Abstract Algebra

Definition of fields is assumed throughout these notes.

> "Algebra is generous; she often gives more than is asked of her."
> – D'Alembert

Groups

1 Definition and examples

Definition 1.1 *A* **group** *is a non-empty set G with an associative binary operation $*$ with the following property:*
(1) **(Identity element)** *There exists an element $e \in G$ such that for all $a \in G$, $e * a = a * e = a$. (Why is it called "e"? This comes from German "Einheit".)*
(2) **(Inverse element)** *For every $a \in G$ there exists $b \in G$ such that $a * b = b * a = e$.*
We often write $(G, *)$ to mean that G is a group with operation $*$.

If F is a field, such as $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, then $(F, +)$ is a group but $(F, \cdot)$ is not. (Justify.) Furthermore, $(F \setminus \{0\}, \cdot)$ is a group. Also, if V is a vector space over F, then $(V, +)$ is a group. (Justify.) Verify that $(\mathbb{Z}, +)$ is a group, but that $(\mathbb{N}, +)$ is not.

We will study the groups abstractly and also group the groups in some natural groups of groups (decide which of the words "group" are technical terms).

Here is a possibly new example: let $G = \{1, -1, i, -i\}$, and let $*$ be multiplication. Then G is a group, and we can write out its multiplication table (**Cayley table**):

	1	-1	i	-i
1	1	-1	i	-i
-1	-1	1	-i	i
i	i	-i	-1	1
-i	-i	i	1	-1

Associativity holds because we know that multiplication of complex numbers is associative. We can clearly find the identity element and an inverse (the inverse?) of each element.

Consider the set H consisting of rotations of the plane around the origin by angles 90°, 180°, 270° and 360°. Verify that H is a group if $*$ is taken to be composition. How many elements does H have? Write its multiplication table. What is its identity element? Can you find a similarity with the previous example?

Exercise 1.2 Let n be a positive integer and let G be the set of all complex numbers whose nth power is 1. Prove that $(G, \cdot)$ is a group. What is its identity element? Can you represent this group graphically?

Exercise 1.3 Let n be a positive integer. Let $G = \{0, \ldots, n-1\}$. For any $a, b \in G$, define $a * b$ to be the remainder of $a + b$ after dividing by n. Prove that $(G, *)$ is a group. What is its identity element? For $a \in G$, what is its inverse? This group is denoted in Math 112 as $\mathbb{Z}_n$ (read: "z n"). Later we will also see the more apt notations $\mathbb{Z}/n\mathbb{Z}$ and $\frac{\mathbb{Z}}{n\mathbb{Z}}$. (COMMENT: this is NEVER "division" by zero!)

For Reed students, who are very familiar with binary properties, it seems best to first narrow down the general possibilities for groups before we look at more examples.

2 What follows immediately from the definition

Theorem 2.1 *Let * be an associative binary operation on a non-empty set G. Then G has at most one element e satisfying the property that for all $a \in G$, $e * a = a * e = a$.*

Proof. If e' is an element of G with $e' * a = a * e' = a$ for all $a \in G$, then

$$e' * e = e \text{ and } e' * e = e'$$

by the defining properties of e and e, whence $e = e'$. □

In particular, a group $(G, *)$ has exactly one element e that acts as an identity element, and it is in fact called **the identity element of** G. Furthermore, **the inverses** are also unique.

Theorem 2.2 *Let $(G, *)$ be a group, $a \in G$. Then there exists a unique element $b \in G$ such that $b * a = a * b = e$.*

Proof. By the inverse element axiom, such an element b exists. Let $c \in G$ such that $c * a = a * c = e$. Then

$$c = c * e = c * (a * b) = (c * a) * b = e * b = b,$$

by associativity and by the property of e. □

This unique inverse element of a is typically denoted as a^{-1}. WARNING: when the operation $*$ is $+$, then the inverse is written $-a$. Beware of confusion.

We also introduce another bit of notation: for $a \in G$, a^0 is the identity element, if n is a positive integer, then a^n is the shorthand for $a * a * \cdots * a$, where a is written n times. Clearly if $n > 0$, then $a^n = a^{n-1} * a = a * a^{n-1}$. WARNING: when the operation $*$ is $+$, then $a * a * \cdots * a$ (with a being written n times) is usually denoted as na. Beware of confusion.

Lemma 2.3 *For any $n \in \mathbb{N}$, $(a^n)^{-1} = (a^{-1})^n$.*

Proof. By definition, $(a^n)^{-1}$ is the unique element of G whose product with a^n in any order is e. But by associativity,

$$\begin{aligned} a^n * (a^{-1})^n &= (a^{n-1} * a) * (a^{-1} * (a^{-1})^{n-1}) \\ &= a^{n-1} * (a * (a^{-1} * (a^{-1})^{n-1})) \\ &= a^{n-1} * ((a * a^{-1}) * (a^{-1})^{n-1})) \\ &= a^{n-1} * (e * (a^{-1})^{n-1})) \\ &= a^{n-1} * (a^{-1})^{n-1}, \end{aligned}$$

which by induction on n equals e (the cases $n = 0$ and $n = 1$ are trivial). Similarly, the product of a^n and $(a^{-1})^n$ in the other order is e. This proves that $(a^{-1})^n$ is the inverse of a^n, which proves the lemma. □

With this, if n is a negative integer, we write a^n to stand for $(a^{-n})^{-1}$.

Theorem 2.4 (Cancellation) *Let $(G, *)$ be a group, $a, b, c \in G$ such that $a * b = a * c$. Then $b = c$.*

Similarly, if $b * a = c * a$, then $b = c$.

Proof. By the axioms and the notation,

$$b = e * b = (a^{-1} * a) * b = a^{-1} * (a * b) = a^{-1} * (a * c) = (a^{-1} * a) * c = e * c = c.$$

The second part is proved similarly. □

Exercise 2.5 Prove that for every $a \in G$, $(a^{-1})^{-1} = a$.

Exercise 2.6 Let $a, b \in G$. Prove that $(a * b)^{-1} = b^{-1} * a^{-1}$.

Exercise 2.7 Let G be a group, $a \in G$. Then the **left translation** or the **left multiplication** by a is the function $L_a : G \to G$ defined by $L_a(x) = a * x$. Prove that L_a is a one-to-one and onto function.

Exercise 2.8 Let G be a group, $a \in G$. Then **the conjugation** by a is the function $C_a : G \to G$ defined by $C_a(x) = a*x*a^{-1}$. Prove that C_a is a one-to-one and onto function and that its inverse is $C_{a^{-1}}$.

3 Bijections

We study our first family of groups.

Exercise 3.1 Let X be a non-empty set and let G be the set of all one-to-one and onto functions $f : X \to X$. (You may need to review what a one-to-one and onto function is.) Then $(G, \circ)$ is a group. Verify. What is the identity element? How do we denote the inverse of $f \in G$?

Definition 3.2 *The group as in the previous exercise is denoted* $\mathbf{S_X}$ *and is called* **the permutation group of** X.

Exercise 3.3 Suppose that X has in addition some built-in topology on it (for example, as a a subset of some $\mathbb{R}^n$, or with a p-adic topology, or with the discrete topology, etc). Let H be the set of all homeomorphisms, i.e., all bicontinuous one-to-one and onto functions $f : X \to X$. Then $(H, \circ)$ is also a group. Verify. What is its identity element?

Exercise 3.4 Let G be the set of all linear one-to-one and onto functions $f : \mathbb{R}^n \to \mathbb{R}^n$. Prove that G is a group under composition. Why does it follow that the set of all invertible $n \times n$ matrices with real entries is a group under multiplication? What is the identity element of this group?

Recall that $f : \mathbb{R}^n \to \mathbb{R}^n$ is **rigid** if for all $x, y \in \mathbb{R}^n$, $||f(x) - f(y)|| = ||x - y||$. Examples of rigid functions: translations, rotations, reflections, glide reflections (what is that?), compositions of these. One can verify that every rigid function is a composition of an orthogonal transformation with a translation.

Let X be a subset of $\mathbb{R}^n$. Consider the subset of the set of all rigid motions of $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $f(X) = X$. It is straightforward to verify that this is a group. We'll call this the **group of rigid motions of $\mathbb{R}^n$ that preserves** X or the **symmetry group of** X.

Example 3.5 Work out the set of all rigid motions of $\mathbb{R}^3$ that preserve a non-square rectangle (a two-dimensional sheet in $\mathbb{R}^3$). Write out its multiplication table.

Example 3.6 Comment on the group D_3 of rigid motions that preserves a regular triangle. Write the multiplication (Cayley) table for D_3. Comment on the corresponding group D_n of a regular n-gon. Can we predict/count at this stage how many elements are in these groups?

Bring some Platonic solids to class. Comment on their groups.

4 Commutativity

In some groups $(G, *)$, $*$ is a commutative operation. Namely, for all $a, b \in G$, $a * b = b * a$. Such a group is called **commutative** or **Abelian**, Abelian in honor of Niels Abel, a Norwegian mathematician from the 19th century. (Read/tell more about him!)

When * is composition of functions, G is rarely commutative. Give examples of commutative and non-commutative groups.

Lemma 4.1 *If $a * b = b * a$, then for all/any one $n \in \mathbb{Z}$, $(a * b)^n = a^n * b^n$.*

Proof. If $n = 0$ or $n = 1$, this holds trivially. Now let $n > 1$. By commutativity, $b^m * a = a * b^m$ for all $m \geq 0$. Then by induction on n,

$$\begin{aligned}(a * b)^n &= (a * b)^{n-1} * (a * b) = (a^{n-1} * b^{n-1}) * (a * b) \\ &= ((a^{n-1} * b^{n-1}) * a) * b = (a^{n-1} * (b^{n-1} * a)) * b \\ &= (a^{n-1} * (a * b^{n-1})) * b = (a^{n-1} * a) * b^{n-1}) * b \\ &= a^n * (b^{n-1} * b) = a^n * b^n.\end{aligned}$$

Thus the lemma holds for all $n \in \mathbb{N}$. If $n < 0$, then by the positive case and commutativity, $(a * b)^n = (b * a)^n = ((b * a)^{-n})^{-1} = (b^{-n} * a^{-n})^{-1}$, whence from Exercise 2.6, $(a * b)^n = (a^{-n})^{-1} * (b^{-n})^{-1}$, which is $a^n * b^n$. □

A partial converse also holds (why is this only a **partial** converse?):

Proposition 4.2 *Let $a, b \in G$ such that $(a * b)^2 = a^2 * b^2$. Then $a * b = b * a$.*

Proof. By assumption,

$$a * b * a * b = (a * b)^2 = a * a * b * b,$$

so that by cancellation, $b * a = a * b$. □

Exercise 4.3 Find a group G, elements $a, b \in G$, and a positive integer n such that $(a * b)^n = a^n * b^n$ yet $a * b \neq b * a$.

Exercise 4.4 (From Gallian, page 55, Exercise 16) In a group, $(a * b)^{-1} = b^{-1} * a^{-1}$. Find an example that shows that it is possible to have $(a * b)^{-2} \neq b^{-2} * a^{-2}$. Find distinct non-identity elements a and b from a non-commutative group with the property that $(a * b)^{-1} = a^{-1} * b^{-1}$. Draw an analogy between the statement $(a * b)^{-1} = b^{-1} * a^{-1}$ and the act of putting on and taking off your socks and shoes.

5 Frequent groups and groups with names

Some groups come in groups (which is the technical term?) and some have names attached to them. This section gives a partial listing of the named groups. Some more names groups will appear later in the notes.

The **trivial** group is the group consisting of only one element.

In Section 3 we defined D_n to be the symmetry group of the regular n-gon. This group is called the **dihedral group** (of order $2n$). This defines D_n only for $n \geq 3$. We declare D_2 to be the symmetry group of a rectangle that is not a square, and D_1 to be the symmetry group of a line segment.

In Exercise 1.3 we defined $\mathbb{Z}_n$ to be the group $\{0, 1, \ldots, n-1\}$ under addition modulo n. (We'll see these groups in greater detail later.)

By U_n we denote the set of elements a in $\mathbb{Z}_n$ for which there exists $b \in \mathbb{Z}_n$ such that $a \cdot b = 1$. Verify that U_n is a group under multiplication. (A word on notation: Gallian writes $U(n)$; this class would rather write U_n for consistency with $\mathbb{Z}_n$.)

The set of all complex numbers whose nth power is 1 forms a group, with the operation being multiplication. (See Exercise 1.2.)

If X is a set, the set $\mathbf{S_X}$ of all bijective functions $X \to X$ is a group under composition (see Definition 3.2), and is called **the permutation group of** X.

In the special case where $X = \{1, \ldots, n\}$, we denote S_X also as S_n. This group is called **the symmetric group of degree** n.

If F is a field, **the general linear group** $\mathrm{GL}\,(n, F)$ over F is the group of all invertible $n \times n$ matrices with entries in F, with the group operation being matrix multiplication. **The special linear group** $\mathrm{SL}\,(n, F)$ over F is the group of all $n \times n$ matrices with entries in F with determinant 1, and with the group operation being matrix multiplication.

We have also seen that **fields** are groups under addition, and it is easy to verify that **vector spaces** are groups under addition.

Let S be a set. The **free group on** S is the set F of all symbols of the form $s_1^{n_1} s_2^{n_2} \cdots s_r^{n_r}$, where $r \in \mathbb{N}$, $s_i \in S$ and $n_i \in \mathbb{Z}$. When $r = 0$, this is the empty symbol, and we denote it e. We make F into a group by concatenation of these symbols, and by making the following (necessary) identifications: $s^n s^m = s^{n+m}$, $s^0 = e$. WARNING: we do NOT require that $st = ts$! Convince yourself that F is a group.

More names appear later in the notes: page 20, ?

6 Group generators

Let $(G, *)$ be a group. We say that G is **generated** by a subset S if every $a \in G$ can be written as $a = a_1^{n_1} * a_2^{n_2} * \cdots * a_k^{n_k}$ for some $a_i \in S$ and some $n_i \in \mathbb{Z}$. If $S = \{s_1, \ldots, s_k\}$, we also say that G is generated by $s_1, \ldots, s_k$. If T is any subset of G or a list of elements in G, we write $\langle T \rangle$ to denote the group contained in G that is generated by T. The operation on this group is $*$ restricted from G. If $G = \langle a \rangle$, we say that G is **cyclic**.

For example,

(1) only the trivial group is generated by the identity element;
(2) $\mathbb{Z}$ is generated by 1 and it is also generated by -1;
(3) $\mathbb{Z}_n$ is generated by 1;
(4) $\mathbb{Z}_7$ is generated by any of its non-identity elements;
(5) $\mathbb{Z}_4$ is not generated by 2;
(6) $\mathbb{Q}$ is not generated by a finite set – justify!
(7) If G is generated by S, then it is also generated by any subset of G that contains S.

Exercise 6.1 Prove that a cyclic group is commutative.

Exercise 6.2 Recall that D_n is the symmetry group of a regular n-gon. Prove that D_n is not generated by one element if $n \geq 2$. Find a minimal generating set for D_n (minimal in the sense that if you remove any element, then the remaining set will not be a generating set).

7 Subgroups

Let $(G, *)$ be a group. A **subgroup** of G is a group $(H, *)$ such that $H \subseteq G$ (and $*$ is the same operation as the one on G). We write $H \leq G$, or $H < G$ when we want to emphasize that in addition $H \neq G$.

Easy examples: $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{Q}, +)$ which is a subgroup of $(\mathbb{R}, +)$ which is a subgroup of $(\mathbb{C}, +)$. On the other hand, $(\mathbb{Q} \setminus \{0\}, \cdot)$ is not a subgroup of $(\mathbb{Q}, +)$ and $\mathbb{Z}_n$ is not a subgroup of $\mathbb{Z}$. (Justify.)

Verify that D_n is a subgroup of S_n.

Let G be the symmetry group of a wallpaper design. The subset of all translations is a subgroup. The subset of all rotations around a point is a subgroup. The subset of all reflections contains some rotations, but no reflections are contained in the subgroup of G generated by translations and rotations. Justify.

Proposition 7.1 *Let H be a subgroup of G and let K be a subgroup of H. Then K is a subgroup of G.*

Proof. Certainly K is a subset of G, and since the group operation on K is inherited from H, and the group operation on H is inherited from G, then the group operation on K is inherited from G. □

Clearly G is a subgroup of G and $\{e\}$ is a subgroup of G. The latter group is meaningfully named **the trivial subgroup**.

Theorem 7.2 *Let $(G, *)$ be a group and H a non-empty subset of G. Then H is a subgroup of G if and only if for all $a, b \in H$, $a * b^{-1} \in H$.*

Proof. If H is a subgroup, then for every $a, b \in H$, $b^{-1} \in H$ and $a * b^{-1} \in H$. So one implication is easy.

Assume that $a * b^{-1} \in H$ for all $a, b \in H$. We need to prove that H is a group with operation $*$. First of all, H is not empty, so there is some a in H, and so by assumption, $e = a*a^{-1} \in H$, and then also $a^{-1} = e*a^{-1} \in H$. If $a, b \in H$, we just proved that $b^{-1} \in H$, so that by the assumption on H, $a * b = a * (b^{-1})^{-1} \in H$. Thus $*$ is a binary operation on H. It is also associative on H since it is associative on the bigger set G. We already proved that the identity element e of G lives in H, and then for all $a \in H$, $a * e = a = e * a$ since this is true for all $a \in G$. Finally, we already proved that the inverses of elements in H live in H. □

Theorem 7.3 *Let $(G, *)$ be a group and H a non-empty subset of G. Then H is a subgroup of G if and only if for all $a, b \in H$, $a * b \in H$ and $a^{-1} \in H$.*

Proof. One implication is easy, as above. For the other direction, let $a, b \in H$. Then by assumption, $b^{-1} \in H$ and hence $a * b^{-1} \in H$. Thus by the previous theorem, H is a subgroup. □

Example 7.4 Let G be the set of all complex numbers of modulus 1. Then G is a subgroup of $(\mathbb{C} \setminus \{0\}, \cdot)$. Also, $\{\pm i, \pm 1\}$ is a subgroup of G.

Example 7.5 Verify that $\mathrm{SL}\,(n, \mathbb{R})$ is a subgroup of $\mathrm{GL}\,(n, \mathbb{R})$.

Exercise 7.6 Find all the possible subgroups of $\mathbb{Z}_6$. Find all the possible subgroups of D_3. Is the number of subgroups the same?

Exercise 7.7 Let G be a group. The **center** of G is the set of all $a \in G$ such that for all $b \in G$, $a * b = b * a$. Prove that the center of a group G is a subgroup of G. Prove that the center is a commutative group.

Exercise 7.8 Prove that the center of a group G is G if and only if G is commutative.

Exercise 7.9 Find the centers of D_3 and D_4.

8 Plane groups

We consider designs covering an infinite plane. For each design, we consider the group of all rigid motions of $\mathbb{R}^3$ that preserves the design (see Section 3). Traditionally such a group is called **the symmetry group of the design**. For example, a blank plane allows arbitrary rotations, reflections, and translations; a plane with one circle but otherwise blank allows no translations but allows infinitely many rotations around the center of the circle and infinitely many reflections through the center of the circle; an infinite checkerboard allows rotations by 90° through the centers of the squares, rotations by 180° through the corners of the squares, translations along the diagonal, reflections through the centers of the squares and along the diagonals, and all compositions of these rigid motions.

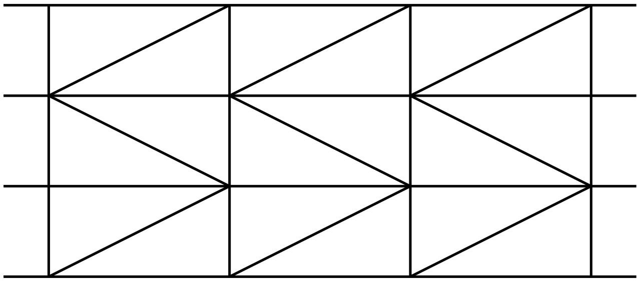
For a pattern to be classified as a plane design, it is not allowed to be without enough of a pattern; instead, we require that there be two linearly independent vectors such that every allowed translation is a translation by a linear combination of the two vectors with integer coefficients. This requirement prevents randomness and too much repetition in the design.

Bring to class some quilts/pictures and figure out all the possible allowed translations, rotations, reflections and glide reflections. An analysis of these groups shows that there are exactly 17 different plane symmetry groups. They vary from the simplest group with only the translations to the most complicated group that has in addition rotations by 60° and reflections. We will not prove the classification of the plane symmetry groups in this class.

Many cultures produced decorative patterns, and for variety's sake used different plane groups. Noted: Alhambra; ancient China; ancient Maya...

Example 8.1 Consider the infinite chessboard pattern of alternate black and white squares. Its symmetry group contains rotations by 90° around the center of each square, and rotations by 180° around each vertex. The symmetry group contains no rotations by 60°. It contains the horizontal and vertical translations t_h and t_v along the grid by 2 units but not translations along the grid by 1 unit only (coloration would not be right). However, it does contain the translation along 1 diagonal. Such a diagonal translation t_d is not in the group generated by t_h and t_v. But $t_h = 2t_d - t_v$, so t_h is in the group generated by t_v and t_d, so **the fundamental translations** are t_d and t_v. The symmetry group of the infinite chessboard further contains reflections along the diagonals of squares and reflections along the horizontal and vertical halves of the squares. It does not contain reflections along the edges. Furthermore, this symmetry group seems to contain a **glide reflection**: compose reflection along the edge by a glide (i.e., a translation) by one vertical unit. Neither the reflection along the edge nor the translation by one vertical unit are in the group, but their composition is, however, this "glide reflection" can be decomposed as the reflection along a legitimate vertical axis with a legitimate translation (check).

Example 8.2 Here is an example of a wallpaper design with a **glide reflection** that cannot be decomposed as a composition of legitimate operations. Can you find the glide reflection in addition to a reflection?



Example 8.3 Now consider the infinite non-competitive chessboard, meaning that all the clearly delineated squares are of the same color. We want to determine if its symmetry group is the same (isomorphic – what is that?) as the symmetry group of the infinite competitive chessboard. Should this be left to the time when we know what isomorphisms are?

Example 8.4 Consider the symmetry group of the infinite honeycomb (with no coloration). Since this group contains rotations by 60°, it cannot be the same as the previous two symmetry groups of the chessboards.

For each wallpaper design one CAN figure out its symmetry group, although at this stage we don't have enough vocabulary to describe these groups succinctly. What is much harder is to prove that there are only 17 possibly such plane groups. We provide a step in this direction.

Theorem 8.5 *The only possible rotations of wallpaper designs are rotations by multiples of 60° and by multiples of 90°. (And not all are possible on every design.)*

Proof. Let u and v be vectors in $\mathbb{R}^2$ such that every translation of the design is translation by a linear combination of u and v with integer coefficients. An elementary consequence of this is that for every real number C there are only finitely many vectors in $\mathbb{R}^2$ of size at most C such that translations by those vectors are allowed. Let R be an allowed rotation of the design. Then translation by Ru is also allowed because $A + Ru = R(R^{-1}A + u)$ and both translations by u and rotations by R preserve the design. Similarly, translations by $R^n u$ are allowed for all n. But $R^n u$ has the same length as u, so by the cardinality assertion there must be only finitely many vectors $R^n u$ as n varies, whence there must be an integer n such that $R^n u = u$. Thus necessarily R is the rotation by $360^\circ k/n$ for some positive integers k, n. Without loss of generality the greatest common divisor of k and n is 1. Say by the Euclidean algorithm, there exist integers a and b such that $ak + bn = 1$. In particular, R^a is rotation by $360^\circ/n$, and it is allowed on the wallpaper design. We now switch notation and we have an allowed rotation R that rotates the wallpaper design by $360^\circ/n$.

Without loss of generality $|u| \le |v|$. If $n \ge 7$, then $|Ru - u| < |u|$. Translations by $R^i(Ru-u)$ are all possible, whence translations by $R(Ru-u)-(Ru-u)$ is possible. If we set $u_0 = u$, $u_1 = Ru-u$, $u_k = Ru_{k-1}-u_{k-1}$, we get a successive sequence of translation vectors of strictly decreasing lengths, contradicting the assumptions. Thus $n \ge 7$ is impossible.

It remains to eliminate the case $n = 5$. It probably helps to draw a regular pentagon for this: the picture tells us that if two non-adjacent sides of the pentagon are translated so that two vertices that shared an edge coincide, then the line segment connecting the other two vertices is much smaller than the line segment from the center of the pentagon to a vertex. In other words, if R denotes rotation by $360°/5 = 72°$, then $Ru - u + R^3u - R^2u$ has much smaller length than u. (In standard coordinates, a numerical approximation for the matrix for $R - I + R^3 - R^2$ is $\begin{bmatrix} -0.690983 & 0.224514 \\ -0.224514 & -0.690983 \end{bmatrix}$.) Thus we get an infinite descending chain of allowed translation vectors, which gives a contradiction. □

This proves that there are only five possible groups of rotations of a wallpaper design with a given center:
(1) the trivial group;
(2) the group generated by the rotation by 60°;
(3) the group generated by the rotation by 90°;
(4) the group generated by the rotation by 120°;
(5) the group generated by the rotation by 180°.

9 Orders of groups and elements

Definition 9.1 *The number of elements of a group G is called* **the order** *of G. We denote it as $|G|$. We call G* **finite** *if it has only finitely many elements; otherwise we call G* **infinite**. *We allow infinite groups, and in this class, we do not differentiate between different infinite cardinalities.*

Definition 9.2 *Let G be a group and $a \in G$. If there is a positive integer n such that $a^n = e$, then we call the smallest such positive integer* **the order** *of a. If no such n exists, we say that a has* **infinite order**. *The order of a is denoted $|a|$.*

Of the named groups (Section 5), which ones are finite?

Work out some examples: $\mathbb{Z}$; $\mathbb{Z}_8$; D_3; the set of all 3×3 matrices under addition; $(\mathbb{Q} \setminus \{0\}, \cdot)$; $(\mathbb{R} \setminus \{0\}, \cdot)$.

Exercise 9.3 Prove that $\mathrm{GL}\,(2, \mathbb{Z}_7)$ is a finite group. Compute the number of elements in $GL(2, \mathbb{Z}_7)$.

Exercise 9.4 Find the orders of all the elements in U_5. Which elements generate U_5? Repeat for U_8.

Exercise 9.5 Give an example of a group with 35 elements. Give two examples of groups with 34 elements.

Exercise 9.6 Let G be a group. Prove or find a counterexample for each of the following:
(i) The subset of G of all elements of order 2 is a subgroup.
(ii) The subset of G of all elements of finite order is a subgroup.
(iii) The subset of G of all elements of order 2 is finite.

Exercise 9.7 Use linear algebra to describe all the diagonalizable elements in $\mathrm{GL}\,(n, \mathbb{C})$ of finite order. If you know the rational or the Jordan canonical form of matrices, describe **all** the elements in $\mathrm{GL}\,(n, \mathbb{C})$ of finite order.

Exercise 9.8 Prove that no group can have exactly two elements of order 2.

Exercise 9.9 Let a, b be elements of finite order in a group G such that $a * b = b * a$. Prove that $|a * b|$ is a factor of $\mathrm{lcm}(|a|, |b|)$. Find an example where $|a * b| < \mathrm{lcm}(|a|, |b|)$ and find an example where $|a * b| = \mathrm{lcm}(|a|, |b|)$. (Hint: try U_{28}, $a = 9$, $b = 11$?)

Exercise 9.10 Prove that for any x in a group G, $|x| = |x^{-1}|$.

Exercise 9.11 Let G be a group, $a, b \in G$. Prove that $|a| = |bab^{-1}|$. (Also handle the infinite order case.)

10 One-generated subgroups

Recall the notation: for any group G and any subset S of G, $\langle S \rangle$ stands for the subgroup generated by the elements of S. This is the smallest subgroup of G that contains S!

If S contains only one element, then $\langle S \rangle$ is said to be **one-generated**, or **cyclic**. Groups $\mathbb{Z}, \mathbb{Z}_n$ are cyclic; D_1 is cyclic but D_2 is not cyclic. Verify that the group of all complex numbers whose nth power is 1 is cyclic for all n – it is generated by $e^{2\pi i/n}$, and it may or may not be generated by $e^{4\pi i/n}$.

The main results of this section have to do with translating results on elements of finite cyclic groups to information about integers. The latter should in principle be more familiar and therefore easier!

Lemma 10.1 *Let G be a group, $a \in G$, and k an integer. If $a^k = e$, then k is a multiple of $|a|$. In particular, if i and j are integers, then $a^i = a^j$ if and only if $i - j$ is a multiple of $|a|$.*

Proof. If $|a| = \infty$, then there is no positive integer k, and therefore no negative integer k, such that $a^k = e$. Thus necessarily $k = 0$, and 0 is a multiple of ∞.

Now let $|a| < \infty$. Without loss of generality $k > 0$. By assumption, $|a|$ is the smallest positive integer such that $a^{|a|} = e$. By the Euclidean algorithm there exist non-negative integers q, r such that $k = q|a| + r$, and $r < |a|$. Then $a^r = a^{k-q|a|} = a^k * a^{-q|a|} = e * (a^{|a|})^{-q} = (e^{-q}) = e$, so that by the definition of $|a|$, $r = 0$. Hence k is a multiple of $|a|$.

Now we prove the second part. If $i - j$ is a multiple of $|a|$, then a^{i-j} is a power of $a^{|a|} = e$, whence $a^j = e * a^j = a^{i-j} * a^j = a^i$. For the converse, if $i = j$, there is nothing to prove. So without loss of generality $i > j$. Then $a^{i-j} * a^j = a^j = e * a^j$, so that by cancellation, $a^{i-j} = e$, and by the first part $i - j$ is a multiple of $|a|$. □

Theorem 10.2 *For every group G and every $a \in G$, $|\langle a \rangle| = |a|$.*

Proof. Let $n < |a|$. By the lemma, $a^0, a^1, \ldots, a^n$ are distinct elements of $\langle a \rangle$, so that $|\langle a \rangle| > n$, whence $|\langle a \rangle| \geq |a|$. Thus without loss of generality $|a| < \infty$. Every element of $\langle a \rangle$ is of the form a^i for some $i \in \mathbb{Z}$. By writing $i = q|a| + r$ for some $q, r \in \mathbb{Z}$ and $0 \leq r < |a|$, $a^i = (a^{|a|})^q * a^r = a^r$, which proves that $\langle a \rangle \subseteq \{a^0, a^1, \ldots, a^{|a|-1}\}$. □

Theorem 10.3 *Let G a be a group, $a \in G$. If a has infinite order, then for all non-zero integers k, a^k also has infinite order. If a has finite order, then*

$$|a^k| = \frac{|a|}{\gcd(k, |a|)}.$$

Proof. The infinite part is easy: there is no positive (and hence no negative) integer n such that $a^n = e$, hence $(a^k)^n \neq e$ for all non-zero integers n.

Now let $|a| < \infty$. Since $k \cdot \frac{|a|}{\gcd(k,|a|)}$ is a multiple of $|a|$, it follows by Lemma 10.1 that $|a^k|$ is a factor of $\frac{|a|}{\gcd(k,|a|)}$. But $e = (a^k)^i = a^{ki}$ implies that ki is a multiple of $|a|$, so that i is a multiple of $\frac{|a|}{\gcd(k,|a|)}$. □

In particular, the order of any element in $\langle a \rangle$ is a factor of $|a| = |\langle a \rangle|$.

Corollary 10.4 *Let G be a group, $a \in G$ of finite order, $k \in \mathbb{Z}_{>0}$. Then $\langle a^k \rangle = \langle a^{\gcd(k,|a|)} \rangle$. (We are not claiming that $a^k = a^{\gcd(k,|a|)}$!)*

Proof. Since k is a multiple of $\gcd(k,|a|)$, it follows that $\langle a^k \rangle \subseteq \langle a^{\gcd(k,|a|)} \rangle$. By the Euclidean algorithm $\gcd(k,|a|) = pk + q|a|$ for some integers p, q. Hence $a^{\gcd(k,|a|)} = (a^k)^p * (a^{|a|})^q = (a^k)^p * e = (a^k)^p \in \langle a^k \rangle$, which proves the other inequality and hence the corollary. □

Corollary 10.5 *Let G be a group, $a \in G$ of finite order, $l, k \in \mathbb{Z}_{>0}$. Then $\langle a^l \rangle = \langle a^k \rangle$ if and only if* $\gcd(k,|a|) = \gcd(l,|a|)$.

Proof. By the previous corollary, $\Leftarrow$ holds. Also, if $\langle a^l \rangle = \langle a^k \rangle$, then $\langle a^{\gcd(k,|a|)} \rangle = \langle a^{\gcd(l,|a|)} \rangle$. Thus $|\langle a^{\gcd(k,|a|)} \rangle| = |\langle a^{\gcd(l,|a|)} \rangle|$, so that by Theorem 10.2, $|a^{\gcd(k,|a|)}| = |a^{\gcd(l,|a|)}|$, whence by Theorem 10.3,

$$\gcd(k,|a|) = \gcd(\gcd(k,|a|),|a|) = \gcd(\gcd(l,|a|),|a|) = \gcd(l,|a|).$$ □

Example 10.6 Find all the generators of $\mathbb{Z}$, $\mathbb{Z}_{15}$, U_5.

Theorem 10.7 *Let a be an element of a group G. Every subgroup of $\langle a \rangle$ is cyclic and is in fact generated by a power of a.*

Proof. Let H be a subgroup of $\langle a \rangle$. If H is trivial, we are done. So we may assume that $H \neq \{e\}$. Let $h \in H \setminus \{e\}$. Then $h = a^n$ for some $n \in \mathbb{Z}$. Necessarily $n \neq 0$. Since $h^{-1} = a^{-n}$ is also in H, by possibly switching h and h^{-1} we may assume that $n > 0$. Let $S = \{m \in \mathbb{Z}_{>0} : a^m \in H\}$. We just proved that S is not empty. Let n be the smallest integer in S. We next prove that $H = \langle a^n \rangle$. Certainly H contains $\langle a^n \rangle$. Let $h \in H$. Write $h = a^m$ for some integer m, and again without loss of generality $m > 0$. By the Euclidean algorithm, $m = qn + r$ for some $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, n-1\}$. Then $a^r = a^{m-qn} = a^m * (a^n)^{-q} \in H$, so that by the assumption that n is the smallest positive integer such that the nth power of a is in H, necessarily $r = 0$. Hence $h = a^m = (a^n)^q \in \langle a^n \rangle$. Since h was arbitrary in H, this proves that $H \subseteq \langle a^n \rangle$. □

Corollary 10.8 *The number of distinct subgroups of $\mathbb{Z}_n$ is the number of distinct divisors of n (including 1 and n).*

Proof. We know that $\mathbb{Z}_n = \langle 1 \rangle$. By the theorem, every subgroup of $\mathbb{Z}_n$ is cyclic, generated by some integer m (additive operation on $\mathbb{Z}_n$). By Corollary 10.4, without loss of generality m is a factor of n. If m and k are distinct factors of n, then by Corollary 10.4, m and k generate distinct subgroups. Thus the number of distinct subgroups of $\mathbb{Z}_n$ is as desired. □

Exercise 10.9 Let $a \in G$ such that for all positive integers m, $a^m = a$. Prove that $a = e$.

Exercise 10.10 Let $m, n \in \mathbb{Z}$. Find a generator for the group $\langle m \rangle \cap \langle n \rangle$.

From now on, we will drop the writing of $*$ – just as we drop the multiplication symbol when it is clear that we are multiplying.

11 The Euler ϕ function – an aside

Here is a little bit of number theory. Define the **Euler phi function** $\phi : \mathbb{Z}_{>0} \to \mathbb{Z}$ as follows:

$$\phi(n) = \begin{cases} 1 & \text{if } n = 1; \\ \text{the number of integers in } \{0, 1, \ldots, n\} \text{ that are relatively prime to } n & \text{otherwise.} \end{cases}$$

We proved above that the number of generators of a finite group $\langle a \rangle$ is $\phi(|a|)$. Namely, $\langle a \rangle$ has $\phi(|a|)$ elements of order $|a|$. Now we want to count all a^k that have order d. By multiplying by an appropriate power of $a^{|a|}$, we only need to count $k \in \{0, 1, \ldots, |a| - 1\}$ such that a^k has order d. By Theorem 10.3, d must be $\frac{|a|}{\gcd(k,|a|)}$, so a factor of $|a|$. Thus we have to count all $k \in \{0, 1, \ldots, |a| - 1\}$ such that $\gcd(k, |a|) = \frac{|a|}{d}$, i.e., such that $\gcd(k, d \cdot \frac{|a|}{d}) = \frac{|a|}{d}$. Thus we are counting all $k \in \{0, 1, \ldots, |a| - 1\}$ that are of the form $l \cdot \frac{|a|}{d}$, where l is an integer relatively prime to d. Thus it suffices to count all the integers l on the interval $[0, \frac{(|a|-1)d}{|a|}]$ that are relatively prime to d, i.e., all the integers l on the interval $[0, d]$, that are relatively prime to d. But this is $\phi(d)$. In other words, we just proved that the number of elements in the finite group $\langle a \rangle$ of order d is

$$\begin{cases} \phi(d) & \text{if } d \text{ divides } |a|; \\ 0 & \text{otherwise (by Theorems 10.3 and 10.7).} \end{cases}$$

Thus in particular for any positive integer n,

$$n = \sum_{d|n} \phi(d).$$

Exercise 11.1 Prove that for any (positive) prime number p and any positive integer n, $\phi(p^n) = p^n - p^{n-1}$.

Exercise 11.2* Define the **Möbius function** $\mu : \mathbb{Z}_{>0} \to \mathbb{Z}$ as

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1; \\ 0 & \text{if } p^2 | n \text{ for some prime } p; \\ (-1)^r & \text{if } n \text{ is a product of } r \text{ distinct primes.} \end{cases}$$

Prove that $\phi(n) = \sum_{d|n} \mu(d) \frac{n}{d}$.

Exercise 11.3 Let m, n be relatively prime positive integers. Prove that $\phi(mn) = \phi(m)\phi(n)$. (A group-theoretic proof follows from Theorem 20.7.)

12 Permutation groups

A **permutation** of a set X is a one-to-one and onto function from X to X. In Section 3 we saw that the set S_X of all permutations of X is a group under composition.

The group of all permutations of $\{1, 2, \ldots, n\}$ is denoted as S_n, and is called **the symmetric group of degree** n (cf. Section 5). The number of elements of S_n is $n!$. Namely, the number of permutations of a set of n elements is $n!$: 1 has n choices to be mapped to, after that 2 has only $n-1$ choices left as 1 already occupied one of the n options; after that 3 has only $n-2$ choices left, etc. Thus the total is the product $n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 1 = n!$. There are various ways of representing elements of S_n, just as there are many ways of representing functions: tabular, formulaic, as a set of ordered pairs, etc. There are two standard group theory notations for elements of S_n, and we start demonstrating the first one: $f \in S_n$ can be represented as

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$$

(Gallian uses square brackets!). If

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, g = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix},$$

then

$$f \circ g = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}, g \circ f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}.$$

There is also the more compact **cyclic notation**: $f = (12)(34)$, $g = (1234)$: meaning that f takes 1 to 2 and 2 cycles around to 1, similarly that f switches 3 and 4, and that g takes 1 to 2, 2 to 3, 3 to 4, and 4 to 1. In cyclic notation, these two functions could have been written also as follows:

$$f = (21)(34) = (21)(43) = (34)(12) = (43)(21), etc., g = (2341) = (3412) = (4123).$$

The cyclic notation $(1)(2)(34)$ describes the function that switches 3 and 4 and leaves 1 and 2 intact. This is also written more briefly as $(1)(2)(34) = (34)$, and it is understood that numbers other than 3 and 4 are left intact. Thus the identity function can be written as $(1) = (1)(2) = (1)(3) = (4)$, etc. Functions can be composed, and in cyclic notation as well, the function on the right is applied first, then the function on the left. Thus for example

$$(12)(34)(1234) = (1)(24)(3),$$

which we also write as (24), and it is implicit that 1 and 3 map to themselves. Similarly,

$$(1234)(12)(34) = (13)(2)(4) = (13).$$

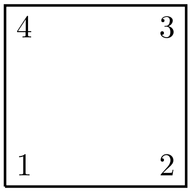
This cyclic notation makes certain structures much easier.

Recall that $D_n \subseteq S_n$. Let's write down the elements of D_3 in cyclic notation, if we label the vertices 1, 2, 3, in counterclockwise order, and the vertex 1 being at the top:

element	in cyclic notation
rotation by 120° counterclockwise	(123)
rotation by 120° clockwise	(132)
vertical flip/reflection	(23)
NW reflection	(12)
NE reflection	(13)
identity	(1)

These are in fact all the possible elements of S_3.

Let's write down all the elements of D_4. For this, we start with a square, labelled as follows:



The three non-trivial rotations are (1234), (13)(24), (1432). The vertical and the horizontal reflections are (12)(34) and (14)(23), respectively, and the two diagonal reflections are (24) (NE-SW diagonal) and (13) (NW-SE diagonal). These elements, together with the identity, account for the 8 group elements. Observe that the cyclic notation makes the computation of the multiplication (Cayley) table much easier.

Exercise 12.1 Fill in the rest of the Cayley table for D_4:

below $\circ$ right	(1)	(13)	(24)	(12)(34)	(14)(23)	(1234)	(13)(24)	(1432)
(1)	(1)	(13)	(24)	(12)(34)	(14)(23)	(1234)	(13)(24)	(1432)
(13)	(13)	(1)	(13)(24)	(1234)	(1432)	(12)(34)	(24)	(14)(23)
(24)	(24)	(13)(24)	(1)					
(12)(34)	(12)(34)							
(14)(23)	(14)(23)							
(1234)	(1234)							
(13)(24)	(13)(24)							
(1432)	(1432)							

Exercise 12.2 Compute all the powers of (123456) and of (123456789).

Definition 12.3 *A permutation in S_n is a* **k-cycle** *if in cyclic notation it has the form $(i_1 i_2 \cdots i_k)$ where $i_1, i_2, \ldots, i_k$ are distinct elements of $\{1, \ldots, n\}$. A permutation in S_n is a* **cycle** *if it is a k-cycle for some k.*

Observe that the identity is a 1-cycle (and the only 1-cycle), and that S_n has k-cycles only for $k = 1, \dots, n$.

Proposition 12.4 *The order of a k-cycle is k.*

Proof. Let $f = (i_1 i_2 \cdots i_k)$ be a k-cycle. By renaming, without loss of generality $f = (12 \cdots k)$. This permutation can be written also algebraically as: $f(x) = x + 1 \bmod k$ (where the remainders are $1, \dots, k$). If $k = 1$, certainly f has order 1. Certainly $f^k(x) = x + k \bmod k = x$, so f^k is the identity, so by Lemma 10.1, k is a multiple of $|f|$. However, if $i \in \{1, \dots, k-1\}$, then $f^i(x) = x + i \bmod k \neq x$, so that f^i is not the identity, which proves the proposition. □

Definition 12.5 *Two cycles $(i_1 i_2 \cdots i_k)$ and $(j_1 j_2 \cdots j_l)$ are* **disjoint** *if the intersection $\{i_1, i_2, \dots, i_k\} \cap \{j_1, j_2, \dots, j_l\}$ is empty.*

Theorem 12.6 *Every permutation of a finite set can be written as a cycle or as a product of pairwise disjoint cycles.*

Proof. Let f be a permutation of $\{1, 2, \dots, n\}$. The theorem is trivially true if $n = 1$. If $f(1) = 1$, then f is a permutation of $\{2, 3, \dots, n\}$, so by induction on n, f can be written as a product of disjoint cycles using only the elements $2, 3, \dots, n$. Now assume that $f(1) \neq 1$. By possibly renaming, without loss of generality $f(1) = 2$. By continuing the renaming, without loss of generality $f(2) = 3$, $f(3) = 4$, etc., $f(k) \in \{1, 2, \dots, k\}$. Since f is one-to-one and since 2 through k are in the image of f, necessarily $f(k) = 1$. Furthermore, for all $i > k$, $f(i) \notin \{1, \dots, k\}$. Thus f maps $\{k+1, \dots n\}$ to itself, and it does so bijectively. Thus by induction on n, f restricted to this smaller set can be written as a product of pairwise disjoint cycles, whence f can be written as $(12 \cdots k)$ times this latter product of pairwise disjoint cycles. □

More generally, we call two permutations α, β in S_n **disjoint** if there exists disjoint subsets A, B of $\{1, \dots, n\}$ such that α restricted to $\{1, \dots, n\} \setminus A$ equals the identity function, and β restricted to $\{1, \dots, n\} \setminus B$ equals the identity function.

Theorem 12.7 *Disjoint permutations commute.*

Proof. Let f, g be disjoint permutations in S_n. Work out why $fg = gf$. (Cf. Exercise 4.4: draw an analogy between the statement $fg = gf$ and the act of putting on and taking off a sock and a sweater.) □

Remark 12.8 Even more: every permutation of a finite set can be written uniquely as a product of pairwise disjoint cycles, where uniqueness is only up to the ordering of the disjoint cycles. Thus we can talk about the **cyclic structure** of a permutation: by this we mean the numerical information on how many k-cycles appear for each k in the writing of the permutation as a product of disjoint cycles.

Theorem 12.9 *The order of a finite product of pairwise disjoint cycles is the least common multiple of the lengths of all the cycles. (Cf. Exercise 9.9.)*

Proof. Let $\alpha_1, \ldots, \alpha_r$ be disjoint cycles in S_n. Let α_i be a k_i-cycle, and let k be the least common multiple of all the k_i. By Exercise 9.9, $(\alpha_1 \cdots \alpha_r)^k = (\alpha_1)^k \cdots (\alpha_r)^k = e$. Thus by Lemma 10.1, k is a multiple of the order d of $\alpha_1 \cdots \alpha_r$. Then by possibly relabelling the indices of the α_i (and we may since the disjoint cycles commute), we may assume that d is not an integer multiple of $k_1, \ldots, k_s$ (and is an integer multiple of the other k_i). Then

$$e = (\alpha_1 \cdots \alpha_r)^d = (\alpha_1)^d \cdots (\alpha_r)^d = (\alpha_1)^d \cdots (\alpha_s)^d.$$

Without loss of generality $\alpha_1 = (12 \cdots k_1)$, and since d is not a multiple of k_1, $(\alpha_1)^d \neq e$, and even the function α_1^d does not take 1 to 1. But since the cycles are disjoint, all α_i for $i > 1$ take 1 to 1, whence $(\alpha_1)^d \cdots (\alpha_s)^d$ does not take 1 to 1 and hence it cannot be identity. This gives a contradiction, so d must be k. □

Remark 12.10 Let's account for all the elements of S_6. We use the theorem to find the form of all the possible elements. In the sequel, a, b, c, d, e, f stand for distinct elements of $\{1, 2, 3, 4, 5, 6\}$.

order/form of elements of that order	number of such elements
order 1: (1)	1
order 2: (ab)	$\binom{6}{2} = 15$
(ab)(cd)	$\frac{1}{2}\binom{6}{2}\binom{4}{2} = 45$
(ab)(cd)(ef)	$\frac{1}{3!}\binom{6}{2}\binom{4}{2} = 15$
order 3: (abc)	$2! \cdot \binom{6}{3} = 40$
(abc)(def)	$\frac{1}{2!}2!2!\binom{6}{3} = 40$
order 4: (abcd)	$3! \cdot \binom{6}{4} = 90$
(abcd)(ef)	$3!\binom{6}{4} = 90$
order 5: (abcde)	$4!\binom{6}{5} = 144$
order 6: (abcdef)	$5! = 120$
(abc)(de)	$2!\binom{6}{3}\binom{3}{2} = 120$
Total:	$6! = 720$

Exercise 12.11 Let f be a product of a k-cycle and an l-cycle. What are the possibilities for the order of f?

Exercise 12.12 Prove that a permutation in S_n is the product of disjoint cycles of the same length if and only if it is a power of a cycle. (Hint: maybe first compute all the powers of (123456) and of (123456789).)

Above we have been talking about every permutation being a product of disjoint cycles. This turned out to be useful for computing orders of elements of S_n. But as far as finding generators of subgroups, the scrambling (un-disjointing) of cycles is also important to understand:

Theorem 12.13 *If $n > 1$, then every permutation in S_n is a product of* 2*-cycles.*

Proof. It suffices to prove that each cycle is a product of 2-cycles. (Justify.) Verify:

$$(123 \cdots k) = (1k) \cdots (14)(13)(12).$$ □

Exercise 12.14 Find all the elements of the subgroup of S_4 generated by (13) and (1234).

Exercise 12.15 Prove that S_n is generated by $(12), (23), \ldots, (n-1, n)$.

Exercise 12.16 Prove that S_n is generated by (12) and $(123 \cdots n)$.

Exercise 12.17 I brought to class the physical puzzle with 20 numbered buttons arranged in a circle. These buttons can be rotated around the circle, plus there is a region where 4 of the buttons can be rearranged with the positions a, b, c, d moving to d, c, b, a. In other words, the puzzle gives us a subgroup of S_{20} generated by (14)(23) and $(1, 2, 3, \ldots, 20)$. The computer program Gap computed that the subgroup of all possible movements has the same cardinality as 20, so at least according to Gap the subgroup is actually the whole S_{20}. Prove that $\langle (14)(23), (1, 2, 3, \ldots, 20) \rangle = S_{20}$.

Lemma 12.18 *The identity element is a product of an even number of* 2*-cycles and is not a product of an odd number of* 2*-cycles.*

Proof. Certainly $e = (12)(12)$ is a product of an even number of 2-cycles. Now suppose that $e = \alpha_1 \cdots \alpha_r$, where each α_i is a 2-cycle. Necessarily $r > 1$. Let a, b, c, d be distinct elements in $\{1, \ldots, n\}$. Observe:

$$\begin{aligned} e &= (ab)(ab), \\ (ab)(bc) &= (ac)(ab), \\ (ac)(bc) &= (bc)(ab), \\ (ab)(cd) &= (cd)(ab). \end{aligned}$$

Thus if $a \in \{1, \ldots, n\}$ appears in some α_i, the observed rewritings above (of expressions on the right to expressions on the left) guarantee the following: either r can be reduced to $r - 2$ (first case) or e can be rewritten as a product of r 2-cycles such that a appears only strictly further to the left than in the original writing of e. If r is reduced to $r - 2$, then by induction on r we are done: $r - 2$ and hence r must be even. If r cannot be reduced to $r - 2$, then it is possible to write e as a product $e = \alpha_1 \cdots \alpha_r$ of r 2-cycles such that a appears only in α_1. But such a product does not fix a, which is a contradiction. □

Definition 12.19 *A permutation is called* **even, resp. odd**, *if it can be written as a product of an even, resp. odd, number of* 2*-cycles.*

The lemma above guarantees that evenness and oddness of permutations are well-defined. The following is easy to prove:

Theorem 12.20 *The set of all even permutations in* S_n *forms a subgroup.* □

The name of this subgroup is **the alternating group of degree** n, and it is denoted A_n.

Exercise 12.21 Compute the order of A_n for $n > 1$.

Exercise 12.22 Prove that a k-cycle is in A_n if and only if k is odd.

Exercise 12.23 Prove that for $n \geq 3$, the center of S_n is $\{e\}$.

13 Group homomorphisms

This section is more general than Gallian's Chapter 6.

Keep in mind the following **"metatheorem"**: You don't get far in understanding a set X by looking at its members. You get farther by observing interactions of its members with each other and with the rest of the world.

In particular, in order to understand groups, it is not enough to just look at its elements, but to also consider functions between groups, at least those functions that preserve the group structure! We will see that such functions (homomorphisms, isomorphisms, actions) give us powerful techniques for studying groups.

Definition 13.1 *A* **group homomorphism** *from a group G to a group G' is a function $\varphi : G \to G'$ such that for all $a, b \in G$, $\varphi(ab) = \varphi(a)\varphi(b)$, where the two implicit group operations are in the two respective groups.*

Examples 13.2 Verify that the following are group homomorphisms.
(1) Let $\mathbb{Z} \to \mathbb{Z}_n$ be defined by $m \mapsto m$.
(2) If H is a subgroup of G, then the inclusion $H \hookrightarrow G$ is a group homomorphism.
(3) If $G = \langle a \rangle$, then $\mathbb{Z} \to G$ defined by $m \mapsto a^m$ is a group homomorphism.
(4) If G is the group of all complex numbers whose nth power is 1, then the function $\mathbb{Z}_n \to G$ defined by $m \mapsto e^{2m\pi i/n}$ is a group homomorphism that is bijective, and the inverse is also a group homomorphism.
(5) Verify that $\det : \mathrm{GL}\,(n, F) \to (F \setminus \{0\})$ is a group homomorphism.
(6) Verify that the trace map from $\mathrm{GL}\,(n, F)$ (or from the set of all $n \times n$ matrices under addition) to (where???) is not a group homomorphism.

Lemma 13.3 *Let $\varphi : G \to G'$ be a group homomorphism. Then*
(1) $\varphi(e) = e$.
(2) $\varphi(a^{-1}) = (\varphi(a))^{-1}$.
(3) For any integer n and any $a \in G$, $\varphi(a^n) = (\varphi(a))^n$.

Proof. For any positive integer m, $\varphi(e) = \varphi(e^m) = (\varphi(e))^m$, so that by Exercise 10.9, $\varphi(e) = e$. The rest is similarly easy. □

Proposition 13.4 *If $\varphi : G \to G'$ is a group homomorphism, and H is a subgroup of G, then $\varphi(H)$ is a subgroup of G'.*

Proof. Easy. □

Exercise 13.5 Let $\varphi : G \to G'$ be a group homomorphism. Then
(1) The **kernel** of φ, i.e., the set $\{a \in G : \varphi(a) = e\}$, is a subgroup of G.
(2) Image of φ is a subgroup of G'.
(3) For any subgroup H of G', $\varphi^{-1}(H) = \{a \in G : \varphi(a) \in H\}$ is a subgroup of G.

Theorem 13.6 *If $\varphi : G \to \varphi(G')$ is a group homomorphism, then for all $a \in G$ of finite order, $|a|$ is a multiple of $|\varphi(a)|$.*

Proof. If $a^d = e$, then $e = \varphi(e) = \varphi(a^d) = (\varphi(a))^d$, so that by Lemma 10.1, d is a multiple of $|\varphi(a)|$. In particular, $\varphi(a)$ also has finite order. □

Exercise 13.7 Let $\varphi : G \to G'$ be a group homomorphism. Prove that φ is injective (=one-to-one) if and only if the only element in G mapping to the identity in G' is the identity of G.

14 Group isomorphisms

Definition 14.1 *A group homomorphism is an* **isomorphism** *if it is a bijective map. Groups G and G' are* **isomorphic** *if there exists a group isomorphism $\varphi : G \to G'$.*

Determine which of the examples in Examples 13.2 are isomorphisms.

If G is isomorphic to G', and if one of the two groups is finite, so is the other, and $|G| = |G'|$.

Proposition 14.2 *The inverse of a bijective group homomorphism is a group homomorphism. (Compare with inverses of continuous functions!!!)*

Proof. Let $\varphi : G \to G'$ be a bijective group homomorphism. Let $a', b' \in G'$. Since φ is bijective, there exist unique $a, b \in G$ such that $\varphi(a) = a'$ and $\varphi(b) = b'$. Then

$$\varphi^{-1}(a'b') = \varphi^{-1}(\varphi(a)\varphi(b)) = \varphi^{-1}(\varphi(ab)) = ab = \varphi^{-1}(a')\varphi^{-1}(b').$$

Since a', b' were arbitrary in G', this proves that φ^{-1} is a group homomorphism. □

Exercise 14.3 Prove that the isomorphism relation on the collection of groups is an equivalence relation (check reflexivity, symmetry, transivitity).

Example 14.4 Let G be the group of 2×2 matrices of the form $\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$, where a varies over the real numbers. Verify that G is a group under matrix multiplication. This verification makes it easy to see (and prove) that G is isomorphic to $(\mathbb{R}, +)$.

Example 14.5 Verify that U_{10} is isomorphic to $\mathbb{Z}_4$ and U_5. (Examine the elements and their orders, which should give an idea on who should map to whom.

Example 14.6 The groups $\mathbb{Z}_4$ and D_2 both have 4 elements, but they are not isomorphic. Observe that $\mathbb{Z}_4$ has two elements of order 4, one of order 2, and one of order 1, whereas D_2 has three elements of order 2 and one element of order 1.

The next theorem now proves the non-isomorphism easily:

Theorem 14.7 *If $\varphi : G \to \varphi(G')$ is a group isomorphism, then for all $a \in G$, $|a| = |\varphi(a)|$.*

Proof. First assume that a has finite order. By Theorem 13.6, $|a|$ is a multiple of $|\varphi(a)|$. In particular, $\varphi(a)$ also has finite order. Since φ^{-1} is a group homomorphism, also $|\varphi(a)|$ is a multiple of $|a|$, whence $|a| = |\varphi(a)|$. Similar reasoning shows that an isomorphism maps elements of infinite order to elements of infinite order. □

Lemma 14.8 *Let n be a positive integer and let G be a cyclic group of order n. Then G is isomorphic to $\mathbb{Z}_n$.*

Proof. Write $G = \langle a \rangle$, and define $\varphi : G \to \mathbb{Z}_n$ by $\varphi(a^k) = k \bmod n$. It is easy to verify that this is a group homomorphism and that it is onto, and hence bijective. □

Remark 14.9 One of the big problems in group theory is to classify all the possible groups. We start by examining groups of small orders.
(1) All groups of order 1 are isomorphic.
(2) All groups of order 2 are isomorphic.
(3) All groups of order 3 are isomorphic. Namely, let a, b be the non-identity elements in such a group. By cancellation, $a^2 \neq a$. If $a^2 = e$, then since left multiplication by a is injective and since $ae = a$, necessarily $ab = b$, whence by cancellation $a = e$, which is a contradiction. So necessarily $a^2 \neq e$, so that $a^2 = b$. But then G is cyclic of order 3, whence by above it is isomorphic to $\mathbb{Z}_3$. (COMMENT: we will soon get to a much more efficient and elegant way of proving that any group of a prime order p is isomorphic to $\mathbb{Z}_p$. For now, we can only do this ad hoc proof.)
(4) If a group G has order 4, it is isomorphic either to $\mathbb{Z}_4$ or to D_2. Namely, let $G = \{e, a, b, c\}$. If any element of G has order 4, it generates a subgroup of order 4, whence it generates all of G, which makes G cyclic and thus isomorphic to $\mathbb{Z}_4$. So we may assume that no element of G has order 4. Thus a, b, c can have orders 2 or 3 (the latter is not really an option, but we haven't proved that yet). Say a has order 3. Then by possibly permuting b and c, $a^2 = b$. It follows that $ab = e$, and so by injectivity of the left multiplication by a, $ac = c$, whence $a = e$, which is a contradiction. So a cannot have order 3, and similarly no element of G can have order 3. Thus a, b, c all have order 2. Again by injectivity of left- and right- multiplication, $ab = c$, $ac = b$, $bc = a$, and now it is easy to see the isomorphism between D_2 and $\mathbb{Z}_4$.

Observe that all groups of order at most 4 are commutative. It is also true that all groups of order 5 are commutative. You have seen non-commutative groups of order 6.

Exercise 14.10 Prove that D_{17} is not isomorphic to $\mathbb{Z}_{34}$.

Exercise 14.11 Let G be the symmetry group of the infinite chessboard, and let G' be the symmetry group of the infinite honeycomb. Prove that G and G' are not isomorphic. (**A few** more proofs like this and you will have that there are at least 17 different plane symmetry groups.)

Exercise 14.12 Let G be the symmetry group of the infinite chessboard, and let G' be the symmetry group of the infinite non-competitive chessboard (where each square is outlined with a square but all the squares are of one color). Prove or disprove: G is isomorphic to G'.

Theorem 14.13 *If $\varphi : G \to \varphi(G')$ is a group isomorphism, then G is commutative (resp., cyclic) if and only if G' is commutative (resp., cyclic).* □

Theorem 14.14 *Let $\varphi : G \to \varphi(G')$ be a group isomorphism and H a subset of G. Then H is a subgroup of G if and only if $\varphi(H)$ is a subgroup of G'.* □

Sometimes, to understand a theorem, it is helpful to find counterexamples to the conclusion of the theorem is one or another hypothesis of the theorem is relaxed. With this in mind, come up with a group homomorphism $\varphi : G \to \varphi(G')$ and H a subset of G such that H is not a subgroup of G but $\varphi(H)$ is a subgroup of G'.

Prove that $\mathbb{Z}_4$ and D_2 are not isomorphic.

Prove that $\mathbb{Z}_4$ is not isomorphic to a subgroup of S_3.

Prove or disprove: $\mathbb{Z}_6$ is isomorphic to a subgroup of S_5.

Exercise 14.15 Let G be the group of **quaternions**: it is the group of 8 elements, generated by two elements x and y satisfying the relations $x^2 = y^2$, $xyx = y$.

(1) List all the elements of G and find their orders.

(2) Let G' be the group of order 8 generated by two elements a and b satisfying the relations: $a^4 = 1$, $b^2 = a^2$, $b^{-1}ab = a^{-1}$. Prove that G and G' are isomorphic.

Definition 14.16 *An isomorphism from a group G to itself is called an* **automorphism**. *The set of all automorphisms of G is denoted* $\mathrm{Aut}\,(G)$.

$\mathrm{Aut}\,(G)$ is a group under composition: it is straightforward to verify that the composition of group homomorphisms is a group homomorphism. In fact, $\mathrm{Aut}\,(G)$ is a subgroup of the symmetric group S_G of all one-to-one and onto functions $G \to G$. (Show that it can be a proper subgroup.)

If G is a cyclic group, an automorphism of G needs to take a generator of G to a generator of G, after which the rest of the automorphism is uniquely determined. Thus $|\mathrm{Aut}\,(\langle a \rangle)|$ is 2 if a has infinite order, and it is $\phi(|a|)$ otherwise.

Exercise 14.17 Prove that conjugation by an element of a group G is an automorphism of G. (See Exercise 2.8.)

Definition 14.18 *A function $G \to G$ is an* **inner automorphism** *if it conjugation by a for some $a \in G$. The set of all inner automorphisms of G is denoted* $\mathrm{Inn}\,(G)$.

Lemma 14.19 $\mathrm{Inn}\,(G)$ *is a subgroup of* $\mathrm{Aut}\,(G)$.

Proof. First of all, by Exercise 14.17 an inner automorphism indeed deserves the "automorphism" name. The inverse of φ is conjugation by a^{-1}, and the conjugation by a composed by conjugation by b is conjugation by ba. Thus $\mathrm{Inn}\,(G)$ is a subgroup. □

In case G is an abelian group, $\mathrm{Inn}\,(G) = \{e\}$, why, but $\mathrm{Aut}\,(G)$ can be much larger.

Theorem 14.20 $\mathrm{Aut}\,(\mathbb{Z}_n) = U_n$.

Proof. An automorphism φ of $\mathbb{Z}_n$ is determined by $\varphi(1)$ as for any integer k, $\varphi(k) = k\varphi(1)$. Since isomorphisms preserve order, $\varphi(1)$ must be a generator of $\mathbb{Z}_n$. We have proved that the generators of $\mathbb{Z}_n$ are those integers $k \in \mathbb{Z}_n$ for which $\gcd(k, n) = 1$. But these k are precisely the elements of U_n. In this way, each element a of U_n gives a distinct automorphism φ_a which is multiplication by a, and these are all the automorphisms of $\mathbb{Z}_n$. Furthermore, it is straightforward to check that

$$\psi : \mathrm{Aut}\,(\mathbb{Z}_n) \to U_n$$

given by $\psi(\varphi_a) = a$ is a group isomorphism. □

Exercise 14.21 Find $\mathrm{Aut}\,(\mathbb{Z})$.

Exercise 14.22 Find $\mathrm{Aut}\,(S_3)$.

Exercise 14.23 Find $\mathrm{Aut}\,(U_7)$.

Exercise 14.24 Let G and H be isomorphic groups. Prove that $\text{Aut}\,(G)$ is isomorphic to $\text{Aut}\,(H)$. Express the isomorphism between $\text{Aut}\,(G)$ and $\text{Aut}\,(H)$ via the isomorphism between G and H.

15 Group actions

Definition 15.1 *Let G be a group. We say that G* **acts** *on a set X if there exists a group homomorphism from G to the group S_X of all the permutations of X.*

Warning: A standard compact notation for this is as follows. Instead of referring explicitly to the group homomorphism $\varphi : G \to S_X$, if $a \in G$ and $x \in X$, we write ax instead of $(\varphi(a))(x)$. However, sometimes the shorthand notation can be hard to parse (see for example the conjugation action below in Definition 15.8).

Example 15.2 S_n acts on $\{1, \ldots, n\}$ via the identity homomorphism $S_n \to S_{\{1,\ldots,n\}}$. It consequenly also acts on $\{1, \ldots, n + m\}$ by ignoring $n + 1, \ldots, n + m$, i.e., for all $g \in S_n$ and all $i = 1, \ldots, m$, $g(n + i) = n + i$.

Example 15.3 In what natural way does D_n act on a regular n-gon?

Example 15.4 Let G be the symmetry group of the infinite chessboard, and let X be the set of all points in the chessboard that are black. Then G acts on X in the obvious way.

In general, if X is a subset of $\mathbb{R}^n$ and G is the set of all rigid motions $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $f(X) = X$, then G, which we already know to be a group, acts on X.

Example 15.5 Let G be a group and H a subgroup. Then H acts on G via $\varphi : H \to S_G$ defined as **left multiplication**:

$$(\varphi(h))(a) = ha.$$

Theorem 15.6 *(Cayley's Theorem) Every group is isomorphic to a group (consisting) of permutations. Every finite group G is isomorphic to a subgroup of $S_{|G|}$.*

Proof. Let G be a group. Then the left multiplication $G \to S_G$ is an isomorphism onto its image. □

The theorem proves that every finite group of order n is isomorphic to a subgroup of S_n. This theorem helps us write elements of groups in a more concrete way.

Exercise 15.7 Write $\mathbb{Z}_4$ and D_2 as subgroups of S_4 (explicitly write the injective group homomorphisms).

Definition 15.8 *Let G be a group and H a subgroup. Then H acts on G via* **conjugation***: $\varphi : H \to S_G$ is defined as*

$$(\varphi(h))(a) = hah^{-1}.$$

By Exercise 14.17, each $\varphi(h)$ is not just a bijective funtion from G to G, it is a group isomorphism. Elements of the form hah^{-1} are called the **conjugates** *of a.*

One needs to verify that φ is a group homomorphism (this is different from saying that $\varphi(h)$ is a group homomorphism).

Certainly conjugation by identity is the identity element of S_G. Conjugation in a commutative group is not much of an action!

Example 15.9 Let $G = D_3$ act on S_3 by conjugation. Observe:

$$\begin{gathered}
(12)(12)(12) = (12),\\
(12)(13)(12) = (23),\\
(12)(23)(12) = (13),\\
(12)(123)(12) = (132),\\
(12)(132)(12) = (123),\\
(13)(12)(13) = (23),\\
(13)(13)(13) = (13),\\
(13)(23)(13) = (12),\\
(13)(123)(13) = (132),\\
(13)(132)(13) = (123),\\
\textit{etc.}
\end{gathered}$$

Find all the orbits, all the stabilizers. Observe that conjugation preserves the cycle structure (see Remark 12.8 and the next exercise).

Exercise 15.10 The goal of this problem is to show that conjugation preserves the cyclic structure of permutations. Let $\alpha \in S_n$.

(1) Let $k \le n$. Prove that $\alpha(123\cdots k)\alpha^{-1} = (\alpha(1)\alpha(2)\cdots\alpha(k))$. (Hint: first explain all the notation?)

(2) Let $\sigma_1, \sigma_2, \ldots, \sigma_l \in S_n$. Prove that

$$\alpha(\sigma_1\sigma_2\cdots\sigma_l)\alpha^{-1} = (\alpha\sigma_1\alpha^{-1})(\alpha\sigma_2\alpha^{-1})\cdots(\alpha\sigma_l\alpha^{-1}).$$

(3) Prove that if $\sigma_1, \sigma_2, \ldots, \sigma_l$ are disjoint cycles, so are $\alpha\sigma_1\alpha^{-1}, \alpha\sigma_2\alpha^{-1}, \ldots, \alpha\sigma_l\alpha^{-1}$.

(4) Suppose that α can be written as a product of disjoint cycles $\alpha = \alpha_1\alpha_2\cdots\alpha_r$. Let α_i be a k_i cycle. Prove that $\beta\alpha\beta^{-1}$ is a product of disjoint cycles $\beta_1\beta_2\cdots\beta_r$, where β_i is a k_i cycle. (We say that conjugation **preserves the cyclic structure** of the permutation; see Remark 12.8.)

Exercise 15.11 (IMPORTANT!) Let G act on X. Let $x \in X$. (In this exercise we use the compact notation.)

(1) Define **the stabilizer of** x to be

$$G_x = \{a \in G : ax = x\}.$$

Prove that G_x is a subgroup of G.

(2) **The orbit of** x is $\{ax : a \in G\}$. Prove that two orbits are either disjoint or identical.

Example 15.12 What are the stabilizers of x under the left multiplication and under conjugation? What are the orbits of the identity element? What are the orbits of elements in S_3?

Example 15.13 Let $m \leq n$ be positive integers. Let S_m act on the set $T = \{A_n, (12)A_n\}$ by left multiplication by elements of S_m on the elements of the sets A_n and $(12)A_n$. (The set $(12)A_n$ is the set $\{(12)a : a \in A_n\}$.) Verify that this is an action! What is the kernel of $S_m \to S_T$? What are the possible orbits of this action; what are the possible stabilizers?

Example 15.14 Let $x_1, \ldots, x_n$ be n variables. We form a set X of functions that are obtained by taking fractions of polynomials in these variables. We let S_n act on X as follows: if $\alpha \in S_n$ and $f(x_1, \ldots, x_n) \in X$, then the action of α on f outputs $f(x_{\alpha(1)}, \ldots, x_{\alpha(n)})$. Examples of functions $f \in X$ that are not changed by any α in S_n are 1, $x_1 + \cdots + x_n$, $x_1^2 + \cdots + x_n^2$, …, $x_1^k + \cdots + x_n^k$, $x_1x_2 + \cdots + x_1x_n + x_2x_3 + \cdots + x_2x_n + \cdots + x_{n-1}x_n$, etc. Can you find n and a polynomial f in $x_1, \ldots, x_n$ such that f is not fixed by some element of S_n but f is fixed by A_n?

16 Cosets and Lagrange's Theorem

Definition 16.1 *Let G be a group and let H be a subgroup of G. For any $a \in G$, the set aH is defined as $\{ah : h \in H\}$, and it is called the* **left coset** *of H containing a. Similarly, the set $Ha = \{ha : h \in H\}$ is called the* **right coset** *of H containing a.*

The two kinds of cosets aH and Ha contain a since H contains e.

Note that the coset aH is the orbit of a under the action of H on G by right multiplication.

We already saw examples of cosets in Example 15.13.

Lemma 16.2 *The cardinality of any coset aH is the same as the cardinality of H.*

Proof. By Exercise 2.7, left multiplication $L_a : G \to G$ is one-to-one, so that the restriction of L_a to H is one-to-one. It is clearly onto aH. □

Lemma 16.3 *Two left cosets are either identical or disjoint.*

Proof. Since left cosets are orbits under the action of H on G by right multiplication, the lemma follows by Exercise 15.11.

For a change, here is an explicit proof for those who did not do the exercise: Let $a, b \in G$ such that $c \in aH \cap bH$. Then $c = ah = bh'$ for some $h, h' \in H$. Hence for all $h'' \in H$, $ah'' = ahh^{-1}h'' = bh'h^{-1}h'' \in bH$, whence $aH \subseteq bH$. Analogously $bH \subseteq aH$, whence any two cosets with an element in common are identical. □

Thus distinct cosets **partition** a group, which by Lemma 16.2 immediately proves the following:

Theorem 16.4 *(Lagrange's Theorem) Let G be a finite group and H a subgroup of G. Then*

$$|G| = |H| \cdot \text{(the number of distinct left cosets of } H\text{)}. \qquad \square$$

Corollary 16.5 *Let G be a group of order p, where p is a prime integer. Then G is isomorphic to $\mathbb{Z}_p$.*

Proof. Let a be a non-identity element of G. Then a generates a non-trivial subgroup of G, whence it generates a subgroup of order p. Necessarily $G = \langle a \rangle$. Thus by Lemma 14.8, G is isomorphic to $\mathbb{Z}_p$. □

Corollary 16.6 *For any a in a finite group G, $|a|$ divides $|G|$.* □

Corollary 16.7 *Let G be a finite group. Then for any $a \in G$, $a^{|G|} = e$.* □

Corollary 16.8 *(Fermat's Little Theorem) For every integer a and every prime p, $a^p \bmod p = a \bmod p$.*

Proof. If $a = 0 \bmod p$, this is obvious. Now assume that $a \neq 0 \bmod p$. Then $a \in U_p$, and U_p is a group of order $p - 1$. Thus by the previous corollary, $a^{p-1} = e$ in U_p, which proves this corollary. □

Remark 16.9 $341 = 31 \cdot 11$ divides $a^{341} - a$ for all integers a. Compare to Fermat's Little Theorem.

Corollary 16.10 *(A bigger Fermat's Little Theorem = Euler's Theorem) For every positive integer n and every integer a that is relatively prime to n, $a^{\phi(n)} \equiv 1 \bmod n$.*

Proof. By assumption $a \in U_n$, so that $a^{\phi(n)} = a^{|U_n|} = 1$, whence the corollary holds. □

Example 16.11 The relative prime assumption is needed for sure. It is not even the case that for all a, $a^{\phi(n)+1} \equiv a \bmod n$. Namely, let $n = 12$, $a = 2$. Then $\phi(12) = 4$, $2^{4+1} = 2^5 = 32 \equiv 8 \neq 2$.

Exercise 16.12 Find **all** the subgroups of A_4. Justify that you found all the subgroups. Is there a subgroup of order d for each divisor d of $|A_4|$?

Proposition 16.13 *Let H be a subgroup of G and let $a, b \in G$. Then*
(1) $aH = H$ if and only if $g \in H$.
(2) $aH = bH$ if and only if $H = a^{-1}bH$.

Proof. □

Exercise 16.14 Let G be a finite commutative group whose order is a multiple of a prime p.

(i) Prove that G is finitely generated.
(ii) Prove that if $a \in G$ has order pm, then G has an element of order p.
(iii) Prove that for $a_1, \ldots, a_n \in G$, $\langle a_1, \ldots, a_n \rangle$ has order that is a multiple of the least common multiple of the orders of the elements $a_1, \ldots, a_n$.
(iv) Prove that $|\langle a_1, \ldots, a_n \rangle|$ is a factor of the product $|a_1|\,|a_2| \cdots |a_n|$.
(v) Prove that G has an element of order p.

17 RSA public key encryption scheme

RSA is short for the discoverers Rivest, Shamir and Adleman.

Pick two very large primes p and q. Let $n = pq$. Pick an integer e that is relatively prime to $\phi(n)$. Compute an integer d such that $de \equiv 1 \bmod \phi(n)$. Make n and d public. People can now send you encrypted messages of size at most $\min\{p, q\}$ each as follows:

If the message to be sent is M, the sender actually sends the encrypted version $M^d \bmod n$.

You receive this encryption N, and you perform:

$N^e \bmod n$.

Verify that $N^e = M$.

Discuss why this works: even if an eavesdropper knows n, finding its two factors p and q is very difficult computationally. Similarly, knowing d does not help much. But you'd better keep e a secret!

Example 17.1 Let $p = 7$, $q = 11$. Then $n = 77$, $\phi(n) = 60$. Let $e = 7$. The Euclidean algorithm gives:

$$\begin{aligned}
60 &= 8 \cdot 7 + 4,\\
7 &= 1 \cdot 4 + 3,\\
4 &= 1 \cdot 3 + 1,
\end{aligned}$$

from which one can deduce, by using first the last row to express 1 as a linear combination of 3 and 4 with integer coefficients, then use second to the last row to then write 1 as a linear combination of 4 and 7 with integer coefficients, then use the first to then write 1 as a linear combination of 7 and 60:

$$1 = 2 \cdot 60 - 17 \cdot 7 = -5 \cdot 60 + 43 \cdot 7.$$

Thus $d = 43$.

A person wants to send you 3 in encrypted form. The message gets encrypted to $3^{43} \bmod 77$:

$$\begin{aligned}
3^{43} &= (3^4)^{10} \cdot 3^3 = 81^{10} \cdot 3^3\\
&\equiv 4^{10} \cdot 3^3 = (4^3)^3 \cdot 4 \cdot 3^3 = 64^3 \cdot 108\\
&\equiv (-13)^3 \cdot 31 = 169 \cdot (-13) \cdot 31 = 169 \cdot (-403)\\
&\equiv 15 \cdot (-18) = -270\\
&\equiv 38.
\end{aligned}$$

When you receive the message 38, you can decrypt it with the secret key $e = 7$ to get

$$\begin{aligned} 38^7 &= (38^2)^3 \cdot 38 = 1444^3 \cdot 38 \\ &\equiv 58^3 \cdot 38 = 58^2 \cdot 58 \cdot 38 = 3364 \cdot 2204 \\ &\equiv 53 \cdot 48 = 2544 \\ &\equiv 3. \end{aligned}$$

18 Stabilizers, orbits

Recall from Exercise 15.11 that if G acts on a set X, then for any $x \in X$, the stabilizer of x is the subgroup $G_x = \{a \in G : ax = x\}$ of G. Also, the orbit of x is $\{ax : a \in G\}$.

Theorem 18.1 *Let G act on X. Let $x \in X$. The cardinality of the orbit of x is the cardinality of the set of cosets of G_x in G.*

Proof. Let $a, b \in G$. Then $ax = bx$ if and only if $x = a^{-1}bx$ if and only if $a^{-1}b \in G_x$, which holds if and only if $G_x = a^{-1}bG_x$, i.e., if and only if $aG_x = bG_x$.

Define $f : \{ax : a \in G\} \to \{aG_x : a \in G\}$ by $ax \mapsto aG_x$. By above, f is one-to-one. Clearly f is onto. This proves the theorem. □

From this and Lagrange's Theorem (Theorem 16.4) we immediately get:

Corollary 18.2 *Assume that G acts on X and that G and X are finite. Then the number of elements in any orbit is a divisor of $|G|$.* □

Proposition 18.3 *Let G act on a set X, let $x \in X$ and $a \in G$. Then $G_{ax} = aG_xa^{-1}$.*

Proof. Let $b \in G_x$. Then $(aba^{-1})(ax) = abx = ax$ since $bx = x$. This proves that $aG_xa^{-1} \subseteq G_{ax}$. Similarly, $a^{-1}G_{ax}a \subseteq G_{a^{-1}(ax)} = G_x$, whence by multiplying on the left by a and on the right by a^{-1}, $G_{ax} \subseteq aG_xa^{-1}$. The proposition follows. □

Theorem 18.4 *Let G be a finite group acting on a finite set X. Then*

$$\textit{the number of orbits} = \frac{1}{|G|} \sum_{a \in G} |\{x \in X : ax = x\}|.$$

Proof. Observe that in the summation $\sum_{a \in G} |\{x \in X : ax = x\}|$, each $x \in X$ is counted $|G_x|$-times. Thus $\sum_{a \in G} |\{x \in X : ax = x\}| = \sum_{x \in X} |G_x|$.

Let Y be a subset of X consisting of exactly one representative from each orbit. By Exercise 15.11, two elements in X either have have identical or disjoint orbits, so that Y is well-defined. By the proposition preceeding this theorem, for any x, y in the same orbit, $|G_x| = |G_y|$. Thus

$$\sum_{a \in G} |\{x \in X : ax = x\}| = \sum_{y \in Y} |G_y||\text{orbit of } y|.$$

By Theorem 18.1 and Theorem 16.4, this is the same as $\sum_{y \in Y} |G_y| \frac{|G|}{|G_y|}$, which is exactly $|G|$ times the number of orbits. This proves the theorem. □

An application of this is to Pólya counting. Pólya used it on some chemistry applications, but we examine very simple models.

Example 18.5 How many different colorings are there of a poster with n equally spaced vertical stripes if we can use q different colors, one for each stripe (and no restriction on color adjacency)? A poster with colors ABC is by rotation by 180° the same as the poster CBA, but this is the only non-trivial rigid motion that converts such a poster to another such a poster. Thus $\mathbb{Z}_2$, the group generated by this rotation R, is the group acting on the set X of all such colorings. By elementary counting, $|X| = q^n$, and

$$|\{x \in X : Rx = x\}| = \begin{cases} q^{n/2} & \text{if } n \text{ is even;} \\ q^{(n+1)/2} & \text{if } n \text{ is odd.} \end{cases}$$

Thus by the theorem, the number of all possible posters is

$$\begin{cases} \frac{1}{2}\left(q^n + q^{n/2}\right) & \text{if } n \text{ is even;} \\ \frac{1}{2}\left(q^n + q^{(n+1)/2}\right) & \text{if } n \text{ is odd.} \end{cases}$$

Example 18.6 How many different colorations are there of an $n \times n$ chessboard with q distinct colors? Note that the group acting on all the possible colorations is $\mathbb{Z}_4$, as it is possible to rotate a chessboard by 90°. By applying the theorem, verify that the total number of possible colorations is

$$\begin{cases} \frac{1}{4}\left(q^{n^2} + 2q^{(n/2)^2} + q^{n^2/2}\right) & \text{if } n \text{ is even;} \\ \frac{1}{4}\left(q^{n^2} + 2q^{(n^2+3)/4} + q^{(n^2+1)/2}\right) & \text{if } n \text{ is odd.} \end{cases}$$

Exercise 18.7 How many different bracelets consisting of n beads can one make, if the beads come in exactly q colors?

19 Centralizer and the class equation

A subgroup H of a group G acts on G via conjugation (see Definition 15.8: for any $x \in G$ and any $a \in H$, $\varphi(a)(x)$ is defined as axa^{-1}. Usually we will have $H = G$.

By Exercise 15.11, the orbits partition G. Being in the same orbit is an equivalence relation.

Definition 19.1 *An orbit of the conjugation action is called a* **conjugacy class**. *Elements of the same conjugacy class are called* **conjugates** *of each other.*

Note that elements of the center $Z(G)$ of G have trivial conjugacy classes, consisting of only one element.

Exercise 19.2 Prove that if two permutations in S_n have the same cyclic structure, then they are conjugates of each other.

Definition 19.3 *Let G be a group acting on itself. For any x in G, the stabilizer of x is called* **the centralizer** *of x. It is denoted as $C_G(x)$.*

Since

$$C_G(x) = \{a \in G \mid axa^{-1} = x\},$$

clearly $C_G(x)$ is the set of all those elements in G that commute with x.

We know that stabilizers are subgroups of G, so that $C_G(x)$ is a subgroup of G.

The motivation behind the name "centralizer" is a sense of proximity of elements of $C_G(x)$ to the center of the group.

By Theorem 18.1 we immediately get:

Corollary 19.4 *Let G be a finite group and $x \in G$. Then the number of conjugates of x (i.e., elements in G of the form axa^{-1} as a varies over G) is the number of cosets of the subgroup $C_G(x)$.* □

We have seen the power of partitioning G with orbits when the action was right multiplication by elements from a subgroup H. In that case we counted the number of cosets, and if G was finite, we established that $|G|$ is the sum of elements of the cosets as we vary over distinct cosets.

Similarly we can partition a finite group G into its conjugacy classes to obtain:

$$|G| = \sum_i \frac{|G|}{|C_G(x_i)|},$$

where x_i vary over representatives of distinct conjugacy classes. This is usually expressed as follows:

Definition 19.5 *If G is a finite group, then*

$$|G| = |Z(G)| + \sum_i \frac{|G|}{|C_G(x_i)|},$$

where x_i vary over representatives of distinct conjugacy classes that have more than 1 element. This is the **class equation** *of G.*

Theorem 19.6 *Let p be a prime integer and n a positive integer. All groups of order p^n have a non-trivial center.*

Proof. Let G be a group of order p^n. By Theorem 16.4 (Lagrange's Theorem), for all $x \in G$, $|C_G(x)|$ and $|G|/|C_G(x)|$ are powers of p (possibly a zeroth power of p). If for some x, $\frac{|G|}{|C_G(x)|}$ is not a multiple of p, it has to be 1, whence $C_G(x) = G$, which proves that $x \in Z(G)$. Then by the class equation, $|Z(G)| = |G| - \sum_i \frac{|G|}{|C_G(x_i)|}$ is a multiple of p. □

20 External direct products/sums

Definition 20.1 *Let $G_1, \ldots, G_k$ be groups. The* **external direct product (or sum)** *of $G_1, \ldots, G_k$ is written as $G_1 \oplus G_2 \oplus \cdots \oplus G_k$. As a set, it equals the Cartesian product $G_1 \times G_2 \times \cdots \times G_k$, but the external direct product also has a binary operation on it:*

$$(g_1, g_2, \ldots, g_k) * (h_1, h_2, \ldots, h_k) = (g_1 * h_1, g_2 * h_2, \ldots, g_k * h_k),$$

where the operation in the ith component is the group operation in G_i. (Duh!) This operation makes the direct product into a group (verify).

Clearly the cardinality of $G_1 \oplus G_2 \oplus \cdots \oplus G_k$ is the product of the cardinalities of the G_i (finite or infinite, even fine-tuned infinite).

The adjective "external" has to do with the fact that G_1 is not a subset of $G_1 \oplus \cdots \oplus G_k$ if $k > 1$. However, obviously $G_1 \oplus \{e\} \oplus \cdots \oplus \{e\}$ is a subgroup of $G_1 \oplus \cdots \oplus G_k$ that is isomorphic to G_1.

Vector space analogy...

Example 20.2 Verify: $\mathbb{Z}_2 \oplus \mathbb{Z}_5 \cong \mathbb{Z}_{10}$.

Example 20.3 Verify: $\mathbb{Z}_2 \oplus \mathbb{Z}_4 \not\cong \mathbb{Z}_8$.

Theorem 20.4 *Let $G_1, \ldots, G_k$ be groups. For $(g_1, g_2, \ldots, g_k) \in G_1 \oplus G_2 \oplus \cdots \oplus G_k$,*

$$|(g_1, g_2, \ldots, g_k)| = \mathrm{lcm}\{|g_1|, \ldots, |g_k|\}.$$

Proof. $(g_1, \ldots, g_k)^l = e$ if and only if $(g_1^l, \ldots, g_k^l) = e$, which holds if and only if for all $j = 1, \ldots, k$, $g_j^l = e$. Thus $|(g_1, g_2, \ldots, g_k)|$ is a multiple of $|g_i|$ for all i, whence it is a multiple of $\mathrm{lcm}\{|g_1|, \ldots, |g_k|\}$. But it couldn't be anything smaller, for otherwise some g_i^l wouldn't be e. □

Lemma 20.5 *Let G and H be finite groups such that $G \oplus H$ is cyclic. Then G and H are cyclic and $|G|$ and $|H|$ are relatively prime.*

Proof. Suppose that $G \oplus H$ is generated by (a, b). Then G is generated by a, so G is cyclic. Similarly $H = \langle b \rangle$ is cyclic. By the previous theorem, $|a| \cdot |b| = |G| \cdot |H| = |G \oplus H| = |(a, b)| = \mathrm{lcm}\{|a|, |b|\}$, so that the orders of a and b are relatively prime. Thus the orders of G and H are relatively prime. □

The main result of this section is a version of the Chinese Remainder Theorem (Theorem 20.6), whose consequence is also the converse of the lemma above. We first give a general set-up.

For any positive integers $n_1, \ldots, n_k$, define the obvious map:

$$\varphi : \mathbb{Z}_{n_1 \cdots n_k} \to \mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2} \oplus \cdots \oplus \mathbb{Z}_{n_k}$$

by $m \in \mathbb{Z}_{n_1 \cdots n_k}$ mapping to the k-tuple $(m \bmod n_1, \ldots, m \bmod n_k)$. This is well-defined and is a group homomorphism. The kernel consists of the images in $\mathbb{Z}_{n_1 \cdots n_k}$ of those integers m for which m is a multiple of n_i for all i. Thus the kernel is the subgroup $\langle n_1 \rangle \cap \cdots \cap \langle n_k \rangle$ in $\mathbb{Z}_{n_1 \cdots n_k}$, which is $\langle \mathrm{lcm}\{n_1, \ldots, n_k\} \rangle$.

For pairwise relatively prime $n_1, \ldots, n_k$, the kernel is $\langle n_1 \cdots n_k \rangle$, so that the obvious map φ above is injective.

Theorem 20.6 *(Criterion for when $\mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is isomorphic to $\mathbb{Z}_{n_1 \cdots n_k}$) $\mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is isomorphic to $\mathbb{Z}_{n_1 \cdots n_k}$ if and only if the n_i are pairwise relatively prime.*

Proof. If $\mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is isomorphic to $\mathbb{Z}_{n_1 \cdots n_k}$, then by Lemma 20.5 and by iduction on k, the n_i are pairwise relatively prime. It remains to prove the other direction. So assume that the n_i are pairwise relatively prime. By the set-up just before this theorem, the obvious map $\varphi : \mathbb{Z}_{n_1 \cdots n_k} \to \mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is injective. For $l = 1, \ldots, k$, set $m_l = n_1 \cdots n_{l-1} n_{l+1} \cdots n_k$. Note that $\gcd(m_1, \ldots, m_l) = n_{l+1} \cdots n_k$. By the Euclidean algorithm and induction on l, we can write $n_{l+1} \cdots n_k$ as a linear combination of $m_1, \ldots, m_l$ with integer coefficients. In particular, there exist integers $a_1, \ldots, a_k$ such that $1 = a_1 m_1 + a_2 m_2 + \cdots + a_k m_k$. Define

$$\psi : \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k} \to \mathbb{Z}_{n_1 \cdots n_k} \text{ by } (b_1, \ldots, b_k) \mapsto a_1 m_1 b_1 + \cdots + a_k m_k b_k.$$

This is a well-defined function: for arbitrary integers $l_1, \ldots, l_k$,

$$\begin{aligned} \psi(b_1 + l_1 n_1, \ldots, b_k + l_k n_k) &= a_1 m_1 (b_1 + l_1 n_1) + \cdots + a_k m_k (b_k + l_k n_k) \\ &\equiv a_1 m_1 b_1 + \cdots + a_k m_k b_k \mod (n_1 \cdots n_k) \\ &= \psi(b_1, \ldots, b_k). \end{aligned}$$

It is a group homomorphism:

$$\begin{aligned} \psi(b_1 + b_1', \ldots, b_k + b_k') &= a_1 m_1 (b_1 + b_1') + \cdots + a_k m_k (b_k + b_k') \\ &= (a_1 m_1 b_1 + \cdots + a_k m_k b_k) + (a_1 m_1 b_1' + \cdots + a_k m_k b_k') \\ &= \psi(b_1, \ldots, b_k) + \psi(b_1', \ldots, b_k'). \end{aligned}$$

The map is surjective because $1 = \psi(1, 1, \ldots, 1)$ by construction. And lastly, ψ is the inverse of φ because

$$\varphi \circ \psi(b_1, \ldots, b_k) = \varphi(a_1 m_1 b_1 + \cdots + a_k m_k b_k).$$

The first component of this in $\mathbb{Z}_{n_1}$ is $a_1 m_1 b_1 + \cdots + a_k m_k b_k \bmod n_1$. As $m_2, \ldots, m_k$ are multiples of n_1, this is the same as $a_1 m_1 b_1 \bmod n_1$, and from the equality $1 = a_1 m_1 + a_2 m_2 + \cdots + a_k m_k$, we deduce that $a_1 m_1 \bmod n_1 = 1$. Thus the first component of $\varphi \circ \psi(b_1, \ldots, b_k)$ equals b_1, and similarly for the other components, so that $\varphi \circ \psi$ is the identity function. Then $\varphi \circ \psi \circ \varphi = \varphi$, and since φ is one-to-one, necessarily $\psi \circ \varphi$ is the identity as well. □

(Discuss the Chinese Remainder Theorem.)

Somewhat more interesting/harder:

Theorem 20.7 *Criterion for when $U_{n_1} \oplus U_{n_2} \oplus \cdots \oplus U_{n_k}$ is isomorphic to $U_{n_1 \cdots n_k}$: If every pair of the n_i is relatively prime, then the isomorphism holds.*

Proof. Let

$$\overline{\varphi} : U_{n_1 \cdots n_k} \to U_{n_1} \oplus U_{n_2} \oplus \cdots \oplus U_{n_k}$$

be the restriction of φ above. We first need to verify that the map is well-defined, namely that the image of $\overline{\varphi}$ is in $U_{n_1} \oplus U_{n_2} \oplus \cdots \oplus U_{n_k}$. Indeed, if m is relatively prime to $n_1 \cdots n_k$, it is relatively prime to each n_i. Thus $\overline{\varphi}$ is well-defined. Since this map is a restriction of an injective map, it is injective. It is easy to verify that it is a group homomorphism (this does not follow from φ being a homomorphism, because the two operations are different), and to prove surjectivity, let $(b_1, \ldots, b_k) \in U_{n_1} \oplus U_{n_2} \oplus \cdots \oplus U_{n_k}$. Then $\psi(b_1, \ldots, b_k)$ is actually in $U_{n_1 \cdots n_k}$, and $\overline{\psi}(b_1, \ldots, b_k) = (b_1, \ldots, b_k)$, which proves that $\overline{\varphi}$ is surjective. □

This allows us to verify that if $n_1, \ldots, n_k$ are pairwise relatively prime positive integers, then $\phi(n_1 \cdots n_k) = \phi(n_1) \cdots \phi(n_k)$ (Euler ϕ function), which we already proved in Exercise 11.3 and used in Section 17. If we know that for any (positive) prime number p and any positive integer n, $\phi(p^n) = p^n - p^{n-1}$, the last corollary enables an easy way to compute the Euler function for any integer whose prime factorization is known (but getting to a prime factorization is another problem!). For example, $\phi(2350) = |U_{2350}| = |U_{2 \cdot 5^2 \cdot 47}| = |U_2 \oplus U_{5^2} \oplus U_{47}| = 1 \cdot 20 \cdot 46 = 920$.

Exercise 20.8 Prove that $\mathbb{Z}_{200} \cong \mathbb{Z}_8 \oplus \mathbb{Z}_{25}$. Prove that $\mathbb{Z}_{200} \not\cong \mathbb{Z}_8 \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_5$.

Exercise 20.9 Assume that $G_i \cong H_i$ (group isomorphisms). Prove that $G_1 \oplus G_2 \oplus \cdots \oplus G_k \cong H_1 \oplus H_2 \oplus \cdots \oplus H_k$.

Exercise 20.10 Find the orders of all the elements in $\mathbb{Z}_4 \oplus U_7$.

Exercise 20.11 Find at least four non-isomorphic groups of order 36. (Prove all claims.)

Exercise 20.12 Let I be a set. For each $i \in I$, let G_i be a group. Define the **external direct product** $\Pi_{i \in I} G_i$ of the $G_i, i \in I$ to be the set consisting of all (ordered) tuples $(a_i : i \in I)$, where for each $i \in I$, $a_i \in G_i$.

(1) Prove that the componentwise group operations make $\Pi_{i \in I} G_i$ into a group.

(2) What is the cardinality of $\Pi_{i \in I} G_i$?

Exercise 20.13 Let I be a set. For each $i \in I$, let G_i be a group. Define the **external direct sum** $\sum_{i \in I} G_i$ of the $G_i, i \in I$ to be the set consisting of all (ordered) tuples $(a_i : i \in I)$, where for each $i \in I$, $a_i \in G_i$, and at most finitely many of the a_i are not equal to the identity element in G_i.

(1) Prove that the componentwise group operations make $\sum_{i \in I} G_i$ into a group.

(2) Prove that $\sum_{i \in I} G_i$ is a subgroup of $\Pi_{i \in I} G_i$.

Remark 20.14 It is not true that $\text{Aut}\,(G \oplus H) \cong \text{Aut}\,(G) \oplus \text{Aut}\,(H)$. For example, try $G = H = \mathbb{Z}_2$. Under what conditions does the isomorphism hold? Verify it for $G = \mathbb{Z}_n$ and $H = \mathbb{Z}_m$ for relatively prime m and n.

Exercise 20.15 Prove that $\text{Inn}\,(G \oplus H) \cong \text{Inn}\,(G) \oplus \text{Inn}\,(H)$.

21 Normal subgroups

We have seen quite a few examples of action. Here is another one: Let G be a group and let X be the set of all its subgroups. Then G acts on X via **conjugation**: for any $H \in X$ and any $a \in G$, the set $aHa^{-1} = \{aha^{-1} : h \in H\}$ is a subgroup of G, and it is called **a conjugate of** H, or more precisely, an **a-conjugate** of H.

Definition 21.1 *A subgroup H of a group G is* **normal** *if it has only one conjugate. We denote this by $H \triangleleft G$.*

Note that a subgroup is normal if and only if $aH = Ha$ for all $a \in G$.

Incidentally, by Exercise 15.11, the stabilizer of a subgroup H under conjugation is a subgroup of G. This stabilizer is called **the normalizer** of H, and it is denoted $\mathbf{N_G(H)}$.

Remark 21.2 It is easy to verify the following:
(1) In a commutative group, every group is normal.
(2) A_n is a normal subgroup of S_n.
(3) The subgroup of rotations in D_n is a normal subgroup of D_n.
(4) If $G_1, \ldots, G_k$ are groups, let $H = G_1 \oplus \cdots \oplus G_{l-1} \oplus \{e_{G_l}\} \oplus G_{l+1} \oplus \cdots \oplus G_k$. Then H is a normal subgroup of $G_1 \oplus \cdots \oplus G_k$.
(5) The subgroup $\langle(12)\rangle$ in S_3 is not normal.
(6) Any subgroup of G contained in $Z(G)$ is normal in G.

Exercise 21.3 Prove that $\mathrm{SL}\,(n, \mathbb{R})$ is a normal subgroup of $\mathrm{GL}\,(n, \mathbb{R})$.

Exercise 21.4 Let $\varphi : G \to H$ be a group homomorphism. Prove that the kernel of φ is a normal subgroup of G.

Exercise 21.5 Prove that if H is a subgroup of G and H has at most two cosets, then H is a normal subgroup.

Exercise 21.6 The goal of this exercise is to prove that A_5 has no normal subgroups other than itself and the trivial subgroup.
(1) Prove that any two 3-cycles in A_5 are conjugate in A_5. (By Exercise 15.10, they are conjugate in S_5, but here you are not allowed to conjugate by odd permutations.)
(2) Prove that (12)(34) has 15 conjugates in A_5.
(3) Prove that all products of two disjoint 2-cycles are conjugate in A_5.
(4) Prove that there are two conjugacy classes of 5-cycles in A_5, each of which has 12 elements.
(5) Prove that A_5 has no normal subgroups other than itself and the trivial subgroup.

22 Factor (or quotient) groups

Definition 22.1 *Let G be a group and let H be a normal subgroup of G. The set of all cosets of H forms a group under the operation $(aH)(bH) = abH$. The notation for this group is G/H or $\frac{G}{H}$, and we read it as* **G mod(ulo) H**.

Why is this group well-defined? In other words, if $aH = a'H$ and $bH = b'H$ for some $a, b, a', b' \in G$, why is $abH = a'b'H$? Work it out (and use that H is normal).

Warning: When the group operation on G is $+$, the notation for the cosets is accordingly additive: $a + H$ rather than aH. In particular, a coset in $\mathbb{Z}$ of a subgroup H is denoted $a + H$. A further warning against confusion: in $\mathbb{Z}$, the subgroup generated by an integer m consists of elements of $\mathbb{Z}$ that are multiples of $\mathbb{Z}$. We write such a set (obviously!!!) as $m\mathbb{Z}$. Thus here, $m\mathbb{Z}$ is a subgroup, and its cosets are subsets of the form $a + m\mathbb{Z}$.

Proposition 22.2 *Let n be a positive integer. Then $n\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.*

Proof. Do it. □

Proposition 22.3 *Let n and m be positive integers. Then $n\mathbb{Z}_m$ is a normal subgroup of $\mathbb{Z}_m$ and $\mathbb{Z}_m/n\mathbb{Z}_m \cong \mathbb{Z}_{\gcd(m,n)}$.*

Proof. Do it. □

Theorem 22.4 *A_4 has no subgroup of order* 6.

Proof. Suppose for contradiction that H is a subgroup of A_4 of order 6. By Exercise 21.5, H is a normal subgroup of A_4. Since the group A_4/H has order 2, for every element $a \in A_4$, $(aH)^2 = H$ (group operation in A_4/H). This means that for all $a \in A_4$, $a^2 \in H$. In particular, if a is a 3-cycle, $a^2 \in H$, so that H contains squares of all 3-cycles, which is all the possible 3-cycles. But A_4 has 8 distinct 3-cycles, and they cannot all be contained in a subgroup with 6 elements. □

Remark 22.5 Let G be a group, H a normal subgroup, and $a \in G$. Then the order of aH in G/H is a factor of the order of a. Verify.

The following was already touched upon in Exercise 16.14; here it is proved more concisely.

Theorem 22.6 *(Cauchy's Theorem for Commutative Groups) Let G be a finite commutative group and let p be a prime integer that divides the order of $|G|$. Then $|G|$ has an element of order p.*

Proof. Let $x \in G \setminus \{e\}$. If the order of x is pm for some integer m, then x^m has order p, and we are done. Now assume that $x \in G$ and that the order t of x is relatively prime to p. Then $\langle x \rangle$ is a normal subgroup of G and $G/\langle x \rangle$ is a finite group of order that is a multiple of p but strictly smaller than the order of G. By induction, there exists $y \in G$ such that the order of $y\langle x \rangle$ in $G/\langle x \rangle$ is p. By the remark above, the order of y is a multiple of p, so that some power of y has order p. □

Theorem 22.7 *(Cauchy's Theorem for Groups) Let G be a finite group and let p be a prime integer that divides the order of $|G|$. Then $|G|$ has an element of order p.*

Proof. Let $x \in G\backslash\{e\}$. Recall that the number of conjugates of x is the same as the number of cosets of the centralizer $C_G(x)$ of x (Definition 19.3). If $x \notin Z(G)$, then $|C_G(x)| < |G|$. If $|C_G(x)|$ is a multiple of p, then by induction $C_G(x)$ has an element of order p, whence G has an element of order p. If instead $|C_G(x)|$ is relatively prime to p for all $x \notin Z(G)$, then by the class equation of G (Definition 19.5), $|Z(G)|$ is a multiple of p, whence by the commutative case, $Z(G)$ has an element of order p, which means that G has an element of order p. □

Observe that this is no longer true if p is not prime: S_3 has no elements of order 6 and D_2 has no elements of order 4.

Exercise 22.8 Let p be a prime, and let G be a finite group. Prove that the order of G is a power of p if and only if every element in G has order a power of p.

Exercise 22.9 Let p be a prime, let G be a group of order p^n, and let H be a subgroup of order p^k for some $k < n$. Prove that G has a subgroup K containing H such that $|K| = p^{k+1}$. (Hint: H contains a normal subgroup of G of order p.)

Theorem 22.10 *If G is a group and H a subgroup of $Z(G)$ such that G/H is cyclic, then G is abelian.*

Proof. First of all, $Z(G)$ is a normal subgroup of G, and H is a normal subgroup of $Z(G)$. (Warning: in general, being a normal subgroup is not a transitive property, but here it works. Say why.) Let $g \in G$ such that $\langle gH\rangle = G/H$. Let $a, b \in G$. Then there exist $i, j \in \mathbb{Z}_{>0}$ such that $aH = g^iH$ and $bH = g^jH$. Thus there exist $\alpha, \beta \in H$ such that $a = g^i\alpha$ and $b = g^j\beta$. Then $ab = g^i\alpha g^j\beta$. Since $H \subseteq Z(G)$, $ab = g^ig^j\alpha\beta = g^{i+j}\beta\alpha = g^jg^i\beta\alpha = g^j\beta g^i\alpha = ba$. □

Theorem 22.11 *Let p be a prime integer. All groups of order p^2 are abelian.*

Proof. Let G be a group of order p^2 and let $a \in G \setminus \{e\}$. By Theorem 19.6, $Z(G)$ is not trivial. If $|Z(G)| = p$, then $G/Z(G)$ is a group of order p, hence cyclic, so by Theorem 22.10, G is abelian, whence $Z(G) = G$. In any case, $Z(G) = G$. □

Exercise 22.12 Let G be a group. Prove that $G/Z(G)$ is isomorphic to $\operatorname{Inn}(G)$.

From now on, the groups (and rings) $\mathbb{Z}_n$ will be written as $\frac{\mathbb{Z}}{n\mathbb{Z}}$ or $\mathbb{Z}/n\mathbb{Z}$.

23 The internal direct product

Let H and K be subgroups of G. By HK we denote the **set** $\{hk : h \in H, k \in K\}$.

Example 23.1 Let $G = S_3$, $H = \langle(12)\rangle$, $K = \langle(13)\rangle$. Then $HK = \{(1), (12), (13), (132)\}$. Observe that HK is not a subgroup here!

Proposition 23.2 *Let H and K be subgroups of G, and assume that H is normal in G. Then HK is a subgroup of G.*

Proof. Let $h, h' \in H$, $k, k' \in K$. Then $(hk)(h'k')^{-1} = hkk'^{-1}h'^{-1} \in Hkk'^{-1}H = H(Hkk'^{-1}) \subseteq HK$. □

Theorem 23.3 *Let G be a group with normal subgroups H and K such that $H \cap K = \{e\}$ and $HK = G$. Then $G \cong H \oplus K$.*

Proof. Each $g \in G$ can be expressed as $g = hk$ for some $h \in H$ and $k \in K$. These g and h are unique, for if $hk = h'k'$ for some $h' \in H$ and $k' \in K$, then $(h')^{-1}h = k'k^{-1} \in H \cap K = \{e\}$, whence $h' = h$ and $k' = k$. Thus we can define the function

$$\varphi : G \to K \oplus H$$

by $\varphi(g) = (h, k)$. We just proved that φ is well defined. Why is φ a group homomorphism? We first claim that for $h \in H$ and $k \in K$, $hk = kh$. Note: $hkh^{-1}k^{-1} = (hkh^{-1})k^{-1} \in K$ since K is normal, and similarly $hkh^{-1}k^{-1} = h(kh^{-1}k^{-1}) \in H$, whence $hkh^{-1}k^{-1} \in H \cap K = \{e\}$, and it follows that $hk = kh$. Now we can prove that φ is a group homomorphism: let $g, g' \in G$, and write $g = hk$, $g' = h'k'$ for some $h, h' \in H$ and $k, k' \in K$. Then $\varphi(gg') = \varphi(hkh'k') = \varphi(hh'kk') = (hh', kk') = (h, k)(h', k') = \varphi(hk)\varphi(h'k') = \varphi(g)\varphi(g')$. It remains to prove that φ is injective and surjective, but that is easy. □

Definition 23.4 *We say that the group G is the* **internal direct product** *of (its subgroups) H and K if H and K are normal subgroups of G, $G = HK$, and $H \cap K = \{e\}$. We write this as $G = H \times K$.*

We say that G is the **internal direct product** *of subgroups $H_1, \ldots, H_k$ if each H_i is normal in G, $G = H_1 \cdots H_k$, and for any $i \in \{1, \ldots, k\}$, $(H_1H_2 \cdots H_i) \cap H_{i+1} = \{e\}$.*

Exercise 23.5 Prove that D_2 is an internal direct product of two non-trivial subgroups.

Exercise 23.6 Prove that S_3 is not an internal direct product of two non-trivial subgroups.

Exercise 23.7 Prove that U_{98} is an internal direct product of $\langle 1 \rangle$ and $\langle 74 \rangle$.

Exercise 23.8 Let $G_1, \ldots, G_k$ be groups, let let $H_i = G_1 \oplus \cdots \oplus G_{l-1} \oplus \{e_{G_l}\} \oplus G_{l+1} \oplus \cdots \oplus G_k$. Prove that $G_1 \oplus \cdots \oplus G_k$ is an internal direct product of $H_1, \ldots, H_k$.

Remark 23.9 If H is a normal subgroup of G and K is a subgroup such that $HK = G$ and $H \cap K = \{e\}$, then G need not be isomorphic to $H \oplus K$. For example, let $G = S_3$, $H = \langle (123) \rangle$, $K = \langle (12) \rangle$, the hypotheses are satisfied. However, we know that elements of H and K do not commute, so there couldn't be any such isomorphism.

However, with the set-up as in the remark, G is called a **semidirect product** of H by K, denoted $G = H \rtimes K$.

Here are a few related facts:

(1) Let G be a group. If H is a normal subgroup of G and K is a subgroup, define $\theta : K \to \mathrm{Aut}\,(H)$ by $\theta(k)(h) = hkh^{-1}$. Prove that θ is a group homomorphism.

(2) With groups H and K and a group homomorphism $\theta : K \to \mathrm{Aut}\,(H)$, define $H \rtimes_\theta K$ as the set of all pairs $(h, k) \in H \times K$ under the binary operation

$$(h, k)(h', k') = (h\theta(k)(h'), kk').$$

Prove that $H \rtimes_\theta K$ is a group.

(3) Let G be a group and H and K subgroups such that there exists a group homomorphism $\theta : K \to \mathrm{Aut}\,(H)$. Under what conditions is $H \rtimes_\theta K$ isomorphic to G?

24 The isomorphism theorems

Theorem 24.1 (First Isomorphism Theorem) *Let $\varphi : G \to H$ be a homomorphism with kernel K. Then K is a normal subgroup of G and $G/K \cong \mathrm{Im}\varphi$.*

Proof. By Exercise 21.4, K is normal. Define $\Phi : G/K \to \mathrm{Im}\varphi$ by $\Phi(aK) = \varphi(a)$. This map is well-defined: if $aK = bK$, then $ab^{-1} \in K$, so $\varphi(ab^{-1}) = e$, whence $\varphi(a) = \varphi(b)$. It is a homomorphism: $\Phi((aK)(bK)) = \Phi(abK) = \varphi(ab) = \varphi(a)\varphi(b) = \Phi(aK)\Phi(bK)$. It is clearly surjective, and it is injective because if $\Phi(aK) = e = \varphi(a)$, then $a \in K$. □

This shows that there is no **significant** difference between a factor group and a homomorphic image of a group homomorphism.

Exercise 24.2 Use the first isomorphism theorem to prove that any two finite cyclic groups of the same order are isomorphic.

Corollary 24.3 *A subgroup H of G is normal if and only if H is the kernel of some group homomorphism.*

Proof. It suffices to prove that if H is normal, then it is the kernel of a group homomorphism. In fact, it is the kernel of the **natural** group homomorphism $\varphi : G \to G/H$. What is the **natural** map here? Verify all details. □

Exercise 24.4 Let H be a normal subgroup of G. Assume that H has n right cosets in G. Prove that for all $a \in G$, $a^n \in H$. Give an example of a non-normal subgroup for which this fails.

Theorem 24.5 (Second Isomorphism Theorem) *Let K and H be subgroups of G, with H normal. Then $K \cap H$ is a normal subgroup of K, and $K/(K \cap H) \cong KH/H$.*

Exercise 24.6 Prove the second isomorphism theorem.

Theorem 24.7 (Third Isomorphism Theorem) *Let K and H be normal subgroups of G, and assume that $K \subseteq H$. Then H/K is a normal subgroup of G/K, and*

$$(G/K)/(H/K) \cong G/H.$$

Exercise 24.8 Prove the third isomorphism theorem.

25 Fundamental Theorem of Finite Abelian Groups

Theorem 25.1 (Fundamental Theorem of Finite Abelian Groups) *A finite commutative group G is isomorphic to $\frac{\mathbb{Z}}{n_1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{n_2\mathbb{Z}} \oplus \cdots \oplus \frac{\mathbb{Z}}{n_k\mathbb{Z}}$ for some positive integers $n_1, n_2, \cdots, n_k$.*

I'll skip the proof now and return to it after ring theory, when we'll have more elegant machinery then. See Corollary 39.3.

We can check out an example now. For example, we start with the group $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$, the shorthand for which is $\mathbb{Z}^3$. This is certainly a commutative group. Thus every subgroup is normal, so we will take

$$G = \frac{\mathbb{Z}^3}{\langle (1,-1,1),(5,1,-5),(-3,-3,29)\rangle}.$$

Then G is certainly a commutative group. It is not clear at all that it is finite. We will rewrite G with different generators, and for this we will suggestively record the relation vectors as columns in a matrix:

$$\begin{bmatrix} 1 & 5 & -3 \\ -1 & 1 & -3 \\ 1 & -5 & 29 \end{bmatrix}.$$

Switching the rows corresponds to switching the generators e_1, e_2, e_3 of G, so the group remains unchanged. Adding an integer multiple of one row to another similarly corresponds replacing one of the generators e_1, e_2, e_3 with the sum of itself plus a multiple of another. This also does not change the group. Furthermore, multiplying by ± 1 does not change the group. (However, multiplying by 2 does change it, as this operation is not invertible!) Thus the suggestive matrix above can be partially row-reduced to

$$\begin{bmatrix} 1 & 5 & -3 \\ 0 & 6 & -6 \\ 0 & -10 & 32 \end{bmatrix}.$$

We can also perform column reductions (say why). Then we get

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & -6 \\ 0 & -10 & 32 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & -6 \\ 0 & 2 & 20 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 20 \\ 0 & 6 & -6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 20 \\ 0 & 0 & -66 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 66 \end{bmatrix}.$$

Thus

$$G \cong \frac{\mathbb{Z}^3}{\langle (1,0,0),(0,2,0),(0,0,66)\rangle} \cong \frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{2\mathbb{Z}} \oplus \frac{\mathbb{Z}}{66\mathbb{Z}} \cong \frac{\mathbb{Z}}{2\mathbb{Z}} \oplus \frac{\mathbb{Z}}{66\mathbb{Z}},$$

which verifies the Fundamental Theorem in this case.

26 Sandpile groups

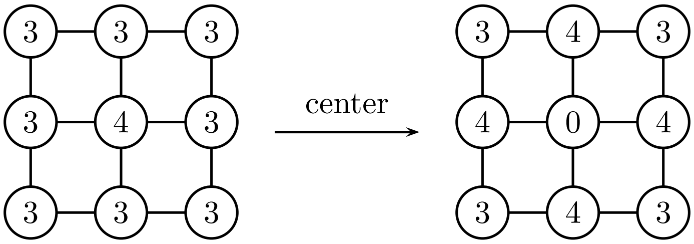
Here is a relatively new family of groups: sandpile groups. It takes a while to describe them. The basic idea is that grains of sand are being dropped on a grid, and when some capacity is reached, the grains start spilling off.

We start with a non-empty graph consisting of finitely many vertices and finitely many edges between the vertices. The edges might be loops, there may be more than one edge between two vertices. We also imagine a special invisible vertex, acting as a sink, that is connected to some vertices by invisible (and possibly multiple) edges. We require that from each vertex it is possible to travel via the edges to get to the sink. (This will ensure ... YOU'll figure out what this ensures!)

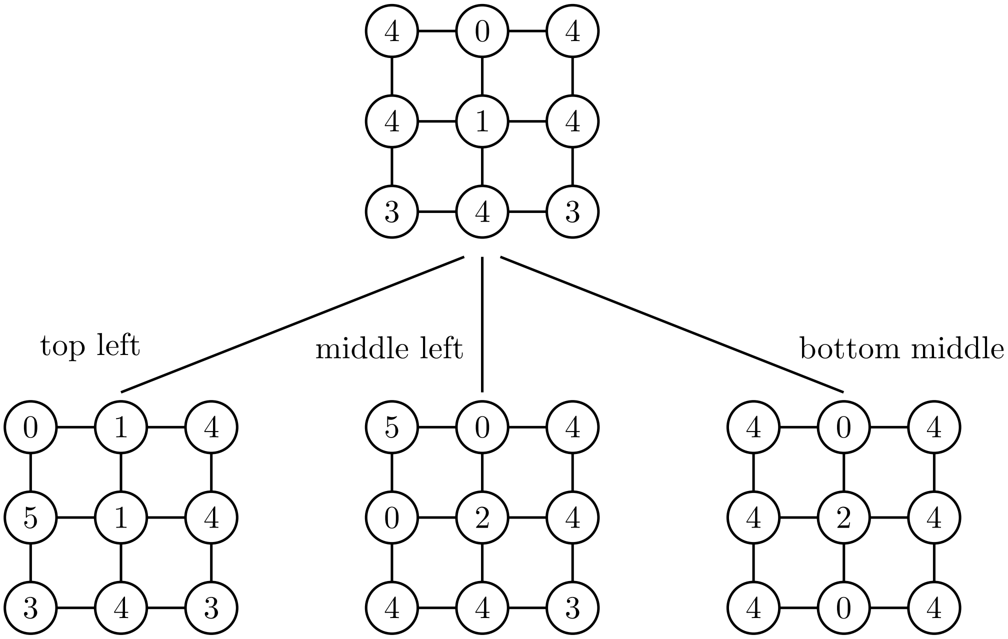
Since each vertex is at least indirectly connected to the sink, the number of edges at each vertex is positive. For a vertex v, the degree $|v|$ denotes the number of edges adjacent to v. Each vertex v can hold exactly $|v|-1$ grains on it. If there are more grains on it, then the grains start spilling off to its neighbors and to the sink, meaning that an equal number of grains spills along each of the incident edges to its neighbors until fewer than $|v|$ grains of sand lie on v. Those grains that move to the sink stay there, but those that topple to the visible vertices may cause overloads on its neighbors, causing spills from there, etc. At any time, there are only finitely many grains of sand around.

We will denote configurations with Greek letters: a configuration α has α_v grains of sand at the vertex v. We will let $+$ denote vertexwise addition of grains of sand on configurations.

We are not yet ready to discuss the definition of the sandpile groups. We first work out an example of this set-up. We start with a 3×3 grid (nine vertices, nine edges). We make each $|v| = 4$, with the sink and some edges invisible (can you picture those?). We load the grid with with 3 grains of sand on each vertex, and then we drop an additional grain on the center point, and start the toppling:



By the rotational and reflection symmetries, it does not matter which of the four vertices filled with 4 grains of sand we start toppling next, as long as the stable configuration obtained from toppling one of them has the same symmetries. We topple the top center vertex, after which we have three options, up to the left-right symmetry:



From each of the sandgrain configurations above there are several ways to proceed. We need to keep going until no more toppling is needed. If the sandpile groups are to make any sense, any order of toppling should produce the same result. This is what we prove next:

Lemma 26.1 *Use the general graph set-up as above. From any loading of the graph with grains of sand, the toppling produces a unique stable configuration in finitely many steps. The stable configuration and the number of topplings do not depend on the order of the topplings.*

Proof. If there is no stable configuration, then some vertex v would have to be overfull infinitely many times, which means that it would spill to its neighbours (adjacent vertices) infinitely many times, which means that each of its neighbours would spill to its neighbours infinitely many times, etc., which means, by the assumption on how the sink is connected, that the sink would receive infinitely many grains of sand. But we only allow finitely many grains of sand, so a stable configuration is always reached in finitely many steps.

We start with some initial configuration. If at most one vertex $v \in V$ has too many grains of sand on it, there is no ambiguity on how to proceed. Now suppose that v and w in V both have too many grains of sand on them. We want to prove that any sequence of topplings that starts with toppling from the vertex v i grains of sand along each adjacent edge ends in the same stable configuration as any sequence of topplings that starts with toppling from the vertex w j grains of sand along each adjacent edge. We use notation $v_1, v_2, \ldots, v_r$ for the sequence of topplings in which in the jth step, we topple from the vertex v_j exactly 1 grain of sand along each edge adjacent to v_j. It is easy to see that if after the $(k-1)$such that toppling in the sequence, the vertices v_k and v_{k+1} are both overfull, then the sequence $v_1, v_2, \ldots, v_r$ and the sequence obtained from this one by switching v_k and v_{k+1} give the same configuration.

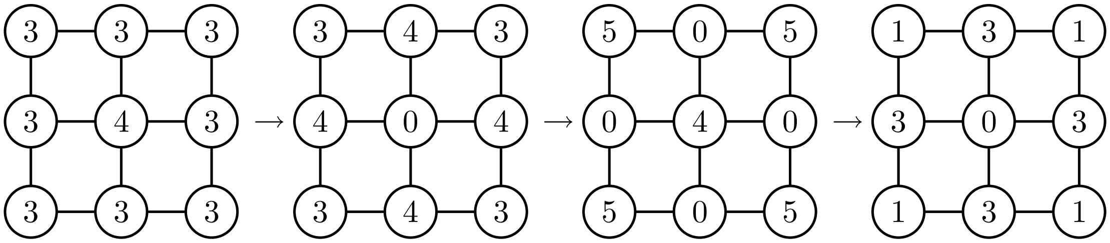
Now let the sequences $v_1, v_2, \ldots, v_r$ and $w_1, \ldots, w_s$ produce stable configurations, and let $v_1 = v_2 = \cdots = v_i = v$, $w_1 = w_2 = \cdots = w_j = w$. Since w is overfull, there exists $k \in \{i+1, \ldots, r\}$ such that $w = v_k$. Let k be the smallest such integer. By repeated use of the end of the previous paragraph, $v_1, \ldots, v_r$ gives the same configuration as $w = v_k, v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_r$. This last sequence starts with the same toppling as $w_1, \ldots, w_s$, and then we only have to compare the shorter sequences of topplings imposed on the configuration obtained after applying the toppling from the vertex w. □

This proves that toppling the grains of sand on the graph in the prescribed manner is a well-defined operation. A slight modification shows that first toppling, then throwing extra grains of sand on top, and then toppling again, produces the same final configuration as if first throwing in the extra grains of sand on top and then toppling everything. This essentially says that the operations of toppling and (adding grains + toppling) is an associative operation on any configuration.

Definition 26.2 *Given a finite graph with a sink as above. The* **full configuration** *of the graph is the loading of each vertex by $|v| - 1$ grains of sand. An* **overfull configuration** *is a loading the full configuration by a finite number of extra grains of sand. Let G be the set of all stable configurations (for this graph/sink) obtained from toppling each overfull configuration. This will be the* **sandpile group**. *(We haven't yet defined the group operation.)*

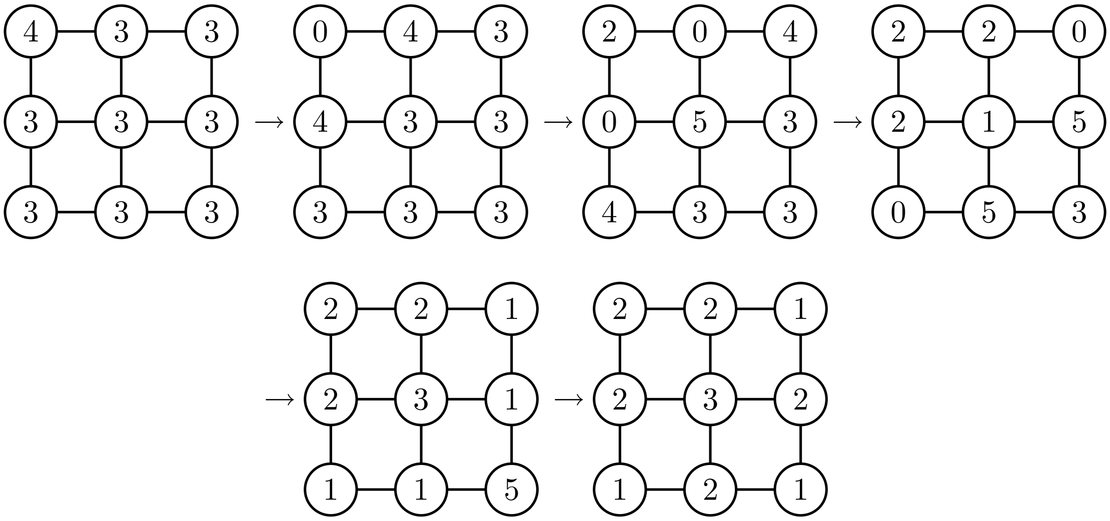
Clearly G is a finite set, with at most $\Pi_v |v|$ elements, as v varies over the vertices of the graph. This is a very rough estimate, however. Note that the constant zero loading is not allowed in G if the graph has at least one edge, in fact, the configuration having an edge between two vertices with zero grains of sand on them is not a possible stable configuration. Namely, one of the vertices adjacent to any edge (even loop) must have at least one other edge, either connecting it to another vertex or to the sink, so that in the full configuration, one of the two vertices will have a positive number of grains of sand on it. When the graph is overfull and we start toppling it, if one of the two vertices topples to be left with no grains of sand, at the same time its adjacent neighbor receives at least one grain of sand, proving the claim.

First we find some elements of G when the graph is the 3×3 grid as above. In particular, we want to finish finding the stable configuration when the graph is overloaded by one grain of sand in the center. By the lemma, we can topple overfull vertices one at a time, or all at the same time, so for speed we proceed with all overfull vertices at the same time.



This is one possible stable configuration, so an element of G. By dropping extra allowable grains of sand on the center or corner positions, this gives us actually $4 \cdot 3^4$ elements of G (4 possibilities for the center, 3 for each of the corners).

We obtain a new stable configuration when we start with the overfull configuration with an extra grain of sand on the top left corner:



We now define the sandpile group operation.

Definition 26.3 *Let G be the sandpile group, i.e., G is the set of all stable configurations obtained by toppling all the overfull configurations. We will typically write the configurations in G by lower case Roman letters. If $a, b \in G$, define $a * b$ to be the unique stable configuration obtained from the configuration $a + b$.*

Since a and b are both obtained from overfull loadings of the graph, and since toppling and (adding grains+toppling) commute, $a * b$ is also obtained from an overfull loading of the graph. Thus $*$ is indeed a binary operation. Furthermore, it is easy to see that $*$ is associative and commutative.

We will prove that $(G, *)$ is a group.

Lemma 26.4 *Let $a, b \in G$. Then there exists $c \in G$ such that $a * c = b$.*

Proof. Let β be the overfull configuration that yields b under toppling. Let t be the stable configuration obtained from 2β. The configuration $\gamma = 3\beta - t - a$ is overfull. Let c be the stable configuration obtained from γ. Then $a * c$ is obtained by toppling $a + \gamma$, which is the same as toppling $3\beta - t = \beta + (2\beta - t)$, which is the same as toppling β, which gives b. □

Warning: In the proof above, observe that $2\beta - t$ does not topple to the constant 0 configuration (in general). Why does $3\beta - t$ topple to v?

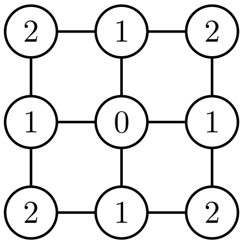
Theorem 26.5 $(G, *)$ *is a commutative group.*

Proof. Since the graph is non-empty, so is G. We already know that $*$ is commutative and associative. Let $a \in G$. By the lemma, there exists $e \in G$ such that $a * e = a$. Also, for any $b \in G$, there exists $c \in G$ such that $a * c = b$. Then

$$b * e = (a * c) * e = (c * a) * e = c * (a * e) = c * a = a * c = b.$$

Thus G contains an identity element e. By the lemma, for each $a \in G$ there exists $b \in G$ such that $a * b = b * a = e$. Thus G contains inverses. □

It is an open question to describe the identity when the graph is an $n \times n$ grid and each vertex has degree 4 (so the four corner vertices have double edges to the sink, and the other $4(n-2)$ side vertices have a single edge to the sink). For each n, the proof of the lemma and the theorem above tell us how to construct the identity. For example, the identity for the 3×3 grid is obtained from the lemma (and the theorem) say by using $a = b$ the stable configuration obtained by toppling the overfull configuration with one extra grain of sand in the center (see page 43). We skip the details, but you can verify that the identity of the sandpile group on this graph is



Sandpile groups are fascinating new groups, and much remains to be proved about them. But it is fun simply just doing the toppling! For some mesmerizing concrete sandpile topplings in action, check out `http://www.cmth.bnl.gov/~maslov/Sandpile.htm`.

27 Rings

Definition 27.1 *A* **ring** *R is a non-empty set with two binary operations, addition, denoted $+$, and multiplication, denoted $\cdot$ or with $\cdot$ omitted, such that the following hold:*
(1) R is a commutative group under $+$; the additive identity is denoted 0.
(2) Multiplication is associative.
(3) Left and right distributive properties hold.

Examples 27.2 $\mathbb{Z}$, $\frac{\mathbb{Z}}{n\mathbb{Z}}$ are rings. The set of all even integers is a ring. Every field is a ring. The set of all $n \times n$ matrices with entries in a ring R is a ring, denoted $M_n(R)$. For any set X, the set of all functions $f : X \to \mathbb{R}$ is a ring. If X has a topology, then the set of all continuous functions $f : X \to \mathbb{R}$ is a ring.

The set of all polynomials in variables in a set T and with coefficients in a ring R form a ring, denoted $R[X : X \in T]$. Recall that a polynomial is an R-linear combination of (only finitely many) monomials in the variables. If T is finite or countable, say $T = \{x_1, \ldots, x_d\}$ or $T = \{x_1, x_2, \ldots\}$, we also write $R[x_1, \ldots, x_d]$ or $R[x_1, x_2, \ldots]$ respectively, for the polynomial ring.

The set of all formal power series in variables in a set T and with coefficients in a ring R form a ring, denoted $R[[X : X \in T]]$. By definition, the formal power series ring has only finitely many terms of any degree (a power series is the limit of elements in the corresponding polynomial ring, with terms in the sequence differing eventually only in higher and higher degrees)! The set of all power series in variables in a set T and with coefficients in $\mathbb{C}$ that converge near 0 form a ring, denoted $\mathbb{C}\{X : X \in T\}$.

Definition 27.3 *A ring R is called* **commutative** *if for all $a, b \in R$, $ab = ba$. An identity of R is an element of R, usually denoted 1, such that for all $a \in R$, $1 \cdot a = a = a \cdot 1$. A ring need not contain an identity. When it does, we call $a \in R$ a* **unit** *if there exists $b \in R$ such that $a \cdot b = b \cdot a = 1$.*

The units of $\mathbb{Z}$ are 1 and -1, and no other. The units in a field are all the non-zero numbers.

Remark 27.4 Let R be a ring.
(1) For all $a \in R$, $0 \cdot a = a \cdot 0 = 0$.
(2) For all $a, b \in R$, $a(-b) = (-a)b = -(ab)$.
(3) For all $a, b \in R$, $(-a)(-b) = ab$.
(4) For all $a, b, c \in R$, $a(b - c) = ab - ac$ and $(b - c)a = ba - ca$.
(5) If R has 1, then $(-1)a = -a$ for all $a \in R$.
(6) 1, if it exists, is unique.
(7) If a is a unit, then the element b such that $ba = ab = 1$ is uniquely determined, it is denoted a^{-1}, and is called the **(multiplicative) inverse** of a. In that case, $(a^{-1})^n$ is the multiplicative inverse of a^n.

More constructions of new rings: If R is a subring of S, R has identity, and T is a subset of R, then $R[T]$, or $R[s : s \in T]$, is the smallest subring of S that contains R and T. It is easy to verify that $R[s : s \in T]$ equals the set of all polynomials in the elements of T, finitely many at a time, with coefficients in R. For example, $\mathbb{Z}[\sqrt{2}] = \{a_0 + a_1\sqrt{2} + a_2(\sqrt{2})^2 + \cdots + a_n(\sqrt{2})^n : n \in \mathbb{Z}_{\geq 0}, a_i \in \mathbb{Z}\} = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$. More examples:

$\mathbb{Z}[\sqrt{2}, \sqrt{3}]$ (not equal to $\mathbb{Z}[\sqrt{6}]$), $\mathbb{Z}[\sqrt{3}, \sqrt[3]{3}]$, $\mathbb{Z}[\sqrt{3}, \sqrt[3]{3}, \sqrt[4]{3}, \sqrt[5]{3}, \ldots]$, $\mathbb{Z}[\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \ldots] = \mathbb{Q}$, etc.

Exercise 27.5 Prove that the set $\mathbb{Z}[i] = \{a + bi : a, b, \in \mathbb{Z}\}$ is a subring of $\mathbb{C}$. The ring $\mathbb{Z}[i]$ is called the ring of **Gaussian integers**. Find, with proof, all the units in $\mathbb{Z}[i]$.

Example 27.6 Find all the units in $\mathbb{Z}[\sqrt{2}]$. Two of the units are 1 and $1 + \sqrt{2}$. Let $a + b\sqrt{2}$ be a unit in $\mathbb{Z}[\sqrt{2}]$, where a, b are integers. We want to find all a, b. Then there exist $c, d \in \mathbb{Z}$ such that $(a + b\sqrt{2})(c + d\sqrt{2}) = 1$. This means that $ac + 2bd = 1$ and $ad + bc = 0$. Necessarily then also $(a - b\sqrt{2})(c - d\sqrt{2}) = 1$, and multiplying the four terms gives us $(a^2 - 2b^2)(c^2 - 2d^2) = 1$. We conclude that $a^2 - 2b^2 = \pm 1$, and in particular, $c + d\sqrt{2}$ is $\pm(a - b\sqrt{2})$. By possibly multiplying through by -1, without loss of generality $a + b\sqrt{2} > 0$. Let n be the largest integer (positive or negative or zero) such that $a + b\sqrt{2} \geq (1 + \sqrt{2})^n$. Then $e + f\sqrt{2} = (a + b\sqrt{2})(1 + \sqrt{2})^{-n}$ is a unit in $\mathbb{Z}[\sqrt{2}]$ that is on the real interval $[1, 1 + \sqrt{2})$: since $(e + f\sqrt{2})(e - f\sqrt{2}) = e^2 - 2f^2 = \pm 1$, then $\sqrt{2} - 1 = \frac{1}{1+\sqrt{2}} < e - f\sqrt{2} \leq 1$. Hence by adding the inequalities, $\sqrt{2} < 2e < 2 + \sqrt{2}$, which forces $e = 1$, whence $1 + f\sqrt{2} \in [1, 1 + \sqrt{2})$ forces $f = 0$, so that $a + b\sqrt{2} = (1 + \sqrt{2})^n$. We just proved that all units in $\mathbb{Z}[\sqrt{2}]$ are, up to sign, powers of $1 + \sqrt{2}$.

Example 27.7 Here is an unusual ring whose glimpses you might have seen in Math 212. Let A be an open subset of $\mathbb{R}^n$, and let F be the set of all differential forms on A. In general, F is not closed under addition (what is a sum of a 1 form with a 0 form?), so we enlarge F to a set R of all (formal) sums of elements of F. Now R is closed under addition. The second binary operation on R is the wedge product: it is associative, distributes over addition, and makes R into a ring, called the **exterior algebra** of differential forms on A. This is not a commutative ring, but it does have identity: the identically-1 0-form.

Exercise 27.8 Let R be a ring with $1 \neq 0$, and let U be the set of all units in R. Prove that U is a group.

Exercise 27.9 Prove that the group of all units in $\mathbb{Z}[\sqrt{2}]$ is isomorphic to $\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$.

28 Some unsurprising definitions

Definition 28.1 Subring. *Subring test.* **Ring homomorphism** *(if the two rings have identities, we typically require that the identity maps to identity).* **Ring isomorphism**. **Direct sum of rings**.

If R is a subring of S, then the inclusion $R \subseteq S$ is a ring homomorphism!

Proposition 28.2 *Let R, S be commutative rings with identity and let $\varphi : R \to S$ be a ring homomorphism with $\varphi(1) = 1$. Let $X_1, \ldots, X_n$ be variables over R, and $s_1, \ldots, s_n \in S$. Then there exists a unique ring homomorphism*

$$\psi : R[X_1, \ldots, X_n] \to S$$

such that $\psi|_R = \varphi$ and $\psi(X_i) = s_i$.

Proof. Every element f in R has a unique representation as a finite sum $f = \sum_\nu r_\nu X_1^{\nu_1} \cdots X_n^{\nu_n}$. Define $\psi(f) = \sum_\nu \varphi(r_\nu) s_1^{\nu_1} \cdots s_n^{\nu_n}$. This clearly acts as expected, and it is a ring homomorphism, partially by the uniqueness of the expressions of f as polynomials in the X_i. □

Exercise 28.3 Let R and S be rings.
(1) Prove that componentwise addition and multiplication make $R \oplus S$ into a ring.
(2) Prove that $R \oplus S$ is commutative if and only if R and S are both commutative.
(3) Prove that $R \oplus S$ has a multiplicative identity if and only if R and S have it.

Exercise 28.4 Let R and S be rings. Prove that R is a (ring-) homomorphic image of $R \oplus S$, i.e., that there exists a surjective ring homomorphism $R \oplus S \to R$.

Exercise 28.5 Prove that there is a ring homomorphism from $\mathbb{Z}$ to any ring.

29 Something new: ideals

Definition 29.1 *A non-empty subset I of a ring R is a* **left (resp. right) ideal** *if I is a group under addition and if for all $i \in i$ and all $r \in R$, $ri \in I$ (resp. $ir \in I$). An ideal is a subset that is a right and a left ideal.*

Remark 29.2 The **kernel of a ring homomorphism**, i.e., the set of all elements in a ring that map to 0, is an ideal.

The kernel of the ring homomorphism $\mathbb{Z} \to \mathbb{Q}$ is 0.

Definition 29.3 *Let S be a subset of a ring R. The ideal of R generated by S is the smallest ideal in R that contains S. It is denoted $\langle S \rangle$, or $(s : s \in S)$. If $S = \{s_1, \ldots, s_k\}$, we also write such an ideal as $(s_1, \ldots, s_k)$. If $k = 1$, we also write $s_1 R$ for (s_1). The empty set generates the ideal* $(0) = \{0\}$.

Note that the ideal $(s_1, \ldots, s_k)$ consists of all elements of the form $\sum_i r_i s_i$, as r_i vary over elements of R.

Example 29.4 Let $R = \mathbb{Z}$. Prove that $(2, 3) = (1)$, or more generally, that $(n_1, \ldots, n_k) = (\gcd(n_1, \ldots, n_k))$.

Exercise 29.5 Prove that every ideal in $\mathbb{Z}$ is generated by (at most) one element.

Example 29.6 Let X, Y, Z, T be variables over $\mathbb{Q}$. By Proposition 28.2, there exists a unique ring homomorphism $\psi : \mathbb{Q}[X, Y, Z] \to \mathbb{Q}[T]$ such that $\psi|_{\mathbb{Q}}$ is identity, $\psi(X) = T$, $\psi(Y) = T^2$, $\psi(Z) = T^3$. Find the kernel of ψ. Certainly $Y - X^2, Z - X^3$ are both in the kernel. We prove next that $(Y - X^2, Z - X^3) = \ker \psi$. Let $f \in \ker \psi$. Write $f = f_0 + f_1 Y + f_2 Z$, where f_0 is a polynomial in X and $f_1, f_2 \in \mathbb{Q}[X, Y, Z]$. Then $f = f_0 + f_1 X^2 + f_2 X^3 + f_1(Y - X^2) + f_2(Z - X^3)$, so $g = f_0 + f_1 X^2 + f_2 X^3 \in \ker \psi$. Note that the Y- and the Z-degrees of g are strictly smaller than the corresponding degrees of f. If we can prove that $g \in (Y - X^2, Z - X^3)$, then we will have proved that $f \in (Y - X^2, Z - X^3)$. Thus by induction it suffices to prove that if $f \in (\ker \psi) \cap \mathbb{Q}[X]$, then $f = 0$. But this is easy! Say why!

Example 29.7 Let X, Y, Z, T be variables over $\mathbb{Q}$. By Proposition 28.2, there exists a unique ring homomorphism $\psi : \mathbb{Q}[X, Y, Z] \to \mathbb{Q}[T]$ such that $\psi|_{\mathbb{Q}}$ is identity, $\psi(X) = T^3$, $\psi(Y) = T^4$, $\psi(Z) = T^5$. Find the kernel of ψ. This example is harder than the previous example! We will prove that $\ker \psi = (X^3 - YZ, Y^2 - XZ, Z^2 - X^2Y)$. One inclusion is easy. For the other inclusion, let $f \in \ker \psi$. We will use a trick similar to the one in the previous example. By possibly adding/subtracting the known elements of the kernel, namely multiples of $X^3 - YZ, Y^2 - XZ$, and $Z^2 - X^2Y$, without loss of generality the X-degree of f is at most 2, the Y-degree of f is at most 1, and the Z-degree of f is at most 1. (Verify/justify!) So $f = a_0 + a_1X + a_2X^2 + a_3XY + a_4X^2Y + a_5XZ + a_6X^2Z + a_7XYZ + a_8X^2YZ$, for some $a_i \in \mathbb{Q}$. Then $0 = \psi(f) = a_0 + a_1T^3 + a_2T^6 + a_3T^7 + a_4T^{10} + a_5T^8 + a_6T^{11} + a_7T^{12} + a_8T^{15}$, which forces all $a_i = 0$.

Exercise 29.8 Let n be a positive integer. Let X, Y be variables over a field F. Prove that the ideal $(X^n, X^{n-1}Y, X^{n-2}Y^2, \ldots, XY^{n-1}, Y^n)$ in $F[X, Y]$ is minimally generated by $n + 1$ elements, i.e., it is not possible to find a generating set of the ideal with fewer elements.

Exercise 29.9 Let R be a commutative ring with 1 and let I be an ideal in R. Let S and T be generating sets of I. Assume that T is finite and S is infinite. Prove that there exists a finite subset $S_0 \subseteq S$ that generates I.

Exercise 29.10 Let F be a field and X, Y variables over F. Prove that the ideal (X, Y) in the ring $F[X, Y]$ has the property that there are no ideals strictly between it and the whole ring.

Exercise 29.11 Let I and J be ideals in a ring R.
(1) Define $I + J$ to be the set of all elements of the form $i + j$, where $i \in I$ and $j \in J$. Prove that $I + J$ is an ideal.
(2) Define $I \cdot J = IJ$ to be the ideal generated by all elements of the form $i \cdot j$, where $i \in I$ and $j \in J$. Show by example that $\{ij : i \in I, j \in J\}$ need not be an ideal. (Hint: start with non-principal ideals, say in $\mathbb{Z}[X]$ or $\mathbb{Q}[X, Y]$.)
(3) Discuss/give some necessary and sufficient conditions for $I \cdot J$ to be a prime ideal.

Exercise 29.12 Let I and J be ideals in a ring R. Prove that $I \cap J$ is an ideal. Give necessary and sufficient conditions for $I \cap J$ to be a prime ideal.

Exercise 29.13 Let I and J be ideals in a ring R and let X be a variable over R. By $IR[X]$ we denote the ideal in $R[X]$ generated by the elements of I (in R). Prove that

$$I \cap J = ((X)(IR[X]) + (1 - X)(JR[X])) \cap R.$$

Exercise 29.14 Let R be the ring of all functions $f : \mathbb{R} \to \mathbb{R}$. List at least four different ideals in R. Find at least one maximal ideal in R. Find a zero-divisor.

30 More that's new: characteristic of a ring

Definition 30.1 *Let R be a ring. Its* **characteristic** *is the least positive integer n such that $nx = 0$ for all $x \in R$, and if no such integer exists, then the characteristic of R is 0.*

The characteristic of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ is 0. The characteristic of $\frac{\mathbb{Z}}{n\mathbb{Z}}$ is n.

Theorem 30.2 (Freshman's dream) *Let R be a ring of prime characteristic p. Then for all $x, y \in R$, and all $e \in \mathbb{N}$, $(x+y)^{p^e} = x^{p^e} + y^{p^e}$.*

Proof. This is trivial if $e = 0$. By induction on e it suffices to prove it for $e = 1$, i.e., it suffices to prove that for all $x, y \in R$, $(x+y)^p = x^p + y^p$. By the binomial formula, $(x+y)^p = \sum_{i=0}^{p} \binom{p}{i} x^i y^{p-i}$. If $i \in \{1, \ldots, p-1\}$, then $\binom{p}{i}$ has p in the numerator but not in the denominator, so that it is an integer multiple of p, whence $\binom{p}{i} x^i y^{p-i} = 0$. □

Theorem 30.3 *Let R be a commutative ring with identity. Let n be its characteristic. Then $\frac{\mathbb{Z}}{n\mathbb{Z}}$ is a subring of R, in a natural way.*

Proof. Easy. □

31 Quotient (or factor) rings

Theorem 31.1 *Let R be a ring, I an ideal of R. By R/I we denote the set of all cosets of I (R is a group under $+$). Then R/I is a ring. If R has identity, so does R/I. If R is commutative, so is R/I.*

Proof. We already know that R/I is a group under $+$. Let $a, b \in R$. Define $(a+I)(b+I) = ab + I$. This is well-defined, associative, it distributes over $+$. □

Theorem 31.2 (First Isomorphism Theorem) *Let $\varphi : R \to S$ be a ring homomorphism with kernel I. Then $R/I \cong \mathrm{Im}\varphi$.*

Proof. ... □

Example 31.3 For any integer n, $n\mathbb{Z}$ is an ideal in $\mathbb{Z}$. Then $\mathbb{Z}/n\mathbb{Z}$ is a ring.

Exercise 31.4 Prove that $\mathbb{C} \cong \mathbb{R}[X]/(X^2+1)$, where X is a variable over $\mathbb{R}$.

Exercise 31.5 Prove that $\mathbb{C} \ncong \mathbb{R}[X]/(X^2-1)$, where X is a variable over $\mathbb{R}$.

Exercise 31.6 Prove that $\mathbb{Z}[i] \cong \mathbb{Z}[X]/(X^2+1)$, where X is a variable over $\mathbb{Z}$.

Example 31.7 Let R be the ring of all polynomials in variable T with real coefficients in which the coefficients of T, T^2 and T^3 are zero. Why is this a ring? Why is it a subring of $\mathbb{R}[T]$. Find I such that $R \cong \mathbb{R}[X, Y, Z, U]/I$.

Example 31.8 Let R be the ring of all polynomials in variables X, Y, Z with real coefficients, restricted to the set of all points $(x, y, z) \in \mathbb{R}^3$ for which $xy = z^3$. What do we mean by that? Note that the polynomials XY and Z^3 do the same thing on all the allowed points. Similarly, the polynomials $XY^2 - X^2$ is the same as $YZ^3 - X^2$. Verify: $R \cong \mathbb{R}[X, Y, Z]/(XY - Z^3)$.

Example 31.9 Read the previous example. Let R be the ring of all polynomials in variables X, Y, Z with real coefficients, restricted to the set of all points $(x, y, z) \in \mathbb{R}^3$ for which $x^4y = x^7$. Verify: $R \cong \mathbb{R}[X, Y, Z]/(XY - X^4)$.

Example 31.10 Let R be the ring of all polynomials in variables X, Y, Z with real coefficients by identifying any two whose difference vanishes on the set of all points $\{(t^3, t^4, t^5) : t \in \mathbb{R}\}$. Verify: $R \cong \mathbb{R}[X, Y, Z]/(X^3 - YZ, Y^2 - XZ, Z^2 - X^2Y)$.

32 (Integral) Domains

Definition 32.1 *Let R be a commutative ring. A* **zero-divisor** *in R is a non-zero element a in R for which there exists a non-zero $b \in R$ such that $ab = 0$.*

Definition 32.2 *An* **integral domain***, or a* **domain***, is a commutative ring R with unity in which there are no zero-divisors.*

Examples 32.3 $\mathbb{Z}$, fields, $\mathbb{Z}[i]$ are domains. If n is a prime integer, $\frac{\mathbb{Z}}{n\mathbb{Z}}$ is a domain; if n is a product of two non-unit, non-zero integers, then $\frac{\mathbb{Z}}{n\mathbb{Z}}$ is not a domain. The ring $M_2(\mathbb{R})$ is not a domain.

Example 32.4 If R is a domain, so is $R[X]$. Namely, if $f = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ and $g = b_0 + b_1X + b_2X^2 + \cdots + b_mX^m$ are two non-zero polynomials, then without loss of generality $a_n \neq 0$ and $b_m \neq 0$. The product fg has degree at most $n + m$, and in fact the coefficient in fg of X^{n+m} is a_nb_m. Since R is a domain, $a_nb_m \neq 0$, whence $fg \neq 0$. This proves that $R[X]$ is a domain.

Exercise 32.5 Modify the argument above to prove that $R[[X]]$ is also a domain. (Obviously you won't be able to trace what happens to the products of the coefficients of highest powers of X.)

Remark 32.6 If R is a domain, then the units in a polynomial ring $R[X]$ over R are the units of R! Say why!

Proposition 32.7 *Let a, b, c be elements of a ring R, and assume that a is not a zero-divisor. Then $ab = ac$ implies that $b = c$.*

Proof. Easy. □

Theorem 32.8 *A finite domain is a field.*

Proof. Let R be a domain that has only finitely many elements. Let $a \in R \setminus \{0\}$. By Proposition 32.7, multiplication by a is injective. By the counting argument, there must be $b \in R$ such that $ab = 1$. □

The finiteness assumption is definitely needed. Try $R = \mathbb{Q}[X]$.

Here are some other fields: $\mathbb{Q}[i]$, $\mathbb{Q}[\sqrt{2}]$, $\mathbb{Q}[2^{1/3}]$, etc. Why are these fields?

Example 32.9 Let $X^2 + X + 1 \in \frac{\mathbb{Z}}{2\mathbb{Z}}[X]$. Prove that $X^2 + X + 1$ has no roots in $\frac{\mathbb{Z}}{2\mathbb{Z}}$. Let $F = \frac{\mathbb{Z}}{2\mathbb{Z}}[X]/(X^2 + X + 1)$. Prove that F is a field with 4 elements.

Exercise 32.10 Prove that $X^3 + X + 1$ has no roots in $\frac{\mathbb{Z}}{2\mathbb{Z}}$. Let $F = \frac{\mathbb{Z}}{2\mathbb{Z}}[X]/(X^3 + X + 1)$. Prove that F is a field with 8 elements.

Exercise 32.11 Prove that there is no field with 6 elements.

Exercise 32.12 Let R and S be (non-zero) rings. Prove that $R \oplus S$ is not a domain.

Theorem 32.13 *The characteristic of a domain is* 0 *or prime.*

Proof. Suppose that the characteristic of a domain R is not 0. Then there exists a positive integer n such that $n \cdot 1 = 0$. Since R is a domain, n cannot be a composite number. □

Exercise 32.14 (In this exercise, you may want to keep in mind the process of building the fraction field $\mathbb{Q}$ out of $\mathbb{Z}$. In fact, the constructed F below is called **the field of fractions of** R.)

Let R be a commutative domain with $1 \neq 0$, and let $S = \{(a, b) : a, b \in R, b \neq 0\}$. For $(a, b), (c, d) \in S$, define $(a, b) \sim (c, d)$ if $ad = bc$.

(i) Prove that $\sim$ is an equivalence relation on S.

(ii) Let F be the set of all equivalence classes in S. (You may want to suggestively write the equivalence class of (a, b) in the form $\frac{a}{b}$. Justify.) Define binary operations $+, \cdot$ on F as

$$(a, b) + (c, d) = (ad + bc, bd), (a, b) \cdot (c, d) = (ac, bd).$$

Prove that $+, \cdot$ are well-defined and that they make F into a field.

(iii) Let $i : R \to F$ be the function $i(r) = (r, 1)$. Prove that i is an injective ring homomorphism.

(iv) Prove that for every $f \in F$ there exists a non-zero $r \in R$ such that $i(r)f \in i(R)$.

33 Prime ideals and maximal ideals

Definition 33.1 *A* **prime ideal** *P in a commutative ring R is a* **proper** *(i.e., $P \neq R$) ideal such that $a, b \in R$ and $ab \in P$ implies that $a \in P$ or $b \in P$.*

Example 33.2 The only prime ideals in $\mathbb{Z}$ are (0) and (p), where p varies over the prime integers. (Hence the term prime ideal!)

Example 33.3 If R is a field, the only prime ideal is (0).

Example 33.4 If X is a variable over $\mathbb{R}$, then (X^2+1), $X-r$, as r varies over $\mathbb{R}$ are all prime ideals.

Example 33.5 Let $R = \mathbb{Q}[X]/(X^2)$. Let P be a prime ideal in R. Certainly $0 = X^2 \in P$, so by the definition of prime ideals, $X \in P$. Now verify that XR is a prime ideal.

Example 33.6 Skim through the table below for what ideals in various rings are prime; some verifications are harder than others, and we will verify one later:

ideal	$\mathbb{Z}[X]$	$\mathbb{Q}[X]$	$\mathbb{R}[X]$	$\mathbb{C}[X]$
(2)	Yes	No	No	No
(X)	Yes	Yes	Yes	Yes
$(X+2)$	Yes	Yes	Yes	Yes
$(X,2)$	Yes	No	No	No
$(X+2i)$	N/A	N/A	N/A	Yes
(X^2+2)	Yes	Yes	Yes	No
$(X+2,X^2+1)$	Yes	No	No	No

Proposition 33.7 *Let R be a commutative ring with unity and let I be an ideal of R. Then I is a prime ideal if and only if R/I is a domain.*

Proof. Do it. □

We already know that $\mathbb{Z}/5\mathbb{Z}$ is a domain, even a field, so by the proposition $5\mathbb{Z}$ is a prime ideal. Furthermore, if we can prove that $\mathbb{Z}[X]/(X+2,X^2+1) \cong \mathbb{Z}/5\mathbb{Z}$, then we'll have established that $(X+2,X^2+1)$ is a prime ideal in $\mathbb{Z}[X]$. Note: every element f in $\mathbb{Z}[X]$ can be written as $f = q(X+2)+r$ for some $q \in \mathbb{Z}[X]$ and some $r \in \mathbb{Z}$ (by the division algorithm for polynomials), so every element in $\mathbb{Z}[X]/(X+2,X^2+1)$ can be represented by an integer r. Note that $5 = (X^2+1) - X(X+2) + 2(X+2)$, so all multiples of 5 are in $(X+2,X^2+1)$. If an integer r is in $(X+2,X^2+1)$, then there exist $a,b \in \mathbb{Z}[X]$ such that $r = a(X+2)+b(X^2+1)$. Now plug in $X=-2$ to get that r is an integer multiple of 5. This proves that $\mathbb{Z}[X]/(X+2,X^2+1) \cong \mathbb{Z}/5\mathbb{Z}$.

Definition 33.8 *A* **maximal ideal** *M in a ring R is a proper ideal such that whenever I is an ideal such that $M \subseteq I$, then either $I = M$ or $I = R$.*

Example 33.9 The maximal ideals in $\mathbb{Z}$ are (p), where p varies over the prime integers.

Example 33.10 If R is a field, the only prime and the only maximal ideal is (0).

Example 33.11 If X is a variable over $\mathbb{R}$, then (X^2+1), $X-r$, as r varies over $\mathbb{R}$ are all maximal ideals.

Proposition 33.12 *Let R be a commutative ring with unity and let I be an ideal of R. Then I is a maximal ideal if and only if R/I is a field.*

Proof. Do it. □

This enables us to construct a lot of different fields, say by starting with $\mathbb{Q}$ or $\mathbb{R}$ or $\frac{\mathbb{Z}}{p\mathbb{Z}}$, adjoining some variables, finding maximal ideals in the corresponding polynomial ring, and then passing to the quotient ring modulo the maximal ideal.

Theorem 33.13 *Every maximal ideal in a commutative ring with identity is a prime ideal.*

Proof. Let M be a maximal ideal in a commutative ring R. Let $a, b \in R$, $ab \in M$. If $a \in M$, we are done. So suppose that $a \notin M$. Let Q be the smallest ideal containing a and M. Every element of Q can be written as an element in (a) plus an element in M (verify!). Then M is properly contained in Q, so that by the definition of maximal ideals, $Q = R$. Hence $1 = m + ra$ for some $m \in M$ and some $r \in R$. Then $b = mb + rab \in M + M = M$, so that $b \in M$, and we are done. □

Theorem 33.14 *Every commutative ring with identity has a maximal ideal. (Use Zorn's lemma.)*

Proof. We first **Zornify**:
(1) Let $\mathcal{S}$ be the set of all proper ideals in the ring.
(2) $\mathcal{S}$ is not empty as it contains the zero ideal.
(3) We impose a partial order on $\mathcal{S}$: if $I, J \in \mathcal{S}$, we say $I \leq J$ if $I \subseteq J$. Verify that this is a partial order.
(4) Let $I_1 \leq I_2 \leq I_3 \leq \cdots$ be a **chain** in $\mathcal{S}$. Set $I = \cup_n I_n$. Certainly I is a subset of R that contains all I_n. It is closed under addition: if $r, s \in I$, then there exist m, n such that $r \in I_m$, $s \in I_n$. Hence r, s are both in the ideal $I_{\max\{m,n\}}$, whence $r + s \in I_{\max\{m,n\}} \subset I$. Note that here we used that the ideals form a chain. One can show even more easily that for any $a \in R$ and any $r \in I$, $ar \in I$. Furthermore, $I \in \mathcal{S}$: $1 \notin I_n$ for all n, therefore $1 \notin 1$. This proves that every chain in $\mathcal{S}$ has an upper bound in $\mathcal{S}$.

These four conditions are enough for applying Zorn's lemma*: Zorn's lemma says that $\mathcal{S}$ has a maximal element, i.e., there exists $M \in \mathcal{S}$ such that for all $I \in \mathcal{S}$, either I and M are incomparable, or else $I \leq M$.

We claim that M is a maximal element in R. First of all, since $M \in \mathcal{S}$, $M \neq R$. Secondly, if I is an ideal in R such that $M \subseteq I$, then either $I \in \mathcal{S}$ or $I \notin \mathcal{S}$. In the latter case, necessarily $I = R$. In the former case, $M \subseteq I$ implies $M \leq I$, but by the definition of M, $I \leq M$, whence $I = M$. This proves that M is a maximal ideal in R. □

34 Division algorithm in polynomial rings

Theorem 34.1 (Division algorithm for F[X]) *Let F be a field and X a variable over F. Given $f, g \in F[X]$, there exist unique polynomials $q, r \in F[X]$ such that $f = qg + r$ and either $r = 0$ or $\deg r < \deg g$.*

Proof. □

Corollary 34.2 *Let F be a field, X a variable over F, $a \in F$, $f(x) \in F[X]$. Then $f(a) = 0$ if and only if $f(x)$ is a multiple of $X - a$.*

Proof. □

* Zorn's lemma is, contrary to its name, not something to prove: either you assume it for your set theory or you don't, and no contradictions arise.

Corollary 34.3 *A polynomial of degree n over a field has at most n zeros.*

Proof. □

Definition 34.4 *A* **principal ideal domain** *is a commutative domain with identity in which each ideal is generated by at most one element.*

We have proved that $\mathbb{Z}$ is a principal ideal domain.

Theorem 34.5 *Let F be a field and X a variable over F. Then the polynomial ring $F[X]$ is a principal ideal domain.*

Proof. Let I be an ideal in $F[X]$. If I is (0), then done. So suppose that I contains a non-zero element. Let g be a non-zero element in I of least possible degree (as a polynomial in X). Claim: $I = (g)$. Certainly $(g) \subseteq I$. Let $f \in I$. By the division algorithm, there exist $q, r \in F[X]$ such that $f = qg + r$ and either $r = 0$ or $\deg r < \deg g$. In any case, $r \in I$, and by the choice of g, necessarily $r = 0$, so that $f \in (g)$. □

Theorem 34.6 *Let R be a principal ideal domain. Then every increasing chain of ideals $I_1 \subseteq I_2 \subseteq \cdots$ stabilizes, meaning that there exists an integer n such that $I_n = I_{n+1} = I_{n+2} = \cdots$.*

Proof. Let $I = \cup_j I_j$. First verify that I is an ideal. Since R is a principal ideal domain, there exists $a \in R$ such that $I = (a)$. Then $a \in I_n$ for some n, whence $I = (a) \subseteq I_n \subseteq I_{n+i} \subseteq I$. □

Example 34.7 Let $f(X) = X^4 + X^3 + X^2 + X + 1 \in \frac{\mathbb{Z}}{2\mathbb{Z}}[X]$. Verify that (f) is a maximal ideal. Verify that $\frac{\mathbb{Z}}{2\mathbb{Z}}[X]/(f)$ is a field with 16 elements.

The following is an easy generalization of the proof of Theorem 34.6:

Exercise 34.8 Let R be a ring. Prove that the following are equivalent:

(i) Every ideal in R is finitely generated.

(ii) Every ascending chain of ideals in R eventually stabilizes.

Example 34.9 $\mathbb{Z}[i]$ is a principal ideal domain. Let I be a non-zero ideal in $\mathbb{Z}[i]$. Let α be a non-zero element in I whose complex norm is smallest possible (why does such α exist)? We claim that $I = (\alpha)$. Let $\beta \in I$. The complex number $\frac{\beta}{\alpha}$ can be written in the form $e + fi$ for some $e, f \in \mathbb{Q}$. Let $n, m \in \mathbb{Z}$ such that $|e - n|, |f - m| \leq 1/2$. Define $\delta = \beta - (n + mi)\alpha \in \mathbb{Z}[i]$. Then

$$\frac{\delta}{\alpha} = \frac{\beta}{\alpha} - (n + mi),$$

and both the imaginary and the real components are at most $1/2$, whence the complex norm of δ/α is strictly smaller than 1. Thus the norm of δ is strictly smaller than the norm of α, which is a contradiction.

35 Irreducibility

Definition 35.1 *Let R be a commutative domain. An* **irreducible** *element in R is a non-zero non-unit r such that whenever $r = ab$ for some $a, b \in R$, then either a or b is a unit in R.*

If R is a commutative domain, an **irreducible polynomial** *in X over R is a non-zero non-unit polynomial $f \in R[X]$ such that whenever $f = gh$ for some $g, h \in R[X]$, then either g or h is a unit in $R[X]$. A non-zero non-unit polynomial f is called* **reducible** *over R if $f = gh$ for some non-units g, h in $R[X]$.*

The prime numbers are irreducible in $\mathbb{Z}$. If r is irreducible in R and X is a variable over R, then r is irreducible in $R[X]$ and $R[[X]]$.

The polynomial $2X^2 + 6$ is reducible over $\mathbb{Z}$ but irreducible over $\mathbb{Q}$. All monic polynomials of degree 1 are irreducible.

Note that irreducible polynomials over R are irreducible elements in $R[X]$.

Theorem 35.2 *Let R be a principal ideal domain, $r \in R$. Then r is irreducible in R if and only if (r) is a maximal ideal in R.*

Proof. $\Longrightarrow$ Let I be any proper ideal in R such that $r \in I$. Since R is a principal ideal domain, $I = (s)$ for some $s \in R$. Therefore r is a multiple of s, so $r = ts$ for some $t \in R$. Since r is irreducible, t must be a unit, so that $s \in (r)$, whence $I = (r)$, so that (r) is a maximal ideal. The other direction is similar. □

This immediately implies:

Theorem 35.3 *Let F be a field, $f \in F[X]$. Then f is irreducible over F if and only if (f) is a maximal ideal in $F[X]$.* □

Corollary 35.4 *If f is irreducible over F, then $F[X]/(f)$ is a field.* □

Theorem 35.5 *Let F be a field, $f \in F[X]$, $\deg f = 2$ or $\deg f = 3$. Then f is reducible over F if and only if f has a zero (root) in F.*

Proof. We have verified in Remark 32.6 that the only units in $F[X]$ are the non-zero elements in F. The polynomial f of degree 2 or 3 is reducible if and only if it has a factor of degree 1, i.e., a factor of the form $aX + b$ for some $a, b \in F$, $a \neq 0$. But this holds if and only if $f(-b/a) = 0$, and $-b/a \in F$. □

Note: $X^2 + 1$ is irreducible over $\mathbb{R}$ but reducible over $\mathbb{C}$. Also, $X^2 - 2$ is irreducible over $\mathbb{Q}$ but reducible over $\mathbb{R}$, or over $\mathbb{Q}[\sqrt{2}]$.

Examples 35.6 We now find some prime ideals in polynomial rings.

(1) Let $R = \mathbb{Q}[X]$. Then $(X^2 + 1), (X - 5), (X)$ are maximal, hence prime, ideals.

(2) Let $S = \mathbb{Q}[X, Y]$. Here $(X^2 + 1), (X - 5), (X)$ are non-maximal prime ideals. To see this, observe that
$$\frac{\mathbb{Q}[X, Y]}{(f(X))} \cong \frac{\mathbb{Q}[X]}{(f(X))}[Y],$$
which is a polynomial ring in variable Y over the field $\mathbb{Q}[X]/(f(X))$, hence a principal ideal domain, which proves the claim.

(3) Let $S = \mathbb{Q}[X,Y]$. Here $(X^2+1,Y), (X-5,Y-7), (X,Y)$ are maximal, hence prime, ideals. Observe that
$$\frac{\mathbb{Q}[X,Y]}{(X^2+1,Y)} \cong \frac{\mathbb{Q}[X]}{(X^2+1)}$$
which is a field. It is even easier to prove that the other two ideals are maximal ideals. $(X-5,Y-7)$ is the kernel of the ring homomorphism $\mathbb{Q}[X,Y] \to \mathbb{Q}$ with $X \mapsto 5$, $Y \mapsto 7$ (why does such a ring homomorphism exist, why is the kernel as specified?); (X,Y) is the kernel of the ring homomorphism $\mathbb{Q}[X,Y] \to \mathbb{Q}$ with $X \mapsto 0$, $Y \mapsto 0$.

Warning: (X^2+1,Y^2+1) is not a prime ideal because $(X-Y)(X+Y) = (X^2+1)-(Y^2+1)$ is in the ideal, but neither factor is. (If $X-Y \in (X^2+1,Y^2+1)$, then $X-Y = a(X^2+1)+b(Y^2+1)$ for some $a,b \in \mathbb{Q}[X,Y]$. Thus under the ring homomorphism $\mathbb{Q}[X,Y] \to \mathbb{C}$, including $\mathbb{Q}$ in $\mathbb{C}$ and sending X to i and Y to $-i$, we get a contradiction.)

In general, it is not easy to decide if an ideal is prime, or if a polynomial is irreducible.

Theorem 35.7 *Let $f \in \mathbb{Z}[X]$. If f is reducible over $\mathbb{Q}$, then it is a product of two non-constant polynomials in $\mathbb{Z}[X]$.*

Proof. Suppose that $f = gh$ for some $g,h \in \mathbb{Q}[X]$, neither of which is a constant. Let m,n be the smallest positive integers such that $mg \in \mathbb{Z}[X]$ and $nh \in \mathbb{Z}[X]$. Let d be the greatest common divisor of the coefficients of mg, and let e be the greatest common divisor of the coefficients of nh. Then $\frac{mn}{de}f = (\frac{m}{d}g)(\frac{n}{e}h)$, and $\frac{m}{d}g, \frac{n}{e}h \in \mathbb{Z}[X]$. Observe that the latter two polynomials have the property that their coefficients generate the ideal $1\mathbb{Z}$. If the coefficients of $\frac{mn}{de}f$ are all multiples of a prime integer p, then by passing modulo p, we get a product of two non-zero polynomials in $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ that equals 0, which is a contradiction by Example 32.4. Thus the coefficients of $\frac{mn}{de}f$ also generate $1\mathbb{Z}$, whence de is a multiple of mn, so that
$$f = \frac{de}{mn}\left(\frac{m}{d}g\right)\left(\frac{n}{e}h\right)$$
is a factorization over $\mathbb{Z}$. □

Theorem 35.8 (Reduction to characteristic p, goes back to Dedekind) *Let $f \in \mathbb{Z}[X]$ with $\deg f \geq 1$. Let p be a prime integer, and let $\overline{f}$ be the polynomial in $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ obtained from f by reducing all coefficients of f modulo p. If $\overline{f}$ is irreducible over $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ and $\deg \overline{f} = \deg f$, then f is irreducible over $\mathbb{Q}$.*

Proof. Suppose that $f = gh$ for some non-constant $g,h \in \mathbb{Q}[X]$. By Theorem 35.7, without loss of generality $g,h \in \mathbb{Z}[X]$. By taking the images of the coefficients modulo $p\mathbb{Z}$, we get that $\overline{f} = \overline{g}\overline{h}$ in $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$. Since $\mathbb{Z}[X]$ and $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ are domains, necessarily
$$\deg \overline{g} + \deg \overline{h} = \deg \overline{f} = \deg f = \deg g + \deg h,$$
which forces $\deg \overline{g} = \deg g$ and $\deg \overline{h} = \deg h$, so that $\overline{g}$ and $\overline{h}$ are non-constant polynomials, which is a contradiction. □

Current computer algorithms for determining irreducibility and computing irreducible components of polynomials are based on this last theorem.

Theorem 35.9 (Eisenstein's criterion) *Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_0 \in \mathbb{Z}[X]$, with $a_i \in \mathbb{Z}$. If there is a prime p such that $p \not| \; a_n$, $p|a_{n-1}, \ldots, p|a_0$, $p^2 \not| \; a_0$, then f is irreducible over $\mathbb{Q}$.*

Proof. Suppose for contradiction that f is reducible. By Theorem 35.7, there exists $g, h \in \mathbb{Z}[X]$ of positive degrees such that $f = gh$. Modulo p, $f(X)$ is $\overline{a}_nX^n$, and is nonzero. Since $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ is a domain, only the leading coefficients of g and h are not multiples of p. Since $p^2 \not| \; a_0$, the constant coefficients in g and h cannot both be multiples of p, but this is a contradiction. □

Exercise 35.10 Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_0 \in \mathbb{Z}[X]$, with $a_i \in \mathbb{Z}$. Prove that if there is a prime p such that $p \not| \; a_0$, $p|a_1, \ldots, p|a_n$, $p^2 \not| \; a_n$, then f is irreducible over $\mathbb{Q}$.

Corollary 35.11 (Irreducibility of the pth cyclotomic polynomial) *For any prime p, the pth cyclotomic polynomial*

$$\Phi_p(X) = \frac{X^p - 1}{X - 1} = X^{p-1} + X^{p-2} + \cdots + X + 1$$

is irreducible over $\mathbb{Q}$.

Proof. Note that $\Phi_p(X+1) = \frac{(X+1)^p - 1}{(X+1) - 1} = X^{p-1} + \binom{p}{1}X^{p-2} + \binom{p}{2}X^{p-3} + \cdots + \binom{p}{p-1}X$. By Eisenstein's criterion, this is irreducible, whence $\Phi_p(X)$ must be irreducible as well. □

Observe: the polynomial $X^n - 1 \in \mathbb{Q}[X]$ factors as $\prod_{k=1}^{n}(X - e^{\frac{2k\pi i}{n}})$. Recall that the set $\{e^{\frac{2k\pi i}{n}} : k = 1, \ldots, n\}$ is a multiplicative subgroup of $\mathbb{C}$, consisting of exactly n elements. By grouping these roots of unity, we get that $X^n - 1 = \prod_{d|n} g_d(X)$, where $g_d(X) = \prod(X - u)$, where u varies over those roots of unity $e^{2k\pi i/n}$ whose order in the group is exactly d (such roots are called **the primitive dth root of unity**, and are independent of n that is a multiple of d). For example,

$$X^2 - 1 = (X-1)(X+1), g_1(X) = X - 1, g_2(X) = X + 1,$$
$$X^3 - 1 = (X-1)(X - e^{2\pi i/3})(X - e^{4\pi i/3}), g_1(X) = X - 1, g_3(X) = X^2 + X + 1,$$
$$X^4 - 1 = (X-1)(X+1)(X-i)(X+i), g_1(X) = X - 1, g_2(X) = X + 1, g_4(X) = X^2 + 1.$$

At least in the examples above, $g_d(X)$ is a polynomial with coefficients in $\mathbb{Z}$ that is irreducible over $\mathbb{Q}$. In particular, if $d = p$ is a prime, clearly $g_d = \Phi_p$, and we have proved that this is irreducible. It is true in general that the g_d are in $\mathbb{Z}[X]$ and are irreducible over $\mathbb{Q}$, but we do not yet have the methods to prove it. In fact, a proof of this general fact is not done in these notes.

Exercise 35.12 Find, with proof, all irreducible factors of $X^4 + 1$ over $\mathbb{Q}$, over $\mathbb{R}$, over $\mathbb{C}$, over $\frac{\mathbb{Z}}{2\mathbb{Z}}$, and over $\frac{\mathbb{Z}}{3\mathbb{Z}}$.

Exercise 35.13 Construct a field of order 27.

36 Unique factorization domains

In $\mathbb{Z}$, we can factor 6 as $2 \cdot 3 = 3 \cdot 2 = (-2) \cdot (-3) = (-3) \cdot (-2)$. In a certain sense, all of these factorizations are the same.

Definition 36.1 *Two elements in a ring R are* **associates** *if one is a unit of R times the other.*

Thus the factorization of 6 above is unique UP TO the order of factorization AND UP TO associates.

Definition 36.2 *A domain D is a* **unique factorization domain** *if every non-zero non-unit element of D can be written as a product of irreducible elements of D, and this factorization is unique up to the order in which the factors are written and up to associates.*

Theorem 36.3 *A principal ideal domain is a unique factorization domain.*

Proof. Let R be a principal ideal domain. Let a be a non-zero non-unit. We need to prove that a is a product of irreducible elements, and that the factors are unique up to order and associates.

Claim: a has an irreducible factor. This is certainly true if a is irreducible. So assume that a is not irreducible. Then $a = a_1 a_2$ for some non-zero non-units a_1, a_2. If a_1 or a_2 is irreducible, the claim is proved. Otherwise there exist non-zero non-units a_{ij} such that $a_i = a_{i1} a_{i2}$. If any of these is irreducible, the claim is proved, otherwise we continue: $a_{ij} = a_{ij1} a_{ij2}$, etc., $a_{i_1 i_2 \cdots i_n} = a_{i_1 i_2 \cdots i_n 1} a_{i_1 i_2 \cdots i_n 2}$. Then $(a) \subsetneq (a_1) \subsetneq (a_{11}) \subsetneq (a_{111}) \subsetneq \cdots$. But by Theorem 34.6, this chain has to stop, say at the nth step. Thus a with n 1s in the subscript must be irreducible, and it is a factor of a.

Claim: a is a product of irreducible elements. By the previous claim we know that $a = a_1 b_1$ for some irreducible element b_1 and some $a_1 \in R$. If a_1 is a unit, we are done, otherwise by the previous claim $a_1 = a_2 b_2$ for some irreducible b_2 and some $a_2 \in R$. We continue in this way to get a strictly increasing chain $(a) \subset (a_1) \subset (a_2) \subset \cdots$ of ideals in R. Since every increasing chain of ideals in R must terminate, this procedure has to stop in the nth step, which forces a_n be irreducible, so that $a = b_1 \cdots b_n a_n$, and each of the $n+1$ factors on the right is irreducible.

Now suppose that $a = a_1 \cdots a_n = b_1 \cdots b_m$ for some irreducible elements a_i, b_j in R. We need to prove that $n = m$ and that up to reordering, a_i is an associate of b_i. If $n = 1$, then $a = a_1$ is irreducible, which forces $m = 1$ and $b_1 = a_1$. So suppose that $n > 1$. By Theorem 35.2, (b_1) is a maximal ideal, so it is a prime ideal, $a_1 \cdots a_n \in (b_1)$, so by induction on n and by the definition of prime ideals, there exists i such that $a_i \in (b_1)$. By possibly reindexing, say $a_1 \in (b_1)$. Write $a_1 = u_1 b_1$, and necessarily u_1 is a unit in R. It follows that a_1 and b_1 are associates, and that $(u_1 a_2) a_3 \cdots a_n = b_2 \cdots b_m$. Note that all $a_3, \cdots a_n, b_2, \cdots, b_m, u_1 a_2$ are irreducible, so that by induction on n, $n = m$, and by reindexing, a_i is an associate of b_i for $i > 2$ and that $u_1 a_2$ is an associate of b_2. Thus a_i is an associate of b_i for all i, and $m = n$. □

Just as in $\mathbb{Z}$, also in a general unique factorization domain R, one can talk about **the greatest common divisor** of finitely many elements $a_1, \ldots, a_n$: this is an element $d \in R$ such that $d \mid a_i$ for all i and such that whenever $e \in R$ and $e \mid a_i$ for all i, then $e \mid d$. Note that d is determined uniquely only up to associates!

Example 36.4 $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain. Namely, $2 \cdot 3 = (1-\sqrt{-5})(1+\sqrt{-5})$. The four listed elements are irreducible:

(1) If $2 = \alpha_1\alpha_2$ for some non-units $\alpha_1, \alpha_2 \in \mathbb{Z}[\sqrt{-5}]$, then multiplication by the complex conjugate produces $4 = (\alpha_1\overline{\alpha}_1)(\alpha_2\overline{\alpha}_2)$. Necessarily $\alpha_i\overline{\alpha}_i = 2$ for each i, but this is not possible!

(2) Similarly 3 is irreducible.

(3) If $1 \pm \sqrt{-5} = \alpha_1\alpha_2$ for some non-units α_i, then as above $6 = (\alpha_1\overline{\alpha}_1)(\alpha_2\overline{\alpha}_2)$, and we get a similar contradiction.

Furthermore, 2 is not an element of ideals $(1 - \sqrt{-5})$ or $(1 + \sqrt{-5})$. Namely, if $2 = \alpha(1 \pm \sqrt{-5})$ for some α, then multiplication by the complex conjugate produces $4 = (\alpha\overline{\alpha})6$, which is not possible as $\alpha\overline{\alpha}$ is an integer.

In particular, this shows that $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain. Indeed, the ideal $(2, 1 - \sqrt{-5})$ is not principal. (This ring is an example of a ring of integers that is not a principal ideal domain.)

Exercise 36.5 Let $R = \mathbb{R}[X, Y]$. Find $\gcd(X^2 - 1, X^3 - 1)$, $\gcd(X, Y)$, $\gcd(2X^2, 3X^3)$.

Exercise 36.6 (This is a generalization of the Euclidean algorithm.) Let R be a principal ideal domain, and let $a_1, \ldots, a_n \in R$. Prove that $(a_1, \ldots, a_n) = (\gcd(a_1, \ldots, a_n))$ (ideals). Prove that the greatest common divisor of $a_1, \ldots, a_n$ can be expressed as an R-linear combination of $a_1, \ldots, a_n$.

Exercise 36.7 Give, with proof, an example of a unique factorization domain R and $a, b \in R$ such that $(a, b) \neq (\gcd(a, b))$ (ideals). Compare with Exercise 36.6.

Proposition 36.8 *Let R be a unique factorization domain. If $r \in R$ is irreducible, then (r) is a prime ideal.*

Proof. First of all, r is not a unit, so $(r) \neq R$. Let $a, b \in R$ such that $ab \in (r)$. Then $ab = cr$ for some $c \in R$. A factorization of a and b into irreducibles and by uniqueness of factorizations shows that an associate of r, and hence r, is a factor of either a or b, whence either $a \in (r)$ or $b \in (r)$. Thus (r) is a prime ideal. □

Corollary 36.9 *Let R be a principal ideal domain. If $r \in R$ is irreducible, then (r) is a maximal ideal.*

Proof. Since every principal ideal domain is a unique factorization domain, by the previous proposition, (r) is a prime ideal. If M is a proper ideal in R containing (r), necessarily $M = (s)$ for some $s \in R$, whence $r = st$ for some $t \in R$. Since r is irreducible and s is not a unit, necessarily t is a unit, whence $M = (s) \subseteq (r)$, so $(r) = M$ is a maximal ideal. □

Exercise 36.10 Let R be a unique factorization domain. As in Exercise 32.14, let F be the field of all fractions $\frac{a}{b}$ with $a, b \in R$ and $b \neq 0$. Let X be a variable over R. Generalize Theorem 35.7 to: $f \in R[X]$ factors as a product of two non-constant polynomials in $R[X]$ if and only if it is reducible over F.

The proof of the following theorem is lengthy, you may want to skip it:

Theorem 36.11 *Let R be a unique factorization domain. Then $R[X]$ is a unique factorization domain.*

Proof. We use Exercise 32.14: let F be the fraction field of R, i.e., the set of all fractions $\frac{a}{b}$ with $a, b \in R$ and $b \neq 0$.

Let f be a non-zero non-unit polynomial in $R[X]$. Since R is a unique factorization domain, there exists $d \in R$ that is the greatest common divisor of the coefficients of f. Write $f = dg$ for some $g \in R[X]$. By assumption, d factors uniquely into irreducibles in R, which are then also irreducibles in $R[X]$. Show that it suffices to prove that g has a unique irreducible factorization.

Suppose that $g = g_1 g_2 \cdots g_n$ and $g = h_1 h_2 \cdots h_m$ are further unfactorable factorizations in $R[X]$. By Exercise 36.10, a further unfactorable factorization of g in $R[X]$ is a further unfactorable factorization of g in $F[X]$. But $F[X]$ is a principal ideal domain, so any further unfactorable factorization of g in $F[X]$ is unique up to order and multiplication by units in $F[X]$. By Remark 32.6, the units in $F[X]$ are the units in F. Thus $n = m$, and, after reindexing, there exist units $f_1, \ldots, f_n \in F$ such that $g_i = f_i h_i$ for all i. By clearing denominators, for some non-zero $r_i, s_i \in R$, $r_i \cdot g_i = s_i \cdot h_i$. Since R is a unique factorization domain, by dividing without loss of generality $\gcd(r_i, s_i) = 1$. The coefficients of $s_i \cdot h_i$ are R-multiples of r_i, and since R is a factorization domain and $\gcd(r_i, s_i) = 1$, then the coefficients of h_i are R-multiples of r_i. It follows that the coefficients of g are R-multiples of r_i, so r_i must be a unit in R and hence in $R[X]$. Similarly, s_i must be a unit in R. Therefore the factorization of g is unique up to order and multiplication by units in $R[X]$ also in $R[X]$. □

Theorem 36.12 (The Hilbert Basis Theorem) *Let R be a ring in which every ideal is finitely generated. Then every ideal in $R[X_1, \ldots, X_n]$ is finitely generated.*

Proof. (This proof is reminiscent of the Gröbner basis manipulations, see Section 37. It suffices to prove that every ideal I in $R[X]$ is finitely generated. Every non-zero element $f \in I$ is a polynomial in X with coefficients in R. Set I_n to be the subideal of I generated by elements of I of degree at most n. Let J_n be the set in R consisting of 0 and all the leading coefficients of all non-zero elements in I_n. Verify that J_n is an ideal in R! By assumption, J_n is finitely generated, say by $r_{n1}, \ldots, r_{nk_n}$. There are correspondingly elements $f_{ni} \in I_n$ such that the leading coefficient of f_{ni} is r_{ni}. We may even assume that $\deg f_{ni} \leq n$ for all i. Let $N_n = \max\{\deg f_{ni} : i = 1, \ldots, k_n\}$. Note that $N_n \leq n$.

Claim: For all $n \geq 1$, $I_n = I_{n-1} + (f_{ni} : i = 1, \ldots k_n)$. Certainly $(f_{ni}) + I_{n-1} \subseteq I_n$. Let $f \in I_n$. If $\deg f < n$, then $f \in I_{n-1}$. Now suppose that $\deg f \geq n$. The leading coefficient c of f can be written as $\sum_i a_i r_{ni}$ for some $a_i \in R$. Then $f - \sum_i a_i f_{ni} X^{\deg f - \deg f_{ni}} \in I_n$ and has degree strictly smaller than f. This proves the claim.

Observe that $J_0 \subseteq J_1 \subseteq J_2 \subseteq \cdots$, so that $J = \cup J_n$ is an ideal in R (why!). Thus it is finitely generated, say by $r_1, \ldots, r_k$, and there are corresponding $f_i \in I$ such that r_i is the leading term of f_i. Let N be the maximum of all the degree of $f_1, \ldots, f_k$.

Claim: For all $n \geq N$, $I_n = (f_1, \ldots, f_k) + I_{n-1}$. Certainly $(f_1, \ldots, f_k) + I_{n-1} \subseteq I_n$. Let $f \in I_n$. If $\deg f < n$, then $f \in I_{n-1}$. Now suppose that $\deg f \geq n$. The leading coefficient c of f can be written as $\sum_i a_i r_i$ for some $a_i \in R$. Then $f - \sum_i a_i f_i X^{\deg f - \deg f_i} \in I$ and has degree strictly smaller than f. This proves the claim.

Thus $I_N = I_{N+1} = I_{N+2} = \cdots$, so $I = I_N$, and so I is generated by $f_1, \ldots, f_k$ and by f_{ni} as n varies from 0 to $N - 1$ and the corresponding i varies from 1 to n_i. □

In particular, every ideal in $\mathbb{Z}[X_1, \ldots, X_n]$ or in $F[X_1, \ldots, X_n]$ where F is a field, is finitely generated. Recall Exercise 34.8 which says that all ideals being finitely generated is equivalent to the stabilization of every ascending chain of ideals in R.

Rings that satisfy the ascending chain condition are called **Noetherian**, in honor of Emmy Noether, who studied them extensively.

Exercise 36.13 Let $\varphi : \mathbb{Q}[X_1, X_2, X_3] \to \mathbb{Q}[Y]$ be the ring homomorphism given by $\varphi(X_1) = Y - Y^2$, $\varphi(X_2) = Y^3$, $\varphi(X_3) = Y^4$. By the Hilbert's Basis Theorem, the kernel of φ is a finitely generated ideal. Find, with proof, a finite set of generators.

37 Monomial orderings (all in exercises)

Exercise 37.1 Let S be the set of all products of powers of variables $X_1, \ldots, X_n$. Define the **lexicographic order** on S as follows: $X_1^{a_1} \cdots X_n^{a_n} \geq X_1^{b_1} \cdots X_n^{b_n}$ if the left-most non-zero entry in $(a_1 - b_1, \ldots, a_n - b_n)$ is positive. Prove that this order is a well-ordering, and that if $s, t, u \in S$ with $s \geq t$, then $us \geq ut$.

Exercise 37.2 Let S be the set of all products of powers of variables $X_1, \ldots, X_n$. Define the **degree lexicographic order** on S as follows: $X_1^{a_1} \cdots X_n^{a_n} \geq X_1^{b_1} \cdots X_n^{b_n}$ if one of the following holds:
(1) $a_1 + \cdots + a_n > b_1 + \cdots + b_n$,
(2) or $a_1 + \cdots + a_n = b_1 + \cdots + b_n$ and the left-most non-zero entry in $(a_1 - b_1, \ldots, a_n - b_n)$ is positive.
Prove that this order is a well-ordering, and that if $s, t, u \in S$ with $s \geq t$, then $us \geq ut$.

Exercise 37.3 Let S be the set of all products of powers of variables $X_1, \ldots, X_n$. Define the **reverse lexicographic order** on S as follows: $X_1^{a_1} \cdots X_n^{a_n} \geq X_1^{b_1} \cdots X_n^{b_n}$ if the right-most non-zero entry in $(a_1 - b_1, \ldots, a_n - b_n)$ is negative. Show that this order need not be a well-ordering, and that if $s, t, u \in S$ with $s \geq t$, then $us \geq ut$.

Exercise 37.4 Let S be the set of all products of powers of variables $X_1, \ldots, X_n$. Define the **degree reverse lexicographic order** on S as follows: $X_1^{a_1} \cdots X_n^{a_n} \geq X_1^{b_1} \cdots X_n^{b_n}$ if one of the following holds:
(1) $a_1 + \cdots + a_n > b_1 + \cdots + b_n$,
(2) or $a_1 + \cdots + a_n = b_1 + \cdots + b_n$ and the right-most non-zero entry in $(a_1 - b_1, \ldots, a_n - b_n)$ is negative.
Prove that this order is a well-ordering, and that if $s, t, u \in S$ with $s \geq t$, then $us \geq ut$.

Exercise 37.5 Prove that if $n \leq 2$, the degree lexicographic and the reverse degree lexicographic orders are identical. Give examples of monomials with $n > 2$ where the four orders do not agree.

Exercise 37.6 Let S be the set of all products of powers of variables $X_1, \ldots, X_n$. A **monomial order** on S is any total order satisfying:
(1) $s \geq 1$ for any $s \in S$;
(2) If $s, t, u \in S$ with $s \geq t$, then $us \geq ut$.
Prove that a monomial order is a well-ordering.

Exercise 37.7 Let F be a field, and let R be the polynomial ring $F[X_1, \dots, X_n]$. Let $\geq$ be a monomial order on the set of all products of variables $X_1, \dots, X_n$. For any non-zero $f \in R$, define the **leading monomial** $\operatorname{lm}(f)$ to be the unique largest monomial $m \in S$ that appears in f with a non-zero coefficient. The **leading coefficient** is denoted by $\operatorname{lc}(f)$, and the **leading term** $\operatorname{lt} f = \operatorname{lc} f \cdot \operatorname{lm} f$.

(1) Compute the leading monomials of $Y^3 - XY$ under the degree reverse lexicographic order and under the lexicographic order.

(2) Let G be a non-empty subset of R. A **reduction step** with respect to $(G, \geq)$ is a procedure which takes as input a polynomial f in R and whose output is a polynomial $f - mg \in R$, where $g \in G$ and the monomial m are chosen so that $\operatorname{lt}(f)$ equals $m\operatorname{lt}(g)$. If there is no such g, the reduction step returns f. A **reduction** with respect to $(G, \geq)$ is a procedure which applies recursively reduction steps to polynomials and stops either when the reduction step returns the zero polynomial or when it returns the polynomial whose leading monomial is not a multiple of the leading monomial of any element of G. Apply reduction to $G = \{Y^3 - XY, X^2 - XY\}$, $f = X^4Y^3$, once in the degree reverse lexicographic order and once in the lexicographic order.

Definition 37.8 *Let R be a polynomial ring, and let $\leq$ be a monomial ordering on the monomials of R. Let I be an ideal in R. A finite set $G \subset I$ is called a* **Gröbner basis** *of I if for every $f \in I$ there exists $g \in G$ such that $\operatorname{lt} f$ is a multiple of $\operatorname{lt} g$.*

Exercise 37.9 Prove that G is a generating set of I.

Exercise 37.10 (Gröbner basis algorithm) Let F be a field, let R be the polynomial ring $F[X_1, \dots, X_n]$, and let $\geq$ be a monomial order on the set of all products of variables $X_1, \dots, X_n$. For any $f, g \in R$, the **S-polynomial** of f and g is

$$S(f,g) = \frac{\operatorname{lcm}(\operatorname{lm} f, \operatorname{lm} g)}{\operatorname{lt} f} f - \frac{\operatorname{lcm}(\operatorname{lm} f, \operatorname{lm} g)}{\operatorname{lt} g} g.$$

Input: A finite generating set G of an ideal I in R.
Output: A Gröbner basis G of I.

```
for all $f$, $g$ $\in$ $G$,
      reduce $S(f,g)$ with respect to $G$
      if the resulting polynomial is not $0$, add it to $G$
repeat as long as any $S(f,g)$ do not reduce to $0$
```

Compute the Gröbner basis of $I = (X^2 - XYZ, Y^2 - Z^2)$ under the degree reverse lexicographic order and under the lexicographic order.

Here is a discussion on the termination of this algorithm. (It does terminate, so it is an algorithm.) At every instance of the loop in the algorithm, we might add a polynomial f to G. We do so only under the condition that $\operatorname{lt} f$ is not a multiple of the leading terms of elements of the current G. Let J be the ideal in $F[X_1, \dots, X_n]$ generated by the leading terms of elements of G. If we had to add infinitely many elements to G, then we'd be at the same time constructing an infinite strictly increasing chain of ideals $J_1 \subsetneq J_2 \subsetneq \cdots \subsetneq J$ (all ideals generated by monomials). But by Hilbert's Basis theorem, J is finitely generated, say by $r_1, \dots, r_k$, so that by Exercise 29.9, there must be a finite subset of the leading terms of elements in G that generate J. But then we couldn't be adding any more subsequent elements to G! So construction of a Gröbner basis terminates after finitely many steps.

In the polynomial ring $F[X]$ in one variable over a field F, the reduction step is simply the division algorithm, and computing the Gröbner basis is finding elements in the ideal of smaller and smaller non-negative degrees. Let the element in the Gröbner basis of smallest degree be called b. By the reduction step in the computation of Gröbner bases, b is unique up to a scalar unit multiple, and reductions of other elements with respect to $\{b\}$ necessarily produce 0 (for otherwise the non-zero elements would have strictly smaller degree than b, contradiction), whence all other elements are multiples of b. Thus b must necessarily be the greatest common divisor of all the generators of the ideal.

We just proved that the division and the Euclidean algorithm in $F[X]$ are applications or special cases of the Gröbner basis algorithm. When more than one variable is involved, the Euclidean algorithm is not available, but the Gröbner basis algorithm is. We expand on this more next:

Exercise 37.11 (Generalization of the division algorithm to several variables) Let $R = F[X_1, \ldots, X_n]$. Let I be an ideal of R and $f \in R$. Let G be a Gröbner basis of I.
(1) Prove that $f \in I$ if and only if the reduction of f with respect to G is 0.
(2) Find I, a generating set G of I, and $f \in I$, such that the reduction of f with respect to G is not 0.

Exercise 37.12 (Product order) Let $\geq_X$ and $\geq_T$ be monomial orders on $F[X_1, \ldots, X_n]$ and $F[T_1, \ldots, T_m]$, respectively. We order monomials in $F[X_1, \ldots, X_n, T_1, \ldots, T_m]$ as follows: $X_1^{a_1} \cdots X_n^{a_n} T_1^{b_1} \cdots T_m^{b_m} \geq X_1^{c_1} \cdots X_n^{c_n} T_1^{d_1} \cdots T_m^{d_m}$ if one of the following holds:

- $T_1^{b_1} \cdots T_m^{b_m} \geq_T T_1^{d_1} \cdots T_m^{d_m}$ and $T_1^{b_1} \cdots T_m^{b_m} \neq T_1^{d_1} \cdots T_m^{d_m}$.
- $T_1^{b_1} \cdots T_m^{b_m} = T_1^{d_1} \cdots T_m^{d_m}$ and $X_1^{a_1} \cdots X_n^{a_n} \geq_X X_1^{c_1} \cdots X_n^{c_n}$.

Prove that $\geq$ is a monomial order.

Exercise 37.13 Let $R = F[X_1, \ldots, X_n]$, $S = R[T_1, \ldots, T_m]$. Let $\geq$ be a monomial order on S such that every monomial in which any T_i appears with a non-zero exponent is larger than any monomial from R.
(1) Prove that the lexicographic order with $T_1 > \cdots > T_m > X_1 > \cdots > X_n$ is a possible order. Show that the degree lexicographic order with $T_1 > \cdots > T_m > X_1 > \cdots > X_n$ is not a possible order. Prove that the product order as in the previous exercise is a possible order.
(2) Let I be an ideal in S, G its Gröbner basis. Prove that $G \cap R$ is a Gröbner basis of $I \cap R$. ($G \cap R$ is still a finite set, $I \cap R$ is an ideal in R.)

A reason why one may want to compute such intersections above is as follows. A general use of Gröbner bases is to find solutions of polynomial systems. If one has a polynomial in one variable, one may be able to factor it, or find solutions numerically, or find field extensions (later in the semester) where solutions exist. But when we start with many polynomials in many variables, it is much harder to find solutions. Let I be an ideal in $F[X_1, \ldots, X_n]$. Under a monomial order on $X_1, \ldots, X_n$ in which any monomial in which some $X_2, \ldots, X_n$ appears is bigger than any monomial in which only X_1 appears, then if $I \cap F[X_1] \neq 0$, a Gröbner basis will contain e non-zero element of $I \cap F[X_1]$. For this one polynomial one can find finitely many solutions, as mentioned, after which one can manipulate the polynomials with X_1 replaced in turn by each of the numerical values to possibly find solutions in the other variables.

Try the ideal $(X^2+Y^2, X-Y^2)$: compute its lexicographic Gröbner basis, and its degree reverse lexicographic Gröbner basis: one gives you a polynomial only in Y, the other doesn't. Compare with the previous exercise.

Exercise 37.14 Prove that it is possible to compute algorithmically the intersection of two ideals in a polynomial ring.

Exercise 37.15 Let $\varphi : F[X_1, \ldots, X_n] \to F[Y_1, \ldots, Y_m]$ be a ring homomorphism. (Here, F is a field, the X_j and the Y_i are variables over F.) Let $R = F[X_1, \ldots, X_n, Y_1, \ldots, Y_m]$, and let $\geq$ be a total monomial order on R, with any monomial containing a Y_i to a positive exponent being larger than any monomial using only the X_j. Let G be the Gröbner basis of the ideal $(X_i - \varphi(X_i) : i = 1, \ldots, n)$.
(1) Prove that $G \cap F[X_1, \ldots, X_n]$ is a generating set of the kernel of φ.
(2) Compute the kernel of $\varphi : \mathbb{Q}[X,Y,Z] \to \mathbb{Q}[T]$, where $\varphi(X) = T^4$, $\varphi(Y) = T^5$, $\varphi(Z) = T^6$.
(3) If you found the previous part too easy, try: Compute the kernel of $\varphi : \mathbb{Q}[X,Y,Z] \to \mathbb{Q}[T]$, where $\varphi(X) = T^4$, $\varphi(Y) = T^5$, $\varphi(Z) = T^7$.

Exercise 37.16 Let $\varphi : F[X_1, \ldots, X_n]/I \to F[Y_1, \ldots, Y_m]/J$ be a ring homomorphism. (Here, F is a field, the X_j and the Y_i are variables over F, I is an ideal in $F[X_1, \ldots, X_n]$ and J is an ideal in $F[Y_1, \ldots, Y_m]$.) Let $R = F[X_1, \ldots, X_n, Y_1, \ldots, Y_m]$, and let $\geq$ be a total monomial order on R, with any monomial containing a Y_i to a positive exponent being larger than any monomial using only the X_j. Let G be the Gröbner basis of the ideal in R generated by all the $X_i - \varphi(X_i)$, by I and by J. Prove that the image of $G \cap F[X_1, \ldots, X_n]$ in $F[X_1, \ldots, X_n]/I$ is a generating set of the kernel of φ.

Exercise 37.17 Let $\varphi : \mathbb{Q}[X_1, X_2, X_3] \to \mathbb{Q}[Y]$ be the ring homomorphism given by $\varphi(X_1) = Y - Y^2$, $\varphi(X_2) = Y^3$, $\varphi(X_3) = Y^4$. By the Hilbert's Basis Theorem, the kernel of φ is a finitely generated ideal.
(1) Spend at least 15 minutes trying to find a finite set of generators by brute force.
(2) Set up the problem with the help of Gröbner bases. Compute the kernel with Gröbner bases.

Let R be a commutative ring, I and $J = (j_1, \ldots, j_k)$ ideals in R. Define $I : J = \{r \in R : rJ \subseteq I\}$.

Exercise 37.18 Let R be a commutative ring, I and $J = (j_1, \ldots, j_k)$ ideals in R. Define $I : J = \{r \in R : rJ \subseteq I\}$. This is referred to as the **colon ideal** of I with J.
(1) Prove that $I : J$ is an ideal of R.
(2) Prove that $I : J = \cap_{a \in J}(I : (a)) = \cap_{i=1}^{k}(I : (j_i))$.
(3) Let $x \in R$. Prove that $I \cap (x) = (x)(I : (x))$.
(4) Assume that intersections and products of ideals are computable in R and that R is a domain. Prove that $I : J$ is computable in R. (In particular, $I : J$ is computable in polynomial rings over fields.)

38 Modules

Modules over rings play the role of vector spaces over fields:

Definition 38.1 *Let R be a ring. A* **left R-module***, or a* **left module over** *R, is a group $(M, +)$ with a function $R \times M \to M$ denoted $(r, m) \mapsto rm$ satisfying:*
(1) $(rs)m = r(sm)$ for all $r, s \in R$, $m \in M$,
(2) $(r + s)m = rm + sm$ for all $r, s \in R$, $m \in M$,
(3) $r(m_1 + m_2) = rm_1 + rm_2$ for all $r \in R$, $m_1, m_2 \in M$,
(4) If $1 \in R$, then $1m = m$ for all $m \in M$.
We will omit the word "left". The function $R \times M \to M$ is referred to as multiplication, or action, of R on M. Indeed, (3) says that this action is the group action of the group $(R, +)$ on M.

Examples 38.2 Every ring is a module over itself. Every ideal in R is an R-module. Every abelian group is a $\mathbb{Z}$-module. Every vector spaces over a field F is an F-module.

Definition 38.3 *A* **generating set** *of an R-module M is a subset S of M such that every $m \in M$ can be written as an R-linear (finite!) combination of elements of S. If S is finite, we say that M is* **finitely generated***.*

Examples 38.4 Every R-module M is generated by M. Of course, we'd rather have small and smaller generating sets. If the ring R has 1, then as an R-module, R is generated by $\{1\}$. The generators of an ideal in R as an R-module are the same as the ideal generators. The generators of an abelian group are also its generators as a $\mathbb{Z}$-module. A basis of a vector space over a field F is a generating set as an F-module.

Remark 38.5 Let M, N be R-modules. The Cartesian product of M and N can be made into an R-module, denoted $M \oplus N$: addition is componentwise, and action/multiplication by elements of R is also componentwise. If S is a generating set of M and T is a generating set of N, then $\{(s, 0), (0, t) : s \in S, t \in T\}$ is a generating set of $M \oplus N$.

Definition 38.6 Submodule. **Module homomorphism**. **Quotient (or factor) module** *M/N if $N \subseteq M$ are R-modules (abelian groups, factor compatible with ring multiplication).* **Kernel, image***, quotient module are all modules.*

Theorem 38.7 (First Isomorphism Theorem, for modules; yet another incarnation) *Let $\varphi : M \to N$ be an R-module homomorphism with kernel K. Then $M/K \cong \mathrm{Im}\varphi$.*

The proof if essentially the same as the proof of the previous two incarnations of this theorem.

Definition 38.8 *Let R be a ring. A* **free** *R-module over R* **of rank** *r is the r-fold direct sum of the R-module R with itself. The nicest generating set of R^r is $\{e_i : i = 1, \ldots, r\}$, where e_i is an r-tuple with 1 in the ith component and 0 elsewhere.*

Remark 38.9 $\{e_i : i = 1, \ldots, r\}$ form a so-called **basis** of R^r: their R-span is the whole module, and if $\sum a_i e_i = 0$ for some $a_i \in R$, then all a_i are 0. But this is not the only possible basis for R^r: say each e_i can be replaced by a unit times e_i; any two e_i can be switched; or e_i can be replaced by $e_i + \sum_{j \neq i} a_j e_j$. (Recall the elementary row operations from linear algebra!)

Theorem 38.10 *Let R be a ring and M a finitely generated R-module. Then M is a homomorphic image of a free R-module of finite rank.*

Proof. Let $\{m_1, \ldots, m_s\}$ be a generating set of M as an R-module. Define

$$\varphi : R^s \to M,$$
$$(a_1, \ldots, a_s) \mapsto \sum a_i m_i.$$

Verify that this is an R-module homomorphism. It is clearly surjective. □

Theorem 38.11 *Let R be a commutative ring with identity satisfying the ascending chain condition, i.e., that every ascending sequence of ideals stabilizes at some point. Then every finitely generated R-module also satisfies the ascending chain condition, i.e., every ascending sequence of submodules of M stabilizes at some point.*

Proof. Let $\varphi : R^s \to M$ be a surjective module homomorphism. Let K be the kernel. Observe that any ascending sequence of submodules of M lifts uniquely to an ascending sequence of submodules of R^s that contain K. It suffices to prove that this latter sequence stabilizes. By induction on s it suffices to prove the subsequent theorem. □

Theorem 38.12 *Let R be a commutative ring, and let M and N be R-modules with the ascending chain condition. Then $M \oplus N$ satisfies the ascending chain condition.*

Proof. So let $K_1 \subseteq K_2 \subseteq \cdots$ be an ascending chain of submodules of $M \oplus N$. Let M_i be the subset of M consisting of all $m \in M$ such that for some $n \in N$, $(m, n) \in K_i$. Define $N_i \subseteq N$ to consist of all those $n \in N$ such that $(0, n) \in K_i$. Observe that M_i is an R-submodule of M, and that $M_1 \subseteq M_2 \subseteq \cdots$. Similarly, N_i is an R-submodule of N and $N_1 \subseteq N_2 \subseteq \cdots$. By assumption there exists i such that $M_i = M_{i+1} = M_{i+2} = \cdots$ and $N_i = N_{i+1} = N_{i+2} = \cdots$. Claim: $K_i = K_{i+1} = K_{i+2} = \cdots$. Let $(m, n) \in K_{i+j}$ for some $j \in \mathbb{N}$. By assumption $m \in M_{i+j} = M_i$, so that there exists $n' \in N$ such that $(m, n') \in K_i$. Then $(0, n - n') = (m, n) - (m, n') \in K_{i+j}$, so that $n - n' \in N_{i+j} = N_i$, whence

$$(m, n) = (0, n - n') + (m, n') \in K_i. \qquad \square$$

Exercise 38.13 (Compare with Exercise 34.8.) Let R be a ring, and M an R-module. Prove that the following are equivalent:
(1) Every submodule of M is finitely generated.
(2) Every ascending chain of submodules in M eventually stabilizes.

Modules which satisfy this condition are called **Noetherian**. In Theorem 38.11 we proved that a finitely generated over a Noetherian ring is Noetherian. In Theorem 38.12 we proved that the direct sum of finitely many Noetherian modules is Noetherian. Clearly a homomorphic image of a Noetherian module is Noetherian. Almost by definition a submodule of a Noetherian module is Noetherian. However, a submodule of a finitely generated module need not be finitely generated! (Example: let $R = \mathbb{Q}[X_1, X_2, \ldots]$, as a module over itself, R is finitely generated by the element 1, but the submodule=ideal $(X_1, X_2, \ldots)$ is not finitely generated.)

39 Finitely generated modules over principal ideal domains

Theorem 39.1 (Fundamental Theorem of Finitely Generated Modules over Principal Ideal Domains) *Let R be a principal ideal domain. Prove that every finitely generated R-module is isomorphic to an R-module of the form $R/(r_1) \oplus \cdots \oplus R/(r_k) \oplus R^l$ for some $r_i \in R$ and some $l \in \mathbb{N}$ such that $r_1|r_2|\cdots|r_k$.*

Proof. Let M be a finitely generated R-module. By Theorem 38.10, M is a homomorphic image of R^s for some integer s. By the First Isomorphism Theorem, $M \cong R^s/K$ for some R-submodule K of R^s. By Theorem 38.12, K is finitely generated, say by elements $k_1, \ldots, k_m$. (Actually we don't need finite generation!) Each of these $k_1, k_2, \ldots$ is an element of R^s, so K can be represented with an $s \times m$ matrix A (or possibly by an $s \times \infty$ matrix A?).

We use the following row (resp. column) operations: switch any two, add a multiple of one to another, and multiply one by a unit in R. Such row operations correspond to switching two generating elements of R^s and M, to replacing one of the generators by itself plus an R-multiple of another, and replacing one by a unit multiple of itself. Thus these operations do not change M. The column operations correspond to similarly allowed operations on the generators of K and do not change K nor M. Furthermore, we may multiply A on the right or on the left by any matrix that is invertible in R: the multiplication on the left reversibly changes the generators of M, and multiplication on the right reversibly changes the generators of K. We apply these operations systematically. The entries of A are in R, and since R is a principal ideal domain, there exists $r_1 \in R$ such that (r_1) is the ideal generated by all the entries.

The workhorse here is that the allowed row and column operations allow us to assume that r_1 is the entry A_{11}. First we prove that there exists an $s \times s$ matrix M with entries in R such that $\det M = 1$ (so that M has an inverse with entries in R) and such that MA is a matrix whose first column is $[r \;\; 0 \;\; \cdots \;\; 0]^T$, where r is necessarily the greatest common divisor of the entries of the first column in A. This is trivial if $s = 1$. Suppose that $s = 2$. By Exercise 36.6, there exists $a_1, a_2 \in R$ such that $a_1 A_{11} + a_2 A_{21} = r$. Then set

$$M = \begin{bmatrix} a_1 & a_2 \\ -\frac{A_{21}}{r} & \frac{A_{11}}{r} \end{bmatrix}.$$

This matrix has entries in R, it has determinant 1, and $MA = [r \;\; 0]^T$. Now suppose that $s > 2$. By induction there exists an $(s-1) \times (s-1)$ matrix M' with entries in R and with determinant 1 such that if A' is the matrix obtained from A by omitting the first row, then $M'A'$ has the first column equal to $\gcd(A_{21}, \ldots, A_{s1})e_1$. Since $r = \gcd(A_{11}, \ldots, A_{s1}) = \gcd(A_{11}, \gcd(A_{21}, \ldots, A_{s1}))$, by the case $s = 2$, there exists a 2×2 matrix M'' with entries in R and of determinant 1 such that

$$M'' \begin{bmatrix} A_{11} \\ \gcd(A_{21}, \ldots, A_{s1}) \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}.$$

Set

$$M = \left[\begin{array}{c|c} M'' & \mathbf{0} \\ \hline \mathbf{0} & I_{s-2} \end{array}\right] \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & M' \end{array}\right].$$

Then M has entries in R, $\det M = 1$, and $MA = [r \;\; 0]^T$. (Multiplication of A by M on the left is allowed.)

By similarly modifying the first row, without loss of generality $A_{1j} = 0$ for $j > 1$. But while modifying the first row, we possibly introduce non-zero elements into the first column! So, did we accomplish anything? Yes, we did! We started with the original matrix A. After the row manipulations (and renaming) we got the entry A_{11} to divide all entries in the first column. After the column operations (and renaming) we got a new entry A_{11} that divides the previous A_{11} and even all entries in the previous row 1. Thus we are building principal ideals that form an increasing chain of ideals. Since R is a principal ideal domain, this procedure must stop, so that after some finite number of iterations (and renaming), A_{11} divides all the entries in the first row and in the first column. By elementary row and column operations we may then assume that $A_{1j} = A_{i1} = 0$ for all $i, j > 1$.

If now A_{11} divides all the entries in A, we are done, otherwise there exist $i, j > 1$ such that A_{ij} is not a multiple of A_{11}. Add the ith column to the first column, at which point the greatest common divisor of the elements of the first column strictly divides A_{11}, so performing the column operations produces a new matrix A for which the ideal (A_{11}) produces a strictly bigger ideal than was given by the old A_{11}. We keep applying this step to produce strictly bigger ideals – but R is a principal ideal domain, so every ascending chain of ideals must stabilize, which means that at some point A_{11} divides all the entries of A.

In the next step, we clear the rest of the first row and column to be 0. After that the obvious manipulation is to work on the $(s-1) \times (r-1)$ submatrix of A that omits the first rows and columns. Thus in finitely many steps we get the matrix A to be:

$$\begin{bmatrix} r_1 & 0 & \cdots & & 0 & 0\cdots0 \\ 0 & r_1 & 0 & \cdots & 0 & 0\cdots0 \\ & & \ddots & & & \\ 0 & & \cdots & & r_k & 0\cdots0 \\ 0 & & \cdots & & 0 & 0\cdots0 \\ & & \ddots & & & \\ 0 & & \cdots & & 0 & 0\cdots0 \end{bmatrix},$$

for some $r_1, \ldots, r_k \in R$, such that $r_1 | r_2 | \cdots | r_k$. (Possibly there are no 0 rows or no 0 columns.)

There is the obvious module isomorphism now:

$$M \cong \frac{R^s}{\langle r_i e_i : i = 1, \ldots, k\rangle} \cong \frac{R}{(r_1)} \oplus \frac{R}{(r_2)} \oplus \cdots \oplus \frac{R}{(r_k)} \oplus R^{s-k}. \qquad \square$$

An immediate corollary is the following promised theorem about commutative groups:

Theorem 39.2 (Fundamental Theorem of Finitely Generated Abelian Groups) *Let G be a finitely generated commutative group. Then $G \cong \frac{\mathbb{Z}}{n_1\mathbb{Z}} \oplus \cdots \oplus \frac{\mathbb{Z}}{n_k\mathbb{Z}} \oplus \mathbb{Z}^l$ for some positive integers $n_1, \ldots, n_k, l$ such that $n_1 | n_2 | \cdots | n_k$.*

Proof. Every commutative group is a $\mathbb{Z}$-module, and every finitely generated commutative group is a finitely generated $\mathbb{Z}$-module. Thus we are done by the previous theorem. □

There are various versions of this Fundamental Theorem in which the n_i are chosen in a special way. Justify the following claims:
(1) **(Invariant factors)**: $n_1|n_2|\cdots|n_k$. (This was done in the proof above!)
(2) **(Elementary divisors)** One can assume that each n_i is a power of a prime integer.
(3) Prove that the invariant factors and the elementary divisors are unique.

Corollary 39.3 (Fundamental Theorem of Finite Abelian Groups) *Let G be a finite commutative group. Then $G \cong \frac{\mathbb{Z}}{n_1\mathbb{Z}} \oplus \cdots \oplus \frac{\mathbb{Z}}{n_k\mathbb{Z}}$ for some positive integers $n_1, \ldots, n_k$.*

Proof. Every finite abelian group is finitely generated, so by the previous theorem, $G \cong \frac{\mathbb{Z}}{n_1\mathbb{Z}} \oplus \cdots \oplus \frac{\mathbb{Z}}{n_{k'}\mathbb{Z}} \oplus \mathbb{Z}^l$. Since G is finite, $k' = k$ and $l = 0$. □

Examples 39.4 Lattice theory, linear algebra, relation modules ??

Exercise 39.5 Let R be a principal ideal domain, and let $r_1, \ldots, r_s \in R$. Prove that there exists an $s \times s$ matrix M with entries in R whose first row or columns is $[r_1 \;\; r_2 \;\; \cdots \;\; r_s]$ and whose determinant is $\gcd(r_1, \ldots, r_s)$. In particular, if $\gcd(r_1, \ldots, r_s) = 1$, then M is an invertible matrix over R, i.e., there exists an $s \times s$ matrix N with entries in R such that $MN = NM$ equals the $s \times s$ identity matrix.

Example 39.6 (Purdue University qualifying exam towards Ph.D. in mathematics, August 1989) Problem: Express $\frac{\mathbb{Z}^3}{\langle f_1, f_2, f_3\rangle}$, where $f_1 = (1, -1, 1)$, $f_2 = (5, 1, -5)$, $f_3 = (-3, -3, 30)$, as a direct sum of cyclic groups.
Solution:

$$\begin{bmatrix} 1 & 5 & -3 \\ -1 & 1 & -3 \\ 1 & -5 & 30 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 5 & -3 \\ 0 & 6 & -6 \\ 0 & -10 & 33 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & -6 \\ 0 & 2 & 21 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 138 \end{bmatrix},$$

which is isomorphic to $\frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{138\mathbb{Z}} \cong \frac{\mathbb{Z}}{138\mathbb{Z}}$.

Example 39.7 (Purdue University qualifying exam towards Ph.D. in mathematics, January 1989) Problem: Let F be a field and X a variable over F. Express $\frac{F[X]^3}{\langle f_1, f_2, f_3\rangle}$, where $f_1 = (X, 1, 0)$, $f_2 = (1, X, 0)$, $f_3 = (0, 0, X - 1)$, as $\frac{F[X]}{(g_1)} \oplus \frac{F[X]}{(g_2)} \oplus \frac{F[X]}{(g_3)}$, where $g_1|g_2|g_3$.
Solution:

$$\begin{bmatrix} X & 1 & 0 \\ 1 & X & 0 \\ 0 & 0 & X-1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & X & 0 \\ X & 1 & 0 \\ 0 & 0 & X-1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-X^2 & 0 \\ 0 & 0 & X-1 \end{bmatrix},$$

which is isomorphic to $\frac{F[X]}{(1)} \oplus \frac{F[X]}{(X-1)} \oplus \frac{F[X]}{(X^2-1)}$.

Example 39.8 (Purdue University qualifying exam towards Ph.D. in mathematics, January 1988) Problem: Let F be a field and X a variable over F. Express $\frac{F[X]^3}{\langle f_1, f_2, f_3\rangle}$, where $f_1 = (X(X-1)(X^2+1), 0, 0)$, $f_2 = (0, X^2(X-1)^2, 0)$, $f_3 = (0, 0, (X-1)^3(X^2+1)^2)$, as $\frac{F[X]}{(g_1)} \oplus \frac{F[X]}{(g_2)} \oplus \frac{F[X]}{(g_3)}$, where $g_1|g_2|g_3$.

Not quite a solution: Obviously we could repeat the procedure in the previous example (which is exactly the procedure outlined in the proof of the Fundamental Theorem of finitely generated modules over principal ideal domains), but we won't. We'll develop some more theory, and then the solution will come running at us. See the next section.

40 The Chinese Remainder Theorem

Theorem 40.1 (The Chinese Remainder Theorem) *Let R be a commutative ring with* 1, *let I_1, I_2 be ideals in R such that $I_1 + I_2 = R$. Then*

$$\frac{R}{I_1 \cap I_2} \cong \frac{R}{I_1} \oplus \frac{R}{I_2}.$$

Proof. Define $\varphi : R \to \frac{R}{I_1} \oplus \frac{R}{I_2}$ by $\varphi(r) = (r + I_1, r + I_2)$. It is easy to verify that this is a ring homomorphism. The kernel is $\{r \in R : r + I_1 = I_1, r + I_2 = I_2\} = \{r \in R : r \in I_1 \cap I_2\} = I_1 \cap I_2$. Let $a, b \in R$. We will prove that $(a + I_1, b + I_2) = \varphi(r)$ for some $r \in R$. This then proves that φ is surjective. We have not yet used that $I_1 + I_2 = R$. This assumption allows us to write $1 = i_1 + i_2$ for some $i_j \in I_j$. Set $r = ai_2 + bi_1$. Then

$$\begin{aligned} \varphi(r) &= (ai_2 + bi_1 + I_1, ai_2 + bi_1 + I_2) \\ &= (a(1 - i_1) + I_1, b(1 - i_2) + I_2) \\ &= (a + I_1, b + I_2), \end{aligned}$$

which proves the claim. Thus by the First Isomorphism Theorem for rings (Theorem 31.2), the theorem follows. □

We can even prove a stronger version:

Theorem 40.2 (The Chinese Remainder Theorem) *Let $I_1, \ldots, I_k$ be ideals in a commutative ring R with identity. Then*

$$\frac{R}{I_1 \cap \cdots \cap I_k} \cong \frac{R}{I_1} \oplus \cdots \oplus \frac{R}{I_k}$$

if and only if for all $i = 1, \ldots, k$, $I_i + \cap_{j \neq i} I_j = R$.

Proof. Assume that for all $i = 1, \ldots, k$, $I_i + \cap_{j \neq i} I_j = R$. Then by Theorem 40.1,

$$\frac{R}{I_1 \cap \cdots \cap I_k} \cong \frac{R}{I_1} \oplus \frac{R}{I_2 \cap \cdots \cap I_k}.$$

But for all $i > 1$, $I_i + \cap_{1 < j \neq i} I_j$ contains $I_i + \cap_{j \neq i} I_j = R$, so that $I_i + \cap_{1 < j \neq i} I_j = R$. Thus by induction on k,

$$\frac{R}{I_1 \cap \cdots \cap I_k} \cong \frac{R}{I_1} \oplus \frac{R}{I_2 \cap \cdots \cap I_k} \cong \frac{R}{I_1} \oplus \cdots \oplus \frac{R}{I_k}.$$

Now assume that

$$\frac{R}{I_1 \cap \cdots \cap I_k} \cong \frac{R}{I_1} \oplus \cdots \oplus \frac{R}{I_k}.$$

Let $\varphi : \frac{R}{I_1 \cap \cdots \cap I_k} \cong \frac{R}{I_1} \oplus \cdots \oplus \frac{R}{I_k}$ be the isomorphism. Note that for all $r \in R$, $\varphi(r) = r\varphi(1)$. Write $\varphi(1) = (a_1 + I_1, \ldots, a_k + I_k)$ for some $a_i \in R$. Since φ is surjective, for each $i = 1, \ldots, k$, there exists $r_i \in R$ such that $\varphi(r_i) = (0, \ldots, 0, 1+I_i, 0, \ldots, 0)$. This means that for all i, $r_i a_i + I_i = 1 + I_i$ and that for all distinct i, j, $r_i a_i + I_j = I_j$. In other words, there exists $b_i \in I_i$ such that $r_i a_i = 1 + b_i$, and $r_i a_i \in \cap_{j \neq i} I_j$. But then $1 = r_i a_i - b_i \in I_i + \cap_{j \neq i} I_j$, which proves that for all i, $I_i + \cap_{j \neq i} I_j = R$. □

Note how this theorem generalizes Theorem 20.6. Also, the proof seems more streamlined here than for the $\mathbb{Z}_n$ version(?).

Now we go back to Example 39.8, using the Chinese Remainder Theorem. First note that

$$X - (X-1) = 1, \qquad (1 - \frac{1}{2}X)(X^2+1) + (\frac{1}{2}X - \frac{1}{2})(X^2 - X) = 1,$$

so that

$$\frac{F[X]}{(X(X-1)(X^2+1))} \cong \frac{F[X]}{(X(X-1))} \oplus \frac{F[X]}{(X^2+1)} \cong \frac{F[X]}{(X)} \oplus \frac{F[X]}{(X-1)} \oplus \frac{F[X]}{(X^2+1)}.$$

Similarly,

$$\begin{aligned}
\frac{F[X]^3}{\langle f_1, f_2, f_3 \rangle} &\cong \frac{F[X]}{(X(X-1)(X^2+1))} \oplus \frac{F[X]}{(X^2(X-1)^2)} \oplus \frac{F[X]}{((X-1)^3(X^2+1)^2)} \\
&\cong \frac{F[X]}{(X)} \oplus \frac{F[X]}{(X-1)} \oplus \frac{F[X]}{(X^2+1)} \\
&\quad \oplus \frac{F[X]}{(X^2)} \oplus \frac{F[X]}{((X-1)^2)} \\
&\quad \oplus \frac{F[X]}{((X-1)^3)} \oplus \frac{F[X]}{((X^2+1)^2)} \\
&\cong \qquad \frac{F[X]}{(X-1)} \\
&\quad \oplus \frac{F[X]}{(X)} \oplus \frac{F[X]}{((X-1)^2)} \oplus \frac{F[X]}{(X^2+1)} \\
&\quad \oplus \frac{F[X]}{(X^2)} \oplus \frac{F[X]}{((X-1)^3)} \oplus \frac{F[X]}{((X^2+1)^2)} \\
&\cong \frac{F[X]}{(X-1)} \oplus \frac{F[X]}{(X(X-1)^2(X^2+1))} \oplus \frac{F[X]}{(X^2(X-1)^3(X^2+1)^2)},
\end{aligned}$$

which is in the form $\frac{F[X]}{(g_1)} \oplus \frac{F[X]}{(g_2)} \oplus \frac{F[X]}{(g_3)}$, where $g_1 | g_2 | g_3$.

41 Fields

Definition of fields is assumed, actually throughout these notes!

Definition 41.1 *A field E is an* **extension field** *of a field F if $F \subseteq E$ and the operations of F are those of E restricted to F. In other words, E is an extension of F if and only if F is a subfield of E.*

Let E be an extension field of F. Then E is clearly an F-vector space, of finite or possibly of infinite vector space dimension.

Definition 41.2 *By $\dim_F(E)$ we denote the vector space dimension of E over F.*

We have a way of generating a lot of extension fields:

$$F \subseteq F[X]/(\text{ an irreducible polynomial}).$$

We often have a way of expressing an extension field as $F[X_1, \dots, X_n]/I$: say, let E be the extension field consisting of all polynomials in $\sqrt{2}$, $\sqrt{3}$, and $\sqrt{5}$ over $\mathbb{Q}$. Map $\mathbb{Q}[X, Y, Z] \to \mathbb{C}$ with $X \mapsto \sqrt{2}$, $Y \mapsto \sqrt{3}$, $Z \mapsto \sqrt{5}$. Certainly $X^2 - 2, Y^2 - 3, Z^2 - 5$ are in the kernel, and they actually generate the kernel. Hence this field is isomorphic to $\mathbb{Q}[X, Y, Z]/(X^2 - 2, Y^2 - 3, Z^2 - 5)$. Another way to write this particular field is also as $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$ (round brackets for taking not just polynomials in these elements, but also fractions with non-zero denominators).

Definition 41.3 *Let F be a field and X a variable over F. We define $F(X)$ to be the set of all fractions of elements in $F[X]$, where the denominator is of course not zero. Verify that $F(X)$ is a field!*

In general, if R is a subring of a ring S and $s_1, \dots, s_n \in R$, then $R[s_1, \dots, s_n]$ is the subset of S consisting of all polynomials in $s_1, \dots, s_n$ with coefficients in R, and this turns out to be a ring. If R and S are both fields, $R(s_1, \dots, s_n)$ is the subset of S consisting of all fractions of polynomials in $s_1, \dots, s_n$ with coefficients in R (the denominator is of course not zero), and this turns out to be a field. However, $R[s_1, \dots, s_n]$ need not be a field even if R and S are fields. For example, if F is a field and X is a variable, then the subring $F[X]$ of $F(X)$ is not a field!

Theorem 41.4 **(Fundamental Theorem of Field Theory)** *Let F be a field and $f(X)$ a non-constant polynomial in $F[X]$. Then there is an extension field E of F in which $f(X)$ has a root (= zero).*

Proof. We have proved that $F[X]$ is a unique factorization domain. Let $p(X)$ be an irreducible factor of $f(X)$. Write $f(X) = p(X)g(X)$ for some $g(X) \in F[X]$. Let $E = F[X]/(p(X))$. We know that E is a field and that it contains F. Let $\overline{X}$ be the image of X in E. Then in E, $f(\overline{X}) = p(\overline{X})g(\overline{X}) = 0$. □

If the root in the theorem is α, we denote this field extension is $F(\alpha)$, and $F(\alpha)$ is isomorphic to $F[X]$ modulo the ideal generated by the irreducible factor of $f(X)$ of which α is the root.

Definition 41.5 *Let $F \subseteq E$ be a field extension, $\alpha \in E$ a root of a non-zero polynomial with coefficients in E. Then the unique monic irreducible polynomial in $F[X]$ that has α as a root is denoted* $\mathrm{Irr}\,(\alpha, F)$*.*

We first need to justify this uniqueness. If $f(X)$ and $g(X)$ are both irreducible elements of $F[X]$ and $f(\alpha) = g(\alpha) = 0$, then if $p(X) = \gcd(f(X), g(X))$ is also in $F[X]$ and $p(\alpha) = 0$ since $p(X)$ is an $F[X]$-linear combination of $f(X)$ and $g(X)$. But $f(X)$ and $g(X)$ are irreducible, so they must be unit multiples of $p(X)$, hence $f(X)$ and $g(X)$ are associates. But if they are both monic, they must be equal.

With the set-up as in Theorem 41.4, if α is a root in E of the non-constant polynomial $f(X) \in F[X]$, then $F(\alpha)$ is a vector space over F, with dimension being the degree d of $\mathrm{Irr}\,(\alpha, F)$, as a basis for the F-vector space $F(\alpha)$ is $\{1, \alpha, \alpha^2, \dots, \alpha^{d-1}\}$.

If we know that all polynomials in $\mathbb{R}[X]$ have all the roots in $\mathbb{C}$, the we immediately get:

Corollary 41.6 *All irreducible polynomials in $\mathbb{R}[X]$ have degree at most* 2. □

Examples 41.7 $\mathrm{Irr}\,(\sqrt{2}, \mathbb{Q}) = X^2 - 2$, $\mathrm{Irr}\,(\sqrt{2}, \mathbb{R}) = X - \sqrt{2}$. Let α be a complex root of $X^4 + X^3 + X^2 + X + 1$, say $\alpha = e^{2\pi i/5}$. Then $\mathrm{Irr}\,(\alpha, \mathbb{Q}) = X^4 + X^3 + X^2 + X + 1$, $\mathrm{Irr}\,(\alpha, \mathbb{R}) = X^2 - 2\cos(2\pi/5)X + 1 = X^2 + \frac{1-\sqrt{5}}{2}X + 1$, $\mathrm{Irr}\,(\alpha, \mathbb{C}) = X - \alpha$. If $\beta = e^{4\pi i/5}$, then $\mathrm{Irr}\,(\beta, \mathbb{Q}) = X^4 + X^3 + X^2 + X + 1$, $\mathrm{Irr}\,(\beta, \mathbb{R}) = X^2 - 2\cos(4\pi/5)X + 1 = X^2 + \frac{1+\sqrt{5}}{2}X + 1$, $\mathrm{Irr}\,(\beta, \mathbb{C}) = X - \beta$. Furthermore, $\mathrm{Irr}\,(\beta, \mathbb{Q}(\alpha)) = X^2 + \frac{1-\sqrt{5}}{2}X + 1$, $\mathrm{Irr}\,(\alpha, \mathbb{Q}(\alpha)) = X - \alpha$.

Another example: let γ be a cube root of unity other than 1. Then $\mathrm{Irr}\,(\gamma, \mathbb{Q}) = X^2 + X + 1$, $\mathrm{Irr}\,(\gamma, \mathbb{Q}[i]) = X^2 + X + 1$.

Lemma 41.8 *Let $F \subseteq E \subseteq K$ be field extensions. Let $\{\alpha_1, \dots, \alpha_n\}$ be a basis of the F-vector space E, and let $\{\beta_1, \dots, \beta_m\}$ be a basis of the E-vector space K. Then $\{\alpha_i\beta_j : i = 1, \dots, n; j = 1, \dots, m\}$ is a basis of the F-vector space K. In particular,*

$$\dim_F(E)\,\dim_E(K) = \dim_F(K).$$

Proof. □

Here is an important corollary:

Corollary 41.9 *Every finite field has cardinality a power of a prime integer.*

Proof. By Theorem 32.13 a finite field F has finite prime characteristic, say p. Then $\mathbb{Z}/p\mathbb{Z}$ is a subfield of F, and F is a vector space over $\mathbb{Z}/p\mathbb{Z}$. Since F is finite, the dimension of this vector space is finite, say n. Then, as a vector space, $F \cong (\mathbb{Z}/p\mathbb{Z})^n$, whence $|F| = p^n$. □

Exercise 41.10 Phrase and prove the lemma for an arbitrary-dimensional version of Lemma 41.8.

The lemma enables us to rephrase certain equations into a mostly numerical argument! We illustrate this on an example.

Let δ be a sixth root of -4 (a complex number). Certainly δ is a zero of $X^6 + 4$, but it is not obvious whether $X^6 + 4$ is irreducible. Clearly $X^6 + 4$ has no linear factors. We could try to write $X^6 + 4$ as a product of two cubics and as a product of a quadratic with a quartic, treat the coefficients as unknowns, multiply the factors together to get $X^6 + 4$, obtain corresponding restrictions/equations on the coefficients, and hopefully get a contradiction that there are no such polynomials with rational coefficients. But we will prove the irreducibility of $X^6 + 4$ in another, more elegant, way. Observe that δ is a root of the polynomial $X^3 + 2i$ or of the polynomial $X^3 - 2i$, so $i \in \mathbb{Q}(\delta)$. Is $X^3 \pm 2i$ irreducible over $\mathbb{Q}[i]$? If we knew that it was irreducible, we'd have the following diagram of fields, with lines denoting that the field on the higher end contains the field at the lower end, and the number at the line gives the vector space dimension of the bigger field over the smaller:

$$\begin{tikzcd}
\mathbb{Q}[\delta] \arrow[d, dash, "3"] \\
\mathbb{Q}[i] \arrow[d, dash, "2"] \\
\mathbb{Q}
\end{tikzcd}$$

This would then show that $\dim_{\mathbb{Q}}\mathbb{Q}[\delta] = 6$, so that $X^6 + 4$ is irreducible over $\mathbb{Q}$. But let's make the same conclusion in another way.

Note that $(\delta^2)^3 = -4$, so that δ^2 is one of the cube roots of -4. Clearly $\text{Irr}\,(\delta^2, \mathbb{Q}) = X^3 + 4$, which is irreducible over $\mathbb{Q}$. Thus we get the following diagram of fields, with lines denoting that the field on the higher end contains the field at the lower end, and the number at the line gives the vector space dimension of the bigger field over the smaller:

$$\begin{tikzcd}
 & \mathbb{Q}[\delta] \arrow[dl, dash] \arrow[dr, dash] & \\
\mathbb{Q}[\delta^2] \arrow[dr, dash, "3"'] & & \mathbb{Q}[i] \arrow[dl, dash, "2"] \\
 & \mathbb{Q} &
\end{tikzcd}$$

By Lemma 41.8, the vector space dimension d of $\mathbb{Q}[\delta]$ over $\mathbb{Q}$ is a multiple of 6. But we also know that δ satisfies a monic polynomial of degree 6 over $\mathbb{Q}$, so that $d \leq 6$. Hence necessarily $d = 6$, which means that $X^6 + 4$ must be irreducible!

(Now isn't this a much more elegant proof than the method of unknown coefficients and multiplying through and equating and solving...???!!!)

Example 41.11 Let $f(X) = X^3 + 2 \in \mathbb{Q}[X]$. Let α, β, γ be all the three complex roots of this. (Justify!) Assume that α is real. Look at the diagram below. Since $f(X)$ is irreducible, the vector space dimensions of $\mathbb{Q}[\alpha]$ and $\mathbb{Q}[\beta]$ over $\mathbb{Q}$ are both 3. But $\beta \notin \mathbb{Q}[\alpha]$ because β is not real, for one thing, so $\mathbb{Q}[\alpha] \neq \mathbb{Q}[\alpha, \beta]$. Justify the rest of the diagram:

$$\begin{tikzcd}
 & \mathbb{Q}[\alpha, \beta, \gamma] = \mathbb{Q}[\alpha, \beta] \arrow[dl, dash, "2"'] \arrow[dr, dash, "2"] & \\
\mathbb{Q}[\beta] \arrow[dr, dash, "3"'] & & \mathbb{Q}[\alpha] \arrow[dl, dash, "3"] \\
 & \mathbb{Q} &
\end{tikzcd}$$

Example 41.12 Let $f(X) = X^6 + 8X^4 + 1 \in \mathbb{Q}[X]$. Let $\alpha \in \mathbb{C}$ be a root of $f(X)$. It is elementary to verify that $X^3 + 8X^2 + 1$ is irreducible. **Challenge:** In how many ways can you determine if $f(X)$ is irreducible?

42 Splitting fields

Definition 42.1 *Let F be a field and $f(X) \in F[X]$ and let E be an extension field of F. We say that $f(X)$* **splits** *in E if $f(X)$ can be factored into linear factors in $E[X]$. We call E the* **splitting field** *of $f(X)$ over F if $f(X)$ splits in E but it splits in no smaller subfield.*

Examples 42.2 Find the splitting fields of $X^2 + 1$ over $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\frac{\mathbb{Z}}{2\mathbb{Z}}$, $\frac{\mathbb{Z}}{3\mathbb{Z}}$.

Theorem 42.3 (Existence of splitting fields) *Let F be a field and f a non-constant polynomial of $F[X]$. Then there exists a splitting field for f over F.*

Proof. We proceed by induction on the degree of f. If $\deg f = 1$, then f splits in F. Now let $\deg f > 1$. By Theorem 41.4, there exists a field extension F' of F in which f has a root α. By the division algorithm, there exists a polynomial $g \in F'[X]$ such that $f = (X - \alpha)g$. By induction on the degree there exists a splitting field E of g. Verify that E is also a splitting field of f. □

For example, $X^2+1 \in \mathbb{Q}[X]$ splits over $\mathbb{C}$, but its splitting fields are: $\mathbb{Q}[i]$, $\mathbb{Q}[X]/(X^2+1)$, $\mathbb{Q}[X,Y]/(X^2+1,Y)$, $\mathbb{Q}[X,Y]/(X^2+1,Y-1)$, etc.

We will prove that any two splitting fields are isomorphic. First we need a lemma.

Lemma 42.4 *$F(\alpha) \cong F(\beta)$ with F as a subset of $F(\alpha)$ mapping identically onto F as a subset of $F(\beta)$ and α mapping to β, if and only if either both α and β are transcendental over F (in the sense that they do not satisfy any algebraic equation over F, as for example π or X do not satisfy any algebraic equation over $\mathbb{Q}$) or else both α and β are the roots of the same irreducible polynomial over F.*

Proof. The transcendental part is clear. So we assume that both α and β are algebraic. Let $f(X) = \mathrm{Irr}\,(\alpha, F)$ and $g(X) = \mathrm{Irr}\,(\beta, F)$. Then $F[X]/(\mathrm{Irr}\,(\alpha, F) \cong F(\alpha) \cong F(\beta) \cong F[X]/(\mathrm{Irr}\,(\beta, F))$, and under this composition isomorphism, elements of F map to themselves. In particular, since the coefficients of $\mathrm{Irr}\,(\alpha, F)$ are in F, $0 = \mathrm{Irr}\,(\alpha, F)$ maps to 0, so it has to be a multiple of $\mathrm{Irr}\,(\beta, F)$, but by irreducibility then these two polynomials have to equal. □

More generally, if $\varphi : F \to F'$ is a field isomorphism, $f(X) \in F[X]$ is irreducible if and only if $\varphi(f)(X) \in F'[X]$ is irreducible. If an extension field E of F contains a root α of $f(X)$ and an extension field E' of F' contains a root β of $\varphi(f)(X)$, then there exists a field isomorphism $F[\alpha] \cong F'[\beta]$.

Theorem 42.5 *Any two splitting fields are isomorphic.*

Proof. Let $f \in F[X]$, say of degree n, and let E, E' be two splitting fields of f. Let $f = u\prod_{i=1}^{n}(X - \alpha_i)$ be a factorization of f in E, and let $f = u\prod_{i=1}^{n}(X - \alpha'_i)$ be a factorization of f in E'. We may assume that α_1 and α'_1 are roots of the same irreducible factor of f over F. We just proved that there exists an isomorphism $\varphi_1 : F[\alpha_1] \to F[\alpha'_1]$

that takes F identically to F and α_1 to α_1'. Suppose we have found an isomorphism $\varphi_k : K = F[\alpha_1, \ldots, \alpha_k] \to K' = F[\alpha_1', \ldots, \alpha_k']$ that takes F identically to F and α_i to α_i'. The coefficients of the **polynomial**(!) $g = \frac{f}{(X-\alpha_1)\cdots(X-\alpha_k)}$ in $K[X]$ map via φ_k to the coefficients of the polynomial $g' = \frac{f}{(X-\alpha_1')\cdots(X-\alpha_k')}$ in $K'[X]$. Their respective splitting fields are E and E'. If $k < n$, α_{k+1} is a root of f, and $\mathrm{Irr}\,(\alpha_{k+1}, K)$ is a factor of f. When φ_k is applied to the coefficients of $\mathrm{Irr}\,(\alpha_{k+1}, K)$, we get an irreducible factor h of g', so by possibly permuting $\alpha_{k+1}', \ldots, \alpha_n'$, without loss of generality α_{k+1}' is a root of h. By the previous lemma, there exists an isomorphism

$$\varphi_{k+1} : F(\alpha_1, \ldots, \alpha_{k+1}) \to F(\alpha_1', \ldots, \alpha_{k+1}')$$

such that φ_{k+1} restricted to $F(\alpha_1, \ldots, \alpha_k)$ is φ_k, and $\varphi_{k+1}(\alpha_{k+1}) = \alpha_{k+1}'$. Thus by induction on k we get an isomorphism $\varphi_n : F(\alpha_1, \ldots, \alpha_n) = E \to F(\alpha_1', \ldots, \alpha_n') = E'$. □

Exercise 42.6 Show that the splitting field of a polynomial $f(X) \in F[X]$ of degree n has vector space dimension at most $n!$ over F. Find an example with $n > 2$ where this dimension is achieved.

43 Derivatives in algebra (optional)

Definition 43.1 *Let F be a field, $f(X) = a_0 + a_1X + \cdots a_nX^n \in F[X]$. The* **derivative** *of $f(X)$ is $f'(X) = a_1 + 2a_2X + 3a_3X^2 + \cdots + na_nX^{n-1}$. (No limits needed, this is well defined!)*

Theorem 43.2 *$f \in F[X]$ has repeated factors in some field extension if and only if* $\gcd(f, f') \neq 1$.

The sum, product, chain rules apply to this derivative as well.

Proof. Let E be a field extension of F such that for some $g, h \in E[X]$, $f = g^2h$. Then $f' = 2gg'h + g^2h'$. If $\gcd(f, f') = 1$ over $F[X]$, then $1 = rf + sf'$ for some $r, s \in F[X]$, but then since g divides f and f' in $E[X]$, it must also divide 1, contradiction.

Now assume that $\gcd(f, f')$ is a non-constant polynomial g. Without loss of generality over some splitting field E of f, $g = \prod_{i=1}^r (X - \alpha_i)$, $f = \prod_{i=1}^n (X - \alpha_i)$. Let $h = \prod_{i=2}^n (X - \alpha_i)$. Then $f = (X - \alpha_1)h$, $f' = h + (X - \alpha_1)h'$. Since $(X - \alpha_1)$ is a factor of f' over E, necessarily $X - \alpha_1$ is also a factor of h, whence it is a double factor of f. □

Example 43.3 Let t, X be variables over $\frac{\mathbb{Z}}{p\mathbb{Z}}$. Then $F = \frac{\mathbb{Z}}{p\mathbb{Z}}(t) \subseteq E = F[X]/(X^p - t)$ is a field extension, and if α is a root of the irreducible $X^p - t$, then $X^p - t = X^p - \alpha^p = (X - \alpha)^p$ has repeated roots. (This phenomenon that an irreducible polynomial has repeated roots only happens in characteristic p.)

44 Finite fields

By Corollary 41.9, every finite field has cardinality p^n for some prime p and some positive integer n. Furthermore, such a field is an extension of $\mathbb{Z}/p\mathbb{Z}$.

Theorem 44.1 *Let p be a prime number and n a positive integer. Any finite field F of cardinality p^n is the splitting field of $X^{p^n} - X \in \frac{\mathbb{Z}}{p\mathbb{Z}}[X]$.*

Proof. We know that $(F\backslash\{0\}, \cdot)$ is a commutative group of cardinality p^n-1. By Lagrange's Theorem Theorem 16.4, for all $x \in F \setminus \{0\}$, $x^{p^n-1} = 1$. Thus for all $x \in F$, $x^{p^n} - x = 0$. Obviously all the p^n distinct elements of F are the roots of this polynomial, so by unique factorization, $X^{p^n} - X = \prod_\alpha (X - \alpha)$, as α varies over the elements of F, and there isn't a smaller subfield containing all these elements α. Thus $X^{p^n} - X$ splits in F. Thus F is the splitting field. □

Exercise 44.2 Prove that if F is a finite field, then $F\setminus\{0\}$ is a cyclic commutative group. An overly generous hint: By the Fundamental Theorem of Finite Abelian groups, $F \setminus \{0\}$ is isomorphic to $\frac{\mathbb{Z}}{n_1\mathbb{Z}} \oplus \cdots \oplus \frac{\mathbb{Z}}{n_k\mathbb{Z}}$ with $n_1|n_2|\cdots|n_k$ (one is in multiplicative and one in additive notation!). In particular, for all $\alpha \in F \setminus \{0\}$, $\alpha^{n_k} = 1$. If $k > 1$, then n_k properly divides $|F| - 1$, and so $X^{n_k} - 1$ is a proper factor of $X^{|F|} - 1$. But the polynomial $X^{n_k} - 1$ has at most n_k roots in F (and in $F \setminus \{0\}$),etc.

Corollary 44.3 *Any two finite fields of the same cardinality are isomorphic.* □

A dumb question: Are any two fields of the same cardinality isomorphic?

Corollary 44.4 *For any prime integer p and any positive integer e there exists a field of cardinality p^e.*

Proof. Take the splitting field of $X^{p^e} - X$ over $\mathbb{Z}/p\mathbb{Z}$. □

Example 44.5 Let $f(X) = X^5 + X^4 + X^3 + X^2 + 1 \in \frac{\mathbb{Z}}{3\mathbb{Z}}[X]$. Then the irreducible factorization of $f(X)$ is $(X^2 + X - 1)(X^3 - X - 1)$. Obviously some roots of $f(X)$ lie in a field of order 9 and some lie in a field of order 27. Let K be the field obtained by adjoining to $\frac{\mathbb{Z}}{3\mathbb{Z}}$ all the roots of $f(X)$. What size is K? (Use Lemma 41.8 to show that it must have $9 \cdot 27$ elements.)

Exercise 44.6 Let $f(X) = X^2 + X + 2, g(X) = X^2 + 2X + 2 \in \frac{\mathbb{Z}}{3\mathbb{Z}}[X]$. Let K be the field obtained by adjoining to $\frac{\mathbb{Z}}{3\mathbb{Z}}$ all the roots of $f(X)$. Let α be one of the roots.
(1) Prove that both f and g are irreducible polynomials.
(2) Verify that $K = F[\alpha]$.
(3) Show that some polynomial in α of degree at most 1 is a root of g.
(4) Conclude that K is the splitting field not just of f but also of g.
(5) Find the splitting field of $X^4 + 1$ over $\mathbb{Z}/3\mathbb{Z}$.

Proposition 44.7 *Let f, g be irreducible polynomials over a finite field F, both of degree n. Let α be root of f in some field extension E of F. Then the splitting field of f and of g is $F(\alpha)$.*

Proof. By Proposition 44.13, the splitting field of f is $F(\alpha)$, and this has cardinality $|F|^n$. Similarly, the splitting field of g is a field of the form $F(\beta)$ of cardinality $|F|^n$. Since any two finite fields of the same cardinality are isomorphic, there exists an isomorphism $\varphi : F(\beta) \to F(\alpha)$. Let $\beta' = \varphi(\beta)$. By Lemma 42.4, $g(\beta') = 0$, so $F(\beta')$ is also a splitting field of g. This field is a subfield of $F(\alpha)$, and since it also has cardinality $|F|^n$, the two fields are identical. □

Theorem 44.8 *Let p be a prime and n a positive integer. Then $X^{p^n} - X$ factors over $\frac{\mathbb{Z}}{p\mathbb{Z}}$ into the product of all monic irreducible polynomials over $\frac{\mathbb{Z}}{p\mathbb{Z}}$ of degree a divisor of n.*

Proof. Let F be a field of order p^n. Then F is the splitting field of $X^{p^n} - X$ over $\frac{\mathbb{Z}}{p\mathbb{Z}}$. Let p be an irreducible factor of $X^{p^n} - X$, and let $\alpha \in F$ be a root of p. Then $\frac{\mathbb{Z}}{p\mathbb{Z}} \subseteq \frac{\mathbb{Z}}{p\mathbb{Z}}(\alpha) \subseteq F$, and by Lemma 41.8, the degree of p is a factor of n.

Now suppose that p is an irreducible polynomial in $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ of degree d which is a factor of n. Let K be a splitting field of p, and let $\alpha \in K$ be a root of p. By Exercise 44.12, $\frac{\mathbb{Z}}{p\mathbb{Z}}(\alpha)$ is a splitting field of p, so $K = \frac{\mathbb{Z}}{p\mathbb{Z}}(\alpha)$. But then K has cardinality p^d, and so K is also the splitting field of $X^{p^d} - X$. Since d divides n, any root of $X^{p^d} - X$ is also a root of $X^{p^n} - X$ (certainly true for 0; if $a \neq 0$ and $a^{p^d-1} = 1$, then $a^{p^{de}-1} = (a^{(p^d-1)(p^{d(e-1)}+p^{d(e-2)}+\cdots+1)} = 1$). Thus since $X^{p^d} - X$ has only distinct roots, $X^{p^d} - X$ must divide $X^{p^n} - X$. Since $\alpha \in K$ is a root of $X^{p^d} - X$ and hence of $X^{p^n} - X$, it follows by the division algorithm for polynomials that $X^{p^n} - X$ is a multiple of f. □

Example 44.9 Over $\frac{\mathbb{Z}}{3\mathbb{Z}}$, Macaulay2 factored $X^{3^n} - X$ as follows:

$$\begin{aligned}
X^3 - X &= (X+1)(X-1)X, \\
X^9 - X &= (X-1)(X+1)X(X^2+X-1)(X^2+1)(X^2-X-1), \\
X^{27} - X &= (X+1)(X-1)X(X^3+X^2-X+1)(X^3-X^2-X-1) \\
&\quad (X^3-X-1)(X^3-X+1)(X^3+X^2+X-1)(X^3-X^2+1) \\
&\quad (X^3-X^2+X+1)(X^3+X^2-1), \\
X^{81} - X &= (X-1)(X+1)X(X^2-X-1)(X^2+1)(X^2+X-1) \\
&\quad (X^4+X^3-X^2-X-1)(X^4+X^3+X^2+X+1)(X^4-X-1) \\
&\quad (X^4-X^2-1)(X^4-X^3+X^2+1)(X^4+X^3+X^2+1) \\
&\quad (X^4+X^3-X+1)(X^4+X^2-1)(X^4+X^3+X^2-X-1) \\
&\quad (X^4+X^3-1)(X^4+X-1)(X^4-X^3-X^2+X-1) \\
&\quad (X^4-X^3+X+1)(X^4+X^2+X+1)(X^4-X^3+X^2+X-1) \\
&\quad (X^4+X^2-X+1)(X^4-X^3+X^2-X+1)(X^4-X^3-1).
\end{aligned}$$

Corollary 44.10 *If p is a prime and d is a positive integer, there exists an irreducible polynomial $f \in \frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ of degree d. More generally, if E is a finite field, there exists an irreducible polynomial in $E[X]$ of degree d.*

Proof. Let $|E| = p^n$. Let K be the splitting field of $X^{p^{nd}} - X$ over E. Since K is a finite field, by Exercise 44.2 there exists $\alpha \in K \setminus \{0\}$ whose powers vary over all the non-zero elements of K. In particular, $K = E[\alpha]$. Let $f = \mathrm{Irr}\,(\alpha, E)$. Then $E[\alpha] = E[X]/(f)$, and necessarily the degree of f is d. □

Exercise 44.11 Let f_n be the number of irreducible polynomials in $\frac{\mathbb{Z}}{p\mathbb{Z}}[X]$ of degree n.
(1) Prove that $p^n = \sum_{d|n} df_d$.
(2) (Possibly skip this part, it is somewhat long or hard.) Prove that

$$nf_n = \sum_{d|n} \mu(d)p^{n/d},$$

where $\mu : \mathbb{Z}_{>0} \to \mathbb{Z}$ is the **Möbius function**:

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1; \\ 0 & \text{if } n \text{ is a multiple of a square of some prime integer;} \\ (-1)^r & \text{if } n = p_1 \cdots p_r \text{ and } p_1, \ldots, p_r \text{ are distinct primes.} \end{cases}$$

(3) Prove that f_n is positive for all positive n.

Contrast this with irreducible polynomials over $\mathbb{R}$, $\mathbb{C}$. How about $\mathbb{Q}$? (Final project for $\mathbb{Q}$?)

Exercise 44.12 Let p be a prime integer, let $E \subseteq K$ be extension fields of $\frac{\mathbb{Z}}{p\mathbb{Z}}$ of finite cardinality, let $f(X) \in E[X]$ and let $\alpha \in K$ be a root of $f(X)$. Prove that $\alpha^{|E|}$ is also a root of $f(X)$. Conclude that if $f(X)$ is irreducible of degree d, then the splitting field of $f(X)$ has cardinality $|E|^d$.

Proposition 44.13 *Let f be an irreducible polynomial over a finite field F. Let α be root of f in some field extension E of F. Then the splitting field of f is $F(\alpha)$.*

Proof. Use Exercise 44.12. □

Exercise 44.14 Let $f = X^4 + X + 1 \in \frac{\mathbb{Z}}{2\mathbb{Z}}[X]$.
(1) Prove that f has no repeated roots.
(2) Prove that $X^2 + X + 1$ is the only irreducible quadratic in $\frac{\mathbb{Z}}{2\mathbb{Z}}[X]$ and that f is not its multiple.
(3) Let α be a root of f. Express all the roots of f as polynomials in α of degree at most 2.

Exercise 44.15 Determine the splitting field of $X^4 + 1$ over $\mathbb{Q}$ and over $\frac{\mathbb{Z}}{3\mathbb{Z}}$.

45 Appendix: Euclidean algorithm for integers

Theorem 45.1 *(Division algorithm) Let $m, n \in \mathbb{Z}$, with $n > 0$. Then there exist unique $q, r \in \mathbb{Z}$ ("q" for* **quotient**, *"r" for* **remainder***) such that*
(1) $m = qn + r$,
(2) $0 \leq r < n$.

Proof. By the Archimedean property of real numbers, there exists $x \in \mathbb{R}$ such that $xn \geq m$. Since n is positive, the set of such x is bounded below. Let x' be the smallest integer such that $x'n \geq m$, and set $q = x' - 1$. Then $q \in \mathbb{Z}$, $(q+1)n \geq m$, $qn < m$, and $r = m - qn$ satisfies all the desired properties. □

Observe that with m, n, r, q as in the theorem, $\gcd(m, n) = \gcd(n, r)$.

Theorem 45.2 *(Euclidean algorithm) Let $n_1, n_2 \in \mathbb{Z}$, with $n_2 > 0$. By repeated use of the Division Algorithm, there exist integers $s, q_1, r_1, \ldots, r_s, r_s$ such that for all $i = 1, \ldots, s$,*
(1) $n_i = q_i n_{i+1} + r_i$,
(2) $0 \leq r_i < n_{i+1}$,
(3) $n_{i+2} = r_i$,
(4) $r_1 > r_2 > \cdots > r_s = 0$.
Then $n_{s+1} = \gcd(n_1, n_2)$.

Proof. The steps in the construction of the q_i, r_i are clear. Since the r_i form a descending chain of non-negative integers, in finitely many steps (in fact, in at most n_2 steps), $r_s = 0$. By the remark after the division algorithm, $\gcd(n_1, n_2) = \gcd(n_2, r_1) = \gcd(r_2, r_3)$, and by proceeding in the same manner, $\gcd(n_1, n_2) = \gcd(n_{s-1}, n_s) = \gcd(n_s, n_{s+1}) = \gcd(n_{s+1}, r_s) = n_{s+1}$. □