Script: Algebra

A. Kresch

Autumn 2022

Contents

0 Review

The basic algebraic structures were defined in the Linear Algebra lecture. We review the definitions here. *A few new statements, which might not be familiar from the Linear Algebra lecture, are presented in italics.*

0.1 Groups

A **group** is a set G with a composition $G \times G \to G$, often denoted by $(a, b) \mapsto a \cdot b$, satisfying:

- (associativity axiom) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, $c \in G$,
- (identity element) there exists $e \in G$ such that $e \cdot a = a = a \cdot e$ for all $a \in G$,
- (inverses) for every $a \in G$, there exists $a' \in G$ with $a' \cdot a = e = a \cdot a'$.

The group is **abelian** if composition is commutative, i.e., $b \cdot a = a \cdot b$ for all a, $b \in G$.

For a group (abelian or not), we have:

- The element e in the definition of group is unique and is called the **identity element** of G.
- Given $a \in G$ the element $a' \in G$ (satisfying $a' \cdot a = e = a \cdot a'$) is unique and is called the **inverse** of a, written a^{-1}.

The composition $G \times G \to G$ may be denoted in other ways (ab, $a * b$, etc.), and especially for abelian groups it is common to use additive notation: composition $a + b$, identity element 0, inverse $-a$.

A **subgroup** of G is a nonempty subset $H \subset G$ that is closed under multiplication (for a, $b \in H$ we have $ab \in H$) and closed under inverse (for $a \in H$ we have $a^{-1} \in H$).

If G' (with a composition $G' \times G' \to G'$) is another group then a **group homomorphism** from G to G' is a map $f \colon G \to G'$ satisfying $f(ab) = f(a)f(b)$ for all a, $b \in G$. The **kernel** of a group homomorphism $f \colon G \to G'$ is the set of elements of G which are mapped to the identity of G'; it is a subgroup of G, denoted by $\ker(f)$, and f is injective if and only if $\ker(f)$ is trivial (i.e., consists of only the identity element of G). The image of f is a subgroup of G'. A **group isomorphism** is a bijective group homomorphism.

0.2 Rings, fields

A **ring** is a set R with a pair of compositions $R \times R \to R$, $(a, b) \mapsto a + b$ (addition) and $R \times R \to R$, $(a, b) \mapsto ab$ (multiplication). These are required to satisfy:

- $(R, +)$ is an abelian group,
- the multiplication satisfies the associativity axiom,
- (distributive axiom) $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$ for all a, b, $c \in R$.

A **multiplicative identity element**, or just **identity element**, is an element $1 \in R$ satisfying $1a = a1 = a$ for all $a \in R$; it is unique if it exists.

The axioms imply: $0a = a0 = 0$ (in any ring); $(-1)a = a(-1) = -a$ (in any ring with identity element).

A ring is **commutative** if multiplication is commutative ($ba = ab$ for all a, $b \in R$).

Given a ring R (commutative or not), a **subring** of R is a subset $Q \subset R$ such that $(Q, +)$ is a subgroup of $(R, +)$ and Q is closed under multiplication.

However if R has identity element $1 \in R$ then we may view R as a ring with identity element. In this context, the definition of subring $Q \subset R$ of the ring with identity element R is different: in addition to requiring that $(Q, +)$ be a subgroup of $(R, +)$ and Q be closed under multiplication we also require $1 \in Q$.

The definition of ring homomorphism also comes in two variants, one for rings (map that is compatible with addition and multiplication) and for rings with identity element (map that is compatible with addition and multiplication and which sends the identity element of the source ring to the identity element of the target ring). To avoid confusion we make the convention that we always work with rings with identity element.

Convention. All rings are rings with identity element. Subrings are required to contain the identity element. Ring homomorphims are homomorphisms of rings with identity element.

The image of a ring homomorphism $f\colon R \to R'$ is a subring of R'. A **ring isomorphism** is a bijective ring homomorphism.

A **unit** in a ring R is an element $a \in R$ that possesses a two-sided multiplicative inverse; this means that there exists $a' \in R$ satisfying $a'a = aa' = 1$. In this case the element a' is unique, and is denoted by a^{-1}. We let $R^\times$ denote the set of units of a ring R and remark that $R^\times$ with composition given by multiplication is a group. We have $0 \in R^\times$ if and only if $1 = 0$ if and only if $R = \{0\}$, a situation that may be described by saying that R is the zero ring.

If R is commutative and satisfies $R^\times = R \smallsetminus \{0\}$ then R is called a **field**. Let K be a field. A subring of K which is itself a field is called a **subfield**. If L is another field then every ring homomorphism $K \to L$ is injective. For this reason, when the source and target rings are fields, we sometimes use the term **embedding** as a synonym for homomorphism.

An **integral domain** is a commutative ring which is not the zero ring, and which satisfies $ab \neq 0$ for every a, $b \in R$ with $a \neq 0$ and $b \neq 0$. In other words, $1 \neq 0$, and $ab = 0$ implies $a = 0$ or $b = 0$. Every field is an integral domain.

If R is an integral domain then the **field of fractions** or **quotient field** of R is

the set $K := R \times (R \setminus \{0\})/\sim$ where $\sim$ is the equivalence relation

$$(a_1, b_1) \sim (a_2, b_2) :\Leftrightarrow a_1 b_2 = a_2 b_1 \tag{1}$$

with the addition and multiplication defined by

$$[(a_1, b_1)] + [(a_2, b_2)] := [(a_1 b_2 + a_2 b_1, b_1 b_2)], \tag{2}$$

$$[(a_1, b_1)][(a_2, b_2)] := [(a_1 a_2, b_1 b_2)]. \tag{3}$$

The map $R \to K$ given by

$$r \mapsto [(r, 1)] \tag{4}$$

is an injective ring homomorphism. *Any ring which admits an injective ring homomorphism to a field is an integral domain.*

Toward the end of Linear Algebra we began to use universal properties to define mathematical objects not by means of elements, but by means of properties that characterize an object uniquely up to a unique isomorphism. If R is an integral domain, then we say that a field K with injective ring homomorphism $R \hookrightarrow K$ satisfies the **universal property of the quotient field**, if for every field L with injective ring homomorphism $f\colon R \hookrightarrow L$ there exists a unique embedding of fields $K \to L$ whose composition with $R \hookrightarrow K$ is f. *The field K defined in* (1)–(3)*, with the homomorphism* (4)*, satisfies the universal property of the quotient field.*

Let R be an integral domain. An element $a \in R$ is **irreducible** if a is a nonzero nonunit which has no nontrivial factorization in R (i.e., if $a = bb'$ with b, $b' \in R$ then $b \in R^\times$ or $b' \in R^\times$). An element $a \in R$ is **prime** if a is a nonzero nonunit with the property, that for all b, $b' \in R$ such that $a \mid bb'$, we have $a \mid b$ or $a \mid b'$. Every prime element is irreducible.

A **Euclidean domain** is an integral domain R such that there exists a map $f\colon R \setminus \{0\} \to \mathbb{N}$ with the following property: for all a, $b \in R$ with $b \neq 0$ there exist q and $r \in R$ such that $a = qb + r$ and either $r = 0$ or $f(r) < f(b)$. Such f is called **Euclidean function**. In a Euclidean domain R, there is a **gcd** of any pair of elements a, $b \in R$: there exists $d \in R$ with $d \mid a$ and $d \mid b$ such that for every $e \in R$ with $e \mid a$ and $e \mid b$ we have $e \mid d$ (and hence d uniquely determined up to multiplication with a unit). In fact, $d = sa + tb$ for some s, $t \in R$. If R is a Euclidean domain and $a \in R$ is irreducible, then for every $b \in R$ with $a \nmid b$ there exists $c \in R$ with $a \mid bc - 1$. In a Euclidean domain, every irreducible element is prime.

A Euclidean domain R enjoys the **unique factorization** property, which may be conveniently formulated as follows. Let P be a subset of the set of irreducible elements of R that contains exactly one element of the set $\{ua \mid u \in R^\times\}$ for every irreducible $a \in R$. Then, for $c \in R \setminus \{0\}$ there exist a unique unit $u \in R^\times$ and a unique function $v\colon P \to \mathbb{N}$ with $\{p \in P \mid v(p) \neq 0\}$ finite, such that

$$c = u \prod_{\substack{p \in P \\ v(p) \neq 0}} p^{v(p)}. \tag{5}$$

The main examples of Euclidean domains are the ring of integers $\mathbb{Z}$ and the ring of polynomials in one variable $K[X]$ over any field K. In these rings we take, by convention, P to be the set of prime numbers $\{2, 3, 5, \dots\}$, respectively the set of monic irreducible polynomials, in (5). The roots of any polynomial $f \in K[X]$ are precisely the $\alpha \in K$ such that $X - \alpha$ divides f. A polynomial of degree n over K can have at most n roots. In particular, the number of nth roots of unity in K is at most n. *The only finite subgroups of $K^\times$ are the groups of nth roots of unity in K for various positive integers n, and they are all cyclic.*

0.3 Modules, vector spaces

Let R be a commutative ring. A **module** over R, or **R-module**, is a set M with additive composition $M \times M \to M$, $(m, n) \mapsto m + n$ and scalar multplication $R \times M \to M$, $(a, m) \mapsto am$, satisfying:

- $(M, +)$ is an abelian group,
- scalar multiplication is associative: $(ab)m = a(bm)$ for all $a, b \in R$ and $m \in M$,
- (distributive axioms) $(a + b)m = am + bm$ and $a(m + n) = am + an$ for all a, $b \in R$ and m, $n \in M$,
- (compatibility with identity element) $1m = m$ for all $m \in M$.

A **vector space** is a module over a field. (So, if K is a field, there is no difference between K-module and K-vector space; the convention is to speak of vector spaces.)

In any R-module M we have $0m = 0$ and $(-1)m = -m$ for all $m \in M$.

A **submodule** of M is a subset $L \subset M$ such that $(L, +)$ is a subgroup of $(M, +)$ and L is closed under scalar multiplication (i.e., satisfies $a\ell \in L$ for all $a \in R$ and $\ell \in L$). In the case of a vector space over a field, these are called **subspaces**.

Let M and M' be R-modules. An **R-module homomorphism** from M to M' is a map $f\colon M \to M'$ such that f is a group homomorphism and $f(am) = af(m)$ for all $a \in R$ and $m \in M$. We may also describe f as **R-linear**. The kernel is a submodule of M, and the image is a submodule of M'. An **R-module isomorphism** is a bijective R-module homomorphism.

Given a submodule $L \subset M$ there is the **quotient module** M/L, whose elements are equivalence classes of M under the equivalence relation $\sim_L$ where $m \sim_L m' :\Leftrightarrow m' - m \in L$. The equivalence class $m + L$ of m is often denoted by $\overline{m}$; addition and scalar multiplication are defined by $\overline{m} + \overline{n} := \overline{m + n}$ and $a\overline{m} := \overline{am}$. The map $m \mapsto \overline{m}$ is a homomorphism, called **canonical homomorphism**. In the case of a vector space over a field, the quotient by a subspace is called **quotient space**, and we have the **canonical linear map** from the vector space to the quotient space.

The **cokernel** of an R-module homomorphism $f\colon M \to M'$ is the quotient module of M' by the image of f.

Let I be a set and let M_i be an R-module for every $i \in I$. Then there are the **direct sum** and **direct product** modules

$$\bigoplus_{i\in I} M_i \subset \prod_{i\in I} M_i.$$

An element of the direct product is an I-tuple $(m_i)_{i\in I}$ with $m_i \in M_i$ for every $i \in I$; addition and scalar multiplication are defined componentwise (so, $(m_i)_{i\in I} + (n_i)_{i\in I} := (m_i + n_i)_{i\in I}$ and $a(m_i)_{i\in I} := (am_i)_{i\in I}$). The direct sum is the submodule of the direct product, consisting of elements $(m_i)_{i\in I}$ such that $\{i \in I \,|\, m_i \neq 0\}$ is finite. If I is finite then the direct sum and direct product are equal. For instance, suppose $n \in \mathbb{N}$ and $I = \{1, \ldots, n\}$. Then the elements are n-tuples $(m_1, \ldots, m_n)$ with $m_i \in M_i$ for $i = 1, \ldots, n$, and $\bigoplus_{i=1}^n M_i$ may also be written as $M_1 \oplus \cdots \oplus M_n$. We commonly write $\bigoplus_{i=1}^n M$ as M^n.

We may view R as a module over itself. Then for $n \in \mathbb{N}$ there is the R-module R^n, and for any set I, the R-module $\bigoplus_{i\in I} R$.

Given an indexed collection $(x_i)_{i\in I}$ of elements of M, we say that $(x_i)_{i\in I}$ **generates** M if every element of M is a finite R-linear combination of the m_i; in symbols, for every $m \in M$ there should exist $r \in \mathbb{N}$, $i_1, \ldots, i_r \in I$, and $a_1, \ldots, a_r \in R$ such that $m = a_1 x_{i_1} + \cdots + a_r x_{i_r}$. We say that $(x_i)_{i\in I}$ is **linearly independent** if there are no nontrivial R-linear relations among the x_i. This means, if $i_1, \ldots, i_r \in I$ are pairwise distinct and $a_1, \ldots, a_r \in R$ are such that $a_1 x_{i_1} + \cdots + a_r x_{i_r} = 0$, then $a_1 = \cdots = a_r = 0$. We call $(x_i)_{i\in I}$ a **basis** of M if $(x_i)_{i\in I}$ generates M and is linearly independent.

To a collection $(x_i)_{i\in I}$ of elements of M there is an associated homomorphism

$$\bigoplus_{i\in I} R \to M, \tag{1}$$

defined by $(a_i)_{i\in I} \mapsto \sum_{i\in I,\, a_i\neq 0} a_i x_i$. *The conditions for $(x_i)_{i\in I}$ to generate M, respectively to be linearly independent, respectively to be a basis of M, can be expressed in terms of equivalent conditions on the homomorphism* (1)*, as follows:*

$$\begin{aligned}
(x_i)_{i\in I} \textit{ generates } M &\quad\Leftrightarrow\quad \textit{the homomorphism } (1) \textit{ is surjective,}\\
(x_i)_{i\in I} \textit{ is linearly independent} &\quad\Leftrightarrow\quad \textit{the homomorphism } (1) \textit{ is injective,}\\
(x_i)_{i\in I} \textit{ is a basis of } M &\quad\Leftrightarrow\quad \textit{the homomorphism } (1) \textit{ is bijective.}
\end{aligned}$$

An R-module is **finitely generated** if there exists a finite collection of generators. An R-module is **free** if there exists a basis. If R is not the zero ring and M is a finitely generated free module, then the cardinality of any two bases of M are

equal; the **rank** of a finitely generated free module M over a nonzero commutative ring is the cardinality of any basis. For any set I the R-module $\bigoplus_{i\in I} R$ is free; it has the **standard basis** $(e_j)_{j\in I}$, where $e_j \in \bigoplus_{i\in I} R$ denotes the I-tuple with jth component 1 and ith component 0 for every $i \neq j$.

Every vector space over a field possesses a basis. Given any pair of bases, one is finite if and only if the other is finite. We call a vector space with a finite basis **finite dimensional**, and the cardinality of any basis its **dimension**. (In this case, dimension is the same as rank, but the convention is to speak of dimension.) A vector space with no finite basis is **infinite dimensional**.

General results about modules over a commutative ring R include the following:
(First Isomorphism Theorem [Homomorphiesatz]) If $\varphi\colon L \to M$ is a surjective homomorphism of R-modules with kernel K then the induced map $L/K \to M$, $\overline{\ell} \mapsto \varphi(\ell)$, is an isomorphism of R-modules;
(Second Isomorphism Theorem [1. Isomorphiesatz]) If N is an R-module with submodules L and M, then $M/(M \cap L) \cong (M + L)/L$, by $\overline{m} \mapsto \overline{m}$ for $m \in M$;
(Third Isomorphism Theorem [2. Isomorphiesatz]) If $L \subset M \subset N$ are R-modules then $(N/L)/(M/L) \cong N/M$ by $\overline{n} + M/L \mapsto \overline{n}$, for $n \in N$.

0.4 English/German terminology

Most vocabulary is similar (Gruppe, etc.); here we indicate some differences:

Euclidean domain: *euklidischer Ring* Euclidean function: *euklidischer Betrag*
gcd: *ggT* identity element: *neutrales Element* monic polynomial: *normiertes Polynom*

1 Rings

In this section we are concerned with properties of commutative rings. (We recall, by convention, *all* rings in this script have an identity element.) Properties concerning factorization in integral domains will be the first topic of study. Building on the unique factorization result for Euclidean rings (§0.2), such as $\mathbb{Z}$ and $K[X]$ where K is any field, we will obtain an equally satisfactory factorization result for the ring $\mathbb{Z}[X]$, and more generally for multivariate polynomial rings $\mathbb{Z}[X_1, \ldots, X_n]$ and $K[X_1, \ldots, X_n]$.

The approach will be different from that taken to deal with Euclidean rings. Special emphasis will be placed on ideals. We recall that a commutative ring R may be viewed as a module over itself (§0.3). Then the submodules are called ideals.

Definition. Let R be a commutative ring. An **ideal** of R is a subset $I \subset R$ which is a submodule, where we view R as a module over itself.

Looking back at the definition of submodule, this translates into the following two requirements:

- $(I, +) \subset (R, +)$ is a subgroup.
- For all $a \in R$ and $x \in I$ we have $ax \in I$.

As examples, there is the **zero ideal** 0, and at the other extreme, the **unit ideal** R. As for general modules, we say an ideal is finitely generated if there exists a finite collection of generators. Another main goal of this section will be to show that for important rings such as $\mathbb{Z}[X_1, \ldots, X_n]$ and $K[X_1, \ldots, X_n]$, *every* ideal is finitely generated.

Ideals are significant because if R is a commutative ring, then the kernel of any ring homomorhism $R \to S$ is an ideal in R. Given an ideal I of a commutative ring R, the quotient R/I inherits a ring structure from that of R, and there is the canonical homomorphism $R \to R/I$, $a \mapsto \overline{a}$, with kernel equal to I.

Ideals which are generated by a single element are important.

Definition. Let R be a commutative ring and $I \subset R$ an ideal. We say I is a **principal ideal** if there exists an element $a \in R$ such that $I = aR$. We call R a **principal ideal domain** (PID for short) if R is an integral domain, in which every ideal is principal.

When clear from context, we may write the ideal aR as (a). So (1) denotes the unit ideal. A finitely generated ideal is one of the form $a_1R + \cdots + a_nR$ for some $n \in \mathbb{N}$ and $a_1, \ldots, a_n \in R$; this may also be written as $(a_1, \ldots, a_n)$ when there is no danger of confusion with other uses of this notation.

1.1 Factorization

Let R be an integral domain. The starting point for questions about factorization in R is the notion of irreducible element, mentioned in §0.2. We may ask, do factorizations into irreducible elements in R always exist? If so, how unique are they?

Another notion reviewed in §0.2 was prime element. There is a notion of prime ideal, with $0 \neq a \in R$ prime if and only if (a) is prime. An ideal I in a commutative ring R is **prime** if $I \subsetneq R$ and for all b, $b' \in R$ with $bb' \in I$, we have $b \in I$ or $b' \in I$. We say I is **maximal** if $I \subsetneq R$ and there exists no ideal J with $I \subsetneq J \subsetneq R$; it is a general fact that every maximal ideal is prime (if I is a maximal ideal and $bb' \in I$ with $b \notin I$, then there exist $c \in R$ and $d \in I$ with $cb + d = 1$, so $b' = cbb' + b'd \in I$).

For an ideal $I \subset R$:

$$\begin{array}{rcl} I \text{ is prime} & \Leftrightarrow & R/I \text{ is an integral domain;} \\ I \text{ is maximal} & \Leftrightarrow & R/I \text{ is a field.} \end{array}$$

Proposition 1.1. *Let R be an integral domain. Consider the following three statements, concerning a nonzero nonunit $a \in R$:*

(i) *For every $b \in R$ with $a \nmid b$ there exists $c \in R$ with $a \mid cb - 1$.*

(ii) *The element a is prime.*

(iii) *The element a is irreducible.*

In general, the implications (i) $\Rightarrow$ (ii) *and* (ii) $\Rightarrow$ (iii) *are valid. If R is a PID, then statements* (i), (ii), *and* (iii) *are equivalent.*

Proof. Since (i) says that aR is a maximal ideal and every maximal ideal is prime, we have (i) $\Rightarrow$ (ii). The implication (ii) $\Rightarrow$ (iii) is a known fact.

If R is a PID, then for $b \in R$ we have the equality of ideals $(a, b) = (d)$ for some $d \in R$, i.e., we have $d \mid a$ and $d \mid b$, with $d = sa + tb$ for some s, $t \in R$. If a is irreducible then d is either a unit or equal to a unit times a. If, further, we have $a \nmid b$, then d cannot be a unit times a. Then d is a unit, and now $-d^{-1}sa = d^{-1}tb - 1$. So, under the assumption that R is a PID, we have established (iii) $\Rightarrow$ (i). $\square$

Proposition 1.2. *Let R be an integral domain, in which there is no infinitely strictly increasing chain of principal ideals. Then every nonzero nonunit of R may be expressed as a finite product of irreducible elements of R.*

The hypothesis is that there is no strictly increasing chain

$$(a_1) \subsetneq (a_2) \subsetneq \cdots \subsetneq (a_n) \subsetneq \cdots \tag{1}$$

of principal ideals. If there does exist a strictly increasing chain (1), then possibly the first ideal (a_1) can be the zero ideal, but no others. No ideal in the chain can be the unit ideal, for we have to have $(a_n) \subsetneq (a_{n+1})$ for every n. The union $I := \bigcup_{n=1}^{\infty}(a_n)$ of all the ideals in the chain is an ideal, but not the unit ideal (if $1 \in I$ then $1 \in (a_n)$ for some n, but we have just remarked, no ideal in the chain can be the unit ideal), and not a principal ideal or even a finitely generated ideal (if $I = (b_1, \ldots, b_m)$ for some $b_1, \ldots, b_m \in R$ then for each i we have $b_i \in (a_{n_i})$ for some n_i, and taking $n := \max(n_1, \ldots, n_m)$ we have a contradiction to $(a_n) \subsetneq (a_{n+1})$).

In the proof of Proposition 1.2 we use some terminology that we will also use later. Let $x \in R$. A **divisor** of x is an element $a \in R$ that satisfies $a \mid x$. A **proper divisor** is a divisor of x which is not a unit, nor equal to x times a unit.

Lemma 1.3. *Assume that R is as in the statement of Proposition 1.2. Then every nonzero nonunit of R possesses an irreducible divisor.*

Proof. Let $a \in R$ be a nonzero nonunit. The result is clear if a is itself irreducible, so let us assume the contrary. Let Y be the set of nonzero nonunits of R, and let us call a function $v\colon Y \to \mathbb{N}$ a *finite multiplicity function* on Y if $\{y \in Y \mid v(y) \neq 0\}$ is finite. There is a natural partial ordering with $v \preceq v'$ if $v(y) \leq v'(y)$ for all $y \in Y$. We define $\mathscr{S}$ to be the partially ordered set of all finite multiplicity functions v on Y such that $\prod_{y\in Y,\, v(y)\neq 0} y^{v(y)}$ is a proper divisor of a. For instance, there exists a factorization $a = bb'$ with b, $b' \in Y$, and then $(y \mapsto \delta_{yb}) \in \mathscr{S}$.

If C is a totally ordered subset of $\mathscr{S}$ then C must be finite. If not, there would exist a sequence $y_1, y_2, \ldots$ of nonzero nonunits of R, such that $\prod_{i=1}^{n} y_i \mid a$ for every $n \in \mathbb{N}$, and we write $a = (\prod_{i=1}^{n} y_i)a_n$ with $a_n \in R$, for every $n \in \mathbb{N}$, to obtain a strictly increasing chain (1), in violation of the hypothesis.

By Zorn's lemma, there exists a maximal element $v \in \mathscr{S}$. We have the proper divisor $\prod_{y\in Y,\, v(y)\neq 0} y^{v(y)}$ of a, i.e., $a = (\prod_{y\in Y,\, v(y)\neq 0} y^{v(y)})y'$ for some nonunit $y' \in Y$. Now we claim that y' is irreducible. If we could write $y' = bb'$ with b, $b' \in Y$, then we would have $(y \mapsto v(y) + \delta_{yb}) \in \mathscr{S}$, in violation of the maximality of $v \in \mathscr{S}$. □

Proof of Proposition 1.2. Let $a \in R$ be a nonzero nonunit. The assertion is clear if a is irreducible, so we suppose the contrary. Now we define Y to be the set of *irreducible* elements of R, and otherwise we copy the definition of the set $\mathscr{S}$ from the proof of Lemma 1.3. By Lemma 1.3 there exists an irreducible divisor b of a, and then $(y \mapsto \delta_{yb}) \in \mathscr{S}$.

The argument that every totally ordered subset of $\mathscr{S}$ is finite may be copied word-for-word from the proof of Lemma 1.3. So we may apply Zorn's lemma and conclude that there exists a maximal element $v \in \mathscr{S}$. We have the proper divisor $\prod_{y\in Y,\, v(y)\neq 0} y^{v(y)}$ of a, i.e.,

$$a = \Big(\prod_{y\in Y,\, v(y)\neq 0} y^{v(y)} \Big) y' \tag{2}$$

for some nonunit $y' \in R$. As in the proof of Lemma 1.3, but taking b to be irreducible and b' to be a nonunit in the supposed factorization $y' = bb'$ (which we may do by Lemma 1.3), we see that y' is irreducible. Now (2) is an expression of a as a finite product of irreducible elements of R. □

Definition. A **UFD** is an integral domain R in which every nonzero nonunit $a \in R$ admits a factorization into irreducible elements

$$a = y_1 y_2 \cdots y_n, \tag{3}$$

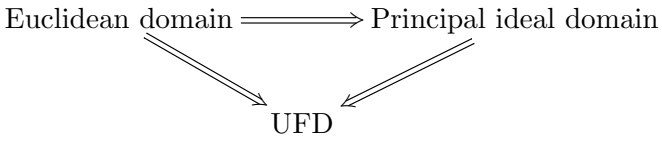
such that if

$$a = z_1 z_2 \cdots z_m \tag{4}$$

is another factorization into irreducible elements, then $m = n$ and there exist a permutation $\sigma \in S_n$ and units $u_1, \ldots, u_n \in R^\times$, such that $u_i y_i = z_{\sigma(i)}$ for all i.

UFD stands for *unique factorization domain*; in German: *faktorieller Ring*. An equivalent formulation of the definition is the one stated in §0.2 involving a subset P of the set of irreducible elements of R, such that for every irreducible $a \in R$ the set P contains exactly one element of the set $\{ua \,|\, u \in R^\times\}$.

Consider the diagram of implications:



(5)

The left-hand implication in (5) was established in the Linear Algebra lecture.

Now we have seen that the top implication in (5) is valid. (This was discussed in Linear Algebra and is reviewed in the Exercises.) The right-hand implication is stated as Proposition 1.4. So we have a strengthening of the result from the Linear Algebra lecture.

Proposition 1.4. *Every PID is a UFD.*

Proof. In a PID there can be no infinitely strictly increasing chain of principal ideals. So by Proposition 1.2, if R is a PID then every nonzero nonunit $a \in R$ admits a factorization into irreducible elements (3). We may suppose that the number of factors n is as small as possible and prove the result by induction on n, the base case $n = 1$ being obvious.

By Proposition 1.1, every irreducible element is prime. So if we have another factorization into irreducibles (4) then we have $y_1 \mid z_i$ for some i, which implies $z_i = u_1 y_1$ for some unit u_1. By the induction hypothesis, from

$$u_1^{-1} y_2 \cdots y_n = \prod_{\substack{j=1 \\ j \neq i}}^{m} z_j$$

it follows that $m = n$ and there exist units $u_2, \ldots, u_n$ and a bijection $\sigma\colon \{2, \ldots, n\} \to \{1, \ldots, i-1, i+1, \ldots, n\}$ such that $u_j y_j = z_{\sigma(j)}$ for $j = 2, \ldots, n$. Extending σ by $\sigma(1) := i$, we have $u_1 y_1 = z_{\sigma(1)}$ as well, with $\sigma \in S_n$ as desired. □

Proposition 1.5. *In a UFD, every irreducible element is prime.*

Proof. Let R be a UFD and $a \in R$ an irreducible element. Suppose that $a \mid bb'$. If b is a unit then $a \mid b'$, while if b' is a unit then $a \mid b$. Suppose that b and b' are nonzero nonunits. There exists $a' \in R$ with $aa' = bb'$, and a' is as well a nonunit. Now we

consider factorizations into irreducible elements

$$\begin{aligned} a' &= y_1' y_2' \cdots y_n', \\ b &= z_1 z_2 \cdots z_\ell, \\ b' &= z_1' z_2' \cdots z_m'. \end{aligned}$$

Then

$$a y_1' y_2' \cdots y_n' = z_1 z_2 \cdots z_\ell z_1' z_2' \cdots z_m'$$

Since R is a UFD, for some unit u we have

$$ua \in \{z_1, \ldots, z_\ell\} \cup \{z_1', \ldots, z_m'\},$$

and it follows that $a \mid b$ or $a \mid b'$. □

Example. Let $R := \mathbb{Z}[\sqrt{10}]$. The element 2 is irreducible: $2 = (a+b\sqrt{10})(c+d\sqrt{10})$ with a, b, c, $d \in \mathbb{Z}$ implies $2 = (a - b\sqrt{10})(c - d\sqrt{10})$, from which we have

$$4 = (a^2 - 10b^2)(c^2 - 10d^2). \tag{6}$$

Each factor on the right-hand side of (6) is excluded from the congruence classes ± 2 modulo 16 (since the squares are 0, 1, 4, 9 modulo 16), so (6) implies $a^2 - 10b^2 = \pm 1$ or $c^2 - 10d^2 = \pm 1$, hence $a + b\sqrt{10}$ or $c + d\sqrt{10}$ is a unit. But 2 is not prime in R, as we have $2 \mid 10 = \sqrt{10} \cdot \sqrt{10}$ but $2 \nmid \sqrt{10}$. Since we have an irreducible element in R which is not prime, Proposition 1.5 tells us that R is not a UFD.

Definition. Let R be an integral domain. A polynomial $f \in R[T]$ is called **primitive** if f is nonzero and the coefficients of f have no common nonunit divisor in R.

Lemma 1.6. *Let R be a UFD with field of fractions K, and let $f \in K[T]$ be a nonzero polynomial. Then there exists $c \in K^\times$, uniquely determined up to multiplication with a unit of R, such that $c^{-1} f$ has coefficients in R and is primitive in $R[T]$. Furthermore, if f has coefficients in R, then we have $c \in R$.*

Proof. Let P be a subset of the set of irreducible elements of R, such that for every irreducible $a \in R$ the set P contains exactly one element of the set $\{ua \mid u \in R^\times\}$. Let $b_i T^{m_i}$ for $i = 1, \ldots, n$ be the nonzero terms of f with $b_i \in K^\times$ and $m_i \in \mathbb{N}$, $m_i \neq m_j$ for $i \neq j$:

$$f = \sum_{i=1}^{n} b_i T^{m_i}.$$

Let us write

$$b_i = u_i \prod_{\substack{p \in P \\ v_i(p) \neq 0}} p^{v_i(p)}$$

for every i, which by the UFD property we can do for unique unit $u_i \in R^\times$ and finite multiplicity function $v_i \colon P \to \mathbb{Z}$.

Define $v_{\min} \colon P \to \mathbb{Z}$,

$$v_{\min}(p) := \min(v_1(p), \ldots, v_n(p)).$$

Now let c be a nonzero element of K, written in its unique factorization as

$$c = u' \prod_{\substack{p \in P \\ v_{\min}(p) - v'(p) \neq 0}} p^{v_{\min}(p) - v'(p)}$$

with $u' \in R^\times$ and $v' \colon P \to \mathbb{Z}$ a finite multiplicity function. Now $c^{-1} f = \sum_{i=1}^n b_i' T^{m_i}$, with

$$b_i' = u_i u'^{-1} \prod_{\substack{p \in P \\ v_i(p) - v_{\min}(p) + v'(p) \neq 0}} p^{v_i(p) - v_{\min}(p) + v'(p)}. \tag{7}$$

We claim:

(i) We have $c^{-1} f \in R[T]$ if and only if v' takes all its values in $\mathbb{N}$.

(ii) Suppose that v' takes all its values in $\mathbb{N}$. Then $c^{-1} f$ is primitive if and only if v' is the zero function.

For (i), if v' takes all its values in $\mathbb{N}$ then the exponents in (7) are all in $\mathbb{N}$. However, if there is $p \in P$ such that $v'(p) < 0$, then we can consider i such that $v_{\min}(p) = v_i(p)$, and then there is a negative exponent in the expression (7) for b_i'. For (ii), we write a nonzero nonunit $d \in R$ as

$$d = z \prod_{\substack{p \in P \\ w(p) \neq 0}} p^{w(p)}$$

with $z \in R^\times$ and $w \colon P \to \mathbb{N}$ a finite multiplicity function. Then d is a common divisor of $b_1', \ldots, b_n'$ if and only if $v_i(p) - v_{\min}(p) + v'(p) \geq w(p)$ for all $p \in P$. If v' is the zero function, this is impossible, while if v' is not the zero function then any d such that $w(p) = v'(p)$ for all p is a common divisor of $b_1', \ldots, b_n'$.

Finally, if f has coefficients in R then the v_i all take their values in $\mathbb{N}$, hence so does $v_{\min}$. □

Lemma 1.7 (Gauss's lemma)**.** *The product of two primitive polynomials over a UFD is again a primitive polynomial.*

For the proof, we will use the compatibility of two operations: passage to the quotient ring (by an ideal) and extension to a polynomial ring. Let R be a commutative ring and I an ideal. The the polynomial ring $R[T]$ contains R as a subring,

and this way I generates an ideal which we may denote by $IR[T]$. The quotient ring R/I is again a commutative ring, so we may consider as well the polynomial ring $(R/I)[T]$. Now we have a ring isomorphism

$$R[T]/IR[T] \cong (R/I)[T], \tag{8}$$

where for $f \in R[T]$, $f = \sum_{i=0}^{n} a_i T^i$ (with $a_i \in R$ for all i), the class $\overline{f}$ on the left is identified with $\sum_{i=0}^{n} \overline{a}_i T^i$ on the right.

To a general homomorphism of commutative rings $\varphi\colon R \to S$ there is a homomorphism $R[T] \to S[T]$, given by $\sum_{i=0}^{n} a_i T^i \mapsto \sum_{i=0}^{n} \varphi(a_i) T^i$, which is injective (respectively surjective) when φ is injective (respectively surjective). Applied to the canonical homomorphism $R \to R/I$, which is surjective, we get a surjective ring homomorphism $R[T] \to (R/I)[T]$. We have $IR[T] = \{\sum_{i=0}^{n} a_i T^i \in R[T] \mid a_i \in I \text{ for all } i\}$. We get the isomorphism (8) by combining these facts and applying the First Isomorphism Theorem.

Proof of Gauss's lemma. Let R be a UFD, and let f and g be primitive polynomials in $R[T]$. If the coefficients of fg have common divisor $a \in R \setminus R^\times$, then for any irreducible divisor y of a, the coefficients of fg also have y as common divisor. So it suffices to show that if $y \in R$ is irreducible, then y is not a common divisor of the coefficients of fg.

For a polynomial in $R[T]$, having coefficients with y as common divisor is equivalent to belonging to the kernel of the homomorphism

$$R[T] \to (R/yR)[T]. \tag{9}$$

By Proposition 1.5, y is prime. This means that R/yR is an integral domain, hence $(R/yR)[T]$ is also an integral domain. Since f and g are primitive, their images under (9) are nonzero. So the same holds for fg. □

Proposition 1.8. *Let R be a UFD with field of fractions K. A primitive polynomial $f \in R[T]$ of positive degree is irreducible in $R[T]$ if and only if f is irreducible in $K[T]$.*

Proof. Suppose that $f = gh$ in $R[T]$ with nonunits g and h. If g or h would have degree 0 then we would have a contradiction to the assumption that f is primitive. So g and h have positive degree. Then $f = gh$ is a nontrivial factorization in $K[T]$ as well.

Suppose that $f = gh$ for some $g, h \in K[T]$, with g and h of positive degree. Now we apply Lemma 1.6 to the polynomials g and h:

$$\begin{aligned} g &= bG && \text{with } b \in K^\times,\ G \in R[T],\ G \text{ primitive,} \\ h &= cH && \text{with } c \in K^\times,\ H \in R[T],\ H \text{ primitive.} \end{aligned}$$

By Gauss's lemma, GH is primitive, and from $f = gh = bcGH$ we have $bc \in R^\times$. Now

$$f = (bcG)H$$

is a nontrivial factorization of f in $R[T]$. □

Proposition 1.9. *Let R be a UFD with field of fractions K. Then the set of irreducible elements of $R[T]$ is the disjoint union of the polynomials of degree 0 which are irreducible as elements of R and the primitive polynomials of positive degree which are irreducible in $K[T]$.*

Proof. Let $f \in R[T]$ be a nonzero nonunit. If f has degree 0, then any factorization $f = gh$ in $R[T]$ must have g and h of degree 0. So for f of degree 0, irreducibility in $R[T]$ is equivalent to irreducibility in R.

If f has positive degree, but is not primitive, then there is a common divisor $a \in R \smallsetminus R^\times$ of the coefficients of f, which lets us write $f = a\tilde{f}$ for some $\tilde{f} \in R[T]$. This is a nontrivial factorization in $R[T]$.

The case f is primitive and of positive degree is treated by Proposition 1.8. □

Theorem 1.10. *Let R be a UFD. Then $R[T]$ is a UFD.*

Proof. Let $f \in R[T]$ be a nonzero nonunit. If f has degree 0, then we know that factorizations of f in $R[T]$ are the same as factorizations of f in R. Since R is a UFD, we have existence and uniqueness (up to multiplication by units and reordering of the factors) of factorizations into irreducible elements in this case.

Let K denote the fraction field of R. If f has positive degree and is primitive, then we apply unique factorization in the Euclidean domain $K[T]$:

$$f = h_1 \cdots h_n$$

with $h_1, \ldots, h_n \in K[T]$ irreducible polynomials. Let us write $h_i = c_i H_i$ with $c_i \in K^\times$ and $H_i \in R[T]$ primitive (Lemma 1.6). Then we have $f = c_1 \cdots c_n H_1 \cdots H_n$. By Gauss's lemma, $H_1 \cdots H_n$ is primitive. Since f is primitive, we have $c_1 \cdots c_n \in R^\times$. So we have $f = (c_1 \cdots c_n H_1) H_2 \cdots H_n$, an expression of f as a product of primitive polynomials of positive degree which are irreducible in $K[T]$. If we have two such expressions, then the uniqueness of factorization in $K[T]$ and the uniqueness up to multiplication by a unit in Lemma 1.6 combine to yield the desired uniqueness of the factorization of f.

Suppose f has positive degree but is not primitive. Then we apply Lemma 1.6:

$$f = cg$$

with $c \in R$ and primitive $g \in R[T]$. Applying the previous cases to c and g, we obtain a factorization of f into irreducible elements of $R[T]$. Consider now an

arbitrary factorization of f into irreducible elements. By rearranging the factors so that all the factors of degree 0 comes before the factors of positive degree, we obtain (taking the respective products) a factorization $f = \tilde{c}\tilde{g}$ where $\tilde{c}$ has degree 0 and $\tilde{g}$ is primitive (by Gauss's lemma). By the uniqueness assertion of Lemma 1.6 there is a unit $u \in R^\times$ such that $\tilde{g} = ug$, and hence $\tilde{c} = u^{-1}c$. We deduce the uniqueness of the factorization of f into irreducible elements of $R[T]$, up to multiplication by units and reordering of the factors, by applying the respective uniqueness for elements of R and for primitive polynomials in $R[T]$ of positive degree. □

Example. $\mathbb{Z}[X]$ is a UFD. To find a factorization into irreducibles of a given nonzero polynomial $f \in \mathbb{Z}[X]$, we first extract the GCD of the coefficients of f, as $f = c\tilde{f}$ with $c \in \mathbb{Z}$ and $\tilde{f} \in \mathbb{Z}[X]$ primitive. Then we factor f in $\mathbb{Q}[X]$, writing each irreducible factor as a primitive polynomial in $\mathbb{Z}[X]$. The result, up to a factor of ± 1, combined with a prime factorization of c, gives a decomposition of f into irreducible factors in $\mathbb{Z}[X]$. For instance, for $f := 8X^4 + 8X^3 + 18X^2 + 16X + 4$, we have $f = 2\tilde{f}$ with primitive $\tilde{f} := 4X^4 + 4X^3 + 9X^2 + 8X + 2$. The factorization of $\tilde{f}$ in $\mathbb{Q}[X]$ would traditionally be written as $\tilde{f} = 4(X + \frac{1}{2})^2(X^2 + 2)$, but we replace each factor by its unique (up to sign) primitive multiple and find $\tilde{f} = (2X + 1)^2(X^2 + 2)$. So we have the decomposition $f = 2(2X + 1)^2(X^2 + 2)$ into irreducible factors in $\mathbb{Z}[X]$.

Example. If K is a field, then by analogy with the previous example but starting with the Euclidean domain $K[T]$ rather than $\mathbb{Z}$, we have the UFD $K[T][X]$, more commonly written $K[T, X]$, with analogous procedure for determining a factorization into irreducibles. Writing $f \in K[T, X]$ as $\sum_{i=1}^{n} f_i X^i$ for some n, with $f_i \in K[T]$, we extract a GCD of the f_i to obtain $f = c\tilde{f}$ with $c \in K[T]$ and $\tilde{f} = \sum \tilde{f}_i X^i$ such that the $\tilde{f}_i \in K[T]$ have no common factors. We proceed as above, by viewing $\tilde{f}$ as an element of $K(T)[X]$ and decomposing into irreducible factors there. Concrete example: $T^3X^4 - T^4X^2 + T^3X^3 - T^4X + T^2X^3 - T^3X = T^2X(X^2 - T)(TX + T + 1)$ in $\mathbb{Q}[X, T]$.

Example. Theorem 1.10 may be iterated. We see that $\mathbb{Z}[X_1, \ldots, X_n]$ is a UFD for every $n \in \mathbb{N}$. If K is a field, then $K[X_1, \ldots, X_n]$ is a UFD for every $n \in \mathbb{N}$.

1.2 Noetherian rings

In Section 1.1 we have seen the ascending chain condition on principal ideals, with relation to the existence of factorizations into irreducible elements (Proposition 1.2). Of great importance is the ascending chain condition for arbitrary ideals.

Definition. A commutative ring R is said to be **Noetherian** if there is no strictly increasing chain of ideals in R.

The condition is named after Emmy Noether (1882–1935). It is equivalent to the finite generation of every ideal in R.

Proposition 1.11. *Let R be a commutative ring. The following conditions are equivalent:*

(i) *R is Noetherian.*

(ii) *Every ideal in R is finitely generated.*

Proof. For the implication (ii) $\Rightarrow$ (i), given a strictly increasing chain of ideals

$$I_1 \subsetneq I_2 \subsetneq \cdots \subsetneq I_n \subsetneq \ldots \tag{1}$$

the union $I := \bigcup_{n=1}^{\infty} I_n$ is also an ideal. Given (ii), I must be finitely generated, and some I_n contains all the elements of a finite set of generators. This contradicts the strict inclusion of I_n in I_{n+1}.

The implication (i) $\Rightarrow$ (ii) makes use of the axiom of choice. Given an ideal J, we let $\varphi\colon \{\text{ideals } I \subset J\} \to J$ be a map satisfying $\varphi(J) = 0$ and $\varphi(I) \in J \smallsetminus I$ for $I \subsetneq J$. Starting with $I_0 := 0$ we obtain an increasing chain of finitely generated ideals by $I_{j+1} := I_j + (\varphi(I_j))$ for $j \in \mathbb{N}$. By (i) this is not a strictly increasing chain of ideals, so for some j we have $I_{j+1} = I_j$. Hence $I_j = J$ is finitely generated. □

Example. Every PID is Noetherian.

Theorem 1.12 (Hilbert Basis Theorem)**.** *Let R be a commutative ring. If R is Noetherian, then so is $R[T]$.*

Proof. Let $I \subset R[T]$ be an ideal. We will show that I is finitely generated.

Define J_n to be the subset of R, consisting of the leading coefficient of every polynomial in I of degree n, together with the element 0. Then J_n is an ideal. We have $J_n \subset J_{n+1}$ for all n. Since R is Noetherian, there exists n such that $J_{n'} = J_n$ for all $n' \geq n$, and each of the ideals $J_0, \ldots, J_n$ is finitely generated, say:

$$J_i = (c_{i1}, \ldots, c_{im_i}), \qquad i = 0, \ldots, n,$$

with $m_i \in \mathbb{N}$ and $f_{i1}, \ldots, f_{im_i} \in I$ of degree i with respective leading coefficients $c_{i1}, \ldots, c_{im_i}$.

Now we claim:

$$I = (f_{01}, \ldots, f_{0m_0}, \ldots, f_{n1}, \ldots, f_{nm_n}).$$

We need to show that every nonzero $f \in I$ is in the ideal generated by the f_{ij}. We do this by induction on $d := \deg f$. The case $d = 0$ is clear, since then $f \in J_0 = (c_{01}, \ldots, c_{0m_0})$, whose generators are among the claimed generators of I.

Suppose $d \geq 1$. We let $c \in R$ denote the leading coefficient of f. Then $c \in J_d = (c_{i1}, \ldots, c_{im_i})$, with $i := \min(d, n)$. So we have

$$c = a_1 c_{i1} + \cdots + a_{m_i} c_{im_i}$$

for some $a_1, \ldots, a_{m_i} \in R$. Therefore,

$$T^{d-i}(a_1 f_{i1} + \cdots + a_{m_i} f_{im_i})$$

also has degree d and leading coefficent c. So $\tilde{f} := f - T^{d-i}(a_1 f_{i1} + \cdots + a_{m_i} f_{im_i})$ is either zero or an element of I of degree less than d. By the induction hypothesis, $\tilde{f}$ belongs to the ideal with the claimed generators. Hence so does f. □

Example. For every $n \in \mathbb{N}$ the rings $\mathbb{Z}[X_1, \ldots, X_n]$ and $K[X_1, \ldots, X_n]$, for any field K, are Noetherian.

If $I \subset R$ is an ideal, then the ideals of R/I are in order-preserving bijective correspondence with the ideals of R containing I. As a consequence, if R is Noetherian, then so is R/I for any ideal I.

Example. The ring $\mathbb{Z}[X_1, \ldots, X_n]/I$ is Noetherian for every $n \in \mathbb{N}$ and ideal $I \subset \mathbb{Z}[X_1, \ldots, X_n]$. In other words, every finitely generated commutative ring is Noetherian. If K is any field, then the ring $K[X_1, \ldots, X_n]/I$ is Noetherian for every $n \in \mathbb{N}$ and ideal $I \subset K[X_1, \ldots, X_n]$, and the same holds with K replaced by any Noetherian commutative ring R.

Proposition 1.13. *Let R be a Noetherian commutative ring, and let M be a finitely generated R-module. Then every submodule of M is finitely generated.*

Proof. Since there is a surjective homomorphism of R-modules $R^n \to M$ for some $n \in \mathbb{N}$, we may reduce to the case $M = R^n$. Now we establish the result by induction on n. The base cases are $n = 0$, which is trivial, and $n = 1$, which is established by Proposition 1.11.

Suppose M' is a submodule of R^n, with $n \geq 2$. We consider the homomorphism $R^n \to R$, given by projection onto the nth factor. The kernel is R^{n-1}, identified with the submodule of R^n of elements with nth component 0. By the induction hypothesis, the image of M' in R and the submodule $M' \cap R^{n-1}$ are finitely generated. We may take S to be a finite subset of M', mapping to generators of the image of M' in R, and T to be a finite set of generators of $M' \cap R^{n-1}$; then M' is generated by $S \cup T$. □

Remark. If R is a PID then any submodule M' of a finitely generated free module M is free, with $\operatorname{rk} M' \leq \operatorname{rk} M$. This fact, which was established for Euclidean domains in the Linear Algebra lecture, only makes use of the PID property. Following the proof of Proposition 1.13, we take S empty if $M' \subset R^{n-1}$ and otherwise consisting of a single element, and T to be a basis of $M' \cap R^{n-1}$; then $S \cup T$ is a basis of M'.

1.3 Smith normal form, modules over a PID

The topics of Smith normal form and structure of finitely-generated modules were discussed in the setting of Euclidean domains in the Linear Algebra lecture. This section presents the generalization to the setting of principal ideal domains. The Smith Algorithm is replaced by an existence proof for the Smith normal form.

Lemma 1.14. *Let R be a PID and $a_1, \ldots, a_n \in R$ be elements that generate the unit ideal. Then there exists an invertible $n \times n$-matrix over R whose first row is $a_1, \ldots, a_n$.*

Proof. We prove this by induction on n. The base case $n = 1$ is obvious. Now suppose $n \geq 2$. The case $a_1 = \cdots = a_{n-1} = 0$ is obvious, so we suppose the contrary. The ideal $(a_1, \ldots, a_{n-1})$ is principal, generated by some nonzero $a' \in R$, and we write $a_j = a'b_j$ with $b_j \in R$ for $j = 1, \ldots, n-1$. So $b_1, \ldots, b_{n-1}$ generate the unit ideal, and by the induction hypothesis there exists an invertible $(n-1)\times(n-1)$-matrix $B = (b_{ij})_{1\leq i,j\leq n-1}$ over R with $b_{1j} = b_j$ for all j.

We have $(a', a_n) = (a_1, \ldots, a_n)$ equal to the unit ideal, so there exist r and $s \in R$ with $ra' + sa_n = 1$. Now we consider the $n \times n$-matrix

$$A := \begin{pmatrix} a_1 & \cdots & a_{n-1} & a_n \\ b_{21} & \cdots & b_{2,n-1} & 0 \\ \vdots & & \vdots & \vdots \\ b_{n-1,1} & \cdots & b_{n-1,n-1} & 0 \\ -sb_1 & \cdots & -sb_{n-1} & r \end{pmatrix}.$$

Evaluating the determinant by Laplace expansion along the last column, we find

$$\det(A) = a_n s \det(B) + ra' \det(B) = \det(B).$$

So A is invertible. □

Theorem 1.15 (Smith Normal Form)**.** *Let R be a PID and $A \in \mathrm{Mat}(m \times n, R)$ a matrix over R. We let r denote the rank of A, as a matrix over the fraction field of R. Then there exist $S \in GL_m(R)$ and $T \in GL_n(R)$, such that SAT has nonzero (i,i)-entry for $i = 1, \ldots, r$ with remaining entries equal to zero, and $a_i \mid a_{i+1}$ for all i, where a_i denotes the (i,i)-entry of SAT (Smith normal form). Moreover, if $S' \in GL_m(R)$ and $T' \in GL_n(R)$ are such that $S'AT'$ is in Smith normal form, with (i,i)-entry a'_i, then there exist units $u_1, \ldots, u_r$ such that $a'_i = u_i a_i$ for all i.*

Proof. We define $q = q(A)$ to be the largest natural number, such that the upper-left $q \times q$-submatrix of A is a diagonal matrix $\mathrm{diag}(a_1, \ldots, a_q)$ with nonzero diagonal entries satisfying $a_i \mid a_{i+1}$ for all i, there are no other nonzero entries in the first q

rows or columns of A, and additionally if $q > 0$ then a_q divides all the remaining matrix entries of A.

We have $q \leq r$, with equality if and only if A is in Smith normal form.

The existence portion of the theorem is proved by descending induction on q. The base case is $q = r$; then we may take S and T to be identity matrices.

If $q < r$, then we define B to be the bottom-right $(m-q)\times(n-q)$-submatrix of A. The matrix B is nonzero. We define $J := (b_{11}, \ldots, b_{1,n-q}, \ldots, b_{m-q,1}, \ldots, b_{m-q,n-q})$, the ideal generated by the entries of B.

For $U \in GL_{m-q}(R)$ and $V \in GL_{n-q}(R)$ we set $\widetilde{B} := UBV$ and consider its entries $\widetilde{B} = (\tilde{b}_{ij})_{1\leq i\leq m-q, 1\leq j\leq n-q}$. The ideal generated by the entries of $\widetilde{B}$ is equal to J (one containment followed immediately from $\widetilde{B} := UBV$, and the other containment, from $B = U^{-1}\widetilde{B}V^{-1}$). Since permutation matrices are invertible, U and V may be chosen so that $\tilde{b}_{11} \neq 0$. There is the obvious containment

$$(\tilde{b}_{11}) \subset J. \tag{1}$$

Suppose we can show, in case the containment (1) is strict, that $U' \in GL_{m-q}(R)$ and $V' \in GL_{n-q}(R)$ may be chosen to yield $\widetilde{B}' := U'\widetilde{B}V'$ with top-left entry $\tilde{b}'_{11}$ satisfying

$$(\tilde{b}_{11}) \subsetneq (\tilde{b}'_{11}). \tag{2}$$

It would follow that U and V above may be chosen to yield equality in (1). Indeed, each strict containment (2) is accompanied by a decrease in the number of factors of $\tilde{b}'_{11}$ in a factorization into irreducibles, in comparison to that of $\tilde{b}_{11}$.

Supposing strict containment (1), with $\tilde{b}_{11} \neq 0$, we divide into three cases:
Case 1: We have $\tilde{b}_{11} \nmid \tilde{b}_{1j}$ for some $j \geq 2$. Let us suppose that the ideal generated by the first row of $\widetilde{B}$ is $(\tilde{b})$, so $(\tilde{b}_{11}) \subsetneq (\tilde{b})$. Writing $\tilde{b}_{1k} = \tilde{b}c_k$ for all k, we have elements $c_1, \ldots, c_{n-q} \in R$ that generate the unit ideal. By Lemma 1.14, there exists $C \in GL_{n-q}(R)$ with first row $c_1, \ldots, c_{n-q}$. Defining $U' := E_{m-q}$ and $V' := C^{-1}$, we obtain $\tilde{b}'_{11} = \tilde{b}$ in the notation of (2).
Case 2: We have $\tilde{b}_{11} \nmid \tilde{b}_{i1}$ for some $i \geq 2$. We proceed as in Case 1.
Case 3: We have $\tilde{b}_{11} \mid \tilde{b}_{i1}$ for all i and $\tilde{b}_{11} \mid \tilde{b}_{1j}$ all j, but $\tilde{b}_{11} \nmid \tilde{b}_{ij}$ for some i, $j \geq 2$. Let us write $\tilde{b}_{i1} = \tilde{b}_{11}c$ with $c \in R$ and perform the row operations to $\widetilde{B}$ of subtacting c times the first row from the ith row and then adding the ith row to the first row. The resulting matrix has $(1,1)$-entry $\tilde{b}_{11}$ and $(1,j)$-entry not divisible by $\tilde{b}_{11}$. We copy from Case 1, replacing E_{m-q} with the product of two elementary matrices corresponding to the two row operations.

Having shown how to achieve (2), it follows that $U \in GL_{m-q}(R)$ and $V \in GL_{n-q}(R)$ may be chosen so that we have equality in (1). This means, $\tilde{b}_{11}$ divides all of the matrix entries of $\widetilde{B}$. Now we let U' (respectively V') be products of

elementary matrices, which correspond to the operations on $\widetilde{B}$ of adding the appropriate multiple of the first row (respectively column) to the other rows (respectively columns), so that $\tilde{b}_{11}$ is the only nonzero entry in the first column (respectively row). Then

$$\widetilde{B}' := U'\widetilde{B}V' = U'UBVV' \tag{3}$$

has the same top-left entry $\tilde{b}_{11}$ as $\widetilde{B}$, which is now the the only nonzero entry in the first row and first column, and still, $\tilde{b}_{11}$ divides all of the matrix entries of $\widetilde{B}'$.

Let $\widetilde{A}'$ be the matrix obtained from A by replacing the bottom-right $(m-q) \times (n-q)$-submatrix with $\widetilde{B}'$. Notice that $\widetilde{A}'$ is related to A by a matrix equation analogous to (3), where the matrices in GL_{m-q} (respectively GL_{n-q}) have been replaced by block diagonal matrices in GL_m (respectively GL_n) with top-left block E_q. We have $q(\widetilde{A}') > q(A)$, so by the induction hypothesis, there exist $\widetilde{S}' \in GL_m(R)$ and $\widetilde{T}' \in GL_n(R)$, such that $\widetilde{S}'\widetilde{A}'\widetilde{T}'$ is in Smith normal form. This establishes the existence portion of the theorem.

The uniqueness of the Smith normal form up to multiplication by units will be proved after the structure theorem for finitely generated modules over a PID. □

Lemma 1.16 (Chinese Remainder Theorem)**.** *Let R be a commutative ring, and let a and b be elements of R that generate the unit ideal. Then the homomorphism $R/(ab) \to R/(a) \oplus R/(b)$, $\bar{x} \mapsto (\bar{x}, \bar{x})$, is an isomorphism of R-modules.*

Proof. By hypothesis there exist c, $d \in R$ with $ca + db = 1$. An inverse map is given by $(\bar{u}, \bar{v}) \mapsto \overline{dbu} + \overline{cav}$. □

Theorem 1.17 (Structure Theorem for Finitely Generated Modules over a PID)**.** *Let R be a PID and M a finitely generated R-module. Then there are uniquely determined natural numbers r and s and nonzero nonunits $a_1, \ldots, a_r \in R$ satisfying $a_i \mid a_{i+1}$ for all i, each uniquely determined up to multiplication with a unit of R, such that M is isomorphic to $R/(a_1) \oplus \cdots \oplus R/(a_r) \oplus R^s$.*

As preparation for the proof, we recall the notion of **torsion module**

$$M_{\text{tors}} := \{m \in M \mid \exists\, 0 \neq a \in R :\ am = 0\},$$

as well as **p-torsion module** for $0 \neq p \in R$

$$M[p] := \{m \in M \mid pm = 0\}$$

and **p-primary torsion module**

$$M(p) := \bigcup_{n=1}^{\infty} M[p^n].$$

By Proposition 1.1, for $p \in R$ irreducible, $\boldsymbol{k}_p := R/(p)$ is a field.

Proof of Theorem 1.17. Let n be the minimal number of elements needed to generate M as R-module and $R^n \to M$ the surjective homomorphism of R-modules corresponding to some sequence of n generators. The kernel is isomorphic to R^r for some $r \in \mathbb{N}$ with $r \leq n$ (see the Remark after Proposition 1.13). We apply the existence portion of Theorem 1.15: letting A denote the $n \times r$-matrix over R corresponding to the inclusion of the kernel in R^n, for some chosen identification of the kernel with R^r, there exist $S \in GL_n(R)$ and $T \in GL_r(R)$ such that SAT is in Smith normal form, with some nonzero entries $a_1, \ldots, a_r$ satisfying $a_i \mid a_{i+1}$ for all i. It follows that M is isomorphic to $R/(a_1) \oplus \cdots \oplus R/(a_r) \oplus R^s$, where $s := n - r$. The a_i are nonunits: if we would have $a_i \in R^\times$ for some i, then by including R^{n-1} in R^n as the submodule of elements with ith component 0, we would find that the composite $R^{n-1} \to R^n \to M$ is surjective, contradicting the minimality of n.

The number s is characterized as the rank of the free module M/M_{tors}, so is uniquely determined. Since n is uniquely determined, so is $r = n - s$.

It remains to show that the elements a_i are uniquely determined up to multiplication by units. For the rest of the proof, we take P to be a subset of the set of irreducible elements of R, such that for every irreducible $a \in R$ the set P contains a unique element of the set $\{ua \mid u \in R^\times\}$. Let $p_1, \ldots, p_t$ be the divisors of a_r in P; then

$$\{p \in P \mid M(p) \neq 0\} = \{p_1, \ldots, p_t\}.$$

We may write

$$a_i = u_i p_1^{m_{i1}} \ldots p_t^{m_{it}} \tag{4}$$

uniquely, with $u_i \in R^\times$ and $m_{ij} \in \mathbb{N}$, and we have

$$m_{1j} \leq m_{2j} \leq \cdots \leq m_{rj} \tag{5}$$

for every j. By repeated application of the Chinese Remainder Theorem we have

$$M \cong \big(R/(p_1^{m_{11}}) \oplus \cdots \oplus R/(p_1^{m_{r1}})\big) \oplus \cdots \oplus \big(R/(p_t^{m_{1t}}) \oplus \cdots \oplus R/(p_t^{m_{rt}})\big) \oplus R^s,$$

and

$$M(p_j) \cong R/(p_j^{m_{1j}}) \oplus \cdots \oplus R/(p_j^{m_{rj}}).$$

It follows that for every j and ℓ,

$$\dim_{\boldsymbol{k}_{p_j}} (M[p_j^\ell]/M[p_j^{\ell-1}]) = |\{i \mid m_{ij} \geq \ell\}|. \tag{6}$$

Combining (4), (5), and (6) gives the remaining uniqueness assertion. □

Proof of the uniqueness portion of Theorem 1.15. Let M be the cokernel of A. Then M is a finitely generated R-module, such that the sequence of nonzero nonunits of the isomorphism type of M (as in Theorem 1.17) is the subsequence of nonunits of

$a_1, \ldots, a_r$. Given $S' \in GL_m(R)$ and $T' \in GL_n(R)$ such that $S'AT'$ is in Smith normal form, with corresponding $a'_1, \ldots, a'_r$, then the subsequence of nonzero nonunits of $a'_1, \ldots, a'_r$ must be equal, up to multiplication by units, with that of $a_1, \ldots, a_r$. The units among $a_1, \ldots, a_r$ and among $a'_1, \ldots, a'_r$ are equal in number. So the uniqueness up to multiplication by units is established. □

1.4 Tensor product of modules

The tensor product of modules, as for vector spaces, is characterized by a universal property. In the case of vector spaces, the existence could be established using the general fact that every vector space has a basis. For modules, we cannot rely on this, since there are modules that are not free. A general construction based on generators and relations takes care of the existence of the tensor product of modules.

The discussion of tensor product is also a reasonable place to introduce some of the *language of category theory*, as this will be useful for stating some of the general properties of the tensor product. A **category** consists of **objects**, for instance, sets with operations (maps) satisfying various axioms, together with **morphisms** between objects, which in the first examples will be homomorphisms. The definition of group and group homomorphism (§0.1) can be formalized as the statement that groups and group homomorphisms form a category, called the **category of groups**. If we make commutativity an additional axiom, then we obtain the **category of abelian groups**. But we may add a different axiom, for instance that the underlying set should be finite, and we get the category of finite groups. We get a category, as long as every object has an identity morphism, and morphisms from and to a given object can be composed, in a manner that is associative and has the property that composition on either side with an identity morphism does nothing.

Similarly (§0.2), there are **categories of rings** and **commutative rings** (by convention, here, always with identity element). Fields form another category. Given a commutative ring R, there is the **category of R-modules** (§0.3).

The morphisms between objects A and B of a category $\mathscr{C}$ are typically denoted by $\mathrm{Hom}_{\mathscr{C}}(A, B)$, or just $\mathrm{Hom}(A, B)$. Invertible morphisms are called **isomorphisms**. Left-, respectively right-cancelable morphisms are called **monomorphisms**, respectively **epimorphisms**. For R-modules, these are just injective, respectively surjective homomorphisms.

There are **functors** between categories which associate objects to objects and morphisms to morphisms, in a manner that is compatible with composition of morphisms and respects identity morphisms. A trivial example is the **identity functor** from any category to itself. Less trivial is the **forgetful functor** from the category of R-modules to the category of abelian groups, sending a R-module to the underlying abelian group; any functor that forgets certain structure may be called forgetful functor. Functors from and to a given category may be composed.

For example, we consider the forgetful functor F from $\mathbb{Z}$-modules to abelian groups. There is a functor G in the other direction, which sends an abelian group A to the same abelian group, with the canonical $\mathbb{Z}$-module structure $\mathbb{Z} \times A \to A$, determined uniquely by the module axioms. We may represent these by arrows:

$$(\text{Ab}) \underset{F}{\overset{G}{\rightleftarrows}} (\mathbb{Z}\text{-Mod}), \tag{1}$$

and we have

$$F \circ G = \text{id}_{(\text{Ab})}, \qquad G \circ F = \text{id}_{(\mathbb{Z}\text{-Mod})}. \tag{2}$$

To express the situation captured by (1)–(2), we say that the categories (Ab) of abelian groups and ($\mathbb{Z}$-Mod) of $\mathbb{Z}$-modules are isomorphic.

Proposition 1.18. *Let R be a commutative ring, and let A and B be R-modules. Then $\text{Hom}(A, B)$ has the structure of R-module, with addition of homomorphisms f, $g \in \text{Hom}(A, B)$ as $f+g\colon A \to B$, $a \mapsto f(a)+g(a)$, and scalar multiplication by $r \in R$ as $rf\colon A \to B$, $a \mapsto r(f(a))$. Furthermore, if $h\colon A' \to A$ is a homomorphism of R-modules then the map $\text{Hom}(A, B) \to \text{Hom}(A', B)$, $f \mapsto f \circ h$, is a homomorphism of R-modules, and if $k\colon B \to B'$ is a homomorphism of R-modules then the map $\text{Hom}(A, B) \to \text{Hom}(A, B')$, $f \mapsto k \circ f$ is a homomorphism of R-modules.*

Proof. We first check that $\text{Hom}(A, B)$ is an abelian group. For f, $g \in \text{Hom}(A, B)$, the map $f + g\colon A \to B$ is indeed a homomorphism of R-modules:

$$\begin{aligned}(f+g)(a+a') &= f(a+a') + g(a+a') \\ &= f(a) + g(a) + f(a') + g(a') = (f+g)(a) + (f+g)(a')\end{aligned}$$

for a, $a' \in A$, and for $r \in R$,

$$(f+g)(ra) = f(ra) + g(ra) = rf(a) + rg(a) = r(f(a) + g(a)) = r(f+g)(a).$$

For f, g, $h \in \text{Hom}(A, B)$, we have

$$\begin{aligned}((f+g)+h)(a) &= (f+g)(a) + h(a) \\ &= f(a) + g(a) + h(a) = f(a) + (g+h)(a) = (f + (g+h))(a),\end{aligned}$$

and similarly, we have $(g + f)(a) = (f + g)(a)$. The identity element is the zero homomorphism, and inverse to f is $-f\colon A \to B$, $a \mapsto -f(a)$.

For $r \in R$ and $f \in \text{Hom}(A, B)$ we check that rf is a homomorphism of R-modules:

$$(rf)(a+a') = rf(a+a') = r(f(a) + f(a')) = rf(a) + rf(a') = (rf)(a) + (rf)(a')$$

for a, $a' \in A$, while for $s \in R$ we have

$$(rf)(sa) = rf(sa) = r(sf(a)) = (rs)f(a) = (sr)f(a) = s(rf(a)) = s(rf)(a).$$

The remaining axioms for R-module are the associativity of scalar multiplication, the distributive axioms, and the compatibility with 1. If r, $s \in R$ then we have $(rs)f\colon A \to B$, $a \mapsto (rs)(f(a)) = rsf(a)$, which is the same as what we get by first acting with s to get $sf\colon A \to B$, $a \mapsto sf(a)$, and then acting with r to get $r(sf)\colon A \to B$, $a \mapsto r(sf(a))$. One of the distributive axioms, $(r+s)f = rf + sf$, is checked by comparing the following two expressions:

$$(r+s)f\colon A \to B, a \mapsto (r+s)(f(a)) = rf(a) + sf(a),$$
$$rf + sf\colon A \to B, a \mapsto (rf)(a) + (sf)(a) = rf(a) + sf(a).$$

The other distributive axiom $r(f+g) = rf + rg$ is also routine to verify, as is the compatiblity with 1.

Let $h\colon A' \to A$ be an R-module homomorphism. Then for f, $g \in \mathrm{Hom}(A, B)$ and $r \in R$,

$$((f+g)\circ h)(a) = (f+g)(h(a)) = f(h(a)) + g(h(a)) = ((f\circ h) + (g\circ h))(a),$$
$$((rf)\circ h)(a) = rf(h(a)) = (r(f\circ h))(a).$$

for all $a \in A$, i.e., we have $(f+g)\circ h = f\circ h + g\circ h$ and $(rf)\circ h = r(f\circ h)$. As well, for an R-module homomorphism $k\colon B \to B'$ the axioms for $f \mapsto k\circ f$ to be an R-module homomorphism are readily verified. □

Example. We have $\mathrm{Hom}(R, A) \xrightarrow{\sim} A$ by $f \mapsto f(1)$. This is an R-module homomorphism (routine), which is injective (if $f(1) = 0$ then $f(r) = f(r1) = r(f(1)) = r0 = 0$ for all $r \in R$) and surjective (for $a \in A$ there is $(r \mapsto ra) \in \mathrm{Hom}(R, A)$ which sends 1 to a). The module $\mathrm{Hom}(A, R)$ is the **dual module** to A, often denoted by A^*. There is the natural map $A \to A^{**}$ to the bidual, sending $a \in A$ to the evaluation map at a; when this is an isomorphism we say that A is a **reflexive** R-module.

We will find it convenient to use the notion of exactness for sequences of modules.

Let R be a commutative ring. A **sequence** of R-modules consists of a collection $(M^i)_{i\in I}$ of R-modules indexed by some integer interval $I \subset \mathbb{Z}$, together with R-module homomorphisms $f^i\colon M^i \to M^{i+1}$ whenever i and $i+1$ belong to I. The sequence is a **complex** if $f^i \circ f^{i-1} = 0$ whenever $i-1$, i, and $i+1$ belong to I. The sequence is **exact** in the ith position if the image of f^{i-1} is equal to the kernel of f^i. An **exact sequence** is a sequence which is exact in the ith position for all i such that $i-1$, i, and $i+1$ belong to I.

Example. Let us take I to consist of four successive integers (it does not matter which), with the module 0 followed by three arbitrary modules M', M, M'':

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \tag{3}$$

Then (3) is exact if and only if f is injective and $\operatorname{im}(f) = \ker(g)$. In this case we call (3) a **left exact** sequence. Similarly,

$$M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0 \tag{4}$$

is exact if and only if $\operatorname{im}(f) = \ker(g)$ and g is surjective. In this case we call (4) a **right exact** sequence. A **short exact sequence** is an exact sequence

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0, \tag{5}$$

i.e., a sequence that is both left exact (ignoring the final 0) and right exact (ignoring the initial 0).

Proposition 1.19. *Let R be a commutative ring, let M', M, and M'' be R-modules, and let $f\colon M' \to M$ and $g\colon M \to M''$ be R-module homomorphisms. Then the following statements are equivalent:*

(i) *The sequence $M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is right exact.*

(ii) *For every R-module N the induced R-module homomorphisms $\operatorname{Hom}(M, N) \to \operatorname{Hom}(M', N)$ and $\operatorname{Hom}(M'', N) \to \operatorname{Hom}(M, N)$ fit into a left exact sequence*

$$0 \to \operatorname{Hom}(M'', N) \to \operatorname{Hom}(M, N) \to \operatorname{Hom}(M', N). \tag{6}$$

Proof. We first show (i) $\Rightarrow$ (ii). Since $g \circ f = 0$, the composite $\operatorname{Hom}(M'', N) \to \operatorname{Hom}(M', N)$ in (6) is 0. If $h \in \operatorname{Hom}(M'', N)$ maps to 0 in $\operatorname{Hom}(M, N)$, then we have $h \circ g = 0$, which since g is surjective implies $h = 0$. So it remains only to show that for $k \in \operatorname{Hom}(M, N)$ satisfying $k \circ f = 0$ there exists $h \in \operatorname{Hom}(M'', N)$ such that $h \circ g = k$. From $k \circ f = 0$ we get an induced homomorphism $\operatorname{coker}(f) \to N$, and from $g \circ f = 0$ we get $\operatorname{coker}(f) \to M''$. By the condition (i), the induced homomorphism $\operatorname{coker}(f) \to M''$ is an isomorphism. Its inverse, composed with $\operatorname{coker}(f) \to N$, is the $h \in \operatorname{Hom}(M'', N)$ that we seek.

For (ii) $\Rightarrow$ (i), the plan is to make strategic choices for N so that the left exactness of (6) gives us what we need to establish (i). First, we take $N := \operatorname{coker}(g)$ to obtain injective $\operatorname{Hom}(M'', \operatorname{coker}(g)) \to \operatorname{Hom}(M, \operatorname{coker}(g))$, sending both the zero homomorphism and the canonical homomorphism to 0. So the zero homomomorphism and the canonical homomorphism are equal, i.e., g is surjective. Next, we take $N := M''$ and, starting with $\mathrm{id}_{M''} \in \operatorname{Hom}(M'', M'')$, obtain that $g \circ f = 0$, i.e., $\operatorname{im}(f) \subset \ker(g)$. Finally, we take $N := \operatorname{coker}(f)$, for which (6) reads

$$0 \to \operatorname{Hom}(M'', \operatorname{coker}(f)) \to \operatorname{Hom}(M, \operatorname{coker}(f)) \to \operatorname{Hom}(M', \operatorname{coker}(f)).$$

The canonical homomorphism in $\operatorname{Hom}(M, \operatorname{coker}(f))$ maps to 0, so by exactness there exists $h \in \operatorname{Hom}(M'', \operatorname{coker}(f))$ such that $h \circ g$ is the canonical homomorphism. It follows that the composite

$$\operatorname{coker}(f) \xrightarrow{\bar{g}} M'' \xrightarrow{h} \operatorname{coker}(f)$$

is the identity morphism, where the left-hand morphism, denoted by $\bar{g}$, is induced from g. So $\bar{g}$ is injective, and hence we have $\mathrm{im}(f) = \ker(g)$. □

The implication (i) ⇒ (ii) of Proposition 1.19 may be expressed by the phrase, $\mathrm{Hom}(-, N)$ *is a left exact functor* for every R-module N. First of all, $\mathrm{Hom}(-, N)$ is a functor, i.e., sends an R-module A to $\mathrm{Hom}(A, N)$, which has the structure of R-module (Proposition 1.18), and sends a homomorphism of R-modules $h\colon A' \to A$ to a homomorphism $\mathrm{Hom}(A, B) \to \mathrm{Hom}(A', B)$ (also by Proposition 1.18). It is easy to check that this is compatible with composition (if $h'\colon A'' \to A$ is another homomorphism then the composite $\mathrm{Hom}(A, B) \to \mathrm{Hom}(A', B) \to \mathrm{Hom}(A'', B)$ sends $f \in \mathrm{Hom}(A, B)$ to $(f \circ h) \circ h' = f \circ (h \circ h')$) and identity maps (composition with id_A induces $\mathrm{id}_{\mathrm{Hom}(A,B)}$). This functor reverses the direction of arrows, a feature which may be made explicit by putting the adjective **contravariant** in front of "functor", to distinguish from **covariant** functors, which preserve the direction of arrows. The phrase "additive functor" denotes further compatibility with addition of homomorphisms: given $\tilde{h}\colon A' \to A$, we have $f \circ (h + \tilde{h}) = f \circ h + f \circ \tilde{h}$ for $f \in \mathrm{Hom}(A, B)$. A contravariant (respectively covariant) additive functor is defined to be **left exact** if it sends a right exact sequence (respectively left exact sequence) to a left exact sequence. This condition is equivalent to every short exact sequence being sent to a left exact sequence. (Every short exact sequence (5) is both left and right exact, so one implication is trivial. For the other implication, right exact sequence (4) may be chopped up into two short exact sequences $0 \to \ker(f) \to M' \to \ker(g) \to 0$ and $0 \to \ker(g) \to M \to M'' \to 0$, sent to respective left exact sequence which may be reassembled into the desired left exact sequence. In case the functor is covariant, "right" needs to be replaced by "left" and kernels by cokernels.)

As well, a contravariant (respectively covariant) additive functor is **right exact** if it sends left exact sequences (respectively right exact sequences) to right exact sequences. As above, an equivalent condition is to send short exact sequences to right exact sequences. If the functor sends short exact sequences to short exact sequences, we say the functor is **exact**; equivalently, arbitrary exact sequences $((M^i)_{i\in I}, (f^i)_i)$ are sent to exact sequences (with the same indexing set in case of a covariant functor, and with reversed indexing set $\{-i \mid i \in I\}$ in case of a contravariant functor).

We recall that a system of generators $(x_i)_{i\in I}$ of an R-module M induces a surjective R-module homomorphism $\bigoplus_{i\in I} R \to M$. Let K denote the kernel. Then we may consider a system of generators $(y_j)_{j\in J}$ of K with corresponding surjective R-module homomorphism $\bigoplus_{j\in J} R \to K$. This information may be put together into a right exact sequence

$$\bigoplus_{j\in J} R \to \bigoplus_{i\in I} R \to M \to 0. \tag{7}$$

Each standard basis element $e_{j'} \in \bigoplus_{j\in J} R$ maps to some element

$$a_1 e_{i_1} + \cdots + a_n e_{i_n} \in \bigoplus_{i\in I} R. \tag{8}$$

Since $a_1 e_{i_1} + \cdots + a_n e_{i_n} \in K$, (8) determines a *relation*

$$a_1 x_{i_1} + \cdots + a_n x_{i_n} = 0 \tag{9}$$

in M. We call a right exact sequence (7), or equivalently the pair consisting of the system of generators $(x_i)_{i\in I}$ of M and system of generators $(y_j)_{j\in J}$ of K, a **presentation of M by generators and relations**.

Example. Let m be a natural number. Then $\mathbb{Z}/m\mathbb{Z}$ is generated by $\bar{1}$, and there is the relation $m \cdot \bar{1} = \bar{0}$ in $\mathbb{Z}/m\mathbb{Z}$. The corresponding presentation by generators and relations is

$$\mathbb{Z} \overset{\cdot m}{\to} \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \to 0 \tag{10}$$

where the second map is the canonical homomorphism.

As well, a module may be *defined* by means of generators and relations. This means, we are given index sets I and J and a collection of elements $(y_j)_{j\in J}$ of $\bigoplus_{i\in I} R$. These determine a homomorphism of R-modules

$$\bigoplus_{j\in J} R \to \bigoplus_{i\in I} R. \tag{11}$$

Then the module M, defined to be the cokernel of the homomorphism (11), fits into a right exact sequence (7). The element in M, image of e_i under the canonical homomorphism from $\bigoplus_{i\in I} R$, may be denoted by $\bar{e}_i$ or x_i, or any other convenient notation. It is common to write the elements y_j in the form of relations (9), rather than the form (8). More generally, a relation as in (9) but with $a'_1 e_{i'_1} + \cdots + a'_m e_{i'_m}$ on the right-hand side leads to $a_1 e_{i_1} + \cdots + a_n e_{i_n} - a'_1 e_{i'_1} - \cdots - a'_m e_{i'_m} \in \bigoplus_{i\in I} R$.

We consider a pair of R-modules M and N. Let us recall the **universal property of the tensor product**. An R-module $M \otimes N$ with R-bilinear map $M \times N \to M \otimes N$ (i.e., map that is R-linear in M for each fixed $n \in N$ and R-linear in N for each fixed $m \in M$) satisfies the universal property of the tensor product if, for every R-module T with R-bilinear map $M \times N \to T$, there exists a unique R-module homomorphism $M \otimes N \to T$ fitting into a commutative diagram (as the dashed arrow)

$$\begin{array}{ccc} M \times N & & \\ \downarrow & \searrow & \\ M \otimes N & \dashrightarrow & T \end{array} \tag{12}$$

Definition. Let R be a commutative ring, and let M and N be R-modules. The **tensor product** is defined, up to a canonical isomorphism, to be an R-module with bilinear map from $M \times N$ satisfying the universal property of the tensor product.

The tensor product is denoted by $M \otimes N$ or $M \otimes_R N$. The image of an element $(m, n) \in M \times N$ under the bilinear map $M \times N \to M \otimes N$ is denoted by $m \otimes n$.

For this to be mathematically valid we need to know that *some* R-module with bilinear map from $M \times N$ satisfies the universal property. This is achieved by a construction with generators and relations. We consider generators indexed by

$$I := M \times N, \tag{13}$$

with the generator indexed by (m, n) denoted by $m \otimes n$, and relations indexed by

$$J := M \times M \times N \sqcup M \times N \times N \sqcup R \times M \times N \sqcup R \times M \times N. \tag{14}$$

This is a disjoint union: an element is a triple, e.g., (m, m', n), together with an index specifying the term in (14), to which we should view the triple as belonging. The respective relations are (with m, $m' \in M$, n, $n' \in N$, $a \in R$)

$$(m + m') \otimes n = m \otimes n + m' \otimes n, \tag{14.1}$$
$$m \otimes (n + n') = m \otimes n + m \otimes n', \tag{14.2}$$
$$a(m \otimes n) = am \otimes n, \tag{14.3}$$
$$a(m \otimes n) = m \otimes an. \tag{14.4}$$

where the term-index is indicated in the equation-numbering.

Example. Let $R := \mathbb{Z}$. Then $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z} = 0$. To see this from the definition, we list the generators

$$0 \otimes 0, \quad 0 \otimes 1, \quad 0 \otimes 2, \quad 1 \otimes 0, \quad 1 \otimes 1, \quad 1 \otimes 2$$

together with some relations

$$\begin{aligned} 0 \otimes n &= 0 \otimes n + 0 \otimes n \qquad (n \in \mathbb{Z}/3\mathbb{Z}), \\ 1 \otimes 0 &= 1 \otimes 0 + 1 \otimes 0, \\ 2(1 \otimes 1) &= 0 \otimes 1, \\ 2(1 \otimes 1) &= 1 \otimes 2, \\ 1 \otimes 1 &= 1 \otimes 2 + 1 \otimes 2. \end{aligned} \tag{15}$$

There are infinitely many relations, but the 7 relations (15) suffice to show that each of the generators is zero in $\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z}$.

Proposition 1.20. *Let R be a commutative ring, and let M and N be R-modules. The module $M \otimes N$, defined by generators* (13) *and relations* (14)*, with the map $M \times N \to M \otimes N$, $(m, n) \mapsto m \otimes n$, satisfies the universal property of the tensor product.*

Proof. We first verify that the map $M \times N \to M \otimes N$ is R-bilinear. Fixing $n \in N$, relations (14.1) and (14.3) tell us that the map $M \to M \otimes N$, $m \mapsto m \otimes n$, is R-linear. Similarly, for fixed $m \in M$, the map $N \to M \otimes N$, $n \mapsto m \otimes n$, is R-linear by relations (14.2) and (14.4).

Now let T be an R-module. By left exactness of $\mathrm{Hom}(-, T)$, and the identification of Hom from a direct sum of copies of R to T with a corresponding product of copies of T, we have the left exact sequence

$$0 \to \mathrm{Hom}(M \otimes N, T) \xrightarrow{f} \prod_{i \in I} T \xrightarrow{g} \prod_{j \in J} T,$$

where f sends $\varphi\colon M \otimes N \to T$ to $(\varphi(m \otimes n))_{(m,n) \in M \times N}$, and where an element $(t_{m,n})_{m \in M,\, n \in N} \in \prod_{i \in I} T$ lies in the kernel of g if and only if for all $a \in R$, m, $m' \in M$, and n, $n' \in N$, we have

$$t_{m+m',n} = t_{m,n} + t_{m',n}, \tag{16.1}$$
$$t_{m,n+n'} = t_{m,n} + t_{m,n'}, \tag{16.2}$$
$$a t_{m,n} = t_{am,n}, \tag{16.3}$$
$$a t_{m,n} = t_{m,an}. \tag{16.4}$$

Now let $M \times N \to T$ be an R-bilinear map, given by $(m, n) \mapsto t_{m,n}$. Then we have (16.1)–(16.4), so $(t_{m,n})_{m \in M,\, n \in N} \in \ker(g)$. It follows that $(t_{m,n})_{m \in M,\, n \in N} = f(\varphi)$ for a unique $\varphi \in \mathrm{Hom}(M \otimes N, T)$. This φ is the unique dashed arrow in (12), and the universal property is established. □

In the Linear Algebra lecture, a different construction of the tensor product was given, adapted to the case of vector spaces over a field where existence of a basis is a general fact. If M and N are free R-modules, M with basis $(m_i)_{i \in I}$ and N with basis $(n_j)_{j \in J}$, then

$$M \otimes N \cong \bigoplus_{(i,j) \in I \times J} R, \tag{17}$$

where $m_i \otimes n_j$ is identifed with standard basis element $e_{i,j}$. We also saw direct constructions for the case that M or N is free. If M is free, with basis $(m_i)_{i \in I}$, then for arbitrary N,

$$M \otimes N \cong \bigoplus_{i \in I} N, \tag{18}$$

with $m_{i'} \otimes n$ for $i' \in I$ and $n \in N$ identified with $(\delta_{ii'} n)_{i \in I}$. Similarly, if N is free, with basis $(n_j)_{j \in J}$, then for arbitrary M,

$$M \otimes N \cong \bigoplus_{j \in J} M, \tag{19}$$

with $m \otimes n_{j'}$ for $m \in M$ and $j' \in J$ identified with $(\delta_{jj'} m)_{j \in J}$. For instance, (17) is shown by verifying the universal property of the tensor product for the module $\bigoplus_{(i,j) \in I \times J} R$ and unique R-bilinear map $M \times N \to \bigoplus_{(i,j) \in I \times J} R$ sending (m_i, n_j) to $e_{i,j}$. This was done for vector spaces in the Linear Algebra lecture; only notational changes are needed to do this for free modules. Similarly, we obtain the isomorphisms (18) and (19) by verifying the universal property of the tensor product for each indicated module and uniquely determined R-bilinear map, making only notational changes to the argument given in the Linear Algebra algebra.

Other isomorphisms that can be established using the universal property include symmetry

$$N \otimes M \cong M \otimes N, \tag{20}$$

by identifying $n \otimes m$ on the left with $m \otimes n$ on the right, associativity

$$(M \otimes N) \otimes P \cong M \otimes (N \otimes P) \tag{21}$$

for any third R-module P, with $(m \otimes n) \otimes p$ identified with $m \otimes (n \otimes p)$, and compatibility with direct sums

$$(M \oplus N) \otimes P \cong (M \otimes P) \oplus (N \otimes P), \tag{22}$$

$$M \otimes (N \oplus P) \cong (M \otimes N) \oplus (M \otimes P), \tag{23}$$

with $(m, n) \otimes p$ identified with $(m \otimes p, n \otimes p)$ in (22) and $m \otimes (n, p)$ with $(m \otimes n, m \otimes p)$ in (23). For vector spaces, isomorphisms (20) and (21) were covered in the Linear Algebra lecture, and the arguments are valid for R-modules. The verification of (22) follows the same pattern as the other verifications: bilinear maps from $(M \oplus N) \times P$ to an arbitrary R-module T by restriction to $M \times P$ and to $N \times P$, stand in bijection with pairs of bilinear maps to T, from $M \times P$ and from $N \times P$, and from this observation it is quickly seen that $(M \otimes P) \oplus (N \otimes P)$ satisfies the universal property of the tensor product. The verification of (23) is similar.

Given R-module homomorphisms $f \colon M \to M'$ and $g \colon N \to N'$, there is an induced R-module homomorphism $f \otimes g \colon M \otimes N \to M' \otimes N'$. This is uniquely determined from the universal property by the R-bilinear map $M \times N \to M' \otimes N'$, $(m, n) \mapsto f(m) \otimes g(n)$, and is functorial (if $f' \colon M' \to M''$ and $g' \colon N' \to N''$ are further R-module homomorphisms, then $(f' \otimes g') \circ (f \otimes g) = (f' \circ f) \otimes (g' \circ g)$, and $\mathrm{id}_M \otimes \mathrm{id}_N = \mathrm{id}_{M \otimes N}$). In particular, we have $f \otimes \mathrm{id}_N \colon M \otimes N \to M' \otimes N$, making $- \otimes N$ into an additive functor. (Additivity: if $\tilde{f} \colon M \to M'$ is another R-module homomorphism, then $(f + \tilde{f}) \otimes \mathrm{id}_N$ sends $m \otimes n$ to $(f(m) + \tilde{f}(m)) \otimes n = f(m) \otimes n + \tilde{f}(m) \otimes n$, as does $f \otimes \mathrm{id}_N + \tilde{f} \otimes \mathrm{id}_N$.)

Lemma 1.21. *Let R be a commutative ring, and let M, N, and P be R-modules. Then we have an isomorphism*

$$\mathrm{Hom}(M \otimes N, P) \xrightarrow{\sim} \mathrm{Hom}(M, \mathrm{Hom}(N, P)) \tag{24}$$

sending $g\colon M \otimes N \to P$ to the map $m \mapsto (n \mapsto g(m \otimes n))$. Given a further R-module M' and R-module homomorphism $f\colon M \to M'$, the isomorphism (24) *and similar isomorphism with M' in place of M fit into a commutative diagram*

$$\begin{array}{ccc} \mathrm{Hom}(M' \otimes N, P) & \xrightarrow{\sim} & \mathrm{Hom}(M', \mathrm{Hom}(N, P)) \\ \downarrow & & \downarrow \\ \mathrm{Hom}(M \otimes N, P) & \xrightarrow{\sim} & \mathrm{Hom}(M, \mathrm{Hom}(N, P)) \end{array} \tag{25}$$

with left-hand vertical map given by composition with $f \otimes \mathrm{id}_N$ and right-hand vertical map given by composition with f.

Proof. For every $m \in M$ we have $(n \mapsto g(m \otimes n)) \in \mathrm{Hom}(N, P)$. We see as well that $m \mapsto (n \mapsto g(m \otimes n))$ is a homomorphism of R-modules $M \to \mathrm{Hom}(N, P)$. So we have a map (24), readily verified to be a homomorphism of R-modules. If $(n \mapsto g(m \otimes n))$ is the zero homomorphism for every $m \in M$, then g must be zero; thus (24) is injective. Given an R-module homomorphism $M \to \mathrm{Hom}(N, P)$, $m \mapsto \varphi_m$, then we have the R-bilinear map $(m, n) \mapsto \varphi_m(n)$, which determines the linear map $m \otimes n \mapsto \varphi_m(n)$. So (24) is also surjective.

For the second assertion, an R-module homomorphism $g'\colon M' \otimes N \to P$, is sent by the top map of diagram (25) to $m' \mapsto (n \mapsto g'(m' \otimes n))$, which by the right-hand vertical map is sent to

$$m \mapsto (n \mapsto g'(f(m) \otimes n)).$$

On the other hand, $f \otimes \mathrm{id}_N$ sends $m \otimes n$ to $f(m) \otimes n$. So g' is sent by the composite of the bottom and left-hand maps in the diagram (25) to $m \mapsto (n \mapsto g'(f(m) \otimes n))$, and thus the diagram commutes. □

Proposition 1.22. *Let R be a commutative ring. For any R-module N, the functor $- \otimes N$ is right exact.*

Proof. Given a right exact sequence (4), we need to show that the sequence

$$M' \otimes N \xrightarrow{f \otimes \mathrm{id}_N} M \otimes N \xrightarrow{g \otimes \mathrm{id}_N} M'' \otimes N \to 0 \tag{26}$$

is right exact. This is equivalent, by Proposition 1.19, to the assertion that for every R-module P, we have the left exact sequence

$$0 \to \mathrm{Hom}(M'' \otimes N, P) \to \mathrm{Hom}(M \otimes N, P) \to \mathrm{Hom}(M' \otimes N, P) \tag{27}$$

with homomorphisms of Hom-modules induced from (26). By Lemma 1.21, each Hom-module in (27) may be identified with a corresponding module of homomorphisms to $\mathrm{Hom}(N,P)$, and the isomorphisms are compatible with the homomorphisms of Hom-modules in

$$0 \to \mathrm{Hom}(M'', \mathrm{Hom}(N,P)) \to \mathrm{Hom}(M, \mathrm{Hom}(N,P)) \to \mathrm{Hom}(M', \mathrm{Hom}(N,P)) \tag{28}$$

But the functor $\mathrm{Hom}(-, \mathrm{Hom}(N,P))$ is left exact, so the sequence (28) is left exact, hence so is the sequence (27), as required. □

Proposition 1.22 may be applied to a module presentation (7) to obtain a right exact sequence

$$\bigoplus_{j\in J} N \to \bigoplus_{i\in I} N \to M \otimes N \to 0. \tag{29}$$

For $j' \in J$ and $n \in N$, the first map in (29) acts by

$$(\delta_{jj'}n)_{j\in J} \mapsto (a_1\delta_{ii_1}n + \cdots + a_n\delta_{ii_n}n)_{i\in I},$$

in the notation of (8), and for $i' \in I$ the second map in (29) sends $(\delta_{ii'}n)_{i\in I}$ to $\bar{e}_{i'} \otimes n$. When neither M nor N is free, so (17)–(19) are inapplicable, (29) offers a description of $M \otimes N$. For example, with $R := \mathbb{Z}$, we obtain from the presentation (10) an isomorphism

$$\mathbb{Z}/m\mathbb{Z} \otimes N \xrightarrow{\sim} N/mN,$$

given by $\bar{1} \otimes n \mapsto \bar{n}$ for $n \in N$.

Let S be a commutative ring and $f\colon R \to S$ a ring homomorphism, a situation that we may express by describing S as a commutative **R-algebra**. Then we may view S as an R-module, with $rs := f(r)s$ for $r \in R$ and $s \in S$. For any R-module M, now, $S \otimes_R M$ has the structure of S-module by

$$s(s' \otimes m) := (ss') \otimes m$$

for s, $s' \in S$ and $m \in M$. (This clearly satisfies the axioms to be an S-module.) If T is another commutative R-algebra, then $S \otimes_R T$ acquires a ring structure with

$$(s \otimes t) \cdot (s' \otimes t') := ss' \otimes tt'.$$

We may view $S \otimes_R T$ as S-algebra (by $S \to S \otimes_R T$, $s \mapsto s \otimes 1$), and as well as T-algebra (by $T \to S \otimes_R T$, $t \mapsto 1 \otimes t$).

Example. We have $S \otimes_R R[X] \cong S[X]$, and more generally:

$$S \otimes_R R[X_1, \ldots, X_n] \cong S[X_1, \ldots, X_n],$$
$$S \otimes_R (R[X_1, \ldots, X_n]/(f_1, \ldots, f_m)) \cong S[X_1, \ldots, X_n]/(g_1, \ldots, g_m),$$

where we use the top isomorphism to obtain from each polynomial $f_i \in R[X_1, \ldots, X_n]$ a polynomial $g_i \in S[X_1, \ldots, X_n]$ (corresponding to $1 \otimes f_i$).

Example. Let R be a PID with field of fractions K. For irreducible $p \in R$, we recall, we have the field $\boldsymbol{k}_p := R/(p)$. To any finitely generated R-module M, the integer s from the Structure Theorem (Theorem 1.17) $M \cong R/(a_1) \oplus \cdots \oplus R/(a_r) \oplus R^s$ may be characterized as

$$\dim_K(K \otimes_R M). \tag{30}$$

Recalling that $a_1, \ldots, a_r$ are determined (up to multiplication by units) by

$$\dim_{\boldsymbol{k}_p}(M[p^\ell]/M[p^{\ell-1}]), \tag{31}$$

for various irreducible $p \in R$ and $\ell \in \mathbb{N}$, with r equal to the maximum of the integers (31) we have a characterization of all the data attached to M by the Structure Theorem, entirely in terms of dimensions (30)–(31) of vector spaces obtained from M over fields associated with R.

1.5 Multilinear algebra for modules

Symmetric and exterior powers make sense for modules over a commutative ring. As for the tensor product, symmetric and exterior powers are defined by means of universal properties coupled with constructions by generators and relations.

The definition of R-bilinear map $M \times N \to T$ generalizes to R-multilinear map $M_1 \times \cdots \times M_n \to T$ for any $n \in \mathbb{N}_{>0}$. (For each $1 \leq j \leq n$ and choice of element from M_i for every $i \neq j$, the map $M_j \to T$ obtained by letting the element from M_j vary should be R-linear.) We restrict immediately to the case where the n modules are all the same. An R-multilinear map $M^n \to T$ is **symmetric** if it is unchanged under arbitrary permutation of the arguments and **alternating** if it vanishes when any two of the arguments are equal. For each kind of multilinear map (symmetric or alternating) there is a corresponding universal property

$$\begin{array}{ccc} M^n & & M^n \\ \downarrow & \searrow & \downarrow \quad \searrow \\ \mathrm{Sym}^n M & \dashrightarrow T & \bigwedge^n M \dashrightarrow T \end{array} \tag{1}$$

A module $\mathrm{Sym}^n M$ with symmetric multilinear map from M^n, respectively a module $\bigwedge^n M$ with alternating multilinear map from M^n, satisfies the respective universal property if for every R-module T with symmetric respectively alternating multilinear map $M^n \to T$ there exists a unique linear map fitting as the dashed arrow into a commutative diagram in (1). The nth **symmetric power** respectively **exterior power** is defined up to a canonical isomorphism to be an R-module with symmetric respectively alternating multilinear map from M^n, satisfying the respective universal property.

There are respective constructions by means of generators and relations. In each case we consider generators indexed by $I := M^n$, with respective notation $m_1 \cdots m_n$

and $m_1 \wedge \cdots \wedge m_n$ for the generator indexed by $(m_1, \ldots, m_n)$. The relations are indexed by a disjoint union J of

- n copies of M^{n+1},
- n copies of $R \times M^n$,
- $\binom{n}{2}$ copies of M^n, respectively of M^{n-1}.

The relation corresponding to $(m_1, \ldots, m_{j-1}, m_j, m_j', m_{j+1}, \ldots, m_n)$ in the jth copy of M^{n+1} is

$$\begin{aligned} m_1 \cdots m_{j-1}(m_j + m_j')m_{j+1} \cdots m_n = m_1 \cdots m_{j-1} m_j m_{j+1} \cdots m_n \\ + m_1 \cdots m_{j-1} m_j' m_{j+1} \cdots m_n, \end{aligned} \tag{2}$$

respectively the same with $\wedge$ inserted. To $(a, m_1, \ldots, m_n)$ in the jth copy of $R \times M^n$:

$$a(m_1 \cdots m_n) = m_1 \cdots m_{j-1}(am_j)m_{j+1} \cdots m_n, \tag{3}$$

respectively the same with $\wedge$ inserted. Finally, to $(m_1, \ldots, m_n)$ in the (i,j)-copy of M^n $(1 \leq i < j \leq n)$,

$$m_1 \cdots m_j \cdots m_i \cdots m_n = m_1 \cdots m_i \cdots m_j \cdots m_n \tag{4}$$

respectively to $(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_n)$ in the (i,j)-copy of M^{n-1},

$$m_1 \wedge \cdots \wedge m_i \wedge \cdots \wedge m_i \wedge \cdots \wedge m_n = 0. \tag{4'}$$

Exactly as in Proposition 1.20, the module defined by these generators and relations, with map from M^n sending $(m_1, \ldots, m_n)$ to $m_1 \cdots m_n$ respectively $m_1 \wedge \cdots \wedge m_n$, satisfies the universal property of the symmetric power, respectively of the exterior power.

As we have seen in Linear Algebra,

$$m_1 \wedge \cdots \wedge m_j \wedge \cdots \wedge m_i \wedge \cdots \wedge m_n = -m_1 \wedge \cdots \wedge m_i \wedge \cdots \wedge m_j \wedge \cdots \wedge m_n. \tag{5}$$

The approach to symmetric and exterior power taken in Linear Algebra, depending on the existence of a finite basis, shows that in case M is finitely generated and free of rank r,

$$\operatorname{Sym}^n M \cong R^{\binom{n+r-1}{n}}, \qquad \bigwedge^n M \cong R^{\binom{r}{n}}.$$

If we omit relations (4)/(4′), then we have relations (2) indexed by n copies of M^{n+1} and (3) indexed by n copies of $R \times M^n$ defining the n-fold tensor product

$$T^n M := M \otimes \cdots \otimes M,$$

defined by a universal property as in (1) but for multilinear maps without the condition to be symmetric or alternating. Then we have homomorphisms

$$T^n M \to \operatorname{Sym}^n M \qquad \text{and} \qquad T^n M \to \bigwedge^n M, \tag{6}$$

sending $m_1 \otimes \cdots \otimes m_n$ to $m_1 \cdots m_n$, respectively to $m_1 \wedge \cdots \wedge m_n$, and these are surjective. When $n = 1$, multilinear is the same as linear, and the condition to be symmetric or alternating is trivial:

$$\operatorname{Sym}^1 M \cong M, \qquad \bigwedge^1 M \cong M, \qquad T^1 M \cong M.$$

In each of these constructions, we may form the direct sum over all $n \in \mathbb{N}$. For $n = 0$ we agree by convention,

$$\operatorname{Sym}^0 M := R, \qquad \bigwedge^0 M := R, \qquad T^0 M := R. \tag{7}$$

Then we define

$$\operatorname{Sym}^\bullet M := \bigoplus_{n=0}^{\infty} \operatorname{Sym}^n M, \qquad \bigwedge^\bullet M := \bigoplus_{n=0}^{\infty} \bigwedge^n M, \qquad T^\bullet M := \bigoplus_{n=0}^{\infty} T^n M. \tag{8}$$

Proposition 1.23. *Let R be a commutative ring and M an R-module. Then $\operatorname{Sym}^\bullet M$, $\bigwedge^\bullet M$, and $T^\bullet M$ defined in (8) are rings, where in each case the multiplication is induced from the maps $M^n \times M^{n'} \to M^{n+n'}$, $((m_1, \dots, m_n), (m'_1, \dots, m'_{n'}) \mapsto (m_1, \dots, m_n, m'_1, \dots, m'_{n'}))$, and where multiplication on either side by elements of the respective copy of R in (7) is defined via the R-module structure. The ring $\operatorname{Sym}^\bullet M$ is commutative, while in the ring $\bigwedge^\bullet M$ we have*

$$m'_1 \wedge \cdots \wedge m'_{n'} \wedge m_1 \wedge \cdots \wedge m_n = (-1)^{nn'} m_1 \wedge \cdots \wedge m_n \wedge m'_1 \wedge \cdots \wedge m'_{n'}. \tag{9}$$

Proof. We start with details about the multiplication induced from $M^n \times M^{n'} \to M^{n+n'}$, in the case of the symmetric algebra (the other cases are similar). Let n and n' be positive integers. For fixed $(m_1, \dots, m_n) \in M^n$, we have $M^{n'} \to M^{n+n'}$, sending $(m'_1, \dots, m'_{n'})$ to $(m_1, \dots, m_n, m'_1, \dots, m'_{n'})$, and hence $M^{n'} \to \operatorname{Sym}^{n+n'} M$,

$$(m'_1, \dots, m'_{n'}) \mapsto m_1 \cdots m_n m'_1 \cdots m'_{n'},$$

which is multilinear and symmetric. By the universal property of the symmetric power, this induces a homomorphism $\operatorname{Sym}^{n'} M \to \operatorname{Sym}^{n+n'} M$. So, we have $M^n \to \operatorname{Hom}(\operatorname{Sym}^{n'} M, \operatorname{Sym}^{n+n'} M)$, and this is multilinear and symmetric. The universal property determines a homomorphism $\operatorname{Sym}^n M \to \operatorname{Hom}(\operatorname{Sym}^{n'} M, \operatorname{Sym}^{n+n'} M)$. Applying Lemma 1.21, we obtain a linear map from $\operatorname{Sym}^n M \otimes \operatorname{Sym}^{n'} M$ to $\operatorname{Sym}^{n+n'} M$, and hence a bilinear map

$$\operatorname{mult}_{n,n'} \colon \operatorname{Sym}^n M \times \operatorname{Sym}^{n'} M \to \operatorname{Sym}^{n+n'} M.$$

Let n'' be another positive integer. We need to show the composites $\mathrm{mult}_{n+n',n''} \circ (\mathrm{mult}_{n,n'} \times \mathrm{id})$ and $\mathrm{mult}_{n,n'+n''} \circ (\mathrm{id} \times \mathrm{mult}_{n',n''})$ (from $\mathrm{Sym}^n M \times \mathrm{Sym}^{n'} M \times \mathrm{Sym}^{n''} M$ to $\mathrm{Sym}^{n+n'+n''} M$) are equal. By multilinearity, this can be checked on elements of the form $((m_1 \cdots m_n), (m'_1 \cdots m'_{n'}), (m''_1 \cdots m''_{n''}))$, and in each case we obtain $m_1 \cdots m_n m'_1 \cdots m'_{n'} m''_1 \cdots m''_{n''}$.

Multiplication in $\mathrm{Sym}^\bullet M$ is commutative, since

$$m'_1 \cdots m'_{n'} m_1 \cdots m_n = m_1 \cdots m_n m'_1 \cdots m'_{n'}$$

in $\mathrm{Sym}^{n+n'} M$. In $\bigwedge^\bullet M$ we apply (5) nn' times to obtain (9). □

Example. Let R be a commutative ring. For $n \in \mathbb{N}$,

$$R[X_1, \ldots, X_n] \cong \mathrm{Sym}^\bullet(R^n),$$

by the homomorphism of R-algebras sending X_i to $e_i \in R^n \cong \mathrm{Sym}^1(R^n)$ for all i.

More generally there is a polynomial ring $\mathrm{Sym}^\bullet(\bigoplus_{s \in S} R)$ for any set S; here the basis element indexed by s of $\mathrm{Sym}^1(\bigoplus_{s \in S} R) \cong \bigoplus_{s \in S} R$ is commonly denoted by X_s. If $S = \{1, \ldots, n\}$ we recover the usual polynomial ring in n variables.

Example. For any R-module M and commutative R-algebra R' the map

$$\mathrm{Hom}_{R\text{-alg}}(\mathrm{Sym}^\bullet M, R') \to \mathrm{Hom}_R(M, R'),$$

where on the left we have homomorphisms of R-algebras, on the right, of R-modules, and the map is given by restriction to $\mathrm{Sym}^1 M \cong M$, is bijective. In particular,

$$\mathrm{Hom}_{R\text{-alg}}\Big(\mathrm{Sym}^\bullet\big(\bigoplus_{s \in S} R\big), R'\Big) \to \{\text{maps } S \to R'\},$$

sending an R-algebra homomorphism f to the to map of sets $s \mapsto f(X_s)$, is bijective.

2 Groups

Abelian groups lie at the heart of most of the algebraic structures used in Linear Algebra, as well as here in Section 1 on rings and modules. For these developments, non-abelian groups such as the symmetric group S_n (with $n \geq 3$) and general linear group $GL_n(R)$ (with $n \geq 2$ and any commutative ring R, not the zero ring) play an important role. Now focus shifts to groups themselves. We develop the basic theory of groups, often with an emphasis on finite groups. If G is a finite group then $|G|$ is called the **order** of G.

Since the symmetric groups S_n are among the most important examples of groups, it is worth recalling the notations used for an element $\sigma \in S_n$:

- (two-line notation) the integers $1, \ldots, n$ are written on the first line, and $\sigma(1)$, $\ldots$, $\sigma(n)$ are written on the second line; e.g.,

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 4 & 5 & 1 & 3 & 6 & 2 \end{bmatrix}.$$

- (one-line notation) $\sigma(1), \ldots, \sigma(n)$ are written on one line:

$$\begin{bmatrix} 7 & 4 & 5 & 1 & 3 & 6 & 2 \end{bmatrix}.$$

- (cycle notation) pairwise disjoint cycles of lengths ≥ 2 are written:

  $(1,7,2,4)(3,5)$, or $(3,5)(1,7,2,4)$ or $(5,3)(4,1,7,2)$ $\ldots$

  (representation is not unique since the cycles can be written in any order, and within each cycle there is an arbitrary choice of the first element).

When writing a permutation in cycle notation we often omit the commas, so for instance $(1,7,2,4)(3,5)$ may be written as $(1724)(35)$.

2.1 Cosets, normal subgroups, quotient groups

Definition. Let G be a group, and let H be a subgroup of G. The **left cosets** of H are the subsets of G of the form $aH = \{ah \,|\, h \in H\}$ for $a \in G$. The **right cosets** of H are the subsets of G of the form $Ha = \{ha \,|\, h \in H\}$ for $a \in G$.

We can see these clearly in an example. The multiplication table of $G := S_3$ is:

id	(12)	(13)	(23)	(123)	(132)
(12)	id	(132)	(123)	(23)	(13)
(13)	(123)	id	(132)	(12)	(23)
(23)	(132)	(123)	id	(13)	(12)
(123)	(13)	(23)	(12)	(132)	id
(132)	(23)	(12)	(13)	id	(123)

If we take $H := \{\mathrm{id}, (12)\}$, then the left H-cosets are the subsets of G appearing in the leftmost two columns of each row of the multiplication table:

$$\{\mathrm{id}, (12)\}, \qquad \{(13), (123)\}, \qquad \{(23), (132)\}.$$

The right H-cosets are read off from the top two rows in each column:

$$\{\mathrm{id}, (12)\}, \qquad \{(13), (132)\}, \qquad \{(23), (123)\}.$$

Notice, for instance, that $\{(13), (123)\}$ is a left H-coset but not a right H-coset. We also see examples of the following general fact:

Proposition 2.1. *Let G be a group and $H \subset G$ a subgroup. Any two left cosets of H are either equal to each other or are disjoint. The union of all left H-cosets is G. As well, any two right cosets of H are either equal to each other or are disjoint, and the union of all right H-cosets is G.*

Proof. Let a, $a' \in G$. If $aH \cap a'H \neq \emptyset$, then there exist h and h' in H such that $ah = a'h'$, and then $a'H = a'h'H = ahH = aH$. Since $a \in aH$, the union of all left H-cosets is G. Similarly, if $Ha \cap Ha' \neq \emptyset$ then $Ha' = Ha$; as well, $a \in Ha$. □

It follows that "belong to the same left H-coset" defines an equivalence relation on G, whose equivalence classes are the left cosets of H. Since the unique left H-coset containing an element $a \in G$ is aH, another element $a' \in G$ belongs to the same left H-coset as a if and only if there exists $h \in H$ such that $a' = ah$. We introduce the notation

$$G/H := \{aH \,|\, a \in G\}$$

for the set of left H-cosets. Similarly, "belong to the same right H-coset" defines an equivalence relation on G; the equivalence classes are the right cosets of H, and two elements a, $a' \in G$ lie in the same right H-coset if and only if there exists $h \in H$ such that $a' = ha$. We introduce the notation

$$H\backslash G := \{Ha \,|\, a \in G\}$$

for the set of right H-cosets.

The bijective map $G \to G$, $g \mapsto g^{-1}$, induces a bijective map from G/H to $H\backslash G$, since for $a \in G$ we have $\{g^{-1} \,|\, g \in aH\} = Ha^{-1}$ and $\{g^{-1} \,|\, g \in Ha\} = a^{-1}H$. It follows that if one of the sets G/H and $H\backslash G$ is finite, then so is the other and they have the same number of elements.

Definition. A subgroup H of a group G is **of finite index** if G/H is finite. If this is the case, the cardinality of G/H is called the **index** of H in G and denoted by $[G : H]$:

$$[G : H] := |G/H|.$$

Proposition 2.2. *Let G be a finite group and H a subgroup. Then we have*

$$|G| = [G : H] \cdot |H|.$$

Proof. By Proposition 2.1, G is the disjoint union of its left cosets. There are $[G : H]$ left cosets, and every left coset has $|H|$ elements. □

Corollary 2.3 (Lagrange's theorem)**.** *If G is a finite group and H is a subgroup, then the order of H divides the order of G.*

Definition. Let G be a group and $H \subset G$ a subgroup. We say that H is a **normal subgroup** of G if for every $a \in G$ we have $aH = Ha$.

Proposition 2.4. *Let G be a group and $H \subset G$ a subgroup. The following statements are equivalent.*

(i) *H is a normal subgroup.*

(ii) *Every left H-coset is a right H-coset, and every right H-coset is a left H-coset.*

Proof. From the definition, we have (i) $\Rightarrow$ (ii). For (ii) $\Rightarrow$ (i), given $a \in G$ we have $a \in aH$ and $a \in Ha$, so we may conclude $aH = Ha$ from Proposition 2.1. □

Remark. The equality $aH = Ha$ of the definition of normal subgroup may be expressed in the equivalent form $aHa^{-1} = H$, i.e., H is invariant under conjugation. In fact, it is equivalent to require $aHa^{-1} \subset H$ for every $a \in G$, i.e., conjugation sends elements of H to elements of H.

The example above shows that $\{\mathrm{id}, (12)\}$ is not a normal subgroup of S_3. But for the subgroup $A_3 = \{\mathrm{id}, (123), (132)\}$, the left H'-cosets are

$$\{\mathrm{id}, (123), (132)\}, \qquad \{(12), (13), (23)\},$$

and these are right H'-cosets as well. So, A_3 is a normal subgroup of S_3.

Proposition 2.5. *Let G be a group and H a subgroup of G. If $[G : H] = 2$ then H is normal.*

Proof. The left H-cosets are H and the complement $G \smallsetminus H$. These are also the right H-cosets. So H is normal by Proposition 2.4. □

Definition. Let H be a normal subgroup of G. Then the **quotient group** of G by H is the set G/H with the multiplication $(aH)(bH) := abH$ for a, $b \in G$. The element aH may be denoted by $\overline{a}$, in which case the multiplication takes the form

$$\overline{a}\,\overline{b} := \overline{ab}. \tag{1}$$

For this to be valid we need to check that (1) is a well-defined map $G/H \times G/H \to G/H$ satisfying group axioms. Suppose a' belongs to the same H-coset as a and b' belongs to the same H-coset as b, i.e., $a' = ah$ and $b' = bk$ with h, $k \in H$. Then $a'b' = ahbk = ab(b^{-1}hb)k$. Since H is normal, we have $b^{-1}hb \in H$. So, $a'b'$ belongs to the same H-coset as ab. Associativity is immediate from (1), the class of the identity of G is the identity of G/H, and $\overline{a^{-1}}$ is inverse to $\overline{a}$.

Given a group G and normal subgroup H of G there is the **canonical homomorphism** $G \to G/H$, $a \mapsto \overline{a}$. The canonical homomorphism is surjective, with kernel equal to H. In fact the kernel of any group homomorphism is normal.

Proposition 2.6. *Let G and G' be groups and $f\colon G \to G'$ a group homomorphism. Then the kernel of f is a normal subgroup of G.*

Proof. Let H denote the kernel of f. We wish to verify that conjugation by an arbitrary element $a \in G$ sends elements of H to elements of H. For $h \in H$ we have

$$f(aha^{-1}) = f(a)f(h)f(a^{-1}) = f(a)f(a^{-1}),$$

which is the identity element of G'. So $aha^{-1} \in H$. □

Example. For $n \in \mathbb{N}$ the sign homomorphism $S_n \to \{\pm 1\}$ has kernel A_n, so A_n is a normal subgroup of S_n. (Since $[S_n : A_n] \leq 2$ this also follows from Proposition 2.5.)

Proposition 2.7. *Let G be a group. Then:*
(First Isomorphism Theorem) If G' is another group and $f\colon G \to G'$ is a surjective group homomorphism with kernel H, then the induced map $G/H \to G'$, $\overline{a} \mapsto f(a)$, is a group isomorphism.
(Second Isomorphism Theorem) If H and K are subgroups of G with K normal, then $H \cap K$ is a normal subgroup of H, the set HK consisting of all products hk with $h \in H$ and $k \in K$ is a subgroup of G, and we have a group isomorphism $H/(H \cap K) \cong HK/K$ given by $\overline{a} \mapsto \overline{a}$ for $a \in H$.
(Third Isomorphism Theorem) If H and K are normal subgroups of G with $K \subset H$, then H/K is a normal subgroup of G/K, and $(G/K)/(H/K) \cong G/H$, by $\overline{a}(H/K) \mapsto \overline{a}$ for $a \in G$.

Proof. In the First Isomorphism Theorem we have a well defined map, since if $a' = ah$ with $h \in H$, then $f(a') = f(a)f(h) = f(a)$. The map is a homomorphism since f is a homomorphism and is surjective since f is surjective. If $f(a)$ is the identity element $e_{G'}$ then $a \in H$, which means that $\overline{a}$ is the identity element of G/H.

For the Second Isomorphism Theorem, using that $hK = Kh$ for all $h \in H$ we see that HK is closed under multiplication and inverse. There is the homomorphism $H \to HK/K$, composite of the canonical homomorphism and the inclusion of the subgroup $H \subset HK$. This is surjective and has kernel equal to $H \cap K$, so we deduce the Second Isomorphism Theorem by applying the First Isomorphism Theorem.

For the Third Isomorphism Theorem, we have a homomorphism $G/K \to G/H$ given by $\overline{a} \mapsto \overline{a}$. This is surjective and has kernel equal to H/K. Again, we apply the First Isomorphism Theorem. □

In the situation of Proposition 2.6, let us suppose that G is finite. Then, by combining the First Isomorphism Theorem with Proposition 2.2, we obtain

$$|G| = |\operatorname{im}(f)| \cdot |\ker(f)|.$$

Proposition 2.8. *Let G and G' be groups, $f\colon G \to G'$ a group homomorphism, and H' a subgroup of G'. Then $f^{-1}(H')$ is a subgroup of G. Furthermore, if H' is normal then $f^{-1}(H')$ is normal.*

Proof. The axioms for $f^{-1}(H')$ to be a subgroup of G are readily verified. If H' is normal, then $f^{-1}(H')$ is the kernel of the composite

$$G \to G' \to G'/H'$$

of the corresponding canonical homomorphism and f, so $f^{-1}(H')$ is normal by Proposition 2.6. □

2.2 Direct and semidirect products

If H and K are groups, then their **product** $H \times K$, also called **direct product**, is the group whose elements are pairs (h, k) with $h \in H$ and $k \in K$, with

$$(h, k)(h', k') := (hh', kk')$$

for h, $h' \in H$ and k, $k' \in K$. There are inclusion and projection homomorphisms to/from $H \times K$. The group $H \times K$ with its pair of projection homomorphisms $H \times K \to H$ and $H \times K \to K$ is universal for groups with pairs of homomorphisms to H and to K.

For abelian groups the product may also be denoted with $\oplus$ (cf. $\mathbb{Z}$-modules, §1.4).

If H and K are groups and we are additionally given a group homomorphism $K \to \mathrm{Aut}(H)$, there is a group denoted by $H \rtimes K$, the **semidirect product** of H and K by the group homomorphism $K \to \mathrm{Aut}(H)$; here, $\mathrm{Aut}(H)$ denotes the group of automorphisms of H. If the group homomorphism is denoted by $k \mapsto \varphi_k$ then $H \rtimes K$ is the group whose elements are pairs (h, k) with $h \in H$ and $k \in K$, with

$$(h, k)(h', k') := (h\varphi_k(h'), kk')$$

for h, $h' \in H$ and k, $k' \in K$. There are inclusion homomorphisms from H and K, as well as a projection homomorphism to K whose kernel is identified with H by its inclusion homomorphism.

Example. We have $\mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$. There is a unique nontrivial homomorphism $\mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/3\mathbb{Z})$, and the corresponding semidirect product $\mathbb{Z}/3\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$ is not isomorphic to $\mathbb{Z}/6\mathbb{Z}$. Indeed, $\mathbb{Z}/3\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$ is a non-abelian group, in fact

$$\mathbb{Z}/3\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z} \cong S_3.$$

Proposition 2.9. *Let G be a group with identity element e_G, and let H and K be subgroups of G. Then:*

(i) *The map* $H \times K \to HK$, $(h,k) \mapsto hk$, *is bijective if and only if* $H \cap K = \{e_G\}$.
(i′) *If* G *is finite with* $|G| = |H| \cdot |K|$ *and* $H \cap K = \{e_G\}$, *then* $HK = G$.
(ii) *We have* $HK = KH$ *if and only if* HK *is a subgroup of* G.
(iii) *If the subgroups* H *and* K *are normal and* $H \cap K = \{e_G\}$, *then* HK *is a normal subgroup of* G *and the map* $H \times K \to HK$, $(h,k) \mapsto hk$, *is an isomorphism.*
(iii′) *If the subgroups* H *and* K *are normal,* $H \cap K = \{e_G\}$, *and* $HK = G$, *then* (iii) *gives an isomorphism* $H \times K \cong G$.
(iv) *If* $Hk = kH$ *for all* $k \in K$, *then* HK *is a subgroup of* G *with* H *a normal subgroup of* HK, *and* $k \mapsto \varphi_k$, $\varphi_k(h) := khk^{-1}$ *(*$h \in H$, $k \in K$*), defines a homomorphism* $K \to \mathrm{Aut}(H)$, *such that we have a surjective homomorphism from the corresponding semidirect product* $H \rtimes K$ *to* HK, *given by* $(h,k) \mapsto hk$, *which is an isomorphism if and only if* $H \cap K = \{e_G\}$.
(iv′) *If* $Hk = kH$ *for all* $k \in K$, $H \cap K = \{e_G\}$, *and* $HK = G$, *then the surjective homomorphism of* (iv) *is an isomorphipsm* $H \rtimes K \cong G$.

In the situation of (iii′) of Proposition 2.9, we may describe G as **direct product** of H and K. In the situation of (iv′) of Proposition 2.9, we may describe G as **semidirect product** of H and K.

Proof. For (i), if $h \in H \cap K$, $h \neq e_G$, then (e_G, e_G), $(h, h^{-1}) \in H \times K$ both map to e_G, while if $H \cap K = \{e_G\}$ and (h,k), $(h',k') \in H \times K$ satisfy $hk = h'k'$ then $h'^{-1}h = k'k^{-1} \in H \cap K$ implies $h = h'$ and $k = k'$. An immediate consequence is (i′). Given $HK = KH$ it follows quickly that HK is a subgroup (e.g., writing $kh' = \tilde{h}\tilde{k}$ for some $\tilde{h} \in H$ and $\tilde{k} \in K$ we have $hkh'k' = h\tilde{h}\tilde{k}k'$); (ii) follows. The key to (iii) (and hence also (iii′)) is that for $h \in H$ and $k \in K$ we have $hkh^{-1}k^{-1} \in H \cap K$. From (ii) we have, in (iv), that HK is a subgroup of G with H a normal subgroup of HK. The remainder of (iv) involves straightforward checking of axioms and a final application of (i); then an immediate consequence is (iv′). □

2.3 Isomorphism classes of small groups

The easiest groups to describe are the **cyclic groups**: G is cyclic if there exists $g \in G$, such that every element of G may be written as g^i for some $i \in \mathbb{Z}$. This means, $\mathbb{Z} \to G$, $i \mapsto g^i$, is a surjective homomorphism. Subgroups of $\mathbb{Z}$ are the same as ideals, so a subgroup of $\mathbb{Z}$ is of the form $a\mathbb{Z}$ for unique $a \in \mathbb{N}$. Then $\mathbb{Z}/a\mathbb{Z} \cong G$ (First Isomorphism Theorem), and we have two cases:

- **infinite cyclic group**, $a = 0$ with $\mathbb{Z} \cong G$;
- **finite cyclic group**, $a > 0$ with G finite of order a.

In both cases we refer to such an element g as a **generator** of G. In an arbitrary group G, for an element $g \in G$ the elements of the form g^i with $i \in \mathbb{Z}$ constitute a subgroup, called the subgroup **generated** by g and denoted by $\langle g \rangle$.

$\lvert G\rvert$	abelian	non-abelian
1	trivial	none
2	$\mathbb{Z}/2\mathbb{Z}$	none
3	$\mathbb{Z}/3\mathbb{Z}$	none
4	$\mathbb{Z}/4\mathbb{Z}$, V_4	none
5	$\mathbb{Z}/5\mathbb{Z}$	none
6	$\mathbb{Z}/6\mathbb{Z}$	S_3
7	$\mathbb{Z}/7\mathbb{Z}$	none
8	$\mathbb{Z}/8\mathbb{Z}$, $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$, $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	D_4, Q_8
9	$\mathbb{Z}/9\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$	none
10	$\mathbb{Z}/10\mathbb{Z}$	D_5

Table 1: Isomorphism types of finite groups of order at most 10

By Lagrange's theorem, if G is a finite group whose order $|G|$ is prime, then G is generated by any of its non-identity elements. This observation lets us immediately fill in four of the entries in the table of isomorphism types of finite groups of order at most 10. The finite abelian groups are classified by Theorem 1.17 (since every abelian group has a canonical structure of a $\mathbb{Z}$-module), which lets us fill in the left-hand column. It is an easy exercise, that if G is a group in which g^2 is equal to the identity element for every $g \in G$, then G is abelian. This lets us see that every group of order 4 is abelian, so there we have the cyclic group and one further unique group up to isomorphism, the **Klein four-group** $V_4 := \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$.

Just as in the first example of a nontrivial semidirect product $\mathbb{Z}/3\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z} \cong S_3$ there is more generally

$$D_n := \mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z},$$

semidirect product for $\mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$, $i \mapsto (\overline{m} \mapsto (-1)^i\overline{m})$. Called **dihedral group**, D_n has order $2n$ for all $n \in \mathbb{N}_{>0}$ and is non-abelian for $n \geq 3$, with $D_3 \cong S_3$.

The completeness of Table 1 thus comes down to the following assertions:

(i) Every non-abelian group of order 8 is isomorphic to D_4 or to the **quaternion group** Q_8 of elements $\{\pm 1, \pm i, \pm j, \pm k\}$ with $i^2 = j^2 = k^2 = -1$ and

$$ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j.$$

(ii) Every group of order 9 is abelian.

(iii) ($p \in \{3, 5\}$): every non-abelian group of order $2p$ is isomorphic to D_p.

For assertion (i), we know that there must be an element $h \in G$ with $H := \langle h \rangle$ of order 4. Suppose there exists k in G, not in H, such that k^2 is the identity

element e_G. Then $K := \langle k \rangle$ satisfies $|K| = 2$ and $H \cap K = \{e_G\}$. So $HK = G$ by Proposition 2.9 (i′). By Proposition 2.5, H is a normal subgroup. Now Proposition 2.9 (iv′) yields $H \rtimes K \cong G$, where the semidirect product is for some nontrivial homomorphism $K \to \mathrm{Aut}(H)$. But $K \cong \mathbb{Z}/2\mathbb{Z}$, $H \cong \mathbb{Z}/4\mathbb{Z}$, and there is a unique nontrivial homomorphism $\mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/4\mathbb{Z})$. So $G \cong D_4$. Notice that we may change 4 to any odd prime p throughout this argument to obtain Proposition 2.10, and in particular assertion (iii).

Proposition 2.10. *For every odd prime p there is a unique isomorphism class of non-abelian groups of order $2p$, that of the dihedral group D_p.*

We are left with treating the case in assertion (i) where we have $H := \langle h \rangle$ of order 4, such that $k^2 \neq e_G$ for every $k \in G \smallsetminus H$. In this case h^2 is the unique non-identity element whose square is identity. Since the condition "non-identity element whose square is identity" is invariant under conjugation, $\langle h^2 \rangle$ is a normal subgroup. We have $G/\langle h^2 \rangle$ of order 4, hence either cyclic or isomorphic to the Klein four-group. If cyclic then a generator $\overline{k}$ would lift to $k \in G$ with $\langle k \rangle$ of order 4, a contradiction since then k^2 would be in $G \smallsetminus H$, with $(k^2)^2 = e_G$. So $G/\langle h^2 \rangle \cong V_4$, and we have $k^2 \neq e_G$ for every lift k of a non-identity element $\overline{k} \in V_4$. Let us choose i, $j \in G$ whose images in V_4 are distinct non-identity elements and define $k := ij$. So $i^2 = j^2 = k^2 = h^2$. Now from $ij = k$ it follows that $kij = j^2$, hence $ki = j$. Now $jki = i^2$, hence $jk = i$. We have $kji = ij^2i = h^2i^2 = e_G$, from which we have $h^2ji = k^2ji = k$. Similarly, we have $h^2kj = i$ and $h^2ik = j$. So $G \cong Q_8$, with h^2 identified with -1.

To make arguments like these easier to follow, we will refer to an element $g \in G$ as having **order** n if the subgroup $\langle g \rangle$ generated by g has order n. We go on with assertion (ii): if $|G| = 9$ with h, $k \in G$ such that $kh \neq hk$, then in particular h and k must each have order 3, with $H := \langle h \rangle$ and $K := \langle k \rangle$ satisfying $H \cap K = \{e_G\}$. By Proposition 2.9 (i′), $HK = G = KH$. So $kh = h^a k^b$ for some a, $b \in \{0, 1, 2\}$ with $(a, b) \neq (1, 1)$. We will obtain a contradiction by excluding every (a, b).

If $a = 0$ or $b = 0$ then we always obtain an immediate contradiction (e.g., $hk = h^2$ would imply $k = h$). For the remaining cases: hk must have order 3, but

- $kh = hk^2$ would imply $(hk)^3 = h(kh)khk = h(hk^2)khk = k$;
- $kh = h^2k$ would imply $(hk)^3 = hkh(kh)k = hkh(h^2k)k = h$;
- $kh = h^2k^2$ would imply $(hk)^3 = h(h^2k^2)khk = hk$.

2.4 Actions of groups on sets

Classically, groups were symmetry groups (e.g., of geometric figures). The axiomatic formulation in terms of sets and maps represents a later development. The link between the classical and modern notions of group is the notion of group action.

Definition. Let G be a group and X a set. An **action** of G on X is a map

$$\begin{aligned} G \times X &\to X, \\ (g, x) &\mapsto g \cdot x, \end{aligned}$$

satisfying:

- $g \cdot (h \cdot x) = (gh) \cdot x$ for every g, $h \in G$ and $x \in X$,
- $e_G \cdot x = x$ for every $x \in X$,

where as usual e_G denotes the identity element of G.

To say that we have an action of G on X we may also say that X is a G**-set**.

A nice geometric example arises by letting the dihedral group D_n act on a regular n-gon in the plane (for any $n \geq 3$), where $\mathbb{Z}/n\mathbb{Z}$ acts by rotations with $\overline{m} \in \mathbb{Z}/m\mathbb{Z}$ acting by counterclockwise rotation by $2\pi m/n$, and $\mathbb{Z}/2\mathbb{Z}$ acts by reflection through a chosen axis of symmetry. In this example G is D_n, but there are several natural choices for X (the n-gon including or not including its interior, or just the set of vertices, or even the whole plane), and each choice leads to an example of group action.

More abstract examples include some actions of an arbitrary group G on itself (meaning that X is G in these examples):

- **left-multiplication**, $g \cdot a := ga$ for all g, $a \in G$;
- **inverse right-multiplication**, $g \cdot a := ag^{-1}$ for all g, $a \in G$;
- **conjugation**, $g \cdot a := gag^{-1}$ for all g, $a \in G$.

Definition. Given a group action $G \times X \to X$, $(g, x) \mapsto g \cdot x$, the **orbits** are the subsets of the form $G \cdot x = \{g \cdot x \,|\, g \in G\}$ for some $x \in X$.

For $x \in X$ we call $G \cdot x$ the orbit of x. The set of all orbits is denoted by $G\backslash X$.

Proposition 2.11. *Let $G \times X \to X$, $(g, x) \mapsto g \cdot x$, be a group action. Any two orbits are either equal to each other or disjoint, and the union of all orbits is X.*

Proof. Let x, $x' \in X$. If $G \cdot x \cap G \cdot x' \neq \emptyset$, then there exist g, $g' \in G$ such that $g \cdot x = g' \cdot x'$, and then $G \cdot x = G \cdot (g \cdot x) = G \cdot (g' \cdot x') = G \cdot x'$. Since $x = e_G \cdot x \in G \cdot x$, the union of all orbits is X. □

The orbits under the conjugation action are called **conjugacy classes**.

Example. For $n \in \mathbb{N}$ there is a bijection between the conjugacy classes of S_n and **partitions** of n, that is, the ways of writing n as a sum of positive integers. This associates to the conjugacy class of $\sigma \in S_n$ the cycle lengths of σ, which may be arranged in weakly decreasing order

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell, \qquad \lambda_1 + \cdots + \lambda_\ell = n, \tag{1}$$

to obtain the partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$. Given $\sigma' \in S_n$ conjugate to σ, then any $\tau \in S_n$ with $\sigma' = \tau\sigma\tau^{-1}$ determines a length-preserving bijection between the cycles of σ and the cycles of σ'. If σ and σ' may be written in cycle notation as

$$\sigma = (i_{11}, \ldots, i_{1\lambda_1})(i_{21}, \ldots, i_{2\lambda_2}) \ldots,$$
$$\sigma' = (i'_{11}, \ldots, i'_{1\lambda_1})(i'_{21}, \ldots, i'_{2\lambda_2}) \ldots,$$

then $\sigma' = \tau\sigma\tau^{-1}$ for $\tau \in S_n$ satisfying $\tau(i_{ab}) = i'_{ab}$ for all a, b.

This example extends to a description of the conjugacy classes of A_n ($n \geq 2$), as above, except that if the cycle lengths of $\sigma \in A_n$ form a **strict** partition consisting of odd **parts** (the λ_j are the parts of λ, and λ is said to be strict if the parts are pairwise distinct, i.e., if the inequalities in (1) are strict) then the set of permutations of cycle type λ is a union of two conjugacy classes in A_n. For instance, A_5 has 5 conjugacy classes: identity, 3-cycles, products of disjoint 2-cycles, and two conjugacy classes of 5-cycles. The respective cardinalities are 1, 20, 15, 12, 12.

For a general group G, a subgroup H of G is normal if and only if H is a union of conjugacy classes. For $G := A_5$, of order 60, if we compare the list of natural numbers dividing 60 with the cardinalities of unions of conjugacy classes containing the identity, we find that a normal subgroup of A_5 can only have order 1 or 60. A nontrivial group G, whose only normal subgroups are G and $\{e_G\}$, is said to be **simple**. So, beyond the easy example of groups of prime order, our first example of a simple group is A_5. A milestone of group theory is the Feit-Thompson theorem, that every non-cyclic finite simple group has even order.

Proposition 2.12. *The alternating group A_n is simple for every $n \geq 5$.*

Proof. We have already seen this for $n = 5$, so we may suppose $n \geq 6$. The set of products of pairs of disjoint 2-cycles

$$P := \{(i,j)(k,\ell) \in A_n \,|\, i,\ j,\ k,\ \ell \text{ pairwise distinct}\}$$

is a conjugacy class in A_n. Every element $\sigma \in A_n$ may be written as a product of elements of P, as we may deduce from an expression of σ as a product of transpositions by repeated application of $(i,j)(i',j') = (i,j)(k,\ell)(k,\ell)(i',j')$. It therefore suffices, given a normal subgroup H containing a non-identity element σ, to exhibit $\tau \in A_n$ such that $\tau\sigma\tau^{-1}\sigma^{-1} \in P$, for then $P \subset H$, and hence $H = A_n$.

If σ has a 2-cycle (i,j) then we may choose $k \notin \{i,j\}$ such that $\sigma(k) \neq k$, and then with $\tau := (i,j,k)$ we have $\tau\sigma\tau^{-1}\sigma^{-1} = (i,k)(j,\sigma(k))$. If σ has no 2-cycle but fixes some i, then we take j such that $\sigma(j) \neq j$, and with $\tau := (i,j,\sigma(j))$ we have $\tau\sigma\tau^{-1}\sigma^{-1} = (i,\sigma^2(j))(j,\sigma(j))$. In the remaining case, there exist i and j such that $j \notin \{\sigma^{-2}(i), \sigma^{-1}(i), i, \sigma(i), \sigma^2(i)\}$, and with $\tau := (i,j)(\sigma(i),\sigma(j))$ we have $\tau\sigma\tau^{-1}\sigma^{-1} = (i,j)(\sigma^2(i),\sigma^2(j))$. □

For the left- and inverse right-multiplication actions of a group on itself there is just one orbit, consisting of everything.

Definition. A group action is **transitive** if it has exactly one orbit.

There are a couple of equivalent ways to formulate the condition for a group action $G \times X \to X$, $(g,x) \mapsto g \cdot x$, to be transitive:

- There exists $x \in X$ such that $G \cdot x = X$;
- $X \neq \emptyset$, and we have $G \cdot x = X$ for every $x \in X$.

Definition. Let $G \times X \to X$, $(g,x) \mapsto g \cdot x$, be a group action. For $x \in X$ the set of elements $g \in G$ which satisfy $g \cdot x = x$ is a subgroup called the **stabilizer** of x and denoted by G_x.

The verification of the subgroup axioms for G_x is immediate (e.g., if $g \cdot x = x$ then $g^{-1} \cdot x = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = x$). To express the condition $g \cdot x = x$ we may say, "g fixes x".

Example. For the conjugation action of a group G on itself the stabilizer of an element $a \in G$ is the subgroup consisting of all elements of G that commuate with a; it is called the **centralizer** of a and denoted by $C(a)$.

The **center** of a group G is the subgroup of elements of G which commute with every element of G. It is denoted by $Z(G)$, or just Z. We have:

$$Z(G) = \bigcap_{a \in G} C(a).$$

Besides the actions of G on itself, above, there are actions of G on sets naturally associated with G, e.g., there are left-multiplication, inverse right-multiplication, and conjugation actions of G on the power set of G (the set of all subsets of G). A stabilizer for the conjugation action on the power set of G is called **normalizer**.

Definition. Let $G \times X \to X$, $(g,x) \mapsto g \cdot x$, and $G \times Y \to Y$, $(g,y) \mapsto g \cdot y$, be group actions. A map $f\colon X \to Y$ is said to be G**-equivariant**, or a G**-map**, if it satisfies

$$f(g \cdot x) = g \cdot f(x)$$

for every $g \in G$ and $x \in X$.

Example. Given an action $G \times X \to X$, $(g, x) \mapsto g \cdot x$, a subset $X' \subset X$ is said to be G**-invariant** if we have $g \cdot x' \in X'$ for every $g \in G$ and $x' \in X'$. This means, the original action $G \times X \to X$ restricts to an action $G \times X' \to X'$. In this case the inclusion map $X' \to X$ is a G-map.

If H is a subgroup of G, then we have this situation for the left-multiplication action on X equal to the power set of G, with $X' := G/H$ (since $g(aH) = (ga)H$ for $g \in G$ and $aH \in G/H$). The induced action of G on G/H is transitive (since the orbit of H is $\{aH \mid a \in G\} = G/H$).

For a general action $G \times X \to X$, $(g, x) \mapsto g \cdot x$, any orbit $G \cdot x$ is G-invariant.

Proposition 2.13. *Let $G \times X \to X$, $(g, x) \mapsto g \cdot x$, be a group action. Then:*
(i) *For $x \in X$ with stabilizer G_x and orbit $G \cdot x$ the map*

$$G \cdot x \to G/G_x,$$
$$g \cdot x \mapsto gG_x,$$

is well-defined and is a bijective G-map.
(ii) *If x, $x' \in X$ belong to the same orbit then the stabilizers G_x and $G_{x'}$ are conjugate to each other; specifically, for $g \in G$ such that $g \cdot x = x'$ we have $G_{x'} = gG_xg^{-1}$.*
(iii) *(orbit formula I) if G is finite then for $x \in X$ with stabilizer G_x and orbit $G \cdot x$ we have*

$$|G \cdot x| = [G : G_x].$$

(iv) *(orbit formula II) if G and X are finite then*

$$|X| = \sum_{G \cdot x \in G \backslash X} [G : G_x].$$

Proof. For (i), we have to check that for a, $a' \in G$ we have $a \cdot x = a' \cdot x$ if and only if a and a' belong to the same left G_x-coset. But $a \cdot x = a' \cdot x$ is equivalent to $a^{-1}a' \in G_x$, which is precisely the condition for a and a' to belong to the same G_x-coset. For (ii) we have, for $h \in G_x$,

$$(ghg^{-1}) \cdot x' = g \cdot (h \cdot x) = g \cdot x = x'.$$

This shows $gG_xg^{-1} \subset G_{x'}$. As well, $g^{-1}G_{x'}g \subset G_x$, and hence $G_{x'} = gG_xg^{-1}$. The formula in (iii) is an immediate consequence of (i). The sum in (iv) is well-defined by (ii), and by combining $|X| = \sum_{G \cdot x \in G \backslash X} |G \cdot x|$ with (iii) we obtain (iv). □

Applied to the conjugation action of a finite group G on itself, we obtain the **class equation**:

$$|G| = |Z(G)| + \sum_{i=1}^{n} [G : C(x_i)]$$

where $x_1, \ldots, x_n$ are representatives of the conjugacy classes of non-central elements of G, with each such conjugacy class represented by x_i for a unique i.

Proposition 2.14 (Cauchy's theorem)**.** *Let G be a finite group and p a prime. If p divides the order of G, then G contains an element of order p.*

Proof. We regard p as fixed and prove the statement for all G by induction on $|G|$. If G is abelian, then the result is clear (by the Structure Theorem or more simply by using a system of generators to express G as quotient of a product of cyclic subgroups), so we may assume that G is non-abelian. If the center Z has order divisible by p then we are done as well, so we may assume that $|Z|$ is not divisible by p. By the class equation some $C(x_i)$ has index not divisible by p and hence order divisible by p. Since $C(x_i)$ is a proper subgroup of G, we may conclude by applying the induction hypothesis. □

To illustrate the class equation and its application in Cauchy's theorem for the group A_5, we have the center, which is trivial, and the remaining conjugacy classes, which have representatives

$$x_1 := (123), \qquad x_2 := (12)(34), \qquad x_3 := (12345), \qquad x_4 := (12354).$$

The centralizers are

$$C(x_1) = \langle x_1 \rangle, \qquad C(x_2) = \{\mathrm{id}, (12)(34), (13)(24), (14)(23)\},$$
$$C(x_3) = \langle x_3 \rangle, \qquad C(x_4) = \langle x_4 \rangle,$$

of respective indices appearing in the class equation $60 = 1 + 20 + 15 + 12 + 12$. For each prime factor $p = 2$, 3, 5 of $|A_5| = 60$ some $C(x_i)$ has order divisible by p.

Proposition 2.15. *Let G be a group and X a set. There is a bijection between actions of G on X and group homomorphisms $G \to \mathrm{Perm}(X)$, sending a G-action $G \times X \to X$, $(g, x) \mapsto g \cdot x$, to $G \to \mathrm{Perm}(X)$, $g \mapsto (x \mapsto g \cdot x)$.*

Proof. We first check that $g \mapsto (x \mapsto g \cdot x)$ is a homomorphism $G \to \mathrm{Perm}(X)$, for any G-action $G \times X \to X$, $(g, x) \mapsto g \cdot x$. For $g \in G$ the map $X \to X$, $x \mapsto g \cdot x$, is bijective, since it has the inverse map $x \mapsto g^{-1} \cdot x$. So we have a map $G \to \mathrm{Perm}(X)$. For g, $g' \in G$ we have the composite $(x \mapsto g \cdot x) \circ (x \mapsto g' \cdot x)$, sending x to $g \cdot (g' \cdot x) = gg' \cdot x$, so the map $G \to \mathrm{Perm}(X)$ is a group homomorphism.

There is a map from the set of group homomorphisms $G \to \mathrm{Perm}(X)$ to the set of G-actions on X, which sends $G \to \mathrm{Perm}(X)$, $g \mapsto \varphi_g \in \mathrm{Perm}(X)$, to $G \times X \to X$, $(g, x) \mapsto \varphi_g(x)$. Notice that $(g, (h, x)) \mapsto \varphi_g(\varphi_h(x)) = \varphi_{gh}(x)$, equal to the image of (gh, x), and $(e_G, x) \mapsto \varphi_{e_G}(x) = x$, so indeed $G \times X \to X$, $(g, x) \mapsto \varphi_g(x)$, is a group action.

The maps, from G-actions on X to homomorphisms $G \to \mathrm{Perm}(X)$ and from homomorphisms $G \to \mathrm{Perm}(X)$ to G-actions on X, are inverse to each other, hence each one is bijective. □

We refer to the group homomorphism $G \to \mathrm{Perm}(X)$ obtained in Proposition 2.15 from a given group action as the associated **permutation representation** of G. The set of elements of G which fix every element of X is a normal subgroup of G, since it is the kernel of the associated permutation representation. We call the group action **faithful** if this is the trivial subgroup, i.e., if the associated permutation representation is injective.

By way of contrast, the stabilizers of elements of X need not be normal subgroups. A group action, for which all the stabilizers are trivial, is called **free**. A group action which is transitive and free is called **simply transitive**.

As application of permutation representation we give two results for finite groups.

Proposition 2.16 (Cayley's theorem)**.** *Every finite group is isomorphic to a subgroup of S_n for some $n \in \mathbb{N}$.*

Proof. We need only observe that a finite group G acts faithfully on itself by left multiplication. The associated permutation representation is therefore an injective homomorphism $G \to \mathrm{Perm}(G) \cong S_{|G|}$. □

The following is a result that, for finite groups, generalizes Proposition 2.5.

Proposition 2.17. *Let G be a finite group, and let H be a subgroup of G of prime index p. If p is the smallest prime factor of $|G|$ then H is normal.*

Proof. To the action by left-multiplication on G/H there is the associated homomorphism $G \to \mathrm{Perm}(G/H) \cong S_p$, which we denote by f. We have $|G| = |\mathrm{im}(f)| \cdot |\ker(f)|$ (§2.1). Since $|\mathrm{im}(f)|$ divides both $|G|$ and $p!$, we have only the possibilities that $\mathrm{im}(f)$ is trivial or of order p. But the action of G on G/H is not the trivial action, so $\mathrm{im}(f)$ cannot be trivial. Hence $|\mathrm{im}(f)| = p$, and $|\ker(f)| = |H|$. Clearly $\ker(f) \subset H$. So H is equal to $\ker(f)$ and is therefore a normal subgroup. □

Example. Proposition 2.17 has immediate applications to the structure of a group G whose order is the product pq of two primes. We suppose $p \leq q$. Let h be an element of G of order q (which exists, by Cauchy's theorem). Then $H := \langle h \rangle$ is normal by Proposition 2.17, and we have the canonical homomorphism $G \to G/H$. If G is not cyclic, then any lift $k \in G$ of a generator $\overline{k}$ of G/H must have order p. We may apply Proposition 2.9 (iv′) to conclude that G is isomorphic to a semidirect product $H \rtimes K$, with $K := \langle k \rangle$, for some homomorphism

$$\mathbb{Z}/p\mathbb{Z} \cong K \to \mathrm{Aut}(H) \cong \mathrm{Aut}(\mathbb{Z}/q\mathbb{Z}) \cong (\mathbb{Z}/q\mathbb{Z})^{\times}.$$

In particular, if $q - 1$ is not divisible by p (notice that this includes the case $p = q$ where the order of G is the square of a prime) then G is abelian.

2.5 The Sylow theorems

Generalizing Cauchy's theorem, we will show that any finite group G has a subgroup whose order is the highest power of p dividing $|G|$, for any prime p.

A **p-group** is a group whose order is a positive power of a prime p. A **p-subgroup** of a group G is a subgroup which is a p-group. A p-subgroup is said to be a **p-Sylow subgroup** if its order is the highest power of p dividing $|G|$.

Theorem 2.18. *Let G be a finite group and p a prime which divides the order of G. Let e be the largest integer, such that p^e divides $|G|$, and let $m := |G|/p^e$. Then:*
(i) *(First Sylow theorem) There exists a p-Sylow subgroup, i.e., a subgroup of G of order p^e.*
(ii) *(Second Sylow theorem) Any pair of p-Sylow subgroups are conjugate to each other, and any p-subgroup sits inside some p-Sylow subgroup.*
(iii) *(Third Sylow theorem) The number n_p of p-Sylow subgroups satisfies $n_p \mid m$ and $n_p \equiv 1 \mod p$.*

The proof of the First Sylow theorem is a modification of the argument of Cauchy's theorem and follows a similar inductive structure. An alternative, more direct proof, relying however on some facts from number theory, is presented in Appendix A.

Proof. We prove the First Sylow theorem for a given prime p by induction on the order of G. If the center Z of G has order divisible by p then we know that Z contains an element h of order p. If $e = 1$ then we are done, while if $e > 1$ then we may consider the canonical homomorphism to the quotient group

$$G \to G/\langle h \rangle$$

and apply the induction hypothesis to $G/\langle h \rangle$. The pre-image of a p-Sylow subgroup of $G/\langle h \rangle$ is a p-Sylow subgroup of G. It remains to treat the case that $|Z|$ is not divisible by p. Then we appeal to the class equation, where some $C(x_i)$ must have index not divisible by p and hence order divisible by p^e. We conclude by applying the induction hypothesis to $C(x_i)$; a p-Sylow subgroup of $C(x_i)$ is a p-Sylow subgroup of G.

We now establish the Second and Third Sylow theorems by considering the conjugation action of G on the set X of p-Sylow subgroups. Let H be a p-Sylow subgroup (which exists by the First Sylow theorem) and let X' be the orbit of H. Then $|X'|$ divides $|G|$. As well we may restrict the G-action on X' to an H-action on X'. Since H is a subgroup, the H-orbit of H is just $\{H\}$. But if $H' \in X'$, $H' \neq H$, then the H-orbit of H' must have more than one element. Indeed, if we had $H'h = hH'$ for every $h \in H$ then we could apply Proposition 2.9 (iv) to conclude that $H'H$ is a subgroup of G with a surjective homomorphism to $H'H$ from some semidirect

product $H' \rtimes H$. Then $H'H$ would be a p-subgroup of G of order bigger than p^e, a contradiction. So in the orbit formula $|X'| = \sum_{H'}[H : H_{H'}]$ (sum over orbit representatives for the H-action on X'), there is just one orbit $\{H\}$ whose cardinality is not divisible by p, and hence $|X'| \equiv 1 \bmod p$. Since $|X'|$ is not divisible by p and divides $|G|$, it follows that $|X'|$ divides m.

It remains only to show that $X' = X$ and that any p-subgroup sits inside some p-Sylow subgroup. Let K be a p-subgroup, and let us consider the K-action on X'. By the orbit formula, there must be some orbit of cardinality 1, say that of $H' \in X'$. Then $H'k = kH'$ for every $k \in K$, and by Proposition 2.9 (iv) as before we have that $H'K$ is a p-subgroup of G. This implies $K \subset H'$. In particular, if K is a p-Sylow subgroup, then $K = H'$, hence $K \in X'$. □

The constraints on the numbers n_p may be combined with the observation that $n_p = 1$ if and only if there is a normal p-Sylow subgroup to deduce results about the structure of groups of given order. For instance, if $|G| = pq^2$ where p and q are distinct primes with $p < q$, then we obtain

$$n_p \mid q^2, \qquad n_p \equiv 1 \bmod p.$$

If $q \not\equiv \pm 1 \bmod p$ then this forces $n_p = 1$, i.e., there is a unique p-Sylow subgroup H, which is normal. If K is a q-Sylow subgroup then K is normal by Proposition 2.17. Now $G \cong H \times K$ by Proposition 2.9 (iii′), hence G is abelian.

Definition. A group G is said to be **solvable** if there exists a finite chain of subgroups

$$\{e_G\} = G_0 \subsetneq G_1 \subset \cdots \subsetneq G_k = G$$

such that G_{j-1} is normal in G_j and G_j/G_{j-1} is abelian, for $j = 1, \ldots, k$.

Any finite group G may be represented as a chain

$$\{e_G\} = G_0 \subsetneq G_1 \subset \cdots \subsetneq G_k = G \tag{1}$$

such that G_{j-1} is normal in G_j and G_j/G_{j-1} is *simple*, for all j. Indeed, among chains of subgroups, each a proper normal subgroup the next, we select one of maximal possible length; then G_j/G_{j-1} is simple for all j, since the pre-image of any nontrivial proper normal subgroup of G_j/G_{j-1} under the canonical homomorphism $G_j \to G_j/G_{j-1}$ could be inserted to make the chain longer. In essence, the quotients G_j/G_{j-1} are the "prime factors" of G; concretely, if

$$\{e_G\} = G'_0 \subsetneq G'_1 \subset \cdots \subsetneq G'_{k'} = G$$

is another chain with simple quotients, then $k' = k$ and there exists a permutation $\sigma \in S_k$ such that G_j/G_{j-1} is isomorphic to $G'_{\sigma(j)}/G'_{\sigma(j)-1}$ for all j, a fact known as the Jordan-Hölder theorem. (It is a nice exercise to give a proof by induction

on $|G|$, by applying Proposition 2.9 (iii) to the subgroups G_{k-1} and $G'_{k'-1}$.) The integer k is called **composition length**, such a chain (1) a **composition series**, and the quotients G_j/G_{j-1}, **composition factors**.

For a finite group, being solvable is equivalent to having composition factors of prime order. For example,

- S_n is solvable for $n \leq 2$, trivially;
- S_3 has composition factors of order 2 and 3, hence is solvable;
- S_4 is solvable: the normal subgroup A_4 has, in turn, a normal subgroup of order 4;
- S_n for $n \geq 5$ has composition series $\{\mathrm{id}\} \subset A_n \subset S_n$, hence is not solvable.

2.6 Generators, relations, free groups

In §2.3 we introduced the subgroup $\langle g \rangle$ generated by an element of g of a group G. This is the set of powers of g, i.e., the image of the homomorphism $\mathbb{Z} \to G$, $i \mapsto g^i$. If we have several elements $g_1, \ldots, g_r$, or even an infinite collection of elements, then we can obtain a subgroup by combining these elements in all possible ways, by which we get not only expressions of the form $g_1^{i_1} \cdots g_r^{i_r}$ but also for instance (assuming $r \geq 2$) $g_1 g_2 g_1^{-1}$, as well as arbitrarily long expressions $g_1 g_2 g_1 \cdots g_2 g_1$.

Definition. Let G be a group and $(g_s)_{s \in S}$ a collection of elements of G, indexed by a set S. The subgroup **generated** by $(g_s)_{s \in S}$ is the smallest subgroup of G containing g_s for every $s \in S$; it is the intersection of all subgroups of G containing g_s for every $s \in S$, and is denoted by $\langle g_s \rangle_{s \in S}$. The case that $\langle g_s \rangle_{s \in S} = G$ is expressed by saying that G is generated by $(g_s)_{s \in S}$, or that $(g_s)_{s \in S}$ is a **system of generators**. We say G is **finitely generated** if there exists a finite system of generators.

The set of all subgroups containing g_s for every $s \in S$ is nonempty (since G contains g_s for every $s \in S$). So the intersection of all subgroups of G containing g_s for every $s \in S$ makes sense, and is contained in every such subgroup, hence is the smallest subgroup of G containing g_s for every $s \in S$.

If $S = \emptyset$ then the condition to contain g_s for every $s \in S$ is trivial, and the intersection of all subgroups is just the identity. So the subgroup generated by the empty set is trivial. If $|S| = 1$ then we recover the subgroup generated by one element as defined in §2.3: we have $g \in \{g^i \mid i \in \mathbb{Z}\}$, and if H is a subgroup of G containing g then $g^i \in H$ for every $i \in \mathbb{Z}$.

For subgroups generated by more than one element we would also like a description as the image of a group homomorphism. But images of products of $\mathbb{Z}$ do

not suffice. For example, the smallest non-abelian group S_3 has (many) systems of generators with two elements. But there is no surjective homomorphism $\mathbb{Z}\times\mathbb{Z} \to S_3$.

Definition. The **free group** on a set S is the group denoted by F_S whose elements are **words** in the alphabet $S \times \{\pm 1\}$ containing no pair of adjacent symbols $(s,1)$, $(s,-1)$, in either order, for $s \in S$. We write $(s,1)$ as s^1, or s, and $(s,-1)$ as s^{-1}. The product is defined by

$$(s_1^{\alpha_1} \cdots s_k^{\alpha_k})(t_1^{\beta_1} \cdots t_\ell^{\beta_\ell}) := s_1^{\alpha_1} \cdots s_{k-j}^{\alpha_{k-j}} t_{j+1}^{\beta_{j+1}} \cdots t_\ell^{\beta_\ell}, \tag{1}$$

where j is the largest natural number, less than or equal to $\min(k,\ell)$, such that

$$t_1^{-\beta_1} = s_k^{\alpha_k}, \qquad \ldots, \qquad t_j^{-\beta_j} = s_{k+1-j}^{\alpha_{k+1-j}}.$$

If $S = \{1, \ldots, n\}$ then F_S is also denoted by F_n.

A word is a sequence of symbols from some alphabet, of some length $\ell \in \mathbb{N}$. The **empty word** is the unique word of length 0. A single symbol may be viewed as a word of length 1. In particular, we may identify $S \times \{\pm 1\}$ with a subset of F_S. We identify S with $S \times \{1\}$ and hence as well with the corresponding subset of F_S.

Proposition 2.19. *For any set S the multiplication in* (1) *defines a group F_S. This group satisfies the universal property, that for every group G and map of sets $f\colon S \to G$ there is a unique group homomorphism $\varphi\colon F_S \to G$, such that $\varphi(s) = f(s)$ for every $s \in S$.*

Proof. The empty word satisfies the axiom of the identity element, and the product of the elements

$$t_1^{\beta_1} \cdots t_\ell^{\beta_\ell} \qquad \text{and} \qquad t_\ell^{-\beta_\ell} \cdots t_1^{-\beta_1},$$

in either order, is the empty word. Consider words $\mathbf{s} := s_1^{\alpha_1} \cdots s_k^{\alpha_k}$ and $\mathbf{t} := t_1^{\beta_1} \cdots t_\ell^{\beta_\ell}$ as in (1), and a third word

$$\mathbf{u} := u_1^{\gamma_1} \cdots u_m^{\gamma_m}.$$

We need to verify associativity $(\mathbf{st})\mathbf{u} = \mathbf{s}(\mathbf{tu})$. This is immediate when any of $\mathbf{s}$, $\mathbf{t}$, and $\mathbf{u}$ is the empty word, so let us suppose not and prove associativity by induction on k. For $k = 1$: if the largest $j \le \min(\ell, m)$ such that $u_1^{-\gamma_1} = t_\ell^{\beta_\ell}, \ldots, u_j^{-\gamma_j} = t_{\ell+1-j}^{\beta_{\ell+1-j}}$ satisfies $j < \ell$, then

$$(\mathbf{st})\mathbf{u} = \begin{cases} t_2^{\beta_2} \cdots t_{\ell-j}^{\beta_{\ell-j}} u_{j+1}^{\gamma_{j+1}} \cdots u_m^{\gamma_m}, & \text{if } t_1^{-\beta_1} = s_1^{\alpha_1}, \\ s_1^{\alpha_1} t_1^{\beta_1} \cdots t_{\ell-j}^{\beta_{\ell-j}} u_{j+1}^{\gamma_{j+1}} \cdots u_m^{\gamma_m}, & \text{otherwise,} \end{cases}$$

and this is equal to $\mathbf{s}(\mathbf{tu})$; if $j = \ell \le m$, then

$$\begin{aligned} (\mathbf{st})\mathbf{u} &= \begin{cases} u_{\ell+2}^{\gamma_{\ell+2}} \cdots u_m^{\gamma_m}, & \text{if } \ell < m \text{ and } u_{\ell+1}^{-\gamma_{\ell+1}} = s_1^{\alpha_1}, \\ s_1^{\alpha_1} u_{\ell+1}^{\gamma_{\ell+1}} \cdots u_m^{\gamma_m}, & \text{otherwise,} \end{cases} \\ &= \mathbf{s}(\mathbf{tu}). \end{aligned}$$

The inductive step $(\mathbf{st})\mathbf{u} = (s_1^{\alpha_1} \cdots s_{k-1}^{\alpha_{k-1}}(s_k^{\alpha_k}\mathbf{t}))\mathbf{u} = \cdots = \mathbf{s}(\mathbf{tu})$ uses the induction hypothesis four times.

Let a group G and a map of sets $f\colon S \to G$ be given. Then we define a map $\varphi\colon F_S \to G$ by sending the empty word to the identity element, s to $f(s)$ and s^{-1} to $f(s)^{-1}$ for all $s \in S$, and inductively on the length k of a word $\mathbf{s} = s_1^{\alpha_1} \cdots s_k^{\alpha_k}$, by

$$\varphi(\mathbf{s}) := f(s_1)^{\alpha_1}\varphi(s_2^{\alpha_2} \cdots s_k^{\alpha_k}). \tag{2}$$

We verify that φ is a homomorphism, i.e., $\varphi(\mathbf{st}) = \varphi(\mathbf{s})\varphi(\mathbf{t})$ where $\mathbf{s}$ and $\mathbf{t}$ are as above, by induction on k. The case $k = 0$ is trivial. For the inductive step, we take j to be as in definition of the product (1). If $j < k$, then

$$\begin{aligned}
\varphi(\mathbf{st}) &= \varphi(s_1^{\alpha_1} \cdots s_{k-j}^{\alpha_{k-j}} t_{j+1}^{\beta_{j+1}} \cdots t_\ell^{\beta_\ell}) \\
&= f(s_1)^{\alpha_1}\varphi(s_2^{\alpha_2} \cdots s_{k-j}^{\alpha_{k-j}} t_{j+1}^{\beta_{j+1}} \cdots t_\ell^{\beta_\ell}) \\
&= f(s_1)^{\alpha_1}\varphi(s_2^{\alpha_2} \cdots s_k^{\alpha_k}\mathbf{t}) \\
&= f(s_1)^{\alpha_1}\varphi(s_2^{\alpha_2} \cdots s_k^{\alpha_k})\varphi(\mathbf{t}) \\
&= \varphi(\mathbf{s})\varphi(\mathbf{t}),
\end{aligned}$$

while if $j = k$ then we have $\varphi(\mathbf{st}) = \varphi(t_{k+1}^{\beta_{k+1}} \cdots t_\ell^{\beta_\ell}) = f(s_1)^{\alpha_1}\varphi(t_k^{\beta_k} \cdots t_\ell^{\beta_\ell})$ by (2) and the fact that $t_k^{\beta_k} = s_1^{-\alpha_1}$, and we conclude as before. So we have exhibited a homomorphism φ. Uniqueness follows by an inductive argument, using (2): a homomorphism which sends s to $f(s)$ must send s^{-1} to $f(s)^{-1}$, and now the property to be a homomorphism forces the value on words of length k to be as in (2). □

The most common free groups that we will work with are the groups F_n for $n \in \mathbb{N}$. In this case it is usual to denote $(j, 1)$ by x_j and $(j, -1)$ by x_j^{-1}. Also, in F_n or F_S we will, as in any group, use g^m to denote the m-fold iterated product of a group element g and g^{-m} to denote the m-fold iterated product of g^{-1} for any positive integer m, as is the usual convention. (And g^0 is just the identity element.)

Example. The free group F_1 is isomorphic to $\mathbb{Z}$. Indeed, a word must be empty, consist of m copies of x_1 (which we would also denote by x_1^m), or consist of m copies of x_1^{-1} (which we would also denote by x_1^{-m}), for a positive integer m. Identifying these with 0, respectively m, respectively $-m$, we obtain $F_1 \cong \mathbb{Z}$.

The groups F_n for $n \geq 2$ are non-abelian, as are F_S for all sets S which are not empty or of cardinality 1.

Proposition 2.20. *Let G be a group, $(g_s)_{s\in S}$ a collection of elements, and $\varphi\colon F_S \to G$ the homomorphism which by the universal property in Proposition 2.19 corresponds to the map $S \to G$, $s \mapsto g_s$. Then*

$$\langle g_s \rangle_{s\in S} = \mathrm{im}(\varphi).$$

In particular, $(g_s)_{s\in S}$ is a system of generators if and only if φ is surjective.

Proof. We have $\varphi(s) = g_s$, so $\mathrm{im}(\varphi)$ contains g_s for every $s \in S$. It remains therefore to show that any subgroup H of G containing g_s for every $s \in S$ must contain the image of φ. Given such H, Proposition 2.19 determines a homomorphism $\psi\colon F_S \to H$ sending $s \in S$ to g_s, and the composition of the inclusion $H \hookrightarrow G$ with ψ is equal to φ. So $\mathrm{im}(\varphi) \subset H$. □

Example. The group F_S is generated by S. Indeed, the identity homomorphism $F_S \to F_S$ sends s to s for every $s \in S$, and the identity homomorphism is surjective.

The kernel of $F_S \to G$ can even in simple situations be quite difficult to manage. For example, for $n \in \mathbb{N}_{>0}$ the dihedral group D_n is generated by 2 elements, an element g corresponding to $(1, 0)$ in the definition of D_n as semidirect product $\mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$, and an element h corresponding to $(0, 1)$, and we know from experience that the information needed to compute in D_n is that g has order n, h has order 2, and we have

$$hg = g^{n-1}h.$$

But the kernel of the homomorphism $F_2 \to D_n$ sending x_1 to g and x_2 to h increases in complexity as n increases. Even in the seemingly simple case $n = 2$ (the Klein four-group) it is a nontrivial exercise to show that the kernel cannot be generated by fewer than 5 elements.

A manageable theory of presentations of groups is possible thanks to a construction known as the normal closure.

Definition. Let F be a group. The **normal closure** of a subset $R \subset F$ is the smallest normal subgroup of F containing R; it is the intersection of all normal subgroups of F containing R and denoted by $\langle R \rangle^F$.

Example. We have

$$F_2/\langle x_1^n, x_2^2, x_1x_2x_1x_2\rangle^{F_2} \cong D_n.$$

We take $F_2 \to D_n$ as above, sending x_1 to g and x_2 to h, and define K to be the kernel. Since K is normal and contains x_1^n, x_2^2, and $x_1x_2x_1x_2 \in K$ we have $\langle x_1^n, x_2^2, x_1x_2x_1x_2\rangle^{F_2} \subset K$. So it suffices to show that in $F_2/\langle x_1^n, x_2^2, x_1x_2x_1x_2\rangle^{F_2}$ the set

$$S := \{\overline{e}_{F_2}, \overline{x}_1, \ldots, \overline{x}_1^{n-1}, \overline{x}_2, \overline{x}_1\overline{x}_2, \ldots, \overline{x}_1^{n-1}\overline{x}_2\}.$$

is closed under multiplication, say on the right, by $\overline{x}_1^{\pm 1}$ and $\overline{x}_2^{\pm 1}$. We have $\overline{x}_1^{-1} = \overline{x}_1^{n-1}$ and $\overline{x}_2^{-1} = \overline{x}_2$, so it suffices to consider just right multiplication by $\overline{x}_1$ and $\overline{x}_2$. For the powers of $\overline{x}_1$ in S, this is in every case immediate or a consequence of $\overline{x}_1^n = \overline{e}_{F_2}$. By combining all three relations $\overline{x}_1^n = \overline{e}_{F_2}$, $\overline{x}_2^2 = \overline{e}_{F_2}$, and $\overline{x}_1\overline{x}_2\overline{x}_1\overline{x}_2 = \overline{e}_{F_2}$, we find

$$\overline{x}_2\overline{x}_1 = \overline{x}_1^{n-1}\overline{x}_2.$$

Finally, $\overline{x}_1^k\overline{x}_2\overline{x}_1 = \overline{x}_1^{k-1}\overline{x}_2$ and $\overline{x}_1^k\overline{x}_2^2 = \overline{x}_1^k$ for $k = 1, \ldots, n-1$.

Definition. We say that a group G is **presented** by generators $(g_s)_{s\in S}$ and relations $R \subset F_S$ if

- $(g_s)_{s\in S}$ is a system of generators for G,
- the kernel of the associated surjective homomorphism $F_S \to G$ is equal to $\langle R\rangle^{F_S}$.

We say G is **finitely presented** if there exists a presentation with R and S finite.

Example. Every finite group is finitely presented. Say G is finite, generated by g_1, …, g_n. We write every $g \in G$ in terms of g_1, …, g_n, thereby obtaining $\mathbf{s}_g \in F_n$ representing g. We form R with a relation that encodes every product gg_j:

$$R := \{\mathbf{s}_g x_j \mathbf{s}_{gg_j}^{-1} \,|\, g \in G,\, 1 \le j \le n\}. \tag{3}$$

Then G is presented by $(g_1, \dots, g_n)$ and R. Indeed, iterating (3) yields

$$x_j^{m_j} \in \langle R\rangle^{F_n}$$

for every j, where m_j is the order of g_j. The rest of the argument is just as in the previous example.

A group G may be defined by generators and relations, by setting

$$G := F_S/\langle R\rangle^{F_S}.$$

We restrict ourselves to the case R and S are finite. Even then it is difficult to determine, e.g., whether G is finite. (Any text on *combinatorial group theory* may be consulted for more about this, including a mathematically precise formulation of "difficult".) We describe an algorithm for transforming a given set of relations; if it brings them into the form (3) then we will have identified G. For simplicity we will assume that the relations are words in x_1, …, x_n (no negative exponents), each of length ≥ 2. (A relation of the form x_i means that x_i can simply be omitted from the set of generators.) If $R = \emptyset$ we have a free group; we henceforth assume $R \ne \emptyset$.

The algorithm works with three tables. In Table 1 we keep a list of expressions $\mathbf{s}_1$, $\mathbf{s}_2$, … (always finitely many), each an element of F_n, in fact a word in which only the symbols x_1, …, x_n appear (no negative exponents). If Table 1 has m rows then Table 2 will be an $m \times n$ matrix (a_{ij}) whose entries are positive integers, or blank; we will read Table 2 as a set of relations

$$R_2 := \{\mathbf{s}_i x_j \mathbf{s}_{a_{ij}}^{-1} \,|\, 1 \le i \le m,\, 1 \le j \le n,\, a_{ij} \in \mathbb{N}_{>0}\}$$

Table 3 will consist of a finite list of triples $(i, \mathbf{u}, j)$ where i and j are integers and $\mathbf{u} \in F_n$ is a word in the symbols x_1, …, x_n. An entry $(i, \mathbf{u}, j)$ from Table 3 will be interpreted, and written, as a relation in G:

$$\overline{\mathbf{s}}_i \overline{\mathbf{u}} = \overline{\mathbf{s}}_j.$$

At the start of the algorithm, Table 1 has a single entry, the empty word, Table 2 is empty, and Table 3 consists of the relations $\overline{\mathbf{s}}_1\overline{\mathbf{r}} = \overline{\mathbf{s}}_1$ for every $\mathbf{r} \in R$:

Table 1	Table 2	Table 3
$\mathbf{s}_1 := e_{F_n}$		$\overline{\mathbf{s}}_1\overline{\mathbf{r}} = \overline{\mathbf{s}}_1 \ (\forall\, \mathbf{r} \in R)$

The **Todd-Coxeter algorithm**, in a simplified form adapted from the presention in M. Artin's text *Algebra*, starts with Tables 1, 2, and 3 as above. The first task is to iterate the following three steps until either the algorithm terminates with failure or iteration leaves the tables unchanged:

- Use the entries from Table 2 to **reduce** entries from Table 3: for every (i, j) entry $a = a_{ij} \in \mathbb{N}_{>0}$ in Table 2 (meaning, a relation $\overline{\mathbf{s}}_i\overline{x}_j = \overline{\mathbf{s}}_a$) we make replacements in Table 3,

$$\begin{aligned} \overline{\mathbf{s}}_i(\overline{x}_j\overline{\mathbf{w}}) &= \overline{\mathbf{s}}_k \qquad \text{with} \qquad \overline{\mathbf{s}}_a\overline{\mathbf{w}} = \overline{\mathbf{s}}_k, \\ \overline{\mathbf{s}}_\ell(\overline{\mathbf{w}}\,\overline{x}_j) &= \overline{\mathbf{s}}_a \qquad \text{with} \qquad \overline{\mathbf{s}}_\ell\overline{\mathbf{w}} = \overline{\mathbf{s}}_i, \end{aligned}$$

  ($\mathbf{w}$ denotes an arbitrary word) until no more replacements can be made.

- Decide whether Table 3 is **self-consistent**. Table 3 is self-consistent if it has no entry of the form $\overline{\mathbf{s}}_a = \overline{\mathbf{s}}_b$ in Table 3 with distinct a and b and has no pair of entries $\overline{\mathbf{s}}_i\overline{x}_j = \overline{\mathbf{s}}_a$ and $\overline{\mathbf{s}}_i\overline{x}_j = \overline{\mathbf{s}}_b$ with distinct a and b.

- If Table 3 is self-consistent then every relation $\overline{\mathbf{s}}_i\overline{x}_j = \overline{\mathbf{s}}_a$ in Table 3 is **moved** to Table 2, as (i, j) entry a; we also delete trivial relations $\overline{\mathbf{s}}_b = \overline{\mathbf{s}}_b$ and any repeated copies of a relation. If Table 3 is not self-consistent then the algorithm **terminates with failure**.

At this point the algorithm **terminates with success** if Table 2 is complete, in the sense of having no blank entries.

If Table 2 still has blank entries, then the next task is to **choose** i and j such that the (i, j) entry of Table 2 is blank and make modifications to the tables. Let m be the number of entries in Table 1. Then we add $\mathbf{s}_{m+1} := \mathbf{s}_i x_j$ to the bottom of Table 1, set the (i, j) entry of Table 2 equal to $m + 1$, and add $\mathbf{s}_{m+1}\mathbf{r} = \mathbf{s}_{m+1}$ to Table 3 for all $\mathbf{r} \in R$. The algorithm continues by returning to the first task, above.

We make some comments on the algorithm. For the first task (iteration of the three steps) there is a quantity, the number of symbols in Table 3, that can only decrease or remain unchanged with each iteration, and the quantity remains unchanged if and only if the tables remain unchanged. Thus we have an easy *a priori* bound on the number of iterations needed to complete the first task.

The second task starts with a choice. One could for instance make the choice each time randomly. It is also possible to devise other schemes for making the choice. So

the algorithm is really a family of algorithms, one for every variation on how the choice is made.

At the conclusion of the second task we go back to the first task. Tables 1 and 2 have gotten longer, and new entries have been added to Table 3. There is no apparent bound on the number of times that algorithm will return to the first task. We do not know how long we have to wait for the algorithm to terminate, or even if it will ever terminate. Even if the algorithm terminates, in case of termination with failure we have learned nothing about G.

If the algorithm terminates with success, let us say with m entries in Table 1, then we have

$$|G| = m \qquad \text{and} \qquad G = \{\overline{\mathbf{s}}_1, \ldots, \overline{\mathbf{s}}_m\}. \tag{4}$$

To see this, we note that the normal closure of the relations in Tables 2 and 3 remains unchanged over the course of the algorithm, and that at termination with success Table 3 is empty, i.e.,

$$\langle R_2 \rangle^{F_n} = \langle R \rangle^{F_n}.$$

We define a homomorphism

$$\psi \colon F_n \to S_m$$

by sending x_j to the permutation σ_j defined by $\sigma_j(a_{ij}) := i$ for $i = 1, \ldots, m$. For $\mathbf{r} \in R$, we obtain from $\mathbf{s}_i \mathbf{r} \mathbf{s}_i^{-1} \in \langle R \rangle^{F_n}$ for all i that $\mathbf{r} \in \ker(\psi)$. Hence

$$\langle R \rangle^{F_n} \subset \ker(\psi).$$

By construction, ψ is the permutation representation of a transitive group action, hence the image of ψ has order at least m. But now we recognize R_2 as having the form (3), so as before we are able to conclude (4).

Example. Suppose that $n = 2$ and $R = \{x_1^2, x_1 x_2^2\}$. Then we start with

Table 1	Table 2	Table 3
$\mathbf{s}_1 := e_{F_2}$		$\overline{\mathbf{s}}_1 \overline{x}_1^2 = \overline{\mathbf{s}}_1$
		$\overline{\mathbf{s}}_1 \overline{x}_1 \overline{x}_2^2 = \overline{\mathbf{s}}_1$

The first task does nothing, so we are at the point of having to make a choice. Of course i must be 1; let us choose $j = 1$. This means, we add $\mathbf{s}_2 := x_1$ to Table 1, set the $(1,1)$ entry of Table 2 equal to 2, and add relations $\overline{\mathbf{s}}_2 \overline{x}_1^2 = \overline{\mathbf{s}}_2$ and $\overline{\mathbf{s}}_2 \overline{x}_1 \overline{x}_2^2 = \overline{\mathbf{s}}_2$ to Table 3. At this point the entry in Table 2 makes reduction possible, as we indicate with squiggly arrows:

Table 1	Table 2	Table 3		
$\mathbf{s}_1 := e_{F_2}$	2	$\overline{\mathbf{s}}_1 \overline{x}_1^2 = \overline{\mathbf{s}}_1$	$\rightsquigarrow$	$\overline{\mathbf{s}}_2 \overline{x}_1 = \overline{\mathbf{s}}_1$
$\mathbf{s}_2 := x_1$		$\overline{\mathbf{s}}_1 \overline{x}_1 \overline{x}_2^2 = \overline{\mathbf{s}}_1$	$\rightsquigarrow$	$\overline{\mathbf{s}}_2 \overline{x}_2^2 = \overline{\mathbf{s}}_1$
		$\overline{\mathbf{s}}_2 \overline{x}_1^2 = \overline{\mathbf{s}}_2$	$\rightsquigarrow$	$\overline{\mathbf{s}}_2 \overline{x}_1 = \overline{\mathbf{s}}_1$
		$\overline{\mathbf{s}}_2 \overline{x}_1 \overline{x}_2^2 = \overline{\mathbf{s}}_2$		

Table 3 is self-consistent. We move $\overline{\mathbf{s}}_2\overline{x}_1 = \overline{\mathbf{s}}_1$ from Table 3 (where it occurs twice) to new $(2,1)$ entry 1 in Table 2 and perform further reductions until we reach

Table 1	Table 2	Table 3
$\mathbf{s}_1 := e_{F_2}$	2	$\overline{\mathbf{s}}_2\overline{x}_2^2 = \overline{\mathbf{s}}_1$
$\mathbf{s}_2 := x_1$	1	$\overline{\mathbf{s}}_1\overline{x}_2^2 = \overline{\mathbf{s}}_2$

and which point no further reductions are possible. So we choose, again, one of the blank entries in Table 2, let us say the $(1,2)$ entry. We add $\mathbf{s}_3 := x_2$ to Table 1, set the $(1,2)$ entry of Table 2 to 3, and add $\overline{\mathbf{s}}_3\overline{x}_1^2 = \overline{\mathbf{s}}_3$ and $\overline{\mathbf{s}}_3\overline{x}_1\overline{x}_2^2 = \overline{\mathbf{s}}_3$ to Table 3. Two reductions are possible:

Table 1	Table 2	Table 3		
$\mathbf{s}_1 := e_{F_2}$	2 3	$\overline{\mathbf{s}}_2\overline{x}_2^2 = \overline{\mathbf{s}}_1$		
$\mathbf{s}_2 := x_1$	1	$\overline{\mathbf{s}}_1\overline{x}_2^2 = \overline{\mathbf{s}}_2$	$\rightsquigarrow$	$\overline{\mathbf{s}}_3\overline{x}_2 = \overline{\mathbf{s}}_2$
$\mathbf{s}_3 := x_2$		$\overline{\mathbf{s}}_3\overline{x}_1^2 = \overline{\mathbf{s}}_3$		
		$\overline{\mathbf{s}}_3\overline{x}_1\overline{x}_2^2 = \overline{\mathbf{s}}_3$	$\rightsquigarrow$	$\overline{\mathbf{s}}_3\overline{x}_1\overline{x}_2 = \overline{\mathbf{s}}_1$

One relation is moved from Table 3 to Table 2, then there are no further reductions.

We choose again, say, the $(2,2)$ entry and leave it to the reader to make the corresponding additions to the tables and carry out the reductions leading to the following, with Table 2 complete and Table 3 empty:

Table 1	Table 2	Table 3
$\mathbf{s}_1 := e_{F_2}$	2 3	
$\mathbf{s}_2 := x_1$	1 4	
$\mathbf{s}_3 := x_2$	4 2	
$\mathbf{s}_4 := x_1x_2$	3 1	

The algorithm terminates with success, and we identify G as being cyclic of order 4.

2.7 Matrix groups

Groups such as $GL_n(K)$, where K is a field, play an important role in Linear Algebra, especially when K is the field of real or complex numbers. Our treatment of matrix groups will be limited to a description of groups of 2×2 and 3×3 orthogonal matrices over $\mathbb{R}$ and a study of their finite subgroups.

The groups $SO(2)$ and $O(2)$ were described in detail in Linear Algebra:

$$SO(2) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \middle| \, \theta \in \mathbb{R} \right\},$$
$$O(2) = SO(2) \cup \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \middle| \, \theta \in \mathbb{R} \right\}.$$

By the usual action of matrices on vectors we have an action of $SO(2)$ on $\mathbb{R}^2$, which restricts to a simply transitive action on the unit circle S^1. This means, once we have chosen a base point on S^1 – a canonical choice is $(1,0)$ – we have an identification

$$SO(2) \cong S^1$$

which is not only a bijective map of sets but also a continuous map (viewing $SO(2)$ as a subset of the space of 2×2 matrices and S^1 as a subset of $\mathbb{R}^2$) and hence a homeomorphism (a bijective continuous map with continuous inverse). Indeed, $SO(2)$ and S^1 are compact, and a bijective continuous map between compact Hausdorff spaces has continuous inverse.

The determinant map

$$O(2) \xrightarrow{\det} \{\pm 1\}$$

is surjective with kernel $SO(2)$. The elements that map to -1 are reflections through lines through the origin. This is the starting point for the analysis of finite subgroups of $O(2)$.

Proposition 2.21. *For a finite subgroup $G \subset O(2)$, we have one of the following:*

(i) *$G \subset SO(2)$ and G is generated by rotation by $2\pi/n$ with $n := |G|$, so G is cyclic of order n, or*

(ii) *$G \not\subset SO(2)$ and G is generated by rotation by $2\pi/n$ with $n := (1/2)|G|$ and a reflection, so G is isomorphic to D_n.*

Proof. For each n the elements $g \in SO(2)$ satisfying $g^n = E_2$ are precisely rotation by multiples of $2\pi/n$, and there are exactly n such rotations forming a cyclic group generated by rotation by $2\pi/n$. If $G \subset SO(2)$, then with $n := |G|$ we have G contained in, and hence since G has n elements, equal to the subgroup generated by rotation by $2\pi/n$. If $G \not\subset SO(2)$ then we define $G_0 := G \cap SO(2)$, the kernel of the surjective determinant homomorphism from G to $\{\pm 1\}$. So, $|G_0| = (1/2)|G|$. By (i), G_0 is generated by rotation by $2\pi/n$, where $n := (1/2)|G|$. Any element of $G \smallsetminus G_0$ is a reflection and together with G_0 generates G. □

For simplicity we restrict our analysis in the case of orthogonal 3×3 matrices to the finite subgroups of $SO(3)$. Every non-identity element of $SO(3)$ is a rotation around a well-defined rotation axis through the origin.

Theorem 2.22. *The nontrivial finite subgroups of $SO(3)$ are:*

(i) *cyclic groups, generated by a rotation by $2\pi/n$ for some $n \geq 2$,*

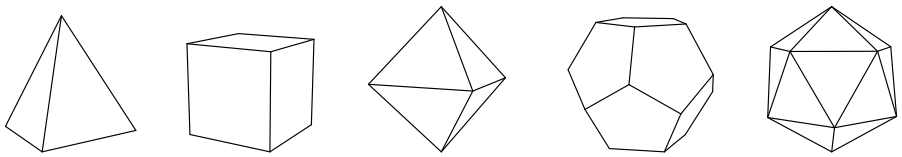
(ii) *dihedral groups, generated by a rotation by $2\pi/n$ for some $n \geq 2$ and a rotation by π, the two generators having orthogonal axes of rotation,*

(iii) *tetrahedral rotation group* $\cong A_4$,

(iv) *octahedral rotation group* $\cong S_4$,

(v) *icosahedral rotation group* $\cong A_5$.

Cases (iii)–(v) refer to symmetry groups of Platonic solids. The Platonic solids:



By a duality symmetry (the tetrahedron is self-dual), pairs of Platonic solids have the same symmetry group. Theorem 2.22 confirms that there are no more regular convex polyhedra than the tetrahedron, cube, octahedron, dodecahedron, and icosahedron (displayed here).

Proof of Theorem 2.22. Given a nontrivial finite subgroup $G \subset SO(3)$ we let P denote the polar set, consisting of all points p on the unit sphere in $\mathbb{R}^3$ such that $\mathbb{R} \cdot p$ is the axis of rotation of some non-identity element of G. There is the usual action of $SO(3)$ on $\mathbb{R}^3$, which restricts to an action of G. We claim, this action sends P to itself, so we get an action of G on P. Indeed, if $p \in P$, meaning for some $E_3 \neq g \in G$ we have $g \cdot p = p$, then for $h \in G$ we have $hgh^{-1}(h \cdot p) = h \cdot p$, so $h \cdot p \in P$.

For $p \in P$ we let $n_p := |G_p|$, so the stabilizer G_p is generated by rotation by $2\pi/n_p$ around $\mathbb{R} \cdot p$. The cardinality of the orbit $m_p := |G \cdot p|$ is given by

$$m_p = \frac{|G|}{n_p}.$$

We have projection maps

$$\begin{array}{ccc} \{(g,p) \in (G \smallsetminus \{E_3\}) \times P \,|\, g \cdot p = p\} & \longrightarrow & P \\ \Big\downarrow{\scriptstyle \text{2-to-1}} & & \\ G \smallsetminus \{E_3\} & & \end{array}$$

As indicated, to every element of $G \smallsetminus \{E_3\}$ there correspond two elements of P, which are the two intersection points of the axis of rotation with the unit sphere. The horizontal map has fibers $G_p \smallsetminus \{E_3\}$. So,

$$2|G| - 2 = \sum_{p \in P} (n_p - 1).$$

We let r denote the number of orbits and choose representatives $p_1, \ldots, p_r$ of the orbits. Then,

$$2|G| - 2 = \sum_{i=1}^{r} (n_{p_i} - 1) m_{p_i}.$$

Dividing by $|G|$,

$$2\big(1 - \frac{1}{|G|}\big) = \sum_{i=1}^{r} \big(1 - \frac{1}{n_{p_i}}\big).$$

The left-hand side is greater than or equal to 1 and strictly less than 2, and the right-hand side is a sum of r numbers, each greater than or equal to $1/2$ and less than 1. So we have $r = 2$ or $r = 3$.

If $r = 2$ then we obtain $2 = m_{p_1} + m_{p_2}$, so $m_{p_1} = m_{p_2} = 1$ and G is cyclic.

Otherwise $r = 3$, and we have

$$1 + \frac{2}{|G|} = \frac{1}{n_{p_1}} + \frac{1}{n_{p_2}} + \frac{1}{n_{p_3}}.$$

We may suppose $n_{p_1} \leq n_{p_2} \leq n_{p_3}$. Then for the sum of reciprocols to be greater than 1 we require $n_{p_1} = 2$, and $n_{p_2} \in \{2, 3\}$. If $n_{p_2} = 2$ then $m_{p_3} = 2$ and $G \cong D_n$ with $n = n_{p_3}$. If $n_{p_2} = 3$ then there are three possibilities:

$$n_{p_3} = 3,\ |G| = 12, \qquad n_{p_3} = 4,\ |G| = 24, \qquad n_{p_3} = 5,\ |G| = 60.$$

These three cases leads to tetrahedral, octahedral, and icosahedral rotation groups, respectively. □

Finally, we obtain a description of $SO(3)$ by relating it to the special unitary group $SU(2)$. In analogy with $SO(2) \cong S^1$ as we saw above, we have a simply transitive action of $SU(2)$ on the unit sphere in $\mathbb{C}^2 \cong \mathbb{R}^4$, i.e., the unit 3-sphere, inducing a homeomorphism $SU(2) \cong S^3$.

Proposition 2.23. *The conjugation action of $SU(2)$ on trace-zero skew-Hermitian 2×2 complex matrices induces an isomorphism $SU(2)/\{\pm E_2\} \cong SO(3)$.*

Proof. This can be seen by a computation in coordinates:

$$\begin{pmatrix} p & -\bar{q} \\ q & \bar{p} \end{pmatrix} \begin{pmatrix} x & -\bar{z} \\ z & \bar{x} \end{pmatrix} \begin{pmatrix} \bar{p} & \bar{q} \\ -q & p \end{pmatrix} = \begin{pmatrix} p\bar{p}x + q\bar{q}\bar{x} - \bar{p}\bar{q}z + pq\bar{z} & p\bar{q}x - p\bar{q}\bar{x} - \bar{q}^2 z - p^2\bar{z} \\ \bar{p}qx - \bar{p}q\bar{x} + \bar{p}^2 z + q^2\bar{z} & q\bar{q}x + p\bar{p}\bar{x} + \bar{p}\bar{q}z - pq\bar{z} \end{pmatrix}.$$

We write i for $\sqrt{-1}$. Substituting $x = it$ and $z = u + iv$ (t, u, $v \in \mathbb{R}$) we obtain

$$\begin{pmatrix} (p\bar{p} - q\bar{q})it + (pq - \bar{p}\bar{q})u - (pq + \bar{p}\bar{q})iv & 2p\bar{q}it - (p^2 + \bar{q}^2)u + (p^2 - \bar{q}^2)iv \\ 2\bar{p}qit + (\bar{p}^2 + q^2)u + (\bar{p}^2 - q^2)iv & -(p\bar{p} - q\bar{q})it - (pq - \bar{p}\bar{q})u + (pq + \bar{p}\bar{q})iv \end{pmatrix}$$

and find, therefore, that the conjugation action is given by

$$\begin{pmatrix} t \\ u \\ v \end{pmatrix} \mapsto \begin{pmatrix} (p\bar{p} - q\bar{q})t + i(-pq + \bar{p}\bar{q})u + (-pq - \bar{p}\bar{q})v \\ i(\bar{p}q - p\bar{q})t + \frac{1}{2}(p^2 + \bar{p}^2 + q^2 + \bar{q}^2)u + \frac{i}{2}(-p^2 + \bar{p}^2 - q^2 + \bar{q}^2)v \\ (\bar{p}q + p\bar{q})t + \frac{i}{2}(p^2 - \bar{p}^2 - q^2 + \bar{q}^2)u + \frac{1}{2}(p^2 + \bar{p}^2 - q^2 - \bar{q}^2)v \end{pmatrix}.$$

So we have the homomorphism $SU(2) \to GL_3(\mathbb{R})$,

$$\begin{pmatrix} p & -\bar{q} \\ q & \bar{p} \end{pmatrix} \mapsto \begin{pmatrix} |p|^2 - |q|^2 & 2\operatorname{Im} pq & -2\operatorname{Re} pq \\ 2\operatorname{Im} p\bar{q} & \operatorname{Re}(p^2 + q^2) & \operatorname{Im}(p^2 + q^2) \\ 2\operatorname{Re} p\bar{q} & \operatorname{Im}(-p^2 + q^2) & \operatorname{Re}(p^2 - q^2) \end{pmatrix},$$

with kernel $\{\pm E_2\}$, and the images of real matrices and diagonal matrices in $SU(2)$ are easily calculated and seen to generate $SO(3)$. □

Combined with the observation that $SU(2)$ is homeomorphic to S^3, we deduce that $SO(3)$ is homeomorphic to three-dimensional **real projective space** (sphere with antipodal points identified).

The results presented in this section represent a small selection of topics on matrix groups and symmetry, taken from M. Artin's *Algebra*.

3 Fields

One of the first tasks in the Linear Algebra and Analysis lectures is to construct the basic fields

$$\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

and prove that $\mathbb{C}$ is algbraically closed.

Here we will add one more field $\overline{\mathbb{Q}}$ to the list, sitting between $\mathbb{Q}$ and $\mathbb{C}$ but neither contained in nor containing $\mathbb{R}$. This will be the field consisting of all roots of polynomials with rational coefficients. This distinction between algebraic numbers $\overline{\mathbb{Q}}$ and transcendental numbers $\mathbb{C} \smallsetminus \overline{\mathbb{Q}}$ will be the starting point for our study of fields and field extensions, after a brief excursion into symmetric polynomials.

3.1 Symmetric polynomials

We need a classical fact about polynomials in n variables, invariant under all permutations of the variables. The **elementary symmetric polynomials** in n variables $X_1, \ldots, X_n$ are

$$e_1, \ldots, e_n \in \mathbb{Z}[X_1, \ldots, X_n],$$

defined by

$$e_k := \sum_{\substack{I \subset \{1,\dots,n\} \\ |I|=k}} \prod_{i \in I} X_i$$

for $k = 1, \dots, n$. We also define $e_0 := 1$ (which is consistent with the empty product being 1, as is the standard convention) and $e_k := 0$ for $k > n$ (which is also consistent with the above definition, the sum being empty in this case). For example, $e_1 = X_1 + \cdots + X_n$ and $e_n = X_1 \cdots X_n$. Clearly, $e_1, \dots, e_n$ are invariant under permutation of the variables.

Proposition 3.1. *Let R be a commutative ring. Then the R-algebra homomorphism*

$$\iota \colon R[E_1, \dots, E_n] \to R[X_1, \dots, X_n]$$

defined by

$$\iota(E_k) := e_k(X_1, \dots, X_n)$$

for all k is an isomorphism onto the subring of $R[X_1, \dots, X_n]$ of polynomials, invariant under all permutations of the variables.

Proof. Homogeneous polynomials are familiar; we let $R[X_1, \dots, X_n]_d$ denote the R-submodule of $R[X_1, \dots, X_n]$ of homogeneous polynomials of degree d, so that

$$R[X_1, \dots, X_n] = \bigoplus_{d \in \mathbb{N}} R[X_1, \dots, X_n]_d. \tag{1}$$

By assigning the weight j to the variable E_j we also have **weighted homogeneous polynomials** of degree d, the R-submodule

$$R[E_1, \dots, E_n]_d \subset R[E_1, \dots, E_n]$$

spanned by the monomials $E_1^{a_1} \cdots E_n^{a_n}$ such that $\sum_j j a_j = d$. The R-modules $R[E_1, \dots, E_n]_d$ and $R[X_1, \dots, X_n]_d$ are finitely generated and free (in fact, any sequence of n positive integers may be used as weights to define weighted homogeneous polynomials and produce a direct sum decomposition as in (1) into finitely generated free submodules), and ι maps $R[E_1, \dots, E_n]_d$ to $R[X_1, \dots, X_n]_d$ for every $d \in \mathbb{N}$.

Let $\sigma \in S_n$. Associated with σ is an R-algebra automorphism of $R[X_1, \dots, X_n]$, sending X_i to $X_{\sigma(i)}$ for all i. This sends $R[X_1, \dots, X_n]_d$ to $R[X_1, \dots, X_n]_d$ for all d, hence an element of $R[X_1, \dots, X_n]$ is invariant under σ if and only if the component of f of degree d is invariant under σ for every d.

To prove the proposition it suffices for every d to exhibit bases of $R[E_1, \dots, E_n]_d$ and of the S_n-invariant submodule of $R[X_1, \dots, X_n]_d$, with respect to which ι is represented by a square matrix that is upper triangular with diagonal entries equal to 1.

Let us index the monomial $E_1^{a_1} \cdots E_n^{a_n}$, for $a_1, \ldots, a_n \in \mathbb{N}$ with $\sum_j j a_j = d$, by

$$(\lambda_1, \ldots, \lambda_n) := (a_1 + \cdots + a_n, a_2 + \cdots + a_n, \ldots, a_n). \tag{2}$$

This transforms the condition $\sum_j j a_j = d$ into $\sum \lambda_j = d$, at the expense of having to impose the additional condition $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. So we have a basis of $R[E_1, \ldots, E_n]_d$ indexed by the set $I_{n,d}$ of $(\lambda_1, \ldots, \lambda_n) \in \mathbb{N}^n$ satisfying

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \qquad \text{and} \qquad \sum_{j=1}^{n} \lambda_j = d. \tag{3}$$

What we see in (3) reminds us of partitions of d, except that we allow some of the λ_j to be 0. The **length** of a partition is the number of positive summands. So, in words, we could describe $I_{n,d}$ as the set of partitions of d of length at most n.

An element

$$\sum_{q_1, \ldots, q_n \in \mathbb{N},\, \sum q_j = d} c_{q_1 \ldots q_n} X_1^{q_1} \cdots X_n^{q_n} \in R[X_1, \ldots, X_n]_d$$

is invariant under permutation of the variables if and only if

$$c_{q_{\sigma(1)} \ldots q_{\sigma(n)}} = c_{q_1 \ldots q_n}$$

for all $\sigma \in S_n$. So, the invariant submodule of $R[X_1, \ldots, X_n]_d$ has a basis indexed by the same set $I_{n,d}$ as above, where to $\lambda = (\lambda_1, \ldots, \lambda_n) \in I_{n,d}$ we associate

$$m_\lambda := \sum_{\substack{q_1, \ldots, q_n \in \mathbb{N} \\ \exists \sigma \in S_n : \forall j\ q_{\sigma(j)} = \lambda_j}} X_1^{q_1} \cdots X_n^{q_n} \in R[X_1, \ldots, X_n]_d.$$

We introduce **dominance order**: for $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{N}^n$,

$$\lambda \trianglelefteq \mu \ :\Leftrightarrow\ \lambda_1 + \cdots + \lambda_k \leq \mu_1 + \cdots + \mu_k \text{ for } k = 1, \ldots, n.$$

This induces a partial order on $I_{n,d}$.

There is also **lexicographic order** on $\mathbb{N}^n$, which is the usual dictionary order, taking natural numbers to be symbols in an alphabet, ordered as usual 0, 1, ... Lexicographic order is a total order which refines dominance order and thus defines a bijection $I_{n,d} \to \{1, \ldots, |I_{n,d}|\}$, such that for $\lambda \trianglelefteq \mu$ the integer associated with λ is less than or equal the integer associated with μ.

With respect to the bases, now indexed by $1, \ldots, |I_{n,d}|$, we obtain from ι a square matrix, which we claim is upper triangular with diagonal entries equal to 1. We consider a monomial $E_1^{a_1} \cdots E_n^{a_n}$, for $a_1, \ldots, a_n \in \mathbb{N}$ with $\sum_j j a_j = d$, and define $(\lambda_1, \ldots, \lambda_n)$ by the formula (2). We will need the following facts about the monomials

in $e_k(X_1, \dots, X_n)$: there is what we will call the "lead monomial" $X_1 \cdots X_k$, for which the sequence of exponents of $X_1, \dots, X_n$ is $(1, \dots, 1, 0, \dots, 0)$ with 1 occuring k times; and there are other monomials, always of the form $X_{j_1} \dots X_{j_k}$ with $1 \leq j_1 < j_2 < \cdots < j_k \leq n$ and $j_i > i$ for some i, i.e., with exponent sequence with 1 occuring k times, otherwise 0, and with least one 0 among the first k entries. Such an exponent sequence is always strictly less in dominance order than the exponent sequence of the lead monomial.

Now ι sends $E_1^{a_1} \cdots E_n^{a_n}$ to a sum of

- the product of lead monomials $X_1^{a_1}(X_1X_2)^{a_2} \cdots (X_1 \cdots X_n)^{a_n}$, and
- other products of monomials, with a_1 monomials taken from $e_1(X_1, \dots, X_n)$, a_2 monomials taken from $e_2(X_1, \dots, X_n)$, and so on, so that at least one of the monomials is not the respective lead monomial.

The product of lead monomials is $X_1^{\lambda_1} \cdots X_n^{\lambda_n}$. All other products contribute a monomial whose exponent sequence is strictly less than $(\lambda_1, \dots, \lambda_n)$ in dominance order. It follows that the matrix representing ι with respect to the bases as above has diagonal entries 1 and is upper triangular. □

As an application, we describe polynomials, whose roots are sums/powers/... of the roots of a given (monic) polynomial.

Proposition 3.2. *Let R be a commutative ring and $n \in \mathbb{N}$. Then there are $\mathscr{P}_{q,n} \in R[T, E_1, \dots, E_n]$ for $q \in R[U]$ and $\Sigma_{\lambda,n} \in R[T, E_1, \dots, E_n]$ for $\lambda = (\lambda_1, \dots, \lambda_n) \in R^n$, such that for every commutative R-algebra S and $b_1, \dots, b_n \in S$, if*

$$f := T^n + b_1T^{n-1} + \cdots + b_n$$

factors in $S[T]$ as $f = \prod_{j=1}^n (T - \alpha_j)$ with $\alpha_1, \dots, \alpha_n \in S$, then the polynomials $\mathscr{P}_q(f) := \mathscr{P}_{q,n}(T, b_1, \dots, b_n)$ and $\Sigma_\lambda(f) := \Sigma_{\lambda,n}(T, b_1, \dots, b_n)$ factor as

$$\mathscr{P}_q(f) = \prod_{j=1}^n (T - q(\alpha_j)) \qquad \text{and} \qquad \Sigma_\lambda(f) = \prod_{\substack{r_1,\dots,r_n \in R \\ \exists \sigma \in S_n \colon \forall j \, r_{\sigma(j)} = \lambda_j}} \big(T - \sum_{j=1}^n r_j \alpha_j\big).$$

Proof. For $1 \leq j \leq n$ the polynomial

$$e_j(q(X_1), \dots, q(X_n)) \in R[X_1, \dots, X_n],$$

is invariant under arbitrary permutation of the variables. We apply Proposition 3.1 to obtain $\mathscr{P}_{q,n,j} \in R[E_1, \dots, E_n]$ such that $\iota(\mathscr{P}_{q,n,j}) = e_j(q(X_1), \dots, q(X_n))$, where ι sends E_i to $e_i(X_1, \dots, X_n)$ for all i. Then we define

$$\mathscr{P}_{q,n} := T^n - \mathscr{P}_{q,n,1}(-E_1, E_2, \dots, (-1)^n E_n)T^{n-1} + \cdots + (-1)^n \mathscr{P}_{q,n,n}(-E_1, E_2, \dots, (-1)^n E_n).$$

Since $b_j = (-1)^j e_j(\alpha_1, \ldots, \alpha_n)$ for $f \in S[T]$ as in the statement of the proposition, we have

$$\begin{aligned}\mathscr{P}_q(f) &= T^n - \mathscr{P}_{q,n,1}(-b_1, \ldots, (-1)^n b_n)T^{n-1} + \cdots \\ &\qquad + (-1)^n \mathscr{P}_{q,n,n}(-b_1, \ldots, (-1)^n b_n) \\ &= T^n - e_1(q(\alpha_1), \ldots, q(\alpha_n))T^{n-1} + \cdots + (-1)^n e_n(q(\alpha_1), \ldots, q(\alpha_n)),\end{aligned}$$

as required. By a similar argument, where instead of $e_j(q(X_1), \ldots, q(X_n))$ we consider the jth elementary symmetric polynomial in the sums $\sum_{j=1}^n r_j X_j$ for all r_1, $\ldots$, $r_n \in R$ such that $(r_1, \ldots, r_n)$ is equal to some permutation of $(\lambda_1, \ldots, \lambda_n)$, we obtain $\Sigma_{\lambda,n} \in R[T, E_1, \ldots, E_n]$ satisfying the claimed property. □

3.2 Algebraic and transcendental field extensions

When we speak of a field extension L/K we mean that K and L are fields, and an embedding of fields $K \to L$ has been fixed, identifying K with a subfield of L. In particular, L acquires a structure of K-vector space, so it makes sense to ask if L is finite-dimensional as K-vector space – in which case we say L/K is a **finite** extension – and if this is the case to speak of the dimension of L as a K-vector space, the **degree** of L/K, denoted by $[L : K]$.

Definition. Let L/K be a field extension. An element $x \in L$ is **algebraic** over K if there exists a nonzero polynomial $f \in K[T]$ such that $f(x) = 0$. Otherwise we say x is **transcendental** over K.

An element $x \in L$ generates a subfield $K(x) \subset L$. If $x \in L$ is algebraic, then the set of $f \in K[T]$ satisfying $f(x) = 0$ is a nonzero ideal of $K[T]$, say, generated by $g \in K[T]$ of some degree d. We enforce the convention that g is monic; then, g is uniquely determined and is the **minimal polynomial** of x. The integer d is the **degree** of x. We have $K(x) = K[x]$, isomorphic to $K[T]/(g)$. So, as a vector space $K(x)$ has the basis 1, x, $\ldots$, x^{d-1}, and we have the compatibility with the notion of degree of a field extension:

$$[K(x) : K] = d.$$

If $x \in L$ is transcendental, then $K[x] \cong K[T]$, with field of fractions $K(x) \cong K(T)$. Generally, if $L = K(x)$ then we call x a **primitive element** for L/K.

We describe a field extension L/K as algebraic if every element of L is algebraic over K, and otherwise we say that L/K is transcendental. For instance, every finite extension is algebraic. Indeed, if $[L : K] = d$ then for $x \in L$ there must be a linear dependence among 1, x, $\ldots$, x^d, which shows that x is algebraic over K of degree at most d.

Definition. Let L/K be a field extension. The **algebraic closure** of K in L is the set of elements of L that are algebraic over K. We say that K is **algebraically**

closed in L if the only elements of L, that are algebraic over K, are the elements of K itself.

Recall, we say that K is algebraically closed if every nonconstant polynomial in $K[T]$ has a root, or equivalently, if every irreducible polynomial in $K[T]$ is linear. The relation between the notions of algebraically closed field and algebraic closure in an extension field is that K is an algebraically closed field if and only if K is algebraically closed in every extension field.

We may also speak of an algebraic closure of a field K, without mentioning any extension field. In that case we mean an algebraic extension $\overline{K}$ of K, such that $\overline{K}$ is algebraically closed. One of the tasks of a more comprehensive treatment of fields is to prove that every field admits an algebraic closure, and any pair of such are isomorphic to each other. (This is explained in Appendix C.) For this brief treatment we will remain content with the observation, which we will present below as Proposition 3.6, that if L/K is a field extension and L is algebraically closed, then the algebraic closure $\overline{K}$ of K in L is an algebraic closure of K. The main example for us is $\overline{\mathbb{Q}}$, the field of **algebraic numbers**, defined as the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$.

Proposition 3.3. *Let L/K be a field extension. The algebraic closure of K in L is a subfield of L.*

Proof. Let $0 \neq x \in L$ be algebraic over K. So there exists $0 \neq f \in K[T]$ such that $f(x) = 0$, let us say

$$f = a_n T^n + \cdots + a_1 T + a_0.$$

Let us define $g := a_0 T^n + a_1 T^{n-1} + \cdots + a_n$. Then $g(x^{-1}) = 0$.

It remains to verify that if x, $y \in L$ are algebraic over K then $x + y$ and xy are algebraic over K. Let us suppose that x has degree d and y has degree e. This means that $K[x]$, as a vector space, is the span of 1, x, …, x^{d-1}, and $K[y]$ is the span of 1, y, …, y^{e-1}. We define M to be the subspace of L spanned by

$$1, x, \ldots, x^{d-1},\ y, xy, \ldots, x^{d-1}y,\ \ldots,\ y^{e-1}, xy^{e-1}, \ldots, x^{d-1}y^{e-1}.$$

For the endomorphisms f, $g\colon M \to M$ defined by $f(z) := (x+y)z$ and $g(z) := xyz$, the Cayley-Hamilton theorem dictates that the respective characteristic polynomials P_f, $P_g \in K[T]$ satisfy

$$P_f(x+y) = P_f(f)(1) = 0 \qquad \text{and} \qquad P_g(xy) = P_g(g)(1) = 0.$$

So $x + y$ and xy are algebraic over K. □

Proposition 3.4. *Let L/K and M/L be finite field extensions. Then M/K is a finite field extension, and*

$$[M : K] = [M : L] \cdot [L : K].$$

Proof. Let $x_1, \ldots, x_d$ be a basis for L as K-vector space, and $y_1, \ldots, y_e$ a basis for M as L-vector space. Then $(x_i y_j)_{1 \leq i \leq d,\, 1 \leq j \leq e}$ is a basis for M as K-vector space. □

We record some elementary observations concerning algebraic closure and embeddings. Let K'/K be a finite field extension, and let $\overline{K}$ be an algebraic closure of K. Then there exists an embedding $K' \to \overline{K}$ over K. We verify this by induction on $d := [K' : K]$, the case $d = 1$ being trivial and, for $d > 1$ and $x \in K' \smallsetminus K$ with minimal polynomial $g \in K[T]$ a choice of $\alpha \in \overline{K}$ satisfying $g(\alpha) = 0$ defines an embedding of $K[x] \cong K[T]/(g)$ in $\overline{K}$, and we may conclude by induction since $[K' : K(x)] < d$. If, instead of an algebraic closure we consider an *arbitrary* field extension L/K, then the same conclusion is valid provided we allow ourselves to replace L by a finite extension. We see, as well, that for an arbitrary field extension L/K the number of embeddings $K' \to L$ over K is less than or equal to $[K' : K]$.

A field extension L/K and collection of elements $(x_s)_{s \in S}$ of L determine a homomorphism of K-algebras $\mathrm{Sym}^\bullet(\bigoplus_{s \in S} K) \to L$, whose image is a subring of L, necessarily an integral domain. The image, which we denote by $K[(x_s)_{s \in S}]$, has a quotient field, which by the universal property of the quotient field is identified with a subfield of L, the field **generated** by $(x_s)_{s \in S}$ over K, denoted by $K((x_s)_{s \in S})$:

$$K \subset K((x_s)_{s \in S}) \subset L.$$

If $S' \subset S$ and we set $K' := K((x_s)_{s \in S'})$ then we have

$$K((x_s)_{s \in S}) = K'((x_s)_{s \in S \smallsetminus S'}).$$

Indeed, we have $K[(x_s)_{s \in S}] \subset K'[(x_s)_{s \in S \smallsetminus S'}] \subset K((x_s)_{s \in S})$.

Proposition 3.5. (i) *Let L/K be a field extension, and let $(x_s)_{s \in S}$ be a collection of elements of L. If x_s is algebraic over K for every $s \in S$, then $K((x_s)_{s \in S})$ is an algebraic extension of K.*
(ii) *If L/K and M/L are algebraic field extensions, then M/K is also algebraic.*

Proof. Every element of $K((x_s)_{s \in S})$ lies in $K((x_s)_{s \in S'})$ for some finite subset $S' \subset S$. Labeling the x_s for $s \in S'$ as $x_1, \ldots, x_n$, we have $K(x_1, \ldots, x_n) = K(x_1)(x_2) \ldots (x_n)$, which can be expressed as a tower of finite extensions over K. By Proposition 3.4, $K(x_1, \ldots, x_n)$ is finite, and hence algebraic, over K, and (i) is established.

For (ii), we know that every element y of M satisfies a polynomial relation with coefficients in L, hence with coefficients in a subfield of the form $K(x_1, \ldots, x_n)$ with $x_1, \ldots, x_n \in L$. We have just seen, $K(x_1, \ldots, x_n)/K$ is finite. Now y is algebraic over $K(x_1, \ldots, x_n)$, hence $K(x_1, \ldots, x_n, y)/K(x_1, \ldots, x_n)$ is finite, and we are again done by Proposition 3.4. □

Proposition 3.6. *Let L/K be a field extension, such that L is algebraically closed. Then the algebraic closure $\overline{K}$ of K in L is an algebraic closure of K.*

Proof. Let $f \in \overline{K}[T]$ be a nonconstant polynomial. Since L is algebraically closed, there exists $x \in L$ such that $f(x) = 0$. Now x is algebraic over $\overline{K}$, which in turn is algebraic over K. By Proposition 3.5 (ii), x is algebraic over K. Hence we have $x \in \overline{K}$. □

Lemma 3.7. *Let L/K be a field extension, and let x, $y \in L$ be transcendental over K. If L is algebraic over $K(x)$, then L is algebraic over $K(y)$.*

Proof. We have y algebraic over $K(x)$, i.e., satisfying a nontrivial polynomial relation with coefficients in $K(x)$. Clearing denominators, we may suppose that the coefficients are in $K[x]$, which means that there is $0 \neq f \in K[X, Y]$ such that $f(x, y) = 0$. Let us write f as

$$c_n X^n + \cdots + c_1 X + c_0.$$

with $c_0, \ldots, c_n \in K[Y]$, and $c_n \neq 0$. We have $n \geq 1$ since y is transcendental over K, and the polynomial relation

$$c_n(y)x^n + \cdots + c_1(y)x + c_0 = 0$$

shows that x is algebraic over $K(y)$. So $K(x, y)$ is algebraic over $K(y)$. Now L, algebraic over $K(x)$, is algebraic over $K(x, y)$ and hence also over $K(y)$. □

Definition. Let L/K be a field extension. A collection of elements $(x_s)_{s\in S}$ of L is **algebraically dependent** over K if there exist $n \in \mathbb{N}$, pairwise distinct elements $s_1, \ldots, s_n \in S$, and nonzero $f \in K[X_1, \ldots, X_n]$ such that $f(x_{s_1}, \ldots, x_{s_n}) = 0$; otherwise, the collection $(x_s)_{s\in S}$ is **algebraically independent** over K. A **transcendence basis** is an algebraically independent collection $(x_s)_{s\in S}$ such that L is algebraic over $K((x_s)_{s\in S})$.

An equivalent condition for $(x_s)_{s\in S}$ to be algebraically independent, is that the corresponding homomorphism of K-algebras $\mathrm{Sym}^\bullet(\bigoplus_{s\in S} K) \to L$ is injective. Given $(x_s)_{s\in S}$, algebraically independent over K, then $y \in L$ is transcendental over $K((x_s)_{s\in S})$ if and only if adjoining y to $(x_s)_{s\in S}$ yields an algebraically independent collection. In complete analogy with linear independent collections of elements in a vector space there is a straightforward argument using Zorn's lemma for the existence of transcendence bases for general field extensions.

Proposition 3.8. *Let L/K be a field extension. Then any algebraically independent collection of elements of L over K may be extended to a transcendence basis of L/K.*

Proof. Suppose $S \subset L$ is algebraically independent over K. The set of subsets of L containing S and algebraically independent over K is partially ordered by inclusion. Given a collection of such subsets, such that for any two, one is contained in the other, their union is again algebraically independent. So by Zorn's lemma, some such subset is maximal, and hence is a transcendence basis. □

Lemma 3.9 (Exchange Lemma)**.** *Let $(x_s)_{s\in S}$ be a transcendence basis for a field extension L/K, and let $y \in L$ be transcendental over $K((x_s)_{s\in S\smallsetminus\{s_0\}})$ for some $s_0 \in S$. Define $\tilde{x}_s$ for $s \in S$ by*

$$\tilde{x}_s := \begin{cases} y, & \text{if } s = s_0, \\ x_s, & \text{otherwise.} \end{cases}$$

Then $(\tilde{x}_s)_{s\in S}$ is a transcendence basis for L over K.

Proof. The transcendence of y over $K' := K((x_s)_{s\in S\smallsetminus\{s_0\}})$ implies the algebraic independence of $(\tilde{x}_s)_{s\in S}$. By Lemma 3.7, L is algebraic over $K'(y) = K((\tilde{x}_s)_{s\in S})$. □

Proposition 3.10. *Let L/K be a field extension. If $(x_s)_{s\in S}$ and $(y_t)_{t\in T}$ are transcendence bases for L over K then S is finite if and only if T is finite. If S and T are finite then they have the same cardinality.*

Proof. Since the elements of a transcendence basis are pairwise distinct, there is no loss of generality in working with transcendence bases as subsets of L. Suppose that S is finite and T is infinite, or finite with $|T| \geq |S|$. Let $n := |S|$. We will show by descending induction on $j := |S \cap T|$, that $|T| = n$. The base case $j = n$, i.e., $S \subset T$, is clear. For the inductive step, we choose $s \in S \smallsetminus T$. There is $t \in T$, transcendental over $K(S \smallsetminus \{s\})$. (If all the elements of T are all algebraic over $K(S \smallsetminus \{s\})$, then L is algebraic over $K(S \smallsetminus \{s\})$, a contradiction.) By the Exchange Lemma,

$$S' := (S \smallsetminus \{s\}) \cup \{t\}$$

is a transcendence basis. We have $|S'| = |S|$ and $|S' \cap T| = |S \cap T| + 1$. So, the induction hypothesis yields $|T| = |S'| = n$. □

Definition. If L has a finite transcendence basis over K then we call the number of elements in any transcendence basis of L over K the **transcendence degree** of L over K and denote this quantity by $\mathrm{trdeg}_K(L)$.

Lemma 3.11. *Let L/K be a field extension. Then:*
(i) *K is algebraically closed in $K(X)$.*
(ii) *If L is algebraic over K, then $L(X)$ is algebraic over $K(X)$, and if furthermore L is finite over K of degree d, then $L(X)$ is finite over $K(X)$ of degree d.*
(iii) *If K is algebraically closed in L, then $K(X)$ is algebraically closed in $L(X)$.*

Proof. For all three parts we will use the observation (Proposition 1.9) that an irreducible polynomial in $K[T]$ is irreducible in $K[T, X]$, and also in $K(X)[T]$.

For (i) we notice that an element of $K(X)$, algebraic over K, would have minimal polynomial in $K[T]$ that is irreducible, hence also irreducible as a polynomial over $K(X)$. An irreducible polynomial with a root is linear.

We establish the first part of (ii) by recognizing that $L(X)$, as an extension of $K(X)$, is generated by the elementsof $L \smallsetminus K$, and applying Proposition 3.5 (i). For the remaining part, we use induction on $[L : K]$ to reduce to the case $L = K(\alpha)$. The minimal polynomial of α is irreducible in $K(X)[T]$, so $[L(X) : K(X)] = [L : K]$.

Part (iii) easily reduces to the case that L is finitely generated as a field extension of K. It suffices to exclude, for each $d \in \mathbb{N}$, $d \geq 2$, the existence of an element of $L(X)$, algebraic over $K(X)$ of degree d. Let $e := \mathrm{trdeg}_K(L)$, and let $x_1, \ldots, x_e$ be a transcendence basis of L over K. Then, besides the inclusion $\sigma^{(0)}$ of L in

$$L((T_j^{(i)})_{1\leq i\leq d, 1\leq j\leq e}) = L(T_1^{(1)}, \ldots, T_e^{(1)}, \ \ldots, \ T_1^{(d)}, \ldots, T_e^{(d)})$$

there exist, for a suitable finite extension $M/L((T_j^{(i)})_{1\leq i\leq d,1\leq j\leq e})$, embeddings

$$\sigma^{(1)}, \ldots, \sigma^{(d)} \colon L \to M$$

extending the embeddings of $K(x_1, \ldots, x_e)$ determined by the condition

$$\sigma^{(i)}(x_j) = T_j^{(i)}$$

for $1 \leq i \leq d$ and $1 \leq j \leq e$. Let $\alpha \in L$, with $\alpha \notin K$. By hypothesis, the minimal polynomial of α over $K(x_1, \ldots, x_e)$ has coefficients, not all in K. Upon applying the embeddings $\sigma^{(i)}$, we obtain elements of M with distinct minimal polynomials over $K(x_1, \ldots, x_e)((T_j^{(i)})_{1\leq i\leq d,1\leq j\leq e})$. In particular,

$$\sigma^{(0)}(\alpha), \ldots, \sigma^{(d)}(\alpha)$$

are all distinct. The embeddings $\sigma^{(0)}, \ldots, \sigma^{(d)}$ extend uniquely to

$$\tau^{(0)}, \ldots, \tau^{(d)} \colon L(X) \to M(X),$$

by $\tau^{(i)}(X) := X$ for all i. An element of $L(X)$, respectively $M(X)$, may be written uniquely as gh^{-1} where g and h are polynomials in $L[X]$, respectively $M[X]$, without common irreducible factor, with h monic. We see this way that for $f \in L(X)$, with $f \notin K(X)$, the elements

$$\tau^{(0)}(f), \ldots, \tau^{(d)}(f)$$

are distinct. In particular, f cannot be algebraic of degree d over $K(X)$. □

Proposition 3.12. *If L/K and M/L are field extensions of finite transcendence degree, then M has finite transcendence degree over K, and*

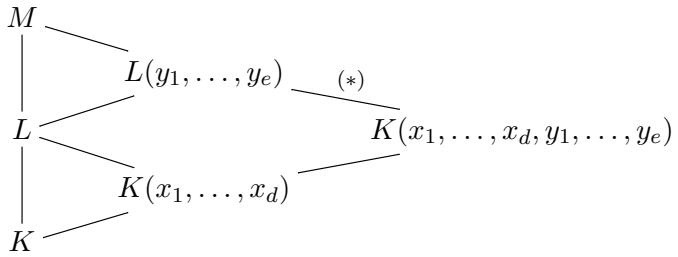
$$\mathrm{trdeg}_K(M) = \mathrm{trdeg}_K(L) + \mathrm{trdeg}_L(M).$$

Specifically, the union of a transcendence basis of L over K and a transcendence basis of M over L is a transcendence basis of M over K.

Proof. Let $x_1, \ldots, x_d$ be a transcendence basis of L over K, and let $y_1, \ldots, y_e$ be a transcendence basis of M over L. Algebraic independence of $x_1, \ldots, x_d, y_1, \ldots, y_e$ is clear, by writing $f \in K[X_1, \ldots, X_d, Y_1, \ldots, Y_e]$ as a sum

$$f = \sum c_{q_1,\ldots,q_e} Y_1^{q_1} \cdots Y_e^{q_e}$$

over a finite subset of $\mathbb{N}^e$, with coefficients $c_{q_1,\ldots,q_e} \in K[X_1, \ldots, X_d]$, and deducing from $f(x_1, \ldots, x_d, y_1, \ldots, y_e) = 0$ that the coefficients $c_{q_1,\ldots,q_e}$ are all zero. It remains to show that M is algebraic over $K(x_1, \ldots, x_d, y_1, \ldots, y_e)$. In the diagram of field extensions



we identify $K(x_1, \ldots, x_d, y_1, \ldots, y_e)$ with $K(x_1, \ldots, x_d)(T_1, \ldots, T_e)$ and $L(y_1, \ldots, y_e)$ with $L(T_1, \ldots, T_e)$ and apply Lemma 3.11 (ii) to deduce that the extension labeled $(*)$ is algebraic. □

Remark. Working with extensions of commutative rings rather than fields, a notion analogous to algebraic closure in an extension field is integral closure. Given commutative rings $R \subset S$, we say that $x \in S$ is **integral** over R if there exists a *monic* polynomial $f \in R[T]$ such that $f(x) = 0$. The **integral closure** of R in S is the set of elements of S integral over R; adapting the proof of Proposition 3.3, we may see that this is a subring of S. We say that R is **integrally closed** in S if the integral closure is just R. When R is a UFD with field of fractions K, and S is an integral domain, it follows from Gauss's lemma that $x \in S$ is integral over R if and only if x is algebraic over K and the minimal polynomial of x over K has coefficients in R. An interesting case for us will be the ring of **algebraic integers**, which is the integral closure of $\mathbb{Z}$ in $\overline{\mathbb{Q}}$. Then, an algebraic number is integral over $\mathbb{Z}$ if and only if its minimal polynomial over $\mathbb{Q}$ has integer coefficients.

3.3 Finite extensions

When L/K is a finite field extension it is possible to take advantage of the fact that L is a finite-dimensional K-vector space to profit from techniques based on linear algebra for finite-dimensional vector spaces. Some of these techniques are valid for finitely generated free modules over commutative rings, so we state some of the definitions and results in this context.

Definition. Let R be a commutative ring, and let S be a commutative R-algebra which is finitely generated and free as an R-module. The **trace** of $x \in S$ is the trace of the R-linear endomorphism $m_x \colon S \to S$, $m_x(y) := xy$, and is denoted by $\mathrm{tr}_{S/R}(x)$ or $\mathrm{tr}(x)$. The **norm** of $x \in S$ is $\det(m_x)$, and is denoted by $N_{S/R}(x)$ or $N(x)$.

Since trace of a sum of two endomorphisms is the sum of their traces and respects scalar multiplication as well, $\mathrm{tr} \colon S \to R$ is a homomorphisms of R-modules. The determinant of a composite of two endomorphisms is the product of determinants: $N(xy) = N(x)N(y)$ for x, $y \in S$. Let us suppose R is not the zero ring (if R is the zero ring then so is S, and tr and N are as well zero) and let d be the rank of S as R-module; then $\mathrm{tr}(a) = da$ and $N(a) = a^d$ for $a \in R$.

Norm and trace behave well in towers.

Proposition 3.13. *Let R be a commutative ring, and let S be a commutative R-algebra which is finitely generated and free as an R-module. Further, let T be an S-algebra which is finitely generated and free as an S-module. Then T is finitely generated and free as an R-module, and we have*

$$\mathrm{tr}_{T/R} = \mathrm{tr}_{S/R} \circ \mathrm{tr}_{T/S} \qquad \textit{and} \qquad N_{T/R} = N_{S/R} \circ N_{T/S}.$$

Proof. We may assume that the rings R, S, and T are nonzero. Let d be the rank and $x_1, \ldots, x_d$ a basis of S as R-module, and let e be the rank and $y_1, \ldots, y_e$ a basis of T as S-module. As in Proposition 3.4, as R-module T is free of rank de with basis

$$x_1y_1, \ldots, x_dy_1, \ \ldots, \ x_1y_e, \ldots, x_dy_e. \tag{1}$$

Say $z \in T$, and let A be the matrix representing multiplication by z with respect to the basis (1). We view A as an $e \times e$ block matrix, with $d \times d$ blocks:

$$A = (A_{ij})_{1 \le i,j \le e}, \qquad A_{ij} \in \mathrm{M}_d(R).$$

Let us write

$$zy_j = b_{1j}y_1 + \cdots + b_{ej}y_e$$

with $b_{1j}, \ldots, b_{ej} \in S$, for every j. Then, we claim, A_{ij} is the matrix representing $S \to S$, multiplication by b_{ij}, with respect to the basis $x_1, \ldots, x_d$. Indeed, if for all h, i, and j we write $b_{ij}x_h = \sum_g c_{ghij}x_g$ with $c_{ghij} \in R$, then for all h and j,

$$zx_hy_j = \sum_{i=1}^{e} b_{ij}x_hy_i = \sum_{i=1}^{e}\sum_{g=1}^{d} c_{ghij}x_gy_i.$$

So the (g,h)-entry of A_{ij} is c_{ghij}.

We have $\mathrm{tr}_{T/S}(z) = b_{11} + \cdots + b_{ee}$, and hence

$$\begin{aligned}\mathrm{tr}_{S/R}(\mathrm{tr}_{T/S}(z)) &= \mathrm{tr}_{S/R}(b_{11}) + \cdots + \mathrm{tr}_{S/R}(b_{ee}) \\ &= \mathrm{tr}(A_{11}) + \cdots + \mathrm{tr}(A_{ee}) = \mathrm{tr}(A) = \mathrm{tr}_{T/R}(z).\end{aligned}$$

As well, the blocks A_{ij} commute with each other, and we may apply the formula for the determinant of a block matrix with commuting blocks to conclude

$$N_{S/R}(N_{T/S}(z)) = N_{S/R}(\det(B)) = \det(A) = N_{T/R}(z),$$

where B denotes the matrix $(b_{ij})_{1\leq i,j\leq e}$. □

A commutative R-algebra S given by an injective ring homomorphism $R \to S$, such that every element of S is integral over R, is called an integral ring extension.

Example. Let $f \in R[T]$ be a monic polynomial of degree $d \geq 1$. Then $R[T]/(f)$ is an integral ring extension of R; as an R-module $R[T]/(f)$ is free of rank d, with basis $1, \overline{T}, \ldots, \overline{T}^{d-1}$.

Proposition 3.14. *Let R be a commutative ring, and let S be a commutative R-algebra which is finitely generated and free as an R-module. We suppose that neither R nor S is the zero ring. Then S is an integral extension of R, and*

$$S^\times = \{x \in S \mid N(x) \in R^\times\}.$$

Proof. We apply the Cayley-Hamilton theorem to the endomorphism $m_x \colon S \to S$ given by multiplication by $x \in S$ as in the proof of Proposition 3.3 to deduce that $P_{m_x}(x) = 0$. So x is integral over R. If $N(x) \in R^\times$, then m_x is bijective, therefore $x \in S^\times$. Conversely, if $x \in S^\times$ then $1 = N(1) = N(xx^{-1}) = N(x)N(x^{-1})$, hence $N(x) \in R^\times$. □

Of particular interest are **number fields**, which are finite extensions of $\mathbb{Q}$, and their subrings that are finite over $\mathbb{Z}$. Concrete examples include quadratic extensions $\mathbb{Q}(\sqrt{d})$, where d is a nonsquare integer. If $d = e^2 d'$ with e, $d' \in \mathbb{Z}$ then $\mathbb{Q}(\sqrt{d}) = \mathbb{Q}(\sqrt{d'})$, there is no loss of generality in supposing that d is squarefree. (An integer is squarefree if it has no nontrivial square integer factor.) The subring $\mathbb{Z}[\sqrt{d}]$ is finite over $\mathbb{Z}$, having basis $1, \sqrt{d}$. The case $d = -1$ is the ring $\mathbb{Z}[\sqrt{-1}]$ of **Gaussian integers**. The non-UFD example in §1.1 is the case $d = 10$. When $d \equiv 2$ or $3 \bmod 4$, $\mathbb{Z}[\sqrt{d}]$ is the integral closure of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{d})$, but when $d \equiv 1 \bmod 4$ the integral closure is the larger ring $\mathbb{Z}[(1+\sqrt{d})/2]$. (For a, $b \in \mathbb{Q}$, $b \neq 0$, the minimal polynomial of $a + b\sqrt{d}$ has coefficients $2a$ and $a^2 - db^2$, and these are integers if and only if a and b are integers, or a and b are a non-integer half-integers and $d \equiv 1 \bmod 4$.)

Proposition 3.15. *Let K be a field of characteristic different from 2. Then every quadratic extension of K is of the form $K(\sqrt{d})$ for some nonsquare $d \in K$, and there is an isomorphism $K(\sqrt{d}) \cong K(\sqrt{d'})$ over K if and only if d' is equal to $e^2 d$ for some $e \in K^\times$.*

For nonsquare $d \in K$, as an abstract field $K(\sqrt{d})$ denotes $K[X]/(X^2 - d)$.

Proof. Suppose L/K is a quadratic field extension. Let $x \in L \smallsetminus K$, so we have the basis 1, x for L as K-vector space. We have

$$x^2 = ax - b$$

for some a, $b \in K$. It follows that

$$\big(x - \frac{a}{2}\big)^2 = \frac{a^2 - 4b}{4}.$$

Hence $d := (a^2 - 4b)/4$ is nonsquare in K, and

$$L \cong K(\sqrt{d}).$$

Given an isomorphism $K(\sqrt{d'}) \cong K(\sqrt{d})$ over K, say, sending $\sqrt{d'}$ to $a + b\sqrt{d}$, we deduce from $d' = (a+b\sqrt{d})^2 = a^2+2ab\sqrt{d}+b^2 d$ that $a = 0$, and we have $d' = b^2 d$. □

Given a field K inside an algebraic closure $\overline{K}$ and a nonzero polynomial $f \in K[T]$, we know that f splits into linear factors over $\overline{K}$:

$$f = c(T - \alpha_1) \cdots (T - \alpha_d) \tag{2}$$

for some $\alpha_1, \ldots, \alpha_d \in \overline{K}$, where c is the leading coefficient and d is the degree of f. In this case, we call the field extension

$$K(\alpha_1, \ldots, \alpha_d)/K$$

a **splitting field** of f. Here, $\overline{K}$ may be replaced by any extension field of K in which f factors as in (2), and we call $K(\alpha_1, \ldots, \alpha_d)/K$ a splitting field of f.

Proposition 3.16 (Primitive element theorem)**.** *Let L/K be a finite field extension. Then the following are equivalent:*

(i) *For some $x \in L$ we have $L = K(x)$.*

(ii) *There are only finitely many subfields of L containing K.*

Furthermore, conditions (i) *and* (ii) *hold when K and L are finite fields, and they hold as well when K and L are fields of characteristic zero.*

Proof. When K and L are finite fields, we may take x to be a generator of the cyclic group $L^\times$ in (i), and (ii) holds trivially. From now on, we suppose that K and L are infinite. Given (i), we let $f \in K[T]$ be the minimal polynomial of x over K. If K' is a subfield of L containing K, then the minimal polynomial $g = T^e + \alpha_1 T^{e-1} + \cdots + \alpha_e \in K'[T]$ of x over K' satisfies $g \mid f$ in $K'[T]$ and hence also in $L[T]$. We claim $K(\alpha_1, \ldots, \alpha_e) = K'$. The containment $\subset$ is obvious, and since g is irreducible in $K(\alpha_1, \ldots, \alpha_e)[T]$,

$$[L : K(\alpha_1, \ldots, \alpha_e)] = e = [L : K'].$$

We have (ii), since there are only finitely many monic polynomials in $L[T]$ dividing f. Given (ii), we only have to take x to lie outside the finitely many proper subfields of L containing K to obtain (i).

When $\mathrm{char}(K) = 0$ and $f \in K[T]$ is an *irreducible* polynomial, the expression (2) always has $\alpha_1, \ldots, \alpha_d$ distinct. (If $\alpha_i = \alpha_j$ for some $i \neq j$, then f and its derivative would have a common factor, contradiction.) Now, an inductive argument shows that the inclusion of K in an algebraic closure $\overline{K}$ extends to precisely $[L : K]$ embeddings

$$\sigma_1, \ldots, \sigma_{[L:K]} \colon L \to \overline{K}.$$

For each $j > 1$ we have a proper subfield $\{y \in L \mid \sigma_j(y) = \sigma_1(y)\}$ of L. We may take x to lie outside of these proper subfields, then $L = K(x)$. □

Proposition 3.17. *Let L/K be a finite field extension with a primitive element $x \in L$, let $f \in K[T]$ be the minimal polynomial of x, and let M/K be a field extension which is a splitting field of f, so M is generated over K by $\alpha_1, \ldots, \alpha_d \in M$ with $d = [L : K]$ and $f = (T - \alpha_1) \ldots (T - \alpha_d)$ in $M[T]$. For $1 \leq i \leq d$ let $\sigma_i \colon L \to M$ be the embedding over K defined by $x \mapsto \alpha_i$. Then for all $y \in L$ we have*

$$\mathrm{tr}_{L/K}(y) = \sum_{i=1}^{d} \sigma_i(y) \qquad \text{and} \qquad N_{L/K}(y) = \prod_{i=1}^{d} \sigma_i(y).$$

Proof. We suppose $d > 1$. It suffices to prove the formulas for all y belonging to a subset of $L^\times$ that spans L as K-vector space and generates $L^\times$ as an abelian group.

We prove the formulas for primitive elements $y \in L$. The minimal polynomial $g \in K[T]$ of y has degree d. With respect to the basis $1, y, \ldots, y^{d-1}$ of L as K-vector space multiplication by y is represented by the companion matrix of g. The formulas follow, then, from the equality

$$g = \prod_{i=1}^{d} (T - \sigma_i(y))$$

in $M[T]$, which holds since the polynomial on the right is monic of degree d, by Proposition 3.2 has coefficients in K, and vanishes when evaluated at y.

Now we show that the primitive elements for L/K span L as K-vector space and generate $L^\times$ as an abelian group. When K is finite, $L^\times$ is cyclic and a generator is primitive, as mentioned in the proof of Proposition 3.16. In this case, that the primitive elements span L as K-vector space is clear if $d = 2$, while for $|L| = p^n$, $n \geq 3$, the set of nonzero nonprimitive elements is contained in the union of the sets of $(p^j - 1)$-th roots of unity for $j = 1, \ldots, n-2$, hence has cardinality at most $\sum_{j=1}^{n-2} p^j$, so the set of primitive elements has to have cardinality $> p^{n-1}$ and hence K-span L. When K is infinite, the argument for (ii) $\Rightarrow$ (i) of Proposition 3.16 shows that a primitive element exists outside of any given proper K-subspace of L; similarly, for any $y \in L^\times$ there is a primitive element $z \in L$ such that $z' := zy^{-1}$ is also primitive, and we have $y = zz'^{-1}$. □

Let L/K be a finite field extension. The trace can be used to define a symmetric K-bilinear form on L:

$$\begin{aligned} L \times L &\to K, \\ (x, y) &\mapsto \mathrm{tr}(xy). \end{aligned}$$

The discriminant of this bilinear form is called the **discriminant** of L/K. As with any bilinear form, the discriminant is defined only up to multiplication by squares of elements of $K^\times$. More generally, we may consider a commutative ring R and commutative R-algebra S which is finitely generated and free as an R-module, and define the discriminant of S/R, which is an element of R defined up to multiplication by squares of units in R.

For example, if $R = \mathbb{Z}$ then the only units are ± 1, and the discriminant is a well-defined integer.

We claim that any finite extension L/K of fields of characteristic zero has nonzero discriminant. By Proposition 3.16, for some $x \in L$ we have $L = K(x)$. With notation as in Proposition 3.17, we have

$$\mathrm{tr}_{L/K}(x^j) = \sum_{i=1}^{d} \alpha_i^j$$

for all $j \in \mathbb{N}$. We introduce

$$A := \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & & \alpha_d \\ \vdots & & \ddots & \vdots \\ \alpha_1^{d-1} & \alpha_2^{d-1} & \cdots & \alpha_d^{d-1} \end{pmatrix}$$

and compute the discriminant of L/K using the basis $1, x, \ldots, x^{d-1}$ of L as K-vector space as

$$\det(\mathrm{tr}_{L/K}(x^{i+j}))_{0 \leq i,j \leq d-1} = \det(A \cdot {}^tA) = \det(A)^2 = \prod_{1 \leq i < j \leq d} (\alpha_j - \alpha_i)^2,$$

where the last equality uses the formula for the Vandermonde determinant. Since $\alpha_1, \ldots, \alpha_d$ are distinct (see the proof of Proposition 3.16), the claim is established.

As an application of discriminant, we show that the integral closure of $\mathbb{Z}$ in any number field K is isomorphic as a $\mathbb{Z}$-module to $\mathbb{Z}^d$, where $d := [K : \mathbb{Q}]$. There exists a subring of K, isomorphic as $\mathbb{Z}$-module to $\mathbb{Z}^d$; such a subring is called an **order** in K. Indeed, by taking $x \in K$ such that $K = \mathbb{Q}(x)$ and replacing x by a suitable nonzero integer multiple, we may arrange that x is integral over $\mathbb{Z}$. Then $\mathbb{Z}[x]$ is an order in K.

Proposition 3.18. *Let K be a number field. The integral closure of $\mathbb{Z}$ in K is isomorphic as a $\mathbb{Z}$-module to $\mathbb{Z}^d$, where $d := [K : \mathbb{Q}]$.*

Proof. It suffices to show that there exists a maximal order in K. Indeed, every order consists of integral elements, and given an order M in K and an integral element $x \in K$, with $x \notin M$, we may adjoin x to M to obtain a strictly larger order.

An order M in K has a well-defined discriminant $D(M) \in \mathbb{Z}$, and the nonvanishing of the discriminant of K over $\mathbb{Q}$ implies $D(M) \neq 0$. If $M \subset M'$ are orders in K, then $D(M) = [M' : M]^2 D(M')$. It follows that an order whose discriminant is minimal in absolute value is a maximal order. □

3.4 Construction by straightedge and compass

Construction by straightedge and compass is a classical topic in Euclidean geometry. Given two distinct points in the plane, for instance, a point of intersection of the circles, centered at each point and passing through the other point, forms an equilateral triangle with the original two points.

One general construction is of the perpendicular line to a given line at a given point. This may be used to construct a square, having the line segment joining two given distinct points as one of its edges. For a regular pentagon, a contruction may be given, based on the fact that the ratio of diagonal to edge length in a regular pentagon is the golden ratio, $(1 + \sqrt{5})/2$ to 1. The cases of regular hexagon and octagon are straightforward; the cases of regular heptagon, which we have skipped, and nonagon will be discussed at the end of this section.

Another general construction is angle bisection. For example, by bisecting an exterior angle of a regular pentagon we may construct a regular decagon.

This brings us to the first of three **classical construction problems**.

- Trisecting an angle: dividing a given angle into three equal parts.
- Duplicating a cube: constructing, from a given side length, the side length corresponding to the cube with twice the volume.
- Squaring a circle: contructing, from a given circle, a square with the same area.

The goal of this section is to characterize all the lengths that are constructible with straightedge and compass from a given length. With the characterization, we will be able to show the impossibility of trisecting a general angle, duplicating a cube, and squaring a circle with straightedge and compass.

Theorem 3.19. *Let two distinct points be given in the plane, and let us use their distance to define the unit distance. Then a length $\alpha \in \mathbb{R}_{\geq 0}$ is constructible with straightedge and compass if and only if α is algebraic and there is a tower of field extensions*

$$\mathbb{Q} = E_0 \subset E_1 \subset \cdots \subset E_{n-1} \subset E_n \subset \mathbb{C}$$

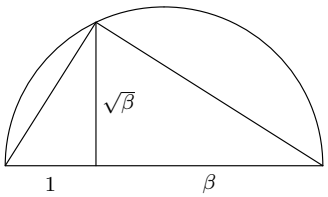
for some $n \in \mathbb{N}$, with each E_k for $k > 0$ a quadratic extension of E_{k-1}, such that $\alpha \in E_n$.

We will more generally describe a negative quanity as constructible if its absolute value is constructible, and a complex value as constructible if its real part and imaginary part are constructible.

Proof. It is easy to see that the sum and difference of constructible complex values are constructible. The analogous claim for product, division by a natural number, and square root reduces easily to the case of real quantities: for β, γ, β', $\gamma' \in \mathbb{R}$ and $N \in \mathbb{N}$,

$$\begin{aligned}(\beta + \gamma\sqrt{-1})(\beta' + \gamma'\sqrt{-1}) &= (\beta\beta' - \gamma\gamma') + (\beta\gamma' + \beta'\gamma)\sqrt{-1},\\ \frac{1}{N}(\beta + \gamma\sqrt{-1}) &= \frac{\beta}{N} + \frac{\gamma}{N}\sqrt{-1},\\ \sqrt{\beta + \gamma\sqrt{-1}} &= \sqrt{\frac{\sqrt{\beta^2+\gamma^2}+\beta}{2}} \pm \sqrt{\frac{\sqrt{\beta^2+\gamma^2}-\beta}{2}}\sqrt{-1},\end{aligned}$$

where the sign in the last formula is $+$ when $\gamma \geq 0$ and otherwise $-$. For real lengths constructions of product and division by a positive integer may be easily given using similar triangles, and for square root we have:



These observations imply that the conditions for α to be constructible are sufficient.

To show they are necessary, we use the initial given pair of points in the plane can be used to define a coordinate system, by taking one of the points as the origin and the other as $(1, 0)$, and we observe that at any point in the construction any pair of lines, line and circle, or pair of circles, will give rise to equations in x and y for the intersection point(s) whose solution requires only addition, multipliciation, division, and square root of coordinates of known points. A given line is assumed to be a line through two points that have already been constructed, and a circle is similarly taken to have center at one and pass through another point that has been constructed.

At each step of a construction a new line or circle is drawn and new intersection points are obtained. If we adjoin all coordinates of constructed points, we obtain a finite tower of quadratic field extensions, starting from $\mathbb{Q}$. Let us write the quadratic extensions as $\mathbb{Q} = E_0 \subset E_1 \subset \cdots \subset E_\ell$. A construction that produces length α yields $\alpha^2 \in E_\ell$. So the claimed condition on α is satisfied (with n equal to ℓ or $\ell + 1$, and in fact with $E_n \subset \mathbb{R}$). □

Example. Let p be a **Fermat prime**, a prime number of the form $2^{2^m} + 1$ for some $m \in \mathbb{N}$. Then we can exhibit a tower of quadratic field extensions, of length $n := 2^m$, from $\mathbb{Q}$ up to $\mathbb{Q}(\gamma)$, where $\gamma := e^{2\pi\sqrt{-1}/p}$. It follows from Theorem 3.19 that $\cos(2\pi/p)$ is constructible, i.e., a regular p-gon is constructible with straightedge and compass. Indeed, $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic of order n; let $a \in \mathbb{Z}$ be such that $\overline{a}$ generates $(\mathbb{Z}/p\mathbb{Z})^\times$, and define $\gamma(j) := \gamma^{a^j}$ for $j \in \mathbb{N}$. We define $E_0 := \mathbb{Q}$ and recursively

$$E_k := E_{k-1}\Bigg(\sum_{j=0}^{2^{n-k}-1} \gamma(2^k j)\Bigg)$$

for $k = 1, \ldots, n$. Using that the field $\mathbb{Q}(\gamma)$ has $\mathbb{Q}$-basis γ, γ^2, $\ldots$, γ^n and automorphism $\sigma\colon \mathbb{Q}(\gamma) \to \mathbb{Q}(\gamma)$, $\sigma(\gamma) := \gamma^a$, of order n, we obtain that E_k is the subfield of $\mathbb{Q}(\gamma)$ of elements fixed by σ^{2^k}, and satisfies $[E_k : \mathbb{Q}] = 2^k$, for all k.

Corollary 3.20. *For a length $\alpha \in \mathbb{R}$ to be constructible, it is necessary that α be an algebraic number, with $[\mathbb{Q}(\alpha) : \mathbb{Q}]$ equal to a power of 2.*

Corollary 3.20 shows $\cos(2\pi/9)$ not constructible (the minimal polynomial is $T^3 - (3/4)T + 1/8$), which implies that the angle $2\pi/3$ cannot be trisected with straightedge and compass. It is a nice exercise to work out the minimal polynomial of $\cos(2\pi/7)$ and show that it also has degree 3. It follows that it is impossible to construct a regular polygon with 7 or 9 sides with straightedge and compass.

Corollary 3.20 shows as well that $\sqrt[3]{2}$ is not constructible, thereby establishing the impossibility of duplicating a cube.

The impossibility of squaring the circle, i.e., the nonconstructibility of $\sqrt{\pi}$, will be established by showing that π is transcendental in the next section.

3.5 Transcendence of e and π

For the transcendence proof we are interested in sums of exponential functions of the form

$$c_1 e^{\alpha_1 x} + \cdots + c_m e^{\alpha_m x},$$

where $\alpha_1, \ldots, \alpha_m$ are pairwise distinct complex numbers.

Theorem 3.21. *Let $u_1, \ldots, u_m \in \mathbb{Z}[T]$ be pairwise distinct irreducible monic polynomials of positive degree, and let $c_1, \ldots, c_m$ be integers, not all zero. Then*

$$c_1\Bigg(\sum_{\substack{\alpha\in\mathbb{C}\\ u_1(\alpha)=0}} e^{\alpha}\Bigg) + \cdots + c_m\Bigg(\sum_{\substack{\alpha\in\mathbb{C}\\ u_m(\alpha)=0}} e^{\alpha}\Bigg) \neq 0.$$

The proof of Theorem 3.21 will occupy the rest of this section. Before we embark on the proof, we point out some consequences.

Corollary 3.22. *The number e is transcendental.*

Proof. We apply Theorem 3.21 to $u_1 = T$, $u_2 = T - 1$, $\ldots$, $u_{n+1} = T - n$ to exclude the possibility that e satisfies a nontrivial polynomial equation with integer coefficients of degree n. □

Corollary 3.23. *The number π is transcendental.*

Proof. Let $v \in \mathbb{Z}[T]$ be irreducible, monic, and $\neq T$, and let $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ be the roots of v. We define $v_k := \Sigma_{(1^k)}(v)$ in the notation of Proposition 3.2, with $(1^k) := (1, \ldots, 1, 0, \ldots, 0) \in \mathbb{Z}^n$ where 1 occurs k times. This means,

$$v_k = \prod_{1\le j_1<\cdots<j_k\le n} \big(T - (\alpha_{j_1} + \cdots + \alpha_{j_k})\big).$$

Let $u_1, \ldots, u_m$ be the irreducible monic factors of the polynomials $v_0, \ldots, v_n$, with $u_i \neq u_j$ for $i \neq j$, and let

$$v_k = u_1^{g_{k1}} \cdots u_m^{g_{km}}$$

be the decomposition into irreducible factors, for all k (with $g_{k\ell} \in \mathbb{N}$ for all k and ℓ, and $g_{k\ell} = 0$ whenever $u_\ell \nmid v_k$). This means, if we define

$$c_\ell := \sum_{k=0}^{n} (-1)^k g_{k\ell},$$

then we have

$$\prod_{j=1}^{n}(e^{\alpha_j x}-1) = (-1)^n \sum_{\ell=1}^{m} c_\ell \sum_{\substack{\alpha\in\mathbb{C}\\ u_\ell(\alpha)=0}} e^{\alpha x}. \tag{1}$$

The left-hand side of (1) is nonzero for all but countably many values of x, hence the c_ℓ are not all zero. By Theorem 3.21, the value at $x = 1$ is nonzero, and hence

$$\{\alpha_1, \ldots, \alpha_n\} \cap 2\pi\sqrt{-1}\mathbb{Z} = \emptyset.$$

Any nonzero algebraic integer appears among $\alpha_1, \ldots, \alpha_n$ for suitable v, so this shows that $\pi\sqrt{-1}$ is transcendental, and hence π is transcendental. □

The proof of Theorem 3.21 splits up into an analytic part and a number-theoretic part. The analytic part begins with the introduction of approximations to the exponential function and suitable error estimate for these. The exponential function e^x is approximated by truncations

$$E_q(x) := \sum_{j=0}^{q} \frac{x^j}{j!}$$

of its Taylor series at 0. We will consider approximations which are certain linear combinations of the E_q.

Let $\underline{r} = (r_0, \ldots, r_N)$ be a sequence of complex numbers, and let us use these as weights to define linear combinations of the E_q:

$$E_{\underline{r}}(x) := r_0E_0(x) + \cdots + r_NE_N(x), \qquad R := r_0 + \cdots + r_N.$$

When $R \neq 0$ it makes sense to consider

$$\frac{E_{\underline{r}}(x)}{R},$$

and ask how well this approximates e^x.

Proposition 3.24. *Given a sequence of complex numbers $\underline{r} = (r_0, \ldots, r_N)$ we define*

$$f_{\underline{r}}(x) := r_0 + r_1x + \frac{r_2}{2!}x^2 + \cdots + \frac{r_N}{N!}x^N.$$

Then

$$|E_{\underline{r}}(x) - Re^x| \leq |x|e^{|x|} \sup_{w\in\mathbb{C},\,|w|\leq|x|} |f_{\underline{r}}(w)|$$

for all $x \in \mathbb{C}$.

The proof uses the following result, which is a straightforward calculation.

Lemma 3.25. *With the notation of Proposition 3.24 we have*

$$E_{\underline{r}}(x) = f_{\underline{r}}(x) + f'_{\underline{r}}(x) + f''_{\underline{r}}(x) + \cdots + f^{(N)}_{\underline{r}}(x),$$
$$\frac{d}{dx} e^{-x} E_{\underline{r}}(x) = -e^{-x} f_{\underline{r}}(x).$$

Proof of Proposition 3.24. When x is real we may directly apply Lemma 3.25:

$$|E_{\underline{r}}(x) - Re^x| = \left| \int_0^x e^{x-t} f_{\underline{r}}(t) dt \right| \leq |x| e^{|x|} \sup_{w \in \mathbb{R},\, |w| \leq |x|} |f_{\underline{r}}(w)|.$$

For the general case, we write $x = |x| x_0$ with $x_0 \in \mathbb{C}$, $|x_0| = 1$, obtain from Lemma 3.25 that $\frac{d}{dt} e^{-x_0 t} E_{\underline{r}}(x_0 t) = -x_0 e^{-x_0 t} f_{\underline{r}}(x_0 t)$, and argue as above. □

Now we consider the polynomial $f_{\underline{r}}$ from Proposition 3.24 and examine the effect of raising $f_{\underline{r}}$ to a high power, thus obtaining from the coefficients new weight-sequences $\underline{r}$. For $n \in \mathbb{N}_{>0}$ we define $\underline{r}^{(n)}$ to be the sequence of complex numbers $(r_0^{(n)}, \ldots, r_{nN}^{(n)})$ defined by the condition

$$f_{\underline{r}^{(n)}} = f_{\underline{r}}^n, \tag{2}$$

and we define

$$R^{(n)} := r_0^{(n)} + \cdots + r_{nN}^{(n)}. \tag{3}$$

Proposition 3.26. *With the notation of Theorem 3.21, and the sequences $\underline{r}^{(n)}$ and numbers $R^{(n)}$ from (2)–(3), there exists $n_0 \in \mathbb{N}_{>0}$ such that*

$$\left| \sum_{j=1}^m c_j \Big(\sum_{u_j(\alpha)=0} E_{\underline{r}^{(n)}}(\alpha) \Big) - R^{(n)} \sum_{j=1}^m c_j \Big(\sum_{u_j(\alpha)=0} e^\alpha \Big) \right| < n!$$

for all $n \geq n_0$.

Proof. We define C to be the largest absolute value of any root of any of the polynomials $u_1, \ldots, u_m$, and set $S := \sup_{w \in \mathbb{C},\, |w| \leq C} |f_{\underline{r}}(w)|$. Then by Proposition 3.24 we have

$$\left| c_j \Big(\sum_{u_j(\alpha)=0} E_{\underline{r}^{(n)}}(\alpha) \Big) - R^{(n)} c_j \Big(\sum_{u_j(\alpha)=0} e^\alpha \Big) \right| \leq |c_j| \deg(u_j) C e^C S^n$$

for $j = 1, \ldots, m$. Summing over j we obtain a bound of the form DS^n for some positive real number D. We have $\lim_{n \to \infty} DS^n/n! = 0$, so there exists n_0 as claimed. □

To make the transition to the number-theoretic part of the proof we give a lemma which involves linear algebra.

Lemma 3.27. *Let $u \in \mathbb{C}[T]$ be a nonconstant polynomial without multiple roots, and let $\beta_1, \ldots, \beta_n \in \mathbb{C}$ be the roots of u. Define $v_j := T^j u$ for $j = 0, \ldots, n-1$. Then $\det(v'_{j-1}(\beta_i))_{1 \le i,j \le n} \neq 0$.*

Proof. We have $v'_j(\beta_i) = \beta_i^j u'(\beta_i)$, so the statement follows from the nonvanishing of the Vandermonde determinant. □

For the number-theoretic part of the proof we will restrict our attention to the case where the polynomial $f_{\underline{r}}$ from Proposition 3.24 has integer coefficients. This implies by Lemma 3.25 that $E_{\underline{r}}(x)$ is a polynomial with integer coefficients. Since the sum of the kth powers of the roots of a monic polynomial with integer coefficients is an integer, as we saw explicitly in the proof of Proposition 3.2, the sum

$$\sum_{u_j(\alpha)=0} E_{\underline{r}}(\alpha)$$

takes an integer value, for every j. The same is true more generally for all the $E_{\underline{r}^{(n)}}$.

We introduce the choice of $f_{\underline{r}}$ that will be used in the number-theoretic part of the proof. By Lemma 3.27, applied to the product $u := u_1 \cdots u_m$, we may find a natural number ℓ less than the degree of u, such that

$$f(x) := x^\ell u(x)$$

satisfies

$$\sum_{j=1}^{m} c_j \sum_{u_j(\alpha)=0} f'(\alpha) \neq 0. \tag{4}$$

Then we take $\underline{r} = (r_0, \ldots, r_N)$, with $N := \ell + \deg(u)$, to satisfy

$$f_{\underline{r}} = f,$$

and $r^{(n)}$ and $R^{(n)}$ to correspond to $f_{\underline{r}}^n$ as in (2)–(3), for all $n \in \mathbb{N}_{>0}$.

Recalling the integer n_0 from Proposition 3.26, we fix n to be a prime number p greater than or equal to n_0 and not a factor of the nonzero integer appearing in (4). We will henceforth write p instead of n as a reminder that we have made the choice of a prime number. Define

$$\tilde{f} := f^p,$$

so that $f_{\underline{r}^{(p)}} = \tilde{f}$.

Lemma 3.28. *The polynomials $\tilde{f}$, $\tilde{f}'$, $\ldots$, $\tilde{f}^{(p-1)}$ are divisible by u.*

Proof. In fact, the kth derivative $\tilde{f}^{(k)}$, for $k \le p-1$, is divisible by u^{p-k}, as can be seen by induction, using the formula for the derivative of a product. □

It results from Lemma 3.28 that if we define

$$\widetilde{E}(x) := \tilde{f}^{(p)}(x) + \tilde{f}^{(p+1)}(x) + \cdots + \tilde{f}^{(pN)}(x)$$

then

$$\sum_{u_j(\alpha)=0} E_{\underline{r}^{(p)}}(\alpha) = \sum_{u_j(\alpha)=0} \widetilde{E}(\alpha)$$

for every j. The coefficient of the pth derivative of a monomial x^s is $p!\binom{s}{p}$. This is a multiple of $p!$, hence so are the coefficients of $\widetilde{E}(x)$. In fact, more is true, as revealed in the next result.

Proposition 3.29. *For every j the quantity*

$$\frac{1}{p!} \sum_{u_j(\alpha)=0} \widetilde{E}(\alpha) \tag{5}$$

is an integer, congruent to $\sum_{u_j(\alpha)=0} f'(\alpha)$ mod p.

Proof. Modulo u we have $(1/p!)\tilde{f}^{(p)}(x)$ equal to $(x^\ell u'(x))^p$. Again modulo u, the latter is equal to $(f'(x))^p$. Raising to the pth power acts as a ring endomorphism on algebraic integers modulo p in a manner that restricts to the trivial action on $\mathbb{Z}/p\mathbb{Z}$, so we are done if we can show that each term $(1/p!)\tilde{f}^{(k)}(x)$ with $k \geq p+1$ contributes trivially to (5) mod p. For any i not divisible by p the coefficient of x^i in $\tilde{f}$ is divisible by p, hence the coefficient of x^{i-k} in $\tilde{f}^{(k)}$ is divisible by $p \cdot p!$, while for i divisible by p we get a multiple of $p \cdot p!$ just from the kth derivative of x^i. □

The proof of Theorem 3.21 is concluded by combining the analytic result Proposition 3.26 with the number-theoretic information in Proposition 3.29. We have

$$\left| \sum_{j=1}^{m} c_j \Bigg(\sum_{u_j(\alpha)=0} E_{\underline{r}^{(p)}}(\alpha) \Bigg) - R^{(p)} \sum_{j=1}^{m} c_j \Bigg(\sum_{u_j(\alpha)=0} e^{\alpha} \Bigg) \right| < p! \tag{6}$$

from Proposition 3.26. On the other hand, Proposition 3.29 shows that

$$\sum_{j=1}^{m} c_j \Bigg(\sum_{u_j(\alpha)=0} E_{\underline{r}^{(p)}}(\alpha) \Bigg) \tag{7}$$

is equal to $p!$ times an integer congruent to $\sum_{j=1}^{m} c_j \sum_{u_j(\alpha)=0} f'(\alpha)$ mod p. So the quantity in (7) is equal to $p!$ times a nonzero integer. This gives $p!$ as a lower bound for the absolute value of the quantity in (7). For this to be compatible with (6) we must have

$$\sum_{j=1}^{m} c_j \Bigg(\sum_{u_j(\alpha)=0} e^{\alpha} \Bigg) \neq 0,$$

and the theorem is proved.

This brief excursion into transcendence theory loosely follows the text *Transzendenz von e und π* by G. Hessenberg.

The transcendence of π (Corollary 3.23), first proved by Ferdinand von Lindemann in 1882, establishes the impossibility of squaring the circle by straightedge and compass. Appendix D gives some historical context surrounding the results in this section.

A Alternate proof of the First Sylow theorem

Number theory is a rich area of mathematics, devoted primarily to the study of the integers. Combinatorics is another area of mathematics, where the focus is mainly on questions of the form, "How many ...?" As with many different areas of mathematics, these two areas are linked, since the answer to a combinatorial question is an integer, and it is then natural to ask about properties of this integer. We do this, with the combinatorial question about subsets of given cardinality of a given finite set and the number-theoretic investigation into the prime factors of an integer. The result (Proposition A.1, below) leads to a quick alternate proof (§A.2) of the First Sylow theorem.

A.1 Prime factors of binomial coefficients

The binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

is significant in algebra, appearing as the coefficient of $x^k y^{n-k}$ in the binomial expansion $(x+y)^n$, and in combinatorics, as the number of subsets of $\{1, \ldots, n\}$ having cardinality k.

The binomial coefficients are assembled into Pascal's triangle, listing those with $n = 0$ on the first line, with $n = 1$ on the second line, and so on:

$$\begin{array}{ll}
(n=0) & 1 \\
(n=1) & 1 \quad 1 \\
(n=2) & 1 \quad 2 \quad 1 \\
(n=3) & 1 \quad 3 \quad 3 \quad 1 \\
(n=4) & 1 \quad 4 \quad 6 \quad 4 \quad 1 \\
(n=5) & 1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
(n=6) & 1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
\cdots &
\end{array}$$

The formula (for $n \in \mathbb{N}_{>0}$)

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

tells us that each entry is the sum of the two entries just above it, so it is easy to fill in further rows.

Thus far, this is familiar from school mathmatics. We enter new territory by exploring the pattern of even and odd entries in Pascal's triangle. Modulo 2 we obtain:

$$\begin{array}{ll}
(n = \quad 0_2) & 1 \\
(n = \quad 1_2) & 1 \quad 1 \\
(n = \quad 10_2) & 1 \quad 0 \quad 1 \\
(n = \quad 11_2) & 1 \quad 1 \quad 1 \quad 1 \\
(n = 100_2) & 1 \quad 0 \quad 0 \quad 0 \quad 1 \\
(n = 101_2) & 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \\
(n = 110_2) & 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1
\end{array}$$

We have used base 2 to write the integer n in the left-hand column.

The following result confirms the pattern that we see above: the binomial coefficient $\binom{n}{k}$ is odd if and only if the algorithm from school to compute the sum of k and $n-k$, applied to the integers in their binary representation, runs without carry. For instance, for $n := 5$ we require a carry when $k = 3$ but not when $k = 4$.

$$\begin{array}{r} {\scriptstyle 1} \\ 11_2 \\ 10_2 \\ \hline 101_2 \end{array} \qquad\qquad \begin{array}{r} \\ 100_2 \\ 1_2 \\ \hline 101_2 \end{array}$$

More generally, divisibility by p is reflected in the addition of integers in their base p expansion for any prime p.

Proposition A.1. *Let p be a prime, and let n and k be natural numbers, with $k \leq n$. Then $\binom{n}{k}$ is relatively prime to p if and only if, for every $\ell \in \mathbb{N}_{>0}$, we have*

$$\left\lfloor \frac{k}{p^\ell} \right\rfloor + \left\lfloor \frac{n-k}{p^\ell} \right\rfloor = \left\lfloor \frac{n}{p^\ell} \right\rfloor. \tag{1}$$

As usual, $\lfloor - \rfloor$ denotes the greatest integer function: for $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the greatest integer, less than or equal to x. We observe, in (1) the left-hand side is always less than or equal to the right-hand side.

We claim that the multiplicity of p is the prime factorization of $n!$ is

$$\sum_{\ell=1}^{\infty} \left\lfloor \frac{n}{p^\ell} \right\rfloor. \tag{2}$$

(The summands become zero as soon as p^ℓ exceeds n, so (2) is equal to the finite sum over the positive integers ℓ satisfying $p^\ell \leq n$.) Notice that the claim implies Proposition A.1. Since $n! = n \cdot (n-1)!$ for $n > 0$ and the multiplicity of p in the product of two nonzero integers is equal to the sum of the multiplicities of p in the factors, the claim is immediate from

$$\left\lfloor \frac{n}{d} \right\rfloor - \left\lfloor \frac{n-1}{d} \right\rfloor = \begin{cases} 1, & \text{if } d \mid n, \\ 0, & \text{if } d \nmid n, \end{cases} \tag{3}$$

for $d \in \mathbb{N}_{>0}$. From the definition of $\lfloor - \rfloor$, (3) is clear.

For example, if we write $n \in \mathbb{N}_{>0}$ as $n = p^e m$, with e, $m \in \mathbb{N}$ and $p \nmid m$, then for $k := p^e$, (1) holds for all ℓ. (For $\ell \leq e$ this is clear, and for $\ell > e$, (3) yields $\lfloor (m-1)/p^{\ell-e} \rfloor = \lfloor m/p^{\ell-e} \rfloor$.) By Proposition A.1, $\binom{n}{p^e}$ is relatively prime to p.

A.2 From Proposition A.1 to the First Sylow theorem

Now let G be a finite group whose order is divisible by p, and let us write $|G| = p^e m$ with e, $m \in \mathbb{N}$ and $p \nmid m$. So, if we define

$$X := \{\text{subsets } S \subset G \mid S \text{ has cardinality } p^e\},$$

then p does not divide $|X|$.

Left-multiplication on subsets of G induces an action of G on X. It follows from the orbit formula that there exists $S \in X$ with stabilizer

$$\{g \in G \mid gS = S\} \tag{1}$$

of order divisible by p^e.

Pick $s_0 \in S$. Clearly the stabilizer (1) is contained in

$$S_0 := \{s s_0^{-1} \mid s \in S\}.$$

Now $|S_0| = |S| = p^e$. It follows that the stabilizer (1) is equal to S_0. So we have exhibited a p-Sylow subgroup of G.

B Epimorphisms, free, amalgmated products of groups

In some texts, monomorphisms and epimorphisms of groups or modules are defined to be injective homomorphisms, respectively surjective homomorphisms. As pointed out in §1.4, the terms monomorphism and epimorphism also refer to cancelability properties. Here we show the equivalence the respective notions.

For a homomorphism of modules over a commutative ring, injectivity is equivalent to left cancelability. One implication is obvious, and for the other implication we just notice that the inclusion homomorphism of a kernel is different from the trivial homomorphism, when the kernel is nontrivial. The same holds, with the same justification, in the setting of groups.

Now suppose we try to show, in the same fashion, that surjectivity is equivalent to right cancelability. For modules this is no problem, since we have the canonical homomorphism to the cokernel, and when the cokernel is nontrivial this is different from the trivial homomorhpism. But if we try to use the same argument for groups, we run into the problem that we only have canonical homomorphisms $H \to H/H'$ for *normal* subgroups $H' \subset H$.

What we need is a more general construction, which to any proper subgroup H' of H associates a group G with a pair of distinct group homomorphsms $H \rightrightarrows G$, such that the two composite homomorphisms

$$H' \to H \rightrightarrows G$$

are equal.

B.1 Free product

For any pair of groups H_1 and H_2, there is a **free product** $H_1 * H_2$, which like the free group is a set of words in an alphabet with a carefully defined product. In this case the alphabet is the disjoint union $(H_1 \smallsetminus \{e_{H_1}\}) \sqcup (H_2 \smallsetminus \{e_{H_2}\})$ of the non-identity elements from the two groups. (In general H_1 and H_2 may have elements in common or even be equal, so we should use notation such as (h, i) or subscripts i, $i \in \{1, 2\}$, to avoid confusion.) The requirement never to have adjacent symbols from the same group defines the set of words S. The rule for multiplying two elements of S is, informally, to concatenate and combine. We will not write this out formally and check the group axioms. Instead, following A. G. Kurosh's text *The Theory of Groups* we will obtain the group structure indirectly using group actions.

We denote the empty word in S by e_S. Now we define a pair of group actions, of H_1 and of H_2, on S. Let $h_1 \in H_1$ be a non-identity element. Then:

$$h_1 \cdot \colon \begin{cases} e_S \mapsto h_1, \\ h_1^{(1)} h_2^{(2)} h_1^{(3)} \cdots \mapsto \begin{cases} h_2^{(2)} h_1^{(3)} \dots & \text{if } h_1^{(1)} = h_1^{-1}, \\ (h_1 h_1^{(1)}) h_2^{(2)} h_1^{(3)} \dots & \text{otherwise,} \end{cases} \\ h_2^{(1)} h_1^{(2)} h_2^{(3)} \cdots \mapsto h_1 h_2^{(1)} h_1^{(2)} h_2^{(3)} \dots . \end{cases} \tag{1}$$

For $h_2 \in H_2$, not the identity:

$$h_2\cdot\colon \begin{cases} e_S \mapsto h_2, \\ h_1^{(1)}h_2^{(2)}h_1^{(3)}\cdots \mapsto h_2h_1^{(1)}h_2^{(2)}h_1^{(3)}\ldots, \\ h_2^{(1)}h_1^{(2)}h_2^{(3)}\cdots \mapsto \begin{cases} h_1^{(2)}h_2^{(3)}\ldots & \text{if } h_2^{(1)} = h_2^{-1}, \\ (h_2h_2^{(1)})h_1^{(2)}h_2^{(3)}\ldots & \text{otherwise.} \end{cases} \end{cases} \tag{2}$$

The verification that these are group actions requires the treatment of several cases, but is straightforward. Since e_S is always sent to a nonempty word, these are faithful group actions.

We let Γ denote the subgroup of $\mathrm{Perm}(S)$ generated by the images of the permutation representations (Proposition 2.15) associated to these group actions. Again using the correspondence between group actions and homomorphisms to $\mathrm{Perm}(S)$ we obtain an action of Γ on S, whose permutation representation is the inclusion of Γ in $\mathrm{Perm}(S)$.

Lemma B.1. *Let $s \in S$, $s \neq e_S$.*
(i) *If $s = h_1^{(1)}h_2^{(2)}h_1^{(3)}\ldots$, then $s = \gamma \cdot e_S$, where $\gamma := (h_1^{(1)}\cdot)(h_2^{(2)}\cdot)(h_1^{(3)}\cdot)\ldots$.*
(ii) *If $s = h_2^{(1)}h_1^{(2)}h_2^{(3)}\ldots$, then $s = \gamma \cdot e_S$, where $\gamma := (h_2^{(1)}\cdot)(h_1^{(2)}\cdot)(h_2^{(3)}\cdot)\ldots$.*

Proof. We prove (i)–(ii) by induction on the length $\ell \in \mathbb{N}_{>0}$ of s. The base case $\ell = 1$ is a direct consequence of (1)–(2). Suppose $\ell > 1$, and let us prove (i). We have $\gamma = (h_1^{(1)}\cdot)\gamma'$ where $\gamma' := (h_2^{(2)}\cdot)(h_1^{(3)}\cdot)\ldots$. Applying the induction hypothesis and (1) we have $\gamma \cdot e_S = h_1^{(1)} \cdot (\gamma' \cdot e_S) = h_1^{(1)} \cdot s' = s$ where $s' := h_2^{(2)}h_1^{(3)}\ldots$. The proof of (ii) is similar. □

Lemma B.2. *Let $\gamma \in \Gamma$, $\gamma \neq \mathrm{id}_S$. Then $\gamma \cdot e_S \neq e_S$.*

Proof. Let γ be represented as a product of generators $h_i\cdot$ of minimal length $\ell \in \mathbb{N}_{>0}$. If this is a product $\ldots(h_i\cdot)(h'_{i'}\cdot)\ldots$ then we claim $i' \neq i$. Indeed, if $i' = i$ then we may use the equality $(h_i\cdot)(h'_i\cdot) = (h_ih'_i)\cdot$ to give a representation of γ as a product of fewer than ℓ generators, which contradicts the minimality of ℓ. By Lemma B.1, $\gamma \cdot e_S$ is a word of length ℓ. In particular, we have $\gamma \cdot e_S \neq e_S$. □

Lemmas B.1 and B.2 show that the action of Γ on S is simply transitive. Hence $\gamma \mapsto \gamma \cdot e_S$ is a bijective map, and in this way we may obtain from Γ a group structure on S having e_S as identity element. A non-identity element $h_1 \in H_1$ is sent to h_1 as word of length 1, and similarly a non-identity element $h_2 \in H_2$ is sent to h_2 as word of length 1. In this way we have, for $i \in \{1, 2\}$, a copy of H_i sitting inside S. And we have $h_is = h_i \cdot s$ as defined in (1)–(2) (since, in the notation of Lemma B.1, $h_is = h_i \cdot (\gamma \cdot e_S)$).

The group S satisfies a universal property for pairs of group homomorphisms, one from H_1 and one from H_2. Let T be a group and let $\psi_1\colon H_1 \to T$ and $\psi_2\colon H_2 \to T$ be group homomorphisms. Then there is a unique group homomorphism $f\colon S \to T$ whose restriction to the copy of H_i in S is ψ_i for $i = 1, 2$:

$$h_1^{(1)}h_2^{(2)}h_1^{(3)}\ldots \overset{f}{\mapsto} \psi_1(h_1^{(1)})\psi_2(h_2^{(2)})\psi_1(h_1^{(3)})\ldots, \tag{3}$$

$$h_2^{(1)}h_1^{(2)}h_2^{(3)}\ldots \overset{f}{\mapsto} \psi_2(h_2^{(1)})\psi_1(h_1^{(2)})\psi_2(h_2^{(3)})\ldots, \tag{4}$$

Since a similar verification was made in Proposition 2.19, we omit the details and just point out the consequence, that thanks to the universal property which characterizes the group S up to canonical isomorphism, we permit ourselves to use the standard notation $H_1 * H_2$ from now on for the group S.

The special case $H_1 = H_2$ is of interest to us. In fact, by comparing the universal property here with that of Proposition 2.19, the reader may deduce

$$\mathbb{Z} * \mathbb{Z} \cong F_2.$$

B.2 Amalgamated product

From now on we fix a group H and consider the free product $H * H$ corresponding to the special case $H_i := H$, $i = 1, 2$, mentioned above. We replace subscripts with underlines, so an element of the alphabet $(H \smallsetminus \{e_H\}) \sqcup (H \smallsetminus \{e_H\})$ will be denoted by $\underline{h}$ (for h from the first copy of $H \smallsetminus \{e_H\}$) or $\underline{\underline{h}}$ (for h from the second copy of $H \smallsetminus \{e_H\}$). Elements of $H * H$ are the empty word e_{H*H} and nonempty words of the form $\underline{h}^{(1)}\underline{\underline{h}}^{(2)}\underline{h}^{(3)}\ldots$ and $\underline{\underline{h}}^{(1)}\underline{h}^{(2)}\underline{\underline{h}}^{(3)}\ldots$.

Now we let H' be a proper subgroup of H. We choose a set $K \subset H$ of right H'-coset representatives,

$$H = \bigsqcup_{k\in K} H'k,$$

subject to the requirement $e_H \in K$. We claim: (i) the quotient of $H * H$ by the normal closure of $\{\underline{g}^{-1}\underline{\underline{g}} \,|\, g \in H' \smallsetminus \{e_H\}\}$ is universal for pairs of homomorphisms from H having the same restriction to H'; (ii) each element of the quotient $H * H/\langle \underline{g}^{-1}\underline{\underline{g}} \,|\, g \in H' \smallsetminus \{e_H\}\rangle^{H*H}$ has a unique representative which is e_{H*H}, a word $\underline{g}$ of length 1 with $g \in H' \smallsetminus \{e_H\}$, or a nonempty word $\underline{h}^{(1)}\underline{\underline{k}}^{(2)}\underline{k}^{(3)}\underline{\underline{k}}^{(4)}\ldots$ or $\underline{\underline{h}}^{(1)}\underline{k}^{(2)}\underline{\underline{k}}^{(3)}\underline{k}^{(4)}\ldots$ with $h^{(1)} \in H \smallsetminus H'$ and $k^{(2)}, k^{(3)}, \ldots \in K \smallsetminus \{e_H\}$.

Given (i) we may denote $H * H/\langle \underline{g}^{-1}\underline{\underline{g}} \,|\, g \in H' \smallsetminus \{e_H\}\rangle^{H*H}$ by $H *_{H'} H$, the **amalgamated product** of H with itself along the subgroup H'. Then we have

$$H \rightrightarrows H *_{H'} H$$

(the composites of the two inclusions of H in $H * H$ with the canonical homomorphism $H * H \to H *_{H'} H$), such that the two composites

$$H' \to H \rightrightarrows H *_{H'} H$$

are the same. By (ii), any element of $H \smallsetminus H'$ has distinct images by the two homomorphisms from H to $H *_{H'} H$. In other words, once we have verified the claims, we will have also verified that every epimorphism of groups is surjective.

Claim (i) is straightforward to verify. For a group T, composition with the canonical homomorphism $H * H \to H * H/\langle \underline{g}^{-1}\underline{\underline{g}} \,|\, g \in H' \smallsetminus \{e_H\}\rangle^{H*H}$ identifies homomorphisms from the quotient to T with homomorphisms $H * H \to T$ with kernel containing $\underline{g}^{-1}\underline{\underline{g}}$ for all $g \in H' \smallsetminus \{e_H\}$. By the universal property from §B.1 the latter set is identified with the set of pairs of homomorphisms $H \to T$ having the same restriction to H'.

The proof of (ii) will use the technique of group action on words. We consider the alphabet $H' \sqcup (K \smallsetminus \{e_H\}) \sqcup (K \smallsetminus \{e_H\})$, where we employ underlines as we have done to distinguish between the two copies of $K \smallsetminus \{e_H\}$. We define $\widetilde{S}$ to be the set of words in this alphabet which are nonempty, with the first symbol taken from H' and all remaining symbols, if any, taken from the rest of the alphabet, alternatingly from the two copies of $(K \smallsetminus \{e_H\})$.

We define a pair of group actions of H on $\widetilde{S}$. The first action, denoted by $h\cdot_1$ for $h \in H$, is defined by

$$\begin{aligned}
g &\overset{h\cdot_1}{\mapsto} \begin{cases} g^* & \text{if } h \in H', \text{ with } g^* := hg, \\ g^*\underline{k}^* & \text{if } hg = g^*k^* \text{ with } g^* \in H',\ k^* \in K \smallsetminus \{e_H\}, \end{cases} \\
g\underline{k}^{(1)}\underline{\underline{k}}^{(2)}\ldots &\overset{h\cdot_1}{\mapsto} \begin{cases} g^*\underline{\underline{k}}^{(2)}\ldots & \text{if } hgk^{(1)} \in H', \text{ with } g^* := hgk^{(1)}, \\ g^*\underline{k}^*\underline{\underline{k}}^{(2)}\ldots & \text{if } hgk^{(1)} = g^*k^* \text{ with } g^* \in H',\ k^* \in K \smallsetminus \{e_H\}, \end{cases} \\
g\underline{\underline{k}}^{(1)}\underline{k}^{(2)}\ldots &\overset{h\cdot_1}{\mapsto} \begin{cases} g^*\underline{\underline{k}}^{(1)}\underline{k}^{(2)}\ldots & \text{if } h \in H', \text{ with } g^* := hg, \\ g^*\underline{k}^*\underline{\underline{k}}^{(1)}\underline{k}^{(2)}\ldots & \text{if } hg = g^*k^* \text{ with } g^* \in H',\ k^* \in K \smallsetminus \{e_H\}. \end{cases}
\end{aligned} \tag{1}$$

The second action, denoted by $h\cdot_2$ for $h \in H$, is defined by

$$\begin{aligned}
g &\overset{h\cdot_2}{\mapsto} \begin{cases} g^* & \text{if } h \in H', \text{ with } g^* := hg, \\ g^*\underline{\underline{k}}^* & \text{if } hg = g^*k^* \text{ with } g^* \in H',\ k^* \in K \smallsetminus \{e_H\}, \end{cases} \\
g\underline{k}^{(1)}\underline{\underline{k}}^{(2)}\ldots &\overset{h\cdot_2}{\mapsto} \begin{cases} g^*\underline{k}^{(1)}\underline{\underline{k}}^{(2)}\ldots & \text{if } h \in H', \text{ with } g^* := hg, \\ g^*\underline{\underline{k}}^*\underline{k}^{(1)}\underline{\underline{k}}^{(2)}\ldots & \text{if } hg = g^*k^* \text{ with } g^* \in H',\ k^* \in K \smallsetminus \{e_H\}, \end{cases} \\
g\underline{\underline{k}}^{(1)}\underline{k}^{(2)}\ldots &\overset{h\cdot_2}{\mapsto} \begin{cases} g^*\underline{k}^{(2)}\ldots & \text{if } hgk^{(1)} \in H', \text{ with } g^* := hgk^{(1)}, \\ g^*\underline{\underline{k}}^*\underline{k}^{(2)}\ldots & \text{if } hgk^{(1)} = g^*k^* \text{ with } g^* \in H',\ k^* \in K \smallsetminus \{e_H\}. \end{cases}
\end{aligned} \tag{2}$$

Verifying that these are group actions, as before, requires consideration of several cases but is straightforward. For $h \in H'$ and $\tilde{s} \in \widetilde{S}$ we have $h \cdot_1 \tilde{s} = h \cdot_2 \tilde{s}$. Therefore

the corresponding pair of homomorphisms $H \to \mathrm{Perm}(\widetilde{S})$ induce a homomorphism $H *_{H'} H \to \mathrm{Perm}(\widetilde{S})$, the permutation representation of an action of $H *_{H'} H$ on $\widetilde{S}$.

The word $e_{\widetilde{S}} := e_H$ will play the role played by the empty word in §B.1.

Lemma B.3. *Let $\tilde{s} \in \widetilde{S}$, $\tilde{s} \neq e_{\widetilde{S}}$.*
(i) *If $\tilde{s} = g\underline{k}^{(1)}\underline{\underline{k}}^{(2)}\dots$, then $\tilde{s} = \bar{s} \cdot e_{\widetilde{S}}$ with $s := \underline{h}^{(1)}\underline{\underline{k}}^{(2)}\dots$, where $h^{(1)} := gk^{(1)}$ (or $h^{(1)} := g$ if $\tilde{s} = g$).*
(ii) *If $\tilde{s} = g\underline{\underline{k}}^{(1)}\underline{k}^{(2)}\dots$, then $\tilde{s} = \bar{s} \cdot e_{\widetilde{S}}$ with $s := \underline{\underline{h}}^{(1)}\underline{k}^{(2)}\dots$, where $h^{(1)} := gk^{(1)}$.*

Proof. We prove (i)–(ii) by induction on the length $\ell \in \mathbb{N}_{>0}$ of $\tilde{s}$. The base cases $\ell = 1$ and $\ell = 2$ are direct consequences of (1)–(2). If $\ell > 2$, then we obtain (i) by writing $s = \underline{h}^{(1)}s'$ with $s' := \underline{\underline{k}}^{(2)}\dots$, applying the induction hypothesis to obtain $e_H\underline{\underline{k}}^{(2)}\dots = \bar{s}'\cdot e_{\widetilde{S}}$, and applying (1) to obtain $\bar{s}\cdot e_{\widetilde{S}} = h^{(1)} \cdot_1 e_H\underline{\underline{k}}^{(2)}\dots = g\underline{k}^{(1)}\underline{\underline{k}}^{(2)}\dots$, and we obtain (ii) in a similar fashion. □

Lemma B.4. *For every non-identity element $\bar{s} \in H *_{H'} H$ we have $\bar{s} \cdot e_{\widetilde{S}} \neq e_{\widetilde{S}}$.*

Proof. Given an element of $H * H$ we introduce, besides the length ℓ, another numerical quantity m, defined to be the largest integer in $\{0, \dots, \ell\}$, so that for every j with $m < j \leq \ell$ the jth symbol belongs to $(K \smallsetminus \{e_H\}) \sqcup (K \smallsetminus \{e_H\})$. We suppose that, to $\bar{s}$, a representative $s \in H * H$ is chosen of minimal length ℓ, and among all representatives of length ℓ we suppose that the quantity m is minimal. We claim, $m \leq 1$. Suppose, to the contrary, $m \geq 2$, and let us suppose that the mth symbol of s is $\underline{h}'$ or $\underline{\underline{h}}'$, with $h' \in H \smallsetminus K$. We write $h' = g'k'$ with $g' \in H' \smallsetminus \{e_H\}$ and $k' \in K$. Now let us suppose that s is $w\,\underline{\underline{h}}\,\underline{h}'\,w'$ or $w\,\underline{h}\,\underline{\underline{h}}'\,w'$, for some w, $w' \in H * H$, with w of length $m-2$ and w' of length $\ell - m$. If $h = g'^{-1}$ or $k' = e_H$ then we obtain a contradiction to the minimality of ℓ, and otherwise we obtain a contradiction to the minimality of m. We do not write out all the cases, but for instance if $h = g'^{-1}$ and $k' \neq e_H$ then we have $s = w\,\underline{\underline{g}}'^{-1}\underline{g}'\underline{k}'\,w'$ or $s = w\,\underline{g}'^{-1}\underline{\underline{g}}'\underline{\underline{k}}'\,w'$, and another representative of the same class in $H *_{H'} H$ is $w\,\underline{k}'\,w'$, respectively $w\,\underline{\underline{k}}'\,w'$, which leads to a contradiction to the minimality of ℓ. So $m \leq 1$, and now an application of Lemma B.3 completes the proof. □

Lemmas B.3 and B.4 imply that $H *_{H'} H$ acts simply transitively on $\widetilde{S}$, and claim (ii) is established.

C Algebraic closure

We record a simple argument for the existence of an algebraic closure of a general field, taken from D. J. H. Garling, *A Course in Galois Theory*. We may also deduce that the algebraic closure is unique up to isomorphism.

C.1 Preliminaries for algebraic closure

Lemma C.1. *Let K be a field and $f \in K[T]$ a monic polynomial. Then there exist a finite extension L/K and $\alpha_1, \ldots, \alpha_d \in L$, such that*

$$f = (T - \alpha_1) \cdots (T - \alpha_d)$$

in $L[T]$.

Proof. This can be achieved by induction on d, the degree of f. The case where f splits into linear factors – which includes the base case $d = 0$ – is trivial, so we may suppose that f has an irreducible factor of g of degree ≥ 2. We set $K' := K[X]/(g)$ (where for this we view g as a polynomial in the variable X). In K' we have $\overline{X}$ satisfying $f(\overline{X}) = 0$, hence in $K'[T]$ we may write

$$f = (T - \overline{X})h$$

for some monic $h \in K'[T]$ of degree $d - 1$. By the induction hypothesis, there exists a finite extension L/K' such that h splits into linear factors in $L[T]$. So f also splits into linear factors in $L[T]$. □

Lemma C.2. *Let K be a field, and let L/K be a field extension, such that every nonconstant polynomial in $K[T]$ admits a factorization in $L[T]$ into linear factors. Then the algebraic closure $\overline{K}$ of K in L is algebraically closed.*

Proof. Let $f \in \overline{K}[T]$ be an irreducible polynomial. Now $\overline{K}[T]/(f)$ is an algebraic extension of $\overline{K}$, with $\overline{T}$ algebraic over $\overline{K}$. Since $\overline{K}$ is algebraic over K, we have $\overline{T}$ algebraic over K by Proposition 3.5 (ii). Let $g \in K[T]$ be the minimal polynomial of $\overline{T}$ and d its degree. By the hypothesis, g admits a factorization

$$g = (T - \alpha_1) \cdots (T - \alpha_d)$$

with $\alpha_1, \ldots, \alpha_d \in L$, but each α_j is algebraic over K, hence lies in $\overline{K}$. We have $g(\overline{T}) = 0$, hence $\overline{T} = \alpha_j$ for some j, and in particular $\overline{T} \in \overline{K}$. This shows $\deg(f) = 1$. □

C.2 Existence of algebraic closure

Proposition C.3. *Every field admits an algebraic closure.*

Let K be a field. We prove the existence of an algebraic closure of K by showing that an algebraic closure can be obtained as a quotient of a polynomial ring over K by a maximal ideal.

Let S be the set of pairs (f, j) where f is a monic irreducible polynomial in $K[T]$ and $j \in \{1, \ldots, \deg(f)\}$. There is the ring $R := \mathrm{Sym}^\bullet(\bigoplus_{s \in S} K)$, a polynomial ring with one variable $X_{f,j}$ for every $(f, j) \in S$.

Now let I be the ideal in R generated by

$$e_j(X_{f,1}, \ldots, X_{f,d}) - (-1)^j a_j \tag{1}$$

for every monic irreducible polynomial

$$f = T^d + a_1 T^{d-1} + \cdots + a_d \tag{2}$$

in $K[T]$, with $d := \deg(f)$. Here, e_j denotes the jth elementary symmetric polynomial in the indicated variables.

We claim, I is a proper ideal of R. For this it is enough to show that every finite collection of elements (1) is contained in a maximal ideal. So we consider pairwise distinct monic irreducible polynomials $f_1, \ldots, f_m \in K[T]$ and exhibit a maximal ideal containing the elements (1) attached to these polynomials and all integers j.

Let $d_i := \deg(f_i)$ for every i. By Lemma C.1 there exists a finite extension L/K such that f_i splits in $L[T]$ into linear factors

$$f_i = (T - \alpha_{i,1}) \cdots (T - \alpha_{i,d_i}),$$

with $\alpha_{i,1}, \ldots, \alpha_{i,d_i} \in L$, for every i. There is a homomorphism

$$R \to L$$

given by $X_{f_i,j} \mapsto \alpha_{i,j}$ for all i and j, and $X_{f,j} \mapsto 0$ for all $f \notin \{f_1, \ldots, f_m\}$. The kernel is a maximal ideal and contains the generators as claimed.

Since I is a proper ideal of R, there exists a maximal ideal $\mathfrak{m}$ of R containing I. Now $R/\mathfrak{m}$ is an extension field of K such that, for every monic irreducible polynomial (2) we have

$$T^d + \overline{a}_1 T^{d-1} + \cdots + \overline{a}_d = (T - \overline{X}_{f,1}) \cdots (T - \overline{X}_{f,d})$$

in $(R/\mathfrak{m})[T]$. By Lemma C.2, the algebraic closure $\overline{K}$ of K in $R/\mathfrak{m}$ is algebraically closed.

C.3 Uniqueness of algebraic closure

Proposition C.4. *Let K be a field, $\overline{K}$ an algebraic closure of K, and L/K an algebraic field extension. Then there exists an embedding $L \to \overline{K}$ over K.*

Proof. Let S be the set of pairs (K', ι) where K' is a subfield of L containing K and ι is an embedding $K' \to \overline{K}$ over K. There is a partial order on S, with $(K', \iota) \preceq (\widetilde{K}', \tilde{\iota})$ if $K' \subset \widetilde{K}'$ and $\tilde{\iota}$ restricts to ι. Every chain (totally ordered subset) in S has an upper bound, namely the union of the subfields and the unique embedding that restricts to each of the embeddings in the chain. So by Zorn's lemma there exists a maximal element (K', ι) in S. If $K' = L$, we are done. But if $K' \neq L$ then we take $x \in L \smallsetminus K'$, so $K'(x)/K'$ is finite, and there exists an embedding of $K'(x) \to \overline{K}$ extending ι, contradicting the maximality of (K', ι). □

Corollary C.5. *Let K be a field, $\overline{K}$ an algebraic closure of K, and L/K an algebraic field extension. If L is algebraically closed, then there exists an isomorphism $L \to \overline{K}$ over K.*

Proof. We apply Proposition C.4 to obtain an embedding $\iota\colon L \to \overline{K}$ over K. Since L is algbraically closed, so is the image $\iota(L)$. Now $\overline{K}$ is an algebraic extension of $\iota(L)$ and hence is equal to $\iota(L)$. So $\iota\colon L \to \overline{K}$ is an isomorphism. □

D The Lindemann theorem

D.1 Historical context

The early study of transcendental numbers focused on approximation by rational numbers. Joseph Liouville (1809-1882) gave a sufficient condition for a number to be transcendental and, with it, the first concrete examples of transcendental numbers.

The modern era of transcendence theory begins with the 1873 proof by Charles Hermite (1822–1901) that e is transcendental. So-called auxiliary functions, coming from suitable approximations, are bounded from above by analytic methods – as in Proposition 3.26 – and, under hypothesis of algebraicity, bounded from below by number-theoretic methods. Incompatibility of the bounds implies transcendence.

We have stated the transcendence of e and π as Corollaries 3.22 and 3.23. By combining their proofs it may be seen that e^α is transcendental for any nonzero algebraic integer α. (It is a good exercise to write out this argument.) A power of an algebraic number is again algebraic, so we obtain the following.

Theorem (Lindemann, 1882)**.** *The number e^α is transcendental for any nonzero algebraic number α.*

Ferdinand von Lindemann (1852–1939) was thus able to deduce that π is transcendental, putting to rest the question, whether squaring the circle is possible with straightedge and compass, which had been open since antiquity. At the same time he announced the following generalization:

Theorem (Lindemann theorem)**.** *If algebraic numbers $\alpha_1, \ldots, \alpha_d$ are $\mathbb{Q}$-linearly independent, then $e^{\alpha_1}, \ldots, e^{\alpha_d}$ are algebraically independent over $\mathbb{Q}$.*

On account of the detailed written treatment given in 1885 by K. Weierstrass, the result is also known as the Lindemann-Weierstrass theorem.

In what follows we will give a proof of the following statement, which we can see to be equivalent to the Lindemann theorem.

Theorem. *If $\beta_1, \ldots, \beta_n$ are pairwise distinct algebraic numbers, then $e^{\beta_1}, \ldots, e^{\beta_n}$ are linearly independent over $\mathbb{Q}$.*

Under the assumption that this statement holds, we obtain the algebraic independence of $e^{\alpha_1}, \ldots, e^{\alpha_d}$ over $\mathbb{Q}$ by expanding any polynomial with $\mathbb{Q}$-coefficients in $e^{\alpha_1}, \ldots, e^{\alpha_d}$ and recognizing each monomial as the exponential of an integer linear combination of $\alpha_1, \ldots, \alpha_d$. To see the implication in the other direction, we take $\alpha_1, \ldots, \alpha_d$ to be a basis of the vector space over the rational numbers spanned by $\beta_1, \ldots, \beta_n$, such that each β_j is an integer linear combination of $\alpha_1, \ldots, \alpha_d$. We may arrange for the integer coefficients to be nonnegative by adding a large multiple of $\alpha_1 + \cdots + \alpha_d$ to $\beta_1, \ldots, \beta_n$. Then, $\mathbb{Q}$-linear combinations of $e^{\beta_1}, \ldots, e^{\beta_n}$ may be expressed as polynomials with $\mathbb{Q}$-coefficients in $e^{\alpha_1}, \ldots, e^{\alpha_d}$.

For a collection of complex numbers, algebraic independence over $\mathbb{Q}$ is equivalent to algebraic independence over $\overline{\mathbb{Q}}$ by Proposition 3.12, applied to $\overline{\mathbb{Q}}/\mathbb{Q}$ and extensions of the form $\overline{\mathbb{Q}}(\beta_1, \ldots, \beta_d)/\overline{\mathbb{Q}}$. We may change $\mathbb{Q}$ to $\overline{\mathbb{Q}}$ throughout the previous paragraph to obtain the following statement, which is again equivalent to the Lindemann theorem.

Theorem. *If $\beta_1, \ldots, \beta_n$ are pairwise distinct algebraic numbers, then $e^{\beta_1}, \ldots, e^{\beta_n}$ are linearly independent over $\overline{\mathbb{Q}}$.*

We will use some input from the theory of analytic functions (§D.2). Then we will obtain a generalization (§D.3) of Theorem 3.21, from which the Lindemann theorem will follow.

Further highlights in the historical development of transcendence theory include:

- (A. Gelfond and indepenently T. Schneider, 1934) If α and β are nonzero algebraic numbers with $\alpha \neq 1$ and $\beta \notin \mathbb{Q}$ then any value of α^β is transcendental.
- (A. Baker, 1966) If $\alpha_1, \ldots, \alpha_n$ are nonzero algebraic numbers, with $\log(\alpha_1)$, $\ldots$, $\log(\alpha_n)$ linearly independent over $\mathbb{Q}$, then 1, $\log(\alpha_1)$, $\ldots$, $\log(\alpha_n)$ are linearly indepenent over $\overline{\mathbb{Q}}$, and effectively so (see below).

The logarithm of a nonzero complex number is only defined up to the addition of an integer multiple of $2\pi\sqrt{-1}$; Baker's theorem applies to any choice of each logarithm in the statement.

Baker's theorem generalizes Lindemann's theorem and the Gelfond-Schneider theorem. Additionally, Baker obtains an effective form of linear independence: for $\beta_0, \ldots, \beta_n \in \overline{\mathbb{Q}}$, of respective degrees $d_0, \ldots, d_n$, and $d := \max(d_0, \ldots, d_n)$,

$$|\beta_0 + \beta_1 \log(\alpha_1) + \cdots + \beta_n \log(\alpha_n)| > \max(H(\beta_0), \ldots, H(\beta_n))^{-C(d, \log(\alpha_1), \ldots, \log(\alpha_n))}.$$

Attached to an algebraic number there is, besides degree, a numerical quantity $H(-)$ called height. The quantity $C(d, \log(\alpha_1), \ldots, \log(\alpha_n))$ is effectively computable.

A statement that generalizes both the Lindemann theorem and the Gelfond-Schneider theorem is the conjecture that if $\alpha_1, \ldots, \alpha_d \in \mathbb{C}$ are $\mathbb{Q}$-linearly independent then the transcendence degree of $\mathbb{Q}(\alpha_1, \ldots, \alpha_d, e^{\alpha_1}, \ldots, e^{\alpha_d})$ over $\mathbb{Q}$ is at least d, made by S. Schanuel in the early 1960s. Schanuel's conjecture remains an important open problem in transcendence theory.

D.2 Analytic preliminaries

Let U be an open subset of $\mathbb{R}$ and $f\colon U \to \mathbb{R}$ a C^∞ function (meaning that f is infinitely differentiable at all points of U). Given a point $x_0 \in U$ there are the Taylor coefficients $a_n := f^{(n)}(x_0)/n!$, and these may be assembled into the Taylor series

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n. \tag{1}$$

Starting with a sequence of real numbers $(a_n)_{n\in\mathbb{N}}$, we may go the other direction and consider the series (1). If for some $r \in \mathbb{R}_{>0}$ we have

$$\limsup_{n\to\infty} |a_n|^{1/n} \leq 1/r, \tag{2}$$

then the series (1) converges to a C^∞ function on the interval $(x_0 - r, x_0 + r)$ whose Taylor coefficients at x_0 are $(a_n)_{n\in\mathbb{N}}$. If we recover f on $(x_0 - r, x_0 + r)$ for some $r \in \mathbb{R}_{>0}$ in this manner, then we say that f is **analytic** at x_0. This implies that f is analytic at x for all $x \in (x_0 - r, x_0 + r)$, and in particular the set of points where f is analytic is open. We say that f is analytic if f is analytic at x_0 for all $x_0 \in U$.

As a variant, we may take f to be complex-valued and repeat the above discussion. A C^∞ function $f\colon U \to \mathbb{C}$ is analytic, respectively analytic at $x_0 \in U$, if and only if its real and imaginary parts are analytic, respectively analytic at $x_0 \in U$.

Now suppose that $U = \mathbb{R}$ and that f (which may be real- or complex-valued) is analytic. If, furthermore, the Taylor coefficients at some $x_0 \in \mathbb{R}$ satisfy

$$\lim_{n\to\infty} |a_n|^{1/n} = 0, \tag{3}$$

then we can take r arbitrarily large in (2), and it follows that (1) recovers f on all of $\mathbb{R}$. In fact, (1) provides a canonical extension of f to a complex-valued function

on $\mathbb{C}$. In this case the function f – or more properly its canonical extension to a complex-valued function on $\mathbb{C}$ – is said to be **entire**. Examples include polynomials, for which $a_n = 0$ for n sufficiently large, and the exponential function e^x, where for $x_0 = 0$ we have $a_n = 1/n!$, and for any $\varepsilon \in \mathbb{R}_{>0}$, if we take $n_0 \in \mathbb{N}$, $n_0 > 2/\varepsilon$, then

$$a_n \leq \frac{1}{n_0^{n-n_0} \cdot n_0!}$$

for all $n \geq n_0$, so $a_n^{1/n} < 2(\varepsilon/2) = \varepsilon$ for all $n \geq \max(n_0, \log(n_0^{n_0}/n_0!)/\log 2)$. Differentiability and analyticity properties of a function $\mathbb{C} \to \mathbb{C}$ requires their own dedicated study, which we do not even begin here. Rather, we just point out that the Taylor coefficients at 0 give an isomorphism between the $\mathbb{C}$-vector space of entire functions and the space of sequences of complex numbers $(a_n)_{n\in\mathbb{N}}$ satisfying (3).

Lemma D.1. *If $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$ are pairwise distinct, then $e^{\alpha_1 x}, \ldots, e^{\alpha_m x}$ are $\mathbb{C}$-linearly independent in the space of entire functions.*

Proof. By the isomorphism to sequences of Taylor coefficients, we are reduced to showing that

$$(\alpha_1^n/n!)_{n\in\mathbb{N}}, \ldots, (\alpha_m^n/n!)_{n\in\mathbb{N}}$$

are linearly independent sequences of complex numbers. In fact,

$$(\alpha_1^n/n!)_{n\in\{0,\ldots,m-1\}}, \ldots, (\alpha_m^n/n!)_{n\in\{0,\ldots,m-1\}}$$

are linearly independent since the Vandermonde determinant

$$\det(\alpha_i^{j-1})_{1\leq i,j\leq m} = \prod_{1\leq i<j\leq m} (\alpha_j - \alpha_i)$$

is nonzero. □

D.3 Proof of the Lindemann theorem

We start by giving a stronger form of Theorem 3.21.

Theorem D.2. *Theorem 3.21 holds for polynomials $u_1, \ldots, u_m$ which are pairwise distinct irreducible monic polynomials with rational coefficients.*

Light modification of the proof of Theorem 3.21 yields a proof of Theorem D.2.

The analytic part of the proof requires no modification. In the number-theoretic part of the proof we will put ourselves in the situation where M is a positive integer, such that $f_{\underline{r}}$ and hence also $E_{\underline{r}}(x)$ have integer coefficients divisible by M^N. Suppose that some u_j has coefficients in $\frac{1}{M}\mathbb{Z}$. Substituting $M^{-1}T$ for T in u_j yields after multiplying by a power of M, a monic polynomial with integer coefficients whose

roots are M times the roots of u_j. It follows that M^k times the sum of the kth powers of the roots of u_j is an integer. In particular, we will have

$$\sum_{u_j(\alpha)=0} E_{\underline{r}}(\alpha) \in \mathbb{Z},$$

and the same will hold for all the $E_{\underline{r}^{(n)}}$.

We define M to be the least common multiple of the denominators of the coefficients of $u := u_1 \cdots u_m$. So, Mu has integer coefficients, and by Gauss's lemma the same holds for each of the polynomials Mu_j. We define ℓ (the exponent of x in the definition of $f(x)$) and $N := \ell + \deg(u)$ exactly as in the proof of Theorem 3.21. Now in the definition of f we insert a factor of M^{N+1}:

$$f(x) := M^{N+1} x^\ell u(x).$$

This ensures that the coefficients of f are multiples of M^N. Thus, we have

$$\sum_{u_j(\alpha)=0} E_{\underline{r}^{(n)}}(\alpha) \in \mathbb{Z},$$

for all j.

The choice of n as a prime number p is made as in the proof of Theorem 3.21, subject to additional requirement $p \nmid M$. Proposition 3.29 remains valid, and the remainder of the proof of Theorem 3.21 goes through unchanged.

Having established Theorem D.2, we prove that if $\beta_1, \ldots, \beta_n$ are pairwise distinct algebraic numbers, then for any rational numbers $c_1, \ldots, c_n$, not all zero, the value of

$$c_1 e^{\beta_1 x} + \cdots + c_n e^{\beta_n x} \tag{1}$$

at $x = 1$ is nonzero. As we have seen in §D.1, this statement is equivalent to the Lindemann theorem.

There is clearly no loss of generality in assuming that the c_j are all integers. Furthermore, there no loss of generality in assuming that the polynomial

$$v := \prod_{j=1}^{n} (T - \beta_j)$$

has rational coefficients. Indeed, the union of the roots of minimal polynomials of $\beta_1, \ldots, \beta_n$ is a finite set $W \subset \mathbb{C}$ containing $\{\beta_1, \ldots, \beta_n\}$. If W contains additional elements then we may add corresponding terms with zero coefficient to (1).

The product

$$p_{c_1,\ldots,c_n} := \prod_{\sigma \in S_n} (c_1 X_{\sigma(1)} + \cdots + c_n X_{\sigma(n)}) \in \mathbb{Z}[X_1, \ldots, X_n],$$

may be expanded as

$$\sum_{F\colon S_n \to \{1,\dots,n\}} \prod_{\sigma \in S_n} c_{F(\sigma)} X_{\sigma(F(\sigma))},$$

where the sum is over all maps from S_n to $\{1, \dots, n\}$. Thus we see that $p_{c_1,\dots,c_n}$ is invariant under permutation of $X_1, \dots, X_n$. Since $p_{c_1,\dots,c_n}$ is homogeneous of degree $n!$ it follows that $p_{c_1,\dots,c_n}$ may be written terms of the basis $(m_\lambda)_{\lambda \in I_{n,n!}}$ introduced in the proof of Proposition 3.1:

$$p_{c_1,\dots,c_n} = \sum_{\lambda \in I_{n,n!}} h_{\lambda;c_1,\dots,c_n} m_\lambda,$$

with $h_{\lambda;c_1,\dots,c_n} \in \mathbb{Z}$.

For $\lambda \in I_{n,n!}$ we have polynomials $\Sigma_\lambda(v)$ from Proposition 3.2. Let $u_1, \dots, u_m$ be the irreducible monic factors of all the polynomials $\Sigma_\lambda(v)$, and let

$$\Sigma_\lambda(v) = u_1^{g_{\lambda 1}} \cdots u_m^{g_{\lambda m}}$$

be the decomposition into irreducible factors.

Lemma D.3. *We have*

$$\prod_{\sigma \in S_n} (c_1 e^{\beta_{\sigma(1)} x} + \cdots + c_n e^{\beta_{\sigma(n)} x}) = \sum_{\lambda \in I_{n,n!}} h_{\lambda;c_1,\dots,c_n} \sum_{\ell=1}^{m} g_{\lambda \ell} \sum_{\substack{\alpha \in \mathbb{C} \\ u_\ell(\alpha)=0}} e^{\alpha x}.$$

Proof. The left-hand side is obtained by evaluating $p_{c_1,\dots,c_n}$ at $(e^{\beta_1 x}, \dots, e^{\beta_n x})$ and hence is equal to

$$\sum_{\lambda \in I_{n,n!}} h_{\lambda;c_1,\dots,c_n} \sum_{\substack{q_1,\dots,q_n \in \mathbb{Z} \\ \exists \sigma \in S_n \colon \forall j \, q_{\sigma(j)} = \lambda_j}} e^{(q_1\beta_1 + \cdots + q_n\beta_n)x}.$$

So the result follows from Proposition 3.2. □

In the equality in Lemma D.3, the left-hand side is a product of nontrivial sums of exponentials. Hence the coefficient of $e^{\alpha x}$ is nonzero for at least one α appearing in the sum on the right-hand side. Theorem D.2 gives the nonvanishing of the value of the right-hand side at $x = 1$ and hence as well of the value at $x = 1$ of the expression in brackets on the left-hand side, for every $\sigma \in S_n$. One of these expressions, corresponding to the identity permutation, is (1). This concludes the proof of the Lindemann theorem.