j)^i_j$$

if the sets over which $i$ and $j$ range is clear. The **entries** $a^i_j$ are from $R$. The set of all $m \times n$ matrices over $R$ can be denoted by

$$\mathrm{M}_{m\times n}(R).$$

This is an abelian group in the obvious way, with addition defined by

$$(a^i_j)^{i<m}_{j<n} + (b^i_j)^{i<m}_{j<n} = (a^i_j + b^i_j)^{i<m}_{j<n}.$$

### 3.1.2. Multiplication of matrices

Given any three elements $m$, $s$, and $n$ of $\omega$, we define **multiplication** as a function from the product $\mathrm{M}_{m\times s}(R) \times \mathrm{M}_{s\times n}(R)$ to $\mathrm{M}_{m\times n}(R)$ by

$$(a^i_j)^{i<m}_{j<s} \cdot (b^j_k)^{j<s}_{k<n} = \Bigl(\sum_{j\in s} a^i_j b^j_k\Bigr)^{i<m}_{k<n}.$$

Then in particular multiplication is a binary operation on each group $\mathrm{M}_{n\times n}(R)$ of *square* matrices. One particular element of this group is

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix},$$

which can be denoted by

$$\mathrm{I}_n.$$

This matrix can also be written as $(\delta^i_j)^{i<n}_{j<n}$, where

$$\delta^i_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases}$$

**Theorem 76.** *For all commutative rings $R$, multiplication of matrices over $R$ is associative and distributes over addition. Also $\mathrm{M}_{n\times n}(R)$ is an associative ring with multiplicative identity $\mathrm{I}_n$.*

The group $\mathrm{M}_{n\times n}(R)^{\times}$ is called the **general linear group** of degree $n$ over $R$; it is also denoted by

$$\mathrm{GL}_n(R).$$

Some elements of $\mathrm{GL}_n(R)$ are picked out by the following.

**Theorem 77.** *For each $n$ in $\omega$, there is an embedding of* $\mathrm{Sym}(n)$ *in* $\mathrm{GL}_n(R)$, *namely*

$$\sigma \mapsto (\delta_j^{\sigma^{-1}(i)})_{j<n}^{i<n}.$$

*Proof.* The given function is evidently injective. It is a homomorphism since

$$(\delta_j^{\sigma^{-1}(i)})_j^i \cdot (\delta_j^{\tau^{-1}(i)})_j^i = \Bigl(\sum_{k<n} \delta_k^{\sigma^{-1}(i)} \cdot \delta_j^{\tau^{-1}(k)}\Bigr)_j^i = (\delta_j^{\tau^{-1}(\sigma^{-1}(i))})_j^i \qquad \square$$

If $R$ is a field, there is an algorithm called **Gauss–Jordan elimination,**learned in linear algebra classes, for determining whether a given element $A$ of $\mathrm{M}_{n\times n}(R)$ is invertible. One systematically performs certain invertible operations on the rows of $A$, attempting to transform it into $\mathrm{I}_n$. These operations are called **elementary row operations,** and they are:

1) interchanging two rows,
2) adding a multiple of one row by an element of $R$ to another, and
3) multiplying a row by an element of $R^{\times}$.

One works through the matrix from left to right, first converting a nonzero element of the first column to 1, and using this to eliminate the other nonzero entries; then continuing with the second column, and so on. One will be successful in transforming $A$ to $\mathrm{I}_n$ if and only if $A$ is indeed invertible. In this case, the same elementary row operations, performed on the rows of $\mathrm{I}_n$, will produce $A^{-1}$. The reason is that performing each of these operations is the same as multiplying from the left by the result of performing the same operation on $\mathrm{I}_n$.

When $R$ is $\mathbb{Z}$, one can instead use the Euclidean algorithm to make one entry in each column of $A$ equal to the *greatest common divisor* of all of the entries in that column. (See page 100.) Then $A$ is invertible if and only if each of these greatest common divisors is 1.

We now develop a method for determining whether a matrix over an arbitrary ring is invertible.

### 3.1.3. Determinants of matrices

Given a commutative ring $R$, we define the function $X \mapsto \det(X)$ from $\mathrm{M}_{n\times n}(R)$ to $R$ by

$$\det((a^i_j)^{i<n}_{j<n}) = \sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma)\prod_{i<n} a^i_{\sigma(i)}.$$

Here $\det(A)$ is the **determinant** of $A$.

**Theorem 78.** *The function $X \mapsto \det(X)$ is a multiplicative homomorphism, that is,*

$$\det(XY) = \det(X)\cdot\det(Y).$$

*Proof.* We shall use the identity

$$\prod_{i<k}\sum_{j<n} f(i,j) = \sum_{\varphi\colon k\to n}\prod_{i<k} f(i,\varphi(i)).$$

Let $A = (a^i_j)^{i<n}_{j<n}$ and $B = (b^i_j)^{i<n}_{j<n}$. Then

$$\begin{aligned}\det(AB) &= \det((\sum_{j<n} a^i_j b^j_k)^{i<n}_{k<n})\\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma)\prod_{i<n}\sum_{j<n} a^i_j b^j_{\sigma(i)}\\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma)\sum_{\varphi\colon n\to n}\prod_{i<n}(a^i_{\varphi(i)} b^{\varphi(i)}_{\sigma(i)})\\ &= \sum_{\varphi\colon n\to n}\prod_{i<n} a^i_{\varphi(i)}\sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma)\prod_{i<n} b^{\varphi(i)}_{\sigma(i)}.\end{aligned}$$

We shall eliminate from the sum those terms in any $\varphi$ that is not injective. Suppose $k<\ell<n$, but $\varphi(k)=\varphi(\ell)$. The function $\sigma\mapsto\sigma\circ\begin{pmatrix}k & \ell\end{pmatrix}$ is a

bijection between $\mathrm{Alt}(n)$ and $\mathrm{Sym}(n) \smallsetminus \mathrm{Alt}(n)$. Writing $\sigma'$ for $\sigma \circ \begin{pmatrix} k & \ell \end{pmatrix}$, we have

$$\sum_{\sigma \in \mathrm{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{i<n} b^{\varphi(i)}_{\sigma(i)} = \sum_{\sigma \in \mathrm{Alt}(n)} \operatorname{sgn}(\sigma) (\prod_{i<n} b^{\varphi(i)}_{\sigma(i)} - \prod_{i<n} b^{\varphi(i)}_{\sigma'(i)}).$$

Each term of the last sum is 0, since $\sigma$ and $\sigma'$ agree on $n \smallsetminus \{k, \ell\}$, while

$$b^{\varphi(k)}_{\sigma(k)} b^{\varphi(\ell)}_{\sigma(\ell)} = b^{\varphi(\ell)}_{\sigma'(\ell)} b^{\varphi(k)}_{\sigma'(k)} = b^{\varphi(k)}_{\sigma'(k)} b^{\varphi(\ell)}_{\sigma'(\ell)}.$$

Therefore, continuing with the computation above, we have

$$\det(AB) = \sum_{\tau \in \mathrm{Sym}(n)} \prod_{i<n} a^i_{\tau(i)} \sum_{\sigma \in \mathrm{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{i<n} b^{\tau(i)}_{\sigma(i)}.$$

Since each $\tau$ in $\mathrm{Sym}(n)$ permutes $n$, we have also

$$\prod_{i<n} b^{\tau(i)}_{\sigma(i)} = \prod_{i<n} b^i_{\sigma\tau^{-1}(i)}, \qquad \operatorname{sgn}(\sigma) = \operatorname{sgn}(\tau) \cdot \operatorname{sgn}(\sigma\tau^{-1}).$$

Putting this all together, we have

$$\begin{aligned}
\det(AB) &= \sum_{\tau \in \mathrm{Sym}(n)} \prod_{i<n} a^i_{\tau(i)} \sum_{\sigma \in \mathrm{Sym}(n)} \operatorname{sgn}(\tau) \operatorname{sgn}(\sigma\tau^{-1}) \prod_{i<n} b^i_{\sigma\tau^{-1}(i)} \\
&= \sum_{\tau \in \mathrm{Sym}(n)} \operatorname{sgn}(\tau) \prod_{i<n} a^i_{\tau(i)} \sum_{\sigma \in \mathrm{Sym}(n)} \operatorname{sgn}(\sigma\tau^{-1}) \prod_{i<n} b^i_{\sigma\tau^{-1}(i)} \\
&= \sum_{\tau \in \mathrm{Sym}(n)} \operatorname{sgn}(\tau) \prod_{i<n} a^i_{\tau(i)} \sum_{\sigma \in \mathrm{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{i<n} b^i_{\sigma(i)} \\
&= \det(A) \cdot \det(B),
\end{aligned}$$

since $\sigma \mapsto \sigma\tau^{-1}$ is a permutation of $\mathrm{Sym}(n)$. □

**Corollary 78.1.** *An element of* $\mathrm{M}_{n\times n}(R)$ *has an inverse only if its determinant is in* $R^\times$*.*

### 3.1.4. Inversion of matrices

Given the commutative ring $R$, we can now characterize the elements of $\mathrm{GL}_n(R)$ among elements of $\mathrm{M}_{n\times n}(R)$ by establishing the converse of Corollary 78.1.

**Theorem 79.** *An element of* $\mathrm{M}_{n\times n}(R)$ *has an inverse if its determinant is in* $R^\times$*.*

*Proof.* Let $A=(a^i_j)^{i<n}_{j<n}$. If $i<n$, then

$$\begin{aligned}\det(A)&=\sum_{\sigma\in\mathrm{Sym}(n)}\operatorname{sgn}(\sigma)\cdot\prod_{\ell<n}a^\ell_{\sigma(\ell)}\\&=\sum_{\sigma\in\mathrm{Sym}(n)}\operatorname{sgn}(\sigma)\cdot a^i_{\sigma(i)}\prod_{\ell\in n\smallsetminus\{i\}}a^\ell_{\sigma(\ell)}\\&=\sum_{j<n}a^i_j\sum_{\substack{\sigma\in\mathrm{Sym}(n)\\ \sigma(i)=j}}\operatorname{sgn}(\sigma)\cdot\prod_{\ell\in n\smallsetminus\{i\}}a^\ell_{\sigma(\ell)}\\&=\sum_{j<n}a^i_jb^j_i,\end{aligned}$$

where in general

$$b^j_k=\sum_{\substack{\sigma\in\mathrm{Sym}(n)\\ \sigma(k)=j}}\operatorname{sgn}(\sigma)\cdot\prod_{\ell\in n\smallsetminus\{k\}}a^\ell_{\sigma(\ell)}.$$

If $i\neq k$, then

$$\begin{aligned}\sum_{j<n}a^i_jb^j_k&=\sum_{j<n}a^i_j\sum_{\substack{\sigma\in\mathrm{Sym}(n)\\ \sigma(k)=j}}\operatorname{sgn}(\sigma)\cdot\prod_{\ell\in n\smallsetminus\{k\}}a^\ell_{\sigma(\ell)}\\&=\sum_{\sigma\in\mathrm{Sym}(n)}\operatorname{sgn}(\sigma)\cdot a^i_{\sigma(k)}\prod_{\ell\in n\smallsetminus\{k\}}a^\ell_{\sigma(\ell)}\\&=\sum_{\sigma\in\mathrm{Sym}(n)}\operatorname{sgn}(\sigma)\cdot a^i_{\sigma(k)}a^i_{\sigma(i)}\prod_{\ell\in n\smallsetminus\{i,k\}}a^\ell_{\sigma(\ell)}=0,\end{aligned}$$

since the map $\sigma\mapsto\sigma\circ\begin{pmatrix}i & k\end{pmatrix}$ is a bijection between $\mathrm{Alt}(n)$ and $\mathrm{Sym}(n)\smallsetminus\mathrm{Alt}(n)$. Thus

$$A\cdot(b^j_k)^{j<n}_{k<n}=(\det(A)\cdot\delta^i_k)^{i<n}_{k<n}.$$

Finally,

$$\begin{aligned}\sum_{j<n} b^i_j a^j_k &= \sum_{j<n} \sum_{\substack{\sigma\in\mathrm{Sym}(n)\\ \sigma(j)=i}} \operatorname{sgn}(\sigma)\cdot \prod_{\ell\in n\smallsetminus\{j\}} a^\ell_{\sigma(\ell)} a^j_k \\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma)\cdot \prod_{\ell\in n\smallsetminus\{\sigma^{-1}(i)\}} a^\ell_{\sigma(\ell)} a^{\sigma^{-1}(i)}_k \\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma)\cdot \prod_{\ell\in n\smallsetminus\{i\}} a^{\sigma^{-1}(\ell)}_\ell a^{\sigma^{-1}(i)}_k,\end{aligned}$$

which is $\det(A)$ if $i = k$, but is otherwise 0, so

$$(b^i_j)^{i<n}_{j<n} A = (\det(A)\delta^i_k)^{i<n}_{k<n}.$$

In particular, if $\det(A)$ is invertible, then so is $A$, and

$$A^{-1} = (\det(A)^{-1} b^j_k)^{j<n}_{k<n}. \qquad \square$$

Thus

$$\mathrm{GL}_n(R) = \{X \in \mathrm{M}_{n\times n}(R)\colon\ \det(X) \in R^\times\}.$$

In the $2 \times 2$ case, if $ad - bc = 1$, we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

### 3.1.5. Modules and vector-spaces

A **module** is a kind of structure with two universes. One of these is the universe of a commutative ring $R$, and other is the universe of an abelian group $M$. Furthermore, there is a function $(x, \boldsymbol{m}) \mapsto x \cdot \boldsymbol{m}$ from $R \times M$ to $M$ such that the function $x \mapsto (\boldsymbol{m} \mapsto x \cdot \boldsymbol{m})$ is a homomorphism from $R$ to $(\mathrm{End}(M), \mathrm{id}_M, \circ)$. Then we can understand $M$ as a group equipped with a certain additional operation for each element of $R$. In this sense, $M$ is a **module over** $R$, or an $R$**-module.**

For example, $R$ is a module over itself. A module over a *field* is called a **vector space.** In this case, the associated homomorphism from $R$ to $(\mathrm{End}(M), \mathrm{id}_M, \circ)$ is an embedding, unless $M$ is the trivial group.

The foregoing definition of modules makes sense, even if $R$ is not commutative; but in that case what we have defined is a **left** module. We restrict our attention to the commutative case.

We further restrict our attention to the case where $M$ is the group $\mathrm{M}_{n\times 1}(R)$ for some $n$ in $\omega$. A typical element of this group can be written as either of

$$\boldsymbol{x}, \qquad (x^i : i < n);$$

thus it can be identified with an element of $R^n$. The group becomes an $R$-module when we make the obvious definition

$$r \cdot \boldsymbol{x} = (r \cdot x^i : i < n).$$

**Theorem 80.** *For every commutative ring $R$, for every $n$ in $\omega$, there is an isomorphism from $\mathrm{GL}_n(R)$ to $\mathrm{Aut}(R^n)$, namely*

$$A \mapsto (\boldsymbol{x} \mapsto A \cdot \boldsymbol{x}). \tag{3.1}$$

*Proof.* By Theorem 76, if $A \in \mathrm{GL}_n(R)$, then the operation $\boldsymbol{x} \mapsto A \cdot \boldsymbol{x}$ is a group endomorphism. Being invertible, it is an group automorphism. By commutativity of $R$ (and the definition of matrix multiplication), for all $r$ in $R$,

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

Hence the function in (3.1) is indeed a homomorphism $h$ from $\mathrm{GL}_n(R)$ to $\mathrm{Aut}(R^n)$. To show that it is a bijection onto $\mathrm{Aut}(R^n)$, we use the notation

$$\mathbf{e}_j = (\delta^i_j : i < n),$$

so that

$$\boldsymbol{x} = \sum_{i<n} x^i \cdot \mathbf{e}_i.$$

If $A = (a^i_j)^{i<n}_{j<n}$, then

$$A \cdot \mathbf{e}_j = (a^i_j : i < n),$$

which is the number-$j$ column of $A$. This shows $\ker(h)$ is trivial. To show that $h$ is surjective onto $\mathrm{Aut}(R^n)$, suppose $f \in \mathrm{Aut}(R^n)$ and $f(\mathbf{e}_i) =$

$(a_i^j : j < n)$. Then

$$\begin{aligned}
f(\boldsymbol{x}) &= f\Big(\sum_{i<n} x_i \cdot \mathbf{e}_i\Big) \\
&= \sum_{i<n} x^i \cdot f(\mathbf{e}_i) \\
&= \sum_{i<n} x^i \cdot (a_i^j : j < n) \\
&= \Big(\sum_{i<n} x^i \cdot a_i^j : j < n\Big) \\
&= A \cdot \boldsymbol{x},
\end{aligned}$$

where $A = (a_j^i)_{j<n}^{i<n}$. Thus $f = h(A)$. □

By composing the isomorphism in the theorem with the embedding of $\mathrm{Sym}(n)$ in $\mathrm{GL}_n(R)$ given by Theorem 77, we obtain the embedding of $\mathrm{Sym}(n)$ in $\mathrm{Aut}(R^n)$ discussed (in case $R = \mathbb{R}$) on page 60 above.

## 3.2. New groups from old

### 3.2.1. Products

If $\mathfrak{A}$ and $\mathfrak{B}$ are two algebras with the same signature, then their **direct product,** denoted by

$$\mathfrak{A} \times \mathfrak{B},$$

is defined in the obvious way: the universe is $A \times B$, and for every $n$ in $\omega$, for every $n$-ary operation-symbol $f$ of the signature of $\mathfrak{A}$ and $\mathfrak{B}$,

$$f^{\mathfrak{A}\times\mathfrak{B}}\big((x_i, y_i) \colon i < n\big) = \big(f^{\mathfrak{A}}(x_i \colon i < n), f^{\mathfrak{B}}(y_i \colon i < n)\big).$$

In the special case where $\mathfrak{A}$ and $\mathfrak{B}$ are groups, we have

$$(x_0, y_0) \cdot^{\mathfrak{A}\times\mathfrak{B}} (x_1, y_1) = (x_0 \cdot^{\mathfrak{A}} x_1, y_0 \cdot^{\mathfrak{B}} y_1),$$

or more simply

$$(x_0, y_0)(x_1, y_1) = (x_0 x_1, y_0 y_1).$$

**Theorem 81.** *The direct product of two*

(a) *groups is a group,*
(b) *associative rings is an associative ring,*
(c) *commutative rings is a commutative ring.*

If $G$ and $H$ are abelian, written additively, then their direct product is usually called a **direct sum,** denoted by

$$G \oplus H.$$

The direct sum $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ is the **Klein four group,** denoted by

$$\mathrm{V}_4$$

(for *Vierergruppe*[1] ). This is the smallest group containing two elements neither of which is a power of the other.

**Theorem 82.** *If $\mathfrak{A}$ and $\mathfrak{B}$ are two algebras with the same signature, then the functions*

$$(x, y) \mapsto x, \qquad (x, y) \mapsto y$$

*are homomorphisms from $\mathfrak{A} \times \mathfrak{B}$ to $\mathfrak{A}$ and $\mathfrak{B}$ respectively.*

**Theorem 83.** *If $\mathfrak{A}$ and $\mathfrak{B}$ are two groups or two associative rings, then the functions*

$$x \mapsto (x, \mathrm{e}), \qquad y \mapsto (\mathrm{e}, y)$$

*are homomorphisms from $\mathfrak{A}$ and $\mathfrak{B}$ respectively to $\mathfrak{A} \times \mathfrak{B}$.*

### 3.2.2. Quotients

The groups $(\mathbb{Z}_n, 0, -, +)$ and the rings $(\mathbb{Z}_n, 0, -, +, 1, \cdot)$ are instances of a general construction.

[1] According to Wikipedia, Klein gave this name to the group in 1884, but the name was later applied to four-person anti-Nazi resistance groups.

Suppose $\sim$ is an equivalence-relation on a set $A$, so that it partitions $A$ into equivalence-classes
$$\{x \in A \colon x \sim a\};$$
each such class can be denoted by an expression like one of the following:
$$a/{\sim}, \qquad [a], \qquad \overline{a}.$$
Each element of an equivalence-class is a **representative** of that class. The **quotient** of $A$ by $\sim$ is the set of equivalence-classes of $A$ with respect to $\sim$; this set can be denoted by
$$A/{\sim}.$$
Suppose for some $n$ in $\omega$ and some set $B$, we have $f \colon A^n \to B$. Then there may or may not be a function $\tilde{f}$ from $(A/{\sim})^n$ to $B$ such that the equation
$$\tilde{f}([x_0], \dots, [x_{n-1}]) = f(x_0, \dots, x_{n-1}) \tag{3.2}$$
is an identity. If there is such a function $\tilde{f}$, then it is unique. In this case, the function $\tilde{f}$ is said to be **well-defined** by the given identity (3.2). Note however that there are no "ill-defined" functions. An ill-defined function would be a nonexistent function. The point is that choosing a function $f$ and writing down the equation (3.2) does not automatically give us a function $\tilde{f}$. To know that there is such a function, we must check that
$$a_0 \sim x_0 \wedge \dots \wedge a_{n-1} \sim x_{n-1} \Rightarrow f(a_0, \dots, a_{n-1}) = f(x_0, \dots, x_{n-1}).$$
When this does hold (for all $a_i$), so that $\tilde{f}$ exists as in (3.2), then
$$\tilde{f} \circ \mathrm{p} = f, \tag{3.3}$$
where p is the function $(x_0, \dots x_{n-1}) \mapsto ([x_0], \dots, [x_{n-1}])$ from $A^n$ to $(A/{\sim})^n$. Another way to express the equation (3.3) is to say that the following diagram **commutes:**

$$\begin{tikzcd}
A^n \arrow[r, "f"] \arrow[d, "\mathrm{p}"'] & B \\
(A/{\sim})^n \arrow[ur, "\tilde{f}"']
\end{tikzcd}$$

Suppose now $\mathfrak{A}$ is an algebra with universe $A$. If for all $n$ in $\omega$, for every distinguished $n$-ary operation $f$ of $\mathfrak{A}$, there is an $n$-ary operation $\tilde{f}$ on $(A/{\sim})^n$ as given by (3.2), then $\sim$ is a **congruence-relation** or **congruence** on $\mathfrak{A}$. In this case, the $\tilde{f}$ are the distinguished operations of a structure with universe $A/{\sim}$. This new structure is the **quotient** of $\mathfrak{A}$ by $\sim$ and can be denoted by

$$\mathfrak{A}/{\sim}.$$

For example, by Theorem 46 on page 56, for each $n$ in $\mathbb{N}$, congruence *modulo* $n$ is a congruence on $(\mathbb{Z}, 0, -, +, 1, \cdot)$. Then the structure $(\mathbb{Z}_n, 0, -, +)$ can be understood as the quotient $(\mathbb{Z}, 0, -, +)/{\sim}$, and $(\mathbb{Z}_n, 0, -, +, 1, \cdot)$ as $(\mathbb{Z}, 0, -, +, 1, \cdot)/{\sim}$. The former quotient is an abelian group by Theorem 48, and the latter quotient is a commutative ring by Theorem 73 on page 81. These theorems are special cases of the next two theorems. In fact the first of these makes verification of Theorem 48 easier.

**Theorem 84.** *Suppose $\sim$ is a congruence-relation on a semigroup $(G, \cdot)$.*

1. *$(G, \cdot)/{\sim}$ is a semigroup.*
2. *If $(G, \cdot)$ expands to a group, then $\sim$ is a congruence-relation on this group, and the quotient of the group by $\sim$ is a group. If the original group is abelian, then so is the quotient.*

**Theorem 85.** *Suppose $(R, 0, -, +, 1, \cdot)$ is an associative ring, and $\sim$ is a congruence-relation on the reduct $(R, +, \cdot)$. Then $\sim$ is a congruence-relation on $(R, 0, -, +, 1, \cdot)$, and the quotient $(R, 0, -, +, 1, \cdot)/{\sim}$ is also an associative ring. If the original ring is commutative, so is the quotient.*

For another example, there is a congruence-relation on $(\mathbb{R}, +)$ given by

$$a \sim b \Leftrightarrow a - b \in \mathbb{Z}.$$

Then there is a well-defined embedding $a \mapsto \exp(2\pi \mathrm{i}\, a)$ of $(\mathbb{R}, 0, -, +)/{\sim}$ in $(\mathbb{C}^{\times}, 1, {}^{-1}, \cdot)$.

### 3.2.3. Subgroups

We defined subgroups of symmetry groups on page 53, and of course subgroups of arbitrary groups are defined the same way. A **subgroup** of a group is just a substructure of the group, when this group is considered as having the full signature $\{\mathrm{e}, {}^{-1}, \cdot\}$. More informally, a subgroup of a group is a subset containing the identity that is closed under multiplication and inversion.

The subset $\mathbb{N}$ of $\mathbb{Q}^+$ contains the identity and is closed under multiplication, but is not closed under inversion, and so it is not a subgroup of $\mathbb{Q}^+$. The subset $\omega$ of $\mathbb{Z}$ contains the additive identity and is closed under addition, but is not closed under additive inversion, and so it is not a subgroup of $\mathbb{Z}$.

**Theorem 86.** *A subset of a group is a subgroup if and only if it is non-empty and closed under the binary operation* $(x, y) \mapsto xy^{-1}$.

If $H$ is a subgroup of $G$, we write

$$H < G.$$

One could write $H \leqslant G$ instead, if one wanted to reserve the expression $H < G$ for the case where $H$ is a *proper* subgroup of $G$. We shall not do this.[2] However, starting on page 160, we shall want an expression for this case: then we shall just have to write

$$H \lneqq G.$$

Meanwhile, we have the following examples.

**Theorem 87.** *1. For all groups $G$,*

$$\{\mathrm{e}\} < G, \qquad G < G.$$

*2. For all groups $G_0$ and $G_1$, if $H_0 < G_0$ and $H_1 < G_1$, then*

$$H_0 \times H_1 < G_0 \times G_1.$$

*3. In particular, for all groups $G$ and $H$,*

$$G \times \{\mathrm{e}\} < G \times H, \qquad \{\mathrm{e}\} \times H < G \times H.$$

[2] I do think it is useful to reserve the notation $A \subset B$ for the case where $A$ is a proper subset of $B$, writing $A \subseteq B$ when $A$ is allowed to be equal to $B$.

4. *For all groups $G$,*

$$\{(x, x) \colon x \in G\} < G \times G.$$

5. *The subset*

$$\{\mathrm{e}, (0\ 1), (2\ 3), (0\ 1)(2\ 3)\}$$

*of* Sym(4) *is a subgroup isomorphic to* $\mathrm{V}_4$.

6. *If* $\sim$ *is a congruence-relation on a group $G$, then*

$$\{x \in G \colon x \sim \mathrm{e}\} < G.$$

It is important to note that the converse of the last part of the theorem is false in general: there are groups $G$ with subgroups $H$ such that for no congruence-relation on $G$ is $H$ the congruence-class of the identity. For example, let $G$ be Sym(3), and let $H$ be the image of Sym(2) in $G$ under the obvious embedding mentioned in §2.2. Then $H$ contains just the identity and $(0\ 1)$. If $\sim$ is a congruence-relation on $G$ such that $(0\ 1) \sim \mathrm{e}$, then

$$(1\ 2)(0\ 1)(1\ 2) \sim (1\ 2)\,\mathrm{e}(1\ 2) \sim \mathrm{e};$$

but $(1\ 2)(0\ 1)(1\ 2) = (0\ 2)$, which is not in $H$. See §3.6 (p. 110) for the full story.

If $f$ is a homomorphism from $G$ to $H$, then the **kernel** of $f$ is the set

$$\{x \in G \colon f(x) = \mathrm{e}\},$$

which can be denoted by $\ker(f)$. The **image** of $f$ is

$$\{y \in H \colon y = f(x) \text{ for some } x \text{ in } G\},$$

that is, $\{f(x) \colon x \in G\}$; this can be denoted by $\mathrm{im}(f)$. For example, considering sgn as a homomorphism from $\mathrm{Sym}(n)$ to $\mathbb{Q}^{\times}$, we have

$$\ker(\mathrm{sgn}) = \mathrm{Alt}(n), \qquad \mathrm{im}(\mathrm{sgn}) = \{\pm 1\}.$$

If $g$ is $(x, y) \mapsto x$ from $G \times H$ to $G$ as in Theorem 82, and $h$ is $x \mapsto (x, \mathrm{e})$ from $G$ to $G \times H$ as in Theorem 83, then

$$\begin{aligned} \ker(g) &= \{\mathrm{e}\} \times H, & \ker(h) &= \{\mathrm{e}\}, \\ \mathrm{im}(g) &= G, & \mathrm{im}(h) &= G \times \{\mathrm{e}\}. \end{aligned}$$

An embedding (that is, an injective homomorphism) is also called a **monomorphism.** A surjective homomorphism is called an **epimorphism.** In the last example, $g$ is an epimorphism, and $h$ is a monomorphism.

**Theorem 88.** *Let $f$ be a homomorphism from $G$ to $H$.*

1. $\ker(f) < G$.
2. *$f$ is a monomorphism if and only if* $\ker(f) = \{\mathrm{e}\}$.
3. $\mathrm{im}(f) < H$.

There is a monomorphism from $\mathbb{R} \oplus \mathbb{R}$ into $\mathrm{M}_{2\times 2}(\mathbb{R})$, namely

$$(x, y) \mapsto \begin{pmatrix} x & y \\ -y & x \end{pmatrix}.$$

One can define $\mathbb{C}$ to be the image of this monomorphism. One shows that $\mathbb{C}$ then is a sub-ring of $\mathrm{M}_{2\times 2}(\mathbb{R})$ and is a field. The elements of $\mathbb{C}$ usually denoted by 1 and $\mathrm{i}$ are given by

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

Then every element of $\mathbb{C}$ is $x + y\mathrm{i}$ for some unique $x$ and $y$ in $\mathbb{R}$. The function $z \mapsto \bar{z}$ is an automorphism of $\mathbb{C}$, where

$$\overline{x + y\mathrm{i}} = x - y\mathrm{i}.$$

There is then a monomorphism from $\mathbb{C} \oplus \mathbb{C}$ into $\mathrm{M}_{2\times 2}(\mathbb{C})$, namely

$$(x, y) \mapsto \begin{pmatrix} x & y \\ -\bar{y} & \bar{x} \end{pmatrix};$$

its image is denoted by

$$\mathbb{H}$$

in honor of its discoverer Hamilton: it consists of the **quaternions.** One shows that $\mathbb{H}$ is a sub-ring of $\mathrm{M}_{2\times 2}(\mathbb{C})$ and that all non-zero elements of $\mathbb{H}$ are invertible, although $\mathbb{H}$ is not commutative. The element of $\mathbb{H}$ usually denoted by $\mathrm{j}$ is given by

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

**Theorem 89.** *An arbitrary intersection of subgroups is a subgroup.*

*Proof.* This is an instance of the general observation that an arbitrary intersection of substructures is a substructure. □

### 3.2.4. Generated subgroups

Given a subset $A$ of (the universe of) a group $G$, we can *close* under the three group-operations, obtaining a subgroup, $\langle A\rangle$. For a formal definition, we let
$$\langle A\rangle = \bigcap \mathcal{S},$$
where $\mathcal{S}$ is the set of all subgroups of $G$ that include $A$. Note that
$$\langle\varnothing\rangle = \{\mathrm{e}\}.$$
The subgroup $\langle A\rangle$ of $G$ is said to be **generated** by $A$, and the elements of $A$ are said to be, collectively, **generators** of $\langle A\rangle$. If $A = \{a_0, \dots, a_{n-1}\}$, then for $\langle A\rangle$ we may write
$$\langle a_0, \dots, a_{n-1}\rangle.$$
In this case, $\langle A\rangle$ is said to be **finitely generated.** If also $n = 1$, then $\langle A\rangle$ is said to be **cyclic.** It is easy to describe cyclic groups as sets, and almost as easy to describe finitely generated *abelian* groups:

**Theorem 90.** *Let $G$ be a group.*

1. *If $a \in G$, then*
$$\langle a\rangle = \{a^n : n \in \mathbb{Z}\}.$$
2. *If $\{a_0, \dots, a_{n-1}\} \subseteq G$, and $G$ is abelian, then*
$$\langle a_0, \dots, a_{n-1}\rangle = \{x_0a_0 + \dots + x_{n-1}a_{n-1} : (x_0, \dots, x_{n-1}) \in \mathbb{Z}^n\}.$$

*Proof.* 1. Let $f$ be the homomorphism $x \mapsto a^x$ from $\mathbb{Z}$ to $G$ as in Theorem 63 (p. 72). We have to show $\langle a\rangle = \mathrm{im}(f)$. Since $a \in \mathrm{im}(f)$, it is now enough, by Theorem 88, to show $\mathrm{im}(f) \subseteq H$ for all subgroups $H$ of $G$ that contain $a$. But for such $H$ we have $a^0 \in H$, and if $a^n \in \langle a\rangle$, then $a^{n\pm1} \in \langle a\rangle$, so by induction, $\mathrm{im}(f) \subseteq H$.

2. The indicated set is a subgroup of $G$ by Theorem 86, and it contains the $a_i$. It remains to note that the indicated set is included in every subgroup of $G$ that contains the $a_i$. $\square$

As examples of cyclic groups, we have $\mathbb{Z}$ and the $\mathbb{Z}_n$. Indeed,

$$\mathbb{Z} = \langle 1 \rangle, \qquad\qquad \mathbb{Z}_n = \langle [1] \rangle.$$

**Theorem 91.** *All subgroups of $\mathbb{Z}$ are cyclic. All nontrivial subgroups of $\mathbb{Z}$ are isomorphic to $\mathbb{Z}$.*

*Proof.* Suppose $G$ is a nontrivial subgroup of $\mathbb{Z}$. Then $G$ has positive elements, so it has a least positive element, $n$. If $a \in G$, then all residues of $a$ *modulo* $n$ belong to $G$. By Theorem 47 (page 56), $a$ has a residue in $n$ (that is, $\{0, \dots, n-1\}$), and so this residue must be $0$. Thus $n \mid a$, so $a \in \langle n \rangle$. Therefore $G = \langle n \rangle$. The function $x \mapsto nx$ from $\mathbb{Z}$ to $\langle n \rangle$ is a surjective homomorphism; that it is injective can be derived from Corollary 22.1 (page 38). $\square$

**Theorem 92.** *If $n$ is a positive integer and $m$ is an arbitrary integer, then*

$$\langle [m] \rangle = \mathbb{Z}_n \iff [m] \in {\mathbb{Z}_n}^{\times}.$$

*Proof.* Each condition means the congruence

$$mx \equiv 1 \pmod{n}$$

is soluble. $\square$

The language of generated subgroups is useful for establishing a basic theorem of number theory. In $\mathbb{Z}$, the relation of dividing is transitive:

$$a \mid b \ \&\ b \mid c \implies a \mid c.$$

This is just because $ax = b$ and $by = c$ imply $axy = c$. A **common divisor** of two integers is just a divisor of each of them. Equivalently, a common divisor of $a$ and $b$ is some $c$ such that

$$\langle a, b \rangle \subseteq \langle c \rangle.$$

Hence it makes sense to speak of a **greatest common divisor** of two integers: it is a common divisor that is divisible by each common divisor. Since 0 divides only itself, it is not a common divisor of two *different* integers. If $a \neq 0$, then $a$ is a greatest common divisor of $a$ and 0. Defining

$$|a| = \begin{cases} a, & \text{if } a \geqslant 0, \\ -a, & \text{if } a < 0, \end{cases}$$

we have

$$c \mid d \ \ \& \ \ d \neq 0 \implies |c| \leqslant |d|\,,$$
$$c \mid d \ \ \& \ \ d \mid c \iff |c| = |d|\,,$$

so if $d$ is a greatest common divisor of $a$ and $b$, then so is $-d$, but nothing else. In this case we denote $|d|$ by

$$\gcd(a, b);$$

this is greater (in the usual sense) than all other common divisors of $a$ and $b$.

**Theorem 93.** *Any two integers $a$ and $b$ have a greatest common divisor, and*

$$\langle a, b\rangle = \langle \gcd(a, b)\rangle,$$

*so that the equation*

$$ax + by = \gcd(a, b)$$

*is soluble.*

*Proof.* By Theorem 91, there is $d$ such that $\langle a, b\rangle = \langle d\rangle$. Since we have

$$c \mid d \iff \langle d\rangle \subseteq \langle c\rangle,$$

it follows that $d$ is a greatest common divisor of $a$ and $b$. Then $\gcd(a, b) = |d|$, so $\langle \gcd(a, b)\rangle = \langle d\rangle$. □

A common divisor of $a$ and $b$ is a common divisor of $|a|$ and $|b|$. The proof of Theorem 91 suggests a way to find greatest common divisors, which is the **Euclidean algorithm,** established in Propositions VII.1 and 2 of the *Elements.* Suppose $a_0$ and $a_1$ are positive integers. We define a sequence $(a_0, a_1, \dots)$ of positive integers by letting $a_{k+2}$ be the residue in $a_{k+1}$ of $a_k$ *modulo* $a_{k+1}$, if this residue is positive; otherwise $a_{k+2}$ is undefined. Then

$$a_{k+1} > a_{k+2},$$

so the sequence must have a last term; this is $\gcd(a_0, a_1)$. When this is 1, then $a_0$ and $a_1$ are said to be **prime to** one another, or **relatively prime.** In this case, by Theorem 93, the equation

$$a_0 x + a_1 y = 1$$

is soluble in $\mathbb{Z}$.

If $a \equiv b \pmod{n}$, then $\gcd(a, n) = \gcd(b, n)$. Hence the following makes sense:

**Theorem 94.** *For all positive integers $n$,*

$$\mathbb{Z}_n{}^{\times} = \{x \in \mathbb{Z}_n \colon \gcd(x, n) = 1\}.$$

*Proof.* By the proof of Theorem 92, $\mathbb{Z}_n{}^{\times}$ consists of those $m$ in $\mathbb{Z}_n$ such that the congruence

$$mx \equiv 1 \pmod{n}$$

is soluble, that is, the equation $mx + ny = 1$ is soluble, so that $\gcd(m, n)$ must be 1. Conversely, if $\gcd(m, n) = 1$, then the equation $mx + ny = 1$ is soluble by Theorem 93. □

For an arbitrary subset $A$ of an arbitrary group, it is not so easy to give a description of the elements of $\langle A \rangle$. We shall do it by means of Theorem 126 on page 130. Meanwhile, we may note some more specific examples:

The subgroup $\langle (0\ 1), (2\ 3) \rangle$ of $\mathrm{Sym}(4)$ is the subgroup given above in Theorem 87 as being isomorphic to $\mathrm{V}_4$.

The subgroup $\langle \mathrm{i}, \mathrm{j} \rangle$ of $\mathbb{H}^{\times}$ is the **quaternion group,** denoted by

$$\mathrm{Q}_8;$$

it has eight elements: $\pm 1$, $\pm \mathrm{i}$, $\pm \mathrm{j}$, and $\pm \mathrm{k}$, where $\mathrm{k} = \mathrm{i}\mathrm{j}$. We consider this group further in the next section (§3.3) and later.

**Theorem 95.** *If $n \geqslant 3$, let*

$$\sigma_n = \begin{pmatrix} 0 & 1 & \dots & n-1 \end{pmatrix},$$
$$\beta = \begin{pmatrix} 1 & n-1 \end{pmatrix} \begin{pmatrix} 2 & n-2 \end{pmatrix} \cdots \begin{pmatrix} m & n-m \end{pmatrix}$$

*in* $\mathrm{Sym}(n)$*, where $m$ is the greatest integer that is less than $n/2$. Then*

$$\mathrm{Dih}(n) = \langle \sigma_n, \beta \rangle = \langle \beta, \beta\sigma_n \rangle.$$

*Proof.* The subset $\{{\sigma_n}^i \beta^j \colon (i,j) \in n \times 2\}$ of $\mathrm{Sym}(n)$ is a subset of $\mathrm{Dih}(n)$ and has $2n$ distinct elements, so by Theorem 53 (p. 61) it must be all of $\mathrm{Dih}(n)$. Moreover $\langle \beta, \beta\sigma_n \rangle < \langle \sigma_n, \beta \rangle$, but also $\langle \sigma_n, \beta \rangle < \langle \beta, \beta\sigma_n \rangle$ since $\sigma = \beta \cdot \beta\sigma_n$. □

Our analysis of $\mathrm{Dih}(n)$ is continued in Theorem 99 below.

In case $n = 0$, the group $\langle a_0, \dots, a_{n-1} \rangle$ should logically be denoted by $\langle\ \rangle$. Probably most people write $\langle \mathrm{e} \rangle$ instead. This is not wrong, but is redundant, since every group contains an identity, and the angle brackets indicate that a group is being given. The practice of these notes will be to write $\{\mathrm{e}\}$.

## 3.3. Order

The **order** of a group is its cardinality. The order of a group $G$ is therefore denoted by

$$|G|.$$

We have examples in Theorems 50 and 53 (pp. 59–61). If $a \in G$, then the order of the cyclic subgroup $\langle a \rangle$ of $G$ is said to be the **order** of $a$ simply and is denoted by

$$|a|.$$

For example, in the quaternion group $\mathrm{Q}_8$ (p. 101 above), we have

$$\langle \mathrm{i} \rangle = \{0, \mathrm{i}, -1, -\mathrm{i}\}, \qquad\qquad |\mathrm{i}| = 4.$$

In the notation of Theorem 95 above,

$$|\sigma_n| = n, \qquad\qquad |\beta| = 2 = |\beta\sigma_n|.$$

For another example, we have the following.

**Theorem 96.** *The order of a finite permutation is the least common multiple of the orders of its disjoint cyclic factors.*

**Theorem 97.** *In a group, if $a$ is an element of finite order $n$, then*

$$\langle a\rangle = \{a^i : i \in n\},$$

*and $x \mapsto a^x$ is a well-defined isomorphism from $\mathbb{Z}_n$ to $\langle a\rangle$, so in particular*

$$a^n = \mathrm{e}\,.$$

*Proof.* Since $\langle a\rangle$ does not have $n+1$ distinct elements, for some $i$ and $j$ we have $0 \leqslant i < j \leqslant n$, but $a^i = a^j$. Therefore $\mathrm{e} = a^{j-i}$, and hence $a^k = a^\ell$ whenever $k \equiv \ell \pmod{j-i}$. Consequently $\langle a\rangle$ has at most $j-i$ elements, that is, $n \leqslant j-i$. Since also $j-i \leqslant n$, we have $n = j-i$, and in particular $a^n = a^{j-i} = \mathrm{e}$. □

For integers $a$ and $b$, the notation $a \mid b$ was defined on page 55.

**Theorem 98.** *The following conditions on positive integers $m$ and $n$ are equivalent.*

1. *$\mathbb{Z}_n$ has a subgroup of order $m$.*
2. *$\mathbb{Z}_n$ has a unique subgroup of order $m$.*
3. *$m \mid n$.*

*Under these conditions, the subgroup is $\langle n/m\rangle$.*

The orders of certain generators of a group may determine the group up to isomorphism. We work out a couple of examples in the next two theorems.

**Theorem 99.** *If $n > 2$, and $G = \langle a, b \rangle$, where*

$$|a| = n, \qquad |b| = 2, \qquad |ab| = 2,$$

*then*

$$G \cong \mathrm{Dih}(n).$$

*Proof.* Assume $n \geqslant 2$. Since $abab = \mathrm{e}$ and $b^{-1} = b$, we have

$$ba = a^{-1}b, \qquad ba^{-1} = ab.$$

Therefore $ba^k = a^{-k}b$ for all integers $k$. This shows

$$G = \{a^i b^j : (i, j) \in n \times 2\}.$$

It remains to show $|G| = 2n$. Suppose

$$a^i b^j = a^k b^\ell,$$

where $(i, j)$ and $(k, \ell)$ are in $n \times 2$. Then

$$a^{i-k} = b^{\ell - j}.$$

If $b^{\ell-j} = \mathrm{e}$, then $\ell = j$ and $i = k$. The alternative is that $b^{\ell-j} = b$. In this case,

$$n \mid 2(i - k).$$

If $n \mid i - k$, then $i = k$ and hence $j = \ell$. The only other possibility is that $n = 2m$ for some $m$, and $i - k = \pm m$, so that $a^m = b$. But then $aa^m aa^m = a^2$, while $abab = \mathrm{e}$, so $n = 2$. □

According to this theorem, if a group with certain abstract properties of $\mathrm{Dih}(n)$ exists, then that group is isomorphic to $\mathrm{Dih}(n)$. In §4.6, we shall develop a way to create a group $G$ with those properties, regardless of whether we know about $\mathrm{Dih}(n)$. Then, using Theorem 99, we shall be able to conclude that $G$ is isomorphic to $\mathrm{Dih}(n)$. This result is Theorem 134 (p. 138).

**Theorem 100.** *If* $G = \langle a, b \rangle$, *where*

$$|a| = 4, \qquad b^2 = a^2, \qquad ba = a^3 b,$$

*then, under an isomorphism taking* $a$ *to* $\mathrm{i}$ *and* $b$ *to* $\mathrm{j}$,

$$G \cong \mathrm{Q}_8.$$

*Proof.* Since $ba = a^3 b$ and $|a| = 4$, we have also

$$ba^{-1} = ba^3 = a^9 b = ab,$$

so we can write every element of $G$ as a product $a^i b^j$ for some $i$ and $j$ in $\mathbb{Z}$. By Theorem 97, since $|a| = 4$, we can require $i \in 4$. Similarly, since $b^2 = a^2$, we can require $j \in 2$. In $\mathrm{Q}_8$, the elements $\mathrm{i}$ and $\mathrm{j}$ have the given properties of $a$ and $b$. Moreover $|\mathrm{Q}_8| = 8$, so that if $(i, j)$ and $(k, \ell)$ are distinct elements of $4 \times 2$, then

$$\mathrm{i}^i \mathrm{j}^j \neq \mathrm{i}^k \mathrm{j}^\ell.$$

Therefore there is a well-defined surjective function $\mathrm{i}^i \mathrm{j}^j \mapsto a^i b^j$ from $\mathrm{Q}_8$ to $G$, and this function is a homomorphism. It remains to show $|G| = 8$. Suppose $(i, j)$ and $(k, \ell)$ are in $4 \times 2$, and

$$a^i b^j = a^k b^\ell.$$

Then $a^{i-k} = b^{\ell-k}$ and hence

$$a^m = b^n$$

for some $n$ in 2 and $m$ in 4. If $n = 0$, then $m = 0$ (since $|a| = 4$), and so $(i, j) = (k, \ell)$. But $a \neq b$ (since $ba = a^3 b$ and $|a| = 4$). Similarly $a^3 \neq b$. Finally, $a^2 \neq b$ (since $b^2 = a^2$ and $|a| = 4$). Thus $n \neq 1$, so $n = 0$. □

As with $\mathrm{Dih}(n)$, so with $\mathrm{Q}_8$, we shall be able to create the group using only the abstract properties just given, in Theorem 135 (p. 139).

## 3.4. Cosets

Suppose $H < G$. If $a \in G$, let

$$aH = \{ax \colon x \in H\},$$
$$Ha = \{xa \colon x \in H\}.$$

Each of the sets $aH$ is a **left coset** of $H$, and the set $\{xH \colon x \in G\}$ of left cosets is denoted by

$$G/H.$$

Each of the sets $Ha$ is a **right coset** of $H$, and the set $\{Hx \colon x \in G\}$ of right cosets is denoted by

$$H\backslash G.$$

Note that $H$ itself is both a left and a right coset of itself.

Sometimes, for each $a$ in $G$, we have $aH = Ha$. For example, this is the case when $G = G_0 \times G_1$, and $H = G_0 \times \{\mathrm{e}\}$, so that, if $a = (g_0, g_1)$, then

$$aH = H \times \{g_1\} = Ha.$$

Sometimes left and right cosets are different, as in the example on page 96, where $G = \mathrm{Sym}(3)$, and $H$ is the image of $\mathrm{Sym}(2)$ in $G$. In this case

$$\begin{aligned} (0\ 2)H &= \{(0\ 2), (0\ 1\ 2)\}, & H(0\ 2) &= \{(0\ 2), (0\ 2\ 1)\}, \\ (1\ 2)H &= \{(1\ 2), (0\ 2\ 1)\}, & H(1\ 2) &= \{(1\ 2), (0\ 1\ 2)\}. \end{aligned}$$

Moreover, there are no other cosets of $H$, besides $H$ itself, by the next theorem; so in the example, no left coset, besides $H$, is a right coset.

**Theorem 101.** *Suppose $H < G$. The left cosets of $H$ in $G$ compose a partition of $G$. Likewise for the right cosets. All cosets of $H$ have the same size; also, $G/H$ and $H\backslash G$ have the same size.*

*Proof.* We have $a \in aH$. Suppose $aH \cap bH \neq \varnothing$. Then $ah = bh_1$ for some $h$ and $h_1$ in $H$, so that $a = bh_1h^{-1}$, which is in $bH$. Thus $a \in bH$, and hence $aH \subseteq bH$. By symmetry of the argument, we have also $bH \subseteq aH$, and therefore $aH = bH$. Hence the left cosets compose a partition of $G$. By symmetry again, the same is true for the right cosets.

All cosets of $H$ have the same size as $H$, since the map $x \mapsto ax$ from $H$ to $aH$ is a bijection with inverse $x \mapsto a^{-1}H$, and likewise $x \mapsto xa$ from $H$ to $Ha$ is a bijection. (One might see this as an application of Cayley's Theorem, Theorem 49, page 57.)

Inversion is a permutation of $G$ taking $aH$ to $Ha^{-1}$, so $G/H$ and $H\backslash G$ must have the same size. □

**Corollary 101.1.** *If $H < G$, then the relation $\sim$ on $G$ defined by*

$$a \sim x \Leftrightarrow aH = xH$$

*is an equivalence-relation, and*

$$G/H = G/{\sim}.$$

**Corollary 101.2.** *If $H < G$ and $aH = Hb$, then $aH = Ha$.*

*Proof.* Under the assumption, $a \in Hb$, so $Ha \subseteq Hb$, and therefore $Ha = Hb$. □

The cardinality of $G/H$ (or of $H\backslash G$) is called the **index** of $H$ in $G$ and can be denoted by

$$[G : H].$$

If $G$ is finite, then by the last theorem,

$$[G : H] = \frac{|G|}{|H|}.$$

However, $[G : H]$ may be finite, even though $G$ is not. In this case, $H$ must also be infinite, and indeed the last equation may be understood to say this, since an infinite cardinal divided by a finite cardinal should still be infinite.

Of the next theorem, we shall be particularly interested in a special case, Lagrange's Theorem, in the next section.

**Theorem 102.** *If $K < H < G$, then $[G : K] = [G : H][H : K]$.*

*Proof.* Every left coset of $K$ is included in a left coset of $H$. Indeed, if $bK \cap aH \neq \varnothing$, then as in the proof of Theorem 101, $bK \subseteq aH$. Moreover, every left coset of $H$ includes the same number of left cosets of $K$. For, the bijection $x \mapsto ax$ that takes $H$ to $aH$ also takes each coset $bK$ of $K$ to a coset $abK$ of $K$. □

## 3.5. Lagrange's Theorem

According to [2, p. 141–2], the following "is implied but not explicitly proved" in a memoir by Lagrange published in 1770–1.

**Theorem 103** (Lagrange)**.** *If $H < G$ and $G$ is finite, then $|H|$ divides $|G|$.*

*Proof.* Use Theorem 102 when $K = \{\mathrm{e}\}$. □

**Corollary 103.1.** *If $G$ is finite and $a \in G$, then $a^{|G|} = \mathrm{e}$.*

*Proof.* $a^{|a|} = \mathrm{e}$ by Theorem 97 (p. 103), and $|a|$ divides $|G|$. □

Cauchy's Theorem (page 147) and its generalization, the first Sylow Theorem (page 154), are partial converses of Lagrange's Theorem.

Meanwhile, some basic results of number theory can be seen as applications of Lagrange's Theorem. First we obtain a classification of certain finite groups. An integer greater than 1 is called **prime** if its only divisors are itself and 1.

**Theorem 104.** *All groups of prime order are cyclic.*

*Proof.* Say $|G| = p$. There is $a$ in $G \smallsetminus \{\mathrm{e}\}$, so $|a| > 1$; but $|a|$ divides $p$, so $|a| = p$, and therefore $G = \langle a \rangle$. □

The following can be obtained as a corollary of Theorem 94 (page 101); but we can obtain it also from Lagrange's Theorem.[3]

[3] This is observed by Timothy Gowers, editor of [12], in a Google+ article of December 21, 2013.

**Theorem 105.** *An integer $p$ that is greater than $1$ is prime if and only if*

$$\mathbb{Z}_p{}^\times = \{1, \dots, p-1\}.$$

*Proof.* Say $1 < a < p$ and $a \in \mathbb{Z}_p{}^\times$, so that $ac \equiv 1 \pmod{p}$ for some $c$. If $ab = p$, then $ab \equiv 0$, so $abc \equiv 0$, hence $b \equiv 0$, which is absurd. Thus $a \nmid p$. Hence, if $\mathbb{Z}_p{}^\times = \{1, \dots, p-1\}$, then $p$ must be prime.

Now suppose $p$ is prime and $1 < a < p$, so that $a \nmid p$. But $\gcd(a, p) \mid p$ and $1 \leqslant \gcd(a, p) \leqslant a$, so $\gcd(a, p) = 1$, and therefore $a \in \mathbb{Z}_p{}^\times$ by Theorem 94.

Alternatively, $\langle a \rangle$ has order greater than 1, so by Lagrange's Theorem this order must be $p$. In particular $ab \equiv 1 \pmod{p}$ for some $b$, so $a \in \mathbb{Z}_p{}^\times$. □

**Theorem 106** (Fermat)**.** *If the prime $p$ is not a factor of $a$, then*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{3.4}$$

*Hence for all integers $a$,*

$$a^p \equiv a \pmod{p}. \tag{3.5}$$

*Proof.* By the previous theorem, if $p \nmid a$, then $[a] \in \mathbb{Z}_p{}^\times$, and this group has order $p-1$, so (3.4) holds by Lagrange's Theorem. Also (3.4) implies (3.5), and the latter holds trivially if $p \mid a$. □

If $n \in \mathbb{N}$, then by Theorem 94, the order of $\mathbb{Z}_n{}^\times$ is the number of elements of $\mathbb{Z}_n$ that are prime to $n$. Let this number be denoted by

$$\phi(n).$$

This then is the number of generators of $\mathbb{Z}_n$, that is, the number of elements $k$ of $\mathbb{Z}_n$ such that $\langle k \rangle = \langle 1 \rangle$. This feature of $\phi(n)$ will be used in Theorem 141 (page 143).

**Theorem 107** (Euler)**.** *If $\gcd(a, n) = 1$, then*

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

*Proof.* If $\gcd(a, n) = 1$, then $[a] \in \mathbb{Z}_n{}^\times$ by Theorem 94. □

## 3.6. Normal subgroups

If $H < G$, we investigate the possibility of defining a multiplication on $G/H$ so that

$$(xH)(yH) = xyH. \tag{3.6}$$

In any case, each member of this equation is a well-defined subset of $G$. The question is when they are the same. Continuing with the example from pages 96 and 106, where $G = \mathrm{Sym}(3)$ and $H = \langle(0\ 1)\rangle$, we have

$$\begin{gathered}(1\ 2)H(1\ 2)H = \{\mathrm{e}, (0\ 1), (0\ 2), (0\ 1\ 2)\},\\ (1\ 2)(1\ 2)H = H = \{\mathrm{e}, (0\ 1))\},\end{gathered}$$

so (3.6) fails in this case.

**Theorem 108.** *Suppose $H < G$. The following are equivalent:*

1. *$G/H$ is a group whose multiplication is given by* (3.6)*.*
2. *Every left coset of $H$ is a right coset.*
3. *$aH = Ha$ for all $a$ in $G$.*
4. *$a^{-1}Ha = H$ for all $a$ in $G$.*

*Proof.* Immediately the last two conditions are equivalent, and they imply the second. The second implies the third, by Corollary 101.2 (p. 107).

Suppose now the first condition holds. For all $h$ in $H$, since $hH = H$, we have

$$aH = \mathrm{e}\,aH = \mathrm{e}\,HaH = hHaH = haH,$$

hence $a^{-1}haH = H$, so $a^{-1}ha \in H$. Thus $a^{-1}Ha \subseteq H$, so $a^{-1}Ha = H$.

Conversely, if the third condition holds, then $(xH)(yH) = xHHy = xHy = xyH$. In this case, the equivalence-relation $\sim$ on $G$ given as in Corollary 101.1 (p. 107) by

$$a \sim x \Leftrightarrow aH = xH$$

is a congruence-relation, and so, by Theorem 84 (p. 94), $G/H$ is a group with respect to the proposed multiplication. □

A subgroup $H$ of $G$ meeting any of these equivalent conditions is called **normal,** and in this case we write

$$H \lhd G.$$

As trivial examples, we have

$$G \lhd G, \qquad\qquad \{\mathrm{e}\} \lhd G.$$

Only slightly less trivially, all subgroups of abelian groups are normal subgroups. More examples arise from the following.

**Theorem 109.** *If $[G : H] = 2$, then $H \lhd G$.*

If $n > 1$, since $[\mathrm{Sym}(n) : \mathrm{Alt}(n)] = 2$, we now have

$$\mathrm{Alt}(n) \lhd \mathrm{Sym}(n).$$

Of course we have this trivially if $n \leqslant 1$.

In general, if $N \lhd G$, then the group $G/N$ is called the **quotient-group** of $G$ by $N$. In this case, we can write the group also as

$$\frac{G}{N}.$$

**Theorem 110.** *If $N \lhd G$ and $H < G$, then $N \cap H \lhd H$. (That is, normality is preserved in subgroups.)*

*Proof.* The defining property of normal subgroups is universal. That is, $N \lhd G$ means that the sentence

$$\forall x\ \forall y\ (x \in N \to yxy^{-1} \in N)$$

is true in the structure $(G, N)$. Therefore the same sentence is true in every substructure of $(G, N)$. If $H < G$, then $(G, N \cap H)$ is a substructure of $(G, N)$. □

For example, if $m < n$, and we identify $\mathrm{Sym}(m)$ with its image in $\mathrm{Sym}(n)$ under $\sigma \mapsto \sigma \cup \mathrm{id}_{n \smallsetminus m}$, then $\mathrm{Sym}(m) \cap \mathrm{Alt}(n) \lhd \mathrm{Sym}(m)$. But then, we already know this, since $\mathrm{Sym}(m) \cap \mathrm{Alt}(n) = \mathrm{Alt}(m)$.

In proving Theorem 95 (p. 102), we showed that every element of $\mathrm{Dih}(n)$ is a product $gh$, where $g \in \langle\sigma_n\rangle$ and $h \in \langle\beta\rangle$. Note that that, since $|\sigma_n| = n$ and $|\mathrm{Dih}(n)| = 2n$, by Theorem 109 we have $\langle\sigma_n\rangle \lhd \mathrm{Dih}(n)$. Thus our result is a special case of the following.

**Lemma 9.** *If $N \lhd G$ and $H < G$, then $\langle N \cup H\rangle = NH$.*

*Proof.* Since

$$N \cup H \subseteq NH \subseteq \langle N \cup H\rangle,$$

it is enough to show $NH < G$. Suppose $n \in N$ and $h \in H$. Then $nh = hh^{-1}nh$. Since $N \lhd \langle N \cup H\rangle$, we have $h^{-1}nh \in N$, so $nh \in HN$. Thus $NH \subseteq HN$, so by symmetry $NH = HN$. Therefore

$$NH(NH)^{-1} = NHH^{-1}N^{-1} = NHHN \subseteq NHN = NNH \subseteq NH,$$

that is, $NH$ is closed under $(x, y) \mapsto xy^{-1}$. Since $NH$ also contains e, it is a subgroup of $G$ by Theorem 86. □

**Theorem 111.** *Suppose $N \lhd G$ and $H < G$ and $N \cap H = \{\mathrm{e}\}$. Then the surjection $(x, y) \mapsto xy$ from $N \times H$ to $NH$ is a bijection, and so the structure of a group is induced on $N \times H$.*

*Proof.* If $g$ and $h$ are in $H$, and $m$ and $n$ are in $N$, and $gm = hn$, then

$$h^{-1}g = nm^{-1},$$

so each side must be e, and hence $g = h$ and $m = n$. □

Multiplication in $NH$ is given by

$$(mg)(nh) = (m \cdot gng^{-1})(gh), \tag{3.7}$$

while multiplication in the direct product $(N, \cdot) \times (H, \cdot)$ is given by

$$(m, g)(n, h) = (m \cdot n, gh).$$

Thus the direct-product structure on $N \times H$ is not necessarily the structure on $N \times H$ given by the theorem. The latter structure is called a **semidirect product** of $N$ and $H$. The group $NH$ is the **internal**

**semidirect product** of $N$ and $H$. Theorem 124 on page 125 below establishes conditions under which this *is* a direct product. Semidirect products are treated abstractly in §5.1 (p. 144). Meanwhile, again in the notation of Theorem 95, we have that $\mathrm{Dih}(n)$ is the internal semidirect product of $\langle\sigma_n\rangle$ and $\langle\beta\rangle$.

**Theorem 112.** *The normal subgroups of a group are precisely the kernels of homomorphisms on the group.*

*Proof.* If $f$ is a homomorphism from $G$ to $H$, then for all $n$ in $\ker(f)$,

$$f(ana^{-1}) = f(a)f(n)f(a)^{-1} = \mathrm{e},$$

so $a(\ker(f))a^{-1} \subseteq \ker(f)$; thus $\ker(f) \lhd G$. Conversely, if $N \lhd G$, then the map $x \mapsto xN$ from $G$ to $G/N$ is a homomorphism with kernel $N$. □

For example, from the homomorphism from $\mathrm{Sym}(4)$ onto $\mathrm{Sym}(3)$ given in Theorem 54 above (p. 62), $\mathrm{Sym}(4)$ has a normal subgroup that contains $(0\ 1)(2\ 3)$, $(0\ 2)(1\ 3)$, and $(0\ 3)(1\ 2)$, along with e. These four elements constitute the subgroup $\langle(0\ 1)(2\ 3),(0\ 2)(1\ 3)\rangle$ of $\mathrm{Sym}(4)$, and this subgroup is isomorphic to $\mathrm{V}_4$. By Theorem 114 on page 115 below, this subgroup is precisely the kernel of the homomorphism in question.

In the proof of the last theorem, the map $x \mapsto xN$ is the **canonical projection** or the **quotient map** of $G$ onto $G/N$; it may be denoted by

$$\pi.$$

**Theorem 113.** *If $f$ is a homomorphism from $G$ to $H$, and $N$ is a normal subgroup of $G$ such that $N < \ker(f)$, then there is a unique homomorphism $\tilde{f}$ from $G/N$ to $H$ such that*

$$f = \tilde{f} \circ \pi,$$

*that is, the following diagram commutes (see page 93).*

$$\begin{tikzcd} G \arrow[r, "\pi"] \arrow[d, "f"'] & G/N \arrow[dl, "\tilde{f}"] \\ H & \end{tikzcd}$$

*Proof.* If $\tilde{f}$ exists, it must be given by

$$\tilde{f}(xN) = f(x).$$

Such $\tilde{f}$ does exist, since if $xN = yN$, then $xy^{-1} \in N$, so $xy^{-1} \in \ker(f)$, hence $f(xy^{-1}) = \mathrm{e}$, and therefore $f(x) = f(y)$. □

**Corollary 113.1** (First Isomorphism Theorem)**.** *Suppose $f$ is a homomorphism from a group $G$ to some other group. Then*

$$G/\ker(f) \cong \operatorname{im}(f).$$

*In particular, if* $\operatorname{im}(f)$ *is finite, then*

$$[G : \ker(f)] = |\operatorname{im}(f)|.$$

*Proof.* Let $N = \ker(f)$; then $\tilde{f}$ is the desired homomorphism. □

For example, letting $f$ be $x \mapsto x + n\mathbb{Z}$ from $\mathbb{Z}$ to $\mathbb{Z}_n$, we have

$$\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n.$$

Another example is Theorem 114 below.

**Corollary 113.2** (Second Isomorphism Theorem)**.** *If $H < G$ and $N \lhd G$, then*

$$\frac{H}{H \cap N} \cong \frac{HN}{N}.$$

*Proof.* The map $h \mapsto hN$ from $H$ to $HN/N$ is surjective with kernel $H \cap N$. So the claim follows by the First Isomorphism Theorem (that is, Corollary 113.1). □

For example, In $\mathbb{Z}$, since $\langle n \rangle \cap \langle m \rangle = \langle \operatorname{lcm}(n, m) \rangle$ and $\langle n \rangle + \langle m \rangle = \langle \gcd(n, m) \rangle$, we have

$$\frac{\langle n \rangle}{\langle \operatorname{lcm}(n, m) \rangle} \cong \frac{\langle \gcd(n, m) \rangle}{\langle m \rangle}.$$

**Corollary 113.3** (Third Isomorphism Theorem). *If $N$ and $K$ are normal subgroups of $G$ and $N < K$, then*

$$K/N \lhd G/N, \qquad \frac{G/N}{K/N} \cong G/K.$$

*Proof.* For the first claim, we have

$$aN\left(\frac{K}{N}\right)(aN)^{-1} = \frac{aKa^{-1}}{N} = \frac{K}{N}$$

since $(aN)(xN)(aN)^{-1} = axa^{-1}N$. By the First Isomorphism Theorem (Corollary 113.1) in case $f$ is $x \mapsto xK$ from $G$ to $G/K$, we have a homomorphism $xN \mapsto xK$ from $G/N$ to $G/K$. The kernel is $\{xN : x \in K\}$, which is just $G/N$. The second claim follows by the First Isomorphism Theorem. □

Another basis result about normal subgroups will be Theorem 163 on page 159. Theorem 113 will be used to prove von Dyck's Theorem (Theorem 133, p. 138). As promised, another application of the First Isomorphism Theorem is the following.

**Theorem 114.** $\langle(0\ 1)(2\ 3), (0\ 2)(1\ 3)\rangle \lhd \mathrm{Alt}(4)$.

*Proof.* Let $f$ be the homomorphism from $\mathrm{Sym}(4)$ to $\mathrm{Sym}(3)$ given in Theorem 54. Then $|\ker(f)| = 4$. We have already noted (p. 113) that

$$\langle(0\ 1)(2\ 3), (0\ 2)(1\ 3)\rangle < \ker(f).$$

Since $\langle(0\ 1)(2\ 3), (0\ 2)(1\ 3)\rangle \cong \mathrm{V}_4$, the Klein four group, it must be equal to $\ker(f)$. Hence $\langle(0\ 1)(2\ 3), (0\ 2)(1\ 3)\rangle \lhd \mathrm{Sym}(4)$. Moreover, this normal subgroup is a subgroup of $\mathrm{Alt}(4)$, and therefore, by Theorem 110, it is a normal subgroup of $\mathrm{Alt}(4)$. □

# 3.7. Classification of finite simple groups

## 3.7.1. Classification

One of the goals of mathematical research is **classification** [12, p. 52]. To classify is to divide into classes. Originally, the word *class* refers to a class of persons in a society. In mathematics, the word is used for collections defined by formulas, as described on page 18 above. To classify a class $\boldsymbol{C}$ of structures is to partition it into subclasses. Such a partitioning corresponds to an equivalence-relation on $\boldsymbol{C}$: the subclasses of $\boldsymbol{C}$ are then the corresponding equivalence-classes.

For example, $\boldsymbol{C}$ might be the class of all structures. We have classified structures according to whether they are algebras or not (p. 42). There is a finer classification, according to the precise signatures of structures. Within the class of structures having the signature $\{\mathrm{e}, {}^{-1}, \cdot\}$ of groups, we have distinguished the subclass consisting of those structures that actually *are* groups.

For the class of groups, or indeed for any class of structures, the finest classification that is of interest to us is the classification determined by the relation of isomorphism. In an abstract sense, merely to specify the relation of isomorphism is to determine a classification of the class in question. But we want to do more. For example, we should like to be able to choose a representative from each isomorphism-class.

We have already done this for sets as such. We have classified sets according to the relation of equipollence, and then we have shown that, within every equipollence-class, there is a unique cardinal (page 51).

For the classification of groups, Cayley's Theorem (page 57) is of use. If $G$ is a group, and $|G| = \kappa$, then $G$ embeds in $\mathrm{Sym}(\kappa)$. Thus the isomorphism-class of $G$ contains a subgroup of $\mathrm{Sym}(\kappa)$. However, it will usually contain more than one subgroup of $\mathrm{Sym}(\kappa)$.

The natural numbers are classified according to whether they are prime. Moreover, every natural number is the product of a unique set of prime powers. We state this formally.

**Theorem 115.** *For every $n$ in $\mathbb{N}$, there is a unique finite set $S$ of prime numbers and a unique function $f$ from $S$ into $\mathbb{N}$ such that*

$$n = \prod_{p \in S} p^{f(p)}.$$

In §4.7 (page 139 below) we are going to be able to give a similar classification of the finitely generated abelian groups, building on the initial distinguishing of certain groups as being cyclic.

### 3.7.2. Finite simple groups

A group is **simple** if it is nontrivial and has no proper nontrivial normal subgroups.[4] In §5.7 (p. 165) below, culminating in the Jordan–Hölder Theorem, we shall see that every finite group can be analyzed as a kind of 'product' of a list of simple groups. In this case, the analysis is not reversible; different finite groups can yield the same list of simple groups. A grand project of group theory has been to classify the finite simple groups. We establish *part* of this classification now. The *abelian* finite simple groups are easy to find:

**Theorem 116.** *The simple abelian groups are precisely the groups isomorphic to $\mathbb{Z}_p$ for some prime number $p$.*

As for nonabelian groups, we already know by Theorem 114 that $\mathrm{Alt}(4)$ is not simple. However, $\mathrm{Alt}(3)$ is simple, being isomorphic to $\mathbb{Z}_3$. Being trivial, $\mathrm{Alt}(2)$ is not simple. We are going to show that $\mathrm{Alt}(n)$ is simple when $n \geqslant 5$.

**Theorem 117.** $\mathrm{Alt}(n)$ *is generated by the* 3*-cycles in* $\mathrm{Sym}(n)$.

[4] In defining simple groups, Hungerford [19, p. 49] omits the condition that they must be nontrivial; but then he immediately states our Theorem 116, which excludes the trivial $\mathbb{Z}_1$ from being simple, because 1 is not prime. Lang [23] gives the nontriviality condition.

*Proof.* The group $\mathrm{Alt}(n)$ is generated by the products $\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}a & c\end{pmatrix}$ and $\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}c & d\end{pmatrix}$, where $a$, $b$, $c$, and $d$ are distinct elements of $n$. But

$$\begin{aligned}\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}a & c\end{pmatrix} &= \begin{pmatrix}a & c & b\end{pmatrix},\\ \begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}c & d\end{pmatrix} &= \begin{pmatrix}b & c & a\end{pmatrix}\begin{pmatrix}c & d & b\end{pmatrix}.\end{aligned}$$

Hence all 3-cycles belong to $\mathrm{Alt}(n)$, and this group is generated by these cycles. □

If $a$ and $b$ belong to an arbitrary group $G$, then the element $aba^{-1}$ of $G$ is called the **conjugate of $b$ by $a$**, and the operation $x \mapsto axa^{-1}$ on $G$ is called **conjugation by $a$**. Conjugation by an element of $G$ is an automorphism of $G$: this is stated formally as Theorem 142 on page 144 below. For now, all we need to know is that, if $N \lhd G$, then conjugates of elements of $N$ by elements of $G$ are elements of $N$.

**Theorem 118.** *Every normal subgroup of* $\mathrm{Alt}(n)$ *containing a* 3*-cycle* is $\mathrm{Alt}(n)$.

*Proof.* By Theorem 117, it is enough to show that for any 3-cycle, every 3-cycle is a conjugate of it. We have

$$\begin{pmatrix}a & b & d\end{pmatrix} = \underbrace{\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}c & d\end{pmatrix}}\begin{pmatrix}c & b & a\end{pmatrix}\underbrace{\begin{pmatrix}c & d\end{pmatrix}\begin{pmatrix}a & b\end{pmatrix}}.$$

Thus, by conjugation, we can change any entry in a 3-cycle's nontrivial orbit. □

**Theorem 119.** $\mathrm{Alt}(n)$ *is simple if* $n > 4$.

*Proof.* Suppose $\mathrm{Alt}(n)$ has normal subgroup $N$ with a nontrivial element $\sigma$. Then $\sigma$ is the product of disjoint cycles, among which are:

1) a cycle of order at least 4; or
2) two cycles of order 3; or
3) transpositions, only one 3-cycle, and no other cycles; or
4) only transpositions.

We show that, in each case, $N$ contains a 3-cycle.

1. Suppose first that $\sigma$ is $\begin{pmatrix}0 & 1 & \dots & k-1\end{pmatrix}\tau$ for some $\tau$ that is disjoint from $\begin{pmatrix}0 & 1 & \dots & k-1\end{pmatrix}$. Then $N$ contains both

$$\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}0 & 1 & \dots & k-1\end{pmatrix}\tau\begin{pmatrix}2 & 1 & 0\end{pmatrix}$$

and $\tau^{-1}\begin{pmatrix}k-1 & \dots & 1 & 0\end{pmatrix}$, and their product is a 3-cycle:

$$\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}0 & 1 & \dots & k-1\end{pmatrix}\tau\begin{pmatrix}2 & 1 & 0\end{pmatrix}\tau^{-1}\begin{pmatrix}k-1 & \dots & 1 & 0\end{pmatrix} = \begin{pmatrix}0 & 1 & 3\end{pmatrix}.$$

2. If $\tau$ is disjoint from $\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}3 & 4 & 5\end{pmatrix}$, then we reduce to the previous case:

$$\begin{pmatrix}0 & 1 & 3\end{pmatrix}\underbrace{\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}3 & 4 & 5\end{pmatrix}}\tau\begin{pmatrix}3 & 1 & 0\end{pmatrix}\tau^{-1}\underbrace{\begin{pmatrix}5 & 4 & 3\end{pmatrix}\begin{pmatrix}2 & 1 & 0\end{pmatrix}} = \begin{pmatrix}0 & 1 & 4 & 2 & 3\end{pmatrix}.$$

3. If $\tau$ is disjoint from $\begin{pmatrix}0 & 1 & 2\end{pmatrix}$ and is the product of transpositions, then

$$\left[\begin{pmatrix}0 & 1 & 2\end{pmatrix}\tau\right]^2 = \begin{pmatrix}2 & 1 & 0\end{pmatrix}.$$

4. Finally, suppose $\tau$ is a product of transpositions disjoint from $\begin{pmatrix}0 & 1\end{pmatrix}$ and $\begin{pmatrix}2 & 3\end{pmatrix}$. Then

$$\begin{pmatrix}0 & 1 & 2\end{pmatrix}\underbrace{\begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}2 & 3\end{pmatrix}\tau}\begin{pmatrix}2 & 1 & 0\end{pmatrix}\underbrace{\tau\begin{pmatrix}3 & 2\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix}} = \begin{pmatrix}0 & 2\end{pmatrix}\begin{pmatrix}1 & 3\end{pmatrix}.$$

Furthermore, since $n > 4$, in $\operatorname{Alt}(n)$ we compute

$$\begin{pmatrix}0 & 2 & 4\end{pmatrix}\underbrace{\begin{pmatrix}0 & 2\end{pmatrix}\begin{pmatrix}1 & 3\end{pmatrix}}\begin{pmatrix}4 & 2 & 0\end{pmatrix}\underbrace{\begin{pmatrix}3 & 1\end{pmatrix}\begin{pmatrix}2 & 0\end{pmatrix}} = \begin{pmatrix}0 & 4 & 2\end{pmatrix}. \qquad \square$$

For the sake of classifying small finite groups in general (in §5.4, page 156), we shall want the following, which assumes $\operatorname{Alt}(n)$ is defined just when $n \geqslant 2$ (see page 76 above).

**Theorem 120.** $\operatorname{Alt}(n)$ *is the unique subgroup of* $\operatorname{Sym}(n)$ *of index* $2$.

# 4. Category theory

## 4.1. Products

There is a simple property of direct products of groups (as defined on page 91) that will turn out to characterize these products. If $G_0$ and $G_1$ are groups, then we know from Theorem 82 on page 92 that for each $i$ in 2, the function

$$(x_0, x_1) \mapsto x_i$$

from $G_0 \times G_1$ to $G_i$ is a homomorphism. It can be called a **coordinate projection** and denoted by

$$\pi_i.$$

**Theorem 121.** *Let $G_0$, $G_1$ and $H$ be groups such that, for each $i$ in 2, there is a homomorphism $f_i$ from $H$ to $G_i$. Then the function*

$$x \mapsto (f_0(x), f_1(x))$$

*from $H$ to $G_0 \times G_1$ is a homomorphism, and it is the unique homomorphism $f$ from $H$ to $G_0 \times G_1$ such that, for each $i$ in 2,*

$$\pi_i f = f_i,$$

*that is, the following diagram commutes:*

$$\begin{tikzcd}
G_0 & G_0 \times G_1 \arrow[l, "\pi_0"'] \arrow[r, "\pi_1"] & G_1 \\
 & H \arrow[ul, "f_0"] \arrow[u, "f"'] \arrow[ur, "f_1"'] &
\end{tikzcd}$$

*If the groups $G_i$ are abelian, then so is $G_0 \times G_1$.*

*Proof.* If $u \in G_0 \times G_1$, then

$$u = (\pi_0(u), \pi_1(u)).$$

Hence, if $f \colon H \to G_0 \times G_1$, then $f(x) = (\pi_0 f(x), \pi_1 f(x))$. In particular then, $f$ is as desired if and only if $f(x) = (f_0(x), f_1(x))$. □

Considering this theorem and its proof, we may see that a more general result can be obtained. This is the porism below. We obtain it by considering an **indexed family** $(G_i \colon i \in I)$ of groups. This is an indexed set in the sense of page 74; we use the word *family* to emphasize that the structure of each $G_i$ will be important. The **direct product** of the indexed family can be denoted by one of

$$\prod_{i \in I} G_i, \qquad \prod (G_i \colon i \in I).$$

This is, first of all, the set whose elements are indexed sets $(x_i \colon i \in I)$ such that $x_i \in G_i$ for each $i$ in $I$. Note a special case: If all of the groups $G_i$ are the same group $G$, then

$$\prod_{i \in I} G = G^I.$$

In case $I = n$, we may write $\prod_{i \in I} G_i$ also as

$$G_0 \times \cdots \times G_{n-1},$$

and a typical element of this as $(x_0, \ldots, x_{n-1})$.

**Theorem 122.** *The direct product $(G_i \colon i \in I)$ of an indexed family of groups is a group under the multiplication given by*

$$(x_i \colon i \in I) \cdot (y_i \colon i \in I) = (x_i \cdot y_i \colon i \in I).$$

*Each of the functions*

$$(x_j \colon j \in I) \mapsto x_i$$

*is a homomorphism from $\prod_{j \in I} G_j$ to $G_i$.*

*Proof.* As for Theorem 75 on page 81 and Theorem 82 on page 92. □

As before, the homomorphisms in the porism are the **coordinate projections,** denoted by

$$\pi_i.$$

**Porism 121.1.** *Suppose $(G_i \colon i \in I)$ is an indexed family of groups, and $H$ is a group, and for each $i$ in $I$ there is a homomorphism from $H$ to $G_i$. Then there is a homomorphism*

$$x \mapsto (f_i(x) \colon i \in I) \tag{4.1}$$

*from $H$ to $\prod_{i\in I} G_i$, and this is the unique homomorphism $f$ from $H$ to $\prod_{i\in I} G_i$ such that, for each $i$ in $I$,*

$$\pi_i f = f_i,$$

*that is, the following diagram commutes:*

$$\begin{tikzcd}
\prod_{j\in I} G_j \arrow[r, "\pi_i"] & G_i \\
H \arrow[u, "f"] \arrow[ur, "f_i"'] &
\end{tikzcd}$$

*If the groups $G_i$ are abelian, then so is $\prod_{i\in I} G_i$.*

If we ignore the actual definition (4.1) of the unique homomorphism $f$, then the porism can be summarized as being that the direct product of an indexed family of groups has a certain **universal property.** Theorem 128 on page 134 below is that the direct product is *characterized* by its universal property. Other constructions characterized by universal properties are:

- the direct sum (next section, namely §4.2);
- the free abelian group and the free group (§4.4);
- the quotient field of an integral domain (§7.5, page 199);
- the polynomial ring (sub-§7.7.1, page 211).

## 4.2. Sums

We now investigate the possibility of reversing the arrows in Theorem 121. If $G_0$ and $G_1$ are arbitrary groups, then we know from Theorem 83 on page 92 that the functions

$$x \mapsto (x, \mathrm{e}), \qquad\qquad x \mapsto (\mathrm{e}, x)$$

are homomorphisms, from $G_0$ and $G_1$ respectively to $G_0 \times G_1$. They can be called the **canonical injections,** denoted respectively by

$$\iota_0, \qquad\qquad \iota_1.$$

**Theorem 123.** *Let $G_0$, $G_1$ and $H$ be abelian groups such that, for each $i$ in $2$, there is a homomorphism $f_i$ from $G_i$ to $H$. Then the function*

$$(x_0, x_1) \mapsto f_0(x_0) + f_1(x_1)$$

*from $G_0 \oplus G_1$ to $H$ is a homomorphism, and it is the unique homomorphism $f$ from $G_0 \oplus G_1$ to $H$ such that, for each $i$ in $2$,*

$$f\iota_i = f_i,$$

*that is, the following diagram commutes:*

$$\begin{tikzcd}
G_0 \arrow[r, "\iota_0"] \arrow[dr, "f_0"'] & G_0 \oplus G_1 \arrow[d, "f"] & G_1 \arrow[l, "\iota_1"'] \arrow[dl, "f_1"] \\
& H &
\end{tikzcd}$$

*Proof.* If $(x_0, x_1) \in G_0 \oplus G_1$, then

$$(x_0, x_1) = \iota_0(x_0) + \iota_1(x_1),$$

so that, if $f$ is a homomorphism on $G_0 \oplus G_1$, then

$$f(x_0, x_1) = f\iota_0(x_0) + f\iota_1(x_1).$$

Hence, if $f$ is as desired, then it must be given by

$$f(x_0, x_1) = f_0(x_0) + f_1(x_1). \tag{4.2}$$

The function so defined is indeed a homomorphism, since

$$\begin{aligned} f((x_0,x_1)+(u_0,u_1)) &= f(x_0+u_0,x_1+u_1) \\ &= f_0(x_0+u_0)+f_1(x_1+u_1) \\ &= f_0(x_0)+f_0(u_0)+f_1(x_1)+f_1(u_1) \\ &= f_0(x_0)+f_1(x_1)+f_0(u_0)+f_1(u_1) \qquad (4{\cdot}3) \\ &= f(x_0,x_1)+f(u_0,u_1), \end{aligned}$$

where (4.3) uses that $H$ is abelian. Moreover, when $f$ is as in (4.2), then

$$f\iota_0(x) = f(x,0) = f_0(x),$$

so $f\iota_0 = f_0$, and similarly $f\iota_1 = f_1$. □

In the proof, the definition of $f$ in (4.2) does not require that the indexed family $(G_i : i \in 2)$ have just two members, but that it have finitely many. Also, as noted, $f$ is a homomorphism because $H$ is abelian; but this condition too can be weakened. Given an arbitrary indexed family $(G_i : i \in I)$ of groups, we have, for each $i$ in $I$, a function $\iota_i$ from $G_i$ to $\sum_{j\in I} G_j$ given by

$$\iota_i(x) = (x_j : j \in I),$$

where

$$x_j = \begin{cases} x, & \text{if } j = i, \\ \mathrm{e}, & \text{otherwise.} \end{cases}$$

The monomorphisms $\iota_i$ are the **canonical injections.**

**Porism 123.1.** *Suppose $(G_i : i < n)$ is a finite indexed family of groups, and $H$ is a group, and for each $i$ in $n$ there is a homomorphism $f_i$ from $G_i$ to $H$. Suppose further that, for all distinct $i$ and $j$ in $n$,*

$$f_i(x)\cdot f_j(y) = f_j(y)\cdot f_i(x).$$

*Then the map*

$$(x_i : i < n) \mapsto \prod_{i<n} f_i(x_i)$$

*from $\prod_{i<n} G_i$ to $H$ is the unique homomorphism $f$ from $\prod_{i<n} G_i$ to $H$ such that, for each $i$ in $n$,*

$$f\iota_i = f_i.$$

We use the porism to establish the next theorem below, which we shall use in characterizing finite nilpotent groups in Theorem 167 on page 161. We need the following observation.

**Lemma 10.** *If $M$ and $N$ are normal subgroups of $G$, and*

$$M \cap N = \{\mathrm{e}\},$$

*then each element of $M$ commutes with each element of $N$, that is, for all $m$ in $M$ and $n$ in $N$,*

$$mn = nm.$$

*Proof.* We can analyze $mnm^{-1}n^{-1}$ both as the element $(mnm^{-1})n^{-1}$ of $N$ and as the element $m(nm^{-1}n^{-1})$ in $M$; so the element is e, and therefore $mn = (m^{-1}n^{-1})^{-1} = nm$. □

**Theorem 124.** *If $(N_i : i < n)$ is a finite indexed family of normal subgroups of a group, and for each $j$ in $n \smallsetminus \{0\}$,*

$$N_0 \cdots N_{j-1} \cap N_j = \{\mathrm{e}\}, \tag{4.4}$$

*then the map*

$$(x_i : i < n) \mapsto \prod_{i<n} x_i \tag{4.5}$$

*from $\prod_{i<n} N_i$ to $N_0 \cdots N_{n-1}$ is an isomorphism.*

*Proof.* Say the $N_i$ are normal subgroups of the group $G$, and let the map in (4.5) be denoted by $h$. Since $N_i \cap N_j = \{\mathrm{e}\}$ whenever $i \neq j$, the last porism and the lemma guarantee that $h$ is a homomorphism and, for each $i$ in $n$, the composition $h\iota_i$ is just the inclusion of $N_i$ in $G$. Then the range of $h$ is $N_0 \cdots N_{n-1}$. To see that $h$ is injective, note that, if $\boldsymbol{m} \in \prod_{i\in n} N_i$ and $h(\boldsymbol{m}) = \mathrm{e}$, then

$${m_{n-1}}^{-1} = \prod_{i<n-1} m_i.$$

The left member is in $N_{n-1}$, and the right is in $N_0 \cdots N_{n-2}$, so each member is e. In particular, $m_{n-1} = \mathrm{e}$, but also, we can repeat the argument to show $m_{n-2} = \mathrm{e}$ and so on. Thus $\boldsymbol{m} = \mathrm{e}$. □

In the theorem, the group $N_0 \cdots N_{n-1}$ is the **internal direct product** of $(N_i : i < n)$. For the result, it is not enough to assume $N_i \cap N_j = \{\mathrm{e}\}$ when $i < j < n$. For example, consider the subgroups $\langle(1,0)\rangle$, $\langle(0,1)\rangle$, and $\langle(1,1)\rangle$ of $\mathrm{V}_4$.

We can generalize Theorem 123 in another sense. Given an arbitrary indexed family $(G_i : i \in I)$ of abelian groups, we define its **direct sum,**

$$\sum_{i \in I} G_i,$$

to consist of the elements $(x_i : i \in I)$ of the direct product $\prod_{i \in I} G_i$ such that the set $\{i \in I : x_i \neq 0\}$ is finite. The direct sum is indeed a group:

**Theorem 125.** *For every indexed family $(G_i : i \in I)$ of abelian groups,*

$$\sum_{i \in I} G_i < \prod_{i \in I} G_i.$$

In case $I = n$, we may write $\sum_{i \in I} G_i$ also as

$$G_0 \oplus \cdots \oplus G_{n-1}.$$

If $I$ is finite, then the direct sum is the same as the direct product. If $I$ is infinite, and the groups $G_i$ are nontrivial for infinitely many $i$ in $I$, then the sum is *not* the same as the direct product. The proof uses the Axiom of Choice, because it involves choosing a nontrivial element from each of infinitely many of the nontrivial groups $G_i$.

**Porism 123.2.** *Suppose $(G_i : i \in I)$ is an indexed family of abelian groups, and $H$ is an abelian group, and for each $i$ in $I$ there is a homomorphism $f_i$ from $G_i$ to $H$. Then the map*

$$x \mapsto \sum_{i \in I} f_i(x_i)$$

*from $\sum_{i \in I} G_i$ to $H$ is the unique homomorphism $f$ from $\sum_{i \in I} G_i$ to $H$ such that, for each $i$ in $I$,*

$$f \iota_i = f_i,$$

*that is, the following diagram commutes:*

$$\begin{tikzcd} G_i \arrow[r, "\iota_i"] \arrow[dr, "f_i"'] & \displaystyle\sum_{j\in I} G_j \arrow[d, "f"] \\ & H \end{tikzcd}$$

## 4.3. *Weak direct products

For completeness, we observe that Theorem 123 can be generalized even further. The **weak direct product** of an indexed family $(G_i : i \in I)$ of arbitrary groups has the same definition as the direct sum in the abelian case; but in the general case we use the notation

$$\prod_{i\in I}^{\mathrm{w}} G_i.$$

So this comprises those elements $(x_i : i \in I)$ of $\prod_{i\in I} G_i$ such that the set $\{i \in I : x_i \neq \mathrm{e}\}$ is finite. For each $i$ in $I$ we have the homomorphism $\iota_i$ from $G_i$ to $\prod_{i\in I}^{\mathrm{w}} G_i$, defined as in the abelian case. Direct products and weak direct products are related as follows.

**Theorem 126.** *Let $(G_i : i \in I)$ be an indexed family of groups. Then*

$$\iota_j[G_j] \lhd \prod_{i\in I}^{\mathrm{w}} G_i, \qquad \prod_{i\in I}^{\mathrm{w}} G_i \lhd \prod_{i\in I} G_i, \qquad \iota_j[G_j] \lhd \prod_{i\in I} G_i.$$

Porism 123.2 can be generalized to some cases of arbitrary groups:

**Porism 123.3.** *Suppose $(G_i : i \in I)$ is an indexed family of groups, and $H$ is a group, and for each $i$ in $I$ there is a homomorphism $f_i$ from $G_i$ to $H$. Suppose further that, for all distinct $i$ and $j$ in $I$,*

$$f_i(x) \cdot f_j(y) = f_j(y) \cdot f_i(x).$$

*Then the map*

$$x \mapsto \prod_{i\in I} f_i(x_i)$$

*from* $\prod^{\mathrm{w}}_{i\in I} G_i$ *to* $H$ *is the unique homomorphism* $f$ *from* $\prod^{\mathrm{w}}_{i\in I} G_i$ *to* $H$ *such that, for each* $i$ *in* $I$,

$$f\iota_i = f_i.$$

**Porism 124.3.** *If* $(N_i : i \in I)$ *is an indexed family of normal subgroups of a group, and for each* $j$ *in* $I$,

$$N_j \cap \left\langle \bigcup_{i\in I\smallsetminus\{j\}} N_i \right\rangle = \{\mathrm{e}\}, \tag{4.6}$$

*then*

$$\left\langle \bigcup_{i\in I} N_i \right\rangle \cong \prod_{i\in I}^{\mathrm{w}} N_i.$$

In this porism, the group $\left\langle \bigcup_{i\in I} N_i \right\rangle$ is called the **internal weak direct product** of the $N_i$.

## 4.4. Free groups

For every index set $I$, the direct sum $\sum_{i\in I} \mathbb{Z}$ is called a **free abelian group on** $I$ for the reason given by the next theorem. To state the theorem, we note that, for every $i$ in $I$, the abelian group $\sum_{i\in I} \mathbb{Z}$ has the element $\iota_i(1)$, which can also be written as $(\delta^i_j : j \in I)$, where

$$\delta^i_j = \begin{cases} 1, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

Let us also use the notation

$$\mathbf{e}^i$$

for $\iota_i(1)$ or $(\delta^i_j : j \in I)$. An arbitrary element of $\sum_{i\in I} \mathbb{Z}$ can then be written as

$$\sum_{i\in I} x_i \mathbf{e}^i.$$

The use of this notation implies that only finitely many of the $x_i$ are different from 0.

**Theorem 125.** *Suppose $G$ is an abelian group, $I$ is a set, and $f$ is a function from $I$ to $G$. Then the map*

$$\sum_{i\in I} x_i\mathbf{e}^i \mapsto \sum_{i\in I} x_i f(i)$$

*from $\sum_{i\in I}\mathbb{Z}$ to $G$ is the unique homomorphism $\tilde{f}$ from $\sum_{i\in I}$ to $G$ such that, for each $i$ in $I$,*

$$\tilde{f}(\mathbf{e}^i) = f(i),$$

*that is, the following diagram commutes, where $\iota$ is the map $i \mapsto \mathbf{e}^i$.*

$$\begin{tikzcd} I \arrow[r, "\iota"] \arrow[d, "f"'] & \sum_{i\in I}\mathbb{Z} \arrow[dl, "\tilde{f}"] \\ G & \end{tikzcd}$$

*In particular, the subgroup $\langle f(i) \colon i \in I\rangle$ of $G$ is isomorphic to a quotient of $\sum_{i\in I}\mathbb{Z}$.*

As a special case, we have that every finitely generated abelian group is isomorphic to a quotient of some $\mathbb{Z}\oplus\cdots\oplus\mathbb{Z}$. Observing this is the first step in classifying the finitely generated abelian groups as in §4.7 (page 139).

Meanwhile, since

$$\sum_{i\in I}\mathbb{Z} = \langle \mathbf{e}^i \colon i \in I\rangle,$$

we can write every element as a finite sum $\sum_{i\in I} x_i\mathbf{e}^i$, as we said. But then, if $x_i > 0$, we can replace $x_i\mathbf{e}^i$ with $x_i$-many copies of $\mathbf{e}^i$, and if $x_j < 0$, we can replace $x_j\mathbf{e}^j$ with $-x_j$-many copies of $-\mathbf{e}^j$. For example,

$$3\mathbf{e}^0 - 2\mathbf{e}^1 = \mathbf{e}^0 + \mathbf{e}^0 + \mathbf{e}^0 - \mathbf{e}^1 - \mathbf{e}^1.$$

In general, every nontrivial element of $\sum_{i\in I}\mathbb{Z}$ is *uniquely* a sum of some copies of the $\mathbf{e}^i$ and the $-\mathbf{e}^j$, if we disregard order, and if we never allow $\mathbf{e}^i$ and $-\mathbf{e}^i$ for the same $i$ to appear in the same sum. If we use multiplicative notation instead, and if we do not disregard order, what we get is not an abelian group, much less a free abelian group; but it is a *free group.*

To be precise, a **word** on $I$ is a finite nonempty string $t_0 t_1 \cdots t_n$, where each entry $t_k$ is either e, or else $a$ or $a^{-1}$ for some $a$ in $I$. A word is **reduced** if $a$ and $a^{-1}$ are never adjacent in it, and e is never adjacent to any other entry. Thus the only reduced word in which e can appear is just the word of length 1 whose only entry is e. The **free group on** $I$, denoted by

$$\mathrm{F}(I),$$

consists of the reduced words on $I$. Multiplication in this group is juxtaposition followed by **reduction,** namely, replacement of each occurrence of $aa^{-1}$ or $a^{-1}a$ with e, and replacement of each occurrence of $x\,\mathrm{e}$ or $\mathrm{e}\,x$ with $x$. Thus, if we write an element $a$ of $I$ as $a^1$, we can express the product of two arbitrary reduced words by the equation

$$(a_m^{\varepsilon(m)} \cdots a_0^{\varepsilon(0)})(b_0^{\zeta(0)} \cdots b_n^{\zeta(n)}) = a_m^{\varepsilon(m)} \cdots a_j^{\varepsilon(j)} b_j^{\zeta(j)} \cdots b_n^{\zeta(n)},$$

where each exponent $\varepsilon(i)$ or $\zeta(i)$ is $\pm 1$, and the equation

$$a_i^{\varepsilon(i)} = b_i^{-\zeta(i)}$$

is true when $i < j$, but false when $i = j$. We consider $I$ as a subset of $\mathrm{F}(I)$. An element of the latter other than e can be written also as

$$a_0{}^{n(0)} \cdots a_m{}^{n(m)},$$

where $a_i$ and $a_{i+1}$ are always distinct elements of $I$, and each $n(i)$ is in $\mathbb{Z} \smallsetminus \{0\}$.

We can now give the following analogue for Theorem 125. This solves the question raised on page 101 above of how to describe the elements of a generated subgroup $\langle A \rangle$ of a given group. The answer is that these elements can be given as reduced words on $A$, although possibly the two different reduced words will stand for the same element of $\langle A \rangle$.

**Theorem 126.** *Suppose $G$ is a group, $I$ is a set, and $f$ is a function from $I$ to $G$. Then the map*

$$a_0{}^{n(0)} \cdots a_m{}^{n(m)} \mapsto f(a_0)^{n(0)} \cdots f(a_m)^{n(m)}$$

*from* $\mathrm{F}(I)$ *to* $G$ *is the unique homomorphism* $\tilde{f}$ *from* $\mathrm{F}(I)$ *to* $G$ *such that*

$$\tilde{f} \upharpoonright I = f,$$

*that is, the following diagram commutes, where* $\iota$ *is the inclusion of* $I$ *in* $\mathrm{F}(I)$*.*

$$\begin{tikzcd} I \arrow[r, "\iota"] \arrow[d, "f"'] & \mathrm{F}(I) \arrow[dl, "\tilde{f}"] \\ G & \end{tikzcd}$$

*In particular, the subgroup* $\langle f(i) \colon i \in I\rangle$ *of* $G$ *is isomorphic to a quotient of* $\mathrm{F}(I)$*.*

## 4.5. *Categories

Suppose $\boldsymbol{C}$ is a class of structures, all having the same signature. For example, $\boldsymbol{C}$ could be the class of all groups, or the class of all abelian groups. If $\mathfrak{A}$ and $\mathfrak{B}$ belong to $\boldsymbol{C}$, we can denote by

$$\mathrm{Hom}(\mathfrak{A}, \mathfrak{B})$$

the set of all homomorphisms from $\mathfrak{A}$ to $\mathfrak{B}$. By Theorem 26 on page 41, if also $\mathfrak{C} \in \boldsymbol{C}$, then

$$(g, f) \mapsto g \circ f \colon \mathrm{Hom}(\mathfrak{B}, \mathfrak{C}) \times \mathrm{Hom}(\mathfrak{A}, \mathfrak{B}) \to \mathrm{Hom}(\mathfrak{A}, \mathfrak{C}).$$

By Theorem 11 on page 29, if $f \in \mathrm{Hom}(\mathfrak{A}, \mathfrak{B})$, $g \in \mathrm{Hom}(\mathfrak{B}, \mathfrak{C})$, and $h \in \mathrm{Hom}(\mathfrak{C}, \mathfrak{D})$, then

$$(h \circ g) \circ f = h \circ (g \circ f). \tag{4.7}$$

By Theorem 27, $\mathrm{Hom}(\mathfrak{A}, \mathfrak{A})$ contains $\mathrm{id}_A$. If $f \in \mathrm{Hom}(\mathfrak{A}, \mathfrak{B})$ and $g \in \mathrm{Hom}(\mathfrak{B}, \mathfrak{C})$, then by Theorem 12,

$$\mathrm{id}_B \circ f = f, \qquad\qquad g \circ \mathrm{id}_B = g. \tag{4.8}$$

Because of these properties, $\boldsymbol{C}$ is called a **category.** Elements of $\boldsymbol{C}$ are called **objects** of the category; elements of each set $\mathrm{Hom}(\mathfrak{A}, \mathfrak{B})$ are called **morphisms** or **arrows** of the category, and specifically morphisms or arrows **from** $\mathfrak{A}$ **to** $\mathfrak{B}$. Strictly, the category is specified by four things:

1) the class $\boldsymbol{C}$,
2) the function $(\mathfrak{A}, \mathfrak{B}) \mapsto \mathrm{Hom}(\mathfrak{A}, \mathfrak{B})$ on $\boldsymbol{C} \times \boldsymbol{C}$,
3) the functions $\circ$, satisfying (4.7);
4) the function $\mathfrak{A} \mapsto \mathrm{id}_A$ on $\boldsymbol{C}$, satisfying (4.8).

The conditions (4.7) and (4.8) can be expressed by means of the following commutative diagrams.

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "f"'] & B \arrow[d, "g"] \\
B \arrow[r, "g"'] \arrow[ur, "\mathrm{id}_B"] & C
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
B \arrow[d, "g"'] & A \arrow[l, "f"'] \arrow[r, "f"] \arrow[d] \arrow[dl, "g\circ f"] & B \arrow[d, "g"] \arrow[dl, "h\circ g"'] \\
C \arrow[r, "h"'] & D & C \arrow[l, "h"]
\end{tikzcd}$$

It is possible to have a category in which the objects are not structures and the arrows are not homomorphisms. For example, if $G$ is a group, then its elements can be considered as objects of a category in which $\mathrm{Hom}(a, b) = \{ba^{-1}\}$, and $c \circ d = cd$, and the function corresponding to $\mathfrak{A} \mapsto \mathrm{id}_A$ is simply the constant function $a \mapsto \mathrm{e}$.

In an arbitrary category, the objects may be denoted by plain capital letters like $A$ and $B$, and the function corresponding to $\mathfrak{A} \mapsto \mathrm{id}_A$ may be denoted simply by $A \mapsto \mathrm{id}_A$. In accordance with Theorems 13 and 27, we say that an element $f$ of $\mathrm{Hom}(A, B)$ is an **isomorphism** if, for some $g$ in $\mathrm{Hom}(B, A)$,

$$g \circ f = \mathrm{id}_A, \qquad\qquad f \circ g = \mathrm{id}_B\,.$$

In this case, $g$ is an **inverse** of $f$.

**Theorem 127.** *In a category, inverses are unique, and the inverse of a morphism has its own inverse, which is that morphism.*

*Proof.* If $g$ and $h$ are inverses of $f$, then

$$g = g \circ \mathrm{id}_B = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{id}_A \circ h = h.$$

The rest is by symmetry of the definition. □

If it exists, then the inverse of $f$ is denoted by

$$f^{-1}.$$

When each object of a category has an associated set, and every arrow from an object with associated set $A$ to an object with associated set $B$ is actually a function from $A$ to $B$, then the category is said to be **concrete.** We shall be interested only in concrete categories. Classes of structures, like $\boldsymbol{C}$ above, can be understood as concrete categories. However, other kinds of concrete categories are possible. For example, there is a concrete category whose objects are topological spaces and whose arrows are continuous functions.

### 4.5.1. Products

Suppose $\boldsymbol{C}$ is a category, and $\mathscr{A}$ is an indexed family $(A_i : i \in I)$ of objects of $\boldsymbol{C}$. If it exists, the *product* of $\mathscr{A}$ in the category is an object with the properties of a direct product of groups given by Porism 121.1 on page 122. For a formal definition, we define a new category, whose objects are the pairs

$$\bigl(B, (f_i : i \in I)\bigr)$$

such that $B$ is an object of $\boldsymbol{C}$ and, for each $i$ in $I$,

$$f_i \in \operatorname{Hom}(B, A_i).$$

An element $h$ of $\operatorname{Hom}(C, B)$ is a morphism from $\bigl(C, (g_i : i \in I)\bigr)$ to $\bigl(B, (f_i : i \in I)\bigr)$ in the new category if, for each $i$ in $I$,

$$f_i \circ h = g_i,$$

that is, the following diagram commutes.

$$\begin{tikzcd} C \arrow[r, "g_i"] \arrow[d, "h"'] & A_i \arrow[d, "\mathrm{id}_{A_i}"] \\ B \arrow[r, "f_i"'] & A_i \end{tikzcd}$$

Suppose, in the new category, there is an object to which there is a *unique* morphism from every other object. This object is called a **product** of $\mathscr{A}$.

By Porism 121.1, if $(G_i : i \in I)$ is an indexed family of groups, then the ordered pair $\bigl(\prod_{i\in I} G_i, (\pi_i : i \in I)\bigr)$ is a product of the indexed family in the category of groups. If the $G_i$ are abelian, then the pair is a product in the category of abelian groups.

**Theorem 128.** *Any two products of the same indexed family of objects in the same category are uniquely isomorphic.*

Thus, if $\mathscr{A}$ is an indexed family $(A_i : i \in I)$ of objects in a category with products, then we may refer to *the* product of $\mathscr{A}$, denoting it by

$$\Bigl(\prod \mathscr{A}, (\pi_i : i \in I)\Bigr).$$

We may still refer to the morphisms $\pi_i$ as **coordinate projections.**

### 4.5.2. Coproducts

Given a category, if we can reverse all of the arrows, and if we reverse composition correspondingly, then we still have a category, called the **dual** or **opposite** of the original category. A *co-product* or *sum* in a category is a product in the dual. Thus, suppose $\boldsymbol{C}$ is a category, and $\mathscr{A}$ is an indexed family $(A_i : i \in I)$ of objects of $\boldsymbol{C}$. We define a new category, whose objects are the pairs

$$\bigl(B, (f_i : i \in I)\bigr)$$

such that $B$ is an object of $\boldsymbol{C}$ and, for each $i$ in $I$,

$$f_i \in \operatorname{Hom}(A_i, B).$$

An element $h$ of $\operatorname{Hom}(B, C)$ is a morphism from $\bigl(B, (f_i : i \in I)\bigr)$ to $\bigl(C, (g_i : i \in I)\bigr)$ in the new category if, for each $i$ in $I$,

$$h \circ f_i = g_i,$$

that is, the following diagram commutes.

$$\begin{tikzcd} C & A_i \arrow[l, "g_i"'] \\ B \arrow[u, "h"] & A_i \arrow[l, "f_i"] \arrow[u, "\mathrm{id}_{A_i}"'] \end{tikzcd}$$

Suppose, in the new category, there is an object from which there is a *unique* morphism to every other object. This object is called a **coproduct** or **sum** of $\mathscr{A}$.

By Porism 123.2, if $(G_i \colon i \in I)$ is an indexed family of abelian groups, then the pair $\left(\sum_{i \in I} G_i, (\iota_i \colon i \in I)\right)$ is its coproduct in the category of abelian groups.

By Theorem 128, coproducts are unique when they exist at all. Thus if $\mathscr{A}$ is an indexed family $(A_i \colon i \in I)$ of objects in a category with coproducts, then we may refer to *the* coproduct of $\mathscr{A}$, denoting it by one of

$$\left(\coprod \mathscr{A}, (\iota_i \colon i \in I)\right), \qquad \left(\sum \mathscr{A}, (\iota_i \colon i \in I)\right).$$

We may still refer to the the morphisms $\iota_i$ as **canonical injections.**

Weak direct products are *not* coproducts in the category of groups. However, this category *has* coproducts, as follows.

The **free product** of an indexed family $(G_i \colon i \in I)$ of groups is the group, denoted by

$$\prod_{i \in I}^{*} G_i,$$

or by

$$G_0 * \cdots * G_{n-1}$$

if $I$ is some $n$ in $\omega$, comprising the string e together with strings $t_0 \cdots t_m$, where each entry $t_i$ is an ordered pair $(g, n(i))$ such that $n(i) \in I$ and $g \in G_{n(i)} \smallsetminus \{\mathrm{e}\}$, and $n(i) \neq n(i{+}1)$. This complicated definition allows for the possibility that $G_i$ might be the same as $G_j$ for some distinct $i$ and $j$; the groups $G_i$ and $G_j$ must be considered as distinct in the formation of the free product. Multiplication on $\prod_{i \in I}^{*} G_i$, as on $\mathrm{F}(I)$, is juxtaposition followed by reduction, so that if $(g, i)$ is followed directly by $(h, i)$, then they are replaced with $(gh, i)$, and all instances of $(\mathrm{e}, i)$ are deleted, or replaced with e if there is no other entry. Each $G_j$ embeds in $\prod^*_{i \in I} G_i$ under $\iota_j$, namely $x \mapsto (x, j)$.

**Theorem 129.** *If $(G_i \colon i \in I)$ is an indexed family of groups, then $\left(\prod^*_{i \in I} G_i, (\iota_i \colon i \in I)\right)$ is its coproduct in the category of groups.*

### 4.5.3. Free objects

Given a *concrete* category $\boldsymbol{C}$ and a set $I$, we define a new category, whose objects are the pairs

$$(f, A),$$

where $A$ is an object of $\boldsymbol{C}$, and $f$ is a function from $I$ to (the associated set of) $A$. An element $h$ of $\mathrm{Hom}(A, B)$ is a morphism from $(f, A)$ to $(g, B)$ in the new category if

$$h \circ f = g,$$

that is, the following diagram commutes.

$$\begin{tikzcd}
I \arrow[r, "f"] \arrow[d, "\mathrm{id}_I"'] & A \arrow[d, "h"] \\
I \arrow[r, "g"'] & B
\end{tikzcd}$$

Suppose, in the new category, from the object $(f, A)$, there is a unique morphism to every other object. Then $A$ is a **free object** on $I$ with respect to $f$.

**Theorem 130.** *In a concrete category $\boldsymbol{C}$, if $A$ is a free object on a set $I$ with respect to a function $f$, and $B$ is a free object on $I$ with respect to $g$, then there is a unique isomorphism $h$ from $A$ to $B$ such that $h \circ f = g$.*

By Theorems 125 and 126, free objects exist in the categories of abelian groups and of arbitrary groups. Another example will be given by Theorem 218 (page 212).

## 4.6. Presentation of groups

We develop a method for describing groups as quotients of free groups. Let us first note that every group *is* (isomorphic to) such a quotient.

**Theorem 131.** *Every group is isomorphic to the quotient of a free group by some normal subgroup.*

*Proof.* By Theorem 126 (page 130), the identity map from $G$ to itself extends to a homomorphism from $\mathrm{F}(G)$ to $G$. Since this homomorphism is surjective, the claim follows by the First Isomorphism Theorem (page 114). □

If $A$ is a subset of some group $G$, on page 98 we defined $\langle A\rangle$ as the intersection of (the set of) subgroups of $G$ that include $A$. We know this intersection is a subgroup of $G$, by Theorem 89. But possibly $\langle A\rangle$ is not a *normal* subgroup of $G$. However, we have the following.

**Theorem 132.** *An arbitrary intersection of normal subgroups is a subgroup.*

Now, given a subset $B$ of a group $G$, we can define

$$\langle\langle B\rangle\rangle = \bigcap \mathcal{N},$$

where $\mathcal{N}$ is the set of all normal subgroups of $G$ that include $B$. If $A$ is an arbitrary set, and $B \subseteq \mathrm{F}(A)$, we define

$$\langle A \mid B\rangle = \mathrm{F}(A)/\langle\langle B\rangle\rangle.$$

This is the group with **generators** $A$ and **relations** $B$. Note however that, strictly, the elements of $A$ as such do not generate the group; rather, the cosets $a\langle\langle B\rangle\rangle$, where $a \in A$, generate the group. But we can understand $a$ as a name for the coset $a\langle\langle B\rangle\rangle$.

Suppose there is a function $f$ from $A$ to a group $G$, and $\tilde{f}$ is the homomorphism from $\mathrm{F}(A)$ to $G$ that extends $f$, and this homomorphism is surjective, and its kernel is $\langle\langle B\rangle\rangle$. By the First Isomorphism Theorem,

$$G \cong \langle A \mid B\rangle.$$

We say in this case that $\langle A \mid B\rangle$ is a **presentation** of $G$. If $A = \{a_0, \dots, a_n\}$, and $B = \{w_0, \dots, w_m\}$, then $\langle A \mid B\rangle$ can be written as

$$\langle a_0, \dots, a_n \mid w_0, \dots, w_m\rangle.$$

Sometimes, instead of $w_i$, one may write $w_i = \mathrm{e}$ or an equivalent equation. Meanwhile, $\mathrm{F}(A)$ can be presented as $\langle A \mid \varnothing\rangle$. In particular $\mathbb{Z}$ can be presented as $\langle a \mid \varnothing\rangle$, but also as $\langle a, b \mid ab^{-1}\rangle$ or $\langle a, b \mid a = b\rangle$. The group $\mathbb{Z}_n$ has the presentation $\langle a \mid a^n\rangle$. More examples are given by the theorems after the next.

**Theorem 133** (von Dyck[1]). *Suppose $G$ is a group, $A$ is a set, $f\colon A \to G$, and $\tilde{f}$ is the induced homomorphism from $\mathrm{F}(A)$ to $G$. Suppose further*

$$B \subseteq \ker \tilde{f}$$

*Then there is a well-defined homomorphism $g$ from $\langle A \mid B\rangle$ to $G$ such that $g(a\langle\langle B\rangle\rangle) = f(a)$ for each $a$ in $A$, that is, the following diagram commutes.*

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d] & G \\
\mathrm{F}(A) \arrow[ur, "\tilde{f}"] \arrow[r, "\pi"'] & \langle A \mid B\rangle \arrow[u, "g"']
\end{tikzcd}$$

*If $G = \langle f(a)\colon a \in A\rangle$, then $g$ is an epimorphism.*

*Proof.* Since $\ker(\tilde{f})$ is a normal subgroup of $\mathrm{F}(A)$ that includes $B$, we have $\langle\langle B\rangle\rangle < \ker \tilde{f}$. Hence $g$ is well-defined by Theorem 113 on page 113. □

**Theorem 134.** *If $n > 2$, then $\mathrm{Dih}(n)$ has the presentation*

$$\langle a, b \mid a^n, b^2, (ab)^2\rangle.$$

*Proof.* Note first that, in the group $\langle a, b \mid a^n, b^2, (ab)^2\rangle$, the order of $a$ must divide $n$, and each of the orders of $b$ and $ab$ must divide 2. Now, by Theorem 95 on page 102, $\mathrm{Dih}(n)$ has elements $\alpha$ and $\beta$ that generate the group and are such that $\alpha^n$, $\beta^2$, and $(\alpha\beta)^2$ are all equal to e. By von

[1] Walther von Dyck (1856–1934) gave an early (1882–3) definition of abstract groups [20, ch. 49, p. 1141].

Dyck's Theorem then, there is an epimorphism from $\langle a, b \mid a^n, b^2, (ab)^2\rangle$ to $\mathrm{Dih}(n)$ taking $a$ to $\alpha$ and $b$ to $\beta$ and hence $ab$ to $\alpha\beta$. Therefore the order of $a$ must be exactly $n$, and the orders of $b$ and of $ab$ must be 2. By Theorem 99 on page 104, the epimorphism onto $\mathrm{Dih}(n)$ must be an isomorphism. □

**Theorem 135.** *The quaternion group* $\mathrm{Q}_8$ *has the presentation*

$$\langle \mathrm{i}, \mathrm{j} \mid \mathrm{i}^4, \mathrm{i}^2\mathrm{j}^2, \mathrm{i}\mathrm{j}\,\mathrm{i}^3\mathrm{j}\rangle,$$

*or equivalently* $\langle \mathrm{i}, \mathrm{j} \mid \mathrm{i}^4 = \mathrm{e},\ \mathrm{i}^2 = \mathrm{j}^2,\ \mathrm{j}\,\mathrm{i} = \mathrm{i}^3\mathrm{j}\rangle$.

*Proof.* Use von Dyck's Theorem and Theorem 100 in the manner of the previous proof. □

Yet another example of a presentation will be given in Theorem 161 on page 157.

## 4.7. Finitely generated abelian groups

We now classify, in the sense of §3.7 (page 116), the abelian groups with finite sets of generators, and in particular the finite abelian groups. A useful application of this will be that the group of units of every finite field is cyclic (Theorem 140).

**Theorem 136.** *If* $(G_i \colon i \in I)$ *is an indexed family of groups, and for each* $i$ *in* $I$, $N_i \lhd G_i$, *then*

$$\prod_{i\in I} N_i \lhd \prod_{i\in I} G_i, \qquad \prod_{i\in I} G_i \Big/ \prod_{i\in I} N_i \cong \prod_{i\in I} \frac{G_i}{N_i}.$$

**Theorem 137.** *For every abelian group* $G$ *on* $n$ *generators, there is a unique element* $k$ *of* $n+1$, *along with positive integers* $d_0$, …, $d_{k-1}$, *where*

$$d_0 \mid d_1 \wedge \cdots \wedge d_{k-2} \mid d_{k-1}, \tag{4.9}$$

*such that*

$$G \cong \mathbb{Z}_{d_0} \oplus \cdots \oplus \mathbb{Z}_{d_{k-1}} \oplus \underbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{n-k}. \tag{4.10}$$

*Proof.* Suppose $G = \langle g^i : i < n \rangle$ and is abelian. Let $F$ be the free abelian group $\sum_{i \in n} \mathbb{Z}$. Using notation from page 128, we have that $F = \langle \mathbf{e}^0, \dots, \mathbf{e}^{n-1} \rangle$, and there is a surjective function

$$\sum_{i \in n} x_i \mathbf{e}^i \mapsto \sum_{i \in n} x_i g^i$$

from $F$ to $G$. Let $N$ be its kernel, so that

$$G \cong F/N.$$

Suppose it should happen to be that $N = \langle d_0 \mathbf{e}^0, \dots, d_{k-1} \mathbf{e}^{k-1} \rangle$. We have

$$F \cong \langle \mathbf{e}^0 \rangle \oplus \cdots \oplus \langle \mathbf{e}^{n-1} \rangle,$$

and under the isomorphism,

$$N \cong \langle d_0 \mathbf{e}^0 \rangle \oplus \cdots \oplus \langle d_{k-1} \mathbf{e}^{k-1} \rangle \oplus \{\mathrm{e}\} \oplus \cdots \oplus \{\mathrm{e}\}.$$

By the lemma then,

$$F/N \cong \frac{\langle \mathbf{e}^0 \rangle}{\langle d_0 \mathbf{e}^0 \rangle} \oplus \cdots \oplus \frac{\langle \mathbf{e}^{k-1} \rangle}{\langle d_{k-1} \mathbf{e}^{k-1} \rangle} \oplus \langle \mathbf{e}^k \rangle \oplus \cdots \oplus \langle \mathbf{e}^{n-1} \rangle,$$

which has the form in (4.10), although (4.9) might not hold. Not every subgroup of $F$ is given to us so neatly, but we shall be able to put it into the desired form, even satisfying (4.9).

We can identify $F$ with $\mathrm{M}_{1 \times n}(\mathbb{Z})$. If $X \in \mathrm{M}_{m \times n}(\mathbb{Z})$, let us denote by $\langle X \rangle$ the subgroup of $F$ generated by the rows of $X$. If $P \in \mathrm{GL}_m(\mathbb{Z})$ and $Q \in \mathrm{GL}_n(\mathbb{Z})$, then

$$\langle X \rangle = \langle PX \rangle, \qquad F/\langle X \rangle \cong F/\langle XQ \rangle.$$

Now we can choose $P$ and $Q$ so as to effect certain row operations (as on page 85) and column operations, respectively. In particular, assuming $m \geqslant n$, for some $P$ we have

$$PX = \begin{pmatrix} U \\ 0 \end{pmatrix},$$

where $U$ is an $n \times n$ **upper triangular** matrix, that is,

$$U = \begin{pmatrix} * & \cdots & * \\ & \ddots & \vdots \\ 0 & & * \end{pmatrix}.$$

Then we may assume $m = n$, so $PX = U$. For some $Q$, the matrix $PXQ$ is **diagonal,** so that

$$PXQ = \begin{pmatrix} d_0 & & 0 \\ & \ddots & \\ 0 & & d_{n-1} \end{pmatrix}.$$

By further adjusting $P$ and $Q$, we may ensure that (4.9) holds, while $d_k = \cdots = d_{n-1} = 0$. Indeed, suppose $b, c \in \mathbb{Z}$ and $\gcd(b, c) = d$. By elementary row and column operations, from a matrix

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

we obtain $\begin{pmatrix} b & 0 \\ c & c \end{pmatrix}$ and then $\begin{pmatrix} d & e \\ 0 & f \end{pmatrix}$, where $e$ and $f$ are multiples of $c$ and hence of $d$; hence, with an invertible column operation, we get

$$\begin{pmatrix} d & 0 \\ 0 & f \end{pmatrix}.$$

where again $d \mid f$. Applying such transformations as needed to pairs of entries in $D$ yields (4.9). The number $k$ is uniquely determined by $X$. We have shown that every subgroup of $F$ is generated by a set of at most $n$ elements. Then we may assume $N = \langle X \rangle$, so that $F/N$ is as desired. □

**Porism 137.1.** *Every subgroup of a free abelian group on $n$ generators is free abelian on $n$ generators or fewer.*

In the theorem, not only is $k$ unique, but the numbers $d_j$ are also unique. This can be established by means of an alternative classification of the finitely generated abelian groups.

**Theorem 138** (Chinese Remainder)**.** *If* $\gcd(m,n)=1$, *then the homomorphism* $x \mapsto (x,x)$ *from* $\mathbb{Z}_{mn}$ *to* $\mathbb{Z}_m \oplus \mathbb{Z}_n$ *is an isomorphism.*

*Proof.* If $x \equiv 0 \pmod{m}$ and $x \equiv 0 \pmod{n}$, then $x \equiv 0 \pmod{mn}$. Hence the given homomorphism is injective. Since $\mathbb{Z}_{mn}$ and $\mathbb{Z}_m \oplus \mathbb{Z}_n$ both have order $mn$, the given homomorphism must also be surjective, by Theorem 43 on page 51. □

The Chinese Remainder Theorem will be generalized as Theorem 194 on page 185. In the usual formulation of the theorem, every system

$$x \equiv a \pmod{m}, \qquad\qquad x \equiv b \pmod{n}$$

of congruences has a unique solution *modulo* $mn$; but this solution is just the inverse image of $(a,b)$ under the isomorphism $x \mapsto (x,x)$.

**Theorem 139.** *For every finite abelian group, there is a unique list* $(p_i : i < k)$ *of primes, where*

$$p_0 \leqslant \dots \leqslant p_{k-1},$$

*there are unique elements* $m(0), \dots, m(k-1)$ *of* $\mathbb{N}$, *and there is a unique* $r$ *in* $\omega$ *such that*

$$G \cong \mathbb{Z}_{{p_0}^{m(0)}} \oplus \dots \oplus \mathbb{Z}_{{p_{k-1}}^{m(k-1)}} \oplus \underbrace{\mathbb{Z} \oplus \dots \oplus \mathbb{Z}}_{r}.$$

*Proof.* To obtain the analysis, apply the Chinese Remainder Theorem to Theorem 137. The analysis is unique, provided it is unique in the case where all of the $p_j$ are the same. But in this case, the analysis is unique, by repeated application of the observation that the order of the group is the highest prime power appearing in the factorization. □

**Theorem 140.** *The group of units of every finite field is cyclic. In particular, if* $p$ *is prime, then*

$${\mathbb{Z}_p}^{\times} \cong \mathbb{Z}_{p-1}.$$

*Proof.* Let $F$ be a finite field. By Theorem 137,

$$F^{\times} \cong \mathbb{Z}_{d_0} \oplus \mathbb{Z}_{d_{k-1}} \oplus \mathbb{Z}_m$$

for some $d(i)$ and $m$ such that

$$d_0 \mid d_i \wedge \cdots \wedge d_{k-1} \mid m.$$

In particular,

$$m \leqslant |F^{\times}|.$$

Also, every element of $F^{\times}$ is a zero of the polynomial $x^m - 1$. But this polynomial can have at most $m$ roots in a field. Thus

$$|F^{\times}| \leqslant m.$$

Hence $|F^{\times}| = m$ and so $F^{\times} \cong \mathbb{Z}_m$. □

If ${\mathbb{Z}_n}^{\times}$ is cyclic, then its generators are called **primitive roots** of $n$; Gauss [9, p. 37] attributes the terminology to Euler. Recall from page 109 the definition

$$\phi(n) = |{\mathbb{Z}_n}^{\times}|.$$

Thus, if ${\mathbb{Z}_n}^{\times}$ is indeed cyclic, it is isomorphic to $\mathbb{Z}_{\phi(n)}$.

**Theorem 141.** *If $n$ has a primitive root $a$, then it has exactly $\phi(\phi(n))$ primitive roots, namely those $a^k$ such that* $\gcd(k, \phi(n)) = 1$.

By Theorem 140, primes have primitive roots. We have to find them by trial. For example, 2 is not a primitive root of 7, but 3 is, by the following computations.

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | (mod 6) |
|---|---|---|---|---|---|---|---|
| $2^k$ | 1 | 2 | $-3$ | 1 | 2 | $-3$ | (mod 7) |
| $3^k$ | 1 | 3 | 2 | $-1$ | $-3$ | $-2$ | (mod 7) |

Then 5 (or $-2$) is the only other primitive root of 7.

# 5. Finite groups

## 5.1. Semidirect products

Recall from page 118 that *conjugation* in a group is an operation $x \mapsto axa^{-1}$ for some element $a$ of the group. The following is reminiscent of Cayley's Theorem (Theorem 49 on page 57), although the homomorphism now need not be an embedding.

**Theorem 142.** *Conjugation in a group is an automorphism. For every group $G$, the function*

$$g \mapsto (x \mapsto gxg^{-1})$$

*from $G$ to $\mathrm{Aut}(G)$ is a homomorphism.*

Conjugation by an arbitrary element of a group is also called an **inner automorphism** of the group. The kernel of the homomorphism in the theorem is the **center** of $G$, denoted by

$$\mathrm{C}(G).$$

We shall generalize this notion in §5.5 (page 158).[1] Meanwhile, it will be useful to have the following generalization of the last theorem.

**Theorem 143.** *For every group $G$, if $N \lhd G$, then there is a homomorphism*

$$g \mapsto (x \mapsto gxg^{-1})$$

*from $G$ to $\mathrm{Aut}(N)$.*

[1] Repeating the process of forming inner automorphisms, we can define a function $\alpha \mapsto G_\alpha$ on the class of ordinals so that $G_0 = G$, and $G_{\alpha'} = \mathrm{Aut}(G_\alpha)$, and if $\beta$ is a limit, then $G_\beta$ is the so-called *direct limit* of $(G_\alpha \colon \alpha < \beta)$. Then for some ordinal $\alpha$, for all ordinals $\beta$, if $\beta \geqslant \alpha$, then $G_\beta = G_\alpha$: Simon Thomas [35] shows this in case $G$ has trivial center; Joel Hamkins [13], in the general case.

In the theorem, let the homomorphism be $g \mapsto \sigma_g$. Suppose also $H < G$, and $N \cap H = \{\mathrm{e}\}$. Then the conditions of Theorem 111 (page 112) are met, and $NH$ is an internal semidirect product. Equation (3.7) describing multiplication on $NH$, namely

$$(mg)(nh) = (m \cdot gng^{-1})(gh),$$

can be rewritten as

$$(mg)(nh) = (m \cdot \sigma_g(n))(gh).$$

**Theorem 144.** *Suppose $N$ and $H$ are groups, and $g \mapsto \sigma_g$ is a homomorphism from $H$ to $\operatorname{Aut}(N)$. Then the set $N \times H$ becomes a group when multiplication is defined by*

$$(m, g)(n, h) = (m \cdot \sigma_g(n), gh).$$

The group given by the theorem is the **semidirect product** of $N$ and $H$ with respect to $\sigma$; it can be denoted by

$$N \rtimes_\sigma H.$$

The bijection in Theorem 111 is an isomorphism from $N \rtimes_\sigma H$ to $NH$ when $\sigma$ is the homomorphism in Theorem 143.

Now recall from Theorem 72 (page 80) that for every associative ring $(R, 1, \cdot)$, the function $x \mapsto \lambda_x$ embeds the ring in $(\operatorname{End}(R), \mathrm{id}_R, \circ)$. From this we obtain the following.

**Theorem 145.** *For every associative ring $(R, 1, \cdot)$, the function*

$$x \mapsto \lambda_x$$

*embeds $(R, \cdot)^\times$ in $\operatorname{Aut}(R)$.*

The embedding is sometimes an isomorphism:

**Theorem 146.** *For all $n$ in $\mathbb{N}$, the function*

$$x \mapsto \lambda_x$$

*is an isomorphism from ${\mathbb{Z}_n}^\times$ to $\operatorname{Aut}(\mathbb{Z}_n)$.*

**Theorem 147.** *If $p$ and $q$ are primes such that*

$$q \mid p-1,$$

*then there is an embedding $\sigma$ of $\mathbb{Z}_q$ in $\operatorname{Aut}(\mathbb{Z}_p)$, and hence there is a semidirect product*

$$\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q,$$

*which is not abelian. If $\tau$ is another embedding of $\mathbb{Z}_q$ in $\operatorname{Aut}(\mathbb{Z}_p)$, then for some $n$ in $\mathbb{Z}_q$, the map*

$$(y,x) \mapsto (y,nx)$$

*is an isomorphism from $\mathbb{Z}_p \rtimes_\tau \mathbb{Z}_q$ to $\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q$.*

*Proof.* The prime $p$ has a primitive root $a$ by Theorem 140 (page 142). Letting $b = a^{(p-1)/q}$, we have an isomorphism $x \mapsto b^x$ from $\mathbb{Z}_q$ to $\langle b \rangle$, and $\langle b \rangle$ is the unique subgroup of ${\mathbb{Z}_p}^\times$ of order $q$ (Theorem 98, page 103). By the last theorem, the map $x \mapsto \lambda_{b^x}$ is an embedding of $\mathbb{Z}_q$ in $\operatorname{Aut}(\mathbb{Z}_p)$. Calling this embedding $\sigma$, we can form

$$\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q.$$

Now suppose $\tau$ is an arbitrary embedding of $\mathbb{Z}_q$ in $\operatorname{Aut}(\mathbb{Z}_p)$. By uniqueness of $\langle b \rangle$ as a subgroup of ${\mathbb{Z}_p}^\times$ of order $q$, the images of $\tau$ and $\sigma$ must be the same, and so $\tau_1 = \lambda_{b^n}$ for some $n$ in ${\mathbb{Z}_q}^\times$, and hence

$$\tau_x = \sigma_{nx}.$$

The function $f$ from $\mathbb{Z}_p \times \mathbb{Z}_q$ to itself given by

$$f(y,x) = (y,nx)$$

is a bijection. If we denote multiplication in $\mathbb{Z}_p \rtimes_\tau \mathbb{Z}_q$ by $\cdot^\tau$, and likewise with $\sigma$ for $\tau$, then

$$\begin{aligned} f\bigl((c,b) \cdot^\tau (y,x)\bigr) &= f(c+\tau_b(y), b+x) \\ &= (c+\sigma_{nb}(y), n(b+x)) \\ &= (c+\sigma_{nb}(y), nb+nx) \\ &= (c,nb) \cdot^\sigma (y,nx) \\ &= f(c,b) \cdot^\sigma f(y,x). \end{aligned}$$

Thus $f$ is an isomorphism from $\mathbb{Z}_p \rtimes_\tau \mathbb{Z}_q$ to $\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q$. $\square$

In case $q = 2$, the group in the theorem is isomorphic to $\mathrm{Dih}(p)$. We investigate groups of order $pq$ a bit more in the next section. The final classification of them will be Theorem 159 on page 156.

## 5.2. Cauchy's Theorem

We can partition a group $G$ into subsets $\{a, a^{-1}\}$. Many of these may indeed have size 2; but $\{\mathrm{e}, \mathrm{e}^{-1}\} = \{\mathrm{e}\}$. Hence, if $G$ is finite of *even* order, we must have $\{a, a^{-1}\} = \{a\}$ for some $a$ other than e. In this case, $a$ has order 2.

We can recast this argument as follows. The function $x \mapsto x^{-1}$ is a permutation $\sigma$ of $G$ as a set. The function $f$ from $\mathbb{Z}_2$ to $\mathrm{Sym}(G)$ given by

$$f_0 = \mathrm{id}_G, \qquad\qquad f_1 = \sigma$$

is a homomorphism. Then $G$ is partitioned by the sets $\{f_x(a) : x \in \mathbb{Z}_2\}$. The size of such a set is 1 or 2. Hence the number of such sets of size 1 is congruent *modulo* 2 to the order of $G$.

Now we can generalize by replacing 2 with an arbitrary prime. Thus we obtain the first promised partial converse of the Lagrange Theorem (page 108). Galois apparently used the following in 1831–2; Cauchy published a proof in 1844 [2, pp. 142–4].

**Theorem 148** (Cauchy)**.** *For all primes $p$, every finite group whose order is a multiple of $p$ has an element of order $p$.*

*Proof (J. H. McKay [27]).* Suppose $G$ is a finite group whose order is divisible by $p$. Let $A$ be the range of the map

$$(x_0, \dots, x_{p-2}) \mapsto \big(x_0, \dots, x_{p-2}, (x_0 \cdots x_{p-2})^{-1}\big).$$

from $G^{p-1}$ to $G^p$. Thus

$$A = \Big\{(x_i : i < p) \in G^p : \prod_{i<p} x_i = \mathrm{e}\Big\}, \qquad |A| = |G^{p-1}|.$$

If $(x_i : i < p) \in A$ and $0 < k < p - 1$, then

$$(x_0 \cdots x_{k-1})^{-1} = x_k \cdots x_{p-1},$$

and so $(x_k, \dots, x_{p-1}, x_0, \dots, x_{k-1}) \in A$. Thus we have a homomorphism $f$ from $\mathbb{Z}_p$ to $\mathrm{Sym}(A)$ given by

$$f_k(x_0, \dots, x_{k-1}, x_k, \dots, x_{p-1}) = (x_k, \dots, x_{p-1}, x_0, \dots, x_{k-1}).$$

Then

$$\begin{gathered} f_k(\boldsymbol{x}) = f_\ell(\boldsymbol{x}) \iff f_{k-\ell}(\boldsymbol{x}) = \boldsymbol{x}, \\ \{k \in \mathbb{Z}_p : f_k(\boldsymbol{x}) = \boldsymbol{x}\} < \mathbb{Z}_p. \end{gathered}$$

Subgroups of $\mathbb{Z}_p$ have order 1 or $p$, and so the set $\{f_k(\boldsymbol{x}) : k \in \mathbb{Z}_p\}$ has size $p$ or 1. Such subsets partition $A$. One of the subsets, namely $\{(\mathrm{e}, \dots, \mathrm{e})\}$, has size 1. Since $|A|$ is a multiple of $p$, there must be $\boldsymbol{x}$ in $A$ different from $(\mathrm{e}, \dots, \mathrm{e})$ such that $f_k(\boldsymbol{x}) = \boldsymbol{x}$ for all $k$ in $\mathbb{Z}_p$. In this case, $\boldsymbol{x}$ must be $(x, \dots, x)$ for some $x$ in $G \smallsetminus \{\mathrm{e}\}$. Thus $x$ has order $p$. □

A $p$-**group** is a group the order of whose every element is a power of $p$.

**Corollary 148.1.** *A finite group is a $p$-group if and only if its order is a power of $p$.*

*Proof.* Let $\ell$ be a prime different from $p$. if $\ell$ divides $|G|$, then $G$ has an element of order $\ell$, so $G$ is not a $p$-group. Conversely, if $g \in G$ and $\ell$ divides $|g|$, then $\ell$ divides $|G|$. □

For example, the trivial group $\{\mathrm{e}\}$ is a $p$-group for every prime $p$. All groups $\mathbb{Z}_{p^k}$, and direct sums of them, are $p$-groups. If $n > 1$, then $\mathrm{Dih}(2^n)$ is a nonabelian 2-group.

By Cauchy's Theorem, the hypothesis of the following is always satisfied.

**Theorem 149.** *Suppose $p$ and $q$ are distinct primes, and $G$ is a group of order $pq$. If $a$ and $b$ are elements of $G$ of orders $p$ and $q$ respectively, then*

$$\langle a \rangle \cap \langle b \rangle = \{\mathrm{e}\}, \qquad G = \langle a \rangle \langle b \rangle.$$

In the theorem, if $\langle a\rangle$ is a *normal* subgroup of $G$, then $G$ is a semidirect product, by Theorem 111 on page 112. If also $\langle b\rangle \lhd G$, then $G$ is actually a direct product, isomorphic to $\mathbb{Z}_p \times \mathbb{Z}_q$. Otherwise, $G$ is not abelian, and by Theorem 147 there is only one possibility. With Theorem 159 on page 156, we shall show that one of $\langle a\rangle$ and $\langle b\rangle$ must be a normal subgroup of $G$, and so $G$ is indeed either a direct or a semidirect product.

## 5.3. Actions of groups

A homomorphism from a group $G$ to the symmetry group of a set $A$ is called an **action** of $G$ on $A$. An alternative characterization of actions is given by the following.

**Theorem 150.** *Let $G$ be a group, and $A$ a set. There is a one-to-one correspondence between*

1. *homomorphisms $g \mapsto (a \mapsto ga)$ from $G$ into* $\operatorname{Sym}(A)$*, and*
2. *functions $(g, a) \mapsto ga$ from $G \times A$ into $A$ such that*

$$\mathrm{e}\,a = a, \qquad (gh)a = g(ha) \tag{5.1}$$

*for all $h$ and $h$ in $G$ and $a$ in $A$.*

*Proof.* If $g \mapsto (a \mapsto ga)$ maps $G$ homomorphically into $\operatorname{Sym}(A)$, then the identities in (5.1) follow. Suppose conversely that these hold. Then, in particular,

$$g(g^{-1}a) = (gg^{-1})a = \mathrm{e}\,a = a$$

and likewise $g^{-1}(ga) = a$, so $a \mapsto g^{-1}a$ is the inverse of $a \mapsto ga$, and the function $g \mapsto (a \mapsto ga)$ does map $G$ into $\operatorname{Sym}(A)$, homomorphically by (5.1). $\square$

Usually it is a function $(g, a) \mapsto ga$ from $G \times A$ to $A$ as in the theorem that is called an action of $G$ on $A$. So in the notation of the proof of Cauchy's Theorem, the function $(k, \boldsymbol{x}) \mapsto f_k(\boldsymbol{x})$ is an action of $\mathbb{Z}_p$ on $A$. Immediately, for any set $A$, the function $(\sigma, x) \mapsto \sigma(x)$ from $\operatorname{Sym}(A) \times A$ to $A$ is an action of $\operatorname{Sym}(A)$ on $A$. Other examples that will be of interest to us are given by the following.

**Theorem 151.** *Let $G$ be a group and $H < G$. Then $G$ acts:*

a) *on itself by $(g, x) \mapsto \lambda_g(x)$ (left multiplication),*

b) *on $G/H$ by $(g, xH) \mapsto gxH$ (left multiplication),*

c) *on itself by $(g, x) \mapsto gxg^{-1}$ (conjugation),*

d) *on $\{xHx^{-1} \colon x \in G\}$ by $(g, K) \mapsto gKg^{-1}$ (conjugation).*

Suppose $(g, x) \mapsto gx$ is an arbitrary action of $G$ on $A$. If $a \in A$, then the subset $\{g \colon ga = a\}$ of $G$ is the **stabilizer** of $a$, denoted by

$$G_a;$$

the subset $\{ga \colon g \in G\}$ of $A$ is the **orbit** of $a$, denoted by

$$Ga.$$

The subset $\{x \colon G_x = G\}$ of $A$ can be denoted by

$$A_0.$$

Note how all of these were used in the proof of Cauchy's Theorem. Also, in the proof we established the appropriate case of the following.

**Theorem 152.** *Suppose a group $G$ acts on a set $A$. Then the orbits of the elements of $A$ under the action are a partition of $A$, that is,*

$$Ga \neq Gb \implies Ga \cap Gb = \varnothing, \qquad \bigcup_{a \in A} Ga = A.$$

*Moreover, for all $a$ in $A$,*

$$G_a < G, \qquad [G : G_a] = |Ga|.$$

*Proof.* Let the action be $(g, x) \mapsto gx$. for the last equation, we establish a bijection between $G/G_a$ and $Ga$ by noting that

$$gG_a = hG_a \iff h^{-1}g \in G_a \iff ga = ha;$$

so the bijection is $gG_a \mapsto ga$. $\square$

**Corollary 152.1.** *If there are only finitely many orbits in $A$ under $G$, then*

$$|A| = |A_0| + \sum_{a\in X}[G : G_a] \tag{5.2}$$

*for some set $X$ of elements of $A$ whose orbits are nontrivial.*

Equation (5.2) is called the **class equation.** We used it implicitly in the proof of Cauchy's Theorem. In fact we used it to derive the appropriate case of the following.

**Theorem 153.** *If $A$ is acted on by a finite $p$-group, then*

$$|A| \equiv |A_0| \pmod p$$

*Proof.* In the class equation, $[G : G_a]$ is a multiple of $p$ in each case. □

### 5.3.1. Centralizers

Suppose $G$ acts on itself by conjugation, and $a \in G$. Then $Ga$ is the **conjugacy class** of $a$, while $G_a$ is the **centralizer** of $a$, denoted by[2]

$$\mathrm{C}_G(a). \tag{5.3}$$

Finally, $G_0$ is the **center** of $G$, denoted by

$$\mathrm{C}(G);$$

this is a normal subgroup of $G$. The class equation for the present case can now be written as

$$|G| = |\mathrm{C}(G)| + \sum_{a\in X}[G : \mathrm{C}_G(a)].$$

**Theorem 154.** *All groups of order $p^2$ are abelian.*

[2]More generally, if $H < G$, then $\mathrm{C}_H(g) = \{h \in H : hgh^{-1} = g\}$.

*Proof.* Let $G$ have order $p^2$. In particular, $G$ is a $p$-group. By Theorem 153, either $\mathrm{C}(G) = G$, in which case $G$ is abelian, or else $|\mathrm{C}(G)| = p$. In the latter case, let $a \in G \smallsetminus \mathrm{C}_G(a)$. Then

$$G = \mathrm{C}(G)\langle a \rangle.$$

But elements of $\mathrm{C}(G)$ commute with all elements of $G$; and $a$ commutes with itself. If the generators commute with one another, the whole group is abelian. Therefore $G$ must be abelian. □

**Porism 154.1.** *Every nontrivial $p$-group has nontrivial center.*

### 5.3.2. Normalizers

If $H < G$, let $G$ act on the set of conjugates of $H$ by conjugation. The stabilizer of $H$ under this action is called the **normalizer** of $H$ in $G$ and is denoted by[3]

$$\mathrm{N}_G(H).$$

Explanation of the name is given by the following.

**Theorem 155.** *If $H < K < G$, then*

$$H \lhd K \iff K < \mathrm{N}_G(H).$$

We establish some technical results for the sake of proving the Sylow Theorems of the next subsection.

**Lemma 11.** *Suppose $H < G$, and let $H$ act on $G/H$ by left multiplication. Then*

$$(G/H)_0 = \mathrm{N}_G(H)/H.$$

*Proof.* Supposing $g \in G$, we have $gH \in (G/H)_0$ if and only if, for all $h$ in $H$,

$$\begin{aligned} hgH &= gH, \\ g^{-1}hgH &= H, \\ g^{-1}hg &\in H. \end{aligned}$$

[3]More generally, if also $K < G$, then $\mathrm{N}_K(H) = \{k \in K \colon kHk^{-1} = H\}$.

Thus

$$\begin{aligned} gH \in (G/H)_0 &\iff g^{-1}Hg = H \\ &\iff g^{-1} \in \mathrm{N}_G(H) \\ &\iff g \in \mathrm{N}_G(H) \\ &\iff gH \in \mathrm{N}_G(H)/H. \end{aligned}$$ □

A $p$**-subgroup** of a group is a subgroup that is a $p$-group. Every group has at least one $p$-subgroup, namely the trivial subgroup $\{\mathrm{e}\}$.

**Lemma 12.** *If $H$ is a $p$-subgroup of $G$, then*

$$[G : H] \equiv [\mathrm{N}_G(H) : H] \pmod{p}.$$

*Proof.* Theorem 153 and the last lemma. □

**Lemma 13.** *If $H$ is a $p$-subgroup of $G$, and $p$ divides $[G : H]$, then for some subgroup $K$ of $G$,*

$$H \lhd K, \qquad [K : H] = p.$$

*Proof.* By the last lemma, $p$ divides $[\mathrm{N}_G(H) : H]$. Since $H \lhd \mathrm{N}_G(H)$, the quotient $\mathrm{N}_G(H)/H$ is a group. By Cauchy's Theorem (Theorem 148), this group has an element $gH$ of order $p$. Then $H\langle g\rangle$ is the desired group $K$. □

Now can start proving the Sylow Theorems.

### 5.3.3. Sylow subgroups

A **Sylow** $p$**-subgroup** of a group is a maximal $p$-subgroup. Then every $p$-subgroup of a finite group $G$ is a subgroup of a Sylow $p$-subgroup of $G$.[4] In particular, since $G$ does have the $p$-subgroup $\{\mathrm{e}\}$, it has at least one Sylow $p$-subgroup. We now establish that the order of every Sylow $p$-subgroup of a finite group is as large as Lagrange's Theorem (page 108) allows it to be.

[4]The same is true for infinite groups $G$, by the version of the Axiom of Choice known as Zorn's Lemma; but we shall not make use of this result.

**Theorem 156** (Sylow I). *If $G$ is a finite group of order $p^n m$, where $\gcd(p, m) = 1$, then every Sylow $p$-subgroup of $G$ has order $p^n$.*

*Proof.* Use the last lemma repeatedly. □

**Porism 156.1.** *If $|G| = p^n m$, where $p \nmid m$, then there is a chain*

$$H_0 < H_1 < \cdots < H_n$$

*of $p$-subgroups of $G$, where*

$$H_0 = \{\mathrm{e}\}, \qquad H_i \lhd H_{i+1}, \qquad [H_{i+1} : H_i] = p.$$

*In particular, $H_n$ is a Sylow $p$-subgroup of $G$. Every $p$-subgroup of $G$ appears on such a chain.*

In the notation of the porism, although $H_i \lhd H_{i+1}$ and $H_{i+1} \lhd H_{i+2}$, we need not have $H_i \lhd H_{i+2}$. For a counterexample, consider $\mathrm{Dih}(4)$:

$$\langle(1\ 3)\rangle \lhd \langle(1\ 3), (0\ 2)\rangle, \qquad \langle(1\ 3), (0\ 2)\rangle \lhd \mathrm{Dih}(4),$$

but $\langle(1\ 3)\rangle \ntriangleleft \mathrm{Dih}(4)$ since

$$(0\ 1\ 2\ 3) \in \mathrm{Dih}(4), \qquad (3\ 2\ 1\ 0)(1\ 3)(0\ 1\ 2\ 3) = (0\ 2).$$

The following is as close as can be to a converse of Lagrange's Theorem.

**Corollary 156.1.** *Suppose $G$ is a finite group. Then $G$ has a subgroup of every order that divides $|G|$, provided that order is a prime power.*

The converse of the first part of the following will be the Second Sylow Theorem.

**Corollary 156.2.** *Every conjugate of every Sylow $p$-subgroup of a finite group is also a Sylow $p$-subgroup. Thus if a finite group has a unique Sylow $p$-subgroup, this must be a normal subgroup.*

To prove the Second Sylow Theorem, we shall use a generalization of Lemma 11.

**Lemma 14.** *Suppose $G$ is a group with subgroups $H$ and $K$. Under the action of $H$ on $G/K$ by left multiplication,*

$$gK \in (G/K)_0 \Leftrightarrow H < gKg^{-1}.$$

*Proof.* The first part of the proof of Lemma 11 shows this. Indeed, for all $g$ in $G$, we have $gK \in (G/K)_0$ if and only if, for all $h$ in $H$,

$$\begin{aligned} hgK &= gK, \\ g^{-1}hgK &= K, \\ g^{-1}hg &\in K, \\ h &\in gKg^{-1}. \end{aligned}$$

□

**Theorem 157** (Sylow II)**.** *All Sylow $p$-subgroups of finite groups are conjugate.*

*Proof.* Say $H$ and $K$ are Sylow $p$-subgroups of $G$. Then $H$ acts on the set $G/K$ by left multiplication. By Theorem 153, since $[G : K]$ is not a multiple of $p$, the set $(G/K)_0$ has an element $aK$. By the lemma, $H < aKa^{-1}$. Then $H = aKa^{-1}$ by the First Sylow Theorem. □

**Theorem 158** (Sylow III)**.** *If $|G| = p^n m$, where $\gcd(p, m) = 1$, and $A$ is the set of Sylow $p$-subgroups of $G$, then*

$$|A| \equiv 1 \pmod{p}, \qquad |A| \text{ divides } m.$$

*Proof.* $G$ acts on $A$ by conjugation, by the First Sylow Theorem (more precisely, Corollary 156.2). Let $H \in A$. By the Second Sylow Theorem, the orbit of $H$ is just $A$. The stabilizer of $H$ is $\mathrm{N}_G(H)$. Since by Theorem 152 the index of the stabilizer is the size of the orbit, we have

$$[G : \mathrm{N}_G(H)] = |A|,$$

and so $|A|$ divides $|G|$. Now suppose also $K \in A$. Then $K$ must be the unique Sylow $p$-subgroup of $\mathrm{N}_G(K)$. Considering $H$ as acting on $A$ by conjugation, we have

$$\begin{aligned} K \in A_0 &\iff H < \mathrm{N}_G(K) \\ &\iff H = K. \end{aligned}$$

Therefore $A_0 = \{H\}$, so by Theorem 153,

$$|A| \equiv 1 \pmod{p}.$$

It now follows that $|A|$ divides $m$. □

## 5.4. *Classification of small groups

We can now complete the work, begun in §5.1 (page 144), of classifying the groups of order $pq$ for primes $p$ and $q$.

**Theorem 159.** *Suppose $p$ and $q$ are distinct primes, with $q < p$, and $G$ is a group of order $pq$. Either*

$$G \cong \mathbb{Z}_p \times \mathbb{Z}_q,$$

*which is cyclic, or else $p \equiv 1 \pmod{q}$ and*

$$G \cong \mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q$$

*for some embedding $\sigma$ of $\mathbb{Z}_q$ in $\mathrm{Aut}(\mathbb{Z}_p)$. In particular, if $q = 2$, then*

$$G \cong \mathrm{Dih}(p).$$

*Proof.* By Cauchy's Theorem, $G$ has elements $a$ and $b$, of orders $p$ and $q$ respectively. Then $\langle a \rangle$ and $\langle b \rangle$ are Sylow subgroups of $G$. Let $A$ be the set of Sylow $p$-subgroups of $G$. By the Third Sylow Theorem, $|A|$ divides $q$. Since $p \nmid q - 1$, we must have $|A| = 1$. Thus $\langle a \rangle$ is the unique Sylow $p$-sugroup of $G$, and so it is a normal subgroup. By Theorems 149 and 111 (pages 148 and 112), $G$ is the semidirect product of $\langle a \rangle$ and $\langle b \rangle$. If it is not actually a direct product, then $\langle b \rangle$ must not be a normal subgroup of $G$, and so $q$ does divide $p - 1$, and the rest follows. □

We now know all groups of order less than 36, but different from 8, 12, 16, 18, 20, 24, 27, 28, 30, and 32.

**Theorem 160.** *Every group of order 8 is isomorphic to one of*

$$\mathbb{Z}_8, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_4, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2, \qquad \mathrm{Dih}(4), \qquad \mathrm{Q}_8.$$

*Proof.* Say $|G| = 8$. If $G$ is abelian, then its possibilities are given by the classification of finitely generated abelian groups (Theorem 137, page 139). Suppose $G$ is not abelian. Then $G$ has an element $a$ of order greater than 2 by Theorem 64 (page 72), and so $|a| = 4$ (since $G \ncong \mathbb{Z}_8$). Then $\langle a\rangle \lhd G$ by Theorem 109 (page 111). Let $b \in G \smallsetminus \langle a\rangle$. Then $b^2$ is either e or $a^2$ (since otherwise $b$ would generate $G$). In the former case, $G = \langle a\rangle \rtimes \langle b\rangle$, so $G \cong \mathrm{Dih}(4)$. In the latter case, $G \cong \mathrm{Q}_8$. □

**Theorem 161.** *The subgroup of* $\mathrm{Sym}(3) \times \mathbb{Z}_4$ *generated by the two elements*

$$\big((0\ 1\ 2), 2\big), \qquad\qquad \big((0\ 1), 1\big)$$

*has order* 12 *and has the presentation*

$$\langle a, b \mid a^6, a^3b^2, bab^{-1}a\rangle.$$

**Lemma 15.** *If* $H < G$*, and* $\sigma$ *is the homomorphism* $g \mapsto (xH \mapsto gxH)$ *from* $G$ *to* $\mathrm{Sym}(G/H)$*, then*

$$\ker(\sigma) < H.$$

**Theorem 162.** *Every group of order* 12 *is isomorphic to one of*

$$\mathbb{Z}_{12}, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_6, \qquad \mathrm{Alt}(4), \qquad \mathrm{Dih}(6), \qquad \langle a, b \mid a^6, a^3b^2, bab^{-1}a\rangle.$$

*Proof.* Suppose $|G| = 12$. By Cauchy's Theorem, $G$ has an element $c$ of order 3. Then $G$ acts on $G/\langle c\rangle$ by left multiplication, which gives us a homomorphism from $G$ to $\mathrm{Sym}(G/\langle c\rangle)$. Since $[G : \langle c\rangle] = 4$, there is a homomorphism from $G$ to $\mathrm{Sym}(4)$. If this is an embedding, then $G \cong \mathrm{Alt}(4)$ by Theorem 120 (page 119). Otherwise, by the lemma, the kernel of the homorphism must be $\langle c\rangle$. In this case,

$$\langle c\rangle \lhd G.$$

Now let $H$ be a Sylow 2-subgroup of $G$. Having order $2^2$, it is abelian (Theorem 154, page 151). If $G$ is not abelian, then the action of $H$ on $\langle c\rangle$ by conjugation must be nontrivial. But since $|\mathrm{Aut}(\langle c\rangle)| = 6$, which is indivisible by the order of $H$, there must be some $d$ in $H$ that commutes with $c$. Then $\langle c, d\rangle \cong \mathbb{Z}_6$. Let $a = cd$, so $\langle a\rangle = \langle c, d\rangle$. Let $b \in G \smallsetminus \langle a\rangle$, so

$$G = \langle a, b\rangle.$$

If $|b| = 2$, then $G \cong \mathrm{Dih}(6)$. In any case, conjugation by $b$ is a nontrivial automorphism of $\langle a \rangle$, and in particular $bab^{-1}$ is a generator of $\langle a \rangle$ different from $a$. There is only one of these, namely $a^{-1}$, so

$$bab^{-1} = a^{-1}. \tag{5.4}$$

Also $b^2 = a^k$ for some $k$ in $\mathbb{Z}_6$. If $k = \pm 1$, then $G = \langle b \rangle$. Suppose $k = \pm 2$. Then $|b| = 6$, so $\langle b \rangle \lhd G$, and therefore

$$ab^{-1}a^{-1} = b. \tag{5.5}$$

From (5.4) we have

$$ab^{-1} = b^{-1}a^{-1}, \qquad ba = a^{-1}b. \tag{5.6}$$

From (5.5) we have $ab^{-1} = ba$, so all members of the equations in (5.6) are equal to one another. In particular,

$$ab^{-1} = a^{-1}b, \qquad ba = b^{-1}a^{-1},$$

which yield $a^2 = b^2$ and $b^2 = a^{-2}$ respectively, contradicting that $|a| = 6$. The only remaining possibility is $k = 3$, which yields the last group listed. □

## 5.5. Nilpotent groups

For a group, what is the next best thing to being abelian? A group $G$ is abelian if and only if $\mathrm{C}(G) = G$. To weaken this condition, we define the **commutator** of two elements $a$ and $b$ of $G$ to be

$$aba^{-1}b^{-1};$$

this can be denoted by

$$[a, b].$$

Then

$$\mathrm{C}(G) = \bigl\{g \in G \colon \forall x\ (x \in G \Rightarrow [g, x] = \mathrm{e})\bigr\}.$$

We now generalize this by defining

$$\mathrm{C}_0(G) = \{\mathrm{e}\}, \quad \mathrm{C}_{n+1}(G) = \Bigl\{g \in G \colon \forall x\ \bigl(x \in G \Rightarrow [g, x] \in \mathrm{C}_n(G)\bigr)\Bigr\}.$$

Then $\mathrm{C}(G) = \mathrm{C}_1(G)$. Also,

$$\mathrm{C}_n(G) = \left\{g \in G \colon \forall \boldsymbol{x} \left(\boldsymbol{x} \in G^n \Rightarrow \Big[\big[\dots[[g, x_0], x_1], \cdots\big], x_{n-1}\Big] = \mathrm{e}\right)\right\}.$$

The following general result will now be useful.

**Theorem 163.** *Suppose $N \lhd G$. Every subgroup $H$ of $G/N$ is of the form $K/N$ for some subgroup $K$ of $G$ of which $N$ is a normal subgroup. Moreover,*

$$K/N \lhd G/N \iff K \lhd G.$$

**Theorem 164.** *For all groups $G$, for all $n$ in $\omega$,*

$$\mathrm{C}_n(G) \lhd G, \tag{5.7}$$

$$\mathrm{C}_n(G) < \mathrm{C}_{n+1}(G), \tag{5.8}$$

$$\mathrm{C}_{n+1}(G)/\mathrm{C}_n(G) = \mathrm{C}(G/\mathrm{C}_n(G)). \tag{5.9}$$

*Proof.* We use induction. Trivially, (5.7) holds when $n = 0$. Suppose it holds when $n = k$. Then the following are equivalent:

$$g \in \mathrm{C}_{k+1}(G),$$
$$\forall x \ \big(x \in G \Rightarrow [g, x] \in \mathrm{C}_k(G)\big),$$
$$\forall x \ \big(x \in G \Rightarrow [g, x]\,\mathrm{C}_k(G) = \mathrm{C}_k(G)\big),$$
$$\forall x \ \big(x \in G \Rightarrow [g\,\mathrm{C}_k(G), x\,\mathrm{C}_k(G)] = \mathrm{C}_k(G)\big),$$
$$g\,\mathrm{C}_k(G) \in \mathrm{C}(G/\mathrm{C}_k(G)).$$

Thus (5.8) and (5.9) hold when $n = k$. In particular,

$$\mathrm{C}_{k+1}(G)/\mathrm{C}_k(G) \lhd G/\mathrm{C}_k(G),$$

and so, by the last theorem, (5.7) holds when $n = k + 1$. □

The sequence $(\mathrm{C}_n(G) \colon n \in \omega)$ may be written out as

$$\{\mathrm{e}\} \lhd \mathrm{C}(G) \lhd \mathrm{C}_2(G) \lhd \mathrm{C}_3(G) \lhd \cdots$$

although strictly this expression is not a noun, but the conjunction of the statements $\{\mathrm{e}\} \lhd \mathrm{C}(G)$, $\mathrm{C}(G) \lhd \mathrm{C}_2(G)$, $\mathrm{C}_2(G) \lhd \mathrm{C}_3(G)$, and so on. By the last theorem (and Theorem 110 on page 111), the relation $\lhd$ on the set $\{\mathrm{C}_n(G) \colon n \in \omega\}$ is indeed transitive. A group is called **nilpotent** if for some $n$ in $\omega$,

$$\mathrm{C}_n(G) = G.$$

So an abelian group is nilpotent, since its center is itself.[5] Other examples of nilpotent groups are given by:

**Theorem 165.** *Finite p-groups are nilpotent.*

*Proof.* If $G$ is a $p$-group and $\mathrm{C}_k(G) \subsetneqq G$, then $G/\mathrm{C}_k(G)$ is a nontrivial $p$-group, so by Porism 154.1 it has a nontrivial center. By Theorem 164 then, $\mathrm{C}_k(G) \subsetneqq \mathrm{C}_{k+1}(G)$. □

The converse fails, because of:

**Theorem 166.** *The direct product of a finite family of nilpotent groups is nilpotent.*

*Proof.* Use Theorem 136 (page 139) and

$$\mathrm{C}(G \times H) = \mathrm{C}(G) \times \mathrm{C}(H).$$

If $\mathrm{C}_n(G) = G$ and $\mathrm{C}_m(H) = H$, then $\mathrm{C}_{\max\{n,m\}}(G \times H) = G \times H$. □

Thus, if all Sylow subgroups of a finite group $G$ are *normal* subgroups, then $G$ must be nilpotent. We now proceed to a partial converse of this result. Given that $G$ is a finite nilpotent group with a Sylow $p$-subgroup $P$ for some prime $p$, we want to show $P \lhd G$, that is, $\mathrm{N}_G(P) = G$.

**Lemma 16.** *If $G$ is a finite group with Sylow p-subgroup $P$, then*

$$\mathrm{N}_G(\mathrm{N}_G(P)) = \mathrm{N}_G(P).$$

[5] Apparently the term *nilpotent* arises for the following reason. If $\mathrm{C}_n(G) = G$ and, for some $g$ in $G$, $f$ is the element $x \mapsto [g, x]$ of the monoid $(G^G, \mathrm{id}_G, \circ)$, then $f^n$ is the constant function $x \mapsto \mathrm{e}$.

*Proof.* Let $N = \mathrm{N}_G(P)$. Supppose $g \in \mathrm{N}_G(N)$, that is,

$$gNg^{-1} = N.$$

Since $P < N$, we have also $gPg^{-1} < N$. But $P \lhd N$, so $P$ is the unique Sylow $p$-subgroup of $N$. Since $gPg^{-1}$ is also a Sylow $p$-subgroup of $N$, we must have $gPg^{-1} = P$. Thus

$$g \in N.$$

We have now proved $\mathrm{N}_G(N) < N$. □

Now, in the notation of the lemma, we want to show that, if $N \lneqq G$, then either $N \lneqq \mathrm{N}_G(N)$, or else $G$ is not finite and nilpotent. We shall use the following.

**Lemma 17.** *If* $\mathrm{C}_n(G) < H$, *then* $\mathrm{C}_{n+1}(G) < \mathrm{N}_G(H)$.

*Proof.* Say $g \in \mathrm{C}_{n+1}(G)$; we show $gHg^{-1} \subseteq H$. But if $h \in H$, then $[g,h] \in \mathrm{C}_n(G)$, so $ghg^{-1} \in \mathrm{C}_n(G)h \subseteq H$. Therefore $gHg^{-1} \subseteq H$. □

**Lemma 18.** *If $G$ is nilpotent, and $H \lneqq G$, then $H \lneqq \mathrm{N}_G(H)$.*

*Proof.* Let $n$ be maximal such that $\mathrm{C}_n(G) < H$. Then $\mathrm{C}_{n+1}(G) \smallsetminus H$ is non-empty, but, by the last lemma, it contains members of $\mathrm{N}_G(H)$. □

**Theorem 167.** *A finite nilpotent group is the direct product of its Sylow subgroups.*

*Proof.* Suppose $G$ is a finite nilpotent group. By Lemmas 16 and 18, every Sylow subgroup of $G$ is a normal subgroup. Suppose the Sylow subgroups of $G$ compose a list $(P_i : i < n)$, where each $P_i$ is a $p_i$-group, and $p_i \neq p_j$ when $i \neq j$. If, for some $i$ in $n$, the product $P_0 \cdots P_{i-1}$ is an internal direct product, then its order is indivisible by $p_i$, and so $P_0 \cdots P_{i-1} \cap P_i = \{\mathrm{e}\}$. Hence, by Theorem 124 (page 125) and induction, each product $P_0 \cdots P_i$ is an internal direct product. Then also the order of $P_0 \cdots P_{n-1}$ is the order of $G$, so the two groups are the same. □

Theorems 165, 166, and 167 give us a classification of the finite nilpotent groups.

## 5.6. Soluble groups

Having defined the commutator of two elements of a group, we define the **commutator subgroup** of a group $G$ to be the subgroup

$$\langle [x,y] \colon (x,y) \in G^2 \rangle$$

generated by the commutators of all pairs of elements of $G$. We denote this subgroup by

$$G'.$$

Its interest arises from the following.

**Theorem 168.** *$G'$ is the smallest of the normal subgroups $N$ of $G$ such that $G/N$ is abelian.*

*Proof.* If $f$ is a homomorphism defined on $G$, then

$$f([x,y]) = [f(x), f(y)].$$

Thus, if $f \in \mathrm{Aut}(G)$, then $f(G') < G'$. In particular, $xG'x^{-1} < G'$ for all $x$ in $G$; so $G' \lhd G$. Suppose $N \lhd G$; then the following are equivalent.

1. $G/N$ is abelian.
2. $N = [x,y]N$ for all $(x,y)$ in $G^2$.
3. $G' < N$. □

We now define the **derived subgroups** $G^{(n)}$ of $G$ by

$$G^{(0)} = G, \qquad G^{(n+1)} = (G^{(n)})'.$$

We have a descending sequence

$$G \rhd G' \rhd G^{(2)} \rhd \cdots$$

The group $G$ is called **soluble** or **solvable** if this sequence reaches $\{e\}$ (after finitely many steps).[6] Immediately, abelian groups are soluble. For

[6] If $f$ is a polynomial in one variable over $\mathbb{Q}$, let $A$ be the set of its zeros in the field $\mathbb{C}$, and let $G = \{\sigma \restriction A \colon \sigma \in \mathrm{Aut}(\mathbb{C})\}$. Then $G < \mathrm{Sym}(A)$, and $G$ is soluble if and only if the elements of $A$ can be obtained from $\mathbb{Q}$ by the field operations and taking $n$th roots for arbitrary $n$ in $\mathbb{N}$.

more examples, let $K$ be a field, and if $n \in \mathbb{N}$, let $G$ be the subgroup of $\mathrm{GL}_n(K)$ consisting of **upper triangular matrices.** So $G$ comprises the matrices

$$\begin{pmatrix} a_0 & & * \\ & \ddots & \\ 0 & & a_{n-1} \end{pmatrix}$$

where $a_0 \cdots a_{n-1} \neq 0$. We have

$$\begin{pmatrix} a_0 & & * \\ & \ddots & \\ 0 & & a_{n-1} \end{pmatrix} \begin{pmatrix} b_0 & & * \\ & \ddots & \\ 0 & & b_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 b_0 & & * \\ & \ddots & \\ 0 & & a_{n-1} b_{n-1} \end{pmatrix}$$

and therefore every element of $G'$ is **unitriangular,** that is, it takes the form of

$$\begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

We also have

$$\begin{pmatrix} 1 & a_1 & & * \\ & 1 & \ddots & \\ & & \ddots & a_{n-1} \\ 0 & & & 1 \end{pmatrix} \begin{pmatrix} 1 & b_1 & & * \\ & 1 & \ddots & \\ & & \ddots & b_{n-1} \\ 0 & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & c_1 & & * \\ & 1 & \ddots & \\ & & \ddots & c_{n-1} \\ 0 & & & 1 \end{pmatrix},$$

where $c_i = a_i + b_i$ in each case, so the elements of $G''$ take the form of

$$\begin{pmatrix} 1 & 0 & & * \\ & 1 & \ddots & \\ & & \ddots & 0 \\ 0 & & & 1 \end{pmatrix}.$$

Proceeding, we find $G^{(n+1)} = \{\mathrm{e}\}$.

**Theorem 169.** *Nilpotent groups are soluble.*

*Proof.* Each quotient $\mathrm{C}_{k+1}(G)/\mathrm{C}_k(G)$ is the center of some group, namely $G/\mathrm{C}_k(G)$, so it is abelian. By Theorem 168 then,

$$\mathrm{C}_{k+1}(G)' < \mathrm{C}_k(G).$$

Suppose $G$ is nilpotent, so that $G = \mathrm{C}_n(G)$ for some $n$ in $\omega$. Then

$$G^{(0)} < \mathrm{C}_n(G).$$

If $G^{(k)} < \mathrm{C}_{n-k}(G)$, then

$$G^{(k+1)} < (\mathrm{C}_{n-k}(G))' < \mathrm{C}_{n-k-1}(G).$$

By induction, $G^{(n)} < \mathrm{C}_0(G) = \{\mathrm{e}\}$. □

The foregoing argument might be summarized in the following commutative diagram, which is built up from left to right, the arrows being inclusions:

$$\begin{tikzcd}
G & G' \arrow[l] & G^{(2)} \arrow[l] & G^{(3)} \arrow[l] & G^{(n)} \arrow[l, dotted] \\
G \arrow[u, leftarrow] & \mathrm{C}_n(G)' \arrow[l] \arrow[u, leftarrow] & \mathrm{C}_{n-1}(G)' \arrow[l] \arrow[u, leftarrow] & \mathrm{C}_{n-2}(G)' \arrow[l] \arrow[u, leftarrow] & \mathrm{C}(G)' \arrow[l, dotted] \arrow[u, leftarrow] \\
\mathrm{C}_n(G) \arrow[u, leftarrow] & \mathrm{C}_{n-1}(G) \arrow[l] \arrow[u, leftarrow] & \mathrm{C}_{n-2}(G) \arrow[l] \arrow[u, leftarrow] & \mathrm{C}_{n-3}(G) \arrow[l] \arrow[u, leftarrow] & \{\mathrm{e}\} \arrow[l, dotted] \arrow[u, leftarrow]
\end{tikzcd}$$

Since $\mathrm{Sym}(3)/\mathrm{Alt}(3)$ is abelian, we have

$$\mathrm{Sym}(3)' < \mathrm{Alt}(3), \qquad \mathrm{Sym}(3)'' < \mathrm{Alt}(3)' = \{\mathrm{e}\},$$

so $\mathrm{Sym}(3)$ is soluble. However,

$$\mathrm{Sym}(3) = \mathrm{Alt}(3) \rtimes \langle (0\ 1) \rangle,$$

the semidirect product of its Sylow subgroups; but the product is not *direct,* so $\mathrm{Sym}(3)$ is not nilpotent.

**Theorem 170.** *Let $H < G$ and $N \lhd G$.*

1. *If $G$ is soluble, then so are $H$ and $G/N$.*

2. *If $N$ and $G/N$ are soluble, then so is $G$.*

*Proof.* 1. $H^{(k)} < G^{(k)}$ and $(G/N)^{(k)} = G^{(k)}N/N$.

2. If $G/N$ is soluble, then $G^{(n)} < N$ for some $n$. If also $N$ is soluble, then $N^{(m)} = \{\mathrm{e}\}$ for some $m$, so $G^{(n+m)} < N^{(m)} = \{\mathrm{e}\}$. □

**Theorem 171.** *Groups with non-abelian simple subgroups are not soluble.*

*Proof.* Suppose $H$ is simple. Since $H' \lhd H$, we have either $H' = \{\mathrm{e}\}$ or $H' = H$. In the former case, $H$ is abelian; in the latter, $H$ is insoluble. □

In particular, Sym(5) is not soluble if $n \geqslant 5$.[7]

## 5.7. Normal series

A **normal series** for a group $G$ is a list $(G_0, \dots, G_n)$ of subgroups, where

$$G = G_0 \rhd G_1 \rhd \dots \rhd G_n = \{\mathrm{e}\}.$$

We do not require $G_k \rhd G_{k+2}$.[8] The quotients $G_k/G_{k+1}$ are called the **factors** of the normal series. The series is called

1) a **composition series,** if the factors are simple;
2) a **soluble series,** if the factors are abelian.

**Theorem 172.** *A group is soluble if and only if it has a soluble series.*

[7] This is why the general 5th-degree polynomial equation is insoluble by radicals.

[8] One may call a normal series a *subnormal series,* reserving the term *normal series* for the case where $G \rhd G_k$ for each $k$. However, we shall not be interested in the distinction recognized by this terminology.

*Proof.* If $G^{(n)} = \{\mathrm{e}\}$, then $(G^{(0)}, \dots, G^{(n)})$ is a soluble series for $G$, by Theorem 168. Suppose conversely $(G_0, \dots, G_n)$ is a soluble series for $G$. Again by Theorem 168, we have ${G_k}' < G_{k+1}$ for each $k$ in $n$. Since also $H < K$ implies $H' < K'$, we have

$$\begin{gathered} G' < G_1, \\ G'' < {G_1}' < G_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ G^{(n)} < {G_1}^{(n-1)} < \dots < {G_{n-1}}' < G_n = \{\mathrm{e}\}. \end{gathered}$$ □

Since not every finite group is soluble, not every finite group has a soluble series. However:

**Theorem 173.** *Every finite group has a composition series.*

*Proof.* Trivially $(\{\mathrm{e}\})$ is a composition series. Every nontrivial finite group $G$ has at least one proper normal subgroup, namely $\{\mathrm{e}\}$. Being finite, $G$ has only finitely many normal subgroups. Therefore $G$ has a maximal proper normal subgroup, $G^*$ (which need not be unique). Then $G/G^*$ is simple, by Theorem 163 (page 159): every normal subgroup of $G/G^*$ is $K/G^*$ for some normal subgroup $K$ of $G$ such that $G^* < K$, and therefore $K$ is either $G^*$ or $G$, so $K/G^*$ is either $\{\mathrm{e}\}$ or $G/G^*$.

Now let $G_0 = G$, and let $G_{k+1} = {G_k}^*$ unless $G_k = \{\mathrm{e}\}$. Since $G$ is finite and $G_k \gneqq G_{k+1}$, we must have $G_n = \{\mathrm{e}\}$ for some $n$. Then $(G_0, \dots, G_n)$ is the desired composition series. □

Two normal series are **equivalent** if they have the same *multiset* of (isomorphism classes of) nontrivial factors. A **multiset** is a set in which repetitions of members are allowed. For a formal definition, we can say a multiset is a pair $(A, f)$, where $A$ is a set and $f \colon A \to \mathbb{N}$. For example, the two series

$$(\mathbb{Z}_{60}, \langle 2\rangle, \langle 6\rangle, \langle 12\rangle, \{\mathrm{e}\}), \qquad (\mathbb{Z}_{60}, \langle 3\rangle, \langle 15\rangle, \langle 15\rangle, \langle 30\rangle, \{\mathrm{e}\})$$

are equivalent, because the factors of the first are isomorphic to $\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_2$, and $\mathbb{Z}_5$ respectively, and the factors of the second are isomorphic $\mathbb{Z}_3$,

$\mathbb{Z}_5$, $\{e\}$, $\mathbb{Z}_2$, and $\mathbb{Z}_2$ respectively, so each series has the same multiset of factors, namely

$$\{\mathbb{Z}_2, \mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5\}.$$

These series are *not* equivalent to $(\mathbb{Z}_{30}, \langle 2\rangle, \langle 6\rangle, \{e\})$, whose factors are $\mathbb{Z}_2$, $\mathbb{Z}_3$, and $\mathbb{Z}_5$.

If, from a normal series for a group, another normal series for the group can be obtained by deleting some terms, then the former series is a **refinement** of the latter. So the series $(\mathbb{Z}_{60}, \langle 2\rangle, \langle 4\rangle, \langle 12\rangle, \{e\})$ is a refinement of $(\mathbb{Z}_{60}, \langle 4\rangle, \langle 12\rangle, \{e\})$. Every normal series is a refinement of a normal series with no trivial factors, and these two series are equivalent. Among normal series with no trivial factors, composition series are *maximal* in that they have no proper refinements. If

$$\begin{aligned}
G = G_0(0) \rhd G_0(1) \rhd G_0(2) \rhd \cdots \rhd G_0(n_0) = \{e\},\\
G = G_1(0) \rhd G_1(1) \rhd G_1(2) \rhd \cdots \rhd G_1(n_1) = \{e\},
\end{aligned}$$

and the two normal series are equivalent and have no trivial factors, this means $n_0 = n_1$, and there is $\sigma$ in $\mathrm{Sym}(n_0)$ such that

$$G_0(i)/G_0(i+1) \cong G_1(\sigma(i))/G_1(\sigma(i)+1)$$

for each $i$ in $n_0$.

**Theorem 174.** *A soluble series for a finite group has a refinement in which the nontrivial factors are cyclic of prime order.*

We now aim to prove Theorem 176 below. The proof will use the following, which is known as the Butterfly Lemma, because the groups that it involves form the commutative diagram in Figure 5.1 (in which arrows are inclusions).

**Lemma 19** (Zassenhaus)**.** *For a group $G$, suppose*

$$N_0 \lhd H_0 < G, \qquad N_1 \lhd H_1 < G,$$

*and let*

$$K = (H_0 \cap N_1)(H_1 \cap N_0), \qquad H = H_0 \cap H_1.$$

$$\begin{tikzcd}
 & H_0 & & H_1 & \\
 & N_0H \arrow[u] & & N_1H \arrow[u] & \\
 & & H \arrow[ul] \arrow[ur] & & \\
 & N_0(H_0 \cap N_1) \arrow[uu] & & N_1(H_1 \cap N_0) \arrow[uu] & \\
N_0 \arrow[ur] & & K \arrow[ul] \arrow[uu] \arrow[ur] & & N_1 \arrow[ul] \\
 & H_1 \cap N_0 \arrow[ul] \arrow[ur] & & H_0 \cap N_1 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Figure 5.1. The Butterfly Lemma

*Then*

$$K \lhd H,$$

*and for each $i$ in $2$, there is a well-defined epimorphism*

$$nh \mapsto Kh$$

*from $N_iH$ to $H/K$ with kernel $N_i(H_i \cap N_{1-i})$. Hence:*

1) *$N_i(H_i \cap N_{1-i}) \lhd N_iH$ for each $i$ in $2$, and*
2) *the two groups $N_iH/N_i(H_i \cap N_{1-i})$ are isomorphic to one another.*

*Proof.* For each $i$ in $2$, we have $H_i \cap N_{1-i} \lhd H$ by Theorem 110 (page 111). Hence $K \lhd H$. If $n, n' \in N_0$ and $h, h' \in H$ and $nh' = n'h$, then

$$h'h^{-1} = n^{-1}n',$$

which is in $N_0 \cap H$ and hence in $K$, so that $Kh = Kh'$. Thus $nh \mapsto Kh$ (where $n \in N_0$ and $h \in H$) is indeed a well-defined homomorphism $f$ from $N_0H$ into $H/K$. It is clear that $f$ is surjective.

Now let $n \in N_0$ and $h \in H$, and suppose $nh \in \ker(f)$, that is,

$$h \in K.$$

Then $h = n_0n_1$ for some $n_0$ in $H_1 \cap N_0$ and $n_1$ in $H_0 \cap N_1$. Hence $nh = nn_0n_1$, which is in $N_0(H_0 \cap N_1)$. Thus

$$nh \in N_0(H_0 \cap N_1).$$

Conversely, suppose this last condition holds. Since $h = n^{-1}nh$, we now have also

$$h \in N_0(H_0 \cap N_1).$$

so $h = n'h'$ for some $n'$ in $N_0$ and some $h'$ in $H_0 \cap N_1$. Then $n' = h(h')^{-1}$, which is in $H(H_0 \cap N_1)$; but this is a subgroup of $H_1$. So $n' \in N_0 \cap H_1$, and therefore $n'h'$, which is $h$, is in $K$, and so $nh \in \ker(f)$. Thus $\ker(f) = N_0(H_0 \cap N_1)$. □

**Theorem 175** (Schreier)**.** *Any two normal series have equivalent refinements.*

*Proof.* Suppose

$$G = G_i(0) \rhd G_i(1) \rhd \cdots \rhd G_i(n_i) = \{\mathrm{e}\},$$

where $i < 2$. In particular then,

$$G_0(j+1) \lhd G_0(j) < G, \qquad G_1(k+1) \lhd G_1(k) < G.$$

Define

$$\begin{aligned} G_0(j,k) &= G_0(j+1) \cdot \big(G_0(j) \cap G_1(k)\big), \\ G_1(j,k) &= G_1(k+1) \cdot \big(G_0(j) \cap G_1(k)\big), \end{aligned}$$

where $(j,k) \in n_0 \times n_1$. Then by the Butterfly Lemma

$$\begin{aligned} G_0(j) = G_0(j,0) \rhd \cdots \rhd G_0(j,n_1) = G_0(j+1), \\ G_1(k) = G_1(0,k) \rhd \cdots \rhd G_1(n_0,k) = G_1(k+1), \end{aligned}$$

giving us normal series that are refinements of the original ones, and also

$$G_0(j,k)/G_0(j,k+1) \cong G_1(j,k)/G_1(j+1,k),$$

so that the two refinements are equivalent. □

**Theorem 176** (Jordan–Hölder)**.** *Any two composition series of a group are equivalent.*

*Proof.* By Schreier's Theorem, any two composition series of a group have equivalent refinements; but every refinement of a composition series is already equivalent to that series. □

Combining this with Theorem 173, we have that every finite group determines a multiset of finite simple groups, and these are just the nontrivial factors of any composition series of the group. Hence arises the interest in the classification of the finite simple groups: it is like studying the prime numbers.

# Part II.

# Rings

# 6. Rings

## 6.1. Rings

We defined associative rings in §2.5 (page 79). Now we define rings in general. If $E$ is an abelian group (written additively), then a **multiplication** on $E$ is a binary operation $\cdot$ that distributes in both senses over addition, so that

$$x \cdot (y + z) = x \cdot y + x \cdot z, \qquad (x + y) \cdot z = x \cdot z + y \cdot z.$$

A **ring** is an abelian group with a multiplication. In particular, if $(R, 1, \cdot)$ is an associative ring, then $(R, \cdot)$ is a ring. However, rings that are not (reducts of) associative rings are also of interest: see the next section.

**Theorem 177.** *Every ring satisfies the identities*

$$(x - y) \cdot z = x \cdot z - y \cdot z, \qquad x \cdot (y - z) = x \cdot y - x \cdot z.$$

*Hence, in particular,*

$$0 \cdot x = 0 = x \cdot 0,$$
$$(-x) \cdot y = -(x \cdot y) = x \cdot (-y).$$

By Theorem 63 (page 72), given an abelian group $E$, we have a homomorphism $n \mapsto (x \mapsto nx)$ from the monoid $(\mathbb{Z}, 1, \cdot)$ to the monoid $(E^E, \mathrm{id}_E, \circ)$. This is actually a homomorphism of associative rings:

**Theorem 178.** *For every abelian group $E$,*

$$n \mapsto (x \mapsto nx) \colon (\mathbb{Z}, 0, -, +, 1, \cdot) \to (\mathrm{End}(E), \mathrm{id}_E, \circ).$$

*In particular,*

$$0x = 0, \qquad 1x = x, \qquad (-1)x = -x.$$

In the theorem, if the abelian group has a multiplication, then

$$0 \cdot x = 0x,$$

where the zeros come from the ring and from $\mathbb{Z}$ respectively. If, further, the multiplication has the identity 1, then

$$1 \cdot x = 1x.$$

More generally, we have

**Theorem 179.** *For every integer $n$, every ring satisfies the identities*

$$(nx) \cdot y = n(x \cdot y) = x \cdot ny.$$

The kernel of the homomorphism in Theorem 178 is $\langle k \rangle$ for some $k$ in $\omega$, by Theorem 91 (page 99). Then $k$ can then be called the **characteristic** of $E$. For example, if $n \in \mathbb{N}$, then $\mathbb{Z}_n$ has characteristic $n$, while $\mathbb{Z}$ has characteristic 0.

**Theorem 180.** *If $(E, 1, \cdot)$ is a ring with a multiplicative identity 1, then*

$$n \mapsto n1 \colon (\mathbb{Z}, 0, -, +, 1, \cdot) \to (E, 1, \cdot).$$

*The kernel of this homomorphism is $\langle k \rangle$, where $k$ is the characteristic of $E$.*

**Theorem 181.** *Every ring embeds in a ring with identity having the same characteristic, and in a ring with identity having characteristic 0.*

*Proof.* Suppose $R$ is a ring of characteristic $n$. Let $A$ be $\mathbb{Z}$ or $\mathbb{Z}_n$, and give $A \oplus R$ the multiplication defined by

$$(m, x)(n, y) = (mn, my + nx + xy);$$

then $(1, 0)$ is an identity, and $x \mapsto (0, x)$ is an embedding. $\square$

## 6.2. Examples

The continuous functions on $\mathbb{R}$ with compact support compose a ring with respect to the operations induced from $\mathbb{R}$. Multiplication in this ring is associative, but there is no identity.

If $n > 1$, then $\langle n \rangle$ is a sub-ring of $\mathbb{Z}$ with no identity.

On page 97 we obtained $\mathbb{H}$ as the sub-ring of $\mathrm{M}_{2\times 2}(\mathbb{C})$ that is the image of $\mathbb{C} \oplus \mathbb{C}$ under the group-homomorphism

$$(x, y) \mapsto \begin{pmatrix} x & y \\ -\bar{y} & \bar{x} \end{pmatrix}.$$

We also defined

$$\mathrm{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

so that every element of $\mathbb{H}$ is $z + w\mathrm{j}$ for some unique $(z, w)$ in $\mathbb{C}^2$. Then $\mathbb{H}$ has the automorphism $z + w\mathrm{j} \mapsto \overline{z + w\mathrm{j}}$, where

$$\overline{z + w\mathrm{j}} = \bar{z} - w\mathrm{j}.$$

then the same construction that creates $\mathbb{H}$ out of $\mathbb{C}$ can be applied to $\mathbb{H}$ itself, yielding the ring $\mathbb{O}$ of **octonions;** but this ring is not associative.

In any ring $(E, \cdot)$, we define

$$[x, y] = x \cdot y - y \cdot x;$$

Then the binary operation $(x, y) \mapsto [x, y]$ is also a multiplication on $E$. This operation can be called the **Lie bracket.** We have

$$[x, x] = 0. \tag{6.1}$$

**Theorem 182.** *In an associative ring,*

$$\big[[x, y], z\big] = \big[x, [y, z]\big] - \big[y, [x, z]\big]. \tag{6.2}$$

The identity (6.2) is called the **Jacobi identity.** A **Lie ring** is a ring whose multiplication has the properties of the Lie bracket given by the identities (6.1) and (6.2). if $(E, 1, \cdot)$ is an associative ring, and b is the Lie bracket in this ring, then $(E, \mathrm{b})$ is a Lie ring. However, we shall see presently that there are Lie rings that do not arise in this way.

If $(E, \cdot)$ is a ring, and $D$ is an element of $\mathrm{End}(E)$ satisfying the **Leibniz rule**

$$D(x \cdot y) = Dx \cdot y + x \cdot Dy,$$

then $D$ is called a **derivation** of $(E, \circ)$. For example, let $\mathrm{C}_\infty(\mathbb{R})$ be the set of all infinitely differentiable functions from $\mathbb{R}$ to itself. This is an associative ring in the obvious way. Then differentiation is a derivation of $\mathrm{C}_\infty(\mathbb{R})$.

**Theorem 183.** *The set of derivations of a ring* $(E, \cdot)$ *is the universe of an abelian subgroup of* $\mathrm{End}(E)$ *and is closed under the bracket*

$$(X, Y) \mapsto X \circ Y - Y \circ X.$$

The abelian group of derivations of a ring $(E, \cdot)$ can be denoted by

$$\mathrm{Der}(E, \cdot).$$

Then $(\mathrm{Der}(E, \cdot), \mathrm{b})$ is a sub-ring of $\mathrm{End}(E), \mathrm{b})$, but is not generally closed under $\circ$.

## 6.3. Associative rings

We know from Theorem 74 (page 81) that an associative ring $(R, 1, \cdot)$ has a group of units, $R^\times$. In particular, in an associative ring, when an element has both a left and a right inverse, they are equal. However, the example on page 81 shows that some ring elements can have right inverses that are not units.

A **zero-divisor** of the ring $R$ is a element $b$ distinct from 0 such that the equations

$$bx = 0, \qquad\qquad yb = 0$$

are soluble in $R$. So zero-divisors are not units. For example, if $m > 1$ and $n > 1$, then $m + \langle mn\rangle$ and $n + \langle mn\rangle$ are zero-divisors in $\mathbb{Z}_{mn}$. The unique element of the trivial ring $\mathbb{Z}_1$ is a unit, but not a zero-divisor.

A commutative ring is an **integral domain** if it has no zero-divisors and $1 \neq 0$. If $n \in \mathbb{N}$, the ring $\mathbb{Z}_n$ is an integral domain if and only if $n$ is prime.[1] Hence the characteristic of an integral domain must be prime or 0. Fields are integral domains, but $\mathbb{Z}$ is an integral domain that is not a field. If $p$ is prime, then, by Theorem 105 (page 109), $\mathbb{Z}_p$ is a field, and as such it is denoted by

$$\mathbb{F}_p.$$

An arbitrary associative ring $R$ such that $R \smallsetminus R^{\times} = \{0\}$ is a **division ring.** So fields are division rings; but $\mathbb{H}$ is a non-commutative division ring.

If $R$ is an associative ring, and $G$ is a group, we can form the direct sum $\sum_{g\in G} R$, which is, first of all, an abelian group. It becomes a module over $R$ (in the sense of sub-§3.1.5, page 89) when we define

$$r \cdot (x_g \colon g \in G) = (r \cdot x_g \colon g \in G)$$

for all $r$ in $R$ and $(x_g \colon g \in G)$ in $\sum_{g\in G} R$. If $g \in G$, we have the canonical injection $\iota_g$ of $R$ in $\sum_{g\in G} R$ as defined on page 124. Let us denote $\iota_g(1)$ also by

$$g.$$

Then

$$(r_g \colon g \in G) = \sum_{g\in G} r_g \cdot g.$$

Thus an element of $\sum_{g\in G} R$ becomes a **formal $R$-linear combination** of elements of $G$. Then multiplication on $\sum_{g\in G} R$ is defined in an obvious way: if $r_i \in R$ and $g_i \in G$ for each $i$ in 2, then

$$(r_0 \cdot g_0)(r_1 \cdot g_1) = r_0 r_1 \cdot g_0 g_1.$$

[1] Lang refers to integral domains as *entire* rings [23, p. 91]. It would appear that integral domains were originally subgroups of $\mathbb{C}$ that are closed under multiplication *and* that include the integers [4, p. 47].

The definition extends to all of $\sum_{g\in G} R$ by distributivity. The resulting ring can be denoted by

$$R(G);$$

it is the **group ring** of $G$ over $R$.

We can do the same construction with monoids, rather than groups. For example, if we start with the free monoid generated by a symbol $X$, we get a **polynomial ring** in one variable, denoted by

$$R[X];$$

this is the ring of formal $R$-linear combinations of powers of $X$. Such combinations can be written as

$$\sum_{k<n} a_k X^k,$$

where $(a_k : k < n) \in R^n$, where $n \in \omega$. In case $n = 0$, the indicated combination is 0; in case $n = m + 1$, the combination can be written as one of

$$\sum_{k=0}^{m} a_k X^k, \qquad\qquad a_0 + a_1 X + a_2 X^2 + \cdots + a_m X^m.$$

This combination too is 0 when each $a_k$ is 0. We could use a second variable, getting for example $R[X, Y]$, which is just $R[X][Y]$. Usually $R$ here is commutative and is in particular a field or at least an integral domain. We shall develop the theory of polynomial rings in §7.7 (page 211), but shall use them meanwhile as examples.

## 6.4. Ideals

Suppose $(R, 0, -, +, \cdot)$ is a ring, and $\sim$ is a congruence-relation on $(R, +, \cdot)$. By Theorem 85 on page 94, $\sim$ is a congruence-relation on the ring. (The theorem is stated for associative rings, but does not require the associativity.) If $A = \{x \in R : x \sim 0\}$, then by Theorem 87 (page 95), $A$ is a *subgroup* of $R$, that is,

$$(A, 0, -, +) < (R, 0, -, +).$$

Similarly, $A$ is even a sub-ring of $R$, that is, in addition to being a subgroup, it is closed under multiplication. We have

$$\begin{aligned} b \sim x &\iff b - x \sim 0 \\ &\iff b - x \in A \\ &\iff b + A = x + A. \end{aligned}$$

In short,

$$b \sim x \iff b + A = x + A.$$

Conversely, given a sub-ring $A$ of $R$, we can use the last equivalence as a definition of $\sim$. Then $\sim$ is an equivalence-relation on $R$ by Corollary 101.1 (page 107), and by this and Theorem 108 (page 110), $\sim$ is even a congruence-relation on $R$ *as a group.* However, $\sim$ need not be a congruence-relation on $R$ as a ring. That is, it may not be possible to define a multiplication on $R/A$ by

$$(x + A)(y + A) = xy + A. \tag{6.3}$$

For example, we cannot use this to define a multiplication on $\mathbb{Q}/\mathbb{Z}$, since for example

$$\frac{1}{2} + \mathbb{Z} = \frac{3}{2} + \mathbb{Z}, \qquad \frac{1}{4} + \mathbb{Z} \neq \frac{3}{4} + \mathbb{Z}.$$

**Theorem 184.** *Suppose $R$ is a ring and $A$ is a sub-ring. The group $R/A$ expands to a ring with multiplication as in* (6.3) *if and only if*

$$r \in R \;\&\; a \in A \implies ra \in A \;\&\; ar \in A. \tag{6.4}$$

*Proof.* If $R/A$ does expand to a ring, and $a \in A$, then $a + A$ is $0$ in this ring, and hence so are $ra + A$ and $ar + A$ by Theorem 177, so that (6.4) holds. Conversely, suppose this holds. If $a+A = x+A$ and $b+A = y+A$, then $A$ contains $a - x$ and $b - y$, so $A$ contains also

$$(a - x) \cdot y + a \cdot (b - y),$$

which is $ab - xy$, so $ab + A = xy + A$. □

Under the equivalent conditions of the theorem, $A$ is called an **ideal** of $R$. The historical reason for the name is suggested in §7.3 (page 187). Meanwhile, he have the following counterpart of Theorem 112 (page 113).

**Theorem 185.** *A sub-ring of a ring $R$ is an ideal of $R$ if and only if it is the kernel of a homomorphism on $R$.*

We can express (6.4) as

$$RA \subseteq A, \qquad\qquad AR \subseteq A.$$

If only one of these holds, then $A$ is called respectively a **left ideal** of $R$ or a **right ideal** of $R$. However, left ideals and right ideals are not kinds of ideals; rather, an ideal is a left ideal that is also a right ideal. One may therefore refer to ideals as **two-sided ideals.**

For example, the set of matrices

$$\begin{bmatrix} * & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ * & 0 & \dots & 0 \end{bmatrix}$$

is a left ideal of $\mathrm{M}_{n\times n}(R)$, but not a right ideal unless $n = 1$. Also, for every element $a$ of an *associative* ring $R$, the subset $Ra$ is a left ideal of $R$, while $RaR$ is a two-sided ideal.

We have the following counterpart to Theorem 113 for groups.

**Theorem 186.** *If $f$ is a homomorphism from a ring $R$ to a ring $S$, and $I$ is a two-sided ideal of $R$ included in* $\ker(f)$*, then there is a unique homomorphism $\tilde{f}$ from $R/I$ to $S$ such that $f = \tilde{f} \circ \pi$.*

Hence the isomorphism theorems, as for groups.

Suppose $(A_i \colon i \in I)$ is an indexed family of left ideals of a ring $R$. Let the abelian subgroup of $R$ generated by $\bigcup_{i\in I} A_i$ be denoted by

$$\sum_{i\in I} A_i;$$

this is the **sum** of the left ideals $A_i$. This must not be confused with the *direct sums* defined in §4.2 (page 123).

Given a *finite* indexed family $(A_0, \ldots, A_{n-1})$ of left ideals of an *associative* ring $R$, we let the abelian subgroup of $R$ generated by

$$\{a_0 \cdots a_{n-1} \colon a_i \in A_i\}$$

be denoted by

$$A_0 \cdots A_{n-1};$$

this is the **product** of the left ideals $A_i$.

**Theorem 187.** *Sums and finite products of left ideals are left ideals; sums and products of two-sided ideals are two-sided ideals. Addition and multiplication of ideals are associative; addition is commutative; multiplication distributes over addition.*

**Theorem 188.** *If $A$ and $B$ are left ideals of a ring, then so is $A \cap B$. If they are two-sided ideals, then $AB \subseteq A \cap B$.*

Usually $AB$ does not include $A \cap B$, since for example $A^2$ might not include $A$; such is the case when $A = 2\mathbb{Z}$, since then $A^2 = 4\mathbb{Z}$.

# 7. Commutative rings

Throughout this chapter, "ring" means commutative ring. We shall often identify properties of $\mathbb{Z}$ and then consider arbitrary rings with these properties. If $R$ is a ring (that is, a commutative ring) with an ideal $I$, and $a + I = b + I$, we may write this as

$$a \equiv b \pmod{I}.$$

## 7.1. Commutative rings

A subset $A$ of a ring $R$ determines the ideal denoted by

$$(A),$$

namely the smallest ideal including $A$. This consists of the $R$**-linear combinations** of elements of $A$, namely the well-defined sums

$$\sum_{a \in A} r_a a,$$

where $r_a \in R$; in particular, $r_a = 0$ for all but finitely many $a$. If $A = \{a_i : i < n\}$, then $(A)$ can be written as one of

$$(a_i : i < n), \qquad\qquad R a_0 + \dots + R a_{n-1}.$$

In particular, if $A = \{a\}$, then $(A)$ is denoted by one of

$$(a), \qquad\qquad Ra$$

and is called a **principal ideal.** Then

$$(a_i : i < n) = (a_0) + \dots + (a_{n-1}).$$

In $\mathbb{Z}$, the ideal $(a)$ is the same as the subgroup $\langle a\rangle$. Therefore every ideal of $\mathbb{Z}$ is principal, by Theorem 91 (page 99). A **principal ideal domain** or **PID** is an integral domain whose every ideal is principal. Thus $\mathbb{Z}$ is a PID, but the polynomial ring $\mathbb{Q}[X, Y]$ is not, since the ideal $(X, Y)$ is not principal.

The following is Proposition VII.30 of Euclid's *Elements.* We are going to be interested in rings besides $\mathbb{Z}$ in which the proof can be carried out. Meanwhile, it will motivate the definition of *prime ideal* below.

**Theorem 189** (Euclid's Lemma)**.** *If $p$ is a prime number, then for all integers $a$ and $b$,*

$$p \mid ab \;\;\&\;\; p \nmid a \implies p \mid b.$$

*Proof.* Given that $p \nmid a$, we know that $\gcd(p, a) = 1$ by the proof of Theorem 105 (page 109; or by the result of this theorem and Theorem 94, page 101). Hence by Theorem 93, we can solve $ax + py = 1$. In this case we obtain

$$abx + pby = b,$$

so if $p \mid ab$, then, since immediately $p \mid pby$, we must have $p \mid b$. □

An ideal of a ring is **proper** if it is not the whole ring. A ring has a unique improper ideal, namely itself, which can be written as

$$(1).$$

Thus an ideal is proper if and only if it does not contain 1. When $A$ is the empty subset of a ring, then the ideal $(A)$, which is $\{0\}$, is usually denoted by

$$(0).$$

This can be counted as a principal ideal. Considering the last theorem, and noting that, in $\mathbb{Z}$,

$$a \mid b \iff b \in (a),$$

we refer to a *proper* ideal $P$ of a ring $R$ as

- **prime,** if for all $a$ and $b$ in $R$,

$$ab \in P \;\;\&\;\; a \notin P \implies b \in P; \tag{7.1}$$

- **maximal,** if for all ideals $J$ of $R$,

$$I \subset J \implies J = R.$$

**Theorem 190.** *Let $R$ be a ring.*

1. *$R$ is an integral domain $\iff$ $(0)$ is a prime ideal.*
2. *$R$ is a field $\iff$ $(0)$ is a maximal ideal.*

*More generally, for an arbitrary ideal $I$ of $R$:*

3. *$R/I$ is an integral domain $\iff$ $I$ is a prime ideal.*
4. *$R/I$ is a field $\iff$ $I$ is a maximal ideal.*

*Proof.* 1. This is immediate from the definitions of integral domain and prime ideal, once we note that $x \in (0)$ means $x = 0$.

2. If $R$ is a field and $(0) \subset I$, then $I \smallsetminus (0)$ contains some $a$, and then $a^{-1} \cdot a \in I$, so $I = R$. Conversely, if $(0)$ is maximal, then for all $a$ in $R \smallsetminus (0)$ we have $(a) = (1)$, so $a$ is invertible.

3. The ideal $(0)$ of $R/I$ is $\{I\}$, and

$$(a+I)(b+I) = I \iff ab \in I.$$

4. By Theorem 163 (page 159), every ideal of $R/I$ is $J/I$ for some subgroup $J$ of $R$. Moreover, this $J$ must be an ideal of $R$. In this case, $J$ is maximal if and only if $J/I$ is a maximal ideal of $R/I$. □

**Corollary 190.1.** *Maximal ideals are prime.*

The prime ideals of $\mathbb{Z}$ are precisely the ideals $(0)$ and $(p)$, where $p$ is prime. Indeed, $(0)$ is prime because $\mathbb{Z}$ is an integral domain, and if $p$ is prime, then $\mathbb{Z}_p$ is the field $\mathbb{F}_p$, so $(p)$ is even maximal. If $n > 1$ and is not prime, so that $n = ab$ for some $a$ and $b$ in $\{2, \dots, n-1\}$, then $a$ and $b$ are zero-divisors in $\mathbb{Z}_n$, so $(n)$ is not prime.

The converse of the corollary fails easily, since $(0)$ is a prime but non-maximal ideal of $\mathbb{Z}$. However, every prime ideal of $\mathbb{Z}$ other than $(0)$ is maximal. This is not the case for $\mathbb{Q}[X, Y]$, which has the prime but non-maximal ideal $(X)$.

In some rings, *every* prime ideal is maximal. Such is the case for fields, since their only proper ideals are $(0)$. It is also the case for *Boolean rings.* A ring is called **Boolean** if it satisfies the identity

$$x^2 = x.$$

In defining ultraproducts in §7.6 (page 200), we shall use the example established by the following:

**Theorem 191.** *if $\Omega$ is a set, then $\mathscr{P}(\Omega)$ is a Boolean ring, where*

$$X \cdot Y = X \cap Y, \qquad X + Y = (X \smallsetminus Y) \cup (Y \smallsetminus X).$$

**Theorem 192.** *Every Boolean ring in which $0 \neq 1$ has characteristic $2$.*

*Proof.* In a Boolean ring, $2x = (2x)^2 = 4x^2 = 4x$, so

$$2x = 0. \qquad \square$$

The following will be generalized by Theorem 213 (page 206).

**Theorem 193.** *In Boolean rings, all prime ideals are maximal.*

*Proof.* In a Boolean ring,

$$x \cdot (x - 1) = x^2 - x = x - x = 0,$$

so every $x$ is a zero-divisor unless $x$ is $0$ or $1$. Therefore there are no Boolean integral domains besides $\{0, 1\}$, which is the field $\mathbb{F}_2$. $\square$

In $\mathbb{Z}$, by Theorem 91, the ideal $(a, b)$ is the principal ideal generated by $\gcd(a, b)$. So $a$ and $b$ are relatively prime if and only if $(a, b) = \mathbb{Z}$. We can write this condition as

$$(a) + (b) = \mathbb{Z}.$$

Then the following generalizes Theorem 138.

**Theorem 194** (Chinese Remainder Theorem). *Suppose $R$ has an indexed family $(I_i : i < n)$ of ideals such that*

$$i < j < n \implies I_i + I_j = R.$$

*The monomorphism*

$$x + \bigcap_{i<n} I_i \mapsto (x + I_i : i < n) \tag{7.2}$$

*from $R/\bigcap_{i<n} I_i$ to $\sum_{i<n} R/I_i$ is an isomorphism. That is, every system*

$$\big(x \equiv a_0 \pmod{I_0}\big) \;\&\; \cdots \;\&\; \big(x \equiv a_{n-1} \pmod{I_{n-1}}\big)$$

*of congruences has a solution in $R$, and the solution is unique* modulo $I_0 \cap \dots \cap I_{n-1}$.

*Proof.* We proceed by induction. The claim is trivially true when $n = 1$. In case $n = 2$, we have $b_0 + b_1 = 1$ for some $b_0$ in $I_0$ and $b_1$ in $I_1$. Then

$$\begin{aligned} b_0 &\equiv 0 \pmod{I_0}, & b_0 &\equiv 1 \pmod{I_1}, \\ b_1 &\equiv 1 \pmod{I_0}, & b_1 &\equiv 0 \pmod{I_1}. \end{aligned}$$

Therefore

$$b_1 a_0 + b_0 a_1 \equiv a_0 \pmod{I_0}, \qquad b_1 a_0 + b_0 a_1 \equiv a_1 \pmod{I_1}.$$

Thus $(a_0 + I_0, a_1 + I_1)$ is in the image of the map in (7.2).

Finally, if the claim holds when $n = m$, then it holds when $n = m+1$ by the proof of the case $n = 2$, once we note that if

$$a_i + b_i = 1$$

for some $a_i$ in $I_i$ and $b_i$ in $I_m$ for each $i$ in $m$, then

$$\prod_{i<m} (a_i + b_i) = 1;$$

but this product is the sum of $\prod_{i<m} a_i$ and an element of $I_m$, and

$$\prod_{i<m} a_i \in \bigcap_{i<m} I_i. \qquad \square$$

## 7.2. Division

As in $\mathbb{Z}$ (page 55), so in an arbitrary ring $R$, an element $a$ is called a **divisor** or **factor** of an element $b$, and $a$ is said to **divide** $b$, and we write

$$a \mid b,$$

if the equation

$$ax = b$$

is soluble in $R$. Two elements of $R$ that divide each other can be called **associates.** Zero is an associate only of itself.

**Theorem 195.** *In any ring:*

1. $a \mid b \iff (b) \subseteq (a)$*;*

2. *$a$ and $b$ are associates if and only if $(a) = (b)$.*

*Suppose $a = bx$.*

3. *If $x$ is a unit, then $a$ and $b$ are associates.*

4. *If $b$ is a zero-divisor or $0$, then so is $a$.*

5. *If $a$ is a unit, then so is $b$.*

For example, in $\mathbb{Z}_6$, the elements 1 and 5 are units; the other non-zero elements are zero-divisors. Of these, 2 and 4 are associates, since

$$2 \cdot 2 \equiv 4, \qquad\qquad 4 \cdot 2 \equiv 2 \pmod{6}; \tag{7.3}$$

but 3 is not an associate of these.

We now distinguish the properties of certain ring-elements that, by Euclid's Lemma (page 182), are the same in $\mathbb{Z}$. In an arbitrary ring $R$, an element $\pi$ that is neither 0 nor a unit is called

- **irreducible,** if for all $a$ and $b$ in $R$,

$$\pi = ab \;\;\&\;\; a \notin R^{\times} \implies b \in R^{\times};$$

- **prime,** if for all $a$ and $b$ in $R$,

$$\pi \mid ab \ \& \ \pi \nmid a \implies \pi \mid b.$$

**Theorem 196.** *A nonzero ring-element* $\pi$ *is*

1) *irreducible* $\iff$ $(\pi)$ *is maximal among the proper principal ideals;*

2) *prime* $\iff$ $(\pi)$ *is prime.*

For example, in $\mathbb{Q}[X, Y]$, the element $X$ is both irreducible and prime, although $(X)$ is not a maximal ideal. However, if $(X) \subseteq (f) \subset \mathbb{Q}[X, Y]$, then $f$ must be constant in $Y$, and then it must have degree 1 in $X$, and then its constant term must be 0; so $f$ is just $aX$ for some $a$ in $\mathbb{Q}^\times$, and thus $(X) = (f)$.

If $\pi$ is irreducible *or* prime, and $\pi = ab$, then $\pi$ is an associate of $a$ or $b$. However, neither irreducibility nor primality implies the other. For example, in $\mathbb{Z}_6$, the element 2 is prime. Indeed, $(2) = \{0, 2, 4\}$, so $\mathbb{Z}_6 \smallsetminus (2) = \{1, 3, 5\}$, and the product of no two of these elements is in (2). Similarly, 4 is prime. However, 2 and 4 are not irreducible, by (7.3) above.

Also, in $\mathbb{C}$ we have

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}). \tag{7.4}$$

The factors 2, 3, and $1 \pm \sqrt{-5}$ are all irreducible in the smallest sub-ring of $\mathbb{C}$ that contains $\sqrt{-5}$, but none of these factors divides another, and so these factors cannot be prime. Details are worked out in the next section.

## 7.3. *Quadratic integers

Every subfield of $\mathbb{C}$ includes $\mathbb{Q}$, and every sub-ring of $\mathbb{C}$ includes $\mathbb{Z}$. If $\omega \in \mathbb{C}$, then the smallest subfield of $\mathbb{C}$ that contains $\omega$ is denoted by

$$\mathbb{Q}(\omega),$$

and the smallest sub-ring of $\mathbb{C}$ that contains $\omega$ is denoted by

$$\mathbb{Z}[\omega].$$

A **squarefree** integer, is an element of $\mathbb{Z}$ different from 1 that is not divisible by the square of a prime number. Suppose $D$ is such. As groups,

- $\mathbb{Z}[\sqrt{D}]$ is the free abelian group $\langle 1, \sqrt{D}\rangle$,
- $\mathbb{Q}(\sqrt{D})$ is the image of $\mathbb{Q} \oplus \mathbb{Q}$ under $(x, y) \mapsto x + y\sqrt{D}$.

If $x = k + n\sqrt{D}$ for some $k$ and $n$ in $\mathbb{Z}$, then

$$(x - k)^2 = n^2 D,$$
$$x^2 - 2kx + k^2 - n^2 D = 0.$$

Thus all elements of $\mathbb{Z}[\sqrt{D}]$ are solutions in $\mathbb{Q}(\sqrt{D})$ of quadratic equations

$$x^2 + bx + c = 0, \tag{7.5}$$

where $b$ and $c$ are in $\mathbb{Z}$, and there is no leading coefficient.[1] Conversely, from school the solutions of (7.5) are

$$x = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

Suppose one of these is in $\mathbb{Q}(\sqrt{D})$. Then $b^2 - 4c = a^2 D$ for some $a$ in $\mathbb{Z}$, so that

$$x = \frac{-b \pm a\sqrt{D}}{2}.$$

If $b$ is odd, then $b^2 - 4c \equiv 1 \pmod 4$, so $a$ must be odd and $D \equiv 1 \pmod 4$. If $b$ is even, then $b^2 - 4c \equiv 0 \pmod 4$, so $a$ is even. Assume now

$$D \not\equiv 1 \pmod 4.$$

Then $\mathbb{Z}[\sqrt{D}]$ consists precisely of the solutions in $\mathbb{Q}(\sqrt{D})$ of equations of the form (7.5). Therefore the elements of $\mathbb{Z}[\sqrt{D}]$ are called the **integers**

[1] If $\xi$ is a solution of such an equation, so that $\xi^2 = -b\xi - c$, David Hilbert referred to the group $\langle 1, \xi\rangle$ as a *number ring* (*Zahlring*) [4, p. 49]. This is apparently the origin of our term *ring*.

**of** $\mathbb{Q}(\sqrt{D})$.[2] In this context, the elements of $\mathbb{Z}$ are the integers of $\mathbb{Q}$, or the **rational integers.** Note that $\mathbb{Z}[\sqrt{D}] \cap \mathbb{Q} = \mathbb{Z}$.

The field $\mathbb{Q}(\sqrt{D})$ has one nontrivial automorphism, namely $z \mapsto z'$, where

$$(x + y\sqrt{D})' = x - y\sqrt{D}.$$

In case $D < 0$, this automorphism is complex conjugation. In any case, we next define a function $N$ from $\mathbb{Q}(\sqrt{D})$ to $\mathbb{Q}$ by

$$N(z) = zz'.$$

Here $N(z)$ can be called the **norm** of $z$. The function $N$ is multiplicative, that is,

$$N(\alpha\beta) = N(\alpha) \cdot N(\beta).$$

Also,

$$N(x + \sqrt{D}y) = x^2 - Dy^2,$$

so $N$ maps $\mathbb{Z}[\sqrt{D}]$ into $\mathbb{Z}$. In particular, if $\alpha$ is a unit of $\mathbb{Z}[\sqrt{D}]$, then $N(\alpha)$ must be a unit of $\mathbb{Z}$, namely $\pm 1$. Conversely, if $N(\alpha) = \pm 1$, this means $\alpha \cdot (\pm\alpha') = 1$, so $\alpha$ is a unit.

If $D < 0$, then $N$ maps $\mathbb{Z}[\sqrt{D}]$ into $\mathbb{N}$, and so $\alpha$ is a unit in $\mathbb{Z}[\sqrt{D}]$ if and only if $N(\alpha) = 1$. Also, $\alpha$ in $\mathbb{Z}[\sqrt{D}]$ is irreducible if and only if it has no divisor $\beta$ such that $1 < N(\beta) < N(\alpha)$ and $N(\beta) \mid N(\alpha)$.

In case $D = -5$ we have

$$\begin{array}{c||c|c|c} x & 2 & 3 & 1 \pm \sqrt{-5} \\ \hline N(x) & 4 & 9 & 6 \end{array}. \tag{7.6}$$

Since no elements of $Z[\sqrt{-5}]$ have norm 2 or 3, the elements 2, 3, and $1 \pm \sqrt{-5}$ are irreducible. However, they are not prime, because each of them divides the product of *two* of the others, but it does not divide *one* of the others, since if $\alpha \mid \beta$, then $N(\alpha) \mid N(\beta)$, but no norm in (7.6) divides another.

[2] In case $D \equiv 1 \pmod 4$, the integers of $\mathbb{Q}(\sqrt{D})$ constitute the ring $\mathbb{Z}[(1 + \sqrt{D})/2]$.

There are however factorizations of the relevant ideals. For example,

$$\begin{aligned}(2,1+\sqrt{-5})(2,1+\sqrt{-5}) = (2,1+\sqrt{-5})&(2,1-\sqrt{-5})\\ &= (4,2+2\sqrt{-5},6) = (2).\end{aligned}$$

Similarly,

$$\begin{aligned}(3) &= (3,1+\sqrt{-5})(3,1-\sqrt{-5}),\\ (1+\sqrt{-5}) &= (2,1+\sqrt{-5})(3,1+\sqrt{-5}),\\ (1-\sqrt{-5}) &= (2,1+\sqrt{-5})(3,1-\sqrt{-5}).\end{aligned}$$

These factorizations are *prime* factorizations. We show this as follows. Every subgroup of $\langle 1,\sqrt{D}\rangle$ has at most two generators, by Porism 137.1 (page 141). When that subgroup is a nonzero ideal $I$ of $\mathbb{Z}[\sqrt{D}]$, then it must have more than one generator as a group, since a cyclic subgroup will not be closed under multiplication by $\sqrt{D}$. For example, since

$$(a+b\sqrt{D})\cdot\sqrt{D} = bD + a\sqrt{D},$$

the ideal $(a+b\sqrt{D})$ is the group

$$\langle a+b\sqrt{D}, bD+a\sqrt{D}\rangle.$$

Let $G$ be the map

$$\begin{pmatrix} a & b\\ c & d\end{pmatrix} \mapsto \langle a+b\sqrt{D}, c+d\sqrt{D}\rangle$$

from $\mathrm{M}_{n\times n}(\mathbb{Z})$ to the set of subgroups of $\mathbb{Z}[\sqrt{D}]$. If $G(X)$ is an ideal, then $\det(X)\neq 0$. Also, $G(X) < G(Y)$ if and only if $X = ZY$ for some $Z$ such that $\det(Z)\neq 0$. Hence $G(X) = G(Y)$ if and only if $X = ZY$ for some $Z$ in $\mathrm{GL}_2(\mathbb{Z})$. By the methods of the proof of Theorem 137 (page 139), every ideal of $\mathbb{Z}[\sqrt{D}]$ has the form

$$\langle a, b+c\sqrt{D}\rangle,$$

where $a > b \geqslant 0$. (This is not a sufficient condition for being an ideal, however.) We have a well-defined function $N$ from the set of subgroups of $\mathbb{Z}[\sqrt{D}]$ to $\mathbb{N}$ given by

$$N\bigl(G(X)\bigr) = |\det(X)|\,.$$

In case $D < 0$, this new function $N$ agrees with the earlier function called $N$ in the sense that

$$\begin{aligned} N\big((a+b\sqrt{D})\big) &= \mathbb{N}\big(\langle a+b\sqrt{D}, bD+a\sqrt{D}\rangle\big) \\ &= \big|a^2-b^2D\big| = a^2-b^2D = N(a+b\sqrt{D}). \end{aligned}$$

If $I$ and $J$ are ideals of $\mathbb{Z}[\sqrt{D}]$ such that $I \subset J \subset \mathbb{Z}[\sqrt{D}]$, then we must have

$$N(J) \mid N(I), \qquad N(I) > N(J) > 1.$$

In case $d = -5$, we compute

$$\begin{aligned} (2, 1+\sqrt{-5}) &= \langle 2, 2\sqrt{-5}, 1+\sqrt{-5}, \sqrt{-5}-5\rangle = \langle 2, 1+\sqrt{-5}\rangle, \\ (3, 1\pm\sqrt{-5}) &= \langle 3, 3\sqrt{-5}, 1\pm\sqrt{-5}, \sqrt{-5}\mp 5\rangle = \langle 3, 1\pm\sqrt{-5}\rangle, \end{aligned}$$

hence

| $I$ | $(2, 1+\sqrt{-5})$ | $(3, 1\pm\sqrt{-5})$ |
|---|---|---|
| $N(I)$ | 2 | 3 |

So these ideals are maximal, hence prime. Ideals of the rings $\mathbb{Z}[\sqrt{D}]$ were originally called *ideal numbers.*

## 7.4. Integral domains

We now consider some rings that are increasingly close to having all of the properties of $\mathbb{Z}$. We start with arbitrary integral domains. We have noted in effect that the following fails in $\mathbb{Z}_6$.

**Lemma 20.** *In an integral domain, if $a$ and $b$ are non-zero associates, and*

$$a = bx,$$

*then $x$ is a unit.*

*Proof.* We have also, for some $y$,

$$b = ay = bxy, \qquad b \cdot (1-xy) = 0, \qquad 1 = xy,$$

since $b \neq 0$ and we are in an integral domain. □

**Theorem 197.** *In an integral domain, prime elements are irreducible.*

*Proof.* If $p$ is prime, and $p = ab$, then $p$ is an associate of $a$ or $b$, so the other is a unit. □

By this and Euclid's Lemma (page 182), the irreducibles of $\mathbb{Z}$ are precisely the primes.

Recall from page 166 that a *multiset* is a pair $(A, f)$, where $f\colon A \to \mathbb{N}$. If $A$ here is a finite subset of a ring, then the product

$$\prod_{a \in A} a^{f(a)}$$

is well-defined (see page 74) and can be called the **product of the multiset.** The components of the proof of the following are found in Euclid, although Gauss's version [9, ¶16] seems to be the first formal statement of the theorem [14, p. 10].

**Theorem 198** (Fundamental Theorem of Arithmetic)**.** *Every element of $\mathbb{N}$ has a unique prime factorization. That is, every natural number is the product of a unique multiset of prime numbers.*

*Proof.* We first show that every integer greater than 1 has a prime factor: this is Propositions VII.31–2 of the *Elements.* Suppose $m > 1$, and let $p$ be the least integer $a$ such that $a \mid m$ and $1 < a$. Then $p$ must be prime.

Now suppose $n > 1$, and every $m$ such that $1 < m < n$ has a prime factorization. If $n$ is prime, then it is its own prime factorization. If $n$ is not prime, then $n = pm$ for some prime $p$, where also $1 < m < n$. By hypothesis $m$ has a prime factorization, and hence so does $n$. Therefore, by induction, every element of $\mathbb{N}$ has a prime factorization.

Prime factorizations are unique by Euclid's Lemma. □

A **unique factorization domain** or **UFD** is an integral domain in which the appropriate formulation of the result of the foregoing theorem holds. Thus, in a UFD, by definition,

1) every nonzero element has an irreducible factorization, that is, every nonzero element is the product of a multiset of irreducibles; and

2) that multiset is unique up to replacement of elements by associates, so that, if

$$\prod_{i<n} \pi_i = \prod_{i<n'} \pi'_i,$$

where the $\pi_i$ and $\pi'_i$ are irreducible, then $n = n'$ and, for some $\sigma$ in $\mathrm{Sym}(n)$, for all $i$ in $n$, $\pi_i$ and $\pi'_{\sigma(i)}$ are associates.

Existence of irreducible factorizations in $\mathbb{Z}$, along with Euclid's Lemma, ensures that those factorizations are unique, so that $\mathbb{Z}$ is a UFD. Conversely, the definition of a UFD is enough to give us Euclid's Lemma:

**Theorem 199.** *In a* UFD, *irreducibles are prime.*

As for $\mathbb{Z}$ (page 100), so for any ring, a **greatest common divisor** of elements $a$ and $b$ is a common divisor of $a$ and $b$ that is a maximum with respect to dividing: that is, it is some $c$ such that $c \mid a$ and $c \mid b$, and for all $x$, if $x \mid a$ and $x \mid b$, then $x \mid c$. There can be more than one greatest common divisor, but they are all associates. Every element of a ring is a greatest common divisor of itself and $0$.

**Theorem 200.** *In a* UFD, *any two nonzero elements have a greatest common divisor.*

*Proof.* We can write the elements as

$$u \prod_{i<n} {\pi_i}^{a(i)}, \qquad v \prod_{i<n} {\pi_i}^{b(i)},$$

where $u$ and $v$ are units and the $\pi_i$ are irreducibles; then the product

$$\prod_{i<n} {\pi_i}^{\min(a(i),b(i))}$$

is a greatest common divisor of the first two elements. □

As in $\mathbb{Z}$, so in an arbitrary PID, more is true, and we shall use this to show that every PID is a UFD. If $a$ and $b$ have a common divisor $d$, then

$$(a, b) \subseteq (d),$$

but we need not have the reverse inclusion, even if $d$ is a greatest common divisor. For example, $\mathbb{Q}[X, Y]$ will be a UFD by Theorem 224 (page 220), and in this ring, $X$ and $Y$ have the greatest common divisor 1, but $(X, Y) \neq 1$. For a PID however, we have the following generalization of Theorem 93 (page 100).

**Theorem 201.** *In a* PID*, any two elements $a$ and $b$ have a greatest common divisor $d$, and*

$$(a, b) = (d),$$

*so that the equation*

$$ax + by = d$$

*is soluble in the ring.*

Now we can generalize Euclid's Lemma.

**Theorem 202.** *In a* PID*, irreducibles are prime.*

*Proof.* Suppose the irreducible $\pi$ divides $ab$ but not $a$. Then 1 is a greatest common divisor of $\pi$ and $a$, and so by the last theorem, $\pi x + ay = 1$ for some $x$ and $y$ in the ring. Now the proof of Euclid's Lemma goes through. □

So now, in a PID, if an element has an irreducible factorization, this factorization is unique. Now, our proof that elements of $\mathbb{N}$ have prime factorizations has two parts. The first part is that every non-unit has a prime factor. The second part can be understood as follows. Suppose some $n_0$ does not have a prime factorization. But $n_0 = p_0 \cdot n_1$ for some prime $p_0$ and some $n_1$. Then $n_1$ in turn must have no prime factorization. Thus $n_1 = p_1 n_2$ for some prime $p_1$ and some $n_2$, and so on. We obtain

$$n_0 > n_1 > n_2 > \cdots, \tag{7·7}$$

which is absurd in $\mathbb{N}$. It follows that $n_0$ must have had a prime factorization.

An arbitrary ring will not have an ordering as $\mathbb{N}$ does, but the relation of divisibility will be an adequate substitute, at least in a PID. Indeed, with the $n_i$ as above, we have

$$(n_0) \subset (n_1) \subset (n_2) \subset \cdots \tag{7.8}$$

This is a strictly ascending chain of ideals. A ring is called **Noetherian** if its every strictly ascending chain of ideals is finite.

**Theorem 203.** *Every* PID *is Noetherian.*

*Proof.* If $I_0 \subseteq I_1 \subseteq \cdots$, then $\bigcup_{i \in \omega} I_i$ is an ideal $(a)$; then $a \in I_n$ for some $n$, so the chain cannot grow beyond $I_n$. □

Now we can adapt to an arbitrary PID the foregoing argument that elements of $\mathbb{N}$ have prime factorizations. In fact that argument can be streamlined. If $n_0$ has no prime factorization, then $n_0 = m_0 \cdot n_1$ for some non-units $m_0$ and $n_1$, where at least $n_1$ has no prime factorization. Again we obtain a descending sequence as in (7.7), hence an ascending sequence as in (7.8).

**Theorem 204.** *Every* PID *is a* UFD.

*Proof.* By Theorem 202, irreducibles in a PID are prime, and therefore irreducible factorizations are unique when they exist. Indeed, if

$$\prod_{i<n} \pi_i = \prod_{i<n'} \pi'_i,$$

where the $\pi_i$ and $\pi'_i$ are irreducible, then, since it divides the right side, $\pi_0$ must divide one of the $\pi'_i$ (because $\pi_0$ is prime). Thus $\pi'_i = u \cdot \pi_0$ for some $u$. Also $u$ must be a unit (because $\pi'_i$ is irreducible and also, being irreducible, $\pi_0$ is not a unit). We may assume $i = 0$. The product $u \cdot \pi'_1$ is an associate of $\pi'_1$ (by Theorem 195) and is therefore also irreducible. Replacing $\pi'_1$ with $u \cdot \pi'_1$, we have

$$\prod_{1 \leqslant i<n} \pi_i = \prod_{1 \leqslant i<n'} \pi'_i,$$

since a PID is an integral domain. By induction, $n = n'$, and for some $\sigma$ in $\operatorname{Sym}(n)$, for all $i$ in $n$, $\pi_i$ and $\pi'_{\sigma(i)}$ are associates.

It remains to show that irreducible factorizations exist in a PID. By the Axiom of Choice, we can well-order the PID. Suppose, if possible, $a \neq 0$ and has no irreducible factorization. Then $a = b \cdot c$ for some non-units $b$ and $c$, where $c$ has no irreducible factorization. We have

$$(a) \subset (c).$$

Now let us denote by $a'$ the *least* such $c$ in the well-ordering. Then we can produce a sequence $(a_i \colon i \in \omega)$, where $a_0$ has no irreducible factorization and, assuming $a_i$ has no irreducible factorization, $a_{i+1} = {a_i}'$. By induction, each $a_i$ does have no irreducible factorization, and so

$$(a_0) \subset (a_1) \subset (a_2) \subset \cdots,$$

which is contrary to the last theorem. Thus every nonzero element of a PID has an irreducible factorization, and this is unique. □

We have thus shown that the Fundamental Theorem of Arithmetic can be founded solely on the status of $\mathbb{Z}$ as a PID. We may now ask further how $\mathbb{Z}$ gets this status. The proof of Theorem 91 can be worked out as follows. The function $x \mapsto |x|$ from $\mathbb{Z}$ to $\omega$ (as defined on page 100) is such that

$$x = 0 \iff |x| = 0.$$

Given an ideal $I$ of $\mathbb{Z}$ that is different from $(0)$, we let $a$ be a nonzero element such that $|a|$ is minimal. If $b \in I$, then

$$|b - ax| < |a|$$

for some $x$ (as for example the $x$ that minimizes $|b - ax|$), and then $|b - ax| = 0$ (since $b - ax \in I$). Then $b = ax$, and hence $b \in (a)$. Therefore $I = (a)$.

A **Euclidean function** on an integral domain $R$ is a function $\partial$ from the ring to $\omega$ such that

$$\partial(x) = 0 \iff x = 0$$

and, for all $a$ in $R \smallsetminus \{0\}$ and $b$ in $R$, the inequality

$$\partial(b - ax) < \partial(a)$$

is soluble in $R$. Thus $x \mapsto |x|$ is a Euclidean function on $\mathbb{Z}$. Actually we need not require the range of a Euclidean function to be a subset of $\omega$; it could be any well-ordered set.

A **Euclidean domain** or ED is an integral domain with a Euclidean function. We now have:

**Theorem 205.** *Every* ED *is a* PID.

Other examples of Euclidean domains include the following.

For any field $K$, the function $f$ on $K$ given by

$$f(x) = \begin{cases} 1, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is a Euclidean function.

If $f$ is a polynomial $\sum_{i=0}^{m} a_i X^i$, where $a_m \neq 0$, then $m$ is $\deg(f)$, the *degree* of $f$. The function $f \mapsto \deg f$ on $K[X]$ will be Euclidean by Theorem 221 (page 215).

The **Gaussian integers** are the elements of $\mathbb{Z}[\sqrt{-1}]$, that is, the integers of $\mathbb{Q}(\sqrt{-1})$ (see §7.3, page 187). Writing $\mathrm{i}$ for $\sqrt{-1}$ as usual, we have that the norm function $z \mapsto |z|^2$ on $\mathbb{Z}[\mathrm{i}]$ is Euclidean, where

$$|x + y\,\mathrm{i}|^2 = x^2 + y^2.$$

Indeed, if $a \in \mathbb{Z}[\mathrm{i}] \smallsetminus \{0\}$ and $b \in \mathbb{Z}[\mathrm{i}]$, then $b/a$ is an element $s + t\,\mathrm{i}$ of $\mathbb{Q}(\mathrm{i})$. There are elements $x$ and $y$ of $\mathbb{Z}$ such that

$$|s - x| \leqslant \frac{1}{2}, \qquad |t - y| \leqslant \frac{1}{2}.$$

Let $q = x + y\,\mathrm{i}$; then

$$\left|\frac{b}{a} - q\right| = |s - x + (t - y)\,\mathrm{i}| \leqslant \frac{\sqrt{2}}{2} < 1$$

and so $|b - aq| < |a|$ (and hence $|b - aq|^2 < |a|^2$).

## 7.5. Localization

We shall now generalize the construction of $\mathbb{Q}$ from $\mathbb{Z}$ that is suggested by Theorem 30 (page 45). A nonempty subset of a ring is called **multiplicative** if it is closed under multiplication. For example, $\mathbb{Z} \smallsetminus \{0\}$ is a multiplicative subset of $\mathbb{Z}$, and more generally, the complement of any prime ideal of any ring is multiplicative.

**Lemma 21.** *If $S$ is a multiplicative subset of a ring $R$, then on $R \times S$ there is an equivalence-relation $\sim$ given by*

$$(a,b) \sim (c,d) \iff (ad - bc) \cdot e = 0 \text{ for some } e \text{ in } S. \tag{7.9}$$

*If $R$ is an integral domain and $0 \notin S$, then*

$$(a,b) \sim (c,d) \iff ad = bc.$$

*Proof.* Reflexivity and symmetry are obvious. For transitivity, note that, if $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$, so that, for some $g$ and $h$ in $S$,

$$0 = (ad - bc)g = adg - bcg, \qquad 0 = (cf - de)h = cfh - deh,$$

then

$$\begin{aligned}(af - be)cdgh &= afcdgh - becdgh \\ &= adgcfh - bcgdeh = bcgcfh - bcgcfh = 0,\end{aligned}$$

so $(a,b) \sim (e,f)$. □

In the notation of the lemma, the equivalence-class of $(a,b)$ is denoted by $a/b$ or

$$\frac{a}{b},$$

and the quotient $R \times S/{\sim}$ is denoted by

$$S^{-1}R.$$

If $0 \in S$, then $S^{-1}R$ has exactly one element. An instance where $R$ is not an integral domain will be considered in the next section (§7.6).

**Theorem 206.** *Suppose $R$ is a ring with multiplicative subset $S$.*

1. *In $S^{-1}R$, if $c \in S$,*
$$\frac{a}{b} = \frac{ac}{bc}.$$
2. *$S^{-1}R$ is a ring in which the operations are given by*
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \qquad \frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd}.$$
3. *There is a ring-homomorphism $\varphi$ from $R$ to $S^{-1}R$ where, for every $a$ in $S$,*
$$\varphi(x) = \frac{xa}{a}.$$
*If $1 \in S$, then $\varphi(x) = x/1$.*

*Suppose in particular $R$ is an integral domain and $0 \notin S$.*

4. *$S^{-1}R$ is an integral domain, and the homomorphism $\varphi$ is an embedding.*
5. *If $S = R \smallsetminus \{0\}$, then $S^{-1}R$ is a field, and if If $\psi$ is an embedding of $R$ in a field $K$, then there is an embedding $\tilde{\psi}$ of $S^{-1}R$ in $K$ such that $\tilde{\psi} \circ \varphi = \psi$.*

When $S$ is the complement of a prime ideal $\mathfrak{p}$, then $S^{-1}R$ is called the **localization** of $R$ at $\mathfrak{p}$ and can be denoted by
$$R_{\mathfrak{p}}.$$
(See Appendix A, page 222, for Fraktur letters like $\mathfrak{p}$.) If $R$ is an integral domain, so that $(0)$ is prime, then localization $R_{(0)}$ (which is a field by the theorem) is the **quotient-field** of $R$. In this case, the last part of the theorem describes the quotient field in terms of a *universal property* in the sense of page 122. However, it is important to note that, if $R$ is not an integral domain, then the homomorphism $x \mapsto x/1$ from $R$ to $R_{\mathfrak{p}}$ might not be an embedding. The following will be generalized as Theorem 214 (page 206 below).

**Theorem 207.** *For every Boolean ring $R$, for every prime ideal $\mathfrak{p}$ of $R$, the homomorphism $x \mapsto x/1$ from $R$ to $R_{\mathfrak{p}}$ is surjective and has kernel $\mathfrak{p}$. Thus*
$$\mathbb{F}_2 \cong R/\mathfrak{p} \cong R_{\mathfrak{p}}.$$

A **local ring** is a ring with a unique maximal ideal. The connection between localizations and local rings is made by the theorem below.

**Lemma 22.** *An ideal* $\mathfrak{m}$ *of a ring* $R$ *is a unique maximal ideal of* $R$ *if and only if*

$$R^{\times} = R \smallsetminus \mathfrak{m}.$$

**Theorem 208.** *The localization* $R_{\mathfrak{p}}$ *of a ring* $R$ *at a prime ideal* $\mathfrak{p}$ *is a local ring whose unique maximal ideal is*

$$\mathfrak{p}R_{\mathfrak{p}},$$

*namely the ideal generated by the image of* $\mathfrak{p}$.

*Proof.* The ideal $\mathfrak{p}R_{\mathfrak{p}}$ consists of those $a/b$ such that $a \in \mathfrak{p}$. In this case, if $c/d = a/b$, then $cb = da$, which is in $\mathfrak{p}$, so $c \in \mathfrak{p}$ since $\mathfrak{p}$ is prime and $b \notin \mathfrak{p}$. Hence for all $x/y$ in $R_{\mathfrak{p}}$,

$$\begin{aligned} x/y \notin R_{\mathfrak{p}}\mathfrak{p} &\iff x \notin \mathfrak{p} \\ &\iff x/y \text{ has an inverse, namely } y/x. \end{aligned}$$

By the lemma, we are done. □

## 7.6. *Ultraproducts of fields

An *ultraproduct* of fields is the quotient of the direct product of a family of fields by a maximal ideal. An algebraic investigation of this construction will involve maximal ideals, prime ideals, localizations, and our theorems about them. First we shall establish the usual tool by which the very existence of maximal ideals is established:

### 7.6.1. Zorn's Lemma

On page 14 we established a Recursion Theorem for $\mathbb{N}$ as an algebra, and hence for $\omega$. Now we establish another such theorem, for arbitrary ordinals, not just $\omega$; but the ordinals are now to be considered as well-ordered sets, not algebras.

**Theorem 209** (Transfinite Recursion)**.** *For all sets $A$, for all ordinals $\alpha$, for all functions $f$ from $\bigcup\{A^\beta : \beta < \alpha\}$ to $A$, there is a unique element*

$$(a_\beta : \beta < \alpha)$$

*of $A^\alpha$ such that, for all $\beta$ in $\alpha$,*

$$f(a_\gamma : \gamma < \beta) = a_\beta.$$

*Proof.* We first prove uniqueness. Suppose, if possible, $(a'_\beta : \beta < \alpha)$ is another element of $A^\alpha$ as desired, and let $\beta$ be minimal such that $a_\beta \neq a'_\beta$. Then

$$(a_\gamma : \gamma < \beta) = (a'_\gamma : \gamma < \beta),$$

so by definition $a_\beta = a'_\beta$. We now prove existence. If the theorem fails for some $\alpha$, let $\alpha$ be minimal such that it fails. Say $f : \bigcup\{A^\beta : \beta < \alpha\} \to A$. By hypothesis, for each $\beta$ in $\alpha$, there is a unique element $(a_\gamma : \gamma < \beta)$ of $A^\beta$ such that, for all $\gamma$ in $\beta$,

$$f(a_\delta : \delta < \gamma) = a_\gamma.$$

As before, $a_\gamma$ is independent of the choice of $\beta$ such that $\gamma < \beta < \alpha$. Then for all $\beta$ in $\alpha$ we are free to define

$$a_\beta = f(a_\gamma : \gamma < \beta).$$

Then the element $(a_\beta : \beta < \alpha)$ of $A^\alpha$ shows that the theorem does not fail for $\alpha$. □

Our proof used the method of the **minimal counterexample:** we showed that there could not be such a counterexample.

We now proceed to Zorn's Lemma. Suppose $\Omega$ is a set and $A \subseteq \mathscr{P}(\Omega)$. Then proper inclusion ($\subset$) is a transitive irreflexive relation on $A$ and on each of its subsets (see Theorems 18 and 19, page 36). A subset $C$ of $A$ is called a **chain** in $A$ if proper inclusion is also a total relation on $C$, so that $C$ is linearly ordered by proper inclusion (see Theorem 20). An **upper bound** of $C$ is a set that includes each element of $C$. In particular, $\bigcup C$ is an upper bound, and every upper bound includes this union. A **maximal element** of $A$ is an element that is not properly included in any other element.

The union of every chain of proper ideals of a ring is itself a proper ideal of the ring. A maximal ideal of the ring is more precisely a maximal element of the set of proper ideals of the ring. By the following, rings do have maximal ideals.

**Theorem 210** (Zorn's Lemma). *For all sets $\Omega$, for all subsets $A$ of $\mathscr{P}(\Omega)$, if $A$ contains an upper bound for each of its chains, then $A$ has a maximal element.*[3]

*Proof.* By the Axiom of Choice, there is a bijection $\alpha \mapsto B_\alpha$ from some cardinal $\kappa$ to $A$. By the Recursion Theorem, there is a function $\alpha \mapsto C_\alpha$ from $\kappa$ to $A$ such that, for all $\alpha$ in $\kappa$, if $\{C_\beta \colon \beta < \alpha\}$ is a chain, and if $\gamma$ is minimal such that $B_\gamma$ is an upper bound of this chain, then

$$C_\alpha = \begin{cases} B_\gamma, & \text{if } B_\gamma \nsubseteq B_\alpha, \\ B_\alpha, & \text{if } B_\gamma \subseteq B_\alpha; \end{cases}$$

in particular, $\{C_\beta \colon \beta \leqslant \alpha\}$ is a chain. If $\{C_\beta \colon \beta < \alpha\}$ is *not* a chain, then we can define $C_\alpha = B_0$. But we never have to do this: for all $\alpha$ in $\kappa$, the set $\{C_\beta \colon \beta < \alpha\}$ *is* a chain, since there can be no minimal counterexample to this assertion. Indeed, if $\alpha$ is minimal such that $\{C_\beta \colon \beta < \alpha\}$ is not a chain, there must be $\beta$ and $\gamma$ in $\alpha$ such that $\gamma < \beta$ and neither of $C_\beta$ and $C_\gamma$ includes the other. But by assumption $\{C_\delta \colon \delta < \beta\}$ is a chain, and so by definition $\{C_\delta \colon \delta \leqslant \beta\}$ is a chain, and in particular one of $C_\beta$ and $C_\gamma$ must include the other.

By a similar argument, $\{C_\alpha \colon \alpha < \kappa\}$ is a chain, so it has an upper bound $D$ in $A$. Suppose for some $\alpha$ we have $D \subseteq B_\alpha$. Then $C_\alpha = B_\alpha$, since otherwise, by definition, $C_\alpha = B_\gamma$ for some $\gamma$ such that $B_\gamma \nsubseteq B_\alpha$: in this

[3] In 1935, Zorn [40] presented this statement for the case where the upper bounds of the chains are actually the unions of the chains. He called the statement the "maximum principle" and suggested that using it would make proofs more algebraic than when the "well-ordering theorem" is used. Probably this theorem is what we have called the Axiom of Choice. Zorn promised to prove, in a later paper, that the maximum principle and the Axiom of Choice are equivalent; but it seems such a paper never appeared. Earlier, in 1922, Kuratowski [21, (42), p. 89] proved "Zorn's Lemma" for the case where the chains in question are well-ordered.

case $C_\alpha \nsubseteq B_\alpha$, so $C_\alpha \nsubseteq D$, which is absurd. Thus $C_\alpha = B_\alpha$, and hence $B_\alpha \subseteq D$, so $D = B_\alpha$. Therefore $D$ is a maximal element of $A$. □

As we said, it follows that rings have maximal ideals. We shall use Zorn's Lemma further to show that there are ideals that are maximal with respect to having certain properties. In our examples, these ideals will turn out to be prime.

### 7.6.2. Boolean rings

Recall that all rings now are commutative rings. For every such ring $R$, the set of its prime ideals is called its **spectrum** and can be denoted by

$$\operatorname{Spec}(R).$$

If $a \in R$, let us use the notation

$$[a] = \{\mathfrak{p} \in \operatorname{Spec}(R) \colon a \notin \mathfrak{p}\}.$$

**Theorem 211.** *For every ring $R$, for all $a$ and $b$ in $R$,*

$$[a] \cap [b] = [ab].$$

*Proof.* Since every $\mathfrak{p}$ in $\operatorname{Spec}(R)$ is prime, we have

$$\begin{aligned} \mathfrak{p} \in [a] \cap [b] &\iff a \notin \mathfrak{p} \ \ \& \ \ b \notin \mathfrak{p} \\ &\iff ab \notin \mathfrak{p} \\ &\iff \mathfrak{p} \in [ab]. \end{aligned}$$

□

As a consequence of the theorem, the spectrum of a ring can be given the **Zariski topology,** in which the sets $[a]$ are basic open sets. This topology is used in algebraic geometry, especially when the ring is one of the polynomial rings defined below in sub-§7.7.1. We are now interested in the case of Boolean rings. We showed in Theorem 191 that the power set of every set can be understood as a Boolean ring in which the operations are defined by

$$X \cdot Y = X \cap Y, \qquad\qquad X + Y = (X \smallsetminus Y) \cup (Y \smallsetminus X).$$

We may abbreviate $(X \smallsetminus Y) \cup (Y \smallsetminus X)$ by

$$X \vartriangle Y;$$

it is the **symmetric difference** of $X$ and $Y$. Immediately from the definition, every sub-ring of a Boolean ring is a Boolean ring. We now show that every Boolean ring embeds in a Boolean ring whose underlying set is the power set of some set. This is an analogue of Cayley's Theorem for groups (page 57) and Theorem 72 for associative rings (page 80).

**Theorem 212** (Stone [33]). *For every Boolean ring $R$, for all $a$ and $b$ in $R$,*

$$[a] \vartriangle [b] = [a+b],$$

*and the map $x \mapsto [x]$ is an embedding of $R$ in $\mathscr{P}(\operatorname{Spec}(R))$.*

*Proof.* By Theorem 192 (page 184), the characteristic of $R$ is at most 2, and so for all $a$ in $R$ we have

$$a \cdot (1+a) = 0.$$

Suppose $\mathfrak{p} \in \operatorname{Spec}(R)$. Since $\mathfrak{p}$ is prime and (like every ideal) contains 0, it must contain $a$ or $1+a$. If $\mathfrak{p}$ contains neither $a$ nor $b$, then it contains the sum of $1+a$ and $1+b$, which is $a+b$. Since the sum of any two elements of the subset $\{a, b, a+b\}$ of $R$ is equal to the third element, every $\mathfrak{p}$ in $\operatorname{Spec}(R)$ contains either one element or all elements of this set. Therefore

$$\begin{aligned}
\mathfrak{p} \in [a+b] &\iff a+b \notin \mathfrak{p} \\
&\iff (a \in \mathfrak{p} \Leftrightarrow b \notin \mathfrak{p}) \\
&\iff (\mathfrak{p} \notin [a] \Leftrightarrow \mathfrak{p} \in [b]) \\
&\iff \mathfrak{p} \in [a] \vartriangle [b].
\end{aligned}$$

By this and the previous theorem, $x \mapsto [x]$ is a homomorphism of Boolean rings. It remains to show that this homomorphism is injective. Say $x \in R \smallsetminus \{0\}$. The union of a chain of ideals of $R$ that do not contain $x$ is an ideal of $R$ that does not contain $x$. Therefore, by Zorn's Lemma, there is an ideal $\mathfrak{m}$ of $R$ that is maximal among those ideals that do not contain $x$. If $a$ and $b$ are not in $\mathfrak{m}$, then by maximality

$$x \in \mathfrak{m} + (a), \qquad\qquad x \in \mathfrak{m} + (b),$$

and therefore

$$x^2 \in \mathfrak{m} + (ab).$$

(We made a similar computation in proving the Chinese Remainder Theorem, page 142.) Since $x^2 \notin \mathfrak{m}$, we must have $ab \notin \mathfrak{m}$. Thus $\mathfrak{m}$ is prime, and so $\mathfrak{m} \in [x]$. In particular, $[x] \neq \varnothing$. □

Equipped with the Zariski topology, the spectrum of a Boolean ring is the **Stone space** of the ring.

### 7.6.3. Regular rings

The Boolean rings are members of a larger class of rings that satisfy the conclusion of Theorem 193 (page 184). We can establish this by first noting that, for every set $\Omega$, there is an isomorphism $U \mapsto \chi_U$ from the Boolean ring $\mathscr{P}(\Omega)$ to the direct power ${\mathbb{F}_2}^{\Omega}$, where

$$\chi_U(i) = \begin{cases} 1, & \text{if } i \in U, \\ 0, & \text{if } i \in \Omega \smallsetminus U. \end{cases}$$

Here $\chi_U$ can be called the **characteristic function** of $U$ (as a subset of $\Omega$). The power ${\mathbb{F}_2}^{\Omega}$ is a special case of the product $\prod_{i\in\Omega} K_i$, where each $K_i$ is a field. If $a \in \prod_{i\in\Omega} K_i$, there is an element $a^*$ of the product given by

$$\pi_i(a^*) = \begin{cases} \pi_i(a)^{-1}, & \text{if } \pi_i(a) \neq 0, \\ 0, & \text{if } \pi_i(a) = 0. \end{cases}$$

Then

$$aa^*a = a.$$

In particular, for every $x$ in the ring $\prod_{i\in\Omega} K_i$ there is $y$ in the ring such that

$$xyx = x.$$

Therefore the ring $\prod_{i\in\Omega} K_i$ is called a **(von Neumann) regular ring.**[4] Thus Boolean rings are also regular rings in this sense, since $xxx = x$ in a Boolean ring. A regular ring can also be understood as a ring in which

$$x \in (x^2)$$

[4] In general, a regular ring need not be commutative; see [19, IX.3, ex. 5, p. 442].

for all $x$ in the ring. The following generalizes Theorem 193 (page 184).

**Theorem 213.** *In regular rings, all prime ideals are maximal.*

*Proof.* If $R$ is a regular ring, and $\mathfrak{p}$ is a prime ideal, then for all $x$ in $R$, for some $y$ in $R$,

$$(xy-1)\cdot x=0,$$

and so at least one of $xy-1$ and $x$ is in $\mathfrak{p}$. Hence if $x+\mathfrak{p}$ is not 0 in $R/\mathfrak{p}$, then $x+\mathfrak{p}$ has the inverse $y+\mathfrak{p}$. Thus $R/\mathfrak{p}$ is a field, so $\mathfrak{p}$ is maximal. □

Now we can generalize Theorem 207 (page 199).

**Theorem 214.** *If $\mathfrak{p}$ is a prime ideal of a regular ring $R$, then there is a well-defined isomorphism*

$$x+\mathfrak{p}\mapsto x/1$$

*from $R/\mathfrak{p}$ to $R_{\mathfrak{p}}$.*

*Proof.* If $a\in R$ and $b\in R\smallsetminus\mathfrak{p}$, and $bcb=b$, then the elements $a/b$ and $ac/1$ of $R_{\mathfrak{p}}$ are equal since

$$(a-bac)b=ab-abcb=ab-ab=0.$$

Thus the homomorphism $x\mapsto x/1$ from $R$ to $R_{\mathfrak{p}}$ guaranteed by Theorem 206 is surjective. By the last theorem, $\mathfrak{p}$ is maximal, and hence $R_{\mathfrak{p}}$ is a field. In particular, if $x\in\mathfrak{p}$, then (as we showed in the proof) $x/1=0/1$, and so, if $y-z=x$, then $y/1=z/1$. Thus the epimorphism $x+\mathfrak{p}\mapsto x/1$ is well-defined. Its kernel then cannot be all of the field $R/\mathfrak{p}$, so the epimorphism must also be an embedding. □

The foregoing two theorems turn out to *characterize* regular rings:

**Theorem 215.** *The following are equivalent conditions on a ring $R$.*[5]

[5] The equivalence of these conditions is part of [11, Thm 1.16, p. 7]. This theorem adds a fourth equivalent condition: "All simple $R$-modules are injective." The proofs given involve module theory, except the proof that, if all prime ideals are maximal, and (7.10) holds, then each localization at a maximal ideal is a field. That proof is reproduced below.

1. *R is regular.*

2. *Every prime ideal of R is maximal, and in R,*

$$x^2 = 0 \implies x = 0. \tag{7.10}$$

3. *The localization $R_{\mathfrak{m}}$ is a field for all maximal ideals $\mathfrak{m}$ of R.*

*Proof.* 1. In regular rings, prime ideals are maximal by Theorem 213. Also, if $xyx = x$, but $x^2 = 0$, then $x = x^2y = 0$.

2. Now suppose every prime ideal of $R$ is maximal, and (7.10) holds. Let $\mathfrak{m}$ be a maximal ideal of $R$. By Theorem 208 (page 200), $\mathfrak{m}R_{\mathfrak{m}}$ is the unique maximal ideal of $R_{\mathfrak{m}}$. By Zorn's Lemma, every prime ideal $\mathfrak{P}$ of $R_{\mathfrak{m}}$ is included in a maximal ideal, which must be $\mathfrak{m}R_{\mathfrak{m}}$. Then $\mathfrak{P} \cap R$ is a prime ideal of $R$ that is included in $\mathfrak{m}$, so it *is* $\mathfrak{m}$, and hence $\mathfrak{P} = \mathfrak{m}R_{\mathfrak{m}}$. Thus this maximal ideal is the unique prime ideal of $R_{\mathfrak{m}}$. However, the set

$$\bigcup_{n\in\mathbb{N}} \{x \in R_{\mathfrak{m}} \colon x^n = 0\}$$

is also a prime ideal of $R_{\mathfrak{m}}$. Therefore it is equal to $\mathfrak{m}R_{\mathfrak{m}}$. Thus for all $r/s$ in $\mathfrak{m}R_{\mathfrak{m}}$, for some $n$ in $\mathbb{N}$, we have $(r/s)^n = 0$, so $r^n/s^n = 0$, and therefore $tr^n = 0$ for some $t$ in $\mathfrak{m}$. In this case, $(tr)^n = 0$, so $tr = 0$, and therefore $r/s = 0$. In short, $\mathfrak{m}R_{\mathfrak{m}} = (0)$. Therefore $R_{\mathfrak{m}}$ is a field.

3. Finally, suppose $R_{\mathfrak{m}}$ is a field for all maximal ideals $\mathfrak{m}$ of $R$. If $x \in R$, define

$$I = \{r \in R \colon rx \in (x^2)\}.$$

This is an ideal of $R$ containing $x$. We shall show that it contains 1. We do this by showing that it is not included in any maximal ideal $\mathfrak{m}$. If $x \notin \mathfrak{m}$, then $I \nsubseteq \mathfrak{m}$. If $x \in \mathfrak{m}$, then $x/1 \notin (R_{\mathfrak{m}})^{\times}$, so, since $R_{\mathfrak{m}}$ is a field, we have $x/1 = 0/1$, and hence

$$rx = 0$$

for some $r$ in $R \smallsetminus \mathfrak{m}$; but $r \in I$. Again $I \nsubseteq \mathfrak{m}$. Thus $I$ must be (1), so $x \in (x^2)$. Therefore $R$ is regular. □

We again consider the special case of a product $\prod \mathcal{K}$, where $\mathcal{K}$ is an indexed family $(K_i : i \in \Omega)$ of fields. Here $\prod \mathcal{K}$ is a regular ring, and $xx^*x = x$ when $x^*$ is defined as above. Hence every sub-ring of $\prod \mathcal{K}$ that is closed under the operation $x \mapsto x^*$ is also a regular ring. We now prove the converse: every regular ring is isomorphic to such a ring.

**Theorem 216.** *For every regular ring $R$, the homomorphism*

$$x \mapsto \big(x + \mathfrak{p} : \mathfrak{p} \in \operatorname{Spec}(R)\big)$$

*is an embedding of $R$ in the product*

$$\prod_{\mathfrak{p} \in \operatorname{Spec}(R)} R/\mathfrak{p}$$

*of fields. The image of this embedding is closed under $x \mapsto x^*$.*

*Proof.* The indicated map is an embedding, just as the map $x \mapsto [x]$ in Stone's Theorem (page 204) is an embedding. Indeed, suppose $R$ is a regular ring, and $x \in R \smallsetminus \{0\}$. Let $\mathfrak{m}$ be maximal among those ideals of $R$ that do not contain 0. If $a$ and $b$ are in $R \smallsetminus \mathfrak{m}$, then

$$\begin{gathered} x \in \big(\mathfrak{m} + (a)\big) \cap \big(\mathfrak{m} + (b)\big), \\ x^2 \in \mathfrak{m} + (ab), \\ x \in \mathfrak{m} + (ab), \end{gathered}$$

so $ab \notin \mathfrak{m}$. Thus $\mathfrak{m}$ is a prime ideal, and $x + \mathfrak{m} \neq 0$ in $R/\mathfrak{m}$. Therefore the map $x \mapsto \big(x + \mathfrak{p} : \mathfrak{p} \in \operatorname{Spec}(R)\big)$ is an embedding: call it $f$. For each $x$ in $R$, there is $y$ in $R$ such that $xyx = y$. For each $\mathfrak{p}$ in $\operatorname{Spec}(R)$, we have

$$\pi_{\mathfrak{p}}(f(x)) = x + \mathfrak{p}.$$

If this is not 0, then, since $f(x) = f(x)f(y)f(x)$, we have

$$y + \mathfrak{p} = (x + \mathfrak{p})^{-1} = \pi_{\mathfrak{p}}(f(x)^*).$$

However, possibly $x + \mathfrak{p} = 0$, while $y + \mathfrak{p} \neq 0$, so that $f(y) \neq f(x)^*$. In this case, letting $z = yxy$, we have

$$xzx = xyxyx = xyx = x, \qquad zxz = yxyxyxy = yxyxy = yxy = z.$$

In short, $xzx = x$ and $zxz = z$. Then

$$x \in \mathfrak{p} \iff z \in \mathfrak{p}, \qquad x \notin \mathfrak{p} \implies (z + \mathfrak{p})^{-1} = x + \mathfrak{p},$$

so $f(z) = f(x)^*$. □

### 7.6.4. Ultraproducts

If $R$ is a Boolean ring, then by Stone's Theorem (page 204), $R$ embeds in $\mathscr{P}(\mathrm{Spec}(R))$. We have also shown

$$\mathscr{P}(\mathrm{Spec}(R)) \cong {\mathbb{F}_2}^{\mathrm{Spec}(R)}.$$

Finally, for each $\mathfrak{p}$ in $\mathrm{Spec}(R)$, by Theorem 207 (page 199), the quotient $R/\mathfrak{p}$ is isomorphic to $\mathbb{F}_2$, and so

$${\mathbb{F}_2}^{\mathrm{Spec}(R)} \cong \prod_{\mathfrak{p} \in \mathrm{Spec}(R)} R/\mathfrak{p}.$$

In this way, Stone's Theorem becomes a special case of the foregoing theorem.

The field $\mathbb{F}_2$ can be considered as a sub*set* of each every field, although not a sub*field* (unless the field has characteristic 2). This observation gives rise to the following.

**Theorem 217.** *For every indexed family $(K_i : i \in \Omega)$ of fields, each ideal $I$ of $\prod_{i \in \Omega} K_i$ is generated by the set*

$$\{aa^* : a \in I\}.$$

*This set is itself an ideal, when considered as a subset of ${\mathbb{F}_2}^{\Omega}$. Hence the map $I \mapsto \{aa^* : a \in I\}$ is a bijection from the set of ideals of $\prod_{i \in \Omega} K_i$ to the set of ideals of ${\mathbb{F}_2}^{\Omega}$.*

*Proof.* We need only check that $\{aa^* : a \in I\}$ is closed under addition in ${\mathbb{F}_2}^{\Omega}$. If $a$ and $b$ are in $I$, then $aa^* = \chi_A$ and $bb^* = \chi_B$ for some subsets $A$ and $B$ of $\Omega$. In ${\mathbb{F}_2}^{\Omega}$, the sum $aa^* + bb^*$ is $\chi_{A \vartriangle B}$, which can be computed in $\prod_{i \in \Omega} K_i$ as

$$\chi_{A \vartriangle B} \cdot (a + b)(a + b)^*;$$

and this is in $I$. □

If $\mathcal{K}$ is an indexed family $(K_i : i \in \Omega)$ of fields, Let $\mathfrak{P}$ be a prime ideal of $\prod \mathcal{K}$. Then the quotient $\prod \mathcal{K}/\mathfrak{P}$ is a field, and this field is called an **ultraproduct** of $\mathcal{K}$. The ideal $\mathfrak{P}$ could be a principal ideal $(a)$. This ideal is equal to $(aa^*)$ and therefore to $(\chi_U)$ for some subset $U$ of $\Omega$. But $(a)$ is maximal, and therefore $U = \Omega \smallsetminus \{i\}$ for some $i$ in $\Omega$. In this case,

$$\prod \mathcal{K}/\mathfrak{P} \cong K_i.$$

However, if $\Omega$ is infinite, then $\mathscr{P}(\Omega)$ has the proper ideal $I$ consisting of the the finite subsets of $\Omega$. Then $\{\chi_U : U \in I\}$ generates a proper ideal of $\prod \mathcal{K}$. If $\mathfrak{P}$ includes this ideal, then $\mathfrak{P}$ is not principal, and the field $\prod \mathcal{K}/\mathfrak{P}$ is called a **nonprincipal ultraproduct** of $\mathcal{K}$. Such ideals $\mathfrak{P}$ exist by Zorn's Lemma.

If $a \in \prod \mathcal{K}$, the subset $\{i \in \Omega : a_i \neq 0\}$ of $\Omega$ can be called the **support** of $a$ and be denoted by

$$\operatorname{supp}(a).$$

In particular, $\operatorname{supp}(\chi_U) = U$. By the last theorem, we have a bijection

$$\mathfrak{P} \mapsto \{\operatorname{supp}(x) : x \in \mathfrak{P}\}$$

from $\operatorname{Spec}(\prod \mathcal{K})$ to $\operatorname{Spec}(\mathscr{P}(\Omega))$. Suppose the image of $\mathfrak{P}$ under this map is $\mathfrak{p}$. Then for all $a$ and $b$ in $\prod \mathcal{K}$ we have, *modulo* $\mathfrak{P}$,

$$a \equiv b \iff \{i \in \Omega : \pi_i(a) \neq \pi_i(b)\} \in \mathfrak{p}.$$

We may think of the elements of $\mathfrak{p}$ as "small" sets; their complements are "large." Then every subset of $\Omega$ is small or large. Two elements of $\prod \mathcal{K}$ are congruent *modulo* $\mathfrak{P}$ if and only if they agree on a large set of indices in $\Omega$. If $\mathfrak{P}$ is the principal ideal $(\Omega \smallsetminus \{i\})$, then the large subsets of $\Omega$ are just those that contain $i$.

Suppose however $\mathfrak{P}$ is nonprincipal. Then all finite subsets of $\Omega$ are small, and all *cofinite* subsets of $\Omega$ are large, and each map $x \mapsto \iota_i(x) + \mathfrak{P}$ from $K_i$ to $\prod \mathcal{K}/\mathfrak{P}$ is the zero map. Thus no one field $K_i$ affects the ultraproduct $\prod \mathcal{K}/\mathfrak{P}$. Rather, the ultraproduct is a kind of "average" of all of the fields $K_i$. Say for example $\Omega$ is the set of prime numbers in $\mathbb{N}$, and for each $p$ in $\Omega$, the field $K_p$ is $\mathbb{F}_p$. Then $\prod \mathcal{K}/\mathfrak{P}$ has characteristic 0, since for each prime $p$, the element $p \cdot 1$ of $\prod_{\ell \in \Omega} \mathbb{F}_\ell$ disagrees with 0 on a large set.

Since in general an ultraproduct $\prod_{i \in \Omega} K_i/\mathfrak{P}$ of fields depends only on $(K_i : i \in \Omega)$ and a prime ideal of $\mathscr{P}(\Omega)$, we can replace the fields $K_i$ with arbitrary structures (all having the same signature). The notion that a nonprincipal ultraproduct is an average of the factors is made precise by the result known as Łoś's Theorem, because it can be extracted from Łoś's 1955 paper [24]. The proof is straightforward, but requires careful attention to logic.

## 7.7. Polynomial rings

### 7.7.1. Universal property

Given a ring $R$, we defined the polynomial ring $R[X]$ on page 177 as the set of formal sums

$$\sum_{i<m} a_i X^i,$$

where $(a_i : i < m) \in R^m$, where $m \in \omega$. This means that, assuming $m \leqslant n$, we have

$$\sum_{i<m} a_i X^i = \sum_{i<n} b_i X^i$$
$$\iff (a_i : i < m) = (b_i : i < m) \ \& \ b_m = 0 \ \& \ \cdots \ \& \ b_{n-1} = 0.$$

We understand $\sum_{i<1} a_i X^i$ to be $a_0$, an element of $R$. Thus $R$ is included in $R[X]$.

We can now define the family of polynomial rings $R[X_0, \dots, X_{n-1}]$ recursively:

$$R[X_0, \dots, X_{n-1}] = \begin{cases} R, & \text{if } n = 0, \\ R[X_0, \dots, X_{k-1}][X_k], & \text{if } n = k+1. \end{cases}$$

These polynomial rings have a certain universal property in the sense of page 122:

**Theorem 218.** *For all rings $R$, for all $n$ in $\omega$, for all rings $S$, for all homomorphisms $\varphi$ from $R$ to $S$, for all $\boldsymbol{a}$ in $S^n$, there is a unique homomorphism $H$ from $R[X_0, \dots, X_{n-1}]$ to $S$ such that*

$$H \restriction R = \varphi, \qquad (H(X_i)\colon i < n) = \boldsymbol{a}.$$

*Proof.* We use induction. The claim is trivially true when $n = 0$. When $n = 1$, given $a$ in $S$, we must have $H \restriction A = \varphi$ and $H(X) = a$ and therefore

$$H\left(\sum_{k<m} b_k X^k\right) = \sum_{k<m} \varphi(b_k) \cdot a^k$$

for all $(b_i\colon i < m)$ in $R^m$, for all $m$ in $\omega$. Thus $H$ is determined on all of $R[X]$. The general inductive step follows in the same way. □

In the notation of the theorem, if $f \in R[X_0, \dots, X_{n-1}]$, then we may denote $H(f)$ by

$$f^{\varphi}(\boldsymbol{a}).$$

if also $\varphi = \mathrm{id}_R$, then $H(f)$ is just

$$f(\boldsymbol{a}).$$

Given a ring $R$, we can define a category (in the sense of §4.5, page 131) whose objects are pairs $(S, \varphi)$, where $S$ is a ring and $\varphi$ is a homomorphism from $R$ to $S$. If $(T, \psi)$ is also in the category, then a morphism from $(S, \varphi)$ to $(T, \psi)$ is a homomorphism $h$ from $S$ to $T$ such that $h \circ \varphi = \psi$.

$$\begin{tikzcd} & S \arrow[dd, "h"] \\ R \arrow[ur, "\varphi"] \arrow[dr, "\psi"'] & \\ & T \end{tikzcd}$$

Then for each $n$ in $\omega$, the pair $(R[X_0, \dots, X_{n-1}], \mathrm{id}_R)$ is an object in this category, and by the last theorem, in the sense of sub-§4.5.3 (page 136), it is a *free object* on $n$, with respect to the map $i \mapsto X_i$ on $n$. Then $R[X_0, \dots, X_{n-1}]$ is uniquely determined (up to isomorphism) by this property, by Theorem 130.

### 7.7.2. Division

If $R$ is a ring, and $f$ is the element $\sum_{i=0}^{n} a_i X^i$ of $R[X]$, and $a_n \neq 0$, then:

- $n$ is called the **degree** of $f$, and we may write
$$\deg(f) = n;$$
- each $a_i$ is a **coefficient** of $f$ and is *the* coefficient of $X^i$;
- $a_n$ is the **leading coefficient** of $f$;
- if this leading coefficient is 1, then $f$ is called **monic.**

We define also
$$\deg(0) = -\infty,$$
and for all $k$ in $\omega$,
$$-\infty + k = -\infty = k - \infty,$$
so that the next lemma makes sense in all cases. We said in §7.4 (page 197) that, if $K$ is a field, then $f \mapsto \deg(f)$ is a Euclidean function on $K[X]$. We now prove this.

**Lemma 23.** *Suppose $f$ and $g$ are polynomials in one variable $X$ over a ring $R$. then*
$$\deg(f + g) \leqslant \max(\deg f, \deg g),$$
*with equality if* $\deg(f) \neq \deg(g)$. *Also*
$$\deg(f \cdot g) \leqslant \deg f + \deg g,$$
*with equality if the product of the leading coefficients of $f$ and $g$ is not* 0. *In particular, if $R$ is an integral domain, then so is $R[X]$.*

**Theorem 219** (Division Algorithm)**.** *If $f$ and $g$ are polynomials in $X$ over a ring $R$, and the leading coefficient of $g$ is* 1*, then*
$$f = q \cdot g + r \tag{7.11}$$
*for some unique $q$ and $r$ in $R[X]$ such that* $\deg(r) < \deg(g)$.

*Proof.* To prove uniqueness, we note that if for each $i$ in 2 we have

$$f_i = q_i \cdot g + r_i,$$

where $q_0 \neq q_1$, and $\deg(r_0)$ and $\deg(r_1)$ are less than $\deg(g)$, then by the lemma

$$\deg(f_0 - f_1) = \deg\bigl((q_0 - q_1) \cdot g + r_0 - r_1\bigr) \geqslant \deg g \geqslant 0,$$

so $f_0 \neq f_1$. To prove existence, if $\deg(f) < \deg(g)$, we let $q = 0$. Suppose $\deg(g) \leqslant \deg(f)$. Given an arbitrary polynomial $h$ over $R$ with leading coefficient $a$ such that $\deg(g) \leqslant \deg(f)$, we define

$$h^* = h - aX^{\deg(h)-\deg(g)} \cdot g.$$

Then $\deg(h^*) < \deg(h)$ and

$$h = aX^{\deg(h)-\deg(g)} \cdot g + h^*.$$

Now define $f_0 = f$, and $f_1 = {f_0}^*$, and so on until $\deg(f_k) < \deg(g)$. Let $a_i$ be the leading coefficient of $f_i$, and let $n_i = \deg(f_i) - \deg(g)$. Then (7.11) holds when $r = f_k$ and

$$q = a_0 X^{n_0} + \cdots + a_{k-1} X^{n_{k-1}}.$$ □

**Corollary 219.1** (Remainder Theorem). *If $c \in R$ and $f \in R[X]$, then*

$$f = q \cdot (X - c) + f(c)$$

*for some unique $q$ in $R[X]$.*

*Proof.* By the Division Algorithm, $f = q \cdot (X - c) + d$ for some unique $q$ in $R[X]$ and $d$ in $R$. Then $f(c) = q(c) \cdot (c - c) + d = d$. □

If $f(c) = 0$, then $c$ is a **zero** of $f$.

**Theorem 220.** *For every polynomial $f$ over a ring, for every $c$ in the ring,*

$$f(c) = 0 \iff (X - c) \mid f.$$

*If the ring an integral domain, and $f \neq 0$, then the number of distinct zeros of $f$ is at most* $\deg(f)$.

*Proof.* By the Remainder Theorem, $c$ is a zero of $f$ if and only if $f = q \cdot (X - c)$ for some $q$. In this case, if the ring is an integral domain, and $d$ is another zero of $f$, then, since $d - c \neq 0$, we must have that $d$ is a zero of $q$. Hence, if $\deg(f) = n$, and $f$ has the distinct zeros $r_0, \dots, r_{m-1}$, then repeated application of the Remainder Theorem yields

$$f = q \cdot (X - r_0) \cdots (X - r_{m-1})$$

for some $q$. If $f \neq 0$, then $q \neq 0$, and $\deg(f) \geqslant m$. □

Recall however from the proof of Theorem 193 (page 184) that every element of a Boolean ring is a zero of $X \cdot (1 + X)$, that is, $X + X^2$; but some Boolean rings have more than two elements. In $\mathbb{Z}_6$, the same polynomial $X + X^2$ has the zeros 0, 2, 3, and 5.

**Theorem 221.** *If $K$ is a field, then $f \mapsto \deg(f)$ is a Euclidean function on $K[X]$.*

*Proof.* Over a field, the Division Algorithm does not require the leading coefficient of the divisor to be 1. □

Thus for all fields $K$, the ring $K[X]$ is a ED, therefore a PID, therefore a UFD.

### 7.7.3. *Multiple zeros

A zero $c$ of a polynomial over an integral domain has **multiplicity** $m$ if the polynomial is a product $g \cdot (X - c)^m$, where $c$ is not a zero of $g$. A zero with multiplicity greater than 1 is a **multiple** zero. Derivations were defined on page 175; they will be useful for recognizing the existence of multiple roots.

**Lemma 24.** *If $\delta$ is a derivation of a ring $R$, then for all $x$ in $R$ and $n$ in $\omega$,*

$$\delta(x^n) = nx^{n-1} \cdot \delta(x).$$

*Proof.* Since

$$\delta(1) = \delta(1 \cdot 1) = \delta(1) \cdot 1 + 1 \cdot \delta(1) = 2 \cdot \delta(1),$$

we have $\delta(1) = 0$, so the claim holds when $n = 0$. If it holds when $n = k$, then

$$\begin{aligned}\delta(x^{k+1}) &= \delta(x) \cdot x^k + x \cdot \delta(x^k)\\ &= \delta(x) \cdot x^k + kx^k \cdot \delta(x) = (k+1) \cdot x^k \cdot \delta(x),\end{aligned}$$

so the claim holds when $n = k + 1$. □

**Theorem 222.** *On a polynomial ring $R[X]$, there is a unique derivation $f \mapsto f'$ such that*

$$X' = 1, \qquad\qquad c' = 0$$

*for all $c$ in $R$. This derivation is given by*

$$\left(\sum_{k=0}^{n} a_k X^k\right)' = \sum_{k=0}^{n-1} (k+1) \cdot a_{k+1} X^k. \tag{7.12}$$

*Proof.* Let $\delta$ be the operation $f \mapsto f'$ on $K[X]$ defined by (7.12). By the lemma and the definition of a derivation, $\delta$ is the only operation that can meet the desired conditions. It remains to show that $\delta$ is indeed a derivation. We have

$$\delta\left(\sum_{k=0}^{n} a_k X^k\right) = \sum_{k=0}^{n} a_k \cdot \delta(X^k).$$

Also

$$\begin{aligned}\delta(X^k X^\ell) &= \delta(X^{k+\ell}) = (k+\ell) \cdot X^{k+\ell-1}\\ &= kX^{k-1}X^\ell + \ell X^k X^{\ell-1} = \delta(X^k) \cdot X^\ell + X^k \cdot \delta(X^\ell).\end{aligned}$$

Therefore $\delta$ is indeed a derivation:

$$\begin{aligned}
&\delta\Biggl(\sum_{k<m} a_k X^k \cdot \sum_{\ell<n} b_\ell X^\ell\Biggr)\\
&= \delta\Biggl(\sum_{k<m}\sum_{\ell<n} a_k X^k \cdot b_\ell X^\ell\Biggr)\\
&= \sum_{k<m}\sum_{\ell<n} a_k b_\ell \cdot \delta(X^k X^\ell)\\
&= \sum_{k<m}\sum_{\ell<n} a_k b_\ell \cdot \bigl(\delta(X^k)\cdot X^\ell + X^k \cdot \delta(X^\ell)\bigr)\\
&= \sum_{k<m}\sum_{\ell<n}\bigl(a_k \cdot \delta(X^k)\cdot b_\ell X^\ell + a_k X^k \cdot b_\ell \cdot \delta(X^\ell)\bigr)\\
&= \sum_{k<m} a_k \cdot \delta(X^k) \cdot \sum_{\ell<n} b_\ell X^\ell + \sum_{k<m} a_k X^k \cdot \sum_{\ell<n} b_\ell \cdot \delta(X^\ell)\\
&= \delta\Biggl(\sum_{k<m} a_k X^k\Biggr) \cdot \sum_{\ell<n} b_\ell X^\ell + \sum_{k<m} a_k X^k \cdot \delta\Biggl(\sum_{\ell<n} b_\ell X^\ell\Biggr). \qquad \square
\end{aligned}$$

In the notation of the theorem, $f'$ is the **derivative** of $f$.

**Lemma 25.** *Let $R$ be an integral domain, and suppose $f \in R[X]$ and $f(c) = 0$. Then $c$ is a multiple zero of $f$ if and only if*

$$f'(c) = 0.$$

*Proof.* Write $f$ as $(X-c)^m \cdot g$, where $g(c) \neq 0$. Then $m \geqslant 1$, so

$$f' = m \cdot (X-c)^{m-1} \cdot g + (X-c)^m \cdot g'.$$

If $m > 1$, then $f'(c) = 0$. If $f'(c) = 0$, then $m \cdot 0^{m-1} \cdot g(c) = 0$, so $0^{m-1} = 0$ and hence $m > 1$. $\square$

If $L$ is a field with subfield $K$, then a polynomial over $K$ may be irreducible over $K$, but not over $L$. For example, $X^2+1$ is irreducible over $\mathbb{Q}$, but not over $\mathbb{Q}(\mathrm{i})$. Likewise, the polynomial may have zeros from $L$, but not $K$. Hence it makes sense to speak of zeros of an irreducible polynomial.

**Theorem 223.** *If $f$ is an irreducible polynomial with multiple zeros over a field $K$, then $K$ has characteristic $p$ for some prime number $p$, and*

$$f = g(X^p)$$

*for some polynomial $g$ over $K$.*

*Proof.* If $f$ has the multiple zero $c$, then by the lemma $X - c$ is a common factor of $f$ and $f'$. Since $f$ is irreducible, itself must be a common factor of $f$ and $f'$, so $f'$ can only be 0, since $\deg(f') < \deg(f)$. Say $f = \sum_{k=0}^{n} a_k X^k$, so $f' = \sum_{k=0}^{n-1}(k+1) \cdot a_{k+1} X^k$. If $f' = 0$, but $a_{k+1} \neq 0$, then $k + 1$ must be 0 in $K$, that is, its image under the homomorphism from $\mathbb{Z}$ to $K$ must be 0. Then this homomorphism has a kernel $\langle p \rangle$ for some prime number $p$. Hence $a_k = 0$ whenever $p \nmid k$, so $f$ can be written as $\sum_{j=0}^{m} a_{pj} X^{pj}$, which is $g(X^p)$, where $g = \sum_{j=0}^{m} a_{pj} X^j$. □

### 7.7.4. Factorization

Throughout this subsection, $R$ is a UFD with quotient field $K$. We know from Theorem 221 that $K[X]$ is a Euclidean domain and therefore a UFD. Now we shall show that $R[X]$ too is a UFD. It will then follow that each of the polynomial rings $R[X_0, \dots, X_{n-1}]$ is a UFD.

A polynomial over $R$ is called **primitive** if 1 is a greatest common divisor of its coefficients. Gauss proved a version of the following for the case where $R$ is $\mathbb{Z}$ [9, ¶42].

**Lemma 26** (Gauss). *The product of primitive polynomials over $R$ is primitive.*

*Proof.* Let $f = \sum_{k=0}^{m} a_k X^k$ and $g = \sum_{k=0}^{n} b_k X^k$. Then

$$fg = \sum_{k=0}^{mn} c_k X^k,$$

where

$$c_k = \sum_{i+j=k} a_i b_j = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0.$$

Suppose $f$ is primitive, but $fg$ is not, so the coefficients $c_k$ have a common prime factor $\pi$. There is some $\ell$ such that $\pi \mid a_i$ when $i < \ell$, but $\pi \nmid a_\ell$. Then $\pi$ divides

$$c_\ell - (a_0 b_\ell + \cdots + a_{\ell-1} b_1),$$

which is $a_\ell b_0$, so $\pi \mid b_0$. Hence $\pi$ divides

$$c_{\ell+1} - (a_0 b_{\ell+1} + \cdots + a_{\ell-1} b_2) - a_{\ell+1} b_0,$$

which is $a_\ell b_1$, so $\pi \mid b_1$, and so on. Thus $g$ is not primitive. □

**Lemma 27.** *Primitive polynomials over $R$ that are associates over $K$ are associates over $R$.*

*Proof.* Suppose $f$ and $g$ are polynomials that are defined over $R$ and are associates over $K$. Then $uf = g$ for some $u$ in $K^\times$, and consequently $bu = a$ for some $a$ and $b$ in $R$, so $af = bg$. If $f$ and $g$ are primitive, then $a$ and $b$ must be associates in $R$, and therefore $u \in R^\times$, so $f$ and $g$ are associates over $R$. □

**Lemma 28.** *Primitive polynomials over $R$ are irreducible over $R$ if and only if they are irreducible over $K$.*

*Proof.* Suppose $f$ and $g$ are polynomials over $K$ such that the product $fg$ is a primitive polynomial over $R$. For some $a$ and $b$ in $K$, the polynomials $af$ and $bg$ have coefficients in $R$ and are primitive over $R$. By Gauss's Lemma, $abfg$ is primitive. Since $fg$ is already primitive, $ab$ must be a unit in $R$. In particular, $abu = 1$ for some $u$ in $R^\times$. Then $af$ and $bug$ are primitive polynomials over $R$ whose product is $fg$.

Now, the units of $K[X]$ are just the polynomials of degree 0, that is, the elements of $K^\times$. In particular,

$$f \in K[X]^\times \iff af \in K[X]^\times.$$

The unit *primitive* elements of $R[X]$ are the elements of $R^\times$. Thus

$$af \in K[X]^\times \iff af \in R[X]^\times.$$

Therefore $fg$ is irreducible over $K$ if and only if over $R$. □

Note however that if $f$ is primitive and irreducible over $R$, and $a$ in $R$ is not a unit or 0, then $af$ is still irreducible over $K$ (since $a$ is a unit in $K$) but not over $R$.

**Theorem 224.** *$R[X]$ is a* UFD.

*Proof.* Every nonzero element of $R[X]$ can be written as $af$, where $a \in R \smallsetminus \{0\}$ and $f$ is primitive. Then $f$ has a prime factorization over $K$ (since $K[X]$ is a Euclidean domain): say $f = f_0 \cdots f_{n-1}$. There are $b_k$ in $K$ such that $b_k f_k$ is a primitive polynomial over $R$. The product of these is still primitive by Gauss's Lemma, so the product of the $b_k$ must be a unit in $R$. We may assume this unit is 1. Thus $f$ has an irreducible factorization

$$(b_0 f_0) \cdots (b_{n-1} f_{n-1})$$

over $R$. Its uniqueness follows from its uniqueness over $K$ and Lemma 27. Since $a$ has a unique irreducible factorization $a_0 \cdots a_{m-1}$, we obtain a unique irreducible factorization of $af$. □

We end with a test for irreducibility.

**Theorem 225** (Eisenstein's Criterion)**.** *If $\pi$ is an irreducible element of $R$ and $f$ is a polynomial*

$$a_0 + a_1 X + \cdots + a_n X^n$$

*over $R$ such that*

$$\pi^2 \nmid a_0, \qquad \pi \mid a_0, \qquad \pi \mid a_1, \qquad \ldots, \qquad \pi \mid a_{n-1}, \qquad \pi \nmid a_n,$$

*then $f$ is irreducible over $K$ and, if primitive, over $R$.*

*Proof.* Suppose $f = gh$, where

$$g = \sum_{k=0}^{n} b_k X^k, \qquad\qquad h = \sum_{k=0}^{n} c_k X^k,$$

all coefficients being from $R$. We may assume $f$ is primitive, so $g$ and $h$ must be primitive. We may assume $\pi$ divides $b_0$, but not $c_0$. Let $\ell$ be such that $\pi \mid b_k$ when $k < \ell$. If $\ell = n$, then (since $g$ is primitive) we must have $b_n \neq 0$, so $\deg(g) = n$. In this case $\deg(h) = 0$, so $h$ is a unit. If $\ell < n$, then, since $\pi \mid a_\ell$, but

$$a_\ell = b_0 c_\ell + b_1 c_{\ell-1} + \cdots + b_\ell c_0,$$

we have $\pi \mid b_\ell$. By induction, $\pi \mid b_k$ whenever $k < n$, so as before $\deg(g) = n$. $\square$

An application is the following.

**Theorem 226.** *If $p$ is a prime number, then the polynomial*

$$1 + X + \cdots + X^{p-1}$$

*is irreducible.*

*Proof.* It is enough to establish the irreducibility of $\sum_{k=0}^{p-1}(X+1)^k$. We have

$$\sum_{k=0}^{p-1}(X+1)^k = \sum_{k=0}^{p-1}\sum_{j=0}^{k}\binom{k}{j}X^j = \sum_{j=0}^{p-1}X^j\sum_{k=j}^{p-1}\binom{k}{j} = \sum_{j=0}^{p-1}X^j\binom{p}{j+1},$$

which meets the Eisenstein Criterion since

$$\binom{p}{1} = p, \qquad \binom{p}{j+1} = \frac{p!}{(p-j-1)!(j+1)!},$$

which is divisible by $p$ if and only if $j < p - 1$. $\square$

# A. The German script

In his encyclopedic *Model Theory* of 1993, Wilfrid Hodges observes [17, Ch. 1, p. 21]:

> Until about a dozen years ago, most model theorists named structures in horrible Fraktur lettering. Recent writers sometimes adopt a notation according to which all structures are named $M$, $M'$, $M^*$, $\bar{M}$, $M_0$, $M_i$ or occasionally $N$. I hope I cause no offence by using a more freewheeling notation.

For Hodges, *structures* (as defined in §1.6 on page 40 above) are denoted by the letters $A$, $B$, $C$, and so forth; he refers to their universes as **domains** and denotes these by $\mathrm{dom}(A)$ and so forth. This practice is convenient if one is using a typewriter (as in the preparation of another of Hodges's books [18], from 1985). In his *Model Theory: An Introduction* of 2002, David Marker [26] uses "calligraphic" letters to denote structures, as distinct from their universes: so $M$ is the universe of $\mathcal{M}$, and $N$ of $\mathcal{N}$. I still prefer the older practice of using capital Fraktur letters for structures:

$$\begin{array}{ccccccccccccc} \mathfrak{A} & \mathfrak{B} & \mathfrak{C} & \mathfrak{D} & \mathfrak{E} & \mathfrak{F} & \mathfrak{G} & \mathfrak{H} & \mathfrak{I} & \mathfrak{J} & \mathfrak{K} & \mathfrak{L} & \mathfrak{M} \\ \mathfrak{N} & \mathfrak{O} & \mathfrak{P} & \mathfrak{Q} & \mathfrak{R} & \mathfrak{S} & \mathfrak{T} & \mathfrak{U} & \mathfrak{V} & \mathfrak{W} & \mathfrak{X} & \mathfrak{Y} & \mathfrak{Z} \end{array}$$

For the record, here are the minuscule Fraktur letters, which are used in this text, starting on page 199, for denoting ideals:

$$\begin{array}{ccccccccccccc} \mathfrak{a} & \mathfrak{b} & \mathfrak{c} & \mathfrak{d} & \mathfrak{e} & \mathfrak{f} & \mathfrak{g} & \mathfrak{h} & \mathfrak{i} & \mathfrak{j} & \mathfrak{k} & \mathfrak{l} & \mathfrak{m} \\ \mathfrak{n} & \mathfrak{o} & \mathfrak{p} & \mathfrak{q} & \mathfrak{r} & \mathfrak{s} & \mathfrak{t} & \mathfrak{u} & \mathfrak{v} & \mathfrak{w} & \mathfrak{x} & \mathfrak{y} & \mathfrak{z} \end{array}$$

A way to write these letters by hand is seen on the page reproduced below from a 1931 textbook [15] on the German language:


Figure A.1. The German alphabet

# Bibliography


[1] Cesare Burali-Forti. A question on transfinite numbers. In van Heijenoort [36], pages 104–12. First published 1897.

[2] Josephine E. Burns. The Foundation Period in the History of Group Theory. *Amer. Math. Monthly*, 20(5):141–148, 1913.

[3] Chen Chung Chang. On unions of chains of models. *Proc. Amer. Math. Soc.*, 10:120–127, 1959.

[4] Harvey Cohn. *Advanced Number Theory.* Dover, New York, 1980. Corrected republication of 1962 edition.

[5] Richard Dedekind. *Essays on the theory of numbers. I: Continuity and irrational numbers. II: The nature and meaning of numbers.* authorized translation by Wooster Woodruff Beman. Dover Publications Inc., New York, 1963.

[6] Euclid. *Euclidis Elementa*, volume I of *Euclidis Opera Omnia.* Teubner, 1883. Edidit et Latine interpretatvs est I. L. Heiberg.

[7] Euclid. *The thirteen books of Euclid's Elements translated from the text of Heiberg. Vol. I: Introduction and Books I, II. Vol. II: Books III–IX. Vol. III: Books X–XIII and Appendix.* Dover Publications Inc., New York, 1956. Translated with introduction and commentary by Thomas L. Heath, 2nd ed.

[8] Euclid. *Euclid's Elements.* Green Lion Press, Santa Fe, NM, 2002. All thirteen books complete in one volume. The Thomas L. Heath translation, edited by Dana Densmore.

[9] Carl Friedrich Gauss. *Disquisitiones Arithmeticae.* Springer-Verlag, New York, 1986. Translated into English by Arthur A. Clarke, revised by William C. Waterhouse.

[10] Carolo Friderico Gauss. *Disquisitiones Arithmeticae.* Gerh. Fleischer Jun., Lipsiae, 1801. Electronic version of the original Latin text from Goettingen State and University Library.

[11] K. R. Goodearl. *von Neumann regular rings*, volume 4 of *Monographs and Studies in Mathematics.* Pitman (Advanced Publishing Program), Boston, Mass., 1979.

[12] Timothy Gowers, June Barrow-Green, and Imre Leader, editors. *The Princeton companion to mathematics.* Princeton University Press, Princeton, NJ, 2008.

[13] Joel David Hamkins. Every group has a terminating transfinite automorphism tower. *Proc. Amer. Math. Soc.*, 126(11):3223–3226, 1998.

[14] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers.* The Clarendon Press Oxford University Press, New York, fifth edition, 1979.

[15] Roe-Merrill S. Heffner. *Brief German Grammar.* D. C. Heath and Company, Boston, 1931.

[16] Leon Henkin. On mathematical induction. *Amer. Math. Monthly*, 67:323–338, 1960.

[17] Wilfrid Hodges. *Model theory*, volume 42 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1993.

[18] Wilfrid Hodges. *Building models by games.* Dover Publications, Mineola, New York, 2006. Original publication, 1985.

[19] Thomas W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1980. Reprint of the 1974 original.

[20] Morris Kline. *Mathematical thought from ancient to modern times.* Oxford University Press, New York, 1972.

[21] Casimir Kuratowski. Une méthode d'élimination des nombres transfinis des raisonnements mathématiques. *Fundamenta Mathematicae*, 3(1):76–108, 1922.

[22] Edmund Landau. *Foundations of Analysis. The Arithmetic of Whole, Rational, Irrational and Complex Numbers.* Chelsea Publishing Company, New York, N.Y., third edition, 1966. Translated by F. Steinhardt; first edition 1951; first German publication, 1929.

[23] Serge Lang. *Algebra.* Addison-Wesley, Reading, Massachusetts, third edition, 1993. Reprinted with corrections, 1997.

[24] Jerzy Łoś. Quelques remarques, théorèmes et problèmes sur les classes définissables d'algèbres. In *Mathematical interpretation of formal systems*, pages 98–113. North-Holland Publishing Co., Amsterdam, 1955.

[25] Jerzy Łoś and Roman Suszko. On the extending of models (IV): Infinite sums of models. *Fund. Math.*, 44:52–60, 1957.

[26] David Marker. *Model theory: an introduction*, volume 217 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 2002.

[27] James H. McKay. Another proof of Cauchy's group theorem. *Amer. Math. Monthly*, 66:119, 1959.

[28] Frank Mittelbach and Michel Goossens. *The LaTeX Companion.* Addison Wesley, Boston, second edition, August 2004. With Johannes Braams, David Carlisle, and Chris Rowley; second printing (with corrections).

[29] Giuseppe Peano. The principles of arithmetic, presented by a new method. In van Heijenoort [36], pages 83–97. First published 1889.

[30] Proclus. *A commentary on the first book of Euclid's* Elements. Princeton Paperbacks. Princeton University Press, Princeton, NJ, 1992. Translated from the Greek and with an introduction and notes by Glenn R. Morrow, reprint of the 1970 edition, with a foreword by Ian Mueller.

[31] Bertrand Russell. Letter to Frege. In van Heijenoort [36], pages 124–5. First published 1902.

[32] Thoralf Skolem. Some remarks on axiomatized set theory. In van Heijenoort [36], pages 290–301. First published 1922.

[33] M. H. Stone. The theory of representations for Boolean algebras. *Trans. Amer. Math. Soc.*, 40(1):37–111, 1936.

[34] Ivor Thomas, editor. *Selections illustrating the history of Greek mathematics. Vol. II. From Aristarchus to Pappus*, volume 362 of *Loeb Classical Library*. Harvard University Press, Cambridge, Mass, 1951. With an English translation by the editor.

[35] Simon Thomas. The automorphism tower problem. *Proc. Amer. Math. Soc.*, 95(2):166–168, 1985.

[36] Jean van Heijenoort, editor. *From Frege to Gödel: A source book in mathematical logic, 1879–1931.* Harvard University Press, Cambridge, MA, 2002.

[37] John von Neumann. An axiomatization of set theory. In van Heijenoort [36], pages 393–413. First published 1925.

[38] John von Neumann. On the introduction of transfinite numbers. In van Heijenoort [36], pages 346–354. First published 1923.

[39] Ernst Zermelo. Investigations in the foundations of set theory I. In van Heijenoort [36], pages 199–215. First published 1908.

[40] Max Zorn. A remark on method in transfinite algebra. *Bull. Amer. Math. Soc.*, 41(10):667–670, 1935.