Groups and Rings

David Pierce

September 30, 2013

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

I wrote the first draft of these notes during a graduate course in algebra at METU in Ankara in 2008–9. I had taught this course also in 2003–4. I revised my notes when teaching the course a third time, in 2009-10.

Section 0.3 (p. 16) is based on part of a course called Non-Standard Analysis, which I gave at the Nesin Mathematics Village, Şirince, in the summer of 2009. I built up Chapter 0 around this section.

For the remaining chapters, the main reference is Hungerford's *Algebra* [8]. This was the suggested text for the course at METU, as well as for the algebra course that I myself took as a graduate student.

Hungerford is inspired by category theory, of which his teacher Saunders Mac Lane was one of the creators. (See §3.3, p. 79 below.) The spirit of category theory is seen for example 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. Inspired myself by model theory, I shall use the term *structure* instead. (See §0.5, p. 24 below.) The objects named here by Hungerford are all structures in the sense of model theory, although not every object in a category is a structure in this sense.

Contents

Note to the reader

Every theorem must have a proof. Some proofs in the present notes are sketchy, if not missing entirely. In such cases, details should be supplied by the reader. No theorem here is expected to be taken on faith. However, for the purposes of an algebra course, some proofs are more important than others. The full development of Chapter 0 would take a course in itself, but is not required for algebra as such.

The material here is taken mainly from Hungerford [8], 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.

0. Mathematical foundations

The full details of this chapter are not strictly part of an algebra course, but are logically presupposed by the course. The main 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[1] of $\upomega$ are the von-Neumann natural numbers, 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 $\{1, \dots, n\}$, which might be denoted by something like $[n]$. This is a needless complication.

0.1. Sets and classes

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 the collection **contains** it.

[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, the $\upomega$ used here is upright, unlike the standard slanted ω (obtained with `\omega`). The latter ω might be used as a variable, although it is not so used in these notes. One could similarly distinguish between the constant $\uppi$ (used for the ratio of the circumference to the diameter of a circle) and the variable π.

[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.

A **set** is a special kind of collection. The properties of sets are given by *axioms*; we shall use a version of the Zermelo–Fraenkel Axioms with the Axiom of Choice [22]. The collection of these axioms is denoted by ZFC. In the logical formalism that we shall use for the these axioms, every element of a set is itself a set. By definition, two sets are **equal** if they have the same elements. There is an empty set—a set with no members—, denoted by $\varnothing$. If a is a set, then there is a set $\{a\}$, with the unique element a. If b is also a set, then there is a set $a \cup b$, whose members are precisely the members of a and the members of b. Thus there are 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 allows us to build up the following infinite sequence:

$$\varnothing, \qquad \{\varnothing\}, \qquad \{\varnothing, \{\varnothing\}\}, \qquad \Bigl\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\Bigr\}, \qquad \dots$$

By definition, these sets are the natural numbers 0, 1, 2, 3, …

As we shall understand them, the ZFC axioms are written in a certain *logic,* whose symbols are:

1) variables, as x, y, and z;
2) the symbol $\in$ denoting the membership relation, so that $x \in y$ means x is a member of y;
3) the Boolean connectives of propositional logic: $\lor$ ("or"), $\land$ ("and"), $\Rightarrow$ ("implies"), $\Leftrightarrow$ ("if and only if"), and $\neg$ ("not");
4) parentheses or brackets;
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 an *atomic formula.* From atomic formulas, other formulas are built up *recursively* by use of the symbols above, according to certain rules. For example, $\neg\, t \in u$ is the formula saying that t is *not* a member of u. We usually abbreviate this formula by

$$t \notin u.$$

Now we can write the **Empty Set Axiom:**

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

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]

$$x \subseteq y.$$

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 given above; in this case we use the abbreviation

$$x = y.$$

Some occurrences of a variable in a formula are *bound.*[4] In particular, if φ is a formula, then so are $\exists x\ \varphi$ and $\forall x\ \varphi$, but all occurrences of x in these two formulas are bound. Occurrences of a variable that are not bound are *free.* If φ is a formula in which only x occurs freely, we may write φ as $\varphi(x)$. If a is a set, then by replacing every free occurrence of x in φ with a, we obtain the formula $\varphi(a)$, which is 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. There is a collection of all sets that satisfy φ. 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 φ.

The definition of equality can also be expressed by the following sentences:

$$\forall x\ \forall y\ \forall z\ \bigl(x = y \Rightarrow (z \in x \Leftrightarrow z \in y)\bigr), \tag{0.1}$$

$$\forall x\ \forall y\ \bigl(\forall z\ (z \in x \Leftrightarrow z \in y) \Rightarrow x = y\bigr). \tag{0.2}$$

[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.

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

That equal sets belong to the same sets is the **Equality Axiom:**

$$\forall x\,\forall y\,\forall z\,\bigl(x = y \Rightarrow (x \in z \Leftrightarrow y \in z)\bigr). \tag{0.3}$$

The meaning of the sentences (0.1) and (0.3) is that equal sets satisfy the same atomic formulas, be they of the form $x \in a$ or $a \in x$. It is then a theorem that equal sets satisfy the same formulas in general:

$$\forall x\,\forall y\,\Bigl(x = y \Rightarrow \bigl(\varphi(x) \Leftrightarrow \varphi(y)\bigr)\Bigr). \tag{0.4}$$

The theorem is proved by *induction* on the complexity of formulas. Such a proof is possible because formulas are defined recursively. See §0.3 below.

It is more usual to take the sentence (0.4) as a logical axiom, of which (0.1) and (0.3) are special cases; but then (0.2) is no longer true by definition or by proof, but must be taken as an axiom, which is called the **Extension Axiom.** The idea behind the name is that having the same members means having the same *extension.*

In any case, all of the sentences (0.1), (0.2), (0.3), and (0.4) 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.

With this understanding, we obtain the sequence 0, 1, 2, ..., described above by starting with the Empty Set Axiom and continuing with the **Adjunction Axiom:**

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

In fact this is not one of Zermelo's original axioms of 1908. It and the Empty Set Axiom have as a consequence

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

This is usually called the **Pairing Axiom** and is one of Zermelo's original axioms. More precisely, Zermelo has an **Elementary Set Axiom,** which consists of the Empty Set Axiom and the Pairing Axiom.[5]

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

We define two classes to be equal if they have the same members. Thus if

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

then the formulas φ and ψ define equal classes. Here too we consider equality as identity.

Similarly, 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.

We now have that *every set is a class.* In particular, every set a is the class $\{x\colon x \in a\}$.

However, *not every class is a set.* For, the class $\{x\colon x \in x\}$ is not a set. If it were a set a, then $a \in a \Leftrightarrow a \notin a$, which is a contradiction. This is the *Russell Paradox* [15].

Every set a has a **union,** which is the class $\{x\colon \exists y\ (x \in y \land y \in a)\}$. This union is denoted by $\bigcup a$. The **Union Axiom** is that this class is a set:

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

Note that $a \cup b = \bigcup\{a, b\}$. The Adjunction Axiom is a consequence of the Union and Pairing Axioms. We use the Union Axiom when considering unions of chains of structures (as on page 38 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 $x \in A \land \varphi(x)$. The **Separation Axiom** is that this class is a set. We may denote this set by $\{x \in A\colon \varphi(x)\}$. Actually Separation is a *scheme* of axioms, one for each singulary formula φ:

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

In most of mathematics, and in particular in these notes, one need not worry 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 the natural numbers as sets. We can talk about sets by means of formulas. Formulas define classes of sets, as above. Some of these classes turn out to be sets themselves; but there is no reason to expect all of them to be sets. Indeed, as we have noted, some of them are not sets. *Sub-classes* of sets are sets; 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 $\mathcal{P}(a)$. The **Power Set Axiom** is that this class is a set:

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

Then $\mathcal{P}(a)$ can be called the **power set** of a. The Power Set Axiom will be of minor importance to us; we shall not actually use it until page 120.

We shall not use the Axiom of Choice to prove anything. However, it can be used to show that some objects that we shall study are interesting (p. 74) or even exist at all (p. 133).

The **Axiom of Infinity** is that the collection $\{0, 1, 2, \dots\}$ of natural numbers is a set. It is not obvious how to formulate this axiom as a sentence of our logic. One approach is to let $\varphi(x)$ be the formula

$$\forall y\ \bigl(0 \in x \land (y \in x \Rightarrow y \cup \{y\} \in x)\bigr)$$

and to declare that the Axiom of Infinity is the sentence $\exists x\ \varphi(x)$. Then by definition

$$\omega = \bigcap\{x\colon \varphi(x)\}. \tag{0.5}$$

In general, $\bigcap a$ is the class

$$\{x\colon \forall y\ (y \in a \Rightarrow x \in y)\}.$$

This class is **intersection** of a. If $b \in a$, then $\bigcap a \subseteq b$, and so $\bigcap a$ is a set by the Separation Axiom. In particular, by the Axiom of Infinity, ω is a set. However, $\bigcap \varnothing$ is the class of all sets.

Our definition of ω does not by itself establish that it has the properties we expect of the natural numbers. We shall do this in §0.4 (p. 22).

For the record, we have now named all of the axioms given by Zermelo in 1908: (I) Extension, (II) Elementary Set, (III) Separation, (IV) Power

Set, (V) Union, and (VI) Choice. Zermelo assumes that equality is identity: we have expressed this as the sentence (0.4) above. In fact Zermelo does not use logical formalism as we have. We prefer to define equality with (0.1) and (0.2) and then use the Axioms of (1) Empty Set, (2) Equality, (3) Adjunction, (4) Separation, (5) Union, (6) Power Set, and (7) Choice. But these two collections of axioms are logically equivalent.

Apparently Zermelo overlooked on axiom, the *Replacement Axiom,* which was supplied in 1922 by Skolem [17] and by Fraenkel.[6] We shall give this axiom in the next section.

An axiom never needed in ordinary mathematics is the *Foundation Axiom.* Stated originally by von Neumann [20], 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. But we shall not use these expressions again in these notes.

0.2. Functions and relations

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. Here the **ordered pair** (x, y) is defined so that

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

[6] I have not been able to consult Fraenkel's original papers. According to van Heijenoort [19, p. 291], Lennes also suggested something like the Replacement Axiom at around the same time (1922) as Skolem and Fraenkel; but Cantor had suggested such an axiom in 1899.

One definition that accomplishes this is $(x, y) = \big\{\{x\}, \{x, y\}\big\}$, but we never actually need the precise definition. An **ordered triple** (x, y, z) can be defined as $\big((x, y), z\big)$, and so forth.

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

$$f(b) = a. \tag{0.6}$$

I assume the reader is familiar with the *kinds* of functions from B to A: injective or one-to-one, surjective or onto, and bijective. 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 B to A, we may write[7]

$$f\colon B \to A. \tag{0.7}$$

If $C \subseteq B$, the class $\{y\colon \exists x\ (x \in C \land y = f(x)\}$ can be written as one of[8]

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

This class is the **image** of C under f. Here this class is a sub-class of A, and so it is a set by the Separation Axiom. By the **Replacement Axiom,** the image of every set under every function is a set. For example, if we are just given a function $n \mapsto G_n$ on ω, by Replacement we have that the class $\{G_n\colon n \in \omega\}$ is a set.

A **singular operation**[9] on A is a function from A to itself; a **binary** operation 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

$$a\ R\ b.$$

[7] Thus, while the symbol f can be understood as a *noun,* the expression $f\colon B \to A$ is a complete *sentence.* We may write "Let $f\colon B \to A$" to mean "Let f be a function from B to A." It would be redundant and even illogical to write "Let $f\colon B \to A$ be a function from B to A"; however, such confusing expressions are common in mathematical writing.

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

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

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 (0.6). I assume the reader is familiar with other kinds of binary relations, such as orderings.

0.3. An axiomatic development of the natural numbers

In §0.1 (p. 8) we sketched an axiomatic approach to set theory. Now we start over with an axiomatic approach to the natural numbers alone. We integrate numbers and sets in the section after this.

For the moment, we forget the definition of ω. 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 $\mathbb{N}$ that has

1) a distinguished **initial element,** denoted by 1 and called **one,** and
2) a distinguished singulary operation of **succession,** namely $n \mapsto n+1$, where $n+1$ is called the **successor** of 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 is sometimes called the **universe** of the structure. If one wants a simple notational distinction between a structure and its universe, and the universe is A, then the structure might be denoted by $\mathfrak{A}$. (Here $\mathfrak{A}$ is the Fraktur version of A. See Appendix A.)

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: $0 \neq n+1$.
2. Succession is injective: if $m+1 = n+1$, then $m = n$.
3. the structure admits **proof by induction,** in the following sense. Suppose A is a subset of the universe with the following two closure properties:
   a) $1 \in A$;

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

Then A must be the whole universe: $A = \mathbb{N}$.

These axioms seem to have been discovered originally by Dedekind [2, II, VI (71), p. 67], although they were also written down by Peano [14] and 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 don't. In the latter case, we may eventually come back to b. Otherwise, we reach an element c, and later a different element d, such that $f(c) = f(d)$. The second of the Peano Axioms would rule out this possibility; the first would ensure that our computations never returned to b. The last 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 [5].

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. However, as we shall note later, the theorem is not just an immediate consequence of induction. The proof uses all three of the Peano Axioms.

Theorem 1 (Recursion). *For every iterative structure (A, b, f), there is a unique **homomorphism** to this structure from $(\mathbb{N}, 1, n \mapsto n + 1)$: that is, there is a unique function h from $\mathbb{N}$ to A such that*

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

Proof. We seek h as 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 $(x, y) \in C$, then either

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

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 $(x, y) \in R$, we may write also

$$x \mathrel{R} y.$$

Since $(1, b) \in B$, we have $1 \mathrel{R} b$. If $n \mathrel{R} y$, then $(n, y) \in C$ for some C in B, but then $C \cup \{(n+1, f(y))\} \in B$ by definition of B, so $(n+1) \mathrel{R} f(y)$. Therefore R is the desired function h, provided it is a *function* from $\mathbb{N}$ to A. Proving this has two stages.

1. For all n in $\mathbb{N}$, there is y in A such that $n \mathrel{R} y$. Indeed, let D be the set of such n. Then we have just seen that $1 \in D$, and if $n \in D$, then $n + 1 \in D$. By induction, $D = \mathbb{N}$.

2. For all n in $\mathbb{N}$, if $n \mathrel{R} y$ and $n \mathrel{R} z$, then $y = z$. Indeed, let E be the set of such n. Suppose $1 \mathrel{R} y$. 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 $(n + 1) \mathrel{R} y$. Then $(n + 1, y) \in C$ for some C in B. Again since 1 is not a successor, we must have $(n + 1, y) = (m + 1, f(v))$ for some (m, v) in C. Since succession is injective, we must have $m = n$. Since $n \in E$, we know v is *unique* such that $n \mathrel{R} v$. Since $y = f(v)$, therefore y is unique such that $(n + 1) \mathrel{R} y$. Thus $n + 1 \in E$. By induction, $E = \mathbb{N}$.

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

Corollary. *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. $H(1) = b$,
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 $\mathbb{N}$

1) the binary operation $(x, y) \mapsto x + y$ of **addition,** and
2) the binary operation $(x, y) \mapsto x \cdot y$ of **multiplication.** (We often write xy for $x \cdot y$.)

The definitions are:

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

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

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

Proof. Induction. □

Theorem 2. *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 3. *Addition on $\mathbb{N}$ allows* ***cancellation:*** *if $n+x=n+y$, then $x=y$.*

Proof. Induction, and injectivity of succession. □

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 4. *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 [10] proves *using induction alone* that $+$ and $\cdot$ exist as given by the recursive definitions above. However, Theorem 3 needs more than induction. Also, the existence of **exponentiation,** as an operation $(x, y) \mapsto x^y$ such that

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

requires more than induction.

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) $m \nless 1$;
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:

1) $\{x \in \mathbb{N} \colon x < 1\} = \varnothing$;
2) $\{x \in \mathbb{N} \colon x < n + 1\} = \{x \in \mathbb{N} \colon x < n\} \cup \{n\}$.

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

Theorem 5. *The relation $<$ is* ***transitive*** *on $\mathbb{N}$, that is, if $k < m$ and $m < n$, then $k < n$.*

Proof. Induction on n. □

Lemma 3. *$m \neq m + 1$.*

Proof. The claim is true when $m = 1$, since 1 is not a successor. Suppose the claim is true when $m = k$, that is, $k \neq k+1$. Then $k+1 \neq (k+1)+1$, by injectivity of succession, so the claim is true when $m = k + 1$. By induction, the claim is true for all m. □

Theorem 6. *The relation $<$ is* ***irreflexive*** *on $\mathbb{N}$: $m \nless m$.*

Proof. The claim is true when $m = 1$, since $m \not< 1$ by definition. Suppose the claim *fails* when $m = k + 1$. This means $k + 1 < k + 1$. Therefore $k + 1 \leqslant k$ by definition. By the previous lemma, $k + 1 < k$. But $k \leqslant k$, so $k < k + 1$ by definition. So $k < k + 1$ and $k + 1 < k$; hence $k < k$ by Theorem 5, that is, the claim fails when $m = k$. By induction, the claim holds for all m. □

Lemma 4. $1 \leqslant m$.

Proof. Induction. □

Lemma 5. *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 $k + 1 = n + 1 < (n + 1) + 1$. If $k < n$, then $k + 1 < n + 1$ by inductive hypothesis, so $k + 1 < (n + 1) + 1$ by transitivity. Thus the claim holds when $m = n + 1$. By induction, the claim holds for all m. □

Theorem 7. *The relation $\leqslant$ is **total** on $\mathbb{N}$: either $k \leqslant m$ or $m \leqslant k$.*

Proof. Induction and the two lemmas. □

Because of Theorems 5, 6, and 7, the set $\mathbb{N}$ is **(strictly) ordered** by the relation $<$.

Theorem 8. *For all m and n in $\mathbb{N}$, we have $m < n$ if and only if the equation*

$$m + x = n \tag{0.8}$$

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 (0.8) is soluble in $\mathbb{N}$. □

Theorem 9. *If $k < \ell$, then*

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

Here the first conclusion is a refinement of Theorem 3; the second yields the following analogue of Theorem 3 for multiplication.

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

Theorem 10. $\mathbb{N}$ *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 $1 \in B$, by the last lemma, 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 $A = \varnothing$. □

0.4. A construction of the natural numbers

Now we recall the definition (0.5) (p. 13) of ω. By this definition, ω contains $\varnothing$ and is closed under the operation $x \mapsto x'$, where

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

Moreover, ω is the *smallest* of the sets with these properties. (Such sets exist by the Axiom of Infinity.) Therefore the iterative structure $(\omega, \varnothing, {}')$ admits induction. We now prove that this structure satisfies the remaining two Peano Axioms.

Lemma 6. *On ω, membership implies inclusion.*

Proof. By induction on n, we prove that, for all k in ω, if $k \in n$, then $k \subseteq n$. The claim is vacuously true when $n = \varnothing$. Suppose it is true when $n = m$. If $k \in m'$, then either $k \in m$ or else $k = m$. In the former case, by inductive hypothesis, $k \subseteq m \subseteq m'$; in the latter case, $k = m \subseteq m'$. Thus the claim is true when $n = m'$. By induction, the claim is true for all n in ω. □

Lemma 7. *In ω, if $k \subset n$, then $k' \subseteq n$.*

Proof. The claim is vacuously true when $n = \varnothing$. Suppose it is true when $n = m$. Say $k \subset m'$. If $k \subseteq m$, then either $k \subset m$, in which case the inductive hypothesis implies, giving us $k' \subseteq m \subseteq m'$,—or else $k = m$, so that $k' = m'$. If $k \nsubseteq m$, then $m \in k$, so by Lemma 6 we have $m \subseteq k \subset m' = m \cup \{m\}$, and therefore $m = k$, so again $k' = m'$. Thus the claim is true when $n = m'$. Therefore the claim holds for all n in ω. □

Lemma 8. *Inclusion is a total ordering of ω.*

Proof. We have to show on ω that, if $k \nsubseteq n$, then $n \subseteq k$. The claim is trivially true when $n = \varnothing$. Suppose it is true when $n = m$. If $k \nsubseteq m'$, then $k \nsubseteq m$, so $m \subseteq k$, but $m \neq k$, so $m \subset k$, and therefore $m' \subseteq k$ by Lemma 7. □

Lemma 9. *Elements of ω are distinct from their successors.*

Proof. We prove that no element of ω has an element that is equal to its successor. This is trivially true for the empty set. Suppose it is true for m. If $k \in m'$, then either $k \in m$, or else $k = m$. In the former case, by inductive hypothesis, $k \neq k'$. In the latter case, if $k = k'$, then $m = k \cup \{k\}$, and in particular $k \in m$, contary to inductive hypothesis. Therefore no element of m' is equal to its successor. This completes the induction. Since every element of ω is an element of its successor, which is in ω, no element of ω is equal to its successor. □

Theorem 11. *The iterative structure $(\omega, \varnothing, {}')$ satisfies the Peano Axioms.*

Proof. We have observed that $(\omega, \varnothing, {}')$ admits induction. Easily too, $\varnothing$ is not a successor. By Lemma 8, if $m \neq n$, we may assume $m \subset n$. By Lemmas 7 and 9, we then have $m' \subseteq n \subset n'$. Thus succession is injective. □

The elements of ω are the **von Neumann natural numbers** [21]. Henceforth we write 0 for $\varnothing$, then 1 for $0'$, and 2 for $1'$, and so on. Thus we identify $\mathbb{N}$ with $\omega \smallsetminus \{\varnothing\}$, so that

$$\begin{aligned}\omega &= \{0\} \cup \mathbb{N},\\ \mathbb{N} &= \{1,2,3,\dots\},\\ \omega &= \{0,1,2,\dots\}.\end{aligned}$$

By the von-Neumann definition, we have

$$0=\varnothing; \qquad 1=\{0\}; \qquad 2=\{0,1\}; \qquad 3=\{0,1,2\}, \quad \dots$$

If $n \in \omega$, then

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

Note that this makes sense even when $n=0$.

0.5. 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 set with n elements. Since the set of functions from a set B to a set A can be denoted by

$$A^B,$$

we have, in particular, that A^n is the set of functions from $\{0,\dots,n-1\}$ into A. We can denote such a function by $(x_0,\dots,x_{n-1})$; that is,

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

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 §0.2. A nullary operation on A can be identified with an element of A.

Generalizing the terminology used at the beginning of §0.3, 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, the signature of an ordered field like $\mathbb{R}$ is $\{<, 0, 1, +, -, \cdot\}$. 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 $s^{\mathfrak{A}}$.

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

$$\begin{gathered} h(f^{\mathfrak{A}}(x_0, \dots, x_{n-1})) = f^{\mathfrak{B}}(h(x_0), \dots, h(x_{n-1})), \\ (x_0, \dots, x_{n-1}) \in R^{\mathfrak{A}} \Rightarrow (h(x_0), \dots, h(x_{n-1})) \in R^{\mathfrak{B}}, \end{gathered} \tag{0.9}$$

for all n-ary operation-symbols f and relation-symbols R of the signature, for all n in ω. 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 $h \colon A \to B$). A homomorphism is an **embedding** if it is injective and if the converse of (0.9) also holds. A surjective embedding is an **isomorphism.** A **substructure** of $\mathfrak{B}$ is a structure $\mathfrak{A}$ of the same signature such that $A \subseteq B$ and the inclusion of A in B is an embedding of $\mathfrak{A}$ in $\mathfrak{B}$.

Part I.

Groups

1. Basic properties of groups and rings

We define both groups and rings in this chapter. We define rings (in §1.6, p. 42), because at the beginning of the next chapter (§2.1, p. 44) we shall define certain groups—namely *general linear groups*—in terms of rings.

1.1. Symmetry 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 id_A, which is the **identity** on A;
2) the singulary operation $f \mapsto f^{-1}$, which is functional **inversion;**
3) the binary operation $(f, g) \mapsto f \circ g$, which is functional **composition.**

The structure $(\mathrm{Sym}(A), \mathrm{id}_A, {}^{-1}, \circ)$ is the **complete group of symmetries** of A. A substructure of this can be called simply a **group of symmetries** of A.

We may use $\mathrm{Sym}(A)$ 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.

In case $n \in \omega$, the notation S_n is also used for $\mathrm{Sym}(n)$. However, when most people write S_n, they probably mean 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. The size of

$\mathrm{Sym}(n)$ or S_n is $n \cdot (n-1) \cdots 2 \cdot 1$, which is denoted by $n!$ and called n **factorial.**

We shall consider the groups $\mathrm{Sym}(n)$ at greater length in §2.7 (p. 62). Meanwhile, it may be worth our while to have a brief look at them now. The group $\mathrm{Sym}(0)$ has a unique element, id_0 (which is itself $\varnothing$ or 0). The group $\mathrm{Sym}(1)$ has the unique element id_1 (which is $\{(0,0)\}$). Suppose $\sigma \in \mathrm{Sym}(n)$ for some n. Then

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

Now, there is no particular reason to list the entries of an ordered pair horizontally. Instead of (x, y), we could write $\begin{pmatrix} x \\ y \end{pmatrix}$. Then we have

$$\sigma = \left\{\begin{pmatrix} 0 \\ \sigma(0) \end{pmatrix}, \dots, \begin{pmatrix} n-1 \\ \sigma(n-1) \end{pmatrix}\right\}.$$

Here the parentheses (the round brackets) serve no particular purpose; we might as well write simply

$$\sigma = \begin{Bmatrix} 0 & \dots & n-1 \\ \sigma(0) & \dots & \sigma(n-1) \end{Bmatrix}.$$

This is a set with n elements, and each of those elements is an ordered pair, here written vertically. In particular, those n elements can be written in a different order; but the entries in a particular element cannot. Thus, with this notation, the same permutation of n can be written in $n!$ different ways, one for each permutation of the columns.

In fact the books that I know of replace the braces (the curly brackets) with parentheses, as in

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

However, this notation is potentially misleading, because it does not stand for a *matrix* such as we shall define in §2.1 (p. 44). In a matrix, the order of the columns (as well as the rows) matters. We could write σ as the ordered n-tuple $\bigl(\sigma(0), \dots, \sigma(n-1)\bigr)$ or the $1 \times n$ matrix

$\bigl(\sigma(0) \quad \cdots \quad \sigma(n-1)\bigr)$; but we shall not do this, because of the potential confusion with a similar notation, to be introduced presently.

In case

$$\sigma = \begin{Bmatrix} 0 & 1 & \cdots & n-2 & n-1 \\ 1 & 2 & \cdots & n-1 & 0 \end{Bmatrix},$$

σ can be called a *cycle.* More generally, if $2 \leqslant m \leqslant n$, then the permutation

$$\begin{Bmatrix} 0 & 1 & \cdots & m-2 & m-1 & m & \cdots & n-1 \\ 1 & 2 & \cdots & m-1 & 0 & m & \cdots & n-1 \end{Bmatrix}$$

is a cycle too, or more precisely an *m-cycle.* For the moment, let us refer to this cycle as σ_m. Then for all k in n, we have

$$\sigma_m(k) = \begin{cases} k+1, & \text{if } k < m-1, \\ 0, & \text{if } k = m-1, \\ k, & \text{if } m \leqslant k < n. \end{cases}$$

In the most general sense, an element σ of $\operatorname{Sym}(n)$ is called an m-**cycle,** or a cycle of **length** m, if, for some τ in $\operatorname{Sym}(n)$, for all k in n,

$$\sigma(\tau(k)) = \begin{cases} \tau(k+1), & \text{if } k < m-1, \\ \tau(0), & \text{if } k = m-1, \\ \tau(k), & \text{if } m \leqslant k < n. \end{cases}$$

In this case

$$\sigma = \begin{Bmatrix} \tau(0) & \tau(1) & \cdots & \tau(m-2) & \tau(m-1) & \tau(m) & \cdots & \tau(n-1) \\ \tau(1) & \tau(2) & \cdots & \tau(m-1) & \tau(0) & \tau(m) & \cdots & \tau(n-1) \end{Bmatrix}.$$

Then $\sigma(\tau(k)) = \tau(\sigma_m(k))$ for all k in n, and so

$$\sigma = \tau \circ \sigma_m \circ \tau^{-1}.$$

We can now write σ neatly as

$$\bigl(\tau(0) \quad \ldots \quad \tau(m-1)\bigr).$$

All this means is that σ takes each entry $\tau(k)$ to the next entry $\tau(k+1)$, except that it takes $\tau(m-1)$ to $\tau(0)$. So the expression above should

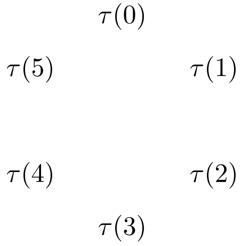


Figure 1.1. A cycle.

be understood, not as a matrix, but rather as a ring, a circle, indeed a *cycle,* as in Figure 1.1 where $m = 6$. In general, the circle can be broken and written in one line in m different ways, as

$$\begin{pmatrix}\tau(i) & \cdots & \tau(m-1) & \tau(0) & \cdots & \tau(i-1)\end{pmatrix}$$

for any i in m.

We have defined m-cycles when $m > 1$. However, we can consider the identity id_n is a 1-cycle. This might be denoted by (0), or even by (i) for any i in m; but I shall use the notation ( ).

Two arbitrary elements σ and τ of $\mathrm{Sym}(n)$ are **disjoint** if, for all k in n,

$$\sigma(k) \neq k \implies \tau(k) = k.$$

In this case, $\sigma \circ \tau = \tau \circ \sigma$, that is, the two permutations **commute.** An arbitrary composite of permutations is also called the **product** of the symmetries. We shall show, as Theorem 55 (p. 62), that every element of $\mathrm{Sym}(n)$ is the product of a unique set of disjoint cycles of length 2 or more.

When n is small, we can just list the elements of $\mathrm{Sym}(n)$:

Sym(2): ( ), (0 1).
Sym(3): ( ), (0 1), (0 2), (1 2), (0 1 2), (0 2 1).
Sym(4): ( ), (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).

For larger n, one might like to have some principles 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 $m \leqslant n$, there is an embedding $\sigma \mapsto \tilde{\sigma}$ of the group $\operatorname{Sym}(m)$ in $\operatorname{Sym}(n)$, where $\tilde{\sigma} = \sigma \cup \mathrm{id}_{n \smallsetminus m}$, so that

$$\tilde{\sigma}(k) = \begin{cases} \sigma(k), & \text{if } k < m, \\ k, & \text{if } m \leqslant k < n. \end{cases}$$

Similarly each $\operatorname{Sym}(n)$ embeds in $\operatorname{Sym}(\omega)$; but the latter has many elements that are not in the image of any $\operatorname{Sym}(n)$.

The main point to observe for now is the following.

Theorem 12. *For all sets A, for all elements f, g, and h of a group of symmetries of A,*

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

1.2. Groups

A **group** is a structure with the properties of a group of symmetries given by the last theorem, Theorem 12. That is, a group is a structure $(G, \mathrm{e}, {}^{-1}, \cdot)$ in which the following equations are identities (that is, are true for all values of the variables):

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

We may say also that these equations are the *axioms* of groups: this means that their universal generalizations $\forall x\ x \cdot \mathrm{e} = x$ and so forth are true by definition in every group.

The operation $\cdot$ here is usually called **multiplication,** and we usually write $g \cdot h$ as gh. The element g^{-1} is the **inverse** of g. The element e is the **identity;** it is sometimes denoted by 1 rather than e. Every element g of G determines a singulary operation λ_g on G, given by

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

Theorem 13 (Cayley). *For every group* $(G, \mathrm{e}, {}^{-1}, \cdot)$ *and every* g *in* G*, the function* λ_g *belongs to* $\mathrm{Sym}(G)$*; moreover, 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. Let $g \in G$. We first establish $\lambda_g \in \mathrm{Sym}(G)$. We have

$$\lambda_{g^{-1}}(\lambda_g(x)) = g^{-1}(gx) = (g^{-1}g)x = \mathrm{e}\,x = x,$$

so $\lambda_{g^{-1}} \circ \lambda_g = \mathrm{id}_G$. Likewise $\lambda_g \circ \lambda_{g^{-1}} = \mathrm{id}_G$. Thus λ_g is invertible and therefore belongs to $\mathrm{Sym}(G)$. Consequently

$$x \mapsto \lambda_x \colon G \to \mathrm{Sym}(G)$$

(recall the notational convention established above on page 15). We now check that $x \mapsto \lambda_x$ is a homomorphism. By what we have already shown,

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

We have also $\lambda_{\mathrm{e}}(x) = \mathrm{e}x = x = \mathrm{id}_G(x)$, so

$$\lambda_{\mathrm{e}} = \mathrm{id}_G,$$

and $\lambda_{gh}(x) = (gh)x = g(hx) = \lambda_g(\lambda_h(x)) = (\lambda_g \circ \lambda_h)(x)$, so

$$\lambda_{gh} = \lambda_g \circ \lambda_h.$$

Thus $x \mapsto \lambda_x$ is indeed a homomorphism from the group $(G, \mathrm{e}, {}^{-1}, \cdot)$ to $(\mathrm{Sym}(G), \mathrm{id}_G, {}^{-1}, \circ)$. It is an embedding, since if $\lambda_g = \lambda_h$, then in particular

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

1.3. The integers and rationals

In this section we define *semigroups* and *monoids.* The structure $(\mathbb{N}, +)$ will be a semigroup, and $(\mathbb{N}, 1, \cdot)$ and $(\omega, 0, +)$ will be monoids. From these, we shall obtain the groups $(\mathbb{Q}^+, 1, {}^{-1}, \cdot)$ and $(\mathbb{Z}, 0, -, +)$ respectively. We then obtain the semigroup $(\mathbb{Q}^+, +)$, from which we obtain the group $(\mathbb{Q}, 0, -, +)$. Then we shall have the monoid $(\mathbb{Q}, 1, \cdot)$. In fact $(\mathbb{Q}, 0, -, +, 1, \cdot)$ will be a *ring* and even a *field,* though the official definitions of these terms will come later.

The structure $(\mathbb{N}, 1, \cdot)$ cannot be given an operation of inversion so that it becomes a group. The structure is however a *monoid.* A **monoid** is a structure $(M, \mathrm{e}, \cdot)$ satisfying the axioms

$$\begin{aligned} x\,\mathrm{e} &= x \\ \mathrm{e}\,x &= x, \\ (xy)z &= x(yz). \end{aligned}$$

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

In general terms, the structure $(G, \mathrm{e}, \cdot)$ is a **reduct** of $(G, \mathrm{e}, {}^{-1}, \cdot)$, and $(G, \mathrm{e}, {}^{-1}, \cdot)$ is an **expansion** of $(M, \mathrm{e}, \cdot)$. The terms *reduct* and *expansion* imply no change in universe of a structure, but only a change in the signature.

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, let us denote by $\mathrm{E}(A)$ the set of functions from A to itself (that is, the set of singulary operations on A). Then $(\mathrm{E}(A), \mathrm{id}_A, \circ)$ is a monoid. However, if A has at least two elements, then $\mathrm{E}(A)$ 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 Theorem 13, $x \mapsto \lambda_x$ is a homomorphism from $(M, \mathrm{e}, \cdot)$ to $(\mathrm{E}(M), \mathrm{id}_M, \circ)$; 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 another monoid, which expands to a group, by the method of fractions learned in school. The following theorem gives a special case of "localization", which will be worked out in full in §6.5 (p. 129):

Theorem 14. *Let $\approx$ be the binary relation on $\mathbb{N} \times \mathbb{N}$ given by*[1]

$$(a, b) \approx (x, y) \Leftrightarrow ay = bx.$$

Then $\approx$ is an equivalence-relation. Let the equivalence-class of (a, b) be denoted by a/b, and let the set of such equivalence-classes be denoted by $\mathbb{Q}^+$. Then $(\mathbb{Q}^+, 1, {}^{-1}, \cdot)$ is a well-defined group according to the rules

$$\begin{gathered} 1 = 1/1, \\ (x/y)^{-1} = y/x, \\ (x/y)(z/w) = (xz)/(yw). \end{gathered}$$

Moreover, $(\mathbb{N}, 1, \cdot)$ embeds in $(\mathbb{Q}^+, 1, \cdot)$ under the map $x \mapsto x/1$.

The set $\mathbb{Q}^+$ in the theorem comprises the **positive rational numbers.** The foregoing theorem is false if we replace the monoid $(\mathbb{N}, 1, \cdot)$ with the monoid $(\mathrm{E}(A), \mathrm{id}_A, \circ)$ for a set A with at least two elements. But the theorem works for $(\omega, 0, +)$. In fact, after appropriate modifications, it will work for $(\mathbb{N}, +)$.

The structure $(\mathbb{N}, +)$ is a *semigroup.* 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 $(\omega, \cdot)$ are not reducts of monoids. Or let S be the set of all operations f on $\mathrm{E}(\omega)$ such that, for all n in ω, $f(n) > n$: then S is closed under composition, so $(S, \circ)$ is a semigroup; but it has no identity.

[1] 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 $\mathbb{N}^4$.

Theorem 15. *Let $\sim$ be the binary relation on $\mathbb{N}\times\mathbb{N}$ given by*

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

Then $\sim$ is an equivalence-relation. Let the equivalence-class of (a,b) be denoted by $a-b$, and let the set of such equivalence-classes be denoted by $\mathbb{Z}$. Then $(\mathbb{Z},0,-,+)$ is a well-defined group according to the rules

$$\begin{gathered}0=1-1,\\ -(x-y)=y-x,\\ (x-y)+(z-w)=(x+z)-(y+w).\end{gathered}$$

Moreover, $(\mathbb{N},+)$ embeds in $(\mathbb{Z},+)$ under the map $x\mapsto(x+1)-1$.

Now we can obtain the set $\mathbb{Q}$ of all rational numbers from $\mathbb{Q}^+$, just as we have obtained $\mathbb{Z}$ from $\mathbb{N}$. To do this, we need addition on $\mathbb{Q}^+$:

Theorem 16. *The set $\mathbb{Q}^+$ is a semigroup with respect to an operation $+$, which can be well defined by*

$$\frac{a}{b}+\frac{x}{y}=\frac{ay+bx}{by}.$$

Then on $\mathbb{Q}^+$,

$$x(y+z)=xy+xz.$$

Now we obtain $\mathbb{Q}$ with its usual addition and multiplication. 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* (see §1.6, p. 42).

1.4. Simplifications

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

Theorem 17. *A semigroup can expand to a group in only one way.*

Proof. Let $(G, \mathrm{e}, {}^{-1}, \cdot)$ be a group. If 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

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

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

Theorem 18. *Any structure satisfying the identities*

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

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

$$\exists z\ \forall x\ \forall y\ \exists u\ (zx = x \wedge uy = z). \tag{1.1}$$

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

Thanks to Theorem 17, 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 19. *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 17 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 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, 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 18, 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 (1.1), but it is. It should be understood as

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

which is a sentence of the general form $\forall\exists$; whereas (1.1) is of the form $\exists\forall\exists$).

Theorem 20. *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 [1, 11]. With this theorem, and with Theorem 20 in place of 19, 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 20 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 §2.2), 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.

1.5. Repeated multiplication

In a semigroup, a product abc is unambiguous: whether it is understood as $(ab)c$ or $a(bc)$, the result is the same. Then $abcd$ is also unambiguous, because $(abc)d$, $(ab)(cd)$, and $a(bcd)$ can be shown to be equal. We are going to show by induction that every product $a_0 \cdots a_{n-1}$ is unambiguous. The main point is to establish the homomorphisms in the last three theorems of this section.

Suppose there is a binary operation $\cdot$ on a set A. We do not assume that the operation is associative. For each n in $\mathbb{N}$, we define a set P_n consisting of certain n-ary operations on A. Our definition is recursive:

1) $P_1 = \{\mathrm{id}_A\}$;
2) P_{n+1} consists of the operations

$$(x_0, \dots, x_n) \mapsto f(x_0, \dots, x_{k-1}) \cdot g(x_k, \dots, x_n),$$

for every f in P_k and g in P_{n+1-k}, for every k in $\mathbb{N}$ such that $k \leqslant n$.

We now distinguish in each P_n a particular element f_n, where

1) f_1 is id_A,
2) f_{n+1} is $(x_0, \dots, x_n) \mapsto f_n(x_0, \dots, x_{n-1}) \cdot x_n$.

So

$$f_n(x_0, \dots, x_{n-1}) = (\cdots (x_0 x_1) x_2 \cdots) x_{n-1}.$$

For example, f_5 is $(x, y, z, u, v) \mapsto (((xy)z)u)v$. But P_5 also contains $(x, y, z, u, v) \mapsto (x(yz))(uv)$. In a semigroup, it is easy to show that this operation is the same as f_5. In general, we have:

Theorem 21. *If A is a semigroup, then, in the notation above, $P_n = \{f_n\}$.*

Proof. The claim is immediately true when $n = 1$. Suppose it is true when $1 \leqslant n \leqslant s$. Each element g of P_{s+1} is therefore

$$(x_0, \dots, x_s) \mapsto f_n(x_0, \dots, x_{n-1}) \cdot f_{s+1-n}(x_n, \dots x_s)$$

for some n, where $1 \leqslant n \leqslant s$. If $n = s$, then g is f_{n+1}. If $n < s$, then

$$\begin{aligned} g(x_0, \dots, x_s) &= f_n(x_0, \dots x_{n-1} \cdot (f_{s-n}(x_n, \dots, x_{s-1}) \cdot x_s) \\ &= (f_n(x_0, \dots x_{n-1} \cdot f_{s-n}(x_n, \dots, x_{s-1})) \cdot x_s \\ &= f_s(x_0, \dots, x_{s-1}) \cdot x_s \\ &= f_{s+1}(x_0, \dots x_s), \end{aligned}$$

so again g is f_{s+1}. By induction, the claim is true for all n in $\mathbb{N}$. □

It follows that, in a semigroup, the product $a_0 \cdots a_{n-1}$ is unambiguous: it is just $g(a_0, \dots, a_{n-1})$ for any element g of P_n, because that element must be the same as f_n. We may write also

$$a_0 \cdots a_{n-1} = \prod_{k=0}^{n-1} a_k = \prod_{k \in n} a_k. \tag{1.2}$$

A group or monoid or semigroup is **abelian** if it satisfies the identity

$$xy = yx.$$

Multiplication on an abelian group is often (though not always) called **addition** and denoted by $+$; in this case, the identity may be denoted by 0, and the group is said to be written additively. This is what we do in the case of $(\omega, 0, +)$, though not $(\mathbb{N}, 1, \cdot)$.

In an abelian group, the product in (1.2) may be written as a sum:

$$a_0 + \cdots + a_{n-1} = \sum_{k=0}^{n-1} a_k = \sum_{k \in n} a_k.$$

We also use the notation

$$\prod_{k \in n} a = a^n, \qquad\qquad \sum_{k \in n} a = na.$$

The set $\mathrm{E}(G)$ in the following was defined in §1.4 (p. 35).

Theorem 22. *Suppose $(G, \cdot)$ is a semigroup, and m and n range over $\mathbb{N}$.*

1. *On G,*
$$x^{m+n} = x^m x^n.$$
That is, if $a \in G$, then
$$x \mapsto a^x \colon (\mathbb{N}, +) \to (G, \cdot).$$
2. *On G,*
$$x^{mn} = (x^m)^n.$$
That is,
$$x \mapsto (y \mapsto y^x) \colon (\mathbb{N}, 1, \cdot) \to (\mathrm{E}(G), \mathrm{id}_A, \circ).$$

Proof. Use induction: $a^{n+1} = a^n \cdot a = a^n \cdot a^1$, and if $a^{n+m} = a^n \cdot a^m$, then
$$a^{n+(m+1)} = a^{(n+m)+1} = a^{n+m} \cdot a = a^n a^m a = a^n a^{m+1}.$$
Also, $a^{n \cdot 1} = a^n = (a^n)^1$, and if $a^{nm} = (a^n)^m$, then
$$a^{n(m+1)} = a^{nm+n} = a^{nm} a^n = (a^n)^m a^n = (a^n)^{m+1}. \qquad \square$$

In a monoid, we define
$$a^0 = \mathrm{e}\,. \tag{1.3}$$
Again, the set $\mathrm{E}(G)$ in the following was defined in §1.4.

Theorem 23. *Suppose $(G, \mathrm{e}, \cdot)$ is a monoid.*

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

In a group, we define
$$a^{-n} = (a^n)^{-1}.$$

Theorem 24. *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. $x \mapsto (y \mapsto y^x) \colon (\mathbb{Z}, 1, \cdot) \to (\mathrm{E}(G), \mathrm{id}_A, \circ).$

1.6. Rings

A homomorphism from a structure to itself is an **endomorphism.** 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 id_E, and multiplication is functional composition. This multiplication distributes in both senses over addition:

$$f(g + h) = fg + fh, \qquad (f + g)h = fh + gh.$$

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** of 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.

For now, we shall refer to associative rings simply as **rings.** (In §5.1 we shall consider rings in a more general sense.) As with a group, so with a ring: an element a determines a singulary operation λ_a on the ring, given by

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

Theorem 25. *The function $x \mapsto \lambda_x$ embeds a ring in the endomorphism ring of its underlying abelian group.*

If, in a ring, the multiplication commutes—

$$xy = yx$$

—then the ring is a **commutative ring.** For example, $\mathbb{Z}$ is a commutative ring.

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

Theorem 26. *In a 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 a ring R is denoted by

$$R^{\times}.$$

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

If R is commutative, and $R^{\times} = R \smallsetminus \{0\}$, then R is a **field.** Multiplication on $\mathbb{Q}^{+}$ can be extended to $\mathbb{Q}$ so that this becomes a field. There are several ways to construct from $\mathbb{Q}$ the field $\mathbb{R}$ of real numbers. Then the field $\mathbb{C}$ can be defined as $\mathbb{R} \times \mathbb{R}$ with the appropriate operations. (See p. 49.) An example of a ring in which some elements have right but not left inverses will be given in §3.1.

2. Groups

2.1. General linear groups

Given a commutative ring R and an element n of ω, we define

$$\mathrm{M}_n(R)$$

as the set of functions from $n \times n$ into R. A typical such function can be written as a **matrix**

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

or as

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

or simply as $(a^i_j)^i_j$ if the set over which i and j range is clear. Here the entries a^i_j are from R. We define an addition on $\mathrm{M}_n(R)$ by

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

We define a multiplication on $\mathrm{M}_n(R)$ by

$$(a^i_j)^{i<n}_{j<n}(b^j_k)^{j<n}_{k<n} = \Big(\sum_{j\in n} a^i_j b^j_k\Big)^{i<n}_{k<n}.$$

One particular element of $\mathrm{M}_n(R)$ is called $(\delta^i_j)^{i<n}_{j<n}$, where

$$\delta^i_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases}$$

so that the element is a certain diagonal matrix, namely

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

Theorem 27. *If R is a commutative ring, then $\mathrm{M}_n(R)$ is a ring with multiplicative identity $(\delta^i_j)^{i<n}_{j<n}$.*

The group $\mathrm{M}_n(R)^{\times}$ is called the **general linear group** of degree n over R; it is also denoted by

$$\mathrm{GL}_n(R).$$

We shall characterize the elements of this group in §2.8. Meanwhile, since

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

we may observe that the element $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of $\mathrm{M}_n(R)$ is invertible if $ad-bc \in R^{\times}$.

2.2. New groups from old

2.2.1. Products

If G and H are two groups, then we can define a multiplication on $G \times H$ termwise:

$$(g_0, h_0)(g_1, h_1) = (g_0g_1, h_0h_1)$$

(that is, $(g_0 \cdot^G g_1, h_0 \cdot^H h_1)$). The result is a group called the **direct product** of G and H and also denoted by

$$G \times H.$$

If G and H are abelian, written additively, then their direct product is usually called a **direct sum,** denoted by

$$G \oplus H.$$

2.2.2. Quotients

Suppose $\sim$ is an equivalence-relation on a set G, so that it partitions G into equivalence-classes

$$\{x \in G \colon x \sim a\};$$

such classes can be denoted by $a/{\sim}$ or $[a]$ or $\overline{a}$. The **quotient** of G by $\sim$ is the set of equivalence-classes with respect to $\sim$; this set can be denoted by

$$G/{\sim}.$$

If, for some n in ω and some set A, we have $f \colon G^n \to A$, and

$$a_0 \sim x_0 \land \cdots \land a_{n-1} \sim x_{n-1} \Rightarrow f(a_0, \dots, a_{n-1}) = f(x_0, \dots, x_{n-1}),$$

then we say there is a **well-defined** function $\tilde{f}$ from $(G/{\sim})^n$ to A given by

$$\tilde{f}([a_0], \dots, [a_{n-1}]) = f(a_0, \dots, a_{n-1}).$$

This terminology is unfortunate, especially when used in the form "the function $([a_0], \dots, [a_{n-1}]) \mapsto f(a_0, \dots, a_{n-1})$ on $G/{\sim}$ is well-defined". Indeed, if this function is *not* well-defined, what this means is that there is no such function at all. But when there *is* such a function, and we call it $\tilde{f}$, then we have

$$\tilde{f} \circ p = f, \tag{2.1}$$

where p is the function $(x_0, \mapsto x_{n-1}) \mapsto ([x_0], \dots, [x_{n-1}])$ on G^n. Another way to express the equation (2.1) is to say that the following diagram **commutes:**

$$\begin{tikzcd}
G^n \arrow[r, "f"] \arrow[d, "p"'] & A \\
(G/{\sim})^n \arrow[ur, "\tilde{f}"'] &
\end{tikzcd}$$

We shall be particularly interested in the case where G is a semigroup. In this case, if there is a well-define multiplication on $G/{\sim}$ given by

$$[x][y] = [xy],$$

then this multiplication is associative, so $G/{\sim}$ is a semigroup. In this case, $\sim$ is called a **congruence-relation** with respect to the multiplication on G.

Theorem 28. *If G is a group, and $\sim$ is a congruence-relation on G, then $G/{\sim}$ is a group.*

For example, if $n \in \omega$, then two integers a and b are **congruent *modulo* n** if $n \mid b - a$; in this case one writes

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

Theorem 29. *If $n \in \omega$, then congruence* modulo *n is a congruence-relation on $\mathbb{Z}$ with respect to addition and multiplication, and the quotient is a commutative ring. If n is prime, then this ring is a field.*

The commutative ring in the theorem can be denoted by

$$\mathbb{Z}_n,$$

though sometimes this expression may denote merely the additive group. Note that $\mathbb{Z}_0$ is isomorphic to $\mathbb{Z}$. The direct sum $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ is the **Klein four group,** denoted by

$$\mathrm{V}$$

(for *Vierergruppe*[1]). This is the smallest group containing two elements neither of which is a power of the other.

There is a congruence-relation on $\mathbb{R}$ with respect to addition given by

$$a \sim b \iff a - b \in \mathbb{Z}.$$

Then there is a well-defined embedding $a \mapsto \exp(2\pi \mathrm{i} a)$ of $\mathbb{R}/{\sim}$ in $\mathbb{C}^{\times}$.

2.2.3. Subgroups

A **subgroup** of a group is just a substructure of a group, when the 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. Every group has both itself and $\{\mathrm{e}\}$ as subgroups. Also $G \times \{\mathrm{e}\}$ and $\{\mathrm{e}\} \times H$ are subgroups of $G \times H$, while $G \times G$ has the subgroup $\{(x, x) \colon x \in G\}$.

[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.

Theorem 30. *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[2]

$$H < G.$$

Theorem 31. *If* $\sim$ *is a congruence-relation on a group* G*, then the* $\sim$*-class of* e *is a subgroup of* G.

It is important to note that the converse of the lemma 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 §1.1. 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 §2.6 (p. 57) 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 $\operatorname{im}(f)$.

An embedding (that is, an injective homomorphism) is also called a **monomorphism.** A surjective homomorphism is called an **epimorphism.**

Theorem 32. *Let* f *be a homomorphism from* G *to* H.

[2] One might write $H \leqslant G$, if one wants to reserve the expression $H < G$ for the case where H is a *proper* subgroup of G. I shall not do this. However, I do think it is important 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.

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 $\mathrm{M}_2(\mathbb{R})$, 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 $\mathrm{M}_n(\mathbb{R})$ and is a field. The elements of $\mathbb{C}$ usually denoted by 1 and 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 $\mathrm{M}_2(\mathbb{C})$, 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 $\mathrm{GL}_2(\mathbb{C})$ 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 j is given by

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

Theorem 33. *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. □

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 $\langle\varnothing\rangle = \{\mathrm{e}\}$.

If $G = \langle A\rangle$, then G is **generated** by A. If $A = \{a_0, \dots, a_{n-1}\}$, we may write

$$\langle a_0, \dots, a_{n-1}\rangle$$

for $\langle A\rangle$, and say that G has the n **generators** a_0, …, a_{n-1}. In particular, G is **finitely generated** in this case. 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: ± 1, $\pm\mathrm{i}$, $\pm\mathrm{j}$, and $\pm\mathrm{k}$, where $\mathrm{k} = \mathrm{ij}$.

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. If one really wants to see something between the angle brackets, again one can write $\langle\varnothing\rangle$.

2.3. Cyclic groups

The **order** of a group is its size (or cardinality). The order of G is therefore denoted by

$$|G|.$$

A group is called **cyclic** if generated by a single element. If a is an element of a group G, then $\langle a\rangle$ is a cyclic subgroup of G, and the **order** of a, denoted by

$$|a|,$$

is defined to be the order of $\langle a\rangle$. In the next theorem, $x \mapsto a^x$ is the homomorphism from $\mathbb{Z}$ to G as in Theorem 24 (p. 41).

Theorem 34. *If a is an element of a group G, then*

$$\langle a\rangle = \mathrm{im}(x \mapsto a^x).$$

Proof. Let f be the homomorphism $x \mapsto a^x$. We have to show $\langle a\rangle = \mathrm{im}(f)$. Since $\langle a\rangle$ is a group, we know that $a^0 \in \langle a\rangle$, and if $a^n \in \langle a\rangle$, then $a^{-n} \in \langle a\rangle$. Also $a \in \langle a\rangle$, so that, if $a^n \in \langle a\rangle$, then $a^{n+1} \in \langle a\rangle$. By induction, $\mathrm{im}(f) \subseteq \langle a\rangle$. Since $a \in \mathrm{im}(f)$, we have $\langle a\rangle \subseteq \mathrm{im}(f)$ by definition of $\langle a\rangle$. □

Theorem 35. *If a is a group-element of finite order, then $a^{|a|} = \mathrm{e}$.*

Proof. The subset $\{\mathrm{e}, a, a^2, \dots, a^{|a|}\}$ of $\langle a\rangle$ has size at most $|a|$. Hence, for some i and j, we have $0 \leqslant i < j \leqslant |a|$, 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, $|a| \leqslant j-i$. Since also $j-i \leqslant |a|$, we have $|a| = j-i$, and in particular $a^{|a|} = a^{j-i} = \mathrm{e}$. □

Theorem 36. *All subgroups of $\mathbb{Z}$ are cyclic. All nontrivial subgroups of $\mathbb{Z}$ are isomorphic to one another.*

Proof. Say $G < \mathbb{Z}$ and $G \neq \langle\ \rangle$. Let m be the least positive element of G. Then $\langle m\rangle < G$.

Let n be an arbitrary element of G. Then $n = km + r$ for some k and r such that $0 \leqslant r < m$. Since $r = n - km$, we must have $r \in G$, so $r = 0$ by minimality of m. Hence $n = km$. Thus $G < \langle m\rangle$. Therefore $G = \langle m\rangle$.

The map $x \mapsto mx$ from $\mathbb{Z}$ to G is an epimorphism, by Theorem 34. The kernel of this map is trivial, simply because $mx = 0 \Rightarrow x = 0$. Therefore the map is an isomorphism, by Theorem 32. □

Theorem 37. *Every cyclic group is isomorphic to some $\mathbb{Z}_n$.*

Proof. Say $G = \langle a\rangle$. By Theorem 36, the epimorphism $x \mapsto a^x$ from $\mathbb{Z}$ to G has kernel $\langle n\rangle$ for some n; therefore

$$a^r = a^s \iff a^{r-s} = \mathrm{e} \iff r - s \in \langle n\rangle \iff n \mid r - s.$$

Hence the map $[x] \mapsto a^x$ is well-defined on $\mathbb{Z}_n$ and has trivial kernel. □

2.4. Cosets

Suppose $H < G$. If $a \in G$, let

$$\begin{aligned} aH &= \{ax \colon x \in H\}, \\ Ha &= \{xa \colon x \in H\}. \end{aligned}$$

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 48, 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 38. *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 13, p. 32.)

Inversion is a permutation of G taking aH to Ha^{-1}, so G/H and $H\backslash G$ must have the same size. □

Corollary. *If $H < G$, then the relation $\sim$ on G defined by*

$$a \sim x \iff aH = xH$$

is an equivalence-relation, and

$$G/H = G/{\sim}.$$

Corollary. *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 size of G/H (or of $H\backslash G$) is called the **index** of H in G and can be denoted by

$$[G : H].$$

This is a *cardinality,* though if it is infinite, we shall not generally be interested in which cardinality it is. 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 39. *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 38, $bK \subseteq aH$. 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. □

The remaining theorems of this section will not be needed later, though the ideas may be useful. In the next theorem and elsewhere, HK has the obvious meaning of $\{xy\colon x \in H \land y \in K\}$. It need not be a group. For example, in Sym(3), if $H = \langle(0\ 1)\rangle$ and $K = \langle(0\ 2)\rangle$, then $HK = \{\mathrm{e}, (0\ 1), (0\ 2), (0\ 2\ 1)\}$, which is not a group.

Theorem 40. *If H and K are finite subgroups of some group, then*

$$|HK| = \frac{|H||K|}{|H \cap K|}.$$

Proof. Since $H \cap K$ is a group by Theorem 33, and $H \cap K \subseteq H$, we have $H \cap K < H$. By Theorem 38, for some n in $\mathbb{N}$, for some a_0, …, a_{n-1} in H, we now have

$$H = a_0(H \cap K) \cup \cdots \cup a_{n-1}(H \cap K),$$

the union being disjoint. Then $|H| = n|H \cap K|$. Also, immediately

$$a_0K \cup \cdots \cup a_{n-1}K \subseteq HK.$$

We have also the reverse inclusion, since if $h \in H$ and $k \in K$, then $h = a_ik_1$ for some i in n and some k_1 in $H \cap K$, so that $hk = a_ik_1k$, which is in a_iK. Thus

$$a_0K \cup \cdots \cup a_{n-1}K = HK.$$

This union is disjoint. For, suppose $a_ik_i = a_jk_j$, where k_i and k_j are in K. Then ${a_j}^{-1}a_i = k_j{k_i}^{-1}$, which belongs both to H and to K. Thus ${a_j}^{-1}a_i \in H \cap K$. Hence we must have $a_i(H \cap K) = a_j(H \cap K)$, so that $a_i = a_j$. So the union above is disjoint, and therefore $|HK| = n|K|$. □

Note that in the foregoing theorem and proof, we have no need to name the group of which H and K are subgroups. If this group is G, then we have $|G| \geqslant |HK|$, and so

$$[G : H] = \frac{|G|}{|H|} \geqslant \frac{|K|}{|H \cap K|} = [K\colon H \cap K].$$

We proved this under the assumption that H and K are finite; but we can do without this assumption as follows:

Theorem 41. *Suppose H and K are subgroups of a group G. Then*

$$[H : H \cap K] \leqslant [G : K]. \tag{2.2}$$

If $[G : K]$ is finite, then it is equal to $[H : H \cap K]$ if and only if $G = HK$.

Proof. In the proof of the last theorem, we showed in effect that the function $x(H \cap K) \mapsto xK$ from $H/(H \cap K)$ to G/K is injective. This gives (2.2). The function is surjective if and only if $G = HK$. □

Theorem 42. *Suppose H and K are subgroups of a group G. Then*

$$[G : H \cap K] \leqslant [G : H][G : K],$$

If $[G : H]$ and $[G : K]$ are finite, then their product is equal to $[G : H \cap K]$ if and only if $G = HK$.

Proof. By Theorems 39 and 41,

$$[G : H \cap K] = [G : H][H : H \cap K] \leqslant [G : H][G : K].$$

Similarly the rest follows. □

2.5. Lagrange's Theorem

Theorem 43 (Lagrange). *If $H < G$ and G is finite, then $|H|$ divides $|G|$.*

Proof. Use Theorem 39 when $K = \langle \mathrm{e} \rangle$. □

Corollary. *Groups of prime order are cyclic.*

Proof. Say $|G| = p$. There is a in $G \smallsetminus \langle \mathrm{e} \rangle$, so $|a| > 1$; but $|a|$ divides p, so $|a| = p$, and therefore $G = \langle a \rangle$. □

Corollary. *If G is finite and $a \in G$, then $a^{|G|} = \mathrm{e}$.*

Proof. $a^{|a|} = \mathrm{e}$ and $|a|$ divides $|G|$. □

The first Sylow Theorem (Theorem 94) is a partial converse of Lagrange's Theorem. An application of Lagrange's Theorem is the remaining two theorems of this section. The theorems are part of number theory; but their proofs can be streamlined with group theory.

Lemma 10. ${\mathbb{Z}_n}^{\times} = \{[x] \in \mathbb{Z}_n \colon \gcd(x, n) = 1\}$.

Proof. $\gcd(m, n) = 1$ if and only if $am + bn = 1$ for some integers a and b; but this just means $[a][m] = 1$ for some a. □

Theorem 44 (Fermat). *If the prime p is not a factor of a, then*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{2.3}$$

Hence for all integers a,

$$a^{p} \equiv a \pmod{p}. \tag{2.4}$$

Proof. By the lemma, the order of ${\mathbb{Z}_p}^{\times}$ is $p - 1$. Hence (2.3) holds if $[a] \in {\mathbb{Z}_p}^{\times}$. Also by the lemma, if $p \nmid a$, then $[a] \in {\mathbb{Z}_p}^{\times}$. This proves the first claim, which implies (2.4) if $p \nmid a$. If $p \mid a$, then (2.4) holds easily. □

If $n \neq 0$, let the order of ${\mathbb{Z}_n}^{\times}$ be denoted by

$$\phi(n).$$

Theorem 45 (Euler). *If $\gcd(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod{n}$.*

Proof. If $\gcd(a, n) = 1$, then by the lemma, $[a] \in {\mathbb{Z}_n}^{\times}$. □

2.6. Normal subgroups

If $H < G$, we investigate the possibility of defining a multiplication on G/H so that

$$(xH)(yH) = xyH. \tag{2.5}$$

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 48 and 52, 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 (2.5) fails in this case.

As a corollary to Theorem 38 (p. 52), we have that the relation $\sim$ on G given by

$$a \sim x \iff aH = xH$$

is an equivalence-relation. Then there is a multiplication on G/H as desired if and only if this equivalence-relation is a congruence-relation (with respect to the multiplication on G). In this case, by Theorem 28 (p. 47), G/H is a group with respect to the proposed multiplication.

Theorem 46. *Suppose $H < G$. The following are equivalent:*

1. *G/H is a group whose multiplication is given by (2.5).*
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 a corollary to Theorem 38.

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 $(xH)(yH) = xHHy = xHy = xyH$. □

A subgroup H of G meeting any of these equivalent conditions is called **normal,** and in this case we write

$$H \lhd G.$$

Of abelian groups, all subgroups are normal. 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 47. *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). □

Theorem 48. *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 30. □

Theorem 49. *Suppose $N \lhd G$ and $H < G$ and $N \cap H = \langle \mathrm{e}\rangle$. Then the surjection $(x, y) \mapsto xy$ from $N \times H$ to NH is a bijection.*

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 $m = n$. □

In the theorem, NH is the **internal semidirect product** of N and H. Note well that the bijection between $N \times H$ and NH need not be an isomorphism of groups, since in $N \times H$

$$(m, g)(n, h) = (mn, gh),$$

while in NH

$$(mg)(nh) = (mgng^{-1})(gh), \tag{2.6}$$

and $mgng^{-1}$ need not be equal to mn, because gng^{-1} need not be equal to n. Theorem 67 on page 76 below establishes conditions under which the bijection between $N \times H$ and NH *is* an isomorphism. Semidirect products in general are treated in §3.6 (p. 88).

Theorem 50. *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. □

In the proof, the map $x \mapsto xN$ is the **canonical projection** or the **quotient map** of G onto G/N; it may be denoted by p or π.

Theorem 51. *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 46).*

$$\begin{tikzcd}
G \arrow[r, "\pi"] \arrow[d, "f"'] & G/N \arrow[dl, "\tilde{f}"] \\
H &
\end{tikzcd}$$

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 (First Isomorphism Theorem). *For every homomorphism f on a group G,*

$$G/\ker(f) \cong \operatorname{im}(f).$$

Proof. Let $N = \ker(f)$; then $\tilde{f}$ is the desired homomorphism. □

Corollary. *If f is a homomorphism from G to H, and N is a normal subgroup of G, and $M \lhd H$, and $f[N] < M$, then there is a homomorphism $\tilde{f}$ from G/N to H/M such that the following diagram commutes:*

$$\begin{tikzcd} G \arrow[r, "\pi"] \arrow[d, "f"'] & G/N \arrow[d, "\tilde{f}"] \\ H \arrow[r, "\pi"'] & H/M \end{tikzcd}$$

Proof. $N < \ker(\pi \circ f)$. □

Theorem 52 (Second Isomorphism). *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 (a corollary to Theorem 51). □

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

$$\frac{\langle n\rangle}{\langle \operatorname{lcm}(n, m)\rangle} \cong \frac{\langle \gcd(n, m)\rangle}{\langle m\rangle}.$$

Theorem 53 (Third Isomorphism). *If N and K are normal subgroups of G and $N < K$, then $K/N \lhd G/N$ and*

$$\frac{G/N}{K/N} \cong G/K.$$

Proof. By (a corollary to) Theorem 51, the map $xN \mapsto xK$ from G/N to G/K is a well-defined epimorphism. The kernel contains xN if and only if $x \in K$, that is, $xN \in K/N$. Again the claim now follows by the First Isomorphism Theorem (a corollary to Theorem 51). □

Theorem 51 will also be used to prove von Dyck's Theorem (Theorem 74, p. 84).

Lemma 11. *If f is an epimorphism from G onto H, then there is a one-to-one correspondence $K \mapsto f[K]$ between subgroups of G that include* $\ker(f)$ *and subgroups of H; under this, normal subgroups correspond.*

$$\begin{tikzcd}
K \arrow[r] \arrow[d] & G \arrow[d, two heads, "f"] \\
f[K] \arrow[r] & H
\end{tikzcd}$$

Theorem 54. *If $N \lhd G$, then every subgroup of G/N is K/N for some subgroup K of G that includes N, and moreover K/N is normal in G/N if and only if K is normal in G.*

$$\begin{tikzcd}
K \arrow[r] \arrow[d] & G \arrow[d, two heads, "f"] \\
K/N \arrow[r] & G/N
\end{tikzcd}$$

Proof. Use the lemma in case H is G/N and f is π. □

2.7. Finite groups

Since every group can be considered as a symmetry group of *itself,* every *finite* group G can be considered as a symmetry group of finite set. In particular, G can be considered as a subgroup of $\mathrm{Sym}(n)$ for some n in ω. As promised on page 30, we now show:

Theorem 55. *Every element of* $\mathrm{Sym}(n)$ *is a composite of disjoint cycles of length at least* 2*, uniquely up to order of factors.*

Proof. Let $\sigma \in \mathrm{Sym}(n)$. If $k \in n$, let

$$[k] = \{\sigma^{\ell}(k) \colon \ell \in \mathbb{Z}\}.$$

Then the sets $[k]$ partition n: we have

$$n = [k_0] \cup \cdots \cup [k_{\ell-1}]$$

for some ℓ, the union being disjoint. If $i \in \ell$, define σ_i by

$$\sigma_i(x) = \begin{cases} \sigma(x), & \text{if } x \in [k_i], \\ x, & \text{otherwise.} \end{cases}$$

If $[k_i]$ has size ℓ_i, then σ_i is the ℓ_i-cycle $\begin{pmatrix} k & \sigma(k) & \cdots & \sigma^{\ell_i-1}(k) \end{pmatrix}$. Finally, σ is the composite of all of the σ_i such that $\ell_i > 1$. □

Theorem 56. *The order of a finite permutation is the least common multiple of the orders of its disjoint cyclic factors.*

A 2-cycle is also called a **transposition.**

Theorem 57. *Every finite permutation is a product of transpositions.*

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

Let the set of 2-element subsets of n by denoted by

$$[n]^2.$$

If $\sigma \in \operatorname{Sym}(n)$, and $\{i, j\} \in [n]^2$, then we can define

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

Thus we have a homomorphism from $\operatorname{Sym}(n)$ to $\operatorname{Sym}([n]^2)$. Understanding n as the subset $\{0, \dots, n-1\}$ of $\mathbb{Q}$, we have a function $X \mapsto q_\sigma(X)$ from $[n]^2$ to $\mathbb{Q}^\times$ given by

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

Then we can define the function $\sigma \mapsto \operatorname{sgn}(\sigma)$ from $\operatorname{Sym}(n)$ into $\mathbb{Q}^\times$ by

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

Theorem 58. *The function* $\sigma \mapsto \operatorname{sgn}(\sigma)$ *is an homomorphism from* $\operatorname{Sym}(n)$ *onto the subgroup* $\langle -1 \rangle$ *of* $\mathbb{Q}^\times$*; it takes every transposition to* -1.

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

$$\begin{aligned} \operatorname{sgn}(\sigma) &= q_\sigma(\{k, \ell\}) \prod_{i \in n \smallsetminus \{k, \ell\}} (q_\sigma(\{i, \ell\}) q_\sigma(\{k, i\})) \\ &= \frac{\ell - k}{k - \ell} \cdot \prod_{i \in n \smallsetminus \{k, \ell\}} \left(\frac{i - k}{i - \ell} \cdot \frac{\ell - i}{k - i} \right) = -1. \end{aligned}$$

If σ and τ are arbitrary elements of $\operatorname{Sym}(n)$, then

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

since τ permutes $[n]^2$. □

The value $\operatorname{sgn}(\sigma)$ can now be called the **signum** of σ; it is 1 if and only if σ is the product of an even number of transpositions. Such a product is itself called **even;** the other permutations, with signum -1, are called **odd.**

The **alternating group** of degree n is the kernel of $\sigma \mapsto \operatorname{sgn}(\sigma)$ on $\operatorname{Sym}(n)$ and is denoted by

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

Hence $\operatorname{Alt}(n) \lhd \operatorname{Sym}(n)$ and $[\operatorname{Sym}(n) : \operatorname{Alt}(n)] = 2$.

A group is **simple** if it has no proper nontrivial normal subgroups. For example, $\mathbb{Z}_n$ is simple just in case $|n|$ is prime. Hence the only simple abelian groups are the $\mathbb{Z}_p$, where p is prime.

Lemma 12. *$\operatorname{Alt}(n)$ is generated by the 3-cycles in $\operatorname{Sym}(n)$.*

Proof. The group $\operatorname{Alt}(n)$ 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 $\operatorname{Alt}(n)$, and this group is generated by these cycles. □

Lemma 13. *$\operatorname{Alt}(n)$ is generated by the 3-cycles $\begin{pmatrix}0 & 1 & k\end{pmatrix}$, where $1 < k < n$.*

Proof. If a, b, and c are distinct elements of $n \smallsetminus \{0,1\}$, then

$$\begin{aligned}\begin{pmatrix}0 & a & b\end{pmatrix} &= \begin{pmatrix}0 & 1 & b\end{pmatrix}\begin{pmatrix}a & 1 & 0\end{pmatrix} = \begin{pmatrix}0 & 1 & b\end{pmatrix}\begin{pmatrix}0 & 1 & a\end{pmatrix}^{-1},\\ \begin{pmatrix}1 & a & b\end{pmatrix} &= \begin{pmatrix}1 & 0 & b\end{pmatrix}\begin{pmatrix}a & 0 & 1\end{pmatrix} = \begin{pmatrix}0 & 1 & b\end{pmatrix}^{-1}\begin{pmatrix}0 & 1 & a\end{pmatrix},\\ \begin{pmatrix}a & b & c\end{pmatrix} &= \begin{pmatrix}c & 1 & 0\end{pmatrix}\begin{pmatrix}0 & a & b\end{pmatrix}\begin{pmatrix}0 & 1 & c\end{pmatrix}.\end{aligned}$$ □

Lemma 14. *Any normal subgroup of $\operatorname{Alt}(n)$ containing a 3-cycle is $\operatorname{Alt}(n)$.*

Proof. We show that every 3-cycle is conjugate in $\operatorname{Alt}(n)$ to a cycle $\begin{pmatrix}0 & 1 & k\end{pmatrix}$. It is enough to note that

$$\begin{pmatrix}a & b & d\end{pmatrix} = \underbrace{\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}c & d\end{pmatrix}}\begin{pmatrix}c & b & a\end{pmatrix}\underbrace{\begin{pmatrix}c & d\end{pmatrix}\begin{pmatrix}a & b\end{pmatrix}}. \qquad \square$$

Lemma 15. *If $n > 4$, then a normal subgroup of $\operatorname{Alt}(n)$ contains a 3-cycle, provided it has a nontrivial element whose factorization into disjoint cycles contains one of the following:*

1. *a cycle of length at least 4;*
2. *two cycles of length 3;*
3. *transpositions, only one 3-cycle, and no other cycles; or*
4. *only transpositions.*

Proof. 1. If $k \geqslant 4$, and σ is disjoint from $\begin{pmatrix}0 & 1 & \dots & k-1\end{pmatrix}$, then

$$\begin{aligned}\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}0 & 1 & \dots & k-1\end{pmatrix}\sigma\begin{pmatrix}2 & 1 & 0\end{pmatrix}\sigma^{-1}\begin{pmatrix}k-1 & \dots & 1 & 0\end{pmatrix}&\\ =\begin{pmatrix}0 & 1 & 3\end{pmatrix}.&\end{aligned}$$

2. If σ is disjoint from $\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}3 & 4 & 5\end{pmatrix}$, then we reduce to the previous case:

$$\begin{aligned}\begin{pmatrix}0 & 1 & 3\end{pmatrix}\underbrace{\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}3 & 4 & 5\end{pmatrix}}\sigma\begin{pmatrix}3 & 1 & 0\end{pmatrix}\sigma^{-1}\underbrace{\begin{pmatrix}5 & 4 & 3\end{pmatrix}\begin{pmatrix}2 & 1 & 0\end{pmatrix}}&\\ =\begin{pmatrix}0 & 1 & 4 & 2 & 3\end{pmatrix}.&\end{aligned}$$

3. If σ is disjoint from $\begin{pmatrix}0 & 1 & 2\end{pmatrix}$ and is the product of transpositions, then

$$\left[\begin{pmatrix}0 & 1 & 2\end{pmatrix}\sigma\right]^2 = \begin{pmatrix}2 & 1 & 0\end{pmatrix}.$$

4. If σ is a product of transpositions disjoint from $\begin{pmatrix}0 & 1\end{pmatrix}$ and $\begin{pmatrix}2 & 3\end{pmatrix}$, then

$$\begin{pmatrix}0 & 1 & 2\end{pmatrix}\underbrace{\begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}2 & 3\end{pmatrix}\sigma}\begin{pmatrix}2 & 1 & 0\end{pmatrix}\underbrace{\sigma\begin{pmatrix}3 & 2\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix}} = \begin{pmatrix}0 & 2\end{pmatrix}\begin{pmatrix}1 & 3\end{pmatrix},$$

$$\begin{pmatrix}0 & 2 & 4\end{pmatrix}\underbrace{\begin{pmatrix}0 & 2\end{pmatrix}\begin{pmatrix}1 & 3\end{pmatrix}}\begin{pmatrix}4 & 2 & 0\end{pmatrix}\underbrace{\begin{pmatrix}3 & 1\end{pmatrix}\begin{pmatrix}2 & 0\end{pmatrix}} = \begin{pmatrix}0 & 4 & 2\end{pmatrix}. \qquad \square$$

Theorem 59. Alt(n) *is simple if and only if* $n \neq 4$.

Proof. Alt(1) and Alt(2) are trivial, and Alt(3) $\cong \mathbb{Z}_3$. The case when $n > 4$ is handled by the previous lemmas. Finally, every element of Alt(4) (in fact, of Sym(4)) can be considered as a permutation of the set

$$\Big\{\{\{0,1\},\{2,3\}\},\{\{0,2\},\{1,3\}\},\{\{0,3\},\{1,2\}\}\Big\}.$$

Thus we get an epimorphism from Alt(4) to Sym(3) whose kernel is therefore a proper nontrivial normal subgroup. □

The normal subgroup of Alt(4) found in the proof is

$$\langle \begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}2 & 3\end{pmatrix}, \begin{pmatrix}0 & 2\end{pmatrix}\begin{pmatrix}1 & 3\end{pmatrix}, \begin{pmatrix}0 & 3\end{pmatrix}\begin{pmatrix}1 & 2\end{pmatrix}\rangle.$$

We can obtain it by considering Alt(4) as the group of rotational symmetries of the regular tetrahedron. The vertices of this tetrahedron can be taken as 4 of the 8 vertices of a cube: say, the vertices with coordinates $(1,1,1)$, $(1,-1,-1)$, $(-1,1,-1)$, and $(-1,-1,1)$. Then a symmetry of the tetrahedron determines a permutation of the 3 coordinate axes, hence an element of Sym(3).

2.8. Determinants

Let R be a commutative ring. We define the function $X \mapsto \det(X)$ from $\mathrm{M}_n(R)$ to R by

$$\det((a^i_j)^{i<n}_{j<n}) = \sum_{\sigma\in\mathrm{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{i<n} a^i_{\sigma(i)}.$$

Theorem 60. *The function* $X \mapsto \det(X)$ *is a multiplicative homomorphism, that is,*

$$\det(XY) = \det(X)\det(Y).$$

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 φ 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 σ' 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 σ and σ' 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 τ in $\mathrm{Sym}(n)$ permutes n, we have also

$$\prod_{i<n} b^{\tau(i)}_{\sigma(i)} = \prod_{i<n} b^i_{\sigma\tau^{-1}(i)}, \qquad \operatorname{sgn}(\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma\tau^{-1}).$$

Putting this all together, we have

$$\begin{aligned}\det(AB) &= \sum_{\tau\in\operatorname{Sym}(n)} \prod_{i<n} a^i_{\tau(i)} \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma\tau^{-1}) \prod_{i<n} b^i_{\sigma\tau^{-1}(i)}\\ &= \sum_{\tau\in\operatorname{Sym}(n)} \operatorname{sgn}(\tau)\prod_{i<n} a^i_{\tau(i)} \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma\tau^{-1}) \prod_{i<n} b^i_{\sigma\tau^{-1}(i)}\\ &= \sum_{\tau\in\operatorname{Sym}(n)} \operatorname{sgn}(\tau)\prod_{i<n} a^i_{\tau(i)} \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{i<n} b^i_{\sigma(i)}\\ &= \det A \det B,\end{aligned}$$

since $\sigma \mapsto \sigma\tau^{-1}$ is a permutation of $\operatorname{Sym}(n)$. □

Corollary. *An element A of $\mathrm{M}_n(R)$ has an inverse only if $\det(A) \in R^\times$.*

Theorem 61. *An element A of $\mathrm{M}_n(R)$ has an inverse if $\det(A) \in R^\times$.*

Proof. Let $A = (a^i_j)^{i<n}_{j<n}$. If $i < n$, then

$$\begin{aligned}\det(A) &= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{\ell<n} a^\ell_{\sigma(\ell)}\\ &= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) a^i_{\sigma(i)} \prod_{\ell\in n\smallsetminus\{i\}} a^\ell_{\sigma(\ell)}\\ &= \sum_{j<n} a^i_j \sum_{\substack{\sigma\in\operatorname{Sym}(n)\\ \sigma(i)=j}} \operatorname{sgn}(\sigma) \prod_{\ell\in n\smallsetminus\{i\}} a^\ell_{\sigma(\ell)}\\ &= \sum_{j<n} a^i_j b^j_i,\end{aligned}$$

where

$$b^j_k = \sum_{\substack{\sigma\in\operatorname{Sym}(n)\\ \sigma(k)=j}} \operatorname{sgn}(\sigma) \prod_{\ell\in n\smallsetminus\{k\}} a^\ell_{\sigma(\ell)}.$$

However, if $i \neq k$, then

$$\begin{aligned}\sum_{j<n} a^i_j b^j_k &= \sum_{j<n} a^i_j \sum_{\substack{\sigma\in\operatorname{Sym}(n)\\ \sigma(k)=j}} \operatorname{sgn}(\sigma) \prod_{\ell\in n\smallsetminus\{k\}} a^\ell_{\sigma(\ell)}\\
&= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) a^i_{\sigma(k)} \prod_{\ell\in n\smallsetminus\{k\}} a^\ell_{\sigma(\ell)}\\
&= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) a^i_{\sigma(k)} a^i_{\sigma(i)} \prod_{\ell\in n\smallsetminus\{i,k\}} a^\ell_{\sigma(\ell)} = 0,\end{aligned}$$

since the map $\sigma \mapsto \sigma \circ \begin{pmatrix} i & k \end{pmatrix}$ is a bijection between $\operatorname{Alt}(n)$ and $\operatorname{Sym}(n) \smallsetminus \operatorname{Alt}(n)$. Thus

$$A(b^j_k)^{j<n}_{k<n} = (\det(A)\delta^i_k)^{i<n}_{k<n}.$$

Finally,

$$\begin{aligned}\sum_{j<n} b^i_j a^j_k &= \sum_{j<n} \sum_{\substack{\sigma\in\operatorname{Sym}(n)\\ \sigma(j)=i}} \operatorname{sgn}(\sigma) \prod_{\ell\in n\smallsetminus\{j\}} a^\ell_{\sigma(\ell)} a^j_k\\
&= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{\ell\in n\smallsetminus\{\sigma^{-1}(i)\}} a^\ell_{\sigma(\ell)} a^{\sigma^{-1}(i)}_k\\
&= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{\ell\in n\smallsetminus\{i\}} a^{\sigma^{-1}(\ell)}_\ell a^{\sigma^{-1}(i)}_k,\end{aligned}$$

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

$$A^{-1} = (\det(A)^{-1} b^j_k)^{j<n}_{k<n}.$$ □

2.9. Dihedral groups

We can consider the elements of n as vertices of a regular n-gon. The group of symmetries of this object is called a **dihedral group** and is denoted by

$$D_n.$$

Formally, this is the subgroup $\langle\sigma_n, \beta\rangle$ of $\operatorname{Sym}(n)$, where as in the last section σ_n is the n-cycle $\begin{pmatrix}0 & 1 & \dots & n-1\end{pmatrix}$, while

$$\beta = \begin{cases} \begin{pmatrix}1 & n-1\end{pmatrix}\begin{pmatrix}2 & n-2\end{pmatrix}\cdots\begin{pmatrix}m-1 & m+1\end{pmatrix}, & \text{if } n = 2m, \\ \begin{pmatrix}1 & n-1\end{pmatrix}\begin{pmatrix}2 & n-2\end{pmatrix}\cdots\begin{pmatrix}m & m+1\end{pmatrix}, & \text{if } n = 2m+1. \end{cases}$$

Note that both β and $\sigma_n\beta$ here have order 2.

Theorem 62. *If $n > 2$, and $G = \langle a, b\rangle$, where $|a| = n$ and $|b| = 2 = |ab|$, then $G \cong D_n$.*

Proof. Assume $n \geqslant 2$. Since $abab = \mathrm{e}$ and $b^{-1} = b$, we have

$$ba = a^{-1}b, \qquad\qquad ba^{-1} = ab.$$

Therefore $ba^k = a^{-k}b$ for all integers k. This shows

$$G = \{a^i b^j \colon (i, j) \in n \times 2\}.$$

It remains to show $|G| = 2n$. Suppose

$$a^i b^j = a^k b^\ell,$$

where (i, j) and (k, ℓ) 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 $aa^maa^m = a^2$, while $abab = \mathrm{e}$, so $n = 2$. □

3. Category theory

3.1. Products and sums

Theorem 63. *Let G_0, G_1 and H be groups. For each i in 2, let π_i be the homomorphism $(x_0, x_1) \mapsto x_i$ from $G_0 \times G_1$ to G_i, and let f_i be a homomorphism from H to G_i. Then there is a homomorphism*

$$x \mapsto (f_0(x), f_1(x))$$

from H to $G_0 \times G_1$, and this 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:

$$\begin{tikzcd}
G_0 & G_0 \times G_1 \arrow[l, "\pi_0"'] \arrow[r, "\pi_1"] & G_1 \\
& H \arrow[ul, "f_0"] \arrow[u, "f"'] \arrow[ur, "f_1"'] &
\end{tikzcd}$$

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 $f(x) = (f_0(x), f_1(x))$. □

We can generalize this theorem by considering an indexed family $(G_i \colon i \in I)$ of groups. The **direct product** of this family is denoted by

$$\prod_{i \in I} G_i.$$

This is, first of all, the set whose elements are $(x_i \colon i \in I)$ (that is, functions $i \mapsto x_i$ on I) such that $x_i \in G_i$ for each i in I. An operation of multiplication on this set is given by

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

Under this multiplication, $\prod_{i \in I} G_i$ becomes a group. If $i \in I$, we define a homomorphism π_i from $\prod_{i \in I} G_i$ to G_i by

$$\pi_i(x_j \colon j \in I) = x_i.$$

In case $I = n$, we may write $\prod_{i \in I} G_i$ also as

$$G_0 \times \dots \times G_{n-1},$$

and a typical element of this as

$$(x_0, \dots, x_{n-1}).$$

To the previous theorem we have:

Porism. *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*

$$x \mapsto (f_i(x) \colon i \in I)$$

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,

$$\pi_i f = f_i.$$

The direct product of a family of abelian groups is an abelian group. When we restrict attention to abelian groups, then we can reverse the arrows in Theorem 63:

Theorem 64. *Let G_0, G_1 and H be abelian groups. Let ι_0 be the homomorphism $x \mapsto (x, 0)$ from G_0 to $G_0 \oplus G_1$, and let ι_1 be $x \mapsto (0, x)$ from G_1 to $G_0 \oplus G_1$. For each i in 2, let f_i be a homomorphism from G_i to H. Then there is a homomorphism*

$$(x_0, x_1) \mapsto f_0(x_0) + f_1(x_1)$$

from $G_0 \oplus G_1$ to H, and this 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. Every element (x_0, x_1) of $G_0 \oplus G_1$ is $\iota_0(x_0) + \iota_1(x_1)$, so that, if f is a homomorphism on $G_0 \oplus G_1$, then

$$f(x_0, x_1) = f\iota_0(x_0) + f\iota_1(x_1). \tag{3.1}$$

Hence f is as desired if and only if $f(x_0, x_1) = f_0(x_0) + f_1(x_1)$. The function so defined is indeed a homomorphism, since

$$\begin{aligned}
f((x_0, x_1) + (u_0, u_1)) &= f(x_0 + u_0, x_1 + u_1) = f_0(x_0 + u_0) + f_1(x_1 + u_1) \\
&= f_0(x_0) + f_0(u_0) + f_1(x_1) + f_1(u_1) \\
&= f_0(x_0) + f_1(x_1) + f_0(u_0) + f_1(u_1) = f(x_0, x_1) + f(u_0, u_1),
\end{aligned}$$

because H is abelian. □

In the proof, the definition of f in (3.1) relies on the *finiteness* of the family $(G_i : i \in 2)$; more precisely, it relies on the finiteness of $\{i \in 2 : x_i \neq e)$. Of an arbitrary indexed family $(G_i : i \in I)$ of groups, we define the ***weak* direct product** to be the subgroup, denoted by

$$\prod_{i \in I}^{\mathrm{w}} G_i,$$

of $\prod_{i \in I} G_i$ comprising those elements $(x_i : i \in I)$ such that $\{i \in I : x_i \neq \mathrm{e}\}$ is finite. We define a homomorphism ι_i from each G_i to $\prod_{j \in I}^{\mathrm{w}} G_j$ by

$$\iota_i(x) = (x_j : j \in I),$$

where

$$x_j = \begin{cases} x, & \text{if } j = i; \\ \mathrm{e}, & \text{otherwise.} \end{cases}$$

If I is finite, then the weak direct product is the same as the (full) direct product. If I is infinite, and the groups G_i are nontrivial for infinitely many i in I, then the weak direct product is *not* the same as the direct product; but the proof uses the Axiom of Choice.

Proving that f as in (3.1) is a *homomorphism* uses that H is abelian. The weak direct product of a family $(G_i \colon i \in I)$ of abelian groups is called the **direct sum** and is denoted by

$$\sum_{i \in I} G_i.$$

In case $I = n$, we may write $\sum_{i \in I} G_i$ also as

$$G_0 \oplus \cdots \oplus G_{n-1}.$$

To the previous theorem we have:

Porism. *Suppose $(G_i \colon i \in I)$ 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,

$$f \iota_i = f_i.$$

Now we can provide an example promised in §1.6. Let E be the abelian group $\sum_{n \in \omega} \mathbb{Z}$. Suppose f is a singulary operation on ω. An element f^* of $\mathrm{End}(E)$ is induced, given by

$$f^*(x_n \colon n \in \omega) = (x_{f(n)} \colon n \in \omega).$$

Then $f^*\iota_{f(n)} = \iota_n$. Let f be the operation $x \mapsto x+1$ on ω, and let g be the operation given by

$$g(x) = \begin{cases} y, & \text{if } f(y) = x, \\ 0, & \text{if } x = 0. \end{cases}$$

Then $gf(x) = x$, so $f^*g^* = (gf)^*$, the identity in $\mathrm{End}(E)$; but g^*f^* is not the identity, since it is $(fg)^*$, and $fg(0) = 1 = fg(1)$.

We have two kinds of products so far, related as follows.

Theorem 65. *Let $(G_i \colon i \in I)$ be an indexed family of groups. Then*

$$\iota_j[G_j] \lhd \prod_{i\in I}^{\mathrm{w}} G_i, \qquad \prod_{i\in I}^{\mathrm{w}} G_i \lhd \prod_{i\in I} G_i, \qquad \iota_j[G_j] \lhd \prod_{i\in I} G_i.$$

Theorem 64 and its porism can be generalized to some cases of arbitrary groups:

Theorem 66. *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 f_i from G_i to H. Suppose further that, for all i and j in I,*

$$f_i(x)f_j(y) = f_j(y)f_i(x).$$

Then the map

$$x \mapsto \prod_{i\in I} f_i(x_i)$$

from $\prod_{i\in I}^{\mathrm{w}} G_i$ to H is the unique homomorphism f from $\prod_{i\in I}^{\mathrm{w}} G_i$ to H such that, for each i in I,

$$f\iota_i = f_i.$$

As a special case of this theorem, we have the next theorem below, by means of the following:

Lemma 16. *If M and N are normal subgroups of G, and*

$$M \cap N = \langle \mathrm{e} \rangle,$$

then each element m of M commutes with each element n of N, that is,

$$mn = nm.$$

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 67. *If $(N_i : i \in I)$ is an indexed family of normal subgroups of a group, and for each j in I,*

$$N_j \cap \Big\langle \bigcup_{i \in I \smallsetminus \{j\}} N_i \Big\rangle = \langle \mathrm{e} \rangle, \tag{3.2}$$

then

$$\Big\langle \bigcup_{i \in I} N_i \Big\rangle \cong \prod_{i \in I}^{\mathrm{w}} N_i.$$

Proof. Say the N_i are normal subgroups of G. Since $N_i \cap N_j = \langle \mathrm{e} \rangle$ whenever $i \neq j$, the last theorem and the lemma guarantee that there is a homomorphism h from $\prod_{i \in I}^{\mathrm{w}} N_i$ into G such that, for each i in I, the composition $h\iota_i$ is just the inclusion of N_i in G. Then the range of h is $\Big\langle \bigcup_{i \in I} N_i \Big\rangle$. To show that h is injective, note that, if $n \in \prod_{i \in I}^{\mathrm{w}} N_i$ and $h(n) = \mathrm{e}$, then, for each j in I, we have

$$n_j{}^{-1} = \prod_{i \in I \smallsetminus \{j\}} n_i.$$

The left member is in N_j, the right in $\Big\langle \bigcup_{i \in I \smallsetminus \{j\}} N_i \Big\rangle$, so each side is e; in particular, $n_j = \mathrm{e}$. Therefore $n = \mathrm{e}$. □

In the conclusion of the theorem, G is the ***internal* weak direct product** of the N_i.

3.2. Free groups

The direct sum $\sum_{i \in I} \mathbb{Z}$ has elements e^i, namely $\iota_i(1)$ or $(\delta^i_j : j \in I)$, where

$$\delta^i_j = \begin{cases} 1, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

An arbitrary element of $\sum_{i\in I}$ is a **'formal sum'**,

$$\sum_{i\in I} x_i\,\mathrm{e}^i\,.$$

Theorem 68. *Suppose G is an abelian group, I is a set, and f is a function from I to G. Then the map*

$$\sum_{i\in I} x_i\,\mathrm{e}^i \mapsto \sum_{i\in I} x_i f(i)$$

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,

$$\tilde{f}(\mathrm{e}^i) = f(i)$$

—that is, the following diagram commutes, where ι is the map $i \mapsto \mathrm{e}^i$:

$$\begin{tikzcd} I \arrow[r, "\iota"] \arrow[d, "f"'] & \sum_{i\in I}\mathbb{Z} \arrow[dl, "\tilde{f}"] \\ G & \end{tikzcd}$$

The direct sum $\sum_{i\in I}\mathbb{Z}$ in the theorem is the **free abelian group** on I with respect to the map $i \mapsto \mathrm{e}^i$. There is also a **free group** on I, which we may denoted by

$$\mathrm{F}(I).$$

This is the group of *reduced words* on I. A **word** on I is a finite nonempty string $t_0t_1\cdots t_n$, 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 (so e can appear only in the string e). We make $\mathrm{F}(I)$ into a group when the multiplication is defined as 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, when an element a of I is written as a^{+1}, we have

$$(a_m^{\epsilon(m)}\cdots a_0^{\epsilon(0)})(b_0^{\zeta(0)}\cdots b_n^{\zeta(n)}) = a_m^{\epsilon(m)}\cdots a_j^{\epsilon(j)} b_j^{\zeta(j)}\cdots b_n^{\zeta(n)},$$

where j is maximal such that, if $i < j$, then $a_i^{\epsilon(i)} = b_i^{-\zeta(i)}$. 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 $\mathbb{Z} \smallsetminus \{0\}$.

Theorem 69. *Suppose G is a group, I is a set, and f is a function from I to G. Then the map*

$$a_0^{\epsilon(0)} \cdots a_n^{\epsilon(n)} \mapsto f(a_0)^{\epsilon(0)} \cdots f(a_n)^{\epsilon(n)}$$

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 ι is the inclusion of I in $\mathrm{F}(I)$:

$$\begin{tikzcd} I \arrow[r, "\iota"] \arrow[d, "f"'] & \mathrm{F}(I) \arrow[dl, "\tilde{f}"] \\ G & \end{tikzcd}$$

The **free product** of a family $(G_i \colon i \in I)$ of groups is the group, denoted by

$$\prod_{i\in I}^{*} G_i,$$

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 (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 ι_j, namely $x \mapsto (x, j)$. We now have the following analogue of the porism to Theorem 64.

Theorem 70. *Let $(G_i : i \in I)$ be an indexed family of groups, and let H be a group. Suppose for each i in I there is a homomorphism f_i from G_i to H. Then there is a homomorphism*

$$(g_0, n(0)) \cdots (g_m, n(m)) \mapsto f_{n(0)}(g_0) \cdots f_{n(n)}(g_n)$$

from $\prod^*_{i \in I} G_i$ to H; this is the unique homomorphism f from $\prod^*_{i \in I} G_i$ to H such that, for each i in I,

$$f \iota_i = f_i$$

—that is, the following diagram commutes:

$$\begin{tikzcd}
G_j \arrow[r, "\iota_j"] \arrow[dr, "f_j"'] & \prod\limits^*_{i \in I} G_i \arrow[d, "f"] \\
& H
\end{tikzcd}$$

3.3. Categories

For any two groups G and H there is a set

$$\mathrm{Hom}(G, H)$$

comprising the homomorphisms from G to H. There is a map

$$(g, f) \mapsto g \circ f$$

from $\mathrm{Hom}(H, K) \times \mathrm{Hom}(G, H)$ to $\mathrm{Hom}(G, K)$, and there is an element id_H of $\mathrm{Hom}(H, H)$, such that

$$\mathrm{id}_H \circ f = f, \quad g \circ \mathrm{id}_H = g, \quad k \circ (g \circ f) = (k \circ g) \circ f$$

whenever $f \in \mathrm{Hom}(G, H)$, $g \in \mathrm{Hom}(H, K)$, and $k \in \mathrm{Hom}(K, L)$. Understood in this way, groups with their homomorphisms compose a prototypical example of a *category.*

A **directed graph** is a certain kind of quadruple

$$(\mathbf{C}_0, \mathbf{C}_1, t, h),$$

where $\mathbf{C}_0$ and $\mathbf{C}_1$ are classes, and t and h are functions from $\mathbf{C}_1$ to $\mathbf{C}_0$. We may refer to each element of $\mathbf{C}_0$ as a **node,** and to each element of $\mathbf{C}_1$ as an **arrow.** If a is an arrow, then $t(a)$ is its **tail,** and $h(a)$ is its **head,** and a is an arrow **from** $t(a)$ **to** $t(b)$. If f is an arrow from A to B, we may express this by writing

$$f\colon A \longrightarrow B \qquad\qquad \text{or} \qquad\qquad A \xrightarrow{f} B.$$

We require the arrows from A to B to compose a *set* (as opposed to a proper class, like the class of all sets that do not contain themselves). We can define

$$\mathbf{C}_2 = \{(f,g) \in {\mathbf{C}_1}^2 \colon t(f) = h(g)\};$$

this is the class of paths of length 2. More generally,

$$\mathbf{C}_{n+1} = \Big\{(f_0,\dots,f_n) \in {\mathbf{G}_1}^{n+1}\colon \bigwedge_{i<n} t(f_i) = h(f_{i+1})\Big\}.$$

The graph above is a **category** if there are

1. a function $A \mapsto \mathrm{id}_A$ from $\mathbf{C}_0$ to $\mathbf{C}_1$, and
2. a function $(f,g) \mapsto f \circ g$ from $\mathbf{C}_2$ to $\mathbf{C}_1$,

such that

$$t(\mathrm{id}_A) = A = h(\mathrm{id}_A), \qquad t(f\circ g) = t(g), \qquad h(f \circ g) = h(f),$$

and also

$$f \circ \mathrm{id}_{t(f)} = f, \qquad \mathrm{id}_{h(g)} \circ g = g, \qquad h \circ (g \circ f) = (h \circ g) \circ f \tag{3.3}$$

whenever these are defined. In particular then, the category is a sextuple

$$(\mathbf{C}_0, \mathbf{C}_1, t, h, \mathrm{id}, \circ), \tag{3.4}$$

meeting the conditions that we have discussed. In this case, the nodes—the elements of $\mathbf{C}_0$—are called **objects.** Conditions (3.3) can be diagrammed as follows.

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "f"'] & B \arrow[d, "g"] \\
B \arrow[r, "g"'] \arrow[ur, "\mathrm{id}_B"] & C
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
B \arrow[d, "g"'] & A \arrow[l, "f"'] \arrow[r, "f"] \arrow[d] \arrow[dl, "g\circ f"] & B \arrow[d, "g"] \arrow[dl, "h\circ g"] \\
C \arrow[r, "h"'] & D & C \arrow[l, "h"]
\end{tikzcd}$$

These are **commutative diagrams** in the sense that any two paths from one vertex to another represent the same arrow.[1] The arrows of a category are also called **morphisms.** The class of morphisms from A to B can be denoted by

$$\mathrm{Hom}(A,B).$$

The morphism $f \circ g$ is the **composite** of f and g.

A category is **concrete** if each of its objects has an underlying set and the morphisms are functions in the way suggested by the notation. For example, the class of sets, with the class of functions, is a concrete category; likewise the class of groups, with homomorphisms, and the class of topological spaces, with continuous functions. However, not all categories are concrete. For example, if G is a group, then its elements can be considered as objects of a category in which $\mathrm{Hom}(a,b) = \{ba^{-1}\}$, $\mathrm{id}_a = 1$, and $c \circ d = cd$.

In a category, a morphism f is an **isomorphism** if

$$g \circ f = \mathrm{id}_{t(f)} \qquad \text{and} \qquad f \circ g = \mathrm{id}_{h(f)}$$

for some morphism g; then g is an **inverse** of f.

Theorem 71. *In a category, inverses are unique.*

Proof. If g and h are inverses of f, then $g = g \circ \mathrm{id}_{h(f)} = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{id}_{t(f)} \circ h = h$. □

If it exists, then the inverse of f is f^{-1}. It is immediate then that $(f^{-1})^{-1} = f$.

Suppose we have an arbitrary category as in (3.4) and an element $(A_i : i \in I)$ or A of $\mathbf{C}_0{}^I$ for some index-set I. If it exists, the **product** of A in the category is an element

$$\Bigl(\prod A, i \mapsto \pi_i\Bigr)$$

[1] One can define commutative diagrams formally. A **diagram** is a homomorphism from a directed graph to a category. One then thinks of the diagram as the graph with its nodes and arrows labelled with their images in the category. The diagram is **commutative** if every path in the graph with the same tail and head is sent to the same arrow in the category.

of $\mathbf{C}_0 \times {\mathbf{C}_1}^I$, where

$$\pi_i \colon \prod A \to A_i$$

for each i in I, such that, whenever $(B, i \mapsto f_i) \in \mathbf{C}_0 \times {\mathbf{C}_1}^I$, where $f_i \colon B \to A_i$ for each i in I, then there is a *unique* morphism f from B to $\prod A$ such that

$$\pi_i \circ f = f_i$$

for each i in I. Again this condition is expressed by a commutative diagram.

$$\begin{tikzcd}
 & \prod A \arrow[d, "\pi_j"] \\
H \arrow[ur, "f"] \arrow[r, "f_j"'] & A_j
\end{tikzcd}$$

The morphisms π_i are the **canonical projections.**

Theorem 72. *Any two products of the same family of objects in the same category are isomorphic.* □

The porism to Theorem 63 is that direct products are products in the category of groups *and* in the category of abelian groups.

Every category has a **dual,** in which the arrows are reversed. To be precise, the dual of $(\mathbf{C}_0, \mathbf{C}_1, t, h, \mathrm{id}, \circ)$ is $(\mathbf{C}_0, \mathbf{C}_1, h, t, \mathrm{id}, \circ')$, where $f \circ' g = g \circ f$. A **co-product** or **sum** in a category is a product in the dual. The co-product of A may be denoted by

$$\Bigl(\coprod A, i \mapsto \iota_i\Bigr) \quad \text{or} \quad \Bigl(\sum A, i \mapsto \iota_i\Bigr);$$

the morphisms ι_i are the **canonical injections.** The relevant commutative diagram is the following.

$$\begin{tikzcd}
A_j \arrow[r, "f_j"] \arrow[d, "\iota_j"'] & H \\
\coprod A \arrow[ur, "f"'] &
\end{tikzcd}$$

Thus the coproduct of an indexed family of objects should be the 'simplest' object that contains all of the 'information' contained in each of the original objects.

The porism to Theorem 64 is that direct sums are coproducts in the category of abelian groups. Theorem 70 is that free products are coproducts in the category of groups.

Suppose F is an object in a concrete category and I is a set. Then F is called **free** on I with respect to a function ι from I to F if for any function f from I to an object B, there is a unique morphism $\tilde{f}$ from F to B such that

$$\tilde{f} \circ \iota = f.$$

That is, the following diagram commutes (where the nodes and arrows, except $\tilde{f}$, are from the category of sets):

$$\begin{tikzcd} I \arrow[r, "\iota"] \arrow[dr, "f"'] & F \arrow[d, "\tilde{f}"] \\ & B \end{tikzcd}$$

Theorem 68 shows that free objects exist in the category of abelian groups; Theorem 69, in the category of groups.

3.4. Presentation of groups

Theorem 73. *Every group is isomorphic to a quotient of a free group.*

Proof. Since every group G is an image of the free group $\mathrm{F}(G)$, the claim follows by the First Isomorphism Theorem (a corollary to Theorem 51). □

Suppose G is a group, A is a set, $f\colon A \to G$, and $G = \langle f(a)\colon a \in A\rangle$. Suppose further $B \subseteq \mathrm{F}(A)$, and N is the intersection of the set of normal subgroups of $\mathrm{F}(A)$ that include B. The quotient F/N, denoted by

$$\langle A \mid B\rangle,$$

is referred to as the group with **generators** A and **relations** B, even though, strictly, F/N here is generated, not by (the elements of) A, but by the cosets aN, where $a \in A$. If there is an isomorphism from $\langle A \mid B\rangle$ to G taking each of these cosets aN to $f(a)$, then $\langle A \mid B\rangle$ is called a **presentation** of G.

In this definition, rather than assuming $A \subseteq G$, we use the map f so as to allow the possibility that f is not injective. Also, if $A = \{a_0, \dots, a_n\}$, and $B = \{w_0, \dots, w_m\}$, then $\langle A \mid B\rangle$ can be written as $\langle a_0, \dots, a_n \mid w_0, \dots, w_m\rangle$.

For example, $\mathrm{F}(A)$ can be presented as $\langle A \mid \varnothing\rangle$, and in particular $\mathbb{Z}$ can be presented as $\langle a \mid \varnothing\rangle$, but also as $\langle a, b \mid ab^{-1}\rangle$. The group $\mathbb{Z}_n$ has the presentation $\langle a \mid a^n\rangle$.

Theorem 74 (von Dyck[2]). *Suppose G is a group, A is a set, and $f\colon A \to G$, and let $\tilde{f}$ be the induced homomorphism from $\mathrm{F}(A)$ to G. Suppose further $B \subseteq \mathrm{F}(A)$ and $\langle A \mid B\rangle = F/N$. If $\tilde{f}(w) = \mathrm{e}$ for each w in B, then there is a well-defined homomorphism g from $\langle A \mid B\rangle$ to G such that $g(aN) = f(a)$ for each a in A. If $G = \langle f(a) \colon a \in A\rangle$, then g is an epimorphism.*

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d] & G \\
\mathrm{F}(A) \arrow[ur, "\tilde{f}"'] \arrow[r, "\pi"'] & \langle A \mid B\rangle \arrow[u, "g"']
\end{tikzcd}$$

Proof. By definition of N, it is included in the kernel of $\tilde{f}$, so g is well-defined by Theorem 51. □

Theorem 75. *If $n > 2$, then D_n has the presentation $\langle a, b \mid a^n, b^2, abab\rangle$.*

Proof. Let $G = \langle a, b \mid a^n, b^2, abab\rangle$. Then the order of (the image of) a in G divides n, and the order of b divides 2. But by von Dyck's Theorem and Theorem 62, G maps onto D_n, and hence n divides the order of a in G, and 2 divides the order of b. Therefore $D_n \cong G$. □

[2] Walther von Dyck (1856–1934) gave an early (1882–3) definition of abstract groups [9, ch. 49, p. 1141].

Theorem 76. *The group* $\langle \mathrm{i}, \mathrm{j} \mid \mathrm{i}^4, \mathrm{i}^2\mathrm{j}^2, \mathrm{iji}^3\mathrm{j}\rangle$ *has order* 8*, and its elements are (the images of)* ± 1*,* $\pm\mathrm{i}$*,* $\pm\mathrm{j}$*,* $\pm\mathrm{k}$*, where* $1 = \mathrm{e}$ *and* $\mathrm{k} = \mathrm{ij}$ *and* $-x = \mathrm{i}^2 x$*.*

Proof. Let the group be called G. In G, we have $\mathrm{j}^2 = \mathrm{i}^{-2} = \mathrm{i}^2$, so $\mathrm{j}^4 = 1$. Hence also $\mathrm{k} = \mathrm{ij} = \mathrm{j}^3\mathrm{i}$, so $\mathrm{i}^3\mathrm{j} = \mathrm{ji}$. This shows that every element of G can be written as $\mathrm{i}^n\mathrm{j}^m$, where $n \in 4$ and $m \in 2$; hence it is one of the given elements. □

3.5. Finitely generated abelian groups

To **classify** a collection of groups is to find a function f such that

$$f(G) = f(H) \iff G \cong H$$

for all groups G and H in the collection. We do this now with the finitely generated abelian groups, and in particular with the finite abelian groups. The next theorem will be needed for Theorem 84.

Theorem 77. *For every abelian group* G *on* n *generators, there is a unique element* k *of* n*, along with positive integers* d_0*, …,* d_{k-1}*, where*

$$d_0 \mid \cdots \mid d_{k-1}, \tag{3.5}$$

such that

$$G \cong \mathbb{Z}_{d_0} \oplus \cdots \oplus \mathbb{Z}_{d_{k-1}} \oplus \underbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{n-k}. \tag{3.6}$$

Proof. Let F be the free abelian group $\sum_{i\in n} \mathbb{Z}$. Then

$$G \cong F/N,$$

where N is the kernel of the induced epimorphism from F onto G. As before, each element of F can be understood as a formal sum $\sum_{i\in n} x_i\,\mathrm{e}^i$. Then F itself is $\langle \mathrm{e}^0, \dots, \mathrm{e}^{n-1}\rangle$. If $N = \langle d_0\,\mathrm{e}^0, \dots, d_{k-1}\,\mathrm{e}^{k-1}\rangle$, then G is as in (3.6). Not every subgroup of F is given to us so neatly, but we can use linear algebra to put it into this form. Every element of F, considered as a formal sum, can be written also as a matrix product:

$$x_0 a^0 + \cdots + x_{n-1} a^{n-1} = \begin{pmatrix} x_0 & \cdots & x_{n-1} \end{pmatrix} \begin{pmatrix} \mathrm{e}^0 \\ \vdots \\ \mathrm{e}^{n-1} \end{pmatrix} = \boldsymbol{x}\mathbf{e}.$$

The generators of a (finitely generated) subgroup of F can be considered as the entries of a column vector, and this column can be considered as the product of a matrix over $\mathbb{Z}$ with $\mathbf{e}$:

$$\begin{pmatrix} x_0^0 \mathrm{e}^0 + \cdots + x_{n-1}^0 \mathrm{e}^{n-1} \\ \vdots \\ x_0^{m-1} \mathrm{e}^0 + \cdots + x_{n-1}^{m-1} \mathrm{e}^{n-1} \end{pmatrix} = \begin{pmatrix} x_0^0 & \dots & x_{n-1}^0 \\ \vdots & \ddots & \vdots \\ x_0^{m-1} & \dots & x_{n-1}^{m-1} \end{pmatrix} \begin{pmatrix} \mathrm{e}^0 \\ \vdots \\ \mathrm{e}^{n-1} \end{pmatrix} = X\mathbf{e}.$$

The subgroup of F generated by the rows of $X\mathbf{e}$ can be denoted by $\langle X\mathbf{e} \rangle$. If P is an $m \times m$ matrix with integer entries, then

$$\langle PX\mathbf{e} \rangle \subseteq \langle X\mathbf{e} \rangle.$$

If also P is *invertible*—that is, $\det(P) = \pm 1$—then

$$\langle PX\mathbf{e} \rangle = \langle X\mathbf{e} \rangle.$$

We can therefore perform the following row-operations on X, without changing the group $\langle X\mathbf{e} \rangle$. We can

1. interchange two rows,
2. multiply a row by -1,
3. add an integer multiple of one row to another.

These operations allow us to perform Gaussian elimination. Adding rows of zeros as necessary, we may also assume that $m \geqslant n$. Then for some invertible integer matrix P, we have

$$PX = \left(\frac{T}{0}\right),$$

where T is an $n \times n$ upper-triangular matrix,

$$T = \begin{pmatrix} * & \cdots & * \\ & \ddots & \vdots \\ 0 & & * \end{pmatrix}.$$

By using also invertible *column*-operations, we can diagonalize T. That is, there are invertible integer matrices P and Q such that

$$PXQ = \begin{pmatrix} D \\ \hline 0 \end{pmatrix},$$

where

$$D = \begin{pmatrix} d_0 & & 0 \\ & \ddots & \\ 0 & & d_{n-1} \end{pmatrix}.$$

We now have

$$\langle X\mathbf{e} \rangle = \langle PXQQ^{-1}\mathbf{e} \rangle = \langle DQ^{-1}\mathbf{e} \rangle \cong \langle D\mathbf{e} \rangle.$$

Working further on D with invertible row- and column- operations, we may assume (3.5) holds, while $d_k = \cdots = d_{n-1} = 0$. Indeed, suppose $b, c \in \mathbb{Z}$ and $\gcd(b, c) = d$. By invertible operations, from

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

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 (3.5). □

Porism. *Every subgroup of a free abelian group on n generators is free abelian on n generators or fewer.*

We can show uniqueness of the numbers d_j by an alternative analysis.

Theorem 78 (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. Its surjectivity follows by counting. □

The Chinese Remainder Theorem will be generalized as Theorem 124. In the usual formulation of the theorem, every system

$$x \equiv a \pmod{m}, \qquad\qquad x \equiv b \pmod{n}$$

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 79. *For every finite abelian group, there are unique primes* p_0, …, p_{k-1}, *not necessarily distinct, along with unique positive integers* $m(0)$, …, $m(k-1)$, *such that*

$$G \cong \mathbb{Z}_{{p_0}^{m(0)}} \oplus \cdots \oplus \mathbb{Z}_{{p_{k-1}}^{m(k-1)}}.$$

Proof. To obtain the analysis, apply the Chinese Remainder Theorem to Theorem 77. 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. □

3.6. Semidirect products

An isomorphism from a structure to itself is an **automorphism.**

Theorem 80. *The automorphisms of a group G compose a subgroup of* $\mathrm{Sym}(G)$.

The subgroup in the theorem is denoted by

$$\mathrm{Aut}(G).$$

Theorem 81. *For every group G, there is a homomorphism*

$$g \mapsto (x \mapsto gxg^{-1})$$

from G to $\mathrm{Aut}(G)$.

An automorphism $x \mapsto gxg^{-1}$ as in the theorem is **conjugation** by g and is an **inner automorphism** of G. The kernel of the homomorphism in the theorem is the **center** of G, denoted by[3]

$$\mathrm{C}(G).$$

Then G is **centerless** if $\mathrm{C}(G)$ is trivial. Repeating the process of forming inner automorphisms, we obtain a chain

$$G \to \mathrm{Aut}(G) \to \mathrm{Aut}(\mathrm{Aut}(G)) \to \cdots ,$$

called the **automorphism tower** of G. The tower reaches a fixed point, perhaps after transfinitely many steps: Simon Thomas [18] shows this in case G is centerless; Joel Hamkins [3], in the general case.

Theorem 82. *For every group G, if $N \lhd G$, then there is a homomorphism*

$$g \mapsto (x \mapsto gxg^{-1})$$

from G to $\mathrm{Aut}(N)$*.*

In the theorem, let the homomorphism be $g \mapsto \sigma_g$. Suppose also $H < G$, and $N \cap H = \langle \mathrm{e} \rangle$. Then the conditions of Theorem 49 are met, and NH is an internal semidirect product. Equation (2.6), describing multiplication on NH, can be rewritten as

$$(mg)(nh) = (m \cdot \sigma_g(n))(gh).$$

Theorem 83. *Suppose N and H are groups, and $g \mapsto \sigma_g$ is a homomorphism from H to* $\mathrm{Aut}(N)$*. Then the set $N \times H$ becomes a group when multiplication is defined by*

$$(m, g)(n, h) = (m \cdot \sigma_g(n), gh).$$

Proof. To check that the multiplication is associative means checking that

$$\lambda_{(m,g)}\lambda_{(n,h)} = \lambda_{(m,g)(n,h)}.$$

[3] An alternative formulation of the center of a group is given and generalized in §4.3.

We can write $\lambda_{(m,g)}$ as $\lambda_m\sigma_g \times \lambda_g$. Then

$$\begin{aligned}\lambda_{(m,g)}\lambda_{(n,h)} = (\lambda_m\sigma_g \times \lambda_g)(\lambda_n\sigma_h \times \lambda_h) &= \lambda_m\sigma_g\lambda_n\sigma_h \times \lambda_g\lambda_h\\ &= \lambda_m\lambda_{\sigma_g(n)}\sigma_g\sigma_h \times \lambda_{gh}\\ &= \lambda_{m\cdot\sigma_g(n)}\sigma_{gh} \times \lambda_{gh}\\ &= \lambda_{(m\cdot\sigma_g(n),gh)}\\ &= \lambda_{(m,g)(n,h)}.\end{aligned}$$

Finally, (e,e) is an identity, and $(\sigma_{h^{-1}}(n^{-1}),h^{-1})$ is an inverse of (n,h). □

The group given by the theorem is the **semidirect product** of N and H with respect to σ; it can be denoted by

$$N \rtimes_\sigma H.$$

The bijection in Theorem 49 is an isomorphism from $N \rtimes_\sigma H$ to NH when σ is as in Theorem 82.

Theorem 84. *If p is prime, then ${\mathbb{Z}_p}^\times \cong \mathbb{Z}_{p-1}$.*

Proof. The group ${\mathbb{Z}_p}^\times$ has order $p-1$ and, by Theorem 77, is isomorphic to

$$\mathbb{Z}_{d_0} \oplus \mathbb{Z}_{d_{k-1}} \oplus \mathbb{Z}_m,$$

where $d_0 \mid \cdots \mid d_{k-1} \mid m$. Hence every element of ${\mathbb{Z}_p}^\times$ is a root of the polynomial x^m-1. But this polynomial can have at most m roots in $\mathbb{Z}_p$, since this is a *field.* Hence $p-1 \leqslant m$, so $m=p-1$, and $k=0$. □

Theorem 85. *The embedding $x \mapsto \lambda_x$ of a ring $(E,\cdot)$ in $(\mathrm{End}(E),\circ)$ restricts to an embedding of $(E,\cdot)^\times$ in $\mathrm{Aut}(E)$. In case E is $\mathbb{Z}_n$, each embedding is an isomorphism. In particular, if a is an element of ${\mathbb{Z}_n}^\times$ of order m, and $m \mid t$, then $\mathbb{Z}_t$ acts on $\mathbb{Z}_n$ by $(x,y) \mapsto a^xy$. Conversely, if some $\mathbb{Z}_t$ acts on $\mathbb{Z}_n$, then the action is so given for some such a.*

Theorem 86. *For every odd prime p, for every prime divisor q of $p-1$, there is a non-abelian semidirect product $\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q$, which is unique up to isomorphism.*

Proof. As ${\mathbb{Z}_p}^\times$ is cyclic, it has a unique subgroup G of order q. As q is prime, every nontrivial element of G is a generator. If $a \in G \smallsetminus \{1\}$, let σ be the homomorphism $x \mapsto (y \mapsto a^x y)$ from $\mathbb{Z}_q$ to $\operatorname{Aut}(\mathbb{Z}_p)$. Then we can form

$$\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q.$$

If $\mathbb{Z}_p \rtimes_\tau \mathbb{Z}_q$ is some other non-abelian semidirect product, then τ_1 is $x \mapsto b \cdot x$ for some b in $G \smallsetminus \{1\}$. But then $b^n = a$ for some n, so there is an isomorphism from $\mathbb{Z}_p \rtimes_\sigma \mathbb{Z}_q$ to $\mathbb{Z}_p \rtimes_\tau \mathbb{Z}_q$ that takes (x, y) to (x, ny). □

Because of its uniqueness, we may refer to the semidirect product of the theorem as

$$\mathbb{Z}_p \rtimes \mathbb{Z}_q.$$

In case $q = 2$, this group is D_p. The next section develops the tools used in §4.2 to show that there is no other way to obtain a group of order pq for distinct primes p and q.

4. Finite groups

4.1. Actions of groups

Theorem 87. *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 $\mathrm{Sym}(A)$, and*
2. *functions $(g, a) \mapsto ga$ from $G \times A$ into A such that*

$$\mathrm{e}\, a = a, \tag{4.1}$$

$$(gh)a = g(ha). \tag{4.2}$$

for all h and h in G and a in A.

Proof. If $g \mapsto (a \mapsto ga)$ maps G homomorphically into $\mathrm{Sym}(A)$, then (4.1) and (4.2) 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 $\mathrm{Sym}(A)$, homomorphically by (4.2). □

Either of two functions that correspond as in the theorem is a **(left) action** of G on A. Examples include the following.

1. A symmetry group of a set acts on the set in the obvious way, by

$$(\sigma, x) \mapsto \sigma(x).$$

2. An arbitrary group G acts on itself by left multiplication:

$$(g, x) \mapsto \lambda_g(x).$$

3. If $H < G$, then G acts on the set G/H by
$$(g, xH) \mapsto gxH.$$

4. Finally, G acts on itself by conjugation:
$$(g, x) \mapsto x \mapsto gxg^{-1}.$$

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
$$A_0.$$
See Appendix B for an alternative development of these notions.

Theorem 88. *Let G act on A by $(g, x) \mapsto gx$.*

1. *The orbits partition A;*
2. *$G_a < G$;*
3. *$[G : G_a] = |Ga|$;*

Proof. For (3), 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 $gG_a \mapsto ga$. □

Corollary. *If there are only finitely many orbits in A under G, then*
$$|A| = |A_0| + \sum_{a \in X} [G : G_a] \tag{4.3}$$
for some set X of elements of A whose orbits are nontrivial.

Equation (4.3) is the **class equation.** For example, suppose G acts on itself by conjugation, and $g \in G$. Then Gg is the **conjugacy class** of g, while G_g is the **centralizer** of g, denoted by[1]

$$\mathrm{C}_G(g). \tag{4.4}$$

Finally, G_0 is the **center** of G, denoted by

$$\mathrm{C}(G).$$

The class equation for the present case can now be written as

$$|G| = |\mathrm{C}(G)| + \sum_{a \in X} [G : \mathrm{C}_G(a)].$$

A **finite p-group** is a finite group whose order is a power of p.

Theorem 89. *If A is acted on by a 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. □

A first application of this theorem is

Theorem 90 (Cauchy)**.** *If p divides $|G|$, then $|g| = p$ for some g in G.*

Proof (J. H. McKay [13]). Suppose p divides $|G|$. We seek a nontrivial solution in G of the equation

$$x^p = \mathrm{e}\,.$$

Let A be the set

$$\{\boldsymbol{x} \in G^p : x_0 \cdots x_{p-1} = \mathrm{e}\};$$

so we seek g in G such that $(g, \dots, g) \in A$ and $g \neq \mathrm{e}$. If $(g_0, \dots, g_{p-1}) \in A$ and $k < p$, then

$$(g_0 \cdots g_{k-1})(g_k \cdots g_{p-1}) = \mathrm{e}, \qquad (g_k \cdots g_{p-1})(g_0 \cdots g_{k-1}) = \mathrm{e},$$

[1] More generally, if $H < G$, then $\mathrm{C}_H(g) = \{h \in H : hgh^{-1} = g\}$.

and therefore

$$(g_k, \dots, g_{p-1}, g_0, \dots, g_{k-1}) \in A.$$

Thus $\mathbb{Z}_p$ acts on A by

$$(k, (g_0, \dots, g_{p-1}) \mapsto (g_k, \dots, g_{p-1}, g_0, \dots, g_{k-1}).$$

With respect to this action,

$$A_0 = \{(g, \dots, g) \colon g^p = \mathrm{e}\};$$

also $\mathbb{Z}_p$ is a finite p-group, Now, the map

$$(g_1, \dots, g_{p-1}) \longmapsto \bigl((g_1 \cdots g_{p-1})^{-1}, g_1, \dots, g_{p-1}\bigr)$$

is a bijection from G^{p-1} onto A, so $|A|$ is a multiple of p; hence $|A_0|$ is a multiple of p, by Theorem 89. Since A_0 contains $(\mathrm{e}, \dots, \mathrm{e})$, it contains some $(g, \dots, g)$, where $|g| = p$. □

Corollary. *A finite group is a p-group if and only if the order of every element is a power of p.*

Proof. If ℓ is a prime dividing $|g|$, then ℓ divides $|G|$. Conversely, if ℓ divides $|G|$, then G has an element of order ℓ. □

Hence an arbitrary group is a ***p*-group** if the order of its every element is a power of p.

Theorem 91. *Every nontrivial p-group has nontrivial center.*

Proof. By Theorem 89,

$$|G| \equiv |\mathrm{C}(G)| \pmod{p},$$

so p divides $|\mathrm{C}(G)|$. Since $\mathrm{C}(G)$ contains at least one element, it contains at least p of them. □

Theorem 92. *All groups of order p^2 are abelian.*

Proof. Let G have order p^2. Then either $\mathrm{C}(G)$ is all of G, or else $|\mathrm{C}(G)| = p$, by the previous theorem. In any case, there is a in G such that

$$G = \langle\{a\} \cup \mathrm{C}(G)\rangle.$$

But elements of $\mathrm{C}(G)$ commute with all elements of G; and powers of a commute with each other (and with elements of $\mathrm{C}(G)$); hence G is abelian. □

Supposing G is an arbitrary group and $H < G$, let A be the set

$$\{gHg^{-1} \colon g \in G\}$$

of conjugates of H. Then G acts on A by conjugation,

$$(g, K) \mapsto gKg^{-1}.$$

The stabilizer of H under this action is the **normalizer** of H in G, denoted by[2]

$$\mathrm{N}_G(H).$$

If $H < K < G$, then

$$H \lhd K \iff K < \mathrm{N}_G(H).$$

Theorem 93. *Suppose G is a group with subgroups H and K. Under the action of H on G/K by left multiplication,*

$$gK \in (G/K)_0 \iff H < gKg^{-1}.$$

In case $H = K$, a finite group,

$$(G/H)_0 = \mathrm{N}_G(H)/H.$$

Proof. We compute:

$$\begin{aligned}
gK \in (G/K)_0 &\iff hgK = gK && \text{for all } h \text{ in } H\\
&\iff g^{-1}hgK = K && \text{for all } h \text{ in } H\\
&\iff g^{-1}hg \in K && \text{for all } h \text{ in } H\\
&\iff h \in gKg^{-1} && \text{for all } h \text{ in } H\\
&\iff H < gKg^{-1}.
\end{aligned}$$

[2] More generally, if also $K < G$, then $\mathrm{N}_K(H) = \{k \in K \colon kHk^{-1} = H\}$.

If H is finite, then

$$H < gHg^{-1} \iff H = gHg^{-1} \iff g \in \mathrm{N}_G(H). \qquad \square$$

A p-**subgroup** of a group is a subgroup that is a p-group.

Lemma 17. *If H is a p-subgroup of G, then*

$$[G : H] \equiv [\mathrm{N}_G(H) : H] \pmod{p}.$$

Proof. Theorems 93 and 89. □

Lemma 18. *If H is a p-subgroup of G, and p divides $[G : H]$, then H is a normal subgroup of some p-subgroup K of G such that $[K : H] = p$.*

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 90, this group has an element gH of order p. So $\langle \{g\} \cup H\rangle$ is the desired K. □

A **Sylow** p-**subgroup** is a maximal p-subgroup. The following is a partial converse to Lagrange's Theorem (Theorem 43).

Theorem 94 (Sylow I)**.** *For every finite group of order $p^n m$, where $p \nmid m$, there is a chain*

$$H_1 < H_2 < \cdots < H_n$$

of subgroups, where $|H_1| = p$ and in each case $H_i \lhd H_{i+1}$ and $[H_{i+1} : H_i] = p$. Every p-subgroup of such a group appears on such a chain. In particular, every p-subgroup is included in a Sylow subgroup, whose index is indivisible by p.

Proof. Cauchy's Theorem (Theorem 90) and repeated application of the last lemma. □

Corollary. *The conjugate of a Sylow p-subgroup is a Sylow p-subgroup. A* unique *Sylow p-subgroup is normal.*

A converse to the corollary is the following.

Theorem 95 (Sylow II). *All Sylow p-subgroups are conjugate.*

Proof. Say H and P are p-subgroups of G, where P is maximal. Then H acts on the set G/P by left multiplication. By Theorem 89, since $[G : P]$ is not a multiple of p, the set $(G/P)_0$ has an element aH. By Theorem 93, $H < aPa^{-1}$. If H is also Sylow, then $H = aPa^{-1}$ by Theorem 94. □

Theorem 96 (Sylow III). *The number of Sylow p-subgroups of a finite group is congruent to* 1 modulo p *and divides the order of the group.*

Proof. Let A be the set of Sylow p-subgroups of a finite group G. Then G acts on A by conjugation. Let $H \in A$. By Theorem 95, the orbit of H is precisely A. The stabilizer of H is $\mathrm{N}_G(H)$. Then by Theorem 88 (3),

$$[G : \mathrm{N}_G(H)] = |A|,$$

so $|A|$ divides $|G|$.

Now consider H as acting on A by conjugation. Then the following are equivalent:

1. $P \in A_0$,
2. $H < \mathrm{N}_G(P)$,
3. H is a Sylow subgroup of $\mathrm{N}_G(P)$,
4. $H = P$,

since $P \lhd \mathrm{N}_G(P)$, so P is the unique Sylow p-subgroup of $\mathrm{N}_G(P)$. Therefore $A_0 = \{H\}$, so by Theorem 89

$$|A| \equiv |A_0| \equiv 1 \pmod{p}.$$ □

4.2. Classification of small groups

We can now complete the work, begun in §3.6, of classifying the groups of order pq for primes p and q.

Lemma 19. *Suppose p and q are distinct primes such that*

$$q \not\equiv 1 \pmod{p}$$

and $|G| = pq$. Then G has a unique Sylow p-subgroup, which is therefore normal.

Proof. Let A be the set of Sylow p-subgroups of G. Then $|A| \equiv 1 \pmod{p}$ by Theorem 96, so $|A|$ is not q or pq; but $|A|$ divides pq; so $|A| = 1$. □

Theorem 97. *Suppose p and q are primes, where $p < q$, so that $p \not\equiv 1 \pmod{q}$, and G is a group of order pq.*

1. *If $q \not\equiv 1 \pmod{p}$, then G is cyclic.*
2. *If $q \equiv 1 \pmod{p}$, then either G is cyclic group, or else G is the unique non-abelian semidirect product $\mathbb{Z}_p \rtimes \mathbb{Z}_q$.*

In particular, every non-abelian group of order $2q$ is isomorphic to D_q.

Proof. By the lemma, G has a normal subgroup N of order q, and N is cyclic by a corollary to Lagrange's Theorem (Theorem 43). By the first Sylow Theorem (Theorem 94), G has a Sylow p-subgroup H, which has order p and is therefore cyclic. Then $N \cap H = \langle \mathrm{e} \rangle$, so $G = NH$ by Theorem 49 and counting.

1. If $q \not\equiv 1 \pmod{p}$, then $H \lhd G$ by the lemma, so $G = N \times H$ by Theorem 67. The product is cyclic by the Chinese Remainder Theorem (Theorem 78).

2. If $q \not\equiv 1 \pmod{p}$, then G might still be $N \times H$; otherwise, G is isomorphic to $\mathbb{Z}_p \rtimes \mathbb{Z}_q$ by Theorem 86. □

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 98. *Every group of order 8 is isomorphic to one of*

$$\mathbb{Z}_8, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_4, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2, \qquad D_4, \qquad \mathrm{Q}_8.$$

Proof. Say $|G| = 8$. If G is abelian, then its possibilities are given by Theorem 77. Suppose G is not abelian. Then G has an element a of order greater than 2 by [8, Exercise I.1.13, p. 30], and so $|a| = 4$ (since $G \ncong \mathbb{Z}_8$). Then $\langle a \rangle \lhd G$ by [8, Exercise I.5.1, p. 45]. Let $b \in G \setminus \langle a \rangle$. 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 $G \cong D_4$. In the latter case, $G \cong \mathrm{Q}_8$. □

Theorem 99. *Every group of order* 12 *is isomorphic to one of*

$$\mathbb{Z}_{12}, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_6, \qquad \mathrm{Alt}(4), \qquad D_6, \qquad \langle a, b \mid a^6, a^3b^2, bab^{-1}a \rangle.$$

Proof. Suppose $|G| = 12$, but G is not abelian. A Sylow 3-subgroup of G has order 3, so it is $\langle a \rangle$ for some a. Then G acts on $G/\langle a \rangle$ by left multiplication, and $[G : \langle a \rangle] = 4$, so there is a homomorphism from G to $\mathrm{Sym}(4)$. If this is an embedding, then $G \cong \mathrm{Alt}(4)$. Assume is is not an embedding. Then the kernel must be $\langle a \rangle$, so $\langle a \rangle \lhd G$.

Let H be a Sylow 2-subgroup of G. Then H is isomorphic to $\mathbb{Z}_4$ or $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. In any case, H has two elements b and c such that none of b, c, or bc is e. Since G is not $\langle a \rangle \times H$, we may assume

$$bab^{-1} = a^2.$$

If also $cac^{-1} = a^2$, then $bcac^{-1}b^{-1} = a$. Thus H has an element that commutes with a. Hence G has a subgroup K isomorphic to $\mathbb{Z}_6$. If $G \setminus K$ has an element of order 2, then $G \cong D_6$; otherwise, G is the last possibility above. □

4.3. Nilpotent groups

For a group, what is the next best thing to being abelian? A group G is abelian if and only if $\mathrm{C}(G) = G$. (See §3.6.) 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

$$\mathrm{C}(G) = \{g \in G \colon \forall x\ [g, x] = \mathrm{e}\}.$$

We now generalize this by defining

$$\mathrm{C}_0(G) = \langle \mathrm{e} \rangle,$$
$$\mathrm{C}_{n+1}(G) = \{g \in G \colon \forall x\ [g, x] \in \mathrm{C}_n(G)\}.$$

Then $\mathrm{C}(G) = \mathrm{C}_1(G)$.

Theorem 100. *Let G be a group.*

1. $\mathrm{C}_n(G) \lhd G$.
2. $\mathrm{C}_n(G) < \mathrm{C}_{n+1}(G)$.
3. $\mathrm{C}_{n+1}(G)/\,\mathrm{C}_n(G) = \mathrm{C}(G/\,\mathrm{C}_n(G))$.

Proof. We use induction to prove 1, and incidentally 2 and 3. Trivially, $\mathrm{C}_0(G) \lhd G$. Suppose $\mathrm{C}_k(G) \lhd G$. Then the following are equivalent:

$$g \in \mathrm{C}_{k+1}(G);$$
$$\forall x\ [g, x] \in \mathrm{C}_k(G);$$
$$\forall x\ gxg^{-1}x^{-1} \in \mathrm{C}_k(G);$$
$$\forall x\ \ \mathrm{C}_k(G)gx = \mathrm{C}_k(G)xg;$$
$$\mathrm{C}_k(G)g \in \mathrm{C}(G/\,\mathrm{C}_k(G)).$$

Thus $\mathrm{C}_k(G) < \mathrm{C}_{k+1}(G)$, and $\mathrm{C}_{k+1}(G)/\,\mathrm{C}_k(G) = \mathrm{C}(G/\,\mathrm{C}_k(G))$; in particular,

$$\mathrm{C}_{k+1}(G)/\,\mathrm{C}_k(G) \lhd G/\,\mathrm{C}_k(G),$$

so $\mathrm{C}_{k+1}(G) \lhd G$. □

The **ascending central series** of G is the sequence $(\mathrm{C}_n(G) \colon n \in \omega)$, usually written out as

$$\langle \mathrm{e} \rangle \lhd \mathrm{C}(G) \lhd \mathrm{C}_2(G) \lhd \mathrm{C}_3(G) \lhd \cdots .$$

A group is called **nilpotent** if the terms in the sequence are eventually the group itself, that is, for some n in ω,

$$\mathrm{C}_n(G) = G.$$

So an abelian group is nilpotent, since its center is itself.

Suppose G is nilpotent, and in particular $\mathrm{C}_n(G) = G$. For some g in G, and let f be the operation $x \mapsto [g, x]$ on G. Writing f^0 for id_G and f^{n+1} for $f \circ f^n$, we have

$$f^0(x) \in G, \quad f(x) \in \mathrm{C}_{n-1}(G), \quad f^2(x) \in \mathrm{C}_{n-2}(G), \quad \dots, \quad f^n(x) = \mathrm{e}\,.$$

Thus f is "nilpotent" in the monoid of operations on G. However, this should not be taken as a sufficient condition for G to be nilpotent.

Examples of nilpotent groups are given by:

Theorem 101. *Finite p-groups are nilpotent.*

Proof. Suppose G is a p-group. If H is a proper normal subgroup of G, then G/H is a nontrivial p-group, so by Theorem 91 it has a nontrivial center. By Theorem 100 the ascending central series of G is strictly increasing, until it reaches G itself. □

The converse fails, because of:

Theorem 102. *A finite direct product of nilpotent groups is nilpotent.*

Proof. Use that

$$\mathrm{C}(G \times H) = \mathrm{C}(G) \times \mathrm{C}(H).$$

If $\mathrm{C}_n(G) = G$ and $\mathrm{C}_m(H) = H$, then $\mathrm{C}_{\max\{n,m\}}(G \times H) = G \times H$. □

We now proceed to the converse of this theorem.

Lemma 20. *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 21. *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 103. *A finite nilpotent group is the direct product of its Sylow subgroups.*

Proof. Suppose G is a finite nilpotent group. We shall show that every Sylow subgroup of G is a normal subgroup. By Theorem 67, the first and second Sylow Theorems (Theorems 94 and 95), and counting, G will be the direct product of its Sylow subgroups.

Suppose then P is a Sylow p-subgroup of G. We shall show that $P \lhd G$. To do this, it is enough to show $\mathrm{N}_G(P) = G$. To do *this,* by the last lemma, it is enough to show $\mathrm{N}_G(\mathrm{N}_G(P)) < \mathrm{N}_G(P)$. To do *this,* note that, as $P \lhd \mathrm{N}_G(P)$, so P is the unique Sylow p-subgroup of $\mathrm{N}_G(P)$. Hence, in particular, for any x in G, if $xPx^{-1} < \mathrm{N}_G(P)$, then $xPx^{-1} = P$, so $x \in \mathrm{N}_G(P)$. But every x in $\mathrm{N}_G(\mathrm{N}_G(P))$ satisfies the hypothesis. □

4.4. Soluble groups

The **commutator subgroup** of a group G is the subgroup

$$\langle [x,y] \colon (x,y) \in G^2 \rangle,$$

which is denoted by

$$G'.$$

Theorem 104. *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

$$f([x,y]) = f(xyx^{-1}y^{-1}) = f(x)f(y)f(x)^{-1}f(y)^{-1} = [f(x),f(y)]. \quad (4.5)$$

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

$$\begin{aligned} G^{(0)} &= G, \\ G^{(n+1)} &= (G^{(n)})'. \end{aligned}$$

We have a descending sequence

$$G \rhd G' \rhd G^{(2)} \rhd \cdots$$

The group G is called **soluble** if this sequence reaches $\langle \mathrm{e} \rangle$ (after finitely many steps).

For examples, let K be a field. 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

$$\begin{pmatrix}1 & a_1 & & *\\ & 1 & \ddots & \\ & & \ddots & a_{n-1}\\ 0 & & & 1\end{pmatrix}\begin{pmatrix}1 & b_1 & & *\\ & 1 & \ddots & \\ & & \ddots & b_{n-1}\\ 0 & & & 1\end{pmatrix}$$
$$=\begin{pmatrix}1 & a_1+b_1 & & *\\ & 1 & \ddots & \\ & & \ddots & a_{n-1}+b_{n-1}\\ 0 & & & 1\end{pmatrix},$$

so the elements of G'' take the form of

$$\begin{pmatrix}1 & 0 & & *\\ & 1 & \ddots & \\ & & \ddots & 0\\ 0 & & & 1\end{pmatrix}.$$

Proceeding, we find $G^{(n+1)} = \langle \mathrm{e} \rangle$.

Theorem 105. *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 104 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 ω. Working left to right, we can build up the following commutative diagram, where arrows are inclusions:

$$\begin{tikzcd}
G \arrow[d] & G' \arrow[l] \arrow[d] & G^{(2)} \arrow[l] \arrow[d] & G^{(3)} \arrow[l] \arrow[d] & G^{(n)} \arrow[l, dotted] \arrow[d]\\
G \arrow[d] & \mathrm{C}_n(G)' \arrow[l] \arrow[d] & \mathrm{C}_{n-1}(G)' \arrow[l] \arrow[d] & \mathrm{C}_{n-2}(G)' \arrow[l] \arrow[d] & \mathrm{C}(G)' \arrow[l, dotted] \arrow[d]\\
\mathrm{C}_n(G) & \mathrm{C}_{n-1}(G) \arrow[l] & \mathrm{C}_{n-2}(G) \arrow[l] & \mathrm{C}_{n-3}(G) \arrow[l] & \langle \mathrm{e} \rangle \arrow[l, dotted]
\end{tikzcd}$$

That is, we know $G^{(0)} < \mathrm{C}_n(G)$; and if $G^{(k)} < \mathrm{C}_{n-k}(G)$ for some k in n, then

$$G^{(k+1)} = (G^{(k)})' < \mathrm{C}_{n-k}(G)' < \mathrm{C}_{n-(k+1)}(G).$$

By induction then, $G^{(n)} < \mathrm{C}_0(G) = \langle \mathrm{e} \rangle$, so $G^{(n)} = \langle \mathrm{e} \rangle$. □

Theorem 106. *Solubility is preserved in subgroups and quotients. If $N \lhd G$, and N and G/N are soluble, then G is soluble.*

Proof. Suppose $f\colon G \to H$. By (4.5), we have $f(G^{(n)}) < H^{(n)}$, with equality is f is surjective. The case where f is an inclusion of G in H shows that subgroups of soluble groups are soluble. The case where f is a quotient map shows that quotients of soluble groups are soluble.

Finally, if $N \lhd G$, then $(G/N)' = G'N/N$. Suppose $(G/N)^{(n)} = \langle \mathrm{e} \rangle$, and $N^{(m)} = \langle \mathrm{e} \rangle$. Then $G^{(n)} < N$ and so $G^{(n+m)} = \langle \mathrm{e} \rangle$. □

Theorem 107. *Groups with non-abelian simple subgroups are not soluble. In particular,* $\mathrm{Sym}(5)$ *is not soluble if $n \geqslant 5$.*

Proof. Suppose H is simple. Since $H' \lhd H$, we have either $H' = \langle \mathrm{e} \rangle$ or $H' = H$. In the former case, H is abelian; in the latter, H is insoluble. □

The last theorem suggests the origin of the notion of solubility of groups: the general 5th-degree polynomial equation

$$a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + x^5 = 0$$

is "insoluble by radicals" precisely because $\mathrm{Sym}(5)$ is an insoluble group.

4.5. Normal series

A **normal series** for a group G is a sequence $(G_n : n \in \omega)$ of subgroups, where $G_{n+1} \lhd G_n$ in each case; the situation can be depicted by

$$G = G_0 \rhd G_1 \rhd G_2 \rhd \cdots$$

(If one wants to distinguish, one may call this a **subnormal series**, normal if each G_i is normal in G.) The **factors** of the normal series are the quotients G_i/G_{i+1}. If $G_n = \langle \mathrm{e} \rangle$ for some n, then the series is called

1. a **composition series,** if the factors are simple;
2. a **soluble series,** if the factors are abelian.

For example, if G is nilpotent, then the series

$$\langle \mathrm{e} \rangle \lhd \mathrm{C}(G) \lhd \mathrm{C}_2(G) \lhd \cdots \lhd G$$

is a soluble series.

Theorem 108. *A group is soluble if and only if it has a soluble series.*

Proof. If the series

$$G \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_n = \langle \mathrm{e} \rangle$$

is soluble, then, by Theorem 104, we have

$$G' < G_1, \quad G'' < {G_1}' < G_2, \quad G''' < {G_1}'' < G_2' < G_3, \quad G^{(n)} = \langle \mathrm{e} \rangle,$$

so G is soluble. Conversely, if G is soluble, then the series

$$G \rhd G' \rhd G^{(2)} \rhd \cdots \rhd \langle \mathrm{e} \rangle$$

is a soluble series. □

So not every group has a soluble series. However:

Theorem 109. *Every finite group has a composition series.*

Proof. A finite group G has a maximal proper normal subgroup N. Then G/N is simple. Indeed, every normal subgroup of G/N is H/N for some normal subgroup H of G such that $N < H$, and therefore H is either N or G.

So we can form $G = G_0 \rhd G_1 \rhd \cdots$, where each G_{n+1} is a maximal proper normal subgroup of G_n. The factors are simple, and, since G is finite, the series must terminate. □

If, from a normal series, another can be got by deleting some terms, then the former is a **refinement** of the latter. As a normal series, a composition series is maximal in that it has no nontrivial refinement, that is, no refinement without trivial factors.

A soluble series for a finite group has a refinement in which the nontrivial factors are cyclic of prime order.

Any normal series is **equivalent** to the series that results when all repeated terms are deleted (so that all trivial factors are removed). Then two normal series

$$G_i(0) \rhd G_i(1) \rhd G_i(2) \rhd \cdots \rhd G_i(n)$$

(where $i < 2$) with no trivial factors are **equivalent** if there is σ in $\operatorname{Sym}(n)$ such that

$$G_0(i)/G_0(i+1) \cong G_1(\sigma(i))/G_1(\sigma(i+1))$$

for each i in n. We now aim to prove Theorem 111 below.

Lemma 22 (Zassenhaus or Butterfly). *Suppose $N_i \lhd H_i < G$ for each i in 2. Let $H = H_0 \cap H_1$. Then:*

1. *$N_i(H_i \cap N_{1-i}) \lhd N_iH$ for each i;*
2. *the two groups $N_iH/N_i(H_i \cap N_{1-i})$ are isomorphic.*

Proof. We have $H_i \cap N_{1-i} \lhd H$. Let

$$K = (H_0 \cap N_1)(H_1 \cap N_0);$$

then $K \lhd H$. The groups we have to work with form the commutative diagram below, arrows being inclusions.

$$\begin{tikzcd}
 & H_0 & & H_1 & \\
 & N_0H \arrow[u] & & N_1H \arrow[u] & \\
 & & H \arrow[ul] \arrow[ur] & & \\
 & N_0(H_0 \cap N_1) \arrow[uu] & & N_1(H_1 \cap N_0) \arrow[uu] & \\
N_0 \arrow[ur] & & K \arrow[uu] \arrow[ul] \arrow[ur] & & N_1 \arrow[ul] \\
 & H_1 \cap N_0 \arrow[ul] \arrow[ur] & & H_0 \cap N_1 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

We exhibit an epimorphism from N_iH onto H/K whose kernel is $N_i(H_i \cap N_{1-i})$. Now, if $n, n' \in N_i$ and $h, h' \in H$ and $nh' = n'h$, then

$$h'h^{-1} = n^{-1}n' \in N_i \cap H < K,$$

so that $Kh = Kh'$. Hence there is a well-defined homomorphism f from N_iH into H/K such that, if $n \in N_i$ and $h \in H$, then

$$f(nh) = Kh.$$

That f is surjective is clear. Moreover, the following are equivalent conditions on such n and h:

1. $nh \in \ker(f)$;
2. $h \in K$;
3. $h = n_0n_1 = n_1n_0$ for some n_i in $H_{1-i} \cap N_i$.

Also, (3) implies that $nh = nn_in_{1-i}$, which is in $N_i(H_i \cap N_{1-i})$; thus

4. $nh \in N_i(H_i \cap N_{1-i})$.

Conversely, suppose (4) holds. Then also $h = n^{-1}nh$, which is also in $N_i(H_i \cap N_{1-i})$, so $h = n'h'$ for some n' in N_i and h' in $N_{1-i} \cap H_i$. Then $n' = h(h')$, which is in $\in H_{1-i}$, so $n' \in N_i \cap H_{1-i}$, and therefore $h \in K$. □

Theorem 110 (Schreier). *Any two normal series have equivalent refinements.*

Proof. Suppose that

$$G = G_i(0) \rhd G_i(1) \rhd \cdots \rhd G_i(n_i) = \langle \mathrm{e} \rangle,$$

where $i < 2$, are normal series for G. In particular,

$$G_i(j+1) \lhd G_i(j) < G.$$

Define

$$G_i(j,k) = G_i(j+1)(G_i(j) \cap G_{1-i}(k)),$$

where $(j,k) \in n_i \times n_{1-i}$. Then

$$\begin{aligned} G_i(j) = G_i(j,0) \rhd G_i(j,1) \rhd \cdots \rhd G_i(j,n_{1-i}-1) \\ \rhd G_i(j,n_{1-i}) = G_i(j+1), \end{aligned}$$

giving us normal series that are refinements of the original ones; but also

$$G_0(j,k)/G_0(j,k+1) \cong G_1(k,j)/G_1(k,j+1)$$

by the Butterfly Lemma. □

Theorem 111 (Jordan–Hölder). *Any two composition series of a group are equivalent.*

Combining this with Theorem 109, we have that every finite group has a uniquely determined set of simple “factors”. Hence the interest in the classification of the finite simple groups.

Part II.

Rings

5. Rings in the most general sense

5.1. Not-necessarily-associative rings

Rings were introduced in §1.6. A more general definition is possible. If E is an abelian group (written additively), then a **multiplication** on E is a binary operation that distributes in both senses over addition. In the most general sense then, a **ring** is an abelian group with a multiplication. The ring is **associative** if the multiplication is associative.

Associative rings are not the only rings of interest. For example, the associative ring $\mathbb{H}$ defined in §2.2 has the automorphism $z+w\mathrm{j} \mapsto \bar{z}-w\mathrm{j}$; then the same constuction 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. Also, if $(E,\cdot)$ is a ring, then there is another multiplication on E, namely b or $(x,y) \mapsto [x,y]$, where

$$[x,y] = x \cdot y - y \cdot x;$$

this multiplication makes E into a **Lie ring,** namely a ring that respects the identity

$$[x,x] = 0$$

along with the **Jacobi identity,**

$$[[x,y],z] = [x,[y,z]] - [y,[x,z]].$$

For example, from the associative ring $(\mathrm{End}(E),\circ)$, we obtain the Lie ring $(\mathrm{End}(E),\mathrm{b})$. Then $\mathrm{End}(E)$ has a subgroup $\mathrm{Der}(E,\cdot)$, which is closed under b, but not generally under $\circ$. Specifically, $\mathrm{Der}(E,\cdot)$ consists of the **derivations** of $(E,\cdot)$, which are the endomorphism D of E respecting the **Leibniz rule,**

$$D(x \cdot y) = Dx \cdot y + x \cdot Dy.$$

In particular, ‘taking the derivative’ on the field of meromorphic functions on $\mathbb{C}$ is a derivation. Derivations will be used in §6.7.

Theorem 112. *Every ring respects 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,

$$\begin{aligned} 0\cdot x = 0 &= x\cdot 0, \qquad (5.1)\\ (-x)\cdot y = -(x\cdot y) &= x\cdot(-y). \end{aligned}$$

A ring is **unital** if it has a multiplicative identity, generally denoted by 1. The result of Theorem 24 can be strengthened when the scope of the theorem is restricted to abelian groups:

Theorem 113. *Let E be an abelian group. Then $n \mapsto (x \mapsto nx)$ is a homomorphism of unital rings from $(\mathbb{Z}, \cdot, 1)$ to $(\mathrm{End}(E), \circ, \mathrm{id}_E)$.*

In a word, we can say that, as a unital ring, $\mathbb{Z}$ **acts** on the endomorphism group of every abelian group. Compare the notion of action defined in §4.1. In the notation of Theorem 113,

$$\begin{aligned} 0x &= 0, \qquad (5.2)\\ 1x &= x,\\ (-1)x &= -x; \qquad (5.3) \end{aligned}$$

here (5.2) is (1.3) written additively; combining it with (5.1), we have

$$0\cdot x = 0x,$$

where the zeros come from the ring and from $\mathbb{Z}$ respectively. More generally, we have

Theorem 114. *For every integer n, every ring respects the identity*

$$(nx)\cdot y = n(x\cdot y) = x\cdot ny.$$

Proof. Induction and (5.3). □

5.2. Associative, not-necessarily-unital rings

Henceforth the word *ring* means associative ring. By Theorem 25, a unital ring also acts on the endomorphism group of the underlying abelian group. We have in particular

$$1 \cdot x = 1x.$$

Again a ring is **commutative** if the multiplication is commutative. As examples of commutative rings with identity, we have $\mathbb{Z}$ and $\mathbb{Z}_n$ (by 29); and if R is a commutative ring with identity, then $\mathrm{M}_n(R)$ is a ring with identity, by Theorem 27. The continuous functions on $\mathbb{R}$ with compact support compose a ring with respect to the operations induced from $\mathbb{R}$: this ring has no identity.

The **characteristic** of a ring $(E, \cdot)$ is the non-negative integer n such that $\mathbb{Z}_n$ is the kernel of the homomorphism $n \mapsto (y \mapsto ny)$ from $\mathbb{Z}$ to $\mathrm{End}(E)$. This kernel is the kernel of $n \mapsto n1$, if $(E, \cdot)$ has an identity. For example, If $0 \leqslant n$, then $\mathbb{Z}_n$ has characteristic n.

Theorem 115. *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. □

5.3. Unital associative rings

Henceforth in the word ring *means ring with identity,* as it did in §1.6. We know from Theorem 26 that a ring R has a group of units, $R^{\times}$. The example in §3.1 shows that some ring elements can have right inverses without being units. However, if a has both a left and a right inverse, then they are the same, since if $ab = 1 = ca$, then

$$c = c1 = c(ab) = (ca)b = 1b = b.$$

A **zero-divisor** of R is a element b distinct from 0 such that the equations $bx = 0$ and $yb = 0$ 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 $1 \neq 0$. So fields are integral domains. But $\mathbb{Z}$ is an integral domain that is not a field. If p is prime, then $\mathbb{Z}_p$ is a field, denoted by $\mathbb{F}_p$.

An arbitrary 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 a 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; we can give it a multiplication as follows. We write an element $(r_g : g \in G)$ of the direct sum as

$$\sum_{g \in G} r_g g;$$

this is a **formal finite R-linear combination** of the elements of G. Then multiplication is defined as one expects: if $r, s \in R$ and $g, h \in G$, then

$$(rg)(sh) = (rs)(gh),$$

and 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

$$\sum_{k=0}^{n} a_k x^k,$$

where $n \in \omega$ and $a_k \in R$. We could use a second variable, getting for example $R[X, Y]$. Usually R here is commutative and is in particular a field.

5.4. Ideals

If A is a sub-ring of R, then we can form the abelian group R/A. We could try to define a multiplication on this by

$$(x + A)(y + A) = xy + A.$$

However, if $x - x' \in A$, and $y - y' \in A$, we need not have $xy - x'y' \in A$.

A **left ideal** of R is a sub-ring I such that

$$RI \subseteq I,$$

that is, $rx \in I$ whenever $r \in R$ and $x \in I$. Likewise, **right** and **two-sided** ideal. For example, the set of matrices

$$\begin{bmatrix} * & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ * & 0 & \dots & 0 \end{bmatrix}$$

is a left ideal of $\mathrm{M}_n(R)$, but not a right ideal unless $n = 1$. Also, Rx is a left ideal of R, while RxR is a two-sided ideal.

Theorem 116. *If I is a two-sided ideal of R, then R/I is a well-defined ring. The kernel of a ring-homomorphism is a two-sided ideal.*

Suppose $(A_i : i \in I)$ 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 §3.1. If in particular $I = n$, 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 117. *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 118. *If A and B are left ideals of a ring, then so is $A \cap B$, and $AB \subseteq A \cap B$.*

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 $A^2 = 4\mathbb{Z}$.

Theorem 119. *If $f \colon R \to S$, a homomorphism of rings, 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.

6. Commutative rings

6.1. Commutative rings

Henceforth, let all rings be commutative, so all ideals are two-sided. 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 $A = \{a\}$, then (A) is denoted by

$$(a)$$

or Ra and is called a **principal ideal.** A **principal ideal domain** or PID is an integral domain whose every ideal is principal. For example, $\mathbb{Z}$ is a PID by Theorem 36. But in the polynomial ring $\mathbb{R}[X, Y]$, the ideal (X, Y) is not principal.

An ideal is proper if and only if it does not contain a unit. A *proper* ideal P is **prime** if

$$ab \in P \implies a \in P \lor b \in P. \tag{6.1}$$

So a ring in which $1 \neq 0$ is an integral domain if and only if (0) is a prime ideal. Compare the definition of prime ideal with the following: a positive integer p is prime if and only if

$$p \mid ab \implies p \mid a \lor p \mid b.$$

We shall address the relation between prime integers and prime ideals in §6.2. Meanwhile, an equivalent formulation of prime ideals is given by the following.

Theorem 120. *A proper ideal P of a ring is prime if and only if, for all ideals I and J of the ring,*

$$IJ \subseteq P \iff I \subseteq P \lor J \subseteq P. \tag{6.2}$$

Proof. The given condition has (6.1) as a special case, since the latter can be written as

$$(a)(b) \subseteq P \implies (a) \subseteq P \lor (b) \subseteq P.$$

Also, if (6.2) fails, so that $IJ \subseteq P$, but $I \smallsetminus P$ contains some a, and $J \smallsetminus P$ contains some b, then $ab \in P$, so (6.1) fails. □

Theorem 121. *A proper ideal P of a ring R is prime if and only if R/P is an integral domain.*

Proof. That I is prime means (6.1), which can be written as

$$(a+I)(b+I) = I \implies a+I = I \lor b+I = I;$$

but this means R/I is integral. □

An ideal is called **maximal** if it is maximal as a proper ideal. A ring is a field if and only if (0) is a maximal ideal. (Note that (0) is in fact the ideal with *no* generators, so it could be written as ( ); but it usually is not.)

Theorem 122. *A proper ideal I of a ring R is maximal if and only if R/I is a field.*

Proof. That R/I is a field means that, if $a \in R \smallsetminus I$, then for some b,

$$ab \in 1 + I.$$

That I is maximal means that, if $a \in R \smallsetminus I$, then

$$I + (a) = R,$$

equivalently, $1 \in I + (a)$, which means that, for some b, $ba - 1 \in I$. □

Corollary. *Maximal ideals are prime.*

The converse fails easily, since the prime ideals of $\mathbb{Z}$ are the ideals (0) and (p), where p is prime, and the latter are maximal, but (0) is not. However, it is not even the case that prime ideals other than (0) are always maximal. For example, $\mathbb{R}[X,Y]$ has the prime ideal (X), which is not maximal.

A ring is **Boolean** if it respects the identity

$$x^2 = x.$$

For example, if Ω is a set, then $\mathcal{P}(\Omega)$ is a Boolean ring, where multiplication is intersection, and addition is the taking of **symmetric differences,** where the symmetric difference of x and y is $x \smallsetminus y) \cup (y \smallsetminus x)$, denoted by $x \bigtriangleup y$.

Theorem 123. *In Boolean rings, all prime ideals are maximal.*

Proof. In a Boolean ring, we have $2x = (2x)^2 = 4x^2 = 4x$, so

$$2x = 0.$$

(Thus nontrivial Boolean rings have characteristic 2.) Hence

$$x(1+x) = x + x^2 = x + x = 0,$$

so x is a zero-divisor unless it or $1 + x$ is 0, that is, unless x is 0 or 1. Therefore there are no Boolean integral domains besides $\mathbb{F}_2$, which is a field. □

In $\mathbb{Z}$, the ideal (a,b) is the principal ideal generated by $\gcd(a,b)$. So a and b are coprime if $(a,b) = \mathbb{Z}$. This condition can be written as $(a)+(b) = \mathbb{Z}$. Then the following generalizes Theorem 78.

Theorem 124 (Chinese Remainder)**.** *Suppose R has an indexed family $(I_i \colon i < n)$ of ideals such that $I_i + I_j = R$ in each case. Let $I = \bigcap_{i<n} I_i$. Then the monomorphism*

$$x + I \mapsto (x + I_0, \dots, x + I_{n-1}) \tag{6.3}$$

from R/I to $\sum_{i<n} R/I_i$ is an isomorphism.

Proof. We proceed by induction. The claim is trivially true when $n = 1$. Proving the inductive step reduces to the proving the claim when $n = 2$. In that case, we have $a_0 + a_1 = 1$ for some a_0 in I_0 and a_1 in I_1. Then

$$a_0 \equiv 1 \pmod{I_1}, \qquad\qquad a_0 \equiv 0 \pmod{I_0},$$

and similarly for a_1. Therefore

$$a_0x_0 + a_1x_1 \equiv x_0 \pmod{I_0}, \qquad a_0x_0 + a_1x_1 \equiv x_1 \pmod{I_1}.$$

Thus $(x_0 + I_0, x_1 + I_1)$ is in the image of the map in (6.3). □

6.2. Factorization

(Recall that all rings are now commutative with identity.) In a ring R, an element a is a **divisor** of b, or a **divides** b, and we write

$$a \mid b,$$

if $ax = b$ for some x in R. Two elements that divide each other are **associates.**

Theorem 125. *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

$$2 \cdot 2 \equiv 4, \qquad\qquad 4 \cdot 2 \equiv 2 \pmod 6; \tag{6.4}$$

but 3 is not an associate of these.

In $\mathbb{Z}$, a **prime number** can be defined as a positive number p with either of two properties:

1. if $p = ab$, then one of a and b is ± 1;
2. if $p \mid ab$, then $p \mid a$ or $p \mid b$.

Easily (2) implies (1), since if $p = ab$, then $p \mid ab$, so that, if also $p \mid b$, then, since $b \mid p$, we have $b = \pm p$, so $a = \pm 1$. Conversely, (1) implies (2), with more difficulty. Indeed, property (1) implies that, if $p \nmid a$, then $\gcd(p, a) = 1$, so $px + ay = 1$ for some x and y. If also $p \mid ab$, but $p \nmid a$, then, since $b = pbx + aby$, we have $p \mid b$.

We let (2) be the defining property of *primes;* and (1), *irreducibles.* More precisely, an element of a ring is **irreducible** if it is not a unit or 0, and its only divisors are associates and units. So the element is irreducible just in case the ideal it generates is maximal amongst the proper principal ideals.

For example, in $\mathbb{R}[X, Y]$, the element X is irreducible, although (X) is not a maximal ideal. However, if $(X) \subseteq (f(X, Y)) \subset \mathbb{R}[X, Y]$, then $f(X, Y)$ must be constant in Y, and then it must have degree 1 in X, and then its constant term must be 0; so $f(X, Y)$ is just aX for some a in $\mathbb{R}^\times$.

An element of a ring is **prime** if it is not 0 and the ideal that it generates is prime in the sense of §6.1.

For example:

1. The primes of $\mathbb{Z}$ are the integers $\pm p$, where p is a prime natural number, and these are just the irreducibles of $\mathbb{Z}$.

2. In $\mathbb{Z}/6\mathbb{Z}$, the element 2 is prime. Indeed, the multiples of 2 are 0, 2, and 4, so the non-multiples are 1, 3, and 5, and the product of no two of these is a multiple of 2. Similarly, 4 is prime. However, 2 and 4 are not irreducible, by (6.4).

3. In $\mathbb{C}$ we have

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}), \tag{6.5}$$

so, because the factors 2, 3, and $1 \pm \sqrt{-5}$ are all irreducible in the smallest sub-ring of $\mathbb{C}$ that contains $\sqrt{-5}$, those factors cannot be prime in that ring. Details are worked out in the next section.

6.3. Some algebraic number theory

Suppose d is a **squarefree** integer, that is, an integer different from 1 that is not divisible by the square of a prime number. The subset $\{x + y\sqrt{d}\colon x, y \in \mathbb{Q}\}$ of $\mathbb{C}$ is a field, denoted by

$$\mathbb{Q}(\sqrt{d}).$$

Define

$$\tau_d = \begin{cases} \sqrt{d}, & \text{if } d \not\equiv 1 \pmod 4, \\ \dfrac{1+\sqrt{d}}{2}, & \text{if } d \equiv 1 \pmod 4. \end{cases}$$

The abelian subgroup $\langle 1, \omega \rangle$ of $\mathbb{Q}(\sqrt{d})$ is a sub-ring, denoted by

$$\mathbb{Z}[\tau_d].$$

Theorem 126. *The elements of $\mathbb{Z}[\tau_d]$ are precisely the solutions in $\mathbb{Q}(\sqrt{d})$ of an equation*

$$x^2 + bx + c = 0,$$

where b and c are in $\mathbb{Z}$.

Proof. From school the solutions of (126) 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. This establishes $x \in \mathbb{Z}[\tau_d]$ in all cases.

Conversely, suppose $x = k+n\tau_d$ for some k and n in $\mathbb{Z}$. If $d \equiv 1 \pmod 4$, then

$$\begin{gathered}2x - 2k - n = n\sqrt{d},\\ 4x^2 - 4(2k+n)x + (2k+n)^2 = n^2d,\\ x^2 - (2k+n)x + k^2 + kn + n^2\frac{1-d}{4} = 0,\end{gathered}$$

while if $d \not\equiv 1 \pmod 4$, then

$$x^2 - 2kx + k^2 - n^2d = 0.$$

In either case, $x \in \mathbb{Z}[\tau_d]$. □

The elements of $\mathbb{Z}[\tau_d]$ are therefore called the **integers** of $\mathbb{Q}(\sqrt{d})$. Since $\mathbb{Z}[\tau_d] \cap \mathbb{Q} = \mathbb{Z}$, we may refer to the elements of $\mathbb{Z}$ as **rational integers.** We have for example (6.5) in $\mathbb{Z}[\tau_{-5}]$; to show that 2, 3 and $1 \pm \tau_{-5}$ are irreducible in this ring, we define, in the general case, the operation $z \mapsto z'$ on $\mathbb{Q}(\sqrt{d})$ by

$$(x + y\sqrt{d})' = x - y\sqrt{d}.$$

This is an *automorphism* of $\mathbb{Q}(\sqrt{d})$. (It is the restriction of complex conjugation, if $d < 0$.) Then we define a **norm** function N from $\mathbb{Q}(\sqrt{d})$ to $\mathbb{Q}$ by

$$N(z) = zz'.$$

Then N is multiplicative, that is,

$$N(\alpha\beta) = N(\alpha)N(\beta).$$

Also,

$$N(x + \tau_d y) = \begin{cases} x^2 - dy^2, & \text{if } d \not\equiv 1 \pmod 4,\\ x^2 + xy + \dfrac{1-d}{4}y^2, & \text{if } d \equiv 1 \pmod 4,\end{cases}$$

so N maps $\mathbb{Z}[\tau_d]$ into $\mathbb{Z}$. If $d < 0$, then it maps $\mathbb{Z}[\tau_d]$ into $\mathbb{N}$. Let us restrict our attention to this case. Here, α is a unit in $\mathbb{Z}[\tau_d]$ if and only if $N(\alpha) = 1$. Therefore α in $\mathbb{Z}[\tau_d]$ is irreducible if and only if it has no divisor β such that $1 < N(\beta) < N(\alpha)$. In case $d = -5$ we have

$$\begin{array}{c||c|c|c} x & 2 & 3 & 1 \pm \tau_{-5} \\ \hline N(x) & 4 & 9 & 6 \end{array}. \tag{6.6}$$

Since no elements of $Z[\tau_{-5}]$ have norm 2 or 3, the elements 2, 3, and $1 \pm \tau_{-5}$ are irreducible.

But they are not prime. Indeed, if $\alpha \mid \beta$, then $N(\alpha) \mid N(\beta)$; but no norm in (6.6) divides another. This is where *ideals* come up. There are factorizations of the relevant ideals:

$$\begin{aligned} (2) &= (2, 1 + \tau_{-5})^2, \\ (3) &= (3, 1 + \tau_{-5})(3, 1 - \tau_{-5}), \\ (1 + \tau_{-5}) &= (2, 1 + \tau_{-5})(3, 1 + \tau_{-5}), \\ (1 - \tau_{-5}) &= (2, 1 + \tau_{-5})(3, 1 - \tau_{-5}). \end{aligned} \tag{6.7}$$

For example,

$$(2, 1 + \tau_{-5})(2, 1 + \tau_{-5}) = (2, 1 + \tau_{-5})(2, 1 - \tau_{-5}) = (4, 2 + 2\tau_{-5}, 6) = (2).$$

The right-hand members of (6.7) are in fact prime factorizations. To see this, we first note that, being a subgroup of $\langle 1, \tau_d \rangle$ on more than one generator, an ideal I of $\mathbb{Z}[\tau_d]$ can be written as $\langle a + b\tau_d, c + d\tau_d \rangle$, where

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(\mathbb{Z}) \cap \mathrm{GL}_2(\mathbb{Q}).$$

Multiplication on the left by a matrix in $\mathrm{GL}_2(\mathbb{Z})$ does not change the ideal. Hence we can define

$$N(I) = \left| \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right|,$$

which is in $\mathbb{N}$. In case $d < 0$, this agrees with the function N defined above in the sense that $N((\alpha)) = N(\alpha)$, because

$$(a + b\tau_d)\langle 1, \tau_d \rangle = \langle a + b\tau_d, db + a\tau_d \rangle.$$

Moreover, if $I \subset J \subset \mathbb{Z}[\tau_d]$, then $N(J) \mid N(I)$ and $N(I) > N(J) > 1$. In case $d = -5$, we compute

$$\begin{aligned}(2, 1 + \tau_{-5}) &= \langle 2, 2\tau_{-5}, 1 + \tau_{-5}, \tau_{-5} - 5\rangle = \langle 2, 1 + \tau_{-5}\rangle,\\ (3, 1 \pm \tau_{-5}) &= \langle 3, 3\tau_{-5}, 1 \pm \tau_{-5}, \tau_{-5} \mp 5\rangle = \langle 3, 1 \pm \tau_{-5}\rangle,\end{aligned}$$

hence

I	$(2, 1 + \tau_{-5})$	$(3, 1 \pm \tau_{-5})$.
$N(I)$	2	3

So these ideals are maximal, hence prime. Ideals of the rings $\mathbb{Z}[\tau_d]$ were originally called **ideal numbers.**

6.4. Integral domains

Theorem 127. *In an integral domain, if a and b are non-zero associates, and $a = bx$, then x is a unit.*

Proof. We have also $b = ay = bxy$, $b(1 - xy) = 0$, $1 = xy$ since $b \neq 0$ and we are in an integral domain. □

Corollary. *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. □

A **unique factorization domain** or UFD is an integral domain whose every non-zero element is 'uniquely' a product of irreducibles. This means that, if

$$\prod_{i<n} \pi_i = \prod_{i<n'} \pi'_i,$$

where the π_i and π'_i are irreducible, then $n = n'$, and (perhaps after re-indexing) π_i and π'_i are associates. Hence:

Theorem 128. *In a UFD, irreducibles are prime.* □

In any ring, a **greatest common divisor** of elements a and b is an element of the set of all divisors 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 is a greatest common divisor of itself and 0.

Theorem 129. *In a UFD, any two elements have a greatest common divisor.*

Proof. If they are nonzero, we can write the elements as

$$u \prod_{i<n} {\pi_i}^{a(i)}, \qquad\qquad v \prod_{i<n} {\pi_i}^{b(i)},$$

where u and v are units and the π_i are irreducibles; a greatest common divisor is then

$$\prod_{i<n} {\pi_i}^{\min(a(i),b(i))}. \qquad \square$$

In a PID, more is true:

Theorem 130. *In a PID, any two elements have a greatest common divisor, which is some linear combination of those elements.*

Proof. If $(a, b) = (c)$, then c is a greatest common divisor of a and b, and $c = ax = by$ for some x and y in the ring. □

Lemma 23. *In a PID, irreducibles are prime.*

Proof. Suppose the irreducible π divides ab but not a. Then a greatest common divisor of π and a is 1; hence $\pi x + ay = 1$ for some x and y in the ring. Then $b = \pi xb + aby$, and π divides each summand, so $\pi \mid b$. □

Lemma 24. *In a PID, irreducible factorizations are unique.*

A ring is **Noetherian** if every strictly ascending chain of ideals is finite.

Theorem 131. *PIDs are 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. □

Lemma 25. *In a PID, every element is a product of irreducibles.*

Proof. A tree of factorizations has no infinite branches. More precisely, let a be an element of a PID. For certain finite binary sequences σ, we define a_σ thus: $a_{()} = a$, and if $a_{(e(0),\dots,e(n-1))}$ can be factorized as bc, where neither b nor c is a unit, then let $a_{(e(0),\dots,e(n-1),0)} = b$ and $a_{(e(0),\dots,e(n-1),1)} = c$; otherwise these are undefined. Then every branch of the tree corresponds to a chain

$$(a_{()}) \subset (a_{(e(0))}) \subset (a_{(e(0),e(1))}) \subset (a_{(e(0),e(1),e(2))}) \subset \cdots,$$

so it must be finite. Therefore the whole tree is finite, and a is the product of the irreducibles found at the end of each branch. □

Theorem 132. *A PID is a UFD.* □

Recall how the Euclidean algorithm for finding greatest common divisors works. To find $\gcd(201, 27)$, compute:

$$\begin{aligned} 201 &= 87 \cdot 2 + 27, \\ 87 &= 27 \cdot 3 + 6, \\ 27 &= 6 \cdot 4 + 3, \\ 6 &= 3 \cdot 2. \end{aligned}$$

So $\gcd(201, 27) = 3$. In general, if $a_0 \geqslant a_1 > 0$, then $\gcd(a_0, a_1) = a_n$, where there is a descending sequence $(a_0, \dots a_n)$ of positive integers such that $a_{k+2} = a_{k+1} \cdot b_k + a_k$ for some b_k. A **Euclidean domain** is then an integral domain in which the Euclidean algorithm works. More precisely, a Euclidean domain is a domain R equipped with a map φ from $R \smallsetminus \{0\}$ to ω such that, and, for all a and b in $R \smallsetminus \{0\}$, one of the following holds:

- there exist q in R and r in $R \smallsetminus \{0\}$ such that $a = qb + r$ and $\varphi(r) < \varphi(b)$, or
- $b \mid a$ and $\varphi(b) \leqslant \varphi(a)$.

For example:

1. $\mathbb{Z}$ is Euclidean with respect to $x \mapsto |x|$;
2. a field, $x \mapsto 0$;
3. a polynomial-ring $K[X]$ over a field K, $f \mapsto \deg f$ (see §6.7).

The **Gaussian integers** are the elements of $\mathbb{Z}[\tau_{-1}]$, where $\tau_{-1} = \sqrt{-1} = \mathrm{i}$ as in §6.3. This domain is Euclidean with respect to the norm function, namely $z \mapsto |z|^2$, where $|x + y\mathrm{i}|^2 = x^2 + y^2$. Indeed, if a and b are nonzero Gaussian integers, then there is a Gaussian integer q such that $|a/b - q| \leqslant \sqrt{2}/2$. Let $r = a - bq$; then $|r|^2 = |b|^2 \cdot |a/b - q|^2 \leqslant |b|^2 / 2$.

Theorem 133. *Euclidean domains are PIDs.*

Proof. An ideal of a Euclidean domain is generated by any non-zero element x such that $\varphi(x)$ is minimal. □

6.5. Localization

A subset of a ring is **multiplicative** if it is closed under multiplication. For example, the complement of a prime ideal is multiplicative.

Lemma 26. *If S is a multiplicative subset of a ring R, then on $R \times S$ there is an equivalence-relation $\sim$ given by*

$$(a, b) \sim (c, d) \iff (ad - bc) \cdot e = 0 \textit{ for some } e \textit{ in } S. \tag{6.8}$$

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

$$\frac{a}{b},$$

and the quotient $R \times S/{\sim}$ is denoted by

$$S^{-1}R.$$

If R is an integral domain, and $0 \notin S$, then (6.8) can be simply

$$(a,b) \sim (c,d) \iff ad - bc = 0.$$

If $0 \in S$, then $S^{-1}R$ has a unique element. An instance where R is not an integral domain will be considered in the next section.

Theorem 134. *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 φ from R to $S^{-1}R$ where, for every a in S,*
$$\varphi(x) = \frac{xa}{a}.$$

Suppose in particular R is an integral domain and $0 \notin S$.

4. *$S^{-1}R$ is an integral domain, and the homomorphism φ is an embedding.*
5. *If $S = R \smallsetminus \{0\}$, then $S^{-1}R$ is a field, and if If ψ is an embedding of R in a field K, then there is an embedding $\tilde{\psi}$ of $S^{-1}R$ in K such that $\tilde{\psi} \circ \varphi = \psi$.*

In the most important case, S is the complement of a prime ideal $\mathfrak{p}$, and then $S^{-1}R$ is called the **localization** of R at $\mathfrak{p}$, denoted by

$$R_{\mathfrak{p}}.$$

If R is an integral domain, so that (0) is prime, then $R_{(0)}$ (which is a field by the theorem) is the **quotient-field** of R. 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 27. *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 135. *The localization of a ring at a prime ideal is a local ring.*

Proof. The ideal generated by the image of $\mathfrak{p}$ in $R_{\mathfrak{p}}$ consists of those a/b such that $a \in \mathfrak{p}$. In this case, if $c/d = a/b$, then $cb = da \in \mathfrak{p}$, so $c \in \mathfrak{p}$ since $\mathfrak{p}$ is prime. Hence the following are equivalent:

1. $x/y \notin R_{\mathfrak{p}}\mathfrak{p}$;
2. $x \notin \mathfrak{p}$;
3. x/y has an inverse, namely y/x.

By the lemma, we are done. □

6.6. Ultraproducts of fields

Suppose $\mathcal{K}$ is an indexed family $(K_i : i \in A)$ of fields. If $a \in \prod \mathcal{K}$, there is an element a^* of $\prod K$ 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

$$aa^*a = a.$$

Because of this, $\prod \mathcal{K}$ is an example of a **regular ring** (in the sense of von Neumann).[1]

Theorem 136. *In a regular ring, all prime ideals are maximal.*

[1] In general, a regular ring need not be commutative; see [8, IX.3, ex. 5, p. 442].

Proof. Let R be a regular integral domain. If $a \in R \smallsetminus \{0\}$, then, since

$$0 = aa^*a - a = a(a^*a - 1),$$

we have $a^*a = 1$. Thus R is a field. □

Theorem 137. *If $\mathfrak{p}$ is a prime ideal of a regular ring R, then*

$$R/\mathfrak{p} \cong R_{\mathfrak{p}},$$

the isomorphism being $x + \mathfrak{p} \mapsto x/1$.

Proof. If $a \in R$ and $b \in R \smallsetminus \mathfrak{p}$, then $a/b = ab^*/1$ since

$$(a - bab^*)b = ab - abb^*b = ab - ab = 0.$$

Thus the homomorphism $x \mapsto x/1$ guaranteed by Theorem 134 is surjective. We also have $a/1 = 0/1$ if and only if $ab = 0$ for some b in $R \smallsetminus \mathfrak{p}$; but the latter implies $ab \in \mathfrak{p}$, so $a \in \mathfrak{p}$ since the ideal is prime. Conversely, if $a \in \mathfrak{p}$, then $a^*a \in \mathfrak{p}$, so $a^*a - 1 \notin \mathfrak{p}$ since the ideal is proper; but $a(a^*a - 1) = 0$, so $a/1 = 0/1$. Therefore the kernel of the homomorphism is $\mathfrak{p}$. □

With $\mathcal{K}$ as above, there is a one-to-one correspondence between ideals of $\prod \mathcal{K}$ and ideals of the Boolean ring $\mathcal{P}(A)$. To define this correspondence, we first define the **support** of an element a of $\prod \mathcal{K}$ to be the set of those i in A such that $\pi_i(a) \neq 0$. We may denote this set by $\operatorname{supp}(a)$. Then

$$\operatorname{supp}(ab) = \operatorname{supp}(a) \cap \operatorname{supp}(b), \qquad \operatorname{supp}(a + b) \subseteq \operatorname{supp}(a) \cup \operatorname{supp}(b).$$

So $x \mapsto \operatorname{supp}(x)$ is not quite a ring-homomorphism from $\prod \mathcal{K}$ to $\mathcal{P}(A)$. However, if I is an ideal of $\prod \mathcal{K}$, then $\operatorname{supp}[I]$ is an ideal of $\mathcal{P}(A)$. Indeed, for every subset B of A, there is an element e_B of $\prod \mathcal{K}$ given by

$$\pi_i(e_B) = \begin{cases} 1, & \text{if } i \in B, \\ 0, & \text{if } i \notin B. \end{cases}$$

Then $\operatorname{supp}(e_B) = B$. If $a \in \prod \mathcal{K}$, and $B = \operatorname{supp}(a)$, then $e_B = aa^*$. If, further, $a \in I$, and $C \subseteq B$, then $e_C = e_C aa^*$, so this is in I and therefore

$C \in \operatorname{supp}[I]$. Also, if B and C are in $\operatorname{supp}[I]$, then $B \triangle C = \operatorname{supp}(e_B - e_C)$, which is in $\operatorname{supp}[I]$. So $\operatorname{supp}[I]$ is indeed an ideal of $\mathcal{P}(A)$. If J is an ideal of $\mathcal{P}(A)$, then $J = \operatorname{supp}[I]$, where I is the ideal of $\prod \mathcal{K}$ generated by those e_B such that $B \in J$. Since every ideal I is generated by those e_B such that $B \in \operatorname{supp}[I]$, we conclude that φ is the claimed bijection.

Let $\mathfrak{p}$ be a prime ideal of $\prod \mathcal{K}$. Then the quotient $\prod \mathcal{K}/\mathfrak{p}$ is a field, called an **ultraproduct** of $\mathcal{K}$. Now, $\mathfrak{p}$ could be principal, in which case $\varphi(\mathfrak{p})$ would be principal; but since it is also maximal, it would have a set $A \smallsetminus \{i\}$ as a generator. In this case $\prod \mathcal{K}/\mathfrak{p} \cong K_i$.

However, $\mathcal{P}(A)$ has the ideal I consisting of the the finite subsets of A. If A itself is infinite, then I is a proper ideal. In this case, if $I \subseteq \operatorname{supp}[\mathfrak{p}]$, then $\mathfrak{p}$ is not principal, and the field $\prod \mathcal{K}/\mathfrak{p}$ is called a **nonprincipal ultraproduct** of $\mathcal{K}$. This is a sort of 'average' of the K_i. In particular, we have

$$\begin{aligned} a \equiv b \pmod{\mathfrak{p}} &\iff a - b \in \mathfrak{p} \\ &\iff \operatorname{supp}(a-b) \in \operatorname{supp}[\mathfrak{p}] \\ &\iff \{i \in A \colon \pi_i(a) \neq \pi_i(b)\} \in \operatorname{supp}[\mathfrak{p}]. \end{aligned}$$

We may think of the elements of $\operatorname{supp}[\mathfrak{p}]$ as 'small' sets; their complements are 'large'. (Then every subset of A is small or large.) So all finite subsets of A are small, and all cofinite subsets of A are large. Then elements of $\prod \mathcal{K}$ represent the same element in the ultraproduct if they agree on a large set.

Say for example A is the set of prime numbers in ω, along with 0, and each K_p has characteristic p. Then $\prod \mathcal{K}/\mathfrak{p}$ has characteristic 0, since for each prime p, the element $p1$ of $\prod \mathcal{K}$ disagrees with 0 on a large set.

The proof that nonprincipal ultraproducts exist uses the Axiom of Choice. ∎

6.7. Factorization of polynomials

Theorem 138. *If R is a ring, then $R[X_0, \dots, X_{n-1}]$ is the unique ring-extension A of R such that, for all rings S, and all homomorphisms φ from R to S, and all $\vec{a}$ in S^n, there is a unique homomorphism $\tilde{\varphi}$ from A to S such that $\tilde{\varphi}|_R = \varphi$ and $\tilde{\varphi}(X^i) = a^i$ in each case.*

An arbitrary element of $R[X]$ can be written

$$\sum_{i \leqslant n} a_i X^i;$$

the **degree** of this is n, if $a_n \neq 0$; then a_n is the **leading coefficient** of the polynomial.

We said in §6.4 that $K[X]$ is a Euclidean domain when equipped with deg. More generally:

Lemma 28. *If f and g are polynomials over R, then:*

- $\deg(f+g) \leqslant \max(\deg f, \deg g)$*;*
- $\deg(f \cdot g) \leqslant \deg f + \deg g$*, with equality if the product of the leading coefficients is not* 0*.*

In particular, if R is an integral domain, then so is $R[X]$.

Proof. The leading coefficient of a product is the product of the leading coefficients. □

Lemma 29 (Division Algorithm)**.** *If f and g are polynomials in X over R, and the leading coefficient of g is* 1*, then*

$$f = qg + r$$

for some unique q and r in $R[X]$ such that $\deg r < \deg g$*.*

Proof. If $\deg g \leqslant \deg f$, and a is the leading coefficient of f, then

$$f = aX^{\deg f - \deg g} \cdot g + (f - aX^{\deg f - \deg g} \cdot g),$$

the second term having degree less than f. Continue as necessary. □

Lemma 30 (Remainder Theorem)**.** *If $c \in R$, then any f in $R[X]$ can be written uniquely as $q(X) \cdot (X - c) + f(c)$.*

Proof. By the Division Algorithm, $f = q(X) \cdot (X - c) + d$ for some d in R; letting X be c yields the claim. □

Theorem 139. *A ring-element c is a zero of a polynomial f if and only if $(X-c) \mid f$. If f is over an integral domain, then the number of its distinct zeros is at most* $\deg f$*.*

Proof. By the Remainder Theorem, c is a zero of f if and only if $f = q(X) \cdot (X-c)$ 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, …, r_{n-1}, then repeated application of the Remainder Theorem yields

$$f = (X - r_0) \cdots (X - r_{m-1}).$$

Then every zero of f is a zero of one of the $X - r_k$, so it must be r_k. □

Recall however from the proof of Theorem 123 that every element of a Boolean ring is a zero of $X(1+X)$, that is, $X + X^2$; but some Boolean rings have more than two elemments. In $\mathbb{Z}_6$, the same polynomial has the zeros 0, 2, 3, and 5.

Theorem 140. *If K is a field, then $K[X]$ is a Euclidean domain whose units are precisely the elements of K.*

Proof. Over a field, the Division Algorithm does not require the leading coefficient of the divisor to be 1. □

A zero c of a polynomial over an integral domain has **multiplicity** m if the polynomial can be written as $g(X) \cdot (X-c)^m$, where c is not a zero of g. A zero with multiplicity greater than 1 is **multiple.** Derivations were defined in §5.1; they will be useful for recognizing the existence of multiple roots.

Lemma 31. *If δ is a derivation of a ring R, then for all x in R and n in ω,*

$$\delta(x^n) = nx^{n-1}\delta(x).$$

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

$$\delta(x^{k+1}) = \delta(x)x^k + x\delta(x^k) = \delta(x)x^k + kx^k\delta(x) = (k+1)x^k\delta(x),$$

so the claim holds when $n = k + 1$. □

Theorem 141. *On a polynomial ring $R[X]$, there is a unique derivation $f \mapsto f'$ such that*

1. $X' = 1$,
2. $c' = 0$ *for all* c *in* R.

This derivation is given by

$$\Bigl(\sum_{k=0}^{n} a_k X^k\Bigr)' = \sum_{k=0}^{n-1} (k+1)a_{k+1}X^k. \tag{6.9}$$

Proof. Uniqueness and (6.9) follow from the lemma and the definition of a derivation. If δ is a derivation, then $\delta(x \cdot (y+z)) = \delta(xy+xz)$. Also, (6.9) does define an endomorphism of the underlying group of $R[X]$ that meets the given conditions. Because

$$\begin{aligned}(X^k)'(X^\ell) + X^k(X^\ell)' &= kX^{k-1}X^\ell + \ell X^k X^{\ell-1}\\ &= (k+\ell)X^{k+\ell+1}\\ &= (X^{k+\ell})',\end{aligned}$$

the additive endomorphism $f \mapsto f'$ of $R[X]$ is a derivation. □

In the notation of the theorem, f' is the **derivative** of f.

Lemma 32. *Say R is an integral domain, $f \in R[X]$ 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

$$f' = m(X - c)^{m-1} \cdot g + (X - c)^m \cdot g'.$$

If $m > 1$, then $f'(c) = 0$. If $f'(c) = 0$, then $m \cdot 0^{m-1} \cdot g(c) = 0$, so $m > 1$. □

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 $\mathbb{R}$, but not over $\mathbb{C}$. Likewise, the polynomial may have zeros from L, but not K. Hence it makes sense to speak of zeros of an irreducible polynomial.

Theorem 142. *Supppose K is a field and $f \in K[X]$.*

1. *If $\gcd(f, f') = 1$, then f has no multiple zeros.*
2. *If f is irreducible, then $\gcd(f, f')$ is 1 or 0.*
3. *If $\gcd(f, f') = 0$, then K has a positive characteristic p, and $f = g(X^p)$ for some polynomial g over K.*

Proof. If $\gcd(f, f') = 1$, then $1 = g \cdot f + h \cdot f'$ for some polynomials g and h, so f and f' can have no common zero. Since $\deg(f') < \deg(f)$ by (6.9), if f is irreducible and $\gcd(f, f') \neq 1$, then $\gcd(f, f') = 0$. The rest also follows from (6.9). □

A polynomial over a UFD is **primitive** if 1 is a greatest common divisor of its coefficients.

Lemma 33 (Gauss). *The product of primitive polynomials 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 the c_k have a common prime factor π, but f is primitive. There is some ℓ such that $\pi \mid a_i$ when $i < \ell$, but $\pi \nmid a_\ell$. Since $\pi \mid c_\ell$, we have $\pi \mid b_0$; then, since $\pi \mid e_{\ell+1}$, we have $\pi \mid b_1$, and so on. So g is not primitive. □

Henceforth let R be a UFD with quotient field K.

Lemma 34. *Primitive polynomials over R that are associated over K are associated over R.*

Proof. If f and g are polynomials defined over R, but associated over K, then they must have the same degree, and so we have $af = bg$ for some a and b in R. If f and g are primitive, then a and b must be associates, so $b = ua$ for some unit in R, and then $f = ug$, so f and g are associates. □

Lemma 35. *Primitive polynomials over R are irreducible over R if and only if irreducible over K.*

Proof. Say f and g are defined over K, but fg is over R and primitive. Then af and bg are over R and primitive for some a and b in R. By a previous lemmma, $abfg$ is primitive; but so is fg, so ab must be a unit in R. Hence a and b are units in R, so f and g are over R. Since units of $R[X]$ are units of $K[X]$, it follows that a primitive polynomial irreducible polynomial over R is still irreducible over K. Also, any non-unit factor of a *primitive* polynomial over R is still not a unit over K, so the polynomial is reducible over K. □

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 143. *$R[X]$ is a UFD.*

Proof. Every element of $R[X]$ can be written as af, where $a \in R$ 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 R such that $a_k f_k$ is a primitive polynomial over R. The product of these is still primitive, so the product of the a_k must be a unit in R, hence each a_k is a unit in R. Thus f has an irreducible factorization over R. Its uniqueness follows from its uniqueness over K and the next-to-last lemma. □

Theorem 144 (Eisenstein's Criterion)**.** *If f is a polynomial $\sum_{k=0}^{n} a_k X^k$ over R, and π is an irreducible element of R such that*

$$\pi^2 \nmid a_0, \qquad \pi \mid a_0, \qquad \pi \mid a_1, \qquad \dots, \qquad \pi \mid a_{n-1}, \qquad \pi \nmid a_n,$$

then f is irreducible over K and, if primitive, over R.

Proof. Suppose $f = gh$, where $g = \sum_{k=0}^{n} b_k X^k$ and $h = \sum_{k=0}^{n} c_k X^k$, all coefficients from R (and some being 0). We may assume f is primitive, so g and h must be primitive. We may assume π divides b_0, but not c_0. Let ℓ 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 $\deg g = n$, and $h = c_0$ and 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 $\deg g = n$. □

An application is the following.

Theorem 145. *If p is prime, then $\sum_{k=0}^{p-1} X^k$ is irreducible.*

Proof. Consider

$$\sum_{k=0}^{p-1}(X+1)^k = \sum_{k=0}^{p-1}\sum_{j=0}^{k}\binom{k}{j}X^j = \sum_{j=0}^{p-1} X^j \sum_{k=j}^{p-1}\binom{k}{j} = \sum_{j=0}^{p-1} X^j \binom{p}{j+1},$$

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 $j < p-1$. □

A. The German script

In his encyclopedic *Model Theory* of 1993, Wilfrid Hodges observes [6, 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.

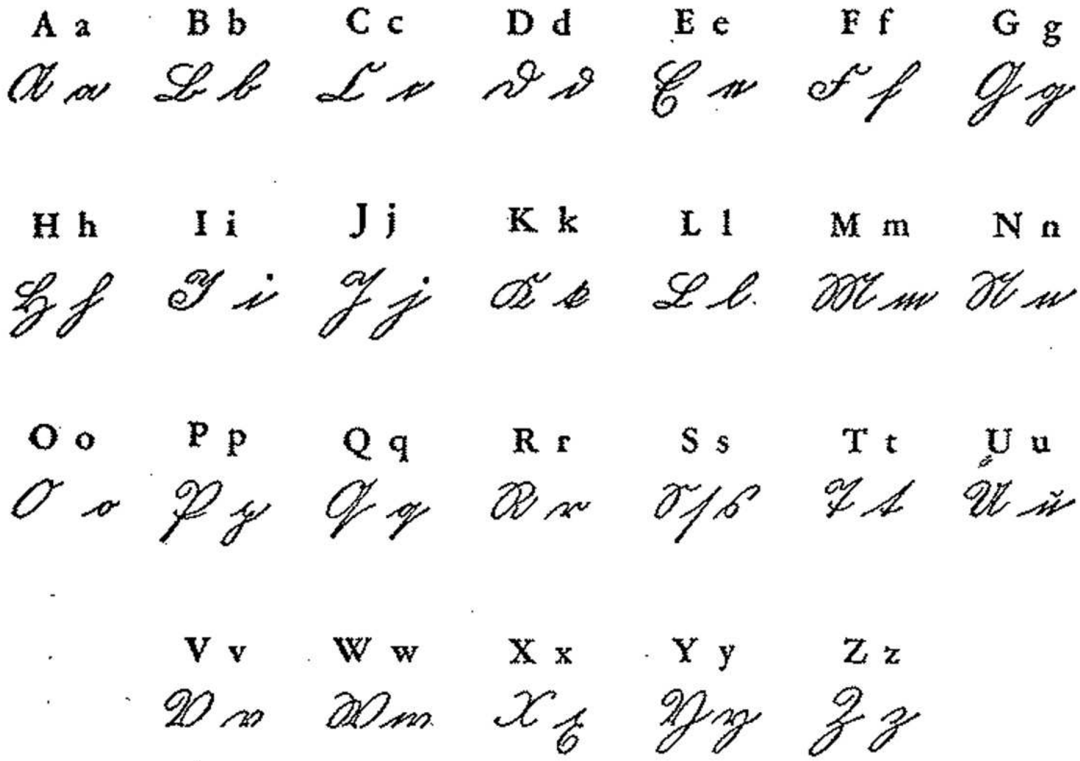
For Hodges, *structures* (as defined in §0.5 above) are denoted by the letters A, B, C, and so forth; he refers to their universes as **domains** and denotes these by $\operatorname{dom}(A)$ and so forth. This practice is convenient if one is using a typewriter (as in the preparation of another of Hodges's books [7], from 1985). In his *Model Theory: An Introduction* of 2002, David Marker [12] 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 also occasionally useful:

$$\begin{array}{ccccccccccccc}
\mathfrak{a} & \mathfrak{b} & \mathfrak{c} & \mathfrak{d} & \mathfrak{e} & \mathfrak{f} & \mathfrak{g} & \mathfrak{h} & \mathfrak{i} & \mathfrak{j} & \mathfrak{k} & \mathfrak{l} & \mathfrak{m} \\
\mathfrak{n} & \mathfrak{o} & \mathfrak{p} & \mathfrak{q} & \mathfrak{r} & \mathfrak{s} & \mathfrak{t} & \mathfrak{u} & \mathfrak{v} & \mathfrak{w} & \mathfrak{x} & \mathfrak{y} & \mathfrak{z}
\end{array}$$

A way to write these letters by hand is seen in a textbook on the German language from 1931 [4]:


A a
B b
C c
D d
E e
F f
G g
H h
I i
J j
K k
L l
M m
N n
O o
P p
Q q
R r
S s
T t
U u
V v
W w
X x
Y y
Z z

B. Group-actions

This chapter is a suggested reference from page 93. The chapter is partially inspired by an expository article [16] by Serre. Suppose a group G acts on a set A by $(g, x) \mapsto gx$. Just as, for an element a of A, we define

$$G_a = \{g \in G \colon ga = a\},$$

so, for an element g of G, we may define

$$A^g = \{x \in A \colon gx = x\} :$$

this is the set of **fixed points** of g. The orbit of a under the action of G is defined by

$$Ga = \{ga \colon g \in G\}.$$

Then $ga = ha \iff gG_a = hG_a$, and therefore

$$|Ga| = [G : G_a],$$

and the sets Ga partition G. We may define

$$A/G = \{Gx \colon x \in A\}.$$

Assume G is finite. For any function φ from G to $\mathbb{R}$ and subset X of G, we define

$$\int_X \varphi = \sum_{g \in X} \frac{\varphi(g)}{|G|}, \qquad \int \varphi = \int_G \varphi.$$

Assume A is also finite, and let χ be the function

$$g \mapsto |A^g|$$

from G to ω.

Lemma 36 (Burnside)**.** $|A/G| = \int \chi$.

Proof. Letting $R = \{(g,x) \in G \times A \colon gx = x\}$, we define π_G as $(g,x) \mapsto g$ from R to G, and π_A as $(g,x) \mapsto x$ from R to A. Then

$$|R| = \sum_{g \in G} |{\pi_G}^{-1}(g)| = \sum_{g \in G} \chi(g),$$

but also

$$|R| = \sum_{x \in A} |G_x| = \sum_{C \in A/G} \sum_{x \in C} |G_x|.$$

But if $C \in A/G$ and $a \in C$, then $C = [G : G_a]$. Hence

$$\sum_{C \in A/G} \sum_{x \in C} |G_x| = \sum_{C \in A/G} \sum_{x \in C} \frac{|G|}{|C|} = \sum_{C \in A/G} |G| = |A/G| \cdot |G|. \qquad \square$$

Now define

$$G_0 = \{g \in G \colon A^g = \varnothing\},$$

the set of elements of G with no fixed points.

Theorem 146 (Jordan). *If $|A/G| = 1$ and $|A| \geqslant 2$, then*

$$G_0 \neq \varnothing.$$

Proof. By the Burnside Lemma, the average size of A^g is 1. Since $A^1 = A$, and $|A| \geqslant 2$, we must have $|A|^g < 1$ for some g in G. $\square$

A stronger result is the following:

Theorem 147 (Cameron–Cohen). *If $|A/G| = 1$ and $|A| \geqslant 2$, then*

$$|G_0| \cdot |A| \geqslant |G|.$$

Proof. The action of G on A induces an action on $A \times A$, and $|(A \times A)^g| = \chi(g)^2$. Now, $(A \times A)/G$ contains the diagonal $G(1,1)$ and at least one other element, so

$$\int \chi^2 \geqslant 2$$

by Burnside's Lemma. Let $n = |A|$. Then for all g in $G \smallsetminus G_0$, we have $1 \leqslant \chi(g) \leqslant n$ and therefore

$$(\chi(g)-1)(\chi(g)-n) \leqslant 0;$$

but $(\chi(g)-1)(\chi(g)-n) = n$ when $g \in G_0$. Consequently,

$$\begin{aligned}\frac{|G_0| \cdot |A|}{|G|} = n\int_{G_0} 1 &= \int_{G_0} (\chi-1)(\chi-n) \\ &\geqslant \int_G (\chi-1)(\chi-n) = \int_G (\chi^2-1) \geqslant 1. \qquad \square\end{aligned}$$

Serre's article gives applications to topology and number-theory.

Bibliography

[1] Chen Chung Chang. On unions of chains of models. *Proc. Amer. Math. Soc.*, 10:120–127, 1959.

[2] 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.

[3] Joel David Hamkins. Every group has a terminating transfinite automorphism tower. *Proc. Amer. Math. Soc.*, 126(11):3223–3226, 1998.

[4] Roe-Merrill S. Heffner. *Brief German Grammar.* D. C. Heath and Company, Boston, 1931.

[5] Leon Henkin. On mathematical induction. *Amer. Math. Monthly*, 67:323–338, 1960.

[6] Wilfrid Hodges. *Model theory*, volume 42 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1993.

[7] Wilfrid Hodges. *Building models by games.* Dover Publications, Mineola, New York, 2006. original publication, 1985.

[8] Thomas W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1980. Reprint of the 1974 original.

[9] Morris Kline. *Mathematical thought from ancient to modern times.* Oxford University Press, New York, 1972.

[10] 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.

[11] Jerzy Łoś and Roman Suszko. On the extending of models (IV): Infinite sums of models. *Fund. Math.*, 44:52–60, 1957.

[12] David Marker. *Model theory: an introduction*, volume 217 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[13] James H. McKay. Another proof of Cauchy's group theorem. *Amer. Math. Monthly*, 66:119, 1959.

[14] Giuseppe Peano. The principles of arithmetic, presented by a new method (1889). In van Heijenoort [19], pages 83–97.

[15] Bertrand Russell. Letter to Frege (1902). In van Heijenoort [19], pages 124–5.

[16] Jean-Pierre Serre. On a theorem of Jordan. *Bull. Amer. Math. Soc. (N.S.)*, 40(4):429–440 (electronic), 2003.

[17] Thoralf Skolem. Some remarks on axiomatized set theory (1922). In van Heijenoort [19], pages 290–301.

[18] Simon Thomas. The automorphism tower problem. *Proc. Amer. Math. Soc.*, 95(2):166–168, 1985.

[19] Jean van Heijenoort, editor. *From Frege to Gödel: A source book in mathematical logic, 1879–1931.* Harvard University Press, Cambridge, MA, 2002.

[20] John von Neumann. An axiomatization of set theory (1925). In van Heijenoort [19], pages 393–413.

[21] John von Neumann. On the introduction of transfinite numbers (1923). In van Heijenoort [19], pages 346–354.

[22] Ernst Zermelo. Investigations in the foundations of set theory I (1908a). In van Heijenoort [19], pages 199–215.