Lecture Note on Algebra

Anthony Hong[1]

March 1, 2024

[1]Thanks to Professor Beheshti Zavareh for her teaching Math5031-32 Algebra I & II; thanks to Albert Peng for his permission to edit his partial note of Math5031 and also J.S. Milne's tex file

Contents

[1] Taken from [3] sec 7.2
[2] Taken from [3] sec 7.2

[3]Taken from David

Chapter 1

Groups

1.1 Recap: Groups, Cosets, and Homomorphisms

Definition 1.1.1 (Group)**.** We define a binary operation (multiplication) $* : G \times G \to G$ on a nonempty set G, and $(G, \cdot)$ is called a **group** if $*$ satisfies the following rules.

(1) the multiplication is closed on G;

(2) associativity of multiplication: $a * (b * c) = (a * b) * c, \forall a, b, c \in G$;

(3) G has an **identity element** (i.e. $\exists e \in G$ s.t. $\forall g \in G : e * g = g * e = g$);

(4) each element $g \in G$ has an **inverse** (i.e. $\exists g^{-1} \in G$ s.t. $g * g^{-1} = g^{-1} * g = e$).

Remark 1.1.2. Several remarks are in order:

1. We will denote $ab = a * b$ and $a^m * a^n = a^{n+m} = a^n * a^m$ and $(a^m)^n = a^{mn} = (a^n)^m$.
2. A **magma** is a tuple $(G, *)$ with (1) above; a **semigroup** is an associative magma, i.e. tuple $(G, *)$ with (1) and (2) above; a **monoid** is a semigroup with an identity element, i.e., tuple $(G, *)$ with (1), (2), and (3) above.
3. Let $(R, +, *)$ be a ring with unity 1. That is, $(R, *)$ is a monoid. An element x is called a **unit** or **invertible element** if it has an inverse, so the set of all invertible elements $U(R)$ is a group, called **group of units** in R.
4. Rules (3) and (4) in definition 1.1.1 are equivalent to the following condition (proof of the equivalence outlined in the exercise 1):

(5) $\forall a, b \in G :$ equations $ax = b, ya = b$ have solutions in G.

Definition 1.1.3 (Abelian Group)**.** A group G is called **Abelian** if $\forall a, b \in G : ab = ba$.

Definition 1.1.4 (Subgroup)**.** A non-empty subset $H \subseteq G$ is a **subgroup**, denoted as $H \leqslant G$, if

(1) $a \in H \implies a^{-1} \in H$

(2) $a, b \in H \implies ab \in H$

Proposition 1.1.5.

1. $H \leqslant G$ implies that H is a group with operation of G (see [9] Theorem 2.1);
2. $H \subseteq G$ is a subgroup iff $e \in H$ and $a, b \in H \implies ab^{-1} \in H$ (see [9] Theorem 2.2).

3. G finite, then a nonempty subset H of G is a subgroup iff $a, b \in H \Rightarrow ab \in H$ (see [9] Corollary 2.4).

Theorem 1.1.6. The inverse and the identity element of a group are both unique.

Proof. Suppose $e, e' \in G$ and $\forall g \in G$ we have

$$e \cdot g = g \cdot e = g \tag{1.1}$$
$$e' \cdot g = g \cdot e' = g \tag{1.2}$$

Putting $g = e$ in (1.2) results in $e = e \cdot e'$ and putting $g = e'$ in (1.1) results in $e \cdot e' = e'$. So $e = e'$. Suppose h and k are inverses of g, so that in particular $hg = e$ and $gk = e$. Then $(hg)k = ek = k$, but $h(gk) = he = h$. But the associativity law tells us $(hg)k = h(gk)$, which says $k = h$. □

Example 1.1.7. The trivial group $G = \{e\}$ with $*$ defined by $e * e = e$. $(\mathbb{C}, \times)$ is not a group. What would the inverse element of 0 be? But if we write $\mathbb{C}^\times$ for the set of nonzero complex numbers then $(\mathbb{C}^\times, \times)$ is a group. Equally the nonzero real numbers or rational numbers under multiplication are groups. Let $\mathrm{GL}(n, \mathbb{C})$ be the set of $n \times n$ invertible matrices over the complex numbers. Then $\mathrm{GL}_n(\mathbb{C})$ with matrix multiplication is a nonabelian group.

Definition 1.1.8 (group homomorphism)**.** Let G, G' be a group. $\phi : G \to G'$ is a **homomorphism** if $\phi(ab) = \phi(a)\phi(b)$ for all $a, b \in G$. f is an **isomorphism** if the homomorphism is bijective, denoted by $G \cong H$. An injective homomorphism is called a **monomorphism**. A surjective homomorphism is called a **epimorophism**. If $G = G'$, we say the homomorphism is an **endomorphism**. If furthermore that endomorphism is also bijective, we say it is an automorphism.

Remark 1.1.9 (Isomorphism is an equiv relation)**.** If $\phi : G \to G'$ is a group isomorphism, i.e., a bijective homomorphism, then its inverse is also an isomorphism. Therefore, if we find the inverse function of a group homomorphism as a function, then that inverse function automatically becomes an isomorphism. This means isomorphism is a symmetric relation on the set of all groups. Isomorphism is also reflexive and transitive, so it's an equivalence relation. The proof of these two are left as exercises. We show the symmetric property: Since ϕ is bijective, there is an inverse function $\phi^{-1} : G' \to G$. Suppose $a, b \in G'$, and we want to show $\phi^{-1}(ab) = \phi^{-1}(a)\phi^{-1}(b)$. Let $x = \phi^{-1}(a)$ and $y = \phi^{-1}(b)$. Since ϕ is a homomorphism, we have $\phi(xy) = \phi(x)\phi(y) = ab$, so $\phi^{-1}(ab) = xy$.

Theorem 1.1.10. Let $f : (G, *) \to (G', \circ)$ be a homomorphism.

1. $f(e) = e'$, where e' is the identity in G';
2. If $a \in G$, then $f\left(a^{-1}\right) = f(a)^{-1}$;
3. If $a \in G$ and $n \in \mathbb{Z}$, then $f\left(a^n\right) = f(a)^n$;
4. $H \leqslant G \Rightarrow f(H) \leqslant G'$ and $H' \leqslant G' \Rightarrow f^{-1}(H') \leqslant G$;

Proof.

1. Applying f to the equation $e = e * e$ gives $f(e) = f(e * e) = f(e) \circ f(e)$. Now multiply each side of the equation by $f(e)^{-1}$ to obtain $e' = f(e)$.
2. Applying f to the equations $a * a^{-1} = e = a^{-1} * a$ gives $f(a) * f\left(a^{-1}\right) = e' = f\left(a^{-1}\right) * f(a)$. It follows from Theorem 1.10, the uniqueness of the inverse, that $f\left(a^{-1}\right) = f(a)^{-1}$.
3. Induction shows $f\left(a^n\right) = f(a)^n$ for all $n \geqslant 0$, and then $f\left(a^{-n}\right) = f\left(\left(a^{-1}\right)^n\right) = f\left(a^{-1}\right)^n = f(a)^{-n}$.
4. $e' \in f(H)$ by 1. Let $x', y' \in f(H)$, then $\exists x, y \in H$ s.t. $f(x) = x', f(y) = y'$. Thus $xy^{-1} \in H \Rightarrow x'y'^{-1} = f(xy^{-1}) \in f(H)$. Now, $e \in f^{-1}(H')$ by 1. Let $x, y \in f^{-1}(H')$. Then $f(xy^{-1}) = f(x)f(y)^{-1} \in H' \Rightarrow xy^{-1} \in f^{-1}(H')$.

□

Example 1.1.11 (Klein-four group)**.** For small groups $(G, *)$ we can completely describe the group operation by drawing a table called a **group table** or **Cayley table**. It is a $n \times n$ matrix whose i, j entry is the group element $g_i g_j$, where $n = |G|$. For example, one can show that $\mathbf{V} = \{1, -1, i, -i\} \subseteq \mathbb{C}$ with multiplication of complex numbers $\cdot$ is a group, where the group table is given below. This is an abelian group. One

$*$	1	-1	i	$-i$
1	1	-1	i	$-i$
-1	-1	1	$-i$	i
i	i	$-i$	-1	1
$-i$	$-i$	i	1	-1

can also show that it is isomorphic to $\{1, (12)(34), (13)(24), (14)(23)\}$ with composition of permutation as multiplication (i.e., as a subgroup of S_4) and also to $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \cong D_2 = \langle a, b | a^2 = b^2 = (ab)^2 = e \rangle$.

Example 1.1.12 (Quaternion group)**.** The quaternion group, Q_8, is defined by

$$Q_8 = \{1, -1, i, -i, j, -j, k, -k\}$$

with product $\cdot$ computed as follows:

$$\begin{gathered} 1 \cdot a = a \cdot 1 = a, \quad \text{for all } a \in Q_8 \\ (-1) \cdot (-1) = 1, \quad (-1) \cdot a = a \cdot (-1) = -a, \quad \text{for all } a \in Q_8 \\ i \cdot i = j \cdot j = k \cdot k = -1 \\ i \cdot j = k, \quad j \cdot i = -k \\ j \cdot k = i, \quad k \cdot j = -i \\ k \cdot i = j, \quad i \cdot k = -j. \end{gathered}$$

It is tedious to check the associative law (it can be proven by a less computational mean), but the other axioms are easily checked. Note that Q_8 is a non-abelian group of order 8.

Example 1.1.13. Consider the set of nonzero real numbers, $\mathbb{R}^*$, with the group operation of multiplication. The identity of this group is 1 and the inverse of any element $a \in \mathbb{R}^*$ is just $1/a$. We will show that

$$\mathbb{Q}^* = \{p/q : p \text{ and } q \text{ are nonzero integers } \}$$

is a subgroup of $\mathbb{R}^*$. The identity of $\mathbb{R}^*$ is 1 ; however, $1 = 1/1$ is the quotient of two nonzero integers. Hence, the identity of $\mathbb{R}^*$ is in $\mathbb{Q}^*$. Given two elements in $\mathbb{Q}^*$, say p/q and r/s, their product pr/qs is also in $\mathbb{Q}^*$. The inverse of any element $p/q \in \mathbb{Q}^*$ is again in $\mathbb{Q}^*$ since $(p/q)^{-1} = q/p$. Since multiplication in $\mathbb{R}^*$ is associative, multiplication in $\mathbb{Q}^*$ is associative.

Example 1.1.14. Let $SL_2(\mathbb{R})$ be the subset of $GL_2(\mathbb{R})$ consisting of matrices of determinant one; that is, a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in $SL_2(\mathbb{R})$ exactly when $ad - bc = 1$. To show that $SL_2(\mathbb{R})$ is a subgroup of the general linear group, we must show that it is a group under matrix multiplication. The 2×2 identity matrix is in $SL_2(\mathbb{R})$, as is the inverse of the matrix A :

$$A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

It remains to show that multiplication is closed; that is, that the product of two matrices of determinant one also has determinant one. We will leave this task as an exercise. The group $SL_2(\mathbb{R})$ is called the **special linear group**.

Example 1.1.15. It is important to realize that a subset H of a group G can be a group without being a subgroup of G. For H to be a subgroup of G, it must inherit the binary operation of G. The set of all 2×2 matrices, $M_2(\mathbb{R})$, forms a group under the operation of addition. The 2×2 general linear group is a subset of $M_2(\mathbb{R})$ and is a group under matrix multiplication, but it is not a subgroup of $M_2(\mathbb{R})$. If we add two invertible matrices, we do not necessarily obtain another invertible matrix. Observe that

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

but the zero matrix is not in $GL_2(\mathbb{R})$.

Two subtleties regarding the binary operation need to be addressed:

Theorem 1.1.16 (associative invariance of bracketing)**.** For each way of bracketing the multiplication of n elements $a_1, \cdots, a_n \in A$, we denote it as

$$\pi_i \left(a_1 \cdot a_2 \cdot \cdots \cdot a_n\right), i = 1, 2, \cdots, N$$

where it can be proved that $N = (2n-2)!/[n!(n-1)!]$. For example, let $n = 3$ and we will have $N = 2$ ways to bracket the three elements: $\pi_1 \left(a_1 \cdot a_2 \cdot a_3\right) = \left(a_1 \cdot a_2\right) \cdot a_3$ and $\pi_2 \left(a_1 \cdot a_2 \cdot a_3\right) = a_1 \cdot \left(a_2 \cdot a_3\right)$. We now claim that these N ways of bracketing are the same if associativity of order 3 holds for the set A (i.e. $\pi_1 \left(a_1 \cdot a_2 \cdot a_3\right) = \pi_2 \left(a_1 \cdot a_2 \cdot a_3\right)$, or $\left(a_1 \cdot a_2\right) \cdot a_3 = a_1 \cdot \left(a_2 \cdot a_3\right)$), and then the notation $a_1 \cdot a_2 \cdot \cdots \cdot a_n$ is well-defined.

Proof. See exercise 1.1-7. □

Theorem 1.1.17 (commutative invariance of permutation)**.** If both associativity and commutativity hold for a binary operation $\cdot$, then permutating the following multiplication in any order results the same

$$a_1 \cdot a_2 \cdot \cdots \cdot a_N$$

Proof. See exercise 1.1-8. □

Definition 1.1.18. If G is a group and $a \in G$, then the **cyclic subgroup generated by** a, denoted by $\langle a \rangle$, is the set of all the powers of a. A group G is called **cyclic** if there is $a \in G$ with $G = \langle a \rangle$; that is, G consists of all the powers of a.

It is plain that $\langle a \rangle$ is, indeed, a subgroup of G. Notice that different elements can generate the same cyclic subgroup. For example, $\langle a \rangle = \left\langle a^{-1} \right\rangle$.

Example 1.1.19. Let $C_n = \left\{e^{2\pi i k/n} : k \in \mathbb{Z}\right\}$, a subset of the complex numbers. This is a group under multiplication: certainly multiplication is a binary operation on this set, for

$$e^{2\pi i k/n} e^{2\pi i l/n} = e^{2\pi i (k+l)/n}$$

which is an element of C_n. You can check the other group axioms. C_n is a cyclic group, because every element is a power of $\zeta = e^{2\pi i/n}$, and ζ has order n so $|C_n| = n$. Any generator of C_n is called a **primitive n-th root of unity**.

Definition 1.1.20. If G is a group and $a \in G$, then the **order of** a is $|\langle a \rangle|$, the number of elements in $\langle a \rangle$.

Theorem 1.1.21. If G is a group and $a \in G$ has finite order m, then m is the smallest positive integer such that $a^m = 1$.

Proof. If $a = 1$, then $m = 1$. If $a \neq 1$, there is an integer $k > 1$ so that $1, a, a^2, \ldots, a^{k-1}$ are distinct elements of G while $a^k = a^i$ for some i with $0 \leqslant i \leqslant k-1$. We claim that $a^k = 1 = a^0$. If $a^k = a^i$ for some $i \geqslant 1$, then $k - i \leqslant k - 1$ and $a^{k-i} = 1$, contradicting the original list $1, a, a^2, \ldots, a^{k-1}$ having no repetitions. It follows that k is the smallest positive integer with $a^k = 1$.

It now suffices to prove that $k = m$; that is, that $\langle a \rangle = \{1, a, a^2, \ldots, a^{k-1}\}$. Clearly $\langle a \rangle \supset \{1, a, a^2, \ldots, a^{k-1}\}$. For the reverse inclusion, let a^l be a power of a. By the division algorithm, $l = qk + r$, where $0 \leqslant r < k$. Hence, $a^l = a^{qk+r} = a^{qk}a^r = a^r$ (because $a^k = 1$), and so $a^l = a^r \in \{1, a, a^2, \ldots, a^{k-1}\}$. □

Theorem 1.1.22. Every subgroup of a cyclic group is cyclic.

Proof. The main tools used in this proof are the division algorithm and the Principle of Well-Ordering. Let G be a cyclic group generated by a and suppose that H is a subgroup of G. If $H = \{e\}$, then trivially H is cyclic. Suppose that H contains some other element g distinct from the identity. Then g can be written as a^n for some integer n. Since H is a subgroup, $g^{-1} = a^{-n}$ must also be in H. Since either n or $-n$ is positive, we can assume that H contains positive powers of a and $n > 0$. Let m be the smallest natural number such that $a^m \in H$. Such an m exists by the Principle of Well-Ordering. We claim that $h = a^m$ is a generator for H. We must show that every $h' \in H$ can be written as a power of h. Since $h' \in H$ and H is a subgroup of $G, h' = a^k$ for some integer k. Using the division algorithm, we can find numbers q and r such that $k = mq + r$ where $0 \leqslant r < m$; hence,

$$a^k = a^{mq+r} = (a^m)^q a^r = h^q a^r.$$

So $a^r = a^k h^{-q}$. Since a^k and h^{-q} are in H, a^r must also be in H. However, m was the smallest positive number such that a^m was in H; consequently, $r = 0$ and so $k = mq$. Therefore,

$$h' = a^k = a^{mq} = h^q$$

and H is generated by h. □

Corollary 1.1.23. The subgroups of $\mathbb{Z}$ are exactly $n\mathbb{Z}$ for $n = 0, 1, 2, \ldots$...

Proof. First, $n\mathbb{Z} = \{\cdots, -2n, -n, 0, n, 2n, \cdots\} = \langle n \rangle$. Then let $H \leqslant \mathbb{Z}$. Since $\mathbb{Z}$ is cyclic, $H = \langle n \rangle$ for some $n \in \mathbb{Z}$ by above theorem. □

Proposition 1.1.24. Let G be a cyclic group of order n and suppose that a is a generator for G. Then $a^k = e$ if and only if n divides k.

Proof. First suppose that $a^k = e$. By the division algorithm, $k = nq + r$ where $0 \leqslant r < n$; hence,

$$e = a^k = a^{nq+r} = a^{nq}a^r = ea^r = a^r.$$

Since the smallest positive integer m such that $a^m = e$ is n, we have $r = 0$. Conversely, if n divides k, then $k = ns$ for some integer s. Consequently,

$$a^k = a^{ns} = (a^n)^s = e^s = e.$$

□

Proposition 1.1.25. An infinite cyclic group $\langle a \rangle \cong \mathbb{Z}$ has exactly two generators $a, -a$. Let G be a cyclic group of order n and suppose that $a \in G$ is a generator of the group. If $b = a^k$, then the order of b is n/d, where $d = \gcd(k, n)$.

Proof. The first statement is trivial. We show the second: we wish to find the smallest integer m such that $e = b^m = a^{km}$. By above proposition, this is the smallest integer m such that n divides km or, equivalently, n/d divides $m(k/d)$. Since d is the greatest common divisor of n and $k, n/d$ and k/d are relatively prime. Hence, for n/d to divide $m(k/d)$ it must divide m. The smallest such m is n/d. □

Theorem 1.1.26. The intersection of any family of subgroups of a group G is again a subgroup of G.

Proof. Let $\{S_i : i \in I\}$ be a family of subgroups of G. Now $1 \in S_i$ for every i, and so $1 \in \bigcap S_i$. If $a, b \in \bigcap S_i$, then $a, b \in S_i$ for every i, and so $ab^{-1} \in S_i$ for every i; hence, $ab^{-1} \in \bigcap S_i$, and $\bigcap S_i \leqslant G$. □

Corollary 1.1.27. If X is a subset of a group G, then **subgroup generated by** X, defined as

$$\langle X \rangle := \bigcap_{X \subseteq H \leqslant G} H$$

is the smallest subgroup H of G containing X, that is, if $X \subset S$ and $S \leqslant G$, then $H \leqslant S$.

Proof. There are subgroups of G containing X; for example, G itself contains X; define H as the intersection of all the subgroups of G which contain X. Note that H is a subgroup, by Theorem 1.1.26, and $X \subset H$. If $S \leqslant G$ and $X \subset S$, then S is one of the subgroups of G being intersected to form H; hence, $H \leqslant S$, and so H is the smallest such subgroup. □

Definition 1.1.28. If X is a nonempty subset of a group G, then a word on X is an element $w \in G$ of the form

$$w = x_1^{e_1} x_2^{e_2} \dots x_n^{e_n},$$

where $x_i \in X, e_i = \pm 1$, and $n \geqslant 1$.

Theorem 1.1.29. Let X be a subset of a group G. If $X = \varnothing$, then $\langle X \rangle = 1$; if X is nonempty, then $\langle X \rangle$ is the set of all the words on X:

$$\langle X \rangle = \{w = x_1^{e_1} x_2^{e_2} \dots x_n^{e_n} | x_i \in X, e_i = \pm 1, n \geqslant 1\}$$

Proof. If $X = \varnothing$, then the subgroup $1 = \{1\}$ contains X, and so $\langle X \rangle = 1$. If X is nonempty, let W denote the set of all the words on X. It is easy to see that W is a subgroup of G containing $X : 1 = x_1^{-1} x_1 \in W$; the inverse of a word is a word; the product of two words is a word. Since $\langle X \rangle$ is the smallest subgroup containing X, we have $\langle X \rangle \subset W$. The reverse inclusion also holds, for every subgroup H containing X must contain every word on X. Therefore, $W \leqslant H$, and W is the smallest subgroup containing X. □

Proposition 1.1.30. Let $\varphi : G \to G$ be a homomorphism. Then $\varphi(\langle X \rangle) = \langle \varphi(X) \rangle$.

Proof. Routine. □

Definition 1.1.31. Let $H \leqslant G, g \in G$. The **right coset** of H in G represented by g is $Hg = \{hg \,|\, h \in H\}$. Similarly, **left coset** is defined as $gH = \{gh \,|\, h \in H\}$.

Example 1.1.32 ([9] Example 2.3)**.** Let G be the additive group of the plane $\mathbb{R}^2$: the elements of G are vectors (x, y), and addition is given by the "parallelogram law": $(x, y) + (x', y') = (x + x', y + y')$. A line ℓ through the origin is the set of all scalar multiples of some nonzero vector $v = (x_0, y_0)$; that is, $\ell = \{rv : r \in \mathbb{R}\}$. It is easy to see that ℓ is a subgroup of G. If $u = (a, b)$ is a vector, then the coset $u + \ell$ is easily seen to be the line parallel to ℓ which contains u.

Example 1.1.33 ([9] Example 2.4)**.** If G is the additive group $\mathbb{Z}$ of all integers, if S is the set of all multiples of an integer $n (S = \langle n \rangle$, the cyclic subgroup generated by $n)$, and if $a \in \mathbb{Z}$, then the coset $a + S = \{a + qn : q \in \mathbb{Z}\} = \{k \in \mathbb{Z} : k \equiv a \bmod n\}$; that is, the coset $a + \langle n \rangle$ is precisely the congruence class $[a]$ of $a \bmod n$.

Proposition 1.1.34. Two observations:

- $Ha = Hb \iff H = Hba^{-1} \iff ba^{-1} \in H$;
- $aH = bH \iff a^{-1}bH = H \iff a^{-1}b \in H$.

Corollary 1.1.35. For two cosets, either $Hg_1 = Hg_2$ or $Hg_1 \cap Hg_2 = \varnothing$ (similar for left cosets).

Proof. Let $a = Hg_1 \cap Hg_2$. Then $a = h_1g_2 = h_2g_2$ and $h_2^{-1}h_1 = g_2g_1^{-1} \implies g_2g_1^{-1} \in H \implies Hg_1 = Hg_2$. □

Example 1.1.36. A right coset is not necessarily a left coset. See [9] Example 2.5.

Proposition 1.1.37. There is a bijection between the set of distinct left cosets of H and distinct right cosets of H: $aH \mapsto Ha^{-1}$.

Proof. $aH = bH \iff a^{-1}b \in H \iff (a^{-1}b)^{-1} \in H \iff b^{-1}a \in H \iff Ha^{-1} = Hb^{-1}$ □

Definition 1.1.38. The **index** of subgroup H in G, $[G : H]$, is the number of distinct right (left) cosets of H in G.

Theorem 1.1.39 (Lagrange's theorem)**.** If G is a finite group and $S \leqslant G$, then $|S|$ divides $|G|$ and $[G : S] = |G|/|S|$, or $|G| = [G : S]|S|$.

Proof. By Corollary 1.1.35, G is partitioned into its right cosets

$$G = St_1 \cup St_2 \cup \cdots \cup St_n,$$

and so $|G| = \sum_{i=1}^{n} |St_i|$. But it is easy to see that $f_i : S \to St_i$, defined by $f_i(s) = st_i$, is a bijection, and so $|St_i| = |S|$ for all i. Thus $|G| = n|S|$, where $n = [G : S]$. □

Corollary 1.1.40. The order of an element of a finite group divides the order of the group.

Proof. The order of an element a of a group G is equal to the order of the cyclic subgroup $\langle a \rangle$ generated by a. Then apply Lagrange's theorem. □

Corollary 1.1.41. If p is a prime and $|G| = p$, then G is a cyclic group.

Proof. Take $a \in G$ with $a \neq 1$. Then the cyclic subgroup $\langle a \rangle$ has more than one element (it contains a and 1), and its order $|\langle a \rangle| > 1$ is a divisor of p. Since p is prime, $|\langle a \rangle| = p = |G|$, and so $\langle a \rangle = G$. □

1.1 EXERCISES

1. By steps i.-iv., prove the equivalence between 1.1.1(1)-(4) and 1.1.1(1),(2)+1.1.2(5):

i. Suppose (1), (2), and (5) are true, show that there exists a left identity element e_l such that $e_la = a$ for any $a \in G$ and show that there exists a left inverse g_l^{-1} for any $g \in G$ such that $g_l^{-1}g = e_l$.
ii. If there is a left inverse element, then there is a right inverse element, and they are the same.
iii. If there is a left identity element, then there is a right identity element, and they are the same.
iv. Show that (1)-(4) imply (5).

2. [3][1.1 ex9] Let $G = \{a + b\sqrt{2} \in \mathbb{R} \mid a, b \in \mathbb{Q}\}$.

i. Prove that G is a group under addition.
ii. Prove that the nonzero elements of G are a group under multiplication. (Hint: "Rationalize the denominators" to find multiplicative inverses.)

3. Prove that a finite group is abelian if and only if its group table is a symmetric matrix.

4. (Cancellation property): suppose $\cdot$ is an internal binary operation for the set A. We say that the operation - is left-cancellative if $\forall a, b \in A : a \cdot b = a \cdot c \Rightarrow b = c$ and rightcancellative if $\forall a, b \in A : b \cdot a = c \cdot a \Rightarrow b = c$. When the operation is both left and right cancellative we simply say it is cancellative. Show that:

i. The cross product of vectors does not obey cancellation law.
ii. Determine when does matrix multiplication obey the cancellation law.
iii. Given a finite set G with an operation ; prove that if $\cdot$ is right and left cancellative and associative and G is closed under, then G is a group.
iv. Observe that an operation $\cdot$ of a group $(G, \cdot)$ obeys left (right) cancellation law iff each row (column) of its group table has elements of itself distinct.

5. Show that for x in a group G, (1) $|x| = 1 \Leftrightarrow x = e$; (2) $x^{-1} = x \Leftrightarrow x^2 = e$.

6. Show that for x in a group G, (1) $|x| = \left|x^{-1}\right|$; (2) $|x| = n \Rightarrow \left|x^k\right| = \frac{n}{(k,n)}$.

7. Prove Theorem 1.1.16.

8. Prove Theorem 1.1.17.

9. [9][p.27 ex2.11] Let $a \in G$ have order $n = mk$, where $m, k \geqslant 1$. Prove that a^k has order m.

10. [9][p.27 ex2.12] Show that

i. every group G of order 4 is isomorphic to either $\mathbb{Z}_4$ or the Klein-four group $\mathbf{V}$ (see example 1.1.11).
ii. If G is a group with $|G| \leqslant 5$, then G is abelian.

11. [9][p.27 ex2.13] If $a \in G$ has order n and k is an integer with $a^k = 1$, then n divides k. Indeed, $\left\{k \in \mathbb{Z} : a^k = 1\right\}$ consists of all the multiplies of n.

12. [9][p.27 ex2.14] If $a \in G$ has finite order and $f : G \to H$ is a homomorphism, then the order of $f(a)$ divides the order of a.

13. [9][p.27 ex2.15] Prove that a group G of even order has an odd number of elements of order 2 (in particular, it has at least one such element). (Hint. If $a \in G$ does not have order 2 , then $a \neq a^{-1}$.)

14. [9][p.27 ex2.17]

i. If $a, b \in G$ commute and if $a^m = 1 = b^n$, then $(ab)^k = 1$, where $k = \operatorname{lcm}\{m, n\}$. (The order of ab may be smaller than k; for example, take $b = a^{-1}$.) Conclude that if a and b have finite order, then ab also has finite order.
ii. Let $G = \mathrm{GL}(2, \mathbb{Q})$ and let $A, B \in G$ be given by

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}.$$

Show that $A^4 = E = B^3$, but that AB has infinite order.

15. [9][p.27 ex2.19] Prove that two cyclic groups are isomorphic if and only if they have the same order.

16. If $K \leqslant H \leqslant G$ with G not necessarily finite, and if $[G : H], [H : K] < \infty$, then $[G : K] < \infty$ and $[G : K] = [H : K][G : H]$.

17. [9][p.27 ex2.16] If $H \leqslant G$ has index 2,then $a^2 \in H$ for every $a \in G$.

18. Suppose $f : G \to G'$ is a homomorphism, show that $f(\langle X \rangle) = \langle f(X) \rangle$ for any subset $X \subseteq G$.

1.2 More Groups

We will present the following groups in this section: $\mathbb{Z}$, $\mathbb{Z}_n$, and $\mathbb{Z}_n^\times$; cyclic groups; symmetric group S_n and alternating group A_n; dihedral group D_n.

1.2.1 $\mathbb{Z}$, $\mathbb{Z}_n$, and $\mathbb{Z}_n^\times$

Definition 1.2.1 (Congruence)**.** Let $n, a, b \in \mathbb{Z}$. We say a is congruent to b modulo (or just mod) n if $a - b$ is divisible by n. In this case we write

$$a \equiv b \ (\mathrm{mod} n)$$

Observe that $a \sim b \Leftrightarrow a \equiv b (\mathrm{mod} n)$ is an equivalence relation. The equivalence class is denoted as $[a]_n$, $[a]$, or $\bar{a}$, called the **congruence class**. We denote the collection of all equivalence classes $[a]_n$ under $\sim$ as $\mathbb{Z}_n$.

Theorem 1.2.2. Define a binary operation $+$ on $\mathbb{Z}_n$ by $[a]_n + [b]_n = [a + b]_n$. Then $(\mathbb{Z}_n, +)$ is a group.

Proof. First we need to check that this really does define a binary operation on $\mathbb{Z}_n$. The potential problem is that an eqivalence class $[a]_n$ can have lots of different representatives, e.g. $[5]_3 = [2]_3$, but our definition of $+$ seems to depend on a specific choice of representative. Couldn't it be that $[a]_n = [a']$ and $[b]_n = [b']_n$ but $[a + b]_n \neq [a' + b']_n$? If so our definition of $+$ wouldn't work - it would not be "welldefined." We need to check that if $[a]_n = [a']_n$ and $[b]_n = [b']_n$ then $[a + b]_n = [a' + b']_n$. Because $[a]_n = [a']_n$, a and a' are congruent $\mathrm{mod} n$ so $a = a' + kn$ for some integer k, and similarly $b = b' + ln$ for some integer I. Therefore

$$\begin{aligned} a + b &= a' + kn + b' + ln \\ &= a' + b' + (k + l)n \end{aligned}$$

so $a + b \equiv a' + b' \bmod n$ and $[a + b]_n = [a' + b']_n$. The group axioms are easy to check. $[0]_n$ is clearly an identity element, $[-a]_n$ is inverse to $[a]_n$, and because $+$ is associative on $\mathbb{Z}$ we have $[a]_n + ([b]_n + [c]_n) = [a]_n + [b + c]_n = [a + b + c]_n$ and $([a]_n + [b]_n) + [c]_n = [a + b]_n + [c]_n = [a + b + c]_n$ so

$$[a]_n + ([b]_n + [c]_n) = ([a]_n + [b]_n) + [c]_n$$

and $+$ is associative on $\mathbb{Z}_n$. □

Theorem 1.2.3. $\mathbb{Z}_n$ is a cyclic group and the generators of $\mathbb{Z}_n$ are the integers r such that $1 \leqslant r < n$ and $\gcd(r, n) = 1$.

Proof. To show $\mathbb{Z}_n$ is cyclic, we only need to show that $\mathbb{Z}_n = \langle x \rangle := \{e, x, \cdots, x^{n-1}\}$ for some $x \in \mathbb{Z}_n$. The choice $x = [1]_n$ would work.
We note that $r = 1 + \cdots + 1$ (r times). Let $b = r$ and $a = 1$ in the prop. 1.1.25 and conclude that the order of r is $\frac{n}{d}$ where $d = \gcd(k, n)$. Since the order of r, a generator of $\mathbb{Z}_n$, is n, we see $\frac{n}{d} = n \Rightarrow d = 1$. □

Example 1.2.4. Let us examine the group $\mathbb{Z}_{16}$. The numbers $1, 3, 5, 7, 9, 11, 13$, and 15 are the elements of $\mathbb{Z}_{16}$ that are relatively prime to 16 . Each of these elements generates $\mathbb{Z}_{16}$. For example,

$$\begin{array}{lll} 1 \cdot 9 = 9 & 2 \cdot 9 = 2 & 3 \cdot 9 = 11 \\ 4 \cdot 9 = 4 & 5 \cdot 9 = 13 & 6 \cdot 9 = 6 \\ 7 \cdot 9 = 15 & 8 \cdot 9 = 8 & 9 \cdot 9 = 1 \\ 10 \cdot 9 = 10 & 11 \cdot 9 = 3 & 12 \cdot 9 = 12 \\ 13 \cdot 9 = 5 & 14 \cdot 9 = 14 & 15 \cdot 9 = 7 \end{array}$$

We can also use the usual multiplication as binary operation on $\mathbb{Z}_n$:

$$[a]_n \times [b]_n = [ab]_n \tag{1.3}$$

Again, we should check that this really defines a binary operation on $\mathbb{Z}_n$: if $[a]_n = [a']_n$ and $[b]_n = [b']_n$ then we need $[ab]_n = [a'b']_n$. This is true because $a = a' + kn$ and $b = b' + \ln$ for some $k, l \in \mathbb{Z}$ so

$$\begin{aligned} ab &= \left(a' + kn\right)\left(b' + ln\right) \\ &= a'b' + n\left(kb' + la' + kln\right) \end{aligned}$$

so $ab \equiv a'b'(\mathrm{mod} n)$ and therefore $[ab]_n = [a'b']_n$. This does not make $(\mathbb{Z}_n, \times)$ into a group, because 0 has no inverse for the operation $\times$.

We notice that $(\mathbb{Z}_n, \times)$ where multiplication $\times$ is given by eq. (1.3) is a monoid with identity $[1]_n$. Therefore, due to Remark 1.1.2, we define $\mathbb{Z}_n^\times$ as the group of units in $\mathbb{Z}_n$, i.e.,

$$\mathbb{Z}_n^\times = \{l \in \mathbb{Z}_n | \gcd(l, n) = 1\}$$

(That's because $[lm]_n = [1]_n \Leftrightarrow lm \equiv 1 \bmod n) \Leftrightarrow \exists q \in \mathbb{Z}$ s.t. $lm - 1 = qn \Leftrightarrow \exists p(= -q) \in \mathbb{Z}$ s.t. $lm + pn = 1$) If $n = p$ is a prime, then

$$\mathbb{Z}_p^\times = \{l \in \mathbb{Z}_n | \gcd(l, p) = 1\} = \{[1], \cdots, [p-1]\}$$

where we note that The greatest common divisor of 0 and any non-zero number is the non-zero number itself (0 is a multiple of every non-zero number).

Example 1.2.5. If G is a cyclic group of order n, i.e., $G \cong \mathbb{Z}_n$, then $\mathrm{Aut}(G) \cong \mathbb{Z}_n^\times$.

Proof. Let $G = \langle x \rangle$ and

$$\begin{aligned} \phi : G &\to G \\ x &\mapsto x^l \end{aligned}$$

for some $0 \leqslant l \leqslant n-1$. Thus $\phi(x^j) = x^{lj}$. Every endomorphism (homomorphism with $G \to G$) is of this form, and we wonder what condition on l can make it an automorphism, i.e., also an isomorphism. In fact, ϕ is an isomorphism iff x^l is a generator of G. By theorem 1.2.3, we see this is the case iff $\gcd(n, l) = 1$. Since $\{l \in \mathbb{Z}_n | \gcd(n, l) = 1\} = \mathbb{Z}_n^\times$, we have an isomorphism:

$$\begin{aligned} \Phi : \mathrm{Aut}(G) &\to \mathbb{Z}_n^\times \\ \phi &\mapsto l \text{ where } \phi(x) = x^l \end{aligned}$$

(For $i = 1, 2$, $\phi_i \mapsto l_i \Rightarrow \phi_i(x) = x^{l_i}$, so $\phi_1 \circ \phi_2(x) = \phi_1(x^{l_2}) = x^{l_1 l_2}$.) □

1.2.2 Cyclic Groups

We begin with definition of **Euler φ-function**. $\varphi(n)$ is defined as the number of non-negative integers less than n that are relatively prime to n. In other words,

$$\varphi(n) = \begin{cases} 1 & \text{if } n = 1 \\ |\{l \in \mathbb{Z}_n : \gcd(l, n) = 1\}| = |\mathbb{Z}_n^\times| & \text{if } n > 1 \end{cases}.$$

Lemma 1.2.6. If $G = \langle a \rangle$ is cyclic of order n, then a^k is also a generator of G if and only if $(k, n) = 1$. Thus the number of generators of G is $\varphi(n)$.

Proof. This is just a restatement of Theorem 1.2.3. □

Lemma 1.2.7. If G is a cyclic group of order n, then there exists a unique subgroup of order d for every divisor d of n.

Proof. If $G = \langle a \rangle$, then $\langle a^{n/d} \rangle$ is a subgroup of order d, by Question 1.1-9. Assume that $S = \langle b \rangle$ is a subgroup of order d (S must be cyclic, by Theorem 1.1.22). Now $b^d = 1$; moreover, $b = a^m$ for some m. By Question 1.1-11, $md = nk$ for some integer k, and $b = a^m = \left(a^{n/d}\right)^k$. Therefore, $\langle b \rangle \leqslant \langle a^{n/d} \rangle$, and this inclusion is equality because both subgroups have order d. □

Theorem 1.2.8. If n is a positive integer, then

$$n = \sum_{d|n} \varphi(d),$$

where the sum is over all divisors d of n with $1 \leqslant d \leqslant n$.

Proof. If $\mathbb{C}$ is a cyclic subgroup of a group G, let $\operatorname{gen}(C)$ denote the set of all its generators. It is clear that G is the disjoint union

$$G = \bigcup \operatorname{gen}(C),$$

where $\mathbb{C}$ ranges over all the cyclic subgroups of G. We have just seen, when G is cyclic of order n, that there is a unique cyclic subgroup C_d of order d for every divisor d of n. Therefore, $n = |G| = \sum_{d|n} |\operatorname{gen}(C_d)|$. In Lemma 1.2.6, however, we saw that $|\operatorname{gen}(C_d)| = \varphi(d)$; the result follows. □

We now characterize finite cyclic groups.

Theorem 1.2.9 (characterization of cyclic group)**.** A group G of order n is cyclic if and only if, for each divisor d of n, there is at most one cyclic subgroup of G having order d.

Proof. If G is cyclic, then the result is Lemma 1.2.7. For the converse, recall from the previous proof that G is the disjoint union $\cup \operatorname{gen}(C)$, where $\mathbb{C}$ ranges over all the cyclic subgroups of G. Hence, $n = |G| = \sum |\operatorname{gen}(C)| \leqslant \sum_{d|n} \varphi(d) = n$, by Theorem 1.2.8. We conclude that G must have a cyclic subgroup of order d for every divisor d of n; in particular, G has a cyclic subgroup of order $d = n$, and so G is cyclic. □

Observe that the condition in Theorem 1.2.9 is satisfied if, for every divisor d of n, there are at most d solutions $x \in G$ of the equation $x^d = 1$ (two cyclic subgroups of order d would contain more than d solutions).

1.2.3 S_n and A_n

If X is a nonempty set, a **permutation** of X is a bijection $\alpha : X \to X$. We denote the set of all permutations of X by S_X. We will focus on the special case $X = 1, \cdots, n$, where S_X is denoted by S_n. Elements in it is of the form $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & \cdots & n-1 & n \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n-1} & \alpha_n \end{smallmatrix}\right)$ where $a_i = \alpha(i)$. S_n is a group, called **symmetric group**, with function composition as multiplication (and we keep the tradition of function composition that permutation of elements is applied from left to right). For example, $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{smallmatrix}\right)$ and $\beta = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right)$ are permutations of $\{1, 2, 3\}$. The product $\alpha\beta$ is $\left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{smallmatrix}\right)$. We compute the product by first applying β and then α:

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

Note that $\beta\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{smallmatrix}\right)$, so that $\alpha\beta \neq \beta\alpha$.

Definition 1.2.10. Let $i_1, i_2, \ldots, i_r$ be distinct integers between 1 and n. If $\alpha \in S_n$ fixes the remaining $n-r$ integers and if

$$\alpha(i_1) = i_2, \alpha(i_2) = i_3, \ldots, \alpha(i_{r-1}) = i_r, \alpha(i_r) = i_1,$$

then α is an r**-cycle**; one also says that α is a cycle of **length** r. Denote α by $(i_1\ i_2\ \cdots\ i_n)$. Every 1-cycle fixes every element of X, and so all 1-cycles are equal to the identity. A 2-cycle, which merely interchanges a pair of elements, is called a **transposition**. Observe that $(1\ 2\ 3\ \cdots\ r-1\ r) = (2\ 3\ \cdots\ r\ 1) = (r\ 1\ \cdots\ r-1)$, so there are exactly r such notations for this r-cycle.

Multiplication is easy when one uses the cycle notation. For example, let us compute $\gamma = \alpha\beta$, where $\alpha = (1\ 2)$ and $\beta = (1\ 3\ 4\ 2)$. Since multiplication is composition of functions, $\gamma(1) = \alpha \circ \beta(1) = \alpha(\beta(1)) = \alpha(3) = 3$; Next, $\gamma(3) = \alpha(\beta(3)) = \alpha(4) = 4$, and $\gamma(4) = \alpha(\beta(4)) = \alpha(2) = 1$. Having returned to 1, we now seek $\gamma(2)$, because 2 is the smallest integer for which γ has not yet been evaluated. We end up with $(1\ 2)(1\ 3\ 4\ 2\ 5) = (1\ 3\ 4)(2\ 5)$. The cycles on the right are disjoint as defined below.

Definition 1.2.11. Two permutations $\alpha, \beta \in S_X$ are **disjoint** if every x moved by one is fixed by the other. In symbols, if $\alpha(x) \neq x$, then $\beta(x) = x$ and if $\beta(y) \neq y$, then $\alpha(y) = y$ (of course, it is possible that there is $z \in X$ with $\alpha(z) = z = \beta(z)$). A family of permutations $\alpha_1, \alpha_2, \ldots, \alpha_m$ is **disjoint** if each pair of them is disjoint. Observe that for $\alpha = (i_1\ i_2\ \cdots\ i_r)$ and $\beta = (j_1\ j_2\ \cdots\ j_s)$, α and β are disjoint if and only if $\{i_1, i_2, \ldots, i_r\} \cap \{j_1, j_2, \ldots, j_s\} = \varnothing$.

The identity of S_n is 1, or (1). To find the inverse of a permutation just write it backwards. If $\tau = (1243)(67)$ then $\tau^{-1} = (76)(3421)$ which can then be rewritten as $\tau^{-1} = (1342)(67)$.

How does one prove this?

First consider a single cycle: $\sigma = (a_1 a_2 \ldots a_k)$. If $b \notin \{a_1, \ldots, a_k\}$, then $\sigma(b) = b$ so $\sigma^{-1}(b) = b$. Thus b shouldn't appear in the inverse. Next $\sigma(a_i) = a_{i+1}$ so $\sigma^{-1}(a_{i+1}) = a_i$. Thus if $\sigma : a_1 \mapsto a_2 \mapsto a_3 \mapsto \cdots \mapsto a_k \mapsto a_1$, then $\sigma^{-1} : a_k \mapsto a_{k-1} \mapsto a_{k-2} \mapsto \cdots \mapsto a_1 \mapsto a_k$. This is precisely the cycle $(a_k, a_{k-1} \ldots, a_2, a_1)$ which is nothing more than σ written backwards.

Now what about a list of cycles? Say $\sigma = \sigma_1 \cdots \sigma_\ell$. Recall that $\sigma^{-1} = (\sigma_1 \cdots \sigma_\ell)^{-1} = \sigma_\ell^{-1} \cdots \sigma_1^{-1}$. So we reverse the list of cycles and then write each one backwards – thus the inverse is just the whole thing written backwards.

One thing to note: This still works even if σ is not written in terms of disjoint cycles.

Proposition 1.2.12. If α and β are disjoint permutations, then $\alpha\beta = \beta\alpha$; that is, α and β commute.

Proof. See [5] Proposition 5.8. □

Now we present results for factorization or permuations.

Theorem 1.2.13. Every permutation $\alpha \in S_n$ is either a cycle or a product of disjoint cycles.

Proof. see [9] Theorem 1.1. □

Theorem 1.2.14. Every permutation $\alpha \in S_n$ is a product of transpositions.

Proof. By Theorem 1.2.13, it is enough to factor cycles: for $n > 1$,

$$\sigma = (a_1\ \ldots\ a_n) = (a_1\ a_n)(a_1\ a_{n-1}) \ldots (a_1\ a_2)$$

□

One can prove that the parity of the number of factors is the same for all factorizations of a permutation a that is, the number of transpositions is always even or odd. We say that a permuation is **even** if it has even parity and is **odd** if it has odd parity. See [9] p.8-9 for more of this.

Corollary 1.2.15. A cycle $\sigma = (a_1 \ \dots \ a_n)$ is even if and only if n is odd.

One of the most important subgroups of S_n is the set of all even permutations, A_n. The group A_n is called the alternating group on n letters.

Theorem 1.2.16. The set A_n is a subgroup of S_n.

Proof. Since the product of two even permutations must also be an even permutation, A_n is closed. The identity is an even permutation and therefore is in A_n. If σ is an even permutation, then

$$\sigma = \sigma_1 \sigma_2 \cdots \sigma_r$$

where σ_i is a transposition and r is even. Since the inverse of any transposition is itself,

$$\sigma^{-1} = \sigma_r \sigma_{r-1} \cdots \sigma_1$$

is also in A_n. □

Proposition 1.2.17. The number of even permutations in $S_n, n \geqslant 2$, is equal to the number of odd permutations; hence, the order of A_n is $n!/2$.

Proof. Let A_n be the set of even permutations in S_n and B_n be the set of odd permutations. If we can show that there is a bijection between these sets, they must contain the same number of elements. Fix a transposition σ in S_n. Since $n \geqslant 2$, such a σ exists. Define

$$\lambda_\sigma : A_n \to B_n$$

by

$$\lambda_\sigma(\tau) = \sigma\tau.$$

Suppose that $\lambda_\sigma(\tau) = \lambda_\sigma(\mu)$. Then $\sigma\tau = \sigma\mu$ and so

$$\tau = \sigma^{-1}\sigma\tau = \sigma^{-1}\sigma\mu = \mu.$$

Therefore, λ_σ is one-to-one. The proof that λ_σ is surjective is left as an exercise. □

Example 1.2.18 (Subgroups of A_4)**.** The group A_4 is the subgroup of S_4 consisting of even permutations. There are twelve elements α_1-α_{12} in A_4: an identity α_1, three permutations written as products of two disjoint cycles α_2-α_4 (each of them having order 2), and eight cycles α_5-α_{12} fixing one element (each of them having order 3). We have the Cayley table of A_4 below (In this table, an entry k inside the table represents α_k. For example, $\alpha_3\alpha_8 = \alpha_6$.)

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}
$(1) = \alpha_1$	1	2	3	4	5	6	7	8	9	10	11	12
$(12)(34) = \alpha_2$	2	1	4	3	6	5	8	7	10	9	12	11
$(13)(24) = \alpha_3$	3	4	1	2	7	8	5	6	11	12	9	10
$(14)(23) = \alpha_4$	4	3	2	1	8	7	6	5	12	11	10	9
$(123) = \alpha_5$	5	8	6	7	9	12	10	11	1	4	2	3
$(243) = \alpha_6$	6	7	5	8	10	11	9	12	2	3	1	4
$(142) = \alpha_7$	7	6	8	5	11	10	12	9	3	2	4	1
$(134) = \alpha_8$	8	5	7	6	12	9	11	10	4	1	3	2
$(132) = \alpha_9$	9	11	12	10	1	3	4	2	5	7	8	6
$(143) = \alpha_{10}$	10	12	11	9	2	4	3	1	6	8	7	5
$(234) = \alpha_{11}$	11	9	10	12	3	1	2	4	7	5	6	8
$(124) = \alpha_{12}$	12	10	9	11	4	2	1	3	8	6	5	7

We will find all subgroups of A_4: since the order of $H \leqslant A_4$ must divide the order of A_4 and $|A_4| = 12 = 1 \times 12 = 3 \times 4 = 2 \times 6$, we see H can have size $1, 2, 3, 4, 6, 12$. H with $|H| = 1$ and 12 are just trivial subgroup and A_4 itself. Thanks to Question 10, we already know the classification of all groups with size smaller than 6: subgroups H with $|H| = 2, 3, 5$ are isomorphic to $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5$ and H with $|H| = 4$ is isomorphic either to $\mathbb{Z}_4$ or $\mathbf{V}$. There is no subgroup of order 6 (proved in the following lemma).

By observations about α_2-α_4 and α_5-α_{12} we made in the beginning, we see subgroups of order 2 are just $\langle\alpha_2\rangle, \cdots, \langle\alpha_4\rangle$; and subgroups of order 3 are just $\langle\alpha_5\rangle, \cdots, \langle\alpha_{12}\rangle$. Since there is no element with order 4 in A_4, subgroup H of order 4 can only be $\mathbf{V}$, which is contained in A_4 as $\{\alpha_1, \cdots, \alpha_4\}$. Our classification is complete.

Lemma 1.2.19. There is no subgroup of index 2 in A_4.

Proof. Suppose a subgroup H of A_4 has index 2, i.e., $|H| = 6$. We will show for each $g \in A_4$ that $g^2 \in H$. If $g \in H$ then clearly $g^2 \in H$. If $g \notin H$ then gH is a left coset of H different from H (since $g \in gH$ and $g \notin H$), so from $[G : H] = 2$ the only left cosets of H are H and gH. Which one is g^2H ? If $g^2H = gH$ then $g^2 \in gH$, so $g^2 = gh$ for some $h \in H$, and that implies $g = h$, so $g \in H$, but that's a contradiction. Therefore $g^2H = H$, so $g^2 \in H$. Every 3-cycle $(a\ b\ c)$ in A_4 is a square: (abc) has order 3, so $(a\ b\ c) = (a\ b\ c)^4 = \big((a\ b\ c)^2\big)^2$. Thus H contains all 3-cycles in A_4, in total 8 of them, which thus contradicts to $|H| = 6$. □

1.2.4 D_n

We from example 1.2.18 see that the Klein-four group $\mathbf{V}$ is a subgroup of A_4 and is thus a subgroup of S_4. We remarked in example 1.1.11 that $\mathbf{V}$ is isomorphic to D_2. We call subgroups of S_n **permuation groups**. In last subsection, we examined alternating groups A_n; now we examine another type of permuation groups, the dihedral groups D_n. Such groups consist of the rigid motions of a regular n-sided polygon or n-gon. For $n = 3, 4, \ldots$, we define the **n-th dihedral group** to be the group of rigid motions of a regular n-gon. We will denote this group by D_n. We can number the vertices of a regular n-gon by $1, 2, \ldots, n$. Notice that there are exactly n choices to replace the first vertex. If we replace the first vertex by k, then the second vertex must be replaced either by vertex $k + 1$ or by vertex $k - 1$; hence, there are $2n$ possible rigid motions of the n-gon. We summarize these results in the following theorem.

Theorem 1.2.20. The dihedral group, D_n, is a subgroup of S_n of order $2n$.

Theorem 1.2.21 (Dihedral group)**.** The group $D_n, n \geqslant 3$, consists of all products of the two elements r and s, where r has order n and s has order 2, and these two elements satisfy the relation $(sr)^2 = 1$.

Proof. The possible motions of a regular n-gon are either reflections or rotations (Figure 1.1).

There are exactly n possible rotations:

$$\text{id}, \frac{360^\circ}{n}, 2 \cdot \frac{360^\circ}{n}, \ldots, (n-1) \cdot \frac{360^\circ}{n}.$$

We will denote the rotation $360^\circ/n$ by r. The rotation r generates all of the other rotations. That is,

$$r^k = k \cdot \frac{360^\circ}{n}$$

Label the n reflections $s_1, s_2, \ldots, s_n$, where s_k is the reflection that leaves vertex k fixed. There are two cases of reflections, depending on whether n is even or odd. If there are an even number of vertices, then two vertices are left fixed by a reflection, and $s_1 = s_{n/2+1}, s_2 = s_{n/2+2}, \ldots, s_{n/2} = s_n$. If there are an odd number of vertices, then only a single vertex is left fixed by a reflection and $s_1, s_2, \ldots, s_n$ are distinct (Figure 1.2).

In either case, the order of each s_k is two. Let $s = s_1$. Then $s^2 = 1$ and $r^n = 1$. Since any rigid motion t of the n-gon replaces the first vertex by the vertex k, the second vertex must be replaced by either $k + 1$ or by

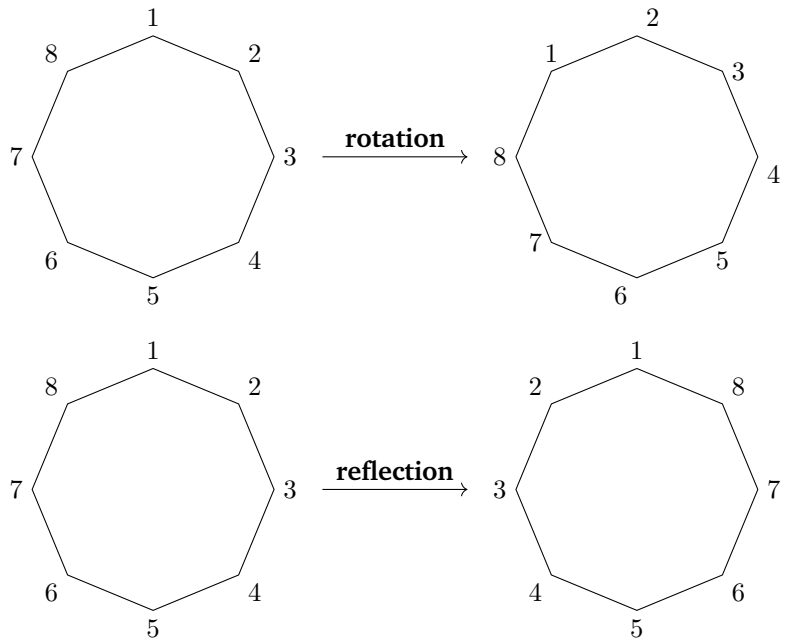


Figure 1.1: Rotations and reflections of a regular n-gon

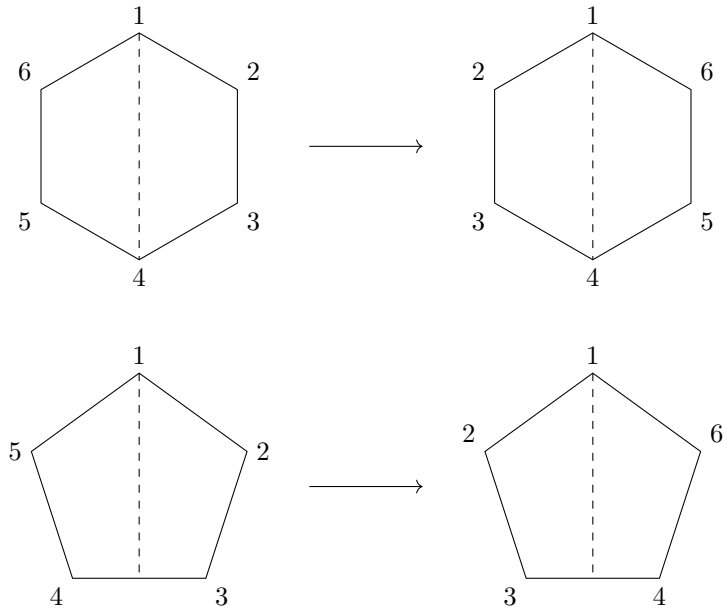


Figure 1.2: Types of reflections of a regular n-gon

$k-1$. If the second vertex is replaced by $k+1$, then $t = r^k$. If the second vertex is replaced by $k-1$, then $t = r^k s$. Hence, r and s generate D_n. That is, D_n consists of all finite products of r and s,

$$D_n = \left\{1, r, r^2, \ldots, r^{n-1}, s, rs, r^2 s, \ldots, r^{n-1} s\right\}.$$

We will leave the proof that $(sr)^2 = 1$ as an exercise. □

Example 1.2.22. The group of rigid motions of a square, D_4, consists of eight elements. With the vertices numbered 1,2,3,4 (Figure 1.3), the rotations are

$$\begin{aligned} r &= (1\ 2\ 3\ 4) \\ r^2 &= (1\ 3)(2\ 4) \\ r^3 &= (1\ 4\ 3\ 2) \\ r^4 &= (1) \end{aligned}$$

and the reflections are

$$\begin{aligned} s_1 &= (2\ 4) \\ s_2 &= (1\ 3). \end{aligned}$$

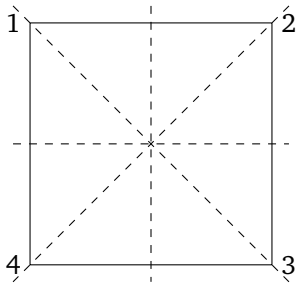


Figure 1.3: The group D_4

The order of D_4 is 8. The remaining two elements are

$$\begin{aligned} rs_1 &= (12)(34) \\ r^3 s_1 &= (14)(23). \end{aligned}$$

A Supplementary Note

One can also analyze group of symmetry of solids. For example, group of rigid motions of a cube is S_4 (Figure 1.4) (see [5] Theorem 5.27). For more on this, including the Planotic solids, see [1] section 6.12.

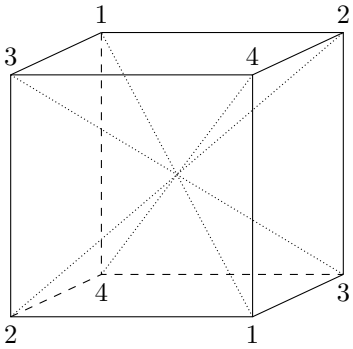


Figure 1.4: cube

1.2 EXERCISES

1. If $1 \leqslant r \leqslant n$, then there are $(1/r)[n(n-1)\dots(n-r+1)]$ r-cycles in S_n.

2. If $\alpha, \beta \in S_n$ are disjoint and $\alpha\beta = 1$, then $\alpha = 1 = \beta$.

3. Let $\alpha \in S_n$ for $n \geqslant 3$. If $\alpha\beta = \beta\alpha$ for all $\beta \in S_n$, prove that α must be the identity permutation; hence, the center of S_n is the trivial subgroup (the center of a group G is defined as $Z(G) = \{g \in G : gx = xg \text{ for all } x \in G\}$.)

4. If $\sigma \in A_n$ and $\tau \in S_n$, show that $\tau^{-1}\sigma\tau \in A_n$.

5. Let $\tau = (a_1, a_2, \dots, a_k)$ be a cycle of length k.

i. Prove that if σ is any permutation, then

$$\sigma\tau\sigma^{-1} = (\sigma(a_1), \sigma(a_2), \dots, \sigma(a_k))$$

is a cycle of length k.

ii. Let μ be a cycle of length k. Prove that there is a permutation σ such that $\sigma\tau\sigma^{-1} = \mu$.

6. [9][p.24 ex2.9]

i. Prove that S_n can be generated by $(1\ 2), (1\ 3), \cdots, (1\ n)$.

ii. Prove that S_n can be generated by $(1\ 2), (2\ 3), \cdots, (i\ i+1), \cdots, (n-1\ n)$.

iii. Prove that S_n can be generated by the two elements $(1\ 2)$ and $(1\ 2\ \cdots n)$.

7. Draw group tables of S_2 and S_3.

8. [9][p.5 ex1.12]

i. Let $\alpha = \begin{pmatrix} i_0 & i_1 & \dots & i_{r-1} \end{pmatrix}$ be an r-cycle. For every $j, k \geqslant 0$, prove that $\alpha^k(i_j) = i_{k+j}$ if subscripts are read modulo r.

ii. Prove that if α is an r-cycle, then $\alpha^r = 1$, but that $\alpha^k \neq 1$ for every positive integer $k < r$.

iii. If $\alpha = \beta_1\beta_2\dots\beta_m$ is a product of disjoint r_i-cycles β_i, then the smallest positive integer l with $\alpha^l = 1$ is the least common multiple of $\{r_1, r_2, \dots, r_m\}$. Therefore, the order of a permutation $\alpha = \beta_1 \cdots \beta_t$, where β_i is an r_i-cycle, is $\operatorname{lcm}\{r_1, \cdots, r_t\}$.

9. By previous question, deduce that each order-3 cycle is a product of 3-cycles.

10. Dihedral group.

i. Show that $D_n = \langle r, s | r^n, s^2, (sr)^2 \rangle = D_n = \langle r, s | r^n, s^2, (rs)^2 \rangle$, that is, $r^n = 1, s^2 = 1, (sr)^2 = 1$ iff $r^n = 1, s^2 = 1, (rs)^2 = 1$.

ii. Show that $r^k s = sr^{-k}$ in D_n.

iii. Prove that the order of $r^k \in D_n$ is $n/\gcd(k, n)$.

11. Show that there is an index-2 subgroup of Dihedral group D_n.

1.3 Normal Subgroups and Quotient Groups

Definition 1.3.1. Subgroup $H \leqslant G$ is **normal**, denoted as $H \trianglelefteq G$, if $\forall g \in G,\ gHg^{-1} \subseteq H$.

Note that $gHg^{-1} = \{ghg^{-1} | h \in H\} \leqslant G$, as $ghg^{-1}(gh'g^{-1})^{-1} \in gHg^{-1}$.

Example 1.3.2.

- If G is an abelian group, then every subgroup of G is normal. The converse is false: see Question 1.3-4.

- $\mathrm{SL}(n,\mathbb{R})$ is a normal subgroup of $\mathrm{GL}(n,\mathbb{R})$: for $A \in \mathrm{GL}(n,\mathbb{R}), B \in \mathrm{SL}(n,\mathbb{R})$ we have $\det(ABA^{-1}) = \det(A)\det(B)\det(A^{-1}) = \det(A)\det(A^{-1}) = 1$.

Proposition 1.3.3 (characterization of normal subgroup)**.** If $H \leqslant G$, then the following are equivalent.

1. $H \trianglelefteq G$;
2. $\forall g \in G,\ gHg^{-1} = H$;
3. $\forall g \in G,\ Hg = gH$;
4. Every right coset of H is a left coset;
5. Every left coset of H is a right coset.

Proof. 1 equiv to 2: the $\Leftarrow$ direction is clear. Conversely, suppose $\forall g \in G,\ gHg^{-1} \subseteq H$, so $g^{-1}H(g^{-1})^{-1} \subseteq H \implies g^{-1}Hg \subseteq H$. Multiply from left and right to cancel, so $H \subseteq gHg^{-1}$. So $gHg^{-1} = H$.

2 equiv to 3: $\forall g \in G,\ gHg^{-1} = H \iff \forall g \in G, h \in H$, there is some $h' \in H$ such that $h' = ghg^{-1} \iff \forall g \in G, h \in H, \exists h' \in H$ s.t. $h'g = gh$.

We prove that 3,4,5 are equivalent.

3 implies 4: we note that 3 is directly stronger than 4, as 4 can be rephrased as: for a right coset Hg, there is some $g' \in G$ such that $Hg = g'H$.

4 implies 3: Suppose $Hg = aH$ for some a. But then $g \in Hg = aH$, and $g \in gH$. So $aH = gH \implies Hg = gH$.

3 implies 5 implies 3: similarly. □

Corollary 1.3.4. Any subgroup of index 2 in any group G is normal.

Proof. $[G : H] = 2 \implies$ two distinct left cosets, H, aH where $a \notin H$. Similarly, H and Ha are distinct right cosets. This gives $H \cap aH = \varnothing, H \cap Ha = \varnothing$, so by 4 in proposition 1.3.3, H is normal. □

If $N \trianglelefteq G$, then the set of cosets of N in G, G/N, form a group under multiplication $(aN)(bN) = abN$. We need to check that

- Well-defined: $aN = a'N$ and $bN = b'N \implies abN = a'b'N$:

$$\begin{aligned} NaNb &= Na\left(a^{-1}Na\right)b \quad \text{(because } N \text{ is normal)} \\ &= N\left(aa^{-1}\right)Nab = NNab = Nab \quad \text{(because } N \leqslant G\text{)}. \end{aligned}$$

  Thus, $NaNb = Nab$, and so the product of two cosets is a coset.

- Group properties easily follow from the group properties of G (associativity, identity $N = N1 = 1N$, and inverse $a^{-1}N$ $(= Na^{-1})$ for aN $(= Na)$.)

Proposition 1.3.5. If $N \trianglelefteq G$, then the **natural map**, or **canonical projection** (i.e., the function $q : G \to G/N$ defined by $q(a) = Na$) is a surjective homomorphism with kernel N.

Proof. The equation $q(a)q(b) = q(ab)$ is just the formula $NaNb = Nab$; hence, q is a homomorphism. If $Na \in G/N$, then $Na = q(a)$, and so v is surjective. Finally, $q(a) = Na = N$ if and only if $a \in N$, so that $N = Ker(q)$. □

We define conjugation $\gamma_a : G \to G$, where $\gamma_a(x) = axa^{-1}$, and call $\gamma_a(x) = axa^{-1}$ a **conjugate of** x in a group G, also denoted as x^a. Moreover, for $g \in G$ we set

$$H^g := gHg^{-1}$$

and say that H^g is a **conjugate of** H in G (more precisely, the conjugate of H by g). For any $K \subseteq G$ set

$$H^K := \left\{ H^k \mid k \in K \right\}.$$

We have now shown in Proposition 1.3.5 that every normal subgroup is the kernel of some homomorphism. Different homomorphisms can have the same kernel. For example, if $a = (1\ 2)$ and $b = (1\ 3)$, then $\gamma_a, \gamma_b : S_3 \to S_3$ are distinct and $Ker(\gamma_a) = 1 = Ker(\gamma_b)$.

The quotient group construction is a generalization of the construction of $\mathbb{Z}_n$ from $\mathbb{Z}$. Recall that if n is a fixed integer, then $[a]$, the congruence class of $a \bmod n$, is the coset $a + \langle n \rangle$. Now $\langle n \rangle \trianglelefteq \mathbb{Z}$, because $\mathbb{Z}$ is abelian, and the quotient group $\mathbb{Z}/\langle n \rangle$ has elements all cosets $a + \langle n \rangle$, where $a \in \mathbb{Z}$, and operation $(a + \langle n \rangle) + (b + \langle n \rangle) = a + b + \langle n \rangle$; in congruence class notation, $[a] + [b] = [a + b]$. Therefore, the quotient group $\mathbb{Z}/\langle n \rangle$ is equal to $\mathbb{Z}_n$, the group of integers modulo n. An arbitrary quotient group G/N is often called $G \bmod N$ because of this example.

1.3 EXERCISES

1. [9][p.31 ex2.29]

i. (H. B. Mann). Let G be a finite group, and let S and T be (not necessarily distinct) nonempty subsets. Prove that either $G = ST$ or $|G| \geqslant |S| + |T|$.
ii. Prove that every element in a finite field F is a sum of two squares.

2. [9][p.31 ex2.32] If $H \leqslant G$, then $H \trianglelefteq G$ if and only if, for all $x, y \in G, xy \in H$ if and only if $yx \in H$.

3. [9][p.31 ex2.33] If $K \leqslant H \leqslant G$ and $K \trianglelefteq G$, then $K \trianglelefteq H$.

4. Every subgroup of an abelian group is normal. This exercise shows that the converse is not true: Let G be the subgroup of $\mathrm{GL}(2, \mathbb{C})$ generated by

$$A = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

i. Find the order of A and B in G.
ii. Show G has order 8 by listing all the elements of G. Show G is is not abelian.
iii. List all elements of oder 2 in G.
iv. Show that every subgroup of G is normal.

5. If N, H_1, H_2 are subgroups of a group G such that $N \trianglelefteq G$ and $H_1 \trianglelefteq H_2$, then show $NH_1 \trianglelefteq NH_2$.

6. Prove that $A_n \trianglelefteq S_n$ for every n by showing that it is an index-2 subgroup (thus $|A_n| = \frac{1}{2}n!$).

7. [9][p.31 ex2.37]

i. The intersection of any family of normal subgroups of a group G is itself a normal subgroup of G. Conclude that if X is a subset of G, then there is a smallest normal subgroup of G which contains X; it is called the normal subgroup generated by X (or the normal closure of X; it is often denoted by $\langle X \rangle^G$).
ii. If $X = \varnothing$, then $\langle X \rangle^G = 1$. If $X \neq \varnothing$, then $\langle X \rangle^G$ is the set of all words on the conjugates of elements in X.
iii. If $gxg^{-1} \in X$ for all $x \in X$ and $g \in G$, then $\langle X \rangle = \langle X \rangle^G \trianglelefteq G$.

8. [9][p.31 ex2.38] If $H, K \trianglelefteq G$, then $\langle H \cup K \rangle \trianglelefteq G$.

9. Suppose $f : G \to G'$ is a homomorphism. Show that $N \trianglelefteq G \Rightarrow f(N) \trianglelefteq G'$; $N' \trianglelefteq G' \Rightarrow f^{-1}(N') \trianglelefteq G$.

10. Finite product (see Definition 1.4.3) and finite intersection of normal subgroups of G are still normal.

11. Suppose $H \leqslant G$ and $N \trianglelefteq G$. Show that $H \cap N \trianglelefteq H$ but not necessarily $H \cap N \trianglelefteq G$ Also note that $H \leqslant N \trianglelefteq G$ does *not* imply $H \trianglelefteq G$; not even $H \trianglelefteq N \trianglelefteq G$ implying $H \trianglelefteq G$. Show such transitivity of normality fails by the counterexample that $K = \langle (1\ 2)(3\ 4) \rangle \trianglelefteq \mathbf{V}$ and $\mathbf{V} \trianglelefteq S_4$ while K is not a subgroup of S_4.

12. (Product formula) If S and T are subgroups of a finite group G, then $|ST||S \cap T| = |S||T|$.

13. Show that conjugacy is an equivalence relation, that is, $x \sim y \iff \exists g \in G$ s.t. $y = x^g := gxg^{-1}$ defines an equivalence relation. We call the equivalence class with respect to this relation **conjugacy class**. Use this definition to show that a subgroup $H \leqslant G$ is normal if and only if it is a union of conjugacy classes of G.

1.4 Isomorphism Theorems

Facts (proofs are left as exercises): for a group homomorphism $\phi : G \to G'$,

1. $Ker(\phi) := \{a \in G | \phi(a) = e_{G'}\} \trianglelefteq G$
2. $\mathrm{Im}(\phi) := \{\phi(a) | a \in G\} \leqslant G'$

Theorem 1.4.1 (1st Isomorphism Theorem)**.** If $f : G \to G'$ is a group homomorphism and $K = Ker(f)$ (so $K \trianglelefteq G$), then

$$G/K \cong \mathrm{Im}(f)$$

Proof. Define $\phi : G/K \to \mathrm{Im}(f)$ by $\phi(aK) = f(a)$. ϕ is well-defined and injective: $aK = bK \iff a^{-1}b \in K = Ker(f) \iff f(a^{-1}b) = f(a)^{-1}f(b) = e \iff f(b) = f(a)$. ϕ is a homomorphism: $\phi(aKer(f)bKer(f)) = \phi(abKer(f))$ since kernel is normal group and that is $f(ab)$. On the other side, $\phi(aKer(f))\phi(bKer(f)) = f(a)f(b)$, so this is homomorphism since f is homomorphism. Lastly, ϕ is surjective: if $b \in \mathrm{Im}(f)$, then $b = f(a)$ for some a. So $\phi(aKer(f)) = b$. □

Example 1.4.2. $\mathrm{SL}(n, \mathbb{R}) \trianglelefteq \mathrm{GL}(n, \mathbb{R})$. Then $\mathrm{GL}(n, \mathbb{R})/\mathrm{SL}(n, \mathbb{R}) \simeq (\mathbb{R} - \{0\}, \cdot)$.

Proof. $f : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R} - \{0\}, A \mapsto \det(A)$. This is a group homomorphism, f is surjective, $Ker(f) = \mathrm{SL}(n, \mathbb{R}) \implies \mathrm{GL}(n, \mathbb{R})/\mathrm{SL}(n, \mathbb{R}) \simeq \mathbb{R} - \{0\}$. □

Definition 1.4.3. For $H, K \leqslant G$, define **product set**

$$HK = \{hk | h \in H, k \in K\}$$

and **inverse set**

$$H^{-1} = \left\{h^{-1} | h \in H\right\}$$

Remark 1.4.4.

1. HK is not necessarily a subgroup of G. For example, consider $G = S_3$ and $H = \{e, (1\ 2)\}$, $K = \{e, (1\ 3)\}$. We have Proposition 1.4.5 (same as [9] Lemma 2.25) instead.
2. Observe that $(AB)^{-1} = B^{-1}A^{-1}$.

Proposition 1.4.5. Let A and B be subgroups of G. Then AB is a subgroup of G if and only if $AB = BA$.

Proof. From $AB \leqslant G$ we get

$$(AB) = (AB)^{-1} = B^{-1}A^{-1} = BA.$$

If $AB = BA$, then

$$(AB)(AB) = A(BA)B = A(AB)B = AABB = AB$$

and

$$(AB)^{-1} = B^{-1}A^{-1} = BA = AB.$$

Thus $AB \leqslant G$. □

Proposition 1.4.6. If $H \leqslant G$ and $N \trianglelefteq G$, then $HN \leqslant G$, $HN = NH$, and HN is the subgroup of G generated by $H \cup N$.

Proof. $HN \leqslant G$: If $a = h_1n_1, b = h_2n_2$, then $ab^{-1} = h_1n_1n_2^{-1}h_2^{-1} = h_1h_2^{-1}h_2n_1n_2^{-1}h_2^{-1}$. Clearly, $n_1n_2^{-1} \in N$ so $h_2n_1n_2^{-1}h_2^{-1} \in N$. Thus, $ab^{-1} \in HN$.

$HN = NH$: We need to first show $HN \subseteq NH$. Let $hn \in HN \implies hnh^{-1} = n' \in N \implies hn = n'h \in NH$, so $HN \subseteq NH$. Similar for other direction.

Clearly, $H, N \subseteq HN \leqslant G$. And for any $K \leqslant G$, let $H, N \subseteq K$. Since K is a subgroup, $\forall n \in N, h \in H, hn \in K$. Thus $HN \leqslant K$ is the smallest subgroup. In particular, HN is the subgroup generated by $H \cup N$. □

Theorem 1.4.7 (2nd Isomorphism Theorem)**.** Let $H \leqslant G, N \trianglelefteq G$. Then $H \cap N \trianglelefteq H$ and

$$H/H \cap N \simeq HN/N$$

Proof. $H \cap N \trianglelefteq H$ due to Question 1.3-11. Let $\phi : H \to HN/N$ be given by $\phi(h) = hN$. The result follows from the first isomorphism theorem after showing the following three facts. We left them as exercises.

- $Ker(\phi) = \{h \in H | hN = N\} = H \cap N$.
- ϕ is surjective: $hnN = hN = \phi(h)$.
- ϕ is homomorphism.

□

Suppose $H_2 \subseteq H_1, H_1, H_2 \trianglelefteq G$. Then we can define a surjective map called the **enlargement of coset**:

$$\phi : \frac{G}{H_2} \to \frac{G}{H_1}; aH_2 \mapsto aH_1$$

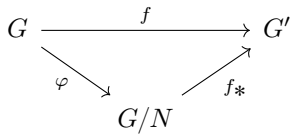
It is well-defined: if $aH_2 = bH_2 \Leftrightarrow b^{-1}a \in H_2 \subseteq H_1 \Rightarrow b^{-1}a \in H_1 \Leftrightarrow aH_1 = bH_2$, then $\phi\left(aH_2\right) = \phi\left(a'H_2\right)$. It is a homomorphism: $\phi\left(aH_2\right)\phi\left(bH_2\right) = \left(aH_1\right)\left(bH_1\right) = aH_1 = \phi\left(abH_2\right)$. It is surjective: for every $aH_1 \in \frac{G}{H_1}$, we have $\phi\left(aH_2\right) = aH_1$. Therefore, by 1st isomorphism theorem, $\frac{G}{H_2}/Ker(\phi) \cong \frac{G}{H_1}$, so G/H_1 is a quotient of G/H_2.

Remark 1.4.8.

(1) Now, let G be a group and $N \trianglelefteq G$. Let $f : G \to G'$ be a homomorphism whose kernel $K = Ker(f)$ contains N. Then we have a composition

$$f_* = \psi \circ \phi : \frac{G}{N} \to \frac{G}{K} \to G'; aN \mapsto aK \mapsto f(a)$$

where $\psi : G/K \to G'$ is the homomorphism from the 1st isomorphism theorem and $\phi : G/N \to G/K$ is the enlargement of coset. This composition g is the unique homomorphism $f_* : G/N \to G'$, said to be **induced by** f, making the following diagram commutative:



As before, φ is the canonical projection.

(2) Now, let G again be a group. Let $f : G \to G'$ be a homomorphism. Consider $N' \trianglelefteq G'$ and $N := f^{-1}(N') \trianglelefteq G$ instead (the normality is justified by Proposition 1.4.14). Consider the composition

$$g = q \circ f : G \to G' \to \frac{G'}{N'}$$

in replace of the homomorphism f in the above commutative diagram, where $q : G' \to G'/N'$ is the canonical projection. Observe that $K = Ker(g) = \{x \in G | f(x) \in N'\} = f^{-1}(N') = N$, so the enlargement $\phi : G/N \to G/K$ degenerates to the identity homomorphism i and the induced map $g_* : \frac{G}{N} \to \frac{G}{K} \to \frac{G'}{N'}; aN \mapsto aK \mapsto g(a)$ becomes the homomorphism in the first isomorphism theorem $g_* = \psi : \frac{G}{N} \to \frac{G'}{N'}; aN \to g(a)$. The map is then injective as ψ is injective.

Theorem 1.4.9 (3rd Isomorphism Theorem)**.** Suppose $K \leqslant N \trianglelefteq G$ and $K \trianglelefteq G$. Then

$$N/K \trianglelefteq G/K \text{ and } (G/K)/(N/K) \simeq G/N$$

Proof. First part follows from definition. Application of the first isomorphism theorem to the enlargement of coset map $\phi : G/K \to G/N$, $\phi(gK) = gN$ will prove the second part (check that $Ker(\phi) = N/K$ and ϕ is surjective). □

Restating the proof that $\phi : G/K \to \mathrm{Im}(f)$, defined in the first isomorphism theorem, is well-defined, we get

Proposition 1.4.10 ([1] Proposition 2.7.1)**.** Let K be the kernel of a homomorphism $\varphi : G \to G'$. Let $b \in G'$, then $\varphi^{-1}(b)$ is called a **fiber**. If $a \in \varphi^{-1}(b)$, then $\varphi^{-1}(b) = aK$, the coset of K containing a. These cosets partition the group G, and they correspond to elements of the image of φ:

$$\begin{aligned} G/K &\longleftrightarrow \mathrm{Im}(\varphi) \\ aK &\longleftrightarrow \varphi(a) \end{aligned}$$

Since $|G/K| = [G : K]$ for finite group G, and $|G/K| = |\mathrm{Im}(\varphi)|$ by the above proposition, we immediately have

Corollary 1.4.11 ([1] Corollary 2.8.13)**.** Let $\varphi : G \to G'$ be a homomorphism of finite groups. Then

- $|G| = |Ker(\varphi)| \cdot |\mathrm{Im}\,\varphi|$;
- $|Ker(\varphi)|$ divides $|G|$;
- $|\mathrm{Im}(\varphi)|$ divides both $|G|$ and $|G'|$.

Proposition 1.4.12. Let $\varphi : G \to G'$ be a homomorphism and $H \leqslant G$. Then the restriction $\varphi|_H : H \to G'$ is also a homomorphism with $Ker(\varphi|_H) = Ker(\varphi) \cap H$ and $\mathrm{Im}(\varphi|_H) = \varphi(H)$.

Remark 1.4.13. By Corollary 1.4.11, we see $|\mathrm{Im}\,\varphi_H| = |\varphi(H)|$ divides $|H|$ and $|G'|$. Therefore, if $|H|$ and $|G'|$ have no common factors, then $|\varphi(H)| = 1 \implies \varphi(H) = e_{G'} \implies \varphi$ is a trivial homomorphism. [1] Example 2.10.3 gives an application of this observation on the sign homomorphism from S_n to $\{\pm 1\} \cong \mathbb{Z}_2$. This will require some readings in permutation matrices that define the sign homomorphism ([1] handles the sign of permutation in a neater way than [9] does).

Proposition 1.4.14 ([1] Proposition 2.10.4)**.** Let $\varphi : G \to \mathcal{G}$ be a homomorphism with kernel K and let $\mathcal{H}$ be a subgroup of $\mathcal{G}$. Denote the inverse image $\varphi^{-1}(\mathcal{H})$ by H. Then H is a subgroup of G that contains K. If $\mathcal{H}$ is a normal subgroup of $\mathcal{G}$, then H is a normal subgroup of G. If φ is surjective and if H is a normal subgroup of G, then $\mathcal{H}$ is a normal subgroup of $\mathcal{G}$.

Theorem 1.4.15 (4th Isomorphism Theorem (Correspondence Theorem))**.** Let $N \trianglelefteq G$, then $\phi : G \to G/N, \phi(g) = gN$ induces a 1-1 correspondence $\Phi : H \to \phi(H) = H/N$ between subgroups of G which contain N and subgroups of G/N:

$$\begin{aligned} \mathcal{S} = \{\text{subgroups of } G \text{ that contain } N\} &\longleftrightarrow \mathcal{S}' = \{\text{subgroups of } G/N\} \\ \text{a subgroup } H \text{ of } G \text{ that contains } N &\longrightarrow \text{its image } \phi(H) = H/N \text{ in } G/N \\ \text{its inverse image } \phi^{-1}(\mathcal{H}) \text{ in } G &\longleftarrow \text{a subgroup } \mathcal{H} \text{ of } G/N \end{aligned}$$

Moreover, if we denote H/N by H^*, then

- For $H_{1,2} \in \mathcal{S}$, $H_1 \leqslant H_2$ if and only if $H_1^* \leqslant H_2^*$, and then $[H_2 : H_1] = [H_2^* : H_1^*]$;
- For $H_{1,2} \in \mathcal{S}$, $H_1 \trianglelefteq H_2$ if and only if $H_1^* \trianglelefteq H_2^*$, and then $H_2/H_1 \cong H_2^*/H_1^*$.

Remark 1.4.16. For the proof of the above theorem, see [9] Theorem 2.28. Also note that [1] Theorem 2.10.5 relaxes the assumption to surjective homomorphism ϕ while getting less interesting results than the case ϕ being the canonical projection.

1.4 EXERCISES

1. [9][p.31 ex2.29] Prove that a homomorphism $f : G \to H$ is an injection if and only if $Ker(f) = 1$.
2. [9][p.37 ex2.48] (Modular Law). Let A, B, and $\mathbb{C}$ be subgroups of G with $A \leqslant B$. If $A \cap C = B \cap C$ and $AC = BC$ (we do not assume that either AC or BC is a subgroup), then $A = B$.
3. [9][p.37 ex2.49] (Dedekind Law). Let H, K, and L be subgroups of G with $H \leqslant L$. Then $HK \cap L = H(K \cap L)$ (we do not assume that either HK or $H(K \cap L)$ is a subgroup).

1.5 Simple and Solvable Groups

Definition 1.5.1. A group G is called **simple** if it has no normal subgroup other than $\{e\}$ and G.

Example 1.5.2. Cyclic groups G of prime order are simple: $|N| \mid |G| = p \implies |N| = 1$ or $p \implies N = G$ or $N = \{e\}$.

Example 1.5.3. Consider the alternating group A_n. By Question 1.3-6, we see $A_n \trianglelefteq S_n$.

$A_2 = \{e\}$ is simple. $A_3 = \{e, (1\ 2\ 3), (1\ 3\ 2)\}$ is cyclic of prime order 3 and is thus simple (apply previous example). A_4 is *not* simple: $\mathbf{V}$ is normal in A_4 because it is the union of conjugacy classes in A_4 (see Question 1.5-1 and Question 1.3-13).

Theorem 1.5.4. A_n is simple if $n \geqslant 5$

Proof. The proof is made up of the following three facts:

(1) $A_n, n \geqslant 5$ is generated by 3-cycles;

(2) Every two 3-cycles are conjugate with each other in A_n: σ_1, σ_2 are 3-cycles, then $\exists \tau \in A_n : \tau\sigma_1\tau^{-1} = \sigma_2$.;

(3) every normal subgroup $N \neq \{e\}$ in A_n has at least one 3-cycle.

Together they prove the statement: suppose $N \neq \{e\}$, and we want to show $N = A_n$. (3) gives a 3-cycle $\sigma_1 \in N$, so $\forall \tau \in A_n$, $\tau\sigma_1\tau^{-1} = \sigma_2 \in N$ as $N \trianglelefteq A_n$. (2) then implies that all 3-cycles are in N. (1) states that $A_n = \langle \text{3-cycles} \rangle$ is the smallest subgroup of A_n containing all 3-cycles, so $N \trianglelefteq A_n$ has to be equal to A_n.

We prove the three facts:

(1): $T = \{(a\ b\ c) \mid 1 \leqslant a < b < c \leqslant n\} \subset A_n$, then $\langle T \rangle \subset A_n$. If

$$\sigma = (a\ b)(c\ d) = \begin{cases} e, & \text{if } \{a,b\} = \{c,d\} \\ (a\ c\ b)(a\ c\ d), & \text{if } a,b,c,d \text{ all distinct} \\ (a\ d\ b) & \text{if } a = c \end{cases}$$

Then $\sigma \in \langle T \rangle \implies A_n \subseteq T$.

(2) is due to a more general theorem, namely Theorem permutations are conjugate iff they have the same cycle structure.

(3): See Exercise 1.5-2. □

Theorem 1.5.5. Permutations $\alpha, \beta \in S_n$ are conjugate if and only if they have the same cycle structure.

Proof. See [9] Theorem 3.5 or Math5031 HW3 Q4. □

Theorem 1.5.6. Jordan-Holder Theorem. If G is any finite group, then there is a unique tower of subgroups

$$\{e\} = N_0 \trianglelefteq N_1 \trianglelefteq \cdots \trianglelefteq N_{k-1} \trianglelefteq N_k = G$$

such that N_i/N_{i-1} is simple.

Definition 1.5.7. A **tower of subgroups**

$$G_m \leqslant G_{m-1} \leqslant \cdots \leqslant G_1 \leqslant G_0 = G$$

is **subnormal** if $G_{i+1} \trianglelefteq G_i$ and **normal** if furthermore $G_i \trianglelefteq G$ for each i. A subnormal series is called **abelian** if each G_i/G_{i+1} is abelian. A group G is called **solvable** if there is an abelian series

$$\{e\} = G_m \leqslant G_{m-1} \leqslant \cdots \leqslant G_1 \leqslant G_0 = G.$$

Example 1.5.8.

- Any abelian group is solvable.
- S_3 is solvable.
- S_4 is solvable.
- $S_n, n \geqslant 5$ is not solvable.
- D_n is not simple and is solvable.

Proof.

- For an abelian group G, any $N \leqslant G$ is normal and abelian, so $N/\{e\}$ is abelian. The factor group G/N is abelian because the natural homomorphism $\phi : G \to G/N$ is surjective.
- $\{e\} \trianglelefteq A_3 \trianglelefteq S_3$. Question 1.3-6 gives $|A_3| = \frac{1}{2}3! = 3$ which is prime, so $A_3 \cong \mathbb{Z}_3$ is abelian. It is also normal in S_3 with index 2, so $S_3/A_3 \cong \mathbb{Z}_2$ is abelian.

- Solvablility of S_4 is due to $\{e\} \trianglelefteq \mathbf{V} \trianglelefteq A_4 \trianglelefteq S_4$. $A_4 \trianglelefteq S_n$ and S_4/A_4 abelian. $\mathbf{V} \trianglelefteq A_4$ (see example 1.5.3) and $\mathbf{V}/\{e\}$ is abelian.
- Let $N \trianglelefteq S_n$. Since $A_n \trianglelefteq S_n$, by 2nd isomorphism theorem, $N \cap A_n \trianglelefteq A_n$. Since A_n for $n \geqslant 5$ is simple, we see $N \cap A_n = \{e\}$ or A_n.

  If $N \cap A_n = A_n$, then $A_n \leqslant N \leqslant S_n \implies N = A_n$ or $N = S_n$ because Question 1.1-16 implies $2 = [S_n : A_n] = [S_n : N][N : A_n]$.

  If $N \cap A_n = \{e\}$ and if $N \neq \{e\}$, then: $\sigma_1, \sigma_2 \neq e, \sigma_1, \sigma_2 \in N$, then $\sigma_1\sigma_2 \in N$, and $\sigma_1\sigma_2 = e$ because $\sigma_1\sigma_2$ is even (so $\sigma_1\sigma_2$ is also in A_n). Thus $N = \{e, \sigma, \sigma^{-1}\}$ and $\sigma^2 = \sigma^{-1}$. σ has order 3, which by Question 1.2-9 implies that it is a product of 3-cycles. But by parts (1) and (2) of theorem 1.5.4, we see $N = A_n$. Therefore, $N = \{e\}, N$, or $S_n \implies S_n, n \geqslant 5$ is not solvable.
- The index-2 subgroup in Question 1.2-11 is the cyclic subgroup generated by the rotation $\langle r \rangle$ and is thus abelian and is also normal in D_n due to corollary 1.3.4. Then $\{e\} \trianglelefteq \langle r \rangle \trianglelefteq D_n$ is the desired abelian subnormal series as $D_n/\langle r \rangle$ is a group of order 2, isomorphic to $\mathbb{Z}_2$.

□

Definition 1.5.9. Let $x, y \in G$. The **commutator** of $x, y := xyx^{-1}y^{-1} = [x, y]$

Note that $[x, y] = e \iff xy = yx$, and $[x, y]^{-1} = [y, x]$. This gives us a notion of how far a group is from abelian.

Definition 1.5.10. G', the **commutator subgroup**, is the subgroup generated by all the commutators $[x, y]$, where $x, y \in G$. $G' = \{[x_1, y_1][x_2, y_2] \cdots [x_k, y_k] \,|\, x_i, y_i \in G\}$

Proposition 1.5.11.

- $G' = \{e\} \iff G$ is ableian
- $G' \trianglelefteq G$
- G/G' is abelian

Proof. Insert gg^{-1} between the elements: $g[xy]g^{-1} = gxg^{-1}gyg^{-1}gx^{-1}g^{-1}gy^{-1}g^{-1} = [gxg^{-1}, gyg^{-1}] \in G'$.

Similarly, $g[x_1, y_1] \cdots [x_k, y_k]g^{-1} = (g[x_1, y_1]g^{-1}) \cdots (g[x_k y_k]g^{-1})$

G/G' is abelian: we want to show that $abG' = baG'$. $a^{-1}b^{-1}ab = [a^{-1}, b^{-1}] \in G'$. So it is true. □

Proposition 1.5.12. If $N \trianglelefteq G$, then G/N is abelian $\iff G' \leqslant N$

Proof. $\implies$: $\forall a, b \in G, G/N$ abelian so $a^{-1}b^{-1}N = b^{-1}a^{-1}N$. Then $aba^{-1}b^{-1} \in N \implies [a, b] \in N \implies G' \leqslant N$

$\impliedby$: $a^{-1}b^{-1}ab = [a^{-1}, b^{-1}] \in G' \subseteq N \implies a^{-1}b^{-1}ab \in N$ □

Example 1.5.13. $(S_n)' = A_n$. See Question 1.5-3.

Let $G^{(0)} := G, G^{(1)} = G', \cdots, G^{(i)} = (G^{(i-1)})'$. $G^{(i+1)} \trianglelefteq G^{(i)}$ and $G^{(i+1)}/G^{(i)}$ is abelian.

Proposition 1.5.14. G is solvable iff $G^{(m)} = \{e\}$ for some $m \geqslant 1$.

Proof. $\impliedby$: $\{e\} = G^{(m)} \trianglelefteq \cdots \trianglelefteq G^{(1)} \trianglelefteq G$ is an abelian tower.

$\implies$: If $\{e\} = G_m \trianglelefteq \cdots \trianglelefteq G_1 \trianglelefteq G_0 = G$ is abelian, then $G_1 \trianglelefteq G_0, G_0/G_1$ abelian $\implies G' \leqslant G_1, G_2 \trianglelefteq G_1, G_1/G_2$ abelian $\implies (G_1)' \leqslant G_2$ implies together that $G^{(2)} \leqslant G_1' \leqslant G_2 \implies G^{(2)} \leqslant G_2$.

By induction, $G^{(i)} \leqslant G_i \forall i, G^{(m)} \leqslant G_m = \{e\}$. □

The following proposition is a good exercise (Math5031 HW2 Q4) for one to review all the isomorphism theorems and various normality theorems.

Proposition 1.5.15. If $N \trianglelefteq G$, then $N, G/N$ are solvable $\iff$ G is solvable.

Proof. G **solvable** $\implies$ N **solvable**:
Let

$$\{e\} = G_m \trianglelefteq G_{m-1} \trianglelefteq \cdots \trianglelefteq G_0 = G$$

be a subnormal series where G_i/G_{i+1} is abelian. Let $N_i = N \cap G_i$. We claim that

$$\{e\} = N \cap \{e\} = N_m \trianglelefteq N_{m-1} \trianglelefteq \cdots N_0 = N \cap G = N$$

is the desired subnormal series where N_i/N_{i+1} is abelian.

We apply Question 1.3-11 three times: $G_i \leqslant G, N \trianglelefteq G \Rightarrow N_i = G_i \cap N \trianglelefteq G_i$ and $G_{i+1} \trianglelefteq G_i \Rightarrow N_i \cap G_{i+1} = N_{i+1} \trianglelefteq G_{i+1}$. Similarly, $N_i \trianglelefteq G_i$ with the third application to $N_i \trianglelefteq G_i, N_{i+1} \trianglelefteq G_{i+1}$, which implies $N_i \cap N_{i+1} = N_{i+1} \trianglelefteq N_i$.

Applying Remark 1.4.8 (2) with homomorphism the inclusion of N_i in G_i, $f = \iota : N_i \hookrightarrow G_i$, $N' = G_{i+1}$, and $N = \iota^{-1}(G_{i+1}) = N_i \cap G_{i+1} = N_{i+1}$, we obtain an injective homomorphism $g_* : N_i/N_{i+1} \to G_i/G_{i+1}$. Thus G_i/G_{i+1} being abelian implies N_i/N_{i+1} being abelian (note that injectivity is necessary for this implication: $\varphi(xy) = \varphi(x)\varphi(y) \overset{\text{abelian codomain}}{=\!=\!=} \varphi(y)\varphi(x) = \varphi(yx) \overset{injectivity}{\implies} xy = yx$).

G **solvable** $\implies$ G/N **solvable**:
Let

$$\{e\} = G_m \trianglelefteq G_{m-1} \trianglelefteq \cdots \trianglelefteq G_0 = G$$

be a normal series where each G_i/G_{i+1} is abelian. Let $H_i = NG_i/N$. Proposition 1.4.6 implies that $NG_{i+1} = G_{i+1}N, NG_i = G_iN$. Notice that $N \subseteq G_{i+1}N \trianglelefteq G_iN$ due to Question 1.3-5. Since $N \subseteq G_{i+1}N \trianglelefteq G_iN, N \trianglelefteq G_iN$, the 3rd isomorphism theorem states that

$$H_{i+1} = \frac{NG_{i+1}}{N} \trianglelefteq \frac{NG_i}{N} = H_i$$

The remaining is to show $\frac{H_i}{H_{i+1}}$ is abelian: first observe that

$$(*): \quad G_iN = G_i\left(G_{i+1}N\right)$$

and then

$$\frac{H_i}{H_{i+1}} = \frac{\frac{NG_i}{N}}{\frac{NG_{i+1}}{N}} \overset{3rd\text{ iso}}{\cong} \frac{G_iN}{G_{i+1}N} \overset{(*)}{=} \frac{G_i\left(G_{i+1}N\right)}{G_{i+1}N} \overset{2nd\text{ iso}}{\cong} \frac{G_i}{G_i \cap G_{i+1}N}$$

where each of the isomorphism theorem's conditions are satisfied (the only nontrivial relationship is $G_{i+1}N \trianglelefteq G_iN$ and is proved above).

$$3^{\text{rd}} : N \subseteq G_{i+1}N \trianglelefteq G_iN, N \trianglelefteq G_iN.$$
$$2^{\text{nd}} : G_i \leqslant G_iN, G_{i+1}N \trianglelefteq G_iN.$$

By enlargement of coset map and $G_{i+1} \subseteq G_i \Rightarrow G_{i+1} \subseteq G_i \cap G_{i+1}N$, we see $\frac{G_i}{G_i \cap G_{i+1}N}$ is isomorphic to a quotient of $\frac{G_i}{G_{i+1}}$, which is abelian, so $\frac{G_i}{G_i \cap G_{i+1}N}$ is abelian (quotient of abelian group is abelian because the canonical projection is a surjective homomorphism).

G/N **solvable and** N **solvable** $\implies$ G **solvable**:
N and G/N are solvable $\Rightarrow$ G is solvable. Suppose

$$\{e\} = N_m \trianglelefteq N_{m-1} \trianglelefteq \cdots \trianglelefteq N_0 = N$$
$$\left\{e_{G/N}\right\} = H_n \trianglelefteq H_{n-1} \trianglelefteq \cdots \trianglelefteq H_0 = \frac{G}{N}$$

Then by 4th isomorphism theorem, for each H_i which is a subgroup of $\frac{G}{N}$, we can find a unique subgroup K_i of G containing N such that $\frac{K_i}{N} = H_i$. Then

$$\{e\} = N_m \trianglelefteq N_{m-1} \trianglelefteq \cdots \trianglelefteq N_0 = N = K_n \trianglelefteq K_{n-1} \trianglelefteq \cdots \trianglelefteq K_0 = G$$

The fact $K_{i+1} \trianglelefteq K_i$ is from properties of the 1 -1 correspondence $\Phi : \{K : A \subseteq K \leqslant G\} \leftrightarrow \left\{\bar{A} = \frac{A}{N} : \frac{A}{N} \leqslant \frac{G}{N}\right\}$. Recall that $A \subseteq B \Leftrightarrow \bar{A} \subseteq \bar{B}$ and $A \trianglelefteq G \Leftrightarrow \bar{A} \trianglelefteq \bar{G}$ where A and B are two subgroups containing N. By the two properties we see

$$K_n \subseteq K_{n-1} \subseteq \cdots \subseteq K_0$$
$$\forall i : K_i \trianglelefteq G$$

Also note that $p \geqslant q \Rightarrow K_p \leqslant K_q$. That's because $K_p \subseteq K_q$ and $K_p \leqslant G$. Thus for each $i = 1$, $K_2 \leqslant G, K_2 \subseteq K_1 \trianglelefteq K_0 = G \Rightarrow K_2 = K_2 \cap K_1 \trianglelefteq K_1$. We set induction hypothesis that $K_{i+1} \trianglelefteq K_i$ then have $K_{i+2} \leqslant K_i, K_{i+2} \subseteq K_{i+1} \trianglelefteq K_i \Rightarrow K_{i+2} = K_{i+2} \cap K_{i+1} \trianglelefteq K_{i+1}$. The induction establishes the series as normal. We now show that K_i/K_{i+1} is abelian due to the third isomorphism theorem (conditions are satisfied: $N = K_0 \subseteq K_{i+1} \trianglelefteq K_i, N = K_0 \trianglelefteq K_i$):

$$\frac{K_i}{K_{i+1}} \cong \frac{\frac{K_i}{N}}{\frac{K_{i+1}}{N}} = \frac{H_i}{H_{i+1}}$$

Therefore, G is also solvable. □

Remark 1.5.16. The proof of a more general nature can be seen in [6] 6.1.1 and 6.1.2, but need an equivalence proof (6.1.5) of their first definition of solvability and the definition we used in class (or used by Serge Lang). 6.1.1 shows that subgroups and homomorphic images of solvable groups are solvable, which implies the $\Rightarrow$ direction of the above statement, because N is normal subgroup of G and G/N is the homomorphic image of the map $\psi : G \to \frac{G}{N}; x \mapsto xN$.

1.5 EXERCISES

1. If G is a group, by a conjugacy class of G we mean all elements of G which are conjugate to a fixed element (so it is an orbit of G for the action of G on G by conjugation).

i. Find all conjugacy classes of A_4.
ii. Show that if $[G : Z(G)] = n$, then every conjugacy class has at most n elements.

2. Use the following steps to show every normal subgroup $N \neq \{e\}$ of $A_n, n \geqslant 5$, contains a 3-cycle. This finishes the proof of the fact that A_n is simple if $n \geqslant 5$.

i. Show that if N contains a permutation of the form $\sigma = (1\ 2\ \cdots\ r)\mu$ (where μ is a product of cycles disjoint from $\{1, 2, \ldots, r\}$) with $r \geqslant 4$, then N contains a 3-cycle by letting $\rho = (1\ 2\ 3)$ and computing $\sigma^{-1}\rho^{-1}\sigma\rho$.
ii. Show that if N contains a permutation of the form $\sigma = (1\ 2\ 3)(4\ 5\ 6)\mu$ (where μ is a product of cycles disjoint from $\{1, 2, \ldots, 6\}$), then N contains a 3-cycle by letting $\rho = (1\ 2\ 4)$ and computing $\sigma^{-1}\rho^{-1}\sigma\rho$.
iii. Show that if N contains a permutation of the form $\sigma = (1\ 2\ 3)\mu$, where μ is a product of 2-cycles a product of 2-cycles which are mutually disjoint and are also disjoint form $\{1, 2, 3\}$, then N contains a 3-cycle by computing σ^2.
iv. Show that if N contains a permutation of the form $\sigma = (1\ 2)(3\ 4)\mu$, where μ is a product of 2-cycles which are mutually disjoint and are also disjoint from $\{1, 2, 3, 4\}$, then N contains a 3-cycle by letting $\rho = (1\ 2\ 3)$, computing $\eta = \sigma^{-1}\rho^{-1}\sigma\rho$ and $\zeta = (1\ 5\ 2)\eta(1\ 2\ 5)$.

Remark: This problem divides into three subcases: (1) the cycle has length $\geqslant 4$ (corresponded to **i**); (2) the cycle has length $\leqslant 3$ (but with at least on of them being 3) (corresponded to **ii** and **iii**); (3) the cycle has length $\leqslant 2$ (corresponded to **iv**). WLOG, each case can be converted to the considerations of the explicit forms given in the above problem.

3. The commutator subgroup of S_n is A_n (Hint: show that every 3-cycle is a commutator, and use the fact that A_n is generated by 2-cycles.)

4. (A simple group of infinite order) Let A_∞ be defined in the following way: identify A_{n-1} with the subgroup of A_n consisting of those permutations which fixes n, and let A_∞ be the union $\bigcup_{n\geqslant 1} A_n$.

i. Show that A_∞ is a group.
ii. Prove A_∞ is a simple group.

1.6 Group Actions

Definition 1.6.1. Let G be a group and X be a set, an **action of G on X** is a function $\alpha : G \times X \to X, (g, x) \mapsto g \cdot x$ such that

- $e \cdot x = x$, $\forall x \in X$.
- $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$, $\forall x_1, x_2 \in X, g \in G$

Note that $\forall g \in X,\ \phi_g : X \to X,\ x \mapsto g \cdot x$ is a permutation. ϕ_g is bijective, as $g \cdot x = g \cdot x' \implies g^{-1} \cdot (g \cdot x) = g^{-1} \cdot (g \cdot x') \implies e \cdot x = e \cdot x'$. Besides, $\forall x \in X, \phi_g(g^{-1} \cdot x) = g \cdot (g^{-1} \cdot x) = x$.

A group action $G \curvearrowright X$ gives rise to a homomorphism $\phi : G \to S_X, g \mapsto \phi_g$ (not necessarily injective): $\phi_{g_1 g_2}(x) = (g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x) = \phi_{g_1} \circ \phi_{g_2}(x)$.

Example 1.6.2.

1. Trivial action. $\forall g \in G, x \in X, g \cdot x = x$.
2. Conjugation on elements of G. $X = G, g \cdot x = gxg^{-1}$.
3. Conjugation on subgroups of G. Let X be set of subgroups of G, $g \in G, H \in X$. Then $g \cdot H = gHg^{-1} \leqslant G$ (for $a, b \in gHg^{-1}$, $a = ghg^{-1}, b = gh'g^{-1} \implies ab = g(hh')g^{-1}$.)
4. Translation on elements of G. $X = G, g \cdot x = gx$.

Theorem 1.6.3 (Cayley's Theorem)**.** Every group is isomorphic to a permutation group.

Proof. Let the set X be G with action by translation (see example 1.6.2). The the homomorphism we constructed above

$$\begin{aligned} \phi : G &\to S_X \\ g &\mapsto \phi_g \end{aligned}$$

gives an isomorphism by restricting S_X to $\mathrm{Im}(\phi)$: it is automatically surjective. Injectivity is becasue:

$$\phi_g = \phi_h \iff \forall x \in G,\ \phi_g(x) = \phi_h(x) \iff gx = hx \iff g = h$$

where the last step is due to cancellation law of the group. □

Definition 1.6.4. Suppose G acts on $X, x \in X$. Then the **stabilizer** is defined as

$$G_x := \{g \in G \mid gx = x\}$$

It is a subgroup of G because

- $e \in G_x$;
- $g \in G_x$ then $g \cdot x = x \Rightarrow x = g^{-1} \cdot (g \cdot x) = g^{-1} \cdot x \Rightarrow g^{-1} \in G_x$.
- $g, g' \in G_x \Rightarrow (gg') \cdot x = g(g'x) = gx = x$.

Definition 1.6.5. We also define an **orbit** of X.

$$O_x = \{gx \mid g \in G\} \subseteq X$$

Note: $x \sim y$ if $y \in O_x$, so $y = gx$ for some g. Thus, any two orbits are either equal or disjoint, and they form a partition of X.

Example 1.6.6. For Example 1.6.2 above, the stabilizer and orbit are

1. Trivial action. $O_x = \{x\}$. $G_x = G$.
2. Conjugation on elements of G. $O_x = \{gxg^{-1} \mid g \in G\}$, the **conjugacy class** of x in G. $G_x = \{g \in G \mid gx = xg\} = N(x) \leqslant G$, the **normalizer** of x.
3. Conjugation on subgroups of G. O_H = all subgroups conjugate to H, $G_H = \{g \in G \mid gHg^{-1} = H\} = \{g \in G \mid gH = Hg\} = N_G(H)$, the **normalizer** of H in G. Note that $H \trianglelefteq N_G(H) \leqslant G$ and is the largest subgroup of G in which H is normal. Also, $H \trianglelefteq G \iff N_G(H) = G$
4. Translation on elements of G. $O_x = \{gx \mid g \in G\} = G$. $G_x = \{g \in G \mid gx = x\} = \{e\}$

Remark 1.6.7. For a subset S of group G, one can define its **centralizer** as

$$C_G(S) = \{g \in G \mid \forall s \in S,\ gs = sg\}$$

and its **normalizer** as

$$N_G(S) = \{g \in G \mid gS = Sg\}.$$

We note that the condition in the normalizer is weaker, so $C_G(S) \subseteq N_G(S)$. If $S = \{x\}$ is a singleton, then the two definitions give the same set, as in Example 1.6.6 (2).

The proof of the following lemma is straightforward:

Lemma 1.6.8. Let N be a normal subgroup of G. Then

1. If N contains an element x, then it contains the conjugacy class $C(x)$ of x.
2. N is a union of conjugacy classes.
3. The order of N is the sum of the orders of the conjugacy classes that it contains.

Definition 1.6.9. For group G, the **center** of G, $Z(G)$, is the set of elements in G commuting with all elements in G:

$$Z(G) = \{g \in G | \forall g' \in G,\ gg' = g'g\}$$

That is, $Z(G) = C_G(G)$.

Proposition 1.6.10.

- Observe that $S_1 \subseteq S_2 \implies C_G(S_2) \subseteq C_G(S_1)$, so $\forall S \subseteq G,\ Z(G) = C_G(G) \subseteq C_G(S)$. In particular, $Z(G) \subseteq C_G(\{x\}) = N_G(\{x\}) = N(x)$ for an element $x \in G$.

- $Z(G) = G \iff G$ abelian
- $Z(G) \trianglelefteq G$

Proof. The first and second statement are trivial.

$Z(G) \leqslant G$: $e \in Z(G)$. $g \in Z(G) \Rightarrow g^{-1} \in Z(G)$ as $g'g^{-1} = g^{-1}g'$, and if $g_1, g_2 \in Z(G)$ then $g_1g_2g' = g_1g'g_2 = g'(g_1g_2)$ so $g_1g_2 \in Z(G)$.

$Z(G) \trianglelefteq G$: let $g \in Z(G)$ and $h \in G$. We want to show that $hgh^{-1} \in Z(G)$. $\underline{hgh^{-1}}g' = \underline{hh^{-1}}gg' = gg'$ but $g'\underline{hg}h^{-1} = g'\underline{gh}h^{-1} = g'g$. Since $hgh^{-1}g' = g'hgh^{-1}$ we see $gg' = g'g$. □

Example 1.6.11. $Z(S_n) = \{e\}, n \geqslant 3$. This is a nontrivial fact. $Z(A_n) = \{e\}, n \geqslant 4$. That's because for $n \geqslant 5$, A_n is simple but $Z(A_n) \trianglelefteq Z(A_n) = \{e\}$ or $Z(A_n) = A_n$. For $n = 4$, find an element not commuting with any element in the Klein-four group $\mathbf{V}$.

Theorem 1.6.12 (Orbit-Stabilizer Theorem)**.** Let X be a G-set, then $\forall x \in X$,

$$|O_x| = [G : G_x], \text{ or } |G| = |O_x||G_x|$$

where we note that $G_x \leqslant G$ as we showed when defining it.

Proof. For the point x, we define

$$\begin{aligned} \phi : O_x = \{gx | g \in G\} &\to \{\text{all left cosets of } G_x\} \\ gx &\mapsto gG_x \end{aligned}$$

Injective: $gG_x = g'G_x \iff g^{-1}g' \in G_x = \{g \in G \mid gx = x\} \iff g^{-1}g'x = x \iff gx = g'x$.

Surjective: clear.

Therefore, $[G : G_x] = \{\text{all left cosets of } G_x\}| = |O_x|$ □

Example 1.6.13. If G acts on the set X of its subgroups, $\{H \mid H \leqslant G\}$, then by example 1.6.6, we have O_H = the set of all subgroups conjugate to H and $G_H = N_G(H)$. Orbit-stabilizer theorem then says $|O_H| = [G : N_G(H)]$. Also notice that $|H|$ divides $|N_G(H)|$, and $|N_G(H)|$ divides $|G|$.

Lemma 1.6.14. An observation: an element x of group G is in the center if and only if its centralizer $C_G(x)$ is the whole group G, and this happens if and only if the conjugacy class $C(x)$ consists of the element x alone. In symbols,

$$x \in Z(G) \iff C_G(x) = G \iff C(x) = x$$

Example 1.6.15. Class Formula is obtained by letting G acts on G via conjugation. If $x \in X = G$, by example 1.6.6, we have stabilizer $G_x = N(x)$ and orbit $O_x = C(x)$. Since orbits O_x give a partition of $X = G$, we see $|G| = \sum_{\text{distinct orbits}} |O_x| \overset{\text{orb-stab thm}}{=\!=\!=} \sum_{\text{distinct orbits}} [G : G_x]$. Also, due to Lemma 1.6.14, we can write that summing all distinct conjugacy classes with more than 1 element:

$$|G| = Z(G) + \underbrace{|C_1| + \cdots + |C_k|}_{\text{distinct conj classes with size} > 1} \tag{1.4}$$

Corollary 1.6.16. If $|G| = p^r$, p prime, then $Z(G) \neq \{e\}$.

Proof. By equation (1.4), we see, if $Z(G) = \{e\}$, we get

$$p^r = 1 + \underbrace{|C_1| + \cdots + |C_k|}_{\text{distinct conj classes with size}>1}.$$

Each $|C_i| = |G|/|G_x|$ is a divisor of $|G| = p^r$, i.e., powers of p, but excluding $p^0 = 1$ since the size of conjugacy classes in above summation is greater than 1. This implies that

$$p^r - \text{sum of some multiples of } p \text{ greater than } 1 = 1,$$

so $p \,|\, 1$, a contradiction. Thus, $Z(G) \neq \{e\}$. □

Corollary 1.6.17. If $|G| = p^r$, then G is not simple.

Proof. The center $Z(G)$ is a nontrivial normal subgroup by corollary 1.6.16 and proposition 1.6.10. □

Corollary 1.6.18. If $|G| = p^2$, then G is ableian.

Proof. If G is not abelian, then $|Z(G)| = p$, so $Z(G)$ is proper subgroup of G. Pick $a \in G - Z(G)$, then $N(a) = \{b \,|\, ab = ba\} \neq G$. However $Z(G)$ is proper subgroup of $N(a)$ and $N(a)$ proper subgroup of G, a contradiction (a in $N(a)$ but not in $Z(G)$).

[1] 7.3.4 claims that G with $|G| = p^2$ is either cyclic or a product of two cyclic groups of order p. □

Corollary 1.6.19. If $|G| = p^r$, then G is solvable.

Proof. Proof by induction on r, $r = 1$ true.

Suppose this holds for $1, ..., r-1$. Consider $Z(G) \trianglelefteq G$ and $Z(G) \neq \{e\}$. Here $|Z(G)|$ and $|G/Z(G)|$ are powers of p. So by hypothesis, $Z(G)$ and $G/Z(G)$ are solvable $\implies$ G also solvable. □

Definition 1.6.20. An action $G \curvearrowright X$ is **transitive** if there is only one orbit, $O_x = X$. Equivalently, $\forall x, y \in X$, $\exists g \in G$ s.t. $g \cdot x = y$.

Definition 1.6.21. An action $G \curvearrowright X$ is **faithful** or **effective** if there is only the identity $e \in G$ that fixes all $x \in X$ (i.e. $\forall x \in X,\ g \cdot x = x$ implies $g = e$). This is equivalent of saying that the homomorphism $\phi : G \to S_X; g \to \phi_g$ is injective or that ϕ is a monomorphism. If X_1 and X_2 are left G-spaces, a mapping $f : X_1 \to X_2$ is called G-**equivariant**, or simply a mapping of left G-spaces, in case

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

for any $g \in G$ and $x \in X_1$. A G-equivariant map $f : X_1 \to X_2$ is called **isomorphism** of left G-spaces in case there exists another G-equivariant map $f' : X_2 \to X_1$ such that $f'f = \mathrm{id}_{X_1}$ and $ff' = \mathrm{id}_{X_2}$. This is equivalent to the condition that f be one-to-one and onto. This definition of isomorphism is the natural one in this context. The reader should note that it is sometimes possible for a group G to operate in several different, nonisomorphic ways on a given set E. As usual, an automorphism of a G-space is a self-isomorphism.

Theorem 1.6.22 (Burnside's Lemma)**.** If G, X finite, X is a G-set, then the number of orbits of the action $G \curvearrowright X$ is $\frac{1}{|G|}\sum_{g \in G} |F_g|$, where F_g is the set of elements of X fixed by g.

Proof. Consider $S = \{(g, x) \,|\, gx = x\} \subset G \times X$. We can count S in two different ways.

1. $\forall g \in G$, there are $|F_g|$ elements fixed by g so $|S| = \sum_{g \in G} |F_g|$.
2. $\forall x \in X$, there are $|G_x|$ elements fixed in x, which equals $|G|/[O_x]$ by the orbit-stabilizer theorem.

So

$$\begin{aligned}\sum_{g\in G}|F_g| &= \sum_{x\in X}\frac{|G|}{|O_x|}\\ &= |G|\left(\underbrace{\frac{1}{|O_{x_1}|}+\cdots+\frac{1}{|O_{x_k}|}}_{\text{the same}}+\cdots\right)\\ &= |G|\sum_{\text{distinct orbits } O_{y_1},O_{y_2}\cdots,}\frac{1}{|O_{y_i}|}|O_{y_i}|\\ &= |G|\times \text{ num distinct orbits}\end{aligned}$$

where for the third equality we notice that $O_x = O_y$ exactly when x and y are both in the same orbit. Thus when going through all X, those in the same orbit will have the same $1/|O_x|$ and there are in total $|O_x|$ of them having this same $1/|O_x|$. □

Corollary 1.6.23. If G acts transitively on X, and $|X| > 1$, then there is $g \in G$ such that $F_g = \varnothing$.

Proof. Burnside's Lemma gives $|G| = \sum_{g\in G}|F_g| = F_e + \sum_{g\neq e}|F_g|$.

If $\forall g,\ |F_g| \geqslant 1$, then $|G| = |X| + \sum_{g\neq e}|F_g| \geqslant |X| + (|G|-1) \implies |X| \leqslant 1$, a contradiction. □

1.6 EXERCISES

1. [9][p.45 ex3.5] Prove that $Z(G_1 \times \cdots \times G_n) = Z(G_1) \times \cdots \times Z(G_n)$.

2. [9][p.45 ex3.6]

i. Prove, for every $a, x \in G$, that $C_G\left(axa^{-1}\right) = aC_G(x)a^{-1}$.
ii. Prove that if $H \leqslant G$ and $h \in H$, then $C_H(h) = C_G(h) \cap H$.

3. [9][p.45 ex3.9]

i. Prove that $N_G\left(aHa^{-1}\right) = aN_G(H)a^{-1}$.
ii. If $H \leqslant K \leqslant G$, then $N_K(H) = N_G(H) \cap K$.
iii. If $H, K \leqslant G$, prove that $N_G(H) \cap N_G(K) \leqslant N_G(H \cap K)$. Give an example in which the inclusion is proper.

1.7 Sylow Theorems

Definition 1.7.1. A group G is a **p-group** if $|G| = p^r$. Since $\operatorname{ord}(a) \,\big|\, p^r$, we see $\forall e \neq a \in G$, a is some multiple of p that is not 1, so $p \,\big|\, \operatorname{ord}(a)$. And if $|G| = p^r m, gcd(m, p) = 1$, $H \leqslant G$, then H is a called a **p-subgroup** if $|H| = p^s$, and H is a **Sylow p-subgroup** if $|H| = p^r$.

Using number of elements to define a subgroup need to be justified by an existence proof, because usually we define subgroup by some form like $\{g \in G \mid p(g)\}$ where $p(\cdot)$ is a statement. This existence proof is the content of the first Sylow theorem.

Theorem 1.7.2 (First Sylow theorem)**.** Suppose $|G| = p^r m$, $r \geqslant 1$, $\gcd(p, m) = 1$. Then G has a subgroup of size p^s for any $0 \leqslant s \leqslant r$.

Lemma 1.7.3. If G is abelian and $p \mid |G|$, then G has an element of order p and thus a subgroup of order p.

Proof. Induction on order of G. If $|G| = p$, there is nothing to prove. Suppose $|G| > p$, Let $e \neq a \in G, t = ord(a)$. Then $H = \langle a \rangle = \{e, a, \cdots, a^{t-1}\} \leqslant G$, so $p^r m = |G| = |H|[G : H] = t \cdot k$. There are two cases:

1. If $p \mid t$, then $\left|\left\langle a^{\frac{t}{p}} \right\rangle\right| = p$.
2. Otherwise, let $n = |G|, n = tn'$ so $p \mid n' = |G/H| < n$. So, by induction hypothesis, G/H has subgroup of order p, so has an element $\bar{b}$ of order p. Consider the canonical projection $\phi : G \to G/H$, so if $\phi(b) = \bar{b}$, then $p \mid ord(b)$. So we can apply case 1 to b and get a subgroup of order p due to the following remark.

□

Remark 1.7.4. If $\phi : G \to G'$ is a group homomorphism and $g \in G$ and $ord(\phi(g)) \mid \underbrace{ord(g)}_{m}$, so $g^m = e \to \phi(g)^m = e$. $(a^k = e \implies ord(a) \mid k)$

Proof of theorem. Recall that class formula states that when G acts on G by conjugation, $|G| = |Z(G)| + \sum[G : G_x]$, summing over distinct orbits with more than 1 element.

Fix p induction on G. If $|G| = p$, we are done. Now, let's have two cases where (1) $p \mid |Z(G)|$ and (2) p doesn't divide $|Z(G)|$.

In case 1, by lemma, $Z(G)$ has subgroup H of order p. Since $H \leqslant Z(G)$ and $Z(G) \trianglelefteq G$, we get $H \trianglelefteq G$ so G/H is a group of size $p^{r-1}m$. So by induction hypothesis G/H has a subgroup of order s for all $0 \leqslant s \leqslant r-1$. Any subgroup of G/H is K/H for $H \leqslant K \leqslant G$. So $|H| = p, |K/H| = p^s \implies |K| = p^{s+1}$. So this holds for $1 \leqslant s+1 \leqslant r$.

In case 2, G is not abelian, and we make two subcases.

1. Suppose $\forall x \notin Z(G), p \mid [G : G_x]$. This case is not possible since $p \mid |G|$ and p doesn't divide $Z(G)$
2. $\exists x \in Z(G), p \nmid [G : G_x] = |G|/|G_x| \implies p^r \mid |G_x|$, and $|G_x| < |G|$. By induction hypothesis, G_x and therefore G has a subgroup of $p^s, 0 \leqslant s \leqslant r$.

□

Theorem 1.7.5 (Second Sylow theorem)**.** If $p \mid |G|$, then

1. Every p subgroup is contained in a Sylow p-subgroup.
2. Any two Sylow p-subgroups are conjugate.

Proof. Assuming proposition 1.7.6, we can show the two claims.
Part 1: $|gPg^{-1}| = |P|$ (this is because $gPg^{-1} \to P;\ k \mapsto g^{-1}kg$ gives an inverse of the map $P \to gPg^{-1};\ k \mapsto gkg^{-1}$), so the conjugate is also a Sylow p-subgroup.

Part 2: P, P' Sylow p-subgroups, then $\exists g$ s.t. $P' \subseteq gPg^{-1}$. Then $|gPg^{-1}| = |P| = p^r$ and $|P'| = r \implies P' = gPg^{-1}$. □

Proposition 1.7.6. If H is a p-subgroup and P is a Sylow p-subgroup, then H is contained in a conjugate of P: $\exists g \in G, H \leqslant gP^{-1}g$

Proof of the proposition. Let S be the set of conjugates of P and H acts on S by conjugation, so that $h \cdot gPg^{-1} := hgPg^{-1}h^{-1}$. Then $S = \sum_{\text{distinct orbits}} |O_s| =$ number of fixed points $+ \sum_{\text{distinct w/ size}>1} |O_s|$.

Now the goal is to show that there $\exists$ a fixed point. Since $|O_s| = [H : H_s]$ and $|H| = p^s$, then $p \,\big|\, |O_s|$.

Here, $|S| = [G : N_G(P)] \implies |S| = \frac{|G|}{|N_G(P)|}$. Since $P \trianglelefteq N_G(P) \leqslant G$ and $p^r \,\big|\, |N_G(P)|$, I get $p \nmid |S|$ and so $p^r \,\big|\, |N_G(P)|$.

Let gPg^{-1} be a fixed point. Then $\forall h \in H, hgPg^{-1}h^{-1} = gPg^{-1} \implies P = g^{-1}h^{-1}gPg^{-1}hg \implies P = g^{-1}h^{-1}gP(g^{-1}h^{-1}g)^{-1} \implies g^{-1}h^{-1}g \in N_G(P)$. So $\forall h \in H \implies g^{-1}Hg \subseteq N_G(P)$.

Let $K = g^{-1}Hg$, $K, P \leqslant N_G(P)$ and $P \trianglelefteq N_G(P)$.

So by the second isomorphism theorem, $KP/P \simeq K/K \cap P \implies |KP| = \frac{|P||K|}{|K \cap P|}$ and $|KP| \,\big|\, |G|$, and $|P||K|$ is a power of $p \implies \frac{|K|}{|K \cap P|} = 1 \implies K \subseteq P \implies g^{-1}Hg \subseteq P \implies H \subseteq gPg^{-1}$. □

Theorem 1.7.7 (Third Sylow theorem)**.** Suppose $|G| = p^r m$ and $\gcd(p, m) = 1$. If $s =$ number of p-Sylow subgroups, then $s \,\big|\, m$ and $s \equiv 1 (\text{mod } p)$.

Proof. By part 2 of the second Sylow theorem, $s =$ number of all conjugates of $P = [G : N_G(P)]$, and $[G : N_G(P)] \,\big|\, |G|$.

To show $s \equiv 1 (\text{mod } p)$, let $H = P$ from proof of the proposition, so that $s =$ number of fixed points + a multiple of p

If gPg^{-1} is a fixed point, then by the proof $P \subseteq gPg^{-1}$, but $|P| = |gPg^{-1}|$ so $P = gPg^{-1}$. So only one fixed point $\implies s \equiv 1 (\text{mod } p)$. □

Corollary 1.7.8. As a corollary of second Sylow theorem, we see a group G has only one Sylow p-subgroup H if and only if that subgroup is normal. In symbols, $s = 1 \iff \forall g \in G,\ gPg^{-1} = P \iff P \trianglelefteq G$.

Corollary 1.7.9. If $|G| = pq$ where p, q are distinct primes and $p \not\equiv 1 (\text{mod } q)$ and $q \not\equiv 1 (\text{mod } p)$. Then G is cyclic.

Proof. Let r_1 be the number of Sylow p-subgroups and r_2 be the number of Sylow q-subgroups. Then $r_1 \,\big|\, pq, r_1 \equiv 1 \mod p \implies r_1 = 1$, and similarly $r_2 = 1$

If $H_1, H_2 \leqslant G$ with $|H_1| = p$ and $|H_2| = q$, then by the note, $H_1, H_2 \trianglelefteq G$.

$H_1 = \{e, a, ..., a^{p-1}\} = \langle a \rangle, H_2 = \{e, b, ..., b^{q-1}\} = \langle b \rangle$. For $aba^{-1} \in H_2$ and $ba^{-1}b^{-1} \in H_1$, $aba^{-1}b^{-1} \in H_1 \cap H_2 = \{e\} \implies ab = ba \implies ord(ab) \in \{1, p, q, pq\}$. So $(ab)^p = a^p b^p = b^p \neq e \implies ord(ab) = pq \implies G = \langle ab \rangle$. □

Example 1.7.10. $|G| = 33 = 3 \times 11$. $3 - 1 = 2$ is relatively prime with 11; $11 - 1 = 10$ is relatively prime with 3. Therefore, $G \cong \mathbb{Z}_{33}$ due to above corollary.

Several observations in summary:

1. Any abelian group is solvable.
2. group with prime order is cyclic, abelian, and thus solvable.
3. group with prime order is simple (see Example 1.5.2).
4. A simple group is solvable iff it is abelian.

Our goal is to show the following theorem:

Theorem 1.7.11. Any group of order < 60 is solvable (note that $|A_5| = 60$).

Our plan:

(1) G prime order $\overset{obs(2)}{\implies}$ we're done.

(2) G not prime order. We want to find a nontrivial $N \trianglelefteq G$ (which also gets us *non*-simplicity) such that $N, G/N$ are solvable, which then implies that G is solvable due to Proposition 1.5.15.

Proposition 1.7.12. If $|G| = n$ and p is the smallest prime divisor of n and $H \leqslant G$ has index p, then $H \trianglelefteq G$.

Proof. If $p = 2$, this is proved before ($[G : H] = 2$ is the smallest prime and index-2 subgroup is normal).

Suppose $H \not\trianglelefteq G$. Then there is $g \in G$ s.t. $gHg^{-1} \neq H$. Let $K = gHg^{-1} \leqslant G$.

By product formula, $|HK| = |H|\frac{|K|}{|H \cap K|}$, where the latter fraction is an integer which divides $|K| = |gHg^{-1}| = |H| = p$ and so divides $|G| = pm$. Then either $\frac{|K|}{|H\cap K|} = 1$ or $\frac{|K|}{|H\cap K|} = p$.

For the first case, $H \cap K = K \implies K \subseteq H \implies gHg^{-1} \subseteq H \implies gHg^{-1} = H$, not true.

For second case, $|HK| = p|H| = |G| \implies HK = G \implies g^{-1} \in HK = HgHg^{-1}$. So for some $h, h' \in H, hgh' = e \implies g = h^{-1}h'^{-1} \in H \implies gHg^{-1} = H$, a contradiction. So $H \trianglelefteq G$. □

Corollary 1.7.13. If $|G| = pq^r$, and p, q are distinct prime and $p < q$. Then G has a nontrivial normal subgroup.

Proof. By First Sylow theorem, there is a Sylow q-subgroup H, so $[G : H] = p$. H is normal from the previous corollary. □

Corollary 1.7.14. If $|G| = pq, p \neq q$, then G has a non-trivial normal subgroup.

Proposition 1.7.15. If $|G| = pq^2$, and p, q are distinct prime, then G has a non-trivial normal subgroup.

Proof. If $p < q$, we are done by previous corollary.

So if $p > q$, let r be the number of Sylow p-subgroups and s be number of Sylow q subgroups.

Goal is to show that $r = 1$ or $s = 1$ since the only Sylow subgroup is normal (corollary 1.7.8).

Since $r \equiv 1 \mod p, r \,|\, |G| = pq^2 \implies r \,|\, q^2$. So either $r = 1, r = q, r = q^2$. If $r = 1$, we are done. $r = q$ is impossible since $q \equiv 1 (\mod p)$ and $p \,|\, q - 1$ but $p > q$. Thus $r = q^2$.

Because $s \equiv 1 \mod q, s \,|\, |G| = pq^2$, we see $s \,|\, p \implies s = 1$ or $s = p$. If $s = 1$, we are done. So assume $s = p$.

Then we have q^2 subgroups H_i of order p and p subgroups K_i of order q^2. Consider $H_1 \cap H_2$. It is a subgroup of H_1 and H_2 and thus $|H_1 \cap H_2| \,|\, |H_1| = p \implies |H_1 \cap H_2| = 1$ or p, so $H_1 \cap H_2 = \{e\}$ or $H_1 = H_2$. Similarly, $|K_1 \cap H_1| \,|\, |H_1| = p \implies |K_1 \cap H_1| = 1$ or p and $|K_1 \cap H_1| \,|\, |K_1| = q^2 \implies |K_1 \cap H_1| = 1, q$, or q^2, so $|H_1 \cap K_1| = 1$ and $H_1 \cap K_1 = \{e\}$. Then $|G| \geqslant 1 + q^2(p-1) + (q^2 - 1)$ (element e, which contributes to 1, is in the common intersection of the Sylow groups. We notice that while we know all the Sylow p-subgroups only have trivial intersection, so each of them contributes $p - 1$ distinct elements. We also know that at least one Sylow q-subgroup contributes $q^2 - 1$ elements distinct from those already contributed by those p-subgroups. We don't know, however, if Sylow q-subgroups intersection trivially, so we only add $(q^2 - 1)$ instead of $p(q^2 - 1)$). Accidentally, the RHS is $1 + q^2(p-1) + (q^2 - 1) = 1 + q^2p - q^2 + q^2 - 1 = q^2p = |G|$ attaining the equality to the LHS, so $s = 1$, and we are done. □

Proposition 1.7.16. If $|G| = pqr$ where p, q, r are distinct prime numbers, then G has a normal Sylow subgroup.

Proof. We assume $p < q < r$ and let $n_p = \#$ of p-Sylow subgroups; $n_q = \#$ of q-Sylow subgroups; $n_r = \#$ of r-Sylow subgroups. Sylow's theorem gives $n_r \mid pq, n_r \equiv 1(\mathrm{mod} r)$. If $n_r = 1$ then we're done. n_r cannot be p or q because $q < r$ and $p < r$, so $n_r = pq$. Sylow's theorem gives $n_q \mid pr, n_q \equiv 1(\mathrm{mod} q)$. If $n_q = 1$ then we're done. n_q cannot be p as $p - 1 < q \Rightarrow q \nmid p - 1$,so $n_q = r$ or pr. Sylow's theorem gives $n_p \mid qr, n_p \equiv 1(\mathrm{mod} p)$. If $n_p = 1$ then we're done. $n_p = q, r$, or qr.

We can count by separating the common identity e. Because intersection of subgroups of prime order is a subgroup of each and divides both primes, we see the intersection can only be e if we assume the two subgroups are not the same (to rule out the case that they have the same prime order). Then $n_r = pq, n_q \geqslant r$, and $n_p \geqslant q$ provide a lower bound of $|G|$:

$$\begin{aligned} |G| &= 1 + n_r(r-1) + n_q(q-1) + n_p(p-1) \\ &\geqslant 1 + pq(r-1) + r(q-1) + q(p-1) \\ &= pqr + (r-1)(q-1) > pqr \end{aligned}$$

which is a contradiction. Thus either $n_r \neq pq \Rightarrow n_r = 1$ (we're done) or $n_p < r \Rightarrow n_p = 1$ (we're done) or $n_p < q \Rightarrow n_p = 1$ (we're done). □

Corollary 1.7.17. Group with order $|G| = 30 = 2 \times 3 \times 5$ has a normal Sylow subgroup.

Corollary 1.7.18. Every group of size $n \leqslant 30$ which is not of prime order is not simple.

Proof. We recall three rules: we have a nontrivial normal subgroup $N \trianglelefteq G$ if

1. $|G| = pq$ with $p \neq q$ (due to Corollary 1.7.14);
2. $|G = pq^2$ (due to Proposition 1.7.15);
3. $|G| = p^r$ (due to Corollary 1.6.17).

Now apply rule 1 to the following group orders:

$$6 = 2 \times 3,\ 10 = 2 \times 5,\ 14 = 2 \times 7,\ 15 = 3 \times 5,\ 21 = 3 \times 7,\ 22 = 2 \times 11,\ 26 = 13 \times 2$$

Apply rule 2 to the following group orders:

$$12 = 2^2 \times 3,\ 18 = 2 \times 3^2,\ 20 = 2^2 \times 5,\ 28 = 2^2 \times 7$$

Apply rule 3 to the following group orders:

$$8 = 2^3,\ 9 = 3^2,\ 16 = 2^4,\ 27 = 3^3$$

There are only two without being checked: $|G| = 30$ and $|G| = 24$. The $|G| = 30$ case is checked by Corollary 1.7.17. We show that group with order 24 has a non-trivial normal subgroup as well now:

Note that $24 = 2^3 \times 3$. Let r be the number of Sylow 2-subgroups and s be the number of Sylow 3-subgroups.

$$\begin{cases} r \equiv 1(\mathrm{mod}\ 2) \\ r \mid 3 \end{cases} \Longrightarrow \begin{cases} r = 1, \text{ so we have normal subgroup} \\ r = 3 \end{cases}$$

So assume $r = 3$, and we have Sylow 2-subgroups $H_1, H_2, H_3,\ |H_i| = 8$. Let $S = \{H_1, H_2, H_3\}$ and G acts on S by conjugation, i.e., $g \cdot H_i = gH_ig^{-1}$.

So there is a homomorphism $\phi : G \to S_3$, the group of permuations of S.

Note that $Ker(\phi) \trianglelefteq G$ and we calim that $Ker(\phi) \neq \{e\}$ or G, so $Ker(\phi)$ is the nontrivial normal subgroup we want to find.

- $Ker\phi \neq \{e\}$: $|G| = 24$, $|S_3| = 6 \implies \phi$ not injective $\implies Ker\phi \neq \{e\}$
- $Ker(\phi) \neq G$: Note that gH_iG^{-1} is still in S due to second Sylow theorem, so $\exists g \in G$ s.t. $gH_1g^{-1} = H_2 \implies g \cdot H_1 \neq H_1 \implies \phi(g) \neq e$, so there is some element in G that is not in the kernel of ϕ.

□

We have finished half of proving that any group of order < 60 is non-simple and solvable. The remaining orders are left as exercise below.

1.7 EXERCISES

1. Show that group of order 36 is non-simple by mimicing the proof for $|G| = 24$.
2. Show that group of order 48 is non-simple by mimicing the proof for $|G| = 24$.
3. Show that group of order 40 is non-simple by counting the number of Sylow 5-subgroups.
4. Show that group of order 56 is non-simple by counting the contributions of distinct elements from each Sylow subgroups.
5. Deduce that group of order < 60 is non-simple.
6. Deduce that group of order < 60 is solvable.

1.8 Products of Groups

1.8.1 Direct Product of Groups

Let G_1, G_2 be groups. Then $G_1 \times G_2 = \{(g_1, g_2) \,|\, g_1 \in G_1, g_2 \in G_2\}$ with $(g_1, g_2)(g_1', g_2') = (g_1g_1', g_2g_2')$ is the direct product of them. The identity element is (e_1, e_2) and the inverse of (g_1, g_2) is $(g_1, g_2)^{-1} = (g_1^{-1}, g_2^{-1})$.

Proposition 1.8.1. Let H and K be subgroups of a group G, and let $f : H \times K \to G$ be the multiplication map, defined by $f(h, k) = hk$. Its image is the set $HK = \{hk \mid h \in H, k \in K\}$.

(a) f is injective if and only if $H \cap K = \{1\}$.

(b) f is a homomorphism from the product group $H \times K$ to G if and only if elements of K commute with elements of $H : hk = kh$.

(c) If H is a normal subgroup of G, then HK is a subgroup of G.

(d) f is an isomorphism from the product group $H \times K$ to G if and only if $H \cap K = \{1\}$, $HK = G$, and also H and K are normal subgroups of G.

It is important to note that the multiplication map may be bijective though it isn't a group homomorphism. This happens, for instance, when $G = S_3$ and $H = \langle x \rangle$ and $K = \langle y \rangle$ where $x = (1\ 2\ 3)$ and $y = (1\ 2)$.

Proof.

(a) If $H \cap K$ contains an element $x \neq 1$, then x^{-1} is in H, and $f\left(x^{-1}, x\right) = 1 = f(1, 1)$, so f is not injective. Suppose that $H \cap K = \{1\}$. Let (h_1, k_1) and (h_2, k_2) be elements of $H \times K$ such that $h_1k_1 = h_2k_2$. We multiply both sides of this equation on the left by h_1^{-1} and on the right by k_2^{-1}, obtaining $k_1k_2^{-1} = h_1^{-1}h_2$. The left side is an element of K and the right side is an element of H. Since $H \cap K = \{1\}, k_1k_2^{-1} = h_1^{-1}h_2 = 1$. Then $k_1 = k_2, h_1 = h_2$, and $(h_1, k_1) = (h_2, k_2)$.

(b) Let (h_1, k_1) and (h_2, k_2) be elements of the product group $H \times K$. The product of these elements in the product group $H \times K$ is (h_1h_2, k_1k_2), and $f(h_1h_2, k_1k_2) = h_1h_2k_1k_2$, while $f(h_1, k_1) f(h_2, k_2) = h_1k_1h_2k_2$. These elements are equal if and only if $h_2k_1 = k_1h_2$.

(c) Suppose that H is a normal subgroup. We note that KH is a union of the left cosets kH with k in K, and that HK is a union of the right cosets Hk. Since H is normal, $kH = Hk$, and therefore $HK = KH$. Closure of HK under multiplication follows, because $HKHK = HHKK = HK$. Also, $(hk)^{-1} = k^{-1}h^{-1}$ is in $KH = HK$. This proves closure of HK under inverses.

(d) Suppose that H and K satisfy the conditions given. Then f is both injective and surjective, so it is bijective. According to (b), it is an isomorphism if and only if $hk = kh$ for all h in H and k in K. Consider the commutator $\left(hkh^{-1}\right)k^{-1} = h\left(kh^{-1}k^{-1}\right)$. Since K is normal, the left side is in K, and since H is normal, the right side is in H. Since $H \cap K = \{1\}, hkh^{-1}k^{-1} = 1$, and $hk = kh$. Conversely, if f is an isomorphism, one may verify the conditions listed in the isomorphic group $H \times K$ instead of in G.

□

Remark 1.8.2. In proof of (d), we saw $H \cap K = \{1\}, H, K \trianglelefteq G \iff \forall h \in H, k \in K,\ hk = kh$.

The condition $\forall h \in H, k \in K,\ hk = kh$ cannot be dropped. We give an example where $G \ncong H \times K$.

Example 1.8.3. $G = S_3$, $H = \{e, (1\ 2\ 3), (1\ 3\ 2)\}, K = \{e, (1\ 2)\}$. $HK = S_3, H \cap K = \{e\}$. But $S_3 \ncong H \times K \simeq Z_3 \times \mathbb{Z}_2$.

Example 1.8.4. One can use the above proposition to classify group of order 4 (a more elementary way is to use the group table, as in Question 10). See [1] Proposition 2.11.5.

We generalize the proudct of two groups:

Let I be an index set. Let $G_i, i \in I$ be groups indexed by I. Then

$$\prod_{i \in I} G_i = \{(x_i)_{i \in I} \,|\, x_i \in G_i\}$$

is the **direct product** of G_i. It is a group with multiplicaiton $(x_i)_{i \in I}(y_i)_{i \in I} = (x_iy_i)_{i \in I}$. For $A_i, i \in I$ abelian, we have the **direct sum**

$$\bigoplus_{i \in I} A_i = \{(a_i)_{i \in I} \,|\, \text{ there are only finitely many non-zero } a_i\} \leqslant \prod_{i \in I} A_i$$

which is an abelian group. For arbitrary groups $G_i, i \in I$, we can similarly define **weak product** as the set of I-tuples of $g_i \in G_i$ with only finitely many non-identity entries.

Let $I = \{1, 2, \cdots, n\}$, i.e., I is finite, then $\bigoplus_{i \in I} G_i = \prod_{i \in I} G_i$. Obviously, for $j = 1, \ldots, n$ the embedding

$$\begin{aligned} \varepsilon_j : G_j &\to \prod_{i \in I} G_i \\ g &\mapsto (1, \ldots, 1, \underbrace{g}_{j\text{-th}}, 1, \ldots, 1). \end{aligned}$$

is an isomorphism from G_j to

$$G_j^* := \{(g_1, \ldots, g_n) \mid g_i = 1 \text{ for } i \neq j\}.$$

For the subgroups $G_1^*, \ldots, G_n^*$ of $G := \prod_{i \in I} G_i$ one has:

- $G = G_1^* \cdots G_n^*$
- $G_i^* \trianglelefteq G, i = 1, \ldots, n$

- $G_i^* \cap \prod_{j \neq i} G_j^* = 1, i = 1, \ldots, n.$

Conversely, we have

Theorem 1.8.5. Let G be a group with subgroups $G_1^*, \ldots, G_n^*$ such that above three properties hold. Then the mapping

$$\alpha : \prod_{i=1}^{n} G_i^* \to G; \quad (g_1, \cdots, g_n) \mapsto g_1 \cdots g_n$$

is an isomorphism.

Proof. See [6] 1.6.1. □

Theorem 1.8.6. Let $G = G_1 \times \cdots \times G_n$.

(a) $Z(G) = Z(G_1) \times \cdots \times Z(G_n)$.

(b) $G' = G_1' \times \cdots \times G_n'$.

(c) Let N be a normal subgroup of G and $N_i = N \cap G_i (i = 1, \ldots, n)$. Suppose that $N = N_1 \times \cdots \times N_n$. Then the mapping

$$\alpha : G = G_1 \times \cdots \times G_n \to G_1/N_1 \times \cdots \times G_n/N_n$$

given by

$$g = (g_1, \ldots, g_n) \mapsto (g_1 N_1, \ldots, g_n N_n)$$

is an epimorphism, with Ker $\alpha = N$. In particular

$$G/N \cong G_1/N_1 \times \cdots \times G_n/N_n$$

(d) If the factors $G_1, \ldots, G_n$ are characteristic subgroups of G, then

$$\text{Aut } G \cong \text{Aut } G_1 \times \cdots \times \text{Aut } G_n.$$

Proof. See [6] 1.6.2 for the rest. □

Theorem 1.8.7. Let G be a group having normal subgroups $H_1, \cdots, H_n$. Then,

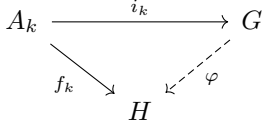
(a) If $G = \langle \bigcup_{i=1}^{n} H_i \rangle$ and, for all j, $1 = H_j \cap \langle \bigcup_{i \neq j} H_i \rangle$, then $G \cong H_1 \times \cdots \times H_n$.

(b) If each $a \in G$ has a unique expression of the form $a = h_1 \cdots h_n$, where each $h_i \in H_i$, then $G \cong H_1 \times \cdots \times H_n$.

Proof. See [9] Exercise 2.75. □

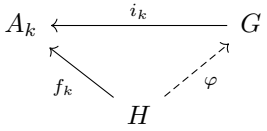
Here are two technical results about direct sums and products that will be useful.

Theorem 1.8.8 (Characteristic property of direct sum)**.** Let G be an abelian group, let $\{A_k\}_{k \in K}$ be a family of abelian groups, and let $\{i_k : A_k \to G\}_{k \in K}$ be a family of homomorphisms. Then $G \cong \bigoplus_{k \in K} A_k$ if and only if, given any abelian group H and any family of homomorphisms $\{f_k : A_k \to H : k \in K\}$, then there exists a unique homomorphism $\varphi : G \to H$ making the following diagrams commute ($\varphi i_k = f_k$):



Proof. See [9] Theorem 10.9. □

Theorem 1.8.9. Let G be an abelian group, let $\{A_k\}_{k\in K}$ be a family of abelian groups, and let $\{i_k : A_k \to G\}_{k\in K}$ be a family of homomorphisms. Then $G \cong \prod_{k\in K} A_k$ if and only if, given any abelian group H and any family of homomorphisms $\{f_k : H \to A_k : k \in K\}$, then there exists a unique homomorphism $\varphi : H \to G$ making the following diagrams commute for all k:



Proof. See [9] Theorem 10.10. □

Proposition 1.8.10.

(i) If $G = \bigoplus A_k$, prove that the maps $i_k : A_k \to G$ in Theorem 10.9 are injections.

(ii) If $G = \prod A_k$, prove that the maps $p_k : G \to A_k$ in Theorem 10.10 are surjections.

Proof. See [9] Exercise 10.4. □

1.8.2 Semi-Direct Product of Groups

We proved in the second isomomorphism theorem that if $K \leqslant G, H \trianglelefteq G$, then $HK \leqslant G$. Then K acts on H by conjugation.

$$\begin{aligned} \phi : K &\to \mathrm{Aut}(H) \\ k &\mapsto \phi_k \end{aligned}$$

where $\phi_k : h \to H;\ h \mapsto khk^{-1}$. It is easy to see that ϕ is a homomorphism.

Definition 1.8.11. Given two groups H and K and homomorphism $\phi : K \to \mathrm{Aut}(H)$, $k \mapsto \phi_k$. Then the set $H \times K$ with operation $(h,k)(h',k') = (h\phi_k(h'), kk')$ is a group, denoted by $H \rtimes K$, the **(external) semi-direct product** of H and K. The identity is (e_H, e_K), as $(e_H, e_K)(h,k) = (e_H\phi_{e_K}(h), k) = (h,k)$. $(h,k)(e_H,e_K) = (h\phi_h(e_H), ke_K) = (h,k)$. Inverse of (h,k) is $(\phi_{k^{-1}}(h^{-1}), k^{-1})$, as $(h,k)(\phi_{k^{-1}}(h^{-1}), k^{-1}) = (h\phi_k(\phi_{k^{-1}})(h^{-1}), e_K) = (e_H, e_K)$.

Fact: If ϕ is the identity homomorphism $\phi_k = e$ on H, then $H \rtimes K \simeq H \times K$.

We have noted in last subsection that $H \times K$ contains copies H and K as normal subgroup. That is, $H \times \{e\} \trianglelefteq H \times K, \{e\} \times K \trianglelefteq H \times K$. We show that this is also the case for semi-direct product:

Proposition 1.8.12. Let H and K be groups with $\phi : K \to \mathrm{Aut(H)}$ a homomorphism. Then the natural function from H to $H \rtimes K$ sending h to (h,e) is an injective group homomorphism and its image is a normal subgroup of $H \rtimes K$.

Proof. Let $f : H \to H \rtimes K; h \mapsto (h, e_K)$ be the function. We show that it is an injective group homomorphism.

It is a homomorphism: let $a, b \in H$.

$$f(a)f(b) = (a, e_K)\,(b, e_K) = (a\phi_{e_K}(b), e_K e_K) = (ab, e_K) = f(ab)$$

It is injective: let $f(h) = (e_H, e_K)$.

$$f(h) = (a, e_K) = (e_H, e_K) \Rightarrow h = e_H \Rightarrow \mathrm{Ker}(f) = \{e_H\}$$

The image of f is

$$\mathrm{Im}(f) = \{f(h) : h \in H\} = \{(h, e_K), h \in H\} = H \times \{e_K\}$$

We show that $H \times \{e_K\} \trianglelefteq H \rtimes K$: let $(a,b) \in H \rtimes K$. Then $(a,b)^{-1} = \left(\phi_{b^{-1}}\left(a^{-1}\right), b^{-1}\right)$ and

$$\begin{aligned}(a,b)(h,e_K)(a,b)^{-1} &= (a,b)(h,e_K)\left(\phi_{b^{-1}}\left(a^{-1}\right), b^{-1}\right) \\ &= (a\phi_b(h), b)\left(\phi_{b^{-1}}\left(a^{-1}\right), b^{-1}\right) = \left(a\phi_b(h)\phi_b\left(\phi_{b^{-1}}\left(a^{-1}\right)\right), bb^{-1}\right) \\ &= \left(a\phi_b(h)\phi_{e_K}\left(a^{-1}\right), e_K\right) = (\underbrace{a}_{\in H}\underbrace{\phi_b(h)}_{\in H}\underbrace{a^{-1}}_{\in H}, e_K) \in H \times \{e_K\}\end{aligned}$$

which shows that $\mathrm{Im}(f) = H \times \{e_K\} \trianglelefteq H \rtimes K$. □

Proposition 1.8.13. If $H, K \leqslant G, H \trianglelefteq G, H \cap K = \{e\}, G = HK$, then we call G **(internal) semi-direct product** of H and K, as we can prove that it is isomorphic to the externel semi-direct product of H and K with respect to conjugation as the homomorphism $k \mapsto Aut(H), k \mapsto \phi_k, \phi_k(h) = khk^{-1}$.

Proposition Proof. $f : H \rtimes K \to G, (h,k) \mapsto hk$. To show f injective, $f(h,k) = e \implies hk = e \implies h, k = e$. Check that it's a homomorphism. □

Corollary 1.8.14. $G = S_3$, $H = \{e, (1\ 2\ 3), (1\ 3\ 2)\} \cong \mathbb{Z}_3$, $K = \{e, (1\ 2) \cong \mathbb{Z}_2\}$. $S_3 \simeq \mathbb{Z}_3 \rtimes \mathbb{Z}_2$. $\mathbb{Z}_3$ has two automomorphisms, id and $f : a \to a^2$ (a is the generator). $\mathbb{Z}_2$ has two elements $[0], [1]$ and should be sent to $\{\mathrm{id}, f\}$. $[0] \mapsto \mathrm{id}$, so $[1] \mapsto f$.

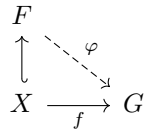
1.8.3 Wreath Product of Groups

see Rotman [9] p.172.

1.9 Free Groups, Free Products, and Group Presentations

We copy almost verbatim from RotmanGroup p.343-349. and p.388-391.

Definition 1.9.1 (Characteristic property of Free Group)**.** If X is a subset of a group F, then F is a **free group** with basis X if, for every group G and every function $f : X \to G$, there exists a unique homomorphism $\varphi : F \to G$ extending f.



We call this **characteristic property of free group**.

We shall see later that X must generate F.

Observe that a basis in a free group behaves precisely as does a basis $B = \{v_1, \ldots, v_m\}$ of a finite-dimensional vector space V. The theorem of linear algebra showing that matrices correspond to linear transformations rests on the fact that if W is any vector space and $w_1, \ldots, w_m \in W$, then there exists a unique linear transformation $T : V \to W$ with $T(v_i) = w_i$ for all i.

The following construction will be used in proving that free groups exist. Let X be a set and let X^{-1} be a set, disjoint from X, for which there is a bijection $X \to X^{-1}$, which we denote by $x \mapsto x^{-1}$. Let X' be a singleton set disjoint from $X \cup X^{-1}$ whose only element is denoted by 1. If $x \in X$, then x^1 may denote x and x^0 may denote 1.

Definition 1.9.2. A word on X is a sequence $w = (a_1, a_2, \ldots)$, where $a_i \in X \cup X^{-1} \cup \{1\}$ for all i, such that all $a_i = 1$ from some point on; that is, there is an integer $n \geqslant 0$ with $a_i = 1$ for all $i > n$. In particular, the constant sequence

$$(1, 1, 1, \ldots)$$

is a word, called the **empty word**, and it is also denoted by 1.

Since words contain only a finite number of letters before they become constant, we use the more suggestive notation for nonempty words:

$$w = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n},$$

where $x_i \in X, \varepsilon_i = +1, -1$, or 0 , and $\varepsilon_n = \pm 1$. Observe that this spelling of a word is unique: two sequences (a_i) and (b_i) are equal if and only if $a_i = b_i$ for all i. The **length** of the empty word is defined to be 0 ; the **length** of $w = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n}$ is defined to be n.

Definition 1.9.3. If $w = x_1^{\varepsilon_1} \ldots x_n^{\varepsilon_n}$ is a word, then its **inverse** is the word $w^{-1} = x_n^{-\varepsilon_n} \ldots x_1^{-\varepsilon_1}$.

Definition 1.9.4. A word w on X is **reduced** if either w is empty or $w = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n}$, where all $x_i \in X$, all $\varepsilon_i = \pm 1$, and x and x^{-1} are never adjacent. The empty word is reduced, and the inverse of a reduced word is reduced.

Definition 1.9.5. Definition. A **subword** of $w = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n}$ is either the empty word or a word of the form $v = x_i^{\varepsilon_i} \ldots x_j^{\varepsilon_j}$, where $1 \leqslant i \leqslant j \leqslant n$.

Thus, v is a subword of w if there are (possibly empty) subwords w' and w'' with $w = w'vw''$. A nonempty word w is reduced if and only if it contains no subwords of the form $x^{\varepsilon} x^{-\varepsilon}$ or x^0.

There is a multiplication of words: if

$$w = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n}, \quad u = y_1^{\delta_1} y_2^{\delta_2} \ldots y_m^{\delta_m},$$

then $wu = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n} y_1^{\delta_1} y_2^{\delta_2} \ldots y_m^{\delta_m}$. This multiplication does not define a product on the set of all reduced words on X because wu need not be reduced (even when both w and u are). One can define a new multiplication of reduced words w and u as the reduced word obtained from wu after cancellations. More precisely, there is a (possibly empty) subword v of w with $w = w'v$ such that v^{-1} is a subword of u with $u = v^{-1}u''$ and such that $w'u''$ is reduced. Define a product of reduced words, called **juxtaposition**, by

$$wu = w'u''.$$

Theorem 1.9.6. Given a set X, there exists a free group F with basis X.

Proof. See [9] Theorem 11.1. □

Corollary 1.9.7. Every group G is a quotient of a free group.

Proof. Construct a set $X = \{x_g : g \in G\}$ so that $f : x_g \mapsto g$ is a bijection $X \to G$. If F is free with basis X, then there is a homomorphism $\varphi : F \to G$ extending f, and φ is a surjection because f is. Therefore, $G \cong F/\ker\varphi$. □

1.9.1 Group Presentations

Definition 1.9.8. Let X be a set and let Δ be a family of words on X. A group G has generators X and relations Δ if $G \cong F/R$, where F is the free group with basis X and R is the normal subgroup of F generated by Δ. The **presentation of** G is denoted as $\langle X \mid \Delta \rangle$.

A relation $r \in \Delta$ is often written as $r = 1$ to convey its significance in the quotient group G being presented.

There are two reasons forcing us to define R as the normal subgroup of F generated by Δ : if $r \in \Delta$ and $w \in F$, then $r = 1$ in G implies $wrw^{-1} = 1$ in G; we wish to form a quotient group.

Example 1.9.9. $G = \mathbb{Z}_6$ has generator x and relation $x^6 = 1$. A free group $F = \langle x \rangle$ on one generator is infinite cyclic, and $\langle x \rangle / \langle x^6 \rangle \cong \mathbb{Z}_6$. A presentation of G is $\langle x \mid x^6 \rangle$.

Another presentation of $\mathbb{Z}_6$ is $\mathbb{Z}_6 = \langle x, y \mid x^3 = 1, y^2 = 1, xyx^{-1}y^{-1} = 1 \rangle$. The inclusion of a commutator as the relator makes the group abelian.

Example 1.9.10. A free abelian group G with basis X has presentation

$$G = \langle X \mid xyx^{-1}y^{-1} = 1 \text{ for all } x, y \in X \rangle;$$

a free group F with basis X has presentation

$$F = \langle X \mid \varnothing \rangle = \langle X \rangle.$$

Example 1.9.11. $X = \{x, y\}$, then $F = \{x^{k_1}y^{r_1} \cdots x^{k_n}y^{r_n} \mid r_n, k_n \in \mathbb{Z}, n > 0\}$.

Proposition 1.9.12. Let G be a free group generated by x, y. G is finitely generated, $H \leqslant G$ generated by $\{yxy^{-1}, y^2xy^{-2}, y^3xy^{-3}, ...\}$. Then H is <u>not</u> finitely generated.

Theorem 1.9.13. Let F and G be free groups with bases X and Y, respectively. Then $F \cong G$ if and only if $|X| = |Y|$.

Proof. See [9] Theorem 11.4. □

Definition 1.9.14. The **rank** of a free group F is the number of elements in a basis of F.

Above theorem says that the rank of F does not depend on the choice of the basis.

Corollary 1.9.15. If F is free with basis X, then F is generated by X.

Proof. See [9] Corollary 11.5. □

Theorem 1.9.16 (Nielsen-Schreier)**.** Every subgroup H of a free group F is itself free.

Proof. See [9] Theorem 11.44. □

1.9.2 Free Abelian Groups

Definition 1.9.17. A **Free abelian group** F is a direct sum of infinite cyclic groups. More precisely, there is a subset $X \subset F$ of elements of infinite order serving as its basis, i.e.,

$$F = \bigoplus_{x \in X} \langle x \rangle \cong \bigoplus_{x \in X} \mathbb{Z}.$$

We allow the possibility $X = \varnothing$, in which case $F = 0$.

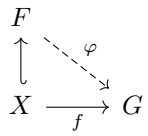
It is easy to see that if X is a basis of a free abelian group F, then each $u \in F$ has a unique expression of the form $u = \sum m_x x$, where $m_x \in \mathbb{Z}$ and $m_x = 0$ for "almost all" $x \in X$; that is, $m_x \neq 0$ for only a finite number of x.

The following theorem justifies "freeness" of the free abelian group (compare to characteristic property of free group where G is arbitrary. G is instead abelian in the following proposition.)

Proposition 1.9.18. Let F be a free abelian group with basis X, let G be any abelian group, and let $f : X \to G$ be any function. Then there is a unique homomorphism $\varphi : F \to G$ extending f; that is,

$$\varphi(x) = f(x) \quad \text{for all} \quad x \in X.$$

Indeed, if $u = \sum m_x x \in F$, then $\varphi(u) = \sum m_x f(u)$.



Proof. If $u \in F$, then uniqueness of the expression $u = \sum m_x x$ shows that $\varphi : u \mapsto \sum m_x f(u)$ is a well defined function. That φ is a homomorphism extending f is obvious; φ is unique because homomorphisms agreeing on a set of generators must be equal. □

As analogs of Corollary 1.9.7 and theorem 1.9.13, we have

Corollary 1.9.19. Every abelian group G is a quotient of a free abelian group.

Proof. See [9] Corollary 10.12. □

Theorem 1.9.20. Too free groups $F = \bigoplus_{x \in X} \langle x \rangle$ and $G = \bigoplus_{y \in Y} \langle y \rangle$ are isomorphic if and only if $|X| = |Y|$.

Proof. See [9] Theorem 10.14. □

Definition 1.9.21. The **rank** of a free abelian group is the cardinal of a basis.

It is clear that if F and G are free abelian, then

$$\operatorname{rank}(F \oplus G) = \operatorname{rank}(F) + \operatorname{rank}(G),$$

for a basis of $F \oplus G$ can be chosen as the union of a basis of F and a basis of G.

Remark 1.9.22. Exercise 11.46 and Theorem 11.6 of Rotman show that a group is free iff it has the projective property. This is the same case for the free abelian group. However, as we have noted, free abelian groups are not free groups (The only free abelian groups that are free groups are the trivial group and the infinite cyclic group). To see that the projective property for abelian group defines the free abelian group, we may note that a free module is projective and free abelian group is a free $\mathbb{Z}$-module. A projective module is free when R is a principal ideal domain like $\mathbb{Z}$.

As an analog of Theorem 1.9.16, we have

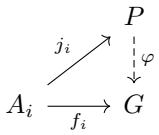
Theorem 1.9.23. Every subgroup H of a free abelian group F of rank n is itself free abelian; moreover, $\operatorname{rank}(H) \leqslant \operatorname{rank}(F)$.

Proof. See [9] Theorem 10.18. □

1.9.3 Free Products

We now generalize the notion of free group to that of free product. As with free groups, free products will be defined with a diagram; that is, they will be defined as solutions to a certain "universal mapping problem." Once existence and uniqueness are settled, then we shall give concrete descriptions of free products in terms of their elements and in terms of presentations.

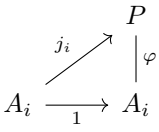
Definition 1.9.24. Let $\{A_i : i \in I\}$ be a family of groups. A free product of the A_i is a group P and a family of homomorphisms $j_i : A_i \to P$ such that, for every group G and every family of homomorphisms $f_i : A_i \to G$, there exists a unique homomorphism $\varphi : P \to G$ with $\varphi j_i = f_i$ for all i.



One should compare this with Theorem 1.8.8, the analogous property of direct sums of abelian groups.

Lemma 1.9.25. If P is a free product of $\{A_i : i \in I\}$, then the homomorphisms j_i are injections.

Proof. For fixed $i \in I$, consider the diagram in which $G = A_i, f_i$ is the identity and, for $k \neq i$, the maps $f_k : A_k \to A_i$ are trivial.



Then $\varphi j_i = 1_{A_i}$, and so j_i is an injection. □

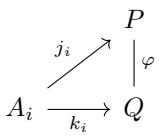
In light of this lemma, the maps $j_i : A_i \to P$ are called the **imbeddings**.

Example 1.9.26. A free group F is a free product of infinite cyclic groups. If X is a basis of F, then $\langle x \rangle$ is infinite cyclic for each $x \in X$; define $j_x : \langle x \rangle \hookrightarrow F$ to be the inclusion. If G is a group, then a function $f : X \to G$ determines a family of homomorphisms $f_x : \langle x \rangle \to G$, namely, $x^n \mapsto f(x)^n$. Also, the unique homomorphism $\varphi : F \to G$ which extends the function f clearly extends each of the homomorphisms f_x; that is, $\varphi j_x = f_x$ for all $x \in X$.

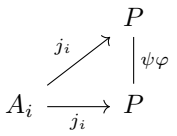
Here is the uniqueness theorem.

Theorem 1.9.27. Let $\{A_i : i \in I\}$ be a family of groups. If P and Q are each a free product of the A_i, then $P \cong Q$.

Proof. Let $j_i : A_i \to P$ and $k_i : A_i \to Q$ be the embeddings. Since P is a free product of the A_i, there is a homomorphism $\varphi : P \to Q$ with $\varphi_i = k_i$ for all i. Similarly, there is a map $\psi : Q \to P$ with $\psi k_i = j_i$ for all i.



Consider the new diagram.



Both $\psi\varphi$ and 1_P are maps making this diagram commute. By hypothesis, there can only be one such map, and so $\psi\varphi = 1_P$. Similarly, $\varphi\psi = 1_Q$, and so $\varphi : P \to Q$ is an isomorphism. □

Because of Theorem 11.50, we may speak of the free product P of $\{A_i : i \in I\}$; it is denoted by

$$P = *_{i\in I} A_i$$

if there are only finitely many A_i 's, one usually denotes the free product by

$$A_1 * \cdots * A_n.$$

Theorem 1.9.28. Given a family $\{A_i : i \in I\}$ of groups, a free product exists.

Proof. See [9] Theorem 11.51. □

For more theories, including the Van Kampen theorem, see Rotman [9] or an algebraic topology text.

1.9.4 Todd-Coxeter Algorithm

See RotmanGroup [9] p.351 or Artin [1] 7.11.

1.10 Abelian Groups

There are two remarks greatly facilitating the study of abelian groups. First, if $a, b \in G$ and $n \in \mathbb{Z}$, then $n(a + b) = na + nb$ (in multiplicative notation, $(ab)^n = a^n b^n$, for a and b commute). Second, if X is a nonempty subset of G, then $\langle X\rangle$ is the set of all linear combinations of elements in X having coefficients in $\mathbb{Z}$.

Definition 1.10.1. If G is an abelian p-group for some prime p, then G is called a **p-primary group**.

Theorem 1.10.2 (Primary decomposition)**.** Every finite abelian group G is a direct sum of p-primary groups.

$$G \cong \bigoplus_{p_i \text{ prime}} G_p.$$

where G_p is the set of all elements a in G such that ord(a) is a power of p, i.e., $\exists r \geqslant 1,\ p^r a = 0$.

Proof. One may see Rotman [9] Theorem 6.1 (which has many references to results in the book). We give a proof here.

Let $\phi : \oplus_{p \text{ prime}} A(p) \to A$ is homomorphism, $(x_p) \mapsto \sum x_p \in A$.

ϕ surjective: $a \in A, ord(a) = m = p_1^{r_1} \cdots p_n^{r_n}$, p_i distinct prime. Then proceed by induction on n. If $n = 1$, then $ord(a) = p_1^{r_1} \implies a \in A(p) \implies a \in \mathrm{Im}(\phi)$. Then for n, $ord(a) = p_1^{r_1} \cdots p_n^{r_n} \iff ap_1^{r_1} \cdots p_n^{r_n} = 0$. So since $p_1^n \cdots p_{n-1}^{r_{n-1}}$ and $p_n^{r_n}$ coprime, $\exists s, t \in \mathbb{Z}$ s.t. $sp_1^n \cdots p_{n-1}^{r_{n-1}} + tp_n^{r_n} = 1$, $asp_n^n \cdots p_{n-1}^{r_{n-1}} + atp_n^{r_n} = a$. Since the two numbers are in $\mathrm{Im}\,\phi$, their sum is in $\mathrm{Im}(\phi)$.

ϕ injective: Suppose $\phi((x_0)) = 0$, and $\exists q, x_q \neq 0$, then $\sum x_p = 0 \implies x_q = -\sum_{p\neq q} x_p \implies x_q = -x_{p_1} - \ldots - -x_{p_n}$. ord$(x_{p_i}) = p_i^{s_i} \implies p_1^{s_1} \cdots p_r^{s_r}(-x_{p_1} - \ldots - x_{p_r}) = 0 \iff q(p_1^{s_1} \cdots p_r^{s_r}) = 0 \implies ord(q) \,|\, p_1^{s_1} \cdots p_r^{s_r}$, a contradiction. □

Example 1.10.3. $G = \mathbb{Q}/\mathbb{Z}$, where $G_p = \{\frac{a}{b} + \mathbb{Z} \,|\, \frac{p^r a}{b} \in \mathbb{Z}\}$ for some r.Then $\frac{p^r a}{b} = c \implies \frac{a}{b} = \frac{c}{p^r}$, so $= \{\frac{c}{p^r} + \mathbb{Z} \,|\, c \in \mathbb{Z}, r \geqslant 0\}$.

Lemma 1.10.4. Let p be a prime. A group G of order p^n is cyclic if and only if it is an abelian group having a unique subgroup of order p. Thus, If A is a finite abelian p-group which is not cyclic, then A has at least 2 subgroups of order p.

Proof. See Rotman [9] Theorem 2.19. □

Theorem 1.10.5 (Cyclic decomposition)**.** A finite abelian p-group is a direct sum of cyclic groups (note that subgroups of p-groups are necessarily p-groups due to Lagrange's theorem, so these cyclic groups are also primary).

Proof. Let $a \in A$ be an element of maximal order. We prove by induction on $|A|$ that there is a $B \leqslant A$ such that $A = \langle a \rangle \oplus B$. This means that if $B_1, B_2 \leqslant A$ s.t. $B_1 \cap B_2 = \{0\}$.

If $|A| = p$, we are done.

Let $ord(a) = p^s$. Then $\langle a \rangle$ has a unique subgroup of order p. Let $\langle b \rangle$ be another subgroup of order p in A s.t. $\langle a \rangle \cap \langle b \rangle = \{0\}$, which exists due to the previous lemma.

Consider $\bar{A} = A/\langle b \rangle, |\bar{A}| = \frac{|A|}{p} < |A|$. Then there is $\bar{a} = a + \langle b \rangle$, an element of maximal order in $\bar{A}$.

By the induction hypothesis, there is a $\bar{B}$ such that $\bar{A} = \langle \bar{a} \rangle \oplus \bar{B}$.

So $\bar{B} \leqslant \bar{A} = A/\langle a \rangle \implies \bar{B} = B/\langle a \rangle$ for $B \leqslant A$ with $\langle a \rangle \subset B_0$. Then $A = \langle a \rangle \oplus B$. □

Corollary 1.10.6 (Basis Theorem)**.** Due to Theorem 1.10.2 and Theorem 1.10.5, every finite abelian group G can be written as

$$G \cong \mathbb{Z}_{p_1^{r_1}} \oplus \mathbb{Z}_{p_2^{r_2}} \oplus \cdots \oplus \mathbb{Z}_{p_m^{r_m}}$$

We will only mention the following result. See its proof in [9] Theorem 6.13 and 6.14, with definitions of elementary divisors, $U_p(n, G)$, and invariant factors.

Theorem 1.10.7 (Fundamental Theorem of Finite Abelian Groups)**.** If G and H are finite abelian groups, then $G \cong H \iff$ for all primes p, they have the same elementary divisors $\iff$ they have the same invariant factors.

We then come to the classification of finitely generated abelian groups. We first need a lemma to separate the torsion and torsion-free parts of the abelian group. We have seen that for $H, K \leqslant G$, we have $G \cong H \times K \iff H, K \trianglelefteq G, H \cap K = \{1\}, HK = G$. For abelian gruoup G, $H, K \trianglelefteq G$ is automatic. Thus, $G \cong H \oplus K \iff H \cap K = \{0\}, H + K = G$.

Lemma 1.10.8. If A is abelian and $B \leqslant A$ such that A/B is a free abelian group, then there is a subgroup $C \leqslant A$ such that $A = B \oplus C$ and $C \cong A/B$.

Proof. Let $\{a_i + B\}_{i \in I}$ be a basis for A/B. Let $C = \langle a_i \rangle \leqslant A$, which is free and thus by Theorem 1.9.13 is isomorphic to A/B. We claim that $A = B \oplus C$:

(1). $B \cap C = \{0\}$: Suppose $\sum_{i \in I} \lambda_i a_i \in B$, then $\sum_{i \in I} \lambda_i a_i + B = B$. Thus, $\sum_{i \in I} \lambda_i (a_i + B) = B$, where B is the 0 of A/B. Then, $\lambda_i = 0 \forall i$.

(2). $A = B + C$: If $a \in A$, then $a + B = \sum_{i \in I} \lambda_i (a + B)$ in A/B, and $a + B = \sum_{i \in I} (\lambda_i a_i) + B$. So $a - \underbrace{\sum_{i \in I} \lambda_i a_i}_{\in C} \in B \implies a \in B + C$. □

Another lemma will be used.

Lemma 1.10.9. Every subgroup of a finitely generated abelian group is finitely generated.

Proof. Let $H \leqslant A, A = \langle a_1, ..., a_n \rangle$, and proceed by induction on n. If $n = 1$, this is cyclic so clearly true.

$n-1 \implies n$: Let $B = \langle a_1, ..., a_{n-1} \rangle \leqslant A$. Then by induction hypothesis, $H \cap B = \langle h_1, ..., h_{n-1} \rangle$ generated by at most $n-1$ elements.

Also, $A/B =< a_n + B >$.

Note that $\frac{H+B}{B} \simeq \frac{H}{H \cap B}$. Since $\frac{H+B}{B} \leqslant \frac{A}{B}$, it is also cyclic, so $\frac{H}{H \cap B}$ cyclic, generated by some $\langle h_n + (H \cap B) \rangle, h_n \in H$.

So $H =< h_1, .., h_n >$, I need to show that they actually generate H. If $h \in H$, then $h + (H \cap B) = \lambda_n h_n + (H \cap B) \implies h - \lambda_n h_n \in (H \cap B) \implies h - \lambda_n h_n = \sum_{i=1}^{n-1} \lambda_i h_i \implies h = \sum_{i=1}^{n} \lambda_i h_i$. □

Definition 1.10.10. Let G be an ableian group. Then

- An element $a \in G$ is **torsion** if ord(a) is finite: $\exists n > 0, na = 0$.
- tG is the set of torsion elements in G, $tG \leqslant G$ since $na = 0, mb = 0 \implies nm(a+b) = 0$.
- G is **torsion-free** if $tG = \{0\}$.
- G is **torsion** if $tG = G$.

Example 1.10.11. $\mathbb{Z}$ is torsion-free. $\mathbb{Z}/m$ is torsion, and any finite abelian group is torsion.

Plan:

By applying Proposition 1.1.30 to the homomorphism $q : G \to G/tG$, we see $G/tG = G/Ker(q) \cong \text{Im}(q)$ is finitely generated if the abelian group G is finitely generated (note that G being abelian ensures tG is normal). Now, Theorem 1.10.12 will show that G/tG is torsion-free. This has a series of consequences:

Theorem 1.10.13 then says G/tG is free abelian, that is, $G/tG \cong \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$. Then Lemma 1.10.8 applies to G to get

$$G \cong tG \oplus F, \quad F \cong G/tG.$$

tG as a subgroup of finitely generated group G is finitely generated due to Lemma 1.10.9. This finitely generated torsion group is then finite by Theorem 1.10.14. Therefore, Theorem 1.10.6 concludes that

$$tG = \mathbb{Z}_{p_1^{r_1}} \oplus \cdots \oplus \mathbb{Z}_{p_m^{r_m}}.$$

Combine the two previous displayed equations to get

$$G \cong tG \oplus F \cong \mathbb{Z}_{p_1^{r_1}} \oplus \cdots \oplus \mathbb{Z}_{p_m^{r_m}} \cong \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}.$$

Theorem 1.10.12. The quotient group G/tG is torsion-free.

Proof. If $n(g + tG) = 0$ in G/tG for some $n \neq 0$, then $ng \in tG$, and so there is $m \neq 0$ with $m(ng) = 0$. Since $mn \neq 0$, we see $g \in tG$, and $g + tG = 0$ in G/tG. Thus, G/tG is torsion-free. □

Theorem 1.10.13. Every finitely generated torsion-free abelian group G is free abelian.

Proof. We prove the theorem by induction on n, where $G = \langle x_1, \ldots, x_n \rangle$. If $n = 1$ and $G \neq 0$, then G is cyclic; $G \cong \mathbb{Z}$ because it is torsion-free.

Define $H = \{g \in G : mg \in \langle x_n \rangle \text{ for some positive integer } m\}$. Now H is a subgroup of G and G/H is torsion-free: if $x \in G$ and $k(x + H) = 0$, then $kx \in H, m(kx) \in \langle x_n \rangle$, and so $x \in H$. Since G/H is a torsion-free group that can be generated by fewer than n elements, it is free abelian, by induction. By Lemma 1.10.8, $G = F \oplus H$, where $F \cong G/H$, and so it suffices to prove that H is cyclic. Note that H is finitely generated, being a summand (and hence a quotient) of the finitely generated group G.

If $g \in H$ and $g \neq 0$, then $mg = kx_n$ for some nonzero integers m and k. It is routine to check that the function $\varphi : H \to \mathbb{Q}$, given by $g \mapsto k/m$, is a well defined injective homomorphism; that is, H is (isomorphic to) a finitely generated subgroup of $\mathbb{Q}$, say, $H = \langle a_1/b_1, \ldots, a_t/b_t \rangle$. If $b = \prod_{i=1}^{t} b_i$, then the map $\psi : H \to \mathbb{Z}$, given by $h \mapsto bh$, is an injection (because H is torsion-free). Therefore, H is isomorphic to a nonzero subgroup of $\mathbb{Z}$, and hence it is infinite cyclic. □

Theorem 1.10.14. Every finitely generated torsion abelian group is finite.

Proof. If $ord(a_i) = m_i$, and $A = \langle a_1, ..., a_k \rangle = \{n_1 a_1 + ... + n_k a_k \,|\, n_i \in \mathbb{Z}\} = \{n_1 a_1 + ... + n_k a_k \,|\, n_1 \in \mathbb{Z}, 0 \leqslant n_i < m_i\}$, which is finite. □

Theorem 1.10.15 (Fundamental Theorem of Finitely Generated Abelian Groups)**.** Every finitely generated abelian group G is a direct sum of primary and infinite cyclic groups, and the number of summands of each kind depends only on G.

Proof. The first past is proved by our plan written before, i.e.,

$$G \cong tG \oplus F \cong \mathbb{Z}_{p_1^{r_1}} \oplus \cdots \oplus \mathbb{Z}_{p_m^{r_m}} \cong \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}.$$

The uniqueness of the number of primary cyclic summands is precisely [9] Theorem 6.11; the number of infinite cyclic summands is just rank(G/tG), and so it, too, depends only on G. □

Proposition 1.10.16. Free abelian groups are torsion-free

Proof. $A = \langle a_i \rangle$. Suppose $b \neq 0 \in A$ s.t. $mb = 0, b = \sum a_i \implies mb = \sum (m\lambda_i)a_i \implies m\lambda = 0 \forall i \implies b = 0$, a contradiction. □

Example 1.10.17. Torsion-free abelian groups are not necessarily free. Consider $\mathbb{Q}$ as an example:

- $\mathbb{Q}$ is torsion-free: let $0 \neq p/q \in \mathbb{Q}$. Suppose $m(p/q) = 0$. Then $mp \overset{p \neq 0}{\implies} m = 0$. Thus, $\nexists m > 0$ s.t. $m(p/q) = 0$. p/q is not torsion. $t\mathbb{Q} = \{0\}$.
- $\mathbb{Q}$ is not free: Any two nonzero rationals linearly independent, i.e., if $a, b \in \mathbb{Q}$, $a \neq 0, b \neq 0$, then $\exists m, n \in \mathbb{Z} - \{0\}$ s.t. $na + mb = 0$. So if $\mathbb{Q}$ were free, it would be free of rank 1 and hence cyclic.

1.11 Classification of Small Groups

For more on classification of small groups, see [9] Chapter4 p.82

By order,

2. $\mathbb{Z}_2$
3. $\mathbb{Z}_3$
4. $\mathbb{Z}_6 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2, \mathbb{Z}_4$

5. $\mathbb{Z}_5$
6. $\mathbb{Z}_2 \oplus \mathbb{Z}_3$. Non-abelian: S_3
7. $\mathbb{Z}_7$
8. $\mathbb{Z}_8, \mathbb{Z}_2 \oplus \mathbb{Z}_4, \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$. Non-abelian: D_4, Q_8
9. $\mathbb{Z}_9, \mathbb{Z}_3 \oplus \mathbb{Z}_3$
10. $\mathbb{Z}_{10} \cong \mathbb{Z}_5 \oplus \mathbb{Z}_2$. Non-abelian: D_5
11. $\mathbb{Z}_{11}$
12. $\mathbb{Z}_{12} \cong \mathbb{Z}_3 \oplus \mathbb{Z}_4, \mathbb{Z}_6 \oplus \mathbb{Z}_2 \cong \mathbb{Z}_3 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$. Non-abelian: $D_6 (\cong \mathbb{Z}_2 \times S_3), A_4, \mathbb{Z}_3 \rtimes \mathbb{Z}_4,$

Chapter 2

Rings

2.1 Rings and Ring Homomorphisms

Definition 2.1.1. A non-empty set R is a **ring** if it is closed under multiplication($\cdot$) and addition $(+)$ on R such that

- $(R, +)$ is an abelian group.
- (associativity) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
- (distributivity) $a \cdot (b + c) = a \cdot b + a \cdot c$, $(b + c) \cdot a = b \cdot a + c \cdot a$.
- There is a **unity**" $1 \in R$ s.t. $\forall a \in R,\ a \cdot 1 = 1 \cdot a = a$.

Proposition 2.1.2.

- Unity is unique. ($1 = 1 \cdot 1' = 1'$)
- $\forall a \in R,\ 0a = 0$. ($0a = (0 + 0)a = 0a + 0a \Rightarrow 0a = 0$)
- $\forall a \in R,\ a0 = 0$. (Similarly)
- $(-a)b = a(-b) = -(ab)$. ($(-a)b + ab = (-a + a)b = 0b = 0 \Rightarrow (-a)b = -(ab)$; similarly, $a(-b) = -(ab)$)
- $-a = (-1)a$. ($1 + (-1) = 0$, $a + (-1)a = a(1 + (-1)) = a0 = 0 \Rightarrow (-1)a = -a$)

Example 2.1.3. $(\mathbb{R}, +, \cdot)$, $(M_n(\mathbb{R}), +, \cdot), (\mathbb{R}[x], +, \cdot), (\mathbb{R}[[x]], +, \cdot)$, which is the ring of formal power series $\{a_0 + a_1x + a_2x^2 + ... \,\big|\, a_i \in \mathbb{R}\}$.

Example 2.1.4. $f : \mathbb{R} \to M_2(\mathbb{R}), r \mapsto \begin{bmatrix} r & 0 \\ 0 & 0 \end{bmatrix}$ does not satisfy $f(1_R) = 1_S$.

Let us see some more classes.

Definition 2.1.5. $S \subseteq R$ is a **subring** if

- $(S, +) \leqslant (R, +)$. (inherits additive group structure)
- $1 \in S$ and S is closed under multiplication. (inherits multiplicative structure)

Definition 2.1.6. An **extension ring (or ring extension)** of a ring R is any ring S of which R is a subring.

For example, the field of rational numbers $\mathbb{Q}$ and the ring of Gaussian integers $\mathbb{Z}[i]$ are extension rings of the ring of integers $\mathbb{Z}$.

For every ring R, the polynomial ring $R[x]$ is a ring extension of R. If S is a ring extension of R, and $a \in S$, the set

$$R[a] = \{f(a) \mid f(x) \in R[x]\},$$

is the smallest subring of S containing R and a, and is a ring extension of R. More generally, given finitely many elements $a_1, \ldots, a_n$ of S, we can consider

$$R\left[a_1, \ldots, a_n\right] = \left\{f\left(a_1, \ldots, a_n\right) \mid f\left(x_1 \ldots, x_n\right) \in R\left[x_1, \ldots, x_n\right]\right\},$$

which is the ring extension of R in S generated by $a_1, \ldots, a_n$.

Definition 2.1.7.

- R is a **division ring** if every $0 \neq a \in R$ is a **unit**, i.e., has a multiplicative inverse a^{-1} such that $a^{-1}a = aa^{-1} = 1$.
- A commutative division ring is a **field**.
- If $a, b \in R, a, b \neq 0$ but $ab = 0$, then a, b are called **zero devisors**. That is, $a \in R$ is a zero divisor if $a \neq 0$ and there is some $b \neq 0$ such that $ab = 0$.
- A nonzero commutative ring, i.e., $\neq \{0\}$, with no zero divisor is an **integral domain**. By the remark below, we see R is an integral domain if $\forall a, b \in R,\ ab = 0 \Rightarrow a = 0$ or $b = 0$.

Remark 2.1.8.

- units cannot be zero divisors: a has a multiplicative inverse a^{-1}. Then suppose $\exists b \neq 0$ such that $ab = 0$. Then $a^{-1}ab = a^{-1}0 = 0 \Rightarrow b = 0$, contradiction.
- a is not a zero divisor $\iff \neg(\exists b \neq 0 \text{ s.t. } ab = 0) \iff \forall b \neq 0,\ ab \neq 0$ (that is $b \neq 0 \to ab \neq 0$) $\iff (ab = 0 \to b = 0)$.

Example 2.1.9.

- $\mathbb{Z}$ is an integral domain
- $\mathbb{Z}_n$ is a field $\iff n$ is prime.

Proof. We prove that $\mathbb{Z}_n$ is a field $\iff n$ is prime. We need to show that n is a prime $\iff$ every $[a] \neq [0]$ has a multiplicative inverse.

$\Leftarrow$: $[a] \neq [0]$ a unit, so by Remark 2.1.8, $[a]$ is not a zero divisor. Then we show that $\gcd(a, n) = 1$. Suppose not, then $d = \gcd(a, n) > 1$ and $[a]\left[\frac{n}{d}\right] = \left[\frac{a}{d}\right] \underbrace{[n]}_{=[0]} = [0]$, which makes $[a]$ a zero divisor. Contradiction.

$\Rightarrow$: Suppose $\gcd(a, n) = 1$. Then $\gcd(a, n) = ax + ny = 1$. Since $ax + ny \neq ax \pmod{n}$, we see $[ax] = [ax + ny] = [1]$. Thus $[a][x] = [1]$, $[x] = [a]^{-1}$. □

Definition 2.1.10. Let R, S be rings, $f : R \to S$ is a **ring homomorphism** if

- $f(a + b) = f(a) + f(b)$. (i.e., $f : (R, +) \to (S, +)$ is a group homomorphism)
- $f(ab) = f(a)f(b)$, $f(1_R) = f(1_S)$. (i.e., multiplicative structure is also preserved)

If f is a bijective ring homomorphism, then it is a **ring isomorphism**.

Remark 2.1.11. We notice that a ring homomorphism is just a group homomorphism (with respect to the additive structure) plus a monoid homomorphism (with respect to the multiplicative structure). The inverse of a group isomorphism is a group isomorphism, and the inverse of a monid isomorphism is a monoid isomorphism. Thus, the inverse of a ring isomomorphism is a ring isomomorphism. In fact, just as in remark 1.1.9, it is an equivalence relation, and if we find an inverse function of a ring homomorphism as a function between sets, the map and its inverse will both automatically be ring isomomorphisms.

2.1.1 Matrix Rings [1]

Fix an arbitrary ring R and let n be a positive integer. Let $M_n(R)$ be the set of all $n \times n$ matrices with entries from R. The element (a_{ij}) of $M_n(R)$ is an $n \times n$ square array of elements of R whose entry in row i and column j is $a_{ij} \in R$. The set of matrices becomes a ring under the usual rules by which matrices of real numbers are added and multiplied. Addition is componentwise: the i, j entry of the matrix $(a_{ij}) + (b_{ij})$ is $a_{ij} + b_{ij}$. The i, j entry of the matrix product $(a_{ij}) \times (b_{ij})$ is $\sum_{k=1}^{n} a_{ik}b_{kj}$ (note that these matrices need to be square in order that multiplication of any two elements be defined). It is a straightforward calculation to check that these operations make $M_n(R)$ into a ring. When R is a field we shall prove that $M_n(R)$ is a ring by less computational means in Part III.

Note that if R is any nontrivial ring (even a commutative one) and $n \geqslant 2$ then $M_n(R)$ is not commutative: if $ab \neq 0$ in R let A be the matrix with a in position 1,1 and zeros elsewhere and let B be the matrix with b in position 1,2 and zeros elsewhere; then ab is the (nonzero) entry in position 1,2 of AB whereas BA is the zero matrix.

These two matrices also show that $M_n(R)$ has zero divisors for all nonzero rings R whenever $n \geqslant 2$.

An element (a_{ij}) of $M_n(R)$ is called a scalar matrix if for some $a \in R, a_{ii} = a$ for all $i \in \{1, \ldots, n\}$ and $a_{ij} = 0$ for all $i \neq j$ (i.e., all diagonal entries equal a and all off-diagonal entries are 0). The set of scalar matrices is a subring of $M_n(R)$. This subring is a copy of R (i.e., is "isomorphic" to R): if the matrix A has the element a along the main diagonal and the matrix B has the element b along the main diagonal then the matrix $A + B$ has $a + b$ along the diagonal and AB has ab along the diagonal (and all other entries 0). If R is commutative, the scalar matrices commute with all elements of $M_n(R)$. If R has a 1, then the scalar matrix with l's down the diagonal (the $n \times n$ identity matrix) is the 1 of $M_n(R)$. In this case the units in $M_n(R)$ are the invertible $n \times n$ matrices and the group of units is denoted $GL_n(R)$, the general linear group of degree n over R.

If S is a subring of R then $M_n(S)$ is a subring of $M_n(R)$. For instance $M_n(\mathbb{Z})$ is a subring of $M_n(\mathbb{Q})$ and $M_n(2\mathbb{Z})$ is a subring of both of these. Another example of a subring of $M_n(R)$ is the set of upper triangular matrices: $\{(a_{ij}) \mid a_{pq} = 0$ whenever $p > q\}$ (the set of matrices all of whose entries below the main diagonal are zero) - one easily checks that the sum and product of upper triangular matrices is upper triangular.

2.1.2 Group Rings [2]

Fix a commutative ring R with identity $1 \neq 0$ and let $G = \{g_1, g_2, \ldots, g_n\}$ be any finite group with group operation written multiplicatively. Define the group ring, RG, of G with coefficients in R to be the set of all formal sums

$$a_1g_1 + a_2g_2 + \cdots + a_ng_n \quad a_i \in R, \quad 1 \leqslant i \leqslant n.$$

If g_1 is the identity of G we shall write a_1g_1 simply as a_1. Similarly, we shall write the element $1g$ for $g \in G$ simply as g. Addition is defined "componentwise"

$$\begin{aligned}(a_1g_1 + a_2g_2 + \cdots + a_ng_n) + (b_1g_1 + b_2g_2 + \cdots + b_ng_n)\\ = (a_1 + b_1)\, g_1 + (a_2 + b_2)\, g_2 + \cdots + (a_n + b_n)\, g_n.\end{aligned}$$

Multiplication is performed by first defining $(ag_i)\,(bg_j) = (ab)g_k$, where the product ab is taken in R and $g_ig_j = g_k$ is the product in the group G. This product is then extended to all formal sums by the distributive laws so that the coefficient of g_k in the product $(a_1g_1 + \cdots + a_ng_n) \times (b_1g_1 + \cdots + b_ng_n)$ is $\sum_{g_ig_j=g_k} a_ib_j$. It is straightforward to check that these operations make RG into a ring (again, commutativity of R is not needed). The associativity of multiplication follows from the associativity of the group operation in G. The ring RG is commutative if and only if G is a commutative group.

[1] Taken from [3] sec 7.2
[2] Taken from [3] sec 7.2

Example 2.1.12. Let $G = D_8$ be the dihedral group of order 8 with the usual generators r, s $\left(r^4 = s^2 = 1\right.$ and $\left. rs = sr^{-1}\right)$ and let $R = \mathbb{Z}$. The elements $\alpha = r + r^2 - 2s$ and $\beta = -3r^2 + rs$ are typical members of $\mathbb{Z}D_8$. Their sum and product are then

$$\begin{aligned}\alpha + \beta &= r - 2r^2 - 2s + rs \\ \alpha\beta &= \left(r + r^2 - 2s\right)\left(-3r^2 + rs\right) \\ &= r\left(-3r^2 + rs\right) + r^2\left(-3r^2 + rs\right) - 2s\left(-3r^2 + rs\right) \\ &= -3r^3 + r^2 - 3 + r^3 s + 6r^2 s - 2r^3 \\ &= -3 - 5r^3 + 7r^2 + r^3 s\end{aligned}$$

The ring R appears in RG as the "constant" formal sums i.e., the R-multiples of the identity of G (note that the definition of the addition and multiplication in RG restricted to these elements is just the addition and multiplication in R). These elements of R commute with all elements of RG. The identity of R is the identity of RG.

The group G also appears in RG (the element g_i appears as $1g_i$ - for example, $r, s \in D_8$ are also elements of the group ring $\mathbb{Z}D_8$ above) - multiplication in the ring RG restricted to G is just the group operation. In particular, each element of G has a multiplicative inverse in the ring RG (namely, its inverse in G). This says that G is a subgroup of the group of units of RG.

If $|G| > 1$ then RG always has zero divisors. For example, let g be any element of G of order $m > 1$. Then

$$(1 - g)\left(1 + g + \cdots + g^{m-1}\right) = 1 - g^m = 1 - 1 = 0$$

so $1 - g$ is a zero divisor (note that by definition of RG neither of the formal sums in the above product is zero).

If S is a subring of R then SG is a subring of RG. For instance, $\mathbb{Z}G$ (called the integral group ring of G) is a subring of $\mathbb{Q}G$ (the rational group ring of G). Furthermore, if H is a subgroup of G then RH is a subring of RG. The set of all elements of RG whose coefficients sum to zero is a subring (without identity). If $|G| > 1$, the set of elements with zero "constant term" (i.e., the coefficient of the identity of G is zero) is not a subring (it is not closed under multiplication).

2.1 EXERCISES

1. [3] 7.1.13. An element x in R is called nilpotent if $x^m = 0$ for some $m \in \mathbb{Z}^+$.

i. Show that if $n = a^k b$ for some integers a and b then $\overline{ab}$ is a nilpotent element of $\mathbb{Z}/n\mathbb{Z}$.
ii. If $a \in \mathbb{Z}$ is an integer, show that the element $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$ is nilpotent if and only if every prime divisor of n is also a divisor of a. In particular, determine the nilpotent elements of $\mathbb{Z}/72\mathbb{Z}$ explicitly.
iii. Let R be the ring of functions from a nonempty set X to a field F. Prove that R contains no nonzero nilpotent elements.

2. [3] 7.1.14. Let x be a nilpotent element of the commutative ring R (cf. the preceding exercise).

i. Prove that x is either zero or a zero divisor.
ii. Prove that rx is nilpotent for all $r \in R$.
iii. Prove that $1 + x$ is a unit in R.
iv. Deduce that the sum of a nilpotent element and a unit is a unit.

2.2 Ideals and Quotient Rings

Definition 2.2.1. $I \subseteq R$ is a **left ideal** if

- $(I, +) \leqslant (R, +)$
- $\forall r \in R, a \in I$, we have $ra \in I$.

A **right ideal** is similarly defined:

- $(I, +) \leqslant (R, +)$
- $\forall r \in R, a \in I$, we have $ar \in I$.

$I \subset R$ is an **ideal** if it is both a left ideal and a right ideal.

We note that since $a \in I$ and $0 \in R$ we have $0a = a0 = 0$ in ideal I. Also, 1 may not be in the ideal. If $1 \in I$, then I is the whole ring R, and we will give it a name soon.

Remark 2.2.2. We will assume that all rings R are commutative rings in this course if not specified, that is, $\forall a, b \in R,\ ab = ba$.

Due to this remark, left and ring ideals are just ideals.

Definition 2.2.3. In any ring R, the multiples of a particular element a form an ideal called the **principal ideal** generated by a. An element b of R is in this ideal if and only if b is a multiple pf a, which is to say, if and only if a divides b in R, denoted by $a \,|\, b$. There are several notations for this principal ideal:

$$(a) = aR = Ra = \{ra \mid r \in R\}$$

Example 2.2.4. The ring R itself is the principal ideal (1), and because of this it is called the **unit ideal**. It is the only ideal that contains a unit of the ring. The set consisting of zero alone is the principal ideal (0), and is called the **zero ideal**. An ideal I is **proper** if it is neither the zero ideal nor the unit ideal.

Definition 2.2.5. The ideal I **generated by a set of elements** $X \subset R$ is the smallest ideal that contains those elements. It is defined as

$$\langle X \rangle := \{r_1 x_1 + \cdots r_k x_k \mid k \geqslant 1, r_i \in R, x_i \in X\}.$$

In particular, for an ideal I and an element $a \in R$, we have

$$\langle a, I \rangle = \{r_1 a + r_2 i \mid r_1, r_2 \in R, i \in I\} = \{ra + i \mid r \in R, i \in I\}$$

Proposition 2.2.6. If $f : R \to S$ is a ring homomorphism, then

(1) $Ker(f)$ is an ideal of R.

(2) If I' is an ideal of S, then $f^{-1}(I')$ is an ideal (as the kernel of $R \to S \to S/I$); however, $f(I)$ for ideal $I \subseteq R$ may not be an ideal. When f is surjective, $f(I)$ is an ideal.

(3) $\mathrm{Im}(f)$ is a subring of S.

(4) If P is a subring of R, then $f(P)$ is a subring; If P' is a subring of S, then $f^{-1}(P')$ is a subring.

Proof. We leave the last two statements as exercises and prove the first two and give an example illustrating when $f(I)$ is not an ideal.

(1) Clearly, $(Ker(f), +) \leqslant (R, +)$. Then consider $a \in Ker(f)$, i.e., $f(a) = 0$, and $r \in R$. Now,

$$f(ar) = f(a)f(r) = 0$$
$$f(ra) = f(r)f(a) = 0.$$

(2) Let $I = f^{-1}(I')$. We know that the preimage of a group homomorphism is a subgroup, so I is an additive subgroup of R. We need to show for $r \in R$ and $a \in I$, we have $ra \in I$. Since I' is an ideal, $f(ra) = f(r)f(a) \in I'$, thus $ra \in f^{-1}(S) = I$. Thus I is an ideal of R. This proved that the preimage of an ideal under a ring homomorphism is an ideal. We now show that the image of an ideal under a surjective ring homomorphism is an ideal. As I is an additive subgroup of R and f is also a group homomorphism, $f(I)$ is an additive subgroup of S. We need to show for $s \in S$ and $f(a) \in f(I)$, we have $sf(a) \in f(I)$. For $s \in S$, because f is surjective, there exists $r \in R$ such that $f(r) = s$. Then $ra \in I$, so

$$sf(a) = f(r)f(a) = f(ra) \in f(I)$$

Thus $f(I)$ is an ideal.

□

Example 2.2.7. Let $i : \mathbb{Z} \to \mathbb{Q}$ be inclusion. Since $\mathbb{Q}$ is a field, ideal I in $\mathbb{Q}$ is either (0) or $(1) = \mathbb{Q}$. We take an ideal $n\mathbb{Z}$ in $\mathbb{Z}$ with $n \neq 0$. Since $i(n\mathbb{Z}) = n\mathbb{Z}$ is not (0) or (1) we see that it is not an ideal.

Let $I \subset R$ be an ideal, then we define $R/I := \{r + I \,\big|\, r \in R\}$, with $(r + I) + (s + I) := (r + s) + I$ and $(r + I)(s + I) = rs + I$.

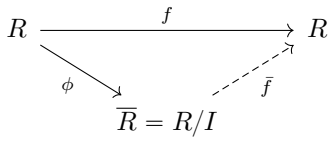
Proof of Well-defined Multiplication. Want to check that $r+I = r'+I$ and $s+I = s'+I \implies rs+I = r's'+I$. $r - r', s - s' \in I$. On the other side, $rs - r's' = r(s - s') + (r - r')s' \in I$, which is true. □

R/I is a ring, called **quotient ring**, with unity $1 + R$ and zero $0 + R$. The **canonical homomorphism** is given by

$$\phi : R \to R/I, \quad r \mapsto r + I$$

where f is clearly surjective and $Ker(\phi) = I$.

Proposition 2.2.8 (Mapping property)**.** Suppose $f : R \to R'$ is a ring homomorphism with $K = Ker(f)$ and $I \subseteq K$ an ideal. Then $\exists!$ homomorphism $\bar{f} : \overline{R} = R/I \to R'$ such that $\bar{f}\phi = f$:



We say f factors through ϕ.

2.2.1 Ring Isomorphism Theorems

If I and J are two ideals in R, we define

$$I + J = \{i + j \mid i \in I, j \in J\}$$

$$IJ = \left\{ \sum_{i=1}^{n} a_i b_i \mid n \geqslant 1, a_i \in I, b_i \in J \right\}$$

Proposition 2.2.9. If I and J are two ideals of R, then their sum $I + J$, intersection $I \cap J$, and product IJ are still ideals. Besides, $IJ \subseteq I \cap J$.

Proof. Exercise. □

We state the isomomorphism theorems for rings without proof (see [1] or [10] if one needs).

Theorem 2.2.10 (First Isomorphism Theorem for Rings)**.** If $f : R \to S$ is a ring homomorphism, then

$$R/\underbrace{Ker(f)}_{\text{an ideal}} \cong \underbrace{\text{Im}(f)}_{\text{a subring}}$$

Theorem 2.2.11 (Second Isomorphism Theorem for Rings)**.** Let R be a ring, and let S be a subring of R, J be an ideal of R.

Then:

- $S + J$ is a subring of R;
- J is an ideal of $S + J$;
- $S \cap J$ is an ideal of S;
- $\frac{S}{S\cap J} \cong \frac{S+J}{J}$.

Theorem 2.2.12 (Third Isomorphism Theorem for Rings)**.** If $I \subset J \subseteq R$, and I, J are ideals in R, then

$$J/I = \{j + I \,|\, j \in J\}$$

is an ideal of R/I and

$$\frac{R/I}{J/I} \cong R/J.$$

Theorem 2.2.13 (Fourth Isomorphism Theorem (Correspondance Theorem) for Rings)**.** Let $\varphi : R \to \mathcal{R}$ be a surjective ring homomorphism with kernel K. There is a bijective correspondence between the set of all ideals of $\mathcal{R}$ and the set of ideals of R that contain K:

$$\{\text{ideals of } R \text{ that contain } K\} \longleftrightarrow \{\text{ideals of } \mathcal{R}\}.$$

This correspondence is defined as follows:

- If I is a ideal of R and if $K \subset I$, the corresponding ideal of $\mathcal{R}$ is $\varphi(I)$.
- If $\mathcal{I}$ is a ideal of $\mathcal{R}$, the corresponding ideal of R is $\varphi^{-1}(\mathcal{I})$.

If the ideal I of R corresponds to the ideal $\mathcal{I}$ of $\mathcal{R}$, the quotient rings R/I and $\mathcal{R}/\mathcal{I}$ are naturally isomorphic.

Note that the inclusion $K \subset I$ is the reverse of the one in the mapping property.

Remark 2.2.14. A more common version is to let the surjective ring homomorphism in the above statement be $\varphi : R \to R/J$ where J is an ideal of R.

2.2 EXERCISES

1. Let I and J be ideals of R.

i. Prove that $I + J$ is the smallest ideal of R containing both I and J.
ii. Prove IJ is an ideal contained in $I \cap J$
iii. Give and example where $IJ \neq I \cap J$.
iv. Prove if $I + J = R$, the $IJ = I \cap J$.

2. For an ideal I of R, let

$$\sqrt{I} = \{x \in R \mid x^n \in I \text{ for some } n \geqslant 1\}.$$

$\sqrt{I}$ is called the radical of I.

i. Show that $\sqrt{I}$ is an ideal of R which contains I.
ii. Show that $\sqrt{IJ} = \sqrt{I \cap J}$ for any two ideals I and J.

For more on radicals of rings, see [2] Chap.8.

3. [3] Ex7.3-6. Decide which of the following are ring homomorphisms from $M_2(\mathbb{Z})$ to $\mathbb{Z}$:

i. $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto a$ (projection onto the 1,1 entry)
ii. $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto a + d$ (the trace of the matrix)
iii. $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto ad - bc$ (the determinant of the matrix).

4. [3] Ex7.3-7. Let $R = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \middle| a, b, d \in \mathbb{Z} \right\}$ be the subring of $M_2(\mathbb{Z})$ of upper triangular matrices. Prove that the map

$$\varphi : R \to \mathbb{Z} \times \mathbb{Z} \text{ defined by } \varphi : \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mapsto (a, d)$$

is a surjective homomorphism and describe its kemel.

5. [3] Ex7.3-8. Decide which of the following are ideals of the ring $\mathbb{Z} \times \mathbb{Z}$:

i. $\{(a, a) \mid a \in \mathbb{Z}\}$
ii. $\{(2a, 2b) \mid a, b \in \mathbb{Z}\}$
iii. $\{(2a, 0) \mid a \in \mathbb{Z}\}$
iv. $\{(a, -a) \mid a \in \mathbb{Z}\}$.

2.3 Maximal Ideals and Prime Ideals

We will first present Zorn's Lemma using a well-written document, which is widely used in many proofs, and then talk about two important ideals, maximal ideals and prime ideals.

2.3.1 Zorn's Lemma

Theorem 2.3.1 (Zorn's lemma)**.** Let S be a partially ordered set. If every totally ordered subset of S has an upper bound, then S contains a maximal element.

To understand Zorn's Lemma, we need to know four terms: partially ordered set, totally ordered subset, upper bound, and maximal element.

A **partial ordering** on a (nonempty) set S is a binary relation on S, denoted $\leqslant$, which satisfies the following properties:

- reflexive: for all $s \in S, s \leqslant s$,
- antisymmetric: if $s \leqslant s'$ and $s' \leqslant s$ then $s = s'$,
- transitive: if $s \leqslant s'$ and $s' \leqslant s''$ then $s \leqslant s''$.

When we fix a partial ordering $\leqslant$ on S, we refer to S (or, more precisely, to the pair $(S, \leqslant)$) as a **partially ordered set**, also abbreviated as **poset**.

It is important to notice that we do not assume all pairs of elements in S are **comparable** under $\leqslant$: for some s and s' we may have neither $s \leqslant s'$ nor $s' \leqslant s$. If all pairs of elements can be compared (that is, for all s and s' in S either $s \leqslant s'$ or $s' \leqslant s$) then we say S is **totally ordered** with respect to $\leqslant$.

Example 2.3.2. The usual ordering relation $\leqslant$ on $\mathbb{R}$ or on $\mathbb{Z}^+$ is a partial ordering of these sets. In fact it is a total ordering on either set. This ordering on $\mathbb{Z}^+$ is the basis for proofs by induction.

Example 2.3.3. On $\mathbb{Z}^+$, declare $a \leqslant b$ if $a \mid b$. This partial ordering on $\mathbb{Z}^+$ is different from the one in previous example and is called ordering by divisibility. It is one of the central relations in number theory. (Proofs about $\mathbb{Z}^+$ in number theory sometimes work not by induction, but by starting on primes, then extending to prime powers, and then extending to all positive integers using prime factorization. Such proofs view $\mathbb{Z}^+$ through the divisibility relation rather than through the usual ordering relation.) Unlike the ordering on $\mathbb{Z}^+$ in previous example, $\mathbb{Z}^+$ is not totally ordered by divisibility: most pairs of integers are not comparable under the divisibility relation. For instance, 3 doesn't divide 5 and 5 doesn't divide 3 . The subset $\{1, 2, 4, 8, 16, \ldots\}$ of powers of 2 is totally ordered under divisibility.

Example 2.3.4. Let S be the set of all subgroups of a given group G. For $H, K \in S$ (that is, H and K are subgroups of G), declare $H \leqslant K$ if H is a subset of K. This is a partial ordering, called ordering by inclusion. It is not a total ordering: for most subgroups H and K neither $H \subset K$ nor $K \subset H$.

One can similarly partially order the subspaces of a vector space or the ideals (or subrings or all subsets) of a commutative ring by inclusion. We shall see this in the next section.

Example 2.3.5. If S is a partially ordered set for the relation $\leqslant$ and $T \subset S$, then the relation $\leqslant$ provides a partial ordering on T. Thus T is a new partially ordered set under $\leqslant$. For instance, the partial ordering by inclusion on the subgroups of a group restricts to a partial ordering on the cyclic subgroups of a group.

Lemma 2.3.6. Let S be a partially ordered set. If $\{s_1, \ldots, s_n\}$ is a finite totally ordered subset of S then there is an s_i such that $s_j \leqslant s_i$ for all $j = 1, \ldots, n$.

Proof. The s_i 's are all comparable to each other; that's what being totally ordered means. Since we're dealing with a finite set of pairwise comparable elements, there will be one that is greater than or equal to them all in the partial ordering on S. The reader can formalize this with a proof by induction on n, or think about the bubble sort algorithm. □

An **upper bound** on a subset T of a partially ordered set S is an $s \in S$ such that $t \leqslant s$ for all $t \in T$. It is important to notice that when we say T has an upper bound in S, we do not assume the upper bound is in T itself; it is just in S.

Example 2.3.7. In $\mathbb{R}$ with its natural ordering, the subset $\mathbb{Z}$ has no upper bound while the subset of negative real numbers has the upper bound 0 (or any positive real). No upper bound on the negative real numbers is a negative real number.

Example 2.3.8. In the proper subgroups of $\mathbb{Z}$ ordered by inclusion, an upper bound on $\{4\mathbb{Z}, 6\mathbb{Z}, 8\mathbb{Z}\}$ is $2\mathbb{Z}$ since $4\mathbb{Z}, 6\mathbb{Z}$, and $8\mathbb{Z}$ all consist entirely of even numbers. (Note $4\mathbb{Z} \subset 2\mathbb{Z}$, not $2\mathbb{Z} \subset 4\mathbb{Z}$.)

A **maximal element** m of a partially ordered set S is an element that is not below any element *to which it is comparable*: for all $s \in S$ to which m is comparable, $s \leqslant m$. Equivalently, m is maximal when the only $s \in S$ satisfying $m \leqslant s$ is $s = m$. This does not mean $s \leqslant m$ for all s in S since we don't insist that maximal elements are actually comparable to every element of S. A partially ordered set could have many maximal elements.

We now return to the statement of Zorn's lemma: If every totally ordered subset of a partially ordered set S has an upper bound, then S contains a maximal element.

All the terms being used here have now been defined. Of course this doesn't mean the statement should be any clearer!

Zorn's lemma is not intuitive, but it turns out to be logically equivalent to more readily appreciated statements from set theory like the Axiom of Choice (which says the Cartesian product of any family of nonempty sets is nonempty) and Well-Ordering Principle (which says every nonempty set has a well-ordering: that means a total ordering in which every nonempty subset has a least element).

2.3.2 Maximal Ideals

The ideals in a commutative ring can be partially ordered by inclusion. The whole ring, which is the unit ideal (1), is obviously maximal for this ordering. But this is boring and useless. Proper ideals that are maximal for inclusion among the proper ideals are called the maximal ideals in the ring. (That is, a maximal ideal is understood to mean a maximal proper ideal.)

Definition 2.3.9 (Maximal Ideals)**.** An ideal $M \subsetneq R$ is called a **maximal ideal** if for any $I \subseteq R$ with $M \subseteq I \subseteq R$, then $I = M$ or $I = R$. That is, the only ideals containing M are M and R.

Proposition 2.3.10. Every nonzero commutative ring contains a maximal ideal.

Proof. Let S be the set of proper ideals in a commutative ring $R \neq 0$. Since the zero ideal (0) is a proper ideal, $S \neq \varnothing$. We partially order S by inclusion.

Let $\{I_\alpha\}_{\alpha \in A}$ be a totally ordered set of proper ideals in R. To write down an upper bound for these ideals in S, it is natural to try their union $I = \bigcup_{\alpha \in A} I_\alpha$. As a set, I certainly contains all the I_α 's, but is I an ideal? We may be hesitant about this, since a union of ideals is not usually an ideal: try $2\mathbf{Z} \cup 3\mathbf{Z}$. But we are dealing with a union of a totally ordered set of ideals, and the total ordering of the ideals will be handy!

If x and y are in I then $x \in I_\alpha$ and $y \in I_\beta$ for two of the ideals I_α and I_β. Since this set of ideals is totally ordered, $I_\alpha \subset I_\beta$ or $I_\beta \subset I_\alpha$. Without loss of generality, $I_\alpha \subset I_\beta$. Therefore x and y are in I_β, so $x \pm y \in I_\beta \subset I$. Hence I is an additive subgroup of R. The reader can check $rx \in I$ for $r \in R$ and $x \in I$, so I is an ideal in R.

Because I contains every I_α, I is an upper bound on the totally ordered subset $\{I_\alpha\}_{\alpha \in A}$ provided it is actually in S : is I a proper ideal? Well, if I is not a proper ideal then $1 \in I$. Since I is the union of the I_α 's, we must have $1 \in I_\alpha$ for some α, but then I_α is not a proper ideal. That is a contradiction, so $1 \notin I$. Thus $I \in S$ and we have shown every totally ordered subset of S has an upper bound in S.

By Zorn's lemma S contains a maximal element. This maximal element is a proper ideal of R that is maximal for inclusion among all proper ideals (not properly contained in any other proper ideal of R). That means it is a maximal ideal of R. □

Corollary 2.3.11. Every proper ideal in a nonzero commutative ring is contained in a maximal ideal.

Proof. Let R be the ring and I be a proper ideal in R. The quotient ring R/I is nonzero, so it contains a maximal ideal by previous theorem. The inverse image of this ideal under the natural reduction map $R \to R/I$ is a maximal ideal of R that contains I. □

Proposition 2.3.12. I is maximal ideal $\iff$ R/I is a field.

Proof. $\implies$: Assume $r + I \neq I$, so $r \notin I$. Let $J = \langle r, I \rangle \subseteq R$ (see Definition 2.2.5). Clearly, $I \subseteq J \subseteq R$. Since J an ideal and I a maximal ideal, we have $I = J$ or $J = R$. Since $r \in J - I$, so $J = R \implies 1 \in J = \langle r, I \rangle \implies 1 = r'r + i$. Thus $1 - rr' \in I \implies (1 + I) = (r + I)(r' + I)$, where $(r' + I)$ is the inverse of $(r + I)$.

$\impliedby$: If R/I is a field and $I \subseteq J \subseteq R$, then J/I is an ideal of R/I. The only proper ideals of a field is $\{0\}$ or itself. Therefore, J/I is (0) or R/I, so $J = I$ or $J = R$. □

While the trick above worth remembering, we have an easier proof of the fact.

Proposition 2.3.13.

(a) Let $\varphi : R \to R'$ be a surjective ring homomorphism, with kernel I. The image R' is a field if and only if I is a maximal ideal.

(b) An ideal I of a ring R is maximal if and only if $\bar{R} = R/I$ is a field.

(c) The zero ideal of a ring R is maximal if and only if R is a field.

Proof. (a): A ring is a field if it contains precisely two ideals,so the Correspondence Theorem asserts that the image of φ is a field if and only if there are two precisely ideals that contain its kernel I. This will be true if and only if I is a maximal ideal.

(b) and (c) follow from (a) by applying to the map $R \to R/I$. □

Corollary 2.3.14. $I = \{0\}$ is a maximal ideal $\iff$ $R = R/\{0\}$ is a field.

2.3.3 Some Terminologies

We review some concepts and also give others. Let R be a commutative ring with unity 1.

1. u is a **unit** $\iff$ $\exists u^{-1}$ s.t. $uu^{-1} = u^{-1}u = 1$ $\iff$ $(u) = (1)$.

2. a **divides** b $\iff$ $a \mid b$ $\iff$ $b = aq$ for some $q \in R$ $\iff$ $b \in (a)$ $\iff$ $(b) \subset (a)$.

3. a and b are **associates** $\iff$ each divides the other $\iff$ $b = ua$ with u a unit (for $b = ua \Rightarrow u^{-1}b = a$; $a = ub \Rightarrow u^{-1}a = b$) $\iff$ $(a) = (b)$ (for $a \mid b \Rightarrow (b) \subset (a)$; $b \mid a \to (a) \subset (b)$.)

4. $0 \neq a$ is **irreducible** $\iff$ a is not a unit, and $a = xy \Rightarrow x =$ unit or $y =$ unit.

5. $0 \neq a$ is **prime** $\iff$ the principal ideal (a) generated by this *nonzero* a is a prime ideal (an ideal $I \subsetneq R$ is prime if $ab \in I$ implies $a \in I$ or $b \in I$.) $\iff$ a is not a unit and $a \mid bc$ implies $a \mid b$ or $a \mid c$ $\iff$ a is not a unit and $bc \in (a)$ implies $b \in (a)$ or $c \in (a)$.

Remark 2.3.15. We want to emphasize that zero cannot be a prime element but (0) is a prime ideal iff R is an integral domain (see Corollary 2.3.19).

2.3.4 Prime Ideals

Definition 2.3.16. If $I \subsetneq R$ is an ideal, we say I is **prime** if $ab \in I \implies a \in I$ or $b \in I$ for $a, b \in R$.

Example 2.3.17. $R = \mathbb{Z}$. Since subgroups of $\mathbb{Z}$ are all of the form $m\mathbb{Z}$, and an ideal is first an additive subgroup of $\mathbb{Z}$, we see that ideals in $\mathbb{Z}$ are of the form $m\mathbb{Z}$ (since each $m\mathbb{Z}$ has $\forall z \in \mathbb{Z}, a \in m\mathbb{Z}$, $za \in m\mathbb{Z}$). Now, we also claim that for positive m, $m\mathbb{Z}$ is a prime ideal iff m is a prime number. Since $m\mathbb{Z} = (-m)\mathbb{Z}$, we then see that each prime ideal in $\mathbb{Z}$ is of the form $(\pm p)\mathbb{Z}$ or (0) (it will be proved at the end of this subsection that the zero ideal is prime iff the ring is an integral domain).

Proof. Let m be positive.

$\implies$: If $m = ab$, and $a, b > 1$, then $ab = m \in m\mathbb{Z}$, but $a, b \notin m\mathbb{Z}$. Contradiction.

$\impliedby$: To show $m\mathbb{Z}$ is prime, we suppose $ab \in m\mathbb{Z} = (m)$, then subsection 2.3.3 2 shows that $m \mid ab$. m being prime number then implies $m \mid a$ or $m \mid b$. □

Proposition 2.3.18.

1. Every maximal ideal is prime.
2. $I \subsetneq R$ is prime $\iff$ R/I is an integral domain.

3. P is a prime ideal $\iff$ $IJ \subseteq P$ implies $I \subseteq P$ or $J \subseteq P$ for ideals $I, J \subseteq R$.

Proof (1): If M is maximal and $ab \in M$ and $a \notin M$, then the ideal generated by a, M, $(a, M) := \{ra + m, m \in M, r \in R\}$ is an ideal where $M \subsetneq \langle a, M\rangle \subset R$. Then $\langle a, M\rangle = R$ since M maximal, so $1 = ra + m$ for some $r \in R, m \in M \implies b = rab + mb$, so $b \in M$. □

Proof (2): $\implies$: If $(a + I)(b + I) = 0$, then $ab + I = 0$, so $ab \in I \implies a \in I$ or $b \in I$, so $a + I = \bar{0}$ or $b + I = \bar{0}$, where $\bar{0}$ is the zero of R/I.

$\impliedby$: If $ab \in I$, then $(a + I)(b + I) = \bar{0}$, so $a + I = \bar{0}$ or $b + I = \bar{0}$, so $a \in I$ or $b \in I$. □

Proof (3): If P is prime and $IJ \subseteq P$ but $I \subsetneq P$ and $J \subsetneq P$, then pick $a \in I \backslash P$ and $b \in J \backslash P$, then $ab \in IJ$ but $ab \notin P$, a contradiction.

Conversely, assume $IJ \subseteq P$ implies $I \subseteq P$ or $J \subseteq P$ for ideals $I, J \subseteq R$. Let $I = (a) = \{ra \mid r \in R\}$ and $J = (b) = \{rb \mid r \in R\}$. Then $IJ = (ab)$ (check this). So $IJ \subseteq P$, so $a \in I \subseteq P$ or $b \in J \subseteq P$, so $a \in P$ or $b \in P$. □

Corollary 2.3.19. $\{0\}$ is a prime ideal $\iff$ R is an integral domain.

Example 2.3.20. $m\mathbb{Z} \subseteq \mathbb{Z}$ is prime $\iff$ $m\mathbb{Z}$ is maximal $\iff$ m is prime.

Proof. Due to Proposition 2.3.18 and Example 2.3.17, we only need to show that $m\mathbb{Z}$ prime implies $m\mathbb{Z}$ maximal.

To show $m\mathbb{Z}$ is maximal, we suppose $m\mathbb{Z} \subset n\mathbb{Z}$, i.e., $(m) \subset (n)$. By subsection 2.3.3, this is equivalent to $n \mid m$. But m is prime so either $n = 1$ or $n = m$. □

2.3 EXERCISES

1. Show that in every finite commutative ring, every prime ideal is maximal.
2. A proper ideal I of R is said the be a **primary ideal** if $ab \in I$ implies $a \in I$ or $b^n \in I$ for some positive integer n.

i. Find all the primary ideals of $\mathbb{Z}$.
ii. Show that if I is a primary ideal, then $\sqrt{I}$ is a prime ideal.

2.4 Product of Rings

Theorem 2.4.1 (Chinese Remainder Theorem)**.** For $0 < m_1, ..., m_n \in \mathbb{Z}, gcd(m_i, m_j) = 1$, then for any $r_1, ..., r_n \in \mathbb{Z}$, the system of equations

$$\begin{cases} x \equiv r_1 \pmod{m_1} \\ \cdots \\ x \equiv r_n \pmod{m_n} \end{cases}$$

has a solution.

Theorem 2.4.2 ((Generalized) Chinese Remainder Theorem)**.** R commutaitve ring. Let $I_1, ..., I_n$, $n \geqslant 2$ be ideals in R such that $I_i + I_j = R$ for every $i, j, i \neq j$. Then for any $r_1, ..., r_n \in R$, there is $x \in R$ s.t. $x - r_i \in I_i \ \forall 1 \leqslant i \leqslant n$.

Remark 2.4.3. We first see it is indeed a generalization:

$$\begin{aligned}\gcd(m_i, m_j) = 1 &\iff 1 = xm_i + ym_j \text{ for } x, y \in \mathbb{Z} \\ &\iff 1 \in \langle m_i \rangle + \langle m_j \rangle \\ &\iff \mathbb{Z} = \langle m_i \rangle + \langle m_j \rangle\end{aligned}$$

proof of the Generalized Chinese Remainder Theorem.
Proceed with induction on n: If $n = 2, I_1 + I_2 = R \implies \exists a_i \in I_i$ s.t. $a_1 + a_2 = 1$. Then let $x = r_1a_1 + r_2a_1$, then $x - r_1 = r_1(a_2 - 1) + r_2a_1 = -r_1a_1 + r_2a_1 \in I_1$. Similar for $x - r_2$.

$n - 1 \implies n$: For $I_1, ..., I_n$, let $J = I_2 \cdots I_n$. Claim: $I + J = R$.

So for $I_1 + I_i = R \forall i \geqslant 2, \exists a_i \in I_1, b_i \in I_i$ s.t. $a_i + b_i = 1 \implies 1 = \prod_{i=2}^n (a_i + b_i) = I_1 + J$. By case 2 of the theorem, $\exists y_1 \in R$ s.t. $y_1 - 1 \in I_1, y_1 - 0 \in J \implies y_1 \in I_2 \cdots I_n$. In a similar way, $\forall 1 \leqslant i \leqslant n$, we find $y_i \in R$ s.t. $y_i - 1 \in I_i$ and $y_i = I_1 \cdots \hat{I_i} \cdot I_n \subseteq I_j \forall j \neq i$. Note that $I \cap J \subseteq IJ$.

Let $x = r_1y_1 + ... + r_ny_n$. Then $x - r_i = r_1y_1 + \cdots r_i(y_i - 1) + \cdots r_ny_n$. Every y_i is in I_i, so this entire expression is in I_i. □

Definition 2.4.4. Let R, S be rings, then **product** of R and S is

$$R \times S = \{(r, s) \,|\, r \in R, s \in S\}$$

where $(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$. and $(r_1, s_1)(r_2, s_2) = (r_1r_2, s_1s_2)$. Its additive identity is $(0, 0)$. Its unity is $(1, 1)$. One can define more general product of rings just like that for groups.

Corollary 2.4.5. If $I_1, ..., I_n$ are ideals of R such that $I_i + I_j = R$ for $i \neq j$. Then

$$\frac{R}{\bigcap_{i=1}^n I_n} \simeq \prod_{i=1}^n R/I_i$$

as isomorphism of rings.

Proof. Define $\phi : R \to \prod_{i=1}^n R/I_i$ by $\phi(r) = (r + I_1, ..., r + I_n)$. ϕ is a ring homomorphism. $Ker(\phi) = \cap_{i=1}^n I_i$.

ϕ surjective: $\forall (r_1 + I_1, ..., r_n + I_n) \in \prod_{i=1}^n R/I_i$, by the Chinese remainder theorem, $\exists x \in R$ s.t. $x + I_i = r_i + I_i$, so by the first isomorphism theorem, we get the result. □

Example 2.4.6. If $R = \mathbb{Z}$, and prime factorization $m = p_1^{r_1} \cdots p_n^{r_n}, I_i = p_i^{r_i}\mathbb{Z}$. Then note that $I_i = p_i^{r_i}\mathbb{Z}, I_i + I_j = \mathbb{Z}$ because $p_i^{r_i}$ and $p_j^{r_j}$ coprimes, which implies $1 = xp_i^{r_i} + yp_j^{r_j} \in I_i + I_j$. Also, $\bigcap_{i=1}^n I_i = m\mathbb{Z}$ because $a \in \bigcap_{i=1}^n I_i \iff a$ is a multiple of all $p_i^{r_i} \iff a$ is a multiple of m. So,

$$\mathbb{Z}/m\mathbb{Z} \simeq \prod_{i=1}^n \mathbb{Z}/p_i^{r_i}\mathbb{Z}$$

as rings. That is,

$$\mathbb{Z}_m \simeq \prod_{i=1}^n \mathbb{Z}_{p_i}^{r_i}$$

as rings. This is a stronger result than the group version.

2.4 EXERCISES

1. [3] Ex7.6-3. Let R and S be rings with identities. Prove that every ideal of $R \times S$ is of the form $I \times J$ where I is an ideal of R and J is an ideal of S.

2. [3] Ex7.6-6. Let $f_1(x), f_2(x), \ldots, f_k(x)$ be polynomials with integer coefficients of the same degree d. Let $n_1, n_2, \ldots, n_k$ be integers which are relatively prime in pairs (i.e., $(n_i, n_j) = 1$ for all $i \neq j$). Use the Chinese Remainder Theorem to prove there exists a polynomial $f(x)$ with integer coefficients and of degree d with

$$f(x) \equiv f_1(x) \bmod n_1, \quad f(x) \equiv f_2(x) \bmod n_2, \quad \ldots, \quad f(x) \equiv f_k(x) \bmod n_k$$

i.e., the coefficients of $f(x)$ agree with the coefficients of $f_i(x) \bmod n_i$. Show that if all the $f_i(x)$ are monic, then $f(x)$ may also be chosen monic. [Apply the Chinese Remainder Theorem in $\mathbb{Z}$ to each of the coefficients separately.]

3. [1] p.378 Ex1.5. Let a and b be relatively prime integers. Prove that there are integers m and n such th $a^m + b^n \equiv 1$ modulo ab.

2.5 Localization

Suppose R is an integral domain. Consider the equivalence relation $\frac{a}{b} \sim \frac{c}{d} \iff ad = bc$. Then, we can mod out by equivalence relationship to get the set of all equivalence classes

$$\left\{ \frac{a}{b} \,\middle|\, a, b \in R, b \neq 0 \right\} / \sim$$

Then we define the ring structure such that for $b, d \neq 0$, $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, $\frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$. There are well-defined. The unity is $\frac{1}{1}$, and the zero is $\frac{0}{1}$. This is a commutative ring as R is commutaitve. Any non-zero element $\frac{a}{b}$ (i.e., $a, b \neq 0$ bc. $\frac{a}{b} = \frac{0}{1} \Leftrightarrow a = 0$) has a multiplicative inverse $\frac{b}{a}$. Thus we get a field, namely the **field of fraction**, or **Quotient field** of R. We will generalize this construction below.

Definition 2.5.1. Suppose R is a commutative ring. Then $S \subset R$ is a **multiplicative subset** if $1 \in S$, $0 \notin S$, and $a, b \in S \implies ab \in S$.

Example 2.5.2.

- For $0 \neq r \in R$, $S = \{1, r, r^2, ...\}$
- $P \subsetneq R$ be a prime ideal and $S = R \backslash P$. Then $a, b \notin P \implies ab \notin P$. Observe that

$$\begin{aligned} P \text{ prime} &\iff (ab \in P \Rightarrow a \in P \text{ or } b \in P) \\ &\iff (a, b \notin P \Rightarrow ab \notin P) \\ &\iff (a, b \in S - P \Rightarrow ab \in S - P) \end{aligned}$$

  $1 \in S - P$ becasue $P \subsetneq R$ (if $1 \notin S - P$, then $1 \in P$ and $P = R$).

Definition 2.5.3. Define $S^{-1}R = \{(r, s) \,|\, r \in R, s \in S\} / \sim$ with the equivalence relationship $(r, s) \simeq (r', s') \iff \exists s'' \in S$ s.t. $s''(rs' - sr') = 0$.

If $0 \in S$, then $(r, s) \simeq (0, 0)$, and everything is in a single equivalence class. That's the reason why we assume $0 \notin S$.

Proposition 2.5.4. $S^{-1}R$ is a commutaitve ring with the operations

$$\frac{r}{s} + \frac{r'}{s'} = \frac{rs' + r's}{ss'}, \quad \frac{r}{s}\frac{r'}{s'} = \frac{rr'}{ss'}.$$

Zero is $\frac{0}{1}$. Unity is $\frac{1}{1}$.

Proof. Addition is well-defined: if $\frac{r}{s} = \frac{r_0}{s_0}$, then $\exists s'', s''(rs_0 - r_0 s) = 0$. Want to check that $\frac{r}{s} + \frac{r'}{s'} = \frac{r_0}{s_0} + \frac{r'}{s'}$. Equivalently,

$$\frac{rs' + r's}{ss'} = \frac{r_0 s' + r' s_0}{s_0 s'} = 0 \iff s''(s_0 s'(rs' + r's) - ss'(r_0 s' + r' s_0)) = s''(\underbrace{s'^2(s_0 r - s r_0)}_{=0} + \underbrace{s_0 s' r' s - s s' r' s_0}_{=0}) = 0.$$

Multiplication's well-definedness is easy to prove. The remaining is left as an exercise. □

There is a natural ring homomorphism defined by

$$\phi : R \to S^{-1}R; r \mapsto \frac{r}{1}.$$

In particular if R is an integral domain, then ϕ is injective:

$$\frac{r}{1} = \frac{r'}{1} \Leftrightarrow \exists s'' \in S \text{ s.t. } s''(r \cdot 1 - 1 \cdot r') = s''(r - r') = 0 \overset{R \text{ int. dom., } 0 \notin S}{\iff} r - r' = 0, r = r'$$

We now see how $S^{-1}R$ generalizes K = field of fractions of R:

$$\frac{r}{s} \sim_{S^{-1}R} \frac{r'}{s} \Leftrightarrow \exists s'' \in S \text{ s.t. } s''(rs' - sr') = s''(r - r') = 0 \overset{R \text{ int. dom., } 0 \notin S}{\iff} rs' - sr' = 0, rs' = sr' \Leftrightarrow \frac{r}{s} \sim_K \frac{r'}{s'}$$

Thus, $S^{-1}R$ is a subring of the field of fractions of R, which we can write as $R \subset S^{-1}R \subset K$, where the first is by the injection $r \to r/1$ and the second is by the inclusion above.

Note that $\phi : R \to S^{-1}R$ also has the property that $\phi(s)$ is invertible for any $s \in S$. Namely $\forall s \in S, \phi(s) = \frac{s}{1}$, so $\frac{s}{1}\frac{1}{s} = \frac{1}{1}$. And if $\psi : R \to R'$ is a ring homomorphism such that $\psi(s)$ invertible in R', then $\exists! f : S^{-1}R \to R'$ such that $f \circ \phi = \psi$

$$\begin{array}{ccc} R & \xrightarrow{\psi} & R' \\ & {\scriptstyle\phi}\searrow \quad \nearrow{\scriptstyle f} & \\ & S^{-1}R & \end{array}$$

$$1 = f(1/1) = f\left(\phi(s)(1/s)\right) = \psi(1) f(1/s) \Rightarrow \psi(s)^{-1} = f(1/s)$$

Example 2.5.5. Assume R is an integral domain.

- If $S = R \backslash \{0\}$, then $S^{-1}R$ is the field of fractions of R.
- If $S = \{1, f, f^2, ..., \}$ where $f \in R$ s.t. $\forall n, f^n \neq 0$. Then

$$R_f = S^{-1}R = \left\{ \frac{a}{f^r} \middle| a \in R, r \geqslant 0 \right\}.$$

- If $P \subsetneq R$ is a prime ideal and $S = R \backslash P$. Then

$$R_P = S^{-1}R = \left\{ \frac{a}{b} \middle| a, b \in R, b \notin P \right\}$$

  R_P is a **local ring**. i.e. it has a unique maximal ideal, which is

$$I = \left\{ \frac{a}{b} \middle| a, b \in R, b \notin P, a \in P \right\}.$$

  It is easy to see I is an ideal. We show it is maximal. Notice that for $\frac{a}{b} + I \in R_P/I$, we have $(\frac{a}{b} + I)(\frac{b}{a} + I) = I$ as long as $a \notin P$. When $a \in P$, $\frac{a}{b} + I = I$. Thus, every nonzero element of R_P/I is a unit. By Corollary 2.3.12, I is maximal.

2.5 EXERCISES

Some useful sources: link.

1. Prove the following claims:

i. The preimage of a prime ideal under a ring homomorphism is a prime ideal.
ii. The preimage of a proper ideal under a surjective ring homomoprhism is a proper ideal.

2. (Correspondence theorem for localization of rings) Let S a multiplicative subset of R not containing 0 , and let $\phi : R \to S^{-1}R$ be the map $\phi(r) = \frac{r}{1}$. For an ideal I in R, let

$$S^{-1}I = \left\{ \frac{i}{s} \,\middle|\, i \in I, s \in S \right\} \subset S^{-1}R.$$

i. Show that $S^{-1}I$ is an ideal of $S^{-1}R$, and $S^{-1}\phi^{-1}(J) = J$ for any ideal J of $S^{-1}R$.
ii. Show the map $P \mapsto S^{-1}P$ gives a one-to-one correspondence between prime ideals of R whose intersection with S is empty and prime ideals of $S^{-1}R$.

3. Let R be a PID and S a multiplicative subset not containing 0 . Show $S^{-1}R$ is a PID.

2.6 PIDs

Definition 2.6.1. Recall from definition 2.2.3 of principal ideal. We say an integral domain R in which every ideal is principal ideal is called a **principal ideal domain**.

Example 2.6.2.

- $\mathbb{Z}$ is PID. Every ideal in $\mathbb{Z}$ is of the form $n\mathbb{Z} = (n)$.
- $\mathbb{R}[x]$ is a PID. If $I \neq \{0\}$ is an ideal and $0 \neq f(x) \in I$ has the smallest degree, then $I = (f)$. If $g \in I$, dividing g by f gives that $g(x) = q(x)f(x) + r(x)$. So $r(x) = 0$ or $\deg(r) < \deg(f)$. Since $r(x) = g(x) - q(x)f(x) \in I$ and f is chosen to have smallest degree, we see $\deg r(x) \geqslant \deg f(x) \implies r = 0 \implies g \in (f)$.
- $\mathbb{R}[x, y]$ is not a PID. $(x, y) = \{f(x, y) \,|\, f(0, 0) = 0\}$ not principal.
- $\mathbb{Z}[x]$ is not a PID. $(x, 2) = \{f(x) \,|\, f(0) \text{ is even}\}$ not principal: $2 \in (x, 2)$. Thus, if $(x, 2) = (a)$, then $\exists r \in \mathbb{Z}[x]$ s.t. $ra = 2$. Either $r = 1, a = 2$ or $r = 2, a = 1$ since the RHS already has the smallest possible degree. $1 \notin (x, 2)$, so $a = 2$. Contradiction.

Definition 2.6.3. Recall from subsection 2.3.3 that

- $a \in R$ is **prime** if (a) is a prime ideal. Equivalently, $a \,|\, bc \implies a \,|\, b$ or $a \,|\, c$.
- $0 \neq a \in R$ is **irreducible** if it is not a unit and if $a = xy$, then x is a unit or y is a unit.

Proposition 2.6.4. If R is an integral domain, a prime element is irreducible.

Proof. If a is prime (so $a \neq 0$) and $a = xy$, then $a \,|\, x$ or $a \,|\, y$, so $x = ax'$ or $y = ay' \implies a = ax'y$ or $a = xay' \implies a(1 - x'y) = 0$ or $a(1 - xy') = 0 \xrightarrow{R \text{ int. dom.}} 1 = x'y$ or $1 = xy'$, so y is a unit or x is a unit. □

Example 2.6.5. Let

$$R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \,|\, a, b \in \mathbb{Z}\} \subseteq \mathbb{C}$$

. It is clear that this is a ring. We claim that

- units of R are ± 1.
- $2, 3, 1 \pm \sqrt{-5}, 2 \pm \sqrt{-5}$ are all irreducibles.
- $3 \in R$ is not prime, so R is not a principal ideal domain due to proposition 2.6.6.
- R is not a unique factorization domain (defn. 2.7.2).

Proof. We first show that the units of R are ± 1. Suppose

$$(a + b\sqrt{-5})(c + d\sqrt{-5}) = 1 = \text{unity}$$

Then

$$\begin{aligned} \text{square it} &\implies (\underbrace{a^2 + 5b^2}_{\in \mathbb{Z}})(\underbrace{c^2 + 5d^2}_{\in \mathbb{Z}}) = 1. \\ &\implies b \text{ and } d \text{ must be } 0 \text{ because in } \mathbb{Z} \text{ unites are } \pm 1. \\ &\implies a, c = \pm 1. \end{aligned}$$

We use the following way to show irredciblility:
We show 3 is irredciblle for example. Suppose not, then

$$3 = xy = (a + b\sqrt{-5})(c + d\sqrt{-5}) \implies 9 = (a^2 + 5b^2)(c^2 + 5d^2)$$

We note that $a^2 + 5b^2, c^2 + 5d^2 \in \{0, 1, 4, 5, 9, 16, \cdots\}$ and their multiplicaiton belongs to $\{0, 1, 4, 5, 9, 16, \cdots\}$, where we note that 0 is obtained by 0×0 or $0 \times$ other, $1, 4, 5, 9$ are obtained by multiplication of 1 with $1, 4, 5, 9$ (so it has to be the case that $a + b\sqrt{-5}$ or $c + d\sqrt{-5}$ is a unit), and 16 is obtained by either 1×16, 16×1, or 4×4. Therefore, 3 is irreducible. In fact, we also proved that 2 is irreducible as $2^2 = 4$.

We show that 3 is not prime:
$9 = (2 + \sqrt{-5})(2 - \sqrt{-5})$ is in $I = (3)$, but $3 \nmid (2 + \sqrt{-5})$ and $3 \nmid 2 - \sqrt{-5}$ since $2 + \sqrt{5} \neq 3(a + b\sqrt{-5})$, for $a, b \in \mathbb{Z}$.

R is not UFD:
That is to say, there are some elements of R that can be written in products that are irreducibles that are not associates. 6 and 9 do:

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$
$$3 \cdot 3 = 9 = (2 + \sqrt{-5})(2 - \sqrt{-5})$$

They are not associates simply because the units are ± 1. □

Proposition 2.6.6. If R is a PID, then irreducible $\implies$ prime.

Proof. Suppose $a \in R$ is irreducible, then it suffices to show that a is a prime ideal. Then the ideal generated by a, $(a) \neq R$ since a is not a unit. So there is a maximal ideal M where $(a) \subseteq M \subsetneq R$.

Since R is a PID, $M = (b)$ for some $b \implies (a) \subseteq (b) \implies a = bc$ for some $\mathbb{C}$. $(b) \neq R$ so b is not a unit. Since a irreducible, $\mathbb{C}$ has to be a unit. So $b = c^{-1}a \implies b \in (a) \implies (b) \subseteq (a)$, so $(a) = (b)$, so (a) maximal and therefore prime. □

Proposition 2.6.7. Every prime ideal is maximal in a PID.

Proof. If $I = (a)$ prime, then $(a) \subseteq M \subsetneq R$ where M is maximal, then let $M = (b) \implies a \in (b) \implies a = bc$. a is prime so it is irreducible, so $\mathbb{C}$ is a unit. So $b \in (a) \implies (a) = (b) \implies (a)$ maximal. □

2.6 EXERCISES

1. Show that $\mathbf{Z}[\sqrt{-5}]$ is not a PID by finding a non-principal ideal.
2. Show the subring $\mathbf{Z}[2i] = \{a + 2bi \mid a, b \in \mathbf{Z}\}$ of the Guassian integers $\mathbf{Z}[i]$ is not a UFD by showing $4 = 2 \cdot 2 = (-2i) \cdot (2i)$ gives two factorization of 4 into product of irreducible elements.

2.7 UFDs and GCDs

We collect some obvious observations and talk about UFDs and GCDs.

Theorem 2.7.1. Let the elements $a, b, c \in R$. Then,

(1) $a \mid 0, 1 \mid a, a \mid a$;

(2) $a \mid 1$ if and only if a is invertible;

(3) if $a \mid b$, then $ac \mid bc$;

(4) if $a \mid b$ and $b \mid c$, then $a \mid c$;

(5) if $c \mid a$ and $c \mid b$, then $c \mid (ax + by)$ for every $x, y \in R$.

Definition 2.7.2. A **unique factorization domain** (UFD) is defined to be an integral domain R in which every non-zero element x of R can be written as a product of a unit u and zero or more irreducible elements p_i of R :

$$x = up_1p_2 \cdots p_n \text{ with } n \geqslant 0$$

and this representation is "unique up to associates and units" in the following sense: if $q_1, \ldots, q_m$ are irreducible elements of R and w is a unit such that

$$x = wq_1q_2 \cdots q_m \text{ with } m \geqslant 0,$$

then $m = n$, and there exists a bijective map $\phi : \{1, \ldots, n\} \to \{1, \ldots, m\}$ such that p_i is associated to $q_{\phi(i)}$ for $i \in \{1, \ldots, n\}$. We note that if there are multiple units in the decomposition, they are first combined to give a single unit by commutativity. We will later in our writing assume that the decomposition into irreducibles $p_1 \cdots p_n$ already include a unit in it if any.

Remark 2.7.3. The condition

$$up_1 \cdots p_n = wq_1 \cdots q_k \Rightarrow k = n, \text{ and } p_i, q_j \text{ associate}$$

is equivalent to the condition

$$p_1 \cdots p_n = q_1 \cdots q_k \Rightarrow k = n, \text{ and } p_i, q_j \text{ associate}$$

Proof.
One should first notice that an irreducible is not a unit by definition. $\Longrightarrow$ direction is direct.
$\Longleftarrow$:

$$\begin{aligned} up_1 \cdots p_n = x &= wq_1 \cdots q_k \\ p_1 \cdots p_m = xu^{-1} = (wq_1 \cdots q_k)w^{-1} &= \underbrace{[wu^{-1}q_1]}_{\text{irreducible}} q_2 \cdots q_k \end{aligned}$$

where $wu^{-1}q_1$ is irreducible becasue

$$wu^{-1}q_1 = xy \Rightarrow \underbrace{q_1}_{\text{irreducible}} = [wu^{-1}x]y \implies wu^{-1}x \text{ a unit or } y \text{ a unit} \implies x \text{ unit or } y \text{ unit}$$

□

We will use this second condition for UFD for now on.

Example 2.7.4. For $\mathbb{Z}$, the units are ± 1. Prime elements are $\{\pm p \mid p \text{ prime}\}$ (see example 2.3.17). $\mathbb{Z}$ is UFD.

Example 2.7.5. $\mathbb{Z}[\sqrt{-5}]$ is not a UFD (see example 2.6.5).

Proposition 2.7.6. Integral Domain R is a UFD if and only if

(1) Every irreducible element is prime.

(2) **Ascedning chain condition on principal ideals** (AACP): R satisfies the ascending chain condition for principle ideals. Namely, if we have

$$(a_1) \subset (a_2) \subseteq \cdots \subseteq (a_m) \subseteq \cdots,$$

then $\exists n$ s.t. $(a_n) = (a_{n+1}) = \cdots$. That is, R does not contain an infinite strictly increasing chain of principal ideals.

Proof.
$\implies$: First assume R is a UFD.

(1). If $a \in R$ irreducible and $a \mid bc$, so for $bc = ax$, write b, c, x as a product of irreducible elements, where $b = q_1 \cdots q_l, c = y_1 \cdots y_t, x = x_1 \cdots x_k$. So $bc = ax \implies q_1 \cdots q_l y_1 \cdots y_t = a x_1 \cdots x_k$. Since R UFD, $\exists q_i$ or y_i associate to a. Assume WLOG $uq_i = a$ for a unit u, so $u^{-1}a = q_i \mid b \implies b = b'u^{-1}a \implies a \mid b$.

(2). $(a) \subseteq (b) \iff b \mid a$. If $(a) \subsetneq (b)$, then $a = bc$, where $\mathbb{C}$ is a non-unit. So the number of irreducible factors of $b <$ the number of irreducible factors of a, so there cannot be infinitely many strict inclusion in the chain.

$\impliedby$: Assume (1) and (2) holds. Suppose an element a factors in two ways into irreducible elements, say $p_1 \cdots p_m = a = q_1 \cdots q_n$, where $m \leqslant n$. If $n = 1$, then $m = 1$ and $p_1 = q_1$. Suppose that $n > 1$. Since p_1 is prime, it divides one of the factors $q_1, \ldots, q_n$, say q_1. Since q_1 is irreducible and since p_1 is not a unit, q_1 and p_1 are associates, say $p_1 = uq_1$, where u is a unit. We move the unit factor over to q_2, replacing q_1 by uq_1 and q_2 by $u^{-1}q_2$. The result is that now $p_1 = q_1$. Then we cancel p_1 and use induction on n.

Uniqueness: Suppose $a = x_1 \cdots x_n = y_1 \cdots y_m$, where x_i, y_j irreducible. Then $y_1 \mid x_1 \cdots x_n$ and y_i prime $\implies y_1 \mid x_i$ for some i. So, $x_i = uy_1$ and x_i irreducible $\implies u$ is a unit, so y_1, x_i associates. □

Theorem 2.7.7. Every PID is a UFD.

Proof. (1) It is proved that every irreducible element is prime in proposition 2.6.6.

(2) If $(a_1) \subset (a_2) \subset \cdots$. Let $I = \bigcup (a_i)$, then it is easy to see I is an ideal. Since R is a PID, we have $I = (b)$. Since $b \in I, \exists i$ s.t. $b \in (a_i)$, so $(b) \subseteq (a_i)$. But $(a_i) \subseteq (b)$, so $(a_i) = (b)$, so $(a_i) = (a_{i+1}) = (a_{i+1}) = \ldots$. □

Definition 2.7.8. If R is an integral domain and $a, b \in R$. Then d is the **greatest common divisor** of a, b if

- $d \mid a$ and $d \mid b$.
- If $d' \mid a$ and $d' \mid b$, then $d' \mid d$

Any two greatest common divisors d and d' are associate elements. The first condition tells us that both d and d' divide a and b, and then the second one tells us that d' divides d and also that d divides d'.

However, a greatest common divisor may not exist. There will often be a common divisor m that is maximal, meaning that a/m and b/m have no proper divisor in common. But this element may fail to satisfy condition (b). For instance, in the ring $\mathbb{Z}[\sqrt{-5}]$ considered above (12.2.3), the elements $a = 6$ and $b = 2 + 2\sqrt{-5}$ are divisible both by 2 and by $1 + \sqrt{-5}$. These are maximal elements among common divisors, but neither one divides the other.

One case in which a greatest common divisor does exist is that a and b have no common factors except units. Then 1 is a greatest common divisor. When this is so, a and b are said to be relatively prime.

We call an integral domain in which any two non-zero elements have a greatest common divisor a GCD domain. We show that UFDs are GCDs:

Proposition 2.7.9 (UFD is GCD)**.** Let $x, y \in R\backslash\{0\}$ for UFD R. Factor x and y into pairwise non-associated irreducible elements:

$$x = p_1^{e_1} \cdots p_n^{e_n},$$
$$y = p_1^{f_1} \cdots p_n^{f_n}.$$

Then one can check that the product $p_1^{r_1} \cdots p_n^{r_n}$ with $r_i := \min\{e_i, f_i\}$ is a greatest common divisor of x and y.

We can generalize the notion of gcd for more elements:

Definition 2.7.10. Let $a_1, a_2, \ldots, a_n$ be nonzero elements of the ring R. An element $d \in R$ is a **greatest common divisor** of $a_1, a_2, \ldots, a_n$ if it possesses the properties

(1) $d \mid a_i$ for $i = 1, 2, \ldots, n$ (d is a common divisor),

(2) $c \mid a_i$ for $i = 1, 2, \ldots, n$ implies that $c \mid d$.

Remark 2.7.11. GCDs can be safely deifned as the rings where any *finite* number of nonzero elements of R admit a greatest common divisor. Just notice that $\gcd(a_1, \cdots, a_{k+1}) = \gcd(\gcd(a_1, \cdots, a_k), a_{k-1})$.

Remark 2.7.12 (gcd unique up to associates)**.** A natural question to ask is whether the elements $a_1, a_2, \ldots, a_n \in R$ can possess two different greatest common divisors. For an answer, suppose that there are two elements d and d' in R satisfying the conditions of Definition refgcd defn.. Then, by (2), we must have $d \mid d'$ as well as $d' \mid d$; according to subsection 2.3.3, this implies that d and d' are associates. Thus, the greatest common divisor of $a_1, a_2, \ldots, a_n$ is unique, whenever it exists, up to arbitrary invertible factors.

We shall find it convenient to denote any greatest common divisor of $a_1, a_2, \ldots, a_n$ by $\gcd(a_1, a_2, \ldots, a_n)$. The next theorem will prove that greatest common divisor of any finite set of nonzero elements can be exprssed as a linear combination. We will first give an example where this fails in UFD:

Example 2.7.13. We give an example of gcd of two elements in a UFD that is not expressible into a linear combination. Let's consider $R[x, y]$. The gcd of x^2y and xy^2 is xy so we are looking for $a, b \in R[x, y]$ such that $ax^2y + bxy^2 = xy \Rightarrow ax + by = 1$. But the constant term of both ax and by is 0 , so the constant term of their sum is also zero. Contradiction.

Theorem 2.7.14. Let $a_1, a_2, \ldots, a_n$ be nonzero elements of the ring R. Then $a_1, a_2, \ldots, a_n$ have a greatest common divisor d, expressible in the form

$$d = r_1a_1 + r_2a_2 + \cdots + r_na_n \quad (r_i \in R),$$

if and only if the ideal $(a_1, a_2, \ldots, a_n)$ is principal.

Proof. Suppose that $d = \gcd(a_1, a_2, \ldots, a_n)$ exists and can be written in the form $d = r_1a_1 + r_2a_2 + \cdots + r_na_n$, with $r_i \in R$. Then the element d lies in the ideal $(a_1, a_2, \ldots, a_n)$, which implies that $(d) \subseteq (a_1, a_2, \ldots, a_n)$.

To obtain the reverse inclusion, observe that since $d = \gcd(a_1, a_2, \ldots, a_n)$, each a_i is a multiple of d; say, $a_i = x_i d$, where $x_i \in R$. Thus, for an arbitrary member $y_1a_1 + y_2a_2 + \cdots + y_na_n$ of the ideal $(a_1, a_2, \ldots, a_n)$, we must have

$$y_1a_1 + y_2a_2 + \cdots + y_na_n = (y_1x_1 + y_2x_2 + \cdots + y_nx_n)\, d \in (d).$$

This fact shows that $(a_1, a_2, \ldots, a_n) \subseteq (d)$, and equality follows. For the converse, let $(a_1, a_2, \ldots, a_n)$ be a principal ideal of R :

$$(a_1, a_2, \ldots, a_n) = (d) \quad (d \in R).$$

Our aim, of course, is to prove that $d = \gcd(a_1, a_2, \ldots, a_n)$. Since each $a_1 \in (d)$, there exist elements b_i in R for which $a_i = b_i d$, whence $d \mid a_i$ for $i = 1, 2, \ldots, n$. It remains only to establish that any common divisor $\mathbb{C}$ of the a_i also divides d. Now, $a_i = s_i c$ for suitable $s_i \in R$. As an element of $(a_1, a_2, \ldots, a_n)$, d must have the form $d = r_1a_1 + r_2a_2 + \cdots + r_na_n$, with r_i in R. This means that

$$d = (r_1s_1 + r_2s_2 + \cdots + r_ns_n)\, c,$$

which is to say that $c \mid d$. Thus, d is a greatest common divisor of $a_1, a_2, \ldots, a_n$ and has the desired representation. □

When $(a_1, a_2, \ldots, a_n) = R$, the elements $a_1, a_2, \ldots, a_n$ must have a common divisor which is an invertible element of R; in this case, we say that $a_1, a_2, \ldots, a_n$ are relatively prime and shall write $\gcd(a_1, a_2, \ldots, a_n) = 1$.

If $a_1, a_2, \ldots, a_n$ are nonzero elements of a principal ideal ring R, then the theorem tells us that $a_1, a_2, \ldots, a_n$ are relatively prime if and only if there exist $r_1, r_2, \ldots, r_n \in R$ such that

$$r_1a_1 + r_2a_2 + \cdots + r_na_n = 1 \quad \text{(Bezout's Identity)}.$$

Proposition 2.7.15. Let a, b, c be elements of the principal ideal ring $\boldsymbol{R}$. $c \mid ab$, with a and $\mathbb{C}$ relatively prime, then $c \mid b$.

Proof. Since a and $\mathbb{C}$ are relatively prime, so that $\gcd(a, c) = 1$, there exint elements $r, s \in R$ satisfying $1 = ra + sc$; hence,

$$b = 1b = rab + scb.$$

As $c \mid ab$ and $c \mid c$, Theorem 2.7.1 (5) guarantees that $c \mid (rab + scb)$, or rather, $c \mid b$. □

Dual to the notion of greatest common divisor there is the idea of a least common multiple, defined below.

Definition 2.7.16. Let $a_1, a_2, \ldots, a_n$ be nonzero elements of a ring R. An element $d \in R$ is a **least common multiple** of $a_1, a_2, \ldots, a_n$ if

(1) $a_i \mid d$ for $i = 1, 2, \ldots, n$ (d is a common multiple),

(2) $a_i \mid c$ for $i = 1, 2, \ldots, n$ implies $d \mid c$.

In brief, an element $d \in R$ is a least common multiple of $a_1, a_2, \ldots, a_1$ if it is a common multiple of $a_1, a_2, \ldots, a_n$ which divides any other common multiple. The reader should note that a least common multiple, in case it exists, is unique apart from the distinction between associates; indeed, if d and d' are both least common multiples of $a_1, a_2, \ldots, a_n$, then $d \mid d'$ and $d' \mid d$; hence, d and d' are associates. We hereafter adopt the standard notation $\operatorname{lcm}(a_1, a_2, \ldots, a_n)$ to represent any least common multiple of $a_1, a_2, \ldots, a_n$. It can be shown that nonzero elements $a_1, a_2, \ldots, a_n$ in any ring R have a least common multiple if and only if the ideal $\cap\,(a_i)$ is principal (see [2] Theorem 6-5). It can also be shown that GCDs are exactly LCMs.

2.7 EXERCISES

1. [1] p.379 EX2.1. Factor the following polynomials into irreducible factors in $\mathbb{F}_p[x]$.

i. $x^3 + x^2 + x + 1, p = 2$,
ii. $x^2 - 3x - 3, p = 5$,
iii. $x^2 + 1, p = 7$

2. [1] p.379 EX2.2. Compute the greatest common divisor of the polynomials $x^6 + x^4 + x^3 + x^2 + x + 1$ and $x^5 + 2x^3 + x^2 + x + 1$ in $\mathbb{Q}[x]$.

3. [1] p.379 EX2.3. How many roots does the polynomial $x^2 - 2$ have, modulo 8?

2.8 Noetherian Rings[3]

We have seen in the proof of PID implying UFD that PIDs has the ascending chain condition:

Definition 2.8.1. A ring R is said to satisfy the ascending chain condition (ACC) if any ascending chain of ideals $I_1 \subset I_2 \subset \cdots$ eventually terminates.

Lemma 2.8.2. A ring R satisfies ACC iff all ideals $I \in R$ are finitely generated.

Proof. Trivial. □

Definition 2.8.3. A ring R is called Noetherian if it satisfies ACC. Note that apart from above lemma, ACC is also equivalent to the condition that every non-empty set of ideals in A has a maximal element.

Theorem 2.8.4 (Hilbert's Basis Theorem)**.** If R is Noetherian, then $R[X]$ is also Noetherian.

Proof. Start with an ideal $J \trianglelefteq R[X]$. Pick $f_1 \in J$ with minimal degree. If $J = (f_1)$, we are done. Otherwise we can pick $f_2 \in J \backslash (f_1)$ with minimal degree. Continuing this, if J is not finitely generated, then there is a nested sequence

$$(f_1) \subsetneq (f_1, f_2) \subsetneq \cdots, \deg f_1 \leqslant \deg f_2 \leqslant \cdots$$

Let a_i be the leading coefficient of f_i, then consider a chain of ideals $(a_1) \subset (a_1, a_2) \subset \cdots$. R is Noetherian, so this sequence must eventually terminates, so in particular there is some $m \in \mathbb{N}$ such that $a_{m+1} \in (a_1, \ldots, a_m)$. So $a_{m+1} = \lambda_1 a_1 + \cdots + \lambda_m a_m$. Now consider

$$g(X) = \sum_{i=1}^{m} \lambda_i X^{\deg f_{m+1} - \deg f_i} f_i$$

So g, f_{m+1} has the same degree and leading coefficient, so $\deg(f_{m+1} - g) < \deg f_{m+1}$. But $f_{m+1} - g \in J$, so since we chose f_{m+1} to have the minimal degree in $J \backslash (f_1, \ldots, f_m)$, $f_{m+1} - g \in (f_1 \ldots, f_m)$, so $f_{m+1} \in (f_1, \ldots, f_m)$, contradiction. □

Corollary 2.8.5. $R[X_1, \ldots, X_n]$ is Noetherian whenever R is.

In particular, $\mathbb{Z}[X_1, \ldots, X_n], \mathbb{F}[X_1, \ldots, X_n]$ are Noetherian (where $\mathbb{F}$ is a field).

Proof. Apply the preceding theorem recursively. □

[3] Taken from David

Example 2.8.6. Let $R = \mathbb{C}[X_1, \ldots, X_n]$. Let $V \subset \mathbb{C}^n$ be of the form

$$V(\mathcal{F}) = \{(a_1, \ldots, a_n) \in \mathbb{C}^n : f(a_1, \ldots, a_n) = 0, \forall f \in \mathcal{F}\}$$

for some (possibly infinite) subset $\mathcal{F} \subset R$. Let

$$I = \left\{ \sum_{i=1}^{m} \lambda_i f_i : m \in \mathbb{N}, \lambda_i \in R, f_i \in \mathcal{F} \right\}$$

Then $I \trianglelefteq R$ and $V(I) = V(\mathcal{F})$, but R is Noetherian by the preceding corollary, so I is finitely generated and thus $V(\mathcal{F})$ can be defined by only finitely many polynomials.

Lemma 2.8.7. Any quotient ring of a Noetherian ring is again Noetherian.

Proof. Suppose R is Noetherian and $I \trianglelefteq R$ is an ideal. Consider a chain of ideals $J_1 \subset J_2 \subset \cdots$ in R/I. But we know the correspondence between the ideals in R/I and the ideals of R containing I, so there are ideals $I_1, I_2, \ldots$ all containing I with $J_i = I_i/I$. But then $I_1 \subset I_2 \subset \ldots$, so there is $N \in \mathbb{N}$ such that for any $m > N$, $I_m = I_N$, hence $J_m = I_m/I = I_N/I = J_N$, hence the sequence eventually terminates, thus R/I is Noetherian. □

Example 2.8.8. 1. The Gaussian integers can be written as $\mathbb{Z}[i] \cong \mathbb{Z}[X]/(X^2 + 1)$ hence is Noetherian.
2. If $R[X]$ is Noetherian, then R is Noetherian since $R \cong R[X]/(X)$, so Hilbert's Basis Theorem is actually an "if and only if".

Example 2.8.9 (Non-example)**.** We shall give examples of a non-Noetherian rings.
1. We consider the ring as the upper limit

$$R = \mathbb{Z}[X_1, X_2, \ldots] = \bigcup_{n \in \mathbb{N}} \mathbb{Z}[X_1, \ldots, X_n]$$

Then $(X_1) \subsetneq (X_1, X_2) \subsetneq \cdots$, so R is not Noetherian.
2. Consider the ring $R \leqslant \mathbb{Q}[X]$ by collecting $R = \{f \in \mathbb{Q}[X] : f(0) \in \mathbb{Z}\}$, then R is obviously a ring with

$$(X) \subsetneq (2^{-1}X) \subsetneq (2^{-2}X) \subsetneq \cdots$$

3. Consider the ring R of infinitely differentiable functions $[-1, 1] \to \mathbb{R}$ under pointwise operations, this is also not Noetherian (exercise).

2.8 EXERCISES

We list some results from Atiyah and MacDonald's *Introduction to Commutative Algebra* (AM) regarding primary decomposition of Neotherian rings. First, an ideal I is called **irreducible** iff $I = J \cap K \Rightarrow (I = J \text{ or } I = K)$.

1. AM Lemma 7.11. In a Noetherian ring A, every ideal is a finite intersection of irreducible ideals.

2. AM Lemma 7.12. In a Noetherian ring A, every irreducible ideal is primary.

3. AM Theorem 7.13. In a Noetherian ring A, every ideal has a primary decomposition.

2.9 Euclidean Domains and Euclid's Algorithms

Remark: rings ⊃ commutative rings ⊃ integral domains ⊃ GCD domains ⊃ UFDs ⊃ PIDs ⊃ Euclidean domains ⊃ fields ⊃ algebraically closed fields.

Definition 2.9.1. An integral domain R is a **Euclidean domain** if there is a map $d : R\backslash\{0\} \longrightarrow \mathbb{Z}_+$ with the following division-with-remainder property:

- Let a and b be elements of R, and suppose that a is not zero. There are elements q and r in R such that $b = aq + r$, and either $r = 0$ or else $d(r) < d(a)$.

Example 2.9.2.

(1) $R = \mathbb{Z}$ with $d(a) = |a|$ is a Euclidean domain.

(2) If $R = F[x]$ where F is a field, then $d(f(x))= \deg(f)$. R with d is a Euclidean domain.

(3) For any field F, we define $\forall a \in F\backslash\{0\},\ d(a) = 0$. Then it is a Euclidean domain.

Proposition 2.9.3. Euclidean domains are PIDs.

Proof. We mimic the proof that the ring of integers is a principal ideal domain once more. Let R be a Euclidean domain with size function σ, and let A be an ideal of R. We must show that A is principal. The zero ideal is principal, so we may assume that A is not the zero ideal. Then A contains a nonzero element. We choose a nonzero element a of A such that $d(a)$ is as small as possible, and we show that A is the principal ideal (a) of multiples of a.

Because A is an ideal and a is in A, any multiple aq with q in R is in A. So $(a) \subset A$. To show that $A \subset (a)$, we take an arbitrary element b of A. We use division with remainder to write $b = aq + r$, where either $r = 0$, or $d(r) < d(a)$. Then b and aq are in A, so $r = b - aq$ is in A too. Since $d(a)$ is minimal, we can't have $d(r) < d(a)$, and it follows that $r = 0$. This shows that a divides b, and hence that b is in the principal ideal (a). Since b is arbitrary, $A \subset (a)$, and therefore $A = (a)$. □

Theorem 2.9.4 (Euclid's division lemma)**.** Given two integers a and b, with $b \neq 0$, there exist unique integers q and r such that

$$a = bq + r, 0 \leqslant r < |b|.$$

In the above theorem, each of the four integers has a name of its own: a is called the dividend, b is called the divisor, q is called the quotient and r is called the remainder.

Theorem 2.9.5 (Euclid's division lemma (half remainder version))**.** For every pair of integers a, b where $b \neq 0$, there exist unique integers q, r such that $a = qb + r$ and $-\frac{|b|}{2} \leqslant r < \frac{|b|}{2}$:

$$\forall a, b \in \mathbb{Z}, b \neq 0 : \exists! q, r \in \mathbb{Z} : a = qb + r, -\frac{|b|}{2} \leqslant r < \frac{|b|}{2}$$

We show that the Gauss integers form a Euclidean domain too and provide the division algorithm for Gauss integers.

Example 2.9.6. $\mathbb{Z}[i] = \{a + bi \,|\, a, b \in \mathbb{Z}\}$ is an Euclidean domain with

$$d : \mathbb{Z}[i] - \{0\} \longrightarrow \mathbb{Z}_+;\ a + bi \mapsto |a + bi| = a^2 + b^2.$$

Proof. d is multiplicative: $d((a + bi)(a' + b'i)) = d((aa' - bb') + (ab' + a'b)i) = (a^2 + b^2)(a'^2 + b'^2) = d(a + bi)d(a' + b'i)$.

(1): If $a = bc$, where $a, b, c \neq 0$, then $d(a) = d(b)d(c) > d(b)$.

(2): Suppose $x, y \in \mathbb{Z}[i]$ and we want to divide x by y.

<u>case 1</u>: if $y = n \in \mathbb{Z}_+$, $x = a + bi$. Then a and b are both integers now. We can write by Theorem 2.9.5 $a = nq + r, r = 0$ or $|r| < \frac{n}{2}$ and $b = nq' + r, r' = 0$ or $|r'| < \frac{n}{2}$. Then $x = a + bi = (nq + r) + i(nq' + r') = n(q + iq') + (r + ir')$, and $d(r + ir') = r^2 + r'^2 < \frac{n^2}{4} + \frac{n^2}{4} = \frac{n^2}{2} < n^2 = d(n)$.

case 2: Now suppose we are dividing x by an arbitary y, and we use the previous result by letting $n = y\bar{y} = d(y) > 0$. So we can divide $x\bar{y}$ by n where

$$x\bar{y} = qn + r, \qquad d(r) < d(n) \implies x\bar{y} = q\bar{y}y + r$$

Then claim that $x = qy + (x - qy)$, where $d(x - qy) < d(y)$. Notice that

$$d(x - qy)d(\bar{y}) = d(x\bar{y} - qy\bar{y}) = d(r) < d(n) = d(y)^2 \implies d(x - qy) < d(y)$$

Thus, this result holds. □

Example 2.9.7. This is not unique. $3 = (1 + i)(1 - i) + 1, d(1) < d(1 - i)$. Also $3 = (2 - i)(1 - i) - i$, $d(-i) < d(1 - i)$

Theorem 2.9.8 (Euclid's Algorithm)**.** If R is a Euclidean Domain, and $a, b \in R \neq 0$, we can find the gcd using the following algorithm

$$\begin{array}{rll} & a = bq_0 + r_0 & \gcd(a, b) = \gcd(b, r_0) \\ \text{if } r_0 \neq 0, & b_0 = r_0q_1 + r_1 & \gcd(b, r_0) = \gcd(r_0, r_1) \\ & r_0 = r_1q_2 + r_2 & \gcd(r_0, r_1) = \gcd(r_1, r_2) \\ & \vdots & \vdots \\ & r_n = r_{n+1}q_{n+2} + r_{n+2} & \gcd(r_n, r_{n+1}) = \gcd(r_{n+1}, r_{n+2}) \\ & r_{n+1} = r_{n+2}q_{n+3} + 0 & \gcd(r_{n+1}, r_{n+2}) = \gcd(r_{n+2}, r_{n+3}) = r_{n+2} \end{array}$$

where the remainder will eventually go to zero as the degree keeps decreasing.

Proof. For example, to verify $\gcd(a, b) = \gcd(b, r_0) = \gcd(b, r - q_0)$ is to show

- $\gcd(a, b) \mid b$, $\gcd(a, b) \mid a - bq_0$.
- $d' \mid b$, $d' \mid a - bq_0 \implies d' \mid \gcd(a, b)$.

That's direct computation. □

2.9 EXERCISES

1. Let $R = \mathbf{Z}[i]$ and $d(a + bi) = a^2 + b^2$. Let $\alpha = 11 + 3i$ and $\beta = 1 + 8i$.

i. Write $\alpha = \beta q + r$ in R with $d(r) < d(\beta)$ using the method we discussed in class.
ii. Find the gcd of α and β by using the Euclidean algorithm.

2. [1] p.379 Ex2.6. Prove that the following rings are Euclidean domains.

i. $\mathbb{Z}[\omega], \omega = e^{2\pi i/3}$
ii. $\mathbb{Z}[\sqrt{-2}]$.

3. $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$ is the field of five elements, with addition and multiplication modulo 5 , isomorphic to $\mathbb{Z}/5\mathbb{Z}$. Find polynomials $q(X), r(X)$ in $\mathbb{F}_5[X]$ such that

$$X^7 + 2X^6 + 3X^5 + 4X^4 + X^2 + 2X + 3 = q(X) \cdot \left(X^2 + 4\right) + r(X)$$

where $r(X)$ has degree at most 1.

2.10 Rings of Formal Power Series

- We will extensively copy from [2] chap.7 and [1]'s section "Factoring Integer Polynomial" for the last four sections of this chapter.

To begin with simpler things, given an arbitrary ring R, let seq R denote the totality of all infinite sequences

$$f = (a_0, a_1, a_2, \ldots, a_k, \ldots)$$

of elements $a_k \in R$. Such sequences are called formal power series, or merely power series, over R. (Our choice of terminology will be justified shortly.)

We intend to introduce suitable operations in the set seq R so that the resulting system forms a ring containing R as a subring. At the outset, it should be made perfectly clear that two power series

$$f = (a_0, a_1, a_2, \ldots) \quad \text{and} \quad g = (b_0, b_1, b_2, \ldots)$$

are considered to be equal if and only if they are equal term by term:

$$f = g \text{ if and only if } a_k = b_k \text{ for all } k \geqslant 0.$$

Now, power series may themselves be added and multiplied as follows:

$$\begin{aligned} f + g &= (a_0 + b_0, a_1 + b_1, \ldots), \\ fg &= (c_0, c_1, c_2, \ldots) \end{aligned}$$

where, for each $k \geqslant 0$, c_k is given by

$$c_k = \sum_{i+j=k} a_i b_j = a_0 b_k + a_1 b_{k-1} + \cdots + a_{k-1} b_1 + a_k b_0.$$

(It is understood that the above summation runs over all integers $i, j \geqslant 0$ subject to the restriction that $i + j = k$.)

A routine check establishes that with these two definitions seq R becomes a ring. To verify a distributive law, for instance, take

$$f = (a_0, a_1, \ldots), \quad g = (b_0, b_1, \ldots), \quad h = (c_0, c_1, \ldots).$$

One finds quickly that

$$f(g + h) = (a_0, a_1, \ldots)(b_0 + c_0, b_1 + c_1, \ldots) = (d_0, d_1, \ldots),$$

where

$$\begin{aligned} d_k &= \sum_{i+j=k} a_i (b_j + c_j) = \sum_{i+j=k} (a_i b_j + a_i c_j) \\ &= \sum_{i+j=k} a_i b_j + \sum_{i+j=k} a_i c_j. \end{aligned}$$

A similar calculation of $fg + fh$ leads to the same general term, so that $f(g + h) = fg + fh$. The rest of the details are left to the reader's care. We simply point out that the sequence $(0, 0, 0, \ldots)$ serves as the zero element of this ring, while the additive inverse of an arbitrary member $(a_0, a_1, a_2, \ldots)$ of seq R is, of course, $(-a_0, -a_1, -a_2, \ldots)$. To summarize what we know so far:

Theorem 2.10.1. The system seq R forms a ring, known as the **ring of (formal) power series over** R. Furthermore, the ring seq R is commutative with unity if and only if the given ring R has these properties.

If S represents the subset of all sequences having 0 for every term beyond the first, that is, the set

$$S = \{(a, 0, 0, \ldots) \mid a \in R\},$$

then it is not particularly difficult to show that S constitutes a subring of seq R which is isomorphic to R; one need only consider the mapping that sends the sequence $(a, 0, 0, \ldots)$ to the element a. In this sense, seq R contains the original ring R as a subring.

Having reached this stage, we shall no longer distinguish between an element $a \in R$ and the special sequence $(a, 0, 0, \ldots)$ of seq R. The elements of R, regarded as power series, are hereafter called **constant series**, or just **constants**.

With the aid of some additional notation, it is possible to represent power series the way we would like them to look. As a first step in this direction, we let ax designate the sequence

$$(0, a, 0, 0, \ldots).$$

That is, ax is the specific member of seq R which has the element a for its second term and 0 for all other terms. More generally, the symbol ax^4, $n \geqslant 1$, will denote the sequence

$$(0, \ldots, 0, a, 0, \ldots),$$

where the element a appears as the $(n + 1)$ st term in this sequence; for example, we have and

$$ax^2 = (0, 0, a, 0, \ldots)$$

$$ax^3 = (0, 0, 0, a, 0, \ldots).$$

By use of these definitions, each power series

$$f = (a_0, a_1, a_2, \ldots, a_n, \ldots)$$

may be uniquely expressed in the form

$$\begin{aligned} f &= (a_0, 0, 0, \ldots) + (0, a_1, 0, \ldots) + \cdots + (0, \ldots, 0, a_n, 0, \ldots) + \cdots \\ &= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots \end{aligned}$$

with the obvious identification of a_0 with the sequence $(a_0, 0, 0, \ldots)$. Thus there is no loss in regarding the power series ring seq R as consisting of all formal expressions

$$f = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots,$$

where the elements $a_0, a_1, \ldots, a_n, \ldots$ (the coefficients of f) lie in R. As a notational device, we shall often write this as $f = \sum a_k x^k$ (the summation symbol is not an actual sum and convergence is not at issue here).

Using sigma notation, the definitions of addition and multiplication of power series assume the form where

$$\begin{aligned} \sum a_k x^k + \sum b_k x^k &= \sum (a_k + b_k)\, x^k, \\ \left(\sum a_k x^k\right)\left(\sum b_k x^k\right) &= \sum c_k x^k, \end{aligned}$$

$$c_k = \sum_{i+j=k} a_i b_j = \sum_{i=0}^{k} a_i b_{k-i}.$$

We should emphasize that, according to our definition, x is simply a new symbol, or indeterminant, totally unrelated to the ring R and in no sense represents an element of R. To indicate the indeterminant x, it is common practice to write $R[[x]]$ for the set seq R, and $f(x)$ for any member of the same. From now on, we shall make exclusive use of this notation.

Remark 2.10.2. If the ring R happens to have a multiplicative identity 1 , many authors will identify the power series $0 + 1x + 0x^2 + 0x^3 + \cdots$ with x thereby treating x itself as a special member of $R[[x]]$; namely, the sequence $x = (0, 1, 0, 0, \ldots)$. From this view, ax becomes an actual product of members of $R[[x]]$:

$$ax = (a, 0, 0, \ldots)(0, 1, 0, 0, \ldots).$$

Concerning the notation of power series, it is customary to omit terms with zero coefficients and to replace $(-a_k)\, x^k$ by $-a_k x^k$. Although x is not to be considered as an element of $R[[x]]$, we shall nonetheless take the liberty of writing the term $1x^k$ as $x^k (k \geqslant 1)$. With these conventions, one should view, for example, the power series

$$1 + x^2 + x^4 + \cdots + x^{2n} + \cdots \in \mathbb{Z}[[x]]$$

as representing the sequence $(1, 0, 1, 0, \ldots)$. An important definition in connection with power series is that of order, given below.

Definition 2.10.3. If $f(x) = \sum a_k x^k$ is a nonzero power series (that is, if not all the $a_k = 0$) in $R[[x]]$, then the smallest integer n such that $a_n \neq 0$ is called the order of $f(x)$ and denoted by ord $f(x)$. Suppose $f(x), g(x) \in R[[x]]$, with ord $f(x) = n$ and ord $g(x) = m$, so that

$$\begin{aligned} f(x) &= a_n x^n + a_{n+1} x^{n+1} + \cdots \quad (a_n \neq 0)\,, \\ g(x) &= b_m x^m + b_{m+1} x^{m+1} + \cdots \quad (b_m \neq 0)\,. \end{aligned}$$

From the definition of multiplication in $R[[x]]$, the reader may easily check that all coefficients of $f(x)g(x)$ up to the $(n + m)$ th are zero, whence

$$f(x)g(x) = a_n b_m x^{n+m} + (a_{n+1} b_m + a_n b_{m+1})\, x^{n+m+1} + \cdots .$$

If we assume that one of a_n or b_m is not a divisor of zero in R, then $a_n b_m \neq 0$ and

$$\operatorname{ord}(f(x)g(x)) = n + m = \operatorname{ord} f(x) + \operatorname{ord} g(x).$$

This certainly holds if R is taken to be an integral domain, or again if R has an identity and one of a_n or b_m is the identity element.

The foregoing argument serves to establish the first part of the next theorem; the proof of the second assertion is left as an exercise.

Theorem 2.10.4. If $f(x)$ and $g(x)$ are nonzero power series in $R[[x]]$, then

(1) either $f(x)g(x) = 0$ or ord $(f(x)g(x)) \geqslant$ ord $f(x)+$ ord $g(x)$, with equality if R is an integral domain;

(2) either $f(x) + g(x) = 0$ or

$$\operatorname{ord}(f(x) + g(x)) \geqslant \min\{\operatorname{ord} f(x), \operatorname{ord} g(x)\}.$$

The notation of order can be used to prove the following corollary.

Corollary 2.10.5. If the ring R is an integral domain, then so also is its power series ring $R[[x]]$.

Proof. We observed earlier that whenever R is a commutative ring with identity, these properties carry over to $R[[x]]$. To see that $R[[x]]$ has no zero divisors, select $f(x) \neq 0, g(x) \neq 0$ in $R[[x]]$. Then,

$$\text{ord } (f(x)g(x)) = \text{ ord } f(x) + \operatorname{ord} g(x) > 0;$$

hence, the product $f(x)g(x)$ cannot be the zero series. □

Although arbitrary power series rings are of some interest, the most important consequences arise on specializing the discussion to power series whose coefficients are taken from a field. These will be seen to form principal ideal domains and, in consequence, unique factorization domains. The following intermediate result is directed towards establishing this fact.

Lemma 2.10.6. Let R be a commutative ring with identity. A formal power series $f(x) = \sum a_k x^k$ is invertible in $R[[x]]$ if and only if the constant term a_0 has an inverse in R.

Proof. If $f(x)g(x) = 1$, where $g(x) = \sum b_k x^k$, then the definition of multiplication in $R[[x]]$ shows that $a_0 b_0 = 1$; hence, a_0 is invertible as an element of R.

For the converse, suppose that the element a_0 has an inverse in R. We proceed inductively to define the coefficients of a power series $\sum b_k x^k$ in $R[[x]]$ which is the inverse of $f(x)$. To do this, simply take $b_0 = a_0^{-1}$ and, assuming $b_1, b_2, \ldots, b_{k-1}$ have already been defined, let

$$b_k = -a_0^{-1}\left(a_1 b_{k-1} + a_2 b_{k-2} + \cdots + a_k b_0\right).$$

Then $a_0 b_0 = 1$, while, for $k \geqslant 1$,

$$c_k = \sum_{i+j=k} a_i b_j = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0 = 0.$$

By our choice of the b_k 's, we evidently must have $\left(\sum a_k x^k\right)\left(\sum b_k x^k\right) = 1$, and so $\sum a_k x^k$ possesses an inverse in $R[[x]]$. □

Corollary 2.10.7. A power series $f(x) = \sum a_k x^k \in F[[x]]$, where F is a field, has an inverse in $F[[x]]$ if and only if its constant term $a_0 \neq 0$. Having dealt with these preliminaries, we are now ready to proceed to describe the ideal structure of $F[[x]]$.

Theorem 2.10.8. For any field F, the power series ring $F[[x]]$ is a principal ideal domain; in fact, the nontrivial ideals of $F[[x]]$ are of the form $\left(x^k\right)$, where $k \in Z_+$.

Proof. Let I be any proper ideal of $F[[x]]$. Either $I = \{0\}$, in which case I is just the principal ideal (0), or else I contains nonzero elements. In the batter event, choose a nonzero power series $f(x) \in I$ of minimal order. Suppose that $f(x)$ is of order k, so that

$$f(x) = a_k x^k + a_{k+1} x^{k+1} + \cdots = x^k\left(a_k + a_{k+1} x + \cdots\right).$$

Since the coefficient $a_k \neq 0$, the previous lemma insures that the power series $a_k + a_{k+1} x + \cdots$ is an invertible element of $F[[x]]$; in other words, $f(x) = x^k g(x)$, where $g(x)$ has an inverse in $F[[x]]$. But, then,

$$x^k = f(x)g(x)^{-1} \in I,$$

which leads to the inclusion $\left(x^k\right) \subseteq I$. On the other hand, take $h(x)$ to be any nonzero power series in I, say of order n. Since $f(x)$ is assumed to have least order among all members of I, it is clear that $k \leqslant n$; thus, $h(x)$ can be written in the form

$$h(x) = x^k\left(b_n x^{n-k} + b_{n+1} x^{n-k+1} + \cdots\right) \in \left(x^k\right).$$

This implies that $I \subseteq \left(x^k\right)$, and the equality $I = \left(x^k\right)$ follows. □

Corollary 2.10.9. The ring $F[[x]]$ is a local ring with (x) as its maximal ideal.

Proof. Inasmuch as the ideals of $F[[x]]$ form a chain

$$F[[x]] \supset (x) \supset \left(x^2\right) \supset \cdots \supset \{0\},$$

the conclusion is obvious. □

Corollary 2.10.10. Any nonzero element $f(x) \in F[[x]]$ can be written in the form $f(x) = g(x)x^k$, where $g(x)$ is invertible and $k \geqslant 0$.

To this we add, for future reference, the following assertion regarding the maximal ideals of a power series ring over a commutative ring with identity.

Theorem 2.10.11. Let R be a commutative ring with identity. There is a one-to-one correspondence between the maximal ideals M of the ring R and the maximal ideals M' of $R[[x]]$ in such a way that M' corresponds to M if and only if M' is generated by M and x; that is, $M' = (M, x)$.

Proof. See [2] Theorem 7-4. □

The **ring of formal Laurent series** in x with coefficients in R is denoted by $R((x))$, and is defined as follows. The elements of $R((x))$ are infinite expressions of the form

$$f(x) = a_r x^r + a_{r+1}x^{r+1} + a_{r+2}x^{r+2} + \cdots$$

in which $r \in \mathbb{Z}$ and $a_n \in R$ for all $n \geqslant r$. That is, a formal Laurent series is a generalization of a formal power series in which finitely many negative exponents are permitted. Addition and multiplication are defined just as for the ring $R[[x]]$ of formal power series, and $R((x))$ is commutative because R is. (I encourage you to check that when multiplying two formal Laurent series the coefficients of the product really are polynomial functions of the coefficients of the factors, and hence are in the ring R. This ensures that the multiplication in $R((x))$ is well-defined.) Note that the ring $R[[x]]$ is a subset of the ring $R((x))$, and that the algebraic operations of these rings agree on the subset $R[[x]]$. If $f(x) \in R((x))$ and $f(x) \neq 0$, then there is a smallest integer n such that $[x^n]\, f(x) \neq 0$; this is called the **index** of $f(x)$ and is denoted by $I(f)$. By convention, the index of 0 is $I(0) := +\infty$. Concerning the existence of multiplicative inverses in $R((x))$, we have the following proposition.

Proposition 2.10.12. Let R be a commutative ring. If R is a field then $R((x))$ is a field.

Proof. Consider a nonzero $f(x) = \sum_{n=I(f)}^{\infty} a_n x^n$ in $R((x))$. Then $a_{I(f)} \neq 0$ so that it is invertible in R, since R is a field. We may write $f(x) = x^{I(f)}g(x)$ with $g(x) = \sum_{n=0}^{\infty} a_{n+I(f)}x^n$, so that $g(x)$ is a formal power series in $R[[x]]$. The coefficient of x^0 in $g(x)$ is $a_{I(f)}$ and, by Lemma 2.10.6, it follows that $g(x)$ is invertible in $R[[x]]$, and hence in $R((x))$. Let $h(x) := x^{-I(f)}g^{-1}(x)$. Then

$$f(x)h(x) = x^{I(f)}g(x)x^{-I(f)}g^{-1}(x) = 1,$$

so that $h(x) = f^{-1}(x)$ and $f(x)$ is invertible in $R((x))$. Therefore, $R((x))$ is a field. □

The inclusions $R[x] \subset R[[x]] \subset R((x))$ and $R[x] \subset R(x)$ have been remarked upon already. In fact, if R is a field then $R(x) \subset R((x))$ as well. Also, the rings $R[[x]]$ and $R(x)$ have a nontrivial intersection, but neither one contains the other. Since we have no pressing need for these facts we will not pause to prove them, but instead relegate them to exercises.

The usual rules of arithmetic hold for all of the rings constructed above, but there are other operations on these rings that have no analogues in $\mathbb{Z}$. Care must be taken with these operations to ensure that they produce well-defined power series. In other words, these operations are not universally defined.

The first of the new operations are formal differentiation and formal integration. Since $R((x))$ contains all the other rings above (if R is a field) we will just define these operations on a typical formal Laurent series $f(x) = \sum_{n=I(f)}^{\infty} a_n x^n$. The formal derivative is always defined as

$$f'(x) := \frac{d}{dx} f(x) := \sum_{n=I(f)}^{\infty} n a_n x^{n-1}.$$

The formal integral is defined only if $\mathbb{Q} \subseteq R$ and $a_{-1} = 0$, in which case

$$\int f(x)dx := \sum_{n \geqslant I(f), n \neq -1} a_n \frac{x^{n+1}}{n+1}.$$

In particular, the formal integral is defined on all of $R[[x]]$ when $\mathbb{Q} \subseteq R$. One can show algebraically from the definitions that the familiar rules of calculus (the Product Rule, Quotient Rule, Chain Rule, Integration by Parts, and so on) continue to hold when all the integrals involved are defined. Concepts of onvergence, sequence, and limit etc. can also be considered in new context. We direct one with further interest to the link.

2.10 EXERCISES

1. Let R be a commutative ring. For $a \in R$ consider the function $\mu_a : R \to R$ defined by $\mu_a(r) := ar$ for all $r \in R$.

i. Show that if R is an integral domain and $a \neq 0$, then $\mu_a : R \to R$ is an injection.
ii. Show that if R is a finite integral domain then R is a field. (The ring $\mathbb{Z}$ of integers is an integral domain which is not a field. Thus, finiteness of R is essential for this problem.)

2. Let R be a commutative ring.

i. Show that if R is an integral domain, then $R[x]$ is an integral domain.
ii. Show that neither of $R[[x]]$ nor $R(x)$ contains the other.
iii. Show that if R is a field then $R(x)$ is a proper subset of $R((x))$.
iv. Find an element of $\mathbb{Z}(x)$ which is not in $\mathbb{Z}((x))$.
v. Show that $R[[x]][y]$ is a proper subset of $R[y][[x]]$.

3. Let $f(x)$ and $g(x)$ be in $R((x))$. Show that

$$\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x).$$

2.11 Polynomial Rings

Power series have so far received all the attention, but our primary concern is with polynomials.

Definition 2.11.1. Let $R[x]$ denote the set of all power series in $R[[x]]$ whose coefficients are zero from some index onward (the particular index varies from series to series):

$$R[x] = \{a_0 + a_1 x + \cdots + a_n x^n \mid a_k \in R; n \geqslant 0\}.$$

An element of $R[x]$ is called a **polynomial (in x) over the ring R.**

In essence, we are defining a polynomial to be a finitely nonzero sequence of elements of R. Thus, the sequence $(1, 1, 1, 0, 0, \ldots)$ would be a polynomial over $\mathbb{Z}_2$, but $(1, 0, 1, 0, \ldots, 1, 0, \ldots)$ would not.

It is easily verified that $R[x]$ constitutes a subring of $R[[x]]$, the socalled **ring of polynomials over** R (in an indeterminant x); indeed, if $f(x) = \sum a_k x^k, g(x) = \sum b_k x^k$ are in $R[x]$, with $a_k = 0$ for all $k \geqslant n$ and $b_k = 0$ for all $k \geqslant m$, then

$$a_k + b_k = 0 \text{ for } k \geqslant \max\{m, n\}$$
$$\sum_{i+j=k} a_i b_j = 0 \text{ for } k \geqslant m + n$$

so that both the sum $f(x) + g(x)$ and product $f(x)g(x)$ belong to $R[x]$. Running parallel to the idea of the order of a power series is that of the degree of a polynomial, which we introduce at this time.

Definition 2.11.2. Given a nonzero polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n \quad (a_n \neq 0)$$

in $R[x]$, we call a_n the **leading coefficient** of $f(x)$; and the integer n, the **degree of the polynomial**. The degree of any nonzero polynomial is therefore a nonnegative integer; no degree is assigned to the zero polynomial. Notice that the polynomials of degree 0 are precisely the nonzero constant polynomials. If R is a ring with identity, a polynomial whose leading coefficient is 1 is said to be a **monic polynomial**.

As a matter of notation, we shall hereafter write $\deg f(x)$ for the degree of any nonzero polynomial $f(x) \in R[x]$.

The result below is similar to that given for power series and its proof is left for the reader to provide; the only change of consequence is that we now use the notion of degree rather than order.

Theorem 2.11.3. If $f(x)$ and $g(x)$ are nonzero polynomials in $R[x]$, then

(1) either $f(x)g(x) = 0$ or $\deg(f(x)g(x)) \leqslant \deg f(x) + \deg g(x)$, with equality whenever R is an integral domain;

(2) either $f(x) + g(x) = 0$ or

$$\deg(f(x) + g(x)) \leqslant \max\{\deg f(x), \deg g(x)\}.$$

Knowing this, one could proceed along the lines of the corollary 2.10.5 to establish

Corollary 2.11.4. If the ring R is an integral domain, then so is its polynomial ring $R[x]$.

Example 2.11.5. As an illustration of what might happen if R has zero divisors, consider Z_8, the ring of integers modulo 8 . Taking

$$f(x) = 1 + 2x, \quad g(x) = 4 + x + 4x^2,$$

we obtain $f(x)g(x) = 4 + x + 6x^2$, so that

$$\deg(f(x)g(x)) = 2 < 1 + 2 = \deg f(x) + \deg g(x).$$

Although many properties of the ring R carry over to the associated polynomial ring $R[x]$, it should be pointed out that for no ring R does $R[x]$ form a field. In fact, when R is a field (or, for that matter, an integral domain), no element of $R[x]$ which has positive degree can possess a multiplicative inverse. For, suppose that $f(x) \in R[x]$, with $\deg f(x) > 0$; if $f(x)g(x) = 1$ for some $g(x)$ in $R[x]$, we could obtain the contradiction

$$0 = \deg 1 = \deg(f(x)g(x)) = \deg f(x) + \deg g(x) \neq 0.$$

The degree of a polynomial is used in the factorization theory of $R[x]$ in much the same way as the absolute value is employed in $\mathbb{Z}$. For, it is through the degree concept that induction can be utilized in $R[x]$ to develop

a polynomial counterpart of the familiar division algorithm. One can subsequently establish that the ring $F[x]$ with coefficients in a field forms a Euclidean domain in which the degree function is taken to be the Euclidean valuation.

Before embarking on this program, we wish to introduce several new ideas. To this purpose, let R be a ring with identity; assume further that R' is any ring containing R as a subring (that is, R' is an extension of R) and let r be an arbitrary element of R'. For each polynomial

$$f(x) = a_0 + a_1x + \cdots + a_nx^n$$

in $R[x]$, we may define $f(r) \in R'$ by taking

$$f(r) = a_0 + a_1r + \cdots + a_nr^n.$$

The element $f(r)$ is said to be the result of **substituting** r for x in $f(x)$. Suffice it to say, the addition and multiplication used in defining $f(r)$ are those of the ring R', not those of $R[x]$.

Now, suppose that $f(x), g(x)$ are polynomials in $R[x]$ and $r \in \operatorname{center}(R')$ (that's bc. [2] does not assume commutativity of the ring here). We leave the reader to prove that if then

$$\begin{aligned} h(x) = f(x) + g(x), \quad k(x) = f(x)g(x), \\ h(r) = f(r) + g(r), \quad k(r) = f(r)g(r). \end{aligned}$$

This being so, it may be concluded that the mapping $\phi_r : R[x] \to R'$ which sends $f(x)$ to $f(r)$ is a homomorphism of $R[x]$ into R'. Such a homomorphism will be called the **substitution homomorphism** determined by r and its image denoted by the symbol $R[r]$:

$$\begin{aligned} R[r] &= \{f(r) \mid f(x) \in R[x]\} \\ &= \left\{a_0 + a_1r + \cdots + a_nr^n \mid a_k \in R; n \geqslant 0\right\}. \end{aligned}$$

It is a simple matter to show that $R[r]$ constitutes a subring of R'; in fact, $R[r]$ is the subring of R' generated by the set $R \cup \{r\}$. (Since R has an identity element $1, 1x = x \in R[x]$, and so $r \in R[r]$.) Notice also that $R[r] = R$ if and only if $r \in R$. The foregoing remarks justify part of the next theorem.

Theorem 2.11.6. Let R be a ring with identity, R' an extension ring of R, and the element $r \in \operatorname{cent} R'$. Then there is a unique homomorphism $\phi_r : R[x] \to R'$ such that $\phi_r(x) = r, \phi_r(a) = a$ for all $a \in R$.

Proof. We need only verify that ϕ_r is unique. Suppose, then, that there is another homomorphism $\tau : R[x] \to R$ satisfy in the indicated conditions and consider any polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$. By assumption, $\tau(a_k) = a_k$ for each coefficient a_k, while $\tau\left(x^k\right) = \tau(x)^k = r^k$. Taking stock of the fact that τ is a homomorphism,

$$\begin{aligned} \tau(f(x)) &= \tau(a_0) + \tau(a_1)\tau(x) + \cdots + \tau(a_n)\tau(x)^n \\ &= a_0 + a_1r + \cdots + a_nr^n = f(r) = \phi_r(f(x)). \end{aligned}$$

This proves that $\tau = \phi_r$, yielding the uniqueness conclusion. Without some commutativity assumption, the above remarks need not hold. For, if we let then

$$h(x) = (x - a)(x - b) = x^2 - (a + b)x + ab,$$

$$h(r) = r^2 - (a + b)r + ab.$$

Lacking the hypothesis that $r \in \operatorname{cent} R'$, it cannot be concluded that

$$(r - a)(r - b) = r^2 - ar - rb + ab$$

will equal $h(r)$; in other words, $h(x) = f(x)g(x)$ does not always imply $h(r) = f(r)g(r)$.

Whenever $f(r) = 0$, we call the element r a root or zero of the polynomial $f(x)$. Of course, a given polynomial $f(x) \in R[x]$ may not have a root in R; we shall see later that when R is a field, there always exists an extension field R' of R in which $f(x)$ possesses a root. It is perhaps appropriate to point out at this time that the problem of obtaining all roots of a polynomial $f(x) \in R[x]$ is equivalent to that of finding all elements $r \in R'$ for which $f(x) \in \ker \phi_r$. □

After this brief digression, let us now state and prove the division algorithm for polynomials.

Theorem 2.11.7 (Division Algorithm (Polynomials))**.** Let R be a commutative ring with identity and f, g be nonzero polynomials in $R[x]$, with the leading coefficient of g an invertible element. Then there exist unique polynomials $q, r \in R[x]$ such that

$$f(x) = q(x)g(x) + r(x),$$

where either $r(x) = 0$ or $\deg r(x) < \deg g(x)$.

Proof. The proof is by induction on the degree of $f(x)$. First, notice that if $f(x) = 0$ or $f(x) \neq 0$ and $\deg f(x) < \deg g(x)$, a representation meeting the requirements of the theorem exists on taking $q(x) = 0, r(x) = f(x)$. Furthermore, if $\deg f(x) = \deg g(x) = 0, f(x)$ and $g(x)$ are both elements of the ring R, and it suffices to let $q(x) = f(x)g(x)^{-1}, r(x) = 0$.

This being so, assume that the theorem is true for polynomials of degree less than n (the induction hypothesis) and let $\deg f(x) = n, \deg g(x) = m$, where $n \geqslant m \geqslant 1$; that is,

$$\begin{aligned} f(x) &= a_0 + a_1 x + \cdots + a_n x^n, \quad a_n \neq 0, \\ g(x) &= b_0 + b_1 x + \cdots + b_m x^m, \quad b_m \neq 0 \quad (n \geqslant m). \end{aligned}$$

Now, the polynomial

$$f_1(x) = f(x) - \left(a_n b_m^{-1}\right) x^{n-m} g(x)$$

lies in $R[x]$ and, since the coefficient of x^n is $a_n - \left(a_n b_m^{-1}\right) b_m = 0$, has degree less than n. By supposition, there are polynomials $q_1(x), r(x) \in R[x]$ such that

$$f_1(x) = q_1(x)g(x) + r(x),$$

where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. Substituting, we obtain the equation

$$\begin{aligned} f(x) &= \left(q_1(x) + \left(a_n b_m^{-1}\right) x^{n-m}\right) g(x) + r(x) \\ &= q(x)g(x) + r(x), \end{aligned}$$

which shows that the desired representation also exists when $\deg f(x) = n$ As for uniqueness, suppose that

$$f(x) = q(x)g(x) + r(x) = q'(x)g(x) + r'(x),$$

where $r(x)$ and $r'(x)$ satisfy the requirements of the theorem. Subtracting we obtain

$$r(x) - r'(x) = \left(q'(x) - q(x)\right) g(x).$$

Since the leading coefficient of $g(x)$ is invertible, it follows that $q'(x) - q(x) = 0$ if and only if $r(x) - r'(x) = 0$. With this in mind, let $q'(x) - q(x) \neq 0$. Knowing that b_m is not a zero divisor of R,

$$\begin{aligned} \deg\left(q'(x) - q(x)\right) g(x) &= \deg\left(q'(x) - q(x)\right) + \deg g(x) \\ &\geqslant \deg g(x) > \deg\left(r(x) - r'(x)\right) \end{aligned}$$

a contradiction; the last inequality relies on the fact that the degrees of $r(x)$ and $r'(x)$ are both less than the degree of $g(x)$. Thus, $q'(x) = q(x)$, which in turn implies that $r'(x) = r(x)$. □

The polynomials $q(x)$ and $r(x)$ appearing in the division algorithm are called, respectively, the **quotient** and **remainder** on dividing $f(x)$ by $g(x)$. In this connection, it is important to observe that if $g(x)$ is a monic polynomial, or if R is taken to be a field, one need not assume that the leading coefficient of $g(x)$ is invertible.

We now come to a series of theorems concerning the factorization properties of $R[x]$.

Theorem 2.11.8 (Remainder Theorem)**.** Let R be a commutative ring with identity. If $f(x) \in R[x]$ and $a \in R$, then there exists a unique polynomial $q(x)$ in $R[x]$ such that $f(x) = (x-a)q(x) + f(a)$.

Proof. All this is scarcely more than an application of the division algorithm to the polynomials $f(x)$ and $x - a$. We then obtain

$$f(x) = (x-a)q(x) + r(x),$$

where $r(x) = 0$ or $\deg r(x) < \deg(x-a) = 1$. It follows in either case that $r(x)$ is a constant polynomial, say $r(x) = r \in R$. Substitution of a for x leads to

$$f(a) = (a-a)q(a) + r(a) = r,$$

as desired. □

Corollary 2.11.9. The polynomial $f(x) \in R[x]$ is divisible by $x - a$ if and only if a is a root of $f(x)$. Let us next show that a polynomial cannot have more roots in an integral domain than its degree.

We give some results without proof:

Theorem 2.11.10 ([2] Theorem 7-9)**.** Let R be an integral domain and $f(x) \in R[x]$ be a nonzero polynomial of degree n. Then $f(x)$ can have at most n distinct roots in R.

Corollary 2.11.11. Let $f(x)$ and $g(x)$ be two nonzero polynomials of degree n over the integral domain R. If there exist $n+1$ distinct elements $a_k \in R (k = 1, 2, \ldots, n+1)$ such that $f(a_k) = g(a_k)$, then $f(x) = g(x)$.

Corollary 2.11.12. Let $f(x) \in R[x]$, where R is an integral domain, and let S be any infinite subset of R. If $f(a) = 0$ for all $a \in S$, then $f(x)$ is the zero polynomial.

Example 2.11.13. Consider the polynomial $x^p - x \in \mathbb{Z}_p[x]$, where p is a prime number. Now, the nonzero elements of $\mathbb{Z}_p$ form a abelian group under multiplication of order $p-1$. Hence, we have $a^{p-1} = 1$, or $a^p = a$ for every $a \neq 0$. This is equally true if $a = 0$. Our example shows that it may very well happen that every element of the underlying ring is a root of a polynomial, yet the polynomial is not zero.

With the Division Algorithm at our disposal, we can prove that the ring $F[x]$ is rich in structure.

Theorem 2.11.14. The polynomial ring $F[x]$, where F is a field, forms a Euclidean domain.

Proof. As has been noted in Corollary 2.11.4, $F[x]$ is an integral domain. Moreover, the function δ defined by $\delta(f(x)) = \deg f(x)$ for any nonzero $f(x) \in F[x]$ is a suitable Euclidean valuation. If $f(x)$ and $g(x)$ are two nonzero polynomials in $F[x]$ Theorem 2.11.3 implies that

$$\begin{aligned}\delta(f(x)g(x)) &= \deg(f(x)g(x)) \\ &= \deg f(x) + \deg g(x) \geqslant \deg f(x) = \delta(f(x)),\end{aligned}$$

since $\deg g(x) \geqslant 0$. Thus, the function δ satisfies the requisite properties of a Euclidean valuation. □

The reader is no doubt anticipating the corollary below.

Corollary 2.11.15. $F[x]$ with F a field is a principal ideal domain; hence, a unique factorization domain.

Since a field is trivially a unique factorization domain, part of the last corollary could be regarded as a special case of the coming theorem.

Theorem 2.11.16. If R is a unique factorization domain, then so is $R[x]$.

Proof. Suppose that $R[x]$ is not a unique factorization domain and let S be the set of all nonconstant polynomials in $R[x]$ which do not have a unique factorization into irreducible elements. Select $f(x) \in S$ to be of minimal degree. We may assume that

$$f(x) = p_1(x)p_2(x)\cdots p_r(x) = q_1(x)q_2(x)\cdots q_s(x),$$

where the $p_i(x)$ and $q_j(x)$ are all irreducible and

$$\begin{aligned} m = \deg p_1(x) &\geqslant \deg p_2(x) \geqslant \cdots \geqslant \deg p_r(x), \\ n = \deg q_1(x) &\geqslant \deg q_2(x) \geqslant \cdots \geqslant \deg q_s(x), \end{aligned}$$

with $n \geqslant m > 0$; it is further evident that no $p_i(x) = uq_j(x)$ for any invertible element u (otherwise, the polynomial obtained on dividing $f(x)$ by $q_j(x)$ will have unique factorization; this implies that $f(x)$ can also be factored uniquely). Let a, b be the leading coefficients of $p_1(x), q_1(x)$, respectively, and define

$$g(x) = af(x) - bp_1(x)x^{n-m}q_2(x)\cdots q_s(x).$$

On one hand, we have

$$\begin{aligned} g(x) &= ap_1(x)p_2(x)\cdots p_r(x) - bp_1(x)x^{n-m}q_2(x)\cdots q_s(x) \\ &= p_1(x)\left(ap_2(x)\cdots p_r(x) - bx^{n-m}q_2(x)\cdots q_s(x)\right), \end{aligned}$$

and, on the other hand,

$$\begin{aligned} g(x) &= aq_1(x)q_2(x)\cdots q_s(x) - bp_1(x)x^{n-m}q_2(x)\cdots q_s(x) \\ &= \left(aq_1(x) - bp_1(x)x^{n-m}\right)q_2(x)\cdots q_s(x). \end{aligned}$$

Now, either $g(x) = 0$, which forces $aq_1(x) = bp_1(x)x^{n-m}$, or else $\deg g(x) < \deg f(x)$. In the latter event, $g(x)$ must possess a unique factorization into irreducibles, some of which are $q_2(x), \ldots, q_s(x)$ and $p_1(x)$. The net result of this is that $p_1(x) \mid g(x)$, but $p_1(x) \nmid q_i(x)$ for $i > 1$, so that

$$p_1(x) \mid \left(aq_1(x) - bp_1(x)x^{n-m}\right)$$

and therefore $p_1(x) \mid aq_1(x)$. In either of the two cases considered, we are able to conclude that $p_1(x)$ divides the product $aq_1(x)$; this being so, $aq_1(x) = p_1(x)h(x)$ for some polynomial $h(x) \in R[x]$. Since R is taken to be a unique factorization domain, a has a unique factorization as a product of irreducible elements of R - hence, of $R[x]$-say, $a = c_1c_2\cdots c_k$, where each c_1 is irreducible in $R[x]$. (The only factorizations of a as an element of $R[x]$ are those it had as an element of R.) Arguing from the representation

$$c_1c_2\cdots c_kq_1(x) = p_1(x)h(x)$$

with $p_1(x)$ an irreducible, it follows that each c_i and, in consequence, the element a divides $h(x)$. But, then,

$$aq_1(x) = p_1(x)ah_1(x)$$

for some $h_1(x)$ in $R[x]$ or, upon canceling, $q_1(x) = p_1(x)h_1(x)$; in other words, $p_1(x) \mid q_1(x)$. Using the irreducibility of $q_1(x)$ as a member of $R[x]$, $p_1(x)$ must be an associate of $q_1(x)$. However, this conflicts with our original assumptions. Thus, we see that $R[x]$ is indeed a unique factorization domain. □

Coming back to the corollary 2.11.15, there is an interesting converse which deserves mention: namely, if R is an integral domain such that the polynomial ring $R[x]$ forms a principal ideal domain, then R is necessarily

a field. In verifying this, the main point to be proved is that any nonzero element $a \in R$ is invertible in R. By virtue of our hypothesis, the ideal generated by x and a must be principal; for instance,

$$(x, a) = (f(x)), \quad 0 \neq f(x) \in R[x].$$

Since both $x, a \in (f(x))$, it follows that

$$a = g(x)f(x), \text{ and } x = h(x)f(x)$$

for suitable $g(x), h(x)$ in $R[x]$. The first of these relations signifies that $\deg f(x) = 0$, say $f(x) = a_0$, and as a result $\deg h(x) = 1$, say $h(x) = b_0 + b_1 x\, (b_1 \neq 0)$. We thus obtain $x = a_0\,(b_0 + b_1 x)$. But this means that the product $a_0 b_1 = 1$, thereby making a_0 (or, equivalently, $f(x)$) an invertible element of R. The implication is that the ideal (x, a) is the entire ring $R[x]$. It is therefore possible to write the identity element in the form

$$1 = xk_1(x) + ak_2(x),$$

with the two polynomials $k_1(x), k_2(x) \in R[x]$. This can only happen if $ac_0 = 1$, where $c_0 \neq 0$ is the constant term of $k_2(x)$. In consequence, the element a has a multiplicative inverse in R, which settles the whole affair.

2.11 EXERCISES

1. Prove that the three additive groups $\mathbb{Z} \times \mathbb{Z}, \mathbb{Z}[i]$, and $\mathbb{Z}[x]/\left(x^2\right)$ are all isomorphic to each other.

i.

ii. Prove that no two of the rings $\mathbb{Z} \times \mathbb{Z}, \mathbb{Z}[i]$, and $\mathbb{Z}[x]/\left(x^2\right)$ are isomorphic to each other.

2. Which of the following ideals of $\mathbb{Z}[x, y]$ are prime? Which are maximal? Justify your answer.

$$(x, y), (x, 3y), \left(x^2 + 1, y\right), \left(x^2 + 1, 3, y\right), \left(x^2 + 1, 5, y\right)$$

3. Determine the maximal ideals of the following rings.

i. $\mathbb{Q}[x]/\left(x^2 - 5x + 6\right)$,

ii. $\mathbb{Q}[x]/\left(x^2 + 4x + 6\right)$.

2.12 Irreducibility

At the heart of all the interesting questions on factorization in $R[x]$ lies the idea of an irreducible polynomial. Unwrapping the definition of irreducible element, we have

Definition 2.12.1. Let R be an integral domain. A nonconstant polynomial $f(x) \in R[x]$ is said to be **irreducible over** R, or is an irreducible polynomial in $R[x]$, if $f(x)$ cannot be expressed as the product of two polynomials (in $R[x]$) of positive degree. Otherwise, $f(x)$ is termed reducible in $R[x]$.

In the case of the principal ideal domain $F[x]$, where F is a field, the irreducible polynomials are precisely the irreducible elements of $F[x]$ (recall that the invertible elements of the polynomial ring $F[x]$ are just the nonzero constant polynomials); by Theorem $5 - 9$, these coincide with the prime elements of $F[x]$. Of the equivalent notions, irreducible polynomial, irreducible element, and prime element, the term "irreducible polynomial" is the one customarily preferred for $F[x]$.

Perhaps we should emphasize that Definition 2.12.1 applies only to polynomials of positive degree; the constant polynomials are neither reducible nor irreducible. Thus, the factorization theory of $F[x]$ concerns only polynomials of degree $\geqslant 1$.

The dependence of Definition 2.12.1 upon the polynomial domain $R[x]$ is essential. It may very well happen that a given polynomial is irreducible when viewed as an element of one domain, yet reducible in another. One such example is the polynomial $x^2 + 1$; it is irreducible in $\mathbb{R}[x]$, but reducible in both $\mathbb{C}[x]$, where $x^2 + 1 = (x + i)(x - i)$, and $\mathbb{Z}_2[x]$, where $x^2 + 1 = (x + 1)(x + 1)$. Thus, to ask merely whether a polynomial is irreducible, without specifying the coefficient ring involved, is incomplete and meaningless.

More often than not, it is a formidable task to decide when a given polynomial is irreducible over a specific ring. If F is a finite field, say one of the fields of integers modulo a prime, we may actually examine all of the possible roots. To cite a simple illustration, the polynomial $f(x) = x^3 + x + 1$ is irreducible in $\mathbb{Z}_2[x]$. If there are any factors of this polynomial, at least one must be linear. But the only possible roots for $f(x)$ are 0 and 1 , yet $f(0) = f(1) = 1 \neq 0$, showing that no roots exist in $\mathbb{Z}_2$.

Example 2.12.2. Any linear polynomial $ax + b, a \neq 0$, is irreducible in $R[x]$, where R is an integral domain. Indeed, since the degree of a product of two polynomials is the sum of the degree of the factors, it follows that a representation

$$ax + b = g(x)h(x), \quad g(x), h(x) \in R[x],$$

with $1 \leqslant \deg g(x), 1 \leqslant \deg h(x)$ is impossible. This signifies that every reducible polynomial has degree at least 2 .

Example 2.12.3. The polynomial $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, where $\mathbb{Q}$ as usual is the field of rational numbers. Otherwise, we would have

$$\begin{aligned} x^2 - 2 &= (ax + b)(cx + d) \\ &= (ac)x^2 + (ad + bc)x + bd, \end{aligned}$$

with the coefficients $a, b, c, d \in Q$. Accordingly,

$$ac = 1, \quad ad + bc = 0, \quad bd = -2,$$

whence $c = 1/a, d = -2/b$. Substituting in the relation $ad + bc = 0$, we obtain

$$0 = -2a/b + b/a = \left(-2a^2 + b^2\right)/ab.$$

Thus, $-2a^2 + b^2 = 0$, or $(b/a)^2 = 2$, which is impossible because $\sqrt{2}$ is not a rational number. Although irreducible in $\mathbb{Q}[x]$, the polynomial $x^2 - 2$ is nonetheless reducible in $\mathbb{R}[x]$; in this case, $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$ and both factors are in $\mathbb{R}[x]$.

For ease of reference more than to present new concepts, let us summarize in the next theorem some of the results of previous sections as applied to the principal ideal domain $F[x]$.

Theorem 2.12.4. If F is a field, the following statements are equivalent:

(1) $f(x)$ is an irreducible polynomial in $F[x]$;

(2) the principal ideal $(f(x))$ is a maximal (prime) ideal of $F[x]$;

(3) the quotient ring $F[x]/(f(x))$ forms a field.

The theorem on prime factorization of polynomials is stated now.

Theorem 2.12.5 (Unique factorization in polynomial ring of a field)**.** Each polynomial $f(x) \in F[x]$ of positive degree is the product of a nonzero element of F and irreducible monic polynomials of $F[x]$. Apart from the order of the factors, this factorization is unique.

Suffice it to say, this theorem can be made more explicit for particular polynomial domains. When we deal with polynomials over the complex numbers, the crucial tool is the Fundamental Theorem of Algebra.

Theorem 2.12.6 (The Fundamental Theorem of Algebra)**.** Let $\mathbb{C}$ be the field of complex numbers. If $f(x) \in \mathbb{C}[x]$ is a polynomial of positive degree, then $f(x)$ has at least one root in $\mathbb{C}$.

Although many proofs of the result are available, none is strictly algebraic in nature; thus, we shall assume the validity of above theorem without proof. The reader will experience little difficulty, however, in establishing the following corollary.

Corollary 2.12.7. If $f(x) \in \mathbb{C}[x]$ is a polynomial of degree $n > 0$, then $f(x)$ can be expressed in $\mathbb{C}[x]$ as a product of n (not necessarily distinct) linear factors.

Another way of stating the corollary above is that the only irreducible polynomials in $\mathbb{C}[x]$ are the linear polynomials. Directing our attention now to the real field, we can obtain the form of the prime factorization in $\mathbb{R}[x]$ (bear in mind that polynomials with coefficients from $\mathbb{R}$ are polynomials in $\mathbb{C}[x]$ and therefore have roots in $\mathbb{C}$).

Corollary 2.12.8. If $f(x) \in \mathbb{R}[x]$ is of positive degree, then $f(x)$ can be factored into linear and irreducible quadratic factors.

Proof. Since $f(x)$ also belongs to $\mathbb{C}[x]$, $f(x)$ factors in $\mathbb{C}[x]$ into a product of linear polynomials $x - c_k, c_k \in C$. If $c_k \in \mathbb{R}$, then $x - c_k \in \mathbb{R}[x]$. Otherwise, $c_k = a + bi$, where $a, b \in \mathbb{R}$ and $b \neq 0$. But the complex roots of real polynomials occur in conjugate pairs (exercise), so that $\bar{c}_k = a - bi$ is also a root of $f(x)$. Thus,

$$(x - c_k)(x - \bar{c}_k) = x^2 - 2ax + \left(a^2 + b^2\right) \in \mathbb{R}[x]$$

is a factor of $f(x)$. The quadratic polynomial $x^2 - 2ax + \left(a^2 + b^2\right)$ is irreducible in $\mathbb{R}[x]$, since any factorization in $\mathbb{R}[x]$ is also valid in $\mathbb{C}[x]$ and $(x - c_k)(x - \bar{c}_k)$ is its unique factorization in $\mathbb{C}[x]$. □

An interesting remark, to be recorded without proof, is that if F is a finite field, the polynomial ring $F[x]$ contains irreducible polynomials of every degree.

This may be a convenient place to introduce the notion of a primitive polynomial.

Definition 2.12.9. Let R be a unique factorization domain. The **content** of a nonconstant polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in R[x]$, denoted by the symbol cont $f(x)$, is defined to be the greatest common divisor of its coefficients :

$$\text{cont } f(x) = \gcd(a_0, a_1, \ldots, a_n).$$

We call $f(x)$ a **primitive polynomial** if cont $f(x) = 1$.

Viewed otherwise, Definition 2.12.9 asserts that a polynomial $f(x) \in R[x]$ is primitive if and only if there is no irreducible element of R which divides all of its coefficients. In this connection, it may be noted that in the domain $F[x]$ of polynomials with coefficients from a field F, every nonconstant polynomial is primitive (indeed, there are no primes in F). The reader should also take care to remember that the notion of greatest common divisor and, in consequence, the content of a polynomial is not determined uniquely, but determined only to within associates.

Given a polynomial $f(x) \in R[x]$ of positive degree, it is possible to write $f(x) = cf_1(x)$, where $c \in R$ and $f_1(x)$ is primitive; simply let $c = \text{cont } f(x)$. To a certain extent this reduces the question of factorization in $R[x]$ (at least, when R is a unique factorization domain) to that of primitive polynomials. By way of specific illustrations, we observe that $f(x) = 3x^3 - 4x + 35$ is a primitive polynomial in $Z[x]$, while $g(x) = 12x^2 + 6x - 3 = 3\left(4x^2 + 2x - 1\right)$ is not a primitive element of the same, since $g(x)$ has content 3.

Here is another new concept: Suppose that I is a (proper) ideal of R, a commutative ring with identity. There is an obvious mapping $v : R[x] \to (R/I)[x]$; for any polynomial $f(x) \in R[x]$ simply apply nat_I to the coefficients of $f(x)$, so that

$$v(f(x)) = (a_0 + I) + (a_1 + I)\, x + \cdots + (a_n + I)\, x^n,$$

or, more briefly, $v(f(x)) = \sum (\mathrm{nat}_I\, a_k)\, x^k$. The reader will encounter no difficulty in verifying that v, defined in this way, is a homomorphism of $R[x]$ onto $(R/I)[x]$, the so-called **reduction homomorphism modulo** I. The polynomial $v(f(x))$ is said to be the **reduction of** $f(x)$ **modulo** I.

In another view, we can derive it from the substitution homomorphism. Let $\psi : R \to S$ be a ring homomorphism. Composing ψ with the inclusion of S as a subring of the polynomial ring $S[x]$, we obtain a homomorphism $\varphi : R \to S[x]$. The substitution principle asserts that there is a unique extension of φ to a homomorphism $\Phi : R[x] \to S[x]$ that sends $x \rightsquigarrow x$. This map operates on the coefficients of a polynomial, while leaving the variable x fixed. If we denote $\psi(a)$ by a', then it sends a polynomial $a_n x^n + \cdots + a_1 x + a_0$ to $a'_n x^n + \cdots + a'_1 x + a'_0$.

A particularly interesting case is that φ is the homomorphism $\mathbb{Z} \to \mathbb{F}_p$ that sends an integer a to its residue $\bar{a}$ modulo p. This map extends to a homomorphism $\Phi : \mathbb{Z}[x] \to \mathbb{F}_p[x]$, defined by

$$\begin{gathered} \Phi :\ \mathbb{Z}[x] \to \mathbb{F}_p[x] \\ f(x) = a_n x^n + \cdots + a_0 \mapsto \bar{a}_n x^n + \cdots + \bar{a}_0 = \bar{f}(x) \end{gathered} \tag{2.1}$$

where $\bar{a}_i$ is the residue class of a_i modulo p. It is natural to call the polynomial $\vec{f}(x)$ the residue of $f(x)$ modulo p.

Another example: Let R be any ring, and let P denote the polynomial ring $R[x]$. One can use the substitution principle to construct an isomorphism

$$R[x, y] \to P[y] = (R[x])[y].$$

This is stated and proved below in Proposition 2.12.10. The domain is the ring of polynomials in two variables x and y, and the range is the ring of polynomials in y whose coefficients are polynomials in x. The statement that these rings are isomorphic is a formalization of the procedure of collecting terms of like degree in y in a polynomial $f(x, y)$. For example,

$$x^4 y + x^3 - 3x^2 y + y^2 + 2 = y^2 + \left(x^4 - 3x^2\right) y + \left(x^3 + 2\right).$$

This procedure can be useful. For one thing, one may end up with a polynomial that is monic in the variable y, as happens in the example above. If so, one can do division with remainder (see Corollary 2.12.11).

Proposition 2.12.10. Let $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_n)$ denote sets of variables. There is a unique isomorphism $R[x, y] \to R[x][y]$, which is the identity on R and which sends the variables to themselves.

This is very elementary, but it would be boring to verify compatibility of multiplication in the two rings directly.

Proof. We note that since R is a subring of $R[x]$ and $R[x]$ is a subring of $R[x][y]$, R is also a subring of $R[x][y]$. Let φ be the inclusion of R into $R[x][y]$. The substitution principle tells us that there is a unique homomorphism $\Phi : R[x, y] \to R[x][y]$, which extends φ and sends the variables x_μ and y_ν wherever we want. So we can send the variables to themselves. The map Φ thus constructed is the required isomorphism. It isn't difficult to see that Φ is bijective. One way to show this would be to use the substitution principle again, to define the inverse map. □

Corollary 2.12.11. Let $f(x, y)$ and $g(x, y)$ be polynomials in two variables, elements of $R[x, y]$. Suppose that, when regarded as a polynomial in y, f is a monic polynomial of degree m. There are uniquely determined polynomials $q(x, y)$ and $r(x, y)$ such that $g = fq + r$, and such that if $r(x, y)$ is not zero, its degree in the variable y is less than m.

Although it might seem to be rather special, the reduction homomorphism will serve us in good stead on several occasions. We make immediate use of it to characterize primitive polynomials.

Theorem 2.12.12. Let R be a unique factorization domain and let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$, with $\deg f(x) > 0$. Then $f(x)$ is a primitive polynomial in $R[x]$ if and only if, for each prime element $p \in R$, the reduction of $f(x)$ modulo the principal ideal (p) is nonzero.

Proof. By definition, the reduction of $f(x)$ modulo (p) is

$$v(f(x)) = (a_0 + (p)) + (a_1 + (p))\, x + \cdots + (a_n + (p))\, x^n.$$

Thus, to say that $v(f(x)) = 0$ for some prime $p \in R$ is equivalent to asserting that $a_k \in (p)$, or rather, $p \mid a_k$ for all k. But the latter condition signifies that cont $f(x) \neq 1$; hence, $f(x)$ is not primitive. □

One of the most crucial facts concerning primitive polynomials is Gauss's Lemma, which we prove next.

Theorem 2.12.13 (Gauss's Lemma)**.** Let R be a unique factorization domain. If $f(x), g(x)$ are both primitive polynomials in $R[x]$, then their product $f(x)g(x)$ is also primitive in $R[x]$.

Proof. Given a prime element $p \in R$, (p) is a prime ideal of R, whence the quotient ring $R' = R/(p)$ forms an integral domain. We next consider the reduction homomorphism v modulo the principal ideal (p). Since $R'[x]$ is an integral domain, it follows that the reduction of $f(x)g(x)$ cannot be the zero polynomial:

$$v(f(x)g(x)) = v(f(x))v(g(x)) \neq 0.$$

The assertion of the theorem is now a direct consequence of our last result. □

Corollary 2.12.14 (Content is multiplicative)**.** If R is a unique factorization domain and $f(x), g(x) \in R[x]$, then

$$\text{cont } (f(x)g(x)) = \text{ cont } f(x) \text{ cont } g(x).$$

Proof. As noted earlier, we can write $f(x) = af_1(x), g(x) = bg_1(x)$, where $a =$ cont $f(x), b =$ cont $g(x)$ and where $f_1(x), g_1(x)$ are primitive in $R[x]$. Therefore, $f(x)g(x) = abf_1(x)g_1(x)$. According to the theorem, the product $f_1(x)g_1(x)$ is a primitive polynomial of $R[x]$. This entails that the content of $f(x)g(x)$ is simply ab, or, what amounts to the same thing, cont $f(x)$ cont $g(x)$. □

Any unique factorization domain R, being an integral domain, possesses a field of quotients (field of fractions) $K = Q_{\text{cl}}(R)$ and we may consider the ring of polynomials $R[x]$ as imbedded in the polynomial ring $K[x]$. The next theorem deals with the relation between the irreducibility of a polynomial in $R[x]$ as compared to its irreducibility when considered as an element of the larger ring $K[x]$. (The classic example of this situation is, of course, the polynomial domain $\mathbb{Z}[x] \subseteq \mathbb{Q}[x]$.) Before concentrating our efforts on this relationship, we require a preliminary lemma.

Lemma 2.12.15. Let R be a unique factorization domain, with field of quotients K. Given a nonconstant polynomial $f(x) \in K[x]$, there exist (nonzero) elements $a, b \in R$ and a primitive polynomial $f_1(x)$ in $R[x]$ such that

$$f(x) = ab^{-1}f_1(x).$$

Furthermore, $f_1(x)$ is unique up to invertible elements of R as factors.

Proof. Inasmuch as K is the field of quotients of R, $f(x)$ can be written in the form

$$f(x) = \left(a_0b_0^{-1}\right) + \left(a_1b_1^{-1}\right) x + \cdots + \left(a_nb_n^{-1}\right) x^n,$$

where $a_i, b_i \in R$ and $b_i \neq 0$. Take b to be any common multiple of the b_i; for instance, $b = b_0 b_1 \cdots b_n$. Then $b \neq 0$ and, since the coefficients of $\boldsymbol{b}f(x)$ all lie in R, we have $bf(x) = g(x) \in R[x]$. Accordingly,

$$f(x) = b^{-1}g(x) = ab^{-1}f_1(x),$$

where $f_1(x) \in R[x]$ is a primitive polynomial and $a = \operatorname{cont} g(x)$. We emphasize that $f_1(x)$ is of the same degree as $f(x)$, so cannot be invertible in $R[x]$.

As for uniqueness, suppose that $f(x) = ab^{-1}f_1(x) = cd^{-1}f_2(x)$ are two representations that satisfy the conditions of the theorem. Then,

$$adf_1(x) = bcf_2(x).$$

Since $f_1(x)$ and $f_2(x)$ are both primitive, the corollary to Gauss's Lemma implies that we must have $ad = ubc$ for some invertible element $u \in R$. In consequence, $f_1(x) = uf_2(x)$, showing that $f_1(x)$ is unique to within invertible factors in R. □

Corollary 2.12.16.

(a) Let f_0 be a primitive polynomial, and let $g \in R[x]$. If f_0 divides g in $K[x]$, then f_0 divides g in $R[x]$. The converse is also true (obviously).

(b) If two polynomials f and g in $R[x]$ have a common nonconstant factor in $K[x]$, they have a common nonconstant factor in $R[x]$.

Proof.
(a) f_0 divides g in $K[x]$, then $g = f_0 q$ where $q \in K[x]$. We want to show that $q \in R[x]$. We write $g = cg_1$, and $q = c'q_1$, with g_1 and q_1 primitives by above theorem. Then $cg_1 = c'f_1q_1$. Gauss's Lemma tells us that f_1q_1 is primitive. Therefore by the uniqueness assertion of above theorem, $c = c'$ and $g_1 = f_1q_1$. Since $g \in R[x]$, $c \in R$, we see $q = cq_1 \in R[x]$.

(b) If the two polynomials f and g in $R[x]$ have a common factor h in $K[x]$ and if we write $h = ch_1$, where h_1 is primitive, then h_1 also divides f and g in $K[x]$, and by (a), h_1 divides both f and g in $R[x]$. □

Theorem 2.12.17. Let R be a unique factorization domain, with field of quotients K. If $f(x) \in R[x]$ is an irreducible primitive polynomial, then it is also irreducible as an element of $K[x]$.

Proof. Assume to the contrary that $f(x)$ is reducible over K. Then, $f(x) = g(x)h(x)$, where the polynomials $g(x), h(x)$ are in $K[x]$ and are of positive degree. By virtue of the lemma just proven,

$$g(x) = ab^{-1}g_1(x), \quad h(x) = cd^{-1}h_1(x),$$

with $a, b, c, d \in R$ and $g_1(x), h_1(x)$ primitive in $R[x]$. Thus,

$$bdf(x) = \operatorname{acg}_1(x)h_1(x).$$

Now, Gauss's Lemma asserts that the product $g_1(x)h_1(x)$ is a primitive polynomial in $R[x]$, whence $f(x)$ and $g_1(x)h_1(x)$ differ by an invertible element of R :

$$f(x) = ug_1(x)h_1(x).$$

Since $\deg g_1(x) = \deg g(x) > 0, \deg h_1(x) = \deg h(x) > 0$, the outcome is a nontrivial factorization of $f(x)$ in $R[x]$, contrary to hypothesis. □

Remark 2.12.18. There is an obvious converse to above theorem viz.: if the primitive polynomial $f(x) \in R[x]$ is irreducible as an element of $K[x]$, it is also irreducible in $R[x]$. This is justified by the fact that $R[x]$ (or an isomorphic copy thereof) appears naturally as a subring of $K[x]$; thus, if $f(x)$ were reducible in $R[x]$, it would obviously be reducible in the larger ring $K[x]$.

Our remarks lead to the following conclusion:

Theorem 2.12.19. Given a primitive polynomial $f(x) \in R[x], R$ a unique factorization domain, $f(x)$ is irreducible in $R[x]$ if and only if $f(x)$ is irreducible in $K[x]$. The irredciblle elements of $R[x]$ are of two types: irreducible elements of R, and primitive elements of $R[x]$ that are irreducible in $K[x]$.

Proposition 2.12.20. If $f \in \mathbb{Z}[X]$ is monic, then every monic factor of f in $\mathbb{Q}[X]$ lies in $\mathbb{Z}[X]$.

Proof. Let g be a monic factor of f in $\mathbb{Q}[X]$, so that $f = gh$ with $h \in \mathbb{Q}[X]$ also monic. Let m, n be the positive integers with the fewest prime factors such that $mg, nh \in \mathbb{Z}[X]$. As in the proof of Gauss's Lemma, if a prime p divides mn, then it divides all the coefficients of at least one of the polynomials mg, nh, say mg, in which case it divides m because g is monic. Now $\frac{m}{p} g \in \mathbb{Z}[X]$, which contradicts the definition of m. □

Theorem 2.12.21. Let R be an integral domain and the nonconstant polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in R[x]$. Suppose that there exists a prime ideal P of R such that

(1) $a_n \notin P$,

(2) $a_k \in P$ for $0 \leqslant k < n$,

(3) $a_0 \notin P^2$.

Then $f(x)$ is irreducible in $R[x]$.

Proof. Assume that, contrary to assertion, $f(x)$ is reducible in $R[x]$; say, $f(x) = g(x)h(x)$ for polynomials $g(x), h(x) \in R[x]$, where

$$\begin{aligned} g(x) &= b_0 + b_1 x + \cdots + b_r x^r \\ h(x) &= c_0 + c_1 x + \cdots + c_s x^s \quad (r + s = n; r, s > 0) \end{aligned}$$

Now consider the reduction of $f(x)$ modulo the ideal P. Invoking hypothesis (2), it can be inferred that

$$v(g(x))v(h(x)) = v(f(x)) = (a_n + P)\, x^n.$$

Since the polynomial ring $(R/P)[x]$ comprises an integral domain, the only possible factorizations of $(a_n + P)\, x^n$ are into linear factors. This being so, a moment's reflection shows that

$$\begin{aligned} v(g(x)) &= (b_r + P)\, x^r, \\ v(h(x)) &= (c_s + P)\, x^s. \end{aligned}$$

This means that the constant terms of these reductions are zero; that is,

$$b_0 + P = c_0 + P = P.$$

Altogether we have proved that both $b_0, c_0 \in P$, revealing at the same time that $a_0 = b_0 c_0 \in P^2$, which is untenable by (3). Accordingly, no such factorization of $f(x)$ can occur, and $f(x)$ is indeed irreducible in $R[x]$. □

Immediately follows is that

Theorem 2.12.22 (Eisenstein Criterion)**.** Let R be a unique factorization domain and K be its field of quotients. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a nonconstant polynomial in $R[x]$. Suppose further that for some prime $p \in R$,

(1) $p \nmid a_n$,

(2) $p \mid a_k$ for $0 \leqslant k < n$,

(3) $p^2 \nmid a_0$.

Then, $f(x)$ is irreducible in $K[x]$.

Proof. We already know that (p) is a prime ideal of R. Taking stock of the theorem, $f(x)$ is an irreducible polynomial of $R[x]$; hence, of $K[x]$ (at this point a direct appeal is made to Theorem 2.12.17). □

This is probably a good time at which to examine some examples.

Example 2.12.23. $x^2 + y^2 + 1$ is irreducible in $\mathbb{C}[x, y]$.

Proof. Let $R = \mathbb{C}[x]$, which is UFD. We then use Eisenstein's criterion to show that $y^2 + (x^2 + 1) = y^2 + a_0$ where $a_0 = x^2 + 1$ is irreducible in $K[y]$ where K is the field of fractions of R, which then implies that it is irreducible in $R[y]$ due to remark 2.12.18. $a_2 = 1, a_1 = 0, a_0 = x^2 + 1$. Let $p = x + i \in R$, then $y^2 + a_0$ is irredciblle as

(1) $x + i \nmid 1$,

(2) $p \mid 0 = 0(x + i)$, $p \mid x^2 + 1 = (x + i)(x - i)$,

(3) $p^2 = (x + i)^2 \nmid a_0 = (x + i)(x - i)$ as $(x + i) \nmid (x - i)$.

□

Example 2.12.24. Consider the monic polynomial

$$f(x) = x^n + a \in \mathbb{Z}[x] \quad (n > 1),$$

where $a \neq \pm 1$ is a nonzero square-free integer. For any prime p dividing a, p is certainly a factor of all the coefficients except the leading one, and our hypothesis ensures that $p^2 \nmid a$. Thus, $f(x)$ fulfils Eisenstein's criterion, and so is irreducible over $\mathbb{Q}$. Incidentally, this example shows that there are irreducible polynomials in $\mathbb{Q}[x]$ of every degree.

On the other hand, notice that $x^4 + 4 = \left(x^2 + 2x + 2\right)\left(x^2 - 2x + 2\right)$; one should not expect Theorem 2.12.21 to lead to a decision in this case, since, of course, 4 fails to be a square-free integer.

Example 2.12.25. Eisenstein's test is not directly applicable to the cyclotonic polynomial

$$f(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1 \in \mathbb{Z}[x], \ p \text{ prime}.$$

because no suitable prime is available. This problem is resolved by the observation that $f(x)$ is irreducible in $\mathbb{Z}[x]$ if and only if $f(x + 1)$ is irreducible. That's because f reducible $\implies f(x) = h(x)g(x) \implies f(x + 1) = h(x + 1)g(x + 1) = u(x)v(x) \implies f(x + 1)$ reducible; $f(x + 1) = u(x)v(x) \implies f(x) = f((x - 1) + 1) = u(x - 1)v(x - 1)$. Now, a simple computation yields

$$f(x + 1) = \sum_{i=0}^{p-1} (x + 1)^i = \sum_{i=0}^{p-1} \sum_{j=0}^{i} \binom{i}{j} x^j = \sum_{j=0}^{p-1} \left(\sum_{i=j}^{p-1} \binom{i}{j} \right) x^j$$

By combinatorial identity $\binom{j}{j} + \cdots + \binom{m}{j} = \binom{m+1}{j+1}$, $c_j = \sum_{i=j}^{p-1} \binom{i}{j} = \binom{p}{j+1} = \frac{p!}{(j+1)!(p-j-1)!}$. Thus $p \,|\, c_j$ for $0 \leqslant j < p-1$. Also, $p \nmid c_{p-1} = \binom{p-1}{p-1} = 1$ and $p^2 \nmid c_0 = \binom{p}{1} = 1$. Eisenstein criterion then concludes that $f(x+1)$ is irreducible, so $f(x)$ irreducible.

2.12 EXERCISES

1. [1] p.379 Ex3.1. Let φ denote the homomorphism $\mathbb{Z}[x] \to \mathbb{R}$ defined by

i. $\varphi(x) = 1 + \sqrt{2}$,
ii. $\varphi(x) = \frac{1}{2} + \sqrt{2}$.

Is the kernel of φ a principal ideal? If so, find a generator.

2. [1] p.379 Ex3.2. Prove that two integer polynomials are relatively prime elements of $\mathbb{Q}[x]$ if and only if the ideal they generate in $\mathbb{Z}[x]$ contains an integer.

3. [1] p.379 Ex3.4. Let x, y, z, w be variables. Prove that $xy - zw$, the determinant of a variable 2×2 matrix, is an irreducible element of the polynomial ring $\mathbb{C}[x, y, z, w]$.

2.13 Factoring Rational and Integer Polynomials

Every monic polynomial $f(x)$ with rational coefficients can be expressed uniquely in the form $p_1 \cdots p_k$, where p_i are monic polynomials that are irreducible elements in the ring $\mathbb{Q}[x]$.

Example 2.13.1. Here are some examples of irreducible elements in $\mathbb{Q}[x]$:

- Linear Polynomials: Any linear polynomial $ax + b$ (with $a \neq 0$) is irreducible in $\mathbb{Q}[x]$ because it cannot be factored further into non-constant polynomials with rational coefficients. For example $2x + 3$ is irreducible.
- Quadratic Polynomials: A quadratic polynomial $ax^2 + bx + c$ is irreducible in $\mathbb{Q}[x]$ if its discriminant $b^2 - 4ac$ is not a perfect square in $\mathbb{Q}$. For example, $x^2 + x + 1$ is irreducible in $\mathbb{Q}[x]$ because its discriminant $1^2 - 4 \cdot 1 \cdot 1 = -3$ is not a perfect square.
- Cubic Polynomials: A cubic polynomial $ax^3 + bx^2 + cx + d$ may be irreducible if it does not have a rational root (which can be checked using the Rational Root Theorem) and cannot be factored into a product of a linear and a quadratic polynomial with rational coefficients. For example, $x^3 + 2x + 1$ is irreducible in $\mathbb{Q}[x]$.

Algorithm for factoring a polynomial in $\mathbb{Q}[x]$: To see this, consider $f \in \mathbb{Q}[x]$. Multiply $f(x)$ by a rational number so that it is monic, and then replace it by $D^{\deg(f)} f\left(\frac{x}{D}\right)$, with D equal to a common denominator for the coefficients of f, to obtain a monic polynomial with integer coefficients. Thus we need consider only polynomials

$$f(x) = x^m + a_1 x^{m-1} + \cdots + a_m, \quad a_i \in \mathbb{Z}.$$

From the fundamental theorem of algebra, we know that f splits completely in $\mathbb{C}[x]$:

$$f(x) = \prod_{i=1}^{m} (x - \alpha_i), \quad \alpha_i \in \mathbb{C}.$$

From the equation

$$0 = f(\alpha_i) = \alpha_i^m + a_1 \alpha_i^{m-1} + \cdots + a_m,$$

it follows that $|\alpha_i|$ is less than some bound depending only on the degree and coefficients of f; in fact,

$$|\alpha_i| \leqslant \max\{1, mB\}, B = \max|a_i|.$$

Now if $g(x)$ is a monic factor of $f(x)$, then its roots in $\mathbb{C}$ are certain of the α_i, and its coefficients are symmetric polynomials in its roots. Therefore, the absolute values of the coefficients of $g(x)$ are bounded in terms of the degree and coefficients of f. Since they are also integers (by proposition2.12.20), we see that there are only finitely many possibilities for $g(x)$. Thus, to find the factors of $f(x)$ we (better PARI) have to do only a finite amount of checking.

Therefore, we need not concern ourselves with the problem of factoring polynomials in the rings $\mathbb{Q}[X]$ or $\mathbb{F}_p[X]$ since PARI knows how to do it. For example, typing `content(6*X^2+18*X-24)` in PARI returns 6, and `factor(6*X^2+18*X-24)` returns $X-1$ and $X+4$, showing that

$$6X^2 + 18X - 24 = 6(X-1)(X+4)$$

in $\mathbb{Q}[X]$. Typing `factormod(X^2+3*X+3,7)` returns $X+4$ and $X+6$, showing that

$$X^2 + 3X + 3 = (X+4)(X+6)$$

in $\mathbb{F}_7[X]$.

```
sage: R.<x> = PolynomialRing(ZZ)
sage: (2*x^2 - 4*x^4 + 14*x^7).content()
2
```

Figure 2.1: SageMath example. See manual.

```
sage: R.<x> = ZZ[]
sage: f = x^4 - 1
sage: f.factor()
(x - 1) * (x + 1) * (x^2 + 1)
```

Figure 2.2: SageMath example. See manual.

More examples can be seen in the link.

We have shown that The polynomial ring $\mathbb{Z}[x]$ is also a unique factorization domain. That is, nonzero polynomial $f(x) \in \mathbb{Z}[x]$ that is not ± 1 can be written as a product

$$f(x) = \pm p_1 \cdots p_m q_1(x) \cdots q_n(x),$$

where p_i are integer primes and $q_j(x)$ are primitive irreducible polynomials. This expression is unique except for the order of the factors.

We have two main tools for studying factoring in $\mathbb{Z}[x]$. The first is the inclusion of the integer polynomial ring into the ring of polynomials with rational coefficients:

$$\mathbb{Z}[x] \subset \mathbb{Q}[x].$$

This can be useful because algebra in the ring $\mathbb{Q}[x]$ is simpler. The second tool is reduction modulo some integer prime p, the homomorphism

$$\psi_p : \mathbb{Z}[x] \to \mathbb{F}_p[x]$$

that sends $x \rightsquigarrow x$. We'll often denote the image $\psi_p(f)$ of an integer polynomial by $\bar{f}$, though this notation is ambiguous because it doesn't mention p. The next lemma should be clear.

Lemma 2.13.2. Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ be an integer polynomial, and let p be an integer prime. The following are equivalent:

- p divides every coefficient a_i of f in $\mathbb{Z}$,
- p divides f in $\mathbb{Z}[x]$,
- f is in the kernel of ψ_p.

The lemma shows that the kernel of ψ_p can be interpreted easily without mentioning the map. But the facts that ψ_p is a homomorphism and that its image $\mathbb{F}_p[x]$ is an integral domain make the interpretation as a kernel useful.

We pose the problem of factoring an integer polynomial

$$f(x) = a_n x^n + \cdots + a_1 x + a_0,$$

with $a_n \neq 0$. Linear factors can be found fairly easily.

Lemma 2.13.3.

(a) If an integer polynomial $b_1 x + b_0$ divides f in $\mathbb{Z}[x]$, then b_1 divides a_n and b_0 divides a_0.

(b) A primitive polynomial $b_1 x + b_0$ divides f in $\mathbb{Z}[x]$ if and only if the rational number $-b_0/b_1$ is a root of f.

(c) A rational root of a monic integer polynomial f is an integer.

Proof. (a) The constant coefficient of a product $(b_1 x + b_0)\left(q_{n-1}x^{n-1} + \cdots + q_0\right)$ is $b_0 q_0$, and if $q_{n-1} \neq 0$, the leading coefficient is $b_1 q_{n-1}$.

(b) According to Corollary 2.12.16, $b_1 x + b_0$ divides f in $\mathbb{Z}[x]$ if and only if it divides f in $\mathbb{Q}[x]$, and this is true if and only if $x + b_0/b_1$ divides f, i.e., $-b_0/b_1$ is a root.

(c) If $\alpha = a/b$ is a root, written with $b > 0$, and if $\gcd(a, b) = 1$, then $bx - a$ is a primitive polynomial that divides the monic polynomial f, so $b = 1$ and α is an integer. □

Corollary 2.13.4 (Rational Root Theorem)**.** Suppose we have a rational $-b_0/b_1$ written in lowest terms so that b_1 and b_0 are relatively prime (i.e., $b_1 x + b_0$ primitive). Thus it is a root of $f \in \mathbb{Z}[x]$ iff $b_1 x + b_0$ divides f due to (b), which by (a) implies that b_1 divides a_n and b_0 divides a_0.

The homomorphism $\psi_p : \mathbb{Z}[x] \to \mathbb{F}_p[x]$ (eq. (2.1)) is useful for explicit factoring, one reason being that there are only finitely many polynomials in $\mathbb{F}_p[x]$ of each degree.

Proposition 2.13.5. Let $f(x) = a_n x^n + \cdots + a_0$ be an integer polynomial, and let p be a prime integer that does not divide the leading coefficient a_n. If the residue $\bar{f}$ of f modulo p is an irreducible element of $\mathbb{F}_p[x]$, then f is an irreducible element of $\mathbb{Q}[x]$.

Proof. We prove the contrapositive, that if f is reducible, then $\bar{f}$ is reducible. Suppose that $f = gh$ is a proper factorization of f in $\mathbb{Q}[x]$. We may assume that g and h are in $\mathbb{Z}[x]$ (cor 2.12.16). Since the factorization in $\mathbb{Q}[x]$ is proper, both g and h have positive degree, and, if $\deg f$ denotes the degree of f, then $\deg f = \deg g + \deg h$.

Since ψ_p is a homomorphism, $\bar{f} = \bar{g}\bar{h}$, so $\deg \bar{f} = \deg \bar{g} + \deg \bar{h}$. For any integer polynomial $p, \deg \bar{p} \leqslant \deg p$. Our assumption on the leading coefficient of f tells us that $\deg \bar{f} = \deg f$. This being so we must have $\deg \bar{g} \cdot = \deg g$ and $\deg \bar{h} = \deg h$. Therefore the factorization $\bar{f} = \bar{g}\bar{h}$ is proper. □

If p divides the leading coefficient of f, then $\bar{f}$ has lower degree, and using reduction modulo p becomes harder.

If we suspect that an integer polynomial is irreducible, we can try reduction modulo p for a small prime, $p = 2$ or 3 for instance, and hope that $\bar{f}$ turns out to be irreducible and of the same degree as f. If so, f will be irreducible too. Unfortunately, there exist irreducible integer polynomials that can be factored modulo every prime p. The polynomial $x^4 - 10x^2 + 1$ is an example. So the method of reduction modulo p may not work. But it does work quite often.

The irreducible polynomials in $\mathbb{F}_p[x]$ can be found by the "sieve" method. The sieve of Eratosthenes is the name given to the following method of determining the prime integers less than a given number n. We list the integers from 2 to n. The first one, 2 , is prime because any proper factor of 2 must be smaller than 2, and there is no smaller integer on our list. We note that 2 is prime, and we cross out the multiples of 2 from our list. Except for 2 itself, they are not prime. The first integer that is left, 3, is a prime because it isn't divisible by any smaller prime. We note that 3 is a prime and then cross out the multiples of 3 from our list. Again, the smallest remaining integer, 5, is a prime, and so on.

The same method will determine the irreducible polynomials in $\mathbb{F}_p[x]$. We list the monic polynomials, degree by degree, and cross out products. For example, the linear polynomials in $\mathbb{F}_2[x]$ are x and $x + 1$. They are irreducible. The polynomials of degree 2 are $x^2, x^2 + x, x^2 + 1$, and $x^2 + x + 1$. The first three have roots in $\mathbb{F}_2$, so they are divisible by x or by $x + 1$. The last one, $x^2 + x + 1$, is the only irreducible polynomial of degree 2 in $\mathbb{F}_2[x]$. The irreducible polynomials of degree $\leqslant 4$ in $\mathbb{F}_2[x]$:

$$x, \quad x+1; \quad x^2+x+1; \quad x^3+x^2+1, \quad x^3+x+1;$$
$$x^4+x^3+1, x^4+x+1, x^4+x^3+x^2+x+1.$$

By trying the polynomials on this list, we can factor polynomials of degree at most 9 in $\mathbb{F}_2[x]$. For example, let's factor $f(x) = x^5 + x^3 + 1$ in $\mathbb{F}_2[x]$. If it factors, there must be an irreducible factor of degree at most 2. Neither 0 nor 1 is a root, so f has no linear factor. There is only one irreducible polynomial of degree 2, namely $p = x^2 + x + 1$. We carry out division with remainder: $f(x) = p(x)\left(x^3 + x^2 + x\right) + (x + 1)$. So p doesn't divide f, and therefore f is irreducible.

Consequently, the integer polynomial $x^5 - 64x^4 + 127x^3 - 200x + 99$ is irreducible in $\mathbb{Q}[x]$, because its residue in $\mathbb{F}_2[x]$ is the irreducible polynomial $x^5 + x^3 + 1$. The monic irreducible polynomials of degree 2 in $\mathbb{F}_3[x]$:

$$x^2 + 1, \quad x^2 + x - 1, \quad x^2 - x - 1.$$

2.13 EXERCISES

1. [1] p.380 Ex4.1. (a) Factor $x^9 - x$ and $x^9 - 1$ in $\mathbb{F}_3[x]$. (b) Factor $x^{16} - x$ in $\mathbb{F}_2[x]$.

2. [1] p.380 Ex4.2. Prove that the following polynomials are irreducible:

i. $x^2 + 1$, in $\mathbb{F}_7[x]$,
ii. $x^3 - 9$, in $\mathbb{F}_{31}[x]$.

3. [1] p.380 Ex4.3. Decide whether or not the polynomial $x^4 + 6x^3 + 9x + 3$ generates a maximal ideal in $\mathbb{Q}[x]$.

4. [1] p.380 Ex4.4. Factor the integer polynomial $x^5 + 2x^4 + 3x^3 + 3x + 5$ modulo 2 , modulo 3 , and in $\mathbb{Q}$.

5. Which of the following polynomials are irreducible in $\mathbb{Q}[x]$?

i. $x^2 + 27x + 213$,
ii. $8x^3 - 6x + 1$,
iii. $x^3 + 6x^2 + 1$
iv. $x^5 - 3x^4 + 3$.

6. [1] p.380 Ex4.5. Factor $x^5 + 5x + 5$ into irreducible factors in $\mathbb{Q}[x]$ and in $\mathbb{F}_2[x]$.

7. [1] p.380 Ex4.10. Factor the following polynomials in $\mathbb{Q}[x]$. (a) $x^2 + 2351x + 125$ (b) $x^3 + 2x^2 + 3x + 1$, (c) $x^4 + 2x^3 + 2x^2 + 2x + 2$, (d) $x^4 + 2x^3 + 3x^2 + 2x + 1$, (e) $x^4 + 2x^3 + x^2 + 2x + 1$, (f) $x^4 + 2x^2 + x + 1$, (g) $x^8 + x^6 + x^4 + x^2 + 1$, (h) $x^6 - 2x^5 - 3x^2 + 9x - 3$, (j) $x^4 + x^2 + 1$, (k) $3x^5 + 6x^4 + 9x^3 + 3x^2 - 1$, (l) $x^5 + x^4 + x^2 + x + 2$.

Chapter 3

Modules

We will extensively copy from [10] for modules.

3.1 Categories and Functors

Definition 3.1.1. A **category** $\mathcal{C}$ consists of three ingredients: a class $\mathrm{obj}(\mathcal{C})$ of **objects**, a set of **morphisms** $\mathrm{Hom}(A,B)$ for every ordered pair (A,B) of objects, and **composition** $\mathrm{Hom}(A,B)\times\mathrm{Hom}(B,C)\to\mathrm{Hom}(A,C)$, denoted by

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

for every ordered triple A,B,C of objects. [We often write $f:A\to B$ or $A\xrightarrow{f}B$ instead of $f\in\mathrm{Hom}(A,B)$.] These ingredients are subject to the following axioms:

(i) the Hom sets are pairwise disjoint; that is, each $f\in\mathrm{Hom}(A,B)$ has a unique **domain** A and a unique **target** B;

(ii) for each object A, there is an **identity morphism** $1_A\in\mathrm{Hom}(A,A)$ such that $f1_A=f$ and $1_Bf=f$ for all $f:A\to B$;

(iii) composition is associative: given morphisms $A\xrightarrow{f}B\xrightarrow{g}C\xrightarrow{h}D$, then

$$h(gf)=(hg)f.$$

Example 3.1.2.

1. **Sets**. The objects in this category are sets (not proper classes), morphisms are functions, and composition is the usual composition of functions.

   It is an axiom of set theory that if A and B are sets, then the class $\mathrm{Hom}(A,B)$ of all functions from A to B is also a set. That Hom sets are pairwise disjoint is just a reflection of the definition of equality of functions, which says that two functions are equal if they have the same domains and the same targets (as well as having the same graphs). For example, if $U\subsetneq X$ is a proper subset of a set X, then the inclusion function $U\to X$ is distinct from the identity function 1_U, for they have different targets. If $f:A\to B$ and $g:C\to D$ are functions, we define their composite $gf:A\to D$ if $B=C$. In contrast, in Analysis, one often says gf is defined when $B\subseteq C$. We do not recognize this; for us, gf is not defined, but gif is defined, where $i:B\to C$ is the inclusion.

2. **Groups**. Objects are groups, morphisms are homomorphisms, and composition is the usual composition (homomorphisms are functions). Part of the verification that **Groups** is a category involves

checking that identity functions are homomorphisms and that the composite of two homomorphisms is itself a homomorphism [one needs to know that if $f \in \mathrm{Hom}(A,B)$ and $g \in \mathrm{Hom}(B,C)$, then $gf \in \mathrm{Hom}(A,C)$].

3. A partially ordered set X can be regarded as the category whose objects are the elements of X, whose Hom sets are either empty or have only one element:

$$\mathrm{Hom}(x,y) = \begin{cases} \varnothing & \text{if } x \not\preceq y \\ \{\iota_y^x\} & \text{if } x \preceq y \end{cases}$$

(the symbol ι_y^x is the unique element in the Hom set when $x \preceq y$), and whose composition is given by $\iota_z^y \iota_y^x = \iota_z^x$. Note that $1_x = \iota_x^x$, by reflexivity, while composition makes sense because $\preceq$ is transitive. The converse is false: if $\mathcal{C}$ is a category with $|\mathrm{Hom}(x,y)| \leqslant 1$ for every $x, y \in \mathrm{obj}(\mathcal{C})$, define $x \preceq y$ if $\mathrm{Hom}(x,y) \neq \varnothing$. Then $\mathcal{C}$ may not be partially ordered because $\preceq$ need not be antisymmetric. The two-point category $\bullet \leftrightarrows \bullet$ having only two nonidentity morphisms is such an example that is not partially ordered.

We insisted, in the definition of category, that each $\mathrm{Hom}(A,B)$ be a set, but we did not say it was nonempty. The category X, where X is a partially ordered set, is an example in which this possibility occurs. [Not every Hom set in a category $\mathcal{C}$ can be empty, for $1_A \in \mathrm{Hom}(A,A)$ for every $A \in \mathrm{obj}(\mathcal{C})$.]

4. Let X be a topological space, and let $\mathcal{U}$ denote its topology; that is, $\mathcal{U}$ is the family of all the open subsets of X. Then $\mathcal{U}$ is a partially ordered set under ordinary inclusion, and so it is a category as in part 3. In this case, we can realize the morphism ι_V^U, when $U \subseteq V$, as the inclusion $i_V^U : U \to V$.

5. View a natural number $n \geqslant 1$ as the partially ordered set whose elements are $0, 1, \ldots, n-1$ and $0 \leqslant 1 \leqslant \cdots \leqslant n-1$. As in part 3, there is a category $\mathbf{n}$ with $\mathrm{obj}(\mathbf{n}) = \{0, 1, \ldots, n-1\}$ and with morphisms $i \to j$ for all $0 \leqslant i \leqslant j \leqslant n-1$.

6. Let S be a set with a relation $\sim$ that is reflexive and transitive, and $\mathcal{C}$ is a category $obj(\mathcal{C})$. $\mathrm{Hom}_{\mathcal{C}}(a,b) = \phi$ if $a \not\sim b$ and $\{(a,b)\}$ if $a \sim b$.

$a \in obj(\mathcal{C}), 1_a = (a,a)$ with composition $(a,b) \in \mathrm{Hom}(a,b), (b,c) \in \mathrm{Hom}(b,c)$ therefore $(b,c)(a,b) = (a,c)$.

7. Let $\mathcal{C}$ be a category, $A \in \mathrm{obj}(\mathcal{C})$ and $\mathcal{C}_A$ be a new catory, where objects are morphism from any object of $\mathcal{C}$ to A.

$$\mathrm{Hom}_{\mathcal{C}_A}(f,g) = \{\sigma \in \mathrm{Hom}_{\mathcal{C}}(B,C) \,\big|\, g\sigma = f\}$$

and $\mathrm{Hom}_{\mathcal{C}_A}(f,g) \times \mathrm{Hom}_{\mathcal{C}_A}(g,h) \to \mathrm{Hom}_{\mathcal{C}_A}(f,h), (\sigma,\alpha) \mapsto \alpha\sigma$. So $h(\alpha\sigma) = (h\alpha)\sigma = g\sigma = f$, and $1_B f = f$.

8. **Top**. Objects are all topological spaces, morphisms are continuous functions, and composition is the usual composition of functions. In checking that **Top** is a category, one must note that identity functions are continuous and that composites of continuous functions are continuous.

9. **Ab**. Objects are abelian groups, morphisms are homomorphisms, and composition is the usual composition.

10. **Rings**. Objects are rings, morphisms are ring homomorphisms, and composition is the usual composition. We assume that all rings R have a unit element 1 , but we do not assume that $1 \neq 0$. (Should $1 = 0$, however, the equation $1r = r$ for all $r \in R$ shows that $R = \{0\}$, because $0r = 0$. In this case, we call R the zero ring.) We agree, as part of the definition, that $\varphi(1) = 1$ for every ring homomorphism φ. Since the inclusion map $S \to R$ of a subring should be a homomorphism, it follows that the unit element 1 in a subring S must be the same as the unit element 1 in R. Category $\mathcal{C}$ with $obj(\mathcal{C})$ the commutative rings is termed **ComRings**

Definition 3.1.3. Let $\mathcal{C}$ be a category, $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$. Then f is an **isomorphism** if it has a two-sided inverse under composition with $g \in \mathrm{Hom}(B, A)$ so that $gf = 1_A, fg = 1_B$. This inverse is unique, and is denoted by f^{-1}. This has the properties that

- $(1_A)^{-1} = 1_A$
- $(fg)^{-1} = g^{-1}f^{-1}$
- $(f^{-1})^{-1} = f$

Example 3.1.4.

- If $\mathcal{C}$ is a set, then isomorphism are bijections.
- $\sim$ on S: (a, b) is an isomorphism $\iff b \sim a$

Definition 3.1.5. $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$ is a **monomorphism** if $\forall C \in \mathrm{obj}(\mathcal{C})$ and $g_1, g_2 \in \mathrm{Hom}_{\mathcal{C}}(A, C)$ with fg_1, fg_1, we have $g_1 = g_2$. f is an **epimorophism** if $\forall C \in \mathrm{obj}(\mathcal{C}), h_1, h_2 \in \mathrm{Hom}_{\mathcal{C}}(B, C)$ with $h_1 f = h_2 f$, we have $h_1 = h_2$

Example 3.1.6.

- For $\mathcal{C}$ a set, a monomorphism is injective and epimorphism is surjective.
- For $S, \sim$, all morphisms are monomorphism and epimorphism.

Definition 3.1.7. A category $\mathcal{S}$ is a subcategory of a category $\mathcal{C}$ if

(i) $\mathrm{obj}(\mathcal{S}) \subseteq \mathrm{obj}(\mathcal{C})$,

(ii) $\mathrm{Hom}_{\mathcal{S}}(A, B) \subseteq \mathrm{Hom}_{\mathcal{C}}(A, B)$ for all $A, B \in \mathrm{obj}(\mathcal{S})$, where we denote Hom sets in $\mathcal{S}$ by $\mathrm{Hom}_{\mathcal{S}}(\square, \square)$,

(iii) if $f \in \mathrm{Hom}_{\mathcal{S}}(A, B)$ and $g \in \mathrm{Hom}_{\mathcal{S}}(B, C)$, then the composite $gf \in \mathrm{Hom}_{\mathcal{S}}(A, C)$ is equal to the composite $gf \in \mathrm{Hom}_{\mathcal{C}}(A, C)$,

(iv) if $A \in \mathrm{obj}(\mathcal{S})$, then the identity $1_A \in \mathrm{Hom}_{\mathcal{S}}(A, A)$ is equal to the identity $1_A \in \mathrm{Hom}_{\mathcal{C}}(A, A)$.

A subcategory $\mathcal{S}$ of $\mathcal{C}$ is a full subcategory if, for all $A, B \in \mathrm{obj}(\mathcal{S})$, we have $\mathrm{Hom}_{\mathcal{S}}(A, B) = \mathrm{Hom}_{\mathcal{C}}(A, B)$.

Example 3.1.8. For example, $\mathbf{Ab}$ is a full subcategory of Groups. Call a category **discrete** if its only morphisms are identity morphisms. If $\mathcal{S}$ is the discrete category with $\mathrm{obj}(\mathcal{S}) = \mathrm{obj}(\mathbf{Sets})$, then $\mathcal{S}$ is a subcategory of Sets that is not a full subcategory. On the other hand, the homotopy category $\mathbf{Htp}$ is not a subcategory of $\mathbf{Top}$, even though $\mathrm{obj}(\mathbf{Htp}) = \mathrm{obj}(\mathbf{Top})$, for morphisms in $\mathbf{Htp}$ are not continuous functions.

If $\mathcal{C}$ is any category and $\mathcal{S} \subseteq \mathrm{obj}(\mathcal{C})$, then the full subcategory generated by $\mathcal{S}$, also denoted by $\mathcal{S}$, is the subcategory with $\mathrm{obj}(\mathcal{S}) = \mathcal{S}$ and with $\mathrm{Hom}_{\mathcal{S}}(A, B) = \mathrm{Hom}_{\mathcal{C}}(A, B)$ for all $A, B \in \mathrm{obj}(\mathcal{S})$. For example, we define the category $\mathbf{Top}_2$ to be the full subcategory of Top generated by all Hausdorff spaces.

Functors are homomorphisms of categories.

Definition 3.1.9. If $\mathcal{C}$ and $\mathcal{D}$ are categories, then a functor $T : \mathcal{C} \to \mathcal{D}$ is a function such that

(i) if $A \in \mathrm{obj}(\mathcal{C})$, then $T(A) \in \mathrm{obj}(\mathcal{D})$,

(ii) if $f : A \to A'$ in $\mathcal{C}$, then $T(f) : T(A) \to T(A')$ in $\mathcal{D}$,

(iii) if $A \xrightarrow{f} A' \xrightarrow{g} A''$ in $\mathcal{C}$, then $T(A) \xrightarrow{T(f)} T(A') \xrightarrow{T(g)} T(A'')$ in $\mathcal{D}$ and $T(gf) = T(g)T(f)$,

(iv) $T(1_A) = 1_{T(A)}$ for every $A \in \mathrm{obj}(\mathcal{C})$.

Example 3.1.10.

(i) If $\mathcal{S}$ is a subcategory of a category $\mathcal{C}$, then the definition of subcategory may be restated to say that the inclusion $I : \mathcal{S} \to \mathcal{C}$ is a functor [this is one reason for the presence of Axiom (iv)].

(ii) If $\mathcal{C}$ is a category, then the **identity functor** $1_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$ is defined by $1_{\mathcal{C}}(A) = A$ for all objects A and $1_{\mathcal{C}}(f) = f$ for all morphisms f.

(iii) If $\mathcal{C}$ is a category and $A \in \mathrm{obj}(\mathcal{C})$, then the Hom **functor** $T_A : \mathcal{C} \to$ Sets, usually denoted by $\mathrm{Hom}(A, \square)$, is defined by

$$T_A(B) = \mathrm{Hom}(A, B) \text{ for all } B \in \mathrm{obj}(\mathcal{C}),$$

and if $f : B \to B'$ in $\mathcal{C}$, then $T_A(f) : \mathrm{Hom}(A, B) \to \mathrm{Hom}\,(A, B')$ is given by

$$T_A(f) : h \mapsto fh.$$

We call $T_A(f) = \mathrm{Hom}(A, f)$ the **induced map**, and we denote it by f_*; thus,

$$f_* : h \mapsto fh.$$

Suppose now that $g : B' \to B''$. Let us compare the functions

$$(gf)_*, g_* f_* : \mathrm{Hom}(A, B) \to \mathrm{Hom}\left(A, B''\right).$$

If $h \in \mathrm{Hom}(A, B)$, i.e., if $h : A \to B$, then

$$(gf)_* : h \mapsto (gf)h;$$

on the other hand, associativity of composition gives

$$g_* f_* : h \mapsto fh \mapsto g(fh) = (gf)h,$$

as desired. Finally, if f is the identity map $1_B : B \to B$, then

$$(1_B)_* : h \mapsto 1_B h = h$$

for all $h \in \mathrm{Hom}(A, B)$, so that $(1_B)_* = 1_{\mathrm{Hom}(A,B)}$.

(iv) A functor $T : \mathbb{Z} \to \mathcal{C}$, where $\mathbb{Z}$ is the category obtained from $\mathbb{Z}$ viewed as a partially ordered set [as in Example 1.3(vi)], is a sequence

$$\cdots \to C_{n+1} \to C_n \to C_{n-1} \to \cdots.$$

(v) Define the **forgetful functor** $U : \mathbf{Groups} \to \mathbf{Sets}$ as follows: $U(G)$ is the underlying set of a group G and $U(f)$ is a homomorphism f regarded as a mere function. Strictly speaking, a group is an ordered pair (G, μ) [where G is its (underlying) set and $\mu : G \times G \to G$ is its operation], and $U((G, \mu)) = G$; the functor U "forgets" the operation and remembers only the set. There are many variants. For example, a ring is an ordered triple (R, α, μ) [where $\alpha : R \times R \to R$ is addition and $\mu : R \times R \to R$ is multiplication], and there are forgetful functors U' : Rings $\to$ **Ab** with $U'(R, \alpha, \mu) = (R, \alpha)$, the additive group of R, and $U''\mathbf{Rings} \to \mathbf{Sets}$ with $U''(R, \alpha, \mu) = R$, the underlying set.

Definition 3.1.11. A **contravariant functor** $T : \mathcal{C} \to \mathcal{D}$, where $\mathcal{C}$ and $\mathcal{D}$ are categories, is a function such that

(i) if $C \in \mathrm{obj}(\mathcal{C})$, then $T(C) \in \mathrm{obj}(\mathcal{D})$,

(ii) if $f : C \to C'$ in $\mathcal{C}$, then $T(f) : T\,(C') \to T(C)$ in $\mathcal{D}$ (note the reversal of arrows),

(iii) if $C \xrightarrow{f} C' \xrightarrow{g} C''$ in $\mathcal{C}$, then $T\,(C'') \xrightarrow{T(g)} T\,(C') \xrightarrow{T(f)} T(C)$ in $\mathcal{D}$ and

$$T(gf) = T(f)T(g),$$

(iv) $T(1_A) = 1_{T(A)}$ for every $A \in \mathrm{obj}(\mathcal{C})$.

To distinguish them from contravariant functors, the functors defined earlier are called covariant functors.

Example 3.1.12. If $\mathcal{C}$ is a category and $B \in \mathrm{obj}(\mathcal{C})$, then the **contravariant Hom functor** $T^B : \mathcal{C} \to \mathbf{Sets}$, usually denoted by $\mathrm{Hom}(\square, B)$, is defined, for all $C \in \mathrm{obj}(\mathcal{C})$, by

$$T^B(C) = \mathrm{Hom}(C, B),$$

and if $f : C \to C'$ in $\mathcal{C}$, then $T^B(f) : \mathrm{Hom}(C', B) \to \mathrm{Hom}(C, B)$ is given by

$$T^B(f) : h \mapsto hf.$$

We also call $T^B(f) = \mathrm{Hom}(f, B)$ the **induced map**, and we denote it by f^*; thus,

$$f^* : h \mapsto hf.$$

Because of the importance of this example, we verify the axioms, showing that $\mathrm{Hom}(\square, B)$ is a (contravariant) functor.

Given homomorphisms

$$C \xrightarrow{f} C' \xrightarrow{g} C'',$$

let us compare the functions

$$(gf)^*, f^*g^* : \mathrm{Hom}\left(C'', B\right) \to \mathrm{Hom}(C, B).$$

If $h \in \mathrm{Hom}(C'', B)$, i.e., if $h : C'' \to B$, then

$$(gf)^* : h \mapsto h(gf)$$

on the other hand,

$$f^*g^* : h \mapsto hg \mapsto (hg)f = h(gf) = (hg)f,$$

as desired. Finally, if f is the identity map $1_C : C \to C$, then

$$(1_C)^* : h \mapsto h1_C = h$$

for all $h \in \mathrm{Hom}(C, B)$, so that $(1_C)^* = 1_{\mathrm{Hom}(C,B)}$.

Definition 3.1.13. For category $\mathcal{C}, I \in \mathrm{obj}(\mathcal{C})$ is **initial** if for any $A \in \mathrm{obj}(\mathcal{C}), \mathrm{Hom}_{\mathcal{C}}(I, A)$ has one element. $F \in \mathrm{obj}(\mathcal{C})$ is **final** if for any $A \in \mathrm{obj}(\mathcal{C})$, then $\mathrm{Hom}_{\mathcal{C}}(A, F)$ has one element.

Example 3.1.14.

- For $\mathcal{C}$ a set, $\varnothing$ is the initial object, any singleton set is a final object.
- For $(S, \sim)$ with $(\mathbb{Z}, \leqslant)$, there is no initial or final object.

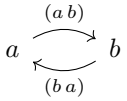
Note: Initial and final objects are unique up to isomorphism.

Example 3.1.15.

- For category of sets, initial object is $\varnothing$ and final object is singleton set.
- For category of groups, initial object is $\{e\}$ and final is also $\{e\}$.
- For category of rings, intial object is $\mathbb{Z}$, final object is $\{0\}$.
- For category of R-modules, initial element is $\{0\}$ and final is $\{0\}$.
- For category of fields, there are no initial and final objects

Definition 3.1.16. A category $\mathcal{C}$ is a **groupoid** if every morphism is an isomorphism.

Example 3.1.17. If $\sim$ on S is an equivalence relation,



Definition 3.1.18. If $A \in \mathrm{obj}(\mathcal{C})$ isomorphisms $\in \mathrm{Hom}(A, A)$ are **automorphism**, they form a group denoted by $Aut(A)$

Fact: A *group* is a *groupoid* of 1 object!

3.2 Modules

Definition 3.2.1. Suppose we have arbitrary ring R and abelian group M such that there is $R \times M \to M$, $(r, m) \mapsto rm$ with distributivity. This is a **left module**, and satisfies the distributivity below:

- $(r + s)m = rm + sm$
- $r(m_1 + m_2) = rm_1 + rm_2$
- $(rs)m = r(sm)$
- $1_R m = m$

Modules also satisfy the following properties:

- $r0_M = 0_M$,
- $0_R m = 0_M$,
- $(-r)m = -(rm)$.

Definition 3.2.2. Let M be an R-module, a subset $N \subset M$ is called a R-**submodule** of M, written as $N \leqslant M$, if $(N, +) \leqslant (M, +)$ and for any $r \in R, n \in N$, we have $r \cdot n \in N$.

Example 3.2.3. 1. If R is a field, then an R-module M is a vector space over R.
2. Let R be a ring and R be a module over R. Submodules are (left) ideals in this case.
3. A $\mathbb{Z}$-module is precisely the same as an abelian group as the scalar multiplication can be uniquely defined by $n \cdot a = a + \cdots + a$ for n many copies of a.
4. Consider the ring $R = \mathbb{F}[X]$ for a field $\mathbb{F}$ and V a vector space over $\mathbb{F}$. Consider $\alpha : V \to V$ an endomorphism. We can make V an R-module over the scalar multiplication $\mathbb{F}[X] \times V \to V$ by $(f, v) \mapsto f(\alpha)(v)$. Note that different choice of α makes V a different module. We sometimes write this as V_α.

There are some general construction methods to produce a module.

Example 3.2.4. 1. For any ring R, R^n is an R-module by $r \cdot (r_1, \ldots, r_n) = (rr_1, \ldots, rr_n)$ for $r, r_i \in R$. In particular, when $n = 1$, R itself is an R-module.
2. If I is an ideal, then I is an R-module by $r \cdot i = ri$ for $r \in R, i \in I$.
3. If I is an ideal, then R/I is an R-module by $r \cdot (s + I) = rs + I$ for $r, s \in R$.
4. If $\phi : R \to S$ is a ring homomorphism, then any S-module M is also an M-module by $r \cdot m = \phi(r) \cdot m$ for $r \in R, m \in M$. In particular, if $R \leqslant S$, then any S-module can be viewed as an R-module.

Example 3.2.5. 1. Any R-submodule of R is an ideal.
2. When R is a field, then an R-module is a vector space, then a submodule is a vector subspace.

Definition 3.2.6. If N is a R-submodule of M, we can form the quotient M/N by taking the quotient group under addition. We can make it as an R-module by specifying the scalar multiplication $r \cdot (m+N) = r \cdot m+N$.

We can check easily that the scalar multiplication defined in this way is well-defined and makes M/N an R-module.

Definition 3.2.7. Let M, N be R-modules, then a function $f : M \to N$ is a homomorphism of R-modules (or **R-module map**) if f is a homomorphism of groups under addition and $\forall r \in R, m \in M, f(r \cdot m) = r \cdot f(m)$. A bijective homomorphism is called an isomorphism, and two R-modules M, N are called isomorphic (written as $M \cong N$) if there is an isomorphism between them.

Example 3.2.8. When R is a field, a homomorphism of R-modules is a linear map.

Isomomorphism Theorems

If $N \subseteq M$ is a submodule, then M/N has the structure of a R-module.

$$r(m + N) := rm + N$$

well-defined: Does $m + N = m' + N \implies r(m + N) = r(m' + N)$?. yes, because $m - m' \in N$ and $r(m - m') \in N$

Isomorphism Theorem 1: If $f : M \to N$ is a R-homomorphism, then

$$M/Ker(f) \simeq \mathrm{Im}(f) \text{ as } R\text{-modules}$$

Isomorphism Theorem 2: If N_1, N_2 are submodules of M, then $N_1 + N_2 := \{x + y \,\big|\, x \in N_1, y \in N_2\}$ is a submodule of M, and $N_1 \cap N_2$ is also a submodule of M, and

$$\frac{N_2}{N_1 \cap N_2} \simeq \frac{N_1 + N_2}{N_1}, \quad f : N_2 \to \frac{N_1 + N_2}{N_1},\ f(n_2) = n_2 + N_1$$

Isomorphism Theorem 3: If $N \subseteq M$ and $K \subseteq N$ are submodules, then N/K is a submodule of M/K, and

$$\frac{M/K}{N/K} \simeq M/N$$

Isomorphism Theorem 4: If $N \subseteq M$ is a submodule, the canonical map $M \to M/N, m \mapsto m + N$ induces a 1-1 correspondence between submodules of M/N and submodules of M containing N

3.3 Finitely Generated Modules

Definition 3.3.1. Let M be an R-module, and $m \in M$, then the submodule Rm generated by m is the smallest R-submodule of M containing m, i.e. $Rm = \{r \cdot m : r \in R\}$.

Definition 3.3.2. Let M be an R-module. M is called cyclic if $M = Rm$ for some $m \in M$. M is finitely generated if $\exists m_1, \ldots, m_n \in M$ such that $Rm_1 + \cdots Rm_n = M$.

Lemma 3.3.3. An R-module M is cyclic iff M is isomorphic as an R-module to R/I for some $I \trianglelefteq R$.

Proof. If M is cyclic, write $M = Rm$, then there is a surjective R-module homomorphism $R \to M$ by $r \mapsto r \cdot m$ so the claim follows by the First Isomorphism Theorem.
Conversely If $M \cong R/I$, then $M \cong R/I = R(1 + I)$. □

Lemma 3.3.4. An R-module M is finitely generated iff there exists a surjective R-module homomorphism from $f : R^n \to M$ for some n.

Proof. If M is finitely generated, then $M = Rm_1 + \cdots + Rm_n$ where $m_i \in M$, so we can take $f(r_1, \ldots, r_n) = r_1m_1 + \cdots + r_nm_n$.
Conversely, if such a map f exists, then $M = Rf(e_1) + \cdots + Rf(e_n)$, then e_i has 1 in i^{th} entry and 0 in j^{th} entry for any $j \neq i$. □

Corollary 3.3.5. The quotient of a finitely generated R-module is a finitely generated R-module.

Proof. Obvious from the preceding lemma. □

Remark 3.3.6. A submodule of a finitely generated R-module needs not be finitely generated. For example, we can take a non-Noetherian ring R itself as an R-module and consider a non-finitely generated ideal of it.

Lemma 3.3.7. Let R be an integral domain, then every R-submodule of a cyclic R-module is cyclic iff R is a PID.

Proof. R itself is a cyclic R-module, so if all R-submodules of it are cyclic, then all of its ideals are generated by one element, so R is a PID.
Conversely, if R is a PID and M is a cyclic R-module, so $M \cong R/I$ for $I \trianglelefteq R$, so the R-submodules of M are in the form J/I for $I \subset J \trianglelefteq R$. Now since R is a PID, J is principal, so J/I is cyclic. □

Theorem 3.3.8. Let R be a PID, and M an R-module. Suppose M is generated by n elements, then any R-submodule N of M can also be generated by at most n elements.

Proof. $n = 1$ is the preceding lemma. For general n, we proceed by induction. Suppose $M = Rx_1 + \cdots Rx_n$. Let $M_i = Rx_1 + \cdots Rx_i$ and $0 = M_0 \leqslant M_1 \leqslant \cdots \leqslant M_n = M$. So we have

$$0 = M_0 \cap N \leqslant M_1 \cap N \leqslant \cdots \leqslant M_n \cap N = N$$

Then the R-module map $M_i \cap N \to M_i/M_{i-1}$ by $m \mapsto m + M_{i-1}$ has kernel $M_{i-1} \cap N$. Hence

$$(M_i \cap N)/(M_{i-1} \cap N) \cong M' \leqslant M_i/M_{i-1}$$

But M_i/M_{i-1} is cyclic by hypothesis, so by preceding lemma, $(M_i \cap N)/(M_{i-1} \cap N)$ is also cyclic and is generated by $y_i + M_{i-1} \cap N$ where $y_i \in M_i \cap N$. Therefore $M_i = M_{i-1} \cap N + Ry_i$. It follows that $M_i \cap N = Ry_1 + \cdots + Ry_i$. In particular, $N = M_n \cap N = Ry_1 + \cdots + Ry_n$, so N is generated by n elements. □

Example 3.3.9. Take $R = \mathbb{Z}$, then we know that any subgroup of $\mathbb{Z}^n$ can be generated by n elements.

3.4 Exact Sequences

Definition 3.4.1. Let R be a ring and M, M', M'' be R-modules. A sequence of R-homomorphism $M' \xrightarrow{f} M \xrightarrow{g} M''$ is called **exact** if $\mathrm{Im}(f) = ker(g)$. More generally, sequence $M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3$ is **exact** if $\mathrm{Im}(f_i) = ker(f_{i+1})$.

Example 3.4.2. The sequence $0 \to M' \xrightarrow{f} M$, is *exact* if and only if f is injective.

Example 3.4.3. The sequence $M \xrightarrow{g} M'' \to 0$ is *exact* if and only if g is surjective

Definition 3.4.4. If $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is an exact sequence, then it is called a **short exact sequence**

Example 3.4.5. If $N \subseteq M$ is a submodule, $0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$.

Proposition 3.4.6. Let $0 \longrightarrow M' \underset{\psi}{\overset{f}{\rightleftarrows}} M \underset{\phi}{\overset{g}{\rightleftarrows}} M'' \longrightarrow 0$ be a short exact sequence of R-modules. Then the following conditions are equivalent.

1. $\exists$ R-homomorphism $\phi : M'' \to M$ s.t. $g \circ \phi = id_{M''}$
2. $\exists$ R-homomorphism $\psi : M \to M'$ s.t. $\psi \circ f = id_{M'}$

and they imply $M \simeq M' \oplus M''$. In this case, we say the sequence **splits**

Example 3.4.7. $R = \mathbb{Z}_4, M = \mathbb{Z}_4, N = \{0, 2\}$. Then $0 \to N \to \mathbb{Z}_4 \to \mathbb{Z}_4/N \to 0$. Notice that $\psi(1) = 0 \implies \psi(2) = 0$ and $\psi(1) = 2 \implies \psi(2) = 0$. Therefore this does not split.

Proof of Proposition. $(1) \implies (2)$: If $m \in M$, then $g(\phi(g(m))) = g(m) \implies g(m - \phi(g(m))) = 0 \implies m - \phi(g(m)) \in ker(g) = \mathrm{Im}(f) \implies \exists! x \in M'$ s.t. $f(x) = m - \phi(g(m))$.

Let $\psi(m) = x$. We need to check that ψ is a R-homomorphism (exercise), and $\psi \circ f = id_{M'}$: if $y \in M'$, let $m = f(y)$. Then $m - \phi(g(m)) = f(y) - \phi(\underbrace{g(f(y))}_{=0}) = f(y)$. By definition of $\psi : \psi(m) = y \implies \psi(f(y)) = y \; \forall y$

$(2) \implies (1)$: Suppose $x \in M''$, then $\exists y \in M$ s.t. $g(y) = x$. Then let $\phi(x) = y - f(\psi(y))$.

This is well-defined: If $y' \in M$ such that $g(y') = x$. I want to check that $y - f(\psi(y)) = y' - f(\psi(y'))$, or $y - y' = f(\psi(y - y'))$. But $g(y - y') = 0$. Since $Ker(g) = \mathrm{Im}(f)$, $\exists z \in M'$ s.t. $y - y' = f(z) \implies f(\psi(y - y')) = f(\psi(f(z))) = f(z) = y - y'$. So ϕ well-defined.

Also $g \circ \phi = id_{M''}$: If $x \in M''$, $\phi(x) = y - f(\psi(y))$ for some $y \in M$ with $g(y) = x$, so $g(\phi(x)) = g(y) - g(f(\psi(y))) = g(y) = x$, since $g \circ f = 0$. Also ϕ is a R-homomorphism, since $\forall r, s \in R, x_1, x_2 \in M'', \phi(rx_1 + sx_2) = r\phi(x_1 + s\phi(x_2))$.

Direct Sum: Define

$$M' \oplus M'' \xrightarrow{\alpha} M, (x, y) \mapsto f(x) + \phi(x)$$

$$M \xrightarrow{\beta} M' \oplus M'', m \mapsto (\psi(m), g(m))$$

Then $\beta \circ \alpha(x, y) = \beta(f(x) + \phi(y)) = (x, y)$, since $\psi \circ \phi = 0$ (Show this as an exercise:) $\square$

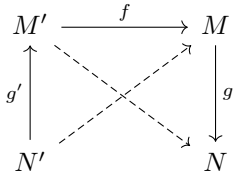
3.5 Hom Functors

Definition 3.5.1. Let M, N be R-module, with $\mathrm{Hom}_R(M, N)$ being the **set of R-homomorphism** $f : M \longrightarrow N$, and $\mathrm{Hom}_R(M, N)$ has the structure of an R-module.

Let $f, g \in \mathrm{Hom}_R(M, N)$ if $f + g \in \mathrm{Hom}_R(M, N)$. Note $(rf)(m) = rf(m), (f + g)(m) = f(m) + g(m)$. We have

$$\mathrm{Hom}_R(M, N) \xrightarrow{-\circ f} \mathrm{Hom}_R(M', N)$$

$$\mathrm{Hom}_R(N, M') \xrightarrow{f \circ -} \mathrm{Hom}_R(N, M)$$

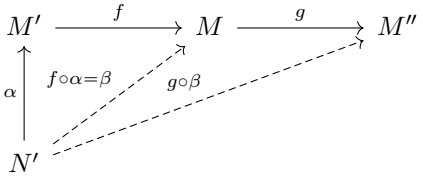


Lemma 3.5.2. If $0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is a short exact sequence of R-modules and N is a R-module, then

$$(1). \quad 0 \to \mathrm{Hom}_R(N, M') \xrightarrow{\psi} \mathrm{Hom}_R(N, M) \xrightarrow{\phi} \mathrm{Hom}_R(N, M'') \text{ exact}$$

$$(2). \quad 0 \longrightarrow \mathrm{Hom}_R(M'', N) \longrightarrow \mathrm{Hom}(M, N) \longrightarrow \mathrm{Hom}(M', N) \text{ exact}$$

Proof.



$\mathrm{Hom}_R(N, M') \to_R \mathrm{Hom}(N, M)$ injective: If $f \circ \alpha = 0$ for some $\alpha \in \mathrm{Hom}_R(N, M')$, then since f injective, $\alpha = 0$.

$\phi \circ \psi = 0 (\implies \mathrm{Im}(\psi) \subset ker(\phi))$: If $\alpha \in \mathrm{Hom}_R(N, M')$, then $\phi \circ \psi(\alpha) = g \circ f \circ \alpha = 0$, where $g \circ f = 0$ since it is exact.

If $\beta \in Ker(\phi)$, then $g \circ \beta = 0$, so for any $x \in N, g(\beta(x)) = 0$, so $\beta(x) \in \mathrm{Im}(f) \implies$ there is a unique $y \in M'$ such that $f(y) = \beta(x)$. Let $\alpha : N \to M'$ be defined by $\alpha(x) = y$, then α is a R-homomorphism (Exercise). And clearly $\beta = f \circ \alpha$, so $\beta \in \mathrm{Im}(\psi)$ □

Remark: If $M' \subseteq M$ is a submodule, then $0 \to M' \to M \to M/M'$ is a short exact sequence. If $g : M \to M''$ is a surjective R homomorphism, then $0 \to ker(g) \to M \to M'' \to 0$ is a short exact sequence.

$$\Rightarrow (x, 0) \in \mathrm{Im}(k) = \left\{\left(-f\left(m'\right), i\left(m'\right)\right)\right\}$$
$$\Rightarrow 0 = i\left(m'\right) \xrightarrow{i \text{ inclusion}} m' = 0 \Rightarrow x = -f\left(m'\right) = 0$$

We then by exactness get R-homomorphism $q : P \to Q$ such that $q \circ \alpha = 1_Q$. If we let $h = q \circ \beta$, then the desired relationship $f = h \circ i$ is obtained from q by composing f on both sides of $q \circ \alpha = 1_Q$, i.e., $h \circ i = q \circ \beta \circ i = q \circ \alpha \circ f = 1_Q \circ f = f$ if $\beta \circ i = \alpha \circ f$, i.e., the square diagram is commutative, which is immediate: $(\beta i - \alpha f)(x) = [(0, i(x)) + \mathrm{Im}(k)] - [(f(x), 0) + \mathrm{Im}(k)] = (0, i(x)) - (f(x), 0) + \mathrm{Im}(k) = (-f(x), i(x)) + + \mathrm{Im}(k) = \underbrace{k(x)}_{\in \mathrm{Im}(k)} + \mathrm{Im}(k) = 0$ in the quotient. Thus, the map $h = q \circ \beta$ concludes.

3.6 Direct Sums and Free Modules

Definition 3.6.1. If $M_1, \ldots, M_n$ are R-modules, then their direct sum $M_1 \oplus \cdots \oplus M_n$ is the set $M_1 \times \cdots \times M_n$ with entry-wise addition and scalar multiplications.

Example 3.6.2. 1. R^n is simply $R \oplus \cdots \oplus R$ of n copies of R.
2. If $M_1, M_2 \leqslant M$, then the R-module homomorphism $M_1 \oplus M_2 \to M$ by $(m_1, m_2) \mapsto m_1 + m_2$ is an isomorphism iff $M_1 \cap M_2 = \varnothing$ and $M_1 + M_2 = M$.

Lemma 3.6.3. If $M = \bigoplus_{i=1}^n M_n$, and $N_1 \leqslant M_i$. Take $N = \bigoplus_{i=1}^n N_i$, then

$$M/N \cong \bigoplus_{i=1}^{n} M_i/N_i$$

Proof. Apply the first isomorphism theorem to the surjective R-module map $\phi : M \to \bigoplus_{i=1}^n M_i/N_i$ by $(m_1, \ldots, m_n) \mapsto (m_1 + N_1, \ldots, m_n + N_n)$. □

Example 3.6.4. Taking $R = \mathbb{Z}$ then $\mathbb{Z}^2 = \mathbb{Z} \oplus \mathbb{Z}$, then we have $(\mathbb{Z} \oplus \mathbb{Z})/(m\mathbb{Z} \oplus n\mathbb{Z}) \cong (\mathbb{Z}/m\mathbb{Z}) \oplus (\mathbb{Z}/n\mathbb{Z})$.

Definition 3.6.5. Let $m_1, \ldots, m_n \in M$. The set $\{m_1, \ldots, m_n\}$ is independent if $r_1m_1 + \cdots + r_nm_n = 0 \implies \forall i, r_i = 0$.

Definition 3.6.6. A subset S of an R-module M generates M freely if S generates M and any function $\psi : S \to N$ for another R-module N extends to an R-module homomorphism $M \to N$.

Note that if such an extension exists then it is necessarily unique.

Definition 3.6.7. A freely-generated R-module is called a free R-module. The corresponding S is called the free basis.

Proposition 3.6.8. For an R-module M and a subset $S = \{m_1, \ldots, m_n\} \subset M$, the followings are equivalent:
1. S generates M freely.
2. S generates M and S is independent.
3. Every $m \in M$ can be written uniquely in the form $m = r_1m_1 + \cdots + r_nm_n$ for $r_1, \ldots, r_n \in R$.
4. The R-module homomorphism $R^n \to M$ by $(r_1, \ldots, r_n) \mapsto r_1m_1 + \ldots r_nm_n$ is an isomorphism.

Proof. $1 \implies 2$: We already knows that S generates M, so it suffices to show that S is independent. Suppose for sake of contradiction that $r_1m_1 + \ldots + r_nm_n = 0$ for some $r_i \in R$ and some r_j is nonzero. Consider the function $\psi : S \to R$ by $m_j \mapsto 1$ and $m_i \mapsto 0$ for any $i \neq j$. Suppose this extends to an R-module map $\theta : M \to R$, then $0 = \theta(0) = \theta(r_1m_1 + \cdots + r_nm_n) = r_j$, contradiction.
Remaining implications $2 \implies 3 \implies 1$ and $3 \iff 4$ are just as easy if not easier. □

Sadly not all R-modules are free. Even if it is, the free basis does not behave like what we expect from a vector space.

Example 3.6.9 (non-example)**.** 1. Suppose we have a nontrivial finite abelian group A, then A is not free as a $\mathbb{Z}$-module since it is not isomorphic to $\mathbb{Z}^n$ which is infinite.
2. The set $\{2, 3\} \subset \mathbb{Z}$ generates $\mathbb{Z}$ as a $\mathbb{Z}$-module, but it is not independent and no subset of it gives a free basis.

Proposition 3.6.10 (Theorem on Invariant of Dimension)**.** Let R be a nonzero ring. If $R^m \cong R^n$ as R-modules, then $m = n$.

We introduce the following general construction: Let R be a ring and $I \trianglelefteq R$ and M is an R-module. We write $IM = \{im : i \in I, m \in M\} \leqslant M$. Then the quotient $M/(IM)$ is an R/I module by $(r + I)(m + IM) = rm + IM$. Also by Zorn's Lemma, for any proper ideal I in a ring R, there is a maximal ideal containing I (this is obvious when R is Noetherian). [1]

Proof. Return to our proof, suppose $R^m \cong R^n$. Choose $I \trianglelefteq R$ maximal, then we have

$$(R/I)^m \cong R^m/(IR^m) \cong R^n/(IR^n) \cong (R/I)^n$$

But R/I is a field, so $m = n$. □

3.7 Projective Module and Injective Module

Definition 3.7.1. If M is a R-module, and $S \subset M$ is a **basis** if $\forall m \in M, m = r_1s_1 + ... + r_ks_k$ in a *unique* way with $r \in R, s \in S$. Equivalently, if $0 = r_1s_1 + ... + r_ks_k$, then $r_1 = ... = r_k = 0$. If $\{s_i\}_{i\in\mathcal{I}}$ is a basis for M, then $M \simeq \bigoplus_{i\in\mathcal{I}} R$. Then, M is **free** is it has a *basis*.

Definition 3.7.2. If R is a ring and P is a R-module, then P is a **projective module** if it satisfies the following:

[1] I think we can prove the proposition without using AC (or equivalence)

1. If g, ϕ are R homomorphism, $\exists \psi : P \to M$, R-homomorphism s.t. $g \circ \psi = \phi$

$$\begin{array}{ccccc} & & P & & \\ & \swarrow^{\exists\psi} & \downarrow{\scriptstyle \phi} & & \\ M & \xrightarrow{g} & M'' & \longrightarrow & 0 \end{array}$$

2. If $0 \to M' \to M \to P \to 0$ is exact, then it splits.
3. There is a R-module N such that $N \oplus P$ is a *free module*.
4. If $0 \to M' \to M \to M''$ is exact, then

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

is exact.

$(1) \implies (2)$. If $0 \to M' \to M \to P \to 0$ is exact, then by (1) $\exists \psi : P \to M$ s.t. $g \circ \psi = id_P$, so the sequence splits

$$\begin{array}{ccccc} & & P & & \\ & \swarrow^{\exists\psi} & \downarrow{\scriptstyle id_P} & & \\ M & \xrightarrow{g} & P & \longrightarrow & 0 \end{array}$$

□

$(2) \implies (3)$. Let $\{x_i\}_{i \in \mathcal{I}}$ be a generating subset of P as a R-module. Then, $g : \bigoplus_{i \in I} R \to P, (r_i)_{i \in I} \mapsto \sum_{i \in I} r_i x_i$. is surjective. Then, $0 \to ker(g) \to \bigoplus_{i \in I} R \to P \to 0$ is a short exact sequence. By (2) this splits, so free R-module $\bigoplus_{i \in I} R \simeq ker(g) \oplus P$. □

$(3) \implies (4)$. It is enough to show that $\mathrm{Hom}(P, M) \to \mathrm{Hom}(P, M'')$ is surjective. If P is free and $(x_i)_{i \in I}$ is a basis for P and let $y_i = \phi(x_i)$ and $z_i \in m$ s.t. $g(z_i) = y_i$. Then let $\psi(x_i) = z_i$ and $\psi(\sum r_i x_i) = \sum r_i z_i$. Then $g \circ \psi = \phi$. If $N \bigoplus P$ is free, then $\tilde{\phi}(r, p) = \phi(p)$ is a R homomorphism, $\exists \tilde{\psi} : N \oplus P \to M$ such that $g \circ \tilde{\psi} = \tilde{\phi}$. Define $\psi : P \to M, \psi(p) = \tilde{\psi}(n, p)$, then $g \circ \psi = \phi$.

$$\begin{array}{ccc} & P & \\ & \swarrow^{\psi} \quad \downarrow{\scriptstyle \phi} & \\ M & \xrightarrow{g} & M'' \end{array} \qquad \Longrightarrow \qquad \begin{array}{ccc} & Q = N \oplus P & \\ & \swarrow^{g\tilde{\psi}} \quad \downarrow{\scriptstyle \tilde{\phi}} & \\ M & \xrightarrow{\quad} & M'' \end{array}$$

□

$(4) \implies (1)$. The surjective map $g : M \to M'$ gives a short exact sequence $0 \to ker(g) \to M \to M'' \to 0$. So by (4) there is a surjective map $\mathrm{Hom}(P, M'') \to \mathrm{Hom}(P, M)$. This is exactly 1. □

Example 3.7.3. $R = \mathbb{Z}_6$. Let $\mathbb{Z}_6$ be a $\mathbb{Z}_6$-module and $I_1 = \{0, 3\}, I_2 = \{0, 2, 4\}$. Then $I_1 \cap I_2 = \{0\}$ and $I_1 + I_2 = \mathbb{Z}_6 \implies \mathbb{Z}_6 = I_1 + I_3$. So by 3, I_1, I_2 are projective modules but not free.

We introduce **injective module**.

Theorem 3.7.4. Let R be a commutative ring and Q a module over R. We show that the following are equivalent:

(a) If M is an R-module, if M' is a submodule of M, and if $f : M' \to Q$ is a R homomorphism, then there is an extension of f to a R-homomorphism $M \to Q$, i.e., there is a R-homomorphism $h : M \to Q$ such that the following diagram is commutative

(b) For any short exact sequence $0 \to M' \to M \to M'' \to 0$, the sequence

$$0 \to \mathrm{Hom_R}\left(\mathrm{M''}, \mathrm{Q}\right) \to \mathrm{Hom_R}(\mathrm{M}, \mathrm{Q}) \to \mathrm{Hom_R}\left(\mathrm{M'}, \mathrm{Q}\right) \to 0$$

is exact.

(c) Every short exact sequence $0 \to Q \to M \to M'' \to 0$ splits.

Proof. These are contents of Rotman's An Introduction to Homological Algebra e2 Proposition 3.25 and 3.26and 3.40. Note that the ring R is commutative.
$(\mathrm{a}) \Rightarrow (\mathrm{b})$: We have shown in class that exactness of $M' \xrightarrow{i} M \xrightarrow{p} M'' \to 0$ gives the exactness of

$$0 \to \mathrm{Hom}_R\left(M'', Q\right) \xrightarrow{p^*=()\circ p} \mathrm{Hom}_R(M, Q) \xrightarrow{i^*=()\circ i} \mathrm{Hom}_R\left(M', Q\right)$$

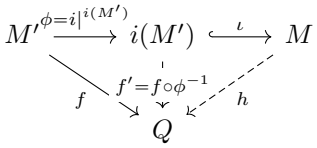
Therefore, to show exactness of

$$(*) \quad 0 \to M' \xrightarrow{i} M \xrightarrow{p} M'' \to 0$$

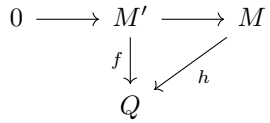
implies

$$(**) \quad 0 \to \mathrm{Hom}_R\left(M'', Q\right) \xrightarrow{p^*=()\circ p} \mathrm{Hom}_R(M, Q) \xrightarrow{i^*=()\circ i} \mathrm{Hom}_R\left(M', Q\right) \to 0$$

we only need to show injectivity of i implies surjectivity of i^*, given Q is an injective module, i.e., (a) is satisfied. Let $f \in \mathrm{Hom}_R(M', Q)$, i.e., $f : M' \to Q$ is an R-homomorphism. Since i is an injective R-homomorphism, $i(M') \subseteq M$ is a submodule, then $\phi = i|^{i(M')}$ denoting the restriction of i on its codomain $i(M')$ is an R-isomorphism. Let l be the inclusion of the submodule of M. Let $f' : i(M') \to Q$ be equal to $f \circ \phi^{-1}$. By (a), there is an R homomorphism $h : M \to Q$ making the following diagram commutative:



Since $i = l \circ \phi$ and $f' = h \circ l$, we see $i^*(h) = h \circ i = h \circ (l \circ \phi) = (h \circ l) \circ \phi = f' \circ \phi = f \circ \phi^{-1} \circ \phi = f$, so there is $h \in \mathrm{Hom}_R(M, Q)$ such that $i^*(h) = f$, proving that i^* is surjective.
(b) $\Rightarrow$ (c) : "(b) $\Rightarrow$ (c) " is like "Im $i \subseteq$ ker p " part of the last exercise we proved: given that $(**)$ is exact, which implies $i^*p^* = (pi)^* = 0$, we ;et $Q = M'$ in $(**)$ and consider the identity homomorphism $f = 1_{M'}$ in $\mathrm{Hom}_R(M', M')$. Then there is some $h \in \mathrm{Hom}_R(M, M')$ such that $i^*(h) = h \circ i = f = 1_{M'}$. Then by the definition/proposition of split, $(*)$ is exact (there is an R-homo $h : M \to M'$ such that $h \circ i$ is the identity).
$(\mathrm{c}) \Rightarrow (\mathrm{a})$: Let $i : M' \to M$ be the inclusion and $f : M' \to Q$ be the given R-homomorphism. We want to show that there is an R-homomorphism $h : M \to Q$ making the following diagram commutative:



Define

$$\begin{aligned} k : M' &\to Q \oplus M \\ x &\mapsto (-f(x), i(x)) \end{aligned}$$

which is clearly an R-homomorphism since f and i are. To obtain an R-homomorphism $h : M \to Q$, we consider using exactness of a sequence of the form $0 \to Q \to P \to \frac{P}{Q} \to 0$ to induce a map $P \to Q$ descending to $h : M \to Q$. Let $P = \frac{Q \oplus M}{\mathrm{Im}(k)}$ be the quotient of $Q \oplus M$ over submodule $\mathrm{Im}(k)$ and define

$$\begin{aligned} \alpha : Q &\to P \\ x &\mapsto (x, 0) + \mathrm{Im}(k) \end{aligned}$$

and

$$\begin{aligned} \beta : M &\to P \\ y &\mapsto (0, y) + \mathrm{Im}(k) \end{aligned}$$

Consider the following diagram:

$$\begin{array}{ccccccccc} & & & & M' & \xrightarrow{i} & M & & \\ & & & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle \beta} & & \\ 0 & \longrightarrow & Q & \xrightarrow{\alpha} & P & \xrightarrow{p} & P/\alpha(Q) & \longrightarrow & 0 \end{array}$$

To show $0 \to Q \xrightarrow{\alpha} P \xrightarrow{p} \frac{P}{\alpha(Q)} \to 0$ is exact (where p is the canonical projection), we need to show that α is injective:

$$\begin{aligned} \alpha(x) = (x, 0) + \mathrm{Im}(k) &= 0 \\ \Rightarrow (x, 0) \in \mathrm{Im}(k) &= \left\{\left(-f\left(m'\right), i\left(m'\right)\right)\right\} \\ \Rightarrow 0 = i\left(m'\right) &\xrightarrow{i \text{ inclusion}} m' = 0 \Rightarrow x = -f\left(m'\right) = 0 \end{aligned}$$

We then by exactness get R-homomorphism $q : P \to Q$ such that $q \circ \alpha = 1_Q$. If we let $h = q \circ \beta$, then the desired relationship $f = h \circ i$ is obtained from q by composing f on both sides of $q \circ \alpha = 1_Q$, i.e., $h \circ i = q \circ \beta \circ i = q \circ \alpha \circ f = 1_Q \circ f = f$ if $\beta \circ i = \alpha \circ f$, i.e., the square diagram is commutative, which is immediate: $(\beta i - \alpha f)(x) = [(0, i(x)) + \mathrm{Im}(k)] - [(f(x), 0) + \mathrm{Im}(k)] = (0, i(x)) - (f(x), 0) + \mathrm{Im}(k) = (-f(x), i(x)) + + \mathrm{Im}(k) = \underbrace{k(x)}_{\in \mathrm{Im}(k)} + \mathrm{Im}(k) = 0$ in the quotient. Thus, the map $h = q \circ \beta$ concludes. □

Example 3.7.5. Let R be a commutative ring and I an ideal of R. By considering the exact sequence $0 \to I \to R \to R/I \to 0$, show that if R/I is a projective R-module, then I is a principal ideal generated by an element a such that $a^2 = a$.

Solution. R is a commutaitve ring. $I \subseteq R$ is an ideal. If R/I is a projective R-module, then $I = (a)$ with $a^2 = a$.

We look at the SES

$$0 \to I \xrightarrow{\rho} R \to R/I \to 0$$

Then projective module R/I gives $\phi : R \to I$ such that $\phi \circ \rho = id_I$. Let $a = \phi(1)$, so $a \in I$. For any $i \in I$, $\phi(\underbrace{\rho(i)}_{\in R}) = i$. Then $i = \phi(i) = i\phi(1) = ia \Rightarrow i \in (a)\ \forall i \in I \Rightarrow I = (a)$. Let $i = a$, then we get $a^2 = a$.

Example 3.7.6. Let R be a commutative ring, and let M be a R-module. Let S be a multiplicative subset of R such that $1 \in S$ and $0 \notin S$. Consider the set of all $\{(m, s), m \in M, s \in S\}$, and show that the relation $(m_1, s_1) \sim (m_2, s_2)$ if there is $s \in S$ such that $s\left(s_2 m_1 - s_1 m_2\right) = 0$ is an equivalence relation. Denote the class of (m, s) by $\frac{m}{s}$, and set

$$S^{-1}M = \{(m, s), m \in M, s \in S\}/\sim$$

(i) Show that $S^{-1}M$ is a module over $S^{-1}R$.

(ii) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules, show that $0 \to S^{-1}M' \to S^{-1}M \to S^{-1}M'' \to 0$ is an exact sequence of $S^{-1}M$-modules.

Solution. (i): $S^{-1}M$ is an abelian group with addition

$$\frac{m_1}{s_1} + \frac{m_2}{s_2} = \frac{s_2 m_1 + s_1 m_2}{s_1 s_2}$$

The definition is commutative: $s_2 m_1 + s_1 m_2 = s_1 m_2 + s_2 m_1, s_1 s_2 = s_2 s_1$. It is also well-defined: let $\frac{m_0}{s_0} = \frac{m_1}{s_1}$ then $s\left(s_0 m_1 - s_1 m_0\right) = 0$ for some $s \in S$. Then

$$\begin{aligned} & s\left(s_0 s_2\left(s_2 m_1 + s_1 m_2\right) - s_1 s_2\left(s_2 m_0 + s_0 m_2\right)\right) \\ &= s\left(s_0 s_2^2 m_1 + s_0 s_2 s_1 m_2 - s_1 s_2^2 m_0 - s_1 s_2 s_0 m_2\right) \\ &= s\left(s_2^2\left(s_0 m_1 - s_1 m_0\right)\right) = 0 \end{aligned}$$

Thus $\frac{s_2 m_1 + s_1 m_2}{s_1 s_2} = \frac{s_2 m_0 + s_0 m_2}{s_0 s_2}$ by element $s \in S$. Similarly, $\frac{m'}{s'} = \frac{m_2}{s_2}$ will give $\frac{s_2 m_1 + s_1 m_2}{s_1 s_2} = \frac{s' m_1 + s_1 m'}{s_1 s'}$. Thus $\frac{s' m_1 + s_1 m'}{s_1 s'} = \frac{s_2 m_1 + s_1 m_2}{s_1 s_2} = \frac{s_2 m_0 + s_0 m_2}{s_0 s_2}$. The definition of addition is then regardless of representatives of the equivalence classes chosen. $S^{-1}M$ is a $S^{-1}R$-module with the scalar multiplication $* : S^{-1}R \times S^{-1}M \to S^{-1}M$ defined by

$$\frac{r}{s} * \frac{m}{s'} = \frac{rm}{ss'}$$

This is well defined: let $\frac{m_0}{s_0} = \frac{m}{s'}$ then $s''\left(s_0 m - s' m_0\right) = 0$ for some $s'' \in S$. Then

$$s''\left(s s_0 r m - s s' r m_0\right) = s''\left(s r\left(s_0 m - s' m_0\right)\right) = 0$$

Thus $\frac{rm}{ss'} = \frac{rm_0}{ss_0}$ by element $s'' \in S$. Similar argument as for addition implies that the definition of scalar multiplication is regardless of representatives of the equivalence classes chosen.

(ii)

We are given the exact sequence

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

with R-homomorphisms f and g. Then f is injective, g is surjective and $\mathrm{Im}(f) = \ker(g)$. Naturally, we define $S^{-1}R$-homomorphisms $p : S^{-1}M' \to S^{-1}M; \frac{m'}{s} \mapsto \frac{f(m')}{s}$ and $q : S^{-1}M \to S^{-1}M''; \frac{m}{s} \mapsto \frac{g(m)}{s}$. Consider the following sequence

$$0 \to S^{-1}M' \xrightarrow{p} S^{-1}M \xrightarrow{q} S^{-1}M'' \to 0$$

Note that $g \circ f = 0 \Rightarrow q \circ p\left(\frac{m'}{s}\right) = \frac{g\left(f\left(m'\right)\right)}{s} = \frac{0}{s} = 0 \Rightarrow q \circ p = 0$. To show it is a short exact sequence, we need to show p injective, q surjective, and $\mathrm{Im}(p) = \ker(q)$: - p injective: Let $\frac{f\left(m_1'\right)}{s_1} = p\left(\frac{m_1'}{s_1}\right) = p\left(\frac{m_2'}{s_2}\right) = \frac{f\left(m_2'\right)}{s_2}$. Then $\exists s \in S$ s.t.

$$\begin{aligned} & s\left(s_2 f\left(m_1'\right) - s_1 f\left(m_2'\right)\right) = 0 \xRightarrow{f R-\text{ homo}} f\left(s\left(s_2 m_1' - s_1 m_2'\right)\right) = 0 \\ & \xRightarrow{\text{finjective}} s\left(s_2 m_1' - s_1 m_2'\right) = 0 \Rightarrow \frac{m_1'}{s_1} = \frac{m_2'}{s_2} \end{aligned}$$

- q surjective: Since g is surjective, we see for $m'' \in M''$ we have $g(m) = m''$ for some $m \in M$, then $q\left(\frac{m}{s}\right) = \frac{g(m)}{s} = \frac{m''}{s}$.

- $\mathrm{Im}(p) = \ker(q)$:

$$\begin{aligned}\ker(q) &= \left\{\frac{m}{s} : q\left(\frac{m}{s}\right) = \frac{g(m)}{s} = \frac{0}{1}\right\} = \left\{\frac{m}{s} : \exists s' \in S \text{ s.t. } s'g(m) = 0\right\} \\ &= \left\{\frac{m}{s} : \exists s' \in S \text{ s.t. } s'm = \underbrace{\ker(g)}_{=\mathrm{Im}(f)}\right\} = \left\{\frac{m}{s} : \exists s' \in S, m' \in M' \text{ s.t. } f\left(m'\right) = s'm\right\} \\ &= \left\{\frac{m}{s} : \exists s' \in S, m' \in M' \text{ s.t. } \frac{f\left(m'\right)}{s's} = \frac{s'm}{s's}\right\} = \left\{p\left(\frac{m'}{s's}\right) : s' \in S, m' \in M'\right\} \\ &= \left\{p\left(\frac{m'}{s''}\right) : s'' \in S, m' \in M'\right\} = \mathrm{Im}(p)\end{aligned}$$

3.8 Tensor Products

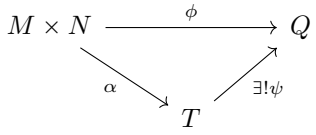
Let R be a ring and M, N be R-modules. Let F be a free module generated by elements $(m, n), m \in M, n \in N$. $F = \{r_1(m_1, n_1) + ... + r_k(m_k, n_k) \,|\, r_i \in R, m_i \in M, n_i \in N\}$. D is the submodule of F generated by elements of the forms below

- $(m_1 + m_2, n) - (m_1, n) - (m_2, n)$,
- $(m, n_1 + n_2) - (m, n_1) - (m, n_2)$
- $(rm, n) - r(m, n)$
- $(m, rn) - r(m, n)$

with $r \in R, m, m_1, m_2 \in M, n, n_1, n_2 \in N$.

Let $T := F/D$ be an R-module. Note there is a map $\alpha : M \times N \longrightarrow T, \alpha(m, n) = (m, n) + D$. This map is bilinear: $\alpha(r_1m_1 + r_2m_2, n) = r_1\alpha(m_1, n) + r_2\alpha(m_2, n)$ and $\alpha(m, r_1n_1 + r_2n_2) = r_1\alpha(m, n_1) + r_2\alpha(m, n_2)$

Proof of above requires us to show $(r_1m_1 + r_2m_2, n) - r_1(m_1, n) - r_2(m_2, n) \in D$. Rewrite expression into $((r_1m_1 + r_2m_2, n) - (r_1m_1, n) - (r_2m_2, n)) + ((r_1m_1, n) - r_1(m_1, n)) + ((r_2m_2, n) - r_2(m_2, n))$



T has the following *universal property*: If Q is a R-module and $\phi : M \times N \longrightarrow Q$ is a bilinear map, then there is a unique R-homomorphism $\psi : T \to \mathbb{Q}$ with $\phi = \psi \circ \alpha$, and define $\psi((r_1(m_1, n_1) + ... + r_k(m_k, n_k)) + D) = r_1\phi(m_1, n_1) + ... + r_k\phi(m_k, n_k)$.

We need to check that ψ is well-defined and is a R-homomorphism. For well-defined, it suffices to show that elements $\in D$.

We denote **tensor product** of M and N as $M \otimes_R N = T = F/D$. Any element is of the form

$$r_1(m_1, n_1) + ... + r_k(m_k, n_k) + D = \underbrace{(r_1m_1, n_1) + ... + (r_km_k, n_k) + D}_{:= r_1m_1 \otimes n_1 + ... + r_km_k \otimes n_k}$$

Proposition 3.8.1. The following properties are satisfied:

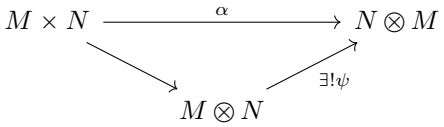
1. $m \otimes (n_1 + n_2) = m \otimes n_1 + m \otimes n_2$
2. $(m_1 + m_2) \otimes n = m_1 \otimes n + m_2 \otimes n$

3. $(rm) \otimes n = r(m \otimes n) = m \otimes (rn)$
4. $0 \otimes n = 0 = m \otimes 0$

Example 3.8.2.
- $\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathbb{Q} = \{0\}$: $a \otimes \frac{b}{c} = a \otimes \frac{bp}{cp} = pa \otimes \frac{b}{cp} = 0 \otimes \frac{b}{cp} = 0$.
- $\mathbb{Z}_2 \otimes \mathbb{Z}_3 = \{0\} : 0 \otimes x = 0, 1 \otimes 0, 2 = 0$. Finally $1 \otimes 1 = 1 \otimes (2+2) = 2 \otimes 1 + 2 \otimes 1 = 0 + 0 = 0$.
- $gcd(m, n) = 1, \mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n = \{0\}$

Proposition 3.8.3. If M, N, P are R-modules, then
- $M \otimes_R N \simeq N \otimes_R M$
- $(M \otimes_R N) \otimes_R P \simeq M \otimes_R (N \otimes_R P)$
- $M \otimes_R (N \oplus P) \simeq M \otimes_R N \bigoplus M \otimes_R P$
- $M \otimes_R R \simeq R \otimes_R M \simeq M$

Proposition 1 Proof. $M \times N \xrightarrow{\alpha} N \otimes M$ is clearly bilinear, $(m, n) \mapsto n \otimes m$



By the universal property, we have R-homomorphism $\psi(m \otimes n) = \alpha(m, n) = n \otimes m$. Conversely, $\exists R$-homomorphism $\phi : N \otimes M \to M \otimes N$, and $n \otimes m \mapsto m \otimes n$, and $\phi \circ \psi$ and $\psi \circ \phi$ are identity maps. □

Proposition 2 Proof. Fix $m \in M$ and define $\alpha_m : N \times P \to (M \otimes N) \otimes P, (n, p) \mapsto (m \otimes n) \otimes p$. Then, α_m is bilinear: $\alpha_m(n, p_1 + p_2) = \alpha_m(n, p_1) + \alpha_m(n, p_2)$. $\alpha_m(n_1 + n_2, p) = \alpha_m(n_1, p) + \alpha_m(n_2, p)$. $\alpha_m(m, p) = r\alpha_m(n, p)$. $\alpha_m(n, rp) - r\alpha_m(n, p)$. Together, this implies that $\exists R$-homomorphism $\psi_m : N \otimes P \longrightarrow (M \otimes N) \otimes P$.

Now, we have a bilinear map $\psi : M \times (N \otimes P) \to (M \otimes N) \otimes P, \psi(m, x) = \psi_m(x)$ and show that this is bilinear.
- $\psi(m, x_1 + x_2) = \psi(m, x_1) + \psi(m, x_2)$
- $\psi(m, rx) = r\psi(m, x)$

So ψ_m is a R-homomorphism. Also $\psi(m_1 + m_2, x) = \psi(m_1, x) + \psi(m_2, x)$ and $\psi(rm, x) = r\psi(m, x)$ so $\psi_{m_1+m_2} = \psi_{m_1} + \psi_{m_2}$.

Since there is a bilinear map, $\exists R$-homomorphism $\gamma : M \otimes (N \otimes P) \to (M \otimes N) \otimes P, m \otimes (n \otimes p) = (m \otimes n) \otimes p$.

Similarly, there is a $R-$homomorphism $\beta : (M \otimes N) \otimes P = M \otimes (N \otimes P), (m \otimes n) \otimes p \mapsto m \otimes (n \otimes p)$. γ, β are inverse maps, so they are isomorphisms. □

Proposition 4 Proof. There is a binear map $M \times R \xrightarrow{\alpha} M, (m, r) \mapsto rm$ bilinear. So there is an R-homomorphism $\psi : M \otimes R \to M, m \otimes r \mapsto rm$. Also there is an R-homomorphism $\phi : M \to M \otimes R, m \mapsto m \otimes 1$. $\psi \circ \phi = id, \phi \circ \psi(m \otimes r) = \phi(rm) = rm \otimes 1 = m \otimes r \implies \phi \circ \psi = id \implies \phi$ isomorphism. □

Example 3.8.4. Consider $R[x] \otimes_R R[x]$, where R is a commutative ring, we claim that $R[x] \otimes R[x] \simeq R[x, y]$.

Let $\phi : R[x] \otimes_R r[x] \to R[x, y]$ be the R-homomorphism induced by the bilinear map $R[x] \times R[x] \longrightarrow R[x, y], (f(x), g(x)) \mapsto f(x)g(y)$.

To define ψ, note that $R[x, y]$ is a free module over R with basis $x^i y^j, 0 \leqslant i, j$. Let $\psi : R[x, y] \to R[x] \otimes_R R[x]$ be such that $\psi(x^i y^j) = x^i \otimes x^j$.

ϕ, ψ are inverse maps: $x^iy^j \xrightarrow{\psi} x^i \otimes x^j \xrightarrow{\phi} x^iy^j$, $f(x) \otimes g(x) = \sum_{i,j} c_{i,j} x^i \otimes x^j, x^i \otimes x^j \xrightarrow{\phi} x^iy^j \xrightarrow{\psi} x^i \otimes x^j$.

Proposition 3.8.5. Let $0 \to M' \to M \to M'' \to 0$ be a short exact sequence of R-modules, and let N be an R module, then

$$M' \otimes_R N \to M \otimes_R N \to M'' \otimes_R N \to 0$$

is exact. Here, $M' \xrightarrow{f} M$ induces $M' \otimes N \xrightarrow{f \otimes id} M \otimes N$, $\sum m_i' \otimes n_i \mapsto \sum f(m_i') \otimes n_i$.

Lemma 3.8.6. Let M, N, Q be R modules, then $\mathrm{Hom}_R(M \otimes_R N, Q) \simeq \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, Q))$.

Corollary 3.8.7. If $Q = R$, $(M \otimes_R N)^\vee \simeq \mathrm{Hom}_R(M, N^\vee)$.

Example 3.8.8. Let k be a field, $R = k[x,y]/(x,y), M = R/(x), N = R/(y)$. Then, $M \otimes_R N = R/(x) \otimes R(y) \simeq R/(x,y)$. Also, $(M \otimes_R N)^\vee \simeq (R/(x,y))^\vee = \mathrm{Hom}_R(R/(x,y), R) = \{0\}$.

Also, $M^\vee = \mathrm{Hom}(R/(x), R) \simeq M, N^\vee = \mathrm{Hom}(R/(y), R) \simeq N$. Consider $\phi : R/(x) \to R, 1 \mapsto \bar{f}, 0 = \bar{x} \mapsto \overline{xf} = 0, f \in k[x,y] \implies xf \in (xy) \implies f \in (y)$.

So $M^\vee \otimes N^\vee \simeq M \otimes N \simeq R/(x,y) \neq \{0\}$.

Proposition Proof using Lemma. If $M' \to M \to M'' \to 0$ is exact, then let Q be an arbitrary R-module and take $\mathrm{Hom}(-, \mathrm{Hom}_R(N, Q))$. Then we have exact sequence

$$0 \to \mathrm{Hom}(M'', \mathrm{Hom}_R(M'', Q)) \to \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, Q)) \to \mathrm{Hom}_R(M', \mathrm{Hom}(N, Q))$$

So we have an exact sequence

$$0 \to \mathrm{Hom}_R(M'' \otimes N, Q) \to \mathrm{Hom}_R(M \otimes N, Q) \to \mathrm{Hom}_R(M' \otimes N, Q)$$

So by homework 9 question, $M' \otimes_R N \to M \otimes_R N \to M'' \otimes_R N \to 0$ is exact. □

Example 3.8.9. Let $0 \to \mathbb{Z} \xrightarrow{f} \mathbb{Z} \to \mathbb{Z}_2$ be a short exact sequence of $\mathbb{Z}$-modules and tensored with $\mathbb{Z}_2$, where $f : a \mapsto 2a$.

Then, $\underbrace{\mathbb{Z} \otimes \mathbb{Z}_2}_{\simeq \mathbb{Z}_2} \to \mathbb{Z} \otimes \mathbb{Z}_2$. [fill in from notes]

Proof of Lemma. Define $\phi : \mathrm{Hom}_R(M \otimes_R N, Q) \to \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, Q))$, where $(\alpha : M \otimes N \to P) \mapsto (\beta : M \to \mathrm{Hom}_R(N, Q))$. $\beta : m \mapsto \beta_m, \beta(n) = \alpha(m \otimes n) \in Q$.

I need to show that β is R-homomorphism, ϕ is R-homomorphism.

β homomorphism: $\beta \in \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, Q))$: Show that $\beta_{r_1m_1 + r_2m_2} = r_1\beta_{m_1} + r_2\beta_{m_2}$. So, $\beta_{r_1m_1+r_2m_2}(n) = \alpha((r_1m_1 + r_2m_2) \otimes n) = \alpha(r_1(m_1 \otimes n) + r_2(m_2 \otimes n))$, and $(r_1\beta_{m_1} + r_2\beta_{m_2})(n) = r_1\alpha(m_1 \otimes n) + r_2\alpha(m_2 \otimes n)$, which is true

ϕ homomorphism shown similarly.

Also define $\psi : \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, Q)) \to \mathrm{Hom}_R(M \otimes_R N, Q)$ with $\beta : M \to \mathrm{Hom}_R(N, Q)$ given. Define bilinear map $M \times N \to Q, (m, n) \mapsto \beta(m)(n)$, this gives a map $\alpha : M \otimes_R N \to Q$.

So ϕ, ψ are inverse maps. □

Definition 3.8.10. A module F is **flat** if for any short exact sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$, the following sequence is exact:

$$0 \to M' \otimes F \xrightarrow{f \otimes id} M \otimes F \xrightarrow{g \otimes id} M'' \otimes F \to 0$$

Equivalently, F is **flat** if for any R-homomorphism $f : M' \to M, M' \otimes F \to M \otimes N$ is injective.

Example 3.8.11. $\mathbb{Z}_2$ is not a flat $\mathbb{Z}$-module. Consider $\mathbb{Z} \to \mathbb{Z}, n \mapsto 2n$. $\mathbb{Z} \otimes \mathbb{Z}_2 \to \mathbb{Z} \otimes \mathbb{Z}_2, a \otimes b \mapsto 2a \otimes b = a \otimes 2b = 0$. Not injective, so this is not flat.

Example 3.8.12. Suppose R is an integral domain:

- Free modules are flat. If F is a free R-module, $F \simeq \bigoplus_{i \in I} R, f : M' \to M$ is an injective map that gives the following injectivity.

$$\begin{array}{ccccccc} M' \otimes F & & M' \otimes (\bigoplus_i R) & & \bigoplus_i M' \otimes R & & \bigoplus_i M' \\ \downarrow{\scriptstyle f \otimes id} & \simeq & \downarrow{\scriptstyle f \otimes id} & \simeq & \downarrow{\scriptstyle \oplus f \otimes id} & \simeq & \downarrow{\scriptstyle \oplus f} \\ M \otimes F & & M \otimes (\bigoplus_i R) & & \bigoplus_i M \otimes R & & \bigoplus_i M \end{array}$$

- More generally, projective modules are flat. If P is projective, $\exists P'$ s.t. for a free module F, $F = P \oplus P'$. Then if $M' \to M$ is injective, then $M' \otimes F \to M \otimes F$ by the previous example. So $M' \otimes P \bigoplus M' \otimes P' \longrightarrow M \otimes P \bigoplus M \otimes P'$ is an injective map $\implies M' \otimes P \to M \otimes P$ is injective.
- Flat module does not necessarily imply projective modules. $\mathbb{Q}$ as a $\mathbb{Z}$-module is flat. [Check 11/29 minute 30 for proof] But $\mathbb{Q}$ is not projective. Suppose $\mathbb{Q} \oplus P'$ is free, then pick a basis and write $(1,0) = \lambda_1 x_1 + ... + \lambda_n x_n$, $x_1, ..., x_n$ part of a basis and $\lambda_1, ..., \lambda_n \in \mathbb{Z}$. Pick N where $N > |\lambda_1|, ..., |\lambda_n|$. Then write $(\frac{1}{N}, 0)$ as a combination of basis elements, where $(\frac{1}{N}, 0) = c_1 x_1 + ... + c_n x_n$, where $c_1, ..., c_n \in \mathbb{Z}$ may be 0. So $(1,0) = Nc_1 x_1 + ... + Nc_n x_n$. If $c_i \neq 0$, then $|Nc_i| > |\lambda_i|$, so they cannot be equal.
- If F is a flat R-module, then it is torsion-free. We need to show that if $0 \neq x \in F$ and $0 \neq r \in R$, then $rx \neq 0$. Let $R \xrightarrow{f} R$, $s \mapsto rs$ be multiplication by r. Then f is injective since R is an integral domain. So, $R \otimes F \xrightarrow{f \otimes id} R \otimes F$ is injective. $0 \neq 1 \otimes x \mapsto r \otimes x = 1 \otimes rx$. So $1 \otimes rx \neq 0, rx \neq 0$

Note: Free $\implies$ Projective $\implies$ Flat $\implies$ Torsion-free

Let $R \xrightarrow{f} S$ be a ring homomorphism.

- Any S-module M has the structure of an R-module, $rm : f(r)m$
- Now, suppose N is a module over R. $N \otimes_R S$ is a R-module which has the structure of S-module, $s(n_1 \otimes s_1) := n_1 \otimes ss_1$

If $\phi : N_1 \to N_2$ is a R-homomorphism, $\phi \otimes id : N_1 \otimes S \to N_2 \otimes_R S$ is a S-homomorphism.

Chapter 4

Fields

We will extensively use J.S. Milne's tex file for field theory and Galois theory.

4.1 Basic Definitions

Note: we will use $R[X]$ instead of $R[x]$ now to emphasize the indeterminant X in the polynomial ring.

Definition 4.1.1. A **field** is a set F with two composition laws $+$ and $\cdot$ such that

1. $(F, +)$ is a commutative group;
2. $(F^\times, \cdot)$, where $F^\times = F \smallsetminus \{0\}$, is a commutative group;
3. the distributive law holds.

Thus, a field is a nonzero commutative ring such that every nonzero element has an inverse. In particular, it is an integral domain. A field contains at least two distinct elements, 0 and 1. The smallest, and one of the most important, fields is $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$.

A **subfield** S of a field F is a subring that is closed under passage to the inverse. It inherits the structure of a field from that on F.

We have shown

Lemma 4.1.2. A nonzero commutative ring R is a field if and only if it has no ideals other than (0) and R.

Example 4.1.3. The following are fields: $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ (p prime).

Definition 4.1.4. A **homomorphism of fields** is simply a homomorphism of rings. Such a homomorphism is always injective, because its kernel is a proper ideal (it doesn't contain 1), which must therefore be zero.

Let F be a field. An **F-algebra** (or **algebra over** F) is a ring R containing F as a subring (so the inclusion map is a homomorphism). A **homomorphism of F-algebras** $\alpha\colon R \to R'$ is a homomorphism of rings such that $\alpha(c) = c$ for every $c \in F$.

Remark 4.1.5. Let F be a field.

The ring $F[X]$ of polynomials in the symbol (or "indeterminate" or "variable") X with coefficients in F is an F-vector space with basis $1, X, \dots, X^n, \dots$, and with the multiplication

$$\left(\sum_i a_i X^i\right)\left(\sum_j b_j X^j\right) = \sum_k \left(\sum_{i+j=k} a_i b_j\right) X^k.$$

The F-algebra $F[X]$ has the following universal property: for any F-algebra R and element r of R, there is a unique homomorphism of F-algebras $\alpha : F[X] \to R$ such that $\alpha(X) = r$.

4.1.1 Generated Subrings and Subfields

An intersection of subrings of a ring is again a ring (this is easy to prove). Let F be a subfield of a field E, and let S be a subset of E. The intersection of all the subrings of E containing F and S is obviously the smallest subring of E containing both F and S. We call it the subring of E **generated by** F **and** S (or **generated over** F **by** S), and we denote it by $F[S]$. When $S = \{\alpha_1, ..., \alpha_n\}$, we write $F[\alpha_1, ..., \alpha_n]$ for $F[S]$. For example, $\mathbb{C} = \mathbb{R}[\sqrt{-1}]$.

Lemma 4.1.6. The ring $F[S]$ consists of the elements of E that can be expressed as finite sums of the form

$$\sum a_{i_1 \cdots i_n} \alpha_1^{i_1} \cdots \alpha_n^{i_n}, \quad a_{i_1 \cdots i_n} \in F, \quad \alpha_i \in S, \quad i_j \in \mathbb{N}. \tag{4.1}$$

Proof. Let R be the set of all such elements. Obviously, R is a subring of E containing F and S and contained in every other such subring. Therefore it equals $F[S]$. □

Example 4.1.7. The ring $\mathbb{Q}[\pi]$, $\pi = 3.14159...$, consists of the real numbers that can be expressed as a finite sum

$$a_0 + a_1\pi + a_2\pi^2 + \cdots + a_n\pi^n, \quad a_i \in \mathbb{Q}.$$

The ring $\mathbb{Q}[i]$ consists of the complex numbers of the form $a + bi$, $a, b \in \mathbb{Q}$.

Note that the expression of an element in the form (4.1) will *not* be unique in general. This is so already in $\mathbb{R}[i]$.

Lemma 4.1.8. Let R be an integral domain containing a subfield F (as a subring). If R is finite-dimensional when regarded as an F-vector space, then it is a field.

Proof. Let α be a nonzero element of R — we have to show that α has an inverse in R. The map $R \to R\colon x \mapsto \alpha x$ is an injective linear map of finite-dimensional F-vector spaces, and is therefore surjective. In particular, there is an element $\beta \in R$ such that $\alpha\beta = 1$. □

Note that the lemma applies to every subring containing F of a finite extension of F.

An intersection of subfields of a field is again a field. Let F be a subfield of a field E, and let S be a subset of E. The intersection of all the subfields of E containing F and S is obviously the smallest subfield of E containing both F and S. We call it the subfield of E **generated by** F **and** S (or **generated over** F **by** S), and we denote it $F(S)$. It is the field of fractions of $F[S]$ in E because this is a subfield of E containing F and S and contained in every other such field. When $S = \{\alpha_1, ..., \alpha_n\}$, we write $F(\alpha_1, ..., \alpha_n)$ for $F(S)$. Thus, $F[\alpha_1, \ldots, \alpha_n]$ consists of all elements of E that can be expressed as polynomials in the α_i with coefficients in F, and $F(\alpha_1, \ldots, \alpha_n)$ consists of all elements of E that can be expressed as a quotient of two such polynomials.

If $\alpha_1, \ldots, \alpha_k \in E$, then

$$\underbrace{F(\alpha_1, \ldots, \alpha_k)}_{\text{a finitely generated extension}} = \text{subfield of } E \text{ generated by } F, \alpha_1, \ldots, \alpha_k$$

$$= \left\{ \frac{f(\alpha_1, \ldots, \alpha_k)}{g(\alpha_1, \ldots, \alpha_k)} \;\middle|\; \begin{array}{l} f, g \in F[x_1, \ldots, x_k] \\ g(\alpha_1, \ldots, \alpha_k) \neq 0 \end{array} \right\}$$

Note that

$$F \subset F(\alpha_1) \subset F(\alpha_1, \alpha_2) \subset \cdots \subset F(\alpha_1, \ldots, \alpha_k) \subset E$$
$$F(\alpha_1, \ldots, \alpha_k) = F(\alpha_1, \ldots, \alpha_{k-1})(\alpha_k)$$

Remark 4.1.9. Lemma 4.1.8 shows that $F[S]$ is already a field if it is finite-dimensional over F, in which case $F(S) = F[S]$.

Example 4.1.10. (a) The field $\mathbb{Q}(\pi)$, $\pi = 3.14\ldots$, consists of the complex numbers that can be expressed as a quotient

$$g(\pi)/h(\pi), \quad g(X), h(X) \in \mathbb{Q}[X], \quad h(X) \neq 0.$$

(b) The ring $\mathbb{Q}[i]$ is already a field.

Example 4.1.11. Suppose E/F is a field extension. $\alpha \in E$. Above definitions of generated subrings and subfields give

$$F[\alpha] = \{b_m\alpha^m + \cdots + b_1\alpha + b_0 | b_i \in F\}$$

and

$$F(\alpha) = \left\{ \frac{b_m\alpha^m + \cdots + b_1\alpha + b_0}{c_r\alpha^r + \cdots + c_1\alpha + c_0} : \ b_i, c_j \in F \text{ and } c_r\alpha^r + \cdots + c_0 \neq 0 \right\}$$

4.1.2 The Characteristic of a Field

One checks easily that the map

$$\mathbb{Z} \to F, \quad n \mapsto n \cdot 1_F \overset{\text{def}}{=} 1_F + 1_F + \cdots + 1_F \quad (n \text{ copies of } 1_F),$$

is a homomorphism of rings. For example,

$$(\underbrace{1_F + \cdots + 1_F}_{m}) + (\underbrace{1_F + \cdots + 1_F}_{n}) = \underbrace{1_F + \cdots + 1_F}_{m+n}$$

because of the associativity of addition. Therefore its kernel is an ideal in $\mathbb{Z}$.

CASE 1: The kernel of the map is (0), so that

$$n \cdot 1_F = 0 \quad (\text{in } F) \implies n = 0 \quad (\text{in } \mathbb{Z}).$$

Nonzero integers map to invertible elements of F under $n \mapsto n \cdot 1_F \colon \mathbb{Z} \to F$, and so this map extends to a homomorphism

$$\mathbb{Q} \hookrightarrow F \colon \frac{m}{n} \mapsto (m \cdot 1_F)(n \cdot 1_F)^{-1}.$$

In this case, F contains a copy of $\mathbb{Q}$, and we say that it has **characteristic zero**.

Thus **characteristic** of a field F is the order of 1, as an element of the additive group F^+, provided that the order is finite. It is the smallest positive integer n such that the sum $1 + \cdots + 1$ of n copies of 1 evaluates to 0. If the order is infinite, that is, $1 + \cdots + 1$ is never 0 in F, the field is then said to have **characteristic zero**. We denote the characteristic of a field by $\mathrm{char}(F)$.

CASE 2: The kernel of the map is $\neq (0)$, so that $n \cdot 1_F = 0$ for some $n \neq 0$. The smallest positive such n will be a prime p (otherwise there will be two nonzero elements in F whose product is zero), and p generates the kernel. Thus, the map $n \mapsto n \cdot 1_F \colon \mathbb{Z} \to F$ defines an isomorphism from $\mathbb{Z}/p\mathbb{Z}$ onto the subring

$$\{m \cdot 1_F \mid m \in \mathbb{Z}\}$$

of F. In this case, F contains a copy of $\mathbb{F}_p$, and we say that it has **characteristic** p.

A field isomorphic to one of the fields $\mathbb{F}_2, \mathbb{F}_3, \mathbb{F}_5, \ldots, \mathbb{Q}$ is called a **prime field**. Every field contains exactly one prime field (as a subfield).

More generally, a commutative ring R is said to have **characteristic** p (resp. 0) if it contains a prime field (as a subring) of characteristic p (resp. 0).[1] Then the prime field is unique and, by definition, contains 1_R. Thus, if R has characteristic $p \neq 0$, then $1_R + \cdots + 1_R = 0$ (p terms).

Let R be a nonzero commutative ring. If R has characteristic $p \neq 0$, then

$$pa \overset{\text{def}}{=} \underbrace{a + \cdots + a}_{p \text{ terms}} = \underbrace{(1_R + \cdots + 1_R)}_{p \text{ terms}} a = 0a = 0$$

for all $a \in R$. Conversely, if $pa = 0$ for all $a \in R$, then R has characteristic p.

Let R be a nonzero commutative ring. The usual proof by induction shows that the binomial theorem

$$(a+b)^m = a^m + \tbinom{m}{1}a^{m-1}b + \tbinom{m}{2}a^{m-2}b^2 + \cdots + b^m$$

holds in R. If p is prime, then it divides

$$\binom{p}{r} \overset{\text{def}}{=} \frac{p!}{r!(p-r)!}$$

for all r with $1 \leqslant r \leqslant p-1$ because it divides the numerator but not the denominator. Therefore, when R has characteristic p,

$$(a+b)^p = a^p + b^p \quad \text{for all } a, b \in R,$$

and so the map $R \to R\colon a \mapsto a^p$ is a homomorphism of rings (even of $\mathbb{F}_p$-algebras). It is called the **Frobenius endomorphism** of R. The map $R \to R\colon a \mapsto a^{p^n}$, $n \geqslant 1$, is the composite of n copies of the Frobenius endomorphism, and so it also is a homomorphism. Therefore,

$$(a_1 + \cdots + a_m)^{p^n} = a_1^{p^n} + \cdots + a_m^{p^n}$$

for all $a_i \in R$.

When F is a field, the Frobenius endomorphism is injective, and hence an automorphism if F is finite.

The **characteristic exponent** of a field F is 1 if F has characteristic 0, and p if F has characteristic $p \neq 0$. Thus, if q is the characteristic exponent of F and $n \geqslant 1$, then $x \mapsto x^{q^n}$ is an isomorphism of F onto a subfield of F (denoted F^{q^n}).

Example 4.1.12. The polynomial ring in one variable $R[X]$ over an integral domain R is an integral domain. The **field of rational fractions in one variable** $R(X)$ is the field of fractions of $R[X]$.

Example 4.1.13. Subfields F of $\mathbb{C}$ have $\operatorname{char}(F) = 0$. $\operatorname{char}(\mathbb{Q}) = 0$. $\operatorname{char}(\mathbb{Z}_p) = p$ with p prime.

Proposition 4.1.14. The characteristic of any field F is either zero or a prime number.

Proof. To avoid confusion, we let $\overline{0}$ and $\overline{1}$ denote the additive and the multiplicative identities in the field F, respectively, and if k is a positive integer, we let $\bar{k}$ denote the sum of k copies of $\overline{1}$. Suppose that the characteristic m is not zero. Then $\overline{1}$ generates a cyclic subgroup H of $F^+$of order m, and $\bar{m} = \overline{0}$. The distinct elements of the cyclic subgroup H generated by $\overline{1}$ are the elements $\bar{k}$ with $k = 0, 1, \ldots, m-1$. Suppose that m isn't prime, say $m = rs$, with $1 < r, s < m$. Then $\bar{r}$ and $\bar{s}$ are in the multiplicative group $F^\times = F - \{0\}$, but the product $\bar{r}\bar{s}$, which is equal to $\overline{0}$, is not in $F^\times$. This contradicts the fact that $F^\times$is a group. Therefore m must be prime. □

[1] A commutative ring has a characteristic if and only if it contains a field as a subring. For example, neither $\mathbb{Z}$ nor $\mathbb{F}_2 \times \mathbb{F}_3$ has a characteristic.

4.2 Field Extensions

Definition 4.2.1. If F is a subfield of E, then E is a called a **field extension** of F. The notation E/F will indicate that E is a field extension of F. We note that a field extension E of F can always be regarded as an F-vector space. Addition is the addition law in E, and scalar multiplication of an element of E by an element of F is obtained by multiplying these two elements in E. The dimension of E, when regarded as an F-vector space, is called the **degree** of the field extension. $[E : F] := \deg(E/F) =$ dim of E as a v.s. over F. An example of finite $[E : F]$ is $\mathbb{C}/\mathbb{R}$ with basis $1, i$; of an infinite one is $\mathbb{R}/\mathbb{Q}$. A field extension E/F is a finite extension if its degree is finite. Extensions of degree 2 are quadratic extensions, those of degree 3 are cubic extensions, and so on.

Example 4.2.2. The field of **Gaussian numbers**

$$\mathbb{Q}(i) \overset{\text{def}}{=} \{a + bi \in \mathbb{C} \mid a, b \in \mathbb{Q}\}$$

has degree 2 over $\mathbb{Q}$ (basis $\{1, i\}$).

The **field of rational fractions in one variable** $F(X)$ has infinite degree over F; in fact, even its subspace $F[X]$ has infinite dimension over F (basis $1, X, X^2, \ldots$).

Definition 4.2.3. If E/F is an extension. $\alpha \in E$. α is **algebraic over** F if there is a non-zero polynomial $0 \neq f(X) \in F[X]$ such that $f(\alpha) = 0$. Elements of E that are not algebraic over F are called **transcendental**. E/F is called an **algebraic extension** if every $\alpha \in E$ is algebraic over F.

Proposition 4.2.4. If $[E : F] < \infty$, then E is algebraic over F.

Proof. If $\alpha \in E$ and $[E : F] = n$, then $1, \alpha, \cdots, \alpha^n$ are linearly independent, so there are $c_0, \cdots, c_n \in F$ such that

$$c_0 + c_1\alpha + \cdots + c_n\alpha^n = 0$$

so if $f(X) = c_0 + c_1X + \cdots + c_nX^n \in F[X]$, then $f(\alpha) = 0$. □

Example 4.2.5. $\mathbb{C}/\mathbb{R}$ is algebraic. Let $z = a + ib$. Let $\bar{z} = a - ib$ be the complex conjugate of z. Note that

$$z\bar{z} = a^2 + b^2, \quad z + \bar{z} = 2a$$

This reminds us of the Viète's Formulas. Consider the polynomial

$$X^2 - 2aX + \left(a^2 + b^2\right)$$

Its roots are

$$X_{1,2} = \frac{2a \pm \sqrt{4a^2 - 4(a^2 + b^2)}}{2} = \frac{2a \pm 2bi}{2} = a \pm bi = z, \bar{z}$$

Thus both z and $\bar{z}$ are roots of the polynomial.

The converse of the proposition is incorrect: $\mathbb{Q} \subset \mathbb{R}$. Those of the form $\sqrt{p}$ with p prime are algebraic over $\mathbb{Q}$. $\left(\sqrt{p}\right)^2 - p = 0$. We will later show that $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \cdots) \subset \mathbb{R}$ gives a non-finite extension.

Proposition 4.2.6 (multiplicativity of degrees)**.** If $F \subset E \subset K$ and $[E : F] = n$ and $[K : E] = m$, then $[K : F] = mn$.

Proof. Let $x_1, \cdots, x_n$ be a basis for E/F and $y_1, \cdots, y_m$ be a basis of K/E. Then $x_iy_j,\ 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m$ is a basis for K/F:

- linear independency: if $\sum_{i,j}\lambda_{ij}x_iy_j = 0$ for $\lambda_{ij}\in F$, then

$$0=\sum_{i,j}\lambda_{ij}x_iy_j=\sum_{j=1}^{m}\underbrace{\left(\sum_{i=1}^{n}\lambda_{ij}x_i\right)}_{\in E}y_j \Rightarrow \sum_{i=1}^{n}\lambda_{ij}x_i=0\ \forall j\Rightarrow \lambda_{ij}=0\ \forall i,j$$

- span: if $z\in K$, then $z=\sum_{1\leqslant j\leqslant m}c_jy_j$ for $c_j\in E$. $c_j=\sum_{1\leqslant i\leqslant n}b_{ij}x_i$. $b_{ij}\in F$, so

$$z=\sum_{j=1}^{m}\sum_{i=1}^{n}b_{ij}x_iy_j$$

□

An extension E of F is said to be **simple** if $E=F(\alpha)$ some $\alpha\in E$. For example, $\mathbb{Q}(\pi)$ and $\mathbb{Q}[i]$ are simple extensions of $\mathbb{Q}$.

Let F and F' be subfields of a field E. The intersection of the subfields of E containing both F and F' is obviously the smallest subfield of E containing both F and F'. We call it the **composite** of F and F' in E, and we denote it by $F\cdot F'$. It can also be described as the subfield of E generated over F by F', or the subfield generated over F' by F:

$$F(F')=F\cdot F'=F'(F).$$

Let $f(X)\in F[X]$ be a monic polynomial of degree m, and let (f) be the ideal generated by f. Consider the quotient ring $F[X]/(f(X))$, and write x for the image of X in $F[X]/(f(X))$, i.e., x is the coset $X+(f(X))$.

(a) The map

$$\begin{aligned}F[X]\to F[x]&=F[X]/(f(X))\\ P(X)\mapsto P(x)&=P(X+(f(X)))=\sum a_i(X+f(X))^i\\ &=\sum a_i(X^i+(f(X)))=\sum a_iX^i+(f(X))=P(X)+(f(X))\end{aligned}$$

is a homomorphism sending $f(X)$ to $f(X)+(f(X))=(f(X))=0_{F[x]}$. Therefore, $f(x)=0$.

(b) The division algorithm shows that every element $g(x)$ of $F[X]/(f)$ is represented by a unique polynomial r of degree $<m$ (g is of the form $P(X)+(f(X))$. By division algorithm, $P(X)=q(X)f(X)+r(X)$. Thus $g=r(X)+q(X)f(X)+(f(X))=r(X)+(f(X))$). Hence each element of $F[x]$ can be expressed uniquely as a sum

$$a_0+a_1x+\cdots+a_{m-1}x^{m-1},\qquad a_i\in F. \tag{4.2}$$

(c) To add two elements, expressed in the form (4.2), simply add the corresponding coefficients.

(d) To multiply two elements expressed in the form (4.2), multiply in the usual way, and use the relation $f(x)=0$ to express the monomials of degree $\geqslant m$ in x in terms of lower degree monomials.

(e) *Now assume that $f(X)$ is irreducible*. Then every nonzero $\alpha\in F[x]$ has an inverse, which can be found as follows. Use (b) to write $\alpha=g(x)$ with $g(X)$ a polynomial of degree $\leqslant m-1$, and apply Euclid's algorithm in $F[X]$ to find polynomials $a(X)$ and $b(X)$ such that

$$a(X)f(X)+b(X)g(X)=d(X)$$

with $d(X)$ the gcd of f and g. In our case, $d(X)$ is 1 because $f(X)$ is irreducible and $\deg g(X)<\deg f(X)$. When we replace X with x, the equality becomes

$$b(x)g(x)=1.$$

Hence $b(x)$ is the inverse of $g(x)$.

We have proved the following statement.

Claim 4.2.7. For a monic irreducible polynomial $f(X)$ of degree m in $F[X]$,

$$F[x] \stackrel{\text{def}}{=} F[X]/(f(X))$$

is a field of degree m over F. Computations in $F[x]$ come down to computations in F.

Note that, because $F[x]$ is a field, $F(x) = F[x]$.[2]

Example 4.2.8. Let $f(X) = X^2 + 1 \in \mathbb{R}[X]$. Then $\mathbb{R}[x]$ has

elements: $a + bx$, $a, b \in \mathbb{R}$;

addition: $(a + bx) + (a' + b'x) = (a + a') + (b + b')x$;

multiplication: $(a + bx)(a' + b'x) = (aa' - bb') + (ab' + a'b)x$;

inverses: in this case, it is possible write down the inverse of $a + bx$ directly.

We usually write i for x and $\mathbb{C}$ for $\mathbb{R}[x]$.

Example 4.2.9. Let $f(X) = X^3 - 3X - 1 \in \mathbb{Q}[X]$. The polynomial is irreducible because its only possible roots in $\mathbb{Q}$ are ± 1 by rational root theorem (or directly by lemma 2.13.3 (c)), but $f(1) \neq 0 \neq f(-1)$. Then $\mathbb{Q}[x]$ is a field. It has basis $\{1, x, x^2\}$ as a $\mathbb{Q}$-vector space. Let

$$\beta = x^4 + 2x^3 + 3 \in \mathbb{Q}[x].$$

Then using that $x^3 - 3x - 1 = 0$, we find that $\beta = 3x^2 + 7x + 5$. This is done by commands below.

```
sage: R.<x> = PolynomialRing(QQ)
sage: f = x^4 + 2*x^3 +3
sage: g = x^3 - 3*x - 1
sage: f.quo_rem(g)

(x + 2, 3*x^2 + 7*x + 5)
```

Because $X^3 - 3X - 1$ is irreducible,

$$\gcd(X^3 - 3X - 1, 3X^2 + 7X + 5) = 1.$$

Euclid's algorithm gives

$$(X^3 - 3X - 1)\left(\tfrac{-7}{37}X + \tfrac{29}{111}\right) + (3X^2 + 7X + 5)\left(\tfrac{7}{111}X^2 - \tfrac{26}{111}X + \tfrac{28}{111}\right) = 1.$$

Hence

$$(3x^2 + 7x + 5)\left(\tfrac{7}{111}x^2 - \tfrac{26}{111}x + \tfrac{28}{111}\right) = 1,$$

and we have found the inverse of β.

We can also do this in PARI: `b=Mod(X^4+2*X^3+3,X^3-3*X-1)` reveals that $\beta = 3x^2 + 7x + 5$ in $\mathbb{Q}[x]$, and `b^(-1)` reveals that $\beta^{-1} = \frac{7}{111}x^2 - \frac{26}{111}x + \frac{28}{111}$.

Let f be a monic irreducible polynomial in $F[X]$. A pair (E, α) consisting of an extension E of F and an $\alpha \in E$ is called[3] a **stem field for** f if $E = F[\alpha]$ and $f(\alpha) = 0$. For example, the pair (E, α) with

[2] Thus, we can denote it by $F(x)$ or by $F[x]$. The former is more common, but I use $F[x]$ to emphasize the fact that its elements are polynomials in x.

[3] Following A.A. Albert (Modern Higher Algebra, 1937) who calls the splitting field of a polynomial its root field.

$E = F[X]/(f) = F[x]$ and $\alpha = x$ is a stem field for f. Let (E, α) be a stem field, and consider the surjective homomorphism of F-algebras

$$F[X] \to E\colon g(X) \mapsto g(\alpha).$$

Its kernel is generated by a nonzero monic polynomial, which divides f, and so must equal it. Therefore the homomorphism defines an F-isomorphism

$$F[x] \to E\colon x \mapsto \alpha, \quad \text{where } F[x] = F[X]/(f).$$

In other words, the stem field (E, α) of f is F-isomorphic to the standard stem field $(F[X]/(f), x)$. It follows that every element of a stem field (E, α) for f can be written uniquely in the form

$$a_0 + a_1\alpha + \cdots + a_{m-1}\alpha^{m-1}, \quad a_i \in F, \quad m = \deg(f),$$

and that arithmetic in $F[\alpha]$ can be performed using the same rules as in $F[x]$. If (E', α') is a second stem field for f, then there is a unique F-isomorphism $E \to E'$ sending α to α'. We sometimes abbreviate "stem field $(F[\alpha], \alpha)$" to "stem field $F[\alpha]$".

4.3 Algebraic and Transcendental Elements

Let F be a field. We view the algebraic and transcendental elements in another way. Recall from the substitution principle that an element α of an extension E of F defines a homomorphism

$$\begin{aligned} \Phi : F[X] &\to E \\ f(X) &\mapsto f(\alpha). \end{aligned}$$

There are two possibilities.

CASE 1: The kernel of the map is (0), so that, for $f \in F[X]$,

$$f(\alpha) = 0 \implies f = 0 \text{ (in } F[X]).$$

In this case, we say that α **transcendental over** F. The homomorphism $F[X] \to F[\alpha]\colon X \mapsto \alpha$ is an isomorphism, and it extends to an isomorphism $F(X) \to F(\alpha)$ of the fields of fractions.

CASE 2: The kernel is $\neq (0)$, so that $g(\alpha) = 0$ for some nonzero $g \in F[X]$. In this case, we say that α is **algebraic over** F. The polynomials g such that $g(\alpha) = 0$ form a nonzero ideal in $F[X]$ (the kernel of the substitution homomorphism),

$$I = \{g(X) \in F[X] \mid g(\alpha) = 0\} \subseteq F[X].$$

Then there is some f generating I as F being a field makes $F[X]$ PID. This f is the monic polynomial of least degree such $f(\alpha) = 0$. We call f the **minimal (**or **minimum) polynomial** of α over F.[4]

f is irreducible: suppose not then $f(X) = p(X)q(X)$. f being monic by definition implies that $0 < \deg(p), \deg(q) < \deg(f)$. f having least degree in I implies that $\Rightarrow p, q \notin I$ so $p(\alpha) \neq 0 \neq q(\alpha)$, which contradicts to the fact that $0 = f(\alpha) = p(\alpha)q(\alpha)$ because two nonzero elements in field (thus an integral domain) E cannot multiply to get 0.

The minimal polynomial is characterized as an element of $F[X]$ by each of the following conditions,

- f is monic, $f(\alpha) = 0$, and f divides every other g in $F[X]$ such that $g(\alpha) = 0$ (that's because $g = qf + r \Rightarrow 0 = g(\alpha) = q(\alpha)f(\alpha) + r(\alpha) = r(\alpha) \Rightarrow r \in I$, but it cannot be the case tht $\deg(r) < \deg(f)$ so r has to be 0);

[4]When we order the polynomials by degree, f is a minimal element of the set of polynomials having α as a root. It is also the *unique* minimal (hence least or minimum) element of the set of *monic* polynomials having α as a root. See Wikipedia: partially ordered set.

- f is the monic polynomial of least degree such that $f(\alpha) = 0$ (this is the first definition we used above);
- f is monic, irreducible, and $f(\alpha) = 0$.

Since f is the generator of the kernel of $\Phi : F[X] \to E$, the first isomorphism theorem implies that $F[x] = F[X]/(f) \cong \mathrm{Im}(\Phi) = F[\alpha]$. Explicitly, this map sends $g(x) = g(X) + (f)$ to $g(\alpha)$. Since $F[x]$ is a field due to claim 4.2.7, so also is $F[\alpha]$,

$$F(\alpha) = F[\alpha].$$

Thus, $F[\alpha]$ is a stem field for f.

Example 4.3.1.

1. $\mathbb{R} \subset \mathbb{C}$. $\alpha = i$. The minimal polynomial is $p = x^2 + 1$. $\mathbb{C} = \mathbb{R}[i] = \{a + bi : a, b \in \mathbb{R}\}$.
2. $\mathbb{Q} \subset \mathbb{R}$. $\alpha = \sqrt[3]{2}$. The minimal polynomial is $x^3 - 2$. $\mathbb{Q}(\alpha) = \{a\alpha^2 + b\alpha + c | a, b, c \in \mathbb{Q}\}$.

Example 4.3.2. Let $\alpha \in \mathbb{C}$ be such that $\alpha^3 - 3\alpha - 1 = 0$. Then $X^3 - 3X - 1$ is monic, irreducible, and has α as a root, and so it is the minimal polynomial of α over $\mathbb{Q}$. The set $\{1, \alpha, \alpha^2\}$ is a basis for $\mathbb{Q}[\alpha]$ over $\mathbb{Q}$. The calculations in Example 4.2.9 show that if β is the element $\alpha^4 + 2\alpha^3 + 3$ of $\mathbb{Q}[\alpha]$, then $\beta = 3\alpha^2 + 7\alpha + 5$, and

$$\beta^{-1} = \tfrac{7}{111}\alpha^2 - \tfrac{26}{111}\alpha + \tfrac{28}{111}.$$

Remark 4.3.3. PARI knows how to compute in $\mathbb{Q}[a]$. For example, `factor(X^4+4)` returns the factorization

$$X^4 + 4 = (X^2 - 2X + 2)(X^2 + 2X + 2)$$

in $\mathbb{Q}[X]$. Now type `F=nfinit(a^2+2*a+2)` to define a number field "F" generated over $\mathbb{Q}$ by a root a of $X^2 + 2X + 2$. Then `nffactor(F,x^4+4)` returns the factorization

$$X^4 + 4 = (X - a - 2)(X - a)(X + a))(X + a + 2),$$

in $\mathbb{Q}[a]$.

A extension E of F is said to be **algebraic** (and E is said to be **algebraic over** F), if all elements of E are algebraic over F, i.e., each element of E has some polynomial over F vanishing it; otherwise it is said to be **transcendental** (and E is said to be **transcendental over** F). Thus, E/F is transcendental if at least one element of E is transcendental over F.

Proposition 4.3.4. Let $E \supset F$ be fields. If E/F is finite, then E is algebraic and finitely generated (as a field) over F; conversely, if E is generated over F by a finite set of algebraic elements, then it is finite (and hence algebraic) over F.

Proof.
$\Longrightarrow$: To say that an element α of E is transcendental over F amounts to saying that its powers $1, \alpha, \alpha^2, \ldots$ are linearly independent over F. Thus, if E is finite over F, then every element of E is algebraic over F. It remains to show that E is finitely generated over F. If $E = F$, then it is generated by the empty set. Otherwise, there exists an $\alpha_1 \in E \smallsetminus F$. If $E \neq F[\alpha_1]$, then there exists an $\alpha_2 \in E \smallsetminus F[\alpha_1]$, and so on. Since

$$[F[\alpha_1]\colon F] < [F[\alpha_1, \alpha_2]\colon F] < \cdots < [E\colon F]$$

this process terminates with $E = F[\alpha_1, \alpha_2, \ldots, \alpha_n]$.

$\Longleftarrow$: Let $E = F(\alpha_1, ..., \alpha_n)$ with $\alpha_1, \alpha_2, \ldots \alpha_n$ algebraic over F. The extension $F(\alpha_1)/F$ is finite because α_1 is algebraic over F, and the extension $F(\alpha_1, \alpha_2)/F(\alpha_1)$ is finite because α_2 is algebraic over F and hence over $F(\alpha_1)$. Thus, by (4.2.6), $F(\alpha_1, \alpha_2)$ is finite over F. Now repeat the argument. □

Corollary 4.3.5.
(a) If E is algebraic over F, then every subring R of E containing F is a field.

(b) Consider fields $L \supset E \supset F$. If L is algebraic over E and E is algebraic over F, then L is algebraic over F.

Proof.
(a) If $\alpha \in R$, then $F[\alpha] \subset R$. But $F[\alpha]$ is a field because α is algebraic (see p. 135), and so R contains α^{-1}.

(b) By assumption, every $\alpha \in L$ is a root of a monic polynomial

$$X^m + a_{m-1}X^{m-1} + \cdots + a_0 \in E[X].$$

Each of the extensions

$$F[a_0, \ldots, a_{m-1}, \alpha] \supset F[a_0, \ldots, a_{m-1}] \supset F[a_0, \ldots, a_{m-2}] \supset \cdots \supset F$$

is generated by a single algebraic element, and so is finite. Therefore $F[a_0, \ldots, a_{m-1}, \alpha]$ is finite over F (see 4.2.6), which implies that α is algebraic over F. □

Example 4.3.6.

$$\mathbb{Q} \subset \underbrace{\mathbb{Q}\left(2^{1/2}, 2^{1/3}, 2^{1/4}, \cdots, 2^{1/n}, \cdots\right)}_{E} \subset \mathbb{R}$$

Note that

$$\mathbb{Q} \subset \mathbb{Q}(2^{1/2}) \subset \mathbb{Q}(2^{1/2}, 2^{1/3}) \subset \mathbb{Q}(2^{1/2}, 2^{1/3}, 2^{1/4}) \subset \cdots \subset \mathbb{R}$$
$$E = \bigcup_n \mathbb{Q}(2^{1/2}, 2^{1/3}, \cdots, 2^{1/n}) \underbrace{\subset}_{\text{subfield}} \mathbb{R}$$

We claim that E is algebraic over $\mathbb{Q}$ but $[E : \mathbb{Q}] = \infty$.

- $\alpha \in E$: Then $\exists n$ s.t. $\alpha \in \mathbb{Q}(2^{1/2}, \cdots, 2^{1/n})$. $2^{1/m}$ is algebraic over $\mathbb{Q}$: $(2^{1/m})^m - 2 = 0$, so $x^m - 2$ vanishes at $2^{1/m}$. By Lemma 4.3.5, we see $[\mathbb{Q}(2^{1/2}, \cdots, 2^{1/n} : \mathbb{Q})] < \infty$ and α is algebraic over $\mathbb{Q}$.
- $[E : \mathbb{Q}] = \infty$: suppose to the contrary $[E : \mathbb{Q}] = r < \infty$. Now look at $\alpha = \frac{1}{r+1}$. Then $f(\alpha) = 0$ where

$$f(x) = \underbrace{x^{r+1} - 2}_{\text{irreducible by Eisenstein}} \in \mathbb{Q}[x]$$

  Thus the degree of minimal polynomial of α is $r + 1$. Then $[\mathbb{Q}(2^{1/2}, \cdots, 2^{\frac{1}{r+1}}) : \mathbb{Q}] \geqslant r + 1$. Contradiction.

4.3.1 Applications

See [4] sections "transcendental numbers" and "constructions with straight-edge and compass" for some interesting discussions.

4.4 Algebraically Closed Fields

Let F be a field. A polynomial is said to **split** in $F[X]$ if it is a product of polynomials of degree at most 1 in $F[X]$.

Proposition 4.4.1. For a field Ω, the following statements are equivalent:

(a) Every nonconstant polynomial in $\Omega[X]$ splits in $\Omega[X]$.

(b) Every nonconstant polynomial in $\Omega[X]$ has at least one root in Ω.

(c) The irreducible polynomials in $\Omega[X]$ are those of degree 1.

(d) Every field of finite degree over Ω equals Ω.

Proof. The implications (a)⇒(b)⇒(c) are obvious.

(c)⇒(a). This follows from the fact that $\Omega[X]$ is a unique factorization domain.

(c)⇒(d). Let E be a finite extension of Ω, and let $\alpha \in E$. The minimal polynomial of α, being irreducible, has degree 1 by (c), and, being monic by definition of min poly, thus has the form $f(X) = X + a$ with $a \in \Omega$. Then $f(\alpha) = 0 \Rightarrow a = -\alpha \in \Omega$, so $\alpha \in \Omega$.

(d)⇒(c). Let f be an irreducible polynomial in $\Omega[X]$. Then $\Omega[X]/(f)$ is an extension of Ω of degree $\deg(f)$ (see 4.3.4), and so $\deg(f) = 1$. □

Definition 4.4.2.
(a) A field Ω is **algebraically closed** if it satisfies the equivalent statements of Proposition 4.4.1.

(b) A field Ω is an **algebraic closure** of a subfield F if it is algebraically closed and algebraic over F.

Example 4.4.3. For example, the fundamental theorem of algebra says that $\mathbb{C}$ is algebraically closed (by characterization (b)). It is an algebraic closure of $\mathbb{R}$.

Proposition 4.4.4. If Ω is algebraic over F and every polynomial $f \in F[X]$ splits in $\Omega[X]$, then Ω is algebraically closed (hence an algebraic closure of F).

Proof. Let f be a nonconstant polynomial in $\Omega[X]$. We have to show that f has a root in Ω. We know (see 4.2.7) that f has a root α in some finite extension Ω' of Ω. Set

$$f = a_n X^n + \cdots + a_0, \quad a_i \in \Omega,$$

and consider the fields

$$F \subset F[a_0, \dots, a_n] \subset F[a_0, \dots, a_n, \alpha].$$

Each extension generated by a finite set of algebraic elements, and hence is finite (4.3.4). Therefore α lies in a finite extension of F (see 4.2.6), and so is algebraic over F (see 4.2.4) — it is a root of a polynomial g with coefficients in F. By assumption, g splits in $\Omega[X]$, and so the roots of g in Ω' all lie in Ω. In particular, $\alpha \in \Omega$. □

Proposition 4.4.5. Let $\Omega \supset F$; then

$$\{\alpha \in \Omega \mid \alpha \text{ algebraic over } F\}$$

is a field.

Proof. If α and β are algebraic over F, then $F[\alpha, \beta]$ is a field (see 4.3.5) of finite degree over F (see 4.3.4). Thus, every element of $F[\alpha, \beta]$ is algebraic over F (see 4.2.4). In particular, $\alpha \pm \beta$, α/β, and $\alpha\beta$ are algebraic over F. □

The field constructed in the proposition is called the **algebraic closure of F in Ω**.

Corollary 4.4.6. Let Ω be an algebraically closed field. For any subfield F of Ω, the algebraic closure E of F in Ω is an algebraic closure of F.

Proof. It is algebraic over F by definition. Every polynomial in $F[X]$ splits in $\Omega[X]$ and has its roots in E, and so splits in $E[X]$. Now apply Proposition 4.4.4. □

Thus, when we admit the fundamental theorem of algebra, every subfield of $\mathbb{C}$ has an algebraic closure (in fact, a canonical algebraic closure).

Theorem 4.4.7. Every field F has an algebraic closure.

Proof. (Emil Artin.) Consider the polynomial ring $F[\dots, x_f, \dots]$ in a family of symbols x_f indexed by the nonconstant monic polynomials $f \in F[X]$. If 1 lies in the ideal I of $F[\dots, x_f, \dots]$ generated by the polynomials $f(x_f)$, then

$$g_1 f_1(x_{f_1}) + \cdots + g_n f_n(x_{f_n}) = 1 \qquad (\text{in } F[\dots, x_f, \dots])$$

for some $g_i \in F[\dots, x_f, \dots]$ and some nonconstant monic $f_i \in F[X]$. Let E be an extension of F such that each f_i, $i = 1, \dots, n$, has a root α_i in E. Under the F-homomorphism $F[\dots, x_f, \dots] \to E$ sending

$$\begin{cases} x_{f_i} \mapsto \alpha_i \\ x_f \mapsto 0, \quad f \notin \{f_1, \dots, f_n\} \end{cases}$$

the above relation becomes $0 = 1$. From this contradiction, we deduce that 1 does not lie in I, and so corollary 2.3.11 applied to $F[\dots, x_f, \dots]/I$ shows that I is contained in a maximal ideal M of $F[\dots, x_f, \dots]$. Let $\Omega = F[\dots, x_f, \dots]/M$. Then Ω is a field containing (a copy of) F in which every nonconstant polynomial in $F[X]$ has at least one root. Repeat the process starting with E_1 instead of F to obtain a field E_2. Continue in this fashion to obtain a sequence of fields

$$F = E_0 \subset E_1 \subset E_2 \subset \cdots,$$

and let $E = \bigcup_i E_i$. Then E is algebraically closed because the coefficients of any nonconstant polynomial g in $E[X]$ lie in E_i for some i, and so g has a root in E_{i+1}. Therefore, the algebraic closure of F in E is an algebraic closure of F (4.4.6). □

4.5 Homomorphisms from simple extensions.

Let F be a field, and let E and E' be fields containing F. Recall that an F-homomorphism is a homomorphism $\varphi\colon E \to E'$ such that $\varphi(a) = a$ for all $a \in F$. Thus an F-homomorphism φ maps a polynomial

$$\sum a_{i_1 \cdots i_m} \alpha_1^{i_1} \cdots \alpha_m^{i_m}, \quad a_{i_1 \cdots i_m} \in F, \quad \alpha_i \in E,$$

to

$$\sum a_{i_1 \cdots i_m} \varphi(\alpha_1)^{i_1} \cdots \varphi(\alpha_m)^{i_m}.$$

An F-**isomorphism** is a bijective F-homomorphism.

An F-homomorphism $E \to E'$ of fields is, in particular, an injective F-linear map of F-vector spaces, and so it is an F-isomorphism if E and E' have the same finite degree over F.

Proposition 4.5.1. Let $F(\alpha)$ be a simple extension of F and Ω a second extension of F.

1. Let α be transcendental over F. For every F-homomorphism $\varphi\colon F(\alpha) \to \Omega$, $\varphi(\alpha)$ is transcendental over F, and the map $\varphi \mapsto \varphi(\alpha)$ defines a one-to-one correspondence

$$\{F\text{-homomorphisms } F(\alpha) \to \Omega\} \leftrightarrow \{\text{elements of } \Omega \text{ transcendental over } F\}.$$

2. Let α be algebraic over F with minimal polynomial $f(X)$. For every F-homomorphism $\varphi\colon F[\alpha] \to \Omega$, $\varphi(\alpha)$ is a root of $f(X)$ in Ω, and the map $\varphi \mapsto \varphi(\alpha)$ defines a one-to-one correspondence

$$\{F\text{-homomorphisms } \varphi\colon F[\alpha] \to \Omega\} \leftrightarrow \{\text{roots of } f \text{ in } \Omega\}.$$

   In particular, the number of such maps is the number of distinct roots of f in Ω.

Proof. (a) To say that α is transcendental over F means that $F[\alpha]$ is isomorphic to the polynomial ring in the symbol α. Therefore, for every $\gamma \in \Omega$, there is a unique F-homomorphism $\varphi\colon F[\alpha] \to \Omega$ such that $\varphi(\alpha) = \gamma$ (see **??**). This φ extends (uniquely) to the field of fractions $F(\alpha)$ of $F[\alpha]$ if and only if nonzero elements of $F[\alpha]$ are sent to nonzero elements of Ω, which is the case if and only if γ is transcendental over F. Thus we see that there are one-to-one correspondences between (a) the F-homomorphisms $F(\alpha) \to \Omega$, (b) the F-homomorphisms $\varphi\colon F[\alpha] \to \Omega$ such that $\varphi(\alpha)$ is transcendental, (c) the transcendental elements of Ω.

(b) Let $f(X) = \sum a_i X^i$, and consider an F-homomorphism $\varphi\colon F[\alpha] \to \Omega$. On applying φ to the equality $\sum a_i \alpha^i = 0$, we obtain the equality $\sum a_i \varphi(\alpha)^i = 0$, which shows that $\varphi(\alpha)$ is a root of $f(X)$ in Ω. Conversely, if $\gamma \in \Omega$ is a root of $f(X)$, then the map $F[X] \to \Omega$, $g(X) \mapsto g(\gamma)$, factors through $F[X]/(f(X))$. When composed with the inverse of the canonical isomorphism $F[X]/(f(X)) \to F[\alpha]$, this becomes a homomorphism $F[\alpha] \to \Omega$ sending α to γ. □

We shall need a slight generalization of this result.

Proposition 4.5.2. Let $F(\alpha)$ be a simple extension of F and $\varphi_0\colon F \to \Omega$ a homomorphism from F into a second field Ω.

1. If α is transcendental over F, then the map $\varphi \mapsto \varphi(\alpha)$ defines a one-to-one correspondence
$$\{\text{extensions } \varphi\colon F(\alpha) \to \Omega \text{ of } \varphi_0\} \leftrightarrow \{\text{elements of } \Omega \text{ transcendental over } \varphi_0(F)\}.$$
2. If α is algebraic over F, with minimal polynomial $f(X)$, then the map $\varphi \mapsto \varphi(\alpha)$ defines a one-to-one correspondence
$$\{\text{extensions } \varphi\colon F[\alpha] \to \Omega \text{ of } \varphi_0\} \leftrightarrow \{\text{roots of } \varphi_0 f \text{ in } \Omega\}.$$
In particular, the number of such maps is the number of distinct roots of $\varphi_0 f$ in Ω.

By $\varphi_0 f$ we mean the polynomial obtained by applying φ_0 to the coefficients of f. By an extension of φ_0 to $F(\alpha)$ we mean a homomorphism $\varphi\colon F(\alpha) \to \Omega$ whose restriction to F is φ_0. The proof of the proposition is essentially the same as that of the preceding proposition (indeed, it is essentially the same proposition).

4.6 Splitting Fields

Let f be a polynomial with coefficients in F. A field E containing F is said to **split** f if f splits in $E[X]$, i.e.,
$$f(X) = a\prod\nolimits_{i=1}^{m}(X - \alpha_i) \text{ with all } \alpha_i \in E.$$
If E splits f and is generated by the roots of f,
$$E = F[\alpha_1, \dots, \alpha_m],$$
then it is called a **splitting** or **root field** for f.

Note that $\prod f_i(X)^{m_i}$ $(m_i \geqslant 1)$ and $\prod f_i(X)$ have the same splitting fields. Note also that f splits in E if it has $\deg(f) - 1$ roots in E because the sum of the roots of f lies in F (if $f = aX^m + a_1X^{m-1} + \cdots$, then Vieta's formula gives $\sum \alpha_i = -a_1/a$).

Example 4.6.1. (a) Let $f(X) = aX^2 + bX + c \in \mathbb{Q}[X]$, and let $\alpha = \sqrt{b^2 - 4ac}$. The subfield $\mathbb{Q}[\alpha]$ of $\mathbb{C}$ is a splitting field for f.

(b) Let $f(X) = X^3 + aX^2 + bX + c \in \mathbb{Q}[X]$ be irreducible, and let $\alpha_1, \alpha_2, \alpha_3$ be its roots in $\mathbb{C}$. Then $\mathbb{Q}[\alpha_1, \alpha_2, \alpha_3] = \mathbb{Q}[\alpha_1, \alpha_2]$ is a splitting field for $f(X)$. Note that $[\mathbb{Q}[\alpha_1]\colon \mathbb{Q}] = 3$ and that $[\mathbb{Q}[\alpha_1, \alpha_2]\colon \mathbb{Q}[\alpha_1]] = 1$ or 2, and so $[\mathbb{Q}[\alpha_1, \alpha_2]\colon \mathbb{Q}] = 3$ or 6. We'll see later that the degree is 3 if and only if the discriminant of $f(X)$ is a square in $\mathbb{Q}$. For example, the discriminant of $X^3 + bX + c$ is $-4b^3 - 27c^2$, and so the splitting field of $X^3 + 10X + 1$ (discriminant -4027) has degree 6 over $\mathbb{Q}$.

Proposition 4.6.2. Every polynomial $f \in F[X]$ has a splitting field E_f, and

$$[E_f : F] \leqslant (\deg f)! \quad (\text{factorial } \deg f).$$

Proof. Let $F_1 = F[\alpha_1]$ be a stem field for some monic irreducible factor of f in $F[X]$. Then $f(\alpha_1) = 0$, and we let $F_2 = F_1[\alpha_2]$ be a stem field for some monic irreducible factor of $f(X)/(X - \alpha_1)$ in $F_1[X]$. Continuing in this fashion, we arrive at a splitting field E_f. Let $n = \deg f$. Then $[F_1 : F] = \deg g_1 \leqslant n$, $[F_2 : F_1] \leqslant n - 1, ...$, and so $[E_f : F] \leqslant n!$. □

Remark 4.6.3. Let F be a field. For a given integer n, there may or may not exist polynomials of degree n in $F[X]$ whose splitting field has degree $n!$ — this depends on F.

Example 4.6.4. (a) Let $f(X) = (X^p - 1)/(X - 1) \in \mathbb{Q}[X]$, p prime. If ζ is one root of f, then the remaining roots are $\zeta^2, \zeta^3, \ldots, \zeta^{p-1}$, and so the splitting field of f is $\mathbb{Q}[\zeta]$.

(b) Let F have characteristic $p \neq 0$, and let $f = X^p - X - a \in F[X]$. If α is one root of f in some extension of F, then the remaining roots are $\alpha + 1, ..., \alpha + p - 1$, and so the splitting field of f is $F[\alpha]$.

(c) If α is one root of $X^n - a$, then the remaining roots are all of the form $\zeta\alpha$, where $\zeta^n = 1$. Therefore, $F[\alpha]$ is a splitting field for $X^n - a$ if and only if F contains all the nth roots of 1 (by which we mean that $X^n - 1$ splits in $F[X]$). Note that if p is the characteristic of F, then $X^p - 1 = (X - 1)^p$, and so F automatically contains all the pth roots of 1.

Proposition 4.6.5. Let $f \in F[X]$. Let E be an extension of F generated by the roots of f in E, and let Ω be an extension of F splitting f.

1. There exists an F-homomorphism $\varphi : E \to \Omega$; the number of such homomorphisms is at most $[E : F]$, and equals $[E : F]$ if f has distinct roots in Ω.
2. If E and Ω are both splitting fields for f, then every F-homomorphism $E \to \Omega$ is an isomorphism. In particular, any two splitting fields for f are F-isomorphic.

As f splits in $\Omega[X]$, $f(X) = a\prod_{i=1}^{\deg(f)}(X - \beta_i)$ with $\beta_1, \beta_2, \ldots \in \Omega$. To say that f has distinct roots in Ω means that $\beta_i \neq \beta_j$ if $i \neq j$.

Proof. We may suppose that f is monic.

We begin with an observation: let F, f, and Ω be as in the statement of the proposition, let L be a subfield of Ω containing F, and let g be a monic factor of f in $L[X]$; as g divides f in $\Omega[X]$, it is a product of certain number of the factors $X - \beta_i$ of f in $\Omega[X]$; in particular, we see that g splits in Ω, and that it has distinct roots in Ω if f does..

(a) By hypothesis, $E = F[\alpha_1, ..., \alpha_m]$ with each α_i a root of $f(X)$ in E. The minimal polynomial of α_1 is an irreducible polynomial f_1 dividing f. From the initial observation with $L = F$, we see that f_1 splits in Ω, and that its roots are distinct if the roots of f are distinct. According to Proposition 4.5.1, there exists an F-homomorphism $\varphi_1 : F[\alpha_1] \to \Omega$, and the number of such homomorphisms is at most $[F[\alpha_1] : F]$, with equality holding when f has distinct roots in Ω.

The minimal polynomial of α_2 over $F[\alpha_1]$ is an irreducible factor f_2 of f in $F[\alpha_1][X]$. On applying the initial observation with $L = \varphi_1 F[\alpha_1]$ and $g = \varphi_1 f_2$, we see that $\varphi_1 f_2$ splits in Ω, and that its roots are distinct if the roots of f are distinct. According to Proposition 4.5.2, each φ_1 extends to a homomorphism $\varphi_2 : F[\alpha_1, \alpha_2] \to \Omega$, and the number of extensions is at most $[F[\alpha_1, \alpha_2] : F[\alpha_1]]$, with equality holding when f has distinct roots in Ω.

On combining these statements we conclude that there exists an F-homomorphism

$$\varphi : F[\alpha_1, \alpha_2] \to \Omega,$$

and that the number of such homomorphisms is at most $[F[\alpha_1, \alpha_2]\colon F]$, with equality holding if f has distinct roots in Ω.

After repeating the argument m times, we obtain (a).

(b) Every F-homomorphism $E \to \Omega$ is injective, and so, if there exists such a homomorphism, then $[E\colon F] \leqslant [\Omega\colon F]$. If E and Ω are both splitting fields for f, then (a) shows that there exist homomorphisms $E \leftrightarrows \Omega$, and so $[E\colon F] = [\Omega\colon F]$. It follows that every F-homomorphism $E \to \Omega$ is an F-isomorphism. □

Corollary 4.6.6. Let E and L be extensions of F, with E finite over F.

1. The number of F-homomorphisms $E \to L$ is at most $[E\colon F]$.
2. There exists a finite extension Ω/L and an F-homomorphism $E \to \Omega$.

Proof. Write $E = F[\alpha_1, \dots, \alpha_m]$, and let $f \in F[X]$ be the product of the minimal polynomials of the α_i; thus E is generated over F by roots of f. Let Ω be a splitting field for f regarded as an element of $L[X]$. The proposition shows that there exists an F-homomorphism $E \to \Omega$, and the number of such homomorphisms is $\leqslant [E\colon F]$. This proves (b), and since an F-homomorphism $E \to L$ can be regarded as an F-homomorphism $E \to \Omega$, it also proves (a). □

Remark 4.6.7. (a) Let $E_1, E_2, \dots, E_m$ be finite extensions of F, and let L be an extension of F. From the corollary we see that there exists a finite extension L_1/L such that L_1 contains an isomorphic image of E_1; then that there exists a finite extension L_2/L_1 such that L_2 contains an isomorphic image of E_2. On continuing in this fashion, we find that there exists a finite extension Ω/L such that Ω contains an isomorphic copy of every E_i.

(b) Let $f \in F[X]$. If E and E' are both splitting fields of f, then we know there exists an F-isomorphism $E \to E'$, but there will in general be no *preferred* such isomorphism. Error and confusion can result if the fields are simply identified. Also, it makes no sense to speak of "the field $F[\alpha]$ generated by a root of f" unless f is irreducible (the fields generated by the roots of two different factors are unrelated). Even when f is irreducible, it makes no sense to speak of "the field $F[\alpha, \beta]$ generated by two roots α, β of f" (the extensions of $F[\alpha]$ generated by the roots of two different factors of f in $F[\alpha][X]$ may be very different).

4.7 Multiple roots

Even when polynomials in $F[X]$ have no common factor in $F[X]$, one might expect that they could acquire a common factor in $\Omega[X]$ for some $\Omega \supset F$. In fact, this doesn't happen — greatest common divisors don't change when the field is extended.

Proposition 4.7.1. Let f and g be polynomials in $F[X]$, and let Ω be an extension of F. If $r(X)$ is the gcd of f and g computed in $F[X]$, then it is also the gcd of f and g in $\Omega[X]$. In particular, distinct monic irreducible polynomials in $F[X]$ do not acquire a common root in any extension of F.

Proof. Let $r_F(X)$ and $r_\Omega(X)$ be the greatest common divisors of f and g in $F[X]$ and $\Omega[X]$ respectively. Certainly $r_F(X) | r_\Omega(X)$ in $\Omega[X]$, but Euclid's algorithm shows that there are polynomials a and b in $F[X]$ such that

$$a(X)f(X) + b(X)g(X) = r_F(X),$$

and so $r_\Omega(X)$ divides $r_F(X)$ in $\Omega[X]$.

For the second statement, note that the hypotheses imply that $\gcd(f, g) = 1$ (in $F[X]$), and so f and g can't acquire a common factor in any extension field. □

The proposition allows us to speak of the greatest common divisor of f and g without reference to a field.

Let $f \in F[X]$. Then f splits into linear factors

$$f(X) = a\prod_{i=1}^{r}(X-\alpha_i)^{m_i},\ \alpha_i \text{ distinct}, m_i \geqslant 1,\ \sum_{i=1}^{r} m_i = \deg(f), \tag{4.3}$$

in $E[X]$ for some extension E of F (see 4.6.2). We say that α_i is a root of f of **multiplicity** m_i in E. If $m_i > 1$, then α_i is said to be a **multiple root** of f, and otherwise it is a **simple root**.

I claim that the unordered sequence of integers $m_1, \dots, m_r$ in (4.3) is independent of the extension E chosen to split f. Certainly, it is unchanged when E is replaced with its subfield $F[\alpha_1, \dots, \alpha_r]$, and so we may suppose that E is a splitting field for f. Let E and E' be splitting fields for F, and suppose that $f(X) = a\prod_{i=1}^{r}(X-\alpha_i)^{m_i}$ in $E[X]$ and $f(X) = a'\prod_{i=1}^{r'}(X-\alpha_i')^{m_i'}$ in $E'[X]$. Let $\varphi\colon E \to E'$ be an F-isomorphism, which exists by (4.6.5b), and extend it to an isomorphism $E[X] \to E'[X]$ by sending X to X. Then φ maps the factorization of f in $E[X]$ onto a factorization

$$f(X) = \varphi(a)\prod_{i=1}^{r}(X-\varphi(\alpha_i))^{m_i}$$

in $E'[X]$. By unique factorization, this coincides with the earlier factorization in $E'[X]$ up to a renumbering of the α_i. Therefore $r = r'$, and

$$\{m_1, \dots, m_r\} = \{m_1', \dots, m_r'\}.$$

We say that f **has a multiple root** when at least one of the $m_i > 1$, and that f has **only simple roots** when all $m_i = 1$. Thus "f has a multiple root" means "f has a multiple root in one, hence every, extension of F splitting f", and similarly for "f has only simple roots".

We wish to determine when a polynomial has a multiple root. If f has a multiple factor in $F[X]$, say $f = \prod f_i(X)^{m_i}$ with some $m_i > 1$, then obviously it will have a multiple root. If $f = \prod f_i$ with the f_i distinct monic irreducible polynomials, then Proposition 4.7.1 shows that f has a multiple root if and only if at least one of the f_i has a multiple root. Thus, it suffices to determine when an *irreducible* polynomial has a multiple root.

Example 4.7.2. Let F be of characteristic $p \neq 0$, and assume that F contains an element a that is not a pth-power, for example, $a = T$ in the field $\mathbb{F}_p(T)$. Then $X^p - a$ is irreducible in $F[X]$, but by 4.1.2 we have $X^p - a = (X-\alpha)^p$ in its splitting field. Thus an irreducible polynomial can have multiple roots.

The derivative of a polynomial $f(X) = \sum a_i X^i$ is defined to be $f'(X) = \sum i a_i X^{i-1}$. The usual rules for differentiating sums and products still hold, but note that in characteristic p the derivative of X^p is zero.

Proposition 4.7.3. For a nonconstant irreducible polynomial f in $F[X]$, the following statements are equivalent:

1. f has a multiple root;
2. $\gcd(f, f') \neq 1$;
3. F has nonzero characteristic p and f is a polynomial in X^p, i.e., of the form
$$f(X) = a_n\,(X^p)^n + a_{n-1}\,(X^p)^{n-1} + \cdots + a_1 X^p + a_0$$
4. all the roots of f are multiple.

Proof. (a) $\Rightarrow$ (b). Let α be a multiple root of f, and write $f = (X-\alpha)^m g(X)$, $m > 1$, in some extension field. Then

$$f'(X) = m(X-\alpha)^{m-1}g(X) + (X-\alpha)^m g'(X). \tag{4.4}$$

Hence f and f' have $X-\alpha$ as a common factor.

(b) $\Rightarrow$ (c). As f is irreducible and $\deg(f') < \deg(f)$,

$$\gcd(f, f') \neq 1 \implies f' = 0.$$

Let $f = a_0 + \cdots + a_d X^d$, $d \geqslant 1$. Then $f' = a_1 + \cdots + i a_i X^{i-1} + \cdots + d a_d X^{d-1}$, which is the zero polynomial if only if F has characteristic $p \neq 0$ and $a_i = 0$ for all i not divisible by p.

(c) $\Rightarrow$ (d). By hypothesis, $f(X) = g(X^p)$ with $g(X) \in F[X]$. Let $g(X) = \prod_i (X-a_i)^{m_i}$ in some extension field. Then each a_i becomes a pth power, say, $a_i = \alpha_i^p$, in some possibly larger extension field. Now

$$f(X) = g(X^p) = \prod\nolimits_i (X^p - a_i)^{m_i} = \prod\nolimits_i (X-\alpha_i)^{pm_i}$$

which shows that every root of $f(X)$ has multiplicity at least p.

(d) $\Rightarrow$ (a). Obvious. □

Proposition 4.7.4. The following conditions on a nonzero polynomial $f \in F[X]$ are equivalent:

1. $\gcd(f, f') = 1$ in $F[X]$;
2. f has only simple roots.

Proof. Let Ω be an extension of F splitting f. From (4.4), p. 142, we see that a root α of f in Ω is multiple if and only if it is also a root of f'.

If $\gcd(f, f') = 1$, then f and f' have no common factor in $\Omega[X]$ (see 4.7.1). In particular, they have no common root, and so f has only simple roots.

If f has only simple roots, then $\gcd(f, f')$ must be the constant polynomial, because otherwise it would have a root in Ω which would then be a common root of f and f'. □

Definition 4.7.5. A polynomial is **separable** if it is nonzero and satisfies the equivalent conditions on (4.7.4).[5]

Remark 4.7.6. Thus a nonconstant irreducible polynomial f is not separable if and only if F has characteristic $p \neq 0$ and f is a polynomial in X^p (see 4.7.3). Let $f = \prod f_i$ with f and the f_i monic and the f_i irreducible; then f is separable if and only if the f_i are distinct and separable. If f is separable as a polynomial in $F[X]$, then it is separable as a polynomial in $E[X]$ for every extension E of F.

Definition 4.7.7. A field F is **perfect** if it has characteristic zero or it has characteristic p and every element of F is a p-th power.

Thus, F is perfect if and only if $F = F^q$, where q is the characteristic exponent of F.

Proposition 4.7.8. A field F is perfect if and only if every irreducible polynomial in $F[X]$ is separable.

Proof. If F has characteristic zero, the statement is obvious, and so we may suppose F has characteristic $p \neq 0$. If F contains an element a that is not a pth power, then $X^p - a$ is irreducible in $F[X]$ but not separable (see 4.7.2). Conversely, if every element of F is a pth power, then every polynomial in X^p with coefficients in F is a pth power in $F[X]$,

$$\textstyle\sum a_i X^{ip} = \left(\sum b_i X^i\right)^p \quad \text{if} \quad a_i = b_i^p,$$

and so it is not irreducible. □

[5] This is Bourbaki's definition. Often (e.g., in the books of Jacobson and in earlier versions of these notes) a polynomial f is said to be separable if each of its irreducible factors has only simple roots.

Example 4.7.9. 1. A finite field F is perfect, because the Frobenius endomorphism $a \mapsto a^p \colon F \to F$ is injective and therefore surjective (by counting).

2. A field that can be written as a union of perfect fields is perfect. Therefore, every field algebraic over $\mathbb{F}_p$ is perfect.
3. Every algebraically closed field is perfect.
4. If F_0 has characteristic $p \neq 0$, then $F = F_0(X)$ is not perfect, because X is not a pth power.

Remark 4.7.10. Let F be a perfect field. We'll see later that every finite extension E/F is simple, i.e., $E = F[\alpha]$ with α a root of a (separable) polynomial $f \in F[X]$ of degree $[E\colon F]$. Thus it follows directly from (4.5.2b) that, for any extension Ω of F, the number of F-homomorphisms $E \to \Omega$ is $\leqslant [E\colon F]$, with equality if and only if f splits in Ω. We can't use this argument here because it would make the exposition circular.

4.1 EXERCISES

1. Let F be a field of characteristic $\neq 2$.

1. Let E be quadratic extension of F; show that
$$S(E) = \{a \in F^\times \mid a \text{ is a square in } E\}$$
is a subgroup of $F^\times$ containing $F^{\times 2}$.
2. Let E and E' be quadratic extensions of F; show that there exists an F-isomorphism $\varphi\colon E \to E'$ if and only if $S(E) = S(E')$.
3. Show that there is an infinite sequence of fields $E_1, E_2, \ldots$ with E_i a quadratic extension of $\mathbb{Q}$ such that E_i is not isomorphic to E_j for $i \neq j$.
4. Let p be an odd prime. Show that, up to isomorphism, there is exactly one field with p^2 elements.

2. (a) Let F be a field of characteristic p. Show that if $X^p - X - a$ is reducible in $F[X]$, then it splits into distinct factors in $F[X]$.

(b) For every prime p, show that $X^p - X - 1$ is irreducible in $\mathbb{Q}[X]$.

3. Construct a splitting field for $X^5 - 2$ over $\mathbb{Q}$. What is its degree over $\mathbb{Q}$?

4. Find a splitting field of $X^{p^m} - 1 \in \mathbb{F}_p[X]$. What is its degree over $\mathbb{F}_p$?

5. Let $f \in F[X]$, where F is a field of characteristic 0. Let $d(X) = \gcd(f, f')$. Show that $g(X) = f(X)d(X)^{-1}$ has the same roots as $f(X)$, and these are all simple roots of $g(X)$.

6. Let $f(X)$ be an irreducible polynomial in $F[X]$, where F has characteristic p. Show that $f(X)$ can be written $f(X) = g(X^{p^e})$ where $g(X)$ is irreducible and separable. Deduce that every root of $f(X)$ has the same multiplicity p^e in any splitting field.

Chapter 5

Galois Theory

In this chapter, we prove the fundamental theorem of Galois theory, which classifies the subfields of the splitting field of a separable polynomial f in terms of the Galois group of f. We also investigate general methods for computing Galois groups.

5.1 Groups of Automorphisms of Fields

Consider fields $E \supset F$. An F-isomorphism $E \to E$ is called an F**-automorphism** of E. The F-automorphisms of E form a group, which we denote $\mathrm{Aut}(E/F)$.

Claim 5.1.1. If $F \subset E$ and $f(X) \in F[X]$ and $\alpha \in E$ is a root of f, then $\phi \in \mathrm{Aut}(E/F)$ sends α to a root of $f(X)$, becasue

$$\begin{aligned} f(X) = a_n X^n + \cdots + aX + a_0 &\Rightarrow a_n\alpha^n + \cdots + a_1\alpha + a_0 = 0 \\ &\Rightarrow \phi\left(a_n\alpha^n + \cdots + a_\alpha\alpha + a\right) = 0 \\ &\Rightarrow a_n\phi(\alpha)^n + \cdots + a_1\phi(\alpha) + a_0 = 0. \end{aligned}$$

Example 5.1.2.

- $\mathbb{R} \subset \mathbb{C}$. $i \in \mathbb{C}$ is a root of $x^+1 \in \mathbb{R}[X]$. Let $\phi \in \mathrm{Aut}(\mathbb{C}/\mathbb{R})$. Then $\phi(i)$ also a root of x^2+1, so $\phi(i) = \pm i$. If $\phi(i) = i$, then it is the identity; if $\phi(i) = -i$, then ϕ is conjugator, i.e., $\phi(a+ib) = a - ib$.

  Therefore, $\mathrm{Aut}(\mathbb{C}/\mathbb{R}) \cong \mathbb{Z}_2$.
- $\mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2})$. We compute $\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$. α is a root of $x^3 - 2 \in \mathbb{Q}[X]$. Other roots are $\omega\alpha$ and $\omega^2\alpha$, where ω is the third root of unity. So $\omega\alpha$, $\omega^2\alpha$ are not in $\mathbb{Q}(\alpha)$. Any $\phi \in \mathrm{Aut}(\mathbb{Q}(\alpha)/\mathbb{Q})$ fixes α., so ϕ fixes $\mathbb{Q}(\alpha)$. So \$ fixes $Q(\alpha) \Rightarrow |\mathrm{Aut}(Q(\alpha)/Q)| = 1$

Example 5.1.3.
(a) There are two obvious automorphisms of $\mathbb{C}$, namely, the identity map and complex conjugation. We'll see later that by using the Axiom of Choice we can construct uncountably many more.

(b) Let $E = \mathbb{C}(X)$. A $\mathbb{C}$-automorphism of E sends X to another generator of E over $\mathbb{C}$. It follows from (**??**) below that these are exactly the elements $\frac{aX+b}{cX+d}$, $ad - bc \neq 0$. Therefore $\mathrm{Aut}(E/\mathbb{C})$ consists of the maps $f(X) \mapsto f\left(\frac{aX+b}{cX+d}\right)$, $ad - bc \neq 0$, and so

$$\mathrm{Aut}(E/\mathbb{C}) \simeq \mathrm{PGL}_2(\mathbb{C}),$$

the group of invertible 2×2 matrices with complex coefficients modulo its centre. Analysts will note that this is the same as the automorphism group of the Riemann sphere. Here is the explanation. The field E of meromorphic functions on the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$ consists of the rational functions in z, i.e., $E = \mathbb{C}(z) \simeq \mathbb{C}(X)$, and the natural map $\mathrm{Aut}(\mathbb{P}^1_{\mathbb{C}}) \to \mathrm{Aut}(E/\mathbb{C})$ is an isomorphism.

(c) The group $\mathrm{Aut}(\mathbb{C}(X_1, X_2)/\mathbb{C})$ is quite complicated — there is a map

$$\mathrm{PGL}_3(\mathbb{C}) = \mathrm{Aut}(\mathbb{P}^2_{\mathbb{C}}) \hookrightarrow \mathrm{Aut}(\mathbb{C}(X_1, X_2)/\mathbb{C}),$$

but this is very far from being surjective. When there are even more variables X, the group is not known. The group $\mathrm{Aut}(\mathbb{C}(X_1, \ldots, X_n)/\mathbb{C})$ is the group of birational automorphisms of projective n-space $\mathbb{P}^n_{\mathbb{C}}$, and is called the **Cremona group.** Its study is part of algebraic geometry (Wikipedia: Cremona group).

In this section, we'll be concerned with the groups $\mathrm{Aut}(E/F)$ when E is a finite extension of F.

Proposition 5.1.4. If E/F is a finite extension, then $|\mathrm{Aut}(E/F)| \leqslant [E : F]$.

Proof. Induction on $r = [E : F]$. We show if $\sigma : F \to F'$ is an F-isomorphism of fields. $F \subset E$, $F' \subset E'$ are field extansions with

$$[E : F] = \left[E' : F'\right] = r,$$

then there are $\leqslant r$ ways to extend σ to an isomorphism $\widetilde{\sigma} : E \to E'$. $r = 1$ case is trivial. We show $1, \cdots, r-1 \implies r$. Pick $\alpha \in E \backslash F$ and let $f(X) \in F(X)$ be the minimal polynomial of α. Let $g = \sigma(f) \in F'[X]$. Then any $\widetilde{\sigma} : E \to E'$ extending σ sends α to a root of g by observation 5.1.1.

$$\deg(g) = \deg(f) = [F(\alpha) : F] =: m$$

so there are $\leqslant m$ choices for $\sigma(\alpha)$. Fix such a choice β. Consider

$$\psi : F(\alpha) \to F'(\beta)$$

with $\psi(\alpha) = \beta$ and

$$F(\alpha) = \{a_{m-1}\alpha^{m-1} + \cdots + a_1\alpha + a_0 \mid a_i \in F\}$$

$$\psi(a_{m-1}\alpha^{m-1} + \cdots + a_1\alpha + a_0) = \sigma(a_{m-1})\beta^{m-1} + \cdots + \sigma(a_1)\beta + a_0$$

Then the extension $E \backslash F(\alpha)$ has degree r/m. By induction hypothesis, there are $\leqslant r/m$ ways to extend ψ to an isomomorphism $E \to E'$. $m \cdot (r/m) = r$. □

Proposition 5.1.5. Let E be a splitting field of a separable polynomial f in $F[X]$; then $|\mathrm{Aut}(E/F)| = [E\colon F]$.

Proof. As f is separable, it has $\deg f$ distinct roots in E. Therefore Proposition 4.6.5 shows that the number of F-homomorphisms $E \to E$ is $[E\colon F]$. Because E is finite over F, all such homomorphisms are isomorphisms. □

Example 5.1.6. Consider a simple extension $E = F[\alpha]$, and let f be a polynomial in $F[X]$ having α as a root. If α is the only root of f in E, then $\mathrm{Aut}(E/F) = 1$ by (4.5.1b). For example, if $\sqrt[3]{2}$ is the real cube root of 2, then $\mathrm{Aut}(\mathbb{Q}[\sqrt[3]{2}]/\mathbb{Q}) = 1$. As another example, let F be a field of characteristic $p \neq 0$, and let a be an element of F that is not a pth power. Let E be a splitting field of $f = X^p - a$. Then f has only one root in E (see 4.7.2), and so $\mathrm{Aut}(E/F) = 1$.

These examples show that, in the statement of the proposition, is necessary that E be a *splitting* field of a *separable* polynomial.

When G is a group of automorphisms of a field E, we set

$$E^G = \mathrm{Inv}(G) = \{\alpha \in E \mid \sigma\alpha = \alpha, \text{ all } \sigma \in G\}.$$

It is a subfield of E, called the subfield of G-**invariants** of E or the **fixed field** of G.

In this section, we'll show that, when E is the splitting field of a separable polynomial in $F[X]$ and $G = \mathrm{Aut}(E/F)$, then the maps

$$M \mapsto \mathrm{Aut}(E/M), \quad H \mapsto \mathrm{Inv}(H)$$

give a one-to-one correspondence between the set of intermediate fields M, $F \subset M \subset E$, and the set of subgroups H of G.

Facts: $M \subseteq E^{\mathrm{Aut}(E/M)}$; $H \leqslant \mathrm{Aut}(E/E^H)$.

Theorem 5.1.7 (E. Artin)**.** Let G be a finite group of automorphisms of a field E, then

$$[E\colon E^G] \leqslant (G\colon 1).$$

Proof. Let $F = E^G$, and let $G = \{\sigma_1, \dots, \sigma_m\}$ with σ_1 the identity map. It suffices to show that every set $\{\alpha_1, \dots, \alpha_n\}$ of elements of E with $n > m$ is linearly dependent over F. For such a set, consider the system of linear equations

$$\begin{aligned} \sigma_1(\alpha_1)X_1 + \cdots + \sigma_1(\alpha_n)X_n &= 0 \\ &\vdots \\ \sigma_m(\alpha_1)X_1 + \cdots + \sigma_m(\alpha_n)X_n &= 0 \end{aligned} \tag{5.1}$$

with coefficients in E. There are m equations and $n > m$ unknowns, and hence there are nontrivial solutions in E. We choose one $(c_1, \dots, c_n)$ having the fewest possible nonzero elements. After renumbering the α_i, we may suppose that $c_1 \neq 0$, and then, after multiplying by a scalar, that $c_1 \in F$. With these normalizations, we'll show that all $c_i \in F$, and so the first equation

$$\alpha_1 c_1 + \cdots + \alpha_n c_n = 0$$

(recall that σ_1 is the identity map) is a linear relation on the α_i.

If not all c_i are in F, then $\sigma_k(c_i) \neq c_i$ for some $k \neq 1$ and $i \neq 1$. On applying σ_k to the system of linear equations

$$\begin{aligned} \sigma_1(\alpha_1)c_1 + \cdots + \sigma_1(\alpha_n)c_n &= 0 \\ &\vdots \\ \sigma_m(\alpha_1)c_1 + \cdots + \sigma_m(\alpha_n)c_n &= 0 \end{aligned}$$

and using that $\{\sigma_k\sigma_1, \dots, \sigma_k\sigma_m\} = \{\sigma_1, \dots, \sigma_m\}$ (σ_k merely permutes the σ_i), we find that

$$(c_1, \sigma_k(c_2), \dots, \sigma_k(c_i), \dots)$$

is also a solution to the system of equations (5.1). On subtracting it from the first solution, we obtain a solution $(0, \dots, c_i - \sigma_k(c_i), \dots)$, which is nonzero (look at the ith entry), but has more zeros than the first solution (look at the first entry) — contradiction. □

Corollary 5.1.8. Let G be a finite group of automorphisms of a field E; then

$$G = \mathrm{Aut}(E/E^G).$$

Proof. As $G \subset \mathrm{Aut}(E/E^G)$, we have inequalities

$$[E\colon E^G] \overset{5.1.7}{\leqslant} (G\colon 1) \leqslant (\mathrm{Aut}(E/E^G)\colon 1) \overset{4.6.6a}{\leqslant} [E\colon E^G].$$

All the inequalities must be equalities, and so $G = \mathrm{Aut}(E/E^G)$. □

5.2 Separable, normal, and Galois extensions

Definition 5.2.1. An algebraic extension E/F is **separable** if the minimal polynomial of every element of E is separable; otherwise, it is **inseparable**.

Thus, an algebraic extension E/F is separable if every irreducible polynomial in $F[X]$ having at least one root in E is separable, and it is inseparable if

- F is nonperfect, and in particular has characteristic $p \neq 0$, *and*
- there is an element α of E whose minimal polynomial is of the form $g(X^p)$, $g \in F[X]$.

See 4.7.5 *et seq*. For example, the extension $\mathbb{F}_p(T)$ of $\mathbb{F}_p(T^p)$ is inseparable extension because T has minimal polynomial $X^p - T^p$.

Definition 5.2.2. An extension E/F is **normal**[1] if it is algebraic and the minimal polynomial of every element of E splits in $E[X]$.

In other words, an algebraic extension E/F is normal if and only if every irreducible polynomial $f \in F[X]$ having at least one root in E splits in $E[X]$.

Let f be a monic irreducible polynomial of degree m in $F[X]$, and let E be an algebraic extension of F. If f has a root in E, so that it is the minimal polynomial of an element of E, then

$$\left.\begin{array}{lcl} E/F \text{ separable} & \Longrightarrow & f \text{ has only simple roots} \\ E/F \text{ normal} & \Longrightarrow & f \text{ splits in } E \end{array}\right\} \Longrightarrow f \text{ has } m \text{ distinct roots in } E.$$

It follows that E/F is separable and normal if and only if the minimal polynomial of every element α of E has $[F[\alpha]\colon F]$ distinct roots in E.

Example 5.2.3. (a) The polynomial $X^3 - 2$ has one real root $\sqrt[3]{2}$ and two nonreal roots in $\mathbb{C}$. Therefore the extension $\mathbb{Q}[\sqrt[3]{2}]/\mathbb{Q}$ (which is separable) is not normal.

(b) The extension $\mathbb{F}_p(T)/\mathbb{F}_p(T^p)$ (which is normal) is not separable because the minimal polynomial of T is not separable.

Theorem 5.2.4. For an extension E/F, the following statements are equivalent:

1. E is the splitting field of a separable polynomial $f \in F[X]$;
2. E is finite over F and $F = E^{\mathrm{Aut}(E/F)}$;
3. $F = E^G$ for some finite group G of automorphisms of E;
4. E is normal, separable, and finite over F.

Proof. (a) $\Rightarrow$ (b). Certainly, E is finite over F. Let $F' = E^{\mathrm{Aut}(E/F)} \supset F$. We have to show that $F' = F$. Note that E is also the splitting field of f regarded as a polynomial with coefficients in F', and that f is still separable when it is regarded in this way. Hence

$$\left|\mathrm{Aut}(E/F')\right| \overset{5.1.5}{=} [E\colon F'] \leqslant [E\colon F] \overset{5.1.5}{=} \left|\mathrm{Aut}(E/F)\right|.$$

According to Corollary 5.1.8, $\mathrm{Aut}(E/F) = \mathrm{Aut}(E/F')$, and so $[E\colon F'] = [E\colon F]$ and $F' = F$.

(b) $\Rightarrow$ (c). Let $G = \mathrm{Aut}(E/F)$. We are given that $F = E^G$, and G is finite because E is finite over F (apply 4.6.6a).

(c) $\Rightarrow$ (d). According to Theorem 5.1.7, $[E\colon F] \leqslant (G\colon 1)$; in particular, E/F is finite. Let $\alpha \in E$, and let f be the minimal polynomial of α; we have to show that f splits into distinct factors in $E[X]$. Let

[1] Bourbaki says "quasi-galoisienne".

$\{\alpha_1 = \alpha, \alpha_2, ..., \alpha_m\}$ be the orbit of α under the action of G on E (so the α_i are distinct elements of E), and let

$$g(X) = \prod_{i=1}^{m}(X - \alpha_i) = X^m + a_1X^{m-1} + \cdots + a_m.$$

The coefficients a_j are symmetric polynomials in the α_i, and each $\sigma \in G$ permutes the α_i, and so $\sigma a_j = a_j$ for all j. Thus $g(X) \in F[X]$. As it is monic and $g(\alpha) = 0$, it is divisible by f (see the definition of minimal polynomial, p. 134). Let $\alpha_i = \sigma\alpha$; on applying σ to the equation $f(\alpha) = 0$ we find that $f(\alpha_i) = 0$. Therefore every α_i is a root of f, and so g divides f. Hence $f = g$, and we conclude that $f(X)$ splits into distinct factors in E.

(d) $\Rightarrow$ (a). Because E has finite degree over F, it is generated over F by a finite number of elements, say, $E = F[\alpha_1, ..., \alpha_m]$, $\alpha_i \in E$, α_i algebraic over F. Let f_i be the minimal polynomial of α_i over F, and let f be the product of the distinct f_i. Because E is normal over F, each f_i splits in E, and so E is the splitting field of f. Because E is separable over F, each f_i is separable, and so f is separable. □

Definition 5.2.5. An extension E/F of fields is **Galois** if it satisfies the equivalent conditions of (5.2.4). When E/F is Galois, $\mathrm{Aut}(E/F)$ is called the **Galois group** of E over F, and it is denoted by $\mathrm{Gal}(E/F)$.

Remark 5.2.6. (a) Let E be Galois over F with Galois group G, and let $\alpha \in E$. The elements $\alpha_1, \alpha_2, ..., \alpha_m$ of the orbit of α under G are called the **conjugates** of α. In the course of proving the theorem we showed that the minimal polynomial of α is $\prod(X - \alpha_i)$, i.e., the conjugates of α are exactly the roots of its minimal polynomial in E.

(b) Let G be a finite group of automorphisms of a field E, and let $F = E^G$. By definition, E is Galois over F. Moreover, $\mathrm{Gal}(E/F) = G$ (apply 5.1.8) and $[E\colon F] = |\mathrm{Gal}(E/F)|$ (apply 5.1.5).

Corollary 5.2.7. Every finite separable extension E of F is contained in a Galois extension.

Proof. Let $E = F[\alpha_1, ..., \alpha_m]$, and let f_i be the minimal polynomial of α_i over F. The product of the distinct f_i is a separable polynomial in $F[X]$ whose splitting field is a Galois extension of F containing E. □

Corollary 5.2.8. Let $E \supset M \supset F$; if E is Galois over F, then it is Galois over M.

Proof. We know E is the splitting field of some separable $f \in F[X]$; it is also the splitting field of f regarded as an element of $M[X]$. □

Remark 5.2.9. An element α of an algebraic extension of F is said to be **separable** over F if its minimal polynomial over F is separable. The proof of Corollary 5.2.7 shows that every finite extension generated by separable elements is separable. Therefore, the elements of an algebraic extension E of F that are separable over F form a subfield E_{sep} of E that is separable over F. When E is finite over F, we let $[E\colon F]_{\text{sep}} = [E_{\text{sep}}\colon F]$ and call it the **separable degree** of E over F.

An algebraic extension E is **purely inseparable** over F if the only elements of E separable over F are the elements of F. If E is a finite extension of F, then E is purely inseparable over E_{sep}. See Jacobson 1964, Chap. I, Section 10, for more on this topic.

Definition 5.2.10. An extension E of F is **cyclic** (resp. **abelian**, resp. **solvable**, etc.) if it is Galois with cyclic (resp. abelian, resp. solvable, etc.) Galois group.

5.3 The fundamental theorem of Galois theory

Let E be an extension of F. A **subextension** of E/F is an extension M/F with $M \subset E$, i.e., a field M with $F \subset M \subset E$. When E is Galois over F, the subextensions of E/F are in one-to-one correspondence with the subgroups of $\mathrm{Gal}(E/F)$. More precisely, there is the following statement.

Theorem 5.3.1 (Fundamental theorem of Galois theory)**.** Let E be a Galois extension of F with Galois group G. The map $H \mapsto E^H$ is a bijection from the set of subgroups of G to the set of subextensions of E/F,

$$\{\text{subgroups } H \text{ of } G\} \overset{1:1}{\leftrightarrow} \{\text{subextensions } F \subset M \subset E\},$$

with inverse $M \mapsto \mathrm{Gal}(E/M)$. Moreover,

1. the correspondence is inclusion-reversing: $H_1 \supset H_2 \iff E^{H_1} \subset E^{H_2}$;
2. indexes equal degrees: $(H_1 : H_2) = [E^{H_2} : E^{H_1}]$;
3. $\sigma H \sigma^{-1} \leftrightarrow \sigma M$, i.e., $E^{\sigma H \sigma^{-1}} = \sigma(E^H)$; $\mathrm{Gal}(E/\sigma M) = \sigma \,\mathrm{Gal}(E/M)\sigma^{-1}$.
4. H is normal in $G \iff E^H$ is normal (hence Galois) over F, in which case
$$\mathrm{Gal}(E^H/F) \simeq G/H.$$

Proof. For the first statement, we have to show that $H \mapsto E^H$ and $M \mapsto \mathrm{Gal}(E/M)$ are inverse maps. Let H be a subgroup of G. Then, Corollary 5.1.8 shows that $\mathrm{Gal}(E/E^H) = H$. Let M/F be a subextension. Then E is Galois over M by (5.2.8), which means that $E^{\mathrm{Gal}(E/M)} = M$.

(a) We have the obvious implications,

$$H_1 \supset H_2 \implies E^{H_1} \subset E^{H_2} \implies \mathrm{Gal}(E/E^{H_1}) \supset \mathrm{Gal}(E/E^{H_2}).$$

As $\mathrm{Gal}(E/E^{H_i}) = H_i$, this proves (a).

(b) Let H be a subgroup of G. According to 5.2.6b,

$$(\mathrm{Gal}(E/E^H) : 1) = [E : E^H].$$

This proves (b) in the case $H_2 = 1$, and the general case follows, using that

$$\begin{aligned}(H_1 : 1) &= (H_1 : H_2)(H_2 : 1)\\ [E : E^{H_1}] &\overset{4.2.6}{=} [E : E^{H_2}][E^{H_2} : E^{H_1}].\end{aligned}$$

(c) For $\tau \in G$ and $\alpha \in E$,

$$\tau\alpha = \alpha \iff \sigma\tau\sigma^{-1}(\sigma\alpha) = \sigma\alpha.$$

Therefore, τ fixes M if and only if $\sigma\tau\sigma^{-1}$ fixes σM, and so $\sigma \,\mathrm{Gal}(E/M)\sigma^{-1} = \mathrm{Gal}(E/\sigma M)$. This shows that $\sigma \,\mathrm{Gal}(E/M)\sigma^{-1}$ corresponds to σM.

(d) Let H be a normal subgroup of G. Because $\sigma H \sigma^{-1} = H$ for all $\sigma \in G$, we must have $\sigma E^H = E^H$ for all $\sigma \in G$, i.e., the action of G on E stabilizes E^H. We therefore have a homomorphism

$$\sigma \mapsto \sigma|E^H : G \to \mathrm{Aut}(E^H/F)$$

whose kernel is H. As $(E^H)^{G/H} = F$, we see that E^H is Galois over F (by Theorem 5.2.4) and that $G/H \simeq \mathrm{Gal}(E^H/F)$ (by 5.2.6b).

Conversely, suppose that M is normal over F, and let $\alpha_1, \dots, \alpha_m$ generate M over F. For $\sigma \in G$, $\sigma\alpha_i$ is a root of the minimal polynomial of α_i over F, and so lies in M. Hence $\sigma M = M$, and this implies that $\sigma H \sigma^{-1} = H$ (by (c)). □

Remark 5.3.2. Let E/F be a Galois extension, so that there is an order reversing bijection between the subextensions of E/F and the subgroups of G. From this, we can read off the following results.

(a) Let $M_1, M_2, \dots, M_r$ be subextensions of E/F, and let H_i be the subgroup corresponding to M_i (i.e., $H_i = \mathrm{Gal}(E/M_i)$). Then (by definition) $M_1M_2\cdots M_r$ is the smallest field containing all M_i; hence it must correspond to the largest subgroup contained in all H_i, which is $\bigcap H_i$. Therefore

$$\mathrm{Gal}(E/M_1\cdots M_r) = H_1 \cap \ldots \cap H_r.$$

(b) Let H be a subgroup of G and let $M = E^H$. The largest normal subgroup contained in H is $N = \bigcap_{\sigma\in G} \sigma H\sigma^{-1}$ (see GT, 4.1), and so E^N is the smallest normal extension of F containing M. Note that, by (a), E^N is the composite of the fields σM. It is called the **normal**, or **Galois**, closure of M in E.

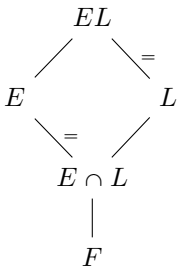
Proposition 5.3.3. Let E and L be extensions of F contained in some common field. If E/F is Galois, then EL/L and $E/E\cap L$ are Galois, and the map

$$\sigma \mapsto \sigma|E\colon\ \mathrm{Gal}(EL/L) \to \mathrm{Gal}(E/E\cap L)$$

is an isomorphism.

Proof. Because E is Galois over F, it is the splitting field of a separable polynomial $f \in F[X]$. Then EL is the splitting field of f over L, and E is the splitting field of f over $E\cap L$. Hence EL/L and $E/E\cap L$ are Galois. Every automorphism σ of EL fixing the elements of L maps roots of f to roots of f, and so $\sigma E = E$. There is therefore a homomorphism

$$\sigma \mapsto \sigma|E\colon\ \mathrm{Gal}(EL/L) \to \mathrm{Gal}(E/E\cap L).$$



If $\sigma \in \mathrm{Gal}(EL/L)$ fixes the elements of E, then it fixes the elements of EL, and hence is the identity map. Thus, $\sigma \mapsto \sigma|E$ is injective. If $\alpha \in E$ is fixed by all $\sigma \in \mathrm{Gal}(EL/L)$, then $\alpha \in E\cap L$. By Corollary 5.1.8,

this implies that the image of $\sigma \mapsto \sigma|E$ is $\mathrm{Gal}(E/E\cap L)$. □

Corollary 5.3.4. Suppose, in the proposition, that L is finite over F. Then

$$[EL\colon F] = \frac{[E\colon F][L\colon F]}{[E\cap L\colon F]}.$$

Proof. According to Proposition 4.2.6,

$$[EL\colon F] = [EL\colon L][L\colon F],$$

but

$$[EL\colon L] \overset{5.3.3}{=} [E\colon E\cap L] \overset{4.2.6}{=} \frac{[E\colon F]}{[E\cap L\colon F]}.$$

□

Proposition 5.3.5. Let E_1 and E_2 be extensions of F contained in some common field. If E_1 and E_2 are Galois over F, then E_1E_2 and $E_1\cap E_2$ are Galois over F, and the map

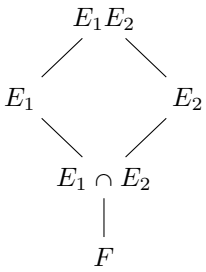
$$\sigma \mapsto (\sigma|E_1, \sigma|E_2)\colon\ \mathrm{Gal}(E_1E_2/F) \to \mathrm{Gal}(E_1/F) \times \mathrm{Gal}(E_2/F)$$

is an isomorphism of $\mathrm{Gal}(E_1E_2/F)$ onto the subgroup

$$H = \{(\sigma_1, \sigma_2) \mid \sigma_1|E_1\cap E_2 = \sigma_2|E_1\cap E_2\}$$

of $\mathrm{Gal}(E_1/F) \times \mathrm{Gal}(E_2/F)$.

PROOF: Let $a \in E_1 \cap E_2$, and let f be its minimal polynomial over F. Then f has $\deg f$ distinct roots in E_1 and $\deg f$ distinct roots in E_2. Since f can have at most $\deg f$ roots in E_1E_2, it follows that it has $\deg f$ distinct roots in $E_1 \cap E_2$. This shows that $E_1 \cap E_2$ is normal and separable over F, and hence Galois (5.2.4). As E_1 and E_2 are Galois over F, they are splitting fields for separable polynomials $f_1, f_2 \in F[X]$. Now E_1E_2 is a splitting field for $\operatorname{lcm}(f_1, f_2)$, and hence it also is Galois over F. The map $\sigma \mapsto (\sigma|E_1, \sigma|E_2)$ is clearly an injective homomorphism, and its image is contained in H. We'll prove that the image is the whole of H by counting.



From the fundamental theorem,

$$\frac{\operatorname{Gal}(E_2/F)}{\operatorname{Gal}(E_2/E_1 \cap E_2)} \simeq \operatorname{Gal}(E_1 \cap E_2/F),$$

and so, for each $\sigma_1 \in \operatorname{Gal}(E_1/F)$, $\sigma_1|E_1 \cap E_2$ has exactly $[E_2 \colon E_1 \cap E_2]$ extensions to an element of $\operatorname{Gal}(E_2/F)$. Therefore,

$$(H \colon 1) = [E_1 \colon F][E_2 \colon E_1 \cap E_2] = \frac{[E_1 \colon F] \cdot [E_2 \colon F]}{[E_1 \cap E_2 \colon F]},$$

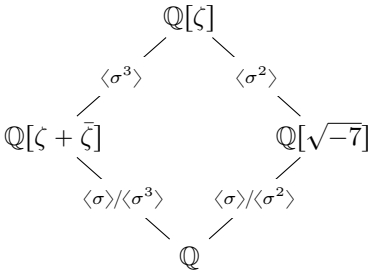
which equals $[E_1E_2 \colon F]$ by (5.3.4). □

Example 5.3.6. We analyse the extension $\mathbb{Q}[\zeta]/\mathbb{Q}$, where ζ is a primitive 7th root of 1, say $\zeta = e^{2\pi i/7}$.

Note that $\mathbb{Q}[\zeta]$ is the splitting field of the polynomial $X^7 - 1$, and that ζ has minimal polynomial

$$X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$$

(see **??**). Therefore, $\mathbb{Q}[\zeta]$ is Galois of degree 6 over $\mathbb{Q}$. For any $\sigma \in \operatorname{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q})$, $\sigma\zeta = \zeta^i$, some i, $1 \leqslant i \leqslant 6$, and the map $\sigma \mapsto i$ defines an isomorphism $\operatorname{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q}) \to (\mathbb{Z}/7\mathbb{Z})^\times$. Let σ be the element of $\operatorname{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q})$ such that $\sigma\zeta = \zeta^3$. Then σ generates $\operatorname{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q})$ because the class of 3 in $(\mathbb{Z}/7\mathbb{Z})^\times$ generates it (the powers of 3 mod 7 are $3, 2, 6, 4, 5, 1$). We investigate the subfields of $\mathbb{Q}[\zeta]$ corresponding to the subgroups $\langle\sigma^3\rangle$ and $\langle\sigma^2\rangle$.



Note that $\sigma^3\zeta = \zeta^6 = \bar{\zeta}$ (complex conjugate of ζ), and so $\zeta + \bar{\zeta} = 2\cos\frac{2\pi}{7}$ is fixed by σ^3. Now $\mathbb{Q}[\zeta] \supset \mathbb{Q}[\zeta]^{\langle\sigma^3\rangle} \supset \mathbb{Q}[\zeta + \bar{\zeta}] \neq \mathbb{Q}$, and so $\mathbb{Q}[\zeta]^{\langle\sigma^3\rangle} = \mathbb{Q}[\zeta + \bar{\zeta}]$ (look at degrees). As $\langle\sigma^3\rangle$ is a normal subgroup of $\langle\sigma\rangle$, $\mathbb{Q}[\zeta + \bar{\zeta}]$ is Galois over $\mathbb{Q}$, with Galois group $\langle\sigma\rangle/\langle\sigma^3\rangle$. The conjugates of $\alpha_1 \overset{\text{def}}{=} \zeta + \bar{\zeta}$ are $\alpha_3 = \zeta^3 + \zeta^{-3}$, $\alpha_2 = \zeta^2 + \zeta^{-2}$. Direct calculation shows that

$$\begin{aligned}
\alpha_1 + \alpha_2 + \alpha_3 &= \textstyle\sum_{i=1}^{6} \zeta^i = -1,\\
\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3 &= -2,\\
\alpha_1\alpha_2\alpha_3 &= (\zeta + \zeta^6)(\zeta^2 + \zeta^5)(\zeta^3 + \zeta^4)\\
&= (\zeta + \zeta^3 + \zeta^4 + \zeta^6)(\zeta^3 + \zeta^4)\\
&= (\zeta^4 + \zeta^6 + 1 + \zeta^2 + \zeta^5 + 1 + \zeta + \zeta^3)\\
&= 1.
\end{aligned}$$

Hence the minimal polynomial[2] of $\zeta + \bar{\zeta}$ is

$$g(X) = X^3 + X^2 - 2X - 1.$$

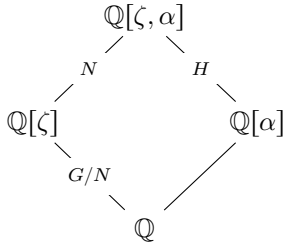
The minimal polynomial of $\cos\frac{2\pi}{7} = \frac{\alpha_1}{2}$ is therefore

$$\frac{g(2X)}{8} = X^3 + X^2/2 - X/2 - 1/8.$$

The subfield of $\mathbb{Q}[\zeta]$ corresponding to $\langle\sigma^2\rangle$ is generated by $\beta = \zeta+\zeta^2+\zeta^4$. Let $\beta' = \sigma\beta$. Then $(\beta-\beta')^2 = -7$. Hence the field fixed by $\langle\sigma^2\rangle$ is $\mathbb{Q}[\sqrt{-7}]$.

Example 5.3.7. We compute the Galois group of a splitting field E of $X^5 - 2 \in \mathbb{Q}[X]$.

Recall from Exercise 3 that $E = \mathbb{Q}[\zeta, \alpha]$ where ζ is a primitive 5th root of 1, and α is a root of $X^5 - 2$. For example, we could take E to be the splitting field of $X^5 - 2$ in $\mathbb{C}$, with $\zeta = e^{2\pi i/5}$ and α equal to the real 5th root of 2. We have the picture at right, and



$$[\mathbb{Q}[\zeta] : \mathbb{Q}] = 4, \quad [\mathbb{Q}[\alpha] : \mathbb{Q}] = 5.$$

Because 4 and 5 are relatively prime,

$$[\mathbb{Q}[\zeta, \alpha] : \mathbb{Q}] = 20.$$

Hence $G = \mathrm{Gal}(\mathbb{Q}[\zeta, \alpha]/\mathbb{Q})$ has order 20, and the subgroups N and H fixing $\mathbb{Q}[\zeta]$ and $\mathbb{Q}[\alpha]$ have orders 5 and 4 respectively. Because $\mathbb{Q}[\zeta]$ is normal over $\mathbb{Q}$ (it is the splitting field of $X^5 - 1$), N is normal in G. Because $\mathbb{Q}[\zeta] \cdot \mathbb{Q}[\alpha] = \mathbb{Q}[\zeta, \alpha]$, we have $H \cap N = 1$, and so $G = N \rtimes_\theta H$. Moreover, $H \simeq G/N \simeq (\mathbb{Z}/5\mathbb{Z})^\times$, which is cyclic, being generated by the class of 2. Let τ be the generator of H corresponding to 2 under this isomorphism, and let σ be a generator of N. Thus $\sigma(\alpha)$ is another root of $X^5 - 2$, which we can take to be $\zeta\alpha$ (after possibly replacing σ by a power). Hence:

$$\begin{cases} \tau\zeta &= \zeta^2 \\ \tau\alpha &= \alpha \end{cases} \quad \begin{cases} \sigma\zeta &= \zeta \\ \sigma\alpha &= \zeta\alpha. \end{cases}$$

Note that $\tau\sigma\tau^{-1}(\alpha) = \tau\sigma\alpha = \tau(\zeta\alpha) = \zeta^2\alpha$ and it fixes ζ; therefore $\tau\sigma\tau^{-1} = \sigma^2$. Thus G has generators σ and τ and defining relations

$$\sigma^5 = 1, \quad \tau^4 = 1, \quad \tau\sigma\tau^{-1} = \sigma^2.$$

The subgroup H has five conjugates, which correspond to the five fields $\mathbb{Q}[\zeta^i\alpha]$,

$$\sigma^i H \sigma^{-i} \leftrightarrow \sigma^i\mathbb{Q}[\alpha] = \mathbb{Q}[\zeta^i\alpha], \qquad 1 \leqslant i \leqslant 5.$$

5.4 The Galois group of a polynomial

If a polynomial $f \in F[X]$ is separable, then its splitting field F_f is Galois over F, and we call $\mathrm{Gal}(F_f/F)$ the **Galois group** G_f of f.

Let $f(X) = \prod_{i=1}^n (X - \alpha_i)$ in a splitting field F_f. We know that the elements of $\mathrm{Gal}(F_f/F)$ map roots of f to roots of f, i.e., they map the set $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ into itself. Being automorphisms, they act as permutations on $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. As the α_i generate F_f over F, an element of $\mathrm{Gal}(F_f/F)$ is uniquely determined by the permutation it defines. Thus G_f can be identified with a subset of $\mathrm{Sym}(\{\alpha_1, \alpha_2, \ldots, \alpha_n\}) \approx S_n$ (symmetric

[2] More directly, on setting $X = \zeta + \bar{\zeta}$ in

$$(X^3 - 3X) + (X^2 - 2) + X + 1$$

one obtains $1 + \zeta + \zeta^2 + \cdots + \zeta^6 = 0$.

group on n symbols). In fact, G_f consists exactly of the permutations σ of $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ such that, for $P \in F[X_1, \dots, X_n]$,

$$P(\alpha_1, \dots, \alpha_n) = 0 \implies P(\sigma\alpha_1, \dots, \sigma\alpha_n) = 0. \tag{5.2}$$

To see this, note that the kernel of the map

$$F[X_1, \dots, X_n] \to F_f, \quad X_i \mapsto \alpha_i, \tag{5.3}$$

consists of the polynomials $P(X_1, \dots, X_n)$ such that $P(\alpha_1, \dots, \alpha_n) = 0$. Let σ be a permutation of the α_i satisfying the condition (5.2). Then the map

$$F[X_1, \dots, X_n] \to F_f, \quad X_i \mapsto \sigma\alpha_i,$$

factors through the map (5.3), and defines an F-isomorphism $F_f \to F_f$, i.e., an element of the Galois group. This shows that every permutation satisfying the condition (5.2) extends uniquely to an element of G_f, and it is obvious that every element of G_f arises in this way.

This gives a description of G_f not mentioning fields or abstract groups, neither of which were available to Galois. Note that it shows again that $(G_f : 1)$, hence $[F_f : F]$, divides $\deg(f)!$.

5.5 Solvability of equations

For a polynomial $f \in F[X]$, we say that $f(X) = 0$ is **solvable in radicals** if its solutions can be obtained by the algebraic operations of addition, subtraction, multiplication, division, and the extraction of mth roots, or, more precisely, if there exists a tower of fields

$$F = F_0 \subset F_1 \subset F_2 \subset \cdots \subset F_m$$

such that

1. $F_i = F_{i-1}[\alpha_i]$, $\alpha_i^{m_i} \in F_{i-1}$;
2. F_m contains a splitting field for f.

Theorem 5.5.1 (Galois, 1832)**.** Let F be a field of characteristic zero, and let $f \in F[X]$. The equation $f(X) = 0$ is solvable in radicals if and only if the Galois group of f is solvable.

We'll prove this later (**??**). Also we'll exhibit polynomials $f(X) \in \mathbb{Q}[X]$ with Galois group S_n, which are therefore not solvable when $n \geqslant 5$ by GT, 4.37.

Remark 5.5.2. When F has characteristic p, the theorem fails for two reasons,

1. f need not be separable, and so not have a Galois group;
2. $X^p - X - a = 0$ need not be solvable in radicals even though it is separable with abelian Galois group (cf. Exercise 2).

If the definition of solvable is changed to allow extensions defined by polynomials of the type in (b) in the chain, then the theorem holds for fields F of characteristic $p \neq 0$ and separable $f \in F[X]$.

5.6 When is $G_f \subset A_n$?

Let σ be a permutation of the set $\{1, 2, \dots, n\}$. The pairs (i, j) with $i < j$ but $\sigma(i) > \sigma(j)$ are called the **inversions** of σ, and σ is said to be **even** or **odd** according as the number of inversions is even or odd. The **signature** of σ, $\operatorname{sign}(\sigma)$, is $+1$ or -1 according as σ is even or odd. We can define the signature of a permutation σ of any set S of n elements by choosing a numbering of the set and identifying σ with a

permutation of $\{1,\dots,n\}$. Then sign is the unique homomorphism $\mathrm{Sym}(S)\to\{\pm1\}$ such that $\mathrm{sign}(\sigma)=-1$ for every transposition. In particular, it is independent of the choice of the numbering. See GT, 4.25.

Now consider a monic polynomial

$$f(X)=X^n+a_1X^{n-1}+\cdots+a_n$$

and let $f(X)=\prod_{i=1}^n(X-\alpha_i)$ in some splitting field. Set

$$\Delta(f)=\prod_{1\leqslant i<j\leqslant n}(\alpha_i-\alpha_j),\qquad D(f)=\Delta(f)^2=\prod_{1\leqslant i<j\leqslant n}(\alpha_i-\alpha_j)^2.$$

The **discriminant** of f is defined to be $D(f)$. Note that $D(f)$ is nonzero if and only if f has only simple roots, i.e., is separable. Let G_f be the Galois group of f, and identify it with a subgroup of $\mathrm{Sym}(\{\alpha_1,\dots,\alpha_n\})$ (as on p. 153).

Proposition 5.6.1. Let $f\in F[X]$ be a separable polynomial, and let $\sigma\in G_f$.

1. $\sigma\Delta(f)=\mathrm{sign}(\sigma)\Delta(f)$, where $\mathrm{sign}(\sigma)$ is the signature of σ.
2. $\sigma D(f)=D(f)$.

Proof. Each inversion of σ introduces a negative sign into $\sigma\Delta(f)$, and so (a) follows from the definition of $\mathrm{sign}(\sigma)$. The equation in (b) is obtained by squaring that in (a). □

While $\Delta(f)$ depends on the choice of the numbering of the roots of f, $D(f)$ does not.

Corollary 5.6.2. Let $f(X)\in F[X]$ be separable of degree n. Let F_f be a splitting field for f and let $G_f=\mathrm{Gal}(F_f/F)$.

1. The discriminant $D(f)\in F$.
2. Assume that F has characteristic $\neq 2$. The subfield of F_f corresponding to $A_n\cap G_f$ is $F[\Delta(f)]$. Hence

$$G_f\subset A_n\iff\Delta(f)\in F\iff D(f)\text{ is a square in }F.$$

Proof. (a) The discriminant of f is an element of F_f fixed by $G_f\overset{\text{def}}{=}\mathrm{Gal}(F_f/F)$, and hence lies in F (by the fundamental theorem).

(b) Because f has simple roots, $\Delta(f)\neq0$, and so the formula $\sigma\Delta(f)=\mathrm{sign}(\sigma)\Delta(f)$ shows that an element of G_f fixes $\Delta(f)$ if and only if it lies in A_n. Thus, under the Galois correspondence,

$$G_f\cap A_n\leftrightarrow F[\Delta(f)].$$

Hence,

$$G_f\cap A_n=G_f\iff F[\Delta(f)]=F.$$

□

The roots of X^2+bX+c are $\frac{-b\pm\sqrt{b^2-4c}}{2}$ and so

$$\Delta(X^2+bX+c)=\sqrt{b^2-4c}\ (\text{or }-\sqrt{b^2-4c}),$$
$$D(X^2+bX+c)=b^2-4c.$$

Similarly,

$$D(X^3+bX+c)=-4b^3-27c^2.$$

By completing the cube, one can put any cubic polynomial in this form (in characteristic $\neq3$).

Although there is a not a universal formula for the roots of f in terms of its coefficients when the $\deg(f) > 4$, there is for its discriminant. However, the formulas for the discriminant rapidly become very complicated, for example, that for $X^5 + aX^4 + bX^3 + cX^2 + dX + e$ has 59 terms. Fortunately, PARI knows them. For example, typing `poldisc(X^3+a*X^2+b*X+c,X)` returns the discriminant of $X^3 + aX^2 + bX + c$, namely,

$$-4ca^3 + b^2a^2 + 18cba + (-4b^3 - 27c^2).$$

Remark 5.6.3. Suppose $F \subset \mathbb{R}$. Then $D(f)$ will not be a square if it is negative. It is known that the sign of $D(f)$ is $(-1)^s$ where $2s$ is the number of nonreal roots of f in $\mathbb{C}$ (see ANT 2.40). Thus if s is odd, then G_f is not contained in A_n. This can be proved more directly by noting that complex conjugation acts on the roots as the product of s disjoint transpositions.

The converse is not true: when s is even, G_f is not necessarily contained in A_n.

5.7 When does G_f act transitively on the roots?

Proposition 5.7.1. Let $f(X) \in F[X]$ be separable. Then $f(X)$ is irreducible if and only if G_f permutes the roots of f transitively.

Proof. $\implies$: If α and β are two roots of $f(X)$ in a splitting field F_f for f, then they both have $f(X)$ as their minimal polynomial, and so $F[\alpha]$ and $F[\beta]$ are both stem fields for f. Hence, there is an F-isomorphism

$$F[\alpha] \simeq F[\beta], \qquad \alpha \leftrightarrow \beta.$$

Write $F_f = F[\alpha_1, \alpha_2, ...]$ with $\alpha_1 = \alpha$ and $\alpha_2, \alpha_3, \dots$ the other roots of $f(X)$. Then the F-homomorphism $\alpha \mapsto \beta\colon F[\alpha] \to F_f$ extends (step by step) to an F-homomorphism $F_f \to F_f$ (use 4.5.2b), which is an F-isomorphism sending α to β.

$\impliedby$: Let $g(X) \in F[X]$ be an irreducible factor of f, and let α be one of its roots. If β is a second root of f, then (by assumption) $\beta = \sigma\alpha$ for some $\sigma \in G_f$. Now, because g has coefficients in F,

$$g(\sigma\alpha) = \sigma g(\alpha) = 0,$$

and so β is also a root of g. Therefore, every root of f is also a root of g, and so $f(X) = g(X)$. □

Note that when $f(X)$ is irreducible of degree n, $n|(G_f\colon 1)$ because $[F[\alpha]\colon F] = n$ and $[F[\alpha]\colon F]$ divides $[F_f\colon F] = (G_f\colon 1)$. Thus G_f is a transitive subgroup of S_n whose order is divisible by n.

5.8 Polynomials of degree at most three

Example 5.8.1. Let $f(X) \in F[X]$ be a polynomial of degree 2. Then f is inseparable $\iff$ F has characteristic 2 and $f(X) = X^2 - a$ for some $a \in F \smallsetminus F^2$. If f is separable, then $G_f = 1 (= A_2)$ or S_2 according as $D(f)$ is a square in F or not.

Example 5.8.2. Let $f(X) \in F[X]$ be a polynomial of degree 3. We can assume f to be irreducible, for otherwise we are essentially back in the previous case. Then f is inseparable if and only if F has characteristic 3 and $f(X) = X^3 - a$ for some $a \in F \smallsetminus F^3$. If f is separable, then G_f is a transitive subgroup of S_3 whose order is divisible by 3. There are only two possibilities: $G_f = A_3$ or S_3 according as $D(f)$ is a square in F or not. Note that A_3 is generated by the cycle (123).

For example, $X^3 - 3X + 1$ is irreducible in $\mathbb{Q}[X]$ by rational root theorem. Its discriminant is $-4(-3)^3 - 27 = 81 = 9^2$, and so its Galois group is A_3.

On the other hand, $X^3 + 3X + 1 \in \mathbb{Q}[X]$ is also irreducible (apply **??**), but its discriminant is -135 which is not a square in $\mathbb{Q}$, and so its Galois group is S_3.

5.9 Quartic polynomials

Let $f(X)$ be a separable quartic polynomial. In order to determine G_f we'll exploit the fact that S_4 has

$$V = \{1, (12)(34), (13)(24), (14)(23)\}$$

as a normal subgroup — it is normal because it contains all elements of type $2+2$ (GT, 4.29). Let E be a splitting field of f, and let $f(X) = \prod(X - \alpha_i)$ in E. We identify the Galois group G_f of f with a subgroup of the symmetric group $\mathrm{Sym}(\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\})$. Consider the partially symmetric elements

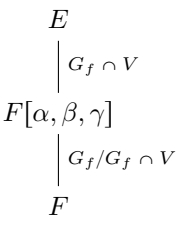
$$\begin{aligned}\alpha &= \alpha_1\alpha_2 + \alpha_3\alpha_4\\ \beta &= \alpha_1\alpha_3 + \alpha_2\alpha_4\\ \gamma &= \alpha_1\alpha_4 + \alpha_2\alpha_3.\end{aligned}$$

They are distinct because the α_i are distinct; for example,

$$\alpha - \beta = \alpha_1(\alpha_2 - \alpha_3) + \alpha_4(\alpha_3 - \alpha_2) = (\alpha_1 - \alpha_4)(\alpha_2 - \alpha_3).$$

The group $\mathrm{Sym}(\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\})$ permutes $\{\alpha, \beta, \gamma\}$ transitively. The stabilizer of each of α, β, γ must therefore be a subgroup of index 3 in S_4, and hence has order 8. For example, the stabilizer of β is $\langle(1234), (13)\rangle$. Groups of order 8 in S_4 are Sylow 2-subgroups. There are three of them, all isomorphic to D_4. By the Sylow theorems, V is contained in a Sylow 2-subgroup; in fact, because the Sylow 2-subgroups are conjugate and V is normal, it is contained in all three. It follows that V is the intersection of the three Sylow 2-subgroups. Each Sylow 2-subgroup fixes exactly one of $\alpha, \beta,$ or γ, and therefore their intersection V is the subgroup of $\mathrm{Sym}(\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\})$ fixing α, β, and γ.

Lemma 5.9.1. The fixed field of $G_f \cap V$ is $F[\alpha, \beta, \gamma]$. Hence $F[\alpha, \beta, \gamma]$ is Galois over F with Galois group $G_f/G_f \cap V$.



Proof. The above discussion shows that the subgroup of G_f of elements fixing $F[\alpha, \beta, \gamma]$ is $G_f \cap V$, and so $E^{G_f \cap V} = F[\alpha, \beta, \gamma]$ by the fundamental theorem of Galois theory. The remaining statements follow from the fundamental theorem using that V is normal. □

Let $M = F[\alpha, \beta, \gamma]$, and let $g(X) = (X - \alpha)(X - \beta)(X - \gamma) \in M[X]$ — it is called the **resolvent cubic** of f. Every permutation of the α_i (*a fortiori*, every element of G_f) merely permutes α, β, γ, and so fixes $g(X)$. Therefore (by the fundamental theorem) $g(X)$ has coefficients in F. More explicitly, we have:

Lemma 5.9.2. The resolvent cubic of $f = X^4 + bX^3 + cX^2 + dX + e$ is

$$g = X^3 - cX^2 + (bd - 4e)X - b^2e + 4ce - d^2.$$

The discriminants of f and g are equal.

sketch of proof. Expand $f = (X - \alpha_1)(X - \alpha_2)(X - \alpha_3)(X - \alpha_4)$ to express b, c, d, e in terms of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. Expand $g = (X - \alpha)(X - \beta)(X - \gamma)$ to express the coefficients of g in terms of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, and substitute to express them in terms of b, c, d, e. □

Now let f be an irreducible separable quartic. Then $G = G_f$ is a transitive subgroup of S_4 whose order is divisible by 4. There are the following possibilities for G:

G	$(G\cap V:1)$	$(G:V\cap G)$
S_4	4	6
A_4	4	3
V	4	1
D_4	4	2
C_4	2	2

$(G\cap V:1)=[E:M]$
$(G:V\cap G)=[M:F]$

The groups of type D_4 are the Sylow 2-subgroups discussed above, and the groups of type C_4 are those generated by cycles of length 4.

We can compute $(G:V\cap G)$ from the resolvent cubic g, because $G/V\cap G=\mathrm{Gal}(M/F)$ and M is the splitting field of g. Once we know $(G:V\cap G)$, we can deduce G except in the case that it is 2. If $[M:F]=2$, then $G\cap V=V$ or C_2. Only the first group acts transitively on the roots of f, and so (from 5.7.1) we see that in this case $G=D_4$ or C_4 according as f is irreducible or not in $M[X]$.

Example 5.9.3. Consider $f(X)=X^4-4X+2\in\mathbb{Q}[X]$. It is irreducible by Eisenstein's criterion (**??**), and its resolvent cubic is $g(X)=X^3-8X-16$, which is irreducible because it has no roots in $\mathbb{F}_5$. The discriminant of $g(X)$ is -4864, which is not a square, and so the Galois group of $g(X)$ is S_3. From the table, we see that the Galois group of $f(X)$ is S_4.

Example 5.9.4. Consider $f(X)=X^4+4X^2+2\in\mathbb{Q}[X]$. It is irreducible by Eisenstein's criterion (**??**), and its resolvent cubic is $(X-4)(X^2-8)$; thus $M=\mathbb{Q}[\sqrt{2}]$. From the table we see that G_f is of type D_4 or C_4, but f factors over M (even as a polynomial in X^2), and hence G_f is of type C_4.

Example 5.9.5. Consider $f(X)=X^4-10X^2+4\in\mathbb{Q}[X]$. It is irreducible in $\mathbb{Q}[X]$ because (by inspection) it is irreducible in $\mathbb{Z}[X]$. Its resolvent cubic is $(X+10)(X+4)(X-4)$, and so G_f is of type V.

Example 5.9.6. Consider $f(X)=X^4-2\in\mathbb{Q}[X]$. It is irreducible by Eisenstein's criterion (**??**), and its resolvent cubic is $g(X)=X^3+8X$. Hence $M=\mathbb{Q}[i\sqrt{2}]$. One can check that f is irreducible over M, and G_f is of type D_4.

Alternatively, analyse the equation as in (5.3.7).

As we explained in (4.3.3), PARI knows how to factor polynomials with coefficients in $\mathbb{Q}[\alpha]$.

Example 5.9.7. (From the web, sci.math.research, search for "final analysis".) Consider $f(X)=X^4-2cX^3-dX^2+2cdX-dc^2\in\mathbb{Z}[X]$ with $a>0$, $b>0$, $c>0$, $a>b$ and $d=a^2-b^2$. Let $r=d/c^2$ and let w be the unique positive real number such that $r=w^3/(w^2+4)$. Let m be the number of roots of $f(X)$ in $\mathbb{Z}$ (counted with multiplicities). The Galois group of f is as follows:

- If $m=0$ and w not rational, then G is S_4.
- If $m=1$ and w not rational then G is S_3.
- If w is rational and w^2+4 is not a square then $G=D_4$.
- If w is rational and w^2+4 is a square then $G=V=C_2\times C_2$.

This covers all possible cases. The hard part was to establish that $m=2$ could never happen.

For a discussion of whether the method of solving a quartic by reducing to a cubic generalizes to other even degrees, see mo149099.

5.10 Examples of polynomials with S_p as Galois group over $\mathbb{Q}$

The next lemma gives a criterion for a subgroup of S_p to be the whole of S_p.

Lemma 5.10.1. For p prime, the symmetric group S_p is generated by any transposition and any p-cycle.

Proof. After renumbering, we may assume that the transposition is $\tau = (12)$, and we may write the p-cycle σ so that 1 occurs in the first position, $\sigma = (1\, i_2 \cdots i_p)$. Now some power of σ will map 1 to 2 and will still be a p-cycle (here is where we use that p is prime). After replacing σ with the power, we have $\sigma = (1\, 2\, j_3 \ldots j_p)$, and after renumbering again, we have $\sigma = (1\, 2\, 3 \ldots p)$. Now

$$(i\ i+1) = \sigma^i(12)\sigma^{-i}$$

(see GT, 4.29) and so lies in the subgroup generated by σ and τ. These transpositions generate S_p. □

Proposition 5.10.2. Let f be an irreducible polynomial of prime degree p in $\mathbb{Q}[X]$. If f splits in $\mathbb{C}$ and has exactly two nonreal roots, then $G_f = S_p$.

Proof. Let E be the splitting field of f in $\mathbb{C}$, and let $\alpha \in E$ be a root of f. Because f is irreducible, $[\mathbb{Q}[\alpha]\colon \mathbb{Q}] = \deg f = p$, and so $p|[E\colon \mathbb{Q}] = (G_f\colon 1)$. Therefore G_f contains an element of order p (Cauchy's theorem, GT, 4.13), but the only elements of order p in S_p are p-cycles (here we use that p is prime again).

Let σ be complex conjugation on $\mathbb{C}$. Then σ transposes the two nonreal roots of $f(X)$ and fixes the rest. Therefore $G_f \subset S_p$ and contains a transposition and a p-cycle, and so is the whole of S_p. □

It remains to construct polynomials satisfying the conditions of the Proposition.

Example 5.10.3. Let $p \geqslant 5$ be a prime number. Choose a positive even integer m and even integers

$$n_1 < n_2 < \cdots < n_{p-2},$$

and let

$$g(X) = (X^2 + m)(X - n_1)...(X - n_{p-2}).$$

The graph of g crosses the x-axis exactly at the points $n_1, \ldots, n_{p-2}$, and it doesn't have a local maximum or minimum at any of those points (because the n_i are simple roots). Thus $e = \min_{g'(x)=0} |g(x)| > 0$, and we can choose an odd positive integer n such that $\frac{2}{n} < e$.

Consider

$$f(X) = g(X) - \frac{2}{n}.$$

As $\frac{2}{n} < e$, the graph of f also crosses the x-axis at exactly $p-2$ points, and so f has exactly two nonreal roots. On the other hand, when we write

$$nf(X) = nX^p + a_1X^{p-1} + \cdots + a_p,$$

the a_i are all even and a_p is not divisible by 2^2, and so Eisenstein's criterion implies that f is irreducible. Over $\mathbb{R}$, f has $p-2$ linear factors and one irreducible quadratic factor, and so it certainly splits over $\mathbb{C}$ (high school algebra). Therefore, the proposition applies to f.[3]

Example 5.10.4. The reader shouldn't think that, in order to have Galois group S_p, a polynomial must have exactly two nonreal roots. For example, the polynomial $X^5 - 5X^3 + 4X - 1$ has Galois group S_5 but all of its roots are real.

[3] If m is taken sufficiently large, then $g(X) - 2$ will have exactly two nonreal roots, i.e., we can take $n = 1$, but the proof is longer (see Jacobson 1964, p. 107, who credits the example to Brauer). The shorter argument in the text was suggested to me by Martin Ward.

5.11 Finite fields

Let $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, the field of p elements. As we noted in §1, every field E of characteristic p contains a copy of $\mathbb{F}_p$, namely, $\{m1_E \mid m \in \mathbb{Z}\}$. No harm results if we identify $\mathbb{F}_p$ with this subfield of E.

Let E be a field of degree n over $\mathbb{F}_p$. Then E has $q = p^n$ elements, and so $E^\times$ is a group of order $q-1$. Therefore the nonzero elements of E are roots of $X^{q-1}-1$, and *all* elements of E are roots of $X^q - X$. Hence E is a splitting field for $X^q - X$, and so any two fields with q elements are isomorphic.

Proposition 5.11.1. Every extension of finite fields is simple.

Proof. Consider $E \supset F$. Then $E^\times$ is a finite subgroup of the multiplicative group of a field, and hence is cyclic (see Exercise **??**). If ζ generates $E^\times$ as a multiplicative group, then certainly $E = F[\zeta]$. □

Now let E be a splitting field of $f(X) = X^q - X$, $q = p^n$. The derivative $f'(X) = -1$, which is relatively prime to $f(X)$ (in fact, to every polynomial), and so $f(X)$ has q distinct roots in E. Let S be the set of its roots. Then S is obviously closed under multiplication and the formation of inverses, but it is also closed under subtraction: if $a^q = a$ and $b^q = b$, then

$$(a-b)^q = a^q - b^q = a - b.$$

Hence S is a field, and so $S = E$. In particular, E has p^n elements.

Proposition 5.11.2. For each power $q = p^n$ of p there exists a field $\mathbb{F}_q$ with q elements. Every such field is a splitting field for $X^q - X$, and so any two are isomorphic. Moreover, $\mathbb{F}_q$ is Galois over $\mathbb{F}_p$ with cyclic Galois group generated by the Frobenius automorphism $\sigma(a) = a^p$.

Proof. Only the final statement remains to be proved. The field $\mathbb{F}_q$ is Galois over $\mathbb{F}_p$ because it is the splitting field of a separable polynomial. We noted in 4.1.2 that $x \overset{\sigma}{\mapsto} x^p$ is an automorphism of $\mathbb{F}_q$. An element a of $\mathbb{F}_q$ is fixed by σ if and only if $a^p = a$, but $\mathbb{F}_p$ consists exactly of such elements, and so the fixed field of $\langle\sigma\rangle$ is $\mathbb{F}_p$. This proves that $\mathbb{F}_q$ is Galois over $\mathbb{F}_p$ and that $\langle\sigma\rangle = \mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p)$ (see 5.2.6b). □

Corollary 5.11.3. Let E be a field with p^n elements. For each divisor m of n, $m \geqslant 0$, E contains exactly one field with p^m elements.

Proof. We know that E is Galois over $\mathbb{F}_p$ and that $\mathrm{Gal}(E/\mathbb{F}_p)$ is the cyclic group of order n generated by σ. The group $\langle\sigma\rangle$ has one subgroup of order n/m for each m dividing n, namely, $\langle\sigma^m\rangle$, and so E has exactly one subfield of degree m over $\mathbb{F}_p$ for each m dividing n, namely, $E^{\langle\sigma^m\rangle}$. Because it has degree m over $\mathbb{F}_p$, $E^{\langle\sigma^m\rangle}$ has p^m elements. □

Corollary 5.11.4. Each monic irreducible polynomial f of degree $d|n$ in $\mathbb{F}_p[X]$ occurs exactly once as a factor of $X^{p^n} - X$; hence, the degree of the splitting field of f is $\leqslant d$.

Proof. First, the factors of $X^{p^n} - X$ are distinct because it has no common factor with its derivative. If $f(X)$ is irreducible of degree d, then $f(X)$ has a root in a field of degree d over $\mathbb{F}_p$. But the splitting field of $X^{p^n} - X$ contains a copy of every field of degree d over $\mathbb{F}_p$ with $d|n$. Hence some root of $X^{p^n} - X$ is also a root of $f(X)$, and therefore $f(X)|X^{p^n} - X$. In particular, f divides $X^{p^d} - X$, and therefore it splits in its splitting field, which has degree d over $\mathbb{F}_p$. □

Proposition 5.11.5. Let $\mathbb{F}$ be an algebraic closure of $\mathbb{F}_p$. Then $\mathbb{F}$ contains exactly one field $\mathbb{F}_{p^n}$ with p^n elements for each integer $n \geqslant 1$, and $\mathbb{F}_{p^n}$ consists of the roots of $X^{p^n} - X$. Moreover,

$$\mathbb{F}_{p^m} \subset \mathbb{F}_{p^n} \iff m|n.$$

The partially ordered set of finite subfields of $\mathbb{F}$ is isomorphic to the set of integers $n \geqslant 1$ partially ordered by divisibility.

Proof. In fact, the set of roots of $X^{p^n} - X$ is a field (see above), with p^n elements, and is the only such subfield. If $\mathbb{F}_{p^m} \subset \mathbb{F}_{p^n}$, say, $[\mathbb{F}_{p^n} : \mathbb{F}_{p^m}] = d$, then $p^n = (p^m)^d = p^{md}$, and so $m|n$; the converse follows from the first statement. The final statement follows from the second statement. □

Proposition 5.11.6. The field $\mathbb{F}_p$ has an algebraic closure $\mathbb{F}$.

Proof. Choose a sequence of integers $1 = n_1 < n_2 < n_3 < \cdots$ such that $n_i | n_{i+1}$ for all i, and every integer n divides some n_i. For example, let $n_i = i!$. Define the fields $\mathbb{F}_{p^{n_i}}$ inductively as follows: $\mathbb{F}_{p^{n_1}} = \mathbb{F}_p$; $\mathbb{F}_{p^{n_i}}$ is the splitting field of $X^{p^{n_i}} - X$ over $\mathbb{F}_{p^{n_{i-1}}}$. Then, $\mathbb{F}_{p^{n_1}} \subset \mathbb{F}_{p^{n_2}} \subset \mathbb{F}_{p^{n_3}} \subset \cdots$, and we define $\mathbb{F} = \bigcup \mathbb{F}_{p^{n_i}}$. As a union of a chain of fields algebraic over $\mathbb{F}_p$, it is again a field algebraic over $\mathbb{F}_p$. Moreover, every polynomial in $\mathbb{F}_p[X]$ splits in $\mathbb{F}$, and so it is an algebraic closure of $\mathbb{F}$ (by 4.4.4). □

Remark 5.11.7. Since the $\mathbb{F}_{p^n}$ are not subsets of a fixed set, forming the union requires explanation. Let S be the disjoint union of the $\mathbb{F}_{p^n}$. For $a, b \in S$, set $a \sim b$ if $a = b$ in one of the $\mathbb{F}_{p^n}$. Then $\sim$ is an equivalence relation, and we let $\mathbb{F} = S/\sim$.

Any two fields with q elements are isomorphic, but not necessarily *canonically* isomorphic. However, once we have chosen an algebraic closure $\mathbb{F}$ of $\mathbb{F}_p$, there is a *unique* subfield of $\mathbb{F}$ with q elements.

PARI factors polynomials modulo p very quickly. Recall that the syntax is
`factormod(f(X),p)`. For example, to obtain a list of all monic polynomials of degree $1, 2,$ or 4 over $\mathbb{F}_5$, ask PARI to factor $X^{625} - X$ modulo 5 (note that $625 = 5^4$).

In one of the few papers published during his lifetime, Galois defined finite fields of arbitrary prime power order and established their basic properties, for example, the existence of a primitive element (Notices A.M.S., Feb. 2003, p. 198). For this reason finite fields are often called **Galois fields** and the field with q elements is often denoted by $\mathrm{GF}(q)$.

5.12 Computing Galois groups over $\mathbb{Q}$

In the remainder of this chapter, I describe a practical method for computing Galois groups over $\mathbb{Q}$ and similar fields. Recall that for a separable polynomial $f \in F[X]$, F_f denotes a splitting field for F, and $G_f = \mathrm{Gal}(F_f/F)$ denotes the Galois group of f. Moreover, G_f permutes the roots $\alpha_1, \dots, \alpha_m$, $m = \deg f$, of f in F_f:

$$G \subset \mathrm{Sym}\{\alpha_1, \dots, \alpha_m\}.$$

The first result generalizes Proposition 5.7.1.

Proposition 5.12.1. Let $f(X)$ be a separable polynomial in $F[X]$, and suppose that the orbits of G_f acting on the roots of f have $m_1, \dots, m_r$ elements respectively. Then f factors as $f = f_1 \cdots f_r$ with f_i irreducible of degree m_i.

Proof. We may suppose that f is monic. Let $\alpha_1, \dots, \alpha_m$, be the roots of $f(X)$ in F_f. The monic factors of $f(X)$ in $F_f[X]$ correspond to subsets S of $\{\alpha_1, \dots, \alpha_m\}$,

$$S \leftrightarrow f_S = \prod_{\alpha \in S} (X - \alpha),$$

and f_S is fixed under the action of G_f (and hence has coefficients in F) if and only if S is stable under G_f. Therefore the irreducible factors of f in $F[X]$ are the polynomials f_s corresponding to minimal subsets S of $\{\alpha_1, \dots, \alpha_m\}$ stable under G_f, but these subsets S are precisely the orbits of G_f in $\{\alpha_1, \dots, \alpha_m\}$. □

Remark 5.12.2. Note that the proof shows the following: let $\{\alpha_1, \dots, \alpha_m\} = \bigcup O_i$ be the decomposition of $\{\alpha_1, \dots, \alpha_m\}$ into a disjoint union of orbits for the group G_f; then

$$f = \prod f_i, \quad \text{where } f_i = \prod_{\alpha_j \in O_i} (X - \alpha_j),$$

is the decomposition of f into a product of irreducible polynomials in $F[X]$.

Now suppose that F is finite, with p^n elements say. Then G_f is a cyclic group generated by the Frobenius automorphism $\sigma\colon x \mapsto x^{p^n}$. When we regard σ as a permutation of the roots of f, then the orbits of σ correspond to the factors in its cycle decomposition (GT, 4.26). Hence, if the degrees of the distinct irreducible factors of f are $m_1, m_2, \dots, m_r$, then σ has a cycle decomposition of type

$$m_1 + \cdots + m_r = \deg f.$$

Proposition 5.12.3. Let R be a unique factorization domain with field of fractions F, and let f be a monic polynomial in $R[X]$. Let P be a prime ideal in R, let $\bar{F} = R/P$, and let $\bar{f}$ be the image of f in $\bar{F}[X]$. Assume that $\bar{f}$ is separable. Then f is separable, and its roots $\alpha_1, \dots, \alpha_m$ lie in some finite extension R' of R. Their reductions $\bar{\alpha}_i$ modulo PR' are the roots of $\bar{f}$, and $G_{\bar{f}} \subset G_f$ when both are identified with subgroups of $\mathrm{Sym}\{\alpha_1, \dots, \alpha_m\} = \mathrm{Sym}\{\bar{\alpha}_1, \dots, \bar{\alpha}_m\}$.

We defer the proof to the end of this section.

On combining these results, we obtain the following theorem.

Theorem 5.12.4 (Dedekind)**.** Let $f(X) \in \mathbb{Z}[X]$ be a monic polynomial of degree m, and let p be a prime such that $f \mod p$ has simple roots (equivalently, $D(f)$ is not divisible by p). Suppose that $\bar{f} = \prod f_i$ with f_i irreducible of degree m_i in $\mathbb{F}_p[X]$. Then G_f contains an element whose cycle decomposition is of type

$$m = m_1 + \cdots + m_r.$$

Example 5.12.5. Consider $X^5 - X - 1$. Modulo 2, this factors as

$$(X^2 + X + 1)(X^3 + X^2 + 1),$$

and modulo 3 it is irreducible. The theorem shows that G_f contains permutations $(ik)(lmn)$ and (12345), and so also $((ik)(lmn))^3 = (ik)$. Therefore $G_f = S_5$ by (5.10.1).

Lemma 5.12.6. A transitive subgroup of $H \subset S_n$ containing a transposition and an $(n-1)$-cycle is equal to S_n.

Proof. After renumbering, we may suppose that the $(n-1)$-cycle is $(123\dots n-1)$. Because of the transitivity, the transposition can be transformed into (in), some $1 \leqslant i \leqslant n-1$. Conjugating (in) by $(123\dots n-1)$ and its powers will transform it into $(1n)$, $(2n)$, …, $(n-1\,n)$, and these elements obviously generate S_n. □

Example 5.12.7. Select separable monic polynomials of degree n, f_1, f_2, f_3 with coefficients in $\mathbb{Z}$ with the following factorizations:

1. f_1 is irreducible modulo 2;
2. $f_2 = (\text{degree } 1)(\text{irreducible of degree } n-1) \mod 3$;
3. $f_3 = (\text{irreducible of degree } 2)(\text{product of } 1 \text{ or } 2 \text{ irreducible polynomials of odd degree}) \bmod 5$.

Take

$$f = -15f_1 + 10f_2 + 6f_3.$$

Then

(i) G_f is transitive (it contains an n-cycle because $f \equiv f_1 \bmod 2$);

(ii) G_f contains a cycle of length $n-1$ (because $f \equiv f_2 \bmod 3$);

(iii) G_f contains a transposition (because $f \equiv f_3 \bmod 5$, and so it contains the product of a transposition with a commuting element of odd order; on raising this to an appropriate odd power, we are left with the transposition). Hence G_f is S_n.

The above results give the following strategy for computing the Galois group of an irreducible polynomial $f \in \mathbb{Q}[X]$. Factor f modulo a sequence of primes p not dividing $D(f)$ to determine the cycle types of the elements in G_f — a difficult theorem in number theory, the effective Chebotarev density theorem, says that if a cycle type occurs in G_f, then this will be seen by looking modulo a set of prime numbers of positive density, and will occur for a prime less than some bound. Now look up a table of transitive subgroups of S_n with order divisible by n and their cycle types. If this doesn't suffice to determine the group, then look at its action on the set of subsets of r roots for some r.

See, Butler and McKay, *The transitive groups of degree up to eleven*, Comm. Algebra 11 (1983), 863–911. This lists all transitive subgroups of S_n, $n \leqslant 11$, and gives the cycle types of their elements and the orbit lengths of the subgroup acting on the r-sets of roots. With few exceptions, these invariants are sufficient to determine the subgroup up to isomorphism.

PARI can compute Galois groups for polynomials of degree $\leqslant 11$ over $\mathbb{Q}$. The syntax is `polgalois(f)` where f is an irreducible polynomial of degree $\leqslant 11$ (or $\leqslant 7$ depending on your setup), and the output is $(n, s, k,$name$)$ where n is the order of the group, s is $+1$ or -1 according as the group is a subgroup of the alternating group or not, and "name" is the name of the group. For example, `polgalois(X^5-5*X^3+4*X-1)` (see 5.10.4) returns the symmetric group S_5, which has order 120, `polgalois(X^11-5*X^3+4*X-1)` returns the symmetric group S_{11}, which has order 39916800, and
`polgalois(X^12-5*X^3+4*X-1)` returns an apology. The reader should use PARI to check the examples 5.9.3–5.9.6.

See also, Soicher and McKay, *Computing Galois groups over the rationals*, J. Number Theory, 20 (1985) 273–281.

5.12.1 Proof of Proposition 5.12.3

We follow the elegant argument in van der Waerden, Modern Algebra, I, §61.

Let $f(X)$ be a separable polynomial in $F[X]$ and $\alpha_1, \dots, \alpha_m$ its roots. Let $T_1, \dots, T_m$ be symbols. For a permutation σ of $\{1, \dots, m\}$, we let σ_α and σ_T respectively denote the corresponding permutations of $\{\alpha_1, \dots, \alpha_m\}$ and $\{T_1, \dots, T_m\}$.

Let

$$\theta = T_1\alpha_1 + \cdots + T_m\alpha_m$$

and

$$f(X,T) = \prod_{\sigma \in S_m} (X - \sigma_T\theta).$$

Clearly $f(X,T)$ is symmetric in the α_i, and so its coefficients lie in F. Let

$$f(X,T) = f_1(X,T)\cdots f_r(X,T) \tag{5.4}$$

be the factorization of $f(X,T)$ into a product of irreducible monic polynomials. Here we use that $F[X,T_1,\dots,T_m]$ is a unique factorization domain (CA 4.10). The permutations σ such that σ_T carries any one of the factors, say $f_1(X,T)$, into itself form a subgroup G of S_m.

Lemma 5.12.8. The map $\sigma \mapsto \sigma_\alpha$ is an isomorphism from G onto G_f.

Proof. In any F-algebra containing the roots of f, the polynomial $f_1(X,T)$ is a product of factors of the form $X-\sigma\theta$. After possibly renumbering the roots of f, we may suppose that $f_1(X,T)$ contains the factor $X-\theta$. Note that $s_T s_\alpha$ leaves θ invariant, i.e., $s_T s_\alpha \theta = \theta$, and so

$$s_\alpha\theta = s_T^{-1}\theta. \tag{5.5}$$

Let σ be a permutation of $\{1,\dots,m\}$. If σ_T leaves $f_1(X,T)$ invariant, then it permutes its roots. Therefore, it maps $X-\theta$ into a linear factor of $f_1(X,T)$. Conversely, if σ_T maps $X-\theta$ into a linear factor of $f_1(X,T)$, then this linear factor will be a common factor of $f_1(X,T)$ and the image of $f_1(X,T)$ under σ_T, which implies that the two are equal, and so σ_T leaves $f_1(X,T)$ invariant. We conclude that σ_T leaves $f_1(X,T)$ invariant if and only if σ_T maps $X-\theta$ into a linear factor of $f_1(X,T)$.

!!In the third paragraph of the proof of Lemma 4.34, θ is algebraic over the field $F(T) =_{def} F(T_1,\dots,T_m)$ with minimal polynomial equal to $f(X,T)$ (regarded as a polynomial in X with coefficients in the field $F(T)$).!!

Again, let σ be a permutation of $\{1,\dots,m\}$. Then $\sigma_\alpha \in G_f$ if and only if it maps $F(T)[\theta]$ isomorphically onto $F(T)[\sigma_\alpha\theta]$, i.e., if and only if θ and $\sigma_\alpha\theta$ have the same minimal polynomial. The minimal polynomial of θ is $f_1(X,T)$, and so this shows that s_α lies in G_f if and only if σ_α leaves $f_1(X,T)$ invariant, i.e., if and only if σ_α maps $X-\theta$ into a linear factor of $f_1(X,T)$.

From the last two paragraphs and (5.5), we see that the condition for σ to lie in G is the same as the condition for σ_α to lie in G_f, which concludes the proof. □

After these preliminaries, we prove Lemma 5.12.3. With the notation of the lemma, let $R' = R[\alpha_1,\dots,\alpha_m]$. Then R' is generated by a finite number of elements, each integral over R, and so it is finite as an R-algebra (CA 6.2). Clearly, the map $a \mapsto \bar{a}\colon R' \to R'/PR'$ sends the roots of f onto the roots of $\bar{f}$. As the latter are distinct, so are the former, and the map is bijective.

A general form of Proposition **??** shows that, in the factorization (5.4), the f_i lie in $R[X,T]$. Hence (5.4) gives a factorization

$$\bar{f}(X,T) = \bar{f}_1(X,T)\cdots\bar{f}_r(X,T)$$

in $\bar{F}[X,T]$. Let $\bar{f}_1(X,T)_1$ be an irreducible factor of $\bar{f}_1(X,T)$. According to Lemma 5.12.8, G_f is the set of permutations σ_α such that σ_T leaves $f_1(X,T)$ invariant, and $G_{\bar{f}}$ is the set of permutations σ_α such that σ_T leaves $\bar{f}_1(X,T)_1$ invariant. Clearly $G_{\bar{f}} \subset G_f$.

For a monic polynomial f of degree n with bounded integers as coefficients, it is expected that the Galois group of f equals S_n with probability 1 as $n \to \infty$. See Bary-Soroker, Kozma, and Gady, Duke Math. J. 169 (2020), 579–598, for precise statements.

5.13 Exercises

Exercise 5.13.1. Let F be a field of characteristic 0. Show that $F(X^2) \cap F(X^2-X) = F$ (intersection inside $F(X)$). [Hint: Find automorphisms σ and τ of $F(X)$, each of order 2, fixing $F(X^2)$ and $F(X^2-X)$ respectively, and show that $\sigma\tau$ has infinite order.]

Exercise 5.13.2. [4] Let p be an odd prime, and let ζ be a primitive pth root of 1 in $\mathbb{C}$. Let $E = \mathbb{Q}[\zeta]$, and let $G = \mathrm{Gal}(E/\mathbb{Q})$; thus $G = (\mathbb{Z}/(p))^\times$. Let H be the subgroup of index 2 in G. Put $\alpha = \sum_{i\in H}\zeta^i$ and $\beta = \sum_{i\in G\setminus H}\zeta^i$. Show:

1. α and β are fixed by H;

[4] This problem shows that every quadratic extension of $\mathbb{Q}$ is contained in a cyclotomic extension of $\mathbb{Q}$. The Kronecker-Weber theorem says that *every* abelian extension of $\mathbb{Q}$ is contained in a cyclotomic extension.

2. if $\sigma \in G\backslash H$, then $\sigma\alpha = \beta$, $\sigma\beta = \alpha$.

Thus α and β are roots of the polynomial $X^2 + X + \alpha\beta \in \mathbb{Q}[X]$. Compute[5] $\alpha\beta$ and show that the fixed field of H is $\mathbb{Q}[\sqrt{p}]$ when $p \equiv 1 \mod 4$ and $\mathbb{Q}[\sqrt{-p}]$ when $p \equiv 3 \mod 4$.

Exercise 5.13.3. Let $M = \mathbb{Q}[\sqrt{2}, \sqrt{3}]$ and $E = M[\sqrt{(\sqrt{2}+2)(\sqrt{3}+3)}]$ (subfields of $\mathbb{R}$).

1. Show that M is Galois over $\mathbb{Q}$ with Galois group the 4-group $C_2 \times C_2$.
2. Show that E is Galois over $\mathbb{Q}$ with Galois group the quaternion group.

Exercise 5.13.4. Let E be a Galois extension of F with Galois group G, and let L be the fixed field of a subgroup H of G. Show that the automomorphism group of L/F is N/H where N is the normalizer of H in G.

Exercise 5.13.5. Let E be a finite extension of F. Show that the order of $\mathrm{Aut}(E/F)$ divides the degree $[E\colon F]$.

Exercise 5.13.6. Find the splitting field of $X^m - 1 \in \mathbb{F}_p[X]$.

Exercise 5.13.7. Find the Galois group of $X^4 - 2X^3 - 8X - 3$ over $\mathbb{Q}$.

Exercise 5.13.8. Find the degree of the splitting field of $X^8 - 2$ over $\mathbb{Q}$.

Exercise 5.13.9. Give an example of a field extension E/F of degree 4 such that there does not exist a field M with $F \subset M \subset E$, $[M\colon F] = 2$.

Exercise 5.13.10. List all irreducible polynomials of degree 3 over $\mathbb{F}_7$ in 10 seconds or less (there are 112).

Exercise 5.13.11. "It is a thought-provoking question that few graduate students would know how to approach the question of determining the Galois group of, say,

$$X^6 + 2X^5 + 3X^4 + 4X^3 + 5X^2 + 6X + 7."$$

[over $\mathbb{Q}$].

1. Can you find it?
2. Can you find it without using the "`polgalois`" command in PARI?

Exercise 5.13.12. Let $f(X) = X^5 + aX + b$, $a, b \in \mathbb{Q}$. Show that $G_f \approx D_5$ (dihedral group) if and only if

1. $f(X)$ is irreducible in $\mathbb{Q}[X]$, and
2. the discriminant $D(f) = 4^4a^5 + 5^5b^4$ of $f(X)$ is a square, and
3. the equation $f(X) = 0$ is solvable by radicals.

Exercise 5.13.13. Show that a polynomial f of degree $n = \prod_{i=1}^{k} p_i^{r_i}$ (the p_i are distinct primes) is irreducible over $\mathbb{F}_p$ if and only if (a) $\gcd(f(X), X^{p^{n/p_i}} - X) = 1$ for all $1 \leqslant i \leqslant k$ and (b) f divides $X^{p^n} - X$ (Rabin irreducibility test[6]).

Exercise 5.13.14. Let $f(X)$ be an irreducible polynomial in $\mathbb{Q}[X]$ with both real and nonreal roots. Show that its Galois group is nonabelian. Can the condition that f is irreducible be dropped?

Exercise 5.13.15. Let F be a Galois extension of $\mathbb{Q}$, and let α be an element of F such that $\alpha F^{\times 2}$ is not fixed by the action of $\mathrm{Gal}(F/\mathbb{Q})$ on $F^\times/F^{\times 2}$. Let $\alpha = \alpha_1, \ldots, \alpha_n$ be the orbit of α under $\mathrm{Gal}(F/\mathbb{Q})$. Show:

1. $F[\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n}]/F$ is Galois with commutative Galois group contained in $(\mathbb{Z}/2\mathbb{Z})^n$.

[5] Schoof suggests computing $\alpha - \beta$ instead.
[6] Rabin, Michael O. Probabilistic algorithms in finite fields. SIAM J. Comput. 9 (1980), no. 2, 273–280.

2. $F[\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n}]/\mathbb{Q}$ is Galois with noncommutative Galois group contained in $(\mathbb{Z}/2\mathbb{Z})^n \rtimes \mathrm{Gal}(F/\mathbb{Q})$. (Cf. mo113794.)

Chapter 6

Linear Algebra and Representation Theory

We refer to Artin's [1] Chapter10 for a short introduction to group representation, Lang's [7] for a comprehensive one, and also Steinberg's [11] for the finite case

Chapter 7

Commutative Ring Theory

Chapter 8

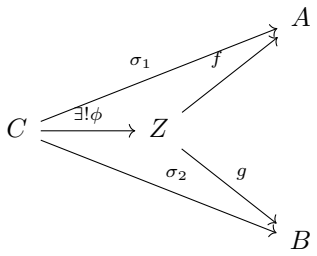
Affine Algebraic Geometry

Chapter 9

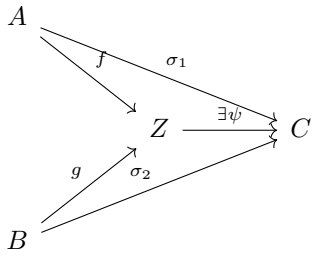
Category Theory

9.1 Product and Coproduct

Definition 9.1.1. Let $\mathcal{C}$ be a category with $A, B \in \mathrm{obj}(\mathcal{C})$. Z is a **product** of A, B if $\exists f \in \mathrm{Hom}(Z, A), g \in \mathrm{Hom}(Z, B)$ such that $\forall C \in \mathrm{obj}(\mathcal{C})$, $\sigma_1 \in \mathrm{Hom}(C, A), \sigma_2 \in \mathrm{Hom}(C, D), \exists! \phi \in \mathrm{Hom}(C, Z)$ s.t. $f \circ \phi = \sigma_1, g \circ \phi = \sigma_2$



Definition 9.1.2. It is a coproduct is the following diagram commutes:



If product (coproduct) of A, B then it is unique up to isomorphism. If Z, Z' coproduct $\psi : Z \to Z', \phi : \mathbb{Z} \to Z$ (replace C with Z' from above). Then $\phi \circ \sigma_2 = g, \psi \circ g = \sigma_2$.

Example 9.1.3. For set A, B, $A \times B$ is the product and the coproduct is the disjoint union $A \sqcup B$. By definition, $\{1, 2\} \sqcup \{2, 3\} = \{1, 2, 2', 3\}$.

Example 9.1.4. For groups G_1, G_2, the product is $G_1 \times G_2$ and the coproduct is free product $G_1 * G_2$ (Note that $G_1 \times G_2$ is only coproduct when it is abelian.)

9.2 Limits

Definition 9.2.1.

Chapter 10

Homological Algebra

Chapter 11

Answer to Selected Problems

Exercises 1.1

1. Exercise 1.1-1

i. Let $a \in G$. By (5), $ya = a$ has a solution $y_0 \in G$. We need that for other $b \in G$, the equation $y_0 b = b$ also holds. This is true because $ax = b$ has a solution $x_0 \in G$, so $b = ax_0 = y_0 a x_0 = y_0(ax_0) = y_0 b$. This shows the existence of a left identity $e_l = y_0$. The existence of a left inverse directly follows from the fact that there is a solution $y \in G$ for $yg = e_l$.

ii. Let a^{-1} be the left inverse of $a \in G$. Let a' be the left inverse of a^{-1}. Thus, $a^{-1}a = e_l$ and $a'a^{-1} = e_l$. On the one hand, $\left(a'a^{-1}\right)\left(aa^{-1}\right) = e_l\left(aa^{-1}\right) = (e_l a)a^{-1} = aa^{-1}$ on the other hand $\left(a'a^{-1}\right)\left(aa^{-1}\right) = a'\left[\left(a^{-1}a\right)a^{-1}\right] = a'\left(e_l a^{-1}\right) = a'a^{-1} = e_l$ so $aa^{-1} = e_l$.

iii. On the one hand, $\left(aa^{-1}\right)a = e_l a = a$. On the other hand, $\left(aa^{-1}\right)a = a\left(a^{-1}a\right) = ae_l$. Thus, $ae_l = a$, which shows that e_l is also the right identity. Therefore, $e_l = e_r = e$, and this in turn elevates "$a^{-1}a = e_l \Rightarrow aa^{-1} = e_l$" to become "$a^{-1}a = e \Rightarrow aa^{-1} = e$."

iv. For the eq. $ax = b$, just take $x = a^{-1}b$ which is in G as $a^{-1} \in G$ and G is closed under multiplication. Similarly, for the eq. $ya = b$, just take $y = ba^{-1} \in G$.

2. Exercise 1.1-6

i. Trivial.

ii. Follows immediately from prop. 1.1.25.

3. Exercise 1.1-7: Let the statement be $p(n)$. We use the strong induction. First we see that $n = 2, 3$ the claim is true. Now assume that for $n \leqslant N-1$ the proposition $p(n)$ is true. To show $p(N)$ is true, we only need to show that for any bracketing $\pi(a_1 \cdot a_2 \cdot \cdots \cdot a_n)$ we have

$$\pi(a_1 \cdot a_2 \cdot \cdots \cdot a_n) = a_1 \cdot (a_2 \cdot \cdots \cdot a_n)$$

where the bracket on the RHS is well-defined by our induction hypothesis. any bracketing $\pi(a_1 \cdot a_2 \cdot \cdots \cdot a_n)$, its last step of computation has to be of the form $b_1 \cdot b_2$ where

$$b_1 = a_1 \cdot a_2 \cdots a_i, b_2 = a_{i+1} \cdot a_{i+2} \cdot \cdots \cdot a_n$$

Since $i, n-i \leqslant N-1$, we by induction hypothesis have them well-defined. To show

$$\begin{aligned}\pi(a_1 \cdot \cdots \cdot a_n) &= (a_1 \cdot a_2 \cdots a_i) \cdot (a_{i+1} \cdot \cdots \cdot a_n) \\ &= a_1 \cdot (a_2 \cdot \cdots \cdot a_n)\end{aligned}$$

we see for $i = 1$ there is nothing to prove, so we assume $i > 1$ and observe

$$\begin{aligned}&\pi(a_1 \cdot a_2 \cdots a_n) \\ &= (a_1 \cdot a_2 \cdot \cdots \cdot a_i) \cdot (a_{i+1} \cdot a_{i+2} \cdot \cdots \cdot a_n) \\ &\overset{IH(i)}{=\!=} (a_1 \cdot (a_2 \cdots a_i)) \cdot (a_{i+1} \cdot a_{i+2} \cdots a_n) \\ &\overset{IH(3)}{=\!=} a_1 \cdot (a_2 \cdots a_i) \cdot (a_{i+1} \cdot a_{i+2} \cdot \cdots \cdot a_n) \\ &\overset{IH(N-1)}{=\!=} a_1 \cdot (a_2 \cdot \cdots \cdot a_i \cdot a_{i+1} \cdot a_{i+2} \cdots a_n) \\ &= a_1 \cdot (a_2 \cdots a_n)\end{aligned}$$

We're done.

4. Exercise 1.1-8: We proceed by weak induction. For $n = 1, 2$ the statement is true. Suppose the statement is true when $n = N-1$. We want to show that permutating $a_1 \cdot a_2 \cdots a_N$, which is $a_{i_1} \cdot a_{i_2} \cdots a_{i_N}$, won't change the result. Suppose the permutation sends N to i_k. Let $\mathbb{C}$ stand

for commutativity and A stand for associativity. Then

$$\begin{aligned}
&a_{i_1}\cdot a_{i_2}\cdot\cdots\cdot a_{i_N}\\
&=\left(a_{i_1}\cdot\cdots\cdot a_{i_{k-1}}\right)\cdot\left[a_{i_k}\cdot\left(a_{i_{k+1}}\cdot\cdots\cdot a_{i_N}\right)\right]\\
&=\left(a_{i_1}\cdot\cdots\cdot a_{i_{k-1}}\right)\cdot\left[a_N\cdot\left(a_{i_{k+1}}\cdot\cdots\cdot a_{i_N}\right)\right]\\
&\overset{C(2)}{=\!=}\left(a_{i_1}\cdot\cdots\cdot a_{i_{k-1}}\right)\cdot\left[\left(a_{i_{k+1}}\cdot\cdots\cdot a_{i_N}\right)\cdot a_N\right]\\
&\overset{A(2)}{=\!=}\left[\left(a_{i_1}\cdot\cdots\cdot a_{i_{k-1}}\right)\cdot\left(a_{i_{k+1}}\cdot\cdots\cdot a_{i_N}\right)\right]\cdot a_N\\
&\overset{\text{Thm1.1.16},A(N-1)}{=\!=\!=\!=}\left(a_{i_1}\cdots a_{i_{k-1}}\cdot a_{i_{k+1}}\cdots a_{i_N}\right)\cdot a_N\\
&\overset{IH}{=\!=}\left(a_1\cdot\cdots\cdot a_{N-1}\right)\cdot a_N\\
&\overset{A(N)}{=\!=}a_1\cdot\cdots\cdot a_{N-1}\cdot a_N
\end{aligned}$$

5. Exercise 1.1-9: let $l=|a^k|:=\min\{m:(a^k)^m=1\}$. Then: (1) $a^{kl}=(a^k)^l=1\Rightarrow kl\geqslant|a|=n=km\Rightarrow l\geqslant m$; (2) $m\geqslant l$ (because $1=a^{km}=(a^k)^m$). They combine to show $l=m$.

6. Exercise 1.1-10: When $n=1$, G is automatically abelian. For $n=2,3,5$ which are primes, G is cyclic and thus abelian. For $n=4$, one can use Cayley table to do the classification to see that G is isomorphic to either $\mathbb{Z}_4$ or the Klein-four group V, both abelian.

7. Exercise 1.1-11: $n=\min\{m:a^m=1\}\Rightarrow k\geqslant n$, $k=np+q\Rightarrow 1=a^k=a^{np+q}=(a^n)^pa^q=a^q$. Since $q<n=\min\{m:a^m=1\}$, we see $q=0$.

8. Exercise 1.1-15: the isomorphism is given by $\phi(x)=y$ and note that isomorphism is bijection.

9. Exercise 1.1-16: We write the distinct cosets of K in G as $\{g_iK\}_{i\in I}$. Thus $G=\amalg_{i\in I}g_iK$. Similarly, we write $K=\amalg_{j\in J}k_jH$. We claim that g_ik_jH are all distinct cosets of H in G. Then, as left cosets form a partition, $[G:H]=\left|\{g_ik_jH\}_{i\in I,j\in J}\right|=|I||J|=[G:K][K:H]$, where we used the fact that $[G:H],[H:K]<\infty$. The claim consists of two parts: (1) every left coset xH of H in G is in $\{g_ik_jH\}_{i\in I,j\in J}$ because it is already clear that each g_ik_jH is a coset of H in G. (2) each g_ik_jH is distinct.
proof of (1): For any left coset xK of K in G, $\exists g_i\in G:xK=g_iK\iff g_i^{-1}x\in H$. Then $g_i^{-1}xH$ is a left coset of subgroup H in K and is one of $\{k_jH\}$: $\exists k_j\in K:g_i^{-1}xH=k_jH\Leftrightarrow\exists h\in H:k_j^{-1}g_i^{-1}x=h$ Thus $x=g_ik_jh$ and $xH=g_ik_jhH=g_ik_jH$.
proof of (2): Suppose not. $g_ik_jH=g_{i'}k_{j'}H$ for some $g_i,k_j,g_i',k_j'\iff(g_ik_j)^{-1}(g_{i'}k_{j'})\in H\subseteq K\Rightarrow g_ik_jK=g_{i'}k_{j'}K\Rightarrow g_iK=g_{i'}K\Rightarrow g_i=g_{i'}$ by distinctiveness in $\{g_iK\}_{i\in I}$. Thus

$$g_i\left(k_jH\right)=g_{i'}\left(k_{j'}H\right)\overset{g_i=g_{i'}}{\Longrightarrow}k_jH=k_{j'}H\Rightarrow k_j=k_{j'}$$

by distinctiveness in $\{k_jH\}_{j\in J}$.

10. Exercise 1.1-17: H has index 2, so there are two left cosets H,aH for some $a\in G$ such that $aH\neq H$, i.e., $a\neq H$. Thus, $a^{-1}\notin H\Rightarrow a^{-1}H\neq H\Rightarrow a^{-1}H=aH\Leftrightarrow(a-1)^{-1}a=a^2\in H$. If $a\in H$, then clearly $a^2\in H$. Therefore, $\forall a\in G$, $a^2\in H$.

11. Exercise 1.1-18: use theorem 1.2.9.

Exercises 1.2

1. Exercise 1.2-6: Since every permutation can be written as product of transpositions, it suffices to show that transpositions can be generated in each of the case. Then note that $(m\ \ k)=(1\ \ m)(1\ \ k)(1\ \ m)$, and

$$\begin{aligned}
(m,k)=&(m,m+d)\\
=&(k-1,k)\dots(m+1,m+2)(m,m+1)\\
&(m+1,m+2)^{-1}\dots(k-1,k)^{-1}\\
=&(k-1,k)\dots(m+1,m+2)(m,m+1)\\
&(m+1,m+2)\dots(k-1,k)
\end{aligned}$$

Thus each of $(i,i+1)$ in the generating set of S_n is further generated by (12) and $(12\cdots n)$, proving the result. For the third claim, just observe that $(i\ i+1)=(1\ 2\ \cdots\ n)^{i-1}(1\ 2)(1\ 2\ \cdots\ n)^{-i+1}$.

2. Exercise 1.2-7: $S_2=\{(1),(1\ 2)\}\cong\mathbb{Z}_2$. Group table of S_3: let $\sigma_0=(1),\sigma_1=(1\ 2\ 3),\sigma_2=(1\ 3\ 2),\sigma_3=(2\ 3),\sigma_4=(1\ 3),\sigma_5=(1\ 2)$

$\circ$	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5
σ_0	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5
σ_1	σ_1	σ_2	σ_0	σ_4	σ_5	σ_3
σ_2	σ_2	σ_0	σ_1	σ_5	σ_3	σ_4
σ_3	σ_3	σ_5	σ_4	σ_0	σ_2	σ_1
σ_4	σ_4	σ_3	σ_5	σ_1	σ_0	σ_2
σ_5	σ_5	σ_4	σ_3	σ_2	σ_1	σ_0

3. Exercise 1.2-9: The permutation ρ has a decomposition as a product of disjoint, hence commuting, (non-trivial) cycles: $\rho=\gamma_1\cdots\gamma_r$. By Question 1.2-iii., The order of ρ is the l.c.m. of the orders of the cycles, so each γ_i has order 3. As

the order of a cycle is its length, this means each γ_i is a 3-cycle.

4. Exercise 1.2-10:

i. $(sr)^2 = 1 \implies (sr)^{-1} = r^{-1}s^{-1} = sr \implies s^{-1}r^{-1}s^{-1} = r \implies s^{-1}r^{-1} = rs \implies (rs)^{-1} = rs$. Vice versa.

ii. $r^k s = sr^{-k}$: start with $(rs)^2 = rsrs = 1 \Leftrightarrow rs = s^{-1}r^{-1} \overset{s^2=e}{=} sr^{-1}$, we see for any $k \in \{0, \cdots, n-1\}$,

$$\begin{aligned} r^k s &= \underbrace{r\cdots rr}_{\#=k} s = \underbrace{r\cdots r}_{\#=k-1} rs \\ &= \underbrace{r\cdots r}_{\#=k-1} sr^{-1} = \underbrace{r\cdots r}_{\#=k-2} (rs)r^{-1} \\ &= \underbrace{r\cdots r}_{\#=k-2} sr^{-1}r^{-1} = \cdots = sr^{-k} \end{aligned}$$

iii. immediately follows from Proposition 1.1.25.

5. Exercise 1.2-11: Let

$$D_n = \left\{e, r, \cdots, r^{n-1}, s, sr, \ldots sr^{n-1}\right\}$$

where r is the rotation and s is the reflection $\left(s^2 = e, r^n = e, (rs)^2 = e\right)$. We note that $H = \left\{e, r, r^2, r^3, \ldots, r^{n-1}\right\} = \langle r\rangle$ is a cyclic subgroup contained in D_n with order n. The complement of it is $H^c = \left\{s, sr, sr^2, sr^3, \ldots sr^{n-1}\right\}$, which has order n as well. Since $H^c = sH$ is the coset of H, H is itself a right coset, and there are no other cosets since they fill the whole group, then the index of H in D_n is 2.

Exercises 1.2

1. Exercise 1.3-4:

i.

$$A^2 = -I,\ A^3 = -A,\ A^4 = I$$

so the order of A in G is 4 .

$$B^2 = -I,\ B^3 = -B,\ B^4 = I$$

so the order of B in G is 4.

ii. We already have six distinct elements $I, -I, A, B, A^3, B^3$ above. G is nonabelian with following two additional elements

$$AB = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix},\ BA = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

iii. By the calculation in **i**, it is obvious that I, A, B, A^3, B^3 don't have order 2, while $-I$ has order 2 as $(-I)^2 = I^2 = I$. We check the rest of the eight:

$$(AB)^2 = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}\begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} = -I \neq I$$

$$(BA)^2 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = -I \neq I$$

Thus, the only element with order 2 is $-I$.

iv. By Lagrange's theorem $8 = |G| = |H|[G : H]$ for subgroup H in G. Hence, except for subgroup $\{e\}$ and G, which are trivial subgroups that are also normal, we only have factorization $8 = 2 \times 4$ or $8 = 4 \times 2$, i.e, $|H| = 2$ with $[G : H] = 4$ or $|H| = 4$ with $[G : H] = 2$. By an example in class that "every subgroup of index 2 in any group is normal" we see subgroup H_1 with $|H_1| = 4$ is normal. The remaining is H_2 with $|H_2| = 2$. Subgroups H_2 with $|H_2| = 2$ must include an identity I and another element x. Counting formula tells us that $2 = |H_2| = |\langle x\rangle|\,[H_2 : \langle x\rangle]$ where $\langle x\rangle$ is the cyclic subgroup generated by x and the order of it is just the order of the element x. The only possible factorization is $2 = 2 \times 1$, so x is an element of order 2 and $H_2 = \langle x\rangle$. For this problem, $x = -I = A^2 = B^2$ by part **i** and part **iii**. Thus, $H_2 = \langle -I\rangle = \{I, -I\}$. To show H_2 is normal in G, we take any $M \in G$ and see that $MIM^{-1} = I \in H_2$; $M(-I)M^{-1} = -MM^{-1} = -I \in H_2$ Therefore, all subgroups of G are normal.

2. Exercise 1.3-5: We want to show that $\forall x \in NH_1, y \in NH_2, yxy^{-1} \in NH_1$. Thus, $x = n_1h_1$ for some $n_1 \in N$ and $h_1 \in H_1$, and $y = n_2h_2$ for some $n_2 \in N$ and $h_2 \in H_2$. Then

$$yxy^{-1} = n_2h_2n_1h_1h_2^{-1}n_2^{-1}$$

Since $h_2n_1 \in NH_2$, we have $h_2n_1h_2^{-1} \in N \Rightarrow \exists n_3 \in N : h_2n_1h_2^{-1} = n_3 \Rightarrow h_2n_1 = n_3h_2$. We call this step exchanging trick, since it gives a new element in the normal subgroup to switch the mul-

tiplication. Thus,

$$\begin{aligned} yxy^{-1} &= n_2n_3h_2h_1h_2^{-1}n_2^{-1} \\ &= \underbrace{n_4}_{=n_2n_3\in N}\ \underbrace{h_2h_1h_2^{-1}}_{=h_1'\in H_1}\ \underbrace{n_5}_{n_2^{-1}\in N} = n_4h_1'n_5 \end{aligned}$$

By the above exchanging trick, we see $h_1'n_5 \in NH_1 \Rightarrow h_1'n_5h_1'^{-1} \in N \Rightarrow \exists n_6 \in N : h_1'n_5h_1'^{-1} = n_6 \Rightarrow h_1'n_5 = n_6h_1'$. Thus,

$$yxy^{-1} = n_4h_1'n_5 = \underbrace{n_4n_6}_{\in N}h_1' \in NH_1$$

3. Exercise 1.3-6: $A_n \longleftrightarrow S_n - A_n$, the set of all odd permuations, by $\sigma \mapsto \sigma(1\ 2)$. Thus, $[S_n : A_n] = 2$, $A_n \trianglelefteq S_n$, and $|A_n| = \frac{1}{2}n!$.

4. Exercise 1.3-12: see [9] Theorem 2.20.

5. Exercise 1.3-13: The relation $x \sim y \iff \exists g \in G$ s.t. $y = x^g := gxg^{-1}$ is reflexive ($x^e = x$); is transitive ($x^g = y, y^h = z \Rightarrow x^{hg} = z$); and is symmetric ($x^g = y \Rightarrow y^{g^{-1}} = x$). Let $H \leqslant G$ be a subgroup. It is normal iff $\forall g \in G,\ gHg^{-1} \subseteq H$, i.e., $\forall h \in H,\ \forall g \in G,\ h^g \in H$, which is just saying that for each $h \in H$, the conjugacy class containing h is contained in H.

Exercises 1.4

Exercises 1.5

1. Exercise 1.5-1.

i. The class equation of G is $|G| = 12 = 1+3+4+4$. The four classes are {e},{(1 2)(3 4), (1 3)(2 4), (1 4)(2 3)},{(1 2 3),(1 4 2),(2 4 3),(1 3 4)},{(1 3 2),(1 4 3),(2 3 4),(1 2 4)}. For a direct derivation without first knowing the result, see Math5031 HW3 Q1 (a).

ii. Let $x \in G$. We first observe a fact: since $Z(G)$ is the set of elements that commute with every element of G and $N(x)$ is the set of elements that commute with x, we get $Z(G) \subseteq N(x)$. Now the center of the group $Z(G)$ is a normal subgroup of G, and we by the counting formula have

$$|G| = |Z(G)|[G : Z(G)] = |Z(G)|n$$

As explained in the first part we by the orbit-stabilizer theorem have

$$|G| = |N(x)||C(x)|$$

for each $x \in G$. above two equations combine to give $|N(x)||C(x)| = |Z(G)|n$ Suppose there is some conjugacy class $C(x)$ with $|C(x)| > n$. Then

$$n|Z(G)| = |N(x)||C(x)| > |N(x)|n \Rightarrow |Z(G)| > |N(x)|$$

which is impossible because $Z(G) \subseteq N(x) \Rightarrow |Z(G)| \leqslant |N(x)|$. Therefore, each conjugacy class has at most n elements.

2. Exercise 1.5-2.

i. We first note that $\sigma^{-1}\rho^{-1}\sigma\rho \in N$ because $\sigma \in N \trianglelefteq A_n \implies \rho^{-1}\sigma\rho \in N$ and $\sigma^{-1} \in N$. We compute

$$\begin{aligned} \sigma^{-1}\rho^{-1}\sigma\rho =& \mu^{-1}(1\ 2\ \dots\ r)^{-1}(1\ 3\ 2)(1\ 2\ \dots\ r)\mu(1\ 2\ 3) \\ \overset{\mu \text{ disjoint; } r\geqslant 4>3}{=\!=\!=}& (1\ 2\ \dots\ r)^{-1}(1\ 3\ 2)(1\ 2\ \dots\ r)(1\ 2\ 3) \\ =& (1\ r\ \dots\ 2)(1\ 3\ 2)(1\ 2\ \dots\ r)(1\ 2\ 3) \\ =& (1\ r\ \dots\ 2)(1\ 3\ 2)(1\ 3\ 2\ 4\ 5\ \dots\ r) \\ =& (1\ r\ \dots\ 2)(3\ 1\ 2\ 4\ 5\ \dots\ r) \\ =& (2\ 3\ r) \end{aligned}$$

Thus N contains a 3-cycle $(2\ 3\ r)$.

ii. Similar to the reasoning in **i**, $x = \sigma^{-1}\rho^{-1}\sigma\rho \in N \trianglelefteq A_n$. We compute

$$\begin{aligned}\sigma^{-1}\rho^{-1}\sigma\rho =&\mu^{-1}(4\ 5\ 6)^{-1}(1\ 2\ 3)^{-1}(1\ 2\ 4)^{-1}\\ &(1\ 2\ 3)(4\ 5\ 6)\mu(1\ 2\ 4)\\ \overset{\mu \text{ disjoint}}{=\!=}&(4\ 6\ 5)(1\ 3\ 2)\,[(1\ 4\ 2)(1\ 2\ 3)]\\ &[(4\ 5\ 6)(1\ 2\ 4)]\\ =&(4\ 6\ 5)(1\ 3\ 2)(2\ 3\ 4)(1\ 2\ 5\ 6\ 4)\\ =&(4\ 6\ 5)(3\ 4\ 1)(1\ 2\ 5\ 6\ 4)\\ =&(3\ 6\ 5\ 4\ 1)(1\ 2\ 5\ 6\ 4)\\ =&(1\ 2\ 4\ 3\ 6)\end{aligned}$$

Then consider $\rho' = (1\ 2\ 4)$ and apply a similar process as **i** to x:

$$\begin{aligned}x^{-1}\rho'^{-1}x\rho' =&(1\ 6\ 3\ 4\ 2)(1\ 4\ 2)\\ &(1\ 2\ 4\ 3\ 6)(1\ 2\ 4) = (2\ 4\ 6)\end{aligned}$$

which is in N as $x \in N \trianglelefteq A_n \Rightarrow \rho'^{-1}x\rho' \in N$ and $\rho'^{-1} \in N$.

iii. In this case $\mu^{-1} = \mu$, so $\mu\mu = 1$. Noticing $\sigma \in N \leqslant A_n$ for the last step, we have

$$\begin{aligned}\sigma^2 =&(1\ 2\ 3)\mu(1\ 2\ 3)\mu\\ \overset{\mu \text{ disjoint}}{=\!=}&(1\ 2\ 3)(1\ 2\ 3)\mu\mu\\ =&(1\ 2\ 3)(1\ 2\ 3) = (1\ 3\ 2) \in N\end{aligned}$$

iv. We compute

$$\begin{aligned}\eta =&\sigma^{-1}\rho^{-1}\sigma\rho\\ =&\mu^{-1}(3\ 4)(1\ 2)(1\ 3\ 2)(1\ 2)(3\ 4)\mu(1\ 2\ 3)\\ \overset{\mu \text{ disjoint}}{=\!=}&(1\ 4)(2\ 3)\end{aligned}$$

and $\zeta = (1\ 5\ 2)\eta(1\ 2\ 5) = (1\ 3)(4\ 5)$. Similar to the reasoning in **i**, we see $\eta \in N$ as $\sigma \in N \trianglelefteq A_n \implies \rho^{-1}\sigma\rho \in N$ and $\sigma^{-1} \in N$. Besides, $\zeta = (1\ 5\ 2)\eta(1\ 2\ 5) = (1\ 5\ 2)\eta(1\ 5\ 2)^{-1} \in N$. Lastly, we observe that $\eta\zeta = (1\ 2\ 3\ 4\ 5)$. This then converts to case **i** for $r = 5$. Thus $(2\ 3\ r) = (2\ 3\ 5)$ is in N.

3. Exercise 1.5-3. We first see two facts: (1) every 3 -cycle (i, j, k) with $i \leqslant j \leqslant k$ is a commutator in 2 -cycles:

$$\begin{aligned}(i,j,k) =&(i,k)(i,j)\\ =&(i,j)(i,k)(i,j)(i,k)\\ =&[(i,j),(i,k)]\end{aligned}$$

(2) A_n is generated by 3-cycles (proved in Example 1.5.3). Immediately from (1) and (2), we see every element of A_n is a product of commutators. We then only need to show that every product of commutators is some element in A_n: each commutator is of the form $[x, y]$ where $x \in S_n, y \in S_n$ can be written as product of transpositions, i.e., $x = \sigma_1\sigma_2\cdots\sigma_k, y = \tau_1\tau_2\cdots\tau_l$ for some integers k, l. We then compute:

$$\begin{aligned}[x,y] = xyx^{-1}y^{-1} =&\sigma_1\sigma_2\cdots\sigma_k\tau_1\tau_2\cdots\tau_l\\ &(\sigma_1\sigma_2\cdots\sigma_k)^{-1}(\tau_1\tau_2\cdots\tau_l)^{-1}\\ =&\sigma_1\sigma_2\cdots\sigma_k\tau_1\tau_2\cdots\tau_l\sigma_k\cdots\sigma_2\sigma_1\tau_l\cdots\tau_2\tau_1\end{aligned}$$

There are in total $2(k + l)$ transpositions. Since $2(k+l)$ is even and products of even permutations are still even permutations, making products of commutators belong to A_n.

4. Exercise 1.5-4. Part one is trivial. Part two: First of all, $A_\infty = \cup_{n\geqslant 1}A_n = \bigcup_{n\geqslant 5}A_n$ simply because $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_5 \subseteq A_6 \cdots$. To show that A_∞ is simple, we need to show that each $N \trianglelefteq A_\infty$ has to be trivial or the whole A_∞. First notice that each A_n is a group and thus a subgroup of the group A_∞, i.e., $A_n \leqslant A_\infty$. Then $N \cap A_n \trianglelefteq A_n$ due to the 2^{nd} isomorphism theorem. When $n \geqslant 5$, this normal subgroup $N \cap A_n$ must be $\{e\}$ or A_n due to the simplicity, i.e., A_n is simple for all $n \geqslant 5$. We analyze the two cases: If $N \cap A_n = A_n$ for some $n \geqslant 5$, then $A_n \subseteq N$. Then for all $m \geqslant n, A_n \subseteq N \cap A_m \Rightarrow N \cap A_m \neq \{e\} \Rightarrow N \cap A_m = A_m \Rightarrow A_\infty = \bigcup_{i\geqslant 5}A_i = \bigcup_{i\geqslant n}A_i = \bigcup_{i\geqslant n}N \cap A_i = N \cap (\bigcup_{i\geqslant n}A_i) \Rightarrow A_\infty \subseteq N$ But $N \trianglelefteq A_\infty \Rightarrow N \subseteq A_\infty$, so $A_\infty = N$. If $N \cap A_n = \{e\}$ for some $n \geqslant 5$. Then for all $m \geqslant n, N \cap A_m$ cannot be A_m as for if $A_m = N \cap A_m$ then $A_n \subseteq A_m = N \cap A_m \Rightarrow A_n \subseteq N \Rightarrow N \cap A_n = A_n \neq \{e\}$ which is a contradiction. Thus, for all $m \geqslant n, N \cap A_m = \{e\}$. Thus, $N = N \cap A_\infty = N \cap (\cup_{i\geqslant 5}A_i) = N \cap N \cap (\cup_{i\geqslant n}A_i) = \cup_{i\geqslant n}N \cap A_i = \cup_{i\geqslant n}\{e\} = \{e\}$. Thus, N is either trivial or the whole group, proving the simplicity of A_∞.

Exercises 1.7

1. Exercise 1.7-1. $36 = 3^2 \times 2^2$.

Let $r = \#$ of Sylow 3-subgroup; $s = \#$ of Sylow 2-subgroup. Then Ttird Sylow theorem implies that $r \,|2^2, 3|\, r-1$, so $r = 1$ (we're done) or $r = 4$; $s \,|3^2, 2|\, s-1$, so $s = 1$ (we're done) or $s = 3$. For $r = 4$ we let $X = \{H_1, H_2, H_3, H_4\}$ be the set of Sylow 3-subgroups, each of which has order $3^2 = 9$. Consider the action of G on X by conjugation, which gives rise to a homomorphism $\phi : G \to S_X$ by sending each g to the permutation defined by multiplication by g. We claim that $Ker(\phi)$ is a nontrivial normal subgroup of G. It is normal. It does not equal to G : second Sylow Theorem implies that $G \overset{\text{conj}}{\curvearrowright} X$ is transitive $\Rightarrow Ker(\phi) \neq G$. It does not equal to $\{e\}$: first Isomorphism theorem implies that $\frac{G}{Ker(\phi)} \cong \operatorname{Im}(\phi) \leqslant S_X$ Since the order of the permutation group of a set with 4 elements $|S_X|$ is $4! = 24$, we see $\left|\frac{G}{Ker(\phi)}\right| = [G : Ker(\phi)] \leqslant 24 < 36 = |G| \Rightarrow Ker(\phi) \neq \{e\}$.

2. Exercise 1.7-2. $48 = 2^4 \times 3$.

Let $r = \#$ of Sylow 2-subgroup; $s = \#$ of Sylow 3-subgroup. Then third Sylow theorem implies that $r|3, 2|r-1$, so $r = 1$ (we're done) or $r = 3$; $s \,|2^4, 3|\, s-1$, so $r = 1$ (we're done) or $s = 4$ or $s = 16$. Sylow 3-subgroups have prime order and trivial intersection. Sylow 2 -subgroups have order 16 with at most 8 elements in common. Then if $s = 16$ we get, by a similar argument of distinct element counting used before,

$$|G| = 48 \geqslant 1 + 16(3-1) + (16-1) + 8 = 56$$

Contradiction, so $s \neq 16$. Suppose $s = 4$. Then we will have a similar argument used for $|G| = 24$ and $|G| = 36$. $G \overset{\text{conj}}{\curvearrowright} X = \{H_1, H_2, H_3, H_4\}$ gives rise to a homomorphism $\phi : G \to S_X$. Second Sylow Theorem shows that $G \overset{\text{conj}}{\curvearrowright} X$ is transitive, so $Ker(\phi) \neq G$, and $\left|\frac{G}{Ker(\phi)}\right| = |\operatorname{Im}(\phi)| \leqslant |S_X| = 24 \Rightarrow |Ker(\phi)| \neq \{e\}$. Thus, $Ker(\phi)$ is a proper normal subgroup of G.

3. Exercise 1.7-3. $40 = 2^3 \times 5$.

Let $r = \#$ of Sylow 2-subgroup; $s = \#$ of Sylow 5-subgroup. Then third Sylow theorem implies that $r|5, 2|r-1$, so $r = 1$ (we're done) or $r = 5$; $s \,|2^3, 5|\, s-1$, but then among $1, 2, 4, 8$, only $s = 1$ satisfies $5 \mid s-1$.$s = 1$ implies that we have only one Sylow 5-subgroup which is then normal.

4. Exercise 1.7-4. $56 = 2^3 \times 7$.

Let $r = \#$ of Sylow 2-subgroup; $s = \#$ of Sylow 7-subgroup. Then third Sylow theorem implies that $r|7, 2|r-1$, so $r = 1$ (we're done) or $r = 7$; $s \,|2^3, 7|\, s-1$, so $r = 1$ (we're done) or $s = 8$. Among the two Sylow subgroups, we have one of them only having a prime order, which is the Sylow 7-subgroups H_i 's, so we can apply the observation that subgroup of prime orders have only trivial intersection to get $H_i \cap H_j = \{e\}$. However, Sylow 2-subgroups K_i 's have order 8 which is not a prime number. Instead $|K_i \cap K_j| \,||K_i| = 8 \Rightarrow |K_i \cap K_j|$ is at most 4 (including e) for distinct i and j. Besides, 7 and 8 are coprime, so K 's and H 's intersect trivially. We take two of the K' 's, say K_1 and K_2, they in total add at least $(8-1)+4$ elements to G :

$$56 = |G| \geqslant 1 + 8(7-1) + (8-1) + 4 = 60$$

A contradiction. Thus, either $r \neq 7 \Rightarrow r = 1$ (we're done) or $s \neq 8 \Rightarrow s = 1$ (we're done).

5. Exercise 1.7-5. We recall the following rules:

1. $|G| = pq$ with p and q distinct primes is not simple (see Corollary 1.7.14);
2. $|G| = pq^2$ with p and q distinct primes is not simple (see Proposition 1.7.15);
3. $|G| = pqr$ with p, q, r distinct primes is not simple (see Proposition 1.7.16);
4. $|G| = p^r$ with p prime and integer $r \geqslant 1$ is not simple (see Corollary 1.6.17);
5. $|G| = pq^r$ with $p < q$ distinct primes is not simple (see Corollary 1.7.13).

It can be easily checked by prime factor decomposition of the orders that only G with $|G| = 36, 40, 48, 56$ cannot be proved to be non-simple using above rules, but we already proved them separately in previous exercises.

6. Exercise 1.7-6. We review our five criteria in the Exercise 1.7-5: (4): $|G| = p^r$ with p prime and integer $r \geqslant 1$ is solvable (see Corollary 1.6.19); (5): $|G| = pq^r$ with $p < q$ distinct primes: the proper normal subgroup N we found in Corollary 1.7.13 is a Sylow q-subgroup. N has order q^n so by (4) it is solvable. Since G/N has order p which is prime we see G/N is cyclic, abelian,

and solvable. Then G is solvable due to Proposition 1.5.15. (1): $|G| = pq$ with p and q distinct primes: special case of (5); (2): $|G| = pq^2$ with p and q distinct primes: when $p < q$ this is a special case of (5); when $p > q$, the proper normal subgroup N we found in Proposition 1.7.15 is a Sylow q-subgroup. N is solvable by (4) and G/N is cyclic, abelian, and solvable, so G is solvable. (3): $|G| = pqr$ with p, q, r distinct primes: again, by Proposition 1.7.16, we get a Sylow subgroup N of p, q, or r, which is a prime group and is solvable. $\left|\frac{G}{N}\right|$ is a product of two primes, so $\frac{G}{N}$ is solvable by (1). Proposition 1.5.15 then concludes that G is solvable. Therefore, all the groups checked to be non-simple by these rules are solvable. We again only need to check G with $|G| = 36, 40, 48, 56$, but this is straightforward: their normal subgroups and factor groups we found when proving their non-simplicity have orders smaller than theirs and are thus shown to be solvable.

Exercises 2.2

1. Ex2.2-1. First two questions are trivial. Last two:

(c) Inspired by part (d), we can let $I = J$ so that the sufficient condition is at least unsatisfied. Let $R = \mathbb{Z}$. All the ideals in $\mathbb{Z}$ are $m\mathbb{Z}$, so let $I = J = 5\mathbb{Z}$. Then $IJ = \{\sum_{i=1}^{n} a_i b_i : a_i \in I, b_i \in J\} = \{\sum_{i=1}^{n} (5k_i)(5l_i) = 25\sum_i^n k_i l_i : k_i, l_i \in \mathbb{Z}\} = 25\mathbb{Z}$. $I \cap J = I = 5\mathbb{Z} \neq 25\mathbb{Z}$. (d)

(d) By part (b), it suffices to show $I \cap J \subseteq IJ$. Since $I + J = R$, in particular $1 = i + j$ for some $i \in I, j \in J$. Then let $a \in I \cap J$ and by commutativity of R see that $a = 1a = i\underset{\in J}{a} + \underset{\in I}{a}j \in IJ$.

2. Ex2.2-2.

i. $\sqrt{I}$ contains I as $x^1 = x \in I$. We show that $\sqrt{I}$ is an ideal of R. - $(\sqrt{I}, +) \leqslant (R, +)$:
Let $x \in \sqrt{I}$, so $x^n \in I$ for some $n \geqslant 1$. Then $-x^n \in (I, +) \leqslant (R, +)$ and thus $(-x)^n$, which is either x^n or $-x^n$, is in I. Thus $-x \in \sqrt{I}$. Let $x, y \in \sqrt{I}$, so $x^m, y^n \in I$ for some $n, m \geqslant 1$. Observe that

$$(x+y)^{n+m} = \sum_{i=0}^{n+m} C_{m+n}^i x^{m+n-i} y^i$$

$$= \sum_{i=0}^{n} C_{m+n}^i x^{m+n-i} y^i + \sum_{i=n+1}^{n+m} C_{m+n}^i x^{m+n-i} y^i$$

$$= \sum_{i=0}^{n} C_{m+n}^i x^{m+n-i} y^i + \sum_{i=1}^{m} C_{m+n}^{n+i} x^{m-i} y^{n+i}$$

$$= \underbrace{x^m}_{\in I} \underbrace{\sum_{i=0}^{n} C_{m+n}^i x^{n-i} y^i}_{\in R} + \underbrace{y^n}_{\in I} \underbrace{\sum_{i=1}^{m} C_{m+n}^{n+i} x^{m-i} y^i}_{\in R}$$

is in the ideal I because each binomial coefficient is an integer, so $x + y \in \sqrt{I}$. - $\forall r \in R, a \in \sqrt{I}$ we have $ar \in \sqrt{I}$:
This is because $a \in \sqrt{I} \Rightarrow \exists n \geqslant 1, a^n \in I \Rightarrow (ra)^n \overset{R \text{ commutative}}{=\!=\!=} r^n a^n \in I$ since I is an ideal. Thus, $ra \in \sqrt{I}$. Therefore, $\sqrt{I}$ is an ideal containing I in R.

ii. Let I and J be two ideals in R. $\sqrt{IJ} \subseteq \sqrt{I \cap J}$: suppose $x \in \sqrt{IJ}$, then $x^n \in IJ$ for some $n \geqslant 1$. We proved in part (b) of last exercise that $IJ \subseteq I \cap J$, so $x^n \in I \cap J$. Then $x \in \sqrt{I \cap J}$. $\sqrt{I \cap J} \subseteq \sqrt{IJ}$: suppose $x \in \sqrt{I \cap J}$, then $x^n \in I \cap J$ for some $n \geqslant 1$. Then $x^{2n} = \underbrace{x^n}_{\in I} \underbrace{x^n}_{\in J} \in IJ$, so $x^{2n} \in IJ$ and $x \in \sqrt{IJ}$.

Exercises 2.3

1. Ex2.4-1. Let the finite commutative ring be R and the prime ideal be I. let $I \subseteq M \subseteq R$ where M is an ideal. If $I \neq M$, that is there is some $x \in M$ not in I, then we want to show that $M = R$, which proves that I is a maximal ideal by definition. Let $J = \langle x, I \rangle = \{rx + i \mid r \in R, i \in I\}$, which is an ideal as we have shown in class, so $I \subseteq J \subseteq R$. Consider the set $S = \{1, x, x^2, \cdots\}$, which as a subset of R should be finite. Thus, elements in S cannot be all distinct, i.e., there are $x^n = x^m$ for some $n < m$. Then observe that

$$x^n \left(1 - x^{m-n}\right) = x^n - x^m = 0 \in I$$

Since I is a prime ideal in R, we see x^n is in I or $1-x^{m-n}$ is in I. We claim that x^n cannot be in I : since $x \notin I$, we see $x^2 \notin I$ because for if $x^2 \in I$ then $x \in I$ (or $x \in I$, which is a duplicate), and $x^3 \notin I$ because for if $x^3 \in I$ then $x \in I$ or $x^2 \in I$. Thus inductively we can show that $x^n \notin I$. Therefore, we are left with $1-x^{m-n} \in I$. Thus, $1 = \underbrace{x^{m-n-1}}_{\in R} x + \underbrace{1-x^{m-n}}_{\in I}$ is an element in J. Then $1 \in J \Rightarrow J = R$. Note that J is the smallest ideal containing I and an element x not in I. Therefore. $x \in M, I \subseteq M \Rightarrow J \subseteq M \overset{J=R}{\Longrightarrow} M = R$.

2. Ex2.4-2.

i. Ideals in $\mathbb{Z}$ are of the form $m\mathbb{Z}$ for integer m. Since primary ideal needs to be proper, we have $m \neq 1$. Next, we claim that $m\mathbb{Z}$ is primary iff $m = p^n$ for some prime number p and some positive integer n. $\Leftarrow$: let $ab \in m\mathbb{Z} = p^n\mathbb{Z}$, then $p^n \mid ab$. Thus $p \mid a$ or $p \mid b$. There are three cases: (1) p does not divide a, then $p^n \mid b$, so $b \in p^n\mathbb{Z}$; (2) p does not divide b, then $p^n \mid a$, so $a \in p^n\mathbb{Z}$; (3) $p \mid a$ and $p \mid b$, then $p^n \mid a^n$ and $p^n \mid b^n$. Therefore $ab \in m\mathbb{Z} = p^n\mathbb{Z} \Rightarrow a \in p^n\mathbb{Z}$ or $b^k \in p^n\mathbb{Z}$ for some positive integer k. $\Rightarrow$: Suppose $m\mathbb{Z}$ is primary. Suppose m is not of the form p^n. Then the prime decomposition of m has at least a q^k as a factor where q is another prime and k is also a positive integer. We will first dela with the case that $m = p^n q^k$ and then see that the general case is similar to the two-factor case. Now just let $a = p^n q^{k-1}$ and $b = q$. Then

$$ab = p^n q^{k-1} q = p^n q^k = m \in m\mathbb{Z}$$

Since $m = p^n q^k$ does not divide $p^n q^{k-1} = a$, so $a \notin m\mathbb{Z}$. We show it also happens that any power of b is also not in $m\mathbb{Z}$ too, which then gives a contradiction to the fact that $m\mathbb{Z}$ is a primary ideal. Let this power be l and observe that $m = p^n q^k$ does not divide $q^l = b^l$, i.e., b^l cannot be a multiple of m and thus does not belong to $m\mathbb{Z}$.
The general case where $m = p^n p_1^{k_1} \cdots p_r^{k_r}$ is similar: let $a = p$ and $b = p^{n-1} p_1^{k_1} \cdots p_r^{k_r}$

ii. Let $ab \in \sqrt{I} = \{x \in R \mid \exists n \geqslant 1, \text{ s.t. } x^n \in I\}$. Thus $(ab)^n \in I$ for some $n \geqslant 1$. $(ab)^n = a^n b^n$ by commutativity of R we assumed in this hw. I being primary implies that either $a^n \in I$ or $b^{nk} \in I$ for some positive integer k. We have $a \in \sqrt{I}$ or $b \in \sqrt{I}$.

Exercises 2.4

Exercises 2.5

1. Ex2.5-1. (1): Let $f : A \to A'$ be a ring homomorphism and I be a prime ideal of A'. Suppose that $xy \in f^{-1}(I)$. Then $f(xy) = f(x)f(y) \in f\left(f^{-1}(I)\right) \subset I$. Since I is prime, $f(x) \in I$ or $f(y) \in I$, thus $x \in f^{-1}(I)$ or $y \in f^{-1}(I)$. Hence $f^{-1}(I)$ is prime.

(2): Let $f : A \to A'$ be a surjective ring homomorphism and let I be a proper ideal of A'. We know that $f^{-1}(I)$ is an ideal by the above. Suppose that $f^{-1}(I)$ is not proper, that is, $f^{-1}(I) = A$. Then $f\left(f^{-1}(I)\right) = f(A) = A'$, but $f\left(f^{-1}(I)\right) = I$ (this equality follows from surjectivity of f), so this is a contradiction as we assumed I is proper.

2. Ex2.5-3. Let J be an ideal of $S^{-1}R$. We have shown in Ex2 that $J = S^{-1}\phi^{-1}(J)$ for the map $\phi : r \mapsto \frac{r}{1}$. Since ϕ is a ring homomorphism, $\phi^{-1}(J)$ is an ideal in R, so $S^{-1}\phi^{-1}(J)$ is an ideal by exercise 1 . Since R is PID, $\phi^{-1}(J) = (a)$ for some $a \in R$, so $J = S^{-1}\phi^{-1}(J) = \left(\frac{a}{1}\right)$ because

- For any ar $\in \phi^{-1}(J) = (a)$ and $s \in S$ we have $\frac{ar}{s} = \frac{a}{1}\frac{r}{s}$ and $\frac{r}{s} \in S^{-1}R$;

- and for any $\frac{r}{s} \in S^{-1}R$ we have $\frac{a}{1}\frac{r}{s} = \frac{ar}{s} \in S^{-1}\phi^{-1}(J)$.

Exercises 2.6

1. Ex2.7-1. Consider the ideal $I = (3, 2+\sqrt{-5})$. Define

$$\begin{aligned} d : \mathbb{Z}[\sqrt{-5}] &\to \mathbb{Z}_+ \\ a + b\sqrt{-5} &\mapsto a^2 + 5b^2 \end{aligned}$$

The function is multiplicative:

$$\begin{aligned} & d((a+b\sqrt{-5})(c+d\sqrt{-5})) \\ =& d(ac-5bd+(bc+ad)\sqrt{-5}) \\ =& (ac-5bd)^2 + 5(bc+ad)^2 \\ =& \left(a^2+5b^2\right)\left(c^2+5d^2\right) \\ =& d(a+b\sqrt{-5})d(c+d\sqrt{-5}) \end{aligned}$$

If $I = (x) = (a+b\sqrt{-5})$ for some $a, b \in \mathbb{Z}$, we have $3 = rx, 2+\sqrt{-5} = tx$ for some $r, t \in \mathbb{Z}[\sqrt{-5}]$. Then $d(3) = d(rx) = d(r)d(x) \Rightarrow d(x) \mid 9 = d(3)$ and $d(2+\sqrt{-5}) = d(tx) = d(t)d(x) \Rightarrow d(x) \mid 9 =$

$d(2+\sqrt{-5})$. Thus, $d(x) \mid 9 \Rightarrow d(x) = 1, 3$, or 9 . Since x is an element in R, we suppose $x = x_1 + x_2\sqrt{-5}$ for $x_1, x_2 \in \mathbb{Z}$. Then $d(x) = x_1^2 + 5x_2^2$. As x_1^2 and x_2^2 are $0, 1, 4\cdots$, so $x_1^2 + 5x_2^2$ can be $0, 1, 4, 5, 9, \cdots$. Thus, $d(x) = 3$ is impossible. $d(x) = 9$ happens when $x_1 = \pm 2$ and $x_2 = \pm 1$ or when $x_1 = \pm 3$ and $x_2 = 0$. When $d(x) = 9$, we see $9 = d(r)d(x) \Rightarrow d(r) = 1, r = \pm 1 \overset{3=rx}{\Longrightarrow} x = \pm 3 \Rightarrow 2 + \sqrt{-5} = \pm 3t$, which is impossible because $2+\sqrt{-5}$ is indivisible by 3 and -3 . Therefore, $d(x) = 1 \Rightarrow x = \pm 1 \Rightarrow (3, 2+\sqrt{-5}) = (x)$ is the whole $\mathbb{Z}[\sqrt{-5}]$. Then $1 \in \mathbb{Z}[\sqrt{-5}] \Rightarrow$ there are $\alpha, \beta \in \mathbb{Z}[\sqrt{-5}]$ s.t.

$$\begin{gathered} 3\alpha + \beta(2+\sqrt{-5}) = 1 \\ (2-\sqrt{-5})(3\alpha + \beta(2+\sqrt{-5})) = 2 - \sqrt{-5} \\ 3(2+\sqrt{-5})\alpha + 9\beta = 2 - \sqrt{-5} \\ 3[(2+\sqrt{-5})\alpha + 3\beta] = 2 - \sqrt{-5} \end{gathered}$$

It then follows that $2 - \sqrt{-5}$ is divisible by 3 , which is a contradiction. Therefore, $I = (3, 2 + \sqrt{-5})$ is not principal.

2. We show $\mathbb{Z}[2i]$ is not UFD by giving the counterexample hinted above:

$$4 = 2 \cdot 2 = (-2i) \cdot (2i)$$

while - $2, 2i, -2i$ are irreducible elements. - 2 and $2i$ are not associates; 2 and $2i$ are not associates.

Let d be the map similarly defined in the last problem:

$$\begin{aligned} d : \mathbb{Z}[2i] = \mathbb{Z}[\sqrt{-4}] &\to \mathbb{Z}_+ \\ a + 2bi &\mapsto a^2 + 4b^2 \end{aligned}$$

Similar to the proof given in last problem, d is multiplicative. We observe that $d(2) = d(2i) = d(-2i) = 4$, so we show that all elements a with $d(a) = 4$ are all irreducible. Suppose $a = bc$ and a is not a unit and $d(a) = 4$. Then $4 = d(a) = d(b)d(c) = 1 \times 4$ or 4×1 or 2×2. Since $a^2 + 4b^2$ can only be $0, 1, 4$ or greater than 4 we see 2×2 is impossible, so either 1×4 or $4 \times 1 \Rightarrow d(b) = 1$ or $d(c) = 1 \Rightarrow b = \pm 1$ is a unit or $c = \pm 1$ is a unit.

Note that a and b associate iff $a = bu$ for a unit u. Since $u = \pm 1$ in $\mathbb{Z}[2i]$ (we have shown this fact for $\mathbb{Z}[\sqrt{-5}]$ and this is similarly true for $\mathbb{Z}[2i]$), we see only a and $-a$ are associate (if not equal). Thus, 2 and $2i$ are not associates; 2 and $2i$ are not associates.

Exercises 2.9

1. Ex2.9-1.

i. The general algorithm is provided in class by two steps: divided by integer and divided by arbitrary $x \in \mathbb{Z}[i]$. To divide α by β we first set $n = d(\beta) = 65$ and divide $\alpha\bar{\beta} = (11+3i)(1-8i) = 35-85i$ by n, which by the algorithm is just dividing real and imaginary parts by n separately ($35 = 65 \cdot 1 - 30$ with $|-30| = 30 < \frac{65}{2} = 32.5$; $-85 = 65\cdot(-1) - 20$ with $|-20| = 20 < \frac{65}{2}$) to get $\alpha\bar{\beta} = nq + s = 65(1+i(-1)) + (35 + i(-20))$. Thus, $q = 1 - i$ and

$$\begin{aligned} \alpha &= q\beta + (\alpha - q\beta) \\ &= (1-i)(1+8i) + (11 + 3i - (1-i)(1+8i)) \\ &= \underbrace{(-i)}_{q}(1+8i) + \underbrace{(2-4i)}_{r} \end{aligned}$$

with $20 = d(r) < d(\beta) = 65$.

ii.

$$\alpha = \underbrace{(1-i)}_{q_0}\underbrace{(1+8i)}_{\beta} + \underbrace{(2-4i)}_{r_0}$$

with $\gcd(\alpha, \beta) = \gcd(1-i, 3+4i)$.

$$\beta = q_1(2-4i) + r_1$$

To compute q_1 and r_1 we divide $\beta\overline{r_0} = (1+8i)(2+4i) = -30+20i$ by $d(2-4i) = 20$: $-30 = 20\cdot(-1) - 10$ with $10 \leqslant \frac{20}{2}$ and $20 = 20\cdot 1 + 0$ with $0 \leqslant \frac{20}{2}$. So, $q_1 = -1+i$. $r_1 = 1+8i-(2-4i)(-1+i) = -1+2i$.

$$\beta = \underbrace{(-1+i)}_{q_1}\underbrace{(2-4i)}_{r_0} + \underbrace{(-1+2i)}_{r_1}$$

with $\gcd(\beta, r_0) = \gcd(r_0, r_1)$

$$r_0 = q_2(-1+2i) + r_2$$

To compute q_2 and r_2 we divide $r_0\overline{r_1} = (2-4i)(-1-2i) = -10$ by $d(-1+2i) = 5$: $-10 = 5\cdot(-2)+0$. So, $q_2 = -2$. $r_2 = 2-4i-(-2)(-1+2i) = 0$.

$$r_0 = \underbrace{(-2)}_{q_2}\underbrace{(-1+2i)}_{r_1} + \underbrace{0}_{r_2}$$

with $\gcd(r_0, r_1) = \gcd(r_1, r_2)$. Thus,

$$\begin{aligned}\gcd(\alpha, \beta) &= \gcd(r_0, r_1) = \gcd(r_1, r_2)\\ &= \gcd(-1+2i, 0) = -1+2i\end{aligned}$$

Bibliography

[1] Artin, Michael. *Algebra*, Pearson, 2nd edition, 2010.

[2] Burton, David M. *A First Course in Rings and Ideals*, Addison-Wesley, 1970.

[3] Dummit, David Steven, and Foote, Richard M. *Abstract Algebra*, Wiley, 2004.

[4] Milne, James S. *Fields and Galois theory*, Courses Notes, Version 5 (2021).

[5] Judson, Thomas. *Abstract Algebra: Theory and Applications*, Virginia Commonwealth University Mathematics, 2009.

[6] Kurzweil, Hans, and Stellmacher, Bernd. *The Theory of Finite Groups: An Introduction*, Springer, 2004.

[7] Lang, Serge. *Algebra*, Springer Science+Business Media, 2012.

[8] Mac Lane, Saunders. *Categories for the Working Mathematician*, Springer Science+Business Media, 2013.

[9] Rotman, Joseph J. *An Introduction to the Theory of Groups*, Springer Science+Business Media, 1995.

[10] Rotman, Joseph J. *An Introduction to Homological Algebra*, Springer Science+Business Media, 2009.

[11] Steinberg, Benjamin. *Representation Theory of Finite Groups: An Introductory Approach*, Springer, 2012.