GROUPS AND RINGS: GROUPS

ALEXEI N. SKOROBOGATOV

Contents

This is the first part of the course. The reader should be familiar with the definitions of a group, of a subgroup, of an abelian group, and with the key examples of finite groups such as cyclic groups C_n and the symmetric groups S_n.

1. Homomorphisms and normal subgroups

1.1. **Homomorphisms, isomorphisms and automorphisms.** A map from a set X to a set Y is a function $f\colon X\to Y$. This is a rule that associates to each element $x \in X$ an element $f(x) \in Y$.

Let G be a group with unit element e. We write the group law $G \times G\to G$ multiplicatively, that is, the group operation sends $g, h \in G$ to $g \cdot h \in G$. The dot between g and h often will be dropped. Recall the axioms of a group:

the *unit element* $e \in G$ satisfies $e \cdot g = g \cdot e = g$ for any $g \in G$;

every element has an *inverse*: for any $g \in G$ there exists an element denoted by $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$;

associativity: for any $a, b, c \in G$ we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

It is known from the first year that the unit element is unique. (If you don't remember this, deduce it from the first axiom.) Given $g \in G$, the inverse g^{-1} is unique. (If you don't remember this, prove it.)

When considering functions $f\colon G\to H$, where G and H are groups, it makes sense to distinguish those respecting the group structure. This means that f should preserve the group law, send the unit element in G to the unit element in H, and send the inverse of each element of G to the inverse of its image.

Date: September 14, 2021.

Definition 1.1. *A function $f : G \to H$ is called a* **homomorphism** *if for any $a, b \in G$ we have $f(ab) = f(a)f(b)$.*

Note that here ab is computed in G, whereas $f(a)f(b)$ is computed in H. So f transforms the group law of G into the group law of H. We do not need to specify that f preserves the unit element and the operation of taking the inverse, as this follows automatically.

Proposition 1.2. *Write e_G for the unit element of G and e_H for the unit element of H. If $f : G \to H$ is a homomorphism, then $f(e_G) = e_H$. For any $g \in G$ the image of the inverse g^{-1} is $f(g^{-1}) = f(g)^{-1}$.*

Proof. Since f is a homomorphism we have $f(e_G) = f(e_G e_G) = f(e_G)f(e_G)$. Multiplying (say, on the left) by the inverse $f(e_G)^{-1}$ we obtain $e_H = f(e_G)$. Next, we have $e_H = f(e_G) = f(gg^{-1}) = f(g)f(g^{-1})$, which implies the second statement. □

Here are some examples of homomorphisms.

Example 1.3. (0) Let $\mathrm{Mat}(n, \mathbb{R})$ be the group of $n \times n$-matrices with real entries with respect to addition of matrices. The trace $\mathrm{tr}\colon \mathrm{Mat}(n, \mathbb{R}) \to \mathbb{R}$ is a homomorphism to the additive group of $\mathbb{R}$.

(1) Let $\mathrm{GL}(n, \mathbb{R})$ be the group of invertible $n \times n$-matrices with real entries with respect to multiplication of matrices. The determinant $\det\colon \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^\times$ is a homomorphism to the multiplicative group $\mathbb{R}^\times = \mathbb{R} \backslash \{0\}$.

(2) Let S_n be the symmetric group of permutations of $\{1, 2, \ldots, n\}$. Then the sign of a permutation $\mathrm{sign}\colon S_n \to \{\pm 1\}$ is a homomorphism to the cyclic group of order 2.

(3) The self-map $G \to G$ that sends g to g^{-1} is an automorphism of G if and only if G is abelian.

(4) Let G_1 be a subgroup of G. Then the identity map $G_1 \to G$ is cleary a homomorphism.

It is immediate to check that if $g\colon H \to K$ is a homomorphism, then the composition $h \circ g\colon G \to H \to K$ is a homomorphism.

Definition 1.4. *A function $f : G \to H$ is called an* **isomorphism** *if it is a homomorphism and a bijection.*

The fact that f is an isomorphism is written as $f : G \xrightarrow{\sim} H$. If there exists an isomorphism $f : G \xrightarrow{\sim} H$, then we say that G and H are isomorphic groups and write $G \cong H$.

Let us write id_G for the identity map $G \to G$, i.e., $\mathrm{id}_G(g) = g$ for any $g \in G$. It is clear that id_G is an isomorphism $G \xrightarrow{\sim} G$.

Exercise 1.5. *If $f\colon G \to H$ is an isomorphism of groups, then (since it is a bijection of sets) we have the inverse map $f^{-1}\colon H \to G$ defined by $f^{-1}(y) = x$ if $y = f(x)$. Prove that f^{-1} is an isomorphism and $f^{-1} \circ f = \mathrm{id}_G$ and $f \circ f^{-1} = \mathrm{id}_H$. Conclude that $G \cong H$ is an equivalence relation on the set of groups.*

Isomorphic groups are indistinguishable as groups.

Here is a standard example of an isomorphism. A cyclic group of order n is a finite set

$$C_n = \{a^0, a, a^2, \ldots, a^{n-1}\}$$

with the unit $e = a^0$ and the group law written multiplicatively: $a^i \cdot a^j = a^k$, where $k \in \{0, 1, \ldots, n-1\}$ is such that $i + j = k + rn$ for some integer r. Now consider the group of residues modulo n denoted by

$$\mathbb{Z}/n = \{\bar{0}, \bar{1}, \bar{2}, \ldots, \overline{n-1}\}$$

with the group law written additively, that is, the sum of $\bar{i}$ and $\bar{j}$ is $\bar{k}$, where $i+j=k+rn$ for some integer r. The function $\bar{i}\mapsto a^i$ is an isomorphism $\mathbb{Z}/n\to C_n$, because it transforms the group law on $\mathbb{Z}/n$ into the group law on C_n. Hence $\mathbb{Z}/n\cong C_n$.

Definition 1.6. *Let G be a group. An isomorphism $f\colon G\xrightarrow{\sim}G$ is called an* **automorphism** *of G.*

By Exercise 1.5, $f^{-1}\colon G\to G$ is also an automorphism of G. The identity map $\mathrm{id}_G\colon G\to G$ is visibly an automorphism of G.

Exercise 1.7. *Prove that the set of all automorphisms of a group G is a group with the unit element id_G and the group law given by the composition of automorphisms. This group is denoted by $\mathrm{Aut}(G)$.*

Thus to an arbitrary group G we associated another group $\mathrm{Aut}(G)$.

Exercise 1.8. *Determine the groups $\mathrm{Aut}(\mathbb{Z})$ and $\mathrm{Aut}(\mathbb{Z}/n)$. (Hint: $\mathrm{Aut}(G)$ sends a generator of a cyclic group to another generator. In the second question you can start with the case when n is a prime number.)*

Example 1.9. Here is an important example of an automorphism. Let G be a group. Take any $g\in G$. The function $G\to G$ given by $x\mapsto gxg^{-1}$ is called the *conjugation by g*. It sends $x\cdot y$ to

$$g(xy)g^{-1}=gx(gg^{-1})yg^{-1}=(gxg^{-1})(gyg^{-1}),$$

so this function is a homomorphism $G\to G$. The conjugation by g is a bijection: the inverse function is the conjugation by g^{-1}. Thus it is an automorphism of G.

To a homomorphism $f\colon G\to H$ we associate its image

$$\mathrm{Im}(f)=f(G)=\{f(x)|x\in G\}\subset H$$

and its kernel

$$\mathrm{Ker}(f)=\{x\in G|f(x)=e_H\}.$$

It is easy to check that f is an injective function if and only if $\mathrm{Ker}(f)=\{e_G\}$.

Proposition 1.10. *Let $f\colon G\to H$ be a homomorphism of groups. Then $\mathrm{Im}(f)$ is a subgroup of H and $\mathrm{Ker}(f)$ is a subgroup of G. Moreover, $\mathrm{Ker}(f)$ is stable under all conjugations, that is, if $x\in G$ is such that $x\in\mathrm{Ker}(f)$, then $gxg^{-1}\in\mathrm{Ker}(f)$ for any $g\in G$.*

Proof. We use Proposition 1.2. Since $f(e_G)=e_H$, the image $\mathrm{Im}(f)$ contains the unit element of H. As $f(x)f(y)=f(xy)$, we see that $\mathrm{Im}(f)$ is closed under the group operation of H. Finally, the inverse of $f(x)\in\mathrm{Im}(f)$ is $f(x^{-1})$ which is also in $\mathrm{Im}(f)$. This proves that $\mathrm{Im}(f)$ is a subgroup of H.

We have $e_G\in\mathrm{Ker}(f)$ and if $f(x)=f(y)=e_H$, then $f(xy)=e_H$. Also, $f(x)=e_H$ implies $f(x^{-1})=f(x)^{-1}=e_H$, so $\mathrm{Ker}(f)$ is a subgroup of G.

Finally, if $f(x)=e_H$, then $f(gxg^{-1})=f(g)f(x)f(g^{-1})=f(g)f(x)f(g)^{-1}=e_H$. This finishes the proof. □

1.2. **Normal subgroups, quotient groups and the isomorphism theorem.** The property of the kernel of a homomorphism to be stable under conjugations is very important.

Definition 1.11. *Let G be a group. A subgroup $S \subset G$ is called* **normal** *if it is stable under the conjugation by any element of G.*

By Proposition 1.10, if $f\colon G \to H$ is a homomorphism, then $\mathrm{Ker}(f)$ is a normal subgroup of G. In Example 1.3 (1) the kernel is the *special linear group* $\mathrm{SL}_n(\mathbb{R})$. In Example 1.3 (2) the kernel is the *alternating group* A_n. These are normal subgroups.

Definition 1.12. *A group G is called* **simple** *if G has no normal subgroups other than $\{e\}$ and G.*

It is easy to see that if p is a prime number, then the cyclic group C_p is simple. This follows from a stronger statement, namely, that C_p has no subgroups at all (normal or not) other than $\{e\}$ and the whole group C_p. In fact, if G is abelian, that is, $xy = yx$ for any $x, y \in G$, then any subgroup of G is normal.

In general, not all subgroups are normal. For example, consider the symmetric group S_3. Let $G \subset S_3$ be the set $\{e, (12)\}$. It is clear that G is a subgroup, but it is not a normal subgroup. Indeed, G is not stable under the conjugation by the element $(13) \in S_3$:

$$(13)(12)(13)^{-1} = (13)(12)(13) = (23) \notin G.$$

The smallest simple non-abelian group is the alternating group A_5. We cannot prove this now but may return to this statement later.

It is a fundamental property of normal subgroups that *any* normal subgroup is the kernel of some homomorphism. To prove this, for any normal subgroup $N \subset G$ we shall construct a homomorphism $f\colon G \to H$ such that $N = \mathrm{Ker}(f)$.

Recall that if $S \subset G$ is a subgroup, then the subsets $gS = \{gs | s \in S\}$ are called the *left cosets* of S, for all $g \in G$. Similarly, the sets $Sg = \{sg | s \in S\}$ are called the *right cosets* of S. It is known from the first year (and is easy to prove) that for $g_1, g_2 \in G$ only two possibilities can occur: either $g_1S = g_2S$ or $g_1S \cap g_2S = \varnothing$. (There is a similar property for the right cosets.) In general, left and right cosets are not the same. For example, the left cosets of S_3 modulo $\{e, (12)\}$ are

$$\{e, (12)\}, \quad \{(13), (123)\}, \quad \{(23), (132)\},$$

whereas the right cosets are

$$\{e, (12)\}, \quad \{(13), (132)\}, \quad \{(23), (123)\}.$$

But if we consider a normal subgroup $N \subset G$, then each left coset gN is equal to the right coset Ng. Indeed, if $x \in N$, then $gx = (gxg^{-1})g \in Ng$; similarly, $xg = g(g^{-1}xg) \in gN$, hence we have an equality of sets $gN = Ng$, for any $g \in G$.

Exercise 1.13. *Let $H \subset G$ be a subgroup. Prove that if $gH = Hg$ for every $g \in G$, then H is a normal subgroup.*

Now we can define a group structure on the set G/N of left cosets gN, for $g \in G$. Indeed, define the composition of g_1N and g_2N as the set of all products of an element from g_1N and an element from g_2N; this set of products is denoted by $(g_1N)(g_2N)$.

Lemma 1.14. *Let N be a normal subgroup of G. For any $g_1, g_2 \in G$ we have*

$$(g_1N)(g_2N) = g_1g_2N. \tag{1.1}$$

Proof. Here is a short proof. We have $NN = N$, because N is a subgroup so is closed under the group operation. Then $(g_1N)(g_2N) = (g_1N)(Ng_2) = g_1Ng_2 = g_1g_2N$. Here we used that $g_2N = Ng_2$.

Here is a more explicit proof. Take any $x, y \in N$. Then $(g_1x)(g_2y) = g_1g_2(g_2^{-1}xg_2)y$. Since N is normal, we have $g_2^{-1}xg_2 \in N$, hence $(g_2^{-1}xg_2)y \in N$ so that $(g_1x)(g_2y) \in g_1g_2N$. But it is clear that every element of g_1g_2N is obtained in this way, namely, by taking $x = e_G$, so we are done. □

Lemma 1.15. *Let N be a normal subgroup of G. The set G/N of left cosets of G modulo N is a group with group law sending g_1N and g_2N to g_1g_2N.*

Proof. The composition of cosets is associative. Indeed, this follows from the associativity in G: the coset $g_1(g_2g_3)N$ equals the coset $(g_1g_2)g_3N$, for all $g_1, g_2, g_3 \in G$. The trivial coset $N = e_GN$ is the unit element of G/N. Finally, the coset $g^{-1}N$ is the inverse of gN. We have checked all the axioms of a group. □

Proposition 1.16. *Let N be a normal subgroup of G. The function $f\colon G \to G/N$ given by $g \mapsto gN$ is a surjective homomorphism of groups with kernel* $\mathrm{Ker}(f) = N$.

Proof. The property (1.1) implies that f is a homomorphism, and is visibly surjective. Since $gN = N$ if and only if $g \in N$, we deduce that $N = \mathrm{Ker}(f)$. □

Definition 1.17. *Let N be a normal subgroup of G. Then G/N is called the* **quotient group** *of G modulo N.*

Exercise 1.18. *List all subgroups of the following groups and determine which of them are normal. For each normal subgroup describe the quotient group.*

$$\mathbb{Z},\ C_n \text{ for } n \geqslant 2,\ S_3,\ D_8 \text{ (the dihedral group of order 8).}$$

Now we go back to the situation when $f\colon G \to H$ is a homomorphism of groups. The following relation between $\mathrm{Ker}(f)$ and $\mathrm{Im}(f) = f(G)$ is fundamental. To state it we note that f is constant on each coset $g\mathrm{Ker}(f)$. This is clear, because $f(x) = e_H$ implies $f(gx) = f(g)$.

Theorem 1.19 (The isomorphism theorem)**.** *Let $f\colon G \to H$ be a homomorphism of groups. The map $g\mathrm{Ker}(f) \mapsto f(g)$ is an isomorphism of groups*

$$G/\mathrm{Ker}(f) \xrightarrow{\sim} f(G).$$

Proof. We know that $\mathrm{Ker}(f)$ is a normal subgroup of G, so $G/\mathrm{Ker}(g)$ is a group.

The rule $g\mathrm{Ker}(f) \mapsto f(g)$ is a function $G/\mathrm{Ker}(f) \to f(G)$, because f is constant on each coset $g\mathrm{Ker}(f)$.

Let us check that this function is a homomorphism. We know that the product of $g_1\mathrm{Ker}(f)$ and $g_2\mathrm{Ker}(f)$ equals $g_1g_2\mathrm{Ker}(f)$. But the image of $g_1g_2\mathrm{Ker}(f)$ is $f(g_1g_2) = f(g_1)f(g_2)$, which is the product of the images of $g_1\mathrm{Ker}(f)$ and $g_2\mathrm{Ker}(f)$, so we are fine.

Our function is visibly surjective onto $\mathrm{Im}(f)$. It is also injective. To check this it is enough to prove that its kernel is the unit element of $G/\mathrm{Ker}(f)$, which is the trivial coset $\mathrm{Ker}(f)$. If a coset $g\mathrm{Ker}(f)$ goes to e_H, we have $f(g) = e_H$, but then $g \in \mathrm{Ker}(f)$, hence $g\mathrm{Ker}(f) = \mathrm{Ker}(f)$.

To conclude, $G/\mathrm{Ker}(f) \to f(G)$ is a bijective homomorphism, hence an isomorphism. □

A homomorphism $f : G \to H$ sends any subgroup $A \subset G$ to a subgroup $f(A) \subset H$. (Indeed, the image $f(A)$ is closed under the group law of H, contains e_H and is closed under taking the inverse.) Conversely, if B is a subgroup of H, then the inverse image

$f^{-1}(B) = \{g \in G | f(g) \in B\}$ is a subgroup of G. (Indeed, if $g_1, g_2 \in G$ are such that $f(g_1) \in B$ and $f(g_2) \in B$, then $f(g_1 g_2) = f(g_1)f(g_2) \in B$ hence $g_1 g_2 \in f^{-1}(B)$. We have $e_G \in f^{-1}(e_H) \subset f^{-1}(B)$. If $g \in f^{-1}(B)$, then $f(g^{-1}) = f(g)^{-1} \in B$, so $g^{-1} \in f^{-1}(B)$.)

We note that if B is normal in H, then $f^{-1}(B)$ is normal in G. Indeed, take any $x \in f^{-1}(B)$. Then $gxg^{-1} \in f^{-1}(B)$ because $f(gxg^{-1}) = f(g)f(x)f(g)^{-1} \in B$. In general, the image of a normal subgroup is not normal (e.g. take any non-normal subgroup $A \subset G$ and consider the identity homomorphism $A \to G$), but this is true for *surjective* homomorphisms. Indeed, let $A \subset G$ be a normal subgroup. We have $f(g)f(A)f(g)^{-1} = f(gAg^{-1}) = f(A)$, and if every element of H can be written as $f(g)$ for some $g \in G$, then this proves that $f(A)$ is a normal subgroup of H.

Proposition 1.20. *Let N be a normal subgroup of G and let $f \colon G \to G/N$ be the surjective homomorphism given by $g \mapsto gN$. If $S \subset G$ is a subgroup containing N, then N is a normal subgroup of S and $f(S) = S/N$ is a subgroup of G/N. Sending S to $f(S)$ a bijection between the subgroups of G containing N and the subgroups of G/N. Moreover, S is normal in G if and only if S/N is normal in G/N.*

Proof. We have $gNg^{-1} = N$ for any $g \in G$, in particular, for $g \in S$, so N is normal in S. As $N = \mathrm{Ker}(f)$, we have $S/N = f(S)$.

Note that S is the disjoint union of the cosets gN, where $g \in S$. To prove that we have a bijection we associate to a subgroup $H \subset G/N$ the subgroup $f^{-1}(H)$ of G; this is a subgroup containing N. The composition

$$S \mapsto f(S) = \{gN | g \in S\} \mapsto f^{-1}(f(S)) = \bigcup_{g \in S} gN = S$$

is the identity map of the set of subgroups of G that contain N. The composition

$$H \mapsto f^{-1}(H) \mapsto f(f^{-1}(H)) = H$$

is the identity map of the set of subgroups of G/N. This establishes the desired bijection. The last claim follows from the remarks before the proposition. □

1.3. **Some group-theoretic constructions.**

1.3.1. *The centre of a group.* The conjugations by the elements of G form a subgroup of $\mathrm{Aut}(G)$, called the group of *inner* automorphisms and denoted by $\mathrm{Inn}(G)$. Indeed, for $a, b \in G$ the conjugation by ab is the conjugation by b followed by the conjugation by a (in this order!) because

$$x \mapsto bxb^{-1} \mapsto a(bxb^{-1})a^{-1} = (ab)x(ab)^{-1}.$$

The unit element of $\mathrm{Aut}(G)$ is the conjugation by e_G. The inverse of the conjugation by g is the conjugation by g^{-1}. Hence $\mathrm{Inn}(G)$ is a subgroup of $\mathrm{Aut}(G)$.

Sending an element $g \in G$ to the conjugation by g is a homomorphism $G \to \mathrm{Aut}(G)$. The image is $\mathrm{Inn}(G)$. The kernel consists of all elements $g \in G$ such that $gxg^{-1} = x$ for any $x \in G$. Equivalently, $gx = xg$, so the kernel is the subset of G consisting of the elements that commute with all elements of G. This set is called the *centre* of G and is denoted by $Z(G)$. Obviously, $Z(G)$ is a normal subgroup of G. Now Theorem 1.19 gives an isomorphism

$$G/Z(G) \xrightarrow{\sim} \mathrm{Inn}(G).$$

It is clear that $Z(G) = G$ if and only if G is abelian.

1.3.2. *The commutator of a group.* Let G be a group. For $a, b \in G$ write $[a, b] = aba^{-1}b^{-1}$ and call this the *commutator* of a and b. Define $[G, G]$ as the smallest subgroup in G containing the commutators $[a, b]$ for all $a, b \in G$. This subgroup is called the *commutator* (or the *derived subgroup*) of G. Clearly, G is abelian if and only if $[G, G] = \{e_G\}$.

Lemma 1.21. *Let G be a group. Then $[G, G]$ is a normal subgroup of G. The quotient group $G/[G, G]$ is abelian.*

Proof. We need to check that $[G, G]$ is stable under conjugations. Indeed,

$$gaba^{-1}b^{-1}g^{-1} = (gag^{-1})(gbg^{-1})(ga^{-1}g^{-1})(gb^{-1}g^{-1}) = (gag^{-1})(gbg^{-1})(gag^{-1})^{-1}(gbg^{-1})^{-1},$$

where we used that $(gag^{-1})^{-1} = ga^{-1}g^{-1}$ and similarly for b. Thus this element is a commutator, so is contained in $[G, G]$. □

Actually, more is true.

Proposition 1.22. *Let N be a normal subgroup of G. Then G/N is abelian if and only if N contains $[G, G]$.*

Proof. A group is abelian if and only if every commutator is the unit element. By the definition of the group law on G/N, this group is abelian if and only if $[a, b] \in N$ for any $a, b \in G$, but this is equivalent to $[G, G] \subset N$. □

Exercise 1.23. *Prove that any subgroup of G that contains $[G, G]$ is normal.*

1.3.3. *The product of groups.* Let A and B be groups. Consider the product of sets $A \times B = \{(a, b) | a \in A, b \in B\}$ and turn it into a group by defining the group law as follows:

$$(a, b) \cdot (a', b') : = (aa', bb').$$

It is clear that (e_A, e_B) is the unit element and (a^{-1}, b^{-1}) is the inverse of (a, b). The associativity in $A \times A$ holds because it holds in A and in B. The group $A \times B$ is called the *product* of A and B.

We have injective maps $i_A \colon A \to A \times B$, $i_A(x) = (x, e_B)$, and $i_B \colon B \to A \times B$, $i_B(y) = (e_A, y)$. Hence, A is isomorphic to its image $i_A(A) = \{(a, e_B) | a \in A\}$ in $A \times B$. Similarly, B is isomorphic to its image $i_B(B) \subset A \times B$. This allows us to think of A and B as subgroups of $A \times B$. We note that each of these subgroups is normal, and their intersection is the unit element (e_A, e_B). Moreover, each element (a, e_B) of $A \subset A \times B$ commutes with each element (e_A, b) of $B \subset A \times B$. In fact, there is a converse statement.

Proposition 1.24. *Let G be a group. Suppose that A and B are normal subgroups of G such that $A \cap B = \{e_G\}$ and $G = AB$ (which means that every element of G can be written as ab, where $a \in A$ and $b \in B$). Then the map $f \colon A \times B \to G$ that sends $a \in A$ and $b \in B$ to their product $ab \in G$ is an isomorphism.*

Proof. By assumption, f is surjective. So it remains to show that f is a homomorphism with kernel $\{(e_A, e_B)\}$. I claim that the subgroups A and B of G commute, which means that for every $a \in A$ and $b \in B$ we have $ab = ba$ in G. Indeed, $bab^{-1} \in A$ since A is normal, hence $bab^{-1}a^{-1} \in A$. But B is also normal, so $ab^{-1}a^{-1} \in B$, hence $bab^{-1}a^{-1} \in B$. By assumption, $A \cap B = \{e_G\}$, thus $bab^{-1}a^{-1} = e_G$, which is equivalent to $ba = ab$. This proves the claim. The claim implies that f is a homomorphism:

$$f\big((a, b) \cdot (a', b')\big) = f\big((aa', bb')\big) = aa'bb' = (ab)(a'b') = f\big((a, b)\big) \cdot f\big((a', b')\big).$$

Finally, Ker(f) consists of the pairs (a, b) such that $ab = e_G$. This implies that $a = b^{-1} \in A \cap B = \{e_G\}$, so $a = e_A$ and $b = e_B$. Thus f has trivial kernel, and therefore is injective. □

Suppose that we are given groups $G_1, \ldots, G_n$. We define the product $G\colon = G_1 \times \ldots \times G_n$ as the set of ordered n-tuples $(g_1, \ldots, g_n)$, where $g_i \in G_i$. Define the group law of G coordinatewise:

$$(g_1, \ldots, g_n) \cdot (h_1, \ldots, h_n)\colon = (g_1h_1, \ldots, g_nh_n).$$

The unit element of G is $(e_{G_1}, \ldots, e_{G_n})$, and the inverse of $(g_1, \ldots, g_n)$ is $(g_1^{-1}, \ldots, g_n^{-1})$. Clearly, G is group. It is called the *product* of groups $G_1, \ldots, G_n$. It can also be obtained by doing the product of two groups $n - 1$ times.

1.3.4. *Abelian groups and p-primary subgroups.* Let us see how products of groups work in the abelian case, which is much simpler than the case of arbitrary groups.

Lemma 1.25. *Let G be an abelian group. If the orders of $a, b \in G$ are finite, then the order of ab is also finite and divides the least common multiple of the orders of a and b. The set of elements of G that have finite order is a subgroup of G.*

Proof. Let m be the order of a, let n be the order of b and let $k =$lcm(a, b). It is clear that $(ab)^k = e$, so the order of ab divides k. Thus the set of elements of G of finite order is closed under the group law, but it is obviously closed under taking the inverse and contains e_G, so it is a subgroup. □

Definition 1.26. *Let G be an abelian group. The set of elements of G of finite order is called* **the torsion subgroup of** *G and is denoted by G_{tors}. If $G = G_{\mathrm{tors}}$, then G is called a* **torsion abelian group***.*

For example, $\mathbb{Q}/\mathbb{Z}$ and C_n are torsion abelian groups, whereas $\mathbb{Q}$ and $\mathbb{Z}$ are not. It is an easy exercise to prove that the quotient of an abelain group G by G_{tors} has no non-zero torsion elements.

By Lemma 1.25 the set of elements in G whose order is a power of a given prime is also a subgroup.

Definition 1.27. *Let G be an abelian group and let p be a prime number. The set of elements $g \in G$ such that the order of g is a power of p is called* **the p-primary subgroup of** *G and is denoted by $G\{p\}$. If $G = G\{p\}$, then G is called a p-***primary torsion abelian group***.*

Corollary 1.28. *Let $n = p_1^{a_1} \ldots p_m^{a_m}$, where $p_1, \ldots, p_m$ are prime numbers and $a_i \geqslant 1$, for $i = 1, \ldots, m$. There is an isomorphism of groups*

$$C_n \cong C_{p_1^{a_1}} \times \ldots \times C_{p_m^{a_m}}.$$

Proof. We proceed by induction on m. If $m = 1$, there is nothing to prove.

Write $G = C_n$ and consider $G\{p_m\}$. Any subgroup of a cyclic group is also cyclic, in particular, $G\{p_m\} \cong C_{p_m^{a_m}}$. Let G' be the set of all elements of G of order coprime to p_m. By Lemma 1.25, G' is a subgroup of G. It is clear that $G\{p_m\} \cap G' = \{e\}$. If we show that $G = G\{p\}G'$, then we will be able to use Proposition 1.24 to conclude that $G \cong G\{p\} \times G'$. Then $|G| = p_m^{a_m}|G'|$, so the order of G' has $m - 1$ prime factors and we finish the proof by applying the induction assumption to G'.

Take any $g \in G$. The order of g can be written as $p^r s$, where $r, s \in \mathbb{Z}$, $r \geqslant 0$, $s \geqslant 1$, and p does not divide s. There are integers k and l such that $1 = kp^r + ls$. Then $g = a^k b^l$, where $a = g^{p^r} \in G'$ and $b = g^s \in G\{p\}$, so we are done. □

1.3.5. *Generators.* Let G be a group. It is easy to see that the intersection of two subgroups of G is also a subgroup. More generally, we have the following lemma.

Lemma 1.29. *Let I be a set. Suppose that for each $i \in I$ we are given a subgroup $H_i \subset G$. Then $H = \cap_{i \in I} H_i$ is a subgroup of G.*

Proof. We need to check that $e_G \in H$ (true, because $e_G \in H_i$ for each $i \in I$), that $g \in H$ implies $g^{-1} \in H$ (true, because $g^{-1} \in H_i$ for each $i \in I$), and that H is closed under the group law of G (true, because each H_i is closed under the group law of G). □

Definition 1.30. *Let G be a group and let $S \subset G$ be a set. The intersection of all subgroups of G that contain S is called* **the subgroup of G generated by** *S. If G is the only subgroup of G that contains S, we say that the elements of S* **generate** *G.*

The subgroup of G generated by S is clearly the smallest subgroup of G that contains S.

Explicitly, the subgroup of G generated by S is the set H consisting of e_G and all finite products $x_1 x_2 x_3 \ldots x_n$, where each factor x_i is either an element of S or an inverse of an element of S, and all cancellations have been done, that is, there is no i such that $x_{i+1} = x_i^{-1}$. The set H is closed under taking products and inverses, so is a subgroup of G (the inverse of $x_1 x_2 x_3 \ldots x_n$ is $x_n^{-1} x_{n-1}^{-1} \ldots x_1^{-1}$). Any subgroup of G that contains S must also contain H, so H is the subgroup of G generated by S.

Some examples. The definition of a (finite or infinite) cyclic group is that it is a group generated by one element. It can be proved that for any $n \geqslant 3$ the symmetric group S_n is generated by two elements. The commutator subgroup $[G, G] \subset G$ is the smallest subgroup generated by the elements of the form $[a, b] = aba^{-1}b^{-1}$, for $a, b \in G$.

Example 1.31. *Suppose that G_1 is generated by n_1 elements, and G_2 is generated by n_2 elements. Prove that $G_1 \times G_2$ is generated by $n_1 + n_2$ elements.*

Definition 1.32. *A group G is called* **finitely generated** *if there is a positive integer n such that G is generated by n elements.*

An important question which we'll look into later is to describe all finitely generated abelian groups. This includes all cyclic groups and their products. The main classification theorem asserts that every finitely generated abelian group is isomorphic to a product of finitely many cyclic groups (which can be finite or infinite).

2. Groups acting on sets

2.1. **Actions, orbits and stabilisers.** In the previous section the notion of a group was introduced abstractly, as a set with a binary operation satisfying certain axioms. This is not how groups were introduced historically. In fact, they were first conceived as symmetry groups, for example the dihedral group D_{2n} is the symmetry group of a regular n-gon in the plane. Interesting examples of groups are the symmetry group of a cube or another Platonic solid. Galois arrived at the notion of a group by considering a polynomial of degree n without multiple roots and attaching to it a certain subgroup of S_n, acting on the roots by permutations, called the Galois group of the polynomial. Infinite groups naturally appear as the symmetry groups of 2- or 3-dimensional crystals. If we know that a group G is the group of all permutations of a (finite or infinite) set X that preserve a certain structure on X (for example, X can be a lattice or a vector space with a scalar product), then this provides us with more tools to study G. This is often insightful for understanding the structure of G.

Let us formally define what we mean by an action of a group G on a set X.

Definition 2.1. *Let G be a group and let X be a set. Let $S(X)$ be the group of bijections $X \to X$ with composition as the group law. An* **action** *of G on X is a homomorphism $G \to S(X)$.*

Thus an action of G on X associates to each $g \in G$ a bijective map $X \to X$, which can be thought of as a permutation of the elements of X. The only condition is that the map $X \to X$ associated to $g_1 g_2$ is the composition of the map associated to g_2 followed by the map associated to g_1 (in this order!). For $g \in G$ and $x \in X$ we write the image of x under the map associated to g as $g(x)$. Thus we can consider an action of G on X as a function from the product set $G \times X$ to X and write it as

$$G \times X \longrightarrow X.$$

Such a function is an action if and only if $(g_1 g_2)(x) = g_1(g_2(x))$ for any $g_1, g_2 \in G$ and $x \in X$. Since $G \to S(X)$ is a homomorphism, Proposition 1.24 implies that e_G acts trivially, i.e., $e_G(x) = x$ for any $x \in X$, and the map $X \to X$ associated to g^{-1} is the inverse of the map associated to g.

Example 2.2. (1) The group $\mathrm{GL}(n, \mathbb{R})$ of invertible matrices acts on the vector space $\mathbb{R}^n$ by linear transformations: if A is a matrix and v is a column vector, then A sends v to Av. In fact, every linear transformation $\mathbb{R}^n \to \mathbb{R}^n$ is given by a matrix, so $\mathrm{GL}(n, \mathbb{R})$ is the group of all automorphisms (=bijective linear transformations) of the vector space $\mathbb{R}^n$.

(2) The group $\mathrm{O}(n, \mathbb{R})$ of orthogonal matrices acts on $\mathbb{R}^n$ preserving the usual scalar product (the dot product). In fact, every linear transformation $\mathbb{R}^n \to \mathbb{R}^n$ which preserves $(x.y)$ is given by an orthogonal matrix, so $\mathrm{O}(n, \mathbb{R})$ is the group of all linear transformations of the vector space $\mathbb{R}^n$ that preserve the dot product.

(3) $S_n = S(X)$ is the group of all permutations of $X = \{1, 2, \ldots, n\}$. Here the set X has no additional structure.

Definition 2.3. *An action of a group G on a set X is* **faithful** *if the map $G \to S(X)$ is injective.*

Equivalently, the kernel of $G \to S(X)$ is trivial, which means that if $g(x) = x$ for every $x \in X$, then $g = e_G$. The action in each of the three examples above is faithful.

Definition 2.4. *Let $G \times X \to X$ be an action of a group G on a set X. The G-***orbit** *of an element $x \in X$ is the subset*

$$G(x) = \{g(x) | g \in G\} \subset X.$$

The **stabiliser** *of x is the subgroup*

$$\mathrm{St}_G(x) = \{g \in G | g(x) = x\} \subset G.$$

Checking group axioms we see that $\mathrm{St}_G(x)$ is indeed a subgroup of G. If G is fixed, we write $\mathrm{St}(x)$ for $\mathrm{St}_G(x)$. It is clear that X is a disjoint union of G-orbits. The following lemma says that the stabilisers of points in the same G-orbit are conjugate in G.

Lemma 2.5. *Let $G \times X \to X$ be an action of a group G on X. Then* $\mathrm{St}(g(x)) = g\mathrm{St}(x)g^{-1}$.

Proof. If $h \in G$ is such that $h(x) = x$, then $(ghg^{-1})(g(x)) = (ghg^{-1}g)(x) = (gh)(x) = g(h(x)) = g(x)$. We obtain that $g\mathrm{St}(x)g^{-1} \subset \mathrm{St}(g(x))$. This holds for any $g \in G$ and any $x \in X$. Thus the inclusion will remain true if we replace x by $g(x)$ and g by g^{-1}. Then we obtain $g^{-1}\mathrm{St}(g(x))g \subset \mathrm{St}(g^{-1}(g(x)))$. Since we have an action, this simplifies as $g^{-1}\mathrm{St}(g(x))g \subset \mathrm{St}(x)$, hence $\mathrm{St}(g(x)) \subset g\mathrm{St}(x)g^{-1}$, and we are done. □

Theorem 2.6 (Orbit–stabiliser theorem)**.** *Let $G \times X \to X$ be an action of a group G on a set X. For any $x \in X$ the map $g \mapsto g(x)$ gives a bijection of the set of left cosets $G/\mathrm{St}(x)$ with the orbit $G(x)$. In particular, if G is a finite group, then $|G(x)| = |G|/|\mathrm{St}(x)|$ for any $x \in X$. If X is a finite set and $X = \cup_{i=1}^{n} G(x_i)$ is a disjoint union of G-orbits, then*

$$|X| = \sum_{i=1}^{n} |G(x_i)| = \sum_{i=1}^{n} [G\colon \mathrm{St}(x_i)], \tag{2.1}$$

where $[G\colon \mathrm{St}(x_i)]$ is the index of $\mathrm{St}(x_i)$ in G.

Proof. The function $G \to G(x)$ given by $g \mapsto g(x)$ is obviously surjective. The inverse image of $g(x)$ is the set of elements $h \in G$ such that $h(x) = g(x)$, which is equivalent to $g^{-1}h \in \mathrm{St}(x)$ and hence also to $h \in g\mathrm{St}(x)$. Thus our function induces a bijection between the set of left cosets $G/\mathrm{St}(x)$ and the orbit $G(x)$. This proves the first statement. The second statement follows from the fact that X is a disjoint union of G-orbits. □

Since the stabilisers of points in a given G-orbit are conjugate in G, they have the same index in G.

2.2. **Applications of the orbit–stabiliser theorem.**

Theorem 2.7 (Cayley)**.** *Let G be a finite group of order n. Then S_n contains a subgroup isomorphic to G.*

Proof. Consider the action of G on itself by left multiplication:

$$G \times G \longrightarrow G, \qquad (a, b) \mapsto ab.$$

This action is faithful, because $ge = e$ implies $g = e$. Hence we have an injective homomorphism $G \to S(G) = S_n$. Its image is isomorphic to G. □

Lagrange's theorem says that the order of any element of a finite group of order n divides n. In general the converse does not hold, that is, if a positive integer m divides n, then not every group of order n contains an element of order m, but this is actually true if m is a prime!

Theorem 2.8 (Cauchy)**.** *Let G be a finite group of order n and let p be a prime factor of n. Then G contains an element of order p.*

Proof. Consider the set

$$G^p = \{(g_1, \ldots, g_p) | g_i \in G,\ i = 1, \ldots, p\}$$

of ordered p-tuples of elements of our group G. The cyclic group C_p of order p acts on G^p by cyclic shifts. Let $X \subset G^p$ be the subset of p-tuples $(g_1, \ldots, g_p)$ such that $g_1 \cdot g_2 \cdot \ldots \cdot g_p = e_G$. Such a p-tuple is uniquely determined by the first $p - 1$ elements, hence $|X| = n^{p-1}$. We claim that X is stable under the action of C_p. For this we need to show that in any group G we have

$$g_1 \cdot g_2 \cdot \ldots \cdot g_p = e_G \quad \Longrightarrow \quad g_p \cdot g_1 \cdot g_2 \cdot \ldots \cdot g_{p-1} = e_G.$$

Indeed, conjugation by g_p turns the first equality into the second equality.

Let us look at the C_p-orbits of X. There can be two kinds of orbits: those of cardinality p and those of cardinality 1, an example of which is the orbit consisting of $(e_G, \ldots, e_G)$. Now (2.1) says that $|X| = n^{p-1} = m + kp$, where m is the number of 1-element orbits and k is the number of p-element orbits. Since p divides n, and $m \geqslant 1$, we see that $m \geqslant p \geqslant 2$, so there is an orbit $(g, g, \ldots, g)$, where $g^p = e_G$ and $g \neq e_G$. The order of g in G is p. □

Definition 2.9. *Let p be a prime. A finite group G is called a p-**group** if the order of G is a power of p.*

Corollary 2.10. *A finite group G is a p-group if and only if the order of every element of G is a power of p.*

Proof. In one direction this follows from Lagrange's theorem and in the other direction from Cauchy's theorem. □

Arbitrary p-groups have very special properties.

Theorem 2.11. *Let G be a p-group, where p is a prime. Then the centre of G is non-trivial, that is, $Z(G) \neq \{e_G\}$.*

Proof. Consider the action of G on itself by conjugations:

$$G \times G \longrightarrow G, \qquad (a, b) \mapsto aba^{-1}.$$

We note that the orbit of g is $\{g\}$ if and only if $g \in Z(G)$. Then (2.1) takes the form

$$p^n = 1 + \ldots + 1 + \sum_{i=1}^{m} [G\colon \mathrm{St}(x_i)],$$

where the terms 1 are the cardinalities of the orbits of the elements of $Z(G)$, which all consist of one element only, and the remaining terms are the cardinalities of orbits which have more than one element. But $|G| = p^n$, hence each index $[G\colon \mathrm{St}(x_i)]$, for $i = 1, \ldots, m$, is a positive power of p. Considering the above equality modulo p we obtain that p divides $|Z(G)|$. However, the unit element e_G is in $Z(G)$, so $Z(G)$ contains at least p elements, hence $|Z(G)| \geqslant 2$. □

Example 2.12. The dihedral group D_8 has 8 elements, so it's a 2-group. By Theorem 2.11 there is an element $g \neq e$ in the centre of D_8. Indeed, think of D_8 as the group of symmetries of a square in $\mathbb{R}^2$ with centre at 0, so that D_8 consists of 4 rotations and 4 symmetries. The map $g(x, y) = (-x, -y)$ (=rotation by 180 degrees) is in the centre of D_8.

Some other properties of p-groups will be discussed in problem sheets.

Definition 2.13. *Let $G \times X \to X$ be an action of a group G on a set X. If X is a G-orbit, i.e., $X = G(x)$ for some $x \in X$, then we say that G acts **transitively** on X.*

For example, the action of G on itself by left multiplication is transitive, whereas the action of G on itself by conjugations is not. Another example of a transitive action is the action of G on G/H, where H is a subgroup of G; here $g \in G$ sends the coset xH to gxH.

Definition 2.14. *Let $G \times X \to X$ be an action of a group G on a set X. An element $x \in X$ is called a **fixed point** of $g \in G$ if $g(x) = x$. We denote by $\mathrm{Fix}(g) \subset X$ the set of fixed points of $g \in G$.*

A fixed point of $g \in G$ is the same as a 1-point orbit of the cyclic group generated by g.

Theorem 2.15 (Jordan). *Let $G \times X \to X$ be a transitive action of a finite group G on a finite set X. Then we have*

$$\sum_{g \in G} |\mathrm{Fix}(g)| = |G|. \tag{2.2}$$

In particular, there is an element $g \in G$ such that $\mathrm{Fix}(g) = \varnothing$.

Proof. To prove the formula we consider the set Y of pairs (g, x), where $g \in G$ and $x \in X$ are such that $g(x) = x$. We count $|Y|$ in two different ways, using the maps

$$G \longleftarrow Y \longrightarrow X$$

that forget one of the coordinates of (g, x). Projecting to G, we write $|Y|$ as the sum over all $g \in G$ of $|\mathrm{Fix}(g)|$. Projecting to X, we write $|Y|$ as the sum over all $x \in X$ of $|\mathrm{St}_G(x)|$. Since G acts transitively on X, by Lemma 2.5 we know that $|\mathrm{St}_G(x)|$ does not depend on x, so that $|\mathrm{St}_G(x)| = |\mathrm{St}_G(x_0)|$ for any chosen point $x_0 \in X$. We have $X = G(x_0)$. By the orbit–stabiliser theorem we have $|G| = |G(x_0)| \cdot |\mathrm{St}_G(x_0)|$. This proves (2.2).

The second statement is clear if $|X| = 1$. Assume $|X| \geqslant 2$. If $|\mathrm{Fix}(g)| \geqslant 1$ for every $g \in G$, then the left hand side has at least $|G|$ elements, and in fact it has more because $\mathrm{Fix}(e_G) = X$ has more than one element. Thus we get a contradiction with (2.2). Hence $\mathrm{Fix}(g) = \varnothing$ for at least one element $g \in G$. □

Corollary 2.16. *Let $G \times X \to X$ be an action of a finite group G on a finite set X. Then the number of G-orbits in X is $|G|^{-1} \sum_{g \in G} |\mathrm{Fix}(g)|$.*

Proof. Write X as a disjoint union of G-orbits, $X = \cup_{i=1}^{n} X_i$. The number of fixed points of $g \in G$ in X is the sum of the numbers of fixed points of g in X_i, for $i = 1, \ldots, n$. For each orbit, the formula in the statement of the corollary gives 1, by Theorem 2.15. Thus for X the value of the formula is n. □

Example 2.17. Check the formula of Corollary 2.16 for the cyclic subgroup $G \subset S_5$ generated by $(12)(345)$ acting on $X = \{1, 2, 3, 4, 5\}$.

3. Finitely generated abelian groups

3.1. **Smith normal form.** This section is a preparation for the main result proved in the next section.

Definition 3.1. *An $(m \times n)$-matrix $A = (a_{ij})$ with entries $a_{ij} \in \mathbb{Z}$ is in* **Smith normal form** *if the following conditions are satisfied.*

(a) *$a_{ij} = 0$ if $i \neq j$ (only diagonal entries can be non-zero).*

(b) *Write $a_i = a_{ii}$. For some integer $k \geqslant 0$ we have $a_i > 0$ for $i \leqslant k$ and $a_i = 0$ for $i > k$.*

(c) *$a_1 | a_2 | a_3 | \ldots | a_k$.*

Condition (c) says that a_1 divides a_2 which divides a_3, and so on. The following result is essentially a linear algebra statement with the added flavour that the entries are integers so we have to be a bit more careful to take care of this.

Theorem 3.2. *Any matrix with integer coefficients can be brought into Smith normal form using row and column operations.*

Recall that the three row operations are (1) switching the i-th and j-th rows, (2) multiplying the i-th row by -1 (and not just by any non-zero number!), (3) replacing the i-th row r_i by $r_i + a r_j$, where $i \neq j$ and $a \in \mathbb{Z}$. There are similar column operations.

We need some lemmas. Let $d(A)$ be the greatest common divisor of the entries a_{ij}. Let $t(A)$ be the smallest non-zero $|a_{ij}|$. Then $d(A)$ divides $t(A)$, so $d(A) \leqslant t(A)$. It is clear that the row and column operations do not change $d(A)$, because $\gcd(r, s)$ =$\gcd(r + as, s)$ for $a \in \mathbb{Z}$. The following lemma is the key observation in the proof of the theorem.

Lemma 3.3. *Any matrix A with integer entries can be transformed using row and column operations into a matrix B such that $t(B) = d(B) = d(A)$.*

Before proving the lemma let us explain the idea on an easier example of a finite set of non-negative integers $S = \{s_1, \dots, s_n\}$. We can replace any s_i by the remainder r, where $s_i = as_j + r$, $0 \leqslant r < s_j$ and $a \in \mathbb{Z}$. This operation does not change $\gcd(S)$. By taking s_j to be the smallest element of S, we can ensure, after finitely many steps, that $\gcd(S)$ is the smallest non-zero element of S. The lemma does the same for matrices.

Proof of Lemma 3.3. We use induction on $t = t(A)$. If $t(A) = 1$, then $t(A) = d(A)$. Assume that the lemma is proved for all matrices M of size $m \times n$ with $d(M) = d(A)$ and $t(M) < t(A)$. Let us prove this for A.

Suppose that $t(A) = |a_{ij}|$. By switching rows and columns, and multiplication of the first row by -1 if necessary, we can assume that $t = t(A) = a_{11}$. If t divides all entries of A, we are done. Otherwise, we can find a_{ij} not divisible by t. If $i = 1$ or $j = 1$, the induction step is easy. If a_{i1} is not divisible by t, we write $a_{i1} = \ell t + c$, where $\ell \in \mathbb{Z}$ and $0 < c < t$, and replace the i-th row by $r_i - \ell r_1$. The resulting matrix has an entry equal to $c < t$, so we have lowered t and can apply the induction assumption. If a_{1i} is not divisible by t, we do a similar column operation.

It remains to deal with the case when t divides all the a_{i1} and a_{1j}. By doing row and column operations we ensure that $a_{i1} = a_{1j} = 0$ for all i and j. Let B be the resulting matrix. If $t(B) < t$, we can apply the induction assumption to B. Suppose that $t(B) = t > d(B) = d(A)$, so there is an entry $x = b_{ij}$ not divisible by t. Let C be the matrix obtained from B by replacing the first row of B by $r_1 + r_i$. This does not change the $(1,1)$-entry (because every entry in the first column is zero, except the $(1,1)$-entry), hence $t(C) \leqslant t$. The $(1, j)$-entry of C is x. The matrix C falls into the easy case treated above, so we do as before and replace the j-column of C by $c_j - \ell c_1$, where $x = \ell t + r$ with $\ell \in \mathbb{Z}$ and $0 < r < t$. The resulting matrix D has the $(1, j)$-entry equal to $r < t$, hence $t(D) < t$ and we can apply the induction assumption to D. This proves the lemma. □

Proof of Theorem 3.2. By Lemma 3.3 we transform A by row and column operations into a matrix B with an entry which divides every other entry of B. Switching rows and columns, and multiplying the first row by -1, if necessary, we can assume that the said entry is $b_{11} = d > 0$. Since b_{11} divides every entry, we can perform row and column operations to ensure that every entry in the first row and the first column, except the $(1,1)$-entry, is zero. The matrix now has the form

$$\begin{pmatrix} d & 0 \\ 0 & M \end{pmatrix},$$

where M is a $(m-1) \times (n-1)$-matrix with entries in $\mathbb{Z}$, all divisible by d. (The two zeros denote the row and column of zeros of the relevant sizes.) We now apply the same arguments to M and finish the proof by induction. □

3.2. **Classification of finitely generated abelian groups.** For abelian groups it customary to write the group law additively, as $x + y$ rather than xy as we did before. We shall follow this convention. For example, an infinite cyclic group is isomorphic to the additive group $\mathbb{Z}$, so we'll write it as $\mathbb{Z}$ and use $+$ for the group law. If $g \in G$, where G is an abelian group, then ng is the shorthand for $g + g + \ldots + g$ (n times).

Definition 3.4. The free abelian group of rank *n is the product of n copies of $\mathbb{Z}$. It is denoted by $\mathbb{Z}^n$.*

In other words, $\mathbb{Z}^n = \{(a_1, \ldots, a_n) | a_i \in \mathbb{Z}\}$ is the set of ordered n-tuples of integers, with coordinate-wise addition as the group law. The unit element $(0, \ldots, 0)$ is written as 0. The inverse of $a = (a_1, \ldots, a_n)$ is $-a = (-a_1, \ldots, -a_n)$. By convention, the group of one element is the free abelian group of rank 0.

Proposition 3.5. *If $\mathbb{Z}^m \cong \mathbb{Z}^n$, then $n = m$. Hence the rank of a free abelian group of finite rank is a well defined integer.*

Proof. Let $f\colon \mathbb{Z}^m \xrightarrow{\sim} \mathbb{Z}^n$ be an isomorphism. Let $g_1, \ldots, g_m$ be the standard generators of $\mathbb{Z}^m$ (the coordinates of g_i are equal to 0 except the i-th coordinate which equals 1) and let $h_1, \ldots, h_n$ be the standard generators of $\mathbb{Z}^n$. Then each $f(g_j)$ is written as a linear combination of $h_1, \ldots, h_n$ with integer coefficients, i.e., $f(g_j) = a_{1j}h_1 + \ldots + a_{nj}h_n$ for some $a_{ij} \in \mathbb{Z}$. Consider the $n \times m$-matrix A whose j-th column is the transpose of $(a_{1j}, \ldots, a_{nj})$. Then f sends $x = (x_1, \ldots, x_m) \in \mathbb{Z}^m$ to A applied to the transpose of $(x_1, \ldots, x_m)$. The linear map $\mathbb{R}^m \to \mathbb{R}^n$ sending v to Av is surjective because its image contains a basis $h_1, \ldots, h_n$ of $\mathbb{R}^n$, as follows from the surjectivity of f. Hence

$$n = \dim_{\mathbb{R}}(\mathbb{R}^n) = \dim_{\mathbb{R}}(A(\mathbb{R}^m)) \leqslant \dim_{\mathbb{R}}(\mathbb{R}^m) = m,$$

where we used the following fact from linear algebra: the dimension of the image of a vector space under a linear map is not greater than the dimension of the source.

The situation is symmetric in n and m, hence $m \leqslant n$ so that $m = n$. □

See Problem Sheet 3 for a counting proof of this fact that does not use linear algebra.

Proposition 3.6. *Any subgroup of $\mathbb{Z}^n$ is isomorphic to $\mathbb{Z}^m$ for some $m \leqslant n$.*

Proof. We use induction on n. For $n = 1$ the statement is clear, because every subgroup of $\mathbb{Z}$ is the set of multiples of a given integer $a \in \mathbb{Z}$, hence is isomorphic either to $\mathbb{Z}$ (if $a \neq 0$) or to the 1-element group (if $a = 0$).

Assume that the statement is true for the subgroups of $\mathbb{Z}^{n-1}$. The function $f\colon \mathbb{Z}^n \to \mathbb{Z}$ which sends $(a_1, \ldots, a_n)$ to a_n is clearly a surjective homomorphism with kernel $\mathrm{Ker}(f) \cong \mathbb{Z}^{n-1}$.

Let G be a subgroup of $\mathbb{Z}^n$. The image $f(G)$ is a subgroup of $\mathbb{Z}$, hence $f(G) = \mathbb{Z}a$ for some $a \in \mathbb{Z}$. If $a = 0$, we have $G \subset \mathrm{Ker}(f) \cong \mathbb{Z}^{n-1}$, so we can conclude by appealing to the induction assumption. Assume $a \neq 0$. Choose $g \in G$ such that $f(g) = a$, that is, $g = (a_1, \ldots, a_n) \in G$ with some $a_1, \ldots, a_{n-1} \in \mathbb{Z}$.

Write $G_0 = G \cap \mathrm{Ker}(f)$. As an intersection of two subgroups of $\mathbb{Z}^n$, this is a subgroup of $\mathbb{Z}^n$. We have $G_0 \subset \mathrm{Ker}(f) \cong \mathbb{Z}^{n-1}$. Since G_0 is isomorphic to a subgroup of $\mathbb{Z}^{n-1}$, by the induction assumption, $G_0 \cong \mathbb{Z}^k$ for some $k \leqslant n-1$. I claim that G is isomorphic to the product of G_0 and the infinite cyclic group $\mathbb{Z}g$ generated by g. It is enough to prove this claim, because then $G \cong \mathbb{Z}^k \times \mathbb{Z} \cong \mathbb{Z}^{k+1}$ and we are done.

The claim is a consequence of Proposition 1.24. Indeed, we work with abelian groups, so all subgroups are normal. We have $G_0 \cap \mathbb{Z}g = \{0\}$. Finally, take any $h \in G$. Then $f(h) = ra$ for some $r \in \mathbb{Z}$, so $h = (h - rg) + rg$, where $h - rg \in G_0$ and $rg \in \mathbb{Z}g$. This proves that every element of G is the sum of an element of G_0 and an element of $\mathbb{Z}g$. This proves the claim. □

Corollary 3.7. *Let G be a finitely generated abelian group. Then there is a surjective homomorphism $f\colon \mathbb{Z}^n \to G$ for some n. We have* $\mathrm{Ker}(f) \cong \mathbb{Z}^m$ *for some $m \leqslant n$.*

Proof. Let $g_1, \dots, g_n$ be a set of elements of G that generates G. This means that G is the only subgroup of G that contains $g_1, \dots, g_n$. Define $f\colon \mathbb{Z}^n \to G$ as the map sending $(a_1, \dots, a_n)$ to $a_1g_1 + \ldots + a_ng_n \in G$. This is clearly a homomorphism. The image $f(\mathbb{Z}^n)$ is a subgroup of G containing $g_1, \dots, g_n$, so $f(\mathbb{Z}^n) = G$. The second claim follows from Proposition 3.6. □

Theorem 3.8. *Every finitely generated abelian group is isomorphic to a product of finitely many cyclic groups.*

Proof. By Corollary 3.7 we need to prove that if $H \cong \mathbb{Z}^m$ is a subgroup of $\mathbb{Z}^n$, then $\mathbb{Z}^n/H$ is isomorphic to a product of finitely many cyclic groups. Since $H \subset \mathbb{Z}^n$ is isomorphic to $\mathbb{Z}^m$, it can be generated by m elements of $\mathbb{Z}^n$. Let us write them as

$$(a_{11}, \dots, a_{1n}), \quad (a_{21}, \dots, a_{2n}), \quad \dots, \quad (a_{m1}, \dots, a_{mn}).$$

Consider the $(m \times n)$-matrix $A = (a_{ij})$. Since H is the subgroup of $\mathbb{Z}^n$ generated by the rows of A, the row operations do not change H. (This is obvious in the case of switching rows and multiplication of a row by -1. As H is generated by $r_1, \dots, r_m$, we see that H is also generated by the same elements with r_i replaced by $r_i + ar_j$ for any $a \in \mathbb{Z}$.) Each column operation is an automorphism $\varphi\colon \mathbb{Z}^n \xrightarrow{\sim} \mathbb{Z}^n$. An automorphism sends a subgroup to an isomorphic subgroup, $\varphi\colon H \xrightarrow{\sim} \varphi(H)$. Thus φ induces a map $gH \mapsto \varphi(g)\varphi(H)$ on left cosets $\mathbb{Z}^n/H \xrightarrow{\sim} \mathbb{Z}^n/\varphi(H)$, which is an isomorphism. We conclude that performing row and column operations on A does not change the quotient group $\mathbb{Z}^n/H$, up to isomorphism.

By Theorem 1.19 we can assume that A is in Smith normal form with non-zero diagonal entries $d_1|d_2|\dots|d_k$. The group generated by the rows of A is

$$H = \mathbb{Z}d_1 \times \mathbb{Z}d_2 \times \ldots \times \mathbb{Z}d_k \times \{0\} \times \ldots \times \{0\} \subset \mathbb{Z}^n.$$

The quotient of $\mathbb{Z}$ by the subgroup $\mathbb{Z}d$, where $d \neq 0$, is isomorphic to the cyclic group C_d of order d. By Question 6 (b) from Problem Sheet 1 we obtain an isomorphism

$$\mathbb{Z}^n/H \cong C_{d_1} \times \ldots \times C_{d_k} \times \mathbb{Z}^{n-k}.$$

This finishes the proof. □

Remark 3.9. Let G be a finitely generated abelian group. By Theorem 3.8 we have an isomorphism $G \cong F \times H$, where F is a finite subgroup of G and $H \subset G$ is a free abelian group of some rank m. We note that $F = G_{\text{tors}}$ so is uniquely defined by G, see Definition 1.26. In contrast, H is not unique (unless G is a free abelian group), because if $x \in F$, $x \neq 0$, and $y \in H$ is infinite, then the order of $x + y$ is also infinite but $x + y \notin H$. However, the rank m of H is well defined, because it equals the rank of the free abelian group $G/F \simeq H$ which is well defined by Proposition 3.5. This non-negative integer is called **the rank** of G.

Corollary 3.10. *Any finite abelian group is isomorphic to the product of its p-primary torsion subgroups.*

Proof. By Theorem 3.8, a finite abelian group is isomorphic to a product of cyclic subgroups. Now use Corollary 1.28. (This can be proved directly using the argument in the proof of Corollary 1.28, so Theorem 3.8 is not needed. I am grateful to Xiang Li for pointing this out.) □

Theorem 3.11. *Every finitely generated abelian group is isomorphic to a product of finitely many infinite cyclic groups and finitely many cyclic groups of prime power order. The number* *of infinite cyclic factors and the number of cyclic factors of order p^r, where p is a prime and r is a positive integer, depend only on the group.*

Proof. Let G be a finitely generated abelian group. Theorem 3.8 gives an isomorphism $G \cong G_{\text{tors}} \times \mathbb{Z}^m$, for some integer $m \geqslant 0$, where G_{tors} is a finite group. By Corollary 3.10 we have $G_{\text{tors}} \cong \prod_p G\{p\}$ where p ranges over the prime factors of $|G_{\text{tors}}|$. Each p-primary torsion subgroup $G\{p\}$ is the set of elements of G whose order is a power of p, so $G\{p\}$ is a well defined subgroup of G. By Theorem 3.8, $G\{p\}$ is isomorphic to a product of cyclic p-groups. It remains to show that the collection of prime powers which are orders of these cyclic p-groups is well defined. This boils down to the following claim: if we have an isomorphism

$$C_{p^{a_1}} \times \ldots \times C_{p^{a_m}} \cong C_{p^{b_1}} \times \ldots \times C_{p^{b_k}}, \tag{3.1}$$

where $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_m \geqslant 1$ and $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_k \geqslant 1$, then $m = k$ and $a_i = b_i$ for all i. Call this group H and let $H[p] = \{x \in H | px = 0\}$. We have $C_{p^a}[p] \cong C_p$, thus $|H[p]| = p^m = p^k$ so $m = k$. Now let $pH = \{px \in H | x \in H\}$. We have $pC_{p^a} \cong C_{p^{a-1}}$, hence pH is isomorphic to

$$C_{p^{a_1-1}} \times \ldots \times C_{p^{a_m-1}} \cong C_{p^{b_1-1}} \times \ldots \times C_{p^{b_k-1}}.$$

Here we can ignore the factors which are 1-element groups, that is, the factors with $a_i = 1$ and $b_j = 1$. Using the previous argument we see that the number of a_i's such that $a_i \geqslant 2$ is equal to the number of b_j's such that $b_j \geqslant 2$. Thus the number of cyclic groups of order p is the same on both sides of (3.1). We apply the same argument to pH and obtain that the number of cyclic groups of order p^2 is the same on both sides of (3.1). Continuing this process, after finitely many steps we prove the claim. □

GROUPS AND RINGS: RINGS

ALEXEI N. SKOROBOGATOV

Three Rings for the Elven-kings under the sky,
Seven for the Dwarf-lords in their halls of stone,
Nine for Mortal Men, doomed to die,
One for the Dark Lord on his dark throne
In the Land of Mordor where the Shadows lie.

Contents

This is the second part of the course.

4. Basic theory of rings

4.1. **Motivation, definitions, examples.** The world of numbers has two operations: addition and multiplication. Groups are objects endowed with one operation, which in concrete situations can be addition, multiplication or any other, provided it satisfies the axioms of a group. But in most mathematical problems we need to deal with two operations at the same time. A ring is a formal algebraic structure, like a group, but with two operations.

Traditionally, these operations are called addition and multiplication, and are denoted by $+$ and $\times$, respectively. There is no symmetry between $+$ and $\times$: we require addition to satisfy group axioms and to be commutative, whereas multiplication does not have to be commutative and the multiplicative inverses do not usually exist.

The crucial examples of rings, as we shall see soon, are as follows.

- The set of integers $\mathbb{Z}$.
- The sets of rational, real or complex numbers $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.
- The set of polynomials with rational coefficients $\mathbb{Q}[t]$.

Date: September 14, 2021.

- The set $M_n(\mathbb{R})$ of $(n \times n)$-matrices with entries in $\mathbb{R}$, where $n \geqslant 2$.
- The binary field $\mathbb{F}_2 = \{0, 1\}$ with binary addition and multiplication.

With the exception of the last one, each of these sets has usual addition and multiplication.

Definition 4.1. *A ring is a set R together with two binary operations, $+$ and $\times$, satisfying the following axioms:*

(1) *$(R, +)$ is an abelian group. It is written additively, so the unit element of $(R, +)$ is denoted by 0 and the inverse of x is denoted by $-x$.*

(2) *Multiplication is* **associative***: for any $a, b, c \in R$ we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.*

(3) *There is a unique* **unit element** *for multiplication, denoted by 1, which satisfies $1 \cdot x = x \cdot 1 = x$ for any $x \in R$.*

(4) **Distributivity***: for any $a, b, c \in R$ we have $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.*

The ring R is closed under $+$ as well as under $\times$. You won't fail to notice that we do not require the existence of multiplicative inverses. This is OK because we want to be able to work with objects like $\mathbb{Z}$ where there are many non-invertible elements (only 1 and -1 have multiplicative inverses).

Let us show how addition and multiplication interact.

Lemma 4.2. *Let R be a ring.*

(i) *For any $x \in R$ we have $x0 = 0x = 0$.*

(ii) *For any $x, y \in R$ we have $(-x)y = x(-y) = -xy$.*

(iii) *If $R \neq \{0\}$, then $1 \neq 0$.*

Proof. We have $0+0 = 0$, hence by axiom (4) we have $x \cdot 0 = x(0+0) = x \cdot 0 + x \cdot 0$. Adding to both sides the additive inverse of $x \cdot 0$ we obtain $x \cdot 0 = 0$. Using the other distributivity law we prove $0 \cdot x = 0$, hence complete the proof of (i).

We have $y + (-y) = 0$, so by axiom (4) we get $x \cdot 0 = x(y + (-y)) = xy + x(-y)$. By part (i) we have $x \cdot 0 = 0$. Adding to both sides the additive inverse of xy we get $-xy = x \cdot (-y)$. A similar proof gives $(-x)y = -xy$.

Assume that $1 = 0$. Now (i) and axiom (3) imply that 0 is the only element of R. □

Henceforth we shall only consider non-zero rings, i.e., rings such that $1 \neq 0$.

Definition 4.3. *A subset of a ring which is a ring under the same operations and the same 1 is called a* **subring**.

Lemma 4.4. *Let S be a non-empty subset of a ring R. Then S is a subring of R if and only if $1 \in S$ and for any $a, b \in S$ we have $a + b \in S$, $ab \in S$ and $-a \in S$.*

Proof. A subring has these properties. Conversely, if S is closed under addition and taking the additive inverse, then $(S, +)$ is a subgroup of $(R, +)$ (by group theory). Associativity and distributivity hold in S because they hold in R. □

Definition 4.5. *A ring R is called* **commutative** *if $xy = yx$ for any $x, y \in R$*

Definition 4.6. *An element $x \in R$ is called* **invertible** *if there are elements $y, z \in R$ such that $xy = 1$ and $zx = 1$.*

Remark 4.7. We have $y = z$. Indeed, $z = z \cdot 1 = z(xy) = (zx)y = 1 \cdot y = y$. One denotes $y = z$ by x^{-1}. The set of all invertible elements of R is denoted by $R^\times$. This set satisfies the group axioms and is called the **multiplicative group** of the ring R.

Definition 4.8. *A ring in which every non-zero element is invertible is called a* **division ring***. A commutative division ring is called a* **field***.*

Thus a field has all the desirable properties: in a field one can add, subtract, multiply and divide (by arbitrary non-zero elements). $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_2$ are fields, whereas $\mathbb{Z}$ and $\mathbb{Q}[t]$ are not. All these rings are commutative. For $n \geqslant 2$ the ring of matrices $M_n(\mathbb{R})$ is not commutative and is not a division ring.

Let us consider more examples of rings.

Example 4.9. Let X be a set and let R be a ring. The set of functions $X \to R$ is a ring with respect to addition and multiplication of functions defined as follows:

$$(f+g)(x) \colon = f(x) + g(x), \quad (fg)(x) \colon = f(x)g(x).$$

The unit element for addition is the function which is identically 0, and the unit element for multiplication is the function which is identically 1.

Example 4.10. For a ring R let $M_n(R)$ be the set of $(n \times n)$-matrices with entries in R. Usual addition and multiplication of matrices make $M_n(R)$ a ring.

Example 4.11. Let A be an abelian group (written additively, as is our convention). Let $\mathrm{End}(A)$ be the set of endomorphisms $A \to A$ (i.e., homomorphisms from A to itself). Define the addition of endomorphisms as the addition of functions, that is, $(f+g)(x) \colon = f(x)+g(x)$. Define the multiplication of endomorphisms as composition. Note that this multiplication is not in general commutative.

4.2. **Homomorphisms, ideals, quotient rings.** In analogy with the theory of groups, where homomorphisms of groups are maps between groups that preserve the group law, we can define homomorphisms of rings.

Definition 4.12. *Let R and S be rings. A function $f \colon R \to S$ is a* **homomorphism** *of rings if*

(1) *$f \colon (R, +) \to (S, +)$ is a homomorphism of abelian groups;*
(2) *$f(xy) = f(x)f(y)$ for all $x, y \in R$;*
(3) *$f(1_R) = 1_S$.*

Here 1_R and 1_S are the unit elements for multiplications, in R and S, respectively. Since in all of our rings we have $1 \neq 0$, any homomorphism is a non-zero map. In other words, a homomorphism of rings is a homomorphism of their additive groups which preserves multiplication and sends 1_R to 1_S.

A subset R' of a ring R is a subring if and only if the tautological map $R' \to R$ is a homomorphism of rings.

Lemma 4.13. *Let $f \colon R \to S$ be a homomorphism of rings. The kernel $\mathrm{Ker}(f)$ is a subgroup of $(R, +)$ which satisfies the following property: for any $x \in \mathrm{Ker}(f)$ and any $r \in R$ we have $xr \in \mathrm{Ker}(f)$ and $rx \in \mathrm{Ker}(f)$.*

Proof. This follows from group theory and Lemma 4.2 (i). □

Here are some examples of homomorphisms.

Example 4.14. Let m be a positive integer. We have a subgroup $m\mathbb{Z}$ of $\mathbb{Z}$. Consider the homomorphism of abelian groups $f \colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ sending $n \in \mathbb{Z}$ to the coset $n + m\mathbb{Z}$ of the subgroup $m\mathbb{Z}$. We can choose a coset representative $\bar{n}$ as the unique integer such that $n-\bar{n}$ is a multiple of m and $0 \leqslant \bar{n} < m$. Then $f(n)$ is the coset of $\bar{n}$.

But $\mathbb{Z}$ is not just a group under addition, it is a ring. Then $\mathbb{Z}/m\mathbb{Z}$ inherits multiplication from $\mathbb{Z}$: this is the operation defined by the rule $(n+m\mathbb{Z})\cdot(k+m\mathbb{Z}) = nk+m\mathbb{Z}$. (This is the same as defining $\bar{n}\cdot\bar{k}$ as $\overline{nk}$.) This operation is well defined, because if n' is any element of $n+m\mathbb{Z}$ and k' is any element of $k+m\mathbb{Z}$, then $n'k'$ differs from nk by a multiple of m, so $nk+m\mathbb{Z} = n'k'+m\mathbb{Z}$ which means that the product of cosets does not depend on the choice of representatives and thus is well defined.

This multiplication makes $\mathbb{Z}/m\mathbb{Z}$ a ring with $\bar{1} = 1+m\mathbb{Z}$ as the unit element for multiplication. (Associativity of multiplication and distributivity hold in $\mathbb{Z}/m\mathbb{Z}$ because they hold in $\mathbb{Z}$.) Now it is clear that $f\colon \mathbb{Z}\to\mathbb{Z}/m\mathbb{Z}$ is a surjective homomorphism of rings.

Example 4.15. Let us consider a similar situation where $\mathbb{Z}$ is replaced by the polynomial ring $\mathbb{Q}[t]$ and m is replaced by a polynomial $p(t)$ of degree at least 1. We follow the same procedure and consider $\mathbb{Q}[t]$ as an abelian group with subgroup $p(t)\mathbb{Q}[t]$ consisting of polynomials divisible by $p(t)$. The quotient group $\mathbb{Q}[t]/p(t)\mathbb{Q}[t]$ inherits multiplication from $\mathbb{Q}[t]$, which turns it into a ring. The coset $1+p(t)\mathbb{Q}[t]$ of the polynomial 1 is the unit element of $\mathbb{Q}[t]/p(t)\mathbb{Q}[t]$. The canonical surjective homomorphism of abelian groups $\mathbb{Q}[t]\to\mathbb{Q}[t]/p(t)\mathbb{Q}[t]$ sending a polynomial $q(t)$ to its coset $q(t)+p(t)\mathbb{Q}[t]$ is then a surjective homomorphism of rings. (Note that $\deg(p(t)) \geqslant 1$ implies that $p(t)\mathbb{Q}[t] \neq \mathbb{Q}[t]$, so the ring $\mathbb{Q}[t]/p(t)\mathbb{Q}[t]$ is non-zero.)

These examples are particular cases of a general construction.

Definition 4.16. *Let R be a ring. A subset $I \subset R$ is called an* **ideal** *if it is a subgroup of $(R,+)$ (with respect to addition) and such that for any $x \in I$ and any $r \in R$ we have $rx \in I$ and $xr \in I$.*

We shall mostly consider commutative rings, and in this case $rx = xr$, so one condition $rx \in I$ is enough. In the non-commutative case what we defined above is called a *two-sided ideal*, whereas if we only require $rx \in I$, then I is called a *left ideal*, and if only require $xr \in I$, then I is called a *right ideal.*

In our definition R is an ideal of R. An ideal not equal to the whole ring is called a *proper* ideal. Another standard example is the zero ideal $\{0\}$.

In the theory of rings, ideals play a role similar to that of normal subgroups in group theory. One common feature is that kernels of homomorphisms of rings are ideals (in the same way as kernels of homomorphisms of groups are normal subgroups). This follows from Lemma 4.13. Note, however, that a proper ideal $I \subsetneq R$ is not a subring of R because it does not contain 1. Indeed, if $1 \in I$ then $r\cdot 1 = r \in I$ for any $r \in R$, so $I = R$.

Another common feature is that for a proper ideal I of a ring R we can define the quotient ring R/I and a canonical surjective homomorphism $f\colon R\to R/I$. Indeed, we take the quotient group R/I of the additive group of R by the subgroup I. Then we turn R/I into a ring like we did in the examples above, namely, the product of cosets $x+I$ and $y+I$ is defined as the coset $xy+I$. A standard verification shows that this operation is well defined. The unit element for multiplication in R/I is $1+I \in R/I$. Associativity of multiplication and distributivity hold in R/I because they hold in R. The surjective homomorphism of additive groups $f\colon R\to R/I$ is thus a homomorphism of rings.

Definition 4.17. *Let R be a ring and let $I \subset R$ be a proper ideal. The quotient abelian group R/I with multiplication inherited from the multiplication on R is a ring called the* **quotient ring** *of R by the ideal I.*

This generalises two examples above. In the case when $R = \mathbb{Z}$ we take the ideal $I = m\mathbb{Z}$, and the quotient ring is $R/I = \mathbb{Z}/m\mathbb{Z}$. In the case $R = \mathbb{Q}[t]$ we take the ideal $I = p(t)\mathbb{Q}[t]$. Such ideals have a special name.

Definition 4.18. *Let R be a commutative ring. Take any $a \in R$ and consider the set of all multiples of a, that is, the set $aR = \{ax|x \in R\}$. This is an ideal in R. An ideal of this form is called a* **principal ideal** *with* **generator** *a.*

It is indeed clear that aR is an ideal: this is a subgroup of $(R, +)$ which is closed under multiplication by arbitrary elements of R. A generator is usually not unique.

As in group theory, we have the following definitions.

Definition 4.19. *A bijective homomorphism of rings $f\colon R \to S$ is called an* **isomorphism**.

A homomorphism of rings $R \to R$ is called an **endomorphism**.

An isomorphism of rings $R \xrightarrow{\sim} R$ is called an **automorphism**.

Continuing the analogy with groups we note that the image of a homomorphism of rings $f\colon R \to S$ is a subring of S. Indeed, $f(R)$ is a subgroup of the additive group of S, contains 1_S and is closed under multiplication. (It is not an ideal unless $f(R) = S$.)

Theorem 4.20 (Isomorphism theorem)**.** *Let $f\colon R \to S$ be a homomorphism of rings. Then the subring $f(R)$ of S is isomorphic to the quotient ring $R/\mathrm{Ker}(f)$.*

Proof. The isomorphism theorem from group theory (Theorem 1.19) says that the map sending $x + \mathrm{Ker}(f)$ to $f(x)$ is an isomorphism of groups under addition $R/\mathrm{Ker}(f) \xrightarrow{\sim} f(R)$. This map respects multiplication and sends 1 to 1, so it is an isomorphism of rings. □

4.3. **Integral domains and fields.**

Definition 4.21. *Let R be a ring. Non-zero elements $a, b \in R$ are called* **zero-divisors** *if $ab = 0$. A commutative ring without zero-divisors is called an* **integral domain**.

Lemma 4.22. *Let R be an integral domain and let $a, b, c \in R$ be such that $a \neq 0$. Then $ab = ac$ if and only if $b = c$.*

Proof. One direction is obvious. So we assume $ab = ac$. Then by distributivity we have $a(b - c) = 0$. Since R has no zero-divisors, we must have $b - c = 0$. □

Lemma 4.23. *Let R be an integral domain and let $a, b \in R$. Then $aR = bR$ if and only if $a = br$, where $r \in R^\times$.*

Proof. Assume $aR = bR$. If $a = 0$, then $b = b \cdot 1 = 0$, so the conclusion is true in this case. Now let $a \neq 0$. We have $a = a \cdot 1 \in aR = bR$. Thus $a = bc$ for some $c \in R$. Similarly, $b = ad$ for some $d \in R$. Thus $a = acd$. By Lemma 4.22 we obtain $cd = 1$, hence $c \in R^\times$.

Conversely, if $a = br$, where $r \in R^\times$, then $aR \subset bR$, but we can also write $b = ar^{-1}$ and this implies $bR \subset aR$, so we are done. □

Proposition 4.24. *Every field is an integral domain.*

Proof. Exercise. □

Theorem 4.25. *Every* **finite** *integral domain is a field.*

Proof. The only thing to check is that every non-zero element is invertible. Let $R = \{r_1, ..., r_n\}$ (distinct elements) be our integral domain. Take any non-zero $r \in R$. Consider $\{r_1 r, ..., r_n r\}$. If for some i and j we have $r_i r = r_j r$ then $r_i = r_j$ by the cancellation property (Lemma 4.22). Therefore $\{r_1 r, ..., r_n r\}$ is a set of n distinct elements of R. Since R has n elements, $\{r_1 r, ..., r_n r\} = R = \{r_1, ..., r_n\}$. Thus any r_i can be written as $r_j r$ for some j. In particular, $1 = r_j r$ for some j, hence $r_j = r^{-1}$. □

Corollary 4.26. *Let n be a positive integer. The ring $\mathbb{Z}/n\mathbb{Z}$ is an integral domain if and only if it is a field, which happens exactly when n is a prime.*

Proof. By Theorem 4.25 it is enough to prove that $\mathbb{Z}/n\mathbb{Z}$ is *not* an integral domain if and only if n is *not* a prime.

If $n = 1$, then $\mathbb{Z}/n\mathbb{Z}$ is the zero ring, hence not an integral domain (by definition).

Now assume $n \geqslant 2$. If $n = ab$, where $a, b \in \mathbb{Z}$, $a > 1$, $b > 1$, then $\bar{a}\bar{b} = \overline{ab} = \bar{n} = 0$, so $\bar{a}$ and $\bar{b}$ are zero-divisors in $\mathbb{Z}/n\mathbb{Z}$, so this is not an integral domain. Conversely, if $\bar{r}$ and $\bar{s}$ are zero-divisors in $\mathbb{Z}/n\mathbb{Z}$, then n divides neither r nor s, but divides rs. But we know that if a prime divides a product of two natural numbers, then it divides one of them. Hence n is not a prime. □

Thus for any prime p we have a finite field with p element. We denote $\mathbb{Z}/p\mathbb{Z}$ by $\mathbb{F}_p$.

Definition 4.27. *A subset K of a field F is called a* **subfield** *of F if K is a field with the same addition and multiplication. In this case, F is called a* **field extension** *of K.*

To check that $K \subset F$ is a subfield of a field F, it is enough to check that for any $a, b \in K$ the elements $a + b, -a, ab$ are in K, and for any non-zero $a \in K$ we have $a^{-1} \in K$.

Proposition 4.28. *For any ring R there is a unique homomorphism of rings $\mathbb{Z} \to R$.*

Proof. A homomorphism of rings $f \colon \mathbb{Z} \to R$ must send 0 to 0 and 1 to 1_R. Then, by definition, $f(2) = 1_R + 1_R$, $f(3) = 1_R + 1_R + 1_R$, and so on, and also $f(-1) = -1_R$, $f(-2) = -(1_R + 1_R)$. By induction, if n is a positive integer, then $f(n)$ is obtained by adding 1_R with itself n times; if n is a negative integer, then $f(n) = -f(-n)$. Thus for any ring there is a unique homomorphism $\mathbb{Z} \to R$, namely, the one defined above. □

Lemma 4.29. *Let R be an integral domain. The kernel of the unique homomorphism $\mathbb{Z} \to R$ is either the zero ideal $\{0\} \subset \mathbb{Z}$ or the principal ideal $p\mathbb{Z}$, where p is a prime.*

Proof. The kernel is $n\mathbb{Z}$ for some $n \in \mathbb{Z}$, because there are no other ideals in $\mathbb{Z}$ (indeed, all subgroups of $(\mathbb{Z}, +)$ are of this form). We have $n = 0$ or $n \geqslant 2$ (because 1 goes to 1_R). By the isomorphism theorem, the image of $\mathbb{Z}$ is a subring of R isomorphic to $\mathbb{Z}/n\mathbb{Z}$. By Corollary 4.26 if $n > 0$, then n is a prime. □

Definition 4.30. *The* **characteristic** *of an integral domain R is the unique non-negative generator of the kernel of a homomorphism $\mathbb{Z} \to R$, so it is 0 or a prime number.*

We denote the characteristic of R by $\mathrm{char}(R)$.

Recall the definition of a vector space over a field from linear algebra. Note that we can talk about vector spaces over any given field, not only over $\mathbb{R}$ as in the first year linear algebra.

Definition 4.31. *Let k be a field. Let V be an abelian group together with an action of the elements of k (called 'scalars') on the elements of V (called 'vectors'), that is, a rule attaching to a scalar $x \in k$ and a vector $v \in V$ a vector $xv \in V$, satisfying the following axioms:*

(1) *$1v = v$ and $x(yv) = (xy)v$ for any $x, y \in k$ and any $v \in V$;*
(2) *$(x + y)v = xv + yv$ for any $x, y \in k$ and any $v \in V$;*
(3) *$x(v + w) = xv + xw$ for any $x \in k$ and any $v, w \in V$.*

Lemma 4.32. *A field extension F of a field k is a vector space over k.*

Proof. The axioms of a vector space obviously hold. □

Theorem 4.33. *Let k be a field. If $\mathrm{char}(k) = 0$, then k contains a unique subfield isomorphic to $\mathbb{Q}$ so k is a vector space over $\mathbb{Q}$. If $\mathrm{char}(k) = p$ (a prime), then k contains a unique subfield isomorphic to $\mathbb{F}_p$, so k is a vector space over $\mathbb{F}_p$.*

Proof. If $\mathrm{char}(k) = 0$, then k contains a subring isomorphic to $\mathbb{Z}$. This is the smallest subring containing 1. Since k is a field, it contains multiplicative inverses of all non-zero elements, hence all ratios of non-zero elements of $\mathbb{Z}$, and this set is a field isomorphic to $\mathbb{Q}$. It remains to use the previous lemma. If $\mathrm{char}(k) = p$, then the statement is clear. □

Corollary 4.34. *Every finite field has p^n elements, where p is a prime and n is a positive integer.*

Proof. Such a field k is a vector space over $\mathbb{F}_p$, by Theorem 4.33. Since k is finite, it is spanned by finitely many vectors. By linear algebra, k has a finite basis $v_1, \ldots, v_n$ for some $n \geqslant 1$. Then every element of k is uniquely written as $a_1v_1 + \ldots + a_nv_n$, where $a_i \in \mathbb{F}_p$ for all $i = 1, \ldots, n$. Hence $|k| = p^n$. □

This prompts an interesting question: for a given prime power p^n, does there exist a field with p^n elements? If yes, how to construct it explicitly? We shall answer both questions later in this course.

4.4. **More on ideals.**

Proposition 4.35. *A commutative ring is a field if and only if the only proper ideal is the zero ideal.*

Proof. If an ideal of a field contains a non-zero element, then it equals to the whole field, because every non-zero element in a field is invertible. Conversely, let R be a commutative ring and let $a \in R$ be a non-zero element. If R has no non-zero proper ideals, then the principal ideal aR equals R. Then $1 = ab$ for some $b \in R$, hence $a \in R^\times$. Thus R is a field. □

Proposition 4.36. *Let $f\colon R \to S$ be a homomorphism of rings and let $J \subset S$ be an ideal. Then $f^{-1}(J)$ is an ideal of R.*

Proof. By group theory, the inverse image of a subgroup is a subgroup, hence $f^{-1}(J)$ is a subgroup of $(R, +)$. If $x \in f^{-1}(J)$ and $r \in R$, then $f(rx) = f(r)f(x)$. Since $f(x) \in J$, we have $f(r)f(x) \in J$, thus $rx \in f^{-1}(J)$. □

Note that the image of an ideal under a homomorphism of rings is not necessarily an ideal. For example, the map sending $n \in \mathbb{Z}$ to $n \in \mathbb{Q}$ is an injective homomorphism $\mathbb{Z} \to \mathbb{Q}$. But $2\mathbb{Z}$ is not an ideal of $\mathbb{Q}$. However, this is true for surjective homomorphisms. (Compare the following with a similar statement for groups, see Proposition 1.20.)

Proposition 4.37. *Let $f\colon R \to S$ be a* **surjective** *homomorphism of rings and let $I \subset R$ be an ideal. Then $f(I)$ is an ideal of S. The maps $I \mapsto f(I)$ and $J \mapsto f^{-1}(J)$ are inverse to each other, so they give a bijection between the ideals of R that contain* $\mathrm{Ker}(f)$ *and the ideals of S.*

Proof. Let $x \in I$. To prove that $f(I)$ is an ideal, we need to check that for any $s \in S$ we have $sf(x) = f(y)$ for some $y \in R$. By the surjectivity of f we find an element $r \in R$ such that $f(r) = s$. Then $y = rx$.

By the previous proposition, for any ideal $J \subset S$ the inverse image $f^{-1}(J)$ is an ideal of R that contains $\mathrm{Ker}(f)$. Since f is surjective, we have $f(f^{-1}(J)) = J$. If $\mathrm{Ker}(f) \subset I$, then $f^{-1}(f(I)) = I$. (The inclusion $I \subset f^{-1}(f(I))$ always holds. For the reverse inclusion note that if $x \in f^{-1}(f(I))$, then $f(x) = f(y)$ for some $y \in I$. Then $x - y \in \mathrm{Ker}(f) \subset I$, hence $x \in I$.) □

Let us define two most important classes of ideals.

Definition 4.38. *Let R be a commutative ring. A proper ideal $I \subset R$ called a* **prime ideal** *if the quotient ring R/I is an integral domain.*

For example, the prime ideals of $\mathbb{Z}$ are $p\mathbb{Z}$, where p is a prime number, and the zero ideal $\{0\}$.

Proposition 4.39. *A proper ideal I of a commutative ring R is prime if and only if for any $x, y \in R$ such that $xy \in I$ we have $x \in I$ or $y \in I$.*

Proof. The property is equivalent to the property that $(x + I)(y + I) = xy + I$ equals I if and only if $x + I = I$ or $y + I = I$. But this is exactly the same as the property that R/I has no zero-divisors. (Note that I is proper, hence R/I is not the zero ring.) □

Definition 4.40. *Let R be a commutative ring. A proper ideal $I \subset R$ called a* **maximal ideal** *if the quotient ring R/I is a field.*

It is clear that every maximal ideal is prime.

Proposition 4.41. *A proper ideal I of a commutative ring R is maximal if and only if there is no proper ideal $J \subset R$ such that $I \subset J$ and $I \neq J$.*

Proof. By Proposition 4.35, the ring R/I is a field if and only if R/I has no non-zero proper ideals. By Proposition 4.37 this is equivalent to the absence of proper ideals of R that strictly contain I. □

The maximal ideals of $\mathbb{Z}$ are $p\mathbb{Z}$, where p is a prime number (but not the zero ideal). This proposition suggests that we can construct fields as quotients of commutative rings by their maximal ideals, much in the same way as the finite fields $\mathbb{F}_p$ are quotients of $\mathbb{Z}$ by $p\mathbb{Z}$. To do this need to enlarge our supply of rings.

5. PID and UFD

5.1. **Polynomial rings.** Let R be an integral domain and let $R[t]$ be the ring of polynomials in one variable t with coefficients in R, with the usual addition and multiplication of polynomials. Every polynomial can be written as

$$p(t) = a_n t^n + \ldots + a_1 t + a_0,$$

where all $a_i \in R$ and $a_n \neq 0$. In this case we call n the degree of $p(t)$ and write $\deg(p(t)) = n$.

Proposition 5.1. *If R is an integral domain, then*

$$\deg(p(t)q(t)) = \deg(p(t)) + \deg(q(t)), \tag{5.1}$$

$R[t]$ is an integral domain, and $R[t]^\times = R^\times$.

Proof. Formula (5.1) is due to the absence of zero-divisors in R, so the product of two leading coefficients is non-zero. It implies that $R[t]$ has no zero-divisors and that the invertible elements of $R[t]$ are the constant polynomials which are invertible elements of R. □

We shall mostly consider polynomial rings with coefficients in a field k. A key feature of the ring $k[t]$ is the possibility to divide with remainder, exactly like in $\mathbb{Z}$.

Proposition 5.2. *Let k be a field. For any polynomials $a(t), b(t) \in k[t]$, where $b(t)$ is non-zero, there exist polynomials $q(t), r(t) \in k[t]$ such that*

$$a(t) = q(t)b(t) + r(t)$$

where either $r(t) = 0$ or $\deg(r(t)) < \deg(b(t))$. These $q(t)$ and $r(t)$ are uniquely determined by $a(t)$ and $b(t)$.

Proof. Let $m = \deg(a(t))$ and $n = \deg(b(t))$. If $m < n$ we let $q(t)$ be the zero polynomial and $r(t) = a(t)$. So assume $m \geqslant n$ and use induction in m. Assume that this is proved for degrees less than m. Write a_m (respectively, b_n) for the leading coefficient of $a(t)$ (respectively, of $b(t)$). The degree of $a(t) - a_m b_n^{-1} t^{m-n} b(t)$ is less than m, so we can apply the induction assumption and finish the proof of the existence part.

If $a(t) = \tilde{q}(t)b(t) + \tilde{r}(t)$, where $\tilde{r}(t) = 0$ or $\deg(\tilde{r}(t)) < n$, then $r(t) - \tilde{r}(t)$ has degree less n but is a multiple of a polynomial of degree n, hence $r(t) = \tilde{r}(t)$. Then $q(t)b(t) = \tilde{q}(t)b(t)$, and this implies $q(t) = \tilde{q}(t)$ by the cancellation property of integral domains. □

An obvious consequence of this is that if a polynomial $p(t) \in k[t]$ has a root $\alpha \in k$, then $p(t) = q(t)(t - \alpha)$ for some $q(t) \in k[t]$. Iterating this we see that a polynomial of degree d can have at most d roots in k.

Definition 5.3. *An integral domain R with a function $\phi\colon R\backslash\{0\} \rightarrow \mathbb{Z}_{\geqslant 0}$ is called a* **Euclidean domain** *if*

(1) *$\phi(xy) \geqslant \phi(x)$ for any non-zero $x, y \in R$;*

(2) *for any $a, b \in R$ there exist $q, r \in R$ such that $a = qb + r$ where $r = 0$ or $\phi(r) < \phi(b)$.*

Examples of Euclidean domains:

- the ring $\mathbb{Z}$ together with $\phi(n) = |n|$;
- the ring $k[t]$, where k is a field, together with the degree function;
- the ring of Gaussian integers $\mathbb{Z}[i] = \{m + ni | m, n \in \mathbb{Z}\}$, where $i = \sqrt{-1}$, with $\phi(m + ni) = m^2 + n^2$;
- the ring of Eisenstein integers $\mathbb{Z}[\zeta] = \{m + n\zeta | m, n \in \mathbb{Z}\}$, where $\zeta = \frac{-1+\sqrt{-3}}{2}$, with $\phi(m + n\zeta) = m^2 - mn + n^2$.

For the last two examples, see Problem Sheet 6.

Definition 5.4. *An integral domain R is called a* **principal ideal domain** *or a* **PID** *if every ideal of R is principal, that is, is of the form aR for some $a \in R$.*

Theorem 5.5. *Any Euclidean domain is a PID.*

Proof. Let $I \subset R$ be a non-zero ideal. Let $b \in I$ be a non-zero element such that $\phi(b)$ is the minimum of $\phi(x)$, $x \in I$, $x \neq 0$. Any $a \in I$ can be written as $a = qb + r$, where $r = 0$ or $\phi(r) < \phi(b)$. But $r = a - qb \in I$ (since I is an ideal) so $\phi(r) < \phi(b)$ is impossible. Thus every element of I is a multiple of b, so $I = bR$. □

5.2. **Factorisation in integral domains.**

Definition 5.6. *Let R be an integral domain. A non-zero element $x \in R\backslash R^\times$ is called an* **irreducible element** *if x is not a product of two elements of $R\backslash R^\times$.*

For example, the irreducible elements in $\mathbb{Z}$ are $\pm p$, where p is a prime number. The irreducible elements of $k[t]$ are called irreducible polynomials.

Lemma 5.7. *Let R be an integral domain. If x is an irreducible element and $a \in R^\times$, then ax is also an irreducible element.*

Proof. Indeed, $ax \neq 0$ cannot be in $R^\times$ because then $x \in R^\times$. Next, if $ax = yz$, where $y, z \in R\backslash R^\times$, then $x = (a^{-1}y)z$. This cannot happen since $a^{-1}y \in R\backslash R^\times$. □

Definition 5.8. *An integral domain R is called a* **unique factorisation domain** *or a* **UFD** *if every element of $R\backslash R^\times$ is a product of finitely many irreducibles, and this decomposition is unique up to changing the order of factors and multiplying the factors by elements of $R^\times$.*

The main theorem of arithmetic says that $\mathbb{Z}$ is a UFD. The polynomial ring $\mathbb{C}[t]$ with coefficients in the field of complex numbers $\mathbb{C}$ is a UFD because every polynomial is uniquely written as $c\prod_{i=1}^n (t - z_i)$, where $c \in \mathbb{C}^\times$ and $z_i \in \mathbb{C}$ for $i = 1, \ldots, n$, up to permutation of factors.

Unique factorisation domains are also sometimes called *factorial* rings.

Our goal is to show that any PID is a UFD.

Definition 5.9. *Let R be an integral domain and let $a, b \in R$. We say that $a \in R$* **divides** *$b \in R$ and write $a|b$ if $b = ra$ for some $r \in R$. An element $a \in R$* **properly divides** *$b \in R$ if $b = ra$ and $r \notin R^\times$. If $b = ra$ for some $r \in R^\times$, then we say that a and b are* **associates***.*

Proposition 5.10. *Let R be a UFD. Then there is no infinite sequence of non-zero elements $r_1, r_2, \ldots$ of R such that r_{n+1} properly divides r_n for each $n \geqslant 1$.*

Proof. Every element dividing an invertible element is invertible. Thus no element properly divides an invertible element, so $r_1 \notin R^\times$. Write $r_1 = a_1 \ldots a_m$, where $a_1, \ldots, a_m$ are irreducibles (possible since R is a UFD). The number of factors m does not depend on the factorisation (m only depends on r_1). Write $m = l(r_1)$. If r_2 properly divides r_1, then $l(r_2) < l(r_1)$. Hence $l(r_1) > l(r_2) > \cdots$ Any decreasing sequence of natural numbers is finite, so no infinite sequence $r_1, r_2, \ldots$ exists. □

Proposition 5.11. *Let R be a UFD. If p is irreducible and $p|ab$ then $p|a$ or $p|b$.*

Proof. If $a \in R^\times$, then $p|b$ (since $p|ab$ implies $ab = pc$ and then $b = pca^{-1}$, for some $c \in R$). So assume that $a, b \notin R^\times$. Then $a = a_1 \ldots a_m$, $b = b_1, \ldots b_n$ for some irreducible elements a_i and b_j with $m \geqslant 1$ and $n \geqslant 1$. Write $a_1 \ldots a_m b_1 \ldots b_n = pc$ for some $c \in R$. If $c \in R^\times$, we have $(c^{-1}a_1)a_2 \ldots a_m b_1 \cdots b_n = p$, which is a contradiction because the number of irreducible factors on both sides is not the same. Otherwise $c = c_1 \ldots c_s$ for some irreducibles $c_1, \ldots, c_s \in R$. Then we have two ways of writing ab as a product of irreducibles

$$a_1 \ldots a_m b_1 \ldots b_n = pc_1 \ldots c_s.$$

Thus p is associated with some a_i or b_j, hence $p|a$ or $p|b$. □

Theorem 5.12. *Let R be an integral domain. Then R is a UFD if and only if the following conditions hold:*

1. *There is no infinite sequence $r_1, r_2, \dots$ of elements of R such that r_{n+1} properly divides r_n for all $n \geqslant 1$.*
2. *For every irreducible element $p \in R$, if $p|ab$, then $p|a$ or $p|b$.*

Proof. By Propositions 5.10 and 5.11, conditions (1) and (2) are satisfied for any UFD.

Conversely, assume R satisfies (1) and (2). For contradiction, suppose that there is a non-zero element r_1 in $R \backslash R^\times$, which cannot be written as a product of irreducibles. Note that r_1 is not irreducible, hence $r_1 = r_2 s_2$, for some $r_2, s_2 \in R \backslash R^\times$. At least one of the factors cannot be written as a product of irreducibles, say r_2. For the same reason as before, we can write $r_2 = r_3 s_3$, with $r_3, s_3 \in R \backslash R^\times$. Continuing in this way, we obtain an infinite sequence $r_1, r_2, r_3, \dots$. Moreover, in this sequence, r_{n+1} properly divides r_n because s_{n+1} is never in $R^\times$. This contradicts condition (1). Hence every non-zero element of $R \backslash R^\times$ can be written as a product of irreducibles.

Now assume that $a_1 \dots a_m = b_1 \dots b_n$, where the a_i and b_j are irreducibles. We can assume that $m \leqslant n$. Since $a_1 | b_1 b_2 \cdots b_n$, by (2) we see that a_1 divides b_j for some j. Reorder the b_j's so that $a_1 | b_1$. Thus $b_1 = a_1 u$ for some $u \in R$, $u \neq 0$. If $u \notin R^\times$, then b_1 cannot be irreducible. Therefore $u \in R^\times$ and hence a_1 and b_1 are associates. If $m = 1$ and $n = 1$, we are done, but if $m = 1$ and $n \geqslant 2$ the cancellation property gives $1 = (ub_2) \dots b_n$, which is impossible. If $m \geqslant 2$ we have $a_2 \dots a_m = (ub_2) \dots b_n$ by the cancellation property. Continue in this way until we get 1 in the left hand side. Since a product of $n - m$ irreducibles cannot equal 1, we must have $m = n$ and, possibly after reordering, a_i and b_i are associates, for $i \geqslant 1$. □

Example 5.13 (Example of a non-UFD)**.** Let

$$R = \{a_0 + a_1 x + \cdots + a_n x^n \mid a_0 \in \mathbb{Z},\ a_i \in \mathbb{Q} \text{ for } i \geqslant 1\}.$$

Clearly $R \subset \mathbb{Q}[x]$ and R is a subring of $\mathbb{Q}[x]$ and also an integral domain. Consider $r_1 = x, r_2 = \frac{1}{2}x, r_3 = \frac{1}{4}x, \dots \in R$ and so $r_n = 2r_{n+1}$ but $\frac{1}{2} \notin R$ and hence $2 \notin R^*$ and $x \notin R^*$ since $\frac{1}{x} \notin \mathbb{Q}[x]$. Thus r_{n+1} properly divides r_n. By Proposition 5.10, R is not a UFD.

Proposition 5.14. *Suppose R is a PID and $I_1 \subset I_2 \subset \cdots$ are ideals in R. Then for some n we have $I_n = I_{n+1} = \cdots$ (One says that any ascending chain of ideals* **stabilises***.)*

Proof. Define

$$I = \bigcup_{n \geqslant 1} I_n.$$

This is a subset of R. We claim that I is an ideal. Given $x, y \in I$ we must show that $x + y$, $-x$, xy are in I. Any $x \in I$ belongs to some I_n. Similarly, any $y \in I$ is in some I_m. Suppose $n \geqslant m$. Then $I_m \subset I_n$. So $x, y \in I_n$ and thus $x + y, -x, xy \in I_n$. Therefore $x + y, -x, xy \in I$. Let $r \in R$ and $x \in I_n$. Then $rx \in I_n$ and therefore $rx \in I$. Thus I is an ideal in R.

By assumption, $I = aR$ for some $a \in R$. Clearly, $a \in I$. Hence, for some $l \geqslant 1$, we have $a \in I_l$. But then $I = aR \subset I_l$. On the other hand, $I_{l+m} \subset I$ for any $m \geqslant 0$, so $I = I_l = I_{l+1} = \dots$ □

Example 5.15. *Assume $R = \mathbb{Z}$. Then $60\mathbb{Z} \subset 30\mathbb{Z} \subset 15\mathbb{Z} \subset 5\mathbb{Z} \subset \mathbb{Z}$.*

Proposition 5.16. *Suppose that R is a PID. Let $p \in R$ be an irreducible element such that $p|ab$. Then $p|a$ or $p|b$.*

Proof. We claim that $I = aR + pR = \{ar_1 + pr_2 \mid r_1, r_2 \in R\}$ is an ideal: if $r \in R$, then $r(ar_1 + pr_2) = a(rr_1) + p(rr_2) \in I$. Since R is a PID, we have $I = dR$ for some $d \in R$.

We have $p = a\cdot 0 + p\cdot 1 \in I$ and so can write $p = dr$ for some $r \in R$. Since p is irreducible, r or d is in $R^\times$. If $r \in R^\times$, then p and d are associates. Since d dividies a, we see that p divides a. If $d \in R^\times$, then $I = dR$ contains $1 = dd^{-1}$. Therefore $1 = at + pu$ for some $t, u \in R$. This implies that $b = abt + bpu$. By assumption, $p|ab$, thus $p|abt + bpu$, so $p|b$. □

Theorem 5.17. *Every PID is a UFD.*

Proof. We will apply Theorem 5.12 whose second condition follows from Proposition 5.16. It remains to prove that there does not exist an infinite sequence $r_1, r_2, \ldots$ such that r_{n+1} properly divides r_n for $n = 1, 2, \ldots$ Indeed, let $r_1, r_2, \ldots$ be such a sequence. This implies that $r_n R \subset r_{n+1} R$ for $n = 1, 2, \ldots$. By Proposition 5.14 there exists $l \geqslant 1$ such that $r_l R = r_{l+1} R \cdots$. But then r_{l+1} and r_l are associates, a contradiction. □

6. Fields

6.1. **Field extensions.** The only proper ideal of any field k is the zero ideal (Proposition 4.35). Thus any homomorphism $k \to R$, where k is a field and R is a ring, is injective. So the only maps between fields are field extensions.

Definition 6.1. *An extension of fields $k \subset K$ is called* **finite** *if K is a finite-dimensional vector space over k. In this case we call $\dim_k(K)$ the* **degree** *of the extension and write $[K : k] = \dim_k(K)$.*

Extensions of degree 2 are called quadratic, extensions of degree 3 are called cubic, etc.

Theorem 6.2. *Let $k \subset F$ and $F \subset K$ be field extensions. Then K is a finite extension of k if and only if F is a finite extension of k and K is a finite extension of F. In this case we have $[K : k] = [K : F][F : k]$.*

Proof. If K is a finite-dimensional vector space over k, then any subspace of K is too, so $\dim_k(F) < \infty$. Any finite set of vectors that spans K as a k-vector space, spans K as an F-vector space, hence $\dim_F(K) < \infty$.

Conversely, suppose that $v_1, \ldots, v_n$ is a basis of F as a k-vector space and that $w_1, \ldots, w_m$ is a basis of K as an F-vector space. We claim that $\{v_i w_j\}$, where $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$, is a basis of K as a k-vector space.

We first show that this set spans K over k. Any element $u \in K$ is written as

$$u = x_1 w_1 + \ldots + x_m w_m, \quad x_j \in F.$$

But the x_j, like all elements of F, are linear combinations of $v_1, \ldots, v_n$ with coefficients in k:

$$x_j = y_{1j} v_1 + \ldots + y_{nj} v_n, \quad y_{ij} \in k.$$

Hence $u = \sum_{i,j} y_{ij} v_i w_j$, so $\{v_i w_j\}$ spans K over k.

It remains to show linear independence of the vectors in $\{v_i w_j\}$. Indeed, if $\sum_{i,j} y_{ij} v_i w_j = 0$ for some $y_{ij} \in k$, then since $w_1, \ldots, w_m$ is a basis of K as an F-vector space, we must have $\sum_{i=1}^n y_{ij} v_i = 0$ for each $j = 1, \ldots, m$. But $v_1, \ldots, v_n$ is a basis of F as a k-vector space, so $y_{ij} = 0$ for all i and j. This proves that K is a finite-dimensional k-vector space of dimension $[K : k] = mn = [K : F][F : k]$. □

6.2. Constructing fields from irreducible polynomials.

Proposition 6.3. *Let R be a PID and let $a \in R$, $a \neq 0$. Then aR is maximal if and only if a is irreducible.*

Proof.

$\Rightarrow$ Assume that $aR \subset R$ is a maximal ideal. Since $aR \neq R$, we have $a \notin R^\times$. Thus if a is not irreducible we have $a = bc$ for $b, c \notin R^\times$. Then $aR \subset bR \subsetneq R$ since $b \notin R^\times$. Since aR is maximal we have $aR = bR$ and so for some $m \in R$ we have $b = am = bcm$ and so $1 = cm$, hence $c \in R^\times$; a contradiction. Therefore a is irreducible.

$\Leftarrow$ Now assume that a is irreducible. In particular, $a \notin R^\times$, so $aR \neq R$. Assume that there exists an ideal J such that $aR \subsetneq J \subsetneq R$. Since R is a PID, $J = bR$ for some $b \in R$. Note that $b \notin R^\times$ because $bR \neq R$. Since $aR \subset bR$, we can write $a = bc$ for some $c \in R$. Also $c \notin R^\times$ because otherwise $aR = bR$ (if $c \in R^\times$ then $c^{-1} \in R$ and so $b = c^{-1}a \in aR$, hence $bR \subset aR$). Thus a is not irreducible; a contradiction. Therefore aR is maximal. □

Corollary 6.4. *If R is a PID and $a \in R$ is irreducible, then R/aR is a field.*

This suggests a method to construct fields which we now explore.

Remark 6.5. Let k be a field, let $R = k[t]$ and let $p(t) \in k[t]$ be an irreducible polynomial of degree d. By Corollary 6.4, $K = k[t]/p(t)k[t]$ is a field. Using Proposition 5.2 we can choose coset representatives to be polynomials of degree at most $d - 1$, then

$$K = \{x_0 + x_1 t + \ldots + x_{d-1}t^{d-1} + p(t)k[t] | x_i \in k\}.$$

By the uniqueness part of Proposition 5.2, each coset has exactly one representative of degree $\leqslant d - 1$, so $\dim_k(K) = d$. Consider the map $k \to K$ sending $x \in k$ to the coset $x + p(t)k[t]$. This maps sends sums to sums and products to products, so it is a homomorphism of fields. Every non-zero homomorphism of fields is injective, so the image of this map is a subfield of K isomorphic to k. Thus K is a field extension of k of degree $[K : k] = \dim_k(K) = d$. Let $\tau \in K$ be the coset $t + p(t)k[t]$. Since k is a subfield of K we can think of $p(t)$ as a polynomial with coefficients in K. Then $p(\tau)$ is the trivial coset $p(t) + p(t)k[t] = p(t)k[t]$; in other words, we have $p(\tau) = 0$ in K. We conclude that *for any irreducible polynomial $p(t) \in k[t]$ there exists a finite field extension $k \subset K$ such that $p(t)$ has a root in K.*

For example, let $R = \mathbb{Q}[t]$ and let $p(t) = t^2 - a$, where a is an integer not divisible by p^2, for any prime p. It is clear that $p(t)$ is irreducible in $\mathbb{Q}[t]$. Then it is immediate to check that sending $x + yt + p(t)\mathbb{Q}[t]$ to $x + y\sqrt{a}$ defines an isomorphism $\mathbb{Q}[t]/p(t)\mathbb{Q}[t] \cong \mathbb{Q}(\sqrt{a})$.

We saw that $p(t) = t^2 - a$ is irreducible by observing that it has no roots. This also works for polynomials of degree 3, but not in higher degrees. (For instance, $(t^2 + 1)(t^2 + 2)$ has no real roots but is not irreducible in $\mathbb{R}[t]$.)

Proposition 6.6. *Let k be a field. A polynomial $f(t) \in k[t]$ of degree 2 or 3 is irreducible if and only if it has no root in k.*

Proof.

$\Leftarrow$ If $f(t)$ is not irreducible, then $f(t) = a(t)b(t)$ with $\deg f(t) = \deg a(t) + \deg b(t)$ and $\deg a(t), \deg b(t) \geqslant 1$ (since polynomials in $k[t]^\times$ have degree 0). Hence $\deg a(t) = 1$ or $\deg b(t) = 1$. Thus a linear polynomial, say $t - \alpha$ divides $f(t)$, so that $f(\alpha) = 0$ for some $\alpha \in k$.

- $\Rightarrow$ If $f(t)$ has a root $\alpha \in k$ then we can write $f(t) = q(t)(t-\alpha) + r(t)$, where $r(t) = 0$ or $\deg r(t) = 0$, so that $r(t)$ is an element of k. But $f(\alpha) = 0$, so $r = 0$. Neither $q(t)$ of degree 2, nor $t - \alpha$ of degree 1, is an element of $k[t]^\times$, so $f(t)$ is not irreducible. □

Example 6.7. Let $k = \mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$. Then $t^2 + t + 1$ has no root in $\mathbb{F}_2$, so is irreducible in $\mathbb{F}_2[t]$. Then $K = \mathbb{F}_2[t]/(t^2+t+1)\mathbb{F}_2[t]$ is a quadratic extension of $\mathbb{F}_2$, so $|K| = 2^2 = 4$. The elements of K are $\{0, 1, \tau, 1+\tau\}$, where τ is the image of $t \in \mathbb{F}_2[t]$, so we have $\tau^2 + \tau + 1 = 0$. Thus the multiplication in K is uniquely determined by the rule $\tau^2 = 1 + \tau$. For example, we have $\tau^{-1} = 1 + \tau$.

We can now prove the existence of fields of order p^2 for any odd prime p. An element a of a field k is called a *non-square* if there is no element $b \in k$ such that $a = b^2$.

Proposition 6.8. *Let p be an odd prime. The field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ contains $(p-1)/2 \geqslant 1$ non-squares. For any non-square $a \in \mathbb{F}_p$, the polynomial $t^2 - a$ is irreducible in $\mathbb{F}_p[t]$ and the field $\mathbb{F}_p[t]/(t^2-a)\mathbb{F}_p[t]$ is a quadratic extension of $\mathbb{F}_p$.*

Proof. By Proposition 6.6 we only need to prove the first statement. The map $x \mapsto x^2$ is a homomorphism $f\colon \mathbb{F}_p^\times \to \mathbb{F}_p^\times$. The element $-1 \in \mathbb{F}_p$ is contained in $\mathrm{Ker}(f)$, but $-1 \neq 1$ because $2 \neq 0$ in $\mathbb{F}_p$ since the characteristic of $\mathbb{F}_p$ is $p \neq 2$. Thus $\mathrm{Ker}(f)$ has exactly two elements 1 and -1, because if there were more, the polynomial $x^2 - 1$ would have more that two roots in $\mathbb{F}_p$, which is absurd. By the isomorphism theorem for groups we conclude that $|\mathrm{Im}(f)| = (p-1)/2$. The set of non-squares in $\mathbb{F}_p$ is the complement to $\mathrm{Im}(f)$ in $\mathbb{F}_p^\times$, hene the result. □

If there was a simple way to construct irreducible polynomials of arbitrary degree over $\mathbb{F}_p$, or at least prove that they exist, then we would immediately deduce the existence of finite fields of arbitrary prime power order. This is not obvious, however, so we need to do some more work to achieve this.

Proposition 6.9. *Let k be a field and let $p(t) \in k[t]$. There exists a finite field extension $k \subset K$ such that $p(t) = c\prod_{i=1}^n (t - \alpha_i)$, where $c \in k^*$ and $\alpha_i \in K$ for $i = 1, \dots, n$.*

Proof. If $\deg(p(t)) = 1$, then $K = k$ does the job. Assume the statement is proved for all polynomials of degree $\leqslant n-1$ and let's prove it for a polynomial $p(t)$ of degree n. Since $k[t]$ is a UFD, we can write $p(t) = cp_1(t)\dots p_m(t)$, where $c \in k^*$ and $p_1(t), \dots, p_m(t)$ are irreducible monic polynomials in $k[t]$. By Remark 6.5 there is a field extension $k \subset k_1$ such that $p_1(t)$ has a root in $k_1[t]$. Then we have $p(t) = (t-\alpha)q(t)$ for some $\alpha \in k_1$ and $q(t) \in k_1[t]$. By induction assumption there is a field extension $k_1 \subset K$ such that $q(t)$ is a product of linear factors in $K[t]$. By Theorem 6.2, the field K is a finite extension of k, so we are done. □

It is possible to prove that there is a "smallest" extension in which a given polynomial decomposes as a product of linear factors and that such an extension is unique up to isomorphism. The proof can be found in most algebra textbooks. We do not need this in this course so we do not give a proof.

6.3. **Existence of finite fields.**

Lemma 6.10. *Let k be a field of characteristic p, where p is a prime. Then for any $x, y \in k$ we have*

$$(x+y)^{p^m} = x^{p^m} + y^{p^m} \tag{6.1}$$

for any $x, y \in k$ and any positive integer m.

Proof. See Problem Sheet 5, Question 8. □

Let k be a field and let $p(t) = a_n t^n + \ldots + a_0 \in k[t]$. Define the *derivative* of $p(t)$ as

$$p'(t) = na_n t^{n-1} + (n-1)a_{n-1}t^{n-2} + \ldots + 2a_2 t + a_1 \ \in \ k[t].$$

Let us stress that this is a formal definition. Although this is exactly the same formula as in analysis, defining the derivative of a polynomial can be made over any field and does not reply on any limiting process. The Leibniz formula still holds: for $p(t), q(t) \in k[t]$ we have

$$(p(t)q(t))' = p'(t)q(t) + p(t)q'(t).$$

Indeed, it is enough to prove this when $p(t) = t^a$ and $q(t) = t^b$, and this is straightforward.

Lemma 6.11. *Let k be a field and let $p(t) = (t-\alpha_1)\ldots(t-\alpha_n)$, where $\alpha_i \in k$ for $i = 1, \ldots, n$. Then $\alpha_i \neq \alpha_j$ for $i \neq j$ if and only if $p(t)$ and $p'(t)$ have no common root.*

Proof. By the Leibniz formula we have

$$p'(t) = \sum_{i=1}^{n} \prod_{\substack{j=1 \\ j\neq i}}^{n} (t-\alpha_j).$$

Hence $p'(\alpha_i) = \prod_{j=1, j\neq i}^{n}(\alpha_i - \alpha_j) \neq 0$ precisely when α_i is a simple root of $p(t)$. □

Theorem 6.12. *Let p be a prime number and let n be a positive integer. There exists a field with p^n elements.*

Proof. Consider the polynomial $t^{p^n} - t$ in $\mathbb{F}_p[t]$. By Proposition 6.9 there is a finite extension $\mathbb{F}_p \subset K$ such that

$$t^{p^n} - t = (t-\alpha_1)\ldots(t-\alpha_{p^n}),$$

where $\alpha_i \in K$ for $i = 1, \ldots, p^n$. Define

$$F = \{\alpha_1, \ldots, \alpha_{p^n}\} \ \subset \ K.$$

We claim that F is a subfield of K. For this we need to show that F is closed under + and ×, and under taking the additive and multiplicative inverses. (Clearly, F contains 0 and 1.) That F is closed under multiplication and taking the multiplicative inverse is clear; that F is closed under addition follows immediately from (6.1). If p is odd, then α is a root of $t^{p^n} - t = 0$ if and only if $-\alpha$ is a root. If $p = 2$, then there is no difference between plus and minus, so the additive inverse of x is x. Thus F is closed under taking the additive inverse. We proved that F is a subfield of K.

By Theorem 4.33, F contains the prime subfield $\mathbb{F}_p$. To complete the proof it remains to show that $\alpha_i \neq \alpha_j$ if $i \neq j$. We do this using Lemma 6.11: the derivative of $t^{p^n} - t$ is $p^n t^{p^n - 1} - 1 = -1$, because $\mathrm{char}(k) = p$. This is a non-zero constant, hence all roots of $t^{p^n} - t$ are simple. Thus $|F| = p^n$. □

It can be proved that all finite fields with p^n elements are isomorphic.