# Algebra Interactive

## 2. Algebra Interactive

## 3. Algebra Interactive

| COLLABORATORS |
|---|

| | | |
|---|---|---|
| | *TITLE* :<br>Algebra Interactive | |

| *ACTION* | *NAME* | *DATE* | *SIGNATURE* |
|---|---|---|---|
| WRITTEN BY | Arjeh M. Cohen, Hans Cuypers, and Hans Sterk | September 6, 2011 | |

| REVISION HISTORY |
|---|

| NUMBER | DATE | DESCRIPTION | NAME |
|---|---|---|---|
| | | | |

# Contents

**1 Arithmetic** **1**

1.1 Divisors and Multiples . . . . . 1

1.2 Euclid’s algorithm . . . . . 8

1.3 Linear Diophantine equations . . . . . 15

1.4 Prime numbers . . . . . 18

1.5 Factorization . . . . . 25

1.6 The $b$-ary number system . . . . . 30

1.7 Exercises . . . . . 32

1.7.1 Divisors and multiples . . . . . 32

1.7.2 Euclid’s algorithm . . . . . 33

1.7.3 Linear Diophantine equations . . . . . 34

1.7.4 Prime numbers . . . . . 35

1.7.5 Factorization . . . . . 35

1.7.6 Number systems . . . . . 36

**2 Modular arithmetic** **37**

2.1 Arithmetic modulo n . . . . . 37

2.2 Linear congruences . . . . . 49

2.3 The Theorems of Fermat and Euler . . . . . 53

2.4 The RSA cryptosystem . . . . . 60

2.5 Exercises . . . . . 61

2.5.1 Arithmetic modulo an integer . . . . . 61

2.5.2 Linear congruences . . . . . 63

2.5.3 The theorems of Fermat and Euler . . . . . 64

2.5.4 The RSA cryptosystem . . . . . 64

3 **Polynomials** **65**
3.1 The notion of a polynomial . . . 65
3.2 Division of polynomials . . . 67
3.3 Polynomial functions . . . 76
3.4 Factorization . . . 79
3.5 Exercises . . . 82
3.5.1 The notion of a polynomial . . . 82
3.5.2 Division of polynomials . . . 82
3.5.3 Polynomial functions . . . 83
3.5.4 Factorization . . . 84

4 **Arithmetic modulo polynomials** **86**
4.1 Congruence modulo a polynomial . . . 86
4.2 The residue class ring . . . 89
4.3 Two special cases . . . 94
4.4 Inverses and fields . . . 96
4.5 Finite fields . . . 98
4.6 Error correcting codes . . . 101
4.7 Exercises . . . 106
4.7.1 Congruence modulo a polynomial . . . 106
4.7.2 The residue class ring . . . 107
4.7.3 Two special cases . . . 108
4.7.4 Inverses and fields . . . 109
4.7.5 Finite fields . . . 110
4.7.6 Error correcting codes . . . 111

5 **Permutations** **112**
5.1 Symmetric Groups . . . 112
5.2 Cycles . . . 115
5.3 Alternating groups . . . 121
5.4 Exercises . . . 125
5.4.1 Symmetric Groups . . . 125
5.4.2 Cycles . . . 125
5.4.3 Alternating groups . . . 127

6 **Monoids and groups** **129**

6.1 Binary operations . . . . . 129

6.2 Monoids and semigroups . . . . . 136

6.3 Invertibility in monoids . . . . . 144

6.4 Groups . . . . . 146

6.5 Cyclic groups . . . . . 152

6.6 Cosets . . . . . 155

6.7 Exercises . . . . . 161

6.7.1 Binary operations . . . . . 161

6.7.2 Monoids and semigroups . . . . . 161

6.7.3 Invertibility in monoids . . . . . 163

6.7.4 Groups . . . . . 163

6.7.5 Cyclic groups . . . . . 164

6.7.6 Cosets . . . . . 165

7 **Rings and fields** **166**

7.1 The structure ring . . . . . 166

7.2 Constructions with rings . . . . . 173

7.3 Domains and fields . . . . . 178

7.4 Fields . . . . . 185

7.5 Ideals . . . . . 195

7.6 Residue class rings . . . . . 203

7.7 Finite fields . . . . . 207

7.8 Exercises . . . . . 217

7.8.1 The structure ring . . . . . 217

7.8.2 Constructions with rings . . . . . 218

7.8.3 Domains and fields . . . . . 219

7.8.4 Fields . . . . . 219

7.8.5 Ideals . . . . . 219

7.8.6 Residue class rings . . . . . 220

7.8.7 Finite fields . . . . . 221

## 7. Algebra Interactive

8 **Groups** 222

8.1 Permutation groups . . . 222

8.2 Orbits . . . 229

8.3 Permutation group algorithms . . . 235

8.4 Automorphisms . . . 241

8.5 Quotient groups . . . 245

8.6 Structure theorems . . . 248

8.7 Exercises . . . 257

8.7.1 Groups . . . 257

8.7.2 Orbits . . . 258

8.7.3 Order . . . 261

8.7.4 Automorphisms . . . 262

8.7.5 Quotient groups . . . 264

8.7.6 Structure theorems . . . 266

# List of Tables

1.1 The primes less than or equal to 1013. . . . . . . . . . . . . . . . . . . . . . . 20
1.2 Eratosthenes' sieve . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
1.3 Prime factorizations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

2.1 Addition table for $\mathbb{Z}/17\mathbb{Z}$. . . . . . . . . . . . . . . . . . . . . . . . . . . 41
2.2 Multiplication table for $\mathbb{Z}/17\mathbb{Z}$. . . . . . . . . . . . . . . . . . . . . . . . 42
2.3 The multiplication table modulo 24 . . . . . . . . . . . . . . . . . . . . . . 46
2.4 Multiplication table for $\mathbb{Z}/17\mathbb{Z}$. . . . . . . . . . . . . . . . . . . . . . . . 48
2.5 Multiplication table modulo 6. . . . . . . . . . . . . . . . . . . . . . . . . . 48
2.6 Euler totient function . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

4.1 The multiplication table of a quotient ring. . . . . . . . . . . . . . . . . . . . 98

6.1 Multiplication in monoids . . . . . . . . . . . . . . . . . . . . . . . . . . . 135

8.1 Groups of order at most 11. . . . . . . . . . . . . . . . . . . . . . . . . . . 255

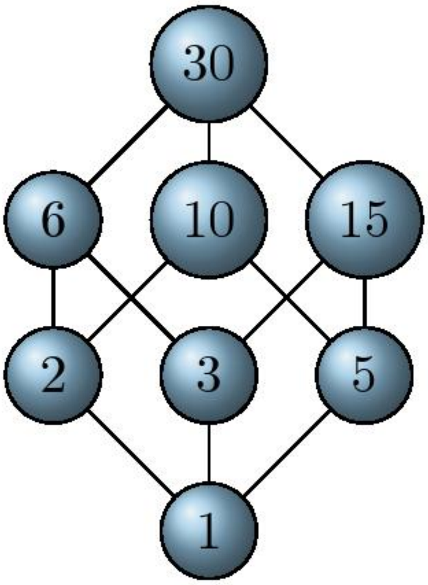
# Chapter 1

# Arithmetic

In this chapter we study properties of the set $\mathbb{Z}$ of integers. We mainly deal with its multiplicative structure and discuss notions such as the greatest common divisor (gcd) and the least common multiple (lcm) of two (or more) integers.

## 1.1 Divisors and Multiples

Let $\mathbb{Z}$ denote the set of integers. We know how to add integers, how to subtract them and how to multiply them. Division is a bit harder.



*A schematic representation of all positive divisors of* 30.

**Definition 1.1.1.** Let $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$.

- We call $b$ a *divisor* of $a$, if there is an integer $q$ such that $a = q \cdot b$.
- If $b$ is a nonzero divisor of $a$ then the (unique) integer $q$ with $a = q \cdot b$ is called the *quotient* of $a$ by $b$ and denoted by $\frac{a}{b}$, $a/b$, or $\mathrm{quot}(a,b)$.

If $b$ is a divisor of $a$, we also say that $b$ *divides* $a$, or $a$ is a *multiple* of $b$, or $a$ is *divisible* by $b$. We write this as $b|a$.

**Example 1.1.2.** If $a = 13$ and $b = 5$ then $b$ does not divide $a$. Indeed, if there were an integer $q$ such that $a = q \cdot b$, then $q$ should be between 2 and 3, so $q = 2$ or $q = 3$. But neither value of $q$ works. For instance, the former choice gives remainder 3 as $a = 2 \cdot b + 3$.

However, if $a = 15$ and $b = 5$ then $b$ does divide $a$, as $a = 3 \cdot b$. So, in the latter case, the quotient of $a$ by $b$ equals 3.

**Example 1.1.3.** For all integers $n$ we find $n - 1$ to be a divisor of $n^2 - 1$.

Indeed, $n^2 - 1 = (n+1) \cdot (n-1)$.

More generally, for all $m > 2$ we have $n^m - 1 = (n-1) \cdot \left(n^{m-1} + n^{m-2} + \cdots + 1\right)$. So, $n - 1$ divides $n^m - 1$.

**Example 1.1.4.** The *even* integers are simply the integers divisible by 2, such as 2, 6, and $-10$. Any even integer can be written in the form $2 \cdot m$ for some integer $m$.

The integers which are not divisible by 2, like 1 and $-7$, are usually called *odd*.

The following observations are straightforward, but very useful.

**Lemma 1.1.5.** *Suppose that a, b and c are integers.*

1. *If $a$ divides $b$, and $b$ divides $c$, then $a$ divides $c$.*
2. *If $a$ divides $b$ and $c$, then $a$ divides $x \cdot b + y \cdot c$ for all integers $x$ and $y$.*
3. *If $b$ is nonzero and if $a$ divides $b$, then $|a| \leq |b|$.*

*Proof.*

**Part (a).**

Suppose $a$ divides $b$, and $b$ divides $c$. Then there exist integers $u$ and $v$ such that $b = u \cdot a$ and $c = v \cdot b$. Consequently, $c = v \cdot (u \cdot a)$. Hence, $c = (v \cdot u) \cdot a$, and so $a$ divides $c$.

**Part (b).**

Suppose that $a$ divides $b$ and $c$. Then there exist integers $u$ and $v$ such that $b = u \cdot a$ and $c = v \cdot a$. So, for all integers $x$ and $y$, we have $x \cdot b + y \cdot c = x \cdot u \cdot a + y \cdot v \cdot a$. But this equals $(x \cdot u + y \cdot v) \cdot a$. Hence, $x \cdot b + y \cdot c$ is a multiple of $a$ for all integers $x$ and $y$.

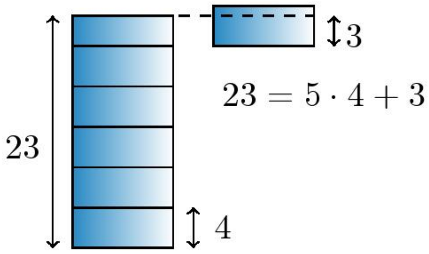
**Part (c).**

Since $a$ divides $b$, there exists an integer $q$ such that $q \cdot a = b$. As $b$ is nonzero, $q$ must be nonzero. From this equality we get $|q| \cdot |a| = |b|$. Since $|q| \geq 1$, we conclude that $|a| \leq |b|$.

□

Clearly, division is not always possible within the integers. Indeed, suppose you need to fit rods of length $b = 4$ one after the other in a box of length $a = 23$. Then you can fit 5 rods in the box, and there will be an open space of length 3. This is an example of *division with remainder*.

Here is a precise statement about division with remainder.



*A division with remainder.*

**Theorem 1.1.6** (Division with Remainder)**.** *If $a \in \mathbb{Z}$ and $b \in \mathbb{Z} \setminus \{0\}$, then there are unique integers $q$ and $r$ such that $a = q \cdot b + r$, $|r| < |b|$, and $a \cdot r \geq 0$.*

*Proof.* In the case where both $a$ and $b$ are positive, the proof is roughly as follows. Find the greatest multiple $q \cdot b$ of $b$ that is less than or equal to $a$; this can be accomplished by starting with $q = 0$ and increasing $q$ by 1 until $a - (q+1) \cdot b < 0$. Then $r = a - q \cdot b$.

A proof follows that proceeds by induction on $|a|$.

**The theorem holds if $|a| = 0$.**

Suppose $|a| = 0$. Then $a = 0$. Clearly, $q = 0$ and $r = 0$ is a solution. To show that this solution is unique, suppose that $q$ and $r$ represent a solution. Then $r = (-q) \cdot b$. If $q \neq 0$, then $|q| \geq 1$, so $|r| \geq |b|$, which contradicts the requirement $|r| < |b|$. Hence $q = 0$. It immediately follows that also $r = 0$. This establishes uniqueness of the solution.

**Existence of $q$ and $r$ for nonnegative $a$ and $b$.**

Suppose that $a$ and $b$ are nonnegative. If $a < b$, then we set $q = 0$ and $r = a$. If $a \geq b$, then $|a - b| < |a|$, so the induction hypothesis implies that there exist integers $q'$ and $r'$ (with $0 \leq r' < b$) such that $a - b = q' \cdot b + r'$. This rewrites to $a = (q' + 1) \cdot b + r'$. Now $q = q' + 1$ and $r = r'$ satisfy the requirements of the theorem.

**Existence of $q$ and $r$ for negative $a$ and positive $b$.**

If $a < 0$, then $-a > 0$, so by the above assertion there are $q'$ and $r'$ with $-a = q' \cdot b + r'$ with $r'$ non-negative and $|r'| < |b|$. But then $a = (-q') \cdot b + (-r')$ with $|-r'| < |b|$ and $a \cdot (-r') \geq 0$. So $q = -q'$ and $r = -r'$ satisfy the requirements of the theorem.

**Existence of $q$ and $r$ for negative $b$.**

If $b$ is negative, then applying one of the two previous assertions to $-a$ and $-b$ yields $q'$ and $r'$ with $-a = q' \cdot (-b) + r'$, where $r'$ satisfies $|r'| < -b$ and $(-a) \cdot r' \geq 0$. If we take $q = -q'$ and $r = -r'$ then $a = q \cdot b + r$ and $|r| < |b|$ and $a \cdot r \geq 0$ as required. We have shown the existence of both $q$ and $r$.

**Uniqueness of $q$ and $r$ for nonzero $a$.**

Suppose that $a = q \cdot b + r$ and $a = q' \cdot b + r'$ with both $|r|$ and $|r'|$ less than $|b|$ and satisfying $a \cdot r \geq 0$ and $a \cdot r' \geq 0$.

Suppose moreover that $r \geq r'$. This restriction is not essential as the roles of $r$ and $r'$ can be interchanged. By subtracting the two equalities we find $r - r' = (q' - q) \cdot b$. Now, since $a$ is nonzero, $r$ and $r'$ have the same sign. But then, as both $r$ and $r'$ are in absolute value less than $|b|$, we find that $r - r' < |b|$. It follows that the integral multiple $(q' - q) \cdot b$ of $b$ satisfies $(q' - q) \cdot b \in [0, |b|)$. This can only happen if $q' - q = 0$. In other words, $q = q'$. It also follows that $r = r'$.

□

**Example 1.1.7.** If $a = 23$ and $b = 7$, then division of $a$ by $b$ yields $23 = 3 \cdot 7 + 2$. So, the quotient of $a = 23$ by $b = 7$ equals 3 and the remainder is 2.

If $a = -23$ and $b = 7$, the quotient and remainder are $q = -3$ and $r = -2$, respectively.

Finally, if $a = -23$ and $b = -7$, the quotient and remainder are $q = 3$ and $r = -2$, respectively.

**Example 1.1.8.** For all integers $n$ greater than 2 the remainder of $n^2 + 1$ divided by $n + 1$ is 2. This follows immediately from the equality $n^2 + 1 = (n + 1) \cdot (n - 1) + 2$.

What is the remainder when $n$ is less than or equal to 2?

**Example 1.1.9.** An odd integer leaves remainder 1 or $-1$ upon division by 2, since these are the only two nonzero integers whose absolute value is less than 2. Any odd integer can therefore be written in the form $2 \cdot m + 1$ or $2 \cdot m - 1$ for some integer $m$. In particular, adding or subtracting 1 from an odd integer gives an even integer. Likewise, adding or subtracting 1 from an even integer produces an odd integer.

**Remark 1.1.10.** The definitions of quotient and remainder as given here are used in many programming languages and computer algebra packages, see for example Java or GAP. However, sometimes slightly different definitions are used. For example, in Mathematica the remainder $r$ of $a$ divided by $b$ is defined by the property that $a = q \cdot b + r$ for some integer $q$ where $|r| < |b|$ and $b \cdot r \geq 0$.

The integer $q$ of the theorem is called the *quotient* of $a$ divided by $b$. It is denoted by $\mathrm{quot}(a,b)$. The integer $r$ is called the *remainder* of $a$ divided by $b$ and will be denoted by $\mathrm{rem}(a,b)$.

The Division with Remainder Theorem states that there exist a quotient $q$ and a remainder $r$, but it does not tell you how to find those two integers. A standard and well-known algorithm is of course *long division*. We describe (a variation of) this algorithm for finding $q$ and $r$.

**Algorithm 1.1.11** (Division and Remainder)**.** • *Input: an integer a and a nonzero integer b.*

- *Output: the quotient q and remainder r of a upon division by b as a list* $[q,r]$.

```
DivisionRemainder := procedure$(a,b)$
local variables
    $q := 0$ , $r$, $x$
while $(q+1) \cdot |b| \leq |a|$ do
    $x := q$ , $q := x+1$
$r := |a| - q \cdot |b|$
if $(a \geq 0) \wedge (b > 0)$
    then
        return
            $[q,r]$
    else
        if $(a \geq 0) \wedge (b < 0)$
            then
                return
                    $[-q,r]$
            else
                if $(a < 0) \wedge (b > 0)$
                    then
                        return
                            $[-q,-r]$
                    else
                        return
                            $[q,-r]$
```

*Proof.*

**Correctness.**

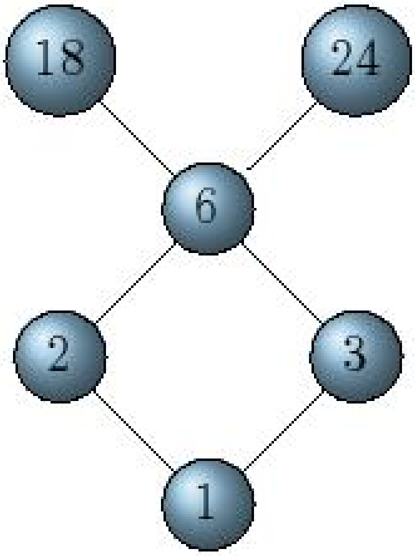
By construction we have $a = q \cdot b + r$. Moreover, as $|q| \cdot |b| \leq |a| < (|q|+1) \cdot |b|$ we find $|r| < |b|$. This proves correctness.

**Termination.**

Since $b$ is nonzero, the while loop will end. Thus the algorithm terminates.

□

For a better understanding of the relations between two or more integers, it is useful to consider the divisors and multiples they have in common.



*Positive common divisors of* 18 *and* 24.

**Definition 1.1.12.** Let $a$ and $b$ be integers.

- An integer $d$ is a *common divisor* of $a$ and $b$ if $d|a$ and $d|b$.
- If $a$ and $b$ are not both zero, the largest common divisor of $a$ and $b$ exists (see below) and is called *the greatest common divisor* of $a$ and $b$.

  We denote the greatest common divisor (gcd) of $a$ and $b$ by $\gcd(a,b)$.
- If the greatest common divisor of $a$ and $b$ equals 1, then $a$ and $b$ are called *relatively prime*.

**Example 1.1.13.** The positive divisors of $a = 24$ are 1, 2, 3, 4, 6, 8, 12, and 24. Those of $b = 15$ are 1, 3, 5, and 15. Hence, the common divisors of $a$ and $b$ are 1 and 3 and their negatives. So the greatest common divisor equals 3.

**Example 1.1.14.** The positive common divisors of $a = 24$ and $b = 16$ are 1, 2, 3, 4, and 8. Hence, the greatest common divisor of $a$ and $b$ equals 8.

**Example 1.1.15.** Suppose that $n > 1$ is an integer. Then any common divisor of $n+1$ and $n-1$ is also a divisor of $n+1-(n-1) = 2$. Hence $\gcd(n+1,n-1) = 2$ if $n$ is odd, and $\gcd(n+1,n-1) = 1$ if $n$ is even.

**Remark 1.1.16.** If $b$ divides $a$, then so does $-b$. For, if $a = q \cdot b$, then $a = (-q) \cdot (-b)$. In particular, any nonzero integer has positive divisors, so $\gcd(a,b) > 0$ if $a$ or $b$ is nonzero.

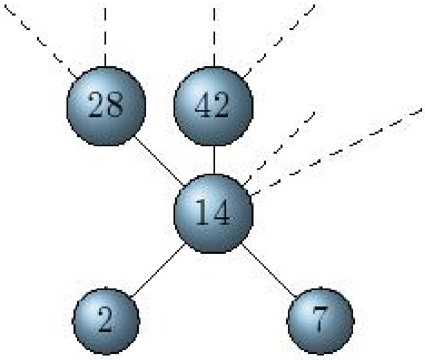
Since the divisors of $a$ coincide with those of $|a|$, we have $\gcd(a,b) = \gcd(|a|,|b|)$.

If $a$ and $b$ are not both 0, their greatest common divisor exists. To see this, first note that the set of common divisors of $a$ and $b$ is certainly bounded above by the largest of $|a|$ and $|b|$ by Properties of Divisors. Since the set is nonempty (1 is in it), it must have a largest element.

For the sake of completeness, we define the greatest common divisor of 0 and 0 to be 0.

The greatest common divisor of more than two integers is defined analogously.

Just like studying common divisors of two integers, we can also consider common multiples of two (or more) integers.



*Some positive common multiples of* 2 *and* 7*.*

**Definition 1.1.17.** Let $a$ and $b$ be nonzero integers.

- The integer $c$ is a *common multiple* of $a$ and $b$ if $c$ is a multiple of $a$ and of $b$ (that is, $a|c$ and $b|c$).
- The smallest positive common multiple of $a$ and $b$ is called the *least common multiple* of $a$ and $b$.

We denote the least common multiple (lcm) of $a$ and $b$ by $\operatorname{lcm}(a,b)$.

**Example 1.1.18.** The first 5 positive multiples of $a = 13$ are 13, 26, 39, 52, and 65.

The first 13 multiples of $b = 5$ are 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, and 65.

So, the only positive common multiple of $a = 13$ and $b = 5$ less than or equal to $a \cdot b$ is 65.

In particular, $\operatorname{lcm}(13,5) = 65$.

For any two nonzero integers $a$ and $b$ there exists a positive common multiple, namely $|a \cdot b|$. As a consequence, the least common multiple of $a$ and $b$ is well defined.

Of course, the least common multiple of more than two integers can be defined in a similar way.

The least common multiple and the greatest common divisor of two integers are closely related.

**Theorem 1.1.19** (Relation between ggd and lcm)**.** *Let $a$ and $b$ be positive integers. Then $a \cdot b = \gcd(a,b) \cdot \operatorname{lcm}(a,b)$.*

*Proof.* Our strategy is to apply division with remainder to $a \cdot b$ and $\operatorname{lcm}(a,b)$, and relate the quotient to $\gcd(a,b)$. Let $q$ be the quotient and let $r$ be the remainder of this division.

First we investigate the remainder $r$. We rewrite $a \cdot b = q \cdot \operatorname{lcm}(a,b) + r$ as $r = a \cdot b - q \cdot \operatorname{lcm}(a,b)$

Since both $a \cdot b$ and $\operatorname{lcm}(a,b)$ are divisible by $a$ and $b$, we infer that the remainder $r$ is also divisible by $a$ and $b$. In other words, $r$ is a common multiple of $a$ and $b$. But $r < \operatorname{lcm}(a,b)$ by the Division with Remainder Theorem, so $r = 0$. Consequently, $a \cdot b = q \cdot \operatorname{lcm}(a,b)$.

Next, we claim that $q$ divides $a$ and $b$. To see this, first let $u$ be such that $\operatorname{lcm}(a,b) = u \cdot b$. Multiplying both sides by $q$ gives $a \cdot b = q \cdot u \cdot b$. As $b$ is nonzero, this equality can be simplified to $a = q \cdot u$, which proves the claim that $q$ divides $a$. The proof that $q$ divides $b$ is entirely similar.

So $q$ is a common divisor of $a$ and $b$. In particular, $q$ is less than or equal to $\gcd(a,b)$.

Finally, we show that $q$ is also greater than or equal to $\gcd(a,b)$.

Since $\gcd(a,b)$ divides both $a$ and $b$, $(a \cdot b)/\gcd(a,b)$ is also a common multiple of $a$ and $b$. As $(a \cdot b)/q$ is the least common multiple of $a$ and $b$, we conclude that $q$ is greater than or equal to $\gcd(a,b)$. Hence $q$ equals $\gcd(a,b)$, which proves the theorem as $a \cdot b = q \cdot \operatorname{lcm}(a,b)$.

□

The above theorem enables us to compute the lcm of two integers from the gcd and vice versa.

**Example 1.1.20.** For $a = 24$ and $b = 15$, we find $\gcd(a,b) = 3$, $\operatorname{lcm}(a,b) = 120$ and $a \cdot b = 360$. We see that $3 \cdot 120 = 360$.

**Example 1.1.21.** Suppose that $n > 1$ is an integer. Then, as we have seen in Example 1.1.15, $\gcd(n+1, n-1) = 2$ if $n$ is odd, and $\gcd(n+1, n-1) = 1$ if $n$ is even. So, $\operatorname{lcm}(n+1, n-1) = \frac{(n+1) \cdot (n-1)}{2}$ if $n$ is odd, and $\operatorname{lcm}(n+1, n-1) = (n+1) \cdot (n-1)$ if $n$ is even.

## 1.2 Euclid's algorithm

The greatest common divisor of two integers $a$ and $b$ can be determined by Euclid's Algorithm, one of the most important algorithms we will encounter. It is based on the observation that, if $a = q \cdot b + r$, then $\gcd(a,b)$ is equal to $\gcd(b,r)$, see Properties of Divisors, where $q = \mathrm{quot}(a,b)$ and $r = \mathrm{rem}(a,b)$.

For simplicity, we will assume the arguments of gcd to be positive. This does not really restrict us when we bear in mind that the arguments of gcd can be replaced by their absolutes in view of .

*Euclid of Alexandria (about 325 BC-265 BC).*

**Algorithm 1.2.1** (Euclid's Algorithm)**.** • *Input: two positive integers a and b.*

- *Output: the gcd of a and b.*

```
GCD := procedure($a,b$)
local variables
    | $c$
while $b > 0$ do
    | $c := a$ , $a := b$ , $b := \mathrm{rem}(c,b)$
return
    | $a$
```

*Proof.* We use three properties of the greatest common divisor of nonnegative integers that follow from Properties of Divisors:

$$\gcd(a,b) = \gcd(b,a) \tag{1.1}$$

$$\gcd(a,b) = \gcd(a, b - k \cdot a) \tag{1.2}$$

(for every integer $k$), and

$$\gcd(a,0) = a \tag{1.3}$$

**Correctness.**

If $a'$ and $b'$ denote the values of $a$ and $b$, respectively, at the end of the body of the while loop, then $a' = b'$ and $b' = a - q \cdot b$, where $q$ is the quotient of division with remainder of $a$ by $b$. By the first two of the three properties, the greatest common divisor is an invariant, that is, $\gcd(a', b') = \gcd(a,b)$. As a consequence, the value of $\gcd(a', b')$ remains unaffected upon changing the arguments. At the end of the while loop, $b' = 0$, so the third property gives that the output $a$ is equal to the initial value of $\gcd(a', b')$.

**Termination.**

The variable $b$ decreases with each step. (By a step we mean a percursion of the full body of the while loop.) After at most $b$ steps we arrive at the point where $b$ equals 0. Then the algorithm ends.

□

**Remark 1.2.2.** The while loop in Euclid's Algorithm can be described rather conveniently in matrix form. Let $q$ be the quotient of division of $a$ by $b$. Then the vector $(a,b)^T$ is replaced by $(b, a - q \cdot b)^T$. We can also write this as the product of the matrix $M$ and the vector $(a,b)^T$, where $M = \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix}$

**Example 1.2.3.** Euclid's Algorithm computes the greatest common divisor of two positive integers. In this example, you can see all the steps of the algorithm.

We compute the greatest common divisor of $a = 123$ and $b = 13$.

In each step of the algorithm we replace (simultaneously) $a$ by $b$, and $b$ by the remainder of $a$ divided by $b$.

The algorithm starts with $a = 123$ and $b = 13$.

Each row of the following table represents a step in the algorithm.

| Step $n$ | $a$ | $b$ |
|---|---|---|
| 0 | 123 | 13 |
| 1 | 13 | 6 |
| 2 | 6 | 1 |
| 3 | 1 | 0 |

Since the value of the second parameter has become 0, the algorithm stops.

We conclude that the greatest common divisor of $a = 123$ and $b = 13$ equals 1.

**Example 1.2.4.** In this example, we compute the greatest common divisor of $a = 56$ and $b = 36$.

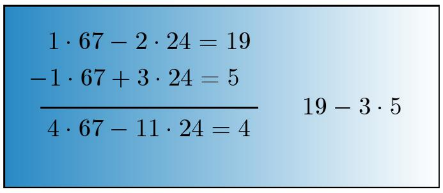
In the following table you find the values of $a$ and $b$ in each step of Euclid's Algorithm.

| Step $n$ | $a$ | $b$ |
|---|---|---|
| 0 | 56 | 36 |
| 1 | 36 | 20 |
| 2 | 20 | 16 |
| 3 | 16 | 4 |
| 4 | 4 | 0 |

Since the value of the second parameter has become 0, the algorithm stops.

We conclude that the greatest common divisor of $a = 56$ and $b = 36$ equals 4.

There is also an extended version of Euclid's Algorithm, which determines, apart from $\gcd(a,b)$, integers $x$ and $y$ such that $a \cdot x + b \cdot y = \gcd(a,b)$. We say that $\gcd(a,b)$ can be expressed as an *integral linear combination* of $a$ and $b$. To find such an integral linear combination for $\gcd(a,b)$, we record at each step of Euclid's Algorithm how to express the intermediate results in the input integers.



$$\begin{array}{r} 1 \cdot 67 - 2 \cdot 24 = 19 \\ -1 \cdot 67 + 3 \cdot 24 = 5 \\ \hline 4 \cdot 67 - 11 \cdot 24 = 4 \end{array} \qquad 19 - 3 \cdot 5$$

*One step in the Extended Euclidean Algorithm applied to* 67 *and* 24*. Using the expressions for the intermediate results* 19 *and* 5*, the next occurring integer,* 4*, can also be expressed in the input values.*

**Algorithm 1.2.5** (Extended Euclidean Algorithm)**.** • *Input: positive integers a and b.*

- *Output: list of integers* $[g,x,y]$ *with* $g = \gcd(a,b)$*, and* $g = x \cdot a + y \cdot b$*.*

```
ExtendedGCD := procedure(a, b)
local variables
    | $a_1$, $b_1$
    | $u := 0$ , $v := 1$ , $x := 1$ , $y := 0$
    | $u_1$, $v_1$, $x_1$, $y_1$
while $b > 0$ do
    | $a_1 := a$ , $b_1 := b$
    | $u_1 := u$ , $v_1 := v$ , $x_1 := x$ , $y_1 := y$
    | $a := b_1$ , $b := \mathrm{rem}(a_1, b_1)$
    | $x := u_1$ , $y := v_1$
    | $u := x_1 - \mathrm{quot}(a_1, b_1) \cdot u_1$ , $y := y_1 - \mathrm{quot}(a_1, b_1) \cdot v_1$
return
    | $[a, x, y]$
```

*Proof.*

**Correctness.**

Find the gcd of $a$ and $b$ using Euclid's Algorithm. In each step of the while-loop of the algorithm the two input values are changed into two new values. These values can be defined recursively by $a_0 = a$ and $b_0 = b$ and for $n \geq 1$ by $a_{n+1} = b_n$ and $b_{n+1} = a_n - \mathrm{quot}(a_n, b_n) \cdot b_n$.

We prove by induction on $n$ that every $a_n$ and $b_n$ can be written as a linear combination of $a$ and $b$ with integer coefficients.

For $n = 0$ this is trivial.

Suppose for some $n$ we have $a_n = x \cdot a + y \cdot b$ and $b_n = u \cdot a + v \cdot b$ for certain integers $x$, $y$, $u$, and $v$. Then after the next step we obtain $a_{n+1} = b_n$ which equals $u \cdot a + v \cdot b$. Thus also $a_{n+1}$ is a linear combination of $a$ and $b$ with integer coefficients.

Furthermore we have $b_{n+1} = a_n - q \cdot b_n$. So, $b_{n+1} = x \cdot a + y \cdot b - q \cdot (u \cdot a + v \cdot b) = (x - q \cdot u) \cdot a + (y - q \cdot v) \cdot b$, where $q = \mathrm{quot}(a_n, b_n)$. In particular, also $b_{n+1}$ is a linear combination of $a$ and $b$ with integer coefficients.

By induction we have proven for all $n$ that $a_n$ and $b_n$ can be written as a linear combination of $a$ and $b$ with integer coefficients.

Since Euclid's algorithm will eventually return the gcd of $a$ and $b$ as $a_n$ for some $n$, the extended Euclidean algorithm will output integers $x$ and $y$ with $\gcd(a, b) = x \cdot a + y \cdot b$.

**Termination.**

As Euclid's Algorithm terminates, so does the extended Euclidean algorithm.

□

**Remark 1.2.6.** Integers $x$ and $y$ satisfying $x \cdot a + y \cdot b = \gcd(a, b)$ are not unique, since, for any integer $t$, we have $(x + t \cdot b) \cdot a + (y - t \cdot a) \cdot b = \gcd(a, b)$.

**Remark 1.2.7.** In terms of matrices, the algorithm can be written somewhat more succinctly. The idea is that in each step the values of the variables are such that the matrix $M = \begin{pmatrix} x & y \\ u & v \end{pmatrix}$ applied to the column vector $\begin{pmatrix} a \\ b \end{pmatrix}$ (the input values) gives the updated values of $a$ and $b$.

At the end, we obtain $\begin{pmatrix} \gcd(a, b) \\ 0 \end{pmatrix} = M \cdot \begin{pmatrix} a \\ b \end{pmatrix}$, with the appropriate matrix $M$. Comparing the first and second entries on both sides of this equality gives $\gcd(a, b) = x \cdot a + y \cdot b$ and $0 = u \cdot a + v \cdot b$, where $x$, $y$, $u$, and $v$ are the suitably updated entries of the matrix $M$.

**Example 1.2.8.** The extended Euclidean algorithm computes the greatest common divisor of two positive integers and expresses it as an integral linear combination of the input. In this example, you can see all the steps of the algorithm.

We compute the greatest common divisor of $a = 123$ and $b = 13$ following the extended Euclidean algorithm.

| Step $n$ | $a$ | $b$ | $x$ | $y$ | $u$ | $v$ |
|---|---|---|---|---|---|---|
| 0 | 123 | 13 | 1 | 0 | 0 | 1 |
| 1 | 13 | 6 | 0 | 1 | 1 | $-9$ |
| 2 | 6 | 1 | 1 | $-9$ | $-2$ | 19 |
| 3 | 1 | 0 | $-2$ | 19 | 13 | $-123$ |

Each row of the following table represents a step in the algorithm.

We conclude that the greatest common divisor of $a = 123$ and $b = 13$ equals 1. From the same table we infer that 1 can be written as $1 = (-2)\cdot 123 + 19\cdot 13$.

The Extended Euclidean Algorithm provides us with the following characterization of the gcd.

**Theorem 1.2.9** (Characterization of the gcd)**.** *The following three statements concerning the positive integers a, b, and d are equivalent.*

1. $\gcd(a,b) = d$.
2. *The integer d is a positive common divisor of a and b such that any common divisor of a and b is a divisor of d.*
3. *d is the least positive integer that can be expressed as $x\cdot a + y\cdot b$ with integers x and y.*

*Proof.*

**The second statement is equivalent to the first.**

To show that the first assertion implies the second, let $d = \gcd(a,b)$. Then $d$ is a common divisor of $a$ and $b$. By the Extended Euclidean Algorithm we have $d = x\cdot a + y\cdot b$ for some integers $x$ and $y$. If $c$ is any common divisor of $a$ and $b$, then it also divides $x\cdot a + y\cdot b = d$, see Properties of Divisors. This proves that the first assertion implies the second.

As for the other way around, suppose that $d$ is as in the second statement. Since $\gcd(a,b)$ is a common divisor of $a$ and $b$ it must divide $d$. On the other hand $d$ cannot be greater than $\gcd(a,b)$. Hence $d$ and $\gcd(a,b)$ must be equal. This proves that the second statement implies the first.

**The third statement is equivalent to the first.**

Let $d = \gcd(a,b)$ and let $e$ be the least positive integer that can be expressed as $x\cdot a + y\cdot b$ with integers $x$ and $y$. We show that $d = e$. Since $d$ is a common divisor of $a$ and $b$ the equality $e = x\cdot a + y\cdot b$ implies that $d$ divides $e$ (see Properties of Divisors). So $d \leq e$. Moreover, as a result of the Extended Euclidean Algorithm, $d$ itself can also be written as an integral linear combination of $a$ and $b$. So $d \geq e$ by the defining property of $e$. Hence $e$ must be equal to $d$. This proves the equivalence.

**Conclusion.**

Since both the second and the third statement of the theorem are equivalent to the first, all three statements are equivalent. This finishes the proof of the theorem.

□

These different characterizations of the gcd, in particular the possibility to express the gcd of two integers $a$ and $b$ as an integral linear combination of $a$ and $b$, will turn out to be very useful in various applications.

The following corollary to the Characterization of the gcd deserves to be stated separately.

**Corollary 1.2.10** (Characterization of Relatively Prime Numbers)**.** *Integers $a$ and $b$ are relatively prime if and only if there exist integers $x$ and $y$ such that $x \cdot a + y \cdot b = 1$.*

*Proof.* Apply the previous Characterization of the gcd with $d = 1$.

□

**Example 1.2.11.** For all natural numbers $m$, $n$, and $k$ with $m < n$, the integers $k^m$ and $k^n - 1$ are relatively prime. For, $k^{n-m} \cdot k^m - 1 \cdot (k^n - 1) = 1$.

**Example 1.2.12.** Suppose that $n$ is a positive integer. Then the greatest common divisor of $n^2 + n + 1$ and $n^2$ equals 1. Indeed, this follows from the equality $n \cdot n^2 - (n-1) \cdot \left(n^2 + n + 1\right) = 1$

A first application of the Characterization of the gcd is the following useful result for deducing divisibility of one integer by another.

**Proposition 1.2.13.** *Let $a$, $b$, and $c$ be integers. If $a$ and $b$ are relatively prime, then $a | b \cdot c$ implies $a | c$.*

*Proof.* Since the gcd of $a$ and $b$ equals 1, Characterization of Relatively Prime Numbers implies that there exist integers $x$ and $y$ such that $x \cdot a + y \cdot b = 1$. Multiplying both sides of this equation by $c$ yields that $x \cdot a \cdot c + y \cdot b \cdot c = c$. Since $a | x \cdot a \cdot c$ and $a | b \cdot c$ (and hence also $a | y \cdot b \cdot c$) we conclude that $a | ((x \cdot a \cdot c) + (y \cdot b \cdot c)) = c$, which proves the proposition.

□

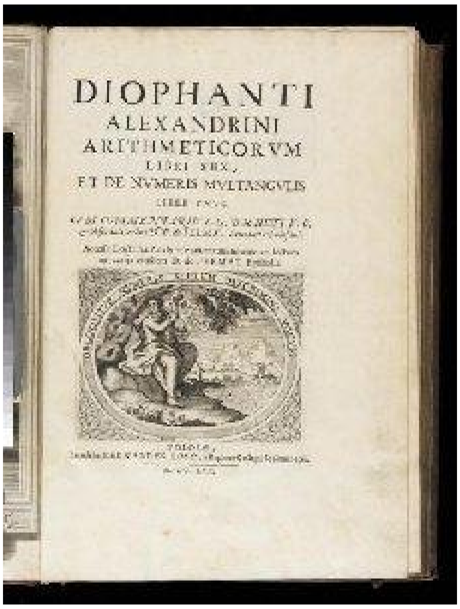
**Example 1.2.14.** The above proposition is a generalization of the following well known statement: The product of two integers is even if and only if at least one of the two integers is even.

## 1.3 Linear Diophantine equations

Let $a$, $b$, and $c$ be integers. A linear equation in the unknowns $x$ and $y$ is an equation of the form $x \cdot a + y \cdot b = c$. If the unknowns $x$ and $y$ are integers, such equations are known as *linear Diophantine equations*.

We will use the Extended Euclidean Algorithm to derive an algorithm for finding all integer pairs $x$, $y$ that satisfy the linear Diophantine equation $x \cdot a + y \cdot b = c$, for given integers $a$, $b$, and $c$.

If we interpret the equation over $\mathbb{Q}$ or $\mathbb{R}$ and if we assume that $b$ is not equal to $0$, then the solutions are all of the form $(x, y) = (x, (c - x \cdot a)/b)$. However, not all of these solutions are integral, and we have to find out which ones are.



DIOPHANTI
ALEXANDRINI
ARITHMETICORVM
LIBRI SEX,
ET DE NVMERIS MVLTANGVLIS
LIBER VNVS.
[illegible]
TOLOSÆ,
[illegible]

*Diophantus' book on Arithmetic. Diophantus' work inspired Fermat to write in the margin of this book his famous last theorem: for $n > 2$ there are no nonzero integers $x$, $y$ and $z$, such that $x^n + y^n = z^n$.*

We first discuss a special case, the *homogeneous equation*, i.e., the case where $c$ equals $0$.

**Lemma 1.3.1.** *If $x \cdot a + y \cdot b = 0$ and $\gcd(a, b) = 1$, then there exists an integer $n$ such that $x = -n \cdot b$ and $y = n \cdot a$.*

*Proof.* Suppose that $x \cdot a + y \cdot b = 0$ and that $\gcd(a, b) = 1$. From $x \cdot a = -b \cdot y$ it follows that $a | b \cdot y$. Since $\gcd(a, b) = 1$, we find $a | y$, see Result on the divisor of a product. This means that there exists an integer $n$ such that $a \cdot n = y$. Substitution of $y$ in the original equation gives $x = -n \cdot b$. This proves the lemma.

□

From Lemma on Diophantine Equation Solving we conclude the following.

**Lemma 1.3.2** (Homogeneous Diophantine Equation Solving)**.** *Suppose that $a$ and $b$ are integers which are not both equal to* 0*. Then the integer solutions to the equation $x \cdot a + y \cdot b = 0$ are given by $x = \frac{-n \cdot b}{d}$ and $y = \frac{n \cdot a}{d}$, where $d = \gcd(a, b)$ and $n \in \mathbb{Z}$.*

*Proof.* First we note that the integers $\frac{a}{d}$ and $\frac{b}{d}$ are relatively prime: Use the Extended Euclidean Algorithm to find a relation of the form $u \cdot a + v \cdot b = d$, divide both sides by $d$, and, finally, apply the Characterization of Relatively Prime Numbers.

Next, we turn to the equation $x \cdot a + y \cdot b = 0$. After dividing both sides of the equation $x \cdot a + y \cdot b = 0$ by $d$, we arrive at the setting of Lemma on Diophantine Equation Solving. Our equation then reads $x \cdot \frac{a}{d} + y \cdot \frac{b}{d} = 0$, where $\gcd\left(\frac{a}{d}, \frac{b}{d}\right) = 1$. Lemma on Diophantine Equation Solving now shows that there exists an integer $n$ such that $x = -n \cdot \frac{b}{d}$ and $y = n \cdot \frac{a}{d}$, as required.

□

**Example 1.3.3.** To find the integral solutions to the equation $24 \cdot x + 15 \cdot y = 0$ we first compute the gcd of 24 and 15. Using for example the Euclid's Algorithm as in Example 1.2.3, we find $\gcd(24, 15) = 3$ By Homogeneous Diophantine Equation Solving, $x = \frac{15 \cdot n}{3} = 5 \cdot n$ and $y = -(\frac{24 \cdot n}{3}) = (-8) \cdot n$ with $n \in \mathbb{Z}$.

We are now ready to solve general linear Diophantine equations of the form $x \cdot a + y \cdot b = c$. We do this in the form of an algorithm.

**Algorithm 1.3.4** (Linear Diophantine Equation Solving Algorithm)**.** - *Input: integers $a$, $b$, and $c$, with $a$ and $b$ not both equal to* 0 *.*

- *Output: set of all integer solutions $(x, y)$ to the Diophantine equation $x \cdot a + y \cdot b = c$.*

```
SolveDiophantine := **procedure**$(a, b, c)$
**local variables**
    $e :=$ extended $-$ gcd $(a, b)$
    $g := e[1]$
    $x_0 := e[2]$
    $y_0 := e[3]$
**if** $g|c$
    **then**
        **return**
            $\left\{\left(\frac{x_0 \cdot c - n \cdot b}{g}, \frac{y_0 \cdot c + n \cdot a}{g}\right) \middle| n \in \mathbb{Z}\right\}$
    **else**
        **return**
            $\varnothing$
```

*Proof.*

**Termination.**

As there are no loops in the algorithm, this is obvious....provided we interpret the returned output set as finite data (instead of returning elements of the set one by one).

**Correctness.**

By definition of the extended gcd algorithm, the value of the variable $g$ is equal to $\gcd(a, b)$.

If there are solutions to the equation $x \cdot a + y \cdot b = c$, then $g$ divides $c$. Indeed, for all integer solutions $x$ and $y$, the integer $g$ divides $x \cdot a + y \cdot b$, which is equal to $c$.

So, suppose that $g$ divides $c$. If $x_0 \cdot a + y_0 \cdot b = g$, then $\frac{c}{g} \cdot x_0 \cdot a + \frac{c}{g} \cdot y_0 \cdot b = c$. So $x_1 = \frac{c}{g} \cdot x_0$ and $y_1 = \frac{c}{g} \cdot y_0$ form a solution to the equation.

If $(x_2, y_2)$ is another solution to the equation $a \cdot x + y \cdot b = c$, then the differences $x_2 - x_1$ and $y_2 - y_1$ form a solution to the so-called homogeneous equation $a \cdot x + y \cdot b = 0$. Hence all solutions of $a \cdot x + y \cdot b = c$, if there are any, are of the form $(x_1, y_1)$ plus a single solution to the homogeneous equation $a \cdot x + y \cdot b = 0$.

From Homogeneous Diophantine Equation Solving we conclude that every solution is of the form $x = \frac{x_0 \cdot c - n \cdot b}{g}$ and $y = \frac{y_0 \cdot c - n \cdot a}{g}$, which proves correctedness of the algorithm.

□

**Example 1.3.5.** Let $a$, $b$, and $c$ be integers. We determine the integral solutions to the equation $24 \cdot x + 15 \cdot y = 63$

Following the Linear Diophantine Equation Solving Algorithm, we use the Extended Euclidean Algorithm to compute the gcd of 24 and 15 and express it as a linear combination of these numbers. We find $\gcd(24, 15) = 3 = 2 \cdot 24 - 3 \cdot 15$. As 3 divides 63, there are solutions.

By the Linear Diophantine Equation Solving Algorithm the general solution to the equation $24 \cdot x + 15 \cdot y = 63$ is now $x = \frac{2 \cdot 63 - n \cdot 15}{3}$ and $y = \frac{(-3) \cdot 63 + n \cdot 24}{3}$, where $n$ runs through $\mathbb{Z}$.

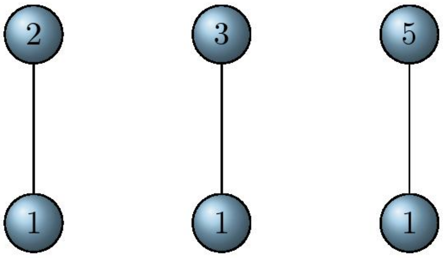
This solution simplifies to $x = 42 - 5 \cdot n$ and $y = -63 + 8 \cdot n$, with $n$ running through $\mathbb{Z}$, the sum of a particular solution and any solution of the homogeneous equation.

Of course, the particular solution $x = 42$ and $y = -63$ could have been found by multiplying both sides of the equation $3 = 2 \cdot 24 - 3 \cdot 15$ by 21.

Note the structure of the solutions in the Linear Diophantine Equation Solving Algorithm: $\left(\frac{x_0 \cdot c}{\gcd(a,b)}, \frac{y_0 \cdot c}{\gcd(a,b)}\right)$ is one particular solution to the equation $x \cdot a + y \cdot b = c$, and all other solutions are obtained by adding all solutions $(x', y')$ of the homogeneous equation $x' \cdot a + y' \cdot b = 0$ to it.

## 1.4 Prime numbers

In this section we discuss prime numbers, the building blocks for the multiplicative structure of the integers. We start with a definition of primes.



*A prime has only 'trivial' divisors.*

**Definition 1.4.1.** A *prime* is an integer $p$ greater than 1 that has no positive divisors other than 1 and $p$ itself.

**Example 1.4.2.** The integer 17 is prime.

The integer 51 is not prime, since it is divisible by 3.

**Example 1.4.3.** Suppose that $n$ is a positive integer such that $2^n - 1$ is prime. Then $n$ itself is prime.

Indeed, if $n$ is the product of two integers $a$ and $b$ (both at least 2), then $2^n - 1 = (2^a)^b - 1$, which is divisible by $2^a - 1$.

The smallest prime number is 2 (and not 1). The first five primes are $2, 3, 5, 7$, and 11, but there are many more.

**Theorem 1.4.4** (Euclid's Theorem)**.** *There are infinitely many primes.*

*Proof.* Suppose that there are only finitely many primes, say $p_1, ..., p_n$, and no others. We will derive a contradiction by showing that there must exist at least one other prime, distinct from all the $p_i$.

Consider the integer $m = 1 + \prod_{i=1}^{n} p_i$ Then $m > 1$. Moreover, for each $i \in \{1, ..., n\}$, the integer $m$ is clearly not divisible by $p_i$. Hence, the smallest divisor $p$ of $m$ greater than 1 is distinct from $p_1, ..., p_n$.

We claim that $p$ is prime. Indeed, any positive divisor $d$ of $p$ is also a divisor of $m$. So, since $p$ is the smallest divisor of $m$ greater than 1, we find $d$ to be equal to either 1 or $p$, which proves our claim. So, we have found a prime $p$ distinct from all the primes $p_1, ...., p_n$. This contradicts the assumption that $p_1, ...., p_n$ are the only primes.

□

**Example 1.4.5.** The primes less than or equal to 1013 are

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 5 | 7 | 11 | 13 | 17 | 19 | 23 | 29 |
| 31 | 37 | 41 | 43 | 47 | 53 | 59 | 61 | 67 | 71 |
| 73 | 79 | 83 | 89 | 97 | 101 | 103 | 107 | 109 | 113 |
| 127 | 131 | 137 | 139 | 149 | 151 | 157 | 163 | 167 | 173 |
| 179 | 181 | 191 | 193 | 197 | 199 | 211 | 223 | 227 | 229 |
| 233 | 239 | 241 | 251 | 257 | 263 | 269 | 271 | 277 | 281 |
| 283 | 293 | 307 | 311 | 313 | 317 | 331 | 337 | 347 | 349 |
| 353 | 359 | 367 | 373 | 379 | 383 | 389 | 397 | 401 | 409 |
| 419 | 421 | 431 | 433 | 439 | 443 | 449 | 457 | 461 | 463 |
| 467 | 479 | 487 | 491 | 499 | 503 | 509 | 521 | 523 | 541 |
| 547 | 557 | 563 | 569 | 571 | 577 | 587 | 593 | 599 | 601 |
| 607 | 613 | 617 | 619 | 631 | 641 | 643 | 647 | 653 | 659 |
| 661 | 673 | 677 | 683 | 691 | 701 | 709 | 719 | 727 | 733 |
| 739 | 743 | 751 | 757 | 761 | 769 | 773 | 787 | 797 | 809 |
| 811 | 821 | 823 | 827 | 829 | 839 | 853 | 857 | 859 | 863 |
| 877 | 881 | 883 | 887 | 907 | 911 | 919 | 929 | 937 | 941 |
| 947 | 953 | 967 | 971 | 977 | 983 | 991 | 997 | 1009 | 1013 |

Table 1.1: The primes less than or equal to 1013.

**Example 1.4.6.** Although there are infinitely many prime numbers, see Euclid's Theorem, the gaps between two consecutive prime numbers can be arbitrarily large.

For example, none of the hundred consecutive integers between $101! + 2$ and $101! + 101$ is prime. A nontrivial divisor (i.e., a divisor greater than 1 and less than the number itself) of $101! + n$, where $n \in \{2, ..., 101\}$, is $n$.

**Example 1.4.7.** Suppose that $L$ is a finite list of primes, for example $[2, 3, 5, 7, 11, 13, 17]$. Put $m = 1 + \prod_{i \in L} i$. According to the proof of the theorem, a new prime occurs among the divisors of $m$, which equals 510511.

The smallest nontrivial positive divisor of 510511 equals 19, a prime not in $L$.

**Remark 1.4.8.** Although there are infinitely many prime numbers, we actually know only a finite number of them. The largest known prime, as of December 2005, is $2^{30402457}-1$. In its decimal representation this number is 9,152,052 digits long. It was found on December 15, 2005, by Curtis Cooper and Steven Boone, two members of a collaborative effort to find primes known as GIMPS. Before finding the prime, Cooper and Boone ran the GIMPS program for 9 years. The GIMPS program searches for so-called Mersenne primes.

*Mersenne primes* are primes of the form $2^n-1$. The prime number $2^{30402457}-1$ is the 43rd known Mersenne prime.

Prime numbers of the form $2^n-1$ are called Mersenne primes, since they were studied first by Marin Mersenne (1588-1648).

*Marin Mersenne (1588-1648).*

By Example 1.4.3, the integer $2^n-1$ can be prime only when $n$ itself is a prime.

A few examples of Mersenne primes are $3=2^2-1$, $7=2^3-1$, $31=2^5-1$ and $127=2^7-1$. Mersenne found that $2^{11}-1$ is not a prime. Can you find its prime divisors?

Eratosthenes' sieve is an algorithm for making the list of all primes less than or equal to some integer $n$.

*Eratosthenes (about 276 BC-194 BC).*

If $M$ is a list of integers and $m$ is an integer, we shall write $M\cup[m]$ for the list obtained by appending $m$ to $M$.

**Algorithm 1.4.9** (Eratosthenes' Sieve)**.** • *Input: a positive integer n.*

- *Output: the list of primes less than or equal to n.*

```
Sieve := **procedure**$(n)$
**local variables**
    $L := \{2,...,n\}$
    $M :=$ list2.nil
    $m$
**while** $L \neq$ list2.nil **do**
    $m := L[1]$ , $L := L \backslash m \cdot \{1,...,n\}$ , $M := M \cup [m]$
**return**
    $M$
```

*Proof.*

**Termination.**

At each step (that is, percursion of the body of the while loop), the length of the list $L$ strictly decreases, so the algorithm will stop after running the while loop at most the length of $L$ times.

**Correctness.**

By construction, the output list $M$ consists of all numbers in $\{2,...,n\}$ that are no multiple of a strictly smaller number. These are precisely the primes less than or equal to $n$.

□

**Example 1.4.10.** We will make a list of all the primes in the interval from 2 to $n = 20$. We use Eratosthenes' Sieve. We start with the complete list of integers from 2 to $n = 20$. See the first row of the table below. Next, in each consecutive row, we remove the proper multiples of the first element for which this has not yet been done.

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 2 | 3 | | 5 | | 7 | | 9 | | 11 | | 13 | | 15 | | 17 | | 19 | |
| 2 | 3 | | 5 | | 7 | | | | 11 | | 13 | | | | 17 | | 19 | |
| 2 | 3 | | 5 | | 7 | | | | 11 | | 13 | | | | 17 | | 19 | |

Table 1.2: Eratosthenes' sieve

We have removed multiples of 2, 3 and 5, respectively.

The numbers in the last row of the table are all prime. They form the set of all primes less than or equal to 20.

**Remark 1.4.11.** The number of runs of the while loop in Eratosthenes' Sieve equals the number of primes in the interval $\{1,...,n\}$. In each run, one has to check less than $n$ integers. So the algorithm takes certainly less than $n^2$ operations. However, the memory use for the algorithm is quite big, as the whole range of numbers from 2 to $n$ has to be in memory at the start of the algorithm.

**Remark 1.4.12.** Eratosthenes' Sieve can also be used as a prime test. However, to avoid problems of big memory use as indicated in Remark on the Running time of Eratosthenes' sieve, one can apply the following straightforward algorithm for verifying if the integer $n$ is prime. Let an integer variable $m$ run from 2 up to $\sqrt{n}$ and check whether $n$ is divisible by $m$. If for some $m$ we find that it divides $n$, then we stop and decide that $n$ is composite, otherwise we decide that $n$ is prime.

Using Eratosthenes' sieve we can find all the primes in the interval $\{1,...,n\}$. The number of such primes can be approximated as follows.

**Theorem 1.4.13** (Prime Number Theorem)**.** *Let* primes$(n)$ *be the number of primes in the interval* $\{1,...,n\}$. *Then we have* $\lim_{n\to\infty^-}\left(\frac{\text{primes}(n)}{\frac{n}{\ln(n)}}\right) = 1$

The Prime Number Theorem is often stated as primes$(n) \approx \frac{n}{\ln(n)}$ when $n$ tends to infinity. The Prime Number Theorem was proved by Hadamard and de la Vallee Poussin in 1896.

*Jacques Hadamard (1865-1963).*

**Example 1.4.14.** To find a large prime, for example a 100-digit number, we can use a random technique. Indeed, if we pick a 100-digit number at random, then by the Prime Number Theorem, the probability of having picked a prime is roughly $\frac{1}{\ln(10^{100})}$. Hence we expect to find a prime after at most $\ln\left(10^{100}\right) < 300$ picks.

Using a fast prime test (which does exist!), this can be easily done by a computer.

**Example 1.4.15** (Secure internet traffic)**.** The software company 'Frames' has finally produced a good operating system. The company wants to produce DVDs with this operating system at plants in the US, Europe, and Australia. All plants have a master copy of the operating system, but before starting the production, they first want to make sure that all these copies are the same.

For security reasons, the company does not want to compare the systems bit by bit over the internet. Indeed, competing companies could get secret information or hackers could corrupt it. So, the president of 'Frames' has asked the mathematics department to come up with a quick and very secure way of checking. The mathematicians' response is the following.

**The procedure.** All plants have high quality equipment at their disposal. First a random prime number $p$ is chosen in the interval between 1 and some integer $a$ which can be represented in the binary system with $n$ bits. So $a$ is approximately equal to $2^n$. Next, each plant transforms the bit-string of the operating system, which has approximately length $b$ say, into a number $x$, and then computes the remainder $r = \text{rem}(x, p)$. Finally the three plants compare the remainders thus obtained. This can be done easily, as these remainders are just numbers between 0 and $p$. If they all find the same remainder, they decide that their copies are the same.

**The security.** Why does this test yield a secure way of checking whether all three copies of the operating system are the same? Suppose that one plant's system is a bit-string representing the number $x$, while another plant's system represents the number $y$. If the bit-strings have length (approximately) $b$, then these numbers $x$ and $y$ have size at most $2^b$. Of course, $x = y$ implies $\text{rem}(x, p) = \text{rem}(y, p)$. This means that the conclusion $x \neq y$ is justified if $\text{rem}(x, p) \neq \text{rem}(y, p)$. So suppose that $\text{rem}(x, p) = \text{rem}(y, p)$. How large is the probability of an error? How large is the probability that $x \neq y$?

In this case $x - y$ must be a nonzero multiple of $p$. So the probability $P$ of a wrong conclusion is at most the quotient of the number of prime divisors of $x - y$ by the number of primes less than $2^n$.

First we analyze the numerator of this quotient. If $k$ is the number of primes that divide the number $z = x - y$, then $z \geq 2^k$. But that implies that $k$ is at most $b$.

Now the denominator. According to the Prime Number Theorem the number of primes less than $2^n$ is approximately $2^n/\ln(2^n)$. So, a good estimate for the denominator is $2^n/n$.

Combining the above, we find that $P$, the probability of declaring $x$ and $y$ to be the same while they are not, is at most $\frac{b \cdot n}{2^n}$.

**A concrete example.** Suppose that the operating system fits on a single DVD of 5 Gigabyte. Then the number $b$ of bits on the DVD equals $5 \cdot 2^{10} \cdot 2^{10} \cdot 2^{10} \cdot 2^3$. So, if we pick the prime $p$ at random between 1 and $2^{200}$, then the probability of declaring $x$ and $y$ to be the same while they are not, is less than $\frac{5 \cdot 2^{33} \cdot 200}{2^{200}}$, which is less than $2^{-153}$.

In a similar way one can analyze the probability of declaring $x$ and $y$ to be not the same, while they are equal.

The next theorem gives a characterization of primes.

**Theorem 1.4.16** (Prime Characterization)**.** *Let $p > 1$. Then $p$ is a prime if and only if, for all integers $b$ and $c$, the condition $p|b \cdot c$ implies that $p|b$ or $p|c$.*

*Proof.*

**If.**

*Proof.* Suppose that $p$ is prime. Assume that $p|b \cdot c$ for some integers $b$ and $c$. If $p|b$ we are done. If $p$ is not a divisor of $b$, then $p$ and $b$ have no common divisors greater than 1 and we can apply Result on the divisor of a product to find that $p$ divides $c$.

□

**Only if.**

*Proof.* If $p$ is not prime, then $p = b\cdot c$ for two integers $b$ and $c$ that are greater than 1 and smaller than $p$. Then $p$ divides the product $b\cdot c$, but divides neither $b$ nor $c$ (as $b$ and $c$ are smaller than $p$). We conclude that if, for all integers $b$ and $c$ the condition $p|b\cdot c$ implies that $p|b$ or $p|c$, then $p$ is a prime.

□

□

**Example 1.4.17.** Suppose $a = b\cdot c$, where $b$ and $c$ are integers. The following fact is well known. If $a$ is even, then so is at least one of $b$ or $c$. It is one implication in the special case $p = 2$ of the theorem.

Prime Characterization has the following useful corollary.

**Corollary 1.4.18.** *If $p$ is a prime and $b_1,\ldots,b_s$ are integers such that $p|\prod_{i=1}^{s} b_i$, then there is an index $i \in \{1,\ldots,s\}$ such that $p|b_i$.*

*Proof.* Let $p$ be a prime and $b_1,\ldots,b_s$ integers providing a counterexample to the corollary with $s$ minimal. Hence $p|\prod_{i=1}^{s} b_i$, but $p$ does not divide $b_i$ for each index $i$.

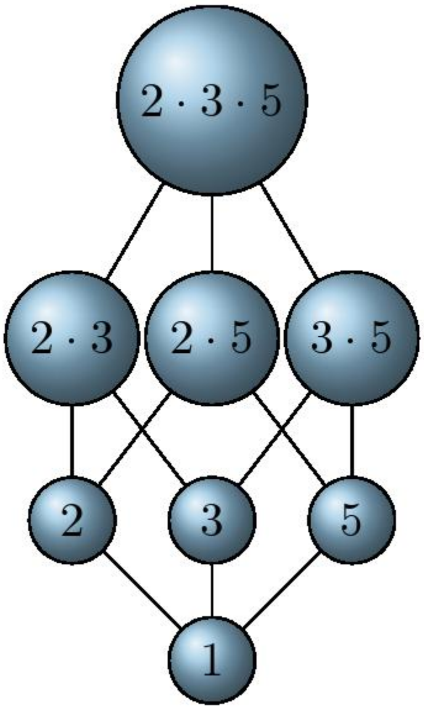
Since $p$ does not divides $b_s$, the Prime Characterization implies that $p$ divides $\prod_{i=1}^{s-1} b_i$. By the minimality of $s$, the integers $b_1,\ldots,b_{s-1}$ do not provide a counterexample to the statement of the corollary. Thus, there is an index $i$ less than $s$ such that $p$ divides $b_i$. This contradicts our assumptions. Hence, no counterexamples exist and we have proven the corollary.

□

**Example 1.4.19.** Let $p$ be a prime, then $p$ does not divide a product of integers, none of which is divisible by $p$. For example, if $i$ is a positive integer less than $p$, then $p$ does not divide $p-i!\cdot i!$.

## 1.5 Factorization

The prime numbers are the building blocks for the multiplicative decomposition of integers. We will now see how integers are built up out of primes.



*Building integers from primes.*

**Theorem 1.5.1** (Unique Factorization)**.** *Every positive integer $a > 1$ can be written as the product of finitely many primes: $a = \prod_{i=1}^{s} p_i$ where s is a positive integer and each $p_i$ is a prime. Up to the order of the factors, this factorization is unique.*

*Proof.* The proof is divided into two steps. Each step is proved by induction on $a$.

**Every integer $a$ is a product of primes.**

The case $a = 2$ is trivial. So suppose that $a$ is at least 3 and that all positive integers less than $a$ can be expressed as a product of primes. If $a$ itself is a prime, then we are done. If $a$ is not a prime, then it has a divisor $b$ such that $1 < b$ and $b < a$ . According to the induction hypothesis, both $b$ and $a/b$ can be written as a product of primes. Explicitly, $b = \prod_{i=1}^{t} p_i$ and $\frac{a}{b} = \prod_{i=1}^{r} q_i$ where $t$ and $r$ are positive integers and all $p_i$ and $q_i$ are primes. But then, as $a = b \cdot (a/b)$, we can write $a$ as the product $a = \prod_{i=1}^{t} p_i \cdot \prod_{i=1}^{r} q_i$. Hence, $a$ is a product of primes.

**The factorization of an integer $a$ is unique (up to order).**

Again the case $a = 2$ is easy. Suppose that $a > 2$, and also suppose that uniqueness of the factorization into primes has been proven for the integers less than $a$.

If $a = \prod_{i=1}^{t} p_i$ and $a = \prod_{i=1}^{r} q_i$ are two ways of expressing $a$ as a product of primes, then it follows that $p_1$ divides $a$. But then $p_1$ also divides $\prod_{i=1}^{r} q_i$.

Using  we conclude that there exists an index $i$ in the set $\{1, \ldots, r\}$ such that $p_1 | q_i$. But then, as $p_1$ and $q_i$ are prime, we have $p_1 = q_i$. Without loss of generality we can assume $i$ to be 1, so $p_1 = q_1$.

Now apply the induction hypothesis to the integer $a/p_1$ with the two expressions as products of primes $\frac{a}{p_1} = \prod_{i=2}^{t} p_i$ and $\frac{a}{p_1} = \prod_{i=2}^{r} q_i$.

These factorizations of $a/p_1$ are the same (up to the order of the factors) and therefore the two factorization of $a$ are also the same.

□

**Example 1.5.2.** Factoring a number into its prime factors is hard! Up to now (2006), the best factorization algorithms can factor numbers consisting of about 100 digits. Factorization of much larger numbers is exceptional. For example, there are numbers with more than 200 digits that have been factorized. One of the more famous examples is the number called RSA-129. In a newspaper article of April, 1994, the following factorization record by A.K. Lenstra, et al. was announced. RSA-129:

$$\begin{gathered}
114381625757888867669235779976146612010218296721242362562561842\,9 \\
35706935245733897830597123563958705058989075147599290026879543541 \\
= \\
3490529510847650949147849619903898133417764638493387843990820577 \\
\times \\
32769132993266709549961988190834461413177642967992942539798288533
\end{gathered}$$

It is not difficult to check that the product of these two factors is indeed the large number: any computer system that can work with these large numbers will confirm it. But it is very hard (indeed many thought it to be unfeasible) to find the factors given the product.

As an indication of how difficult this is, you should try to calculate how many years it would cost to find the above factorization using the obvious algorithm of trying all integers less than the number to be factored. You may assume that the multiplication of two numbers of 130 digits takes about $1/100000$-th of a second. There remains the problem of checking that these two numbers are prime. By means of Eratosthenes' Sieve, this would take a very long time. However there exist primality tests that can check if a 130 digit number is prime in a reasonable amount of time. In 2002, Agrawal, Kayal, and Saxena came up with an algorithm that, for input a prime number $p$, gives a proof of primality in time a polynomial function of the input length, the logarithm of $p$.

| | |
|---|---|
| 2 | $2^1$ |
| 3 | $3^1$ |
| 4 | $2^2$ |
| 5 | $5^1$ |
| 6 | $2^1 \cdot 3^1$ |
| 7 | $7^1$ |
| 8 | $2^3$ |
| 9 | $3^2$ |
| 10 | $2^1 \cdot 5^1$ |
| 11 | $11^1$ |
| 12 | $2^2 \cdot 3^1$ |
| 13 | $13^1$ |
| 14 | $2^1 \cdot 7^1$ |
| 15 | $3^1 \cdot 5^1$ |
| 16 | $2^4$ |
| 17 | $17^1$ |
| 18 | $2^1 \cdot 3^2$ |
| 19 | $19^1$ |
| 20 | $2^2 \cdot 5^1$ |

Table 1.3: Prime factorizations

**Example 1.5.3.** The prime factorizations of the integers between 2 and 20 are

**Remark 1.5.4.** If $a$ is a square, then $\mathrm{ord}_p(a)$ is even for each prime $p$. Using this observation it is not difficult to prove that the square root of 2 is not *rational*, i.e., it is not in $\mathbb{Q}$. This means that there are no integers $a$ and $b$ with $b \neq 0$ such that $\left(\frac{a}{b}\right)^2 = 2$. For, if such $a$ and $b$ exist, then $2 \cdot b^2 = a^2$ and so $\mathrm{ord}_2(2 \cdot b^2) = \mathrm{ord}_2(a^2)$. But $\mathrm{ord}_2(2 \cdot b^2)$ is odd and $\mathrm{ord}_2(a^2)$ is even, a contradiction. Therefore, the assumption that $a$ and $b$ with $\left(\frac{a}{b}\right)^2 = 2$ exist is false.

The same method implies that any $n$-th root of a prime numer is not rational. Indeed, suppose $q$ is a prime and $n$ is at least 2. If $a$ and $b$ are two integers with $\frac{a}{b} = q^{1/n}$, then $\left(\frac{a}{b}\right)^n = q$. So $q \cdot b^n = a^n$ and hence $\mathrm{ord}_q(q \cdot b^n) = \mathrm{ord}_q(a^n)$. But $\mathrm{ord}_q(q \cdot b^n)$ equals $1 + n \cdot \mathrm{ord}_q(b)$, a multiple of $n$ plus 1, while $\mathrm{ord}_q(a^n)$ equals $n \cdot \mathrm{ord}_q(a)$, a multiple of $n$. This is a contradiction.

**Remark 1.5.5.** There also exist arithmetic systems in which uniqueness of factorizations is not guaranteed. For example, in the system $R$ of numbers of the form $a + b \cdot \sqrt{-5}$ with $a, b \in \mathbb{Z}$ we can express 6 in two essentially different ways: $6 = 3 \cdot 2 = \left(1 + \sqrt{-5}\right) \cdot \left(1 - \sqrt{-5}\right)$. The system $R$ is an example of a ring, an algebraic structure with properties similar to those of $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$.

For a non-zero integer $a$, we denote the number of times that the prime $p$ occurs in its factorization by $\mathrm{ord}_p(a)$. So $\mathrm{ord}_p(a)$ is the maximum of all integers $n$ for which $a$ is divisible by $p^n$.

The factorization into primes of $a$ can be written as

$$a = \prod_{p \in \mathbb{P}} p^{\mathrm{ord}_p(a)} \tag{1.4}$$

Here the product is taken over the set $\mathbb{P}$ of all primes. Note however, that only a finite number of factors is distinct from 1.

By definition, a product that has the empty set as index set (the empty product) is 1. With this convention the equality also holds for $a = 1$.

Here is an explicit description of the gcd and lcm of two integers in terms of their prime factorizations.

**Theorem 1.5.6.** *If a and b are positive integers, then*

$$\gcd(a,b) = \prod_{p \in \mathbb{P}} p^{min\left(\mathrm{ord}_p(a),\mathrm{ord}_p(b)\right)} \tag{1.5}$$

*and*

$$\mathrm{lcm}(a,b) = \prod_{p \in \mathbb{P}} p^{max\left(\mathrm{ord}_p(a),\mathrm{ord}_p(b)\right)} \tag{1.6}$$

*In particular we have*

$$a \cdot b = \gcd(a,b) \cdot \mathrm{lcm}(a,b) \tag{1.7}$$

*Proof.* We prove the first equality: For each prime $p$ we certainly have: $min\left(\mathrm{ord}_p(a),\mathrm{ord}_p(b)\right) \leq \mathrm{ord}_p(a)$ and $min\left(\mathrm{ord}_p(a),\mathrm{ord}_p(b)\right) \leq \mathrm{ord}_p(b)$. Hence the right-hand side of the equality $\gcd(a,b) = \prod_{p \in \mathbb{P}} p^{min\left(\mathrm{ord}_p(a),\mathrm{ord}_p(b)\right)}$ is a common divisor of $a$ and $b$. In particular, by the Characterization of the gcd, we find that the right-hand side divides $\gcd(a,b)$.

On the other hand, if for some prime $p$ we have $\mathrm{ord}_p(\gcd(a,b)) = m$, then $p^m$ divides both $a$ and $b$. Therefore, $m \leq \mathrm{ord}_p(a)$ and $m \leq \mathrm{ord}_p(b)$.

Hence the left-hand side of the equation $\gcd(a,b) = \prod_{p \in \mathbb{P}} p^{min\left(\mathrm{ord}_p(a),\mathrm{ord}_p(b)\right)}$ is a divisor of the right-hand side.

Combining the above the equality follows.

The proof of the second equality is left to the reader.

The third statement is a direct consequence of the first two, when you take into account that, for any two integers, their sum is equal to the sum of their maximum and their minimum. In Relation between ggd and lcm another proof of this statement is given.

□

**Example 1.5.7.** Suppose that $a$ is a positive integer and that $p^n$ divides $a$ for some prime number $p$ and positive integer $n$. Choose $n$ maximal with this property, so $n = \mathrm{ord}_p(a)$. Then the binomial coefficient $\left(\begin{array}{c} a \\ p^n \end{array}\right)$ is not divisible by $p$.

Indeed, the binomial coefficient $\left(\begin{array}{c} a \\ p^n \end{array}\right)$ can be written as the quotient of $\prod_{i=0}^{p^n-1}(a-i)$ by $(p^n)!$.

Now for all positive integers $b$ with $b \leq p^n$ we find that $\mathrm{ord}_p(b)$ equals $\mathrm{ord}_p(a-b)$. So every factor $p$ in the numerator is canceled by a factor $p$ in the denominator.

**Example 1.5.8.** Given the integers $a$ and $b$ we can express them as a product of primes. Indeed, we can factor $a = 345$ and $b = 246$ as $a = 3 \cdot 5 \cdot 23$ and $b = 2 \cdot 3 \cdot 41$

Moreover, $\gcd(a,b) = 3$ and $\mathrm{lcm}(a,b) = 2 \cdot 3 \cdot 5 \cdot 23 \cdot 41$

Each of the factors in the above products is prime. You can check this with the Prime test of Eratothenes.

The prime factorization is very well suited for studying the multiplicative structure of the integers. However, it is not so convenient to study the additive structure.

## 1.6 The $b$-ary number system

We commonly represent integers in the *decimal system*. But there are also other systems, like the *binary system* which is heavily used in computer science. The decimal and binary system are two examples in a series.

**Definition 1.6.1** ($b$-ary representation)**.** Let $b > 1$ be an integer. A $b$-ary representation, or representation with respect to base $b$, of an integer $a \geq 0$ is a sequence of numbers $a_0, \ldots, a_k$ with $0 \leq a_i < b$ (the *digits*), such that $a = \sum_{i=0}^{k} a_i \cdot b^i$

We write $a = [a_k, \ldots, a_0]_b$. We speak of the *b-ary number system.*

**Remark 1.6.2.** Besides the binary system, the octal (base 8) and hexadecimal (base 16) systems are often used in computer science.

In base 8 we use the digits 0 to 7, but in base 16 we need more digits. Apart from the digits 0 to 9, it is customary to use the symbols $A$, $B$, $C$, $D$, $E$, $F$ to represent the decimal numbers 10, 11, 12, 13, 14, and 15, respectively.

Thus, the integer 123 is represented as $[7\mathrm{B}]_{16}$.

In the $b$-ary number system, every positive number can be written in precisely one way.

**Theorem 1.6.3.** *Let $b > 1$ be an integer. Every integer $a \geq 0$ has a b-ary representation. Furthermore, this representation is unique if $a > 0$ and if we require that $a_k \neq 0$ for the 'most significant' (i.e., left most) digit in $a = [a_k, ..., a_0]_b$.*

*Proof.* The proof consists of two parts. In both we proceed by induction on $a$.

**Existence: the number $a$ has a $b$-ary representation.**

For $a = 0$, a $b$-ary representation is $[0]_b$. Now suppose that $a > 0$ and that the existence assertion is true for all non-negative integers less than $a$. Let $r$ be the remainder of division of $a$ by $b$. Then $0 \leq r$ and $r < b$. Moreover, $b|a-r$. Since $\frac{a-r}{b} < a$, we can apply the induction hypothesis. We find that there are digits $a_0, ..., a_k$ satisfying $\frac{a-r}{b} = \sum_{i=0}^{k} a_i \cdot b^i$ Rewriting this expression as $a = r + \sum_{i=0}^{k} a_i \cdot b^{i+1}$ we find that $a = [a_k, ..., a_0, r]_b$.

**Uniqueness of the representation.**

Suppose that $a = [a_k, ..., a_0]_b$ and also $a = [c_l, ..., c_0]_b$ are both $b$-ary representations of $a$. By the assumption on the most significant digit we have $a_k \neq 0$ and $c_l \neq 0$. According to the first representation, the remainder when $a$ is divided by $b$ is equal to $a_0$ and, according to the second, it equals $c_0$. Hence $a_0 = c_0$. If $a < b$, then $a = a_0$ and we are finished. Otherwise, we apply the induction hypothesis to the number $\frac{a-a_0}{b}$, which is smaller than $a$. It has representations $[c_l, ..., c_1]_b$ and $[a_k, ..., a_1]_b$ in the $b$-ary number system. So, by the induction hypothesis, $k = l$ and $a_i = c_i$ for all $i \in \{1, ..., k\}$. As we already proved $a_0 = c_0$, this establishes that the two representations are the same.

□

**Example 1.6.4.** The proof of Theorem on b-ary Representation provides an algorithm for computing the $b$-ary representation of the integer $a$ (which is given in the decimal system). Suppose $a = 1238$ and $b = 7$. The last symbol in the string representing $a$ equals $\mathrm{rem}(a, b)$, while the string before the last symbol is the representation of $\mathrm{quot}(a, b)$.

We begin with the empty string. At each step of the algorithm we insert the remainder $\mathrm{rem}(a, b)$ at the beginning of the string and replace $a$ by $\mathrm{quot}(a, b)$.

The algorithm starts with $a = 1238$ and stops when $a$ is equal to 0.

Each row of the following table represents a step in the algorithm.

| $n$ | $a_n = \mathrm{quot}(a_{n-1}, b)$ | $\mathrm{rem}(a_{n-1}, b)$ |
|---|---|---|
| 1 | 176 | 6 |
| 2 | 25 | 1 |
| 3 | 3 | 4 |
| 4 | 0 | 3 |

The algorithm has finished! The $b$-ary representation, where $b = 7$, of $a = 1238$ equals $[3416]_7$.

# 1.7 Exercises

## 1.7.1 Divisors and multiples

**Exercise 1.7.1.** Determine the remainder of $a$ divided by $b$ for each of the following pairs $a$, $b$.

1. 480, 175;
2. 5621, 192;
3. 983675, 105120.

**Exercise 1.7.2.** Suppose that $a$ and $b$ are nonzero integers. Prove that if $a$ divides $b$ and $b$ divides $a$, then $a = b$ or $a = -b$.

**Exercise 1.7.3.** Show that if $a$ divides $b$ and $c$ divides $d$, then $a \cdot c$ divides $b \cdot d$.

**Exercise 1.7.4.** Use induction to prove that 10 divides $3^{4 \cdot n} - 1$ for all positive integers $n$.

**Exercise 1.7.5.** Use induction to prove that, if $a$ and $b$ are integers, $a - b$ divides $a^n - b^n$ for every positive integer $n$.

**Exercise 1.7.6.** Determine the gcd and lcm of $a$ and $b$ for each of the following pairs $a$, $b$.

1. 48, 15;
2. 21, 19;
3. 75, 105.

**Exercise 1.7.7.** Suppose that $a$ and $b$ are nonzero relatively prime integers and suppose that $c$ is a divisor of $a$. Prove that $c$ and $b$ are relatively prime.

**Exercise 1.7.8.** Show that the following three properties hold for the greatest common divisor. Here, $a$, $b$ and $k$ are integers.

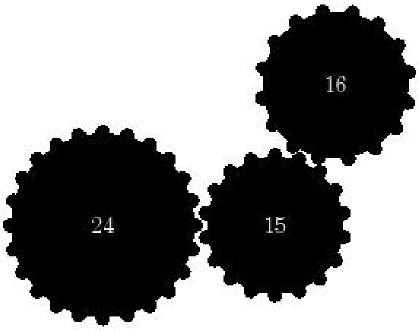
1. $\gcd(a,b) = \gcd(b,a)$
2. $\gcd(a,b) = \gcd(a, b - k \cdot a)$
3. $\gcd(a,0) = |a|$

**Exercise 1.7.9.** For any positive integer $n$ divide $10^{3 \cdot n}$ by $10^n - 1$ and find the remainder.

**Exercise 1.7.10.** If $n$ is a positive integer, determine the possibilities for the greatest common divisor of $n$ and $n^2+3$, and also provide examples.

**Exercise 1.7.11.** Three cogwheels with 24, 15, and 16 cogs, respectively, touch as shown.

What is the smallest positive number of times you have to turn the left-hand cogwheel (with 24 cogs) before the right-hand cogwheel (with 16 cogs) is back in its original position? What is the smallest positive number of times you have to turn the left-hand cogwheel before all three wheels are back in their original position?



*Three cogs*

**Exercise 1.7.12.** Prove that the square of an odd integer is again odd, where 'odd' means 'not divisible by 2' or, equivalently, 'having remainder 1 upon division by 2'. Show that the remainder of division by 4 of the square of an odd integer is 1. Does the last statement hold if we replace 4 by 8? And by 16?

**Exercise 1.7.13.** Suppose that $a$, $b$, and $c$ are integers. If $c$ divides $a$ and $b$, it also divides $\text{rem}(a,b)$. Prove this.

**Exercise 1.7.14.** If $c$ is a common multiple of the integers $a$ and $b$, then $c$ is a multiple of $\text{lcm}(a,b)$. Prove this.

## 1.7.2 Euclid's algorithm

**Exercise 1.7.15.** Determine the gcd of each of the following pairs of numbers, and write this gcd as a linear combination of the given numbers:

1. 480, 175;
2. 5621, 192;
3. 983675, 105120.

**Exercise 1.7.16.** Show that, for all positive integers $x$ and $y$, and nonnegative $z$, we have $\gcd(z\cdot x, z\cdot y) = z\cdot \gcd(x,y)$

**Exercise 1.7.17.** Suppose that $d$ is the nonzero gcd of $a$ and $b$. Prove that $a/d$ and $b/d$ are relatively prime.

**Exercise 1.7.18.** Let $a$, $b$, and $c$ be integers. Show that $\gcd(a,b,c) = \gcd(\gcd(a,b),c)$

**Exercise 1.7.19.** Let $a$, $b$ and $c$ be integers. Prove that there are integers $x$, $y$, and $z$ such that $\gcd(a,b,c) = x\cdot a + y\cdot b + z\cdot c$

**Exercise 1.7.20.** Let $a$ be a rational number such that both $18\cdot a$ and $25\cdot a$ are integers. Show that $a$ itself is an integer.

**Exercise 1.7.21.** Let $a$, $b$, and $c$ be nonzero integers.

Determine the set of all integers that can be expressed in the form $x\cdot a + y\cdot b + z\cdot c$ with $x$, $y$, and $z$ integers.

**Exercise 1.7.22.** Determine the gcd of each of the following pairs of numbers, and write each gcd as a linear combination of the given numbers:

1. 5672, 234;
2. 5311, 121;
3. 32125, 1012.

**Exercise 1.7.23.** Suppose $a$ is a rational number such that $45\cdot a$ and $36\cdot a$ are integers. Is $a$ necessarily an integer? And what if $20\cdot a$ is also known to be an integer?

### 1.7.3 Linear Diophantine equations

**Exercise 1.7.24.** Find all integer solutions $x$ and $y$ to the following Diophantine equations.

1. $22\cdot x + 32\cdot y = 12$
2. $12\cdot x + 25\cdot y = 11$
3. $24\cdot x + 36\cdot y = 18$

**Exercise 1.7.25.** In how many ways can you pay 50 eurocents using only 5 eurocent and 20 eurocent coins? Can you do it with exactly 7 coins?

**Exercise 1.7.26.** Find all integers $x$, $y$, and $z$ that satisfy the two equations $x + y + 3\cdot z = 19$ and $x + 2\cdot y + 5\cdot z = 29$ simultaneously. Also, determine all solutions with $x$, $y$, and $z$ positive.

### 1.7.4 Prime numbers

**Exercise 1.7.27.** Determine all primes of the form $n^2 - 4$, where $n$ is an integer.

**Exercise 1.7.28.** Determine all primes $p$ and $q$ satisfying $p \cdot q = 4 \cdot p + 7 \cdot q$.

**Exercise 1.7.29.** Prove that there exist infinitely many primes of the form $4 \cdot n + 3$, where $n$ is a positive integer.

**Exercise 1.7.30.** Let $p > 1$ be an integer. Prove that $p$ is a prime if and only if for every integer $a$ either $\gcd(p,a) = 1$ or $\gcd(p,a) = p$.

**Exercise 1.7.31.** Let $p$ be a prime and let $a$ be a positive multiple of $p$. Show that there exists a positive integer $n$ such that $a/p^n$ is an integer and $\gcd(p, a/p^n) = 1$.

**Exercise 1.7.32.** Determine all primes less than 100.

**Exercise 1.7.33.** Determine all primes of the form $n^3 + 1$, with $n$ an integer.

**Exercise 1.7.34.** Which of the following integers is prime: 187, 287, 387, 487, or 587?

**Exercise 1.7.35.** Let $n$ be an integer greater than 1, and let $p$ be the smallest divisor of $n$ greater than 1. Prove that $p$ is prime.

### 1.7.5 Factorization

**Exercise 1.7.36.** Determine the prime factorization of the integers 111, 143, 724, and 1011.

**Exercise 1.7.37.** Prove that the cube root of 17 is not rational.

**Exercise 1.7.38.** Prove that 5 is the only prime $p$ such that $3 \cdot p + 1$ is a square.

**Exercise 1.7.39.** The musical pitch of each note corresponds to its frequency, which is expressed in Hertz. If you double the frequency, you find a note an octave higher. If you change the frequency by a factor $3/2$, you obtain a note which is a so-called fifth higher. Starting from a given note, you can construct notes which are one, two, etc., octaves higher. Similarly, you can construct notes which are one, two, etc., fifths higher. Show that these two series of notes have no note in common, except the note you started with.

**Exercise 1.7.40.** Suppose that $a$ and $b$ are coprime positive integers and that the positive integer $n$ is a multiple of both $a$ and $b$. Show that $n$ is a multiple of $a \cdot b$.

**Exercise 1.7.41.** Determine $\gcd\left(2^3 \cdot 3^5 \cdot 7^2, 2^4 \cdot 5 \cdot 5 \cdot 11\right)$ and $\operatorname{lcm}\left(2^3 \cdot 3^5 \cdot 7^2, 2^4 \cdot 5 \cdot 5 \cdot 11\right)$.

**Exercise 1.7.42.** Determine $\gcd\left(4^3 \cdot 6^5 \cdot 7^2, 8^4 \cdot 10^5 \cdot 11\right)$.

**Exercise 1.7.43.** Determine $\gcd\left(2^4 \cdot 3^2 \cdot 5 \cdot 7^6 \cdot 11, 2^2 \cdot 3^2 \cdot 5^3 \cdot 11\right)$.

**Exercise 1.7.44.** How many different positive divisor does 1000 have? And how many 10.000.000?

**Exercise 1.7.45.** What are the gcd and lcm of the following integers:

1. $2^3 \cdot 5^7 \cdot 11$ and $2^2 \cdot 3^4 \cdot 5^2 \cdot 11^4$;
2. $2^1 \cdot 3^3 \cdot 5^2$ and $2^2 \cdot 3^4 \cdot 5 \cdot 11$;
3. $3^2 \cdot 4^5 \cdot 7^2$ and $2^3 \cdot 3^2 \cdot 6^5 \cdot 7^2$.

**Exercise 1.7.46.** Prove the following identity: $\gcd\left(a^2, b^2\right) = (\gcd(a,b))^2$.

## 1.7.6 Number systems

**Exercise 1.7.47.** Compute the 7-ary representation of the following integers given in their decimal representation: 12373, 32147, and 7231.

**Exercise 1.7.48.** Write an algorithm that converts numbers given in the decimal system to the binary system and vice versa.

**Exercise 1.7.49.** Compute the 3-ary representation of the following integers given in their decimal representation: 12373, 32147, and 7231.

**Exercise 1.7.50.** Which $b$-ary system would you use to weigh all possible weights between 1 and 40 with just four standard weights on a balance?

**Exercise 1.7.51.** The decimal representation of an integer $n$ is $[\text{abcabc}]_{10}$, where $a, b$ and $c$ are elements from $\{0, ..., 9\}$.

Prove that 7, 11, and 13 are divisors of $n$.

**Exercise 1.7.52.** The integers 1222, 124211, 2113 and 4121 are given in their decimal representation.

Give the representation in base 2, 4, and 8, respectively.

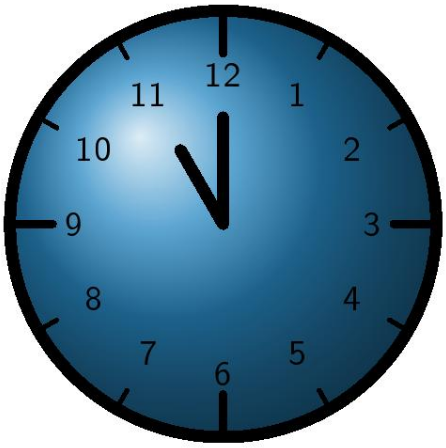
# Chapter 2

# Modular arithmetic

It frequently happens that we prefer to ignore multiples of a given number when we do calculations. Just think of the days in the week or the hours in a day; in the first case we ignore multiples of seven, in the second case multiples of 12 or 24. In this chapter we will describe this ‘arithmetic modulo n’. As an application we will describe the RSA cryptosystem.

## 2.1 Arithmetic modulo n

Clock arithmetic is an example of arithmetic modulo an integer, which is 24 in this case. Suppose that the time is 15:00 hours. If 20 hours pass by, then it will be 11:00 hours. In terms of modular arithmetic, we say that 15 + 20 equals 11 modulo 24. Here, modulo means ‘up to a multiple of’. On the other hand, if 83 hours elapse, then it will be 2 o’clock in the morning. In modular arithmetic, 15 + 83 equals 2 modulo 24. We look at the time of day as a quantity determined up to a multiple of 24.



*Clock arithmetic*

We will analyze arithmetic modulo an integer.

**Definition 2.1.1.** Let $n$ be an integer. On the set $\mathbb{Z}$ of integers we define the relation *congruence modulo n* as follows: $a$ and $b$ are *congruent modulo n* if and only if $n|a-b$.

We write $a \equiv b \pmod{n}$ to denote that $a$ and $b$ are congruent modulo $n$. If $a$ and $b$ are congruent modulo $n$, we also say that $a$ is congruent to $b$ modulo $n$, or that $a$ is equal to $b$ modulo $n$.

**Example 2.1.2.** If $a = 342$, $b = 241$, and $n = 17$, then $a$ is not congruent to $b$ modulo $n$. Indeed $a-b = 101$ is not divisible by $n = 17$.

However, if $a = 342$, $b = 240$, and $n = 17$, then $a$ is congruent to $b$ modulo $n$. Indeed, $a-b = 102$ is divisible by $n = 17$.

**Proposition 2.1.3.** *Let n be an integer. The relation congruence modulo n is reflexive, symmetric, and transitive; in particular, it is an equivalence relation.*
*For nonzero n, there are exactly n distinct equivalence classes:*

$$n\cdot\mathbb{Z}, 1+n\cdot\mathbb{Z}, \ldots, n-1+n\cdot\mathbb{Z} \tag{2.1}$$

*The set of equivalence classes of* $\mathbb{Z}$ *modulo n is denoted by* $\mathbb{Z}/n\mathbb{Z}$.

*Proof.* We need to verify that the relation is reflexive, symmetric, and transitive. This implies congruence modulo $n$ to be an equivalence relation. The other statements of the proposition follow easily.

**The relation is reflexive.**

Let $a$ be an integer. Then $a \equiv a \pmod{n}$ as $n$ divides $a-a = 0$.

**The relation is symmetric.**

Suppose that $a$ and $b$ are integers with $a \equiv b \pmod{n}$. Then $n$ divides $a-b$, and hence also $b-a$. Thus $b \equiv a \pmod{n}$.

**The relation is transitive.**

If $a$, $b$, and $c$ are integers with $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $n$ divides both $a-b$ and $b-c$. But then $n$ is also a divisor of $a-b+b-c = a-c$ and so $a \equiv c \pmod{n}$.

□

**Example 2.1.4.** As congruence modulo $n$ is an equivalence relation, its equivalence classes partition the set $\mathbb{Z}$ of all integers.

For example, the relation modulo 2 partitions the integers into two classes, the even numbers and the odd numbers.

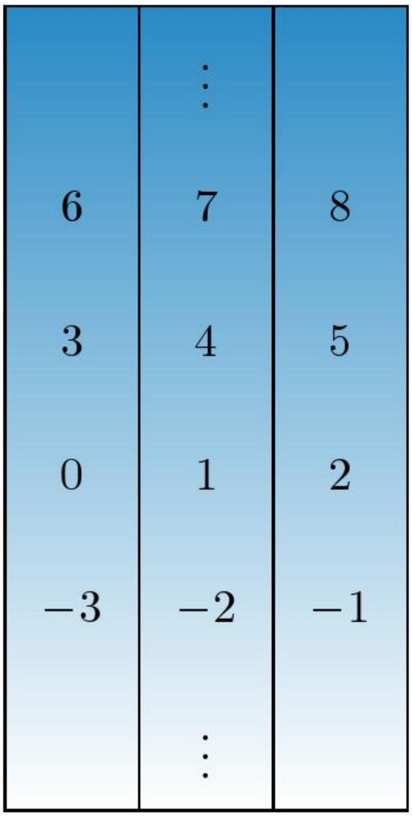
**Remark 2.1.5.** In the Definition of quot and rem [5], the notation $\text{rem}(a,n)$ for the remainder $r$ of the division of $a$ by $n$ is introduced. Observe that $r$ is congruent to $a$ modulo $n$. The remainder $r$ is a natural representative of the set of all elements congruent to $a$ modulo $n$.

If $n$ equals 0, then $a$ is only congruent to itself modulo $n$.

Congruence modulo $n$ is the same relation as congruence modulo $-n$. So, when studying congruence modulo $n$, we may take $n$ to be non-negative without loss of generality.

The set $k+n\cdot\mathbb{Z}$ consists of all integers of the form $k+n\cdot m$ where $m$ is an integer. It is the equivalence class of congruence modulo $n$ containing the integer $k$ and will also be denoted by $k(\text{mod } n)$.

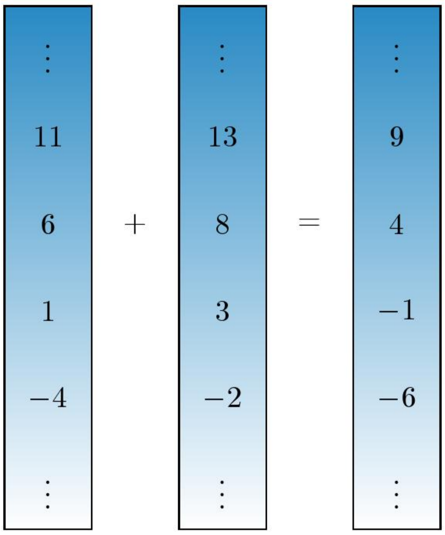
The integer $k$ is a representative of this equivalence class. If no confusion arises, we will also denote the class $k(\text{mod } n)$ by $k$ itself.



|  |  |  |
|---|---|---|
|  | $\vdots$ |  |
| 6 | 7 | 8 |
| 3 | 4 | 5 |
| 0 | 1 | 2 |
| $-3$ | $-2$ | $-1$ |
|  | $\vdots$ |  |

*Congruence modulo 3 splits the integers in three disjoint subsets. These subsets are represented by columns. Integers in the same subset differ by a multiple of 3.*

Let $n$ be an integer. Consider $\mathbb{Z}/n\mathbb{Z}$, the set of equivalence classes of $\mathbb{Z}$ modulo $n$. Addition and multiplication with these classes can be defined in the following way.



*Addition of congruence classes is defined in terms of representatives. For instance, to add the two congruence classes modulo* 5 *above take any representatives in each of these classes, say* 6 *in the first and* 3 *in the second. Then their sum,* 9, *is a representative of the sum of the two classes.*

**Theorem 2.1.6** (Addition and Multiplication)**.** *On* $\mathbb{Z}/n\mathbb{Z}$ *we define two so-called binary operations, an* addition *and a* multiplication, *by:*

- *Addition:* $a(\mathrm{mod}\ n) + b(\mathrm{mod}\ n) = a + b(\mathrm{mod}\ n)$.
- *Multiplication:* $a(\mathrm{mod}\ n) \cdot b(\mathrm{mod}\ n) = a \cdot b(\mathrm{mod}\ n)$.

*Both operations are well defined.*

*Proof.* We have to verify that the definitions of addition and multiplication are consistent. That is, if $x \equiv x' \pmod{n}$ and $y \equiv y' \pmod{n}$, then $x + y \equiv x' + y' \pmod{n}$ and $x \cdot y \equiv x' \cdot y' \pmod{n}$. For then, the outcome of an addition or multiplication is independent of the chosen representatives. Well, $x \equiv x' \pmod{n}$ means that there exists an integer $a$ such that $x - x' = n \cdot a$. Similarly, $y \equiv y' \pmod{n}$ means that there exists an integer $b$ such that $y - y' = n \cdot b$.

**Addition.**

The above implies $(x + y) - (x' + y') = x - x' + y - y' = n \cdot a + n \cdot b = n \cdot (a + b)$. Hence $x + y \equiv x' + y' \pmod{n}$.

**Multiplication.**

By the above we find $x \cdot y - x' \cdot y' = x \cdot (y - y') + (x - x') \cdot y' = n \cdot b \cdot x + n \cdot a \cdot y' = n \cdot (b \cdot x + a \cdot y')$. Hence $x \cdot y \equiv x' \cdot y' \pmod{n}$.

□

**Example 2.1.7** (Tables for modular addition and multiplication)**.** Here is the addition table for $\mathbb{Z}/17\mathbb{Z}$.

| + | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 |
| 2 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 |
| 3 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 |
| 4 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 |
| 5 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 |
| 6 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 |
| 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 8 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 9 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 10 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 11 | 11 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 12 | 12 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 13 | 13 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 14 | 14 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 15 | 15 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 16 | 16 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

Table 2.1: Addition table for $\mathbb{Z}/17\mathbb{Z}$.

Below is the multiplication table for $\mathbb{Z}/17\mathbb{Z}$.

In computations modulo $n$ the following properties of the two operations addition and multiplication are often tacitly used. They look quite straightforward and are easy to use in practice. But since we have constructed a new arithmetical structure, they actually do require proofs. Here is a list of the properties we mean.

| $\cdot$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 2 | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 |
| 3 | 0 | 3 | 6 | 9 | 12 | 15 | 1 | 4 | 7 | 10 | 13 | 16 | 2 | 5 | 8 | 11 | 14 |
| 4 | 0 | 4 | 8 | 12 | 16 | 3 | 7 | 11 | 15 | 2 | 6 | 10 | 14 | 1 | 5 | 9 | 13 |
| 5 | 0 | 5 | 10 | 15 | 3 | 8 | 13 | 1 | 6 | 11 | 16 | 4 | 9 | 14 | 2 | 7 | 12 |
| 6 | 0 | 6 | 12 | 1 | 7 | 13 | 2 | 8 | 14 | 3 | 9 | 15 | 4 | 10 | 16 | 5 | 11 |
| 7 | 0 | 7 | 14 | 4 | 11 | 1 | 8 | 15 | 5 | 12 | 2 | 9 | 16 | 6 | 13 | 3 | 10 |
| 8 | 0 | 8 | 16 | 7 | 15 | 6 | 14 | 5 | 13 | 4 | 12 | 3 | 11 | 2 | 10 | 1 | 9 |
| 9 | 0 | 9 | 1 | 10 | 2 | 11 | 3 | 12 | 4 | 13 | 5 | 14 | 6 | 15 | 7 | 16 | 8 |
| 10 | 0 | 10 | 3 | 13 | 6 | 16 | 9 | 2 | 12 | 5 | 15 | 8 | 1 | 11 | 4 | 14 | 7 |
| 11 | 0 | 11 | 5 | 16 | 10 | 4 | 15 | 9 | 3 | 14 | 8 | 2 | 13 | 7 | 1 | 12 | 6 |
| 12 | 0 | 12 | 7 | 2 | 14 | 9 | 4 | 16 | 11 | 6 | 1 | 13 | 8 | 3 | 15 | 10 | 5 |
| 13 | 0 | 13 | 9 | 5 | 1 | 14 | 10 | 6 | 2 | 15 | 11 | 7 | 3 | 16 | 12 | 8 | 4 |
| 14 | 0 | 14 | 11 | 8 | 5 | 2 | 16 | 13 | 10 | 7 | 4 | 1 | 15 | 12 | 9 | 6 | 3 |
| 15 | 0 | 15 | 13 | 11 | 9 | 7 | 5 | 3 | 1 | 16 | 14 | 12 | 10 | 8 | 6 | 4 | 2 |
| 16 | 0 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

Table 2.2: Multiplication table for $\mathbb{Z}/17\mathbb{Z}$.

**Proposition 2.1.8** (Properties of Modular Arithmetic)**.** *Let n be an integer bigger than* 1*. For all integers a, b, and c, we have the following equalities.*

- *Commutativity of addition:* $a(\text{mod } n)+b(\text{mod } n) = b(\text{mod } n)+a(\text{mod } n)$
- *Commutativity of multiplication:* $a(\text{mod } n)\cdot b(\text{mod } n) = b(\text{mod } n)\cdot a(\text{mod } n)$
- *Associativity of addition:* $((a(\text{mod } n))+(b(\text{mod } n)))+c(\text{mod } n) = a(\text{mod } n)+(((b(\text{mod } n))+(c(\text{mod } n))))$
- *Associativity of multiplication:* $(a(\text{mod } n)\cdot b(\text{mod } n))\cdot c(\text{mod } n) = a(\text{mod } n)\cdot(b(\text{mod } n)\cdot c(\text{mod } n))$
- *Distributivity of multiplication over addition:* $a(\text{mod } n)\cdot(b(\text{mod } n)+c(\text{mod } n)) = a(\text{mod } n)\cdot b(\text{mod } n)+a(\text{mod } n)\cdot c(\text{mod } n)$

*Proof.* The laws hold for integers. For instance, in the case of commutativity, we have $a+b = b+a$. Now apply the Modular Addition and Multiplication Theorem to both sides. The commutativity for $\mathbb{Z}/n\mathbb{Z}$ follows. The proofs of the other equalities are similar.

□

**Example 2.1.9** (Solving equations)**.** Calculations modulo an integer can sometimes be used to show that an equation has no integer solutions. By working in $\mathbb{Z}/4\mathbb{Z}$, for example, we can show that 1203 cannot be written as a sum of two (integer) squares. For, in $\mathbb{Z}/4\mathbb{Z}$, the set of squares is $\{0,1\}$. This is easily verified by squaring each of the four elements of $\mathbb{Z}/4\mathbb{Z}$. Indeed, $(0(\text{mod } 4))^2 = 0(\text{mod } 4)$, $(1(\text{mod } 4))^2 = 1(\text{mod } 4)$, $(2(\text{mod } 4))^2 = 0(\text{mod } 4)$ and $(3(\text{mod } 4))^2 = 1(\text{mod } 4)$.

Now if $m$ and $n$ are integral, then $m^2+n^2(\text{mod } 4) = m^2(\text{mod } 4)+n^2(\text{mod } 4)$, and, by the above, this sum can only take the values $0(\text{mod } 4)$, $1(\text{mod } 4)$, or $2(\text{mod } 4)$. So $m^2+n^2$ is not equal to 3 plus a multiple of 4. In particular, 1203 cannot be written as the sum of two squares.

**Example 2.1.10** (The nine test)**.** Suppose that $a = [a_k,...,a_0]_{10}$ is the usual decimal representation of $a$. The well-known nine test

$$(9|a) \Leftrightarrow (9|((a_k)+...+(a_0))) \tag{2.2}$$

is based on modular arithmetic. In order to see this, we work modulo 9.

Since $10 \equiv 1 \pmod 9$, we find $10^n \equiv 1 \pmod 9$ for all nonnegative integers $n$. As $[a_k,...,a_0]_{10} = a_k\cdot 10^k+...+a_0\cdot 10^0$ reduction modulo 9 implies that $a \equiv a_k+...+a_0 \pmod 9$. Thus $9|a$ if and only if $9|((a_k)+...+(a_0))$.

**Example 2.1.11** (Trigonometric arguments)**.** When playing with a calculator, you may have noticed that $\sin(10^a)$ gives the same value for all values of $a$ bigger than 2, at least when the argument expresses the number of degrees of an angle. The explanation is that $10^a$ is the same number modulo 360 for each of these values of $a$. Check this!

**Example 2.1.12** (Calculating with powers)**.** Modular arithmetic can greatly reduce the amount of work when computing divisibility properties of expressions involving powers. By way of example, we show that $10^9+1$ is divisible by 19. Working modulo 19 we start with $10^2 \equiv 5 \pmod{19}$. Squaring this equation, we find $10^4 \equiv 6 \pmod{19}$. Similarly we get $10^8 \equiv -2 \pmod{19}$ and $10^9 \equiv -1 \pmod{19}$. But then we deduce that $10^9+1 \equiv 0 \pmod{19}$, which implies that $19|((10^9)+1)$.

A *neutral* element for the addition is $0(\text{mod } n)$. Indeed, $a(\text{mod } n)+0 = a(\text{mod } n)$ and $0+a(\text{mod } n) = a(\text{mod } n)$. The *opposite* of $a(\text{mod } n) \in \mathbb{Z}/n\mathbb{Z}$ is $-a(\text{mod } n)$, the unique element $b$ such that $a(\text{mod } n)+b(\text{mod } n) = 0$.

A *neutral* element for the multiplication is $1(\text{mod } n)$, as $a(\text{mod } n)\cdot 1(\text{mod } n) = a(\text{mod } n)$ and $1(\text{mod } n)\cdot a(\text{mod } n) = a(\text{mod } n)$.

The set $\mathbb{Z}/n\mathbb{Z}$ together with addition and multiplication is an example of a quotient ring, an algebraic structure to be discussed in the theory of rings and fields.

In $\mathbb{Z}/n\mathbb{Z}$ we can add, multiply, and subtract. But how about division? Does every nonzero element have an inverse?

**Definition 2.1.13.** An element $a \in \mathbb{Z}/n\mathbb{Z}$ is called *invertible* if there is an element $b$, called *inverse* of $a$, such that $a\cdot b = 1$.

If $a$ is invertible, its inverse (which is unique, as follows from Uniqueness of the Inverse) will be denoted by $a^{-1}$.

The set of all invertible elements in $\mathbb{Z}/n\mathbb{Z}$ will be denoted by $\mathbb{Z}/n\mathbb{Z}^{\times}$. This set is also called the *multiplicative group* of $\mathbb{Z}/n\mathbb{Z}$.

**Example 2.1.14.** In $\mathbb{Z}/18\mathbb{Z}$ the element $5(\text{mod } 18)$ is invertible. Indeed, since $2\cdot 18 - 7\cdot 5 = 1$, the inverse of $5(\text{mod } 18)$ is $7(\text{mod } 18)$. The element $6(\text{mod } 18)$ is not invertible, since any multiple of 6 is either congruent to 0, 6, or 12 modulo 18.

**Remark 2.1.15** (Uniqueness of the Inverse)**.** Multiplicative inverses are unique, i.e., every invertible element has exactly one inverse. For, if

$$a(\text{mod } n)\cdot b(\text{mod } n) = a(\text{mod } n)\cdot c(\text{mod } n) = 1 \tag{2.3}$$

then

$$b(\text{mod } n) = b(\text{mod } n)\cdot a(\text{mod } n)\cdot c(\text{mod } n) = a(\text{mod } n)\cdot b(\text{mod } n)\cdot c(\text{mod } n) = c(\text{mod } n) \tag{2.4}$$

An integer $a$ will be called *invertible modulo n* if its class $a(\text{mod } n)$ is invertible.

In $\mathbb{Z}$ division is not always possible. Some nonzero elements do have an inverse, others don't. The following theorem tells us precisely which elements of $\mathbb{Z}/n\mathbb{Z}$ have an inverse.

**Theorem 2.1.16** (Characterization of Modular Invertibility)**.** *Let $n > 1$ and $a \in \mathbb{Z}$.*

1. *The class $a(\text{mod } n)$ in $\mathbb{Z}/n\mathbb{Z}$ has a multiplicative inverse if and only if* $\gcd(a,n) = 1$.
2. *If $a$ and $n$ are relatively prime, then the inverse of $a(\text{mod } n)$ is the class* $\text{extended} - \gcd(a,n)_2(\text{mod } n)$.
3. *In $\mathbb{Z}/n\mathbb{Z}$, every class distinct from 0 has an inverse if and only if n is prime.*

*Proof.* The second and third statement of the theorem are straightforward consequences of the first and its proof. So, we only prove the first. There are two parts to the proof.

**If.**

If $\gcd(a,n) = 1$, then, from the Extended Euclidean Algorithm, it follows that there are integers $x$ and $y$ such that $a\cdot x + n\cdot y = 1$. In $\mathbb{Z}/n\mathbb{Z}$ this translates to $a(\text{mod } n)\cdot x(\text{mod } n) + 0 = 1$. In particular, $x(\text{mod } n)$ is the inverse of $a(\text{mod } n)$.

Notice that $x$ indeed coincides with $\text{extended} - \gcd(a,n)_2$ modulo $n$, which proves the second statement.

**Only if.**

If $a(\bmod n)$ has an inverse $b(\bmod n)$ in $\mathbb{Z}/n\mathbb{Z}$, then there exists an integer $x$ with $a\cdot b+x\cdot n=1$. So, by the Characterization of the gcd, we find $\gcd(a,n)=1$.

□

**Example 2.1.17.** The invertible elements in $\mathbb{Z}/2^n\mathbb{Z}$ are the classes $x(\bmod 2^n)$ for which $x$ is an odd integer.

Indeed, the gcd of $x$ and $2^n$ equals 1 if and only if $x$ is odd.

An arithmetical system such as $\mathbb{Z}/p\mathbb{Z}$ with $p$ prime, in which every element not equal to 0 has a multiplicative inverse, is called a *field*, just like $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$.

Suppose that $n$ and $a$ are integers with $n>1$ and $\gcd(a,n)=1$. The Characterization of Modular Invertibility not only gives the existence of the inverse of $a(\bmod n)$ in $\mathbb{Z}/n\mathbb{Z}$, but also a way to compute this inverse.

**Algorithm 2.1.18** (Modular Inverse)**.** • *Input: integers $n>1$ and $a$.*

- *Output: the inverse of the class $a(\bmod n)$ of $a$ in $\mathbb{Z}/n\mathbb{Z}$ if it exists, and 0 otherwise.*

```
Inverse := procedure($a, n$)
local variables
    | $E :=$ extended − gcd($a, n$)
if $E_1 = 1$
    then
        return
            | $E_2(\bmod n)$
    else
        return
            | 0
```

*Proof.*

**Termination.**

By the absence of loops this is obvious.

**Correctness.**

Obvious by part (b) of the Characterization of Modular Invertibility.

□

**Example 2.1.19.** Consider $a=24$ and $n=35$. Then $a$ and $n$ are relative prime. So $a(\bmod n)$ has an inverse. To find the inverse of $a(\bmod n)$, we apply the Extended Euclidean Algorithm. This gives the following expression of 1 as a linear combination of $a$ and $n$:

$$1=35\cdot 11-24\cdot 16 \tag{2.5}$$

We deduce that the inverse of $a(\bmod n)$ equals $-16(\bmod n)$.

Besides invertible elements in $\mathbb{Z}/n\mathbb{Z}$, which can be viewed as divisors of 1, see Definition of inverse, one can also consider the divisors of 0.

**Definition 2.1.20.** An element $a \in \mathbb{Z}/n\mathbb{Z}$ not equal to 0 is called a *zero divisor* if there is a nonzero element $b$ such that $a \cdot b = 0$.

**Example 2.1.21.** The zero divisors in $\mathbb{Z}/24\mathbb{Z}$ are those elements for which one finds a 0 in the corresponding row (or column) of the multiplication table. These are the elements $2 \pmod{24}$, $4 \pmod{24}$, $6 \pmod{24}$, $8 \pmod{24}$, $9 \pmod{24}$, $10 \pmod{24}$, $12 \pmod{24}$, $14 \pmod{24}$, $15 \pmod{24}$, $16 \pmod{24}$, $18 \pmod{24}$, $20 \pmod{24}$, $21 \pmod{24}$, and $22 \pmod{24}$.

**The multiplication table modulo 24**

| · | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 2 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| 3 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 0 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 0 | 3 | 6 | 9 | 12 | 15 | 18 | 21 |
| 4 | 4 | 8 | 12 | 16 | 20 | 0 | 4 | 8 | 12 | 16 | 20 | 0 | 4 | 8 | 12 | 16 | 20 | 0 | 4 | 8 | 12 | 16 | 20 |
| 5 | 5 | 10 | 15 | 20 | 1 | 6 | 11 | 16 | 21 | 2 | 7 | 12 | 17 | 22 | 3 | 8 | 13 | 18 | 23 | 4 | 9 | 14 | 19 |
| 6 | 6 | 12 | 18 | 0 | 6 | 12 | 18 | 0 | 6 | 12 | 18 | 0 | 6 | 12 | 18 | 0 | 6 | 12 | 18 | 0 | 6 | 12 | 18 |
| 7 | 7 | 14 | 21 | 4 | 11 | 18 | 1 | 8 | 15 | 22 | 5 | 12 | 19 | 2 | 9 | 16 | 23 | 6 | 13 | 20 | 3 | 10 | 17 |
| 8 | 8 | 16 | 0 | 8 | 16 | 0 | 8 | 16 | 0 | 8 | 16 | 0 | 8 | 16 | 0 | 8 | 16 | 0 | 8 | 16 | 0 | 8 | 16 |
| 9 | 9 | 18 | 3 | 12 | 21 | 6 | 15 | 0 | 9 | 18 | 3 | 12 | 21 | 6 | 15 | 0 | 9 | 18 | 3 | 12 | 21 | 6 | 15 |
| 10 | 10 | 20 | 6 | 16 | 2 | 12 | 22 | 8 | 18 | 4 | 14 | 0 | 10 | 20 | 6 | 16 | 2 | 12 | 22 | 8 | 18 | 4 | 14 |
| 11 | 11 | 22 | 9 | 20 | 7 | 18 | 5 | 16 | 3 | 14 | 1 | 12 | 23 | 10 | 21 | 8 | 19 | 6 | 17 | 4 | 15 | 2 | 13 |
| 12 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 | 0 | 12 |
| 13 | 13 | 2 | 15 | 4 | 17 | 6 | 19 | 8 | 21 | 10 | 23 | 12 | 1 | 14 | 3 | 16 | 5 | 18 | 7 | 20 | 9 | 22 | 11 |
| 14 | 14 | 4 | 18 | 8 | 22 | 12 | 2 | 16 | 6 | 20 | 10 | 0 | 14 | 4 | 18 | 8 | 22 | 12 | 2 | 16 | 6 | 20 | 10 |
| 15 | 15 | 6 | 21 | 12 | 3 | 18 | 9 | 0 | 15 | 6 | 21 | 12 | 3 | 18 | 9 | 0 | 15 | 6 | 21 | 12 | 3 | 18 | 9 |
| 16 | 16 | 8 | 0 | 16 | 8 | 0 | 16 | 8 | 0 | 16 | 8 | 0 | 16 | 8 | 0 | 16 | 8 | 0 | 16 | 8 | 0 | 16 | 8 |
| 17 | 17 | 10 | 3 | 20 | 13 | 6 | 23 | 16 | 9 | 2 | 19 | 12 | 5 | 22 | 15 | 8 | 1 | 18 | 11 | 4 | 21 | 14 | 7 |
| 18 | 18 | 12 | 6 | 0 | 18 | 12 | 6 | 0 | 18 | 12 | 6 | 0 | 18 | 12 | 6 | 0 | 18 | 12 | 6 | 0 | 18 | 12 | 6 |
| 19 | 19 | 14 | 9 | 4 | 23 | 18 | 13 | 8 | 3 | 22 | 17 | 12 | 7 | 2 | 21 | 16 | 11 | 6 | 1 | 20 | 15 | 10 | 5 |
| 20 | 20 | 16 | 12 | 8 | 4 | 0 | 20 | 16 | 12 | 8 | 4 | 0 | 20 | 16 | 12 | 8 | 4 | 0 | 20 | 16 | 12 | 8 | 4 |
| 21 | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 0 | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 0 | 21 | 18 | 15 | 12 | 9 | 6 | 3 |
| 22 | 22 | 20 | 18 | 16 | 14 | 12 | 10 | 8 | 6 | 4 | 2 | 0 | 22 | 20 | 18 | 16 | 14 | 12 | 10 | 8 | 6 | 4 | 2 |
| 23 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

Table 2.3: The multiplication table modulo 24

The following theorem tells us which elements of $\mathbb{Z}/n\mathbb{Z}$ are zero divisors. They turn out to be those nonzero elements which are not invertible. Hence a nonzero element in $\mathbb{Z}/n\mathbb{Z}$ is either invertible or a zero divisor.

**Theorem 2.1.22** (Zero Divisor Characterization)**.** *Let $n > 1$ and $a \in \mathbb{Z}$.*

1. *The class $a(\bmod n)$ in $\mathbb{Z}/n\mathbb{Z}$ is a zero divisor if and only if $\gcd(a,n) > 1$ and $a(\bmod n)$ is nonzero.*
2. *The residue ring $\mathbb{Z}/n\mathbb{Z}$ has no zero divisors if and only if $n$ is prime.*

*Proof.* The second statement of the theorem is a straightforward consequence of the first. So, we only prove the first. There are two parts to the proof.

**If.**

Suppose that $\gcd(a,n) > 1$, and set $b = n/\gcd(a,n)$. Then the class $b(\bmod n)$ of $b$ is nonzero, but $a \cdot b$ is a multiple of $n$ and so $a \cdot b(\bmod n) = 0$. This translates to $a(\bmod n) \cdot b(\bmod n) = 0$ in $\mathbb{Z}/n\mathbb{Z}$. In particular, $a(\bmod n)$ is a zero divisor.

**Only if.**

If $a(\bmod n)$ is a zero divisor, then it is nonzero and there is a nonzero element $b(\bmod n)$ in $\mathbb{Z}/n\mathbb{Z}$ with $a(\bmod n) \cdot b(\bmod n) = 0$. So, for the representative $b_0$ of $b(\bmod n)$ in $\{1,\ldots,n-1\}$, we find that $a \cdot b_0$ is a common multiple of $a$ and $n$. In particular, $\operatorname{lcm}(a,n) < a \cdot b_0$, which is certainly less than $a \cdot n$. Now the Relation between ggd and lcm implies that $\gcd(a,n) > 1$.

□

**Example 2.1.23.** Below you find the multiplication table of $\mathbb{Z}/17\mathbb{Z} \setminus \{0\}$. As you can see, it contains no entry with a 0, which implies that $\mathbb{Z}/17\mathbb{Z}$ has no zero divisors. Moreover, as each row and column contains a 1, each nonzero element of $\mathbb{Z}/17\mathbb{Z}$ is invertible.

Since an element $a(\bmod n)$ of $\mathbb{Z}/n\mathbb{Z}$ is either 0, a zero divisor, or invertible, the Modular Inverse Algorithm for computing inverses in $\mathbb{Z}/n\mathbb{Z}$ also provides us with a way to check whether an arbitrary element of $\mathbb{Z}/n\mathbb{Z}$ is a zero divisor.

Let $n$ be an integer. Inside $\mathbb{Z}/n\mathbb{Z}$, we can distinguish the set of invertible elements and the set of zero divisors. The set of invertible elements is closed under multiplication, the set of zero divisors together with 0 is even closed under multiplication by arbitrary elements.

**Lemma 2.1.24.** *Let $n$ be an integer with $n > 1$.*

1. *If $a$ and $b$ are elements in $\mathbb{Z}/n\mathbb{Z}^\times$, then their product $a \cdot b$ is invertible and therefore also in $\mathbb{Z}/n\mathbb{Z}^\times$. The inverse of $a \cdot b$ is given by $b^{-1} \cdot a^{-1}$.*
2. *If $a$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$ and $b$ an arbitrary element, then $a \cdot b$ is either $0$ or a zero divisor.*

| · | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 2 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 |
| 3 | 3 | 6 | 9 | 12 | 15 | 1 | 4 | 7 | 10 | 13 | 16 | 2 | 5 | 8 | 11 | 14 |
| 4 | 4 | 8 | 12 | 16 | 3 | 7 | 11 | 15 | 2 | 6 | 10 | 14 | 1 | 5 | 9 | 13 |
| 5 | 5 | 10 | 15 | 3 | 8 | 13 | 1 | 6 | 11 | 16 | 4 | 9 | 14 | 2 | 7 | 12 |
| 6 | 6 | 12 | 1 | 7 | 13 | 2 | 8 | 14 | 3 | 9 | 15 | 4 | 10 | 16 | 5 | 11 |
| 7 | 7 | 14 | 4 | 11 | 1 | 8 | 15 | 5 | 12 | 2 | 9 | 16 | 6 | 13 | 3 | 10 |
| 8 | 8 | 16 | 7 | 15 | 6 | 14 | 5 | 13 | 4 | 12 | 3 | 11 | 2 | 10 | 1 | 9 |
| 9 | 9 | 1 | 10 | 2 | 11 | 3 | 12 | 4 | 13 | 5 | 14 | 6 | 15 | 7 | 16 | 8 |
| 10 | 10 | 3 | 13 | 6 | 16 | 9 | 2 | 12 | 5 | 15 | 8 | 1 | 11 | 4 | 14 | 7 |
| 11 | 11 | 5 | 16 | 10 | 4 | 15 | 9 | 3 | 14 | 8 | 2 | 13 | 7 | 1 | 12 | 6 |
| 12 | 12 | 7 | 2 | 14 | 9 | 4 | 16 | 11 | 6 | 1 | 13 | 8 | 3 | 15 | 10 | 5 |
| 13 | 13 | 9 | 5 | 1 | 14 | 10 | 6 | 2 | 15 | 11 | 7 | 3 | 16 | 12 | 8 | 4 |
| 14 | 14 | 11 | 8 | 5 | 2 | 16 | 13 | 10 | 7 | 4 | 1 | 15 | 12 | 9 | 6 | 3 |
| 15 | 15 | 13 | 11 | 9 | 7 | 5 | 3 | 1 | 16 | 14 | 12 | 10 | 8 | 6 | 4 | 2 |
| 16 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

Table 2.4: Multiplication table for $\mathbb{Z}/17\mathbb{Z}$.

*Proof.* Assume that $a$ and $b$ are elements in $\mathbb{Z}/n\mathbb{Z}^{\times}$. As $(a\cdot b)\cdot(b^{-1}\cdot a^{-1}) = a\cdot a^{-1} = 1$ the inverse of $a\cdot b$ is $b^{-1}\cdot a^{-1}$. This establishes the first assertion.

If $a$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$, then there is a nonzero element $c$ with $a\cdot c$ equal to 0. But then $a\cdot b\cdot c$ is also equal to 0. So $a\cdot b$ is 0 or a zero divisor.

□

**Example 2.1.25.** The zero divisors in $\mathbb{Z}/6\mathbb{Z}$ are those elements for which 0 occurs in the corresponding row (or column) of the multiplication table. The invertible elements are the elements for which 1 occurs in the corresponding row (or column).

| · | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 | 5 |
| 2 | 2 | 4 | 0 | 2 | 4 |
| 3 | 3 | 0 | 3 | 0 | 3 |
| 4 | 4 | 2 | 0 | 4 | 2 |
| 5 | 5 | 4 | 3 | 2 | 1 |

Table 2.5: Multiplication table modulo 6.

So, the zero divisors are the classes of 2, 3, and 4, while the invertible elements are the classes of 1 and 5.

Notice that $5^2(\text{mod } n) = 1(\text{mod } n)$. So indeed, the set of invertible elements is closed under multiplication.

## 2.2 Linear congruences

In addition to the linear equation

$$a \cdot x = b \tag{2.6}$$

with integer coefficients $a$ and $b$ in the single unkown $x$, we study, for positive integers $n$, the related equation

$$a \cdot x \equiv b \pmod{n} \tag{2.7}$$

in the unknown $x$. Such equation is called a *linear congruence*. It is closely related to the equation

$$a \cdot x = b \tag{2.8}$$

where $a$ and $b$ are elements of $\mathbb{Z}/n\mathbb{Z}$ and the unknown $x$ is also in $\mathbb{Z}/n\mathbb{Z}$.

Solving such a linear congruence or the related equation in $\mathbb{Z}/n\mathbb{Z}$ is based on solving

$$a \cdot x + n \cdot y = b \tag{2.9}$$

in the unknown $x$ and $y$; see Linear Diophantine Equation Solving Algorithm. The results of Linear Diophantine Equation Solving Algorithm can easily be translated to the present situation. As a result we obtain the following algorithm for solving linear congruences.

**Algorithm 2.2.1** (Linear Congruence)**.** • *Input: integers a, b, and a positive integer n*

- *Output: the set of all classes x modulo n satisfying the equation* $a \cdot x \equiv b \pmod{n}$

```
SolveLinCong := procedure$(a,b,n)$
local variables
    $E :=$ extended − gcd $(a,n)$
    $g := E_1$
    $z := E_2$
if $g|b$
    then
        return
            $\left\{z \cdot \frac{b}{g} + k \cdot \frac{n}{g} \pmod{n} \;\middle|\; k \in \mathbb{Z}/n\mathbb{Z}\right\}$
    else
        return
            $\varnothing$
```

*Proof.*

**Termination.**

Obvious in the absence of loops.

**Correctness.**

For each integer solution $x$ to the linear congruence $a \cdot x \equiv b \pmod{n}$, there is an integer $y$ such that the pair $x$, $y$ is a solution to the linear Diophantine equation $a \cdot x + n \cdot z = b$, and vice versa. So, the correctness of the algorithm follows from the correctness of Linear Diophantine Equation Solving Algorithm for solving linear Diophantine equations.

□

**Remark 2.2.2.** In the terminology of the Linear Congruence Algorithm, the solutions of the related equation $a \cdot x = b$ over $\mathbb{Z}/n\mathbb{Z}$ are the elements of the set

$$\left\{ z \cdot \frac{b}{g} + k \cdot \frac{n}{g} (\bmod\ n) \;\middle|\; k \in \mathbb{Z}/n\mathbb{Z} \right\} \tag{2.10}$$

Observe that there are exactly $g$ distinct solutions.

**Example 2.2.3.** In order to find all solutions to the congruence $24 \cdot x \equiv 12 \pmod{15}$ we first compute the gcd of 24 and 15. Using the Extended Euclidean Algorithm we find

$$\gcd(24, 15) = 3 = 2 \cdot 24 - 3 \cdot 15 \tag{2.11}$$

Now 3 divides 12, so the solution set is

$$\{(2 \cdot 12 + k \cdot 15)/3 \mid k \in \mathbb{Z}\} \tag{2.12}$$

Instead of using the algorithm, we can also use the expression of the gcd as a linear combination of 24 and 15 to argue what the solution is. To this end, multiply both sides of the equality $3 = 2 \cdot 24 - 3 \cdot 15$ by 4. This gives $12 = 8 \cdot 24 - 12 \cdot 15$.

So, a solution of the congruence is $x = 8 (\bmod\ 15)$. Other solutions can be found by adding multiples of $15/3 (\bmod\ 15)$ to this particular solution.

So, the complete set of solutions for $x$ consists of the classes $3 (\bmod\ 15)$, $8 (\bmod\ 15)$, and $13 (\bmod\ 15)$.

We extend the study of a single congruence to a method for solving special systems of congruences.

**Theorem 2.2.4** (Chinese Remainder Theorem)**.** *Suppose that $n_1$, ..., $n_k$ are pairwise coprime integers. Then for all integers $a_1$, ..., $a_k$ the system of linear congruences*

$$x \equiv a_i \pmod{n_i} \tag{2.13}$$

*with $i \in \{1,...,k\}$ has a solution.*
*Indeed, the integer*

$$x = \sum_{i=1}^{k} a_i \cdot y_i \cdot \frac{n}{n_i} \tag{2.14}$$

*where for each i we have*

$$y_i = \mathrm{extended}-\mathrm{gcd}\left(\frac{n}{n_i}, n_i\right)_3 \tag{2.15}$$

*satisfies all congruences.*
*Any two solutions to the system of congruences are congruent modulo the product $\prod_{i=1}^{k} n_i$.*

*Proof.* The proof consists of two parts.

**Existence of a solution.**

Let $n$ be equal to $\prod_{i=1}^{k} n_i$. Then, by the assumption that all the $n_i$ are coprime we find that for each $i$ the greatest common divisor of $n_i$ and $\frac{n}{n_i}$ equals 1. Thus by the Extended Euclidean Algorithm we can find $x_i$ and $y_i$ with $x_i \cdot n_i + y_i \cdot \frac{n}{n_i} = 1$. Since $x_i \cdot n_i + y_i \cdot \frac{n}{n_i} = 1$, we find that $a_i \cdot y_i \cdot \frac{n}{n_i}$ is equal to $a_i$ if we compute modulo $n_i$, and equal to 0 if we compute modulo $n_j$ where $n_i \neq n_j$. This clearly implies that $x = \sum_{i=1}^{k}(a_i \cdot y_i \cdot \frac{n}{n_i})$ satisfies $x \equiv a_i \pmod{n_i}$ for all $i$. So we have found that $x$ is a solution. This solution is not unique. Indeed, for any integer $a$, the integer $x + a \cdot n$ is also a solution.

**Uniqueness modulo $n$.**

Suppose that, besides $x$, also $y$ is a solution to the system of congruences. Then for each $i$ we find that the integer $n_i$ divides the difference $x - y$. By the observation that, if two coprime integers divide an integer, then so does their product, this implies that $x - y$ is a common multiple of all the $n_i$, and thus a multiple of the least common multiple of the $n_i$, which equals $n$. This proves that up to multiples of $n$ there is only one solution.

□

**Example 2.2.5.** Suppose that $a$, $b$, $m$, and $n$ are integers. We indicate how to find the common integral solutions $x$ to the linear congruences $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$.

Consider the case where $a = 13$, $b = 5$, $m = 14$, and $n = 17$.

Of course, adding multiples of $m \cdot n = 238$ to any solution will provide other solutions. Therefore we can restrict our attention to solutions in the interval $\{0, \ldots, 237\}$.

The positive integers $x$ in$\{0, \ldots, 237\}$ satisfying $x \equiv 13 \pmod{14}$ are

$$13, 27, 41, 55, 69, 83, 97, 111, 125, 139, 153, 167, 181, 195, 209, 223, 237 \tag{2.16}$$

The positive integers $x$ in $\{0, \ldots, 237\}$ satisfying $x \equiv 5 \pmod{17}$ are

$$5, 22, 39, 56, 73, 90, 107, 124, 141, 158, 175, 192, 209, 226 \tag{2.17}$$

So, modulo 238, the unique common solution to both congruences is 209.

Here is another way of making the last statement of Chinese Remainder Theorem: If $x$ is a solution, then the set of all solutions is the set $x(\mathrm{mod}\ \prod_{i=1}^{k} n_i)$.

The Chinese Remainder Theorem can be turned into an algorithm to solve systems of linear congruences.

**Algorithm 2.2.6** (Chinese Remainder Algorithm)**.** • *Input: distinct and pairwise coprime integers $n_1$, ..., $n_k$, as well as integers $a_1$, ..., $a_k$.*

- *Output: a common solution $x$ to the congruences $x \equiv a_i \pmod{n_i}$.*

```
ChineseRemainder := procedure($n_1, \ldots, n_k, a_1, \ldots, a_k$)
local variables
    $i$
    $y_1, \ldots, y_k$
    $n := \prod_{i=1}^{k} n_i$
for $i := 1$ while $i \leq k$ with step $i := i + 1$ do
    $y_i := \mathrm{extended} - \mathrm{gcd}\left(\frac{n}{n_i}, n_i\right)_3$

return
    $\sum_{i=1}^{k} a_i \cdot y_i \cdot \frac{n}{n_i}$
```

*Proof.*

**Termination.**

Obvious.

**Correctness.**

This follows immediately from the Chinese Remainder Theorem.

□

## 2.3 The Theorems of Fermat and Euler

Let $p$ be a prime. Consider $\mathbb{Z}/p\mathbb{Z}$, the set of equivalence classes of $\mathbb{Z}$ modulo $p$. In $\mathbb{Z}/p\mathbb{Z}$ we can add, subtract, multiply, and divide by elements which are not 0. Moreover, it contains no zero divisors. So $\mathbb{Z}/p\mathbb{Z}$ has very nice properties. These are used in the proof of the following important result.

*Pierre de Fermat (1601-1665)*

**Theorem 2.3.1** (Fermat's Little Theorem)**.** *Let p be a prime. For every integer a we have*

$$a^p \equiv a \pmod{p} \tag{2.18}$$

*In particular, if a is not in* $0 (\mathrm{mod}\ p)$ *then*

$$a^{p-1} \equiv 1 \pmod{p} \tag{2.19}$$

*Equivalently, for all elements a in* $\mathbb{Z}/p\mathbb{Z}$ *we have*

$$a^p = a \tag{2.20}$$

*For nonzero elements a we have*

$$a^{p-1} = 1 \tag{2.21}$$

*Proof.* Although the statements on integers and on classes are easily seen to be equivalent, we present a proof for each of these. Let $p$ be a prime.

**For every integer $a$ we have $a^p \equiv a \pmod{p}$.**

For nonnegative $a$ we give a proof by induction on $a$.

For $a$ equal to 0 the statement is trivial. Now assume that, for some $a \geq 0$, we have $a^p \equiv a \pmod{p}$. By Newton's Binomium, we find that $(a+1)^p$ equals $\sum_{i=0}^{p} \left( \begin{array}{c} p \\ i \end{array} \right) \cdot a^i$. Recall that the binomial coefficient is determined by $\left(\begin{array}{c} p \\ i \end{array}\right) = \frac{p!}{(p-i)!\cdot i!}$. Thus, for $i$ not equal to 0 or $p$, the numerator of this fraction is divisible by the prime $p$, whereas the denominator is not. We conclude that, for $i$ not equal to 0 or $p$, the binomial coefficient $\left(\begin{array}{c} p \\ i \end{array}\right)$ is divisible by $p$.

As a result we find that $(a+1)^p \equiv a^p + 1 \pmod p$. Now, from the hypothesis $a^p \equiv a \pmod p$ we conclude that

$$(a+1)^p \equiv a+1 \pmod p \tag{2.22}$$

This proves the theorem for all nonnegative $a$.

If $a$ is negative, then, by the above, $(-a)^p \equiv -a \pmod p$. If $p$ is odd, we immediately deduce $a^p \equiv a \pmod p$. If $p$ is even, then it is 2 and the above implies that $a^p \equiv -a \pmod p$. But as $-a \equiv a \pmod 2$, we again find that $a^p \equiv a \pmod p$. This proves the assertion for all integers $a$.

**For all elements $a$ in $\mathbb{Z}/p\mathbb{Z}$ we have $a^p = a$.**

For $a$ equal to 0 the statements are trivial. Thus assume that $a$ is nonzero. Consider the set $\mathbb{Z}/p\mathbb{Z}^\times$ of nonzero (and hence invertible) elements of $\mathbb{Z}/p\mathbb{Z}$.

Consider the map

$$M_a\colon \mathbb{Z}/p\mathbb{Z}^\times \to \mathbb{Z}/p\mathbb{Z}^\times, b \longmapsto a \cdot b \tag{2.23}$$

that is, multiplication by $a$. As $\mathbb{Z}/p\mathbb{Z}$ contains no zero divisors, see Characterization of Modular Invertibility, the map is well defined. Moreover, this map is bijective. Indeed, its inverse is $M_{a^{-1}}$, multiplication by $a^{-1}$. As a result we see that the product of all elements in $\mathbb{Z}/p\mathbb{Z}^\times$ is not only equal to $\prod_{z\in\mathbb{Z}/p\mathbb{Z}^\times} z$, but also to $\prod_{z\in\mathbb{Z}/p\mathbb{Z}^\times} (M_a(z))$. The products are taken over the same set. The order in which the elements are multiplied might differ, but that does not affect the result. The latter product equals

$$\prod_{z\in\mathbb{Z}/p\mathbb{Z}^\times} ((a\cdot z)) = a^{p-1} \cdot \prod_{z\in\mathbb{Z}/p\mathbb{Z}^\times} z \tag{2.24}$$

By Characterization of Modular Invertibility the product $\prod_{z\in\mathbb{Z}/p\mathbb{Z}^\times} z$ is nonzero and hence invertible, see Invertibility of Products. Therefore, $a^{p-1} = 1$. Multiplying both sides of the equation by $a$ proves the assertion.

The other statements in Fermat's Little Theorem follow easily from the above assertions.

□

**Example 2.3.2.** As 7 is prime, Fermat's Little Theorem implies that $2^6 \equiv 1 \pmod 7$. Indeed, $2^6 = 64 = 9\cdot 7 + 1$.

**Example 2.3.3.** The integer $1234^{1234} - 2$ is divisible by 7.

Indeed, if we compute modulo 7, then we find $1234 \equiv 2 \pmod 7$. Moreover, by Fermat's Little Theorem we have $2^6 \equiv 1 \pmod 7$, so integer2.eqsmod $\left(1234^{1234}, 2^{1234}, 2^{6\cdot 205+4}, 2^4, 2, 7\right)$.

**Remark 2.3.4.** Pierre de Fermat (1601-1665) was a French magistrate who was very interested in mathematics. He is especially known for the statement that there are no nonzero integers $x, y, z$ with $x^n + y^n = z^n$ when $n$ is an integer greater than 2. For $n = 2$ there are lots of solutions.

Fermat wrote this statement in the margin of a book and claimed to have proved it; see also Diophatus' book on Arithmetic. Although many mathematicians have tried to prove this statement, it took more than 300 years before a rigorous proof was found. In 1994, Andrew Wiles finally came up with a proof, that uses very deep and advanced mathematics. Whether Fermat really proved the statement remains unclear.

Fermat's Little Theorem states that the multiplicative group $\mathbb{Z}/p\mathbb{Z}^\times$, where $p$ is a prime, contains precisely $p - 1$ elements. For arbitrary positive $n$, the number of elements in the multiplicative group $\mathbb{Z}/n\mathbb{Z}^\times$ is given by the so-called *Euler totient function*.

**Definition 2.3.5** (Euler totient function)**.** The Euler totient function $\Phi\colon \mathbb{N} \to \mathbb{N}$ is defined by $\Phi(n) = \left|\mathbb{Z}/n\mathbb{Z}^\times\right|$ for all $n \in \mathbb{N}$ with $n > 1$, and by $\Phi(1) = 1$.

**Example 2.3.6.** Below the values of the Euler totient function are listed for all positive integers up to 20.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Phi(n)$ | 1 | 1 | 2 | 2 | 4 | 2 | 6 | 4 | 6 | 4 | 10 | 3 | 12 | 6 | 8 | 8 | 16 | 6 | 18 | 8 |

Table 2.6: Euler totient function

**Theorem 2.3.7** (Euler Totient)**.** *The Euler totient function satisfies the following properties.*

1. *Suppose that n and m are positive integers. If* $\gcd(n,m) = 1$*, then*
$$\Phi(n \cdot m) = \Phi(n) \cdot \Phi(m) \tag{2.25}$$
2. *If p is a prime and n a positive integer, then*
$$\Phi(p^n) = p^n - p^{n-1} \tag{2.26}$$
3. *If a is a positive integer with distinct prime divisors* $p_1, \ldots, p_s$ *and prime factorization* $a = \prod_{i=1}^{s}(p_i)^{n_i}$ *then*
$$\Phi(a) = \prod_{i=1}^{s}\left((p_i)^{n_i} - (p_i)^{n_i-1}\right) \tag{2.27}$$
4. *The Euler Totient function satisfies the follwoing recursion:*
$$\Phi(1) = 1 \tag{2.28}$$
*and*
$$\Phi(n) = n - \sum_{d \in \{d \in \mathbb{N} | d | n\}} (\Phi(d)) \tag{2.29}$$

*Proof.*

**Part (a).**

Suppose that $n$ and $m$ are two positive integers which are coprime. If $a$ and $b$ are two integers congruent modulo $n \cdot m$, then they are also congruent modulo $n$ and modulo $m$.

Moreover, if an integer $a$ is relatively prime to $n \cdot m$, then clearly $a$ is also relatively prime to both $n$ and $m$. Consequently, the map $F : \mathbb{Z}/n \cdot m\mathbb{Z}^\times \to \mathbb{Z}/n\mathbb{Z}^\times \times \mathbb{Z}/m\mathbb{Z}^\times$ defined by $F(a(\text{mod } n \cdot m)) = (a(\text{mod } n), a(\text{mod } m))$ is well defined.

The Chinese Remainder Theorem implies that for each pair $(b(\text{mod } n), c(\text{mod } m))$ in $\mathbb{Z}/n\mathbb{Z}^\times \times \mathbb{Z}/m\mathbb{Z}^\times$ there is one and only one class $a(\text{mod } n \cdot m)$ of $\mathbb{Z}/n \cdot m\mathbb{Z}^\times$ which is mapped onto the pair $(b(\text{mod } n), c(\text{mod } m))$ by $F$. This proves that $F$ is a bijection. So $\mathbb{Z}/n \cdot m\mathbb{Z}^\times$ and $\mathbb{Z}/n\mathbb{Z}^\times \times \mathbb{Z}/m\mathbb{Z}^\times$ have the same number of elements. This proves that $\Phi(n \cdot m) = \Phi(n) \cdot \Phi(m)$.

**Part (b).**

Suppose that $p$ is a prime and $n$ a positive integer. The integers $a$ which are not relatively prime to $p^n$ are exactly the multiples of $p$. As there are $p^{n-1}$ multiples of $p$ in $\{1,...,p^n\}$, we find $\Phi(p^n) = p^n - p^{n-1}$.

**Part (c).**

Part (c) is a direct consequence of the two other statements.

**Part (d).**

The first part is obvious, so we concentrate on proving the second Part.

The set $\{1,...,n\}$ is the disjoint union of the sets $V(n,d) = \{m \in \{1,...,n\} \,|\, \gcd(m,n) = d\}$ where $d$ runs through the set of positive divisors of $n$ (in which case also $\frac{n}{d}$ runs through the set of positive divisors of $n$).

For multiples $m,n$ of $d$, we have $\gcd(m,n) = d$ if and only if $\gcd\left(\frac{m}{d},\frac{n}{d}\right) = 1$. The set $V(n,d)$ therefore also equals $d \cdot V\left(\frac{n}{d},1\right)$.

But $|V(m,1)| = \Phi(m)$, so $V(n,d)$ contains precisely $\Phi(\frac{n}{d})$ elements. Consequently, $n = \sum_{d\in\{d\in\mathbb{N}|d|n\}} \left(\Phi(\frac{n}{d})\right) = \sum_{d\in\{d\in\mathbb{N}|d|n\}} \left(\Phi(d)\right)$.

Taking apart the summand $\Phi(n)$ (occurring for $d = n$), and bringing the remaining summation to the other side, we find the required formula.

□

**Example 2.3.8.** By the Euler Totient Theorem we find:

$$\Phi(100) = \Phi(2^2 \cdot 5^2) = \Phi(2^2) \cdot \Phi(5^2) = \left(2^2 - 2\right) \cdot \left(5^2 - 5\right) = 40 \qquad (2.30)$$

**Example 2.3.9.** The number of invertible elements in $\mathbb{Z}/6\mathbb{Z}$ can be computed with the formula of Part (4) of the theorem.

$\Phi(6) = 6 - \Phi(1) - \Phi(2) - \Phi(3) = 6 - 1 - 1 - 2 = 2$

Let $n$ be a prime. Then $\Phi(n) = n - 1$. So, by Fermat's Little Theorem we have $(a(\mathrm{mod}\ n))^{\Phi(n)} = 1(\mathrm{mod}\ n)$ for all integers $a$ that are not a multiple of $n$.

This statement can be generalized to arbitrary $n$.

*Leonard Euler*

**Theorem 2.3.10** (Euler's Theorem)**.** *Suppose $n$ is an integer with $n \geq 2$. Let $a$ be an element of $\mathbb{Z}/n\mathbb{Z}^{\times}$. Then $a^{\Phi(n)} = 1$.*

*Proof.* The proof of the theorem almost literally follows the second proof of Fermat's Little Theorem.

Suppose $a$ in $\mathbb{Z}/n\mathbb{Z}^{\times}$. Consider the map

$$M_a \colon \mathbb{Z}/n\mathbb{Z}^{\times} \to \mathbb{Z}/n\mathbb{Z}^{\times}, z \longmapsto a \cdot z \tag{2.31}$$

In other words, $M_a$ is multiplication by $a$. By the Invertibility of Products, this map is well defined. Moreover, the map is bijective. Indeed, its inverse is given by $M_{a^{-1}}$, multiplication by $a^{-1}$. As a result we see that the product of all elements in $\mathbb{Z}/n\mathbb{Z}^{\times}$ equals not only $\prod_{z \in \mathbb{Z}/n\mathbb{Z}^{\times}} z$ but also $\prod_{z \in \mathbb{Z}/n\mathbb{Z}^{\times}} (M_a(z))$. The products are over the same set of elements. They are just taken in different order, but that does not influence the result. In other words, the products are equal. But the latter product equals $\prod_{z \in \mathbb{Z}/n\mathbb{Z}^{\times}} (a \cdot z) = a^{\Phi(p)} \cdot \prod_{z \in \mathbb{Z}/n\mathbb{Z}^{\times}} z$. By Invertibility of Products the product $\prod_{z \in \mathbb{Z}/n\mathbb{Z}^{\times}} z$ is invertible, so, multiplying both sides of the above equation by its inverse, we find $a^{\Phi(n)} = 1$. This proves the theorem.

□

**Example 2.3.11.** The set $\mathbb{Z}/15\mathbb{Z}^{\times}$ contains 8 elements, one of them being $7 (\bmod 15)$. For this element we have integer2.eqsmod $\left(7^8, 49^4, 4^4, 1^2, 1, 15\right)$

This in accordance with Euler's Theorem.

Let $n$ be an integer. The *order* of an element $a$ in $\mathbb{Z}/n\mathbb{Z}^{\times}$ is the smallest positive integer $m$ such that $a^m = 1$. By Euler's Theorem the order of $a$ exists and is at most $\Phi(n)$. More precise statements on the order of elements in $\mathbb{Z}/n\mathbb{Z}^{\times}$ can be found in the following result.

**Theorem 2.3.12** (Orders)**.** *Let $n$ be an integer greater than 1.*

1. *If $a \in \mathbb{Z}/n\mathbb{Z}$ satifies $a^m = 1$ for some positive integer $m$, then $a$ is invertible and its order divides $m$.*
2. *For all elements $a$ in $\mathbb{Z}/n\mathbb{Z}^{\times}$ the order of $a$ is a divisor of $\Phi(n)$.*
3. *If $\mathbb{Z}/n\mathbb{Z}$ contains an element $a$ of order $n-1$, then $n$ is prime.*

*Proof.*

**Part (a).**

Suppose $a \in \mathbb{Z}/n\mathbb{Z}$ satifies $a^m = 1$ for some integer $m$. Then, since $a \cdot a^{m-1} = 1$, the element $a$ is invertible with inverse $a^{m-1}$.

Let $k$ be the order of $a$, and set $q = \mathrm{quot}(m,k)$ and $r = \mathrm{rem}(m,k)$. Then $(a(\mathrm{mod}\ n))^r$ equals $(a(\mathrm{mod}\ n))^{m-q\cdot k} = (a(\mathrm{mod}\ n))^m \cdot \left((a(\mathrm{mod}\ n))^k\right)^{-q}$, which is equal to 1. By the definition of order, the above implies that $r$ is equal to 0, which proves the first part of the theorem.

**Part (b).**

The second part follows immediately from the first statement of the theorem and Euler's Theorem.

**Part (c).**

As for the last statement, $\Phi(n) = n-1$ if and only if all integers between 0 and $n-1$ have greatest common divisor 1 with $n$. This implies that $n$ is prime.

□

**Example 2.3.13.** The element $7(\mathrm{mod}\ 15)$ of $\mathbb{Z}/15\mathbb{Z}$ satisfies integer2.eqsmod $(7^4, 49^2, 4^2, 1, 15)$

Hence its order divides 8, which is the order of $\mathbb{Z}/15\mathbb{Z}^{\times}$.

**Remark 2.3.14.** Fermat's Little Theorem and the Theorem on orders form a basis for various prime tests. Suppose, for example, that given some large integer $n$ one wants to decide whether $n$ is prime. Choosing a random integer $a$ one can check whether $a^{n-1} \equiv 1 \pmod{n}$.

If this is *not* the case, one can conclude that $a$ is composite. However, when $a^{n-1} \equiv 1 \pmod{n}$, one is still not able to decide that $n$ is prime, but one has at least a good chance that it is. Repeating this test a couple of times increases the probability of a correct answer to the question whether $n$ is prime.

However, there are composite integers $n$, so-called *Carmichael numbers*, for which it is very likely that the test will indicate that $n$ is prime. A Carmichael number is a composite integer $n$ such that $a^{n-1} \equiv 1 \pmod{n}$ for all integers $a$ with $\gcd(a,n) = 1$. (If $\gcd(a,n) > 1$, then $a(\mathrm{mod}\ n)$ is not invertible, so $\gcd(a,n) \neq 1$.) The only Carmichael number less than 1000 is 561.

**Definition 2.3.15.** An element $a$ from $\mathbb{Z}/p\mathbb{Z}$ is called a *primitive element* of $\mathbb{Z}/p\mathbb{Z}$ if every element of $\mathbb{Z}/p\mathbb{Z}^{\times}$ is a power of $a$.

**Example 2.3.16.** The element 2 is a primitive element in $\mathbb{Z}/11\mathbb{Z}^{\times}$. Indeed its powers are $2^1 = 2, 2^2 = 4, 2^3 = 8, 2^4 = 5, 2^5 = 10, 2^6 = 9, 2^7 = 7, 2^8 = 3, 2^9 = 6, 2^{10} = 1$. It is not primitive in $\mathbb{Z}/7\mathbb{Z}^{\times}$ as $2^3 = 1$ in $\mathbb{Z}/7\mathbb{Z}^{\times}$.

For every prime $p$ there exist primitive elements; but we cannot say a priori which ones.

**Theorem 2.3.17.** *For each prime $p$ there exists a primitive element in $\mathbb{Z}/p\mathbb{Z}$.*

## 2.4 The RSA cryptosystem

Suppose that you want to buy your favorite book or music CD at an internet book or record shop. To submit the order to the shop, you are required to supply various private data, such as your name, home address and credit card information. However, if you send this information unprotected over the internet, it can be intercepted by unreliable persons. To secure your personal data, the internet shop makes use of so-called *public-key cryptography*.

This means the following. The shop supplies every customer with a (public) function $E$. With this function the customer encrypts his or her personal data, denoted by data, into $E(\text{data})$. The customer then sends the encrypted message $E(\text{data})$ to the shop.

Besides the encryption function $E$ the shop also has a (secret) decryption function $D$ which can be used to decrypt the message $E(\text{data})$. This means that $E$ and $D$ have the property that $D(E(\text{data})) = \text{data}$. The idea is that, in case one does not know $D$, it is hard (or almost impossible) to discover data from the encrypted message $E(\text{data})$. Only the trusted shop can find the personal information in data by applying $D$ to $E(\text{data})$.

We discuss the RSA cryptosystem, an example of a public-key crypto system. The RSA cryptosystem (RSA stands for Rivest, Shamir, and Adleman, the three mathematicians who designed the system) is a modern cryptosystem based on modular arithmetic. The basis for the RSA cryptosystem is Euler's Theorem. Its security is based on the difficulty of factoring large integers.

In the RSA cryptosystem the data to be encrypted is assumed to be an integer, $x$ say. (If the data is computer data, one may view the string of bits representing the data as the binary representation of the integer $x$.)

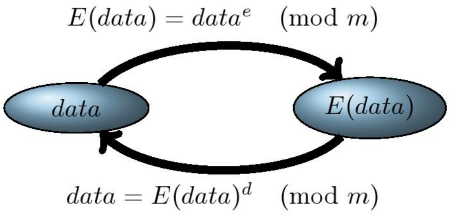
The encryption function $E$, which is public, makes use of two integers, the *modulus* $m$, which is the product of two primes, and the *encoding number* $e$. These two integers are usually called the *public keys*. The *secret key* is a number $d$, called the decoding number, which is used for the decoding function $D$.

**Definition 2.4.1** (RSA Decription and Encryption)**.** Suppose that $p$ and $q$ are distinct primes. Let $m = p \cdot q$ and $d$ and $e$ be two integers such that $d \cdot e \equiv 1 \pmod{(p-1)\cdot(q-1)}$.

Then the encryption function $E$ and decryption function $D$ of an RSA cryptosystem are defined by

- $E(x) = \text{rem}(x^e, m)$;
- $D(x) = \text{rem}(x^d, m)$.

The RSA cryptosystem enables the owner of the decryption function $D$ to recover an encrypted message, provided the input integer $x$ is not too large. In practice, this can easily be achieved by splitting the input for the encryption in small separated pieces and subsequently applying $D$ and $E$ to the individual pieces.



**Theorem 2.4.2** (RSA Decoding)**.** *Suppose that $x$ is a positive integer less than both $p$ and $q$. Then $D(E(x)) = x$.*

*Proof.* Suppose that $x$ is a positive integer less than both $p$ and $q$. Then $D(E(x)) \equiv x^{d \cdot e} \pmod m$. By Euler's Theorem we have $x^{(p-1)\cdot(q-1)} \equiv 1 \pmod m$. As $d \cdot e \equiv 1 \pmod{(p-1)\cdot(q-1)}$, we even have $x^{d \cdot e} \equiv x \pmod m$. Since $x$ is less than both $p$ and $q$, it is certainly less than $m$. In particular, we find $x$ to be equal to $D(E(x))$.

□

How secure is RSA? The security of RSA depends of course on the difficulty of computing the decoding number $d$. To find this number it is necessary to know the two primes $p$ and $q$. Once you know these primes it is a piece of cake to find $d$. But, as noticed in the section on Example 1.5.2, factoring the modulus $m = p \cdot q$ into $p$ and $q$ is an extremely time-consuming task (provided $p$ and $q$ are chosen sufficiently large): if one chooses two very big primes $p$ and $q$, then, with current methods, it is almost impossible to find the factorization of the modulus $m = p \cdot q$.

So, at the moment, the RSA cryptosystem is believed to provide excellent security. But it remains unclear whether there exist fast methods to crack the code or not.

## 2.5 Exercises

### 2.5.1 Arithmetic modulo an integer

**Exercise 2.5.1.** Show that if $a$ and $b$ leave the same remainder on division by $n$, then $a \equiv b \pmod n$.

**Exercise 2.5.2.** Show that if $a$ and $b$ are congruent modulo $m$, then $a^2$ and $b^2$ are congruent modulo $m$.

Give an example to show that $a^2$ and $b^2$ are not necessarily congruent modulo $m^2$.

**Exercise 2.5.3.** If $a$ is congruent to 2 modulo 5, then to which of the integers 0, 1, 2, 3, 4 is $a^3 - 3 \cdot a + 1$ congruent?

**Exercise 2.5.4.** Suppose that the positive integers $a$ and $b$ leave remainders 3 and 4, respectively, on division by 7. Use modular arithmetic to show that $a \cdot b$ leaves remainder 5 on division by 7.

**Exercise 2.5.5.** Divisibility by 4 of a number which is written in the decimal system can be tested as follows: the number is divisible by 4 if and only if the number formed by the two last digits is divisible by 4.

Prove this statement.

**Exercise 2.5.6.** Formulate an 8-test (i.e., a test for deciding divisibility by 8) for numbers in the decimal system.

How does one decide divisibility by 8 for a binary number?

**Exercise 2.5.7.** Formulate a test and prove its correctness for divisibility by $a-1$ in the $a$-ary system.

**Exercise 2.5.8.** Prove that $n^4 + n^2 + 1$ is divisible by 3 if $n > 0$ is not divisible by 3.

**Exercise 2.5.9.** Prove the following statements:

1. $13 | 10^6 - 1$.
2. $17 | ((10^8) + 1)$.
3. If $n \not\equiv 0 \pmod 5$, then $n^4 + 64$ is not prime.
4. The number $2^{1000} + 5$ is divisible by 3.
5. For every $n > 0$ we find that 3 is a divisor of $2^{2 \cdot n} - 1$.

**Exercise 2.5.10.** Determine the multiplicative inverses of the given elements or show that this inverse does not exist.

1. $3 \in \mathbb{Z}/37\mathbb{Z}$;
2. $4 \in \mathbb{Z}/14\mathbb{Z}$.

**Exercise 2.5.11.** Fermat conjectured that numbers of the form $2^{2^n} + 1$ are prime. For $n = 5$ this conjecture does not hold. Prove, with the help of the following observations, that $641 | ((2^{2^5}) + 1)$.

1. $641 = 2^9 + 2^7 + 1$ and so integer2.eqsmod$\left(2^7 \cdot 1, 2^7 \cdot \left(2^2 + 1\right), -1, 641\right)$.

2. $2^4 \equiv -(5^4) \pmod{641}$.

**Exercise 2.5.12.** The binomial coefficient $\left(\begin{array}{c} p \\ k \end{array}\right)$ (pronounce: $p$ choose $k$) equals $\frac{p\cdot(p-1)\cdot\ldots\cdot(p-k)}{k\cdot(k-1)\cdot\ldots\cdot 2\cdot 1}$

If $p$ is prime and $0 < k < p$, then the binomial coefficient $\left(\begin{array}{c} p \\ k \end{array}\right)$ is divisible by $p$. Prove this! In addition show that for all $x$ and $y$ in $\mathbb{Z}/p\mathbb{Z}$ the equality $(x+y)^p = x^p + y^p$ holds.

**Exercise 2.5.13.** What are the invertible elements of $\mathbb{Z}/n\mathbb{Z}$ where $n$ is an element of $\{2,6,12\}$?

**Exercise 2.5.14.** Let $p$ be a prime. What are the invertible elements of $\mathbb{Z}/p^2\mathbb{Z}$?

**Exercise 2.5.15.** Which integers are congruent to 7 modulo 17: 1734, 1127 or 1251?

**Exercise 2.5.16.** Which integers represent an invertible congruence class modulo 17 and which a zero divisor: 1734, 1127, 1251?

**Exercise 2.5.17.** Find for each of the following statements a counterexample.

If $a$ is an invertible element in $\mathbb{Z}/n\mathbb{Z}$, and $b$ an arbitrary nonzero element, then $a\cdot b$ is invertible.

If $a$ and $b$ are invertible elements in $\mathbb{Z}/n\mathbb{Z}$, then $a+b$ is invertible.

If $a$ and $b$ are zero divisors in $\mathbb{Z}/n\mathbb{Z}$, then $a+b$ is also a zero divisor.

**Exercise 2.5.18.** Let $p$ and $q$ be distinct primes. What are the invertible elements of $\mathbb{Z}/p\cdot q\mathbb{Z}$?

## 2.5.2 Linear congruences

**Exercise 2.5.19.** Solve each of the following linear congruences:

1. $2\cdot x \equiv 37 \pmod{21}$
2. $5\cdot x \equiv 15 \pmod{25}$
3. $3\cdot x \equiv 7 \pmod{18}$

**Exercise 2.5.20.** Solve the following system of linear congruences: $2\cdot x \equiv 37 \pmod{5}$ and $3\cdot x \equiv 48 \pmod{7}$

**Exercise 2.5.21.** Solve the following system of linear congruences: $x + y \equiv 6 \pmod{11}$ and $2\cdot x - y \equiv 8 \pmod{11}$

**Exercise 2.5.22.** Find the smallest positive $x$ equal to 15 modulo 37 and 13 modulo 42.

Similarly, find the smallest positive $x$ equal to 17 modulo 42 and 13 modulo 49.

## 2.5.3 The theorems of Fermat and Euler

**Exercise 2.5.23.** Is the converse of Fermat's Little Theorem,

'if $x^{p-1} \equiv 1 \pmod{p}$ for all $x$ not equal to $0(\mathrm{mod}\ p)$, then $p$ is a prime'

also true?

**Exercise 2.5.24.** Determine the following remainders: $\mathrm{rem}(12312^{112311}, 7)$, $\mathrm{rem}(13452^{5323}, 5)$ and $\mathrm{rem}(5332^{11322}, 11)$.

**Exercise 2.5.25.** The hypothesis that an integer $n$ is prime if and only if it satisfies the condition that $2^n - 2$ is divisible by $n$ is called the 'Chinese Hypothesis'. Leibniz, a famous mathematician from the 17th-18th century, believed to have proved that this congruence indeed implies that $n$ is prime. However, although this condition is necessary for $n$ to be prime, it is not sufficient. For example, $2^{341} - 2$ is divisible by 341, but $341 = 11 \cdot 31$ is composite.

Prove that $2^{341} - 2$ is indeed divisible by 341.

**Exercise 2.5.26.** What value does the Euler totient function take on the integers 334, 231, and 133?

**Exercise 2.5.27.** How many zero divisors has $\mathbb{Z}/n\mathbb{Z}$?

**Exercise 2.5.28.** What is the order of $2(\mathrm{mod}\ 35)$ in $\mathbb{Z}/35\mathbb{Z}$? And of $4(\mathrm{mod}\ 35)$?

**Exercise 2.5.29.** Suppose that $x$ is an element of order $\Phi(n)$ in $\mathbb{Z}/n\mathbb{Z}$. Then every invertible element of $\mathbb{Z}/n\mathbb{Z}$ is a power of $x$. Prove this!

## 2.5.4 The RSA cryptosystem

**Exercise 2.5.30.** Consider the RSA cryptosystem with modulus 2623 and with encoding number $v = 37$.

If we represent the letters a, b, c, ..., z by the numbers 01, 02, ..., 26, respectively, and a space by 00, then try to decode the following text, where in each group of four figures a pair of these symbols is encoded:

0249 1133 1279 1744 0248 1188 1220 1357 1357.

**Exercise 2.5.31.** Consider the RSA cryptosystem with modulus 2623 and with encoding number $v = 37$.

If we represent the letters a, b, c, ..., z by the numbers 01, 02, ..., 26, respectively, and a space by 00, then how do you encode the text 'math is beautiful'?

# Chapter 3

# Polynomials

In this chapter we extend calculation with integers to calculation with polynomials, expressions in which, beside scalars (from $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Z}/n\mathbb{Z}$) also an indeterminate occurs. You have already seen polynomials when solving quadratic equations and plotting graphs of quadratic functions. For polynomials we introduce concepts related to the concepts divisor, gcd, etc., which we have introduced for integers.

## 3.1 The notion of a polynomial

Let $R$ be one of the rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/n\mathbb{Z}$.

**Definition 3.1.1** (Polynomials)**.** A *polynomial* over $R$ in the *indeterminate* $X$ is an expression of the form $a_0 + a_1 \cdot X + ... + a_n \cdot X^n$, where $n \in \mathbb{N}$, $a_0, ..., a_n \in R$ and $X$ is an indeterminate.

**Remark 3.1.2.** The following notions are connected to the definition.

- The name of the indeterminate chosen here is $X$. However, it could be any free symbol, that is, any symbol to which no meaning or value has been assigned.
- The elements $a_0, ..., a_n$ are called the *coefficients* of the polynomial.

  Given the name of the indeterminate, the polynomial is uniquely determined by the assignment of a coefficient $a_k$ to each natural number $k$ in such a way that $a_k$ is nonzero for only finitely many $k$.
- The polynomial is built up from *terms* of the form $a_k \cdot X^k$ where $k \in \mathbb{N}$.
- The powers $X^k$ of $X$, for which the coefficient $a_k$ is nonzero, are called the *monomials* of the polynomial.

**Remark 3.1.3.** The summation symbols in a polynomial express the fact that the order of the terms in the summation is immaterial. For instance, $a_0 + a_1 \cdot X + ... + a_n \cdot X^n = a_n \cdot X^n + ... + a_1 \cdot X + a_0$.

**Example 3.1.4.** Consider the polynomial $X^3+3\cdot X^2+X-2$. The coefficients are integers, so we can view the polynomial as an element of $\mathbb{Z}[X]$. As such, its terms are $X^3$, $3\cdot X^2$, $X$, and $-2$. Its monomials are $X^3$, $X^2$, and $X$.

If the ring of coefficients is $\mathbb{Z}/3\mathbb{Z}$, then the expression $3\cdot X^2$ disappears and so $X^2$ is no longer a monomial of the polynomial.

When speaking about a polynomial in $X$ over $R$, we refer to a polynomial with coefficients in $R$ in the indeterminate $X$.

We also say polynomial in $X$, or over $R$, or just polynomial if no confusion is possible about the ring of coefficients or the indeterminate $X$.

We write $R[X]$ for the set of all polynomials over $R$ in the indeterminate $X$.

Two polynomials in $R[X]$ are equal if the corresponding coefficients are equal. Polynomials of the form $a$ with $a\in R$ are called *constant*.

Using the *summation notation* we also write $a_0+a_1\cdot X+...+a_n\cdot X^n=\sum_{k=0}^{n} a_k\cdot X^k$.

A polynomial in $X$ is often denoted by a symbol like $a$, but sometimes also by $a(X)$ to emphasize the dependence on $X$.

Let $a=a_0+a_1\cdot X+...+a_n\cdot X^n$ and $b=b_0+b_1\cdot X+...+b_m\cdot X^m$ be two polynomials in $R[X]$. To define their sum and product it is convenient to assume $m=n$. This can always be achieved by adding terms of the form $0\cdot X^k$.

**Definition 3.1.5.** The set of polynomials $R[X]$ provided with the addition and multiplication specified below is called a *polynomial ring*.

- The *sum* of the polynomials $a$ and $b$ is the polynomial $a+b=\sum_{k=0}^{m}(a_k+b_k)\cdot X^k$.
- The *product* of the two polynomials $a$ and $b$ is the polynomial $a\cdot b=c_0+c_1\cdot X+...+c_{2\cdot m}\cdot X^{2\cdot m}$ where $c_k=a_0\cdot b_k+a_1\cdot b_{k-1}+...+a_k\cdot b_0$.

**Remark 3.1.6.** The definition of the product looks rather complicated, but becomes easier to grasp once you realize that it comes down to expanding the product of $a$ and $b$ as usual and replacing products like $c\cdot X^m\cdot d\cdot X^n$ by $c\cdot d\cdot X^{m+n}$, where $c$ and $d$ are elements of the ring $R$.

**Example 3.1.7.** Let $a=X^3+2\cdot X+1$ and $b=X^2+3\cdot X+2$.

Inside $\mathbb{R}[X]$ we have $a+b=X^3+X^2+5\cdot X+3$ and $a\cdot b=X^5+3\cdot X^4+4\cdot X^3+6\cdot X^2+7\cdot X+2$.

However, inside $\mathbb{Z}/3\mathbb{Z}[X]$ we have $a+b=X^3+X^2+2\cdot X$ and $a\cdot b=X^5+X^3+X^2+X-1$.

**Example 3.1.8.** The product rule allows us to write some very long polynomials very concisely. For instance, the left-hand side of the following equation only needs a few symbols, but, when fully written out as a polynomial, the right-hand side needs, in general, $n+1$ terms:

$$(1+X)^n=\sum_{k=0}^{n}\left(\begin{array}{c} n \\ k \end{array}\right)\cdot X^k \tag{3.1}$$

**Remark 3.1.9.** The sum rule allows us to repeat terms with the same monomials in an expression of a polynomial. For instance, the monomial $X^2$ occurs twice at the left-hand side of the following equation, but only once at the right-hand side: $X+2\cdot X^2+3\cdot X^3-4\cdot X^2 = X+(-2)\cdot X^2+3\cdot X^3$.

Polynomial rings have an arithmetic structure that shows many similarities with the integers. For instance, the following rules hold for polynomials (for all $a$, $b$, $c$ in $R[X]$).

- $a+b=b+a$ (commutativity of addition);
- $a\cdot b=b\cdot a$ (commutativity of multiplication);
- $(a+b)+c=a+((b+c))$ (associativity of addition);
- $(a\cdot b)\cdot c=a\cdot b\cdot c$ (associativity of multiplication);
- $a\cdot(b+c)=a\cdot b+a\cdot c$ (distributivity of multiplication over addition).

The proofs of these rules are not difficult, but some of them involve quite a bit of writing. By way of example, the commutativity of multiplication follows directly from the equality $a_0\cdot b_k+a_1\cdot b_{k-1}+...+a_k\cdot b_0=b_0\cdot a_k+b_1\cdot a_{k-1}+...+b_k\cdot a_0$ (the expression on the right-hand side is, apart from the order of the factors in each term, the expression on the left-hand side read backwards), where the left-hand side is the $k$-th coefficient of $a\cdot b$, and the right-hand side is the $k$-th coefficient of $b\cdot a$.

For polynomials, we will discuss division with remainder, gcd, and more notions that are already familiar for the integers.

## 3.2 Division of polynomials

Let $R$ be one of the rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/n\mathbb{Z}$.

**Definition 3.2.1.** Let $a=a_0+a_1\cdot X+...+a_n\cdot X^n$ be a polynomial in $R[X]$ with $a_n\neq 0$. We call

- $a_n\cdot X^n$ the *leading term* and $a_n$ the *leading coefficient* of $a$. The leading term of $a$ is denoted by $\mathrm{lt}(a)$ and the leading coefficient by $\mathrm{lc}(a)$.
- $n$ the *degree* of the polynomial $a$. The degree of $a$ is denoted $\mathrm{degree}(a)$.

**Example 3.2.2.** Consider the polynomial $X^3+3\cdot X^2+X-2$ over $\mathbb{Z}$. It has degree 3 and its terms are $X^3$, $3\cdot X^2$, $X$, and $-2$. The leading term is $X^3$ and the leading coefficient is 1.

If all the coefficients of a polynomial $a$ are equal to 0, then $a=0$ (the zero polynomial). It is practical to define the degree of the zero polynomial to be $-\infty$.

A polynomial of degree 1 is also called a *linear polynomial*. A polynomial is said to be *monic* if its leading coefficient is equal to 1.

Suppose that $R$ has no nonzero elements whose product is 0. If the nonzero polynomial $a$ has leading coefficient $a_n$ and the nonzero polynomial $b_m$ has leading coefficient $b$, then the leading coefficient of $a \cdot b$ is $a_n \cdot b_m$, as follows from the definition of the product. In that case we have the following results.

**Theorem 3.2.3** (Degree Formulas)**.** *Let $R$ be a field and $a$ and $b$ polynomials over $R$ in $X$. Then the following assertions hold.*

1. $\text{degree}\,(a\cdot b) = \text{degree}\,(a) + \text{degree}\,(b)$ .
2. $\text{degree}\,(a+b) \leq max\,(\text{degree}\,(a)\,, \text{degree}\,(b))$.
3. *If $a\cdot b = 0$, then $a = 0$ or $b = 0$.*

*Proof.* The first part of the proof is obvious from the above. Note that the statement also holds if $a$ and/or $b$ is the zero polynomial. Here, we use obvious rules like $-\infty + m = -\infty$ for any integer $m$.

The second part of the proof is a direct consequence of the definition of addition of polynomials.

In order to prove the third part, suppose that $a\cdot b = 0$. Then, according to the first assertion, the degree of $a$ or $b$ is $-\infty$, and hence $a$ or $b$ equals zero.

□

For the polynomial ring $R[X]$, where $R$ is a field, like $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Z}/p\mathbb{Z}$, with $p$ prime, we introduce, similarly to the integer case, division with remainder. In the integer case this involves the absolute value as a kind of measure. For polynomials, the appropriate measure is the degree.

We start with the more general situation where $R$ is an arbitrary ring.

**Definition 3.2.4.** Suppose that $a$ and $b$ are polynomials in $R[X]$, where $R$ is a field. The polynomial $b$ is called a *divisor* of $a$ if there exists a polynomial $q \in R[X]$ such that $a = q\cdot b$. We use the notation $b|a$ to denote that $b$ divides $a$.

**Example 3.2.5.** The polynomial $X^2-1$ is a divisor of $X^6-1$. Indeed, we have $X^6-1 = \left(X^2-1\right)\cdot\left(X^4+X^2+1\right)$.

**Example 3.2.6.** In the definition of divisor we restrict to fields in order to avoid various problems. For instance, in $\mathbb{Z}/9\mathbb{Z}$ the two equalities $6\cdot X^6 = 3\cdot X^2\cdot 5\cdot X^4$ and $6\cdot X^6 = 3\cdot X^2\cdot 2\cdot X^4$ show that a quotient need not be unique.

Instead of $b$ is a divisor of $a$, we also say that $a$ is a *multiple* of $b$, or $a$ is *divisible* by $b$, or $b$ is a *factor* of $a$, or $b$ *divides* $a$.

There is a division algorithm for polynomials that is much like the one for integers. It can be used to determine both quotient and remainder. For this algorithm to work, however, we need the ring of coefficients to be a field.

**Theorem 3.2.7** (Division with Remainder)**.** *Let $R$ be a field and suppose that $a$ and $b$ are two polynomials in $R[X]$ with $b \neq 0$. Then there are polynomials $q$ (the quotient) and $r$ (the remainder) such that $a = q \cdot b + r$ and* $\text{degree}(r) < \text{degree}(b)$.
*The polynomials $q$ and $r$ are uniquely determined. They are called the* quotient *and* remainder *of $a$ divided by $b$ and are denoted by* $\text{quot}(a,b)$ *and* $\text{rem}(a,b)$, *respectively, just like for integers.*
*If $b \neq 0$ divides $a$, the quotient is denoted by $\frac{a}{b}$ or $a/b$.*

*Proof.* (Compare this proof with the proof of Properties of Divisors for integers.)

The proof is divided into two parts, one part for existence, the other for uniqueness.

**There exist polynomials $q$ and $r$ as in the theorem.**

Let $n$ be the degree of $a$ and $m$ the degree of $b$. If $n < m$, then $q = 0$ and $r = a$ satisfy the requirements. Assume therefore that $n \geq m$. As $b \neq 0$, we have $m \geq 0$, so $n \geq 0$, and therefore, $a \neq 0$.

We proceed to prove the assertion by induction on $n$.

First assume that $n = 0$, i.e., $a$ is constant. Then also $m = 0$ and $b$ is constant. In this case, $q = a/b$ and $r = 0$ fulfill the requirements.

Now suppose that $n > 0$ and that (the induction hypothesis) the existence of polynomials $q$ and $r$ has been proved for polynomials of degree at most $n-1$. Let $a_n$ be the leading coefficient of $a$ and $b_m$ the leading coefficient of $b$. Consider the polynomial $a' = a - \frac{a_n}{b_m} \cdot b \cdot X^{n-m}$. The leading term of the polynomial subtracted from $a$ has been chosen so that the degree of $a'$ is less than $n$. According to the induction hypothesis there are polynomials $q'$ and $r'$ with $a' = q' \cdot b + r'$ where the degree of $r'$ is less than $m$. Now set $q = q' + \frac{a_n}{b_m} \cdot X^{n-m}$ and $r = r'$. Then $q$ and $r$ satisfy the requirements of the theorem.

**The polynomials $q$ and $r$ are uniquely determined by the existence requirements of the theorem.**

Suppose that $a = q \cdot b + r$ with $\text{degree}(r) < \text{degree}(b)$ and also $a = q' \cdot b + r'$ with $\text{degree}(r') < \text{degree}(b)$ for certain polynomials $q$, $r$, $q'$, and $r'$.

Subtracting these two expressions of $a$ yields: $0 = (q - q') \cdot a + r - r'$. In particular, $(q - q') \cdot a = r' - r$. By Part 2 of the Degree Formulas, the degree of $r' - r$ is less than the degree of $a$, so, by Part 1 of the Degree Formulas, both sides of the equality must be equal to 0. In particular, $r' - r = 0$ and, as $a \neq 0$, also $q = q'$.

□

**Remark 3.2.8.** At various places in the proof of Division with Remainder Theorem we made use of the fact that in the field $R$ every nonzero element has an inverse.

**Example 3.2.9.** To determine the quotient $q$ and the remainder $r$ when dividing $a = 2 \cdot X^4 + X$ by $b = X^2 + 1$ in $\mathbb{Q}[X]$ we need the following steps.

- Compare the leading terms of $a$ and $b$. Subtract $2 \cdot X^2 \cdot b$ from $a$ in order to cancel the leading term of $a$: $a - 2 \cdot X^2 \cdot b = 2 \cdot X^4 + X - 2 \cdot X^2 \cdot \left(X^2 + 1\right) = -2 \cdot X^2 + X$. From this step we conclude that $2 \cdot X^2$ is a term of the quotient $q$. We now have $a = 2 \cdot X^2 \cdot b + \left(-2 \cdot X^2 + X\right)$. Since the degree of $(-2) \cdot X^2 + X$ is not less than the degree of $b$ we need a further step.
- Compare the leading terms of $(-2) \cdot X^2 + X$ and $b$ and subtract $(-2) \cdot b$ from $(-2) \cdot X^2 + X$. This yields $2 \cdot X^2 + X + 2 \cdot \left(X^2 + 1\right) = X + 2$. The resulting polynomial has degree less than the degree of $b$, so the division stops here. We conclude that the quotient $q$ satisfies $q = 2 \cdot X^2 - 2$ and the remainder $r$ satisfies $r = X + 2$.

It is easy to verify the identity $a = q \cdot b + r$, i.e., $2 \cdot X^4 + X = \left(2 \cdot X^2 - 2\right) \cdot \left(X^2 + 1\right) + (X + 2)$.

The Division with Remainder Theorem states that there exist a quotient $q$ and a remainder $r$, but it does not tell you how to find those two polynomials. As for the integers, a standard and well-known algorithm is *long division*. We describe (a variation of) this algorithm for finding $q$ and $r$.

**Algorithm 3.2.10** (Polynomial Division and Remainder)**.** • *Input: a polynomial a and a nonzero polynomial b, both in the indeterminate X, and with coefficients in a field.*

- *Output: the quotient q and remainder r of a upon division by b as a list* $[q, r]$.

```
PolyDivisionRemainder := **procedure**$(a, b)$
**local variables**
    $q := 0$ , $r := a$
    $n :=$ degree$(a)$ , $m :=$ degree$(b)$
**while** $n \geq m$ **do**
    $q := q + \frac{\mathrm{lc}(r)}{\mathrm{lc}(b)} \cdot X^{\mathrm{degree}(r) - \mathrm{degree}(b)}$
    $r := r - \frac{\mathrm{lc}(r)}{\mathrm{lc}(b)} \cdot X^{\mathrm{degree}(r) - \mathrm{degree}(b)} \cdot b$ , $n :=$ degree$(r)$
**return**
    $[q, r]$
```

*Proof.*

**Correctness.**

By construction we have $a = q \cdot b + r$ in each step of the while loop. Moreover, after termination the degree of $r$ is less than the degree of $b$. This proves correctness.

**Termination.**

Since the degree of $r$ decreases in each step of the while loop, this loop will end. Thus the algorithm terminates.

□

The following definitions are analogous to those for integers.

**Definition 3.2.11.** Let $R$ be a field and let $a, b \in R[X]$.

- A *common divisor* of $a$ and $b$ is a polynomial which divides both $a$ and $b$.
- A common divisor $d$ is called *greatest common divisor* (gcd) if, moreover, every common divisor of $a, b$ (not both zero) is a divisor of $d$.
- A *common multiple* of $a$ and $b$ is a polynomial which is divisible by both $a$ and $b$.
- A *least common multiple* (lcm) of $a$ and $b$ is a common multiple of $a$ and $b$ of minimal degree at least 0.

**Remark 3.2.12.** It is not obvious from the definition that gcd's exist. Existence would have been evident, however, if the definition had been: a common divisor of $a$ and $b$ of maximal degree (similar to the definition of common divisor for two integers). Both definitions will be shown to be equivalent, but the given definition turns out to be more convenient to set up the theory. Existence will be shown in Existence and Uniqueness of gcd.

**Remark 3.2.13.** A gcd is not unique: multiplying by a nonzero constant also provides a gcd. If we speak of *the gcd* of $a$ and $b$ we mean a gcd of $a$ and $b$ with leading coefficient equal to 1. This gcd is also denoted by $\gcd(a, b)$. Uniqueness of the gcd follows from the Existence and Uniqueness of gcd.

The concept gcd of $a$ and $b$ is only meaningful when the polynomials $a$ and $b$ are not both equal to the zero polynomial.

Two polynomials are called *relatively prime* if their gcd equals 1.

**Theorem 3.2.14** (Existence and Uniqueness of gcd)**.** *Suppose that $R$ is a field and $a$ and $b$ are polynomials in $R[X]$, which are not both the zero polynomial. Then a greatest common divisor of $a$ and $b$ exists, and, moreover, if $c$ and $d$ are two greatest common divisors of the polynomials $a, b$, then there is a constant $q \neq 0$ such that $q \cdot c = d$.*

*Proof.* The proof is divided into two parts, one part for existence, one part for uniqueness.

**There exists a gcd for $a$ and $b$.**

We show that a gcd in $R[X]$ can be found among the polynomials of the form $x \cdot a + y \cdot b$, where $x$ and $y$ are also polynomials. The polynomials $x \cdot a + y \cdot b$ are obviously divisible by every common divisor of $a$ and $b$. Let $d$ be a nonzero polynomial of the form $x \cdot a + y \cdot b$ of minimal degree. Then $d$ turns out to be a gcd. Every common divisor of $a$ and $b$ clearly divides $d$, it remains to show that $d$ divides $a$ and $b$. Take any $x \cdot a + y \cdot b$ and divide by $d$. This produces a relation $x \cdot a + y \cdot b = q \cdot d + r$, where the degree of $r$ is less than the degree of $d$. From this relation we infer that $r$ is also of the form $u \cdot a + v \cdot b$, so that $r$ must be 0 by the minimality of the degree of $d$. So $d$ divides any $x \cdot a + y \cdot b$, and in particular $a$ and $b$. So $d$ is a gcd of $a$ and $b$.

**Two gcd's of $a$ and $b$ differ by a nonzero constant factor.**

From the fact that $c$ and $d$ are both gcd's of $a$ and $b$, it follows that $c$ divides $d$ and that $d$ divides $c$. The former means that there is a polynomial $q$ with $d = c \cdot q$. Since $d$ also divides $c$, the Degree Formulas show that the degree of $q$ is 0. This means that $q$ is a nonzero constant.

□

**Example 3.2.15.** Consider the polynomials $f = 2 \cdot X^2 - 3 \cdot X - 2$ and $g = 4 \cdot X^2 - 1$. Viewed as polynomials over $\mathbb{Z}$, the polynomial $2 \cdot X + 1$ is a gcd of $f$ and $g$ and there is no monic gcd. Viewed as a polynomials over $\mathbb{Q}$ the polynomial $X + \frac{1}{2}$ is *the* gcd of $f$ and $g$.

The gcd of two polynomials can be determined similarly to the computation of the gcd for integers. It is of importance to factorization of polynomials, which in turn is useful for solving systems of polynomial equations.

In the following we will use, without explicit mentioning it, the following easy to prove facts: $\gcd(a,b) = \gcd(b,a)$, $\gcd(a,b) = \gcd(a, b - k \cdot a)$ (for every polynomial $k$), $\gcd(a,0) = a$.

**Algorithm 3.2.16** (Euclid's Algorithm for Polynomials)**.** • *Input: two polynomials a and b in $R[X]$, not both zero, where R is a field.*

- *Output: the gcd of a and b.*

```
PolyGCD := procedure(a, b)
local variables
  | c
while degree(b) > -1 do
  | c := a , a := b , b := rem(c, b)
return
  | $\frac{a}{\mathrm{lc}(a)}$
```

*Proof.*

**Termination.**

As degree $(b)$ goes strictly down at each step, termination is guaranteed.

**Correctness.**

Let $a_0$ and $b_0$ denote the input values of $a$ and $b$, respectively. Then the values of $a$ and $b$ at the end of each loop satisfy $\gcd(a,b) = \gcd(a_0,b_0)$ In computer science terms, this is an invariant of the algorithm. At the end we have $b = 0$ and so $a = \gcd(a,0) = \gcd(a_0,b_0)$. Division by $\operatorname{lc}(a)$ makes the gcd monic.

□

**Example 3.2.17.** In the spirit of the algorithm, we compute the gcd of $X^4 - 1$ and $X^6 - 1$:

$$\gcd\left(X^4-1,X^6-1\right)=\gcd\left(X^6-1,X^4-1\right)=\gcd\left(X^4-1,X^2-1\right)=\gcd\left(X^2-1,0\right)=X^2-1 \tag{3.2}$$

As for the integers, there is an extended version of the Euclidean algorithm, with which we can find polynomials $x$ and $y$ with $x\cdot a + y\cdot b = \gcd(a,b)$.

**Algorithm 3.2.18** (Extended Euclidean Algorithm for Polynomials)**.** • *Input: polynomials a and b over a field R, at least one of which is not zero.*

- *Output: list of polynomials* $\gcd(a,b)$*, x, y such that* $\gcd(a,b) = x\cdot a + y\cdot b$*.*

```
PolyExtendedGCD := procedure($a,b$)
local variables
    $a_1$, $b_1$
    $u := 0$ , $v := 1$
    $x := 1$ , $y := 0$
    $u_1$, $v_1$, $x_1$, $y_1$
while degree$(b) > -1$ do
    $a_1 := a$ , $b_1 := b$ , $u_1 := u$ , $v_1 := v$ , $x_1 := x$ , $y_1 := y$
    $a := b_1$ , $b := \operatorname{rem}(a_1,b_1)$ , $x := u_1$ , $y := v_1$
    $u := x_1 - \operatorname{quot}(a_1,b_1)\cdot u_1$ , $y := y_1 - \operatorname{quot}(a_1,b_1)\cdot v_1$
return
    $\left[\frac{a}{\operatorname{lc}(a)}, \frac{x}{\operatorname{lc}(a)}, \frac{y}{\operatorname{lc}(a)}\right]$
```

*Proof.*

**Termination.**

As $\operatorname{degree}(b)$ goes strictly down at each step, termination is guaranteed.

**Correctness.**

Let $a_0$ and $b_0$ denote the input values of $a$ and $b$, respectively. Then the values of $a$ and $b$ at the end of each loop satisfy $a = x\cdot a_0 + y\cdot b_0$ and $b = u\cdot a_0 + v\cdot b_0$. In computer science terms, these equations are invariants of the algorithm. Since the assignments involving $a$ and $b$ are as in Euclid's Algorithm for Polynomials, at the end we have $b = 0$ and $a = \gcd(a_0,b_0)$. The above equality for $a$ then gives the required expression of a gcd as a linear combination of $a_0$ and $b_0$. In order to obtain the corresponding expression for *the* gcd, the three output polynomials are divided by $\operatorname{lc}(a)$.

Although we do not use the equality involving $u$ and $v$, it is worth noting that, at the end of the algorithm, it gives a linear combination of $a_0$ and $b_0$ that is equal to 0.

□

**Example 3.2.19.** A convenient way to interpret the assignments in the algorithm is by means of matrix multiplication. To this end we put the key variables into a matrix as follows. $\begin{pmatrix} a & x & y \\ b & u & v \end{pmatrix}$. In terms of this matrix, the loop of the algorithm sees to it that it is multiplied from the left by the matrix $\begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix}$, where $q = \text{quot}(a,b)$.

For instance, for the extended gcd of the polynomials $X^4 - 1$ and $X^6 - 1$ the computations would consist of multiplying the $2 \times 3$ matrix from the left by the matrix with the $q$ entry for $q$ equal to, respectively,

- 0, the quotient of $X^4 - 1$ after division by $X^6 - 1$,
- $X^2$, the quotient of $X^6 - 1$ after division by $X^4 - 1$,
- $X^2$, the quotient of $X^4 - 1$ upon division by $X^2 - 1$.

Now the product of these three matrices is

$$\begin{pmatrix} 0 & 1 \\ 1 & -(X^2) \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & -(X^2) \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -(X^2) & 1 \\ X^4 + 1 & -(X^2) \end{pmatrix}.$$

Since at the outset $x$, $y$, $u$, $v$ build up the identity matrix, the resulting matrix contains the final values of $x$ and $y$ in the top row. Thus the gcd can be expressed as $X^2 - 1 = \left(-(X^2)\right) \cdot \left(X^4 - 1\right) + 1 \cdot \left(X^6 - 1\right)$.

The greatest common divisor (gcd) of two positive integers is the greatest among all divisors, both in the absolute sense and with respect to the (partial) ordering given by division. Here follows a similar characterization for polynomials, where the degree measures the size.

**Theorem 3.2.20** (Degree Maximality of the gcd)**.** *Suppose that $R$ is a field. Let $a$, $b$, and $c$ be polynomials in $R[X]$. If $a$ and $b$ are not both zero and $c$ is a common divisor of $a$ and $b$ of maximal degree, then $c$ is a greatest common divisor of $a$ and $b$.*

*Proof.* If $d$ is the gcd of $a$ and $b$, then by the Extended Euclidean Algorithm for Polynomials there are polynomials $p$ and $q$ with $d = p \cdot a + q \cdot b$. Thus the common divisor $c$ of $a$ and $b$ is also a divisor of $d$. As the degree of $d$ is less than or equal to the degree of $c$, this implies that $c$ is a scalar multiple of $d$ and hence also a greatest common divisor of $a$ and $b$.

□

**Example 3.2.21.** In $\mathbb{R}[X]$, the polynomial $X-1$ divides both $X^8-1$ and $X^{12}-1$, but so does $X^2+1$, so, by the Degree Maximality of the gcd, it is not a gcd of the two polynomials.

**Remark 3.2.22.** For polynomials and integers, the notions degree and absolute value play comparable roles. These rings are both instances of *Euclidean rings*, algebraic structures for which there exists a measure with comparable properties.

The Extended Euclidean Algorithm for Polynomials provides us with the following characterization of the gcd.

**Theorem 3.2.23** (Characterization of the gcd of Polynomials)**.** *Let $a$ and $b$ be two nonzero polynomials in $R[X]$, where $R$ is a field. Then the following three statements are equivalent.*

1. $\mathrm{gcd}(a,b)=d$.
2. *The polynomial $d$ is a monic common divisor of $a$ and $b$ of maximal degree.*
3. *$d$ is a monic polynomial of least nonnegative degree that can be expressed as $x\cdot a+y\cdot b$ with $x$ and $y$ polynomials in $R[X]$.*

*Proof.* The proof is divided into two steps.

**The second statement is equivalent to the first.**

This follows immediately from Degree Maximality of the gcd.

**The third statement is equivalent to the first.**

Let $d=\mathrm{gcd}(a,b)$ and let $e$ be a polynomial of least nonnegative degree that can be expressed as $x\cdot a+y\cdot b$ with $x$ and $y$ in $R[X]$. We show that $d=e$. Since $d$ is a common divisor of $a$ and $b$, the equality $e=x\cdot a+y\cdot b$ implies that $d$ divides $e$. So $\mathrm{degree}(d)\le\mathrm{degree}(e)$. Moreover, as a result of the Extended Euclidean Algorithm for Polynomials, $d$ itself can also be written as a combination of $a$ and $b$. So $\mathrm{degree}(e)\le\mathrm{degree}(d)$ by the defining property of $e$. Hence $e$ must be a scalar multiple of $d$. As both polynomials have leading coefficient 1, they are equal. This proves the equivalence.

Since both the second as well as the third statement of the theorem are equivalent to the first, all three statements are equivalent. This finishes the proof of the theorem.

□

**Example 3.2.24.** To see that the polynomials $X^5+1$ and $X^3-1$ have gcd equal to 1, it suffices to verify the following equality and apply the Characterization of the gcd of Polynomials:

$$\left(1+X+\left(-(X^2)\right)\right)\cdot\left(X^5+1\right)+\left(-1+X+\left(-(X^2)\right)+\left(-(X^3)\right)+X^4\right)\cdot\left(X^3-1\right)=2 \tag{3.3}$$

These different characterizations of the gcd, in particular the possibility of expressing the gcd of two polynomials $a$ and $b$ as a combination of $a$ and $b$, will turn out to be very useful in all kinds of applications.

## 3.3 Polynomial functions

We connect our formal definition of a polynomial with the more common notion of a polynomial function. Let $R$ be one of the rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/n\mathbb{Z}$. When we refer to $R$ as a field, we mean to restrict the choice to $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}/n\mathbb{Z}$ with $n$ prime. In these cases (and only these) each nonzero element has an inverse.

**Definition 3.3.1.** Let $a(X)=a_0+\ldots+a_m\cdot X^m$ be a polynomial in $R[X]$. By replacing the variable $X$ in the polynomial $a(X)$ by an element $r$ of $R$, we find the element $a(r)=a_0+a_1\cdot r+\ldots+a_m\cdot r^m$. In this way we obtain a function $a\colon R\to R, r\longmapsto a(r)$ called the *polynomial function* of $a$. An element $r$ of the ring $R$ is called a *zero* of $a(X)$ if $a(r)=0$.

**Example 3.3.2.** Consider the polynomials $X^3$ and $X$ in $\mathbb{Z}/2\mathbb{Z}[X]$. The polynomial function of each of these polynomials is the identity map on $\mathbb{Z}/2\mathbb{Z}$.

**Remark 3.3.3.** The set of polynomial functions is useful for many applications, especially because they are functions which are easy to represent, to manipulate and to use for approximations of other, more complicated, functions.

By way of example, on the next page, we construct polynomial functions with prescribed behaviour.

**Remark 3.3.4.** It is also customary to speak of *root* of a polynomial, instead of zero of a polynomial. The notion is in accordance with expressions like cube root of 2, which refers to the positive real number that is a zero of the real polynomial $X^3-2$ in $\mathbb{R}[X]$.

Zeros of a polynomial are related to linear factors (that is, factors of degree 1).

**Theorem 3.3.5** (Characterization of the Zeros of a Polynomial)**.** *Let $R$ be a field and $f\in R[X]$.*

1. *An element $x\in R$ is a zero of $f$ if and only if $X-x$ divides $f$.*
2. *If $f$ is a polynomial of degree $n$, then $f$ has at most $n$ distinct zeros.*

*Proof.* Let $x \in R$. Dividing $f$ by $X - x$ yields $f = (X - x) \cdot q + r$ with $r$ of degree at most zero and hence in $R$.

Evaluating both sides at $x$ gives $f(x) = r$. Consequently, $f(x) = 0$ if and only if $X - x$ divides $f$.

Suppose that $f$ is a polynomial with distinct zeros $x_1, x_2, \ldots, x_t$. We claim that the product $\prod_{i=1}^{n} (X - x_i)$ is a divisor of $f$. For, $f(x_1) = 0$ implies that there is a polynomial $g_1$ such that $f = (X - x_1) \cdot g_1$. Now $f(x_2) = 0$ is equivalent to $(x_2 - x_1) \cdot g_1(x_2) = 0$. But $x_2 - x_1 \neq 0$ and so $g_1(x_2) = 0$, and hence $X - x_2$ divides $g_1$. This implies that $(X - x_1) \cdot (X - x_2)$ divides $f$. Continuing this way, we obtain a proof of the claim.

If $f$ has degree $n$, then, by the Degree Formulas, every divisor of it has degree at most $n$, so the claim implies that $f$ has at most $n$ different zeros.

□

**Remark 3.3.6.** Another proof of the second statement of the theorem (and the claim used in the proof) will follow from Characterization of Relative Prime Polynomials.

**Example 3.3.7.** Suppose that $m$ and $n$ are positive integers with $m$ dividing $n$. We consider polynomials over $\mathbb{C}$. Now $X^m - 1$ divides $X^n - 1$. This means that any $m$-th root of unity (i.e., a complex number whose $m$-th power is equal to one) is a zero of $X^n - 1$. By dividing $X^n - 1$ by the gcd of all $X^m - 1$, for $m$ a proper divisor of $n$, we find the monic polynomial all of whose zeros are primitive $n$-th roots of unity; here, *primitive* means that these roots are no $m$-th roots of unity for any proper divisor of $n$. For example, $X^6 = \left(X^2 - X + 1\right) \cdot \left(X^2 + X + 1\right) \cdot (X + 1) \cdot (X - 1)$ where $X^2 - X + 1$ is the product to the two linear factors corresponding to the primitive 6-th roots of unity, $X^2 + X + 1$ is the product to the two linear factors corresponding to the primitive third roots of unity, $X + 1$ the linear factor corresponding to $-1$, the primitive second root of 1, and $X - 1$ the linear factor corresponding to 1, the primitive first root of 1.

*Interpolation* concerns the question of finding a function that has prescribed values at a given number of points. In the polynomial context we are of course looking for polynomial functions. Given $n$ points $x_1, \ldots, x_n \in R$, and $n$ prescribed values $a_1, \ldots, a_s \in R$, does a polynomial function $f\colon R \to R$ exist that interpolates the values $a_i$ on $x_i$?

**Theorem 3.3.8** (Lagrange Interpolation)**.** *Let $n$ be a positive integer and $R$ a field. Suppose that $n$ distinct elements $x_1, \ldots, x_n \in R$ and $n$ required values $a_1, \ldots, a_n \in \mathbb{R}$ are given. Then there is a unique polynomial function $f\colon R \to R$ of degree at most $n - 1$ with $f(x_i) = a_i$ for all $i$.*

*Proof.* Let $f_i$ be the polynomial

$$\frac{1}{X - x_i} \cdot \prod_{j=1}^{n} (X - x_j) \tag{3.4}$$

Then $f_i(x_i)$ is nonzero and $f_i(x_j) = 0$ for $i \neq j$.

But then

$$\sum_{j=1}^{n} \frac{a_j}{f_j(a_j)} \cdot f_j \tag{3.5}$$

is a polynomial that satisfies the conditions of the Theorem.

Now assume that both $f$ and $g$ do satisfy the condition of the Theorem. Then $f - g$ is polynomial of degree at most $n-1$ with at least $n$ distinct zeros. But then the Characterization of the Zeros of a Polynomial implies that $f - g$ is the zero polynomial. So $f = g$.

□

**Example 3.3.9.** An example of a polynomial $f \in \mathbb{R}[X]$ such that the corresponding function $f\colon \mathbb{R} \to \mathbb{R}$ satisfies $f(1) = 2$ and $f(2) = 5$, is $f(X) = X^2 + 1$ but also $3 \cdot X - 1$ One can look for such a polynomial as follows. Choose a degree, preferably equal to the number of interpolation points minus 1; but let us now take 2. Then write $f(X) = f_0 + f_1 \cdot X + f_2 \cdot X^2$ and substitute the given values. This leads to the following system of linear equations: $f_0 + f_1 \cdot 1 + f_2 \cdot 1^2 = 2$ $f_0 + f_1 \cdot 2 + f_2 \cdot 2^2 = 5$ Solving these equations gives $f_0 = 2 \cdot r - 1$ $f_1 = (-3) \cdot r + 3$ and $f_2 = r$ with $r \in \mathbb{R}$. This shows that there are many polynomials with the required properties. No polynomials of degree $d$ with $d \leq 0$ will do the job, exactly one polynomial of degree $d \leq 1$ works (with $r = 0$), and there is an infinite number of solutions of degree $d \geq 2$. This is in accordance with the Lagrange Interpolation, applied for $n = 2$.

The so-called Fundamental Theorem of Algebra says that every polynomial over $\mathbb{C}$ has a zero. Equivalently: every polynomial in $\mathbb{C}[X]$ is a product of linear factors. We shall not prove this fact. Giving a proof is hard and requires a rigorous treatment of $\mathbb{C}$.

**Theorem 3.3.10** (Fundamental Theorem of Algebra)**.** *Every polynomial over $\mathbb{C}$ has a zero.*

**Remark 3.3.11.** Equivalent to the Fundamental Theorem of Algebra is the following statement: every polynomial in $\mathbb{C}[X]$ is a product of linear factors. This is immediate by the Characterization of the Zeros of a Polynomial.

We can use this fact to find factors of polynomials over $\mathbb{R}$. Let $f$ be a polynomial over $\mathbb{R}$. Then we can consider $f$ as a polynomial over $\mathbb{C}$. In particular, $f$ will have a (complex) zero, $x$ say. If $x$ is real, then $f$ is divisible by $X - x$. If $x$ is not real, then its complex conjugate $\overline{x}$ is also a zero of $f$. Indeed, as all coefficients of $f$ are real we have $f(\overline{x}) = \overline{f(x)} = \overline{0} = 0$

So, if $x$ is not real, then $f$ is divisible by the linear complex polynomials $X - x$ and $X - \overline{x}$ and therefore also by the real polynomial $(X - x) \cdot (X - \overline{x}) = X^2 - 2 \cdot \mathrm{Re}(x) + x \cdot \overline{x}$

We conclude that a real polynomial always has a factor of degree one or two.

## 3.4 Factorization

In the following $R$ is, without explicit mention of the contrary, always a field, like $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Z}/p\mathbb{Z}$ with $p$ prime. These arithmetic systems have in common that every nonzero element has a multiplicative inverse.

Here is the counterpart in the setting of polynomial rings of primality.

**Definition 3.4.1** (Irreducibility)**.** A polynomial $f \in R[X]$ is called *irreducible* if degree $(f) > 0$ and if the only nonconstant polynomials $g$ with $g|f$ have the same degree as $f$; in other words, if $f$ is not a constant and if its only divisors are the constants and the constant multiples of $f$. If $f$ is not irreducible, then $f$ is called *reducible*.

**Example 3.4.2.** By definition, all polynomials of degree 1 are irreducible. Clearly, such a statement is no longer true for polynomials of higher degree.

For instance, the only irreducible polynomials of $\mathbb{Z}/2\mathbb{Z}[X]$ of degrees 2 and 3 are $X^2+X+1$, $X^3+X+1$, and $X^3+X^2+1$.

We shall study factorizations of a polynomial, that is, ways to write the polynomial as a product of polynomials of smaller degree.

With the help of the Fundamental Theorem of Algebra, we can determinte which polynomials over $\mathbb{R}$ and $\mathbb{C}$ are irreducible.

**Theorem 3.4.3** (Classification of Real and Complex Irreducible Polynomials)**.** *A complex polynomial $f \in \mathbb{C}[X]$ is irreducible if and only if its degree is 1.*
*If a real polynomial $f \in \mathbb{R}[X]$ is irreducible, then its degree is 1 or 2.*
*The real polynomial $a \cdot X^2 + b \cdot X + c \in \mathbb{R}[X]$ of degree 2 is irreducible if and only if $b^2 - 4 \cdot a \cdot c < 0$.*

*Proof.* As we have seen in Example 3.4.2, a complex polynomial is always divisible by a linear polynomial. So indeed, a complex polynomial is irreducible if and only if it is linear.

As we have seen in Example 3.4.2, a real polynomial of positive degree is always divisible by a linear or a degree 2 polynomial. So, if it is irreducible, then it has degree at most 2. Moreover, if its degree is 2, then it is irreducible if and only if it has no real zeros. The latter is equivalent to the discriminant being negative.

□

**Example 3.4.4.** The polynomial $a \cdot X^2 + b \cdot X + c \in \mathbb{R}[X]$ with $a \neq 0$ and $b^2 - 4 \cdot a \cdot c \geq 0$ is reducible. It equals the product $a \cdot \left(X - \frac{-b+\sqrt{b^2-4 \cdot a \cdot c}}{2 \cdot a}\right) \cdot \left(X - \frac{-b-\sqrt{b^2-4 \cdot a \cdot c}}{2 \cdot a}\right)$

**Example 3.4.5.** The theorem states that the polynomial $a\cdot X^2+b\cdot X+c\in\mathbb{R}[X]$ with $a\neq 0$ and $b^2-4\cdot a\cdot c<0$ is irreducible. But when viewed as a complex polynomial it is reducible and equals the product $a\cdot\left(X-\frac{-b+i\cdot\sqrt{4\cdot a\cdot c-b^2}}{2\cdot a}\right)\cdot\left(X-\frac{-b-i\cdot\sqrt{4\cdot a\cdot c-b^2}}{2\cdot a}\right)$

Let $R$ be a field. The following result for polynomials parallels the characterization of relative prime integers.

**Lemma 3.4.6** (Characterization of Relative Prime Polynomials)**.** *Two polynomials $f$ and $g$ (not both zero) in $R[X]$ are relatively prime if and only if there exist polynomials $a$ and $b$ such that $a\cdot f+b\cdot g=1$.*

*Proof.*

**If.**

From a relation $a\cdot f+b\cdot g=1$ we infer that a common divisor of $f$ and $g$ must be a divisor of the left-hand side $a\cdot f+b\cdot g$ and therefore of $1$. So the gcd of $f$ and $g$ is $1$. This proves the 'if' part.

**Only if.**

The 'only if' implication is an immediate consequence of the Extended Euclidean Algorithm for Polynomials.

□

Compare the next theorem with the similar Result on the divisor of a product.

**Proposition 3.4.7.** *If $f$ and $g$ are relatively prime, then $f|g\cdot h$ implies $f|h$.*
*If $p$ is an irreducible polynomial and $b_1,\ldots,b_s$ are polynomials such that $p|b_1\cdot\ldots\cdot b_s$, then there is an index $i\in\{1,\ldots,s\}$ with $p|b_i$.*

*Proof.* By the Extended Euclidean Algorithm for Polynomials, there exist polynomials $a$ and $b$ with $a\cdot f+b\cdot g=1$. Multiplying this relation by $h$ yields $a\cdot f\cdot h+b\cdot g\cdot h=h$

Since $f|a\cdot f\cdot h$ and $f|b\cdot g\cdot h$, it follows that $f|h$.

This proves the first part of the theorem. The second follows immediately.

□

The Result on divisors of a product leads to unique factorization of polynomials.

**Theorem 3.4.8** (Unique Factorization)**.** *Let $R$ be a field. Every nonconstant polynomial $f \in R[X]$ can be written as the product of a finite number of irreducible polynomials: $f = p_1 \cdot \ldots \cdot p_s$ for some positive integer $s$, and irreducible polynomials $p_i$ where $i \in \{1,\ldots,s\}$.*
*This way of writing is unique up to the order of the irreducible factors and up to multiplication by constants.*

*Proof.* The proof is divided into two parts: existence and uniqueness.

**The polynomial $f$ can be written as a product of irreducible factors.**

We show by induction on the degree of $f$ that $f$ can be written as a product of irreducible factors.

If the degree of $f$ equals 1, then $f$ itself is obviously irreducible and we are done.

Now suppose that the degree of $f$ is greater than 1. The induction hypothesis says that every polynomial of degree less than degree$(f)$ can be written as a product of irreducible factors. If $f$ is irreducible, we are done. If not, then $f$ has a divisor $g$ such that both $g$ and $f/g$ have degree less than the degree of $f$. The induction hypothesis implies that both $g$ and $f/g$ can be written as a product of irreducible factors. But then, as $f = (f/g)\cdot g$, we find that $f$ itself is also a product of irreducible polynomials.

**The factorization of $f$ into irreducible factors is unique up to order and multiplication by constants.**

Again we use induction on the degree $n$ of $f$.

The case $n = 1$ is easy and left to the reader.

Now suppose that $n > 1$, and suppose that uniqueness has been shown for polynomials of degree less than $n$. Suppose $f = p_1 \cdot \ldots \cdot p_s$ and $f = q_1 \cdot \ldots \cdot q_t$ are two possible ways of writing $f$ as a product of irreducible factors. From Result on divisors of a product we conclude that there exists an index $k \in \{1,\ldots,t\}$ such that $p_s$ divides $q_k$. Without loss of generality we can assume $k$ to be equal to $t$ and, as we may multiply by constants, that $p_s = q_t$. Applying the induction hypothesis to the polynomial $\frac{f}{p_s}$ with the two ways of writing it as a product of irreducible factors:$\frac{f}{p_s} = p_1 \cdot \ldots \cdot p_{s-1}$ and $\frac{f}{p_s} = q_1 \cdot \ldots \cdot q_{t-1}$ yields that these factorizations are equal (up to the order of the factors and multiplications by constants). Clearly this implies that the two factorizations of $f$ are also equal (up to the order of the factors and multiplications by constants).

□

**Example 3.4.9.** The factorization in irreducibles of $X^4 - 1$ in $\mathbb{Q}[X]$ is $\left(X^2+1\right)\cdot(X+1)\cdot(X-1)$.

The first factor is irreducible since it has degree at most two and no rational zeros. Considered as a polynomial over $\mathbb{C}$, the factorization of $X^4-1$ is $(X+i)\cdot(X-i)\cdot(X+1)\cdot(X-1)$.

Considered as a polynomial over $\mathbb{Z}/2\mathbb{Z}$, the factorization is $(X+1)^4$.

**Example 3.4.10.** As for integers (compare with the example on the factorization record), it is not difficult to verify a factorization. However, it is not always as easy to check whether the found factors are irreducible. A proof that a polynomial $f \in \mathbb{Q}[X]$ with integer coefficients is irreducible, can often be given by computing modulo $p$ for a prime number $p$. If the polynomial is irreducible modulo $p$, then it is also irreducible over $\mathbb{Q}$. However, the converse does not hold. There are polynomials $f \in \mathbb{Z}[X]$ which are irreducible over $\mathbb{Q}$ but reducible modulo each prime $p$. An example is $f(X) = X^4+1$. Modulo 2 it factors as $(X+1)^4$ and modulo 3 as $\left(X^2-X-1\right)\cdot\left(X^2+X-1\right)$. It carries too far to show that $X^4+1$ factors modulo every prime.

# 3.5 Exercises

## 3.5.1 The notion of a polynomial

**Exercise 3.5.1.** Find the sum and product of the following polynomials.

- $X^3+2{\cdot}X^2-X+1$ and $X^2+2{\cdot}X-1$ over $\mathbb{Q}$;
- $X^3+2{\cdot}X^2-X+1$ and $X^2+2{\cdot}X-1$ over $\mathbb{Z}/3\mathbb{Z}$;
- $X^3+X-1$ and $X^2-X-2$ over $\mathbb{Q}$;
- $X^3+X-1$ and $X^2-X-2$ over $\mathbb{Z}/3\mathbb{Z}$.

**Exercise 3.5.2.** Show that for any prime $p$ and any polynomial $a_0+a_1{\cdot}X+\ldots+a_{n-1}{\cdot}X^{n-1}+a_n{\cdot}X^n$ in $\mathbb{Z}/p\mathbb{Z}[X]$, we have $\left(a_0+a_1{\cdot}X+\ldots+a_{n-1}{\cdot}X^{n-1}+a_n{\cdot}X^n\right)^p = a_0+a_1{\cdot}X^p+\ldots+a_{n-1}{\cdot}X^{p\cdot(n-1)}+a_n{\cdot}X^{p\cdot n}$

## 3.5.2 Division of polynomials

**Exercise 3.5.3.** Determine the gcd of each of the following pairs of polynomials and write each gcd as a combination of the given polynomials.

- $X^2+1$ and $X^3+1$ as polynomials over $\mathbb{Q}$;
- $X^2+1$ and $X^3+1$ as polynomials over $\mathbb{Z}/2\mathbb{Z}$;
- $X^2-X+1$ and $X^3+X+2$ as polynomials over $\mathbb{Z}/3\mathbb{Z}$.

**Exercise 3.5.4.** Suppose that the polynomials $a$ and $b$ have integer coefficients and that $b$ is monic, i.e., has leading coefficient 1. Prove that the quotient $q$ and remainder $r$ of division of $a$ by $b$ in $\mathbb{Q}[X]$ also belong to $\mathbb{Z}[X]$.

**Exercise 3.5.5.** Analogously to the definition of the gcd of two polynomials one can define the gcd of more than two (nonzero) polynomials.

Indeed, the gcd of a set of polynomials is a polynomial with leading coefficient 1 and the property that it is divisible by every common divisor of the polynomials in the set.

Let $a$, $b$, and $c$ be three nonzero polynomials with coefficients in $\mathbb{Q}$.

- Show that $\gcd(a,b,c) = \gcd(a,\gcd(b,c))$
- Show that $a,b,c$ are relatively prime (have gcd 1 ) if and only if there exist polynomials $p,q,r$ such that $p\cdot a+q\cdot b+r\cdot c=1$.

**Exercise 3.5.6.** Let $a$, $b$, and $c$ be polynomials in $X$. Prove the following:

If $a$ divides $b$ and $c$, then $a$ divides $b+d\cdot c$ for every polynomial $d$.

**Exercise 3.5.7.** Let $a$, $b$, and $c$ be polynomials in $X$. Prove the following:

If $a$ divides $b$ and $b$ divides $c$, then $a$ divides $c$.

**Exercise 3.5.8.** Determine the quotient and remainder of $a$ upon division by $b$, where $a$ and $b$ are as below.

1. $a=X^4+3\cdot X^2+X+1$ and $b=X^2+X+1$ in $\mathbb{Q}[X]$;
2. $a=X^4+3\cdot X^2+X+1$ and $b=X^2+X+1$ in $\mathbb{Z}/2\mathbb{Z}[X]$;
3. $a=X^4+3\cdot X^2+X+1$ and $b=X^2+X+1$ in $\mathbb{Z}/3\mathbb{Z}[X]$.

**Exercise 3.5.9.** Let $a$ and $b$ be polynomials in $X$ over the field $R$. The gcd of $a$ and $b$ can be written as $p\cdot a+q\cdot b$ for some polynomials $p$ and $q$. Show that every polynomial that can be written as $p\cdot a+q\cdot b$ with $p$ and $q$ polynomials over $R$, and divides $a$ and $b$, is a gcd of $a$ and $b$.

**Exercise 3.5.10.** Determine polynomials $a$ and $b$ in $\mathbb{Q}[X]$ such that $a\cdot\left(X^2+1\right)+b\cdot\left(X^3-X+1\right)=X-1$

**Exercise 3.5.11.** Determine polynomials $a$ and $b$ in $\mathbb{Z}/2\mathbb{Z}[X]$ such that $a\cdot\left(X^2+1\right)+b\cdot\left(X^3-X+1\right)=X-1$

### 3.5.3 Polynomial functions

**Exercise 3.5.12.** Find all zeros of each of the following polynomials

1. $X^2+2\cdot X+2$ in $\mathbb{Z}/5\mathbb{Z}[X]$;

2. $X^2+X+1$ in $\mathbb{Z}/24\mathbb{Z}[X]$;
3. $X\cdot(X+1)\cdot(X+2)$ in $\mathbb{Z}/12\mathbb{Z}[X]$;
4. $2\cdot X^2+13\cdot X+9$ in $\mathbb{Z}/33\mathbb{Z}[X]$.

**Exercise 3.5.13.** Let $f$ be a polynomial in $\mathbb{Z}[X]$ of degree at least 1.

Prove that $f(n)$ cannot be a prime for each $n\in\mathbb{Z}$.

**Exercise 3.5.14.** Find all polynomials $p\in\mathbb{Q}[X]$ that satisfy $p(x)=p(-x)$ for any $x$ in $\mathbb{Q}$.

**Exercise 3.5.15.** Find all polynomials $p\in\mathbb{Z}/2\mathbb{Z}[X]$ that satisfy $p(x)=p(-x)$ for any $x$ in $\mathbb{Z}/2\mathbb{Z}$.

What happens if we replace $\mathbb{Z}/2\mathbb{Z}$ by $\mathbb{Z}/6\mathbb{Z}$?

### 3.5.4 Factorization

**Exercise 3.5.16.** Consider the polynomial $a=a_0+a_1\cdot X+...+a_{n-1}\cdot X^{n-1}+a_n\cdot X^n$ in $\mathbb{Z}[X]$, with $a_n\neq 0$.

1. Prove: If $r\in\mathbb{Z}$ is a zero of $a$, then $r$ is a divisor of $a_0$.
2. Suppose that $r,s\in\mathbb{Z}$ are relatively prime and that $r/s$ is a root in $\mathbb{Q}$ of $a$. Prove that $s$ divides $a_n$ and that $r$ divides $a_0$.
3. Find all rational roots of the polynomial $15-32\cdot X+3\cdot X^2+2\cdot X^3$.

**Exercise 3.5.17.** Consider the ring $\mathbb{Z}/3\mathbb{Z}[X]$ of polynomials in $X$ with integer coefficients modulo 3.

1. How many polynomials of degree $n$ are there in $\mathbb{Z}/3\mathbb{Z}[X]$?
2. Determine all irreducible polynomials in $\mathbb{Z}/3\mathbb{Z}[X]$ of degrees 2 and 3.

**Exercise 3.5.18.** Verify the identity of polynomials $\left(X^2-1\right)^2+(2\cdot X)^2=\left(X^2+1\right)^2$

A Pythagorean triple is a triple of positive integers $r,s$ and $t$ such that $r^2+s^2=t^2$ According to the Pythagorean theorem, these triples occur as sides of right triangles.

By substituting rational numbers $p/q$ for $X$ show how to produce Pythagorean triples from the identity $\left(X^2-1\right)^2+(2\cdot X)^2=\left(X^2+1\right)^2$.

**Exercise 3.5.19.** Suppose the polynomials $f(X)$ and $g(X)$ over $\mathbb{Q}$ have greatest common divisor $d(X)$. Fix $a$ in $\mathbb{Q}$ and replace every occurrence of $X$ in $f$ and $g$ by $X+a$. For instance, if $a=2$ then $X^2+X-1$ changes into $(X+2)^2+(X+2)-1$.

Prove that the gcd of the new polynomials $f(X+a)$ and $g(X+a)$ is $d(X+a)$.

**Exercise 3.5.20.** Show that the polynomials $X-1$ and $X^2+X+1$ over $\mathbb{Q}$ are relatively prime.

Use the Extended Euclidean Algorithm for Polynomials to find constants $a, b, c$ such that $\frac{3}{X^3-1} = \frac{a}{X-1} + \frac{b \cdot X + c}{X^2+X+1}$.

**Exercise 3.5.21.** Let $R$ be one of the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, $\mathbb{Z}/p\mathbb{Z}$ with $p$ prime. Prove that there are infinitely many irreducible polynomials in $R[X]$.

**Exercise 3.5.22.** Determine all irreducible polynomials $p$ and $q$ in $\mathbb{Z}[X]$ that satisfy the equation $\left(X^2+1\right) \cdot p + (X+2) \cdot q = p \cdot q$.

# Chapter 4

# Arithmetic modulo polynomials

One step beyond arithmetic modulo an integer, is arithmetic 'modulo a polynomial' (or several polynomials). Here polynomials that differ by multiples of a fixed polynomial are considered equivalent. This construction gives us arithmetical systems that are important in, for example, coding theory and cryptology. In this chapter, $R$ is always one of the sets $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C},\mathbb{Z}/n\mathbb{Z}$ where $n>1$, with the usual addition and multiplication, unless explicitly stated otherwise.

## 4.1 Congruence modulo a polynomial

We know computation modulo a fixed integer $n$. Here we will do something similar, but with polynomials instead of integers. Thus we work with elements of polynomial rings $R[X]$, with $R$ a ring like one of $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C},\mathbb{Z}/n\mathbb{Z}$ with $n>1$.

Often, but not always, we will require that $R$ be a field, that is, a ring in which every nonzero element is a divisor of 1. Of the above rings, $\mathbb{Q},\mathbb{R},\mathbb{C},\mathbb{Z}/n\mathbb{Z}$, with $n$ a prime, are fields.

**Definition 4.1.1.** Let $d$ be a polynomial in $R[X]$. We define the relation *congruence modulo d* on $R[X]$ as follows. The polynomials $a,b\in R[X]$ are congruent modulo $d$ (notation: $a\equiv b \pmod{d}$) if there exists a polynomial $q\in R[X]$ such that $a-b=q\cdot d$; in other words if $a$ and $b$ differ by a multiple of $d$.

**Example 4.1.2.** Consider the constant 2. In $\mathbb{Q}[X]$ every polynomial is congruent to 0 modulo 2. However, in $\mathbb{Z}[X]$ a polynomial is congruent to 0 modulo 2 if and only if each of its coefficients is even.

Consider the polynomial $d=3\cdot X-1$ in $\mathbb{R}[X]$. By the Characterization of the Zeros of a Polynomial a polynomial in $\mathbb{R}[X]$ is congruent to 0 modulo $d$ if and only if its value at $1/3$ (as a polynomial function) is 0.

Our goal will be to port as many results as possible from the arithmetic modulo an integer to the arithmetic modulo a polynomial. The following theorem tells us that, to begin with, the most important property (the division into residue classes) is preserved.

**Theorem 4.1.3.** *Congruence modulo d is an equivalence relation on $R[X]$.*

*Proof.* To show that congruence modulo $d$ is an equivalence relation, we have to verify that this relation is reflexive, symmetric, and transitive.

**Congruence modulo $d$ is reflexive.**

This follows from the fact that for every polynomial $a$ we have: $a-a=0\cdot d$.

**Congruence modulo $d$ is symmetric.**

If $a$ and $b$ are congruent modulo $d$, i.e., if $a-b=q\cdot d$ for some polynomial $q$, then rewriting this equality as $b-a=(-q)\cdot d$ shows that $b$ and $a$ are also congruent modulo $d$.

**Congruence modulo $d$ is transitive.**

If $a$ is congruent to $b$ modulo $d$ and $b$ is congruent to $c$ modulo $d$, then there exist polynomials $q$ and $p$ with $a-b=q\cdot d$ and $b-c=p\cdot d$. Adding these equalities yields $a-c=(q+p)\cdot d$. This shows that $a$ and $c$ are congruent modulo $d$.

□

**Example 4.1.4.** Consider the polynomial $d=3\cdot X-1$ in $\mathbb{Q}[X]$. By Characterization of the Zeros of a Polynomial two polynomials in $\mathbb{R}[X]$ are congruent 0 modulo $d$ if and only if their values at $1/3$ (as a polynomial function) are equal. So the equivalence classes are in bijective correspondence with $\mathbb{Q}$, the set of possible values of the polynomial function of $d$.

We introduce some notation for the equivalence classes of congruence modulo $d$.

**Definition 4.1.5.** By $(d)R[X]$ we denote the set $\{f\in R[X]\,|\,\exists g.f=g\cdot d\}$

The equivalence class $\{f\in R[X]\,|\,\exists g.f=a+g\cdot d\}$, containing the polynomial $a$, is called the *residue class* modulo $d$ of $a$ and is denoted by $a+(d)R[X]$. The set of residue classes modulo $d$ is denoted by $R[X]/(d)R[X]$. This set is called the *residue class ring* or *quotient ring* modulo $d$.

**Example 4.1.6.** In $\mathbb{Q}[X]$, the polynomials $X^6$ and 1 represent the same residue class modulo $X^2-X+1$. Indeed, $X^6-1=(X-1)\cdot\left(X^2+X+1\right)\cdot\left(X^2-X+1\right)\cdot(X+1)$ from which we deduce that $X^6-1$ is divisible by $X^2-X+1$.

Other notations for the residue class modulo $d$ containing the polynomial $a$ are:

- $a$, when it is clear we mean the residue class,

- or $a+(d)R[X]$.

In these notations, naturally, $a$ is the most obvious representative from the residue class $a+(d)R[X]$, but not necessarily the only one. For any $g \in R[X]$ the polynomial $a+g\cdot d$ is also a representative of this class.

The notation $R[X]/(d)R[X]$ is similar to the notation $\mathbb{Z}/n\mathbb{Z}$ introduced in Congruence is an Equivalence Relation.

Suppose that $R$ is a field and $d \in R[X]$. Then every residue class modulo $d$ contains a canonical representative:

**Theorem 4.1.7.** *If $d \in R[X]$ is a polynomial of degree $n > 0$, then every residue class modulo $d$ has a unique representative of degree less than $n$. This unique representative is the remainder obtained when dividing an arbitrary representative of the class by $d$.*

*Proof.* Let $a+(d)R[X]$ be the class of $a$ modulo $d$. The proof is divided into two parts. Together they imply the theorem.

**There exists a representative of $a+(d)R[X]$ of degree smaller than $n$.**

Division with remainder leads to an equality $a = q\cdot d + r$ where $r$ is a polynomial of degree less than $n$. Rewriting the equality as $a - r = q\cdot d$ shows that $a$ and $r$ are congruent modulo $d$. Hence $r$ is a representative of degree less than $n$ of the residue class of $a$.

**The class of $a$ modulo $d$ contains at most one element of degree less than $n$.**

Suppose that both $a$ and $b$ are representatives of degree less than $n$ of the same residue class modulo $d$. Then $a-b = q\cdot d$ for some polynomial $q$. Since the degrees of both $a$ and $b$ are less than the degree of $d$, the degree of the left-hand side is less than $n$. But the degree of the right-hand side can only be less than $n$ if $q$ is the zero polynomial. In particular, $a = b$.

□

**Example 4.1.8.** Consider the residue classes modulo $X^2+1$ in $\mathbb{Z}/3\mathbb{Z}[X]$. According to the Theorem on the Representative of Congruence Classes, every residue class has its own unique representative of degree at most 1. Conversely, every polynomial of degree at most 1 represents a different class. Since there are precisely nine polynomials in $\mathbb{Z}/3\mathbb{Z}[X]$ of degree at most 1, we find exactly nine residue classes. Below we list their representatives of degree at most 1.

$$0, 1, 2, X, 1+X, 2+X, 2\cdot X, 1+2\cdot X, 2+2\cdot X \tag{4.1}$$

In practice we will often use the short notation, like $1+X$, not only for the representative, but also to denote the congruence class. Naturally, we prefer it to the long expression $1+X+(X^2+1)\mathbb{Z}/3\mathbb{Z}[X]$ whenever no confusion is imminent.

## 4.2 The residue class ring

Suppose that $R$ is a ring. Let $d$ be a polynomial in $R[X]$. In this section we describe how to add and multiply residue classes in $R[X]/(d)R[X]$.

We use addition and multiplication for the operations of taking sum and product, respectively.

**Definition 4.2.1.** The *sum* and *product* of the residue classes $a+(d)R[X]$ and $b+(d)R[X]$ in $R[X]/(d)R[X]$ are defined as follows.

- Sum: $(a+((d)R[X]))+((b+((d)R[X]))) = (a+b)+(d)R[X]$;
- Product: $(a+(d)R[X])\cdot(b+(d)R[X]) = (a\cdot b)+(d)R[X]$.

**Proposition 4.2.2.** *Sum and product on $R[X]/(d)R[X]$ are well-defined.*

*Proof.* We need to verify that a different choice of representatives leads to the same residue class for the sum (and the product).

**The sum is well defined.**

Suppose that $a$ and $a'$ are both representatives of the same residue class and also that $b$ and $b'$ represent a single class. Then there are polynomials $p$ and $q$ with $a-a' = p\cdot d$ and $b-b' = q\cdot d$. Addition leads to the equality $(a+b)-(a'+b') = (p+q)\cdot d$. This implies that $a+b$ and $a'+b'$ belong to the same residue class modulo $d$. Hence addition is well defined.

**The product is well defined.**

The check is similar to the one for addition.

□

**Example 4.2.3.** Consider the polynomials $a = X^3+3\cdot X^2+1$, $b = X^2+2\cdot X-1$, and $d = X^2+X+1$ in $\mathbb{Q}[X]$. Then inside $\mathbb{Q}[X]/(d)\mathbb{Q}[X]$ we find

$$\begin{aligned}(a+((d)\mathbb{Q}[X]))+((b+((d)\mathbb{Q}[X]))) &=\\ (a+b)+(d)\mathbb{Q}[X] &=\\ X^3+3\cdot X^2+1+X^2+2\cdot X-1+(d)\mathbb{Q}[X] &=\\ X^3+4\cdot X^2+2\cdot X+(d)\mathbb{Q}[X] &=\\ (-2)\cdot X-3+(d)\mathbb{Q}[X]&.\end{aligned}$$

The product modulo $d$ equals

$$\begin{aligned}(a+(d)\mathbb{Q}[X])\cdot(b+(d)\mathbb{Q}[X]) &=\\ (a\cdot b)+(d)\mathbb{Q}[X] &=\\ \left(X^3+3\cdot X^2+1\right)\cdot\left(X^2+2\cdot X-1\right)+(d)\mathbb{Q}[X] &=\\ -1+2\cdot X-2\cdot X^2+5\cdot X^3+5\cdot X^4+X^5+(d)\mathbb{Q}[X] &=\\ 5+8\cdot X+(d)\mathbb{Q}[X]&.\end{aligned}$$

Let $R$ be a ring and let $d \in R[X]$. The usual arithmetical rules imply the rules below for addition and multiplication modulo $d$. First we identify two special elements.

- The element $0+(d)R[X]$ is called the *zero element* of $R[X]/(d)R[X]$ and
- the element $1+(d)R[X]$ is called the *unity* or *unit element*.

We often simply denote these elements by 0 and 1, respectively.

**Theorem 4.2.4** (Arithmetical Rules)**.** *For arbitrary $a \in R[X]/(d)R[X]$ we have*

- $a+0=a$ *and* $0+a=a$*;*
- $a\cdot 0=0$ *and* $0\cdot a=0$*;*
- $a\cdot 1=a$ *and* $1\cdot a=a$*;*
- *there exists a unique $b \in R[X]/(d)R[X]$ with $a+b=0$.*

  *The element $b$ is called the* opposite *of $a$ and is written as $-a$. It is also the unique element with $b+a=0$.*

*Proof.* The proofs follow from the corresponding arithmetical rules for addition and multiplication of polynomials. By way of illustration, we prove two equalities.

**For all $a$ we have $a\cdot 0=0$.**

Choose a representative $a'$ from the residue class $a$. Then $a\cdot(0+(d)R[X])=a'\cdot 0+(d)R[X]$ according to the definition of multiplication. The multiplication in $R$ yields $a'\cdot 0=0$, so that we find $a'\cdot 0+(d)R[X]=0+(d)R[X]=(d)R[X]=0$. Hence $a\cdot 0=0$.

**Each element has a unique opposite.**

Given a class $a$ choose a representative $a'$ in it. Now take $b$ to be the class of $-a'$. Then the sum of $a$ and $b$ is the class of $a'+(-a')$, i.e., the class of 0. This establishes that there is at least one opposite.

The proof that there is at most one opposite reads as follows. Suppose that the class $c$ is also an opposite of $a$. Choose a representative $c'$. As $a+c=0$, we find $a'+c'$ to be divisible by $d$. But this implies that $-a'$ and $c'$ are congruent modulo $d$. In particular, their classes coincide: $b=c$.

□

**Example 4.2.5.** Let $R=\mathbb{Z}/2\mathbb{Z}$ and $d=X^3+X+1$. Then the residue class $a$ of $X$ in $R[X]/(d)R[X]$ satisfies $a^7=1$. Indeed, $X^7-1=d\cdot\left(X^4+X^2+X+1\right)$ in $R[X]$, so $a^7-1=0\cdot\left(a^4+a^2+a+1\right)=0$.

Some more rules are given in the theorem below.

**Theorem 4.2.6** (General Arithmetical Rules)**.** *For all a, b, and c in $R[X]/(d)R[X]$ the following equalities hold.*

- $a+b=b+a$ *(commutativity of addition);*
- $a\cdot b=b\cdot a$ *(commutativity of multiplication);*
- $(a+b)+c=a+((b+c))$ *(associativity of addition);*
- $(a\cdot b)\cdot c=a\cdot(b\cdot c)$ *(associativity of multiplication);*
- $a\cdot(b+c)=a\cdot b+a\cdot c$ *(distributivity of multiplication over addition).*

*Proof.* The proofs of arithmetical rules for computing modulo a polynomial follow from the corresponding arithmetical rules for addition and multiplication of polynomials.

□

**Example 4.2.7.** When computing modulo a polynomial, it is of importance to note in which order the computations are carried out. Taking a clever route can gain a lot of time. For example, let $a\in\mathbb{R}[X]/(X^2+1)\mathbb{R}[X]$ be the equivalence class containing the element

$$\left(X^3+1\right)^{27}\cdot\left(X^2+X+1\right)^{35}$$

and suppose that the question is to find a representative of degree at most 1 for $a$. Evidently, it is a lot of work to first work out the product and then find the remainder after division by $X^2+1$. A considerable reduction of the computational work is achieved by the following method, in which we make clever use of the relation for the class $x$ of $X$: $x^2=-1$

Using this relation we compute $\left(x^3+1\right)^{27}\cdot\left(x^2+x+1\right)^{35}=(-x+1)^{27}\cdot(-1+x+1)^{35}=(-x+1)\cdot\left((-x+1)^2\right)^{13}\cdot x^{35}=(-x+1)\cdot(-2\cdot x)^{13}\cdot x^{35}=(-x+1)\cdot(-2)^{13}\cdot x^{48}=2^{13}\cdot x-2^{13}$

So a representative of $a$ is $2^{13}\cdot X-2^{13}$. Verify yourself how the arithmetical rules were used.

Let $R$ be a ring and consider the the restriction of the residue class map to $R$, i.e., the map

$$j\colon R\to R[X]/(d)R[X], a\longmapsto a+(d)R[X] \tag{4.2}$$

**Lemma 4.2.8.** *The map $j$ is injective if $R$ is a field and $d\in R[X]$ is a polynomial of positive degree.*

*Proof.* Suppose that $a,b \in R$ satisfy $j(a) = j(b)$. We then have $j(a-b) = j(0)$. Therefore it suffices to check that if $c \in R$ satisfies $j(c) = 0$, then $c = 0$. Now both $c$ and 0 are representatives of the residue class $j(c)$ having degree less than 1, and hence less than the degree of $d$. As $d$ has positive degree, Theorem on the Representative of Congruence Classes implies $c = 0$.

□

**Example 4.2.9.** Let $R = \mathbb{R}$, the real numbers, and take $d = X^2+1$. Then the residue class ring $\mathbb{R}[X]/(d)\mathbb{R}[X]$ is a description of the complex numbers $\mathbb{C}$, with the role of the complex number $i$ being played by $X + (d)\mathbb{R}[X]$. Indeed, $(X+(d)R[X])^2 = X^2+(d)R[X] = -1+(d)R[X]$ If you let the complex number $a+b\cdot i$ correspond to the class of $a+b\cdot X$, you get the precise correspondence. Here, $j$ is the usual embedding of the real numbers into the complex numbers.

**Remark 4.2.10.** Clearly, the condition that the degree of $d$ be positive is necessary.

Let $R = \mathbb{Z}/6\mathbb{Z}$ and $d = 3\cdot X+1$. Then $j(2) = j(0)$, so $j$ is not injective. This shows that the lemma does not hold if the condition that $R$ be a field is removed.

The injectivity of $j$ tells us that within $R[X]/(d)R[X]$ we find the copy $j(R)$ of $R$, where the term copy refers not only to the bijective correspondence between the sets $R$ and $j(R)$, but also refers to the fact that $j$ respects the operations addition and multiplication.

Let $R$ be a field and $d$ a polynomial of degree $n > 0$ in $R[X]$. The residue class ring $R[X]/(d)R[X]$ carries a vector space structure as follows.

**Theorem 4.2.11.** *The residue class ring $S = R[X]/(d)R[X]$ is a vector space of dimension $n$ over $R$, with*

- *the addition of the ring $S$,*
- *scalar multiplication of the scalar $r \in R$ and the vector $g \in S$ given by the product $r\cdot g$ in the ring $S$.*

*The residue classes of $1, X, \ldots, X^{n-1}$ form a basis of $S$.*

*Proof.* The proof is divided into three steps.

**$S$ is a vector space.**

First we specify the zero vector and the opposite of a vector:

- The zero vector is the class of the zero polynomial.
- The opposite of a vector coincides with the opposite of that element in the ring $S$.

The arithmetical rules for the ring $S$ imply that all the axioms of a vector space over $R$ are satisfied. For example, the 'scalar' $r \in R$ and the 'vectors' $f, g \in S$ satisfy $r \cdot (f+g) = r \cdot f + r \cdot g$.

**The residue classes of $1, X, \ldots, X^{n-1}$ in $S$ span $S$.**

By Division with Remainder Theorem each residue class contains an element of degree at most $n-1$ which can be written as a linear combination of $1, X, \ldots, X^{n-1}$.

**The residue classes of $1, X, \ldots, X^{n-1}$ in $S$ are linearly independent vectors.**

Let $f$ be any linear combination of the elements $1, X, \ldots, X^{n-1}$. Then $f$ is a polynomial of degree less than $n$. If $f$ equals 0 modulo $d$, then $f$ is a multiple of $d$, so, by the Degree Formulas, $\mathrm{degree}(f) \geq \mathrm{degree}(d)$, a contradiction as $\mathrm{degree}(d) = n$. This proves that the vectors are linearly independent.

□

**Example 4.2.12.** Given is the residue class ring $S = \mathbb{Z}/2\mathbb{Z}[X]/(d)\mathbb{Z}/2\mathbb{Z}[X]$, where $d = X^3 + X + 1$. A basis for $S$ as a vector space over $\mathbb{Z}/2\mathbb{Z}$ is $1, X, X^2$. (Notice that, here, we have used the powers of $X$ to denote residue classes in $S$.) With respect to this basis, multiplication by $X$ is a linear map on $S$ expressed by the matrix $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

Let $R$ be a field and $d \in R[X]$ a polynomial of degree $n > 0$. The unique representatives of degree less than $n$ of the various classes in $R[X]/(d)R[X]$ form a subspace $R[X]_{<n}$ of the vector space $R[X]$. A complement is formed by the multiples of $d$:

**Theorem 4.2.13.** *The ring $R[X]$ has the following vector space decomposition: $R[X] = R[X]_{<n} + (d)R[X]$ Furthermore, the map $R[X] \to R[X]_{<n}, f \longmapsto \mathrm{rem}(f,d)$ is the linear projection onto $R[X]_{<n}$ with kernel $(d)R[X]$.*

*Proof.* Division with Remainder Theorem by $d$ shows that every polynomial $f$ can be written in a unique way as the sum of a multiple of $d$ and a polynomial of degree less than $n$ (the remainder). This establishes the first claim.

The map $f \longmapsto \mathrm{rem}(f,d)$ is linear. Indeed, if division with remainder applied to the polynomials $f$ and $g$ yields equalities $f = q \cdot d + r$ and $g = p \cdot d + s$, then for all $a$ and $b$ in $R$ we have $a \cdot f + b \cdot g = (a \cdot q + b \cdot p) \cdot d + (a \cdot r + b \cdot s)$, so that $\mathrm{rem}(a \cdot f + b \cdot g, d) = a \cdot \mathrm{rem}(f,d) + b \cdot \mathrm{rem}(g,d)$.

The kernel of the map consists of course of all multiples of $d$, and the image of the map is precisely $R[X]_{<n}$. Indeed, every polynomial in $R[X]_{<n}$ occurs as remainder upon division by $d$ of that polynomial itself.

□

**Example 4.2.14.** Let $R = \mathbb{Z}/2\mathbb{Z}$ and $d = X^2 + X + 1 \in R[X]$. The matrix of the map $R[X]_{<5} \to R[X]/(d)R[X], f \longmapsto f + (d)R[X]$ with respect to the basis $1, X, X^2, X^3, X^4$ of $R[X]_{<5}$ and the basis $1 + (d)R[X], X + (d)R[X]$ of $R[X]/(d)R[X]$ is $\begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$.

## 4.3 Two special cases

We consider two special cases of computations modulo a polynomial. The first special case is closely related to $n$-th-order approximations of real valued functions.

Consider the map $f \longmapsto \text{rem}\left(f, X^{n+1}\right)$ for polynomials $f$ in $\mathbb{R}[X]$. In terms of polynomial functions $f$ from $\mathbb{R}$ to $\mathbb{R}$, the image of this map corresponds to an approximation of $f$ around 0 of order $n$. We can transfer this principle to arbitrary, sufficiently often differentiable functions.

Let $f$ be a real-valued function defined on an interval containing $0 \in \mathbb{R}$ and sufficiently often differentiable. Then the polynomial $a = a_0 + a_1 \cdot X + \ldots + a_n \cdot X^n$ is called the $n$-th-order approximation of $f$ around 0 if $f(x) = a(x) + O\left(x^{n+1}\right)$ for $x \to 0$.

Recall from Analysis or Calculus that this big Oh notation means that there are positive real constants $C$ and epsilon such that $|f(x) - a(x)| \leq C \cdot \left|x^{n+1}\right|$ for all $x$ with $|x| <$ epsilon.

Such an $n$-th-order approximation is unique; in fact it consists of the first $n+1$ terms of the Taylor series of $f$ around 0.

**Theorem 4.3.1** (Taylor Approximation)**.** *Let $f$ be a continuous $n$-times differentiable real-valued function. Then the polynomial $F = f(0) + \frac{f^{(1)}(0)}{1!} \cdot X + \ldots + \frac{f^{(n)}(0)}{n!} \cdot X^n$ in $\mathbb{R}[X]$ is the $n$-th order approximation of $f$ around 0. Furthermore, if $G$ is an $n$-th order approximation of a function $g$, then* $\text{rem}(F \cdot G, X^{n+1})$ *and* $\text{rem}(F + G, X^{n+1})$ *are the $n$-th-order approximations of $f \cdot g$ and $f + g$, respectively.*

*Proof.* We only give a sketch of the proof. The polynomial function $x \longmapsto F(x)$ is the first part of the Taylor series expansion of $f$. From Calculus or Analysis it follows that there exists a real-valued function $h$ satisfying $f(x) = F(x) + x^{n+1} \cdot h(x)$ for $x$ in the neighbourhood of 0.

From this we conclude that $F$ is an $n$-th-order approximation of $f$ around 0.

Considering the second part of the theorem, suppose $g(x) = G(x) + O\left(x^{n+1}\right)$ for $x$ going to 0.

Then we have $f \cdot g - F \cdot G(x) = f \cdot g - F \cdot g(x) + F \cdot g - F \cdot G(x) = (f(x) - F(x)) \cdot g(x) + F(x) \cdot (g(x) - G(x)) = O\left(x^{n+1}\right) \cdot g(x) + F(x) \cdot O\left(x^{n+1}\right) = O\left(x^{n+1}\right)$ for $x$ going to 0.

So $F \cdot G$ is indeed the $n$-th-order approximation of $f \cdot g$ around 0.

The proof for $f + g$ is simpler. Do it yourself.

□

**Example 4.3.2.** The second-order approximation of the function $x \longmapsto e^x$ around 0 is the function $x \longmapsto 1+x+x^2/2$.

The second-order approximation of the function $x \longmapsto \sin(x)$ is the function $x \longmapsto x$.

But then the second order approximation of the product function $x \longmapsto \sin(x) \cdot e^x$ equals the function $x \longmapsto x+x^2$, which is the remainder of the division of $x \cdot \left(1+x+x^2/2\right)$ by $x^3$.

The second special case to discuss is arithmetic modulo the constant polynomial $n$ (greater than 0) in the polynomial ring $\mathbb{Z}[X]$. Two polynomials in $\mathbb{Z}[X]$ are congruent modulo $n$ if and only if for each $i$, the coefficients of $X^i$ differ by a multiple of $n$. Therefore, each residue class has a representative all of whose coefficients lie in $\{0,1,...,n-1\}$. This is similar for polynomials over $\mathbb{Z}/n\mathbb{Z}$. The relation is clarified by the following map.

$I\colon \mathbb{Z}[X]/(n)\mathbb{Z}[X] \to \mathbb{Z}/n\mathbb{Z}[X], a_0+a_1 \cdot X+...+a_m \cdot X^m+(n)\mathbb{Z}[X] \longmapsto a_0(\text{mod }) + a_1(\text{mod }) \cdot X + ... + a_m(\text{mod }) \cdot X^m$.

Since this map is constructed using representatives, we have to check that the result does not depend on the representatives chosen.

**Theorem 4.3.3.** *The map I is well defined and has the following properties.*

- *It is a bijection.*
- *It respects addition:* $I(a+b) = I(a)+I(b)$.
- *It respects the zeros:* $I(0+n) = 0$.
- *It respects multiplication:* $I(a \cdot b) = I(a) \cdot I(b)$.
- *It respects the units:* $I(1+n) = 1$.

*Proof.*

***I* is well defined.**

Let $a = a_0+a_1 \cdot X+...+a_m \cdot X^m$ and $b = b_0+b_1 \cdot X+...+b_m \cdot X^m$ be two polynomials that are congruent modulo $n$ (according to the convention in Chapter 3 we may assume the highest power of a monomial in both $a$ and $b$ to be equal to $m$). Then $a$ and $b$ differ by a multiple of $n$ for $i = 0,1,...,m$. This implies that $a_i \equiv b_i \pmod{n}$ for $i = 0,1,...,m$. So our definition does not depend on the representative $a$ or $b$ that we have chosen.

***I* respects addition.**

Suppose that $a = a_0+a_1 \cdot X+...+a_m \cdot X^m$ and $b = b_0+b_1 \cdot X+...+b_k \cdot X^k$ are elements of $\mathbb{Z}[X]$. Then, adding some powers of $X$, we can assume that $k = m$. Now $I(a+b+(n)\mathbb{Z}[X])$ equals $((a_0)+(b_0))+(a_1+b_1) \cdot X+...+(a_m+b_m) \cdot X^m$ in $\mathbb{Z}/n\mathbb{Z}[X]$, which is equal to $a_0 + a_1 \cdot X+...+a_m \cdot X^m + (b_0+b_1 \cdot X+...+b_m \cdot X^m)$.

But the latter is equal to $I(a+(n)\mathbb{Z}[X]) + I(b+(n)\mathbb{Z}[X])$.

**$I$ respects zeros.**

Indeed, $I\left(0+(n)\mathbb{Z}\left[X\right]\right)=0$.

**$I$ respects multiplication.**

The proof is similar to the proof of the fact that $I$ respects addition.

**$I$ respects units.**

Indeed, $I\left(1+(n)\mathbb{Z}\left[X\right]\right)=1$.

**$I$ is a bijection.**

Suppose that $a$ and $b$ are in $\mathbb{Z}\left[X\right]$ and satisfy $I\left(a\right)=I\left(b\right)$. As $I$ respects addition and scalar multiplication, $I\left(a-b\right)=0$. But then it is straightforward to check that $a-b=0$ modulo $(n)\mathbb{Z}\left[X\right]$ and hence $a=b$ modulo $(n)\mathbb{Z}\left[X\right]$.

□

**Example 4.3.4.** The image of $3+6\cdot X+8\cdot X^2+2\cdot X^3-88\cdot X^4\in\mathbb{Z}\left[X\right]/(5)\mathbb{Z}\left[X\right]$ under the map $I$ of the theorem is $3+X+3\cdot X^2+2\cdot X^3+2\cdot X^4\in\mathbb{Z}/5\mathbb{Z}\left[X\right]$.

The conclusion of the above result is that the arithmetic in $\mathbb{Z}\left[X\right]/(n)\mathbb{Z}\left[X\right]$ is nothing but the arithmetic in $\mathbb{Z}/n\mathbb{Z}\left[X\right]$. In mathematical jargon: The two arithmetical structures are isomorphic (i.e., equal of form).

## 4.4 Inverses and fields

Let $R$ be a ring like $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C}$, or $\mathbb{Z}/n\mathbb{Z}$ and $d$ a polynomial in $R\left[X\right]$. In the newly constructed arithmetical system $R\left[X\right]/(d)R\left[X\right]$ we have not yet considered division, since it comes with various complications.

**Definition 4.4.1.** Suppose that $d$ is a nonconstant polynomial in $R\left[X\right]$. Then $f\in R\left[X\right]/(d)R\left[X\right]$ is called *invertible* with respect to multiplication if there exists a $g\in R\left[X\right]/(d)R\left[X\right]$ satisfying $f\cdot g=1$. Such an element $g$ is called an *inverse* of $f$ and is denoted by $\frac{1}{f}$, $1/f$, or $f^{-1}$.

**Remark 4.4.2.** Suppose that $f$ is an invertible residue class in $R\left[X\right]/(d)R\left[X\right]$ and both $g$ and $h$ are inverses of $f$. Then $g=g\cdot 1=g\cdot(f\cdot h)=(g\cdot f)\cdot h=1\cdot h=h$ Therefore, $f$ has a unique inverse.

To guarantee the existence of inverses in $R$, we assume that $R$ is a field (think of $\mathbb{Q},\mathbb{R},\mathbb{C}$ or $\mathbb{Z}/p\mathbb{Z}$ with $p$ a prime). Let $d$ be a polynomial in $R\left[X\right]$ of positive degree.

The following characterization of the invertible elements in $R\left[X\right]/(d)R\left[X\right]$ yields also a way of computing inverses with the help of the Extended Euclidean Algorithm for Polynomials.

**Theorem 4.4.3** (Characterization of Invertibility in Residue Class Rings)**.** *Let $a$ be a polynomial in $R[X]$. Then the residue class $a+(d)R[X]$ in $R[X]/(d)R[X]$ has an inverse if and only if* $\gcd(a,d)=1$.

*Proof.*

**If.**

If the residue class $a+(d)R[X]$ has inverse $b+(d)R[X]$, then $a\cdot b=1+(d)R[X]$. Hence there is a polynomial $p$ with $a\cdot b+p\cdot d=1$.

According to the Result on divisors of a product, $\gcd(a,d)=1$.

**Only if.**

If $\gcd(a,d)=1$, then the Extended Euclidean Algorithm for Polynomials produces polynomials $b$ and $p$ such that $a\cdot b+p\cdot d=1$. But then $b$ represents an inverse of the residue class $a+(d)R[X]$.

□

**Example 4.4.4.** We take $R=\mathbb{R}$ and $d=X^n$ with $n>0$. Then a class represented by the polynomial $a$ is invertible in $R[X]/(d)R[X]$ if and only if the constant term of $a$ differs from 0.

Characterization of Invertibility in Residue Class Rings allows us to construct new fields.

**Corollary 4.4.5** (Characterization of Fields among Residue Class Rinigs)**.** *Let $R$ be a field and $d$ an irreducible polynomial in $R[X]$. Then $S=R[X]/(d)R[X]$ is a field, i.e., every nonzero element in $S$ has an inverse.*

*Proof.* Consider a residue class different from 0 and let $a$ denote a representative of this class. Then $a$ is not a multiple of $d$.

Since $d$ is irreducible, $\gcd(a,d)$ equals 1 or $d$. As $a$ is nonzero modulo $d$, the second possibility is excluded. So $\gcd(a,d)=1$, and, by Characterization of Invertibility in Residue Class Rings, the class of $a$ is invertible.

We conclude that all nonzero elements in $S$ are invertible and $S$ is indeed a field.

□

**Example 4.4.6.** We take $R=\mathbb{Z}/2\mathbb{Z}$ and $d=X^2+X+1$. Then $R[X]/(d)R[X]$ contains the following four elements: 0, 1, $a$ and $a+1$, where $a=X+(d)R[X]$.

| $\cdot$ | $0$ | $1$ | $a$ | $a+1$ |
|---|---|---|---|---|
| $0$ | $0$ | $0$ | $0$ | $0$ |
| $1$ | $0$ | $1$ | $a$ | $a+1$ |
| $a$ | $0$ | $a$ | $a+1$ | $1$ |
| $a+1$ | $0$ | $a+1$ | $1$ | $a$ |

Table 4.1: The multiplication table of a quotient ring.

The multiplication table for the four elements from $R[X]/(d)R[X]$ is as follows:

The table shows that $a$ and $a+1$ are each other's inverses. Compare this table with the multiplication table of $\mathbb{Z}/4\mathbb{Z}$. In $\mathbb{Z}/4\mathbb{Z}$ there is no element $b$ with $2\cdot b = 1$. The element 2 of $\mathbb{Z}/4\mathbb{Z}$ has no inverse. Therefore, the arithmetical system on 4 elements we have just constructed is fundamentally different from $\mathbb{Z}/4\mathbb{Z}$.

## 4.5 Finite fields

Up to now we have encountered the following finite fields, where $p$ a prime. $\mathbb{Z}/p\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z}[X]/(d)\mathbb{Z}/p\mathbb{Z}[X]$ with $d$ an irreducible polynomial.

The theory of finite fields tells us that these are the only finite fields. This will not be shown here, but is postpond to later. Nevertheless, we state the main result on finite fields.

**Theorem 4.5.1** (Classification of Finite Fields)**.** *For each prime p and positive integer n there exists an irreducible polynomial d of degree n in* $\mathbb{Z}/p\mathbb{Z}[X]$. *The residue class ring* $\mathbb{Z}/p\mathbb{Z}[X]/(d)\mathbb{Z}/p\mathbb{Z}[X]$ *is a finite field.*
*Any finite field can be constructed in this way.*

**Example 4.5.2.** In order to construct a field of 9 elements, we have to find an irreducible polynomial of degree 2 over $\mathbb{Z}/3\mathbb{Z}$. The monic irreducible polynomials of degree 2 are $X^2 + X + 1, X^2 - X - 1, X^2 + 1$. So, we can construct a field of 9 elements by taking the residue class ring $S = \mathbb{Z}/3\mathbb{Z}[X]/(d)\mathbb{Z}/3\mathbb{Z}[X]$ where $d = X^2 + 1$. One of the special properties of finite fields is their uniqueness. For example, had we taken one of the other two irreducible polynomials of degree 2, we would essentially have obtained the same field.

Although we do not prove the Classification of Finite Fields at this moment, we will investigate the finite fields somewhat closer. First, we determine the cardinality of such fields.

Let $p$ be a prime number and $n$ a positive integer.

**Theorem 4.5.3.** *If $d$ is an irreducible polynomial over $\mathbb{Z}/p\mathbb{Z}$ of degree $n$, then $\mathbb{Z}/p\mathbb{Z}[X]/(d)\mathbb{Z}/p\mathbb{Z}[X]$ is a field with exactly $p^n$ elements.*
*Moreover, this field is the unique field with $p^n$ elements.*

*Proof.* According to Characterization of Fields among Residue Class Rinigs, the residue class ring $S = \mathbb{Z}/p\mathbb{Z}[X]/(d)\mathbb{Z}/p\mathbb{Z}[X]$ is a field. On the other hand, $S$ is a vector space over $\mathbb{Z}/p\mathbb{Z}$ of dimension $n$ (see the ). There are exactly $p$ possible coefficients for every basis vector, so this leads to $p^n$ elements.

Uniqueness of the field will not be proven here. This will be discussed in later chapters.

□

**Example 4.5.4.** Let $f = X^3 + X + 1$ be a polynomial in $\mathbb{Z}/2\mathbb{Z}[X]$. The residue class ring $\mathbb{Z}/2\mathbb{Z}[X]/(f)\mathbb{Z}/2\mathbb{Z}[X]$ has 8 elements. We present the mutiplication table of the 7 nonzero elements. Here $a$ represents the class of $X$ modulo $f$.

| $\cdot$ | $1$ | $a$ | $1+a$ | $a^2$ | $a^2+1$ | $a^2+a$ | $a^2+a+1$ |
|---|---|---|---|---|---|---|---|
| $1$ | $1$ | $a$ | $1+a$ | $a^2$ | $a^2+1$ | $a^2+a$ | $a^2+a+1$ |
| $a$ | $a$ | $a^2$ | $a^2+a$ | $1+a$ | $1$ | $a^2+a+1$ | $a^2+1$ |
| $1+a$ | $1+a$ | $a^2+a$ | $a^2+1$ | $a^2+a+1$ | $a^2$ | $1$ | $a$ |
| $a^2$ | $a^2$ | $1+a$ | $a^2+a+1$ | $a^2+a$ | $a$ | $a^2+1$ | $1$ |
| $a^2+1$ | $a^2+1$ | $1$ | $a^2$ | $a$ | $a^2+a+1$ | $1+a$ | $a^2+a$ |
| $a^2+a$ | $a^2+a$ | $a^2+a+1$ | $1$ | $a^2+1$ | $1+a$ | $a$ | $a^2$ |
| $a^2+a+1$ | $a^2+a+1$ | $a^2+1$ | $a$ | $1$ | $a^2+a$ | $a^2$ | $1+a$ |

Notice that in each row (and each column) of the table one finds a 1, implying that each element has an inverse. So, $\mathbb{Z}/2\mathbb{Z}[X]/(f)\mathbb{Z}/2\mathbb{Z}[X]$ is a field and $f$ is an irreducible polynomial.

Let $p$ be a prime, $n$ a positive integer, and $d$ an irreducible polynomial in $\mathbb{Z}/p\mathbb{Z}$ of degree $n$. We are concerned with the finite field $S = \mathbb{Z}/p\mathbb{Z}[X]/(d)\mathbb{Z}/p\mathbb{Z}[X]$.

**Theorem 4.5.5.** *Write $q = p^n$ for the cardinality of $S$. Then, for each $a, b \in S$,*

1. $a + a + ... + a = 0$ *(with $p$ terms);*
2. $(a+b)^p = a^p + b^p$*;*
3. $a^q = a$ *(Fermat's Little Theorem).*

*Proof.* We prove the three parts of the theorem separately.

**Part 1.** $a + a + ... + a = 0$ **(with** $p$ **terms).**

We have $a + a + ... + a = (1 + 1 + ... + 1) \cdot a = p \cdot a = 0$.

**Part 2.** $(a+b)^p = a^p + b^p$**.**

Expand $(a+b)^p$ by means of Newton's Binomium. As each binomial coefficient $\begin{pmatrix} p \\ i \end{pmatrix}$ with $i$ different from 0 and $p$ is zero modulo $p$ (see the proof of Fermat's Little Theorem), we find $(a+b)^p = a^p + b^p$.

**Part 3.** $a^q = a$**.**

The proof we give here is similar to the second proof of Fermat's Little Theorem.

For $a = 0$ the statements are trivial. Assume that $a$ is nonzero. Consider the set $S^\times$ of invertible (that is, nonzero, because $S$ is a field) elements from $S$. On it, we define the map $M_a = S^\times \to S^\times, b \longmapsto a \cdot b$ multiplication by $a$. This map is bijective. Indeed, its inverse equals $M_{a^{-1}}$, multiplication by the inverse of $a$. As a result we see that the product of all elements in $S^\times$ equals not only $\prod_{b \in S^\times} b$ but also $\prod_{b \in S^\times} (M_a(b))$ as here the order of the factors in the product is all that has changed. The latter product equals $\prod_{b \in S^\times} ((a \cdot b)) = a^{q-1} \cdot \prod_{b \in S^\times} b$ As the product is nonzero, it is invertible. Dividing by this product, we deduce that $a^{q-1} = 1$. Multiplying both sides of the equation with $a$ proves the assertion.

□

The first identity of Special Identities in Finite Fields can also be written as $p \cdot a = 0$. In mathematical jargon, it is referred to by saying that the *characteristic* of $S$ is $p$.

The second identity is also called the Freshman's Dream, as it concurs with the outcome of ordinary power expansions by many freshmen who forget about cross products.

The third identity is just Fermat's Little Theorem for finite fields! (Note that the proof does not use the particular construction of the field $S$.)

Special Identities in Finite Fields implies that every nonzero element in a field $S$ with $q$ elements raised to the power $q - 1$ is equal to 1.

An element of $S$ having no smaller (positive) power equal to 1 is called *primitive*. In general, for $a$ in $S$, the smallest positive number $l$ satisfying $a^l = 1$ is called the *order* of $a$. So a nonzero element of $S$ is primitive if its order is $q-1$.

Without proof we state:

**Theorem 4.5.6.** *Every finite field has a primitive element.*

## 4.6 Error correcting codes

In RSA Decription and Encryption we introduced the RSA cryptosystem. Using this system, one can transform sensitive information into a code that is hard (if not impossible) for outsiders to crack. On the opposite side, however, transportation of data can lead to unwanted errors. So, it is often necessary to secure the information to be sent in such a way that errors can be detected or even corrected.

**Definition 4.6.1** (Coding theory)**.** Coding theory is the branch of mathematics where one considers ideas that make it possible to encode information in such a way that errors, occurred during transmission or caused by other reasons, are corrected.

**Example 4.6.2** (CD and DVD)**.** A Game, music or video is stored on a CD or DVD in the form of a code. Using a laser beam, the CD player reads the information on the disc and converts it into information transmitted to the viewer or listener. However, the player can make real errors in reading: there can be scratches or little pieces of dirt on the disc, the laser beam just misses the right place on the disc, and so on. Nevertheless we want the music to be replayed as well as possible. We want the CD player to correct its reading errors. The game, video or music has to be stored on disc in such a way that the player can correct its errors.

**Example 4.6.3** (Satellite)**.** Satellites hang above the earth. Information, for example, a TV program, is sent from one place on earth to the satellite, which sends it back to other places on earth. In this way we can follow important events live on TV. However, the signals going to and coming from the satellite suffer from noise. The TV watcher does not want to notice the damage to the live images.

**Example 4.6.4** (Fax and email)**.** Faxes and e-mail messages are transmitted via telephone lines throughout the world. Telephone lines also suffer from noise. This can cause a fax to be damaged. The fax has to be protected against this.

**Example 4.6.5** (Parity check)**.** A trivial way to secure your information is to keep copies of it. A somewhat more advanced way is to include control characters in your information. Suppose that your information is a string of zeros and ones. Now add at each 8-th position a control character equal to 0 or 1 such that the sum of the control character and the seven preceding characters are even. So,

$$110110011010001110011 \tag{4.3}$$

is transformed into

$$1101100011010001111100111 \tag{4.4}$$

If at most one mistake occurs in each substring of eight characters, these errors can be detected, but not corrected.

**Example 4.6.6** (ISBN)**.** Each book is given a number, the so-called International Standard Book Number, abbreviated to ISBN. The ISBN consists of 10 symbols. The first 9 symbols are digits giving information on the book, like the year and place it is published. The last symbol is a check symbol and is either a digit or the symbol $x$ (representing 10). If the ISBN of a book is $a_1,\ldots,a_9,b$, then the following relation is satisfied. $a_0+2\cdot a_1+\ldots+9\cdot a_9 \equiv b \pmod{11}$. If one of the symbols is incorrect, then the above equality is violated. This makes it possible to detect an error.

We come now to a mathematical description of coding theory.

**Definition 4.6.7.** Let $V$ be a vector space over $\mathbb{Z}/p\mathbb{Z}$ with $p$ a prime.

A *code* in $V$ is a set of vectors in $V$. The vectors of a code are called *code words*. A *linear code* in $V$ is a linear subspace of $V$. If $C$ is a linear code of dimension $k$ in the $n$-dimensional vector space $V$, then $C$ is referred to as an $(n,k)$-code.

**Example 4.6.8.** We consider the numbers $0,\ldots,15$ in their binary representation (see $b$-ary representation), i.e., sequences of length 4, each element of which is either 0 or 1. So 0 is represented as $[0,0,0,0]_2$, 7 by $[0,1,1,1]_2$ and 13 by $[1,1,0,1]_2$.

A mistake in reading such a string causes a wrong number to be read. The following can help to prevent this. We encode these numbers by vectors in $(\mathbb{Z}/2\mathbb{Z})^7$. Such a vector is often written, in short, as a word in the alphabet $\{0,1\}$:

$(0,0,1,0,0,1,1)$ is written as $0,0,1,0,0,1,1$.

The first 4 coordinates form the binary notation of the number. The remaining 3 positions are filled in the following way:

Note that the 16 vectors form indeed a vector space. Caution: the vector space addition in $(\mathbb{Z}/2\mathbb{Z})^7$ does not correspond to the addition of the numbers connected to the vectors. The following property is crucial for its coding capacity: any two vectors differ in at least 3 positions. So if we make at most one reading error, for example, we read 1101110 instead of 1101010, we can still decide that we are dealing with the number 13. Indeed, the vectors for all the other numbers differ in at least 2 positions from 1101110. Therefore, we are able to correct one reading error. We say that the code above for the numbers $0,\ldots,15$ is a 1-error correcting code. If at most one error is made, we can correct it. A complication is that we do not know a priori how many reading errors have been made. If 6 errors are possible, the original could have been any number.

| | |
|---|---|
| 0 | 0,0,0,0,0,0,0 |
| 1 | 0,0,0,1,0,1,1 |
| 2 | 0,0,1,0,1,0,1 |
| 3 | 0,0,1,1,1,1,0 |
| 4 | 0,1,0,0,1,1,0 |
| 5 | 0,1,0,1,1,0,1 |
| 6 | 0,1,1,0,0,1,1 |
| 7 | 0,1,1,1,0,0,0 |
| 8 | 1,0,0,0,1,1,1 |
| 9 | 1,0,0,1,1,0,0 |
| 10 | 1,0,1,0,0,1,0 |
| 11 | 1,0,1,1,0,0,1 |
| 12 | 1,1,0,0,0,0,1 |
| 13 | 1,1,0,1,0,1,0 |
| 14 | 1,1,1,0,1,0,0 |
| 15 | 1,1,1,1,1,1,1 |

Now we address the real ‘coding’ aspects.

**Definition 4.6.9.** Let $C$ be a code in the vector space $V$. The *distance* between two vectors from $V$ is the number of coordinate positions at which the two vectors differ. The *minimal distance* of $C$ is the minimum taken over all distances between any two different code words from $C$.

*Proof.* We show that the distance delta as defined indeed satisfies the axioms for a distance function with values in $\mathbb{N}$, viz., $\text{delta}\,(v,w)=0$ if and only if $v=w$, symmetry: $\text{delta}\,(v,w)=\text{delta}\,(w,v)$, and the triangle inequality: $\text{delta}\,(v,w)+\text{delta}\,(w,u)\geq\text{delta}\,(v,u)$, where $u$, $v$, and $w$ belong to $V$.

$\text{delta}\,(v,w)=0$ **if and only if** $v=w$**.**

Clearly, $v$ and $w$ differ in zero positions if and only if they coincide.

**Symmetry:** $\text{delta}\,(v,w)=\text{delta}\,(w,v)$**.**

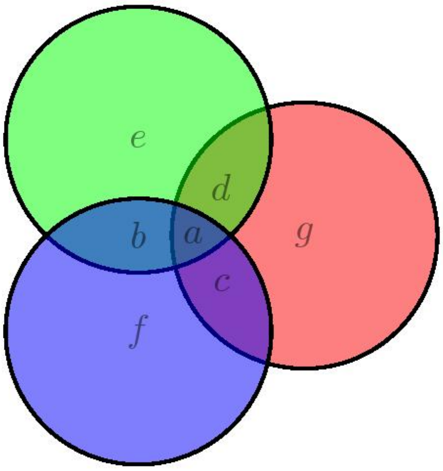
The number of positions in which $v$ and $w$ differ is obviously the same as the number of positions in which $w$ and $v$ differ.

**Triangle inequality:** $\text{delta}\,(v,w)+\text{delta}\,(w,u)\geq\text{delta}\,(v,u)$**.**

Let $S$ be the set of positions in which $v$ and $w$ differ and let $T$ denote the set of positions in which $w$ and $u$ differ. Then $v$ and $u$ differ only in positions within $S\cup T$. In particular, $\text{delta}\,(v,u)\leq|S\cup T|$. As $|S\cup T|\leq|S|+|T|$, $|S|=\text{delta}\,(v,w)$, and $|T|=\text{delta}\,(w,u)$, this implies the triangle inequality.

□

**Example 4.6.10.** The code in Example 4.6.8 can also be depicted graphically. Let $x$ be a number in $\{0,...,15\}$. In the diagram below we fill the positions $a,b,c,d$ with zeros and ones in such a way that $[a,b,c,d]_2$ forms the binary notation of $x$. We then fill the positions $e,f,g$ with zeros and ones in such a way that any circle contains an even number of zeros. Now the code word for the number $x$ is $a,b,c,d,e,f,g$. The figure can also be used for a given vector $r$ in $(\mathbb{Z}/2\mathbb{Z})^7$ to determine the numbers $x$ for which the code word differs in at most one position from $r$. Indeed, given $r$, change at most one position in such a way that we get an even number of ones in each circle. Then the number $x$ is the number with binary notation $[a,b,c,d]_2$.



If the minimal distance of a code $C$ is equal to $d$, then any word differing in at most $d-1$ positions from a code word $w$, is either equal to $w$ or not a code word. Therefore minimal distance $d$ implies perfect detection of at most $d-1$ errors. If $d > 2\cdot e$, it is possible to correct $e$ errors. Indeed, using the triangle inequality we find that a word $v$ at distance at most e from a code word $w$, has distance greater than $e$ to any code word distinct from $w$.

The smaller the length and the larger its minimal distance the more useful the code is. In the remainder of this section we will describe a method for constructing useful error-correcting codes with the help of polynomials.

In the world of digital communication, the binary number system is used a lot. In most applications and examples we confine ourselves to codes in vector spaces over $\mathbb{Z}/2\mathbb{Z}$. In these vector spaces, scalar multiplication is very simple: there are only two scalars, 0 and 1. These codes are known as *binary codes*.

Suppose that $p$ is a prime. In the polynomial ring $\mathbb{Z}/p\mathbb{Z}[X]$ we consider the polynomial $X^n-1$ with $n>1$ and the residue class ring $S = \mathbb{Z}/p\mathbb{Z}[X]/(X^n-1)\mathbb{Z}/p\mathbb{Z}[X]$. This ring has the structure of a vector space over the field $\mathbb{Z}/p\mathbb{Z}$ with basis $1,...,X^{n-1}$, cf. . So each element of $S$ can be represented by the vector of coefficients with respect to this basis, and vice versa:

$a = a_0 + a_1\cdot X + ... + a_{n-1}\cdot X^{n-1} + (X^n-1)\mathbb{Z}/p\mathbb{Z}[X]$ corresponds bijectively to $a = (a_0, a_1, ..., a_{n-1})$

The polynomial $X^n - 1$ is reducible for $n > 1$: it is divisible by $X - 1$.

**Definition 4.6.11.** Let $g$ be a divisor of $X^n - 1$ over $\mathbb{Z}/p\mathbb{Z}$. The image under the linear map $\mathbb{Z}/p\mathbb{Z}[X] \to S, a \longmapsto a \cdot g + (X^n - 1)\mathbb{Z}/p\mathbb{Z}[X]$ is called the *cyclic code* of length $n$ generated by $g$.

**Example 4.6.12.** The polynomial $X^7 - 1$ over $\mathbb{Z}/2\mathbb{Z}$ is the following product of irreducible polynomials: $(X+1) \cdot (X^3 + X + 1) \cdot (X^3 + X^2 + 1)$ If $g = X^3 + X + 1$, then the cyclic code generated by $g$ is a linear $(7,4)$-code. Compare this code with the code discussed in Example 4.6.8.

Let $l$ be the degree of $g$ and write $k = n - l$. The elements $g, ..., X^{k-1} \cdot g$ form a basis for the image space $C$ of the map from . So the dimension of $C$ is equal to $k$. The space $C$ is called the *code generated by* $g$ . The polynomial $g$ is known as the *generator* of $C$. The quotient $(X^n - 1)/g$ is called the *check polynomial* of $C$.

We use the map of  to convert a so-called *information vector* from $(\mathbb{Z}/p\mathbb{Z})^k$ into a code word in $C$. This is done as follows:

- Let $a = (a_0, a_0, ..., a_{k-1})$ be a vector in $(\mathbb{Z}/p\mathbb{Z})^k$.
- Form the polynomal $a = a_0 + a_1 \cdot X + ... + a_{k-1} \cdot X^{k-1}$ in $\mathbb{Z}/p\mathbb{Z}[X]$.
- Determine the representative $c$ of the class $a \cdot g + (X^n - 1)\mathbb{Z}/2\mathbb{Z}[X]$ of lowest degree, that is, $c = \text{rem}(a \cdot g, X^n - 1)$, cf. Division with Remainder Theorem.
- This polynomial $c$ corresponds to a code word $c$. Thus, the information vector $a$ is transformed into the code word $c$.

Let $C$ be a cyclic $(n,k)$-code with generator $g$. We present a way to estimate how useful the cyclic code generated by $g$ is.

Naturally, it is important to be able to find the information vector corresponding to a code word. For this, the check polynomial $h = (X^n - 1)/g$ is used.

**Theorem 4.6.13** (Cyclic Decoding Theorem)**.** *Let $C$ be a cyclic code of length $n$ generated by $g$ and let $h = (X^n - 1)/g$ be the check polynomial of $g$. If $c$ is a code word, viewed as a polynomial of degree at most $n - 1$, then the information vector corresponding to the code word $c$ equals* $-(\text{rem}(c \cdot h, X^n))$.

**Example 4.6.14.** Take $g = (X+1) \cdot (X^3 + X + 1) \in \mathbb{Z}/2\mathbb{Z}[X]$ to be a generator of a cyclic code of length 7. The corresponding check polynomial is $h = X^3 + X^2 + 1$. Now, choose an information vector, say, $a = X$. It maps to code word $c = \text{rem}(a \cdot g, X^7 - 1) = X^5 + X^4 + X^3 + X$ Since $c \cdot h = X^8 + X$, the polynomial of minimal degree in $c \cdot h + (X^7)\mathbb{Z}/2\mathbb{Z}[X]$ is $X$, which coincides with $a$.

*Proof.* Consider $c \in C$ as a polynomial. Suppose that $c$ comes from the information vector $a$, also considered as a polynomial, of degree at most $k-1$. Then $c = a \cdot g + m$ for a polynomial $m \in (X^n - 1)\mathbb{Z}/p\mathbb{Z}[X]$. By the Degree Formulas, the degrees of $c$ and of $a \cdot g$ are at most $n-1$. Therefore the degree of $m$ is at most $n-1$, too, and so $m = 0$. In particular, $c = a \cdot g$, and we obtain the following relation between $c$ and $a$: $c \cdot h = a \cdot g \cdot h = a \cdot (X^n - 1) = X^n \cdot a - a$

After Division with Remainder Theorem, we conclude $-a = \text{rem}(c \cdot h, X^n)$.

□

Let $d$ and $g$ be polynomials in the polynomial ring $R[X]$. We will consider the residue class ring $S = R[X]/(d)R[X]$. For an element $s \in S$ the substitution of $s$ for $X$ in $g$ gives the element $g(s)$ of $S$, see the .

If $g$ equals $d$ and $s$ is the class of $X$ modulo $d$, then $g(s) = 0$. In this particular case, the image of $X$ in $S$ is a zero of $g$ in $S$, cf. Characterization of the Zeros of a Polynomial.

The following result shows how useful codes can be built by means of modular polynomial arithmetic. The code $C$ of our interest is a cyclic $(n,k)$ code with generator polynomial $g$.

**Theorem 4.6.15** (BCH bound)**.** *Set $d = X^n - 1$ and write $S = \mathbb{Z}/p\mathbb{Z}[X]/(X^n - 1)\mathbb{Z}/p\mathbb{Z}[X]$, where $p$ is a prime. Suppose that $g$ is a divisor of $d$ in $\mathbb{Z}/p\mathbb{Z}[X]$. Let $a$ be the residue class of $X$ in $S$.*
*If the set $J$ of all positive integers $j$ with $g\left(a^j\right) = 0$ contains a sequence of $m$ consecutive integers, then the minimal distance of the $(n,k)$-code $C$ generated by $g$ is at least $m+1$.*

**Example 4.6.16.** Take for $g$ the polynomial $X^3 + X + 1$ in $\mathbb{Z}/2\mathbb{Z}[X]$. Then $g$ divides $X^7 - 1$ and accordingly we consider the binary cyclic code of length 7 generated by $g$. According to the BCH bound, the minimum distance of the code $C$ generated by $g$ is at least 3. Indeed, if $a$ is the residue class of $X$ modulo $X^7 - 1$, then both $a$ and $a^2$ are roots of $g$. So BCH bound can be applied with $p = 2$ and $m = 2$. Note that 3 is also the minimum distance of $C$.

By choosing the generating polynomial in a clever way, codes can be constructed that correct multiple errors. BCH stands for Bose, Ray-Chaudhuri, and Hocquenghem, the three mathematicians who discovered the bound.

## 4.7 Exercises

### 4.7.1 Congruence modulo a polynomial

**Exercise 4.7.1.** Determine in each of the following cases whether the polynomials $a$ and $b$ are congruent modulo $c$.

1. $a = X^3, b = 1, c = X^2 + X + 1$ as polynomials over $\mathbb{Q}$.
2. $a = X^4 + X + 2, b = X + 3, c = X + 1$ as polynomials over $\mathbb{Z}/5\mathbb{Z}$.
3. $a = \left(X^3 + X + 1\right)^5, b = \left(X^2 + 2 \cdot X\right)^5, c = X - 1$ over $\mathbb{Q}$.

**Exercise 4.7.2.** In each of the following cases, the polynomials $a$ are $d$ given. Find a representative of the residue class of $a$ modulo $d$ whose degree is less than the degree of $d$.

1. $a = X^4, d = X^2 + X + 1$ in $\mathbb{Q}[X]$,
2. $a = X^4 + X^2 + 1, d = X^2 + X + 1$ in $\mathbb{Z}/2\mathbb{Z}[X]$.

**Exercise 4.7.3.** Determine representatives for all congruence classes for each of the following residue class rings.

1. $\mathbb{Z}/2\mathbb{Z}[X]/(X^3 + 1)\mathbb{Z}/2\mathbb{Z}[X]$,
2. $\mathbb{Q}[X]/(X - 1)\mathbb{Q}[X]$,
3. $\mathbb{R}[X]/(2)\mathbb{R}[X]$.

## 4.7.2 The residue class ring

**Exercise 4.7.4.** Consider the residue class $a$ of $X$ in $S = \mathbb{Z}/2\mathbb{Z}[X]/(X^2 + X + 1)\mathbb{Z}/2\mathbb{Z}[X]$.

1. Describe the elements of $S$ in terms of ‘polynomials’ in $a$.
2. Compose a multiplication table for $S$.
3. Show that $a^{17} = a + 1$.

**Exercise 4.7.5.** Let $a \in \mathbb{R}$. We define the map eval: $\mathbb{R}[X]/(X - a)\mathbb{R}[X] \to \mathbb{R}$ by $\mathrm{eval}\,(f + (X - a)\mathbb{R}[X]) = f(a)$.

1. Show that this map is well defined.
2. Show that eval is a bijection.

**Exercise 4.7.6.** We define the two maps $f_+$ and $f_-$ from $\mathbb{Q}[X]/(X^2 - 2)\mathbb{Q}[X]$ to $\mathbb{Q} + \mathbb{Q} \cdot \sqrt{2}$ in the following way. For any residue class $g + (X^2 - 2)\mathbb{Q}[X]$ we have

$f_+\left(g + (X^2 - 2)\mathbb{Q}[X]\right) = g\left(\sqrt{2}\right)$ and

$f_-\left(g + (X^2 - 2)\mathbb{Q}[X]\right) = g\left(-(\sqrt{2})\right)$

1. Show that $f_+$ and $f_-$ are well defined, i.e., the description of the maps does not depend on the choice of representative from an equivalence class.

2. Show that $f_+$ and $f_-$ are both injective.
3. Show that both $f_+$ and $f_-$ are both surjective.
4. Show that, for all $a$, $b$ in $\mathbb{Q}[X]/(X^2-2)\mathbb{Q}[X]$, $f_+(a+b) = f_+(a)+f_+(b)$ $f_+(a\cdot b) = f_+(a)\cdot f_+(b)$ $f_-(a+b) = f_-(a)+f_-(b)$ $f_-(a\cdot b) = f_-(a)\cdot f_-(b)$

Both maps give a way to associate the residue class ring $\mathbb{Q}[X]/(X^2-2)\mathbb{Q}[X]$ to $\mathbb{Q}+\mathbb{Q}\cdot\sqrt{2}$.

**Exercise 4.7.7.** Find the representative of degree less than 5 of the residue class of

$$(1+X)\cdot\left(1+X^3\right)\cdot\left(1+X^4\right)\cdot\left(1+X^5\right)$$

in $\mathbb{Z}/2\mathbb{Z}[X]/(X^5)\mathbb{Z}/2\mathbb{Z}[X]$.

**Exercise 4.7.8.** The polynomial $f$ in $\mathbb{Q}[X]$ satisfies the relation $\left(X^3+1\right)\cdot f + a\cdot\left(X^2+1\right) = X^3-1$ for some polynomial $a$ in $\mathbb{Q}[X]$. Determine the remainder upon division of $f$ by $X^2+1$.

**Exercise 4.7.9.** Let $R$ denote one of the fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/p\mathbb{Z}$ where $p$ is a prime.

Let $c, d$ be a pair of polynomials in $R[X]$ of degrees $m$ and $n$, respectively. Suppose that $c$ and $d$ are relatively prime.

Show that for any $a$ and $b$ in $R[X]$ there is exactly one polynomial in $R[X]$ of degree less than $m\cdot n$ that at the same time is equal to $a$ modulo $c$ and equal to $b$ modulo $d$. This is the *Chinese Remainder Theorem* for polynomials.

**Exercise 4.7.10.** Write an algorithm that, given two polynomials $c$ and $d$ that are relative prime and have degree $n$ and $m$, respectively, and two polynomials $a$ and $b$, computes the unique polynomial $f$ of degree less than $n\cdot m$ which is equal to 0 modulo both $c$ and $d$.

For existence and uniqueness of this polynomial we refer to .

### 4.7.3 Two special cases

**Exercise 4.7.11.** Determine the first 3 terms of the Taylor series around 0 of each of the following functions in $x$ by computation modulo $x^4$.

1. $\frac{1}{1+x}$
2. $\frac{1}{1+x+x^2}$
3. $\frac{1}{\cos(x)}$

**Exercise 4.7.12.** Determine the first 3 terms of the Taylor series around 0 of each of the following functions in $x$ by computation modulo $x^4$.

1. $\frac{1}{1-x}$
2. $\frac{1}{1-x+x^3}$

The smaller the length and the larger its minimal distance the more useful the code is. In the remainder of this section we will describe a method for constructing useful error-correcting codes with the help of polynomials.

In the world of digital communication, the binary number system is used a lot. In most applications and examples we confine ourselves to codes in vector spaces over $\mathbb{Z}/2\mathbb{Z}$. In these vector spaces, scalar multiplication is very simple: there are only two scalars, 0 and 1. These codes are known as *binary codes*.

## 4.7.4 Inverses and fields

**Exercise 4.7.13.** Consider the classes of $a = 1+X$ and $b = 1+2\cdot X$ in the ring $\mathbb{Z}/3\mathbb{Z}[X]/(X^2+1)\mathbb{Z}/3\mathbb{Z}[X]$.

Solve the following equation for $z$: $a\cdot z = b$.

**Exercise 4.7.14.** Consider the element $a = X+(X^3+X+1)\mathbb{Q}[X]$ in $\mathbb{Q}[X]/(X^2+X+1)\mathbb{Q}[X]$.

1. Show that $X^3+X+1$ is irreducible in $\mathbb{Q}[X]$. Conclude that $\mathbb{Q}[X]/(X^2+X+1)\mathbb{Q}[X]$ is a field.
2. Write $\frac{1}{a}$ as $p+q\cdot a+r\cdot a^2$ with $p,q,r \in \mathbb{Q}$.
3. Write $\frac{1}{a+2}$ as $p+q\cdot a+r\cdot a^2$ with $p,q,r \in \mathbb{Q}$.
4. Same question for $\frac{1}{a^2+a+1}$.

**Exercise 4.7.15.** Let $R$ be a field and $f$ and $d$ be polynomials in $R[X]$.

Prove or disprove:

1. If $f|d$, then $f$ is invertible in $R[X]/(d)R[X]$.
2. If the degree of $d$ is larger than 1 and $R = \mathbb{Z}$, then $R[X]/(d)R[X]$ is infinite.
3. If $a$ and $b$ are elements from $R[X]/(d)R[X]$ with $a\cdot b = 0$, but $a$ nor $b$ are equal to 0, then both $a$ and $b$ are not invertible.
4. If $a$, $b$, and $c$ are elements from $R[X]/(d)R[X]$ with $a\cdot b = a\cdot c$, then $b = c$.
5. If $a$, $b$, and $c$ are elements from $R[X]/(d)R[X]$ with $a\cdot b = a\cdot c$ and $a$ is invertible, then $b = c$.
6. If $a^4 = 0$ for some element $a$ in $R[X]/(d)R[X]$, then $1-a$ is invertible.

**Exercise 4.7.16.** Suppose that $R$ is a field. If $d \in R[X]$ is a polynomial of degree 1, then the map $R \to R[X]/(d)R[X], a \longmapsto a+(d)R[X]$ is bijective. Prove this.

**Exercise 4.7.17.** Let $K$ be one of the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/p\mathbb{Z}$ with $p$ prime.

1. Let $f, g \in K[X]$ with $f$ irreducible and let $a$ be the class of $X$ in $K[X]/(f)K[X]$. Show that $f|g$ if and only if $a$ is a zero of $g$, where we view $g$ as a polynomial with coefficients in $K[X]/(f)K[X]$.
2. Apply the divisibility criterion of the previous part to the polynomials $f = X^2+X+1$ and $g = X^6 - X^3 + 1$ over the ring $\mathbb{Z}/2\mathbb{Z}$ to find out whether $f$ divides $g$.

**Exercise 4.7.18.** Let $R$ be a ring. A polynomial in $R[X]$ is called *monic* if its leading coefficient equals 1.

1. If $d$ is a monic polynomial in $R[X]$ of positive degree $n$, then each residue class in $R[X]/(d)R[X]$ contains an element of degree smaller than $n$. Prove this.
2. Let $R$ be equal to $\mathbb{Z}/4\mathbb{Z}$ and $d$ the polynomial $2 \cdot X$. Verify that the class of $X$ in $R[X]/(d)R[X]$ does not contain an element of degree 0.

### 4.7.5 Finite fields

**Exercise 4.7.19.** Let $d = X^4+X+1 \in \mathbb{Z}/2\mathbb{Z}[X]$ and write $S = \mathbb{Z}/2\mathbb{Z}[X]/(d)\mathbb{Z}/2\mathbb{Z}[X]$.

1. Prove that $d$ is irreducible.
2. Determine the addition and multiplication table for the field $S$.
3. Find a subfield of $S$ of order 4. Here, a subfield of $S$ is a subset $Y$ such that inverses of nonzero members of $Y$, and products and sums of arbitrary members of $Y$, again belong to $Y$.

**Exercise 4.7.20.** Let $K = \mathbb{Z}/2\mathbb{Z}[X]/(d)\mathbb{Z}/2\mathbb{Z}[X]$, where $d = X^3+X+1$ and let $a$ be the class of $X$ modulo $d$.

1. Show that the polynomial $X^3+X+1$ in $\mathbb{Z}/2\mathbb{Z}[X]$ is irreducible and conclude that $K$ is a field with 8 elements.
2. Show that $((X^3)+X+1)|((X^7)+1)$ and that $K = \{0, 1, a, a^2, a^3, a^4, a^5, a^6\}$.
3. The element $a$ is a zero of $X^3+X+1$ (viewed as polynomial in $K$). Express all zeros as powers of $a$.
4. Find the zeros of $X^3+X^2+1$.

**Exercise 4.7.21.** Let $d = X^3+X+1 \in \mathbb{Z}/2\mathbb{Z}[X]$ and write $S = \mathbb{Z}/2\mathbb{Z}[X]/(d)\mathbb{Z}/2\mathbb{Z}[X]$.

1. Prove that $d$ is irreducible and conclude that $S$ is a field.
2. Show that each nonzero element of $S$ is a power of $X+(d)\mathbb{Z}/2\mathbb{Z}[X]$.

### 4.7.6 Error correcting codes

**Exercise 4.7.22.** Let $g$ be the polynomial $X^3+X^2+1$ over the field with 2 elements. Then $g$ is a divisor of $X^7-1$. Determine all codewords in the cyclic code generated by $g$.

**Exercise 4.7.23.** Suppose that $C$ is a code in $(\mathbb{Z}/2\mathbb{Z})^n$ that has minimal distance $d$ with $d \geq 2\cdot e+1$.

Show that $C$ contains at most $\frac{2^n}{\Sigma_{i=0}^{e}\left(\left(\begin{array}{c} n \\ i \end{array}\right)\right)}$ codewords.

# Chapter 5

# Permutations

In this chapter we study the bijections of a (finite) set $X$ into itself. The analysis of these bijections, called permutations, yields a tool to describe symmetry. This includes, for example, the description of the symmetries of a quadrangle as bijections of the 4 vertices. This description will appear in detail in [?], where we deal with groups of permutations.

## 5.1 Symmetric Groups

Let $X$ and $Y$ be sets. We recall some basic adjectives for maps from $X$ to $Y$.

**Definition 5.1.1.** A map $f\colon X \to Y$ is called

- *injective* if $f(x) = f(x')$ implies $x = x'$, for all $x, x' \in X$;
- *surjective* if, for every $y \in Y$, there exists an element $x \in X$ with $y = f(x)$;
- *bijective* if it is both injective and surjective.

**Example 5.1.2** (The exponential function)**.** The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = \mathrm{e}^x$ is injective.

Namely, if $f(x) = f(y)$, then $\mathrm{e}^x = \mathrm{e}^y$ and thus $\mathrm{e}^{x-y} = 1$. This is only possible for $x - y = 0$, hence $x = y$.

The function is not surjective, since $f(x) > 0$ for all $x$.

If we consider $f$ as a function of $\mathbb{R}$ to $(0,\infty)$, then $f$ is bijective. The inverse function then is the natural logarithm.

**Example 5.1.3** (A quadratic function)**.** Let $P$ denote the set of positive real numbers. The function $f = P \to P$ given by $f(x) = x^2$ is bijective.

If $P$ is replaced by $\mathbb{R}$, the real numbers, then $f$ is neither injective nor surjective.

If $P$ is replaced by $\mathbb{N}$, the natural numbers, then $f$ is injective but not surjective.

If $P$ is replaced by $\mathbb{C}$, the complex numbers, then $f$ is surjective but not injective.

We are mainly concerned with bijections of a finite set $X$ to itself. Often we work with the set $X$ of integers from 1 to $n$, thus $X = Y = \{1,...,n\}$. There is no loss of generality, since we will see soon that there is no essential difference in the naming of the elements.

The advantage of the natural numbers as names of the elements of $X$ is twofold:

- they have a natural ordering (this is convenient since we often intend to write the elements in a row);
- there is an infinite number of them (in contrast with, for example, the letters of the alphabet).

We will use no arithmetic properties of the natural numbers (as names of elements of $X$) apart from the ordering.

Let $X$ be a set. We introduce permutations of $X$ and describe multiplication of permutations as composition of maps.

**Definition 5.1.4.** • A bijection of $X$ to itself is also called a *permutation* of $X$. The set of all permutations of $X$ is denoted by $\mathrm{Sym}(X)$, the *symmetric group* on $X$.

- The product of two permutations $g,h$ in $\mathrm{Sym}(X)$ is defined as the composition $g \circ h$ of $g$ and $h$. Thus, for all $x \in X$, we have $g \cdot h(x) = g(h(x))$.
- If $X = \{1,...,n\}$, we also write $\mathrm{Sym}_n$ instead of $\mathrm{Sym}(X)$. Furthermore, a permutation $f$ of $X$ is often given by

$$[f(1), f(2), ..., f(n)] \tag{5.1}$$

called the *list notation* or as

$$\begin{pmatrix} 1 & 2 & ... & n \\ f(1) & f(2) & ... & f(n) \end{pmatrix} \tag{5.2}$$

called the *matrix notation* for $f$.

The product of two permutations in $\mathrm{Sym}(X)$ is again a permutation and hence an element of $\mathrm{Sym}(X)$. (Prove this!)

The identity map $\mathrm{id}\colon X \to X$ plays a special role: $g = g \cdot \mathrm{id}$ and $g = \mathrm{id} \cdot g$, for all $g$ in $\mathrm{Sym}(X)$. The inverse of $g$, denoted by $g^{-1}$, is again a permutation and satisfies $g^{-1} \cdot g = \mathrm{id}$ and $g \cdot g^{-1} = \mathrm{id}$. We call id the identity element for the product on $\mathrm{Sym}(X)$. We often use 1 or $e$ to denote the identity element. For every positive integer $m$, we denote by $g^m$ the product of $m$ factors $g$. Instead of $\left(g^{-1}\right)^m$ we also write $g^{-m}$.

We call $\mathrm{Sym}_n$ the *symmetric group* of degree $n$. The symmetric group is an instance of the structure group that will be discussed in [?].

**Example 5.1.5.** Let $g$ and $h$ be the permutations of $\{1,...,4\}$ with $g(1)=2$, $g(2)=3$, $g(3)=1$, $g(4)=4$, and $h(1)=1$, $h(2)=3$, $h(3)=4$, $h(4)=2$. So $g=[2,3,1,4]$ and $h=[1,3,4,2]$.

Then $g\cdot h$ is the permutation with $g\cdot h(1)=g(1)=2$, $g\cdot h(2)=g(3)=1$, $g\cdot h(3)=g(4)=4$, and $g\cdot h(4)=g(2)=3$, so $g\cdot h=[2,1,4,3]$.

Similarly, $h\cdot g$ is the permutation with $h\cdot g(1)=h(2)=3$, $h\cdot g(2)=h(3)=4$, $h\cdot g(3)=h(1)=1$, and $h\cdot g(4)=h(4)=2$, so $h\cdot g=[3,4,1,2]$.

In particular, $g\cdot h$ and $h\cdot g$ are not the same. The official terminology is that $g$ and $h$ do not commute.

The inverse of $g$ is the map that sends 1 to 3, 2 to 1, 3 to 2, and 4 to 4, so $g^{-1}=[3,1,2,4]$.

**Remark 5.1.6.** Sometimes the product $g\cdot h$ of two permutations is defined the other way around: as $h\circ g$. In other words, the product is the right composition of functions instead of left composition.

Right composition is convenient when writing mappings at the right-hand side of their arguments: for $x\in X$, the element $(x)\,g\cdot h$ is then as well the image under $g\cdot h$ of $x$ as the image under $h$ of the image under $g$ of $x$. In formula: $(x)\,g\cdot h=((x)\,g)\,h$.

Right composition is standard in the computer algebra packages GAP and Magma. One should be aware of this fact!

A permutation of $\{1,...,n\}$ can be described in matrix notation by a 2 by $n$ matrix with the numbers $1,...,n$ in the first row and the images of $1,2,...,n$ (in that order) in the second row. Since there are $n!$ possibilities to fill the second row, the following theorem holds.

**Theorem 5.1.7.** $\mathrm{Sym}_n$ *has exactly* $n!$ *elements.*

**Example 5.1.8.** $\mathrm{Sym}_3$ has the following 6 elements:

$$\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix},\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix},\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix},\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix},\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix},\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix} \tag{5.3}$$

Instead of the conventional matrix notation, we also write permutations as lists. In the so-called list notation we leave out the first row, since that row is always the same. Here are the 6 permutations again in list notation:

$$[1,2,3]\,,[1,3,2]\,,[2,1,3]\,,[2,3,1]\,,[3,1,2]\,,[3,2,1] \tag{5.4}$$

The first row of the 2 by $n$ matrix describing a permutation in $\mathrm{Sym}_n$, is always $1,2,...,n$ and hence yields no essential information. Nevertheless, the matrix notation is useful for calculating products and inverses.

- Product: To calculate $g{\cdot}h$ for two permutations $g,h$ in $\mathrm{Sym}_n$, we first look up, for each $i \in \{1,...,n\}$, the value $h(i)$, then we look for this value in the first row of the $g$ matrix; below this entry you find $g{\cdot}h(i)$.
- Inverse: If $g$ is written as the 2 by $n$ matrix $M$, then the inverse of $g$ is described by the matrix obtained from $M$ by interchanging the two rows and sorting the columns in such a way that the first row is again $1,2,...,n$.

**Definition 5.1.9** (Order of a Permutation)**.** The order of a permutation $g$ is the smallest positive integer $m$ such that $g^m = \mathrm{e}$.

**Example 5.1.10.** • The order of the identity is 1.

- The order of the permutation $[2,1,3]$ (in list notation) in $\mathrm{Sym}_3$ is 2.
- The order of the permutation $g = [2,3,4,1]$ (in list notation) in $\mathrm{Sym}_4$ is $4: g^2 = [3,4,1,2], g^3 = [4,1,2,3], g^4 = e$.

**Remark 5.1.11.** Of course we must justify that the notion order makes sense. If $g$ is a permutation in $\mathrm{Sym}_n$, then the permutations $g, g^2, g^3, ...$ can not all be distinct, because there are only finitely many permutations in $\mathrm{Sym}_n$ ($n!$ to be precise). So there must exist positive numbers $r < s$ such that $g^r = g^s$. Since $g$ is a bijection, we find $g^{s-r} = 1$. So there exist positive numbers $m$ with $g^m = 1$, and in particular a smallest such number. Therefore each permutation $g$ has a well-defined *order*: the smallest positive integer $m$ with $g^m = 1$.

## 5.2 Cycles

Let $g$ be a permutation of $\mathrm{Sym}_n$. We distinguish between the points which are moved and the points which are fixed by $g$.

**Definition 5.2.1.** • The *fixed points* of $g$ in $X$ are the elements of $x$ of $X$ for which $g(x) = x$ holds. The set of all fixed points is denoted by $\mathrm{permutation1.fix}(g,X)$.

- The *support* of $g$ is the complement in $X$ of $\mathrm{permutation1.fix}(g,X)$. Notation: $\mathrm{Support}(g)$.
- If $g(x) = x$, i.e., $x$ is a fixed point of $g$, then we say that $x$ is *fixed* by $g$.
- If $g(x) \neq x$, i.e., $x$ is in the support of $g$, we say that $g$ *moves* $x$.

**Remark 5.2.2.** • The fixed points of a permutation $g$ form the set $\mathrm{permutation1.fix}(g) = \{x \in X | g(x) = x\}$. The notation refers to the verb ‘to fix’.

- The support of $g$ equals $\mathrm{Support}(g) = \{x \in X | g(x) \neq x\}$. Support refers to the subset where ‘really something happens’.

Cycles are elements in $\mathrm{Sym}(n)$ of special importance.

**Definition 5.2.3.** Let $g \in \mathrm{Sym}(n)$ be a permutation with $\mathrm{Support}(g) = \{a_1, ..., a_m\}$, where the $a_i$ are pairwise distinct. We say $g$ is an $m$-cycle if $g(a_i) = g(a_{i+1})$ for all $i \in \{1, ..., m-1\}$ and $g(a_m) = a_1$. For such a cycle $g$ we also use the cycle notation $(a_1, ..., a_m)$.

2-cycles are called *transpositions*.

**Example 5.2.4.** • In $\mathrm{Sym}(3)$ all elements are cycles. The identity element e is a 0- or 1-cycle, the other elements are 2 - or 3-cycles: $(1,2), (1,3), (2,3), (1,2,3)$ and $(1,3,2)$. No two of these 5 cycles are disjoint.

- In $\mathrm{Sym}(4)$, the element (in list notation) $[2,1,4,3]$ is not a cycle, but it is the product $(1,2) \cdot (3,4)$ of the transpositions $(1,2)$ and $(3,4)$.

**Remark 5.2.5.** • The cycle notation of a permutation $g$ does not tell us in which $\mathrm{Sym}_n$ we are working in. This is in contrast to the list and matrix notation. So $(1,2)$ might belong to $\mathrm{Sym}_2$ just as well as to $\mathrm{Sym}_3$. This yields no real confusion because of the natural identification of $\mathrm{Sym}_{n-1}$ with the part of $\mathrm{Sym}_n$ consisting of all permutations fixing $n$ : $\mathrm{Sym}_{n-1} = \{g \in \mathrm{Sym}_n | g(n) = n\}$.

- The composition of permutations in $\mathrm{Sym}_n$ (where $n > 2$ ) is not commutative. This means that the products $g \cdot h$ and $h \cdot g$ are not always the same. If $g \cdot h = h \cdot g$, then we say that $g$ and $h$ commute. Two cycles $c$ and $d$ are called disjoint if the intersection of their supports is empty. Two disjoint cycles always commute. (Prove this!) A cycle $(a_1, a_2, ..., a_n)$ also commutes with its inverse $(a_n, ..., a_2, a_1)$

Every element in $\mathrm{Sym}_n$ is a product of cycles. Even more is true:

**Theorem 5.2.6** (Disjoint Cycle Decomposition)**.** *Every permutation in* $\mathrm{Sym}_n$ *is a product of disjoint cycles. This product is unique up to rearrangement of the factors.*

*Proof.* First we show that every $g$ in $\mathrm{Sym}_n$ can be written as a product of disjoint cycles (the existence). Then we prove the uniqueness of this product. Both parts are proved by induction.

**Every permutation is a product of disjoint cycles.**

We use induction with respect to the number of elements in the support of the permutation $g$. If the support of $g$ is empty, then $g$ is 1, the identity element, a 0-cycle. We regard this as an empty product of cycles.

Now assume that for some number $k > 0$ any element $g$ with $|\mathrm{Support}(g)| \leq k$ can be written as a product of disjoint cycles. Let $g$ be an element with $k$ elements in its support. Fix an element $x$ in $\mathrm{Support}(g)$. We try to 'split off' a cycle containing $x$. We set $a_0 = x$ and $a_i = g(a_{i-1})$ for $i > 0$. Let $m$ denote the smallest positive integer for which $a_m = x$ and consider the cycle $c = (a_1, a_2, ..., a_m)$. Its support is a subset of $\mathrm{Support}(g)$. So the permutation $h = g \cdot c^{-1}$ fixes all points that are fixed by $g$ as well as the points $a_i$, with $i < m+1$. Indeed, $h(a_i) = g \cdot c^{-1}(a_i) = g(a_{i-1}) = a_i$. This implies that the support of $h$ is contained in $\mathrm{Support}(g) \setminus$

$\{a_1,a_2,...,a_m\}$. By the induction assumption we may write $h$ as a product of disjoint cycles $c_1,c_2,...,c_k$. The support of these cycles is contained in Support$(h)$ and therefore disjoint from $\{a_1,a_2,...,a_m\}$. But then $g = h\cdot c = c_1\cdot c_2\cdot...\cdot c_k\cdot c$ is a product of disjoint cycles. By induction we have finished the first part of the proof.

**The disjoint product decomposition is unique up to permutation of the cycles.**

Assume that $g$ is the product of the disjoint cycles $c_1,c_2,...,c_k$ and at the same time of the disjoint cycles $d_1,d_2,...,d_l$, all of length at least 2. We prove the uniqueness by induction on $k$. The case $k=0$ is trivial. So assume that $k>0$. Then Support$(g)$ is not empty and we can find an element $x$ in Support$(g)$. As $x$ is not fixed by $g$, there exist cycles $c_i$ and $d_j$ which do not fix $x$. Without loss of generality we may suppose that $x\in$ Support$(c_1)$ and $x\in$ Support$(d_1)$. For every $m\in\mathbb{N}$, we have $(c_1)^m(x) = g^m(x) = (d_1)^m(x)$. In particular $c_1 = d_1$. But then also $c_2\cdot...\cdot c_k = (c_1)^{-1}\cdot g = (d_1)^{-1}\cdot g = d_2\cdot...\cdot d_l$. The induction hypothesis yields that $k-1 = l-1$ and, possibly after renumbering of the indices, $c_i = d_i$ for all $i$ from 1 to $k$. This proves the proposition.

□

**Example 5.2.7.** The proof actually shows how to find the disjoint cycles decomposition of a permutation. Consider the permutation (in list notation) $g = [8,4,1,6,7,2,5,3]$ in $\mathrm{Sym}_8$. The following steps lead to the disjoint cycles decomposition.

- Choose an element in the support of $g$, for example 1. Now construct the cycle
$$\left(1,g(1),g^2(1),...\right).$$
In this case this cycle is $(1,8,3)$. On $\{1,3,8\}$ the permutation $g$ and the cycle $(1,8,3)$ coincide.

- Next, choose an element in the support of $g$, but outside $\{1,3,8\}$, for example 2. Construct the cycle $\left(2,g(2),g^2(2),...\right)$. In the case at hand, this cycle is $(2,4,6)$. Then $g$ and $(1,8,3)\cdot(2,4,6)$ coincide on $\{1,2,3,4,6,8\}$.

- Choose an element in the support of $g$ but outside $\{1,2,3,4,6,8\}$, say 5. Construct the cycle $\left(5,g(5),g^2(5),...\right)$, i.e., $(5,7)$. Then $g$ and $(1,8,3)\cdot(2,4,6)\cdot(5,7)$ coincide on $\{1,2,3,4,5,6,7,8\}$ and we are done.

Note that the three cycles $(1,8,3),(2,4,6),(5,7)$ commute, so that $g$ can also be written as $(5,7)\cdot(1,8,3)\cdot(2,4,6)$ or as $(2,4,6)\cdot(5,7)\cdot(1,8,3)$, etc.

If a permutation is written as a product of disjoint cycles, we say that it is given in disjoint cycles form or disjoint cycles notation. 1-cycles are usually left out in this notation.

The above proposition justifies the following definition:

**Definition 5.2.8.** The cycle structure of a permutation $g$ is the (unordered) sequence of the cycle lengths in an expression of $g$ as a product of disjoint cycles.

So, rephrasing the above Disjoint Cycle Decomposition, we can say that every permutation has a unique cycle structure.

The choice $X = \{1,\ldots,n\}$ fixes the set $X$ under consideration. Suppose someone chooses a different numbering of the elements in $X$. How do we compare two permutations of $X$ with respect to these two numberings?

There is a permutation $h$ of $X$, which changes our numbering in the new one; so $h$ can be used as a change of names. We describe a given permutation $g$ with respect to the new numbering as follows. First, we apply the 'back-transformation' $h^{-1}$ to our own numbering, then we apply $g$, and, finally, we use $h$ again to translate back to the other numbering. As a formula, with respect to the new numbering, the transformation $g$ 'reads' $h\cdot g\cdot h^{-1}$. The map $g \longmapsto h\cdot g\cdot h^{-1}$ is called conjugation with $h$. The cycle decomposition of $g$ yields a nice way to calculate the effect of conjugation with a permutation $h$:

**Lemma 5.2.9.** *Let $h$ be a permutation in* $\mathrm{Sym}_n$.

- *For every cycle $(a_1,\ldots,a_m)$ in $\mathrm{Sym}_n$ we have $h\cdot(a_1,\ldots,a_m)\cdot h^{-1} = (h(a_1),\ldots,h(a_m))$.*
- *If $g_1$, ..., $g_k$ are in $\mathrm{Sym}_n$, then $h\cdot g_1\cdot\ldots\cdot g_k\cdot h^{-1} = h\cdot g_1\cdot h^{-1}\cdot\ldots\cdot h\cdot g_k\cdot h^{-1}$. In particular, if the $g_i$ are (disjoint) cycles, then $h\cdot g_1\cdot\ldots\cdot g_k\cdot h^{-1} = h\cdot g_1\cdot h^{-1}\cdot\ldots\cdot h\cdot g_k\cdot h^{-1}$ is the product of the (disjoint) cycles $h\cdot g_i\cdot h^{-1}$.*

*Proof.* The proof of both items in the lemma are easy verifications if you take the following approach.

**Conjugation of a cycle.**

Let $h$ be the cycle $(a_1,\ldots,a_m)$. We compute $h\cdot g\cdot h^{-1}(x)$ by distinguishing two cases.

- If $x = h(a_i)$ for some $i$ in $\{1,\ldots,m\}$, then
$$h\cdot g\cdot h^{-1}(x) = h\cdot g\cdot h^{-1}(h(a_i)) = h\cdot g(a_i) = h(a_{i+1}) \tag{5.5}$$
where we set $a_{m+1} = a_1$.
- If $x$ is not equal to $h(a_i)$ for all $i$ in $\{1,\ldots,m\}$. Then $h(x)$ is not in the support of $g$ and consequently
$$h\cdot g\cdot h^{-1}(x) = h\cdot g\left(h^{-1}(x)\right) = h\left(h^{-1}(x)\right) = x \tag{5.6}$$

We conclude that

$$h\cdot(a_1,\ldots,a_m)\cdot h^{-1} = (h(a_1),\ldots,h(a_m)) \tag{5.7}$$

**Conjugation of a product of permutations.**

The second item of the lemma follows once you realize that in the product $h\cdot g_1\cdot h^{-1}\cdot\ldots\cdot h\cdot g_m\cdot h^{-1}$ the pairs $h^{-1}\cdot h$ cancel, so that

$$h\cdot g_1\cdot h^{-1}\cdot\ldots\cdot h\cdot g_m\cdot h^{-1} = h\cdot g_1\cdot\ldots\cdot g_m\cdot h^{-1} \tag{5.8}$$

is what remains. In particular, if the $g_i$ are disjoint cycles, then

$$h\cdot g_1\cdot\ldots\cdot g_m\cdot h^{-1} \tag{5.9}$$

is also a product of disjoint cycles of the form $h\cdot g_i\cdot h^{-1}$.

□

**Example 5.2.10.** Consider an equilateral triangle with vertices $A$, $B$ and $C$. The reflection in the line $L$ through $B$ and the midpoint of the edge on $A$ and $C$ induces a permutation of the three vertices: $A \longmapsto C$, $B \longmapsto B$, $C \longmapsto A$. If we name the three vertices $1,2,3$ for $A,B,C$, respectively, then we can describe the reflection by the permutation $(1,3)$. The rotation through 120 degrees around the midpoint of the triangle is also a permutation of the three vertices. This rotation is described by the permutation $(1,3,2)$. If we choose other names for the vertices, for example $1,3,2$ for $A,B,C$, then the description of the reflection and the rotation change. The reflection is then for example described by $(1,2)$ and the rotation by $(1,2,3)$. This renumbering may be achieved by the permutation $k = (2,3)$. Indeed, we see that $k\cdot(1,2)\cdot k^{-1} = (1,2)$ and $k\cdot(1,3,2)\cdot k^{-1} = (1,2,3)$. Conjugation is similar to basis transformation in linear algebra.

It follows from the Conjugation Formulas that any two conjugate permutations (one permutation can be obtained from the other by conjugation) have the same cycle structure. The converse also holds.

**Theorem 5.2.11.** *Two elements $g$ and $h$ in $\mathrm{Sym}_n$ have the same cycle type if and only if there exists a permutation $k$ in $\mathrm{Sym}_n$ with $g = k\cdot h\cdot k^{-1}$.*

*Proof.*

**If.**

This implication follows directly from the Conjugation Formulas.

**Only if.**

We write both $g$ and $h$ as a product of disjoint cycles $s_i$ and $t_j$, respectively, all of length at least 2. Since $g$ and $h$ have the same cycle structure, we can write $g = s_1\cdot s_2\cdot\ldots\cdot s_k$ and $h = t_1\cdot t_2\cdot\ldots\cdot t_k$ in such a way that $s_i$ and $t_i$ have equal length for all $i$. Suppose $s_i = \left(s_{i,1}, s_{i,2}, \ldots, s_{i,k_i}\right)$

and $t_i = \left(t_{i,1}, t_{i,2}, \ldots, t_{i,k_i}\right)$. Denote by $u$ a permutation with $u\left(s_{i,j}\right) = t_{i,j}$ for all $i$ from 1 to $k$ and $j$ from 1 to $k_i$. This is possible since the supports of the $s_i$ are disjoint as well as the supports of the $t_i$. (Notice that there may be more than one permutation $u$ satisfying these requirements.) The Conjugation Formulas yield that $u \cdot g \cdot u^{-1} = h$.

□

**Example 5.2.12.** In $\mathrm{Sym}_4$ the permutations (in list notation) $g = [2,1,4,3]$ and $h = [3,4,1,2]$ are conjugate, since both have the cycle structure $2,2$ : $g = (1,2) \cdot (3,4)$ and $h = (1,3) \cdot (2,4)$. A permutation $k$ such that $k \cdot g \cdot k^{-1} = h$ is $k = [1,3,2,4]$. In disjoint cycles notation this is $(2,3)$.

Transpositions play an important role among permutations.

**Theorem 5.2.13.** *Let $n \geq 2$. Every element of $\mathrm{Sym}_n$ is the product of (not necessarily disjoint) transpositions.*

*Proof.* Since every permutation in $\mathrm{Sym}_n$ can be written as a product of disjoint cycles , it suffices to show that every cycle is a product of 2-cycles.

Now every $m$-cycle $(a_1, \ldots, a_m)$, is equal to the product

$$\prod_{i=1}^{m-1} \left(\left(a_i, a_{i+1}\right)\right) \tag{5.10}$$

and the proof is complete.

□

**Example 5.2.14.** Let $a = \{a_1, \ldots, a_n\}$ be a list of $n$ integers. The algorithm ‘Bubble sort’ ranks the elements of $a$ with respect to increasing value. The algorithm works as follows. Take an element $a_i$ of the list, compare it with the predecessor $a_{i-1}$, and switch both elements if $a_i$ is less than $a_{i-1}$. First, $i$ decreases from $n$ to 2 . Then the least element is in the first position of the list. Now one repeats the procedure, but only with $i$ decreasing from $n$ to 3. By this time the second least element is in the second position. And so forth. Finally, the algorithm yields a sorted list.

The switch of two elements of the list is a transposition $(i-1, i)$ applied to the positions $i-1$ and $i$ of the two elements in the list.

If $a$ is filled with the numbers from 1 to $n$, it is the list form of a permutation. So we can consider $a$ to be a permutation.

Applying the transpositions $(i-1, i)$ to the positions in the sequence $a$ boils down to a multiplication of the permutation $a$ with the transposition $(i-1, i)$. As Bubble Sort sorts $a$ to the standard list $[1,...,n]$, which represents the identity map, we find that there are transpositions $t_1$, ..., $t_k$ with

$$a{\cdot}t_1{\cdot}...{\cdot}t_k = 1 \tag{5.11}$$

which implies that

$$a = t_k{\cdot}...{\cdot}t_1 \tag{5.12}$$

Hence we may write each permutation as a product of transpositions, in particular even of transpositions of the form $(i-1,i)$. This yields again a proof of the theorem.

## 5.3 Alternating groups

From the theory in Section 5.2, every permutation can be written as a product of transpositions. To be able to distinguish between products of even and odd length, we need the following result.

**Theorem 5.3.1.** *If a permutation is written in two ways as a product of transpositions, then both products have even length or both products have odd length.*

*Proof.* Suppose that the permutation $g$ can be written both as the product of transpositions $c_1{\cdot}...{\cdot}c_k$ with $k$ even, as the product of transpositions $d_1{\cdot}...{\cdot}d_m$ with $m$ odd. Then $1 = c_1{\cdot}...{\cdot}c_k{\cdot}{d_m}^{-1}{\cdot}...{\cdot}{d_1}^{-1}$ expresses the identity as the product of an odd number of transpositions. We will show that this is impossible.

So assume that the identity element 1 is a product of an odd number of transpositions. We choose such a product $1 = t_1{\cdot}...{\cdot}t_m$ with $m$ minimal subject to being odd. It is obvious that $m > 0$.

**We may assume that $t_1 = (1,2)$.**

If $t_1 = (i,j)$, we can conjugate left-hand side and right-hand side by $(1,i)\cdot(2,j)$.

**We may assume that there is some $l > 0$ with $t_1$ up to $t_l$ all moving $1$, that is, $t_i = (1,a_i)$ for all $i \leq l$, and that $t_{l+1}$ up to $t_m$ all fix $1$.**

Applying the formulas $(a,b)\cdot(1,c) = (1,c)\cdot(a,b)$ and $(a,b)\cdot(1,b) = (1,a)\cdot(a,b)$, where $1,a,b$ and $c$ are different numbers in $\{1,...,n\}$, we can shift all transpositions which contain 1 to the front without violating the minimality of $m$.

**There is an index $i$ with $i \in \{2,...,l\}$ such that $t_i = t_1$.**

We must have $t_1\cdot t_2{\cdot}...{\cdot}t_l\,(1) = 1$. Therefore $2 = t_1\,(1)$ lies is in the support of $t_2{\cdot}...{\cdot}t_l$, and at least one of the $a_i$ with $i > 1$ is equal to 2.

**Final contradiction.**

We have $t_i = t_1 = {t_1}^{-1}$, and, because of minimality of $m$, also $t_2 \neq t_1$. Hence, $1 = t_1 \cdot \ldots \cdot t_m = t_1 \cdot (t_2 \cdot \ldots \cdot t_{i-1}) \cdot (t_1)^{-1} \cdot t_{i+1} \cdot \ldots \cdot t_m = s_2 \cdot \ldots \cdot s_{i-1} \cdot t_{i+1} \cdot \ldots \cdot t_m$, where $s_j = t_1 \cdot t_j \cdot (t_1)^{-1}$ for $j \in \{2, \ldots, i-1\}$ is also a transposition. We have written 1 as a product of $m-2$ transpositions. This contradicts the minimality of $m$.

□

In other words, no permutation can be written both as a product of transpositions of even length and as such a product of odd length. So if one product involves an even (odd) number of factors, then all products involve an even (odd) number of factors.

We saw that no permutation can be written both as a product of transpositions of even length and as such a product of odd length. So if one product involves an even (odd) number of factors, then all products involve an even (odd) number of factors. This justifies the following definition.

**Definition 5.3.2.** Let $g$ be an element of $\mathrm{Sym}_n$. The sign (signum) of $g$, denoted by $\mathrm{sign}(g)$, is defined as

- 1 if $g$ can be written as a product of an even number of 2-cycles, and
- $-1$ if $g$ can be written as a product of an odd number of 2-cycles.

We say that $g$ is even if $\mathrm{sign}(g) = 1$ and odd if $\mathrm{sign}(g) = -1$.

The sign is multiplicative.

**Theorem 5.3.3.** *For all permutations $g$, $h$ in $\mathrm{Sym}_n$, we have $\mathrm{sign}(g \cdot h) = \mathrm{sign}(g) \cdot \mathrm{sign}(h)$.*

*Proof.* Let $g$ and $h$ be elements of $\mathrm{Sym}_n$.

- If one of the permutations is even and the other is odd, then $g \cdot h$ can obviously be written as the product of an odd number of transpositions and is therefore odd.
- If $g$ and $h$ are both even or both odd, then the product $g \cdot h$ can be written as the product of an even number of transpositions so that $g \cdot h$ is even.

□

**Remark 5.3.4.** • The sign of a permutation and its inverse are the same. There are various ways to see this, one of which is based on the multiplicative property of the sign. Since $g \cdot g^{-1} = \mathrm{id}$, we find

$$\mathrm{sign}(g)\cdot\mathrm{sign}(g^{-1}) = \mathrm{sign}(g\cdot g^{-1}) = \mathrm{sign}(1) = 1 \tag{5.13}$$

so that

$$\mathrm{sign}(g) = \mathrm{sign}(g^{-1}) \tag{5.14}$$

- Every $m$-cycle $(a_1, ..., a_m)$ can be written as the product of $m-1$ transpositions: $(a_1, ..., a_m) = (a_1, a_2)\cdot(a_2, a_3)\cdot ...\cdot(a_{m-1}, a_m)$. Since transpositions are odd, the multiplicativity of the sign implies that the sign of an $m$ - cycle is $-1^{m-1}$, i.e., a cycle of even length is odd and a cycle of odd length is even.

We also say that the sign map is a multiplicative map from $\mathrm{Sym}_n$ to $\{1, -1\}$. (The notion morphism explores this view further in a general context.)

The previous theorem implies the following way of determining the sign.

**Corollary 5.3.5.** *If a permutation g is written as a product of cycles, then* $\mathrm{sign}(g) = (-1)^w$*, where w is the number of cycles of even length.*

*Proof.* Since the sign map is a multiplicative mapping, the sign of $g$ is the product of the signs of every factor. Now a cycle of odd length has sign 1, so we only need to count the number of cycles of even length.

□

Permutations and the sign of permutations occur in the explicit expression for determinants. If $A$ is an $n$ by $n$ matrix with entries $A_{ij}$ then the determinant $\det(A)$ is the sum over all $n!$ permutations $g$ in $\mathrm{Sym}_n$ of the products $\mathrm{sign}(g)\cdot A_{1g(1)}\cdot A_{2g(2)}\cdot ...\cdot A_{ng(n)}$, i.e.,

$$\det(A) = \sum_{g\in\mathrm{Sym}_n} \mathrm{sign}(g)\cdot A_{1g(1)}\cdot A_{2g(2)}\cdot ...\cdot A_{ng(n)} \tag{5.15}$$

In the case of a 2 by 2 matrix $A$ we find two terms:

- $A_{11}\cdot A_{22}$ corresponding to the identity permutation, which has sign 1, and
- $-A_{12}\cdot A_{21}$ corresponding to the permutation $(1,2)$, which has sign $-1$.

Summing yields the familiar formula $\det(A) = A_{11}\cdot A_{22} - A_{12}\cdot A_{21}$.

It is still easy to write down the explicit 6 term formula for a 3 by 3 determinant, but since $n!$ grows so rapidly, the formula becomes quite impractical for computations if $n$ gets large. For computations of determinants more practical methods are available derived from the above formula. Such methods are discussed in courses on linear algebra.

The fact that the sign map is multiplicative implies that products and inverses of even permutations are even. This gives rise to the following definition.

**Definition 5.3.6.** By $\mathrm{Alt}_n$ we denote the set of even permutations in $\mathrm{Sym}_n$. We call $\mathrm{Alt}_n$ the alternating group on $n$ letters.

**Example 5.3.7.** For $n = 3$, the even permutations are (in cycle notation): $\mathrm{e}, (2,3,1)$ and $(3,1,2)$.

**Remark 5.3.8.** The set of even permutations is closed with respect to taking products and inverse elements.

There are just as many even permutations as there are odd permutations in $\mathrm{Sym}_n$.

There are just as many even as odd permutations in $\mathrm{Sym}_n$.

**Theorem 5.3.9.** *For $n > 1$, the alternating group $\mathrm{Alt}_n$ contains precisely $\frac{n!}{2}$ elements.*

*Proof.* An element $g$ of $\mathrm{Sym}_n$ is even (respectively, odd), if and only if the product $g\cdot(1,2)$ is odd (respectively, even). Hence the map $g \longmapsto g\cdot(1,2)$ defines a bijection between the even and the odd elements of $\mathrm{Sym}_n$. But then precisely half of the $n!$ elements of $\mathrm{Sym}_n$ are even.

□

3-cycles are the smallest nontrivial even cycles. They are the building blocks for even permutations:

**Theorem 5.3.10.** *Every even permutation is a product of* 3*-cycles.*

*Proof.* Every element of $\mathrm{Alt}(X)$ is a product of an even number of transpositions. Hence it suffices to prove that each product of two transpositions, different from the identity element, can be written as a product of 3-cycles.

Let $(a,b)$ and $(c,d)$ be two different transpositions.

- If $a,b,c$ and $d$ are pairwise distinct, then

$$(a,b)\cdot(c,d) = (a,b)\cdot(b,c)\cdot(b,c)\cdot(c,d) = (a,b,c)\cdot(b,c,d) \qquad (5.16)$$

- Without loss of generality we are left with the case where $a,b,d$ are pairwise distinct and $b = c$. But then

$$(a,b)\cdot(b,d) = (a,b,d) \qquad (5.17)$$

This proves the theorem.

□

## 5.4 Exercises

### 5.4.1 Symmetric Groups

**Exercise 5.4.1.** In $\mathrm{Sym}_6$ we choose the permutations $a = (1,2,3)\,, b = (2,3,4,5,6)$ and $c = (1,4,6,3)$.

- Calculate $a^{-1}, a\cdot b\cdot c, a\cdot b\cdot c^2, c^{-1}\cdot b$ and $(a\cdot c\cdot b)^{-1}$.
- Calculate the sign of each of the above permutations.

### 5.4.2 Cycles

**Exercise 5.4.2.** Let $g$ be a permutation in $\mathrm{Sym}_n$. Show that if $i \in \mathrm{Support}(g)$, then $g(i) \in \mathrm{Support}(g)$.

**Exercise 5.4.3.** How many elements of $\mathrm{Sym}_5$ have the cycle structure $2,3$?

**Exercise 5.4.4.** Let $g$ be the permutation $(1,2,3)\cdot(2,3,4)\cdot(3,4,5)\cdot(4,5,6)\cdot(5,6,7)\cdot(6,7,8)\cdot(7,8,9)$ in $\mathrm{Sym}_6$.

- Write $g$ as a product of disjoint cycles.
- Calculate the fixed points of $g$.
- Write $g^{-1}$ as a product of disjoint cycles.
- Is $g$ even?

**Exercise 5.4.5.** Let $n$ be an integer greater than 2. Suppose $a$ has an inverse modulo $n$. Label the elements of the set $S$ of residue classes $1,2,\ldots,n-1$ in $\frac{\mathbb{Z}}{n\cdot\mathbb{Z}}$ in the obvious way with the integers $1,2,\ldots,n-1$.

- Show that multiplication by $a$ defines a permutation $p$ of $1,2,\ldots,n-1$. For $a = 2$ and $n = 9$ write the corresponding permutation as a product of disjoint cycles. Can you read off the smallest positive integer $m$ such that $a^m = \mathrm{rem}(1,n)$?
- Suppose $p$ is written as a product of disjoint cycles. Prove that the cycles fall into two categories: One consisting of cycles all of whose entries are invertible mod $n$ and one consisting of cycles all of whose entries are not invertible modulo $n$.

**Exercise 5.4.6.** Let $R$ be the residue class ring $\mathbb{Z}/3\mathbb{Z}[X]/(X^2+1)\mathbb{Z}/3\mathbb{Z}[X]$ and let $a$ be the class of $X$. Then $a$ is an invertible element of $R$. Show that multiplication by $a$ produces a permutation of these elements. Write this permutation as a product of disjoint cycles. What is its cycle structure?

**Exercise 5.4.7.** • If the permutations $g$ and $h$ in $\mathrm{Sym}_n$ have disjoint supports, then $g$ and $h$ commute, i.e., $g \cdot h = h \cdot g$. Prove this.

- Suppose that the permutations $g$ and $h$ in $\mathrm{Sym}_n$ commute. Prove that $(g \cdot h)^m = g^m \cdot h^m$ for all positive numbers $m$.
- Suppose that the permutations $g$ and $h$ in $\mathrm{Sym}_n$ have disjoint supports. Prove that $(g \cdot h)^m = 1$ for some positive number $m$ implies that $g^m = 1$ and $h^m = 1$.
- If the permutation has order $t$ and if $g^m$ = id for some positive number $m$, show that $t$ divides $m$. In particular, if $c$ is a $t$-cycle and $c^m$ = id for some positive number $m$, then $m$ is divisible by $t$.
- Prove that if the permutation $g$ has cycle structure $m_1, ..., m_r$, then the order of $g$ equals $\mathrm{lcm}\,(m_1, ..., m_r)$.

**Exercise 5.4.8.** • Prove that for $n > 4$ every permutation in $\mathrm{Sym}_n$ can be written as a product of 4-cycles.

- Prove that for $n > 5$ every even permutation can be written as a product of 5-cycles.

**Exercise 5.4.9.** Let $a = (1,2,3)(4,7,9)(5,6)$. Determine an element $b$ in $\mathrm{Sym}_9$ such that $b \cdot a \cdot b^{-1} = (9,8,7)(6,5,4)(3,2)$.

**Exercise 5.4.10.** Let $g$ be an element of $\mathrm{Sym}_n$ with $n > 2$.

- If $g$ commutes with the transposition $(i,j)$, where $i \neq j$, then $g(i) \in \{i,j\}$. Prove this.
- Show that $g \cdot i = i$, whenever $g$ commutes with the transpositions $(i,j)$ and $(i,k)$, where $i,j,k$ are mutually distinct.
- Prove that the identity map is the only permutation in $\mathrm{Sym}_n$ that commutes with all elements of $\mathrm{Sym}_n$.

**Exercise 5.4.11.** Write all elements of $\mathrm{Alt}_4$. as products of disjoint cycles.

**Exercise 5.4.12.** Let $a = (1,2)$ and $b = (2,...,n)$.

- Calculate $b \cdot a \cdot b^{-1}$.
- Calculate $b^k \cdot a \cdot b^{-k}$, for $k \in \mathbb{N}$.
- Prove that every element of $\mathrm{Sym}_n$ can be written as a product of elements from $\{a, b, b^{-1}\}$.

### 5.4.3 Alternating groups

**Exercise 5.4.13.** For $g$ in $\mathrm{Sym}_n$, we define a matrix $M$ by $M_{ij} = 1$ if $i = g(j)$, and $M_{ij} = 0$ otherwise. The matrix $M$ is called the permutation matrix of $g$.

- Calculate the permutation matrices for the 6 permutations of $\mathrm{Sym}_3$.
- Prove: If $g$, $h$ are permutations in $\mathrm{Sym}_n$ with associated permutation matrices $M$ and $N$, then the permutation matrix of $g \cdot h$ is $M \cdot N$.
- Prove: If $g$ is a transposition, then $\det(M) = -1$.
- Show that $\mathrm{sign}(g) = \det(M)$.

**Exercise 5.4.14.** Label the vertices of a quadrangle with the numbers 1 to 4.

- Which permutation of the four vertices describes the rotation through + 90 TODO b0 whose center is the middle point of the quadrangle? And which one describes the reflection in the diagonal through the vertices 1 and 3 ?
- Determine the permutations $g$ of $\mathrm{Sym}_4$ satisfying: If $\{i, j\}$ is an edge of the quadrangle, then so is $\{g(i), g(j)\}$.
- Describe each of the permutations of the above part in geometric terms as a reflection or a rotation. Which of these permutations are even?

**Exercise 5.4.15.** Prove that the determinant of a square matrix $A$ and the determinant of its transpose $A^T$ are equal, i.e., prove $\det A = \det A^T$.

**Exercise 5.4.16.** Put the numbers $1, 2, 3, 4$ into a 2 by 2 matrix as follows. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

- Suppose you are allowed to interchange two columns or two rows. Which permutations of $\mathrm{Sym}_4$ can you get using these moves repeatedly? What if you allow as extra type of move a reflection in the diagonal of the matrix?
- Suppose you are allowed to do the following types of moves: Choose a column or row and interchange the two entries. What permutations do you get this way?
- Now consider the 3 by 3 matrix $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$. Individual moves are: Choose two rows (or two columns) and interchange them. Show that you can label each resulting permutation with a pair of permutations from $\mathrm{Sym}_3 \times \mathrm{Sym}_3$. Conclude that you get 36 permutations.
- Experiment with a 3 by 3 matrix, where a single move consists of shifting the entries of an individual column or row cyclically. With the techniques of Chapter 8, you will be able to deal with such problems effectively.

**Exercise 5.4.17.** Label the vertices of a regular tetrahedron with the integers $1, 2, 3, 4$ (see figure). Consider the following moves: For each face of the tetrahedron the corresponding move consists of turning the face 120 degrees clockwise or counter clockwise and moving the labels accordingly (so the vertex opposite the face remains fixed). After applying a number of moves, we read off the resulting permutation $g$ in the obvious way: $g(i)$ is the new label of vertex $i$.

- List the 8 moves as permutations.
- Suppose, after a number of moves, we have obtained the permutation $g$. Show that applying a move $h$ leads to the permutation $g\cdot h^{-1}$.
- Which permutations of $1,2,3,4$ can you get by using these moves?

# Chapter 6

# Monoids and groups

In previous chapters we have considered several sets with operations, like addition and multiplication, defined on them. Such an enriched set is often called a structure. In this chapter we start with a more systematic approach to structures. The title of this chapter refers to the two most important ones we shall deal with here.

## 6.1 Binary operations

The map that takes an integer to its negative is a unary operation on $\mathbb{Z}$, while addition and multiplication are binary operations on $\mathbb{Z}$ in the following sense.

**Definition 6.1.1** (Operations)**.** Let $V$ be a set.

- A *unary operation* is a map $V \to V$.
- A *binary operation* is a map $V \times V \to V$.
- For each natural number $n$, an *n-ary operation* is a map $V^n \to V$.

A set together with a number of operations defined on it is called a *structure*.

**Example 6.1.2** (Unary operation)**.** Of course, any map from a set to itself is a unary operation. But bear in mind maps like these:

- On $\mathbb{Z}$, the map $x \mapsto -x$.
- On $\mathrm{Sym}_n$, the map $g \mapsto g^{-1}$.

**Example 6.1.3** (Binary operation)**.** Also for binary operations there are many possibilities. Some of these are quite natural, like projection onto the first or second coordinate. But examples of interest to us are: addition, multiplication, and subtraction.

**Example 6.1.4.** The $+$ operator is usually considered as a binary operation, but it can also be viewed as a 3-ary operator in the following interpretation: $+(3,5,7) = 3+5+7$.

**Remark 6.1.5.** There also are *nullary* operations. Since $V^0$ is viewed as a singleton (a set consisting of a single element), nullary operations are distinguished elements of $V$. For instance, zero and one of $\mathbb{Z}/n\mathbb{Z}$ and of $\mathbb{R}$, and the identity of $\mathrm{Sym}_n$. These elements are distinguished by properties with respect to other operations. For instance, the effect of adding a zero element to any element is nil. Most of the time such elements will have special names, like identity element.

For a binary operations like $+\colon V \times V \to V$ and $\cdot : V \times V \to V$ we often use *infix* notation:

$$+(a,b) = a+b \tag{6.1}$$

$$\cdot(a,b) = a\cdot b \tag{6.2}$$

This is in accordance with the familiar notation $\cdot$ for multiplication and $+$ for addition in, for example, $\mathbb{Q}$.

Most binary operations in which we are interested distinguish themselves from arbitrary ones in that they have the following property.

**Definition 6.1.6.** A binary operation $\cdot : V \times V \to V$ is called *associative* if, for all $a,b,c \in V$, we have $a\cdot(b\cdot c) = (a\cdot b)\cdot c$.

**Example 6.1.7** (Arithmetic operations)**.** The addition and multiplication on $\mathbb{Z},\mathbb{Q},\mathbb{R}$, and $\mathbb{C}$ are associative.

Subtraction is not.

**Exercise 6.1.8.** For which integers $n > 1$ is subtraction on $\mathbb{Z}/n\mathbb{Z}$ associative?

- None. Not quite.
- All. No, associativity for $n = 10$ would imply $1-2-3 = \mathrm{rem}(1-(2-3),10)$.
- $n = 2$ True; this is the only value of $n$ for which subtraction coincides with addition.

**Remark 6.1.9.** The advantage of using the infix notation becomes obvious from the following comparison of two ways of writing the associativity law for the binary operation $\cdot$:

- For the usual (prefix) notation: $\cdot(a,\cdot(b,c)) = \cdot(\cdot(a,b),c)$.
- For the infix notation: $a\cdot(b\cdot c) = (a\cdot b)\cdot c$.

**Example 6.1.10.** Each polynomial $f \in \mathbb{Z}[X,Y]$ determines a binary operation on $\mathbb{Z}$, mapping $(a,b) \in \mathbb{Z}\times\mathbb{Z}$ to $f(a,b) \in \mathbb{Z}$. For instance, let $f$ in $\mathbb{Z}[X,Y]$ be $X\cdot Y + X + Y$. We test whether the corresponding binary map is associative. To this end, we expand both

$f(x,f(y,z))$ and $f(f(x,y),z)$, for integers $x$, $y$, and $z$ and compare the two. The first expansion gives $f(x,f(y,z)) = f(x,y\cdot z+y+z) = x\cdot y\cdot z+x\cdot y+x\cdot z+y\cdot z+x+y+z$ and the second expansion gives the same result.

Therefore the operation defined by this particular $f$ is associative.

It is well known that brackets are superfluous for the binary operations addition and multiplication on real numbers or integers. This is true for arbitrary associative operations. Indeed, for an associative binary operation brackets are superfluous.

**Theorem 6.1.11.** *If a binary operation is associative, then each positioning of brackets leads to the same result.*

*Proof.* Consider a product $x$ involving the factors $x_1,x_2,\ldots,x_n$, but with an arbitrary positioning of the brackets. We will show, by induction on $n$, that $x$ is equal to the product involving the same factors but with the right-most bracketing, that is, $x_1\cdot(x_2\cdot(x_3\cdot(\ldots\cdot x_n)))$.

If $n \leq 2$ then there is only one possible bracketing, which is the right-most one.

Suppose, therefore, $n \geq 3$. If $x$ is of the form $x_1\cdot y$ with $y$ a product involving the factors $x_2,x_3,\ldots,x_n$, then we can apply the induction hypothesis to $y$. Replacing $y$ in $x$ by its right-most bracketing gives the right-most bracketing of $x$.

If $x$ is not of the form described in the previous paragraph, then $x$ is of the form $(x_1\cdot y)\cdot z$ where for some index $i$ less than $n$ the element $y$ is a product involving the factors $x_2,x_3,\ldots,x_i$ and $z$ is a product involving the factors $x_{i+1},x_{i+2},\ldots,x_n$. But then, by the associative law, $(x_1\cdot y)\cdot z = x_1\cdot(y\cdot z)$, and we are back in the previous case.

□

**Remark 6.1.12.** The number of ways to position $n$ pairs of brackets in a product of $n+1$ variables equals the so-called $n$-th *Catalan number*.

The first Catalan number is equal to 1. The second is equal to 2, and the third Catalan number equals 5 corresponding to the following five ways of placing brackets in an expression with four variables: $a\cdot(b\cdot(c\cdot d))$, $a\cdot((b\cdot c)\cdot d)$, $(a\cdot b)\cdot(c\cdot d)$, $(a\cdot(b\cdot c))\cdot d$, and $((a\cdot b)\cdot c)\cdot d$.

The $n$-th Catalan number is given by the formula $\frac{1}{n+1}\cdot\left(\begin{array}{c}2\cdot n\\ n\end{array}\right)$.

Can you give a proof?

**Example 6.1.13.** Consider $(a\cdot(((b\cdot c)\cdot d)\cdot e))\cdot f$. We can use the associative law to change bracket positionings so as to obtain the rightmost bracketing $a\cdot(b\cdot(c\cdot(d\cdot(e\cdot f))))$:

$$\begin{array}{r}(a\cdot(((b\cdot c)\cdot d)\cdot e))\cdot f =\\ a\cdot((((b\cdot c)\cdot d)\cdot e)\cdot f) =\\ a\cdot(((b\cdot c)\cdot d)\cdot(e\cdot f)) =\\ a\cdot((b\cdot c)\cdot(d\cdot(e\cdot f))) =\\ a\cdot(b\cdot(c\cdot(d\cdot(e\cdot f)))).\end{array} \tag{6.3}$$

The Theorem on brackets for associative operations indeed implies that it is not necessary to use brackets for associative binary operations. Therefore we will often omit the brackets.

Besides the ordinary addition and multiplication, composition of maps from a set $X$ to itself is a very important binary operation.

If $X$ is a set, we write $Maps(X)$ for the set of all maps $X \to X$.

**Theorem 6.1.14** (Composition is associative)**.** *Suppose that $X$ is a set and $f,g,h \in Maps(X)$.*
*The composition of $f$ and $g$, notation $f \circ g$, is the map $X \to X$ given by*

$$f \circ g(x) = f(g(x)) \tag{6.4}$$

*for $x \in X$.*
*Composition is a binary associative operation on $Maps(X)$.*

*Proof.* For each $x \in X$ we have $(f \circ g) \circ h(x) = (f \circ g)(h(x)) = f(g(h(x))) = f(g \circ h(x)) = f \circ (g \circ h)(x)$.

□

**Remark 6.1.15.** In Monoids and semi groups as Maps we shall see a kind of converse to Composition is associative: every associative operation can be viewed as coming from composition of maps.

The case where $X$ is a finite set is dealt with in more detail below.

Suppose that $X$ is a finite set and its elements are labeled $1,2,\ldots,n$. Then an element $g$ of $Maps(X)$ is fully specified by the list $[g(1),g(2),\ldots,g(n)]$ of length $n$ whose $i$-th element equals the image of $i$ under $g$, as the image of every member of $X$ is specified.

**Example 6.1.16** (Maps on five elements)**.** Let $X$ be the set $\{1,2,3,4,5\}$. By a list of length 5 with elements from $X$ we indicate the map $X \to X$ sending element $i$ to the $i$-th element of the list.

For example, $f = [2,2,1,3,3]$ is the map $f = X \to X$ mapping 1 and 2 to 1, mapping 3 to 1, and mapping 4 and 5 to 3.

If, in addition, $g = [4,3,2,3,3]$ and $h = [5,1,4,2,2]$, then $f \circ g = [3,1,2,1,1]$, and so $(f \circ g) \circ h = [1,3,1,1,1]$. On the other hand, $(g \circ h) = [3,4,3,3,3]$, and so $f \circ (g \circ h) = [1,3,1,1,1]$. So, indeed, $(f \circ g) \circ h = f \circ (g \circ h)$.

**Remark 6.1.17.** For the finite set $X = \{1,\ldots,n\}$, both permutations of $X$ and elements of $Maps(X)$ are given by lists. In fact permutations are members of $Maps(X)$ such that every member of $X$ occurs in the list.

Observe that $|Maps(X)| = n^n$ and $|\mathrm{Sym}(X)| = n!$.

Using associativity we can define a very basic structure.

**Definition 6.1.18.** A *semigroup* is a structure $[S,\cdot]$ consisting of a set $S$ and a binary associative operation $\cdot$ on $S$, called multiplication.

**Example 6.1.19** (Structures encountered so far)**.** Each of the following structures is a semigroup: $[\mathbb{Z},+]$, $[\mathbb{Z},\cdot]$, $[\mathbb{Z}/n\mathbb{Z},+]$, $[\mathbb{Z}/n\mathbb{Z},\cdot]$, $[\mathbb{R}[X],+]$, $[\mathbb{R}[X],\cdot]$ $[\mathbb{R}[X]/(f)\mathbb{R}[X],+]$, and $[\mathbb{R}[X]/(f)\mathbb{R}[X],\cdot]$.

**Example 6.1.20** (A variation)**.** $[2\cdot\mathbb{Z},+]$ and $[2\cdot\mathbb{Z},\cdot]$ are also semigroups.

**Example 6.1.21** (Maps)**.** Let $X$ be a set. The set $Maps(X)$ of all maps $X\to X$ is a semigroup with respect to composition of maps.

**Example 6.1.22** (Words)**.** The set $\mathrm{Words}(A)$ of all words over a given alphabet $A$ with concatenation of words is a semigroup.

Here, a word over an alphabet is a sequence of elements from that alphabet. The similarity with usual words, in which case the alphabet is the usual alphabet $\{\mathrm{a},\mathrm{b},...,\mathrm{z}\}$ is a reason for calling these elements words. An alternative name for words is strings.

The concatenation of two words is the act of putting the two words behind each other so as to make a new word. For example the concatenation of the words semi and group leads to the word semigroup. Using the infix notation with symbol $o$ for the operation, we have

$$\mathrm{conca}o\mathrm{tenation} = \mathrm{concatenation} \tag{6.5}$$

Associativity implies that we can write words without brackets!

Semantically, it is not always clear that associativity holds. Consider the word

$$\text{Mathematical\_Knowledge\_Management} \tag{6.6}$$

which is an element of the semigroup $\mathrm{Words}(A)$ where $A$ consists of the English language alphabet and the space, represented here by an underscore _. Now a mathematician might be the appropriate person to deal with

$$\text{Mathematical\_(Knowledge\_Management)} \tag{6.7}$$

whereas a computer scientist might be the better expert for

$$\text{(Mathematical\_Knowledge)\_Management} \tag{6.8}$$

**Remark 6.1.23.** Notice that the ordinary addition of integers is the multiplication of the semigroup $[\mathbb{Z},+]$. Although this terminology may seem confusing, it indicates that the standard operations of addition and of ordinary multiplication have quite a lot in common.

More advanced structures, as we will see later on, usually consist of a semigroup with some additional operations.

When considering a semigroup $[S,\cdot]$, we often speak of the semigroup $S$ if it is clear what the associative multiplication $\cdot$ is.

The sets $Maps(X)$ with composition are an important class of semigroups.

**Definition 6.1.24.** An *identity* in a semigroup $S$ is an element $e$ of $S$ with the property that, for all $a$ in $S$, we have $e\cdot a = a$ and $a\cdot e = a$.

**Example 6.1.25** (Addition)**.** The semigroups $\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ with addition have identity 0.

**Example 6.1.26** (Multiplication)**.** The semigroups $\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ with multiplication have identity $1$.

**Example 6.1.27** (Maps)**.** The identity map $X \to X, x \mapsto x$ is the identity of the semigroup $Maps(X)$.

**Example 6.1.28** (Words)**.** The empty word, that is, the sequence with no letters, is the identity element of the semigroup Words $(A)$ of all words over a given alphabet $A$ with concatenation of words.

We denote the empty word by $\varepsilon$. Of course we have to make sure that this symbol is not an element of $A$!

Instead of an identity, one also speaks of a *unit* element.

**Lemma 6.1.29** (Uniqueness of the identity)**.** *A semigroup has at most one identity.*

*Proof.* Suppose that $e$ and $f$ are identities of the semigroup $[S,\cdot]$. Then $e = e\cdot f = f$.

□

**Remark 6.1.30.** The proof of Uniqueness of the identity does not need the full force of the hypotheses that the two elements are identities. We have only used the facts that $e$ is a left identity and that $f$ is a right identity.

Can you give a proof using only that $e$ is a right identity and that $f$ is a left identity?

**Remark 6.1.31.** The lemma says that there are either no identities in a semigroup or there is exactly one. Both cases occur:

- The semigroup $[\mathbb{N},\cdot]$ of all natural numbers with multiplication has identity $1$.
- The semigroup $[2\cdot\mathbb{N},\cdot]$ of all even natural numbers with multiplication does not have an identity.

Semigroups with an identity are special and have therefore been given a special name:

**Definition 6.1.32.** A structure $[M,\cdot,e]$ in which $[M,\cdot]$ is a semigroup with identity $e$ is called a *monoid*.

**Example 6.1.33** (The usual monoids)**.** Consider the usual arithmetic operations on sets like $\mathbb{N},\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C},\mathbb{Z}/n\mathbb{Z},\mathbb{R},\ldots$. With respect to both addition and multiplication, these sets are semigroups having an identity element, so they are monoids. We call them the additive and multiplicative monoids, respectively.

**Example 6.1.34** (Matrices)**.** There are two natural ways to make the set $M_n(\mathbb{R})$ of all real $n\times n$-matrices a monoid:

- The monoid multiplication is matrix multiplication. The identity element is the identity matrix.
- The monoid multiplication is matrix addition. The identity element is the zero matrix (all entries of the matrix are equal to 0).

**Example 6.1.35** (The symmetric group)**.** The symmetric group $\mathrm{Sym}_n$ with composition as its binary operation and the identity as the identity element, is a monoid.

**Example 6.1.36** (Words)**.** In computer science, the set $\mathrm{Words}(A)$ of all words over the alphabet $A$, as already considered in Example 6.1.22 and Example 6.1.28, is a well-known object of study. It is a monoid with respect to concatenation, whose identity is the 'empty word', the word consisting of 0 letters. Notation $\varepsilon$.

**Example 6.1.37.** We determine all monoids having 2 elements. The identity element is denoted by 1.

Let $[A,\cdot,1]$ be a monoid with two elements. Suppose that $a$ is the unique element of $M$ different from 1. Then for $a\cdot a$ we have only two possibilities. Either $a\cdot a = 1$ or $a\cdot a = a$. This determines the multiplication $\cdot$ completely and we find two multiplication tables for $M$. They give rise to two distinct monoids, denoted by $C_{1,1}$ and $C_{0,2}$. Their multiplication tables are as follows.

| $C_{1,1}$ | 1 | $a$ |
|---|---|---|
| 1 | 1 | $a$ |
| $a$ | $a$ | $a$ |

Table 6.1: Multiplication in monoids

| $C_{0,2}$ | 1 | $a$ |
|---|---|---|
| 1 | 1 | $a$ |
| $a$ | $a$ | 1 |

Both monoids can be realized on the set $\mathbb{Z}/2\mathbb{Z}$. Indeed, addition (with the identity being 0) leads to $C_{0,2}$; multiplication, (with the identity element being 1) leads to $C_{1,1}$.

Since a monoid has only one identity element (as stated in Uniqueness of the identity) we can speak of *the* identity of a monoid.

## 6.2 Monoids and semigroups

There are various constructions of new semigroups and monoids from known ones. The first we discuss is the direct product.

**Theorem 6.2.1** (Direct products)**.** *Let* $[M_1,\cdot_1]$ *and* $[M_2,\cdot_2]$ *be two semigroups. We define a multiplication* $\cdot$ *on* $M_1 \times M_2$*, the Cartesian product of* $M_1$ *and* $M_2$*, as follows:* $(x_1,x_2)\cdot(y_1,y_2) = (x_1\cdot_1 y_1, x_2\cdot_2 y_2)$.
*The resulting structure is again a semigroup, usually called the* direct product*, and denoted by* $M_1 \times M_2$.
*If* $M_1$ *and* $M_2$ *are monoids, then so is* $M_1 \times M_2$*. The identity element of the product is* $(e_1,e_2)$ *where* $e_i$ *is the identity of* $M_i$.

*Proof.* We need to show that the multiplication is associative. This is a direct consequence of the associativity of the multiplications of the two component semigroups, as the following sequence of equalities shows.

$$\begin{aligned}(x_1,x_2)\cdot(y_1,y_2)\cdot(z_1,z_2) =\\ (x_1\cdot_1 y_1, x_2\cdot_2 y_2)\cdot(z_1,z_2) =\\ (x_1\cdot_1 y_1\cdot_1 z_1, x_2\cdot_2 y_2\cdot_2 z_2) =\\ (x_1\cdot_1 y_1\cdot_1 z_1, x_2\cdot_2 y_2\cdot_2 z_2) =\\ (x_1,x_2)\cdot(y_1\cdot_1 z_1, y_2\cdot_2 z_2) =\\ (x_1,x_2)\cdot(y_1,y_2)\cdot(z_1,z_2)\end{aligned} \tag{6.9}$$

□

**Example 6.2.2.** We write out the direct product of the two monoids from Example 6.1.37. These are $C_{0,2}$ and $C_{1,1}$. Their multiplications are given by the following tables.

| $\cdot$ | 1 | $a$ |
|---|---|---|
| 1 | 1 | $a$ |
| $a$ | $a$ | 1 |

| $\cdot$ | 1 | $b$ |
|---|---|---|
| 1 | 1 | $b$ |
| $b$ | $b$ | $b$ |

Their direct product is the monoid on four elements given by the multiplication table below.

| $\cdot$ | $(1,1)$ | $(1,b)$ | $(a,1)$ | $(a,b)$ |
|---|---|---|---|---|
| $(1,1)$ | $(1,1)$ | $(1,b)$ | $(a,1)$ | $(a,b)$ |
| $(1,b)$ | $(1,b)$ | $(1,b)$ | $(a,b)$ | $(a,b)$ |
| $(a,1)$ | $(a,1)$ | $(a,b)$ | $(1,1)$ | $(1,b)$ |
| $(a,b)$ | $(a,b)$ | $(a,b)$ | $(1,b)$ | $(1,b)$ |

**Remark 6.2.3.** The direct product construction can be carried out with more than two monoids.

For example, if we take $M = [\mathbb{R}, +]$, then the additive structure of the vector space $\mathbb{R}^n$ can be obtained as a direct product of $n$ copies of $M$.

**Definition 6.2.4.** Let $[M, \cdot, 1]$ be a monoid. A subset $W$ of $M$ is said to be *closed* under the multiplication $\cdot$ if, for all $a, b \in W$, the product $a \cdot b$ belongs to $W$.

A *submonoid* of $M$ is a subset $W$ of $M$ closed under multiplication and containing $1$.

**Example 6.2.5** (The symmetric and alternating groups)**.** The alternating group $\mathrm{Alt}_n$ is a submonoid of the symmetric group $\mathrm{Sym}_n$. For, the product of two even permutations is again even and the identity map is even.

Also, for $m > n$, the monoid $\mathrm{Sym}_n$ can be viewed as the submonoid of $\mathrm{Sym}_m$ consisting of all permutations fixing $n+1, n+2, ..., m$.

Both $\mathrm{Sym}_n$ and $\mathrm{Alt}_n$ are submonoids of the monoid $\mathrm{Maps}\left(\{1, ..., n\}\right)$ of all maps of $\{1, ..., n\}$ to itself.

**Example 6.2.6** (Polynomial rings)**.** The set of elements of $\mathbb{R}[X]$ which take the value $0$ at some fixed element $a$ form a submonoid of $[\mathbb{R}[X], +, 0]$.

The set of elements of $\mathbb{R}[X]$ which take the value $1$ at some fixed element form a submonoid of $[\mathbb{R}[X], \cdot, 1]$.

**Example 6.2.7** (Matrix rings)**.** The matrices in $M = M_n(\mathbb{R})$ with determinant $1$ form a submonoid, denoted by $SL(n, \mathbb{R})$, of the monoid defined on $M$ by matrix multiplication. Indeed, if $A, B \in SL(n, \mathbb{R})$, then $\det(A \cdot B) = \det(A) \cdot \det(B) = 1$. Moreover, the identity matrix also has determinant $1$.

A second submonoid of $M$ is formed by the set of matrices whose determinant is not equal to $0$. This submonoid is denoted by $GL(n, \mathbb{R})$. Notice that $SL(n, \mathbb{R})$ is also a submonoid of $GL(n, \mathbb{R})$.

**Example 6.2.8** (The integers)**.** The set of even integers $2 \cdot \mathbb{Z}$ is closed under addition and multiplication. The even integers form a submonoid of $\mathbb{Z}$ with respect to addition, but not with respect to multiplication, as $1$ is not even.

**Remark 6.2.9.** A similar definition as in Definition of submonoid can be made for semigroups, except that assertions about the identity element should be removed. This remark will apply more often:

- More notions for monoids apply to semigroups;

- similar notions apply to other structures, to be defined later (such as groups, rings, and fields).

If $W$ is a submonoid of $M$, then the restriction of $\cdot$ to $W$ defines a monoid $[W, \cdot \downarrow W \times W, 1]$, which is called the monoid induced on $W$ by $\cdot$.

The following theorem shows that the intersection of submonoids of a monoid is again a submonoid.

**Theorem 6.2.10.** *If $C$ is a collection of submonoids of $M$, then $\bigcap_{c \in C} c$ is also a submonoid of $M$.*

*Proof.* Let $K = \bigcap_{c \in C} c$. In order to establish that $K$ is a submonoid, we need to prove the following two assertions.

**The identity element belongs to $K$.**

Every element of $C$ contains the identity $1$. Hence, $K$ contains it.

**$K$ is closed under multiplication.**

If $a$ and $b$ are elements of $M$, then they are elements of each $H \in C$. Thus, $a \cdot b \in H$ for every $H \in C$, whence $a \cdot b \in K$.

□

The Theorem on intersection of submonoids shows in particular that, if $W$ and $W'$ are submonoids of $M$, then also $W \cap W'$ is a submonoid of $M$.

If $D$ is a subset of the monoid $M$, then $\langle D \rangle_M$ is defined to be the set of elements of $M$ that are products of elements of $D$. The empty product is (by definition) the unit $e$ of $M$; in this way, $\langle D \rangle_M$ is a submonoid of $M$. This submonoid is called the submonoid (of $M$) *generated* by $D$. The elements of $D$ are called the *generators* of $\langle D \rangle_M$. We say a monoid is *cyclic*, if it can be generated by a single element.

Here is an abstract characterization of $\langle D \rangle_M$.

**Theorem 6.2.11.** *If $D$ is a subset of the monoid $M$ then*

1. *$\langle D \rangle_M$ is the smallest submonoid of $M$ containing $D$;*
2. *$\langle D \rangle_M = \bigcap_{c \in C} c$, where $C$ is the collection of all submonoids of $M$ containing $D$.*

*Proof.*

**$\langle D\rangle_M$ is the smallest submonoid of $M$ containing $D$.**

We need to show that every submonoid of $M$ containing $D$ also contains $\langle D\rangle_M$. Let $W$ be a submonoid of $M$ containing $D$. Since elements of $D$ belong to $W$, their products are in $W$. Hence, $\langle D\rangle_M$ is contained in $W$.

**$\langle D\rangle_M = \bigcap_{c\in C} c$, where $C$ is the collection of all submonoids of $M$ containing $D$.**

Let $R = \bigcap_{c\in C} c$. It is a submonoid, as we have seen in Theorem on intersection of submonoids. Part 1 implies that $\langle D\rangle_M$ is contained in $R$. But, as the submonoid generated by $D$ belongs to $C$, the intersection $R$ is also contained in $\langle D\rangle_M$, whence $\langle D\rangle_M = R$.

□

**Example 6.2.12** (Generation by a single element.)**.** Since every positive integer $n$ can be written as the sum of $n$ times 1, the element 1 of the monoid $[\mathbb{N},+,0]$ generates the whole monoid. This implies that $\mathbb{N}$ is cyclic.

**Example 6.2.13** (Not finitely generated)**.** The monoid $[\mathbb{Z},\cdot,1]$ can be generated by the set of all prime numbers together with 1, but not by a proper subset of this set. Actually, any generating set of this monoid should contain either $+p$ or $-p$ for every prime $p$. Thus $\mathbb{Z}$ is not finitely generated.

The following algorithm determines the submonoid of a given monoid $M$ generated by a given subset $D$. We shall use the notation $x\cdot N$ for $\{x\cdot n \mid n\in N\}$.

**Algorithm 6.2.14.** - *Input: a subset D of a finite monoid M (whose multiplication · is considered given).*

- *Output:* $\langle D\rangle_M$.

```
MonoidGeneratedBy := procedure($D$)
local variables
    $S$
    $N$
$S := \{1\}$ $N := D$ while $N \neq \varnothing$ do
    $S := S \cup N$ , $N := \bigcup_{c\in\{x\cdot N \mid x\in D\}} c \setminus S$
return
    $S$
```

As a third construction method for monoids, we consider the set of words over some alphabet $A$ with concatenation as operation.

**Definition 6.2.15** (Free monoid)**.** Let $A$ be a set of symbols. The free monoid on $A$ is the structure $[\mathrm{Words}(A), o, \varepsilon]$, where

- $\mathrm{Words}(A)$ is the set of all words in the alphabet $A$;

- $o$ is the concatenation of words
- $\varepsilon$ is the empty string.

**Example 6.2.16.** If we take for $A$ the set $\{0,...,9\}$, then the elements from the free monoid on $A$ form the set of natural numbers in their natural representation. However, the product in this monoid is different from the ordinary multiplication $\cdot$. For example the product $o(3124,532) = 3124532$ and hence not equal to $3124 \cdot 532$.

Another difference is that distinct elements may represent the same natural number; for instance, $00087, 087$, and $87$.

**Example 6.2.17.** The free monoid $M$ on a single letter, say $A = \{c\}$, has as elements $c^0, c^1, c^2, ...$

This monoid has the same shape as $[\mathbb{N}, 0, +]$. The map $c^n \mapsto n$ establishes the correspondence. Multiplication in $M$ corresponds to addition in the exponent of $c$ and thus to addition in $\mathbb{N}$.

Our next issue is how to express the fact that two monoids may appear in different guises but are essentially the same.

The standard notion for comparing structures is that of homomorphism.

**Definition 6.2.18** (Homomorphism)**.** Let $S_1$ and $S_2$ be two structures with $n_i$-ary operations $*_{i,1}$ and $*_{i,2}$, respectively (where $i$ runs through a finite set).

A *homomorphism* between these structures is a map $f\colon S_1 \to S_2$ respecting all operations, i.e., for all $i$ we have

$$f(*_{i,1}(a_1, \cdots, a_{n_i})) = *_{i,2}(f(a_1), \cdots, f(a_{n_i})) \qquad (6.10)$$

If $f$ is bijective, then we call $f$ an *isomorphism.*

In particular, for monoids $[M_1, \cdot_1, e_1]$ and $[M_2, \cdot_2, e_2]$ this means the following.

A homomorphism between $M_1$ and $M_2$ is a map $f = M_1 \to M_2$ with the following properties.

- $f(e_1) = e_2$.
- for all $a, b$: $f(a \cdot_1 b) = f(a) \cdot_2 f(b)$.

**Example 6.2.19.** Suppose that all elements of the monoid $M$ can be expressed as products of a single element, say $c$. So $M = \left\{c^0, c, c^2, c^3, ...\right\}$. Then the monoid is said to be generated by $c$. It is cyclic.

Define a map $f\colon \mathbb{N} \to M$ by $f(n) = c^n$. Then we have $f(n+m) = c^{n+m} = c^n \cdot c^m = f(n) \cdot f(m)$. Also, $f(0) = 1$. Hence $f$ is a homomorphism of monoids.

Clearly, $f$ is surjective. But it need not be injective. If $M$ is a free monoid, then the map $f$ is also injective.

Another example of a homomorphism of monoids is the length function for a free monoid. Indeed, if $M$ is a free monoid over an alphabet $A$, then the length function $L$ from $M$ to $\mathbb{N}$ satisfies $L(\varnothing) = 0$ and $L(xoy) = L(x) + L(y)$. If $A$ has size 1, this length function is the inverse of the homomorphism $f$.

If two structures are isomorphic (that is, there is an isomorphism from one to the other), then they are of the 'same shape' (morph = shape).

An isomorphism $S_1 \to S_1$ (that is, with both domain and target structure the same) is called an *automorphism* of $S_1$.

**Remark 6.2.20.** The notion of homomorphism of semigroups is similar; the condition about the identity element is dropped, of course.

Notions like homomorphisms, isomorphisms, and automorphisms (see below) exist for all structures. We shall encounter them again when we discuss rings, groups, and fields.

**Theorem 6.2.21** (Isomorphisms of monoids)**.** *If $f\colon M_1 \to M_2$ is an isomorphism of monoids, then*

1. *the cardinalities of $M_1$ and $M_2$ are equal;*
2. *the inverse map $f^{-1}\colon M_2 \to M_1$ is also an isomorphism of monoids.*

*Moreover, if $g\colon M_2 \to M_3$ then $g \circ f\colon M_1 \to M_3$ is also an isomorphism.*

*Proof.*

**The cardinalities of $M_1$ and $M_2$ are equal;**

This follows from the fact that $f$ is a bijection.

**The inverse map $f^{-1}\colon M_2 \to M_1$ is also an isomorphism of monoids.**

Suppose that $M_1$ and $M_2$ are two monoids and that $f$ is an isomorphism from $M_1$ to $M_2$.

Since $f$ is an isomorphism, $f(e_1)$ is the identity $e_2$ of $M_2$. Consequently, $f^{-1}(e_2) = f^{-1}(f(e_1)) = e_1$, the identity of $M_1$.

Now suppose that $a'$ and $b'$ are elements in $M_2$. Since $f$ is a bijection there exist unique elements $a$ and $b$ in $M_1$ with $a' = f(a)$ and $b' = f(b)$. Then $f(a \cdot b) = a' \cdot b'$. Thus we also have that $f^{-1}(a' \cdot b') = a \cdot b = f^{-1}(a') \cdot f^{-1}(b')$.

We have shown that $f^{-1}$ is also an isomorphism.

**$g \circ f\colon M_1 \to M_3$ is an isomorphism.**

We have

$g \circ f(e_1) = g(f(e_1)) = g(e_2) = e_3$ and $g \circ f(x \cdot y) = g(f(x \cdot y)) = g(f(x) \cdot f(y)) = g(f(x)) \cdot g(f(y)) = g \circ f(x) \cdot g \circ f(y)$.

□

| $\cdot$ | $1$ | $a$ |
|---|---|---|
| $1$ | $1$ | $a$ |
| $a$ | $a$ | $1$ |

| $\cdot$ | $1$ | $b$ |
|---|---|---|
| $1$ | $1$ | $b$ |
| $b$ | $b$ | $b$ |

**Example 6.2.22.** Consider the monoids $C_{1,1}$ and $C_{0,2}$, from Example 6.1.37, given by the following multiplication tables.

Both have size 2. But they are not isomorphic. For otherwise, there would be an isomorphism: $f\colon C_{1,1}\to C_{0,2}$ with $f(1)=1$. Hence, as $f$ is bijective, also $f(a)=a$. But then we would have $1=f(1)=f\left(a^2\right)=f\left(a^2\right)=b^2=b$, a contradiction.

A monoid that can be generated by a single element is called *cyclic*. Let $k,n\in\mathbb{N}$ with $n>0$. An example of a cyclic monoid with generator $c$ is the monoid defined on the set $\left\{c^i \,\middle|\, i\in\{0,\ldots,k+n-1\}\right\}$ by means of the following multiplication rules.

- $c^j\cdot c^i=c^{j+i}$ if $j+i<k+n$;
- $c^j\cdot c^i=c^{k+\mathrm{rem}(j+i-k,n)}$, if $j+i\geq k+n$;
- $c^0=1$ is the identity.

We refer to this monoid as $C_{k,n}$.

Clearly, $C_{k,n}$ is cyclic with generator $c$.

**Theorem 6.2.23** (Characterization of cyclic monoids)**.** *Every cyclic monoid is isomorphic with either* $[\mathbb{N},+,0]$ *or with* $C_{k,l}$ *for certain* $k,l\in\mathbb{N}$.

*Proof.* Suppose that $[C,\cdot,1]$ is a cyclic monoid generated by the element $g$ of $C$. We make the following case distinction.

- There are $k<l$ with $g^k=g^l$. Let $k$ and $l$ be the smallest pair (in lexicographical order) with this property. Then $C=\left\{g^i \,\middle|\, i\in\{0,\ldots,l-1\}\right\}$. Indeed, for all $t\geq 0$, we have $c^{l+t}=c^{k+t}$. Put $n=l-k$. We shall establish that the map $C_{k,n}\to C$ sending $c^i$ to $g^i$, is an isomorphism. Clearly, it is a bijection. In $C$, we have $g^{k+m\cdot n}=g^k$ for all $m$. So, for all $i,j\in\mathbb{N}$ with $k+m\cdot n\leq i+j\leq k+(m+1)\cdot n$, we have $g^i\cdot g^j=g^{i+j-k-m\cdot n}\cdot g^{k+m\cdot n}=g^{i+j-m\cdot n}$. Therefore the powers of $g$ in $C$ satisfy the multiplication laws of $C_{k,n}$. This proves that the bijection $C_{k,n}\to C$ is a homomorphism of monoids. As it is also a bijection, it is an isomorphism.

- There are no such $k$ and $l$ . The map $\mathbb{N} \to C$ given by $n \mapsto c^n$ is readily seen to be an isomorphism from $[\mathbb{N}, +, 0]$ to the monoid $C$.

□

**Remark 6.2.24.** If you think of $C_{k,n}$ in the following way, the reason for the name cyclic becomes clear. First there is the beginning piece of the monoid consisting of $\mathrm{e}, c, c^2, \ldots, c^k$. Then comes the cyclic part, consisting of $c^k, c^{k+1}, c^{k+2}, \ldots, c^{k+n-1}, c^{k+n} = c^k$. At the end of this list we are back at the element $c^k$. After that the cyclic part repeats itself: $c^{k+n+1} = c^{k+1}, c^{k+n+2} = c^{k+2}, \ldots$

We list some properties of the cyclic monoid $C_{k,n}$:

- $|C_{k,n}| = k+n$.
- For every $m \in \mathbb{N}$ with $m > 0$, there are precisely $m$ nonisomorphic cyclic monoids with $m$ elements, viz., $C_{m-k,k}$ for $k = 1, \ldots, m$.
- If $k > 0$, then no element of $C_{k,n}$ but 1 is invertible.
- In $C_{0,n}$ every element is invertible (in other words, $C_{0,n}$ is a group, see later).

If $M$ is a monoid, then it can be viewed as a submonoid of a monoid of maps, see Example 6.1.22.

**Theorem 6.2.25.** 1. *If $M$ is a semigroup, then the map $L\colon M \to Maps(M)$ given by $L_m = x \mapsto m \cdot x$ is a homomorphism of semigroups.*

2. *If, in addition, $M$ is a monoid, then the map $L$ is an injective homomorphism of monoids.*

   *In particular, each monoid $M$ is isomorphic to a submonoid of $Maps(M)$.*

*Proof.*

**The map $L$ is a homomorphism.**

Suppose $x$, $y$ belong to $M$. Then, for each $z$ in $M$, $L_{x\cdot y}(z) = x \cdot y(z) = x \cdot y \cdot z = L_x(y \cdot z) = L_x(L_y(z)) = L_x \cdot L_y(z)$ Consequently, $L_{x\cdot y} = L_x \cdot L_y$, proving that $L$ is a homomorphism of semigroups.

**Suppose that $M$ is a monoid with identity $e$. Then $L$ is an injective homomorphism of monoids.**

Suppose that $x$ and $y$ are elements of $M$ with the same image in $Maps(M)$ under $L$.

Then $x = x \cdot e = L_x(e) = L_y(e) = y \cdot e = y$ so $L$ is injective.

The final assertion follows as $M$ is isomorphic to its image under $L$.

□

**Example 6.2.26.** Consider the multiplicative monoid of $\mathbb{Z}/4\mathbb{Z}$. Multiplication by 1 is the identity map. Multiplication by 0 maps all four element to zero, so equals $[0,0,0,0]$. Here the index $i$ of the list corresponds to the element $i$ modulo 4 of $\mathbb{Z}/4\mathbb{Z}$. Multiplication by 2 equals $[2,0,2,0]$ and Multiplication by 3 equals $[3,2,1,0]$. Verify that the composition of each of these four elements is again one of these four! This expresses the fact that the image of $\mathbb{Z}/4\mathbb{Z}$ under the map $L$ is a submonoid of $Maps(\mathbb{Z}/4\mathbb{Z})$.

**Remark 6.2.27.** For the semigroup with carrier set $\{a,b\}$ and multiplication given by

| $\cdot$ | $a$ | $b$ |
|---|---|---|
| $a$ | $a$ | $a$ |
| $b$ | $a$ | $a$ |

the map $L$ is not injective.

Nevertheless, every semigroup $S$ is a sub-semigroup of $Maps(X)$ for $X$ the union of the carrier set of $S$ and a disjoint singleton $\{e\}$.

This can be shown by extending the multiplication on $S$ to a multiplication on $X$ by demanding that $e$ be the identity of $X$. This multiplication turns $X$ into a monoid. Now apply the theorem to $X$ to conclude that $X$ is a submonoid of $Maps(X)$. As $S$ is a sub-semigroup of $X$, it is also a sub-semigroup of $Maps(X)$.

## 6.3 Invertibility in monoids

**Definition 6.3.1** (Inverse)**.** In a monoid with identity element 1 an element $h$ is called the *inverse* of the element $g$ if $g \cdot h = h \cdot g = 1$.

An element is called *invertible* if it has an inverse.

**Example 6.3.2** (Addition of integers)**.** Consider $[\mathbb{Z},+,0]$. In this monoid every element has an inverse: The inverse of $a$ is $-a$.

**Example 6.3.3** (Multiplication of integers)**.** Consider $[\mathbb{Z},\cdot,1]$. In this monoid only 1 and $-1$ have an inverse; they are their own inverses.

**Example 6.3.4** (Modular arithmetic)**.** Consider $[\mathbb{Z}/10\mathbb{Z},\cdot,1]$. In this monoid only the elements $1,3,7,9$ have an inverse. They are invertible because $3 \cdot 7 \equiv 1 \pmod{10}$ and $9 \cdot 9 \equiv 1 \pmod{10}$.

**Example 6.3.5** (Matrices)**.** Thanks to Cramer's rule we know that exactly those real $n$ by $n$ matrices have an inverse with respect to matrix multiplication that have a nonzero determinant.

**Example 6.3.6** (Monoid of Maps)**.** Consider the monoid Maps $(\{1,...,n\})$ of all maps $\{1,...,n\} \to \{1,...,n\}$, in which multiplication is composition of functions and the identity map is the identity element. In this monoid an element is invertible if and only if it is a permutation.

**Example 6.3.7** (Polyonomials modulo a given polynomial)**.** In $\mathbb{Q}[X]/(d)\mathbb{Q}[X]$, where $d$ is some polynomial in $\mathbb{Q}[X]$, an element $f \in \mathbb{Q}[X]$ represents an invertible element in the multiplicative monoid of $\mathbb{Q}[X]$ if and only if $\gcd(f,d) = 1$.

If an element is invertible, then we can 'divide by it', which just means multiplying with the inverse. But be aware, if multiplication is not commutative, then left and right division might be different.

**Theorem 6.3.8** (Cancellation law)**.** *Let $x, y, z$ be elements of a monoid. If $x$ is invertible and $x \cdot y = x \cdot z$, then $y = z$.*

*Proof.* Suppose that $x, y, z$ are elements of the monoid with $x \cdot y = x \cdot z$, and suppose that $x$ is invertible with inverse $u$. Multiplying both sides of the equality by $u$, we find $u \cdot (x \cdot y) = u \cdot (x \cdot z)$.

Since $\cdot$ is associative, the definition of inverse gives: $y = 1 \cdot y = (u \cdot x) \cdot y = u \cdot (x \cdot y) = u \cdot (x \cdot z) = (u \cdot x) \cdot z = 1 \cdot z = z$.

Hence $y = z$.

□

**Example 6.3.9.** Cancellation does not apply to any three elements in a monoid: For $x = 4$, $y = 5$, and $z = 3$ in $[\mathbb{Z}/8\mathbb{Z}, \cdot, 1]$ we have $x \cdot y = x \cdot z$, but $y \neq z$.

The inverse of an element from a monoid need not exist, but if it does, it is unique:

**Corollary 6.3.10.** *Every element of a monoid has at most one inverse.*

*Proof.* If both $y$ and $z$ are inverses of $x$, then $x \cdot y = e = x \cdot z$. Now apply the Cancellation law to conclude that $y = z$.

□

**Example 6.3.11.**
- In the monoid $[\mathbb{N}, +, 0]$, the element 1 has no inverse.
- In the monoid $[\mathbb{Z}, +, 0]$, the element 1 has inverse $-1$.

The inverse of an invertible element $g$ is denoted by $g^{-1}$.

**Theorem 6.3.12.** *Suppose that $[M, \cdot, e]$ is a monoid. Then*

1. *e is invertible;*
2. *if g and h are invertible, then also $g \cdot h$ is invertible with inverse $h^{-1} \cdot g^{-1}$;*
3. *if g is invertible, then $g^{-1}$ is also invertible with inverse g;*
4. *the subset of invertible elements of M is a submonoid in which every element is invertible.*

*Proof.*

**Part** 1

$e \cdot e = e$, so $e^{-1} = e$.

**Part** 2

$(g \cdot h) \cdot (h^{-1} \cdot g^{-1}) = g \cdot (h \cdot h^{-1}) \cdot g^{-1} = g \cdot e \cdot g^{-1} = g \cdot g^{-1} = e$, so $(g \cdot h)^{-1} = h^{-1} \cdot g^{-1}$.

**Part** 3

$g \cdot g^{-1} = g^{-1} \cdot g = e$, so $g^{-1^{-1}} = g$.

**Part** 4

Follows from the previous parts.

□

**Example 6.3.13.** Consider the monoid of all maps from $\{1, ..., n\}$ to itself. The set of invertible elements in this monoid is $\mathrm{Sym}_n$, which is also a monoid.

**Example 6.3.14.** The invertible elements of the multiplicative monoid of $\mathbb{Z}/8\mathbb{Z}$ are $1, 3, 5, 7$. Each of these elements is its own inverse. In particular, this submonoid of $\mathbb{Z}/8\mathbb{Z}$ is not cyclic.

The theorem implies that if $g$ is invertible, then $g^n$ is invertible for positive $n$. The inverse of $g^n$ is $\left(g^{-1}\right)^n$ and is denoted by $g^{-n}$.

## 6.4 Groups

Monoids in which every element has an inverse deserve a special name.

**Definition 6.4.1** (Definition of a group)**.** A *group* is a structure $\left[G, \cdot, e, x \mapsto x^{-1}\right]$, consisting of a set $G$, a binary associative multiplication $\cdot$ with identity element $e$ and a unary operation $x \mapsto x^{-1}$ such that $x^{-1}$ is an inverse of $x$.

**Example 6.4.2** (The integers)**.** $[\mathbb{Z}, +, 0, z \mapsto -z]$ is the additive group of $\mathbb{Z}$.

**Example 6.4.3** (Multiplicative groups in arithmetic)**.** In $\mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$ every nonzero element has an inverse with respect to multiplication. So on $\mathbb{Q} \setminus \{0\}$, $\mathbb{R} \setminus \{0\}$, and $\mathbb{C} \setminus \{0\}$ we have a group structure with multiplication being the ordinary multiplication.

**Example 6.4.4** (The multiplicative group of $\mathbb{Z}/p\mathbb{Z}$)**.** Suppose that $p$ is a prime. Then multiplication defines a group on $\mathbb{Z}/p\mathbb{Z} \setminus \{0\}$. Indeed, since $p$ is prime, every element has an inverse.

**Example 6.4.5** (Polynomials)**.** Consider the monoid $\mathbb{Q}[X]$. The structure $[\mathbb{Q}[X], +, 0, a \mapsto -a]$ is a group. Multiplication does not define a group structure on $\mathbb{Q}$, since $X$ has no inverse.

**Example 6.4.6** (Modular polynomial arithmetic)**.** Let $R$ be the ring $\mathbb{Q}[X]/(X^2+1)\mathbb{Q}[X]$. Then $R$ is a field, as the polynomial $X^2+1$ is irreducible. Thus every nonzero element has a multiplicative inverse, so multiplication defines a group on $\mathbb{R} \setminus \{0\}$.

**Example 6.4.7** (Symmetric and Alternating groups)**.** Consider the monoids $\mathrm{Sym}_n$ and $\mathrm{Alt}_n$ consisting of all, respectively, all even permutations. In these monoids, every element has an inverse. So both these monoids are also groups. This of course justifies the names symmetric and alternating group.

**Example 6.4.8** (Square matrices)**.** Let $GL(n, \mathbb{R})$ denote the set of $n$ by $n$ matrices with real coefficients and nonzero determinant. Every element in $M_n(\mathbb{R})$ with non-zero determinant has an inverse with respect to matrix multiplication. Hence $GL(n, \mathbb{R})$ is a group, called the *general linear group*. The subset $SL(n, \mathbb{R})$ of matrices of determinant 1 also is a group, called the *special linear group*.

**Example 6.4.9** (The dihedral groups)**.** These are the groups $D_n$ of symmetries of a regular $n$-gon. Consider a regular $n$-gon Gamma. A rotation over $\frac{2 \cdot k \cdot \pi}{n}$ is a symmetry of Gamma. Also a reflection in a line through the center and a vertex of Gamma or the middle of an edge of Gamma is a symmetry. The $n$ different rotations (including the identity) and $n$ different reflections form a group, denoted $D_n$.

This group is called the *dihedral group* of order $2 \cdot n$.

**Example 6.4.10** (The invertible elements of a monoid)**.** By Invertibility in Monoids, invertible elements of a monoid $M$ form a group, usually denoted by $M^{\times}$.

**Remark 6.4.11.** Just as with monoids, we often talk about a group $G$ without mentioning all binary and unary operations. Sometimes we indicate with a single word what type of operation we are considering.

For example, the additive group of the integers is understood to be the group defined on the monoid $[\mathbb{Z}, +, 0]$, whose inverse map is $z \mapsto -z$.

**Remark 6.4.12.** Since groups are monoids, the properties that we have derived for monoids so far, also hold for groups.

Notation introduced for monoids will also apply to groups. For example, a group is called *commutative* (or *abelian* after the mathematician Abel) if the corresponding monoid is commutative, i.e., the multiplication is commutative.

**Remark 6.4.13.** Note that if $\left[G,\cdot,e,x\mapsto x^{-1}\right]$ is a group, then $[G,\cdot,e]$ is a monoid. Since every element of a monoid has at most one inverse, we could also have defined a group as a monoid in which every element has an inverse.

We discuss some constructions of groups. Since groups are also monoids, we can consider the same constructions as in the previous section on monoids. In particular we will deal with the direct product of two groups.

**Definition 6.4.14** (Direct Product of Groups)**.** If $G$ and $H$ are groups, their direct product as monoids $G\times H$ is a group. It is called the *direct product* of $G$ and $H$.

Likewise, the product of several groups can be defined. The direct product of $n$ copies of the same group $G$ is denoted by $G^n$.

*Proof.* By Direct products, this direct product is a monoid. But each element has an inverse: the inverse of the element $(a,b)$ of $G\times H$ is equal to $\left(a^{-1},b^{-1}\right)$.

□

**Example 6.4.15.** The direct product $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$ of two copies of the additive group $\mathbb{Z}/2\mathbb{Z}$ consists of $e=(0,0)$, $a=(1,0)$, $b=(0,1)$, and $c=(1,1)$.

Each nonidentity element has order 2 (that is, $x^2=e$). Moreover, $a\cdot b=c$, $a\cdot c=b$, and $b\cdot c=a$. Don't be confused by the multiplicative notation for the binary operation!

**Remark 6.4.16.** The direct product construction can be considered associative in the sense that $G\times G\times G=G\times G\times G=G\times G\times G$.

Of course, these equalities are considered to be the natural identifications.

Just like submonoids, we can define subgroups.

**Definition 6.4.17.** A subset $H$ of a group $G$ is called a *subgroup* if $H$ is a submonoid of the monoid $G$ and the inverse of every element in $H$ is again in $H$. Thus, $H$ is a subgroup of $G$ if the following holds.

- $e\in H$, where $e$ is the identity of $G$;
- $a\cdot b\in H$ and $a^{-1}\in H$ for all $a,b\in H$.

**Example 6.4.18** (Some subgroups of the additive group of the integers)**.** Consider the subset $n\cdot\mathbb{Z}$ of $\mathbb{Z}$. This set is closed under addition. Moreover, it contains 0 and for each element $x$ also its opposite $-x$. So, $n\cdot\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

**Example 6.4.19** (Some subgroups of polynomial rings)**.** Let $R$ be a ring like $\mathbb{Z}$, $\mathbb{Q}$ $\mathbb{R}$ or $\mathbb{C}$. For each natural number $n$, the set of elements of degree at most $n$ form a subgroup of $[R[X], +, , ]$.

Also, the set of all elements in $R[X]$ that take the value 0 at some fixed $x$ form a subgroup of the additive group on $R[X]$.

**Example 6.4.20** (Permutation groups)**.** A *permutation group* is by definition a subgroup of $\text{Sym}(X)$ for some set $X$. We will often consider the finite case and take $X = \{1, ..., n\}$. We find $\text{Sym}_{n-1}$ to be a subgroup of $\text{Sym}_n$. If $m$ is less than $n$, then we can think of $\text{Sym}_m$ as consisting of those permutations in $\text{Sym}_n$ that fix all $x$ with $x > m$. So, we can view $\text{Sym}_m$ as a subgroup of $\text{Sym}_n$. Similarly we can view $\text{Alt}_m$ as a subgroup of $\text{Alt}_n$.

**Example 6.4.21** (Some subgroups of the group of invertible matrices)**.** Recall that $SL(n, \mathbb{R})$ denotes the set of real $n$ by $n$ matrices with determinant 1. Each element in $SL(n, \mathbb{R})$ has an inverse with respect to matrix multiplication. This inverse has determinant 1. Also, the product of two elements of $SL(n, \mathbb{R})$ has determinant 1. Hence $SL(n, \mathbb{R})$ is a subgroup of $GL(n, \mathbb{R})$, called the special linear group.

The subset of matrices of determinant $-1$ or 1 also forms a subgroup of $GL(n, \mathbb{R})$. The subset of upper (or lower) triangular matrices of $GL(n, \mathbb{R})$ or $SL(n, \mathbb{R})$ is closed under multiplication and inverses and hence a subgroup of $GL(n, \mathbb{R})$ or $SL(n, \mathbb{R})$, respectively.

**Example 6.4.22** (The rotations in the dihedral group)**.** The rotations in the dihedral group form a subgroup. Consider a regular $n$-gon Gamma. The rotations over $\frac{2 \cdot k \cdot \pi}{n}$, $k = 0, ..., n-1$, around the center of form a subgroup with $n$ elements of $D_n$.

Just like submonoids are monoids, subgroups are themselves groups: A subgroup contains the identity element, is closed with respect to taking products and contains the inverse of every one of its elements.

In the context of groups we also have the notion ‘generated by’.

**Definition 6.4.23.** Let $D$ be a subset of a group $G$. The set of all products $g_1 \cdot g_2 \cdot \ldots \cdot g_n$ where $n$ is a natural number and $g_i$ an element or the inverse of an element of $D$, is a subgroup of $G$, called the subgroup *generated by* $D$ and denoted $\langle D \rangle_G$.

If the the group $G$ is clear from the context, one often writes $\langle D \rangle$ instead of $\langle D \rangle_G$.

If $G = \langle D \rangle_G$, then we say that $G$ is generated by $D$. A group is called finitely generated if the group is generated by a finite set.

We call a group *cyclic* if it can be generated by a single element.

**Example 6.4.24** (Some groups generated by one element)**.** A group generated by a single element $g$ is cyclic and consists of the (not necessarily distinct) positive and negative powers of $g$: $..., g^{-2}, g^{-1}, g^0 = 1, g^1, g^2, ....$

The group $[\mathbb{Z}, +, 0, ]$ or $[\mathbb{Z}/n\mathbb{Z}, +, 0, ]$ is cyclic. It can be generated by 1 and by $-1$.

The group $\mathbb{Z}/10\mathbb{Z}^\times$ of invertible elements in $\mathbb{Z}/10\mathbb{Z}$ is cyclic. It can be generated by the element 3.

**Example 6.4.25** (Generators for symmetric and alternating groups)**.** Every element of $\mathrm{Sym}_n$ is a product of transpositions, see Every permutation is a product of transpositions. Thus $\mathrm{Sym}_n$ is generated by its transpositions. The even elements of $\mathrm{Sym}_n$ can be written as products of 3-cycles, see Every even permutation is a product of 3-cycles. Hence $\mathrm{Alt}_n$ is generated by its 3-cycles.

**Example 6.4.26** (Two generating reflections for $D_n$)**.** Consider $D_n$ the group of symmetries of a regular $n$-gon. If $r$ and $s$ denote two reflections in $D$ whose reflection lines make an angle of $\frac{\pi}{n}$, then their product is a rotation over $\frac{\pi \cdot 2}{n}$. Hence we have the following equalities where 1 denotes the identity map. $r^2 = \mathrm{e} = s^2$, $(r \cdot s)^n = 1$. Now it is straightforward to check that the elements of $D_n$ are $1, r, r \cdot s, r \cdot s \cdot r, \ldots, r \cdot s \cdot r \cdot \ldots \cdot s \cdot r, s$. (Can you find out which one of these is a reflection and which one is a rotation?) So, the group $D_n$ is generated by $r$ and $s$.

The subgroup of a group $G$ generated by a set $D$ equals the submonoid generated by $D \cup D^{-1}$, where $D^{-1}$ is the set of all $d^{-1}$ for $d \in D$.

The two results for monoids on intersections and monoids generated by subsets have their analogues for groups.

**Theorem 6.4.27.** *Let $C$ be a collection of subgroups of a group $G$. Then $\bigcap_{c \in C} c$ is also a subgroup of $G$.*

*Proof.* Let $K$ be $\langle G \rangle_D$. Then by the result on intersections of submonoids we find that $K$ is a submonoid of $G$. It remains to check that for every $k \in K$, also the inverse $k^{-1}$ is in $K$.

Since $k$ is an element of every $H$ in $C$, also $\langle G \rangle_D$ is in $H$ for every $H$ in $C$ (this is because $H$ is a subgroup). Hence $\langle G \rangle_D$ is in the intersection $K$ of all $H$ in $C$.

□

Now the abstract characterization of $\langle G \rangle_D$.

**Theorem 6.4.28.** *Let $G$ be a group and $D$ a subset of $G$. Then*

- *$\langle G \rangle_D$ is the smallest subgroup of $G$ containing $D$;*
- *$\langle G \rangle_D = \bigcap_{c \in C} c$, where $C$ is the collection of all subgroups of $G$ that contain $D$.*

*Proof.* Compare this with the proof of Charactrization of the submonoid generated by a set.

□

**Example 6.4.29.** In the additive group of the integers, the multiples $n\cdot\mathbb{Z}$ of a fixed number $n$ form a subgroup. The intersection of $m\cdot\mathbb{Z}$ and $n\cdot\mathbb{Z}$ is the subgroup $\operatorname{lcm}(m,n)\cdot\mathbb{Z}$.

We now consider three special types of subgroups.

**Theorem 6.4.30.** *Let $G$ be a group and $X$ a subset of $G$. Then each of the following three subsets of $G$ is a subgroup of $G$.*

- *The centralizer of $X$ in $G$, i.e., the subset of all $g \in G$ with $g\cdot x = x\cdot g$ for all $x$ of $X$.*
- *The normalizer of $X$ in $G$ , i.e., the subset of all $g \in G$ with $g\cdot X\cdot g^{-1} = X$.*
- *The center of $G$, i.e., the centralizer of $G$ itself.*

*Proof.* We prove that the centralizer $C$ in $G$ of a set $X$ is a subgroup of $G$. The other cases are left to the reader.

**The centralizer $C$ contains the unit element 1.**

For each $x \in X$ we have $1\cdot x = x = x\cdot 1$. Hence $1 \in C$.

**$C$ is closed under multiplication.**

Suppose $g$ and $h$ are both in $C$. Then they centralize $X$, that is, $g\cdot x = x\cdot g$ and $h\cdot x = x\cdot h$ for all $x \in X$. Consequently, $g\cdot h\cdot x = g\cdot h\cdot x = g\cdot x\cdot h = x\cdot g\cdot h = x\cdot g\cdot h$ and we find that $g\cdot h$ also centralizes $X$, and so belongs to $C$.

**$C$ is closed under inversion.**

Suppose $g$ belongs to $C$. Then $g\cdot x = x\cdot g$ for all $x$ in $X$. Multiplying this equality from the right and left with $g^{-1}$ we find $x\cdot g^{-1} = g^{-1}\cdot x$. Since this holds for all $x$ in $X$, we conclude that $g^{-1}$ also centralizes $X$.

Hence the centralizer of $X$ is a subgroup of $G$.

□

**Example 6.4.31** (Commutative groups)**.** If $G$ is commutative, then centralizer, normalizer of any subset and center of $G$ are all three equal to $G$.

**Example 6.4.32** (Symmetric groups)**.** The center of $\mathrm{Sym}(n)$ is trivial if $n > 2$. It only consists of the identity. Indeed, if $c$ is an element of the center, it has to commute with the transposition $(1,2)$. Hence $c(2) = c((1,2)(1)) = (1,2)(c(1))$. Since $c(2)$ and $c(1)$ are distinct, we find that $c(1)$ is in the support of $(1,2)$. The same reasoning with $(i,j)$ instead of $(1,2)$ yields that $c(i)$ is in the support of $(i,j)$. Varying $j$ implies that $c(i) = i$.

How about the case $n = 2$?

**Example 6.4.33** (General linear groups)**.** Consider the group $GL(n,\mathbb{R})$. The center of this group coincides with the set of diagonal matrices with nonzero determinant. The proof of this fact is similar to the Example 6.4.32. Fix a basis $B$ consisting of $b_1,\ldots,b_n$. Let $P_{i,j}$ be the linear map that interchanges the basis vectors $b_i$ and $b_j$ and fixes all other basis elements of $B$.

If $c$ is an element in the center, then it commutes with all $P_{i,j}$. Suppose $n > 2$ and let $k$ be different from $i, j$. Then $c(b_k) = c(P_{i,j}(b_k)) = P_{i,j}(c(b_k))$. So $c(b_k)$ is contained in the 1-eigenspace of $P_{i,j}$. Similarly we obtain that $c(b_k)$ is contained in the 1-eigenspace of $Q_{i,j}$, the linear map that fixes all $b_k$ except for $b_i$ and $b_j$, and acts on these two elements as follows: $Q_{i,j}(b_j) = -(b_i)$ and $Q_{i,j}(b_i) = b_j$. Thus $c(b_k)$ is contained in the space generated by $B\setminus\{b_i,b_j\}$. Varying the $i$ and $j$, we easily find that $c(b_k)$ is a scalar multiple of $b_k$. Hence $c$ has a diagonal matrix with respect to the basis $B$. But $B$ was chosen to be an arbitrary basis. Hence, each vector is an eigenvector of $c$. It follows that $c$ has only one eigenvalue, and so $c$ is multiplication by a scalar.

Again the case $n = 2$ is left to the reader.

**Remark 6.4.34.** The normalizer of a subset $X$ of $G$ contains the centralizer as a subgroup of $G$. Both the normalizer and the centralizer of $X$ contain the center of the group.

The center of a group is commutative.

## 6.5 Cyclic groups

Cyclic groups, just like cyclic monoids, are well understood.

**Theorem 6.5.1.** *Let $G = \left[\left\{g^k \middle| k\in\mathbb{Z}\right\},\cdot,e,x\mapsto x^{-1}\right]$ be a cyclic group of size $n$ with generator $g$. If $n$ is infinite, then $G$ is isomorphic to $[\mathbb{Z},+,0,x\mapsto -x]$. If $n$ is finite, then $G$ is isomorphic to $[\mathbb{Z}/n\mathbb{Z},+,0,x\mapsto -x]$.*

*Proof.*

**The map $f\colon \mathbb{Z}\to G$ given by $f(i) = g^i$ is a homomorphism of groups.**

For any $i, j\in\mathbb{Z}$, $g^{i+j} = g^i\cdot g^j$ and $g^0 = e$.

**If there is no positive integer $n$ such that $g^n = e$, then $f$ is a bijection.**

Clearly, $f$ is surjective. Suppose there are distinct positive integers $i, j$ such that $g^i = g^j$. Then, for $n = i - j$, we have $g^n = g^{i-j} = e$. Thus, $f$ is injective as well, and hence an isomorphism.

**Otherwise, if $n$ is the minimal positive integer $n$ with $g^n = e$, then $G$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}$.**

By the choice of $n$, the elements $g^i$ for $i = 0,1,...,n-1$, are all distinct. Now for any $m = q \cdot n + r$, with $q$ the quotient and $r$ the remainder of $m$ divided by $n$, we have $f(m) = g^m = g^{q \cdot n + r} = g^{n \cdot q} \cdot g^r = g^r = f(r)$. In particular, the map $f' = \mathbb{Z}/n\mathbb{Z} \to G$ given by $f'(m + n \cdot \mathbb{Z}) = g^m$ is well defined. It is straightforward to check that $f'$ is an isomorphism of groups.

□

**Example 6.5.2.** Another incarnation of the finite cyclic group $C_n$ of order $n$ is the subgroup of $\mathrm{Sym}_n$ generated by $(1,2,...,n)$.

**Remark 6.5.3.** In the case where the cyclic group $G$ is finite of order $n$, it is isomorphic to the monoid $C_{0,n}$ defined and studied in the Characterization of cyclic monoids. In other words, $C_n$ is isomorphic to $C_{0,n}$.

The size of a finite group or monoid is often referred to as its *order*.

A cyclic group of order $n$ is denoted by $C_n$. If $n$ is finite, then we also use $\mathbb{Z}/n\mathbb{Z}$ for a cyclic group of order $n$, as $C_n$ is isomorphic to the additive group of $\mathbb{Z}/n\mathbb{Z}$.

**Definition 6.5.4.** If $G$ is a group and $g \in G$, then the order of $g$ is the smallest positive integer $m$ with $g^m = e$. If no positive integer $m$ with $g^m = e$ exists, we say that the order of $g$ is infinite.

**Example 6.5.5.** • The order of the identity element is 1.

- The order of the permutation $(1,2,3)$ in $\mathrm{Sym}_3$ is 3. Indeed, $((1,2,3))^3 = 1$ but $((1,2,3))^2 = (1,3,2)$ which is not equal to 1.

- The order of the complex number $i$ is 4.

- The order of 2 in $\mathbb{Z}/5\mathbb{Z}^\times$ is 4, as follows from $2^2 = 4$, $2^3 = 3$, and $2^4 = 1$.

**Example 6.5.6.** • The order of $g$ in $G$ is the size of the subgroup $\langle g \rangle_G$ of $G$ generated by $g$.

- The notion of order introduced here generalizes the notion of the order of a permutation defined in Order of a Permutation.

**Remark 6.5.7.** The order of an element $g$ of a group $G$ is equal to the order of the subgroup of $G$ generated by $g$. Both are equal to the size of the set $\{e, g, g^2, ...\}$.

For cyclic groups we can give more detailed information on the order of its elements:

**Theorem 6.5.8.** *Let $G$ be a cyclic group of order $n$ with generator $g$.*

1. *Every subgroup of $G$ is cyclic.*
2. $\langle g^k \rangle_G = \langle g^d \rangle_G$ *for* $d = \gcd(n,k)$; *it is a subgroup of order* $n/d$.
3. $g^k$ *generates $G$ if and only if* $\gcd(k,n) = 1$.

*Proof.* Let $G$ be a cyclic group of order $n$ with generator $g$.

**Every subgroup of $G$ is cyclic.**

Let $H$ be a subgroup of $G$, and suppose that $k$ is the smallest positive integer such that $g^k$ is in $H$. Suppose now that $g^l$ is also in $H$ for some positive integer $l$. By the extended Euclidean algorithm, there exist integers $a$ and $b$ such that $m = \gcd(k,l)$ can be expressed as $a \cdot k + b \cdot l$. But then $g^m$, being equal to $\left(g^k\right)^a \cdot \left(g^l\right)^b$, is an element of $H$. By the choice of $k$, we find that $k = m$ and that $l$ is a multiple of $k$. In particular, $g^l$ is an element of $\langle g^k \rangle$. This proves that $H = \langle g^k \rangle$.

**$\langle g^k \rangle_G = \langle g^d \rangle_G$ for $d = \gcd(n,k)$; it is a subgroup of order $\frac{n}{d}$.**

Let $d = \gcd(k,n)$. By the extended Euclidean algorithm there is a relation $d = a \cdot k + b \cdot n$. So, for every $l$ we have the relation $d \cdot l = a \cdot k \cdot l + b \cdot n \cdot l$. This implies that every power of $g^d$ is also a power of $g^k$. On the other hand, as $d$ divides $k$, every power of $g^k$ is also a power of $g^d$. This shows that

$g^k$ and $g^d$ generate the same subgroup of $G$. Since $d$ divides $n$, the first power of $g^d$ equal to 1 is $\frac{n}{d}$. Therefore, the subgroup $\langle g^d \rangle$ of $G$ has order $\frac{n}{d}$.

**$g^k$ generates $G$ if and only if $\gcd(k,n) = 1$.**

In view of Part 2 and the fact that the value of $\gcd(k,n)$ does not change after replacement of $k$ by $\gcd(k,n)$, we may assume that $k$ divides $n$.

But then, by the second assertion of Part 2, the order of $\langle g^k \rangle$ is $\frac{n}{k}$. This is equal to $n$ if and only if $k$ is equal to 1. Part 3 follows since the subgroup $\langle g^k \rangle$ of $G$ coincides with $G$ if and only if its order is equal to $n$.

□

**Example 6.5.9.** If $g$ is an element of a cyclic group $G$ of order $n$, then the order of $g$ is a divisor of $n$. We can use this to show that 2 generates the multiplicative group of $\mathbb{Z}/101\mathbb{Z}$: Since 101 is prime, the group $\mathbb{Z}/101\mathbb{Z}^\times$ of invertible elements of $\mathbb{Z}/101\mathbb{Z}$ contains 100 elements. By Fermat‘s little theorem we also have that $2^{100} \equiv 1 \pmod{101}$. Thus the order of 2 is a divisor of 100. Easy computations show that

- $2^{10} \equiv 14 \pmod{101}$,
- $2^{20} \equiv -6 \pmod{101}$, and
- $2^{50} = (-6) \cdot (-6) \cdot 14 = 504 \equiv -1 \pmod{101}$.

Hence the order of 2 is neither a divisor of 50 nor of 20, and so it is 100.

The above implies that the number of generators in a cyclic group of order $n$ equals $\Phi(n)$. (Here $\Phi$ denotes the Euler indicator).

We use this in the following characterization

**Theorem 6.5.10** (Characterisation of cyclic groups)**.** *Let $G$ be a finite group of order $n$.*
*The group $G$ is cyclic if for each proper divisor $m$ of $n$, there are exactly $m$ elements $g$ in $G$ with $g^m = 1$.*

*Proof.* Denote by $\mathrm{psi}\,(m)$ the number of elements $g$ in $G$ of order $m$. Then we have

$$\mathrm{psi}\,(1) = 1 \tag{6.11}$$

and

$$\mathrm{psi}\,(m) = m - \sum_{d \in \{d \in \mathbb{N} | d | m\}} \mathrm{psi}\,(d) \tag{6.12}$$

This implies that psi satisfies the same recursion as the Euler Totient function, see Euler Totient Theorem. In particular, psi $=$ euler.

But that implies that $G$ contains $\Phi(n) > 0$ elements of order $n$ and hence $G$ is cyclic.

□

## 6.6 Cosets

Let $G$ be a group and $H$ a subgroup of $G$. For $g \in G$, we write

$$g \cdot H = \{g \cdot h | h \in H\} \tag{6.13}$$

**Lemma 6.6.1.** *Let $\sim$ be the relation on $G$ given by $g \sim k$ if and only if $k^{-1} \cdot g \in H$.*

- *The relation $\sim$ is an equivalence relation.*
- *The $\sim$-equivalence classes are the sets $g \cdot H$ with $g \in G$.*

*Proof.*

**$\sim$ is an equivalence relation.**

We need to establish that $\sim$ is reflexive, symmetric, and transitive.

- **Reflexivity:** $g \sim g$, since $g^{-1} \cdot g = 1_H$.
- **Symmetry:** If $g \sim k$, then $k^{-1} \cdot g \in H$. But then $k^{-1} \cdot g^{-1} = g^{-1} \cdot k$ is also in $H$ and we find $k \sim g$.
- **Transitivity:** If $g \sim k$ and $k \sim h$, then $k^{-1} \cdot g, h^{-1} \cdot k \in H$. But then also $h^{-1} \cdot k \cdot k^{-1} \cdot g = h^{-1} \cdot g \in H$. Hence $g \sim h$.

**The $\sim$-equivalence classes are the sets $g \cdot H$ with $g \in G$.**

Let $k, g \in G$. Then $k \sim g$ is equivalent to the existence of $h \in H$ such that $g^{-1} \cdot k = h$, which in turn is equivalent to $k = g \cdot h$ for some element $h$ of $H$, and so can be rewritten as $k \in g \cdot H$.

□

**Example 6.6.2.** Let $G$ be the cyclic (additive) group $[\mathbb{Z}/n\mathbb{Z}, +, 0, -]$ of order $n = p \cdot q$, and let $H$ be the subgroup generated by the residue class of the integer $q$ in $\mathbb{Z}/n\mathbb{Z}$. Then $H$ has order $p$, and $r \sim s$ if and only if $q$ divides $r - s$. In particular, the equivalence class of $r$ consists of all residue classes in $\mathbb{Z}/n\mathbb{Z}$ of $s \in \mathbb{Z}$ such that $s \equiv r \pmod{q}$.

Taking the specific values $n = 15, p = 5, q = 3$, we find $H = \{0, 3, 6, 9, 12\}$ and the equivalence classes are: $H$, $1 + H = \{1, 4, 7, 10, 13\}$, $2 + H = \{2, 5, 8, 11, 14\}$.

The $\sim$-equivalence classes of an equivalence relation partition $G$. These $\sim$-equivalence classes are so important that they deserve a special name:

**Definition 6.6.3** (Definition of Cosets)**.** The $\sim$-equivalence classes $g \cdot H$ with $g \in G$, are called the *left cosets* of $H$ in $G$.

The set of all left cosets $g \cdot H$ of $H$ in $G$ is denoted by $G/H$.

For $g \in G$, we write

$$H \cdot g = \{h \cdot g | h \in H\} \tag{6.14}$$

This set is called the *right coset* of $H$ containing $g$.

The right cosets of $H$ partition $G$. The set of all right cosets of $H$ in $G$ is denoted by $G\backslash\backslash H$.

**Example 6.6.4.** Some example of cosets are:

- Cosets of a subspace of a vector space. Let $V$ be a real vector space. The linear subspaces of $V$ are subgroups of the additive group on $V$. The left (and right) cosets of a fixed 1 - dimensional linear subspace $L$ of $V$ are those lines in $V$ that are parallel to $L$.
- The left cosets of $\mathrm{Sym}_2$ of $\mathrm{Sym}_3$
  - $H$,
  - $(2,3) \cdot H = \{(2,3), (1,3,2)\}$, and
  - $(1,3) \cdot H = \{(1,3), (1,2,3)\}$.

- The right cosets of $G\backslash\backslash H$ of $H$
  - $H$,
  - $H\cdot(2,3)=\{(2,3),(1,2,3)\}$, and
  - $H\cdot(1,3)=\{(1,3),(1,3,2)\}$.

We introduce computations with cosets. It is a preparation for the construction of the quotient group.

Let $G$ be a group. If $X,Y$ are subsets of the group $G$ and $a,b\in G$, then we write $a\cdot X=\{a\cdot x|x\in G\}$, $X\cdot a\cdot Y=\{x\cdot a\cdot y\mid xy\in X\times Y\}$, $a\cdot X\cdot b\cdot Y=\{a\cdot x\cdot b\cdot y\mid xy\in X\times Y\}$, etc.

Let $H$ be a subgroup of $G$.

**Theorem 6.6.5.** *Suppose that H is a subgroup of G. Then, for all $a,b\in G$*

*i.* $a\cdot(b\cdot H)=(a\cdot b)\cdot H$;

*ii.* $(a\cdot H)\cdot H=a\cdot H$;

*iii. If $a\in b\cdot H$, then $a\cdot H=b\cdot H$.*

*Proof.*

$a\cdot(b\cdot H)=(a\cdot b)\cdot H$

$a\cdot(b\cdot H)=a\cdot\{b\cdot x|x\in H\}=\{a\cdot(b\cdot x)|x\in H\}=a\cdot b\cdot H.$

$a\cdot H\cdot H=a\cdot H$**.**

Since $H$ contains the identity element, we clearly have $H\subset H\cdot H$. But $H$, being a subgroup of $G$, is closed under multiplication, so $H\cdot H=H$. The assertion follows by left multiplication with $a$.

**If** $a\in b\cdot H$**, then** $a\cdot H=b\cdot H$**.**

Suppose $a\in b\cdot H$. Then $a\cdot H$ is a subset of $b\cdot H$. But, as these cosets are classes of an equivalence relation, they coincide.

□

**Example 6.6.6.** Let $G$ be the symmetric group $\mathrm{Sym}_3$. The subgroup

$$H=\{\mathrm{e},(1,2,3),(1,3,2)\}$$

of order 3 is a normal subgroup. It has index 2. In fact, more generally, whenever $H$ is a subgroup of $G$ of index 2, it is a normal subgroup. For then, for $g\in G$, either $g\in H$ and so $g\cdot H=H=H\cdot g$ or or not, in which case $g\cdot H=G\setminus H=H\cdot g$.

**Example 6.6.7.** Let $G$ be the symmetric group $\mathrm{Sym}_4$. The subgroup $H = \{\mathrm{e}, (1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\}$ of order 4 is a normal subgroup. For, it is a subgroup and it is the union of two conjugacy classes. It has index 6 in $G$.

**Remark 6.6.8.** Group multiplication induces a monoid structure on $P(G)$, the collection of all subsets of $G$. The identity element is $\{1_G\}$. The expressions $X \cdot a \cdot Y$ and $a \cdot X \cdot b \cdot Y$ discussed earlier can be seen as elements of the monoid.

The cosets of $H$ in $G$ are the equivalence classes of the equivalence relation, $\sim$, called 'congruence modulo $H$' on the set $G$ given by $\sim$ if and only if $b^{-1} \cdot a \in H$. Observe that $\sim (a,b)$ if and only if $a \cdot H = b \cdot H$.

The following result is a very important consequence of the fact that the left cosets of a subgroup partition a group.

**Theorem 6.6.9** (Lagrange's theorem)**.** *Let $G$ be a finite group and $H$ a subgroup of $G$. Then $|G/H| = |G| / |H|$. In particular, $|H|$ divides $|G|$.*

*Proof.* The (left or right) cosets of a subgroup $H$ of $G$ are the equivalence classes of the equivalence relation $\sim$. Thus, these cosets partition the set $G$. As each coset contains $|H|$ elements we find that $|H|$ divides $|G|$.

□

**Example 6.6.10.** Let $G$ be the cyclic (additive) group $\mathbb{Z}/n\mathbb{Z}$ of order $n = p \cdot q$, and $H$ the subgroup generated by the residue class $q \in \mathbb{Z}/n\mathbb{Z}$. Then $H$ has order $p$, and $|G/H| = q$. The cosets are of the form $r + H$ with $r = 0, \ldots, q-1$.

**Remark 6.6.11.** The converse of Lagrange‘s theorem does not hold! The group $\mathrm{Alt}_5$ has 60 elements, but does not have a subgroup containing 30 elements.

Here is the proof. Suppose that $H$ is a subgroup of order 30. Then there must be a 3-cycle $g$ which is not contained in $H$, since the 3-cycles generate $\mathrm{Alt}_5$. Every (left) coset of $H$ contains 30 elements, so there are only two of them. As soon as an element $a$ is not contained in $H$, then $H$ and $a \cdot H$ are the two left cosets. Apply this observation to the elements $g$ and $g^2$: the group $\mathrm{Alt}_5$ is partitioned into the two left cosets $H$ and $g \cdot H$, but also into the two left cosets $H$ and $g^2 \cdot H$. But $g^2$ is not in $H$, because then $g^4$, which equals $g$, would also be in $H$. This implies that $g \cdot H = g^2 \cdot H$. We conclude that the element $g^2$ is in $g \cdot H$, which in turn implies that $g$ is in $H$. This is a contradiction.

**Remark 6.6.12.** The analog of Lagrange's theorem for right cosets also holds: $|H \backslash\backslash G| = |G| / |H| = |G/H|$.

If $H$ is a subgroup of $G$, then the quotient $|G| / |H|$ is called the *index* of $H$ in $G$.

In Order of an Element we saw that the order of an element $g$ of $G$ is equal to the order of the subgroup of $G$ generated by $g$. Thus we find:

**Corollary 6.6.13.** *If $G$ is a finite group and $g \in G$, then the order of $g$ divides $|G|$.*

*Proof.* Let $g$ be an element of the finite group $G$. Then the order of $g$ equals the number of elements in the subgroup $\langle g \rangle$ of $G$. In particular, Lagrange's Theorem implies that the order of $g$ divides $|G|$.

□

The following famous result is a second corollary to Lagrange's theorem.

**Theorem 6.6.14** (Fermat's little theorem)**.** *If $p$ is a prime number, then the multiplicative monoid $\mathbb{Z}/p\mathbb{Z} \setminus \{0\}$ is a group. So, for all $n$ not divisible by $p$, we have $n^{p-1} \equiv 1 \pmod{p}$.*

*Proof.* Since $p$ is a prime, we find $\mathbb{Z}/p\mathbb{Z}^{\times}$ to be a (multiplicative) group of order $p-1$. Hence the order of every element $x$ is a divisor of $p-1$, so that $x^{p-1} = 1$. This just says that for every $n$ which is not divisible by $p$ we have that $n^{p}$ equals 1 modulo $p$.

□

**Example 6.6.15.** The multiplicative group of $\mathbb{Z}/p\mathbb{Z}$ is actually cyclic, as we shall see later. However, there is no closed expression known for a generator of this group. The residue of the integer 2 in $\mathbb{Z}/p\mathbb{Z}$ is a generator when $p = 3$ or $p = 5$ but not when $p = 7$. In the latter case, 3 is a generator.

In general, left cosets need not coincide with right cosets. If they do, we have a case that deserves special attention. Let $G$ be a group.

**Theorem 6.6.16** (Normality)**.** *Let $H$ be a subgroup of $G$. The following assertions are equivalent.*

1. *$g \cdot H = H \cdot g$ for every $g \in G$.*
2. *$g \cdot h \cdot g^{-1} \in H$ for every $g \in G$ and $h \in H$.*

*If $H$ satisfies these properties, it is called a* normal subgroup *of $G$.*

*Proof.*

1 **implies** 2.

Suppose $g\cdot H = H\cdot g$ for every $g \in G$. Then for each $h \in H$ we have an element $h' \in H$ with $g\cdot h = h'\cdot g$. So, $g\cdot h\cdot g^{-1} = h'$, proving $g\cdot h\cdot g^{-1} \in H$.

2 **implies** 1.

Suppose for all $g \in G$ and $h \in H$ we have $g\cdot h\cdot g^{-1} \in H$. Then $g\cdot h$ can be written in the form $h'\cdot g$ for some $h' \in H$. This shows that $g\cdot H$ is contained in $H\cdot g$. Now apply this for $g^{-1}$. Hence $g^{-1}\cdot H$ is contained in $H\cdot g^{-1}$. Multiplying by $g$ from the left and by $g^{-1}$ from the right yields that $H\cdot g$ is contained in $g\cdot H$.

□

**Example 6.6.17.** Some examples of normal subgroups in familair groups:

- **Symmetric groups:**

  The alternating group $\mathrm{Alt}_n$ is a normal subgroup of $\mathrm{Sym}_n$: If $h$ is even, then $g\cdot h\cdot g^{-1}$ is an even element of $\mathrm{Sym}_n$ for each $g$.

- **Linear groups:**

  $SL(n,\mathbb{R})$ is a normal subgroup of $GL(n,\mathbb{R})$: If $\det(A) = 1$, then for every invertible matrix $B$, the product $B\cdot A\cdot B^{-1}$ has determinant 1.

- **The center of a group:** The center of a group is a normal subgroup since all its elements commute with every element in the group.

- **Commutative groups:** Suppose that $G$ is a commutative group and $H$ is a subgroup. Then for every $g \in G$ and $h \in H$, we have $g\cdot h\cdot g^{-1} = h$, so $H$ is a normal subgroup of $G$. This shows that every subgroup of a commutative group is normal.

Normal subgroups and their cosets play a special role with respect to homomorphisms.

**Theorem 6.6.18** (Normal subgroups and Kernels of homomorphisms)**.** *Let $f\colon G \to H$ be a group homomorphism.*
*Let $N$ be the kernel of $f$. Then $N$ is a normal subgroup of $G$.*
*Moreover, if $g \in G$, then the preimage of $f(g)$ is the coset $g\cdot N$ of $N$.*

*Proof.* For $g$ in $G$ and $n$ in $N$ we have

$$f\left(g\cdot n\cdot g^{-1}\right) = f(g)\cdot f(n)\cdot f\left(g^{-1}\right) = f(g)\cdot f\left(g^{-1}\right) = f(1) = 1 \tag{6.15}$$

which implies $g\cdot n\cdot g^{-1}$ to be an element of $N$. So $N$ is indeed a normal subgroup.

Now consider elements $g$ and $h$ of $G$ with $f(g) = f(h)$. Then

$$(f(g))^{-1} \cdot f(h) = f\left(g^{-1} \cdot h\right) = 1 \tag{6.16}$$

and hence, $g^{-1} \cdot h$ is in the kernel of $f$, which is $N$. But that implies that $h$ is inside the coset $g \cdot N$ of $N$.

Clearly all elements of this coset are mapped to $f(g)$ and we find the coset to be the full preimage of $f(g)$.

□

# 6.7 Exercises

## 6.7.1 Binary operations

**Exercise 6.7.1.** Show that for an associative and commutative binary operation $\cdot$ the products $((a \cdot a) \cdot ((b \cdot a) \cdot b))$ and $(a \cdot (a \cdot (b \cdot (a \cdot b))))$ are equal.

**Exercise 6.7.2.** Write an algorithm that takes as input an $n \times n$ multiplication table and that checks for associativity and commutativity of the multiplication.

## 6.7.2 Monoids and semigroups

**Exercise 6.7.3.** Prove: if in a monoid every element $x$ satisfies $x^2 = 1$, then the monoid is commutative.

Let $x$, $y$ be elements of the monoid. By hypothesis, $(x \cdot y)^2 = 1$, then the monoid is commutative.

**Exercise 6.7.4.** Let $S$ be a semigroup. We can extend $S$ with an identity element $e$, which is not in $S$, to a monoid $[S \cup \{e\}, e, \cdot_1]$. How should the multiplication on $S \cup \{e\}$ be defined in order to make this work?

What happens if $S$ contains an identity element?

**Exercise 6.7.5.** Which of the two monoids on 2 elements, $\mathbb{Z}/2\mathbb{Z}$ with addition or with multiplication, is the extension of a semigroup with an identity element?

**Exercise 6.7.6.** Show that the direct product of two monoids is again a monoid.

**Exercise 6.7.7.** Find two submonoids of $\mathbb{Z}/6\mathbb{Z}$ such that their union is not a submonoid.

**Exercise 6.7.8.** If $S_i$ is a submonoid of the monoid $M_i$ for each $i \in \{1,2\}$, then $S_1 \times S_2$ is a submonoid of $M_1 \times M_2$. Prove this.

**Exercise 6.7.9.** Suppose $a_1, b_1, a_2, b_2, ..., a_n, b_n$ are elements of $\mathbb{Q}$, and $p$ is an integer greater than $|b_1 \cdot b_2 \cdot ... \cdot b_n|$.

- Show that $\frac{1}{p}$ is not contained in the submonoid of (Q,+,0) generated by $a_1, b_1, a_2, b_2, ..., a_n, b_n$.
- Prove that $\mathbb{Q}$ is not finitely generated.

**Exercise 6.7.10.** Let $X$ be a nonempty set. If $M$ is a monoid with unit element $e$, then we can define a monoid structure on the set $F$ of all maps from $X$ to $M$ as follows.

- If $f$ and $g$ are in $F$, then their product $f \cdot g$ is defined by $f \cdot g\,(x) = f\,(x) \cdot g\,(x)$
- The constant map $x \mapsto e$ is the identity element.

Prove this.

**Exercise 6.7.11.** Let $M$ be a cyclic monoid generated by the element $c$. Suppose that $c^2 \neq e$, $c^2 \neq c^6$, and $c^4 \neq c^8$. With which cyclic monoid $C_{k,l}$ is $M$ isomorphic?

**Exercise 6.7.12.** Let $M$ be the cyclic monoid generated by $c$ and isomorphic to $C_{k,l}$. Write an algorithm that rewrites every power of $c$ to a power of $c$ whose exponent $i$ satisfies $i \leq k + l$.

**Exercise 6.7.13.** Suppose that $f = M \to N$ is a homomorphism. Prove that the image of $f$ is a submonoid of $N$ and that the kernel of $f$, i.e., $\{m \in M | f\,(m) = 1_N\}$, is a submonoid of $M$.

**Exercise 6.7.14.** Determine up to isomorphism all monoids on three elements.

**Exercise 6.7.15.** On $R$ we define the operation $*$ by $x * y = x + y - x \cdot y$.

1. Is $*$ commutative?
2. Is $*$ associative?
3. Is there an identity element in $R$ with respect to $*$?

**Exercise 6.7.16.** Consider the monoid $M$ consisting of $n$ by $n$ matrices over a ring $r$ whose multiplication is matrix multiplication. Which of the following sets are submonoids?

1. The set consisting of only the zero matrix.
2. The set consisting of only the identity matrix.
3. The set of all matrices with determinant 1.
4. The set of matrices with trace 0.
5. The set of upper triangular matrices.

**Exercise 6.7.17.** Determine, for every $m \in \{3, 4, 5\}$, the integers $k$ and $l$ such that the submonoid of $[\mathbb{Z}/m\mathbb{Z}, \cdot, 1]$ generated by 2 is isomorphic to $C_{k,l}$.

**Exercise 6.7.18.** Prove that the monoid $[\mathbb{Z}/8\mathbb{Z}, \cdot, 1]$ cannot be generated by less than 3 elements.

Prove that it can be generated by 3 elements.

**Exercise 6.7.19.** Let $[M, \cdot, 1]$ be a monoid. Define a new multiplication $\cdot$ on $M$ by $x \cdot y = x \cdot y$. Prove that $[M, \cdot, 1]$ is also a monoid.

### 6.7.3 Invertibility in monoids

**Exercise 6.7.20.** Prove that in the monoid $[\mathbb{Z}/n\mathbb{Z}, \cdot, 1]$ an element$m$ has an inverse if and only if $\gcd(m,n) = 1$.

**Exercise 6.7.21.** Let $M_1$, $M_2$ be monoids. Prove that the invertible elements of $M_1 \times M_2$ are of the form $(m_1, m_2)$ with $m_1$ invertible in $M_1$ and $m_2$ invertible in $M_2$.

**Exercise 6.7.22.** What are the invertible elements of $C_{k,l}$?

**Exercise 6.7.23** (Exercise 32)**.** Determine the invertible elements of the following monoids.

1. $[\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \cdot, 1]$.
2. The multiplicative monoid of $\mathbb{Q}[X]/(X^2)\mathbb{Q}[X]$.
3. The multiplicative monoid of $\mathbb{Z}/16\mathbb{Z}$.

**Exercise 6.7.24.** Consider the group $G$ of invertible elements in the multiplicative monoid of $\mathbb{Z}/26\mathbb{Z}$.

1. How many elements does $G$ have?
2. $G$ is cyclic. Find all possible single generators.

### 6.7.4 Groups

**Exercise 6.7.25.** Is the following true? If $G$ is a group of order $n$, and $m$ is a positive divisor of $n$ strictly smaller than $n$, then $G$ contains an element of order $m$.

**Exercise 6.7.26.** Let $G$ be a finite group. Show that each element of $G$ appears exactly once in each column and each row of the multiplication table (also called Cayley-table) of $G$.

**Exercise 6.7.27.** Let $I$ be the identity matrix of size $n$, i.e., the $n$ by $n$ matrix with ones on the diagonal and zeros outside the diagonal. For any matrix $A$ we denote by $A^T$ the transposed matrix of $A$.

Let $R$ be a commutative ring. Prove that the set $O(n,R) = \{x \in GL(n,R) | A \cdot A^T = I\}$ is a subgroup of $GL(n,R)$.

**Exercise 6.7.28.** Prove that the groups $C_2 \times C_3$ and $C_6$ are isomorphic.

Show that these two groups are not isomorphic to $\mathrm{Sym}_3$.

**Exercise 6.7.29.**

1. Show that the map $f = \mathbb{Z} \to \mathbb{Z}, (x,y) \mapsto x - 2 \cdot y$ is a morphism of the additive groups. What is the image of this homomorphism?

2. Let $G$ be a group and $g$ an element of $G$. Prove that the map $f = \mathbb{Z} \to G, k \mapsto g^{2 \cdot k}$ is a homomorphism of groups. What is the image of $f$ if the order of $g$ equals 6 or 7, respectively?

3. Determine all homomorphisms of the additive group $\mathbb{Z}/4\mathbb{Z}$ to itself. Which of these are isomorphisms?

4. If $f = G \to K$ and $h = K \to L$ are homomorphisms of groups, then the composition $h \circ f = G \to L$ is also a homomorphism of groups. Prove this. Deduce furthermore that if $G$ is isomorphic with $K$ and $K$ with $L$, then $G$ is isomorphic with $L$.

### 6.7.5 Cyclic groups

**Exercise 6.7.30.** Determine the order of the element $(1,2)(3,4,5)$ in $\mathrm{Sym}_5$.

Prove that, in general, the order of a permutation equals the least common multiple of the cycle lengths occurring in a disjoint cycle decomposition.

**Exercise 6.7.31.** Let $G$ be a group and $H$ a nonempty finite subset of $G$ closed under multiplication. Prove the following statements.

1. For $h$ in $H$, the elements $h^1, h^{2\mathrm{n}}, h^3, \cdots$ are not all distinct.

2. The identity element belongs to $H$.

3. Every element of $H$ has finite order.

4. $H$ is a subgroup of $G$.

**Exercise 6.7.32.** Let $G$ be a finite group of order $m$ Let $g$ be in $G$. Suppose that for each prime divisor $p$ of $m$ the element $g^{\frac{m}{p}}$ is not the identity. Prove that the group $G$ is generated by $g$.

**Exercise 6.7.33.** Let $G$ be a cyclic group with generator $g$.

1. Show that the map $f = \mathbb{Z} \to G, k \mapsto g^k$ is a homomorphism of groups.

2. Suppose that $G$ has order $n$. Show that the map $f = \mathbb{Z}/m\mathbb{Z} \to G, k \mapsto g^k$ is well defined and is an isomorphism of groups.

**Exercise 6.7.34.** Consider the additive group $G = \mathbb{Z} \times \mathbb{Z}$.

1. Prove that this group is not cyclic, but can be generated by the elements $(2,3)$ and $(3,4)$.

2. Prove that $(a,b)$ and $(c,d)$ generate the group if and only if $a\cdot d - b\cdot c \in \{1,-1\}$.

**Exercise 6.7.35.** Let $G$ be a group of order 4. Prove the following statements.

1. If $G$ contains an element of order 4, then $G$ is cyclic and isomorphic to $C_4$.
2. If $G$ contains no element of order 4, then $G$ is commutative and isomorphic to $C_2 \times C_2$.

**Exercise 6.7.36.** Let $p$ be a prime. Then the multiplicative group $\mathbb{Z}/p\mathbb{Z}^\times$ is cyclic. (This will be proved in [?].) Write an algorithm that determines a generator for $\mathbb{Z}/p\mathbb{Z}^\times$. Determine all odd primes $p$ less than 10.000 such that 2 is a generator for this group. (It is a conjecture of Artin that there are infinitely many primes $p$ for which 2 generates the group $\mathbb{Z}/p\mathbb{Z}^\times$. Although very likely, as of April, 2011, it is not known to be true.)

### 6.7.6 Cosets

**Exercise 6.7.37.** Determine the left and right cosets of $\mathrm{Sym}_3$ in $\mathrm{Sym}_4$.

**Exercise 6.7.38.** The kernel of a group homomorphism is a normal subgroup as follows from Normal subgroups and Kernels of homomorphisms. Can you provide such a homomorphism for the normal subgroups discussed in Example 6.6.17?

**Exercise 6.7.39.** Suppose $G$ and $H$ are finite groups admitting a surjective homomorphism from $G$ to $H$ .

Show the the order of $H$ divides the order of $G$.

**Exercise 6.7.40.** Suppose $G$ is a group and $H$ a subgroup of index 2. Prove that $H$ is normal in $G$.

Is the same true if the index equals 3? Give a proof or a counterexample.

**Exercise 6.7.41.** Suppose $G$ is a group and $H$ a subgroup.

If $K$ is a subgroup of $G$ normalising $H$, i.e., contained in the normaliser of $H$, then $H\cdot K$ is a subgroup of $G$. Prove this.

# Chapter 7

# Rings and fields

We continue the study of structures. Having dealt with two basic examples, monoids and groups, we now focus on two structures in which they play a significant role: rings and fields.

## 7.1 The structure ring

Multiplication turns each of the sets $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C},\mathbb{Z}[X],\mathbb{Q}[X],\mathbb{R}[X],\mathbb{C}[X]$ into monoids, whereas addition defines a group structure. These two structures are combined in the notion of a ring.

**Definition 7.1.1.** A *ring* is a structure $[R,+,0,-,\cdot,1]$ consisting of a set $R$ for which $[R,+,0,-]$ is a commutative group and $[R,\cdot,1]$ is a monoid, in such a way that the following laws hold for all $x,y,z\in R$:

- $x\cdot(y+z)=x\cdot y+x\cdot z$ (left distributivity);
- $(y+z)\cdot x=y\cdot x+z\cdot x$ (right distributivity).

The ring is called *commutative* if the monoid $[R,\cdot,1]$ is commutative.

**Example 7.1.2** (Usual arithmetic)**.** Each of $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C}$, with the usual addition and multiplication, is a commutative ring.

**Example 7.1.3** (Modular arithmetic)**.** Addition and multiplication as defined in the Modular Addition and Multiplication Theorem determine a commutative ring structure on $\mathbb{Z}/n\mathbb{Z}$. The zero element is the class of 0, the identity element is the class of 1.

**Example 7.1.4** (Polynomial rings)**.** Let $R$ be one of the rings $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C}$ or $\mathbb{Z}/n\mathbb{Z}$. Then $R[X]$, with the usual addition and multiplication is a commutative ring.

**Example 7.1.5** (Residue class rings)**.** If $R$ is a commutative ring as in Example 7.1.4 and $f$ is a polynomial in $R[X]$, then $R[X]/(f)R[X]$, as defined in Operations modulo a Polynomial, is a commutative ring. The zero element is $0+(f)R[X]$, the identity element is $1+(f)R[X]$.

**Example 7.1.6** (The Gaussian integers)**.** The subset $R=\mathbb{Z}+\mathbb{Z}\cdot i$ of the complex numbers is a ring with the usual addition and multiplication, with zero element $0=0+0\cdot i$ and identity element $1=1+0\cdot i$. Most ring properties, like associativity of the multiplication, are 'inherited' from the ring $\mathbb{C}$: since they hold in the complex numbers they hold a fortiori in the subset $R$. A crucial issue for $R$ to be a ring, is that $R$ is closed with respect to the operations. For instance, $(a+b\cdot i)\cdot(c+d\cdot i)=a\cdot c-b\cdot d+(a\cdot d+b\cdot c)\cdot i$ shows that the set $R$ is closed with respect to multiplication, since $a\cdot c-b\cdot d$ and $a\cdot d+b\cdot c$ are integers if $a,b,c,d$ are. The ring $R$ is called the ring of *Gaussian integers*.

**Example 7.1.7** (Matrix rings)**.** Let $R$ be a ring. Then the following structure is a ring: $S=[M_n(R),+,0,-,\cdot,1]$, where $M_n(R)$ is the set of $n$ by $n$ matrices with coefficients in $R$, where 0 is short for the zero matrix, 1 is short for the identity matrix, $+$ denotes matrix addition and $\cdot$ denotes matrix multiplication. If $n>1$, it is easy, and left to the reader, to find matrices $A,B$ such that $A\cdot B$ and $B\cdot A$ are distinct. Thus, $S$ is not commutative for $n>1$ even if $R$ is commutative.

**Example 7.1.8** (The quaternions)**.** Take $1,i,j,k$ to be a set of four vectors (think of a standard basis) of the 4-dimensional real vector space $\mathbb{H}=\mathbb{R}\cdot 1+\mathbb{R}\cdot i+\mathbb{R}\cdot j+\mathbb{R}\cdot k$. On $\mathbb{H}$ we define the operations $+$ and $\cdot$ as follows. For $x=a\cdot 1+b\cdot i+c\cdot j+d\cdot k$ and $x'=a'\cdot 1+b'\cdot i+c'\cdot j+d'\cdot k$ let $x+x'$ be the vector sum of $x$ and $x'$ and set $x\cdot x'=p\cdot 1+q\cdot i+r\cdot j+s\cdot k$ where $p=a\cdot a'-b\cdot b'-c\cdot c'-d\cdot d'$, $q=a\cdot b'+b\cdot a'+c\cdot d'-d\cdot c'$, $r=a\cdot c'-b\cdot d'+c\cdot a'+d\cdot b'$, and $s=a\cdot d'+b\cdot c'-c\cdot b'+d\cdot a'$. Now $\mathbb{H}$ is a ring. (It is quite tedious to check associativity, etc.) Since $i\cdot j=k=-j\cdot i$, the ring is not commutative. The ring $\mathbb{H}$ is called the *ring of real quaternions*.

**Example 7.1.9.** Here is an application of the ring of the Gaussian integers $R=\mathbb{Z}+\mathbb{Z}\cdot i$. Suppose that the integers $k$ and $l$ can both be written as sums of two squares of integers: $k=a^2+b^2$ and $l=c^2+d^2$. Then the product $k\cdot l$ is also a sum of squares. You may find it hard to show this from scratch. Here is how the ring $R$ comes into play: $k=(a+b\cdot i)\cdot(a-b\cdot i)$ and $l=(c+d\cdot i)\cdot(c-d\cdot i)$ so $k\cdot l=(a+b\cdot i)\cdot(c+d\cdot i)\cdot(a-b\cdot i)\cdot(c-d\cdot i)$. Expanding the product of the first two factors gives $a\cdot c-b\cdot d+(a\cdot d+b\cdot c)\cdot i$ and the product of the last two factors is $a\cdot c-b\cdot d-(a\cdot d+b\cdot c)\cdot i$. This yields $k\cdot l=(a\cdot c-b\cdot d)^2+(a\cdot d+b\cdot c)^2$.

**Example 7.1.10.** An argument, similar to the one in Example 7.1.9, using the quaternions can be used to show that if two integers can be written as sums of four squares of integers, then so can their product. The equality $(a+b\cdot i+c\cdot j+d\cdot k)\cdot(a-b\cdot i-c\cdot j-d\cdot k)=a^2+b^2+c^2+d^2$ plays a role in the proof.

Notation and terminology for a ring $[R,+,0,-,\cdot,1]$:

- $+$ is called the addition,
- $\cdot$ is called the multiplication (the symbol $\cdot$ is often omitted),
- 0 is the zero element,
- 1 is the identity element of the ring.

**Definition 7.1.11.** A *subring* of a ring $[R,+,0,-,\cdot,1]$ is a subset $S$ of $R$ containing $0$ and $1$ such that, whenever $x,y\in S$, we have $x+y$, $-x$ and $x\cdot y\in S$.

In other words, a subring is a subset $S$ of $R$ closed under all operations of $R$.

**Example 7.1.12** (Usual arithmetic)**.** The ring $\mathbb{Z}$ is a subring of each of the rings $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C}$. It is the smallest possible subring.

The ring $\mathbb{Q}$ is a subring of $\mathbb{R}$ and of $\mathbb{C}$.

The ring $\mathbb{R}$ is a subring of $\mathbb{C}$.

**Example 7.1.13** (Modular arithmetic)**.** Let $m$ be an integer, $m>1$. Suppose that $S$ is a subring of $\mathbb{Z}/m\mathbb{Z}$. Then $S$ contains $1$, hence each integer multiple of $1$, hence the whole ring $\mathbb{Z}/m\mathbb{Z}$. Therefore, the only subring of $\mathbb{Z}/m\mathbb{Z}$ is the ring itself. In other words, there are no proper subrings.

**Example 7.1.14** (Polynomial rings)**.** The coefficient ring $R$ is a subring of $R[X]$.

Also, the polynomials in which only even powers of $X$ occur, form a subring of $R[X]$.

**Example 7.1.15** (Residue class rings)**.** Let $R$ be a commutative ring and let $f$ be a monic polynomial in $R[X]$ (so its leading coefficient is equal to one). If the degree of $f$ is positive, then, by a Lemma on Coefficient Ring, $R$ is a subring of $R[X]/(f)$.

**Example 7.1.16** (The Gaussian integers)**.** The ring $\mathbb{Z}$ is a subring of $\mathbb{R}=\mathbb{Z}+\mathbb{Z}\cdot i$ .

**Example 7.1.17** (Matrix rings)**.** The upper triangular matrices form a subring of $M_n(R)$.

**Example 7.1.18** (The Quaternions)**.** The subset $\mathbb{R}\cdot 1$ of $\mathbb{H}$ is a subring. In fact $1$ is the identity element, and the ring is just a copy of $\mathbb{R}$. The symbol $1$ is often left out from $\mathbb{R}\cdot 1$ so as to interpret $\mathbb{R}$ as a subring of $\mathbb{H}$.

Also $\mathbb{R}+\mathbb{R}\cdot i$ is a subring, and so are $\mathbb{R}+\mathbb{R}\cdot j$ and $\mathbb{R}+\mathbb{R}\cdot k$.

**Remark 7.1.19.** For the set $S$ to be a subring of a given ring it suffices that $0,1,x-y$, and $x\cdot y$ are in $S$ for all $x\in S$ and $y\in S$. Indeed, then $0-x=-x$ is also in $S$, and similarly for $x+y=x-(-y)$. A subring, supplied with the restrictions of all operations of the ambient ring $R$, is itself a ring.

For instance, as $x\cdot(y+z)=x\cdot y+x\cdot z$ holds for all $x,y,z\in R$, it also holds for all elements in the subset $S$ of $R$.

In other words, $[S,+,0,-]$ and $[S,\cdot,1]$, where $+$ and $\cdot$ are the restrictions to $S\times S$, are again a group and a monoid, respectively.

Let $R$ be a ring. Addition defines a group structure on $R$. So every element $a$ has an inverse with respect to the addition. This inverse is denoted $-a$ and is called the *opposite* of $a$.

**Theorem 7.1.20.** *The following properties hold for all $a,b \in R$.*

1. $a\cdot 0 = 0\cdot a = 0$;
2. $a\cdot(-b) = -(a\cdot b) = (-a)\cdot b$;
3. $(-a)\cdot(-b) = a\cdot b$;
4. $(-1)\cdot a = -a$.

*Proof.* We prove each of the fours parts individually.

$a\cdot 0 = 0\cdot a = 0$**.**

By left distributivity and the role of the zero element we can write $a\cdot 0 + a\cdot 0 = a\cdot(0+0) = a\cdot 0 = a\cdot 0 + 0$ so that $a\cdot 0 + a\cdot 0 = a\cdot 0 + 0$. The Cancellation law for groups allows us to conclude: $a\cdot 0 = 0$, as required. Similarly one shows $0\cdot a = 0$.

$a\cdot(-b) = -(a\cdot b) = (-a)\cdot b$**.**

Clearly, $a\cdot(b-b) = a\cdot 0 = 0$. Using distributivity to expand the left-hand side, we find: $a\cdot b + a\cdot(-b) = 0$, from which we derive $-a\cdot b = a\cdot(-b)$.

The other equality is proved similarly.

$(-a)\cdot(-b) = a\cdot b$**.**

By the previous assertion, $(-a)\cdot(-b) = -(-a)\cdot b$, so $(-a)\cdot(-b)$ is the inverse of $(-a)\cdot b$. But from the previous part we also conclude that $a\cdot b$ is the inverse of $(-a)\cdot b$. Since (additive) inverses are unique, we are done.

$(-1)\cdot a = -a$**.**

By distributivity, $a + (-1)\cdot a = 1\cdot a + (-1)\cdot a = (1+-1)\cdot a = 0\cdot a = 0$. Therefore $(-1)\cdot a$ is the additive inverse of $a$, i.e., $-a = (-1)\cdot a$.

□

**Example 7.1.21.** The ring laws lead to rules for calculations which are familiar from the usual examples.

For instance, if, in a product, one factor is 0, then the whole product is 0.

Another example: $(-(a_1))\cdot(-(a_2))\cdot\ldots\cdot(-(a_n)) = (-1)^n\cdot a_1\cdot a_2\cdot\ldots\cdot a_n$.

Recall that a ring $R$ is a monoid with respect to multiplication. It is not necessarily the case that every (nonzero) element has an inverse with respect to multiplication. Those elements of $R$ that do have an inverse are called the *invertible* elements of $R$. The inverse of $a$ in $R$ is denoted by $a^{-1}$.

**Theorem 7.1.22.** *The invertible elements of R form a multiplicative group (i.e., a group with respect to multiplication). This group is denoted by* $R^{\times}$.

*Proof.* This is a direct consequence of Arithmetic Properties in Rings.

□

**Example 7.1.23** (Usual arithmetic)**.** $\mathbb{Z}^{\times} = \{1, -1\}$. Every nonzero element of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ is invertible. (In other words, these rings are fields.)

**Example 7.1.24** (Modular arithmetic)**.** $\mathbb{Z}/n\mathbb{Z}^{\times}$ consists of the classes $m + n\cdot\mathbb{Z}$ of $\mathbb{Z}/n\mathbb{Z}$ for which $m$ is an integer such that $\gcd(m,n) = 1$.

**Example 7.1.25** (Polynomials rings)**.** $\mathbb{Z}[X]^{\times} = \{1, -1\}$ and $\mathbb{Q}[X]^{\times} = \mathbb{Q} \setminus \{0\}$. Similarly for $\mathbb{R}$ and $\mathbb{C}$. To prove these statements you will need to involve the degree. We leave this to the reader.

**Example 7.1.26** (Residue class rings)**.** If $R = \mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, and $f$ is a polynomial in $R[X]$ of positive degree, then $R[X]/(f)R[X]^{\times}$ consists of the residue classes of those polynomials $g$ in $R[X]$ for which $\gcd(g, f) = 1$.

If $R = \mathbb{Z}$, then it is harder to describe the invertible elements of $R[X]/(f)R[X]$ for general $f$.

**Example 7.1.27** (Gaussian integers)**.** $(\mathbb{Z} + (\mathbb{Z}\cdot i))^{\times} = \{1, -1, i, -i\}$. If $a + b\cdot i$ is invertible, then there exists an element $c + d\cdot i$ such that $(a + b\cdot i)\cdot(c + d\cdot i) = 1$. Using the property $|z|\cdot|w| = |z\cdot w|$ for the absolute value of complex numbers, we infer that $\left(a^2 + b^2\right)\cdot\left(c^2 + d^2\right) = 1$. Since $a, b, c, d$ are integers, we find that the integer $a^2 + b^2$ divides 1. The conclusion is that $a + b\cdot i$ must be one of the four elements $1, -1, i, -i$, as stated.

**Example 7.1.28** (Matrix rings)**.** Let $R$ be a commutative ring. The invertible elements of $M_n(R)$ are those matrices whose determinant is invertible in $R$. This follows from Cramer's rule, which expresses the inverse of a matrix in terms of minors (elements of $R$) and the inverse of the determinant.

**Example 7.1.29** (The Quaternions)**.** $\mathbb{H}^{\times} = \{x \in \mathbb{H} | N(x) \in \mathbb{R}^{\times}\}$ where $N(x) = a^2 + b^2 + c^2 + d^2$ if $x = a\cdot 1 + b\cdot i + c\cdot j + d\cdot k$. As for the proof: write $C(x) = a\cdot 1 - (b\cdot i + c\cdot j - d\cdot k)$. Then $C(x)\cdot x = N(x)$ So, if $N(x)$ is invertible, then $(N(x))^{-1}\cdot C(x)$ is the inverse of $x$.

Also, if $N(x)$ is not invertible, it is zero and so $x$ is zero or a zero divisor; in particular, it is not invertible.

Let $R$ and $R'$ be rings.

**Definition 7.1.30.** A map $f\colon R \to R'$ is called a *(ring) homomorphism* from $R$ to $R'$ if $f$ is

- a homomorphism of additive groups $R \to R$, and
- a homomorphism $R \to R$ of multiplicative monoids.

Let $f\colon R \to R'$ be a homomorphism.

- The *kernel* of $f$ is the set $\{a \in R | f(a) = 0\}$. It is denoted by $\mathrm{Ker}(f)$.
- The *image* of $f$ is the set $f(R) = \{f(a) | a \in R\}$. It is denoted by $\mathrm{Im}(f)$.

**Example 7.1.31** (Left multiplication by an integer)**.** The map $f\colon \mathbb{Z} \to \mathbb{Z}$ given by $f(a) = 3\cdot a$ is not a homomorphism of rings because $f(1)$ is not equal to 1.

**Example 7.1.32** (Modding out an integer)**.** The map $f\colon \mathbb{Z} \to \mathbb{Z}/6\mathbb{Z}$ given by $f(a) = a + 6\cdot\mathbb{Z}$ is a homomorphism of rings:

- $f(0) = 0 + 6\cdot\mathbb{Z}$,
- $f(1) = 1 + 6\cdot\mathbb{Z}$,
- $f(a+b) = a + b + 6\cdot\mathbb{Z} = a + 6\cdot\mathbb{Z} + (b + 6\cdot\mathbb{Z}) = f(a) + f(b)$, and
- $f(a\cdot b) = a\cdot b + 6\cdot\mathbb{Z} = (a + 6\cdot\mathbb{Z})\cdot(b + 6\cdot\mathbb{Z}) = f(a)\cdot f(b)$.

The kernel of the map is exactly $6\cdot\mathbb{Z}$.

**Example 7.1.33** (Complex conjugation)**.** Complex conjugation is a homomorphism of rings $\mathbb{C} \to \mathbb{C}$. In fact, it is an isomorphism: It is its own inverse.

**Example 7.1.34** (Modding out a divisor)**.** The map $f\colon \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ given by $f(a + 6\cdot\mathbb{Z}) = a + 2\cdot\mathbb{Z}$ is a homomorphism of rings. The kernel consists of all $a + 6\cdot\mathbb{Z}$ for $a \in \mathbb{Z}$ such that $a = \mathrm{rem}(0,2)$. The kernel is therefore $\{0,2,4\}$. It is not hard to verify that $f$ is surjective.

**Example 7.1.35** (Modding out a polynomial)**.** The homomorphism $f\colon \mathbb{Q}[X] \to \mathbb{Q}[X]/(X^2)\mathbb{Q}[X]$ which sends a polynomial to its class modulo $X^2$ is a homomorphism. This is easily verified. The kernel of this homomorphism consists of all polynomials that are divisible by $X^2$. The homomorphism is surjective, so the image is the whole ring $\mathbb{Q}[X]/(X^2)\mathbb{Q}[X]$.

**Example 7.1.36** (Gaussian numbers as polynomial residues)**.** The map $f\colon \mathbb{Q}[X]/(X^2+1)\mathbb{Q}[X]$ where $i$ is the usual imaginary number (square root of $-1$), is defined by $g + (X^2+1)\mathbb{Q}[X] \mapsto g(i)$, for every residue class $g + (X^2+1)\mathbb{Q}[X] \in \mathbb{Q}[X]/(X^2+1)\mathbb{Q}[X]$. Observe that it is indeed well defined. This follows from the fact that if $g - h$ is divisible by $X^2+1$, then $g(i) = h(i)$. So $g + (X^2+1)\mathbb{Q}[X] = h + (X^2+1)\mathbb{Q}[X]$ implies $g(i) = h(i)$. This map is in fact an isomorphism. The inverse map is given by $a + b\cdot i \mapsto a + b\cdot X + (X^2+1)R$, as can be easily checked.

**Example 7.1.37** (From quaternions to matrices)**.** The map $f\colon \mathbb{H} \to M_2(\mathbb{C})$ given by $a + b\cdot i + c\cdot j + d\cdot k \mapsto \begin{pmatrix} a+b\cdot i & c+d\cdot i \\ -c+d\cdot i & a-b\cdot i \end{pmatrix}$ is a homomorphism of noncommutative rings. Its kernel is $\{0\}$.

**Example 7.1.38** (Subrings)**.** If $S$ is a subring of the ring $R$, then the map $f\colon S\to R, a\mapsto a$ is a homomorphism, the so-called *inclusion map*. It is usually convenient to view the inclusion of rings as a homomorphism in this way.

Apparently, a map $f\colon R\to R$ is a homomorphism if the following conditions are satisfied.

- $f(0)=0$;
- $f(a+b)=f(a)+f(b)$;
- $f(1)=1$;
- $f(a\cdot b)=f(a)\cdot f(b)$.

The first of these four conditions follows directly from the second.

**Theorem 7.1.39.** *Let $R$ and $S$ be rings. For every homomorphism $f\colon R\to S$ the following holds.*

1. *If $a\in R$ is invertible, then so is $f(a)$, in which case its inverse is $f\left(a^{-1}\right)$.*
2. *If $f$ is an isomorphism, then so is its inverse $f^{-1}\colon S\to R$.*
3. *The image* $\mathrm{Im}(f)$ *is a subring of $S$.*
4. *The kernel* $\mathrm{Ker}(f)$ *is an additive subgroup of $R$. If $a$ is in* $\mathrm{Ker}(f)$*, then $r\cdot a$ is in* $\mathrm{Ker}(f)$ *for all $r\in R$.*
5. *The map $f$ is injective if and only if* $\mathrm{Ker}(f)=0$.

*Proof.*

**If $a\in R$ is invertible, then so is $f(a)$, in which case its inverse is $f\left(a^{-1}\right)$.**

If $a$ is invertible with inverse $b$, then $a\cdot b=1$. Applying $f$ gives $f(a\cdot b)=f(1)=1$ and so $f(a)\cdot f(b)=1$, i.e., $f(a)$ is invertible and its inverse is $f(b)=f\left(a^{-1}\right)$.

**If $f$ is an isomorphism, then so is its inverse $f^{-1}\colon R\to S$.**

Suppose that $f$ is a bijection from $R$ to $S$ . Then we show that $f^{-1}$ respects multiplication and leave other details to the reader. Suppose $a',b'\in S$. As $f$ is surjective, there are $a,b\in R$ such that $f(a)=a'$ and $f(b)=b'$. The fact that $f$ is a homomorphism implies $f(a\cdot b)=f(a)\cdot f(b)$. Applying $f^{-1}$ to both sides gives $a\cdot b=f^{-1}(f(a)\cdot f(b))$. Substituting $f^{-1}(a')$ for $a$ and $f^{-1}(b')$ for $b$ in the left-hand side, and $a'$ for $f(a)$ and $b'$ for $f(b)$ in the right-hand side, we find $f^{-1}(a')\cdot f^{-1}(b')=f^{-1}(a'\cdot b')$, as required.

**$f(R)$ is a subring of $S$.**

This is direct from the conditions given before the theorem.

**$\mathrm{Ker}(f)$ is an additive subgroup of $R$. If $a$ is in $\mathrm{Ker}(f)$, then $r \cdot a$ is in $\mathrm{Ker}(f)$ for all $r \in R$.**

$\mathrm{Ker}(f)$ is an additive subgroup of $R$. This follows directly from the conditions given before the theorem. Let $r \in R$ and $a \in \mathrm{Ker}(f)$. Then $f(r \cdot a) = f(r) \cdot f(a) = f(r) \cdot 0 = 0$, whence $r \cdot a \in \mathrm{Ker}(f)$.

**$f$ is injective if and only if $\mathrm{Ker}(f) = 0$.**

If $f$ is injective and $a$ belongs to $\mathrm{Ker}(f)$, then $f(a) = 0 = f(0)$, and injectivity implies $a = 0$. Conversely, if $\mathrm{Ker}(f) = \{0\}$, and $f(a) = f(b)$, then $f(a-b) = 0$ so that $a - b = 0$ and $a = b$.

□

**Example 7.1.40** (The identity)**.** For any ring $R$, the identity map $R \to R$ is an isomorphism, which is its own inverse.

**Example 7.1.41** (Modding out an integer)**.** The homomorphism $f\colon \mathbb{Z} \to \mathbb{Z}/6\mathbb{Z}$ given by $f(a) = a + 6 \cdot \mathbb{Z}$ is not injective as its kernel is $6 \cdot \mathbb{Z}$. The invertible elements of $\mathbb{Z}$ are $1, -1$ ; they are mapped onto the invertible elements of $\mathbb{Z}/6\mathbb{Z}$.

**Example 7.1.42** (Complex conjugation)**.** Complex conjugation is an isomorphism of rings $\mathbb{C} \to \mathbb{C}$.

**Example 7.1.43** (Modding out a polynomial)**.** The homomorphism $f\colon \mathbb{Q}[X] \to \mathbb{Q}[X]/(X^2)\mathbb{Q}[X]$ is surjective, but not injective. The residue class of $1 + X$ is invertible in $\mathbb{Q}[X]/(X^2)\mathbb{Q}[X]$ (its inverse is the class of $1 - X$ ), but its inverse image does not contain an invertible element in $\mathbb{Q}$.

**Example 7.1.44** (Gaussian numbers as polynomial residues)**.** The map $f\colon \mathbb{Q}[X]/(X^2+1)\mathbb{Q}[X]$ where $i$ is the usual imaginary number (square root of $-1$), is defined by $g + (X^2+1)\mathbb{Q}[X] \mapsto g(i)$, for every residue class $g + \left(X^2+1\right) \in R[X]/(X^2+1)R[X]$. This map is an isomorphism. It demonstrates that two completely different looking rings may nevertheless carry the same ring structure.

## 7.2 Constructions with rings

Let $[R, +, 0, -, \cdot, 1]$ and $[R', +, 0', -, \cdot, 1']$ be rings. Just like the product of two monoids (respectively, groups) is a monoid (respectively, group), the product of two rings is a ring.

**Theorem 7.2.1.** *The direct product $R \times R'$ with coordinatewise addition and multiplication and with zero element $(0, 0')$ and identity $(1, 1')$ is a ring.*

*Proof.* The proof is a routine verification. Here are the different parts.

$[R\times R',\cdot,(1,1')]$ **is a monoid.**

Application of the direct product construction for monoids.

$[R\times R',+,(0,0'),-]$**, where the binary operation** $-$ **is defined coordinatewise, is a commutative group.**

The structure is a group by application of Direct Product of Groups.

Observe that the direct product of two commutative groups is again commutative: $(a,a')+(b,b')=(a+b,a'+b')=(b+a,b'+a')=(b,b')+(a,a')$

**The structure is left distributive.**

By Cartesian addition, Cartesian multiplication, and left distributivity for $R$ and $R'$, we have

$$\begin{aligned}(a,a')\cdot((b,b')+(c,c'))=(a,a')\cdot(b+c,b'+c')=\\(a\cdot(b+c),a'\cdot(b'+c'))=(a\cdot b+a\cdot c,a'\cdot b'+a'\cdot c')=\\(a\cdot b,a'\cdot b')+(a\cdot c,a'\cdot c')=\\(a,a')\cdot(b,b')+(a,a')\cdot(c,c').\end{aligned}.$$

**Right distributivity.**

Just like left distributivity.

□

**Example 7.2.2** (Sample computation)**.** In the product $\mathbb{Z}\times\mathbb{Z}/6\mathbb{Z}$ we have $(3,4)\cdot(2,3)=(6,0)$ and $(3,4)+(2,3)=(5,1)$.

**Remark 7.2.3** (Multiple direct products)**.** The process of taking direct products can be repeated to obtain rings like $R\times S\times T$, or the $n$-fold product of a ring with itself: $R^n=R\times R\times\ldots\times R$ ($n$ factors). There is of course the question whether, say $R\times S\times T$ and $R\times S\times T$ yield the same result. The answer is 'yes' in the sense that they are isomorphic.

The ring is called the direct product of $R$ and $S$ and often denoted by $R\times S$ (instead of the full information with multiplication, addition, zero, and unit).

**Example 7.2.4** (Chinese Remainder Theorem)**.** The Chinese Remainder Theorem can be nicely phrased in terms of direct products: If $m$ and $n$ are positive integers greater than 1 with $\gcd(m,n)=1$, then $\mathbb{Z}/(m\cdot n)\mathbb{Z}$ is isomorphic with $\mathbb{Z}/m\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}$; the isomorphism is given by the map $a(\mathrm{mod}\ m\cdot n)\mapsto(a(\mathrm{mod}\ m),a(\mathrm{mod}\ n))$.

Hence, given an element $x=(b(\mathrm{mod}\ m),c(\mathrm{mod}\ n))$ in $\mathbb{Z}/m\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}$ there is a unique element in $\mathbb{Z}/(m\cdot n)\mathbb{Z}$ that is mapped onto $x$.

**Theorem 7.2.5.** $R\times S^\times=R^\times\times S^\times$.

*Proof.*

**If $a \in R$ has inverse $b$ and $a' \in R'$ has inverse $b'$, then $(a,a')$ has inverse $(b,b')$.**

$(a,a') \cdot (b,b') = (a \cdot b, a' \cdot b') = (1,1)$, and similarly for $(b,b') \cdot (a,a')$.

**Conversely, if $(a,a')$ has inverse $(c,c')$, then $a$ has inverse $c$ and $a'$ has inverse $c'$.**

By the same kind of equalities as in the proof of the previous assertion.

□

**Example 7.2.6.** In the direct product $\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$, the invertible elements are $(1,1)$, $(1,5)$, $(-1,1)$, and $(-1,5)$, i.e., all elements in which both the first and the second coordinate are invertible.

**Example 7.2.7.** The multiplicative formula for the Euler Totient Function, see Euler Totient Theorem, can be explained by a combination of the Chinese Remainder Theorem and the Invertible Elements in Direct Products of Rings with $R = \mathbb{Z}/m\mathbb{Z}$ and $S = \mathbb{Z}/n\mathbb{Z}$ for positive integers $m$ and $n$ such that $\gcd(m,n) = 1$. We have

$$\Phi(m \cdot n) = \left|\mathbb{Z}/m \cdot n\mathbb{Z}^{\times}\right| = \left|\mathbb{Z}/m\mathbb{Z}^{\times} \times \mathbb{Z}/n\mathbb{Z}^{\times}\right| = \left|\mathbb{Z}/m\mathbb{Z}^{\times}\right| \cdot \left|\mathbb{Z}/n\mathbb{Z}^{\times}\right| = \Phi(m) \cdot \Phi(n) \quad (7.1)$$

The notion of generators, known for monoids, is similar for rings.

**Theorem 7.2.8.** *Let $R$ be a ring.*
*If $C$ is a collection of subrings of $R$, then $\bigcap_{c \in C} c$ is also a subring of $R$.*

*Proof.* Let $S$ denote the intersection $\bigcap_{c \in C} c$ of which we must prove that it is a subring of $\mathbb{R}$. We verify the conditions for $S$ to be a subring.

$0, 1 \in S$**.**

Since each $H \in C$ is a subring, we have $0, 1 \in H$. Hence $0, 1$ belong to the intersection over all $H$, that is, to $S$.

**The set $S$ is closed under multiplication and addition.**

Suppose $a, b \in S$. Then, for each $H \in S$, we have $a, b \in H$, whence (as $H$ is a subring) $a + b \in H$. It follows that $a + b \in S$.

The proof for multiplication is very similar and therefore omitted.

□

**Remark 7.2.9.** • If $C$ is the empty collection, the intersection over $C$ is taken to be $R$.

- The theorem is the analog for rings of the result for monoids. In fact, the result holds for any structure. The proof remains basically the same: if each substructure of a collection is closed under all operations, then so is the intersection. For this reason, when dealing with fields later on, we shall not treat the result any more as a separate theorem.

Intersection of subrings is a subring shows that the smallest subring containing a given set $D$ exists: it is the intersection of all subring containing $D$. Therefore, the definition below makes sense.

**Definition 7.2.10.** Let $D$ be a subset of a ring $R$. The smallest subring of $R$ that contains $D$, denoted $\langle D\rangle_R$ is called the subring *generated by* $D$.

Explicitly, the subring $\langle D\rangle_R$ of a ring $R$ consists of all finite sums of products of elements from $D$ or $-D = \{x \in R | -x \in D\}$ including 0 (the empty sum) and 1 (the empty product). If a ring can be generated by finitely many elements, it is called *finitely generated*.

**Example 7.2.11** (Usual arithmetic)**.** The ring $\mathbb{Z}$ is generated by the empty set. For, $0, 1$ always belong to a subring; but then also $-1$ (because the additive structure is a group) and $2 = 1+1$. Now, by induction, $n = n-1+1$ belongs to the subring, and hence also its additive inverse $-n$.

The ring $\mathbb{Q}$ is not even finitely generated (that is, generated by a finite subset): to see this, use that there are infinitely many primes and study the possible denominators of elements from a finitely generated subring.

Similarly, neither $\mathbb{R}$ nor $\mathbb{C}$ are finitely generated.

**Example 7.2.12** (Modular arithmetic)**.** The ring $\mathbb{Z}/n\mathbb{Z}$ is generated by the empty set.

**Example 7.2.13** (Polynomials rings)**.** The ring $\mathbb{Z}[X]$ is generated by $X$. More generally, if $R$ is $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, its polynomial ring $R[X]$ is generated by $R \cup \{X\}$.

**Example 7.2.14** (Residue class rings)**.** If $R$ is $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, and $f$ is a polynomial in $R[X]$, then $R[X]/(f)R[X]$ is generated by $R \cup \{x\}$, where $x$ is the residue class of $X$.

**Example 7.2.15** (The Gaussian integers)**.** The ring $R = \mathbb{Z} + \mathbb{Z}\cdot i$ of Gaussian integers is generated by $i$.

**Example 7.2.16** (Matrix rings)**.** The matrix ring $M_n(\mathbb{R})$ is generated by all upper and lower triangular matrices. It is even generated by all upper triangular matrices and permutation matrices.

**Example 7.2.17** (The Quaternions)**.** The ring of quaternions $\mathbb{H} = \mathbb{R} + \mathbb{R}\cdot i + \mathbb{R}\cdot j + \mathbb{R}\cdot k$ is generated by $\mathbb{R} \cup \{i, j\}$.

**Remark 7.2.18.** The subring of a ring $R$ generated by the empty set is the same as the subring generated by 0 and 1, since these two elements belong to any subring.

Let $R$ be a ring and let $X$ be an indeterminate. By $R[X]$ we denote the set of all polynomials in $X$ with coefficients in $R$, compare Definition of polynomial ring.

Let $a = a_0 + a_1 \cdot X + ... + a_n \cdot X^n$ and $b = b_0 + b_1 \cdot X + ... + b_m \cdot X^m$ be two elements of $R[X]$. By adding, if necessary, some terms $0 \cdot X^k$ we may assume $n = m$.

The sum of these polynomials is $a + b = a_0 + b_0 + (a_1 + b_1) \cdot X + ... + (a_n + b_n) \cdot X^n$

The product of these polynomials is $a \cdot b = c_0 + c_1 \cdot X + ... + c_{n+m} \cdot X^{n+m}$ where $c_k = a_0 \cdot b_k + a_1 \cdot b_{k-1} + ... + a_k \cdot b_0$.

The symbol $\cdot$ is often omitted.

**Theorem 7.2.19.** *The sum and product of polynomials define the structure of a commutative ring on the set $R[X]$ of all polynomials in X with coefficients in R. The zero element is the zero polynomial* $0$*; the identity element is the polynomial* $1$.

*Proof.* We must prove that $[R[X], +, 0, -]$ is a commutative group, that $[R[X], \cdot, 1]$ is a commutative monoid and that distributivity holds.

Since most verifications are very similar, we restrict to one typical verification, that of left distributivity.

Let $a = a_0 + a_1 \cdot X + ... + a_n \cdot X^n$, $b = b_0 + b_1 \cdot X + ... + b_m \cdot X^m$, and $c = c_0 + c_1 \cdot X + ... + c_l \cdot X^l$ be three polynomials. The coefficient of $X^k$ in $a \cdot (b + c)$ equals

$$a_0 \cdot (b_k + c_k) + a_1 \cdot (b_{k-1} + c_{k-1}) \cdot ... \cdot a_k \cdot (b_0 + c_0)$$

and can be rewritten by commutativity and distributivity in $R$ as

$$((a_0 \cdot b_k) + (a_1 \cdot b_{k-1}) + ... + (a_k \cdot b_0)) + (((a_0 \cdot c_k) + (a_1 \cdot c_{k-1}) + ... + (a_k \cdot c_0)))$$

which is the coefficient of $X^k$ in $a \cdot b + a \cdot c$.

□

**Example 7.2.20.** Let $R$ be a ring and take $S$ to be the polynomial ring $R[X]$. Then the polynomial ring $S$ in the indeterminate $Y$ is the same as the ring $R[X,Y]$ of polynomials in the two indeterminates $X, Y$. So its elements are of the form $\sum_{(i,j) \in \mathbb{N} \times \mathbb{N}} a_{i,j} \cdot X^i \cdot Y^j$, with $a_{i,j} \in R$, nonzero for only a finite number of pairs $(i, j)$. The element $X \cdot Y$ is equal to the product $Y \cdot X$. This emphasizes that there are two ways to build this ring with indeterminates $X$ and $Y$ from $R$: As $R[X][Y]$ and as $R[Y][X]$. To emphasize the symmetry in $X$ and $Y$, we usually write $R[X,Y]$ for this ring.

**Example 7.2.21.** Notions like degree are of course valid for all polynomial rings. But weird things may happen if the coefficient ring $R$ is not a field: $(2 \cdot X) \cdot (2 \cdot X) = 0$ in $\mathbb{Z}/4\mathbb{Z}[X]$. Here the degree of the product of two polynomials of degree 2 is not 4.

The ring $R[X]$ is called the *polynomial ring* over $R$ in the *indeterminate* $X$. The ring $R$ is called the *coefficient ring* of $R[X]$.

## 7.3 Domains and fields

Although some of the definitions and results presented in this and the following sections are valid for general rings, we concentrate on commutative rings. So, from now on, unless the contrary is stated explicitly, all rings will be considered to be commutative.

**Definition 7.3.1.** Let $R$ be a commutative ring.

- An element $x$ of $R$ is called a *multiple* of an element $y$ if there exists $z \in R$ such that $x = y \cdot z$.
- A *zero divisor* in $R$ is an element $a$ of $R$ for which there exists $b \in R \setminus \{0\}$ with $a \cdot b = 0$.
- A ring without nonzero zero divisors is called a *domain*.

**Example 7.3.2** (Modular arithmetic)**.** In $\mathbb{Z}/6\mathbb{Z}$, the element $2+6\cdot\mathbb{Z}$ is a multiple of $4+6\cdot\mathbb{Z}$: $(4+6\cdot\mathbb{Z})\cdot(2+6\cdot\mathbb{Z}) = 2+6\cdot\mathbb{Z}$.

**Example 7.3.3** (Residue class rings)**.** Let $R = \mathbb{Q}[X]$. If $f \in R$ is irreducible, then any nonzero polynomial $g \in R$ of degree less than the degree of $f$ has an invertible residue class in $R/(f)R$, and so $g+(f)R$ divides 1.

**Example 7.3.4** (The Gaussian integers)**.** In $\mathbb{Z}+\mathbb{Z}\cdot i$, the element $1+i$ is a divisor of 2 since $(1+i)\cdot(1-i) = 2$.

The ring $\mathbb{Z}+\mathbb{Z}\cdot i$ is a domain. For, suppose $(a+b\cdot i)\cdot(c+d\cdot i) = 0$. Multiply both sides with $(a-b\cdot i)\cdot(c-d\cdot i)$, to obtain the equation $\left(a^2+b^2\right)\cdot\left(c^2+d^2\right) = 0$ involving integers only, from which it is clear that $a = b = 0$ or $c = d = 0$, i.e., $a+b\cdot i = 0$ or $c+d\cdot i = 0$.

**Remark 7.3.5.** The notions introduced in Definition of divisor, zero divisor and domain generalize the familiar notions of divisor and multiple in the integers and in polynomial rings.

If $x$ is a multiple of $y$, then $y$ is also called a divisor of $x$.

**Theorem 7.3.6.** *Let $R$ be a commutative ring.*

1. *A zero divisor is never invertible.*
2. *The ring $R$ is a domain if and only if for all $a$ and $b$ in $R$ we have that $a\cdot b = 0$ implies $a = 0$ or $b = 0$.*

*Proof.*

**A zero divisor of $R$ is never invertible.**

Suppose that $a$ is an invertible element of $R$ and suppose that $b$ is an element such that $a\cdot b = 0$. Multiply the latter equality on both sides by $a^{-1}$ to obtain $a^{-1}\cdot(a\cdot b) = 0$. Using the associativity of multiplication gives $b = (a^{-1}\cdot a\cdot b) = a^{-1}\cdot(a)\cdot b = 0$, so $b = 0$. In particular, $a$ is not a zero divisor.

**The ring $R$ is a domain if and only if for all $a$ and $b$ in $R$ $a \cdot b = 0$ implies $a = 0$ or $b = 0$.**

This is a restatement of the definition of domain.

□

**Example 7.3.7** (Usual arithmetic)**.** The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all domains. It is sufficient to note that $\mathbb{C}$ is a domain, since then a forteriori all of its subrings are domains.

**Example 7.3.8** (Modular arithmetic)**.** The ring $\mathbb{Z}/n\mathbb{Z}$ is a domain if and only if $n$ is a prime.

**Example 7.3.9** (Polynomial rings)**.** The polynomial ring $R[X]$ is a domain if and only if $R$ is a domain. See Polynomial rings over a domain are domains.

**Example 7.3.10** (Residue class rings)**.** Let $R$ be a field. Then the residue class ring $R[X]/(f)R[X]$ is a domain, if and only if the polynomial $f$ is irreducible in $R[X]$. So, $R[X]/(f)R[X]$ is a domain if and only if it is a field.

**Example 7.3.11** (The Gaussian integers)**.** The Gaussian integers $\mathbb{Z} + \mathbb{Z} \cdot i$ is a domain. It is a subring of the domain $\mathbb{C}$.

**Remark 7.3.12.** It is a common misconception to think that each element of a ring would be either a zero divisors or an invertible element. The element 4 in the ring $\mathbb{Z}$ is an example of an element that is neither invertible nor a zero divisor.

**Example 7.3.13.** Suppose $x \in \mathbb{R}$ (the reals) is a solution of the equation $x^5 - 8 \cdot x^4 + 16 \cdot x^3 + 3 \cdot x^2 - 14 \cdot x = 4$. Using more advanced methods than treated so far, real polynomials can be factored. Bringing 4 to the left hand side and factoring the resulting polynomial, we find $\left(x^2 - 3 \cdot x - 1\right) \cdot \left(x^3 - 5 \cdot x^2 + 2 \cdot x + 4\right) = 0$. Since $\mathbb{R}$ is a domain, we conclude $x^2 - 3 \cdot x = 1$ or $x^3 - 5 \cdot x^2 + 2 \cdot x = -4$. So $x$ is a solution of one of two equations of smaller degree.

Clearly, a subring of a domain is a domain.

Here is a way to construct a domain out of a given one.

**Theorem 7.3.14.** *If $R$ is a domain, then so is $R[X]$.*

*Proof.* Suppose that $f, g$ are nonzero polynomials in $R$ such that $f \cdot g = 0$. Let $m = \text{degree}(f)$ and $n = \text{degree}(g)$. Then $m$ and $n$ are non-negative integers (since $f$ and $g$ are nonzero). The corresponding top coefficients of $f$ and $g$ are nonzero, so (as $R$ is a domain) the coefficient of $X^{m+n}$ in $f \cdot g$ is nonzero, showing that $f \cdot g \neq 0$. Therefore, $R[X]$ is a domain.

□

**Example 7.3.15.** Let $R$ be a domain. By applying the proposition twice, we see that $R[X,Y]$ is a domain.

**Remark 7.3.16** (Converse)**.** Since $R$ is a subring of $R[X]$, for the latter to be a domain it is of course necessary that $R$ be a domain.

The following property is an important reason why domains are good to work with.

**Theorem 7.3.17** (Cancellation law for domains)**.** *Let $R$ be a domain. If $a$ is a nonzero element of $R$, then $a\cdot x = a\cdot y$ implies $x = y$.*

*Proof.* From $a\cdot x = a\cdot y$ we deduce $a\cdot (x-y) = 0$. Since $R$ is a domain and $a$ is nonzero, we conclude that $x-y = 0$, i.e., $x = y$.

□

**Example 7.3.18.** Suppose $x \in \mathbb{Z}+\mathbb{Z}\cdot i$, the ring of Gaussian integers, is a solution of the equation $(2+i)\cdot x = 5$. Then, by the cancellation law, the equation is equivalent to the one obtained by left multiplication with $2-i$ on both sides: $5\cdot x = 5\cdot (2-i)$. Applying the cancellation law once more, we find $x = 2-i$.

**Remark 7.3.19** (Converse)**.** If a ring has zero divisors, the cancellation law need not hold. For instance, in $\mathbb{Z}/6\mathbb{Z}$, we have $2\cdot 2 = 2\cdot 5$, but $2 \neq 5$.

We now give a more formal approach to fields than before.

**Definition 7.3.20.** A *field* is a commutative ring in which every nonzero element has a multiplicative inverse.

**Example 7.3.21** (Usual arithmetic)**.** The ring $\mathbb{Z}$ is not a field: most of its elements are not invertible.

On the other hand, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields.

**Example 7.3.22** (Modular arithmetic)**.** The ring $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if $n$ is a prime number.

**Example 7.3.23** (Polynomial rings)**.** The rings $\mathbb{Q}[X], \mathbb{R}[X], \mathbb{C}[X]$ are not fields: $X$ does not have an inverse.

**Example 7.3.24** (Residue class rings)**.** If $R = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, or $\mathbb{Z}/p\mathbb{Z}$ for some prime $p$, and $f$ is a polynomial in $R[X]$, then $R[X]/(f)R[X]$ is a field if and only $f$ is irreducible in $R[X]$.

**Example 7.3.25** (The Gaussian integers)**.** The ring $\mathbb{R} = \mathbb{Z}+\mathbb{Z}\cdot i$ of Gaussian integers is not a field. For instance, the element $1+i$ has no inverse: if $a+b\cdot i$ were its inverse, then $2\cdot a+2\cdot b\cdot i = (a+b\cdot i)\cdot 2 = (a+b\cdot i)\cdot(1+i)\cdot(1-i) = 1-i$, whence $2\cdot a = 1$, which contradicts $a \in \mathbb{Z}$. The variation $\mathbb{Q}+\mathbb{Q}\cdot i$ however, is a field. Can you find the inverse of an arbitrary nonzero element?

Since by definition every nonzero element is invertible, the nonzero elements of a field $K$ form a group with respect to the multiplication: $K^{\times} = K \setminus \{0\}$.

Since an invertible element cannot be a zero divisor, every field is a domain. The converse is not necessarily true: $\mathbb{Z}$ is a domain but not a field. For finite domains, however, the converse does hold.

**Theorem 7.3.26.** *Every finite domain is a field.*

*Proof.* Let $R$ be a finite domain and $a$ a nonzero element of $R$. We need to show that $a$ is invertible. To this end, consider left multiplication by $a$, that is, the map $L_a\colon R \to R, x \mapsto a \cdot x$.

Since $R$ is a domain, it follows from the Cancellation law for domains that $L_a$ is injective. Since $R$ is a finite set, the pigeon hole principle says that the map is necessarily surjective. In particular, there exists $y \in R$ such that $L_a(y) = 1$. This means $a \cdot y = 1$, as required for $a$ to be invertible in $R$.

□

**Example 7.3.27.** Consider the ring $R = \mathbb{Z}/3\mathbb{Z} + \mathbb{Z}/3\mathbb{Z} \cdot i$, where $i$ is the square root of $-1$; so $i^2 = -1$.

$R$ is a field. To see this, suppose that $x = a + b \cdot i$ and $y = c + d \cdot i$, with $a, b, c, d \in \mathbb{Z}/3\mathbb{Z}$ satisfy $x \cdot y = 0$. Multiplying this equation by $(a - b \cdot i) \cdot (c - d \cdot i)$, we find $\left(a^2 + b^2\right) \cdot \left(c^2 + d^2\right) = 0$. Both factors are in $\mathbb{Z}/3\mathbb{Z}$, which is (a field and hence) a domain. Therefore, at least one of them is zero, say the first (the argument for the second is similar). This means $a^2 = -(b^2)$, that is, $a = b = 0$, as is easily checked within $\mathbb{Z}/3\mathbb{Z}$ and $x = 0$. We conclude that $R$ is a finite domain, whence a field.

Let $F$ be a field. The following definitions are completely standard; compare them with those for monoids, groups, and rings given so far.

**Definition 7.3.28.** A *subfield* of the field $F$ is a subring of $F$ which is closed under inverses of nonzero elements. If $X$ is a subset of $F$, the subfield of $F$ generated by $X$ is the intersection of all subfields containing $X$.

**Example 7.3.29** (Usual arithmetic)**.** $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and $\mathbb{R}$ is a subfield of $\mathbb{C}$.

**Example 7.3.30** (Modular arithmetic)**.** There are no proper subfields of $\mathbb{Z}/p\mathbb{Z}$ for $p$ prime. For any subfield contains 1 and hence all its multiples and thus the complete field $\mathbb{Z}/p\mathbb{Z}$.

**Example 7.3.31** (Polynomial rings)**.** $\mathbb{Q}$ is a subring of the polynomial ring $\mathbb{Q}[X]$. Since $\mathbb{Q}$ by itself is a field, one might speak of a subfield here, although the ambient ring $\mathbb{Q}[X]$ is not a field.

Later we shall see how to "extend" the domain $\mathbb{Q}[X]$ to a field. Similar remarks hold for $\mathbb{R}$ and $\mathbb{C}$ instead of $\mathbb{Q}$.

**Example 7.3.32** (Residue class rings)**.** Let $f = X^4 - 2$ be a polynomial in $\mathbb{Q}[X]$ and consider $F = \mathbb{Q}[X]/(f)\mathbb{Q}[X]$. Since $f$ is irreducible in $\mathbb{Q}[X]$, this is a field. Now consider the element $b = X^2 + (f)\mathbb{Q}[X]$ of $F$. The subfield of $F$ generated by $\mathbb{Q}$ and $b$ is $K = \mathbb{Q} + \mathbb{Q}\cdot b$. (To see this, notice that $b^2 = 2$ and $b^{-1} = b/2$.) Thus, the field $F$, which is a 4-dimensional vector space over $\mathbb{Q}$ has a subfield $K$, which is a 2-dimensional linear space of $\mathbb{Q}$.

**Example 7.3.33** (The Gaussian numbers)**.** The field $\mathbb{Q} + \mathbb{Q}\cdot i$ is a 2-dimensional vector space over $\mathbb{Q}$. An obvious subfield is $\mathbb{Q}$. This is the only proper subfield of $\mathbb{Q} + \mathbb{Q}\cdot i$, as will become clear later, from the fact that any subfield contains $\mathbb{Q}$.

**Remark 7.3.34.** By now we assume that you are aware from previous cases like monoids and rings that the intersection of any collection of subfields is a subfield. This fact is of course used in the definition.

**Remark 7.3.35.** Another description (again, as usual) of the subfield generated by $X$ is that it is the smallest subfield containing $X$.

We now focus on subfields of the field $\mathbb{C}$ of complex numbers. Let $K$ be a subfield of $\mathbb{C}$. If $a$ is an element of $K$, and $f, g$ are polynomials in $\mathbb{Q}[X]$, then $\frac{f(a)}{g(a)}$ is an element of $K$ whenever $g(a) \neq 0$. The set of all these fractions makes up the smallest subfield of $\mathbb{C}$ that contains $a$ and $L$. Of course, instead of polynomials in $\mathbb{Q}$, we could have chosen $f$ and $g$ with coefficients in any subfield $L$ of $K$. In general we obtain the following.

**Theorem 7.3.36.** *If $a \in \mathbb{C}$ and $L$ a subfield of $\mathbb{C}$, then $K = \{\frac{f(a)}{g(a)} \in \mathbb{C} | \left((f,g) \in L[X]^2\right) \wedge (g(a) \neq 0)\}$ is the subfield of $\mathbb{C}$ generated by $a$ and $L$.*

*Proof.* We must show two things:

**$K$ is a subfield.**

- $K$ contains the elements 0 (take $f = 0$ and $g = 1$ ) and 1 (take $f = 1$ and $g = 1$).
- $K$ is closed under addition: $\frac{f(a)}{g(a)} + \frac{h(a)}{l(a)} = \frac{f(a)\cdot l(a) + g(a)\cdot h(a)}{g(a)\cdot l\cdot a}$; here the polynomials $f\cdot l + g\cdot h$ and $g\cdot l$ are used.
- $K$ is closed under multiplication: $\frac{f(a)}{g(a)}\cdot\frac{h(a)}{l(a)} = \frac{f(a)\cdot h(a)}{g(a)\cdot l(a)}$; here the polynomials $f\cdot h$ and $g\cdot l$ are used.
- The additive inverse of $\frac{f(a)}{g(a)}$ is $\frac{-(f(a))}{g(a)}$; here the polynomials $-f$ and $g$ are used.
- Every nonzero element in $K$ has its multiplicative inverse in $K$: the inverse of $\frac{f(a)}{g(a)}$ is $\frac{g(a)}{f(a)}$ . Note that $f(a) \neq 0$ .

**$K$ is the smallest subfield of $\mathbb{C}$ containing $a$ and $L$.**

To show that $K$ is the smallest field containing $a$ and $L$, we note that any field containing $a$ and $L$ also contains $f(a)$ for every polynomial $f \in L[X]$, since $f(a)$ arises by repeated addition and multiplication starting from $a$ and elements of $L$. But if the subfield contains $f(a)$ and $g(a)$, with nonzero $g(a)$, then it also contains the product of $f(a)$ and the inverse $\frac{1}{g(a)}$, that is, the quotient $\frac{f(a)}{g(a)}$. In conclusion, the subfield must contain $K$.

□

The subfield of $\mathbb{C}$ generated by $a$ and the subfield $L$ is often denoted by $L(a)$.

More generally, if $K$ is a field, $L$ a subfield of $K$ and $a$ an element (or set of elements) of $K$, then the subfield of $K$ generated by $L$ and $a$ is denoted by $L(a)$.

**Example 7.3.37.** Let $a$ be the (positive) square root of 2 in $\mathbb{R}$. Thus, $a = \sqrt{2}$. We will determine the subfield of $\mathbb{R}$ generated by $a$. Since $a^2 = 2$, for every polynomial $f \in \mathbb{Q}[X]$, the number $f(a)$ is of the form $c + d \cdot a$. So $\mathbb{Q}(a)$ consists of the quotients $\frac{c+d\cdot a}{g+h\cdot a}$, with $c, d, g, h \in \mathbb{Q}$. These expressions can be simplified even further: multiply numerator and denominator by $g - h \cdot a$ to conclude that $\mathbb{Q}(a) = \{x + y \cdot a | (x, y) \in \mathbb{Q} \times \mathbb{Q}\}$. In other terms, $\mathbb{Q}\left(\sqrt{2}\right) = \mathbb{Q} + \mathbb{Q} \cdot \sqrt{2}$.

**Remark 7.3.38.** The field $K$ is readily seen to be a vector space over $L$. If there is a polynomial $h \in L[X]$ such that $h(a) = 0$, then $K$ is a finite-dimensional vector space. If there is no such polynomial, then $K$ is an infinite-dimensional vector space over $L$. For instance, there is no polynomial in $\mathbb{Q}[X]$ having $\pi$ as a zero (nontrivial; we give no proof here!), and so $K$ is infinite-dimensional if $a = \pi$ and $L = \mathbb{Q}$.

Let $R$ be a domain. On the set of pairs $(t, n)$ from $R$ with $n \neq 0$, we define an equivalence relation eaq (equal as quotient): $((t, n) \operatorname{eaq} (t', n')) \Leftrightarrow (t \cdot n' = t' \cdot n)$

We call $t$ the numerator and $n$ the denominator of the pair $(t, n)$. Denote the equivalence class containing $(t, n)$ by $\frac{t}{n}$, and the set of equivalence classes by $Q(R)$. Addition and multiplication on these classes are defined as follows:

- addition: $\frac{t}{n} + \frac{s}{m} = \frac{n\cdot s + t\cdot m}{n\cdot m}$;
- zero element: $\frac{0}{1}$;
- multiplication: $\frac{t}{n} \cdot \frac{s}{m} = \frac{t\cdot s}{n\cdot m}$;
- idenitity element: $\frac{1}{1}$.

It is readily checked that these operations are well defined and that $Q(R)$ is a ring. Even more is true:

**Theorem 7.3.39.** *Let $R$ be a domain. The structure $Q(R)$, with operations defined as above, is a field.*
*This field is called the field of fractions of $R$.*

*Proof.* The first three parts of the proof suffice to establish that $Q(R)$ is a ring, the last part that it is a field.

$[Q(R),+,0,x\mapsto -x]$ **is an additive group.**

Let $a,b,c,d,e,f\in R$, with $d,e,f\neq 0$. Then, by associativity of $+$ on $R$, we have $\frac{a}{d}+\left(((\frac{b}{e})+(\frac{c}{f}))\right)=\frac{a}{d}+(b\cdot f+c\cdot e/e\cdot f)=\frac{a\cdot e\cdot f+b\cdot d\cdot f+c\cdot d\cdot e}{d\cdot e\cdot f}=\frac{a\cdot e+b\cdot d}{d\cdot e}+\frac{c}{f}=((\frac{a}{d})+(\frac{b}{e}))+\frac{c}{f}$ We have shown that $+$ is associative on $Q(R)$. We leave the (easier) verifications that $\frac{0}{n}$ is the zero element, that $\frac{-t}{n}$ is the inverse of $\frac{t}{n}$ and that $+$ is commutative to the reader.

$[Q(R),\cdot,1]$ **is a commutative monoid.**

Let $a,b,c,d,e,f\in R$, with $d,e,f\neq 0$. Then, by associativity of $\cdot$ on $R$, we have $(\frac{a}{d}\cdot\frac{b}{e})\cdot\frac{c}{f}=\frac{a\cdot b}{d\cdot e}\cdot\frac{c}{f}=\frac{(a\cdot b)\cdot c}{(d\cdot e)\cdot f}=\frac{a\cdot(b\cdot c)}{d\cdot(e\cdot f)}=\frac{a}{d}\cdot\frac{b\cdot c}{e\cdot f}=\frac{a}{d}\cdot(\frac{b}{e}\cdot\frac{c}{f})$. We have shown that $\cdot$ is associative on $Q(R)$. We leave the (easier) verifications that $\cdot$ is commutative and that $\frac{1}{1}$ is the identity element to the reader.

**Distributivity.**

Let $a,b,c,d,e,f\in R$, with $d,e,f\neq 0$. Then, by distributivity of $R$, we have $\frac{a}{d}\cdot\left(\frac{b}{e}+\frac{c}{f}\right)=\frac{a}{d}\cdot\frac{b\cdot f+c\cdot e}{e\cdot f}=\frac{((a\cdot b\cdot f)+(a\cdot c\cdot e))}{d\cdot e\cdot f}=\frac{((a\cdot d\cdot b\cdot f)+(a\cdot d\cdot c\cdot e))}{d^2\cdot e\cdot f}=\frac{a\cdot b}{d\cdot e}+\frac{a\cdot c}{d\cdot f}=\frac{a}{d}\cdot\frac{b}{e}+\frac{a}{d}\cdot\frac{c}{f}$. We have shown left distributivity. In view of commutativity of $\cdot$, there is no need to prove right distributivity.

**Each nonzero element of $Q(R)$ has a multiplicative inverse.**

Let $\frac{a}{r}$ be a nonzero element of $Q(R)$. Then $a$ is a nonzero element of $R$ and so $\frac{r}{a}$ belongs to $Q(R)$ and $\frac{\frac{a}{r}\cdot r}{a}=\frac{a\cdot r}{r\cdot a}=\frac{1}{1}$. This establishes that $\frac{a}{r}$ is invertible in $Q(R)$ with inverse $\frac{r}{a}$.

□

The map $R\to Q(R),x\mapsto\frac{x}{1}$ is an injective homomorphism of rings. Thus, $R$ may be viewed as a subring of $Q(R)$.

**Remark 7.3.40.** Note that the addition and multiplication of $\frac{t}{n}$ and $\frac{t'}{n'}$, with $n$ and $n'$ nonzero, is well defined because $R$ is a domain. For, in the product $\frac{t}{n}\cdot\frac{t'}{n'}=\frac{t\cdot t'}{n\cdot n'}$ we have a nonzero denominator since both $n$ and $n'$ are nonzero. Similarly for addition.

**Example 7.3.41** (The integers)**.** The field of fractions of the integers is the field of rational numbers.

In this case there is a unique representative $(t,n)$ for each class with the properties

- $\gcd(t,n)=1$;
- $n>0$.

It is obtained from an arbitrary representative by dividing both numerator and denominator by their common gcd, and also by $-1$ if necessary to obtain a positive denominator.

**Example 7.3.42** (The Gaussian integers)**.** Let $R=\mathbb{Z}+\mathbb{Z}\cdot i$ where $i=\sqrt{-1}$. We claim $Q(\mathbb{Z}+\mathbb{Z}\cdot i)=\mathbb{Q}+\mathbb{Q}\cdot i$. For, $\frac{a+b\cdot i}{c+d\cdot i}=\frac{a\cdot c-b\cdot d}{c^2+d^2}+\frac{a\cdot d+b\cdot c}{c^2+d^2}\cdot i$.

**Example 7.3.43** (Polynomial rings)**.** Let $K$ be a field, then the ring $K[X]$ is a domain, and we can form its fraction field. This fraction field is denoted by $Q(X)$, and called *rational functions field* over $K$ in $X$. This elements of this field can be described as: $\frac{f(X)}{g(X)}$ with $g(X)\neq 0$.

**Remark 7.3.44.** Suppose that we know how to work with elements of a domain $R$ on computer. Can we work with elements of $Q(R)$? Clearly, a fraction $\frac{t}{n}$ can be represented by the pair $(t,n)$, and the given formulas work for defining product and addition in terms of the operations for $R$. Equality amongst fractions also requires a computation: $\frac{t}{n}=\frac{t'}{n'}$ is verified by determining whether $t\cdot n'=t'\cdot n$ holds.

## 7.4 Fields

Let $K$ be a field. Every subfield of $K$ contains 0 and 1, and so it also contains $1+...+1$ and $-1-1-...-1$.

The subfield therefore contains all integral multiples of 1 and $-1$ as well as all fractions of these multiples (as long as the denominator is nonzero). These elements make up a subfield themselves.

**Theorem 7.4.1.** *A field generated by the empty set (or by* 0 *and* 1*), is isomorphic with* $\mathbb{Q}$ *or* $\mathbb{Z}/p\mathbb{Z}$ *for some prime number* $p$*. In particular, every field contains a subfield isomorphic with* $\mathbb{Q}$ *or* $\mathbb{Z}/p\mathbb{Z}$ *for some prime number* $p$*.*

*Proof.* Let $L$ be the subfield of $K$ generated by the empty set $\varnothing$. Then it containts 0 and 1, and so it is also generated by these two elements. For every positive integer $m$ the element $m\cdot 1=1+1+...+1$ ($m$ terms) belongs to $L$, and therefore also the element $(-m)\cdot 1=-m\cdot 1=m\cdot(-1)$.

Put $A=\{x\in\mathbb{N}\,|\,(x>0)\wedge(x\cdot 1=0)\}$. We distinguish according to $A$ being the empty set or not.

**If $A=\varnothing$, then $L$ is isomorphic to $\mathbb{Q}$**

The map $\mathbb{Z}\to L$ that sends $m$ to $m\cdot 1$ is an injective homomorphism. It is easy to see that this map extends to an injective homomorphism $\mathbb{Q}\to L$, $\frac{m}{n}\mapsto\frac{m\cdot 1}{n\cdot 1}$. This map identifies $\mathbb{Q}$ with $L$.

**If $A$ is not empty, then it contains a smallest positive element $p$. Then $L$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}$.**

Since 0 and 1 are distinct, $p > 1$. If $p$ were not prime, then there exist positive integers $b, c < p$ such that $b \cdot c = p$. It follows that $(b \cdot 1) \cdot (c \cdot 1) = (b \cdot c) \cdot 1 = p \cdot 1 = 0$ so that at least one of $b \cdot 1, c \cdot 1$ equals 0, contradicting the minimality of $p$. But then the obvious map $\mathbb{Z}/p\mathbb{Z} \to L$ is injective and maps $\mathbb{Z}/p\mathbb{Z}$ isomorphically onto $L$.

□

**Example 7.4.2** (Usual arithmetic)**.** The ring $\mathbb{Q}$ of rational numbers has no proper subfields. In case of $\mathbb{R}$, $\mathbb{C}$, or any subfield of $\mathbb{C}$, the smallest subfield is $\mathbb{Q}$.

**Example 7.4.3** (Modular arithmetic)**.** The field $\mathbb{Z}/p\mathbb{Z}$ has no smaller subfields.

**Example 7.4.4** (Rational fields)**.** $\mathbb{Q}$ is the smallest subfield of $\mathbb{Q}(X)$ and of $\mathbb{R}(X)$.

**Example 7.4.5** (Residue class fields)**.** If $p$ is a prime number, $\mathbb{Z}/p\mathbb{Z}$ is the smallest subfield of $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$ where $f$ is irreducible in $\mathbb{Z}/p\mathbb{Z}[X]$.

**Example 7.4.6** (The Gaussian numbers)**.** The smallest subfield of $\mathbb{Q} + \mathbb{Q} \cdot i$ is $\mathbb{Q}$.

We consider the smallest subfield $L$ of $K$ (it is generated by 0 and 1).

**Definition 7.4.7.** If $L$ is isomorphic with $\mathbb{Q}$, then $K$ is said to have characteristic 0. If $L$ is isomorphic with $\mathbb{Z}/p\mathbb{Z}$, then $K$ is said to have characteristic $p$.

**Example 7.4.8** (Usual arithmetic)**.** The characteristic of $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$, or any subfield of $\mathbb{C}$ is 0.

**Example 7.4.9** (Modular arithmetic)**.** Of course, $\mathbb{Z}/p\mathbb{Z}$ has characteristic $p$.

**Example 7.4.10** (Fields of rational functions)**.** If $R$ is a field, then the characteristic of the field of rational functions $R(X)$ is equal to the characteristic of $R$.

**Example 7.4.11** (Residue class fields)**.** If $F$ is a field and $f$ an irreducible polynomial in $F$ of positive degree, then the residue class ring $F[X]/(f)F[X]$ is a field whose characteristic is that of $F$.

**Example 7.4.12** (The Gaussian numbers)**.** The characteristic of $\mathbb{Q} + \mathbb{Q} \cdot i$ is 0.

By the above theorem, the characteristic of a field is either zero or a prime number.

Let $K$ be a field. The next theorem gives a connection between linear algebra (see the prerequisites) and elements of a field extension.

**Theorem 7.4.13.** *If $L$ is a subfield of the field $K$, then the following two statements hold.*

1. *$K$ is a vector space over $L$.*
2. *For each $x \in K$, multiplication with $x$ is a linear transformation of this vector space over $L$.*

*Proof.*

**$K$ is a vector space over $L$.**

$K$ is a ring and addition on $K$ is a commutative group structure. Scalar multiplication $L \times K \to K$ is given by ordinary multiplication in $K$. We need to verify the following laws.

- For $x \in K$, we have $1 \cdot x = x$; this holds because of the multiplication laws in $K$.
- Associativity: For $x, y \in L$ and $z \in K$ , we have $x \cdot y \cdot z = x \cdot y \cdot z$ simply because $K$ is associative.
- Distributivity: For $x \in L$ and $y, z \in K$, we have $x \cdot (y+z) = x \cdot y + x \cdot z$ simply because of distributivity in $K$.
- Distributivity: For $x, y \in L$ and $z \in K$ , we have $(x+y) \cdot z = x \cdot z + y \cdot z$ again because of distributivity in $K$.

**For each $x \in K$, multiplication with $x$ is a linear transformation of this vector space over $L$.**

Let $x \in K$. By $L_x$ we denote left multiplication with $x$ on $K$. The fact that $L_x$ is a linear transformation of the vector space $K$ over $L$ follows from

- $L_x$ respect vector addition because of distributivity: $L_x(y+z) = x \cdot (y+z) = x \cdot y + x \cdot z = L_x(y) + L_x(z)$ for $y, z \in K$,
- and $L_x$ respects scalar multiplication because of commutativity: $L_x(y \cdot z) = x \cdot (y \cdot z) = y \cdot (x \cdot z) = y \cdot L_x(z)$ for $y \in L$ and $z \in K$.

□

**Example 7.4.14** **($\mathbb{R} \subset \mathbb{C}$).** This corresponds to the familiar view of $\mathbb{C}$ as the 'complex plane', a 2-dimensional vector space over $\mathbb{R}$ with basis $1, i$.

**Example 7.4.15** **($\mathbb{Q} \subset \mathbb{R}$).** This is an infinite-dimensional vector space. For instance, the numbers $\sqrt{p}$, for $p$ prime numbers in $\mathbb{N}$, form an infinite set of linearly independent elements. But not a basis, as elements such as e and $\pi$ and are still not in their linear span.

**Example 7.4.16** **($\mathbb{Z}/2\mathbb{Z} \subset \mathbb{Z}/2\mathbb{Z}[X]/(X^2+X+1)$).** This is the situation described before; we are dealing here with a 2-dimensional vector space over $\mathbb{Z}/2\mathbb{Z}$, and so with a field of 4 elements.

Here is a consequence of the previous theorem for finite fields: their orders form a proper subset of the natural numbers.

**Corollary 7.4.17.** *If $F$ is a finite field, then there is a prime $p$ and a natural number $n$ such that $|F| = p^n$.*

*Proof.* By the first theorem of this section, the subfield generated by the empty set is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ for some prime $p$. By the previous theorem, $F$ inherits the structure of a vector space over $\mathbb{Z}/p\mathbb{Z}$. If the dimension of this vector space is $n$, then every element of $F$ can be uniquely represented as a $\mathbb{Z}/p\mathbb{Z}$ linear combination of $n$ given basis vectors, and so the number of elements of $F$ is $p^n$.

□

The fact that, for every prime power, there is a field of that order, has been stated before. Later we shall prove this as well as the fact that all fields of a given order are isomorphic.

**Example 7.4.18.** Suppose that $K$ is a field of order 4. Then $L = \{0,1\}$ is a subfield of order 2. Take $y \in K \setminus L$. The theorem tells us that $K$ is a 2-dimensional vector space over $L$, and so $1, y$ is a basis of $K$ over $L$. In particular, there are $a, b \in L$ such that $y^2 = a + b \cdot y$. Now consider the linear transformation $x \mapsto y \cdot x$ of $K$. It has matrix $\begin{pmatrix} 0 & a \\ 1 & b \end{pmatrix}$ with respect to the basis 1, $y$. As $y$ must be invertible, we have $a \neq 0$. But then $a = 1$. There remain two possibilities for $b$. Suppose $b = 0$. Then $y^2 = 1$. But from this we deduce $(y+1)^2 = 0$, and so $y + 1 = 0$, that is, $y = 1$, a contradiction with $y \notin L$. Hence $b = 1$, and so $y$ satisfies $y^2 = y + 1$. We conclude that $K = \{0, 1, y, y+1\}$ with the multiplication determined by the rule $y^2 = y + 1$.

The above argument gives a glimpse of why there is just one field of order 4.

Here is another way of interpreting the result. The element $y$ is a zero of the irreducible polynomial $X^2 + X + 1$. Thus, it behaves in the same way as the residue of $X$ in the field $L[X]/(X^2 + X + 1)L[X]$. In fact, $K$ is isomorphic with this field.

Many properties of the polynomial ring $K[X]$ discussed before for special fields like $K = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ and $\mathbb{Z}/p\mathbb{Z}$ with $p$ a prime, are in fact valid for arbitrary fields $K$. For instance,

- division with remainder,
- Euclid‘s algorithm,
- gcd and lcm,
- unique factorization.

Proofs can be copied verbatim, so we shall not repeat them. An important consequence is that we can compute modulo a polynomial $d$ in $K[X]$ and construct the residue class ring $K[X]/(d)K[X]$. This allows us to construct new fields.

**Theorem 7.4.19.** *Let $K$ be a field and $d$ a polynomial in $K[X]$.*

1. *The residue class $a + (d)K[X]$ has an inverse in $K[X]/(d)K[X]$ if and only if* $\gcd(a, d) = 1$.
2. *If $d$ is irreducible in $K[X]$, then $K[X]/(d)K[X]$ is a field.*

*Proof.*

**Part 1.**

If the residue class $a+(d)K[X]\in K[X]/(d)K[X]$ has inverse $b+(d)K[X]$, then $a\cdot b=1$ (mod $d$ ). Hence there is a polynomial $p$ with $a\cdot b+p\cdot d=1$.

But that implies that $\gcd(a,d)=1$.

On the other hand, if $\gcd(a,d)=1$, then the extended Euclidean algorithm leads to a method for finding polynomials $b$ and $p$ such that $a\cdot b+p\cdot d=1$. But then $b$ represents an inverse of the residue class $a+(d)K[X]$.

**Part 2.**

By the first statement, every nonzero element in $K[X]/(d)K[X]$ has an inverse.

□

**Example 7.4.20.** We take $K$ any field and $d=X^n$ with $n>1$. Then the residue class of a polynomial $a$ is invertible in $K[X]/(d)K[X]$ if and only if $a_1$, the constant term of $a$, differs from 0.

Let $K$ be a field. In the sequel we need the following general result, which extends a previous lemma.

**Lemma 7.4.21.** *Let $g\in K[X]$.*

1. *If $x\in K$ is a zero of $g$, then $X-x$ divides $g$.*
2. *If $g$ has degree $n$, then $g$ has at most $n$ zeros in $K$.*

*Proof.* By parts.

**If $x\in K$ is a zero of $g$, then $X-x$ divides $g$.**

Computing $\mathrm{rem}(g,X-x)$, we find the constant $g(x)$, which is zero by the assumption that $x$ is a zero of $g$. Hence $X-x$ divides $g$.

**If $g$ has degree $n$, then $g$ has at most $n$ zeros in $K$.**

By the first part of this lemma, each zero $x$ of $g$ corresponds to a linear factor $X-x$, and so distinct zeros correspond to distinct linear factors. Since $g$ has degree $n$, it can have at most $n$ distinct linear factors.

□

**Example 7.4.22** (Fewer zeros than the degree)**.** Consider $X^2+1$ in $\mathbb{Q}[X]$. Since there is no element in $\mathbb{Q}$ squaring to $-1$, there are no zeros of $X^2+1$ in $\mathbb{Q}$. Since each non-constant proper divisor of $X^2+1$ must have degree 1, the above theorem implies that this polynomial is irreducible in $\mathbb{Q}[X]$.

**Example 7.4.23** (The Fundamental Theorem of Algebra)**.** The fundamental theorem of algebra says that every polynomial in $\mathbb{C}$ has a zero. Equivalently: every polynomial in $\mathbb{C}$ is a product of linear factors. We shall give no proof of this fact. One reason is that it is hard, another that we have given no rigorous treatment of $\mathbb{C}$ anyway.

**Remark 7.4.24** (Converse)**.** Consider $X^3-X\in\mathbb{Z}/6\mathbb{Z}$. It has more than 3 zeros in $\mathbb{Z}/6\mathbb{Z}$. Apparently, for the lemma to hold it is essential that the coefficient ring is a domain.

Let $K$ be a field and $S$ a ring. Homomorphisms can be used to construct subfields.

**Theorem 7.4.25.** *$f\colon K\to S$ is a ring homomorphism.*

1. *The homomorphism $f$ is injective.*
2. *The image of $f$ is a subring of $S$ isomorphic to $K$.*
3. *If $S=K$, then $\{x\in K|f(x)=x\}$ is a subfield of $K$.*

*Proof.* By parts.

**The homomorphism $f$ is injective.**

Suppose that $x\in K$ satisfies $f(x)=0$. By an [?] it suffices to show $x=0$. If $x\neq 0$, then $x$ is invertible. But then $1=f\left(x^{-1}\cdot x\right)=f\left(x^{-1}\right)\cdot f\cdot x=f\left(x^{-1}\right)\cdot 0=0$ a contradiction. Hence $x=0$.

**The image of $f$ is a subring of $S$ isomorphic to $K$.**

This is a direct consequence of the first part.

**If $S=K$, then $\{x\in K|f(x)=x\}$ is a subfield of $K$.**

Put $L=\{x\in K|f(x)=x\}$. Clearly, $f(0)=0$ and $f(1)=1$, so $0,1\in L$. Suppose that $x,y\in L$. Then, as $f$ is a homomorphism,

- $f(-x)=f(0-x)=f(0)-f(x)=0-x=-x$,
- $1=f(1)=f\left(x\cdot x^{-1}\right)=f(x)\cdot f\left(x^{-1}\right)=x\cdot f\left(x^{-1}\right)$, so $f\left(x^{-1}\right)=x^{-1}$,
- $f(x\cdot y)=f(x)\cdot f(y)=x\cdot y$,
- $f(x+y)=f(x)+f(y)=x+y$,

whence $-x, x^{-1}, x\cdot y, x+y \in L$. This suffices to establish that $L$ is a subfield of $K$.

□

**Example 7.4.26** (Usual arithmetic)**.** The embedding of $\mathbb{Q}$ in $\mathbb{R}$ and of $\mathbb{R}$ in $\mathbb{C}$ are homomorphisms of fields. Complex conjugation is a homomorphism $c\colon \mathbb{C} \to \mathbb{C}$. The subfield $\{x \in \mathbb{C} | c(x) = x\}$ coincides with $\mathbb{R}$.

**Example 7.4.27** (Modular arithmetic)**.** By definition, for each field $K$ of characteristic $p$, there is an injective morphim $\mathbb{Z}/p\mathbb{Z} \to K$.

**Example 7.4.28** (Rational function field)**.** Let $R$ be a field, and $h \in R[X]$. Then the map $R(X) \to R(X), \frac{f(X)}{g(X)} \mapsto \frac{f(h(X))}{g(h(X))}$ is a homomorphism. Its image is the subfield of $\mathbb{R}(X)$ of all fractions of polynomials that can be written as a polynomial in $h$.

**Example 7.4.29** (The Gaussian numbers)**.** On the Gaussian number field $\mathbb{Q}+\mathbb{Q}\cdot i$, we have complex conjugation: $c\colon \mathbb{Q}+\mathbb{Q}\cdot i \to \mathbb{Q}+\mathbb{Q}\cdot i, a+b\cdot i \mapsto a-b\cdot i$. The subfield $\{x \in \mathbb{C} | c(x) = x\}$ coincides with $\mathbb{Q}$.

**Remark 7.4.30.**

- A homomorphism of fields is nothing but a homomorphism of the underlying rings. Observe that if $x$ is invertible, a homomorphism of rings takes $x^{-1}$ to the inverse of the image of $x$.
- A homomorphism of fields $f\colon K \to S$ need not be surjective, not even when $K = S$. For instance, let $K$ and $S$ be the rational functions field $\mathbb{Z}/2\mathbb{Z}(X)$. Then $X$ is not in the image of the map $x \mapsto x^2$.
- If $K$ is finite and of characteristic $p$, then, by the pigeon hole principle, the homomorphism $x \mapsto x^p$ is surjective and hence an isomorphism.
- In fact, the fixed points of a homomorphism $R \to R$ of rings also form a subring of $\mathbb{R}$.

The subfield in Part 3 of Theorem on Field Homomorphisms is called the *fixed field* of the homomorphism $f$. A *fixed point* of $f$ is an element $x \in K$ such that $f(x) = x$. Thus, the fixed field of $f$ consists of all fixed points of $f$.

We apply the above result to the case where $K$ has positive characteristic.

**Theorem 7.4.31.** *Suppose that $K$ is a field of characteristic $p > 0$. Let $L$ be the smallest subfield of $K$ (isomorphic to $\mathbb{Z}/p\mathbb{Z}$) and let $q$ be a power of $p$. Then the following statements hold.*

1. $(x+y)^q = x^q + y^q$ *for all* $x, y \in K$.
2. *The map* $x \to x^q$ *is a homomorphism* $K \to K$.
3. *For each* $g \in L[X]$, *we have* $g(X^p) = (g(X))^p$.
4. *The subset* $\{x \in K | x^q = x\}$ *is a finite subfield of* $K$.
5. $\{x \in K | x^p = x\} = L$.

*Proof.* By parts:

$(x+y)^q = x^q + y^q$ **for all** $x, y \in K$**.**

By Newton‘s binomium, and the fact that all but the two extreme binomial coefficients are zero, $(x+y)^p = x^p + x^p$. To prove the equation with $q$ instead of $p$, we can use induction on the number $a$ such that $q = p^a$. Above we have establshed the case $a = 1$. Suppose we have dealt with the case $a-1$. Then, using the induction hypothesis and $\frac{q}{p} = p^{a-1}$, we find $(x+y)^q = \left((x+y)^{\frac{q}{p}}\right)^p = \left(x^{\frac{q}{p}} + y^{\frac{q}{p}}\right)^p = x^q + y^q$.

**The map** $x \to x^q$ **is a homomorphism** $K \to K$**.**

We need to verify:

- $(x \cdot y)^q = x^q \cdot y^q$.
- $(x+y)^q = x^q + y^q$.
- $0^q = 0, 1^q = 1$.

The first and third statement are obvious. The second has just been proved in Part 1 and the fact that $x^p = x$ for $x \in \mathbb{Z}/p\mathbb{Z}$ (known as Fermat’s Little Theorem).

**For each** $g \in L[X]$**, we have** $g(X^p) = (g(X))^p$**.**

$x \mapsto x^q$ is a homomorphism by Part 2.

**The subset** $\{x \in K | x^q = x\}$ **is a finite subfield of** $K$**.**

The subfield result follows from the Theorem on Field Homomorphisms. Finiteness follows from Zeros of Polynomials.

$\{x \in K | x^p = x\} = L$.

Write $M = \{x \in K | x^p = x\}$. By Part 4, $M$ is a subfield of $K$. Clearly, the smallest subfield $L$ of $K$ is contained in $M$, so we only need show that $M$ has no more than $p$ elements. But elements of $M$ are zeros of the polynomial $X^p - X$, and so there are at most $p$ solutions by Zeros of Polynomials.

□

**Example 7.4.32.** Consider the polynomial $f = X^4 + X + 1 \in \mathbb{Z}/2\mathbb{Z}[X]$. Since it is irreducible, the residue class ring $K = \mathbb{Z}/2\mathbb{Z}[X]/(f)\mathbb{Z}/2\mathbb{Z}[X]$ is a field. It has order 16. The map $x \mapsto x^4$ is a homomorphism $K \to K$. We wish to determine its fixed field $M = \{x \in K | x^4 = x\}$. Put $y = X + (f)K[X]$. Suppose $g = a \cdot y^3 + b \cdot y^2 + c \cdot y + d \in M$ Then, using $y^4 = y + 1$, $y^8 = y^2 + 1$, and $y^{12} = (y+1) \cdot \left(y^2+1\right) = y^3 + y^2 + y + 1$, we find $g^4 = a \cdot y^3 + (a+b) \cdot y^2 + (a+c) \cdot y + (a+b+c+d)$ From $g^4 = g$ we derive $a = 0, b = c$. Thus, $M = \left\{0, 1, y^2+y, y^2+y+1\right\}$, a subfield of order 4.

**Remark 7.4.33.** Part 3 need not hold if we replace $L$ by an arbitrary field of characteristic $p$. For instance, let $L$ be the rational function field $\mathbb{Z}/p\mathbb{Z}(Y)$. Then the polynomial $g(X) = Y \cdot X$ satisfies $(g(X))^p = (Y \cdot X)^p = Y^p \cdot X^p$ whereas $g(X^p) = Y \cdot X^p$.

**Remark 7.4.34.** If $K$ is finite, of order say $q$, it may happen that, for different powers $r, s$ of $p$, the maps $x \mapsto x^r$ and $x \mapsto x^s$ are identical. For instance, $r = 1 = p^0$ and $s = q$ both represent the identity on $K$.

**Definition 7.4.35.** Complex numbers that are zeros of nonzero polynomials in $\mathbb{Q}$ are called *algebraic*.

**Example 7.4.36** ($\sqrt{3}$)**.** Clearly, $\sqrt{3}$ is a zero of $X^2 - 3$. So it is algebraic.

**Example 7.4.37** ($\mathrm{e}^{\frac{2\cdot\pi\cdot i}{5}}$)**.** $\mathrm{e}^{\frac{2\cdot\pi\cdot i}{5}}$ is a zero of $X^5 - 1 = 0$. But it is not a zero of the linear factor $X - 1$, so it is a zero of the quotient: $X^4 + X^3 + X^2 + X + 1$.

**Example 7.4.38** ($2\cdot\cos\left(\frac{2\cdot\pi}{5}\right)$)**.** The number $2\cdot\cos\left(\frac{2\cdot\pi}{5}\right)$ is equal to $\mathrm{e}^{\frac{2\cdot\pi\cdot i}{5}} + \mathrm{e}^{\frac{(-2)\cdot\pi\cdot i}{5}}$ and also to $\frac{-1+\sqrt{5}}{2}$. Put $a = \mathrm{e}^{\frac{2\cdot i}{5}}$. Then, as we have seen in the previous example, $a^4 + a^3 + a^2 + a + 1 = 0$. Multiply by $a^{-2}$ and replace $a^2 + a^{-2}$ by $\left((a + (a^{-1}))\right)^2 - 2$. Then we have $\left((a + (a^{-1}))\right)^2 - 2 + \left(a + a^{-1}\right) + 1 = 0$, from which we conclude that $2\cdot\cos\left(\frac{2\cdot\pi}{5}\right) = a + a^{-1}$ is a zero of $X^2 + X - 1$.

**Remark 7.4.39.**
- Note that a polynomial of $\mathbb{C}[X]$ lies in $\mathbb{Q}[X]$ if and only if it has rational coefficients.
- An algebraic number is characterised by the fact that it generates a subfield of $\mathbb{C}$ that is finite-dimensional, when viewed as a vector space over $\mathbb{Q}$. For instance, e and $\pi$ are known not to be algebraic (although the proof is not easy).
- If $a$ is algebraic, then there is a polynomial of minimal degree of which $a$ is a zero. For, if $f$ and $g$ are both nonzero polynomials of which $a$ is a zero, then so is $\gcd(f, g)$.

- The notion of algebraic element exists for any field $K$ with a subfield $L$: an element of $K$ is called algebraic over $L$ if it is a zero of a nonzero polynomial in $K$.

If $x$ is algebraic, then $\mathbb{Q}(x)$ has finite dimension as a vector space over $\mathbb{Q}$. The converse is also true.

**Theorem 7.4.40.** *The set of all algebraic numbers in $\mathbb{C}$ is a subfield of $\mathbb{C}$.*

**Remark 7.4.41.** The crux of the matter is the following fact: Given two polynomials $f, g \in \mathbb{Q}[X]$, there are polynomials $h, k \in \mathbb{Q}[X]$, such that

- the sum of each root of $f$ and each root of $g$ is a root of $h$,
- the product of each root of $f$ and each root of $g$ is a root of $k$.

The proof of these statements is beyond the scope of these notes. But constructions of such polynomials were given in several examples.

**Example 7.4.42** ($\sqrt{3}+1$)**.** The number $\sqrt{3}+1$ is a zero of the polynomial $X^2-2\cdot X-2$.

**Example 7.4.43** ($\sqrt{3}+\sqrt{2}$)**.** $\sqrt{3}+\sqrt{2}$ is a zero of the polynomial $X^4-10\cdot X^2+1$.

We show how to find such a polynomial for $a=\sqrt{3}+\sqrt{2}$.

We look for a $\mathbb{Q}$-linear relation between the powers of $a$. First form $a^2=5+2\cdot\sqrt{6}$. The three elements $1, a, a^2$ are written as $\mathbb{Q}$-linear combinations of the independent elements $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$. Because we cannot yet expect a linear relation, we calculate the next power: $a^3=9\cdot\sqrt{3}+11\cdot\sqrt{2}$. Still no linear relation, so we continue: $a^4=49+20\cdot\sqrt{6}$. But now $a^4=10\cdot a^2-1$, so $a$ is a root of $X^4-10\cdot X^2+1=0$.

**Example 7.4.44** ($2^{\frac{1}{3}}+2^{\frac{1}{2}}$)**.** The number $2^{\frac{1}{3}}+2^{\frac{1}{2}}$ is a zero of the polynomial $X^6-6\cdot X^4-4\cdot X^3+12\cdot X^2-24\cdot X-4$.

We show how to find a polynomial of which $b=\sqrt[3]{2}+\sqrt{2}$ is a root. Computing powers of $b$, we find $\mathbb{Q}$-linear combinations of powers of $2^{\frac{1}{6}}$. Therefore, we determine a $7\times 6$ matrix whose rows are the powers of $b$, written out with respect to the basis $1, 2^{\frac{1}{6}}, \ldots, 2^{\frac{5}{6}}$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 2 & 2 & 6 & 6 & 2 \\ 0 & 0 & 4 & 0 & 2 & 8 \\ 1 & 2 & 8 & 4 & 0 & 4 \\ 0 & 2 & 0 & 4 & 1 & 0 \end{pmatrix} \tag{7.2}$$

Next, we look for a linear relation between the rows. This amounts to finding a vector in the kernel of the transposed matrix. As a row vector, this is $(-4,-24,12,-4,-6,0,1)$, which means that the polynomial $f = X^6 - 6{\cdot}X^4 - 4{\cdot}X^3 + 12{\cdot}X^2 - 24{\cdot}X - 4$ is as required. It is straightforward now to verify $f(b) = 0$.

## 7.5 Ideals

Ideals appear in the study of ring homomorphisms. They are very useful in the study of polynomial equations, and in the construction of rings by means of residue classes, in much the same way we have seen them come about in modular and polynomial arithmetic. As before, we only consider commutative rings. So, let $R$ be a commutative ring.

**Definition 7.5.1** (Definition of Ideal)**.** A nonempty subset $I$ of $R$ is an *ideal* of $R$ if, for all $a, b \in I$ and all $r \in R$ we have $a + r{\cdot}b \in I$.

An equivalent definition for $I$ to be an ideal is the following:

- $0 \in I$;
- for all $a \in I$ and $b \in I$ we have $a + b \in I$;
- for all $a \in I$ and $r \in R$ we have $r{\cdot}a \in I$;

**Example 7.5.2** (Usual arithmetic)**.** In the ring of integers $\mathbb{Z}$, the subset $n{\cdot}\mathbb{Z}$ of all multiples of $n$ is an ideal: if $a{\cdot}n$ and $b{\cdot}n$ are multiples of $n$, then $a{\cdot}n + b{\cdot}n = (a+b)\cdot n$ is a multiple of $n$. If furthermore $r$ is in $\mathbb{Z}$ and $a{\cdot}n$ is a multiple of $n$ then $r{\cdot}(a{\cdot}n) = (r{\cdot}a){\cdot}n$ is a multiple of $n$.

**Example 7.5.3** (Modular arithmetic)**.** In the ring $\mathbb{Z}/n\mathbb{Z}$, where $n$ is a multiple of $m \in \mathbb{Z}$, the set of all residue classes of multiples of $m$ is an ideal of $\mathbb{Z}/n\mathbb{Z}$, denoted again by $(m)\mathbb{Z}/n\mathbb{Z}$ or by $m{\cdot}\mathbb{Z}/n\mathbb{Z}$.

**Example 7.5.4** (Polynomial rings)**.** In the polynomial ring $R[X]$, the multiples of a given polynomial $f$ form an ideal.

In $\mathbb{Z}[X]$ the subset $\{f \in \mathbb{Z}[X] \,|\, f(2) = 0\}$ is an ideal:

- if $f(2) = 0$ and $g(2) = 0$, then $((f+g))\cdot 2 = f(2) + g(2) = 0 + 0 = 0$, and
- if $f(2) = 0$ and $r$ is an element of $\mathbb{Z}[X]$, then $r{\cdot}f(2) = r(2)\cdot f(2) = 0$.

**Example 7.5.5** (Residue class rings)**.** In the polynomial ring $R[X]/(f)R[X]$, the set of all multiples of the residue class of a divisor $g$ of $f$ is an ideal, denoted by $(g)R[X]/(f)R[X]$ or $g{\cdot}R[X]/(f)R[X]$.

**Example 7.5.6** (The Gaussian integers)**.** In the ring $R = \mathbb{Z} + \mathbb{Z}{\cdot}i$, the set of all elements $a + b{\cdot}i$ with $a \equiv b \pmod 2$ is an ideal.

**Remark 7.5.7.** Suppose $R$ is a non-commutative ring. Then there are three notions of ideal:

- **Left ideal:** A nonempty subset $I$ of $R$ such that, for all $a, b \in I$ and all $r \in R$, we have $a + b \in I$ and $r{\cdot}a \in I$.

- **Right ideal:** A nonempty subset $I$ of $R$ such that, for all $a,b \in I$ and all $r \in R$, we have $a+b \in I$ and $a\cdot r \in I$.
- **Two-sided ideal:** A subset of $R$ that is both left and right ideal.

Each ideal contains 0. The subsets $\{0\}$ and $R$ of $R$ are both ideals of $R$.

If a subset $V$ of $R$ is contained in an ideal $I$, then every combination $r_1\cdot v_1+r_2\cdot v_2+\ldots+r_n\cdot v_n$, with $r_1,r_2,\ldots,r_n \in R$ and $v_1,v_2,\ldots,v_n \in V$, also belongs to $I$. In fact, all these combinations form an ideal themselves.

**Theorem 7.5.8.** *Let $V$ be a nonempty subset of $R$. The subset of $R$ consisting of all combinations of the form $r_1\cdot v_1+r_1\cdot v_1+\ldots+r_n\cdot v_n$ with $r_1,\ldots,r_n \in R$ and $v_1,\ldots,v_n \in V$, is an ideal of $R$.*

*Proof.* Let $M$ be the indicated subset of $R$. We show that $M$ satisfies the three defining properties of an ideal.

**$0 \in M$.**

Taking $n=1$, $r_1=0$, and $v_1$ any element of $V$, we find $0=r_1\cdot v_1$ to be an element of $M$.

**If $x,y \in M$, then $x+y \in M$.**

Suppose that $r=r_1\cdot v_1+r_2\cdot v_2+\ldots+r_n\cdot v_n$ and $s=s_1\cdot w_1+s_2\cdot w_2+\ldots+s_m\cdot w_m$ are elements of $M$, with all $v_i$ and $w_j$ in $V$. Then $r+s=r_1\cdot v_1+r_2\cdot v_2+\ldots+r_n\cdot v_n+s_1\cdot w_1+s_2\cdot w_2+\ldots+s_n\cdot w_n$ also belongs to $M$.

**For each $r \in R$ and $m \in M$, we have $r\cdot m \in M$.**

If $m=r_1\cdot v_1+r_2\cdot v_2+\ldots+r_n\cdot v_n$ is an element of $M$, with $v_i \in V$, then for $r \in R$, we have $r\cdot m=r\cdot r_1\cdot v_1+r\cdot r_2\cdot v_2+\ldots+r\cdot r_n\cdot v_n$, which obviously belongs to $M$.

□

**Example 7.5.9.** Let $a \in R$ and put $V=\{a\}$. The ideal of the theorem is the set of all multiples of $a$; in formula: $\{r\cdot a \mid r \in R\}$. In the cases $R=\mathbb{Z}$ and $R=\mathbb{Q}$, these are exactly the elements equivalent to 0 modulo $a$. We shall see shortly that this is no coincidence. Notation: $a\cdot R$ or $(a)R$, as usual for, e.g., $R=\mathbb{Z}$ and $\mathbb{Q}[X]$.

**Example 7.5.10.** Different sets of generators $V$ may lead to the same ideal. For example take $V=\left\{X^2\cdot Y-1, X\cdot Y^2-1\right\}$ and $W=\left\{X-Y, X^3-1\right\}$ in the ring $R=\mathbb{Q}[X,Y]$. Then $\{V\}R=\{W\}R$. To see this, we write $X-Y=Y\cdot\left(X^2\cdot Y-1\right)-X\cdot\left(X\cdot Y^2-1\right)$ and $X^3-1=\left(X^2\cdot Y+1\right)\cdot\left(X^2\cdot Y-1\right)-X^3\cdot\left(X\cdot Y^2-1\right)$ from which we derive that $V$ is contained in $\{W\}R$, which implies that $\{V\}R$ is contained in $\{W\}R$.

Conversely, $X^2\cdot Y-1=\left(-(X^2)\right)\cdot(X-Y)+X^3-1$ and $X\cdot Y^2-1=\left(-X\cdot Y-X^2\right)\cdot(X-Y)+X^3-1$ whence the equality $\{V\}R=\{W\}R$.

**Example 7.5.11.** Suppose that $v_1,...,v_n \in R[X,Y]$ are polynomials. Then $v_1(x,y) = ... = v_n(x,y) = 0$ is a set of equations with unknown $x,y \in R$. Now, for any polynomial $f \in R[X,Y]$, we also have $f(x,y) = 0$. The reason is that $f$, being in $\{[v_1,...,v_n]\}R[X,Y]$, can be written as $r_1 \cdot v_1 + r_2 \cdot v_2 + ... + r_n \cdot v_n$ for suitable $r_1,...,r_n \in R[X,Y]$, so that $f(x,y) = r_1 \cdot v_1(x,y) + r_2 \cdot v_2(x,y) + ... + r_n \cdot v_n(x,y) = 0$.

This means that we can try and derive a lot of "easier" equations from the given one as a first step to solve the set of equations. For example, suppose that we have $v_1 = X^2 \cdot Y - 1$ and $v_2 = X \cdot Y^2 - 1$, so that the system of equations is $x^2 \cdot y = 1, x \cdot y^2 = 1$. Then also $f = Y \cdot v_1 - X \cdot v_2 = X - Y$ belongs to the ideal generated by $v_1$ and $v_2$, and so we also have $x = y$. Substituting this result in $v_2(x,y) = 0$, we find $x^3 = 1$, which is readily solved.

Of course, ad hoc methods may lead to the same result here. The indicated method however is part of an algorithm that works in all cases to bring the set of equations in a better form.

**Remark 7.5.12.** If $V$ is a subset of $R$, then the ideal generated by $V$ could also be defined as the intersection of all ideals containing $V$.

To see that the ideal defined in the theorem is exactly that, note that, if $I$ is an ideal containing $V$, then $I$ contains $V$. This implies that $V$ is contained in the intersection of all ideals containing $V$. On the other hand, the ideal defined in the theorem clearly contains $V$ and so coincides with the intersection.

It is called the ideal generated by $V$. Notation $\{V\}R$ or $\{V\}R$.

Let $R$ be a ring. Just like with subrings and submonoids, we can also describe generation of ideals by means of intersections.

**Theorem 7.5.13.** *If $C$ is a collection of ideals of $R$, then $\bigcap_{c \in C} c$ is also an ideal of $R$.*

*Proof.* Write $M = \bigcap_{c \in C} c$. We verify three criteria that suffice for $M$ to be an ideal.

$0 \in M$**.**

Each $I \in C$ is an ideal and so contains 0. Hence $0 \in \bigcap_{c \in C} c$, and so $0 \in M$.

**If** $x,y \in M$**, then** $x + y \in M$**.**

Each $I \in C$ is an ideal containing both $x$ and $y$, whence $x + y$, and so $x + y \in \bigcap_{c \in C} c$.

**For each** $r \in R$ **and** $m \in M$**, we have** $r \cdot m \in M$**.**

Each $I \in C$ is an ideal containing $m$ and hence $r \cdot m$. Therefore $\bigcap_{c \in C} c$ also contains $r \cdot m$.

□

**Example 7.5.14** (Usual arithmetic)**.** In the ring of integers $\mathbb{Z}$, the intersection of the ideals $m\cdot\mathbb{Z}$ and $n\cdot\mathbb{Z}$, for given integers $m,n$, is the ideal generated by $\operatorname{lcm}(m,n)$. For, this is clear if at least one of $m,n$ is zero. Otherwise, if $a \in (m)\mathbb{Z} \cap (n)\mathbb{Z}$, then $a$ is a multiple of both $m$ and $n$, and hence also of $\operatorname{lcm}(m,n)$. Thus, $a$ is in the ideal $\{\operatorname{lcm}(m,n)\}\mathbb{Z}$. This proves that the intersection $(m)\mathbb{Z} \cap (n)\mathbb{Z}$ is contained in the ideal $\{\operatorname{lcm}(m,n)\}\mathbb{Z}$. The other inclusion is obvious.

**Example 7.5.15** (Modular arithmetic)**.** In the ring $\mathbb{Z}/n\mathbb{Z}$, the intersection of the ideals $(g)\mathbb{Z}/n\mathbb{Z}$ and $(h)\mathbb{Z}/n\mathbb{Z}$ is $(\operatorname{lcm}(g,h))\mathbb{Z}/n\mathbb{Z}$.

This follows by a similar reasoning as used in the previous Example 7.5.14.

**Example 7.5.16** (Polynomial rings)**.** Just as for integers, the intersection of $(f)R[X]$ and $(g)R[X]$ is $(\operatorname{lcm}(f,g))R[X]$.

**Example 7.5.17** (Residue class rings)**.** Let $d$ be a polynomial in $R[X]$. In the residue class ring $R[X]/(d)R[X]$, just as for modular arithmetic, the intersection of $(f)R[X]/(d)R[X]$ and $(g)R[X]/(d)R[X]$ is $(\operatorname{lcm}(f,g))R[X]/(d)R[X]$.

**Example 7.5.18** (The Gaussian integers)**.** In the ring $R = \mathbb{Z} + \mathbb{Z}\cdot i$, the intersection of the ideals $1+i$ and 2 is $1+i$, as $2 = (1-i)\cdot(1+i)$.

Suppose that $V$ is a subset of $R$. We claim that $\{V\}R$, the ideal generated by $V$, coincides with $M$, the intersection over all ideals containing $V$. As noted , the ideal $\{V\}R$ is contained in $M$. But also, $\{V\}R$ contains $V$, so is one of the ideals over which the intersection forming $M$ is taken, so $M$ is contained in $\{V\}R$. Thus, $\{V\}R = M$.

In a ring $R$, the complete ring itself is an ideal.

The following is a characterization of this special ideal.

**Theorem 7.5.19** (Characterization of the ring as an ideal)**.** *Suppose that $I$ is an ideal of $R$. The following are equivalent.*

1. *$I = R$.*
2. *$1 \in I$.*
3. *$I$ contains an invertible element.*
4. *There are $v_1,\ldots,v_n \in I$ and $r_1,\ldots,r_n \in R$ such that $1 = r_1\cdot v_1 + \ldots + r_n\cdot v_n$.*

*Proof.*

**Part** 1 **implies Part** 2**.**

Suppose $I = R$. Then obviously, as $1 \in R$, also $1 \in I$.

**Part** 2 **implies Part** 3**.**

Clearly, 1 is an invertible element of $I$.

**Part 3 implies Part 4.**

Assume that $v$ is an element of $I$ with inverse $r$. Then $1 = r \cdot v$ is an expression as required in Assertion 4.

**Part 4 implies Part 1.**

Suppose that Assertion 4 holds: there are $v_1, v_2, \ldots, v_n \in I$ and $r_1, r_2, \ldots, r_n \in R$ such that $1 = r_1 \cdot v_1 + r_2 \cdot v_2 + \ldots + r_n \cdot v_n$. By the theorem on the previous page, the right-hand side belongs to $I$. As this expression is equal to 1, the identity element also belongs to $I$.

□

**Example 7.5.20.** Let $R$ be a field and $I$ an ideal of $R$ distinct from 0. Then there is an element in $I \setminus \{0\}$, which must be invertible (as $R$ is a field). By the theorem, $I = R$. We conclude that in fields there are no proper nonzero ideals.

**Example 7.5.21.** Suppose that $v_1, v_2, \ldots, v_n \in \mathbb{R}[X,Y]$ are polynomials and consider the corresponding set of equations (cf. the Characterization of the ring as an ideal). $v_1(x,y) = v_2(x,y) = \ldots = v_n(x,y) = 0$ with unknown $x, y \in R$.

If 1 belongs to the ideal generated by the $v_i$, then there are no solutions. For then 1 can be written as $r_1 \cdot v_1 + r_2 \cdot v_2 + \ldots + r_n \cdot v_n$ for suitable $r_1, r_2, \ldots, r_n \in \mathbb{R}[X,Y]$, so that the existence of a solution $(x,y) \in R^2$ would lead to $1 = 1(x,y) = r_1 \cdot v_1(x,y) + r_2 \cdot v_2(x,y) + \ldots + r_n \cdot v_n(x,y) = 0$, a contradiction.

For example, suppose that we have $v_1 = X^2 \cdot Y - 1, v_2 = X \cdot Y^2 - 1, v_3 = X - Y - 1$ Then also $1 = Y \cdot v_1 - X \cdot v_2 - v_3$ belongs to the ideal generated by $v_1, v_2, v_3$, and so the system $x^2 \cdot y = 1, x \cdot y^2 = 1, x - y = 1$ has no solutions.

We encountered generation as a means of constructing ideals. Here we discuss two more ways of obtaining ideals. Let $R$ be a commutative ring. For subsets $X$ and $Y$ of $R$, the sum $X + Y$ is the subset $\{a + b | (a,b) \in X \times Y\}$ of $R$.

**Theorem 7.5.22.** *If $I$ and $J$ are ideals of $R$, then the sum $I + J$ is an ideal of $R$.*

*Proof.* We verify the three laws for $I + J$ to be an ideal.

**$0 \in I + J$.**

Clearly, $0 = 0 + 0 \in I + J$.

**If $u, u' \in I + J$, then $u + u' \in I + J$.**

Suppose $u, u' \in I + J$. Then there are $x, x' \in I$ and $y, y' \in J$ such that $u = x + y$ and $u' = x' + y' \in J$. By commutativity of +, we find $u + u' = x + x' + y + y' \in I + J$.

**If $r \in R$ and $u \in I+J$, then $r \cdot u \in I+J$.**

Suppose $r \in R$ and $u = x+y \in I+J$ with $x \in I$ and $y \in J$. Then $r \cdot u = r \cdot x + r \cdot y \in I+J$.

□

**Example 7.5.23** (Usual arithmetic)**.** In the ring of integers $\mathbb{Z}$, the sum of the ideals $(m)\mathbb{Z}$ and $(n)\mathbb{Z}$, for given integers $m,n$, is the ideal $(\gcd(m,n))\mathbb{Z}$. To see this, let $a$ and $b$ be integers such that $a \cdot m + b \cdot n = \gcd(m,n)$ (they can be found by means of the Extended Euclidean Algorithm). This equality shows that $\gcd(m,n)$, and therefore every multiple of it, belongs to the ideal generated by $m$ and $n$. This shows that the ideal $(\gcd(m,n))\mathbb{Z}$ is contained in the ideal $(m)\mathbb{Z} + (n)\mathbb{Z}$. On the other hand, every element $c \cdot m + d \cdot n$ of the sum ideal $(m)\mathbb{Z} + (n)\mathbb{Z}$ is a multiple of $\gcd(m,n)$, since both $m$ and $n$ are multiples of this gcd.

**Example 7.5.24** (Modular arithmetic)**.** Fix a nonzero integer $d > 1$. Suppose $m,n$ are integers representing residue classes of the ring $\mathbb{Z}/\mathrm{d}\mathbb{Z}$. If $m$ and $n$ divide $d$, then the sum of the ideals $(m)\mathbb{Z}/\mathrm{d}\mathbb{Z}$ and $(n)\mathbb{Z}/\mathrm{d}\mathbb{Z}$ of $\mathbb{Z}/d\mathbb{Z}$ is the ideal $(\gcd(m,n))\mathbb{Z}/d\mathbb{Z}$.

**Example 7.5.25** (Polynomial rings)**.** Just as for integers, in the polynomial ring $R[X]$, with $R$ a field, the sum of the ideals $f$ and $g$ equals the ideal $(\gcd(f,g))R[X]$ whenever $f,g \neq 0$.

**Example 7.5.26** (Residue class rings)**.** Let $d$ be a polynomial in $R[X]$, where $R$ a field. In the residue class ring $R[X]/(d)R[X]$, just as for modular arithmetic, the sum of $(f)R[X]/(d)R[X]$ and $(g)R[X]/(d)R[X]$ is $(\gcd(f,g))R[X]/(d)R[X]$.

**Example 7.5.27** (The Gaussian integers)**.** In the ring $R = \mathbb{Z} + \mathbb{Z} \cdot i$, the sum of the ideals $(1+i)$ and $(1-i)$ is $(1+i)R$, as $1-i = (-i) \cdot (1+i)$.

**Remark 7.5.28.** The ideal $I+J$ can also be described as the ideal generated by $I$ and $J$.

Let $S$ also be a commutative ring. The kernel of a homomorphism $f \colon R \to S$ is the subset $\{x \in R | f(x) = 0\}$ of $R$.

**Theorem 7.5.29** (The Kernel of a Ring Homomorphism is an Ideal)**.** *If $f \colon R \to S$ is a homomorphism of rings, then the kernel* $\operatorname{Ker}(f)$ *is an ideal of $R$.*

*Proof.* We verify the three laws for $\operatorname{Ker}(f)$ to be an ideal:

**$0 \in \operatorname{Ker}(f)$.**

Clearly, $f(0) = 0$, so $0 \in \operatorname{Ker}(f)$.

**If $u,v \in \operatorname{Ker}(f)$, then $u+v \in \operatorname{Ker}(f)$.**

Suppose $u,v \in \operatorname{Ker}(f)$. Then $f(u+v) = f(u) + f(v) = 0+0 = 0$, so $u+v \in \operatorname{Ker}(f)$.

**If $r \in R$ and $u \in \mathrm{Ker}\,(f)$ , then $r{\cdot}u \in \mathrm{Ker}\,(f)$.**

Suppose $r \in R$ and $u \in \mathrm{Ker}\,(f)$. Then $f\,(r{\cdot}u) = f\,(r) \cdot f\,(u) = f\,(r) \cdot 0 = 0$, so $r{\cdot}u \in \mathrm{Ker}\,(f)$.

□

**Example 7.5.30** (Usual arithmetic)**.** The kernel of the natural homomorphism $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ is the ideal $(m)\mathbb{Z}$.

**Example 7.5.31** (Modular arithmetic)**.** If $m$ divides $n$, then there is a homomorphism $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}, x + (n)\mathbb{Z} \mapsto x + (m)\mathbb{Z}$ Its kernel is the ideal generated by the residue class of $m$.

**Example 7.5.32** (Polynomial rings)**.** Fix $x \in \mathbb{Q}$. The kernel of the homomorphism $\mathbb{Q}\,[X] \to \mathbb{Q}, f\,(X) \mapsto f\,(x)$ is the ideal generated by $X - x$. Prove this!

**Example 7.5.33** (Residue class rings)**.** Just like the modular arithmetic case: If $f, g$ are polynomials in $\mathbb{Q}\,[X]$ such that $g$ divides $f$, then there is a homomorphism $\mathbb{Q}\,[X]\,/(f)\mathbb{Q}\,[X] \to \mathbb{Q}\,[X]\,/(g)\mathbb{Q}\,[X]\,, h + (f)\mathbb{Q}\,[X] \mapsto h + (g)\mathbb{Q}\,[X]$ Its kernel is generated by the residue class of $g$.

**Example 7.5.34** (The Gaussian integers)**.** The map $f = \mathbb{Z} + \mathbb{Z}{\cdot}i \to \mathbb{Z}/2\mathbb{Z}, a + b{\cdot}i \mapsto a + b$ is a homomorphism. Check:

- $f\,(1) = f\,(0 + i) = 1$.
- $f\,((a + b{\cdot}i) \cdot (c + d{\cdot}i)) = f\,(a{\cdot}c - b{\cdot}d + (a{\cdot}d + b{\cdot}c) \cdot i) = a{\cdot}c + b{\cdot}d + a{\cdot}d + b{\cdot}c = (a + b) \cdot (c + d) = f\,(a + b{\cdot}i) \cdot f\,(c + d{\cdot}i)$.

Its kernel is the ideal generated by $1 + i$.

**Remark 7.5.35.** The theorem [?] is crucial in what follows. It will be used to describe the image ring $\mathrm{Im}\,(f)$ fully in terms of $R$.

We shall see later that every proper ideal of $R$ can be seen as the kernel of some homomorphism.

For a positive integer $n$, the ring $\mathbb{Z}/n\mathbb{Z}$ is a domain if and only if $n$ is a prime. This notion of prime will be generalized to arbitrary ideals. Later, the notion of residue classes will be extended beyond $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Q}\,[X]\,/(d)\mathbb{Q}\,[X]$ to residue class rings with respect to arbitrary ideals, and it will turn out that primality has the same role as for $\mathbb{Z}/n\mathbb{Z}$.

Let $R$ be a commutative ring and let $I$ be an ideal of $R$. We say that $I$ is proper if it is not equal to $R$.

**Definition 7.5.36.** • $I$ is called a prime ideal if it is proper and, for all $a, b \in I$ the equation $a{\cdot}b = 0$ implies $a \in I$ or $b \in I$.

- $I$ is called maximal if it is proper and if there exists no proper ideal strictly containing $I$.

**Example 7.5.37** (Usual arithmetic)**.** A simple example of a prime ideal is the ideal 0 in the ring $\mathbb{Z}$ : if $a{\cdot}b = 0$, then of course $a = 0$ or $b = 0$. In fact, the same argument shows that in any domain the zero ideal is prime. The ideal $(6)\mathbb{Z}$ of multiples of 6 in $\mathbb{Z}$ is not a prime ideal:

$2 \cdot 3 = 6$ and neither 2 nor 3 is a multiple of 6. For every prime number $p$ the ideal $(p)\mathbb{Z}$ of multiples of $p$ in the ring $\mathbb{Z}$ is maximal: if an ideal $J$ strictly contains $(p)\mathbb{Z}$, then it contains an integer $m$ which is not a multiple of $p$. But then $p$ and $m$ are relatively prime and, by the Extended Euclidean Algorithm, there is a relation $a \cdot m + b \cdot p = 1$. But this implies that 1 is contained in the ideal $J$ and that $J = \mathbb{Z}$. Hence each ideal that strictly contains $p$ coincides with $\mathbb{Z}$, so $(p)\mathbb{Z}$ is maximal. The ideal $(0)\mathbb{Z}$ of $\mathbb{Z}$ is prime but not maximal: for example, the ideal $(2)\mathbb{Z}$ is proper and contains $(0)\mathbb{Z}$.

**Example 7.5.38** (Modular arithmetic)**.** The ideal $(m)\mathbb{Z}/n\mathbb{Z}$ in $\mathbb{Z}/n\mathbb{Z}$ is prime if and only if $\gcd(m,n)$ is a prime number. If $n$ is prime, then, as $\mathbb{Z}/n\mathbb{Z}$ is a field, the only maximal ideal of $\mathbb{Z}/n\mathbb{Z}$ is $(0)\mathbb{Z}/n\mathbb{Z}$. If $p$ is a proper prime divisor of $n$, then $(p)\mathbb{Z}/n\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}/n\mathbb{Z}$.

**Example 7.5.39** (Polynomial rings)**.** In $\mathbb{Z}[X]$ the ideal $(X)\mathbb{Z}[X]$ is prime: if the product of two polynomials is divisible by $X$, then at least one of them is already divisible by $X$.

In the ring $\mathbb{R}[X]$, the ideal $(X)\mathbb{R}[X]$ is maximal: if the ideal $J$ strictly contains $(X)\mathbb{R}[X]$, then it contains a polynomial $f$ with a nonzero constant term $a$. But then it follows that the ideal $J$ contains $a$ itself and so also the element 1. We conclude from Characterization of the ring as an ideal that $J = R$.

**Example 7.5.40** (Residue class rings)**.** This case is very similar to modular arithmetic. For example, the ideal generated by $X^2+1$ is prime in $\mathbb{Q}[X]/(X^4-1)\mathbb{Q}[X]$, but not in $\mathbb{C}[X]/(X^4-1)\mathbb{C}[X]$. The same ideal is maximal in $\mathbb{Q}[X]/(X^4-1)\mathbb{Q}[X]$, and not in $\mathbb{C}[X]/(X^4-1)\mathbb{C}[X]$. In the latter case, $X-i$ and $X+i$ are two maximal ideals containing $X^2+1$.

**Example 7.5.41** (The Gaussian integers)**.** In the ring $R = \mathbb{Z} + \mathbb{Z} \cdot i$, the ideal generated by $i-2$ is prime and maximal. Of course this requires an argument. The ideal generated by 2 is not prime: $(1-i) \cdot (1+i) = 2$.

**Example 7.5.42.** To show that an ideal $M$ is maximal, one often reasons as follows: suppose that there is an ideal $J$ that strictly contains $M$. Then try to show (using that there are elements in the ideal $J$ that are not contained in $M$) that $J$ contains the identity element and therefore equals the whole ring.

Although the definitions of prime and maximal ideals look very different, there are important connections between the two notions. For instance one implies the other.

**Theorem 7.5.43.** *A maximal ideal is prime.*

*Proof.* Suppose that $M$ is a maximal ideal of the commutative ring $R$. Let $a$ and $b$ be elements of $R$ such that $a \cdot b \in M$.

If neither $a$ nor $b$ belongs to $M$, then $R = a + M$ and $R = b + M$, because of the maximality of $M$. This implies the existence of elements $r, s \in R$ and $m, n \in M$ such that $1 = a \cdot r + m, 1 = b \cdot s + n$.

Multiplying left-hand sides and right-hand sides yields $1 = a\cdot b\cdot r\cdot s + a\cdot r\cdot n + b\cdot s\cdot m + m\cdot n$.

As $a\cdot b, m, n \in M$, we find $1 \in M$, a contradiction.

Hence $a$ or $b$ belongs to $M$, proving that $M$ is a prime ideal.

□

**Example 7.5.44.** If $R$ is a field, then the only proper ideal of $R$ is $\{0\}$. It is both maximal and prime.

**Remark 7.5.45** (Converse)**.** The converse does not hold. If $R = \mathbb{Z}[X]$, then the ideal of $R$ generated by 2 and the ideal of $R$ generated by $X$ are prime ideals; but they are not maximal, the ideal of $R$ generated by both of these being a bigger proper ideal.

## 7.6 Residue class rings

In this section, arithmetic modulo an integer $n$ or modulo a polynomial $d$ is generalized to arithmetic modulo an ideal.

Let $I$ be an ideal in the commutative ring $R$. Two elements $a, b \in R$ are called congruent modulo $I$ if their difference $a-b$ belongs to $I$. Notation if $I$ is clear from the context: $a \equiv b$

**Theorem 7.6.1.** *Congruence modulo I is an equivalence relation.*

*Proof.* To show that the relation is indeed an equivalence relation we have to check that the relation is reflexive, symmetric, and transitive.

**Congruence is reflexive**

$a \equiv a$, since $a - a = 0$ and so belongs to $I$.

**Congruence is symmetric**

Suppose

$a \equiv b$. Then $a-b$ belongs to $I$ and hence so does $-(a-b) = b-a$. But this means $b \equiv a$.

**Congruence is transitive**

Suppose

$a \equiv b$ and $b \equiv c$. Then $a-b$ and $b-c$ belong to $I$ and then so does their sum $a-b+b-c = a-c$. But this means that $a \equiv c$.

□

**Example 7.6.2.** If $R = \mathbb{R}[X]$ and $I = \{\{X - x, Y - y\}\}R$ for certain $x, y \in \mathbb{R}$, then $f$ and $g$ are congruent modulo $I$ in $R$ if and only if $f(x,y) = g(x,y)$.

**Remark 7.6.3.** Equivalence modulo $I$ generalizes both

- Congruence mod $n$ in $\mathbb{Z}$. For $a, b \in \mathbb{Z}$ we have $a \equiv b \pmod{n}$ if and only if $a$ and $b$ are congruent modulo $(n)\mathbb{Z}$.
- Congruence mod $d$ in $\mathbb{Q}[X]$. For $a, b \in \mathbb{Q}$ we have $a \equiv b$ if and only if $a - b \in (d)\mathbb{Q}[X]$.

An equivalence class is called a residue class. The set of all residue classes is denoted by $R/I$. An element of $R/I$ is denoted by $a + I$ when we are precise, and simply by $a$ if there is no danger of confusion.

**Theorem 7.6.4.** *The set $R/I$ inherits from $R$ the following ring structure:*

- *addition:* $a + I + (b + I) = a + b + I$,
- *multiplication:* $(a + I) \cdot (b + I) = a \cdot b + I$,
- *identity element:* $1 + I$,
- *zero element:* $0 + I$.

*Proof.* The definitions involve implicitly the choices of representatives, so we need to check that they do not depend on these choices.

Suppose $a' + I = a + I$ and $b' + I = b + I$. Then $a' = a + r$ and $b' = b + s$ for some $r, s \in I$. Now both $a' + b' - (a + b) = r + s$ and $a' \cdot b' - a \cdot b = a \cdot s + r \cdot b + r \cdot s$ clearly belong to $I$. We conclude that $a' + b' + I = a + b + I$ and $a' \cdot b' + I = a \cdot b + I$, so that addition and multiplication are well defined.

It remains to check the definitions of the ring axioms. These are routine checks and are left to the reader.

□

**Example 7.6.5.** Let $R = \mathbb{Z}/4\mathbb{Z}[X]/\{[2, X^2]\}\mathbb{Z}/4\mathbb{Z}[X]$. Its elements are (represented by) $0, 1, X, X + 1$. The product $X \cdot (X + 1)$ is (represented by) $X$, for $X \cdot (X + 1) - X$ is equal to $X^2$, which belongs to $\{[2, X^2]\}\mathbb{Z}/4\mathbb{Z}[X]$. We write down the multiplication table of this ring.

| $\cdot$ | $0$ | $1$ | $X$ | $1+X$ |
|---|---|---|---|---|
| $0$ | $0$ | $0$ | $0$ | $0$ |
| $1$ | $0$ | $1$ | $X$ | $1+X$ |
| $X$ | $0$ | $X$ | $0$ | $X$ |
| $1+X$ | $0$ | $1+X$ | $X$ | $1$ |

Instead of $0+I$, we also write just $I$. In particular, we might work with the identifications $0=0+I=I$, which exemplify computing modulo $I$: as if all elements of $I$ are equal to zero.

The ring $R/I$ is called the residue class ring or quotient ring of $R$ modulo $I$.

Homomorphisms relate rings modulo an ideal. Let $R$ and $S$ be commutative rings. The image of a homomorphism $R \to S$ can be entirely described in terms of $R$.

**Theorem 7.6.6** (First isomorphism theorem)**.** *If $f\colon R \to S$ is a homomorphism of rings, then $R/\mathrm{Ker}\,(f)$ is isomorphic to the image $\mathrm{Im}\,(f)$.*

*Proof.* Put $I=\mathrm{Ker}\,(f)$. By The Kernel of a Ring Homomorphism is an Ideal, this is an ideal of $R$. We shall prove the following two assertions.

**There is a homomorphism $f'\colon R/I \to S$ such that, for each $x \in R$ we have $f\,(x+I)=f\,(x)$.**

The map $f$ is determined by the requirement $f\,(x+I)=f\,(x)$. It needs to be verified that $f$ is well defined. For, if $x+I=y+I$, then $x-y \in I$ and, as $I=\mathrm{Ker}\,(f)$, we find $f\,(x-y)=0$. As $f$ is a homomorphism, it follows that $f\,(x)=f\,(y)$. Thus, indeed, the definition of $f$ does not depend on the choice of $y \in x+I$. It is easy to see that $f'$ is a homomorphism.

**The homomorphism $f'$ is injective.**

By The Kernel of a Ring Homomorphism is an Ideal, it suffices to prove that $\mathrm{Ker}\,(f')=\{0\}$. Suppose, to this end, that $x+I \in \mathrm{Ker}\,(f')$. Then $f\,(x)=0$, and so $x \in \mathrm{Ker}\,(f)$, which is $I$. Consequently, $x+I=I=0 \in R/I$. So indeed, $\mathrm{Ker}\,(f')=0$.

As the images of $f$ and $f'$ coincide, the above two statements prove the theorem.

□

**Example 7.6.7** (Usual arithmetic)**.** The kernel of the natural homomorphism $f\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is $(n)\mathbb{Z}$. This homomorphism is surjective, and has kernel $n$. Application of the theorem now gives the obvious fact that $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to $\mathbb{Z}/(n)\mathbb{Z}$.

**Example 7.6.8** (Modular arithmetic)**.** If $n$ is a multiple of $m$, then there is a homomorphism $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}, x+(n)\mathbb{Z} \mapsto x+(m)\mathbb{Z}$. This homomorphism is surjective and its kernel is $(m)\mathbb{Z}/n\mathbb{Z}$, so the theorem implies that that there is an isomorphism $\mathbb{Z}/n\mathbb{Z}/(m) \to \mathbb{Z}/m\mathbb{Z}$.

**Example 7.6.9** (Polynomial rings)**.** Let $x \in \mathbb{C}$. Consider the homomorphism $f\colon \mathbb{Q}\,[X] \to \mathbb{Q}\,[x]\,, g \mapsto g\,(x)$. Clearly, $f$ is surjective. Let us determine its kernel. Observe that $g \in \mathbb{Q}\,[X]$ lies in $\mathrm{Ker}\,(f)$ if and only if $x$ is a zero of $g$. Thus, $\mathrm{Ker}\,(f)=0$ if $x$ is not algebraic. Otherwise, there is a unique polynomial $d \in \mathbb{Q}$ with leading coefficient 1 of minimal degree such that $d\,(x)=0$. By the Extended Euclidean Algorithm for Polynomials it readily follows that $\mathrm{Ker}\,(f)=(d)\mathbb{Q}\,[X]$.

**Example 7.6.10** (Residue class rings)**.** Similarly to the modular arithmetic case, we find, for $g$ a divisor of $f$, that $\mathbb{Q}\,[X]\,/(f)\mathbb{Q}\,[X]\,/(g) \to \mathbb{Q}\,[X]\,/(g)\mathbb{Q}\,[X]$ is an isomorphism.

**Example 7.6.11** (The Gaussian integers)**.** Consider the homomorphism $\mathbb{Z}[X] \to \mathbb{Z}+\mathbb{Z}\cdot i, f(X) \mapsto f(i)$. Clearly, $X^2+1$ is in the kernel of this homomorphism. On the other hand $\mathbb{Z}[X]/(X^2+1)\mathbb{Z}[X]/$ is readily seen to be isomorphic to $\mathbb{Z}+\mathbb{Z}\cdot i$. A close analysis of the proof of the theorem gives that the kernel of the homomorphism must coincide with $(X^2+1)\mathbb{Z}[X]$.

Let $R$ be a commutative ring with ideal $I$. Here is what prime ideals and maximal ideals mean in the context of quotient rings:

**Theorem 7.6.12.** *The quotient ring $R/I$ is*

- *a domain if and only if the ideal $I$ is prime;*
- *a field if and only if the ideal $I$ is maximal.*

*Proof.* Observe that for two elements $a, b \in R$ the following holds: $a\cdot b \in I$ if and only if $(a+I)\cdot(b+I) = I$. For, the left hand side is equal to $a\cdot b+I$

There are four assertions to be verified.

**$I$ prime implies $R/I$ is a domain.**

Suppose that $I$ is a prime ideal. We need to show that the quotient ring has no zero divisors. Suppose that $a+I$ and $b+I$ are elements whose product is the zero element: $(a+I)\cdot(b+I) = I$. This comes down to $a\cdot b$ belonging to $I$. As $I$ is a prime ideal $a$ or $b$ belongs to $I$. In other words: $a+I = I$ or $b+I = I$. This shows that $R/I$ is a domain.

**$R/I$ is a domain implies $I$ prime Suppose that $R/I$ is a domain.**

If $a\cdot b \in I$, then it follows that $a+I = I$ or $b+I = I$. But this means $a \in I$ or $b \in I$.

**If $I$ is maximal then $R/I$ is a field.**

Suppose that $I$ is a maximal ideal. Let $a+I$ be a nonzero element of $R/I$; that is, $a$ does not belong to $I$. Then the ideal $a+I$ is strictly bigger than $I$. Maximality of $I$ implies $a+I = R$. In particular, there exist $b \in R$ and $c \in I$ with $a\cdot b+c = 1$. Thus, $a\cdot b+I = 1+I$, from which we derive $(a+I)\cdot(b+I) = 1+I$. Hence $a+I$ is invertible in $R/I$. This establishes that $R/I$ is a field.

**$R/I$ is a field implies that $I$ is maximal.**

Conversely, suppose that $R/I$ is a field. Let $J$ be an ideal of $R$ strictly containing $I$. Then there is $a \in J\setminus I$, so $a+I \neq I$. Thus being nonzero, $a+I$ has a multiplicative inverse: for some $b \in R$ we have $(a+I)\cdot(b+I) = 1+I$. But then $a\cdot b-1 \in I$, that is, $1 \in a$ + I, so $1 \in J$, whence $J = R$. This establishes that $I$ is a maximal ideal.

□

**Example 7.6.13.** Consider $R = \mathbb{Z}[X]$ and $f = X^2 + 1$ in $R$. Then $(f)R$ is a prime ideal, since $f$ is irreducible in $\mathbb{Q}$, and so $R/(f)R$ is a domain. Observe that $R/(f)R$ is isomorphic to the Gaussian integers, which were shown page 166 be a domain. The ideal $(f)R$ is not maximal, as $R/(f)R$, the Gaussian integers, do not form a field.

The quotient ring obtained from $R$ by modding out the ideal $\{X+1\}R/(f)R$ leads to the residue class ring $\mathbb{Z}+\mathbb{Z}\cdot i/(i+1)\mathbb{Z}+\mathbb{Z}\cdot i$, which is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, a field. Therefore, $\{X+1\}R/(f)R$ is a maximal ideal.

**Remark 7.6.14.** The theorem generalises two known cases:

- Ideals in $\mathbb{Z}$. The ideal $n$ of $\mathbb{Z}$ is prime if and only if $n$ is prime; then $\mathbb{Z}/(n)\mathbb{Z}$ is a field, so then $(n)\mathbb{Z}$ is even maximal.
- Ideals in polynomial rings. Similar to the previous case, each ideal of $K$, where $K$ is a field, is of the form $f$ for some polynomial $f$. The ideal is prime if and only if it is maximal, in which case $f$ is irreducible.

In the next section, we shall take a closer look at finite fields.

## 7.7 Finite fields

For $p$ a prime number and $f$ an irreducible polynomial of degree $n$ in $\mathbb{Z}/p\mathbb{Z}[X]$, the quotient ring $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$ is a field with $p^n$ elements. We will see that any field is essentially of this form.

Let $F$ be a finite field of order $q$. By Order of a finite field, we know that $q = p^a$, the power of a prime number $p$. We start our investigation of $F$ with yet another (but more general) version of Fermat's little theorem.

**Theorem 7.7.1** (Fermat's little theorem)**.** *Each $x \in F$ satisfies the equation*

$$x^q = x \tag{7.3}$$

*In particular, we have*

$$X^q - X = \prod_{x \in F} (X - x) \tag{7.4}$$

*Proof.* If $x = 0$, then clearly, $x^q = x$. Suppose, therefore, $x \neq 0$.

Since $F \setminus \{0\}$ consists of invertible elements, the multiplicative group $F^\times$ of $F$ has order $q-1$. Now $x$ belongs to this group. By Order of an element, we find that $x^{q-1} = 1$. The required equation follows when we multiply both sides by $x$.

The above implies that for each $x$ in $F$ the linear polynomial $X-x$ is a factor of $X^q - X$. As $\gcd(X-x, X-y) = 1$ for $x, y \in F$ with $x \neq y$, their product $\prod_{x\in F}(X-x)$ divides $X^q - X$. But both polynomials are of degree $q$, and their leading coefficients are both 1, so they are equal.

□

**Example 7.7.2** (Fields of order 9)**.** Each element of a field of order 9 is a zero of the polynomial $X^9 - X \in \mathbb{Z}/3\mathbb{Z}[X]$. The elements 0, 1, and 2 of $\mathbb{Z}/3\mathbb{Z}$ are zeros of this polynomial and correspond to the linear factors $X, X-1, X-2$. Dividing out these factors, we find a polynomial of degree 6 that factors into a product of three quadratic polynomials as follows:

$$\left(X^2+X+2\right)\cdot\left(X^2+2\cdot X+2\right)\cdot\left(X^2+1\right) \tag{7.5}$$

Each of these factors can be used to define a field of order 9.

In the Classification of finite fields we shall see that they all lead to the same field up to isomorphism. That means that the fields $\mathbb{Z}/3\mathbb{Z}[X]/(X^2+X+2)\mathbb{Z}/3\mathbb{Z}[X]$, $\mathbb{Z}/3\mathbb{Z}[X]/(X^2+2\cdot X+2)\mathbb{Z}/3\mathbb{Z}[X]$, and $\mathbb{Z}/3\mathbb{Z}[X]/(X^2+1)\mathbb{Z}/3\mathbb{Z}[X]$ are isomorphic to each other.

On the other hand, Fermat's little theorem says that if we pick one of these fields, say the first $F = \mathbb{Z}/3\mathbb{Z}[X]/(X^2+X+2)\mathbb{Z}/3\mathbb{Z}[X]$, then all three quadratic factors are reducible, considered as polynomials over $F$.

Let $x$ denote the residue class in $F$ of $X$. Then, by construction, $x$ is a zero of $X^2+X+2$. Since, by Frobenius Automorphisms taking third powers is an automorphism of $F$ preserving this quadratic polynomial, $x^3 = -x + -1$ is the other zero. Computing remainders we find $x^4 = 2$. Consequently, $x^2$ is a zero of the third quadratic polynomial, and, by an argument as before, so is its third power $x^6 = -(x^2)$. We have accounted for all powers $x^i$ that may occur except for $i = 5, 7$. It is readily checked that they are zeros of the second quadratic polynomial and that one is the third power of the other.

**Example 7.7.3.** 
- In Frobenius Automorphisms, we saw that, for any power $r = p^b$ of $p$, the subset $\{x \in F | x^r = x\}$ is a subfield of $F$. Apparently,
  - for $r = q$, the subfield coincides with $F$;
  - the subfield only depends on the value of $\mathrm{rem}(b, a)$, where $q = p^a$.
- Note that $x^{q-1} = 1$ for nonzero $x$ in $F$.

We derive some more properties of finite fields. But first a lemma.

**Lemma 7.7.4.** *Suppose that m and n are positive integers. Then*

$$\gcd(X^m - 1, X^n - 1) = X^{\gcd(m,n)} - 1 \tag{7.6}$$

*In particular, m divides n if and only if $X^m - 1$ divides $X^n - 1$.*

*Proof.* We first prove the special case: if $m$ divides $n$ then $X^m - 1$ divides $X^n - 1$.

So suppose that $m$ divides $n$. Then the following identity holds:

$$X^n - 1 = (X^m - 1) \cdot \left(X^{n-m} + X^{n-2m} + \ldots + X^m + 1\right) \tag{7.7}$$

In particular $X^m - 1$ divides $X^n - 1$.

Next we derive the general case:

$$\gcd\left(X^m - 1, X^n - 1\right) = X^{\gcd(m,n)} - 1 \tag{7.8}$$

Suppose $n = q \cdot m + r$ Then, by the above,

$$\gcd\left(X^m - 1, X^n - 1\right) = \gcd\left(X^m - 1, X^n - 1 - X^r \cdot (X^{q \cdot m} - 1)\right) = \gcd\left(X^m - 1, X^r - 1\right) \tag{7.9}$$

Thus, $n$ can be replaced by the remainder of division of $n$ by $m$. But this is the first step of Euclid's algorithm, which can be repeated and repeated, until one of the arguments of the gcd is $X^{\gcd(m,n)} - 1$, and the other 0.

This proves the lemma.

□

We use Fermat's little theorem to prove the following result, announced before.

**Theorem 7.7.5.** *The multiplicative group of a finite field of order $q$ is cyclic of order $q - 1$.*
*In particular, every finite field contains a primitive element.*

*Proof.* Let $F$ be a finite field of order $q$. By Order of a finite field there is a prime number $p$ and a positive integer $a$ such that $q = p^a$.

Suppose $m$ is a natural number dividing $q - 1$. We show that the number of elements $x \in F^\times$ with $x^m = 1$ equals $m$. It is precisely the number of solutions of $X^m - 1$ in $F$. As $m | q - 1$, the polynomial $X^m - 1$ divides $X^{q-1} - 1$, whence also $X^q - X$, see Lemma on polynomials. By a Fermat's little theorem, the latter polynomial decomposes into a product of linear factors in $F$. But then its divisor $X^m - 1$ is also a product of $m$ linear factors. Hence, there are $m$ solutions of $X^m - 1$ in $F$. In other words, the number of elements $x \in F^\times$ with $x^m = 1$ equals $m$.

Finally, we apply the Characterisation of cyclic groups to conclude that $F^\times$ is cyclic.

Any generator of this cyclic group is then a primitive element for $F$.

□

**Example 7.7.6.** Suppose that $K$ is a field of order $32$. Then $K^\times$ is a group of order $31$. Each element distinct from $1$ in $K^\times$ has order $31$, as its order is a divisor of $31$ and distinct from $1$.

Consider the polynomial $f = X^{31} - 1$. In $\mathbb{Z}/2\mathbb{Z}[X]$, the polynomial $f$ factors into

$$f = (1+X)\cdot\left(1+X^2+X^5\right)\cdot\left(1+X^3+X^5\right)\cdot\left(1+X+X^2+X^3+X^5\right)\cdot \\ \left(1+X+X^2+X^4+X^5\right)\cdot\left(1+X+X^3+X^4+X^5\right)\cdot\left(1+X^2+X^3+X^4+X^5\right).$$

Let $a$ be an element of $K$ which is a zero of $1+X+X^2+X^3+X^5$. Then an elementary calculation shows that $a^2$ is also a zero of this polynomial. Indeed, this follows from $1+X^2+X^4+X^6+X^8+X^{10} = \left(1+X+X^2+X^3+X^4+X^5\right)^2$. The five zeros of the polynomial are therefore $a, a^2, a^4, a^8, a^{16}$. This result could also have been derived by applying Frobenius Automorphisms with the Frobenius map $x \mapsto x^2$.

By the theorem, there are always primitive elements in finite fields. If $g$ is a primitive element of the finite field $F$, then the elements can be easily enumerated by their exponents with respect to $g$ : $F = \{0\} \cup \left\{g^i \middle| i \in \{0,\ldots,q-2\}\right\}$. When written in this form, multiplication on the nonzero elements of $F$ is given by modular arithmetic, with modulus $q-1$. This is very efficient, but addition is less convenient. Thus, we have the opposite to the usual form, where addition is a minor effort, but multiplication is harder.

The following algorithm checks whether an element is primitive, it is used in the second algorithm which provides us with a prmitive element.

**Algorithm 7.7.7** (Is Primitive?). • *Input: an element a in a field F of order q for which the prime divisors $p_1$, ..., $p_k$ of $q-1$ are known.*

- *Output: true if a is primitive, false otherwise.*

```
IsPrimitive := procedure($a$)
local variables
    $t := 1$
    $k$
    $p_1, \ldots, p_k$
while $\left(a^{\frac{q-1}{p_t}} \neq 1\right) \wedge (t \leq k)$ do
    $t := t+1$

if $t > n$
    then
        return
            $true$
    else
        return
            $false$
```

**Algorithm 7.7.8** (Primitive element). • *Input: field F of order q for which the prime divisors $p_1$, ..., $p_k$ of $q-1$ are known.*

- *Output: primitive element a of F .*

```
PrimitiveElement := procedure($F$)
local variables
    $a :=$ RandomElement$(F)$
while $\neg$(IsPrimitive$(a)$) do
    $a :=$ RandomElement$(F)$

return
    $a$
```

**Theorem 7.7.9** (Characterization of subfields of finite fields)**.** *Let $F$ be a field of order $q = p^n$, where $p$ is a prime. Suppose $K$ is a subfield of $F$. Then*

$$K = \{x \in F | x^r = x\} \tag{7.10}$$

*for some $r = p^m$, where $m$ divides $n$.*
*The subfield $K$ is thus the unique subfield of order $r$.*

*Proof.* Suppose $K$ is a subfield of $F$. Then the order of $K$ equals $r = p^m$ for some $m$. By Fermat's little theorem every element $x$ of $K$ is a root of $X^r - X$. As this polynomial has at most $r$ roots, we can write $K$ as the set $\{x \in F | x^r = x\}$.

Moreover, we also see that $X^r - X$ divides $X^q - X$ and hence $X^{r-1} - 1$ divides $X^{q-1} - 1$. Now Lemma on polynomials implies that $r - 1 = p^m - 1$ divides $q - 1 = p^n - 1$. Applying Lemma on polynomials once more, we find that $m$ divides $n$.

The uniqueness of $K$ follows immediately.

□

Suppose $K$ is a subfield of the field $F$ and $a$ is an element of $F$. Then $K(a)$ denotes the smallest subfield of $F$ containing $K$ and $a$.

We will give a description of $K(a)$ as a quotient of a polynomial ring over $K$. To this end we introduce the concept of minimum polynomial:

**Definition 7.7.10.** Suppose $K$ is a subfield of the field $F$ and $a$ is an element of $F$. Then a polynomial $f$ in $K[X]$ is called a *minimal polynomial* for $a$ if $f$ is a monic polynomial of minimal degree having $a$ as root.

**Example 7.7.11.** In some cases a minimal polynomial does exist, in other cases not.

For example, $X^2 + 1$ is a minimal polynomial over $\mathbb{R}$ for $i \in \mathbb{C}$. The minimal polynomial for $\sqrt[3]{2}$ over $\mathbb{Q}$ is $X^3 - 2$.

The elements in $\mathbb{C}$ that do have a minimal polynomial over $\mathbb{Q}$ are precisely the algebraic elements in $\mathbb{C}$. So, the elements $\pi$ and e do not have a minimal polynomial over $\mathbb{Q}$. Indeed, these elements are not algebraic.

**Theorem 7.7.12** (Uniqueness and irreducibility of the minimal polynomial)**.** *If an element $a$ has a minimal polynomial over the field $K$ then this polynomial is the unique minimal polynomial for $a$, it is irreducible, and divides every other polynomial that has $a$ as a root.*

*Proof.* Let $f$ be a minimal polynomial for $a$. If $g$ is another minimal polynomial for $a$, then both polynomials are of the same degree. Moreover $a$ is also a root of $f-g$. As this polynomial is of lower degree than $f$, it has to be 0. So $f=g$, proving uniqueness of the minimal polynomial.

If the polynomial $f$ can be written as a product $g\cdot h$ with $g$ and $h$ both monic and of positive degree, then $a$ is a root of $g$ or $h$, both contradicting that $f$ is a minimal polynomial of $a$. This proves $f$ to be irreducible.

If $g$ is a polynomial with $a$ as a root, then the remainder of $g$ divided by $f$ has also root $a$. As the degree of this remainder is smaller than the degree of $f$, we find a contradiction, unless the remainder is 0. This proves that $f$ divides $g$.

□

If $K$ is a subfield of the finite field $F$, then, by Characterization of subfields of finite fields, there is a prime power $q$ such that the order of $K$ equals $q$ and the order of $F$ is $q^n$ for some $n$.

The Frobenius map

$$\mathrm{phi}\colon F\to F, x\mapsto x^q \tag{7.11}$$

fixes precisely those elements that are in $K$.

Now let $a$ be an element in $F$ with minimal polynomial $f$. Then we find that elements $a$, $\mathrm{phi}\,(a)$, $\mathrm{phi}^2\,(a)$, ... are all roots of $f$.

Denote by $t$ the smallest positive integer with $\mathrm{phi}^t\,(a)$ equal to $a$. Such $t$ exists, as $\mathrm{phi}^n\,(a)=a$, as follows from Fermat's little theorem.

The elements $\mathrm{phi}\,(a)$, $\mathrm{phi}^2\,(a)$,...,$\mathrm{phi}^t\,(a)$ are all distinct roots of $f$. So $f$ has degree at least $t$. On the other hand, $\prod_{i=1}^{t}\left(X-a^{q^i}\right)$ is a polynomial invariant under the Frobenius automorphism phi. This implies the following result:

**Proposition 7.7.13.** *Let $a$ be an element from the finite field $F$.
The minimal polynomial of $a$ equals*

$$\prod_{i=1}^{t}\left(X-a^{q^{i}}\right) \tag{7.12}$$

*where $t$ is the smallest positive integer with $a^{q^t}$ equal to $a$.*

The above result also provides an algorithm to find the minimal polynomials.

**Algorithm 7.7.14** (Minimal polynomial)**.** • *Input: element $a$ in a finite field $F$.*

- *Output: minimal polynomial of $a$ over the a subfield of order $q$ of $F$.*

```
MininalPolynomial := procedure($a, q$)
local variables
    | $t := 1$
    | $f := X - a$
while $\neg\left(a^{q^t} = a\right)$ do
    | $f := f \cdot \left(X - a^{q^t}\right)$ , $t := t + 1$

return
    | $f$
```

We are now in a position to give the following description of the field $K(a)$ in terms of the minimal polynomial of $a$.

**Theorem 7.7.15** (Subfields and minimal polynomials)**.** *Let $K$ be be subfield of $F$.
Suppose $a$ is an element of $F$ with minimal polynomial $m$ over $K$.
Then $K(a)$ is isomorphic to $K[X]/(m)K[X]$.*

*Proof.* Consider the map

$$\text{rho}\colon K[X]/(m)K[X] \to K(a), f+(m)K[X] \mapsto f(a) \tag{7.13}$$

Notice that this map is well defined.

The map rho is homomorphims of fields. The image of rho contains $a$ and is contained in $K(a)$. Hence $K[X]/(m)K[X]$ is isomorphic to its image $K(a)$ under the map rho.

□

To test whether an element is in a subfield can be done using the following algorithm.

**Algorithm 7.7.16** (Subfield membership test)**.** • *Input: elements $a$ and $b$ in a finite field $F$ and a subfield $K$.*

- *Output: true if*

  *$b$ is an element of the a subfield $K(a)$ of $F$, false otherwise.*

```
Membership := procedure$(a,b)$
local variables
    $m :=$ MinimalPolynomial$(a,|K|)$
    $d :=$ degree$(m)$
if $b^{q^d} = b$
    then
        return
            $true$
    else
        return
            $false$
```

By $\mathrm{Irr}(d,p)$ we denote the set of all monic and irreducible polynomials $f \in \mathbb{Z}/p\mathbb{Z}[X]$ of degree $d$.

**Theorem 7.7.17** (Product of irreducible polynomials)**.** *If $f \in \mathbb{Z}/p\mathbb{Z}[X]$ is an irreducible polynomial of degree $n$, then $f$ divides $X^{p^n} - X$.*
*More precisely,*

$$X^{p^n} - X = \prod_{f \in \{g \in \mathrm{Irr}(d,p) \mid d|n\}} f \qquad (7.14)$$

*Proof.* Let $f$ be an irreducible polynomial in $\mathbb{Z}/p\mathbb{Z}[X]$ of degree $d$. Then consider the field $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$. Inside this field we find that the element $X + (f)\mathbb{Z}/p\mathbb{Z}[X]$ is a zero of $f$. This implies that over $\mathbb{Z}/p\mathbb{Z}(X)$ the gcd of $f$ and $X^{p^d} - X$ is nonzero. As $f$ is irreducible, this gcd equals $f$, from which we deduce that $f$ is a divisor of $X^{p^d} - X$.

Now Lemma on polynomials implies that $f$ divides $X^{p^n} - X$ if and only if $d$ divides $n$.

On the other hand, if $d$ divides $n$ then any irreducible polynomial $f$ of degree $d$ divides $X^{p^n} - X$ and hence has a root $r$ in $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$. As the multiplicity of this root $r$ as zero of $X^{p^n} - X$ is one, we find that $f^2$ does not divide $X^{p^n} - X$. So indeed, as the leading coefficient of $X^{p^n} - X$ is 1, we find that $X^{p^n} - X$ is equal to the product of all irreducibles in $\mathrm{Irr}(d,p)$ with $d$ dividing $n$.

□

**Example 7.7.18.** In $\mathbb{Z}/2\mathbb{Z}[X]$, the polynomial $X^4 - X$ factors as the following product of irreducibles:

$$X^4 - X = X \cdot (X-1) \cdot \left(X^2 + X + 1\right) \tag{7.15}$$

proving that $X^2 + X + 1$ is the unique irreducible polynomial of degree 2 in $\mathbb{Z}/2\mathbb{Z}[X]$.

In $\mathbb{Z}/3\mathbb{Z}[X]$ we find

$$X^9 - X = X \cdot (X-1) \cdot (X+1) \cdot \left(X^2 + 1\right) \cdot \left(X^2 - X - 1\right) \cdot \left(X^2 + X - 1\right) \tag{7.16}$$

We deduce that there are exactly three monic irreducible polynomials of degree 2 in $\mathbb{Z}/3\mathbb{Z}[X]$, namely, $X^2 + 1$, $X^2 - X - 1$ and $X^2 + X - 1$.

In general, since $X^{p^2} - X$ is the product of all monic irreducible polynomials of degree 1 and 2 from $\mathbb{Z}/p\mathbb{Z}[X]$, and the fact that there are $p$ monic polynomials of degree 1, there are exactly $\frac{p^2-p}{2}$ monic irreducible polynomials of degree 2 in $\mathbb{Z}/p\mathbb{Z}[X]$.

The results obtained so far lead to the following theorem, which is the main result of this section.

**Theorem 7.7.19** (Classification of finite fields)**.** *For every prime p and positive integer n there exists a field F of order $p^n$. This field is unique up to isomorphism.*
*In particular, there exists an irreducible polynomial f in $\mathbb{Z}/p\mathbb{Z}[X]$ of degree n, and, for any such polynomial f, the field F is isomorphic with $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$.*

*Proof.* We prove existence and uniqueness.

**Existence of $F$.**

To show that for every prime power $p^n$ there exists a field with exactly $p^n$ elements, it suffices to construct a finite field $F$ in which $X^{p^n} - X$ factors into linear factors: for then the subfield of $F$ of elements $x$ satisfying $x^{p^n} = x$ has $p^n$ elements.

So we start with $h = X^q - X$ in $\mathbb{Z}/p\mathbb{Z}[X]$. (Here $q = p^n$.) If $h$ factors into linear factors, we are done. If not, then choose an irreducible factor $h_1$ of $h$ and consider the field $K_1 = \mathbb{Z}/p\mathbb{Z}[X]/(h_1)\mathbb{Z}/p\mathbb{Z}[X]$. In this field, $X + (h_1)\mathbb{Z}/p\mathbb{Z}[X]$ is a zero of $h_1$, so $h_1$, whence also $h_1$, has a linear factor in $K_1[X]$. (Notice the new role of the indeterminate $X$; we index it by 1 in order to distinguish it from the previous $X$.) If $h$ does not completely factor into linear factors in $K_1[X]$, then choose an irreducible factor $h_2$ of $h$ (irreducible in $K_1[X]$), and construct $K_2 = K_1[X]/(h_2)K_1[X]$, etc. Since the number of linear factors increases in every step, this process must terminate and produces a field containing $\mathbb{Z}/p\mathbb{Z}$ in which $h$ factors into linear factors, proving the existence of a field of order $q$.

**Existence of $f$.**

Let $a$ be a primitive element of $F$ and $f$ its minimal polynomial over the prime field $\mathbb{Z}/p\mathbb{Z}$. This polynomial is irreducible. Moreover, by Subfields and minimal polynomials we find that $\mathbb{Z}/p\mathbb{Z}(a)$, which clearly equals $F$, is isomorphic to $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$. This implies that, indeed, there exists an irreducible polynomial $f$ of degree $n$ such that $F$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}[X]/(f)\mathbb{Z}/p\mathbb{Z}[X]$.

**Uniqueness.**

Let $g$ be any other irreducible polynomial of degree $n$ in $\mathbb{Z}/p\mathbb{Z}[X]$. Then, by Product of irreducible polynomials, $g$ divides $X^q - X$. In particular, as the latter polynomial factors into linear terms over $F$, we can find a root $x$ in $F$ of $g$. This implies that the map

$$\text{rho}\colon \mathbb{Z}/p\mathbb{Z}[X]/(g)\mathbb{Z}/p\mathbb{Z}[X] \to K, k+(g)\mathbb{Z}/p\mathbb{Z}[X] \mapsto k(x) \tag{7.17}$$

is well defined. This map is an injective homomorphism of the field into $F$. As both fields have the same order, it is an isomorphism.

□

**Example 7.7.20.** To construct a field of order $81 = 3^4$, we look for an irreducible polynomial $f$ of degree 4 in $\mathbb{Z}/3\mathbb{Z}[X]$. According to the theory, $f$ is a divisor of the polynomial $X^{81} - X$. We first divide out the roots belonging to the subfield of order 9: $\frac{X^{81}-X}{X^9-X} = X^{72} + X^{64} + X^{56} + X^{48} + X^{40} + X^{32} + X^{24} + X^{16} + X^8 + 1$. This polynomial will factor into 18 irreducible polynomials of degree 4. We find one by trial and error: Creating a degree 4 polynomial and checking that it is relatively prime with $X^9 - X$. The 18 choices for $f$ that may arise are:

$X^4 - X^2 - 1$, $X^4 - X^2 - X + 1$, $X^4 - X^3 + X^2 + 1$, $X^4 + X^3 - X + 1$, $X^4 + X^3 + X^2 - X - 1$, $X^4 + X^2 - 1$, $X^4 - X^3 - 1$, $X^4 + X - 1$, $X^4 + X^3 - 1$, $X^4 - X^3 + X + 1$, $X^4 - X^3 + X^2 + X - 1$, $X^4 + X^2 + X + 1$, $X^4 - X^3 - X^2 + X - 1$, $X^4 - X^3 + X^2 - X + 1$, $X^4 + X^3 - X^2 - X - 1$, $X^4 + X^3 + X^2 + X + 1$, $X^4 - X - 1$, and $X^4 + X^3 + X^2 + 1$.

We end this section with an algorithm testing irreducibility of a polynomial.

**Algorithm 7.7.21** (Testing irreducibility)**.** • *Input: polynomial $f$ of degree $n$ in the polynomial ring $\mathbb{Z}/p\mathbb{Z}[X]$*

- *Output: true if $f$ is irreducible and false otherwise.*

```
Irreducible := procedure($f$)
local variables
    $t := 1$
while $\gcd\left(f, X^{p^t} - X\right) = 1$ do
    $t := t + 1$

if $t = n$
    then
        return
            $true$
    else
        return
            $false$
```

*Proof.*

**Termination.**

The while-loop will certainly stop when $t$ reaches the value $n$; see Product of irreducible polynomials.

**Correctness.**

If $f$ is reducible, then it will be divisible of some irreducible polynomial of degree $t$ less than $n$. This implies, by Product of irreducible polynomials that the gcd of $f$ and $X^{p^t} - X$ is not 1. In this case the algorithm will return $false$.

□

## 7.8 Exercises

### 7.8.1 The structure ring

**Exercise 7.8.1.** Determine in each of the following cases whether the indicated set is a subring of $\mathbb{C}$.

1. $\left\{x + y \cdot i \cdot \sqrt{2} \;\middle|\; x, y \in \mathbb{Z}\right\}$
2. $\left\{x + y \cdot \sqrt[3]{2} \;\middle|\; x, y \in \mathbb{Z}\right\}$.

**Exercise 7.8.2.** Let $R$ be a ring and let $f\colon R \to R$ be a ring homomorphism. Prove that the subset $S$ of $R$ consiting of the elements $r$ with $f(r) = r$ form a subring of $R$.

**Exercise 7.8.3.** Let $a$ be an invertible element in the ring $R$ and let $f\colon R \to R$ be defined by $f(r) = a \cdot r \cdot a^{-1}$ for all $r$ in $R$. Prove that $f$ is an isomorphism and determine its inverse.

**Exercise 7.8.4.** For which elements $a$ in the ring $R$ is the map $f\colon R \to R$ given by $f(r) = a \cdot r \cdot a$ for all $r$ in $R$ a ring homomorphism?

**Exercise 7.8.5.** Let $S$ be a nonempty set and $R$ a ring. Show that the set of all maps from $S$ to $R$ is a ring, where the sum and product of two elements $f$ and $g$ is defined as follows:

$$(f+g)(r) = f(r) + g(r) \tag{7.18}$$

$$(f \cdot g)(r) = f(r) \cdot g(r) \tag{7.19}$$

for all $r$ in $R$.

**Exercise 7.8.6.** Prove that the following maps are ring homomorphisms.

1. $f\colon \mathbb{Q}[X] \to \mathbb{Q}, f \mapsto f(2)$, evaluation in 2.
2. $g\colon \mathbb{Z}[X] \to \mathbb{Z}/p\mathbb{Z}[X]$, given by reduction of the coefficients of the polynomial modulo $p$.

**Exercise 7.8.7.** Let $R$, $S$ and $T$ be rings and let $f\colon R \to S$ and $g\colon S \to T$ be ring homomorphisms.

1. Show that the composition $g \circ f\colon R \to T$ of $f$ and $g$ is a ring homomorphism.
2. Show that $\mathrm{Ker}(g \circ f)$ equals $\mathrm{Ker}(f)$ if $g$ is a ring isomorphism.
3. Show that $\mathrm{Ker}(g)$ equals $f(\mathrm{Ker}(g \circ f))$ if $f$ is a ring isomorphism.

**Exercise 7.8.8.** Show that the map $F\colon \mathbb{Z}[X] \to \mathbb{Q}, f \mapsto f\left(\frac{1}{2}\right)$ is a homomorphism of rings. What is its image and what its kernel?

**Exercise 7.8.9.** If one replaces every $X$ in a polynomial $f$ in $\mathbb{Q}[X]$ by $a \cdot X + b$, then the new polynomial can be written as $f(a \cdot X + b)$ and is again an element from $\mathbb{Q}[X]$.

Let $F\colon \mathbb{Q}[X] \to \mathbb{Q}[X], f \mapsto f(2 \cdot X + 3)$.

Show that $F$ is an isomorphism of rings and determine its inverse.

## 7.8.2 Constructions with rings

**Exercise 7.8.10.** Prove that $\mathbb{Q}$ is not finitely generated.

**Exercise 7.8.11.** Provide two nonisomorphic rings with 4 elements.

Give a proof that they are nonisomorphic!

### 7.8.3 Domains and fields

**Exercise 7.8.12.** Prove the converse of the Cancellation law for domains: if for all nonzero $a$ the implication holds that for all elements $r$ and $s$ in $R$ we have that $a \cdot r = a \cdot s$ implies that $r = s$, then $R$ is a domain.

**Exercise 7.8.13.** Let $F$ be a field and phi: $F \to F$ an isomorphism. Then the subset of $F$ consisting of all elements $x$ with phi$(x) = x$ is a subfield of $F$ . Prove this.

**Exercise 7.8.14.** What is the set of zero divisors in the cartesian product of two rings $R$ and $S$?

**Exercise 7.8.15.** Prove or disprove:

1. A subring of a domain is a domain.
2. A subring of a field is a field.

### 7.8.4 Fields

**Exercise 7.8.16.** What is the dimension of $\mathbb{Q}\left(\sqrt[3]{2}\right)$ as a vector space over $\mathbb{Q}$? Give a basis for this vector space.

**Exercise 7.8.17.** Prove that $\frac{1}{\sqrt{2}} + \sqrt{3}$ and $\sqrt[3]{3} - \sqrt{3}$ are algebraic numbers.

**Exercise 7.8.18.** Determine all isomorphism from the field $\mathbb{Q}\left(i \cdot \sqrt[4]{2}\right)$ to itself.

Find for each such isomorphism the field of fixed points.

**Exercise 7.8.19.** Let $K$ be a field and $R$ a subring of $K$. Then $R$ is a domain.

Let $Q$ be the field of fractions of $R$.

Prove that $Q$ is isomorphic with a subfield of $K$.

### 7.8.5 Ideals

**Exercise 7.8.20.** In each of the following cases determine the ideal generated by $V$ in the ring $R$.

1. $V = \{2\}$ in $R = \mathbb{Z}/6\mathbb{Z}$;
2. $V = \{2,3\}$ in $R = \mathbb{Z}/6\mathbb{Z}$;
3. $V = \{2\}$ in $R = \mathbb{Z}/8\mathbb{Z}$;

4. $V = \{2, X\}$ in $R = \mathbb{Z}[X]$.

**Exercise 7.8.21.** Let $I$ be the subset of $\mathbb{Z}[X]$ consisting of all polynomials $f$ with $f(0)$ divisible by $n$ for some fixed integer $n$.

Prove that $I$ is an ideal of $\mathbb{Z}[X]$.

**Exercise 7.8.22.** Let $f\colon R \to S$ be a homomorphism between rings and suppose $J$ is an ideal of $S$.

Prove that the set $f^{-1}(J)$ is an ideal of $R$.

**Exercise 7.8.23.** In the ring $R = \mathbb{Z} + \mathbb{Z}\cdot i$, the ideal generated by $i - 2$ is prime and maximal.

Prove this.

**Exercise 7.8.24.** Which of the following ideals is maximal or prime?

1. The ideal generated by 3 and $X^2$ in $\mathbb{Z}[X]$.
2. The ideal generated by 3 and $X$ in $\mathbb{Z}[X]$.
3. The ideal generated by $X^2 + 1$ in $\mathbb{Q}[X]$.

**Exercise 7.8.25.** Let $I$ be the subset of $\mathbb{Z}[X]$ consisting of all polynomials $f$ with $f(0)$ divisible by 5

Prove that $I$ is a maximal ideal of $\mathbb{Z}[X]$.

### 7.8.6 Residue class rings

**Exercise 7.8.26.** Which of the following ideals is maximal or prime?

1. The ideal generated by 3 and $X^2$ in $\mathbb{Z}[X]$.
2. The ideal generated by 3 and $X$ in $\mathbb{Z}[X]$.
3. The ideal generated by $X^2 + 1$ in $\mathbb{Q}[X]$.

**Exercise 7.8.27.** Let $I$ be the subset of $\mathbb{Z}[X]$ consisting of all polynomials $f$ with $f(0)$ divisible by 3

Prove that $\mathbb{Z}[X]/I$ is a field.

**Exercise 7.8.28.** Consider the residue class ring $R = \mathbb{Z}/4\mathbb{Z}[X]/\{2, X^2\}\mathbb{Z}/4\mathbb{Z}[X]$.

How many elements does this quotient ring have?

Is this ring is isomorphic with $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ or $\mathbb{Z}/4\mathbb{Z}$? Or to none of these? Give a proof for your answer.

**Exercise 7.8.29.** Show that the quotient ring $\mathbb{Q}[X]/(X^2+X+2)\mathbb{Q}[X]$ is isomorphic to $\mathbb{Q}\left(\sqrt{-7}\right)$

**Exercise 7.8.30.** Consider the homomorphism $\mathbb{Q}[X] \to \mathbb{Q}\left(\sqrt{2}\right), f(X) \mapsto f\left(\sqrt{2}\right)$. Prove that the kernel of this homomorphism is a maximal ideal.

### 7.8.7 Finite fields

**Exercise 7.8.31.** Determine all the primitive elements in the fields $\mathbb{Z}/5\mathbb{Z}$, $\mathbb{Z}/7\mathbb{Z}$ and $\mathbb{Z}/11\mathbb{Z}$.

**Exercise 7.8.32.** What is the minimal polynomial over $\mathbb{Q}$ of the following complex numbers: $1+i$, $2+\sqrt[3]{2}$ and $2-i\cdot\sqrt{3}$.

**Exercise 7.8.33.** Construct fields of order 27, 32 and 125.

**Exercise 7.8.34.** How many subfields has a field of order 64?

# Chapter 8

# Groups

Groups have been introduced in Definition of a group as abstract sets with some operations. However, groups often appear as transformations mapping a set to itself. For example, in the group of real invertible $n \times n$-matrices, each element determines a bijective linear map $\mathbb{R}^n \to \mathbb{R}^n$. Such group actions on a set enable us to analyze the group in a concrete setting. But it is also a means of unveiling the symmetries of the structures on that set. In this chapter, we look into the way a group can be represented by letting it act on a set or a structure.

## 8.1 Permutation groups

Let $X$ be a set. Recall that $\mathrm{Sym}(X)$ denotes the group of all bijections from $X$ to $X$ itself, multiplication being composition.

**Definition 8.1.1** (Definition of a permutation group)**.** A *permutation group* on $X$ is a subgroup of $\mathrm{Sym}(X)$.

- The elements of the set $X$ are called *points*.
- The size of $X$, denoted by $|X|$, is called the *degree* of the permutation group.

**Example 8.1.2.** Let $X$ be the set $\{1,\ldots,n\}$. Then of course the symmetric and alternating group on $X$ act as permutation groups on $X$. But there are many more permutation groups on $X$.

For example, the cyclic group of order $m$, where $m$ is at most $n$ can be seen as a permutation group: it is generated by $(1,\ldots,m)$.

The following definition expresses what it means to view an arbitrary group as a group of permutations.

**Definition 8.1.3** (Permutation representation)**.** A *permutation representation* of a group $G$ on a set $X$ is a homomorphism of groups $H \to \mathrm{Sym}(X)$.

The size of $X$, denoted by $|X|$, is called the *degree of the permutation representation*.

Given a permutation representation of a group $G$ on a set $X$, the group $G$ is said to *act* on $X$. We also speak of the *action* of $G$ on $X$.

- If $f$ is a permutation representation of the group $G$ on $X$, and $g \in G$, then we often write $g(x)$ for the image of $x$ under $g$, instead of the more complete expression $f(g)(x)$.
- If $f\colon H \to \mathrm{Sym}(X)$ is injective, then $f$ determines an isomorphism $H \to \mathrm{Im}(f)$. In this case, identifying $H$ with its image $\mathrm{Im}(f)$ under $f$, we sometimes call the group $H$ itself a permutation group on $X$.

**Example 8.1.4** ($\mathrm{Sym}_3$ acting by conjugation)**.** Consider the group $\mathrm{Sym}_3$ acting on its elements by conjugation. If the numbers $1,2,3,4,5,6$ correspond to the following elements $(),(1,2),(2,3),(1,3),(1,2,3),(1,3,2)$ of $\mathrm{Sym}_3$, then the image of this group in $\mathrm{Sym}_6$ consists of the following permutations $\{(),(2,3)(5,6),(1,2)(5,6),(1,3)(5,6),(1,2,3),(1,3,2)\}$.

**Example 8.1.5** (The action of $\mathrm{Sym}_5$ on pairs)**.** The group $\mathrm{Sym}_5$ is a permutation group on the set $\{1,2,3,4,5\}$. But $\mathrm{Sym}_5$ also permutes pairs from this set. For instance, the permutation $(1,2,3)$ transforms the pair $\{2,4\}$ into the pair $\{3,4\}$. Let $X$ be the set consisting of all subsets of $\{1,2,3,4,5\}$ of size 2 (there are 10 such pairs). Because each element $g$ of $\mathrm{Sym}_5$ is injective, $\{g(a),g(b)\}$ has size 2 whenever $\{a,b\}$ is a subset of $\{1,2,3,4,5\}$ of size 2. Hence we can define, for each $g \in \mathrm{Sym}_5$, the following map:

$$\mathrm{OnPairs}(g) : X \to X, \{a,b\} \mapsto \{g(a),g(b)\} \tag{8.1}$$

The map $\mathrm{OnPairs}(g)$ is bijective, the inverse of $\mathrm{OnPairs}(g)$ being $\mathrm{OnPairs}\left(g^{-1}\right)$. Indeed, for an arbitrary element $\{a,b\}$ of $X$ we have

$$\begin{aligned}\mathrm{OnPairs}(g)\cdot\mathrm{OnPairs}\left(g^{-1}\right)(\{a,b\}) &= \mathrm{OnPairs}(g)\left(\mathrm{OnPairs}\left(g^{-1}\right)(\{a,b\})\right)\\ &= \mathrm{OnPairs}(g)\left(\mathrm{OnPairs}\left(g^{-1}\right)\left(\{g^{-1}(a),g^{-1}(b)\}\right)\right)\\ &= \left\{g\left(g^{-1}(a)\right),g\left(g^{-1}(b)\right)\right\}\\ &= \{a,b\}\end{aligned} \tag{8.2}$$

and similarly

$$\mathrm{OnPairs}\left(g^{-1}\right)\cdot\mathrm{OnPairs}(g)(\{a,b\}) = \{a,b\} \tag{8.3}$$

Finally, we check that the map $g \mapsto \mathrm{OnPairs}(g)$ is a homomorphism $\mathrm{Sym}_5 \to \mathrm{Sym}(X)$, so that we indeed have a permutation representation. Let $g$, $h$ be arbitrary elements of $\mathrm{Sym}_5$. We need to verify that $\mathrm{OnPairs}(g\cdot h) = \mathrm{OnPairs}(g)\cdot\mathrm{OnPairs}(h)$, that is, that left and right hand side represent the same bijection. This is straightforward: for each unordered pair $\{a,b\}$ in $X$ we have

$$\begin{aligned} \mathrm{OnPairs}(g)\cdot\mathrm{OnPairs}(h)(\{a,b\}) &= \mathrm{OnPairs}(g)(\mathrm{OnPairs}(h)(\{a,b\})) \\ &= \mathrm{OnPairs}(g)(\{h(a),h(b)\}) \\ &= \{g(h(a)),g(h(b))\} \\ &= \{g\cdot h(a),g\cdot h(a)\} \\ &= \mathrm{OnPairs}(g\cdot h)(\{a,b\}) \end{aligned} \tag{8.4}$$

If we name the subsets of $\{1,2,3,4,5\}$ of size 2 by letters as follows: $a=\{1,2\}$, $b=\{1,3\}$, $c=\{1,4\}$, $d=\{1,5\}$, $e=\{2,3\}$, $f=\{2,4\}$, $g=\{2,5\}$, $h=\{3,4\}$, $i=\{3,5\}$, $j=\{4,5\}$. Then we can represent the elements from $\mathrm{Sym}_5$ as permutations of these letters. For example:

$$(1,2)\mapsto(b,e)(c,f)(d,g) \tag{8.5}$$

$$(1,2,3)\mapsto(a,e,b)(c,f,h)(d,g,i) \tag{8.6}$$

and

$$(1,2,3,4,5)\mapsto(a,e,h,j,d)(b,f,i,c,g) \tag{8.7}$$

Of course we can restrict the action on the pairs of $\{1,2,3,4,5\}$ to any subgroup of $\mathrm{Sym}_5$. In particular to $\mathrm{Alt}_5$.

**Example 8.1.6** (The action of $\mathrm{Sym}_n$ on subsets)**.** We generalise the above example as follows. Let $K$ be the set of all subsets of $\{1,...,n\}$ of size $k$. Thus, $|K|=\left(\begin{array}{c} n \\ k \end{array}\right)$. Each permutation $g$ in $\mathrm{Sym}_n$ acts on $K$ as follows. The set $X$ in $K$ is mapped to the set

$$\{g(x)|x\in X\} \tag{8.8}$$

This defines a bijection $g_K\colon K\to K$, and so $g_K$ is an element of $\mathrm{Sym}(K)$. The map that assigns to $g$ in $\mathrm{Sym}_n$ the element $g_K$ of $\mathrm{Sym}(K)$, is a homomorphism $\mathrm{Sym}_n\to\mathrm{Sym}(K)$ and hence a permutation representation of $\mathrm{Sym}_n$.

Of course we can restrict the action on $K$ to any subgroup of $\mathrm{Sym}_n$. This way, for instance, we obtain also permutation representations of $\mathrm{Alt}_n$.

**Example 8.1.7.** The general linear group $GL(n,\mathbb{R})$ acts on the set of vectors in $\mathbb{R}^n$. Indeed, if $A$ is in $GL(n,\mathbb{R})$, then $A\colon\mathbb{R}^n\to\mathbb{R}^n$ is an invertible map.

**Example 8.1.8.** The dihedral group $D_n$ acts on the vertices of the regular $n$-gon. Each element of the group $D_n$ of symmetries of the regular $n$-gon permutes the $n$ vertices of the $n$-gon. If we number these vertices 1 to $n$ (counter clockwise), then a rotation over $\frac{2\cdot\pi}{n}$ induces the $n$-cycle $(1,2,...,n)$ on these vertices. A reflection in the axis through the center of the $n$-gon and the vertex 1 induces the permutation $(2,n)(3,n-1)\ldots\left(\frac{n}{2},\frac{n}{2}+1\right)$ in case $n$ is even, and

$(2,n)(3,n-1)\dots\left(\frac{n-1}{2},\frac{(n+1)}{2}\right)$ in case $n$ is odd. This yields a permutation representation of $D_n$ into $\mathrm{Sym}_n$.

**Remark 8.1.9.** An important way of describing a permutation group is as a subgroup of $\mathrm{Sym}(X)$ generated by a list $A$ of permutations of $X$. This means that the subgroup is the smallest set of permutations of $X$ containing $A$, closed under multiplication and taking inverses, and containing the identity element.

Let $G$ be a group. Three fundamental examples of permutation representations of $G$ into $\mathrm{Sym}(X)$ with $X = G$ are:

- The *left regular* representation $L_g\colon G \to G, x \mapsto g\cdot x$. Here, the map $L_g$ is left multiplication by $g$ on $G$.
- The *right regular* representation $R_g\colon G \to G, x \mapsto x\cdot g^{-1}$. Here, the map $R_g$ is right multiplication by $g^{-1}$ on $G$.
- The *conjugation* representation $C_g\colon G \to G, x \mapsto g\cdot x\cdot g^{-1}$. Here, the map $C_g$ is conjugation by $g$ on $G$.

In contrast to the usual notation, we write $L_g$ for the image of $g$ under $L$. This way, the expressions, which are maps themselves, are better readible when applied to an element of $G$: $L_g(h)$ is preferred to $L(g)(h)$.

Similarly, we prefer the notations $R_g(h)$ and $C_g(h)$ to $R(g)(h)$ and $C(g)(h)$, respectively.

**Theorem 8.1.10.** *The maps $L,R,C$ are permutation representations of $G$ on $G$.*

*Proof.* In order to prove that $L$, $R$, and $C$ are permutation representations of $G$ on $X = G$, we proceed in two steps.

**The maps $L_g, R_g, C_g$ are bijections, so they belong to $\mathrm{Sym}(G)$.**

If $L_g(x) = L_g(y)$, then $g\cdot x = g\cdot y$, so, by the cancellation law, $x = y$. Hence $L$ is injective.

If $x \in G$, then also $g^{-1}\cdot x \in G$, and $L_g\left(g^{-1}\cdot x\right) = x$. Thus, $L$ is also surjective.

We conclude that $L_g$ is a bijection. Therefore, it belongs to $\mathrm{Sym}(G)$. The proofs for $R$ and $C$ are similar.

**The maps $L,R,C$ are morphisms $G \to \mathrm{Sym}(G)$, so they are permutation representations.**

We need to verify that, for each $g,h \in G$, we have $L_{g\cdot h} = L_g\cdot L_h$. This is indeed the case as, for each $x \in G$,

$$L_{g\cdot h}(x) = (g\cdot h)\cdot x = g\cdot(h\cdot x) = L_g(L_h(x)) = L_g\cdot L_h(x) \tag{8.9}$$

The proofs for $R$ and $C$ are similar.

□

**Example 8.1.11** (The left regular represenation of $\mathrm{Sym}_3$)**.** Let $G$ be the group $\mathrm{Sym}_3$. Then $G$ consists of six elements: $e = 1$, $y = (1,2,3)$, $z = (1,3,2)$, $a = (1,2)$, $b = (2,3)$, $c = (1,3)$.

The representation $L\colon\ G \to \mathrm{Sym}(G)$ is written out explicitly as permutations on $\{a,b,c,e,y,z\}$.

- $L_a = (a,e)\,(b,y)\,(c,z)$
- $L_b = (b,e)\,(a,z)\,(c,y)$
- $L_c = (c,e)\,(a,y)\,(b,z)$
- $L_y = (y,z,e)\,(a,c,b)$
- $L_z = (e,z,y)\,(a,b,c)$.

Note that the multiplication of $G$ can be easily recovered from the list. For instance $L_c\,(a) = y$ means $c\cdot a = y$.

**Example 8.1.12** (The right regular represenation of $\mathrm{Sym}_3$)**.** With the notation of Example 8.1.11 we can express right multiplication as the following permutations

- $R_a = (a,e)\,(b,z)\,(c,y)$
- $R_b = (b,e)\,(a,y)\,(c,z)$
- $R_c = (c,e)\,(a,z)\,(b,y)$
- $R_z = (z,y,e)\,(a,b,c)$
- $R_y = (e,y,z)\,(a,c,b)$

**Example 8.1.13** (The conjugation representation of $\mathrm{Sym}_3$)**.** With the notation of Example 8.1.11 we can express conjugation by the following permutations

- $C_a = (b,c)\,(y,z)$
- $C_b = (a,c)\,(y,z)$
- $C_c = (a,b)\,(y,z)$
- $C_y = (a,b,c)$
- $C_z = (a,c,b)$

Next we study the kernels of these representations. We recall that the center of $G$ is the subgroup $\mathbf{Z}(G) = \{d \in G | \forall g. (g \in G) \Rightarrow (d \cdot g = g \cdot d)\}$ of $G$.

**Theorem 8.1.14.** *The kernels of $L$ and $R$ are trivial.*
*The kernel of $C$ is the center of $G$.*

*Proof.* The following three steps suffice.

**The homomorphisms $L$ and $R$ are injective.**

If $L_g = L_{1_G}$, then, in particular, $g = g \cdot 1_G = L_g(1_G) = L_{1_G}(1_G) = 1_G \cdot 1_G = 1_G$. Therefore $g = 1_G$. This shows that the kernel of the homomorphism $L\colon G \to \mathrm{Sym}(G)$ is trivial. It follows that the homomorphism $L\colon G \to \mathrm{Sym}(G)$ is injective. The proof for $R$ is similar.

**The kernel of the homomorphism $C$ is the center of $G$.**

If $C_g = C_{1_G}$, then, for all $x \in G$, we have $L_g(x) = L_{1_G}(x)$, so $g \cdot x \cdot g^{-1} = x$. Thus, $g$ belongs to the kernel of $C$ if and only if, for all $x \in G$, we have $g \cdot x = x \cdot g$, that is, if and only if $g \in \mathbf{Z}(G)$.

□

**Example 8.1.15.** Let $G$ be the group $\mathrm{Sym}_3$. Then $G$ consists of six elements: $e = 1$, $y = (1,2,3)$, $z = (1,3,2)$, $a = (1,2)$, $b = (1,3)$, $c = (2,3)$. By the theorem, the left and right regular representations are injective. What about $C$? By the theorem, $C_g$ is the identity if and only if it commutes with every element of $\mathrm{Sym}_3$. Since each conjugacy class distinct from $1_G$ consists of more than one element, we find $\mathbf{Z}(G) = \mathrm{Ker}(C) = \{1_G\}$.

Since the left regular representation $L$ has a trivial kernel, every group $G$ is isomorphic with its image under $L$ and hence with a subgroup of some symmetric group.

**Theorem 8.1.16.** *Every group is isomorphic with a permutation group.*

*Proof.* As $L$ is injective, the group $G$ is isomorphic with its image $\{L_g | g \in G\}$ in $\mathrm{Sym}(G)$. The image is a permutation group.

□

The permutation representations $L, R$, and $C$ all have degree $|G|$. There do exist methods for constructing lower-degree permutation representations.

**Remark 8.1.17.** Although the theorem asserts that we can write every group as a group of permutations, it does not give us a practical presentation of the group. Think of the symmetric group on $n$ letters: the proof of the theorem realizes $\mathrm{Sym}_n$ as a group of permutations of $n!$ letters, while the natural permutation presentation of this group is on $n$ letters.

The following theorem shows two ways of obtaining a permutation representations from a given permutation representation.

The restriction of a map $f\colon X \to Y$ to a subset $Z$ of $X$ is the map

$$f\downarrow Z\colon Z \to Y, h \mapsto f(h) \tag{8.10}$$

**Theorem 8.1.18.** *Let $f\colon G \to \mathrm{Sym}(X)$ be a permutation representation.*

*i. If $H$ is a subgroup of $G$, then the restriction $f\downarrow H$ of $f$ to $H$ is also a permutation representation.*

*ii. Let $Y$ be a subset of $X$ such that for all $g \in G$ and all $y \in Y$, also $f(g)(y) \in Y$. Then every $g \in G$ determines by restriction to $Y$ a bijection $f(g)\downarrow Y$ of $Y$. The resulting map $G \to \mathrm{Sym}(Y), g \mapsto f(g)\downarrow Y$ is a permutation representation.*

*Proof.*

**If $H$ is a subgroup of $G$, then the restriction $f\downarrow H$ of $f$ to $H$ is also a permutation representation.**

Let $g,h \in H$. Since $g,h \in G$ and $f$ is a morphism of groups, we have $f(g\cdot h) = f(g)\cdot f(h)$, so the restriction $f\downarrow H$ of $f$ to $H$ is also a morphism of groups.

**Let $Y$ be a subset of $X$ such that for all $g \in G$ and all $y \in Y$, also $f(g)(y) \in Y$. Then every $g \in G$ determines by restriction to $Y$ a bijection $f(g)\downarrow Y$ of $Y$. The resulting map $G \to \mathrm{Sym}(Y), g \mapsto f(g)\downarrow Y$ is a permutation representation.**

We first show that $f(g)\downarrow Y$ is a bijection for $g \in G$.

Let $y \in Y$. Then there is an element $x \in X$ with $f(g)(x) = y$. Since $y \in Y$ and $g$ has inverse $g^{-1}$ in $G$, we have $f\left(g^{-1}\right)(y) \in Y$, and so $x \in Y$. Since $f(g)\downarrow Y(x) = y$, this proves that $f(g)\downarrow Y$ is surjective.

Clearly, $f(g)\downarrow Y$ is injective because $f(g)$ is.

It remains to show that the map $g \mapsto f(g)\downarrow Y$ is a morphism. Let $h \in G$ and $y \in Y$. Then $f(g\cdot h)\downarrow Y(y) = f(g\cdot h)(y) = f(g)(f(h)(y)) = f(g)\downarrow Y(f(h)\downarrow Y(y)) = f(g)\downarrow Y\cdot f(h)\downarrow Y(y)$ proving $f(g\cdot h)\downarrow Y = f(g)\downarrow Y\cdot f(h)\downarrow Y$.

□

**Example 8.1.19** (The general linear group and special linear group acting on vectors)**.** The group $G = GL(n,\mathbb{R})$ of real invertible $n\times n$-matrices acts as a permutation group on the set of vectors of $\mathbb{R}^n$: the element $g$ of $G$ is mapped to the bijection $g\colon \mathbb{R}^n \to \mathbb{R}^n, v \mapsto g(v)$, an element of $\mathrm{Sym}(\mathbb{R}^n)$. See a previous example. As the zero vector is fixed by all matrices in $G$, the group $G$ is also a permutation group on the set of nonzero vectors. Also the special linear group $SL(n,\mathbb{R})$, which consists of the $n\times n$-matrices with determinant 1, acts on the nonzero vectors in $\mathbb{R}^n$.

**Example 8.1.20** (A subgroup of $\mathrm{Sym}(X)$ acting on all subsets of $X$)**.** Suppose $G$ acts on $X$. Then there is an action of $G$ on the set of all subsets of $X$. If $Y = \{x_1, \ldots, x_k\}$ is a subset of $X$ of size $k$, its image under the permutation $g$ is $g(Y) = \{g(x_1), \ldots, g(x_k)\}$. Verify that this defines a permutation representation indeed! Let $Z$ be the set of all subsets of $X$ of size 2. Then $Z$ is $G$-invariant, that is, for each $g$ in $G$, the image $g(Y)$ of a 2-set $Y$ is again a 2-set. Thus, we find a permutation representation of $G$ on $Z$.

A set $Y$ as in the theorem is called *invariant* under $G$. We also say $Y$ is $G$-invariant or, if $G$ is clear from the context, just invariant.

## 8.2 Orbits

**Definition 8.2.1.** Let $G \to \mathrm{Sym}(X)$ be a permutation representation of the group $G$ on $X$.

Suppose $x, y \in X$. If there is $g \in G$ with $g(x) = y$, then we say that $x$ and $y$ are *in the same orbit*, and write $\mathrm{itso}(x,y)$.

- The relation itso, being in the same orbit, is an equivalence relation.
- Its equivalence classes are called *orbits* of $G$ on the set $X$.
- The group $G$ is said to be *transitive* on $X$ if it has only one orbit on $X$.

**Proposition 8.2.2.** *The relation* itso*, being in the same orbit, is an equivalence relation.*

*Proof.* In order to show that itso is an equivalence relation on $X$ we verify the three basic properties of an equivalence relation.

itso **is reflexive.**

For $g$ equal to the identity, we have $g(x) = x$, whence $x$ is in the same orbit as $x$.

itso **is symmetric.**

Suppose that $x$, $y$ are in the same $G$-orbit. Then $g(x) = y$, for some element $g$ of $G$. Consequently $g^{-1}(y) = x$, so $\mathrm{itso}(x,y)$.

itso **is transitive.**

Suppose $\mathrm{itso}(x,y)$ and $\mathrm{itso}(y,z)$. Then there are elements $g$, $h$ of $G$ such that $g(x) = y$ and $h(y) = z$. Then $h \cdot g(x) = z$, and so $\mathrm{itso}(x,z)$.

□

**Example 8.2.3** (The symmetric group on 3 letters acting on pairs of elements)**.** Let $X$ be the set consisting of the 15 subsets of $\text{Sym}(3)$ having exactly two elements. The map $L\colon \text{Sym}(3) \to \text{Sym}(X)$ with $L_g(\{a,b\}) = \{g(a), g(b)\}$ is a permutation representation (it was treated once before ). The orbit of $\{e,(1,2)\}$ consists of 3 elements: $\{1,(1,2)\}$, $\{(1,3),(1,2,3)\}$, and $\{(2,3),(1,3,2)\}$. Can you describe the other orbits of the given action $L$ ?

**Example 8.2.4** (The general linear group on vectors)**.** The group $GL(n,\mathbb{R})$, with $n > 1$, is not transitive on the set of all vectors of $\mathbb{R}^n$. For, the zero vector 0 can only be transformed into itself. The group is transitive on the set of nonzero vectors: if $v_1$ and $w_1$ are two such vectors, then $v_1$ (respectively $w_1$) can be extended to a basis $v_1,...,v_n$ (respectively $w_1,...,w_n$ ) of $\mathbb{R}^n$ and determine an invertible linear map $a\colon \mathbb{R}^n \to \mathbb{R}^n$ by $a\cdot(r_1\cdot v_1 + ... + r_n\cdot v_n) = r_1\cdot w_1 + ... + r_n\cdot w_n$. The map $a$ belongs to $GL(n,\mathbb{R})$ and satisfies $A\cdot v_1 = w_1$, and so indeed $v_1$ and $w_1$ are in the same orbit. Conclusion: there are precisely two orbits, viz., $\mathbb{R}^n \setminus \{0\}$ and $\{0\}$.

**Example 8.2.5** (Conjugation of the symmetric group on itself)**.** The orbits in a group $G$ of the group $G$ acting by conjugation on itself are the so-called *conjugacy classes*. Since $\{1\}$ is a single orbit, the action is transitive only if $G$ is the trivial group 1. We determine the conjugacy classes of $G = \text{Sym}(3)$. They are $\{1\},\{(1,2),(1,3),(2,3)\},\{(1,2,3),(1,3,2)\}$. More generally, for $\text{Sym}(X)$ the conjugacy classes consist of all elements of a given cycle type, see the Conjugation Theorem. Above, the cycle structure $1,1,1$ belongs to 1 (the identity), the type $2,1$ to the class of $(1,2)$, and the type 3 to the class of $(1,2,3)$. The cycle structures are nothing but the partitions of $n$. For $n = 4$, the partitions are $4,31,22,211,1111$. Representative elements from the corresponding conjugacy classes are: $(1,2,3,4)$, $(1,2,3)$, $(1,2)(3,4)$, $(1,2)$, 1.

**Example 8.2.6.** Let $G$ be the permutation group on $X = \{1,...,8\}$ generated by $(1,3)(2,4)$, $(5,6)$, $(2,7)$, and $(1,8)$.

The orbits of $G$ are $\{1,3,8\}$, $\{2,4,7\}$ and $\{5,6\}$.

Clearly, these sets partition $X$. Moreover, they are invariant under the generators of $G$, and hence also under $G$ itself. Using the generators of $G$ one can also easily check that the sets are contained in single $G$-orbits.

The fact that itso is an equivalence relation implies that a $G$-orbit is equal to $Gx = \{g(x) \mid g \in G\}$ for any point $x$ in this orbit.

The observation that itso is an equivalence relation leads to the following algorithm for a permutation group $G$ on a finite set $X$.

**Algorithm 8.2.7** (Orbit algorithm)**.** • *Input: a set $B$ of generators of $G$ and an element $x$ of $X$;*

- *Output: the $G$-orbit of $x$.*

```
Orbit := procedure(G,x)
local variables
    | O
    | L
    | N
O := {x} L := {x} while L ≠ ∅ do
    | N := {g (a)| (g,a) ∈ B × L} , L := N \ O , O := L ∪ O
return
    | O
```

*Proof.*

**Termination**

The subset $O$ increases by the set $L$ disjoint to $O$ at each pass of the while loop. As these are subsets of the finite set $X$, we must have $L = \varnothing$ at the end of some while loop pass. Hence termination is guaranteed.

**Correctness**

If $y \in Gx$, then there are $b_1,...,b_n \in B$ such that $y = b_1 \cdot ... \cdot b_n(x)$. By construction, $O$ is invariant under each of the elements in $B$, so also $b_1 \cdot ... \cdot b_n(x)$ belongs to $O$. In particular, $Gx$ is contained in the output $O$.

□

The behaviour of a permutation representation $f\colon G \to \mathrm{Sym}(X)$ can be recorded inside $G$. The first step is to relate a point $x$ of $X$ to a particular subgroup of $G$.

**Definition 8.2.8.** If $x \in X$, then the *stabilizer* of $x$ in $G$ is the subgroup $G_x$ of $G$ given by $G_x = \{g \in G | g(x) = x\}$

If $g(x) = x$, then $g$ is said to *fix* or to *stabilize* $x$.

**Example 8.2.9** ($G = \mathrm{Sym}_5$ acting on $\{1,2,3,4,5\}$)**.** The stabilizer of 3 consists of all permutations $g$ with $g(3) = 3$. These are all permutations of $\{1,2,4,5\}$. Hence, the stabilizer is $\mathrm{Sym}(\{1,2,4,5\})$, which is isomorphic with $\mathrm{Sym}_4$.

**Example 8.2.10** ($G = \mathrm{Sym}_5$ acting on the set $X$ of subsets of $\{1,2,3,4,5\}$ of size 2)**.** The stabilizer of the set $\{4,5\}$ consists of all elements $g$ of $\mathrm{Sym}_5$ with $g(4) \in \{4,5\}$ and $g(5) \in \{4,5\}$. In the disjoint cycle decomposition of such an element $g$, we find either the cycle $(4,5)$, or no cycle at all in which 4 or 5 occurs. Thus, such an element $g$ is either of the form $h$ or $h\cdot(4,5)$, for some $h \in \mathrm{Sym}_3$. Hence, the stabilizer of $\{4,5\}$ is the subgroup $\mathrm{Sym}_3 \times \mathrm{Sym}(\{4,5\})$. More precisely, the stabilizer is the image of the natural morphism $\mathrm{Sym}_3 \times \mathrm{Sym}(\{4,5\}) \to \mathrm{Sym}_5, [g,h] \mapsto g\cdot h$. Thus, the stabilizer has order $6\cdot 2 = 12$.

**Example 8.2.11** ($G = GL(3,\mathbb{R})$ acting on vectors)**.** Let $x$ be the first standard basis vector. Then $G_x$ is the subgroup of $G$ of all invertible matrices of the form $\begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix}$.

**Example 8.2.12** (Conjugation)**.** Let $x \in G$. Then the stabilizer of $x$ in $G$ under conjugaction is $C_G(x)$, the subgroup of $G$ of all elements $g$ with $g \cdot x = x \cdot g$. This subgroup is called the centralizer of $x$ in $G$. Observe that $C_G(1) = G$.

**Example 8.2.13** ($G = D_n$ on the vertices of a regular $n$-gon)**.** Let $G$ be the group $D_n$ acting on the $n$ vertices of a regular $n$-gon. Let $x$ be a vertex. Among the $n$ rotations in $G$ only the identity fixes $x$. The only reflection fixing $x$ is the reflection in the axis through $x$ and the center of the $n$-gon. So in this case $G_x$ consists of two elements.

**Remark 8.2.14.** The notation $G_x$ does not explicitly use $f$. But the stabilizer does depend on it. For instance, if $G = \mathrm{Sym}_4$ and $x = (1,2)$, then

- $G_x = 1$ if $f = L$, left multiplication (or $R$, right inverse multiplication);
- $G_x = \{1, (1,2), (3,4), (1,2)(3,4)\}$ if $f = C$, conjugation.

If a group $G$ acts as a permutation group on a set $X$, we can associate to each point $x$ in $X$ the stabilizer in $G$ of $x$. The next step is to construct permutation representations from within $G$. Suppose $H$ is a subgroup of $G$. We shall construct a transitive permutation representation of $G$ on $G/H$ with $H$ as point stabilizer.

**Theorem 8.2.15.** *For $g \in G$, let $L_g$ be the map $G/H \to G/H$ specified by $\forall h. L_g(h \cdot H) = g \cdot h \cdot H$. Then $L \colon G \to \mathrm{Sym}(G/H), g \mapsto L_g$ is a transitive permutation representation. Moreover, the stabilizer in G of the element H of $G/H$ is H.*

*Proof.* Regarding the map $L : G \to \mathrm{Sym}(G/H), g \mapsto L(g)$ given by $L_g \cdot x \cdot H = g \cdot x \cdot H$, we need to show the following.

- For each $g \in G$, the image $L_g$ is a bijection $G/H \to G/H$. As, for each $x \in G$, $L_{g^{-1}}(L_g(x \cdot H)) = g^{-1} \cdot g \cdot x \cdot H = x \cdot H$, we have ${L_g}^{-1} = L_{g^{-1}}$. So indeed, $L$ is a bijection.
- The map $L$ is a morphism of groups. Let $g, h \in G$. We need to show $L_g \cdot L_h = L_{g \cdot h}$. For each $k \cdot H$ we have $L_{g \cdot h} \cdot k \cdot H = (g \cdot h) \cdot k \cdot H = L_g(h \cdot k \cdot H) = L_g(L_h(k \cdot H)) = L_g \cdot L_h(k \cdot H)$. So the map $L$ is indeed a morphism. In other words, $G \to \mathrm{Sym}(G/H), g \mapsto L(g)$ is a permutation representation.
- The permutation representation $L$ is transitive. Let $g \cdot H$ and $h \cdot H$ be two elements of $G/H$. Then $L_{h \cdot g^{-1}}$ maps $g \cdot H$ onto $h \cdot H$.
- The stabilizer of the element $H$ of $G/H$ coincides with $H$. The stabilizer is $K = \{k \in G | k \cdot H = H\}$. If $k \in K$, then there are $h, h' \in H$ with $k \cdot h = h'$, and so $k = h' \cdot h^{-1} \in H$, proving $K \subset H$. Conversely, if $h \in H$, then $h \cdot H = H$, so $h \in K$, proving $H \subset K$. Hence $H = K$.

□

**Example 8.2.16.** The kernel $K$ of $L$ need not be trivial: If $G=\mathbb{Z}$ and $H=3\cdot\mathbb{Z}$, then the kernel is equal to $3\cdot\mathbb{Z}$. For, $L\cdot(m+3\cdot\mathbb{Z})=n+m+3\cdot\mathbb{Z}$ describes the action on the cosets and it is clear that $L\cdot(m+3\cdot\mathbb{Z})=m+3\cdot\mathbb{Z}$ holds for all $m$ if and only if $n\in 3\cdot\mathbb{Z}$. Thus, $K=H$. It is true, however, that the kernel is always a subgroup of $H$. Do you see why?

**Example 8.2.17.** The construction is a generalisation of the left regular representation of $G$. The latter is the special case where $H=\{1\}$, the trivial subgroup of $G$. If $g\cdot H$ is a left coset of $H$, then the stabilizer of $g\cdot H$ is the conjugate of $H$ by $g$, i.e. the stabilizer equals $g\cdot H\cdot g^{-1}=\left\{g\cdot h\cdot g^{-1}\middle|\, h\in H\right\}$. Indeed, for each element $g\cdot h\cdot g^{-1}$, with $h\in H$, we have $g\cdot h\cdot g^{-1}\cdot g\cdot H=g\cdot h\cdot H=g\cdot H$. On the other hand, if $k\in G$ satisfies $k\cdot g\cdot H=g\cdot H$, then there is an $h\in H$ with $k\cdot g=g\cdot h$, from which we deduce $k=g\cdot h\cdot g^{-1}$.

We are now ready for the final step. It will establish that any transitive permutation representation $G\to\mathrm{Sym}(G/H)$ can be identified with the permutation representation $L$ as above for $H$ the stabilizer of an element $x$ of $X$.

Let $f\colon G\to\mathrm{Sym}(X)$ be a permutation representation. Fix $x\in X$. We can identify $X$ with the set of cosets of $G$ with respect to the stabilizer of an element of $X$, provided $f$ is transitive.

**Theorem 8.2.18.** *Suppose that $f$ is transitive. Then the map $t\colon G/G_x\to X, g\cdot G_x\mapsto f(g)(x)$ is a well-defined bijection and satisfies $f(h)\circ t=t\circ L_h$ for every $h\in G$. If, moreover, $G$ is finite, then $|G|=|G_x|\cdot|X|$.*

*Proof.* Write $H=G_x$. We split the proof in the following steps.

**The map $t$ is well defined.**

Suppose $g$ and $g'$ are in the same coset $g\cdot H$. Then there is $h\in H$, such that $g'=g\cdot h$. As $H=G_x$, we have $f(h)(x)=x$, whence $f(g')(x)=f(g\cdot h)(x)=f(g)\cdot f(h)(x)=f(g)(x)$. This proves that the assigment $g\cdot H\mapsto f(g)(x)$ does not depend on the choice of $g'\in g\cdot H$.

**The map $t$ is injective.**

Suppose $g,g'\in G$ satisfy $t(g\cdot H)=t(g'\cdot H)$. Then $f(g)(x)=f(g)(x)$, so $x=f\left(g^{-1}\cdot g'\right)(x)$, that is, $g^{-1}\cdot g'\in G_x$. Since $G_x=H$, this shows $g^{-1}\cdot g'\cdot H=H$, and so $g'\cdot H=g\cdot H$. Hence $t$ is injective.

**The map $t$ is surjective.**

Let $y\in X$. As $f$ is transitive, there is $g\in G$ with $y=g(x)$. But then $t(g\cdot H)=y$. Hence $t$ is surjective.

**For each $h\in G$, we have $f(h)\circ t=t\circ L_h$.**

Let $g\in G$. Then $t\circ L_h(g\cdot H)=t(h\cdot g\cdot H)=f(h\cdot g)(x)=f(h)\circ f(g)(x)=f(h)(f(g)(x))=f(h)(t(g\cdot H))=f(h)\circ t(g\cdot H)$. Hence the assertion.

**If $G$ is finite, then $\frac{|G|}{|H|} = |X|$.**

If $G$ is finite, then Lagrange's theorem gives that $|G/H| = \frac{|G|}{|H|}$. As $t$ is a bijection, we also have $|G/H| = |X|$.

□

**Example 8.2.19** ($G = \mathrm{Sym}_5$ on pairs of $\{1,2,3,4,5\}$)**.** The stabilizer in $\mathrm{Sym}_5$ of $\{4,5\}$ is equal to a group $H$ isomorphic to $\mathrm{Sym}_3 \times \langle(4,5)\rangle$. The index $|G/H|$ of this subgroup in $\mathrm{Sym}_5$ is equal to $\frac{5!}{3!\cdot 2!} = 10$. This is equal to the number of subsets of $\{1,...,5\}$ of size 2. Under the bijection of Identification of orbit with cosets, the coset $(1,2,3,4,5)\cdot H$ is mapped onto $\{1,5\}$. This image can be computed by use of any element from the coset; for example

- $(1,2,3,4,5)$ maps $\{4,5\}$ to $(1,2,3,4,5)(\{4,5\}) = \{1,5\}$;
- but also $(1,2,3,4,5)\cdot(4,5)\cdot(1,2,3)(\{4,5\}) = \{1,5\}$.

**Example 8.2.20** (Existence of elements of order 2 in a group of order 10)**.** Let $G$ be a group of order 10. It acts by left multiplication on the set $X$ consisting of the 45 subsets of $G$ of size 2. By the theorem, the number of elements in an orbit is a divisor of $|G|$ and hence equal to $1,2,5$, or 10. An orbit cannot be a singleton (do you see why?). As $|X|$ is odd, there must be an orbit of size 5. The stabilizer of an element from this orbit as order 2. This establishes that $G$ has a subgroup, and thus also an element, of order 2. Later, we shall repeat this argument in greater generality to show that if $p$ is a prime dividing $|G|$, there is an element of order $p$ in $G$.

**Example 8.2.21** (Conjugation)**.** Let $x \in G$. Then the centralizer $C_G(x)$ of $x$ in $G$ is the stabilizer of $x$ in the conjugation action. Hence the number of conjugates of $x$ equals the index of $C_G(x)$ in $G$ and is a divisor of the order of $G$.

**Example 8.2.22** (The dihedral group)**.** Let $D$ be the dihedral group of order $2\cdot n$, acting on the $n$ vertices of a regular $n$-gon. The group $D$ is transitive on the $n$ vertices. So, the stabilizer of a vertex in $D$ consists of $\frac{|D|}{n} = 2$ elements. One element is the identity element, the other is a reflection with respect to a line through the vertex.

Let $B$ be an orbit of the permutation representation $f\colon G \to \mathrm{Sym}(G/H)$ of the finite group $G$. If $x$ is an element of $B$, then, by the theorem, $\frac{|G|}{|G_x|} = |B|$. In particular, the number of elements of $B$ is a divisor of $|G|$, and the order $|G|$ of the stabilizer of $x$ does not depend on the choice of $x$ in $B$.

The condition $t\cdot L = h\cdot t$ for each $h \in G$ is often phrased as 'the map $t$ commutes with the group action'. It means that $t$ does not just establish an identification of $G/G_x$ and $X$ as sets, but even of the permutation representations $L$ on $G/G_x$ and $f$ on $X$.

The last assertion of the theorem says that, for finite groups $G$, the degree of a transitive permutation representation is equal to the index of a point stabilizer in $G$.

## 8.3 Permutation group algorithms

Consider a permutation representation of $G$ on a set $X$. By Identification of orbit with cosets, the order of a permutation group $G$ can be determined once we know the order of the stabilizer $G_x$ in $G$ of a point $x \in X$. Since $G_x$ is a smaller permutation group if the $G$-orbit of $x$ is nontrivial, the result leads to a recursive computation.

**Definition 8.3.1.** A *base* for $G$ on $X$ is a sequence $B$ of elements of $X$ such that the stabilizer $G_B$ in $G$ of each element of $B$ is the trivial group.

**Example 8.3.2** (Bases for the groups $\mathrm{Sym}(X)$ and $\mathrm{Alt}(X)$)**.** Let $G$ be a subgroup of $\mathrm{Sym}(X)$. Then the sequence $[1,...,n-1]$ is a base for $G$. If $G = \mathrm{Sym}_n$, then we cannot replace the base by a smaller one. If $G = \mathrm{Alt}_n$, then $[1,...,n-2]$ is a base. For, the only nontrivial permutation in $\mathrm{Sym}_n$ stabilizing each of the elements $1,...,n-2$ is the transposition $(n-1,n)$. But this is element is odd and so does not belong to $\mathrm{Alt}_n$.

**Example 8.3.3** ($GL(V)$)**.** Let $V$ be a vector space of dimension $n$. Consider $G = GL(V)$ acting on the vectors of the vector space $V$. If $v_1,...,v_n$ is a base of $V$, then $[v_1,...,v_n]$ is a base of $G$ acting on $V$. For, a linear transformation fixing a basis of $V$ is the identity.

**Example 8.3.4** ($D_n$)**.** Consider $G$ to be $D_n$ acting on the $n$ vertices of the $n$-gon. Any two vertices that are not opposite form a base for $G$.

If $G$ is a subgroup of $\mathrm{Sym}_n$ and $B = [b_1,\cdots,b_m]$ is a base for $G$ acting on $X = \{1,...,n\}$, then the order of $G$ is equal to the size of the $G$-orbit of $B$. Alternatively, we can determine the order as follows, where $Gx$ stands for the orbit of $G$ on $x$.

**Theorem 8.3.5** (Order theorem)**.** *If $B = [b_1,...,b_m]$is a base for $G$ acting on $X$, then* $|G| = \prod_{i=1}^{m-1}\left(\left|G_{[b_1,\cdots,b_i]}b_{i+1}\right|\right)$

*Proof.* $|G| = \left|G_{b_1}\right| \cdot |Gb_1| = \left|G_{b_1,b_2}\right| \cdot \left|G_{b_1}b_2\right| \cdot |Gb_1| = \left|G_{b_1,b_2,...,b_{n-1}}b_n\right| \cdot ... \cdot \left|G_{b_1}b_2\right| \cdot |Gb_1|$.

□

**Example 8.3.6.** Let $F$ be a finite field of size $q$ and put $V = F^3$ for the 3-dimensional vector space over $F$. Consider $G = GL(V)$, the group of all invertible linear mappings on $V$, acting on the vectors of the vector space $V$. By Example 8.1.7, we know that the $q^3-1$ nonzero vectors form an orbit.

Moreover, in Example 8.2.4 we noticed that every basis of $V$ is a base for the action of $G$ on the nonzero vecotors of $V$.

Since $G$ is transitive on bases (indeed the matrix whose columns consist of the vectors of a basis $B$ maps the standard bases to $B$), its order equals the number of distinct bases. But that implies that $|G| = \left(q^3-1\right) \cdot \left(q^3-q\right) \cdot \left(q^3-q^2\right)$

Can you find the order of $GL(n,F)$, the group of invertible $n \times n$ matrices with coefficients in $F$?

A handicap in applying this theorem to a group generated by a set of permutations is that we have no way (yet) of determining the stabilizer. This is taken care of by Schreier's lemma.

A permutation group $G$ on $X = \{1,\ldots,n\}$ can be conveniently represented by a (small) generating set of permutations. Most algorithms for permutation groups take such a generating set as input. Let $S$ be a generating set for $G$.

**Definition 8.3.7.** A *Schreier tree* with root $x$ for a list $S$ of generating elements of $G$ is a tree rooted at $x$ with the following properties:

- Its vertices are the elements of the orbit $Gx$.
- Each edge $i,j$ with $i$ closer to the root $x$ than $j$ is labeled by a generator $s \in S$ such that $s(i) = j$.

**Example 8.3.8.** Consider the permutation group $G = \langle a,b \rangle$ where $a = (1,2)(3,4)$ and $b = (1,3)(2,4)$. The following graph describes the action of both $a$ and $b$ fully. A Schreier tree with root 1 results from deletion of any one of the four edges.

**Remark 8.3.9.** For a given permutation group $G$ acting on a set $X$, and generated by a set $S$ of permuations, one can draw a labeled directed graph, in which an edge $[x,y]$ is labeled $g$, if $g \in S$ and $g$ satisfies $g(x) = y$. Now the orbits of $G$ on $X$ are the connected components of this graph. Moreover, a spanning tree of the component containing $x$ is a Schreier tree for $S$.

A Schreier tree $T$ can be represented by a triple $[V,L,W]$, where $V$ is the ordered list of vertices of $T$ starting with the root $x$ of the tree, $L$ is a list of labels, starting with a dummy 0 which is followed by elements form $S$ or their inverses, and the third list $W$ also starts with 0 which is followed by vertices from the tree. The elements in the three lists are ordered in such a way that the unique neighbour of a vertex $v$ in $T$, being at position $i \geq 2$ in $V$, on the path to the root $x$ is the vertex $w$ at the same position $i$ in $W$. The edge on $v$ and $w$ carries the label $s$, where $s$ is the element at position $i$ in $L$. The triple is called the *Schreier data* for the tree $T$.

We present an algorithm to find Schreier trees. This algorithm is a slight extension of the Orbit algorithm.

**Algorithm 8.3.10** (Schreier tree algorithm)**.** • *Input: list $S$ of generators of $G$ and $x \in X$.*

- *Output: Schreier data for a Schreier tree for $S$ with root $x$.*

```
SchreierTree := **procedure**$(S,x)$
**local variables**
    pnt, $j$, im, bpnt $:= x$ , gens $:= S$ , $J := \{1,...,Length(B)\}$ , orbit := [bpnt]
    new := [bpnt] , newest, svect := [0] , bpnts := [0]
**while** new $\neq \varnothing$ **do**
    **for** $i_1 := 1$ **while** $i_1 \leq Length(\text{new})$ **with step** $i_1 := i_1 + 1$ **do**
        **for** $i_2 := 1$ **while** $i_2 \leq Length(\text{gens})$ **with step** $i_2 := i_2 + 1$ **do**
            pnt $:= i_{1\,\text{new}}$ , $j := i_{2J}$ , im := $\text{pnt}^{j_{\text{gens}}}$ , newest $:= \varnothing$
            **if** $\neg$ (im $\in$ orbit)
                **then**
                    Add (orbit, im) , Add (newest, im) , Add $(\text{svect}, j_{\text{gens}})$ , Add (bpnts, pnt)

            im := $\text{pnt}^{\left(j_{\text{gens}}\right)^{-1}}$
            **if** $\neg$ (im $\in$ orbit)
                **then**
                    Add (orbit, im) , Add (newest, im) , Add (svect, $-j$) , Add (bpnts, pnt)

    new := newest
**return**
    [orbit, svect, bpnts]
```

*Proof.* The proof is similar to the proof of the orbit algorithm.

**Termination**

Since $X$ is a finite set, and the vertex set of $T$ is a subset of $X$ which increases strictly at each pass of Step 3 with nonempty $N$, termination is guaranteed.

**After termination, $T$ is a tree with the right labels.**

By construction, the vertex set of $T$ is the $G$-orbit of $x$. See the orbit algorithm.

□

**Example 8.3.11.** Suppose the group $G$ is generated by the list $S = [a,b]$ where $a = (1,2)(3,4)$ and $b = (1,3)(2,4)$. We create a Schreier tree following the Schreier tree algorithm. Take $x = 1$ and set orbit $= [1]$. Now create a new generation of elements in orbit. This yields the new elements 2 and 3 which can be added to orbit, which becomes svect $= [1,2,3]$. The lists svect and bpnt both starting with a zero are now extended by to svect $= [0,a,b]$ and bpnt $= [0,1,1]$. The new points are now 2 and 3.

Applying both $a$ and $b$ yields only one new point, namely 4 as the image of 2 under $b$. We add this element to the list and obtain svect $= [1,2,3,4]$, svect $= [0,a,b,b]$ and bpnt $= [0,1,1,2]$.

Clearly no new elements will be added in the next step, so the algorithm stops.

Let $S$ be a generating set for the group $G$ and $T$ a Schreier tree for $S$ with root $x$. If $a \in Gx$, then $a$ is a vertex of $T$. Hence there is a unique path from $x$ to $a$ in the tree. This path is helpful in finding an element in $G$ mapping $x$ to $a$.

**Definition 8.3.12.** Let $G$ is a permutation group acting transtively on the set $X$. Suppose $S$ is a generating set of permutations for $G$ and $T$ a Schreier tree with respect to $S$ rooted at the point $x \in X$. If the labels of the edges in the unique path from $x$ to $a$ are $b_1, \ldots, b_k$, respectively, then the element $t_a = b_k \cdot \ldots \cdot b_1$ of $G$ satisfies $t_a(x) = a$. The map $t\colon X \to G$ obtained in this way is the *Schreier transversal* for G (determined by $T$).

**Example 8.3.13.** Consider the permutation group $G = \langle a, b\rangle$ where $a = (1,2)(3,4)$ and $b = (1,3)(2,4)$. Take the following Schreier tree with root $1$. We compute the various transversals for the Schreier tree from Example 8.3.11: $t_1 = 1$, $\mathrm{t}_2 = a$, $\mathrm{t}_3 = b$, and $t_4 = b \cdot a$.

Schreier transversals will turn out to be useful tools to construct generators for stabilizer subgroups.

**Algorithm 8.3.14** (Schreier transversal)**.** • *Input: Schreier data $D$ for a Schreier tree $T$ and a vertex $v$ of the tree.*

- *The image $t_v$of $v$ under the Schreier transversal of $T$.*

```
SchreierTransversal := procedure($D, v$)
local variables
    $V := D[1]$ , $L := D[2]$ , $W := D[3]$
    $i := 1$ , $p := v$ , $t := 1$ , root $:= V[1]$
while $p \neq$ root do
    while $V[i] \neq p$ do
        $i := i + 1$

    $t := L[i] \cdot t$ , $p := W[i]$ , $i := 1$
return
    $t$
```

Let $S$ be a generating set of permutations for the group $G$ acting on the set $X$. Let $T$ be a Schreier tree with respect to $S$ rooted at the point $x$ of $X$. Let $V$ be the vertex set of $T$. Finally let $t\colon V \to G$ be the Schreier transversal for $G$ determined by $T$.

**Theorem 8.3.15** (Schreier's lemma)**.** *The stabilizer $G_x$ is generated by* $\left\{ \left(t_{s(v)}\right)^{-1} \cdot s \cdot t_v \,\middle|\, (s,v) \in S \times V \right\}$.

*Proof.* Let $M$ be the set $\left\{ \left(t_{s(v)}\right)^{-1} \cdot s \cdot t_v \,\middle|\, sv \in S \times V \right\}$.

**$M$ is contained in $G_x$.**

We show $\left(t_{g(v)}\right)^{-1} \cdot g \cdot t_v \in G$ for all $g \in G_x$. Then certainly for all $g \in S$ the statement is true.

Indeed, $\left(t_{g(v)}\right)^{-1} \cdot g \cdot t_v(x) = \left(t_{g(v)}\right)^{-1} \cdot g(v) = x$.

***M* generates $G_x$.**

Suppose $g \in G_x$. Then $g \in G$ and so it can be written as a product of elements of $S$ and their inverses. Thus, $g = g_r \cdot \ldots \cdot g_1$ with $g_i \in A \cup A^{-1}$. We will show with induction on $r$, that $g$ can be written as a product of elements from $M$.

If $r = 1$, then $g \in S$ and $g(x) = x$. As $t_x = 1$, we find $g = \left(t_{g(x)}\right)^{-1} \cdot g \cdot t_x$, and so $g \in M$, as required.

Assume, therefore, $r > 1$. Let $j$ be the maximal index such that

$$x, g_1(x), g_2 \cdot g_1(x), \ldots, g_j \cdot \ldots \cdot g_2 \cdot g_1(x)$$

is a path in $T$ with the labels $g_1, g_2, \ldots, g_j$. Observe that $j < r$ as $T$ has no cycles. Put $a = g_j \cdot \ldots \cdot g_2 \cdot g_1(x)$. Then $t_a = g_j \cdot \ldots \cdot g_2 \cdot g_1$. Now consider the element $g \cdot \left(\left(t_{h(a)}\right)^{-1} \cdot h \cdot t_a\right)^{-1}$, where $h = g_{j+1}$. Rewrite this element as $g \cdot \left(\left(t_{h(a)}\right)^{-1} \cdot h \cdot t_a\right)^{-1} = g_r \cdot \ldots \cdot g_{j+2} \cdot t_{h(a)}$. Now repeat the above argument on this element; it also belongs to $G_x$. As $t_{h(a)}$ corresponds to a path in the Schreier tree $T$, we find an expression of the form

$$g \cdot \left(\left(t_{h(a)}\right)^{-1} \cdot h \cdot t_a\right)^{-1} \cdot \left(\left(t_{h'(a')}\right)^{-1} \cdot h' \cdot t_{a'}\right)^{-1} = g_r \cdot \ldots \cdot g_{j'+2} \cdot t_{h'(a')}$$

with $j' > j$. Thus, we can repeat the argument at most $r - j$ times; each time the head $g_r \cdot \ldots \cdot g_{j'+2}$ becomes shorter and shorter. We finish with a Schreier element $t$ in the right hand side. However, since the left hand side is a product of $g$ and (inverses of) elements stabilizing $x$, also $t$ belongs to $G_x$. But this is only possible if $t = 1$. Consequently, $g$ is a product of elements from $M$ and their inverses. Hence the theorem follows.

□

**Example 8.3.16.** Consider the permutation group $G = \langle a, b \rangle$ where $a = (1,2)(3,4)$ and $b = (1,3)(2,4)$. Take the Schreier tree with root 1 and edges $(1,2,a)$, $(1,3,b)$, $(2,4,b)$. Consequently, $t_1 = 1$, $t_2 = a$, $t_3 = b$, $t_4 = b \cdot a$.

Using this knowledge we compute the generators for $G$ that are indicated by Schreier's lemma. $(t_2)^{-1} \cdot a \cdot t_1 = 1$, $(t_3)^{-1} \cdot b \cdot t_1 = 1$, $(t_1)^{-1} \cdot a \cdot t_2 = 1$, $(t_4)^{-1} \cdot b \cdot t_2 = 1$, $(t_4)^{-1} \cdot a \cdot t_3 = 1$, $(t_1)^{-1} \cdot b \cdot t_3 = 1$, $(t_3)^{-1} \cdot a \cdot t_4 = 1$, $(t_2)^{-1} \cdot b \cdot t_4 = 1$. We conclude that $G_1 = \{1\}$ and $|G| = 4$.

Often, many of the $|X| \cdot |A|$ generators of $G$ are redundant. Unfortunately, we cannot say in advance which.

**Algorithm 8.3.17** (Stabilizer Algorithm)**.** • *Input: list of generators S of a permutation group G acting on a set X and a point x.*

- *Output: List of generators for the stabilizer in G of x.*

```
Stabilizer := procedure($S$,$x$)
local variables
    Tree := SchreierTree($S$,$x$) , Vertices := Tree[1] , $i$, $j$, Stab, $s$, $v$, $t$, $t_1$
while $i \leq Length(S)$ do
    while $j \leq Length(V)$ do
        $s := S[i]$ , $v := V[j]$ , $t :=$ SchreierTransversal(Tree,$v$)
        $t_1 :=$ SchreierTransversal(Tree,$s(v)$)
        if $(t_1)^{-1} \cdot s \cdot t \neq 1$
            then
                Stab := Stab $\cup \left[\left[(t_1)^{-1} \cdot s \cdot t\right]\right]$
        $j := j+1$
    $i := i+1$
return
    Stab
```

The algorithms presented so far enable us to compute the order of a permutation group $G$ acting on a finite set $X$, once we are given a set of generating permutations for $G$.

For convenience we assume the set $X$ to be the set $\{1,...,n\}$.

**Algorithm 8.3.18** (Order algorithm). • *Input: a list S of generating permutations for the permutation group G on* $\{1,...,n\}$.

- *Output: the order of G.*

```
Order := procedure($S$)
local variables
    order := 1
    gens := $S$
    $i := 1$
while gens $\neq \varnothing$ do
    order := order$\cdot$ Length(Orbit(gens,$i$)) , gens := stabilizer(gens,$i$) , $i := i+1$
return
    order
```

*Proof.*

**Termination**

The algorithm stops since the set of points fixed by $G$ becomes larger every time one passes Step 2. Eventually the stabilizer of all these points will be trivial and the points form a basis.

**Correctness**

The Order theorem implies that the output of the algorithm is the order of $G$.

□

## 8.4 Automorphisms

A regular triangle looks more symmetric than a nonequilateral triangle in the plane. The notion of symmetry can be attached to any mathematical object or set with some additional structure. This structure need not necessarily be algebraic, but can also be, for example, a graph. An *isomorphism* mapping the structure into itself is called an *automorphism*. The set of all automorphisms of a structure is a group with respect to composition of maps. This group represents the symmetry of the structure. We will study automorphism groups of various structures. Such symmetry groups are important for determining and investigating regular structures in nature, like molecules and crystals.

We recall that a graph consists of a vertex set $V$ and an edge set $E$, whose elements are subsets of $V$ of size 2.

**Definition 8.4.1** (Automorphisms)**.** • An *automorphism* of a graph $(V,E)$ is a bijective map $f\colon V \to V$ satisfying if $\{v,w\} \in E$ then $\{g(v),g(w)\} \in E$.

- Let $K$ be a ring, field, group or monoid. An *automorphism* of $K$ is an isomorphism $K \to K$.

**Example 8.4.2** (Inner automorphisms of a group)**.** Let $G$ be a group. For $g \in G$, conjugation by $g$, that is, the map $x \mapsto g\cdot x\cdot g^{-1}$, is an automorphism of $G$. These automorphisms are also called *inner* automorphisms of the group $G$.

**Example 8.4.3** (Automorphisms of a finite field)**.** Suppose $p$ is a prime. By Frobenius Automorphisms, the map $x \mapsto x^p$ is an automorphism of a field of characteristic $p$.

**Example 8.4.4** (Automorphisms of the rational numbers)**.** There is exactly one automorphism of the field $\mathbb{Q}$ : the identity.

For an automorphism $\mathrm{phi}\colon \mathbb{Q} \to \mathbb{Q}$, we have $\mathrm{phi}(1) = 1$, so $\mathrm{phi}(2) = \mathrm{phi}(1+1) = 1+1 = 2$, etc. By induction, $\mathrm{phi}(m) = m$ for positive integers $m$. From $\mathrm{phi}(0) = 0$ it follows that $\mathrm{phi}(0) = \mathrm{phi}(m+(-m)) = \mathrm{phi}(m) + \mathrm{phi}(-m) = m + \mathrm{phi}(-m) = 0$ and so $\mathrm{phi}(-m) = -m$ for all positive integers $m$ (here use we that is an automorphism of the additive group of $\mathbb{Q}$).

For $\frac{a}{b} \in \mathbb{Q}$, with $b$ positive, $a = b\cdot \frac{a}{b}$. This implies $a = \mathrm{phi}(a) = \mathrm{phi}\left(b\cdot \frac{a}{b}\right) = \mathrm{phi}(b) \cdot \mathrm{phi}\left(\frac{a}{b}\right) = b\cdot \mathrm{phi}\left(\frac{a}{b}\right)$. In particular, $\mathrm{phi}\left(\frac{a}{b}\right) = \frac{a}{b}$.

If $\mathbb{Q}$ is a subfield of the field $K$ then the same argument shows that every automorphism of $K$ fixes all elements in $\mathbb{Q}$.

**Example 8.4.5** (Automorphisms of the cyclic group of order $n$)**.** Let $C$ be a group of order $n$ generated by $g$. An automorphism of $C$ is determined by the image of $g$, which, must be of the form $g^j$ for an integer $j$ with $\gcd(j,n) = 1$. For, otherwise the element $g^j$ does not have the same order as $g$. On the other hand, for each such exponent $j$ prime to $n$ the map $g \mapsto g^j$ is an automorphism.

**Remark 8.4.6.** There is some ‘asymmetry’ between the definition of automorphism for graph on the one hand and group, ring, field, etc., on the other. This is not necessary. One could define a morphism of graphs $(V,E) \to (V',E')$ as a map $f\colon V \to V'$ such that $\{f(x),f(y)\} \in E'$ whenever $\{x,y\} \in E$. Then an isomorphism of graphs is a bijective morphism whose inverse is also a morphism (in contrast to the ring case, this requirement is necessary), and an automorphism of the graph $(V,E)$ is an isomorphism $(V,E) \to (V,E)$. We stayed away from this approach as we do not use the notions any further.

**Theorem 8.4.7.** *Let $K$ be a graph, a ring, a field, a group, or a monoid. The set of all automorphisms of $K$ is a subgroup of* $\mathrm{Sym}(K)$*. It is denoted by* $Aut(K)$ *and is called the* automorphism group *of $K$.*

*Proof.* Automorphisms of $K$ are bijective and so belong to $\mathrm{Sym}(K)$.

The subset of all automorphisms is not empty as the identity is an automorphism.

If $g$ is an automorphism of $K$, then so is $g^{-1}$ (by definition if $K$ is a graph, by a Isomorphisms of monoids if $K$ for the other structures).

Likewise, if $g$ and $h$ are automorphisms of $K$, then so is the composition $g \cdot h$.

□

**Example 8.4.8. The automorphism group of a regular n-gon in the plane.** We have already met the example of the group $D_n$, which is the group of symmetries of a regular $n$ - gon in the plane. We have seen that this group is also a subgroup of the automorphism group of the $n$ - gon as a graph. In fact, it is the full automorphism group of the graph. Prove this!

**Example 8.4.9** (The Petersen graph)**.** Let $P$ be the Petersen graph. The vertices of $P$ can be identified with the pairs of elements from $\{1,2,3,4,5\}$. Two vertices $\{x,y\}$ and $\{u,v\}$ are adjacent if and only if their intersection $\{x,y\} \cap \{u,v\}$ is empty. The group $\mathrm{Sym}_5$ acts on the set $\{1,2,3,4,5\}$, but also on the vertex set of $P$. For, if $g \in \mathrm{Sym}_5$, then the map $g_2$ defined by $g_2(\{x,y\}) = \{g(x), g(y)\}$ defines a permutation of the 10 vertices of $P$. See Example 8.4.8

This implies that the automorphism group $G$ of $P$ contains a subgroup, denoted by $H$, isomorphic to $\mathrm{Sym}_5$. This subgroup acts transitively on the vertex set of the graph.

The triple $B$ consisting of the vertices $\{1,2\}, \{1,3\}, \{2,4\}$ is a basis for $G$. For, if an element of $G$ fixes these vertices, then it also fixes the unique common neighbour $\{4,5\}$ of $\{1,2\}$ and $\{1,3\}$ and, similarly, $\{3,5\}$ the unique common neighbour of $\{1,2\}$ and $\{2,4\}$.

Since each further vertex of the Petersen graph is connected with a unique vertex from the pentagon with vertices $\{1,2\}$, $\{1,3\}$, $\{2,4\}$, $\{4,5\}$ and $\{3,5\}$, the element fixes all vertices of $P$. This argument establishes that the stabilizer in $G$ of $B$ is indeed trivial.

The $G$-orbit of $B$ contains at least the 120 images of $B$ under the group $H$. But as the Petersen graph contains precisely 30 ordered edges, and for each such edge, there are only 4 points nonadjacent to any vertex of the edge, the $G$-orbit of $B$ contains at most 120 images of $B$. We can conclude that the order of $G$ equals 120. In particular, $G$ equals $H$ and is isomorphic to $\mathrm{Sym}_5$.

**Example 8.4.10** (The cyclic group of order $n$)**.** Let $C$ be a group of order $n$ generated by $g$. By Example 8.4.5, the order of $Aut(C)$ is the Euler indicator $\Phi(n)$ of $n$. The group $Aut(C)$ is commutative but need not be cyclic: a counterexample occurs for $n = 8$.

**Example 8.4.11** (Symmetries of the 5-gon)**.** When we look at the regular pentagon in the plane, we can consider symmetries in two ways:

- as automorphisms of the Euclidean plane (rotations, reflections, etc.) that leave invariant the pentagon;
- as a group of permutations of the graph with vertex set

$$\{1,2,3,4,5\} \tag{8.11}$$

  and edge set

$$\{\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{1,5\}\} \tag{8.12}$$

Naturally the former symmetry group (subject to more restrictions) is contained in the latter. Remarkably enough, the two groups coincide. They are both the dihedral group $D_5$ of order 10. The elements of order 5 correspond to rotations around the origin with angle a multiple of 72 degrees, and the elements of order 2 to reflections in an axis through the center and one of the vertices of the pentagon.

We will have a closer look at automorphisms of fields.

**Theorem 8.4.12.** *Let $K$ be a subfield of $L$ and let $f \in K[X]$ be an irreducible polynomial. If $x, y$ are two roots of $f$ in $L$, then there is an isomorphism $h\colon K(x) \to K(y)$ with $h(x) = y$.*

*Proof.* Write $L = K[X]/\{f\}K[X]$ and consider the maps $r\colon L \to K(x), g \mapsto g(x)$ and $s = L \to K(y), g \mapsto g(y)$.

These maps are well defined since $x$ and $y$ are roots of $f$.

We proceed in three steps.

**The maps $r$ and $s$ are surjective.**

First we claim that $K(x)$ consists of the elements $g(x)$ with $g \in K[X]$. For if $g(x) \neq 0$, then $g$ is not divisible by $f$ as $f$ has $x$ as a root. Thus, there are polynomials $a, b$ in $K$ with $a \cdot f + b \cdot g = 1$. Substitution of $x$ for $X$ yields: $(g(x))^{-1} = b(x)$. Therefore, the inverse of $g(x)$ also belongs to $K$. Thus, the expressions of the form $g \cdot x$, where $x \in K$ form a subfield of $K(x)$ containing $x$. As $K(x)$ is the smallest field containing $x$, the claim follows.

**The maps $r$ and $s$ are isomorphisms, so $L \cong K(x)$ and $L \cong K(y)$.**

Consider the substitution map $K[X] \to K(x), g \mapsto g(x)$. It is easily seen to be a morphism. By the first part of the proof, it is surjective. Its kernel is the ideal generated by $f$, since $x$ is a zero of $f$ and the latter is irreducible. The First isomorphism theorem then gives that there is an isomorphism as required. The proof for $s$ is similar.

**The map $s \circ r^{-1}$ is the required isomorphism.**

By the previous part, the composition $r \circ s^{-1}$ is an isomorphism $K(x) \to K(y)$.

□

**Example 8.4.13** (Gaussian numbers)**.** $K = \mathbb{Q}(i)$, where $i = \sqrt{-1}$. Each element of $K$ is of the form $a + b \cdot i$ with $a, b \in \mathbb{Q}$. The map $\mathbb{Q}(i) \to \mathbb{Q}(i), a + b \cdot i \mapsto a - b \cdot i$ is an automorphism. This follows from the rules for complex conjugation. The square of this map is equal to the identity. In fact, the group of automorphisms consists of the identity and the conjugation map. In order to see this, note that, for each automorphism $s$, we have $s(a + b \cdot i) = a + b \cdot s(i)$, so the automorphism is fully determined by the image of $i$. Now $i^2 = -1$, so $i$ is a root of $X^2 + 1$. Also $-i$ is a root of this polynomial. Both roots of $X^2 + 1$ correspond indeed to an automorphism of $K$, namely $i$ corresponds to the identity and $-i$ corresponds to $c$. So $Aut(K)$ is a group of order two and hence isomorphic to $C_2$. The possible automorphisms are apparently connected to the zeros of the polynomial $X^2 + 1$.

**Example 8.4.14** (Cubes roots of 2)**.** Consider $K = \mathbb{Q}(x)$, where $x = \sqrt[3]{2}$. Let $s$ be an automorphism of $K$. If it fixes $x$, then it is obviously the identity. If $s$ is not the identity, it must move $x$ to another solution of $X^3 - 2$. These solutions do not exist in $K$. An intuitive way of seeing this runs as follows: the other solutions are $\mathrm{e}^{\frac{2 \cdot \pi \cdot i}{3}} \cdot x$ and $\mathrm{e}^{\frac{4 \cdot \pi \cdot i}{3}} \cdot x$, and these are complex imaginary numbers, whereas $K$ is a subfield of $\mathbb{R}$. Thus, $s$ must fix $x$ and, since $x$ generates $K$, the automorphism group of $\mathbb{Q}(x)$ is trivial. In particular, it is strictly smaller than the dimension of $\mathbb{Q}(x)$ over $\mathbb{Q}$.

The isomorphism constructed in [?] fixes every element of the subfield $K$ of $L$.

On the other hand, each isomorphism $K(x) \to K(y)$ which fixes $K$ elementwise, is determined by the image of $x$. Hence we can determine automorphism groups of finite fields.

**Theorem 8.4.15** (Automorphisms of finite fields)**.** *Let $p$ be a prime number and $q = p^a$ a power of $p$. If $K$ is a finite field of order $q$, then $Aut(K)$ is a cyclic group of order $a$ generated by the map $K \to K, x \mapsto x^p$.*

*Proof.* Let $z \in K$ be a primitive element of $K$.

An automorphism of $K$ is determined by the image of $z$.

Let $f$ be the minimal polynomial of $z$ over $\mathbb{Z}/p\mathbb{Z}$. Then $f$ has degree $a$.

Let $g$ be an automorphism of $K$. Then $g(z)$ is a zero of $f$. Hence there are at most $a$ possibilities for $g(z)$.

On the other hand, $a$ possibilities occur: $z, z^p, z^{p^2}, \ldots, z^{p^{a-1}}$.

So $Aut(K)$ is a cyclic group of order $a$ generated by the automorphism sending $z$ to $z^p$.

□

**Example 8.4.16** (The field of order 8)**.** Consider $K = \mathbb{Z}/2\mathbb{Z}[X]/(X^3+X+1)\mathbb{Z}/2\mathbb{Z}[X]$. The polynomial $X^3+X+1 \in \mathbb{Z}/2\mathbb{Z}[X]$ is irreducible, so $K$ is a field of order 8.

Put $x = X + (X^3+X+1)\mathbb{Z}/2\mathbb{Z}[X]$, so that $K = \mathbb{Z}/2\mathbb{Z}(x)$.

The polynomial $X^3+X+1$ has 3 roots in $K$, viz., $x, x^2$, and $x+x^2$. Each of them leads to an automorphism. For example, the root $x^2$ corresponds to the map sigma: $K \to K$ sending $x$ to $x^2$. That is, sigma $\left(a+b\cdot x+c\cdot x^2\right) = a+b\cdot x^2+c\cdot x^4 = a+c\cdot x+(b+c)\cdot x^2 = \left(a+b\cdot x+c\cdot x^2\right)^2$. (Verify!) The automorphism sigma$^2$ satisfies sigma$^2(x) = s\left(x^2\right) = x^4 = x+x^2$. Apparently, this is the automorphism sending $x$ to the third root of $X^3+X+1$.

The group of automorphisms of $K$ has order 3, and hence is isomorphic to $C_3$.

## 8.5 Quotient groups

We will introduce computations modulo a normal subgroup and the corresponding construction of the quotient group.

Let $G$ be a group and let $N$ be a normal subgroup of $G$. The notions of left and right coset (a set of the form $g\cdot N$) and right coset (a set of the form $N\cdot g$) of $N$ in $G$ coincide since normal subgroups satisfy $g\cdot N = N\cdot g$ for all $g \in G$. Thus, we can just speak of cosets.

**Theorem 8.5.1** (Multiplying cosets of normal subgroups)**.** *Suppose that $N$ is a normal subgroup of $G$. Then, for all $a, b \in G$ we have $a\cdot N\cdot b\cdot N = a\cdot b\cdot N$;*

*Proof.* $a\cdot N\cdot b\cdot N = a\cdot N\cdot b\cdot N = a\cdot b\cdot N\cdot N = a\cdot b\cdot N\cdot N = a\cdot b\cdot N$

□

**Example 8.5.2.** Let $G$ be the symmetric group $\mathrm{Sym}(3)$. The subgroup $H = \langle(1,2,3)\,(1,3,2)\rangle$ of order 3 is a normal subgroup. It has index 2.

More generally, whenever $H$ is a subgroup of $G$ of index 2, it is a normal subgroup. For then, for $g \in G$, either $g \in H$ and so $g\cdot H = H = H\cdot g$ or or not, in which case $g\cdot H = G\setminus H = H\cdot g$.

**Example 8.5.3.** Let $G$ be the group of all motions in the plane. The subgroup $T$ of all translations of the plane is a normal subgroup. Fix a point $p$ of the plane. The subgroup $H$ of $G$ of all elements fixing the point $p$ is a complement of $T$ in the sense that

- $H \cap T = \{1\}$.
- $G = H \cdot T$.

As a consequence, setwise $G$ can be identified with the Cartesian product of $H$ and $T$. But groupwise, it is not the direct product of these two groups.

Due to Multiplying cosets of normal subgroups, the set $G/N$ of cosets admits a group structure.

**Definition 8.5.4** (Definition of quotient group)**.** We call the group $G/N$ with

- multiplication: $g \cdot N \cdot g' \cdot N = g \cdot g' \cdot N$
- unit: $N$
- inverse: $g \cdot N \to g^{-1} \cdot N$

the *quotient group* of $G$ with respect to $N$.

**Remark 8.5.5.** Normal subgroups play the same role for groups as ideals do for rings. The procedure for making a quotient group is similar to the construction of a residue class ring.

**Example 8.5.6.** If $G$ is a commutative group, then each subgroup $H$ of $G$ is a normal subgroup. Thus, the quotient group $G/H$ always exists. Moreover, it is commutative.

**Example 8.5.7.** The additive group of $\mathbb{Q}$ is commutative. Therefore, the subgroup $\mathbb{Z}$ is a normal subgroup of $\mathbb{Q}$. The cosets of $\mathbb{Z}$ in $\mathbb{Q}$ are the sets of the form $\frac{a}{b} + \mathbb{Z}$, where $a, b$ in $\mathbb{Z}$ and $b \neq 0$. For example, $\frac{1}{2} + \mathbb{Z}$. Computing in the quotient $\mathbb{Q}/\mathbb{Z}$ comes down to 'computing modulo integers'. For example $\frac{3}{4} + \mathbb{Z} + \left(\frac{5}{6} + \mathbb{Z}\right) = \frac{7}{12} + \mathbb{Z}$.

Computing modulo a normal subgroup behaves well, as becomes clear by the following result.

**Theorem 8.5.8.** *Let $N$ be a normal subgroup of the group $G$. The map* phi$: G \to G/N, g \mapsto g \cdot N$ *is a surjective homomorphism with kernel $N$.*

*Proof.* Clearly, phi is surjective. Moreover phi is a homomorphism of groups, Indeed, for all $g, h \in G$ we have phi$(g \cdot h) = g \cdot h \cdot N = g \cdot N \cdot h \cdot N =$ phi$(g) \cdot$ phi$(h)$.

The kernel of phi consists of the elements $g \in G$ satisfying phi$(g) = N$, that is, $g \cdot N = N$. Since $g \cdot N = N$ is equivalent to $g \in N$, we find the kernel of phi to be equal to $N$. .

□

**Example 8.5.9.** Let $G$ be the set of all $2\times 2$ matrices with entries from a field $F$ of the form $\begin{pmatrix}1 & x\\ 0 & y\end{pmatrix}$, where $x$ is an arbitrary element of $F$ and $y$ a nonzero element of $F$. Then $G$ is a subgroup of $GL(2,F)$. The subgroup $N$ of all matrices of the form $\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}$ is a normal subgroup of $G$. The quotient group $G/N$ is isomorphic to the multiplicative group on $F\setminus\{0\}$. Observe that $N$ is the kernel of the determinant, viewed as a homomorphism.

In Normal subgroups and Kernels of homomorphisms it was shown that the kernel of a group homomorphism is a normal subgroup. A normal subgroup is the kernel of a hommorphism states the converse, namely that every normal subgroup is the kernel of a homomorphism.

Let $f\colon G\to H$ be a surjective group homomorphism with kernel $N$. According to a previous proposition, $N$ is a normal subgroup of $G$.

**Theorem 8.5.10** (First isomorphism theorem for groups)**.** *If $G$ and $H$ are groups and $f\colon G\to H$ is a surjective homomorphism with kernel $N$, then the map $f'\colon G/N\to H$ defined by $f'(g\cdot N)=f(g)$ is an isomorphism.*

*Proof.* The important steps in the proof are the following two.

**The map $f'$ is well defined.**

Suppose $g'\in g\cdot N$. Then there is $n\in N$ with $g'=g\cdot n$. Consequently, $f(g')=f(g\cdot n)=f(g)\cdot f(n)=f(g)\cdot 1=f(g)$. Thus $f(g\cdot N)$ does not depend on the choice of $g'\in g\cdot N$.

**$f'$ is injective.**

Suppose $g\in G$ satisfies $f'(g\cdot N)=1$. Then $f(g)=1$, so $g\in N$, whence $g\cdot N=N$, which is the identity element of $G/N$. We have shown that $\mathrm{Ker}(f')$ is trivial, so, $f'$ is injective.

□

**Example 8.5.11** (Cyclic groups)**.** The classification of cyclic groups can be handled easily with the theorem. Because $G$ is cyclic, there exists $g\in G$ with $\langle\{g\}\rangle_G=G$. Consider the map $f\colon\mathbb{Z}\to G, i\mapsto g^i$. It is a surjective homomorphism with kernel $n\cdot\mathbb{Z}$ for some non-negative integer $n$. The assertion that every cyclic group is isomorphic to either $\mathbb{Z}$ (the case where $n=0$) or $C_n$ for some positive integer $n$ now follows directly from the First isomorphism theorem for groups applied to $f$.

**Example 8.5.12** (Different groups with same quotient and kernel)**.** Let $G$ be a group and $N$ a normal subgroup of $G$ distinct from 1 and from $G$. The groups $G/N$ and $N$ are both smaller than $G$. A lot of information about $G$ can be obtained from study of these two smaller groups. However, the exact structure of $G$ is not completely determined by $G/N$ and $N$. For instance, the groups $C_4$ and $C_2\times C_2$ both have a normal subgroup isomorphic with $C_2$, and in both cases the quotient group is isomorphic with $C_2$.

**Example 8.5.13** (The quotient of the symmetric group by the alternating group)**.** The group $\mathrm{Sym}_n/\mathrm{Alt}_n$ is isomorphic with $C_2$. For, the map permutation1.sign: $\mathrm{Sym}_n \to \{1,-1\}$ is a surjective homomorphism of groups with kernel $\mathrm{Alt}_n$. Here, $\{1,-1\}$ is the group of invertible elements of the monoid $[\mathbb{Z},\cdot,1]$. This group is isomorphic with $C_2$.

**Example 8.5.14** (The general linear group)**.** The quotient group $GL(n,\mathbb{R})/SL(n,\mathbb{R})$ is isomorphic to the multiplicative group $\mathbb{R}^\times$. The subgroup $SL(n,\mathbb{R})$ is the kernel of the determinant map linalg1.determinant: $GL(n,\mathbb{R}) \to \mathbb{R}^\times$.

## 8.6 Structure theorems

We introduce some common (series of) groups, some of which occur in the Classification of groups of order at most 11

**Definition 8.6.1** (Dihedral and quaternion groups)**.**
- The *dihedral group* of order $2\cdot n$ is the group $D_n$ generated by two elements $a$ and $b$ with multiplication determined by $b^n = a^2 = 1$ and $a\cdot b = b^{n-1}\cdot a$.
- The *quaternion group* is the group of order 8 consisting of the following invertible quaternions. $Q_8 = \{1,-1,i,-i,j,-j,k,-k\}$.

**Example 8.6.2.** The group $D_n$ has been introduced in Example 6.4.9 as the symmetry group of the regular $n$-gon in the plane. The element $b$ is clockwise rotation over $\frac{2\cdot\pi}{n}$ degrees. The element $a$ is a reflection with mirror through a vertex. The corresponding permutation representation is described in the Example 8.1.8.

**Example 8.6.3** (The quaternion group as a permutation group)**.** Left multiplication in the quaternion group gives the transitive permutation representation determined by the following assignments:

$$i \mapsto (1,i,-1,-i)\,(j,k,-j,-k)\,,j \mapsto (1,j,-1,-j)\,(i,-k,-i,k)\,,k \mapsto (1,k,-1,-k)\,(i,j,-i,-j,k)\,.$$

Replacing the elements by numbers $\{1,...,8\}$, a more usual description is obtained.

**Example 8.6.4** (The quaternion group by means of matrices)**.** An injective morphism $Q \to GL(2,\mathbb{C})$ is determined by $i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, j \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Verify that this forces $k \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$.

**Remark 8.6.5.** The groups introduced in Dihedral and quaternion groups are mutually non-isomorphic. Since $|D_n| = 2\cdot n$ and $|Q_8| = 8$, the only two groups which might be isomorphic to each other are $D_4$ and $Q_8$. But they are not: $D_4$ has only two elements of order 4, viz., $b$ and $b^3$, whereas $Q_8$ has 6 elements of order 4, namely all but 1 and $-1$.

We will present some powerful structure theorems about finite groups, starting with some properties of finite commutative groups.

**Lemma 8.6.6.** *Let $G$ be a group.*

*i. If every non-identity element of $G$ has order $2$, then $G$ is commutative.*

*ii. If $G$ is a commutative group and $n$ is a natural number, then $T(G,n) = \{x \in G | x^n = 1_G\}$ is a subgroup of $G$. It is called the $n$-torsion of $G$.*

*iii. Each finite commutative group is isomorphic to a direct product of commutative groups of prime power order.*

*Proof.*

**If every non-identity element of $G$ has order $2$, then $G$ is commutative.**

Suppose $x, y \in G$. Since $x$ and $y$ have order 1 or 2, we have $x^{-1} = x$ and $y^{-1} = y$. Consequently, $x^{-1} \cdot y^{-1} \cdot x \cdot y = x \cdot y \cdot x \cdot y = (x \cdot y)^2 = \mathrm{e}$. Multiplying the extreme sides by $y \cdot x$ from the left, we find $x \cdot y = y \cdot x$.

**If $G$ is a commutative group and $n$ is a natural number, then $T(G,n) = \{x \in G | x^n = 1_G\}$ is a normal subgroup of $G$.**

Clearly, the identity element of $G$ belongs to $T(G,n)$. Suppose $a, b \in T(G,n)$. Then, because $G$ is commutative, $(a \cdot b)^n = a^n \cdot b^n = 1_G$ and $\left(a^{-1}\right)^n = (a^n)^{-1} = {1_G}^{-1} = 1_G$. Thus, $a \cdot b, a^{-1} \in T(G,n)$, proving that $T(G,n)$ is a subgroup of $G$.

The subgroup $T(G,n)$ is normal in $G$, as the order of an element $g$ in $G$ is invariant under conjugation.

**Each finite commutative group is isomorphic to a direct product of finite commutative groups of prime power order.**

Suppose that $G$ is a finite commutative group of order $|G| = a \cdot b$ where $a$ and $b$ are positive integers which are mutually prime. Then $T(G,a)$ and $T(G,b)$ are normal subgroups of $G$ intersecting in the identity element. If $g$ is an element in $T(G,a)$ and $h$ of $T(G,b)$, then $g \cdot h \cdot g^{-1} \cdot h^{-1}$ is contained in both $T(G,a)$ and $T(G,b)$ and hence trivial. This implies that $g \cdot h = h \cdot g$. So, $G = T(G,a) \times T(G,b)$.

□

**Example 8.6.7.** The group ${C_2}^n$ has only elements of order 2 and 1.

**Example 8.6.8.** There exist commutative groups of the same prime power order that are not isomorphic. Indeed, a cyclic group of order $p^n$, where $p$ is prime and $n$ is at least 2 is not isomorpic to the direct product of $n$ copies of the cyclic group of order $p$.

**Remark 8.6.9.** Of course there are commutative groups with elements that do not have order 2. Indeed, any cyclic group is commutative.

We prove two more preliminary results for the important theorems that will appear soon.

**Lemma 8.6.10.** *Let $G$ be a finite group.*

*i. If $|G|$ is a prime, then $G$ is cyclic.*

*ii. If $q$ is the highest power of the prime $p$ dividing $|G|$ and $S$ is a normal subgroup of $G$ of order $q$, then every subgroup of $G$ whose order is a power of $p$ is a subgroup of $S$.*

*Proof.*

**If $|G|$ is a prime, then $G$ is cyclic.**

By the Lagrange's theorem, there is $g \in G$ of order $p = |G|$. But then the subgroup of $G$ generated by $g$ has the same size as $G$ and so coincides with $G$. Therefore $g$ is a generator of $G$ and $G \cong C_p$.

**If $q$ is the highest power of the prime $p$ dividing $|G|$ and $S$ is a normal subgroup of $G$ of order $q$, then every subgroup of $G$ whose order is a power of $p$ is a subgroup of $S$.**

Let $K$ be a subgroup of $G$ whose order is a power of $p$. Since $S$ is a normal subgroup of $G$, the product $S \cdot K$ is a subgroup of $G$. (See .) Clearly, $S$ is also a normal subgroup of $S \cdot K$. By First isomorphism theorem for groups, the quotient group $S \cdot K/S$ is isomorphic to $K/S \cap K$. But by Lagrange's theorem the order of this group is a divisor of $|K|$ and so a power of $p$. Consequently, the order of $S \cdot K$ is also a power of $p$. But this group contains $S$ of order $q$, the highest power of $p$ occurring in $|G|$, so we must have $S \cdot K = S$, proving that $K$ is contained in $S$.

□

**Example 8.6.11.** If $p$ is a prime, all nontrivial elements of the group $(C_p)^n$ have order $p$.

**Example 8.6.12.** The group $D_p$ with $p$ a prime has order $2 \cdot p$ and contains reflections of order 2 and rotations of order $p$.

The subgroup of the $p$ rotations is normal in $D_p$ and contains all elements of order $p$.

Sylow's theorem presented below is a very powerful result, with which we can analyse the structure of finite groups. It enables us for example to obtain a Classification of finite commutative groups, as well as a Classification of groups of order at most 11.

**Theorem 8.6.13** (Sylow's theorem)**.** *Let $G$ be a finite group, $p$ a prime number dividing $|G|$, and let $q$ be the highest power of $p$ dividing $|G|$.*

*i. $G$ has an element of order $p$.*

*ii. $G$ has a subgroup of order $q$.*

*iii. If $H$ is a subgroup of $G$ whose order is a power of $p$, then $H$ is a subgroup of a subgroup of $G$ of order $q$.*

*iv. Any two subgroups of $G$ of order $q$ are conjugate.*

*Proof.* Let $G$ be a finite group, $p$ a prime number dividing $|G|$, and let $q$ be the highest power of $p$ dividing $|G|$.

**$G$ has an element of order $p$.**

This is a direct consequence of the second statement. Indeed, by Lagrange's theorem, any non trivial element of a subgroup of order $q$ in $G$ has order a (nontrivial) divisor of $q$. So, an approprate power of the element has order $p$.

**$G$ has a subgroup of order $q$.**

Recall that $|G| = q \cdot m$ with $\gcd(q,m) = 1$. If $m = 1$, we can take $G$ itself to be the required subgroup. Hence the result for $m = 1$.

We proceed by induction on $|G|$. Assume the truth of the assertion for all groups of order smaller than $|G|$. Consider the set $X$ of all subsets of $G$ of size $q$. The group $G$ acts on $X$ by left multiplication: the element $g \in G$ carries the subset $Y$ of $G$ to $\{g(y) | y \in Y\}$. Now $|X| = \left(\begin{array}{c} m \cdot q \\ q \end{array}\right)$, which, by a binomial argument is nonzero mod $p$. Hence, there is an orbit of $G$ on $X$ of size not divisible by $p$. So, every element $Y$ of this orbit in $X$ has a stabilizer, say $S$, in $G$ of order divisible by $q$. On the other hand, $S$ cannot be all of $G$, as left multiplication is transitive on $G$, and so left multiplication by $G$ does not leave invariant the subset $Y$ of size $q$. Hence the induction hypothesis applies to $S$, yielding that it contains a subgroup of order $q$; but then so does $G$.

**If $H$ is a subgroup of $G$ whose order is a power of $p$, then $H$ is a subgroup of a subgroup of $G$ of order $q$.**

Consider the collection $T$ of subgroups of $G$ of order $q$. Take $S$ a subgroup as in Part 2. Then $S \in T$, so $T$ is non-empty. The group $G$ acts on $T$ by conjugation.

Restrict this action to the subgroup $H$ and consider its orbits. By Lagrange's theorem, each nontrivial $H$-orbit has size a multiple of $p$. Suppose that $H$ fixes a member $M$ of $T$ in its conjugation action. Then $H$ is contained in the normalizer $N$ in $G$ of $M$. Since $M$ is a normal subgroup of $N$ and $M$ is a subgroup of $N$ of order $q$, by Lemma on subgroups of prime power order implies that $H$ is a subgroup of $M$.

Let us now take $S$ for $H$ in the previous argument. Then $S$ stabilizes $S$ in the conjugation action, and so the argument applies, giving that $S$ is a subgroup of $M$. But both are of order $q$, so they coincide. We have found that $S$ has only one fixed point. Since all other $S$-orbits have sizes a multiple of $p$, it follows that the size of $T$ is 1 modulo $p$.

Coming back to the arbitrary subgroup $H$ of $G$ of order a power of $p$, we see that it must fix a member of $T$ in its conjugation action because otherwise the size of $T$ would be a multiple of $p$, contradicting that is 1 modulo $p$. By the above, this shows that $H$ is contained in the subgroup $M$ of $G$ of order $q$.

**Any two subgroups of $G$ of order $q$ are conjugate.**

Take $S$ as in Part 2. Let $U$ be the $G$-orbit containing $S$ (in the collection $T$ of subgroups of $G$ of order $q$). Since $S$ is the only fixed member of $U$, the size of $U$ is 1 modulo $p$. Let $M$ be an arbitrary subgroup of $G$ of order $q$. If $M$ does not fix a member of $U$, then the size of $U$, being a union of nontrivial $M$-orbits, is a multiple of $p$, a contradiction. Hence $M$ fixes a member of $U$, which, by the above argument, must coincide with $M$. In particular, $M$ is in the same $G$-orbit as $S$.

□

**Example 8.6.14.** Consider the group $\mathrm{Sym}_5$. This group has order $120 = 2^3 \cdot 3 \cdot 5$.

The subgroup generated by the permutations $(1,2,3,4)$ and $(1,2)(3,4)$ is a Sylow 2-subgroup of order 8. It is isomorphic to a dihedral group of order 8.

All Sylow 2-subgroups of $\mathrm{Sym}_5$ are conjugate to this subgroup. There are exactly 15 Sylow 2-subgroups in $\mathrm{Sym}(5)$.

Each 3-cycle generates a Sylow 3-subgroup and each 5-cycle a Sylow 5-subgroup. The number of Sylow 3-subgroups equals 10, the number of Sylow 5-subgroups is equal to 6.

**Remark 8.6.15.** Of course there are commutative groups with elements that do not have order 2. Indeed, any cyclic group is commutative.

A subgroup of $G$ of order $q$ is called a *Sylow $p$-subgroup* of $G$.

Notice that if $S$ is a Sylow $p$-subgroup of $G$, then so is $g \cdot S \cdot g^{-1}$ for any $g$ in $G$. Sylow's theorem implies that all Sylow $p$-subgroups can be obtained in this way.

**Corollary 8.6.16.** *Let $G$ be a finite group and $p$ a prime number dividing $|G|$. The number of Sylow $p$-subgroups of $G$ is a divisor of $|G|$ and equal to $1$ modulo $p$.*

*Proof.* The size 1 modulo $p$ is immediate from the arguments in Parts 3 and 4 of the proof of Sylow's theorem.

The fact that the number of Sylow $p$-subgroups of $G$ divides the order of $G$ follows from the Identification of orbit with cosets and assertion 4 of Sylow's theorem.

□

**Example 8.6.17.** Let $G$ be a group of order 100. Then $G$ is not simple. (Here simple meens that it does not have any normal subgroup, except for the trivial normal subgoups being the subgroup containing only the identy element or the whole group.) This can be shown as follows.

Let $S$ be a Sylow-5 subgroup. Then $S$ has order 25. The number of Sylow-5 subgroups is a divisor of $\frac{100}{25}$ and equal to 1 modulo 5. This implies that $S$ is the only Sylow-5 subgroup of $G$. In particular, $S$ is a normal subgroup of $G$.

**Example 8.6.18.** The number of Sylow 2-subgroups of $\mathrm{Sym}_5$ is equal to 15, see Example 8.6.14, which divides 120 the order of $\mathrm{Sym}_5$, and is equal to 1 modulo 2.

Also the number of Sylow 3-subgroups and Sylow 5-subgroups, viz. 10 and 6, are divisors of 120 and are equal to 1 modulo 3 and 5, respectively. See Example 8.6.14.

**Remark 8.6.19.** In a commutative group the number of Sylow $p$-subgroups is one for every divisor $p$ of the order of the group.

The converse is not true: The dihedral group $D_4$ of order 8 but is not commutative and has exactly one Sylow 2-subgroup, viz., $D_4$ itself.

**Theorem 8.6.20** (Classification of finite commutative groups)**.** *Each finite commutative group is isomorphic to a direct product of cyclic groups of prime power order.*

*Proof.* Let $G$ be a commutative finite group.

**Suppose that $|G|$ is a power of the prime $p$ and that $C$ is a maximal cyclic subgroup of $G$. If $G$ is not cyclic, then there is a subgroup $N$ of $G$ of order $p$ with $C \cap N = \{1_G\}$.**

Let $c$ in $G$ be a generator of $C$. Write $r$ for the order of $c$. By Lagrange's theorem each element has order a power of $p$. So $s$ is of the form $p^k$ for some natural number $k$.

$C$ is a normal subgroup of $G$ (as $G$ is commutative), so there is a commutative quotient group $G/C$.

Take $d \in G \setminus C$ such that its image in $G/C$ is of maximal order, say $q$.

Now $d^q$ belongs to $C$ and therefore, $d^q = c^s$ for some natural number $s$. Since $r$ is the largest order occurring in $G$, we have $1_G = d^r = (d^q)^{\frac{r}{q}} = (c^s)^{\frac{r}{q}} = c^{\frac{s \cdot r}{q}}$ which implies that $r$ divides $\frac{s \cdot r}{q}$. Therefore, $q$ divides $s$. Since $q$ is a positive power of $p$, this implies that $p$ divides $s$. Consider now the element $x = d^{\frac{q}{p}} \cdot c^{\frac{-s}{p}}$. It satisfies $x^p = d^q \cdot c^{-s} = 1_G$, so it has order $p$. In view of the definition of $q$, the element $d^{\frac{q}{p}}$ does not belong to $C$, so neither does $x$. Hence the subgroup $N = \langle x \rangle_G$ is as required.

**Suppose that $|G|$ is a power of the prime $p$ and $C$ is a cyclic subgroup of $G$ of maximal order. Then there is a subgroup $D$ of $G$ such that $G = C \times D$.**

We prove the assertion by induction on $|G|$. If $|G| = 1$, there is nothing to show. (If $|G| = p$, the assertion follows from an earlier assertion, but we do not need this here.)

Let $N$ be a subgroup as in the previous assertion. That is, it has order $p$ and meets $C$ only in $1_G$.

In the quotient group $G/N$, the image of $C$ under the quotient morphism is again a cyclic subgroup of maximal order. But the size of $G/N$ is strictly less than $|G|$, so by the induction hypothesis, there is a subgroup $K$ of $G/N$ such that $G/N = C/N \times K$. Let $D$ be the full inverse image of $K$ in $G$. Then $C \cap D$ maps onto $C/N \cap K$, which is the identity according to a property of the direct product. Hence $C \cap D \subset C \cap N$. But, by construction of $N$, we have $C \cap N = \{1_G\}$. This establishes $C \cap D = \{1_G\}$.

Furthermore, the subgroup $C \cdot D$ of $G$ maps surjectively onto $G/N$ as its image contains both $C/N$ and $K$, and contains the kernel $N$ of the quotient map, so it must coincide with $G$. This shows that $G$ is indeed the direct product of $C$ and $D$.

**Suppose that $|G|$ is of prime power order. Then $G$ is a direct product of cyclic groups.**

By the previous assertion and induction on the size of $G$.

The theorem now follows from the combination of the fact that commutative groups are a direct product of groups of prime power, Properties of commutative groups, and the last assertion.

□

**Example 8.6.21** (Commutative groups of order 12)**.** A commutative group $G$ of order 12 is isomorphic to either $C_4 \times C_3$ or $C_2 \times C_2 \times C_3$. Observe that $C_{12}$ is isomorphic to the first of these.

We have gathered enough knowledge to determine all groups of order at most 11. We do this up to isomorphism: for each isomorphism class, we give one representative.

**Theorem 8.6.22** (Classification of groups of order at most 11)**.** *The table below contains, up to isomorphism, all groups of order at most* 11.

*Proof.*

**No two groups from the table are isomorphic.**

This is easily verified by use of the following remarks:

- a commutative group is not isomorphic with a non-commutative group;

| order | group | number |
|---|---|---|
| 1 | $\{e\}$ | 1 |
| 2 | $C_2$ | 1 |
| 3 | $C_3$ | 1 |
| 4 | $C_4$ , ${C_2}^2$ | 2 |
| 5 | $C_5$ | 1 |
| 6 | $C_6$, $D_6$, | 2 |
| 7 | $C_5$ | 1 |
| 8 | $C_8$ , $C_4 \times C_2$ , ${C_2}^3$ , $Q_8$ , $D_8$ | 5 |
| 9 | $C_9$ , ${C_3}^2$ | 2 |
| 10 | $C_{10}$ , $D_{10}$ | 2 |
| 11 | $C_{11}$ | 1 |

Table 8.1: Groups of order at most 11.

- two isomorphic groups have the same number of elements of a given order.

We now determine the isomorphism types of the groups of order $2,3,4,5,6,7,8,9,10,11$.

**If $|G|$ is a prime, then $G$ is cyclic.**

This follows from the Lemma on subgroups of prime power order.

This handles the cases where the order $G$ is equal to 2, 2, 5, 7, or 11.

**Suppose $|G| = 4$. If $G$ is not cyclic, then $G$ is isomorphic to ${C_2}^2$.**

By each element distinct from e has order 2. By Properties of commutative groups, we find $G$ to be commutative and to be isomorphic to a direct product of two cyclic groups of order 2.

**Suppose $|G| = 6$. If $G$ is not cyclic, then $G$ is isomorphic to $D_6$.**

Suppose that $G$ is not cyclic. By Lagrange's theorem, the elements of $G$ have order 2 or 3.

By the Sylow's theorem, the number of elements of order 2 equals 1 or 3. If there is only one such element, say $a$ then for every element $b$ we have $b \cdot a \cdot b^{-1} = a$ and hence $b \cdot a = a \cdot b$. But then we find that $(a \cdot b)^2 = a^2 \cdot b^2 = b^{-1}$, and $(a \cdot b)^3 = a^3 \cdot b^3 = a$ are not the identity element and hence the element $a \cdot b$ is an element of order 6. This contradicts our assumptions.

Hence we can assume that there are three elements of order 2 in $G$. Moreover, by Sylow's theorem, the group acts transitively on the set of these three elements. This permutation representation of $G$ provides us with a homomorphism from $G$ into $\mathrm{Sym}_3$. If this representation is an isomorphism, then $G$ is isomorphic with $\mathrm{Sym}_3$. Otherwise, the kernel is equal to $\langle a \rangle$. But then $\langle a \rangle$ is a normal subgroup of $G$, contradicting that $a$ has three conjugates.

**Suppose $G$ has order 8 and is not cyclic. Then $G$ is isomorphic to $C_4 \times C_2$ ,**

**${C_2}^3$ , $Q_8$ , or $D_4$.**

Each element of $G$ has order $1,2$, or $4$. If $G$ is commutative then it is a direct product of cyclic groups and hence isomorphic to $C_4 \times C_2$ or ${C_2}^3$.

So assume that $G$ is not commutative. This implies that $G$ contains an element $b$ of order $4$. Choose an element $a$ not commuting with $b$. Note that $G = \langle a, b \rangle$. As $\langle b \rangle$ has index $2$ in $G$, it is a normal subgroup of $G$. In particular, $a \cdot b \cdot a^{-1} \in \langle b \rangle$. As $a$ does not commute with $b$ and any conjugate of $b$ has order $4$, we find $a \cdot b \cdot a^{-1} = b^{-1}$. Now assume that the element $a$ can be chosen in such a way that its order is 2 Then consider the action of $G$ by left multiplication on the $4$ (left) cosets of the subgroup $\langle a \rangle$ of $G$. These cosets are $\langle a \rangle, b \cdot \langle a \rangle, b^2 \cdot \langle a \rangle, b^3 \cdot \langle a \rangle$. Numbering these cosets $1,2,3,4$, respectively, we find $L(b) = (1,2,3,4)$ and $L(a) = (2,4)$. Hence, the image of $G$ in $\mathrm{Sym}_4$ is isomorphic to the group generated by these two permutations. This latter group is isomorphic to $D_4$, the dihedral group of order $8$. As this order equals the order of $G$, we find the two groups to be isomorphic.

It remains the case that there is no element of order $2$ in $G$ that does not commute with $b$. In particular, any element not in $\langle b \rangle$ has order $4$. Pick an element $a$ of order $4$, not in $\langle b \rangle$. Then $a^2 = b^2$, the only element of order $2$ in $G$. Moreover, as we already noticed above, $a \cdot b \cdot a^{-1} = b^{-1}$. The map $f \colon G \to Q_8$ with $f(a) = i$ and $f(b) = j$ is now easily seen to be an isomorphism between $G$ and the quaternion group $Q_8$.

**If $G$ has order $9$ and is not cyclic, then $G \cong {C_3}^2$.**

Each element distinct from $e$ has order $3$. Let $a$ be such an element and consider the permutation representation $L \colon G \to \mathrm{Sym}(G/\langle a \rangle)$. Its kernel is contained in $\langle a \rangle$. On the other hand it cannot be trivial, for otherwise, the image of $G$ under $L$ would be a subgroup of $\mathrm{Sym}(G/\langle a \rangle)$ of order $9$, and so, by Lagrange's theorem, $9$ would divide the order of $\mathrm{Sym}(G/\langle a \rangle)$, which is $6$. Hence, the kernel of $L$ is $\langle a \rangle$. Consequently, $\langle a \rangle$ is a normal subgroup of $G$. In particular, the conjugacy class $C$ of $a$ is contained in $\left\{e, a, a^2\right\}$. Clearly, $e$ cannot be conjugate to $a$. Therefore, $C$ has at most $2$ elements. But, by Identification of orbit with cosets, the number of elements of $C$ is a divisor of $9$, so $C$ consist only of the element $a$. So, for each element $b \in G$, we have $b \cdot a \cdot b^{-1} = a$, that is, $a \cdot b = b \cdot a$. In other words, $a$ lies in the center of $G$. As the element $a$ was chosen arbitrarily, this implies that $G$ is commutative and hence isomorphic with $C_3 \times C_3$; see Classification of finite commutative groups.

**Let $G$ be a group of order $10$. If $G$ is not cyclic, then it is isomorphic with $D_5$**

Suppose that $G$ has order $10$ and is not cyclic. Then it contains an element $a$ of order $5$ and an element $b$ of order $2$. The group $G$ is not commutative, for otherwise, it would be generated by $a \cdot b$ and hence cyclic. So, $a$ and $b$ do not commute. The subgroup $\langle a \rangle$ has index $2$ in $G$ and is a normal subgroup. In particular, $b \cdot a \cdot b^{-1}$ belongs to $\langle a \rangle$. This means that $b \cdot a \cdot b^{-1} = a^k$ for some $k \in \{2,3,4\}$. But then

$$a = b \cdot (b \cdot a) \cdot b^{-1} \cdot b^{-1} = b \cdot a^k \cdot b^{-1} = \left(a^k\right)^k \tag{8.13}$$

from which we deduce that $k = 4$.

This implies that $G$ is indeed isomorphic to the dihedral group $D_5$.

□

**Example 8.6.23** (Groups of order $2\cdot p$ with $p$ prime)**.** A group $G$ of order $2\cdot p$, with $p$ prime, contains an element $a$ of order $p$ and an involution (element of order 2) $b$. The subgroup $\langle G\rangle_{[a]}$ is normal in $G$. If the element $b$ commutes with $a$, then $G$ is cyclic and hence isomorphic to $C_{2\cdot p}$. If $b$ does not commute with $a$, then $b\cdot a\cdot b^{-1} = a^k$ for some $k$. But then $a = b\cdot b\cdot a\cdot b^{-1}\cdot b^{-1} = b\cdot a^k\cdot b^{-1} = \left(a^k\right)^k = a^{k^2}$. But that means that $k^2 \equiv 1 \pmod p$ and hence $k \equiv -1 \pmod p$. In particular, $b\cdot a\cdot b^{-1} = a^{-1}$ and $G$ is isomorphic to $D_p$.

**Example 8.6.24** (Groups of order 12)**.** We have already met a cyclic group $C_{12}$, a direct product of cyclic groups $C_2\times C_6$, the dihedral group $D_6$, the direct product $C_2\times \mathrm{Sym}_3$ and the alternating group $\mathrm{Alt}_4$. Up to isomorphism, there is one more group of order 12. Let groupname1.generalized_quaternion_group (3) be the subgroup of $SL(2,\mathbb{C})$ generated by the following two matrices: $A = \begin{pmatrix} x & 0 \\ 0 & x^{-1}\end{pmatrix}$ and $B = \begin{pmatrix} 0 & -1 \\ 1 & 0\end{pmatrix}$, where $x$ equals $\frac{-1+i\cdot\sqrt{3}}{2}$. The element $A$ is of order 3 and the element $B$ is of order 4. Furthermore, $B\cdot A = A^2\cdot B$. Hence, every element in groupname1.generalized_quaternion_group (3) can be written as $A^k\cdot B^l$, where $k$ is in $\{0,1,2\}$ and $l$ in $\{0,1,2,3\}$. In particular, groupname1.generalized_quaternion_group (3) has order 12. The group groupname1.generalized_quaternion_group (3) is not commutative and contains elements of order 4. Hence, it is not isomorphic to one of the examples above. It is not very easy to prove that each group of order 12 is isomorphic to one of the examples mentioned here.

## 8.7 Exercises

### 8.7.1 Groups

**Exercise 8.7.1.** Determine in $GL(2,R)$

1. a matrix $A$ mapping the vector $(1,0)^T$ onto $(0,1)^T$,
2. a matrix $B$ mapping $(1,0)^T$ onto $(1,1)^T$, and
3. a matrix $C$ mapping $(0,1)^T$ onto $(1,1)^T$.

**Exercise 8.7.2.** Let $G$ be a group and suppose $g$ is in $G$. In analogy with the map $L_g$ (left multiplication by $g$), we define a map $R'_g = G\to G$ by: $R'_g(h) = h\cdot g$ (for $h$ in $G$ ).

1. Prove that for each $g$ in $G$ the map $R'_g$ is in $\mathrm{Sym}(G)$.
2. Does the map from $G$ to $\mathrm{Sym}(G)$, given by $g\mapsto R'_g$, define a morphism?

**Exercise 8.7.3.** Let $G$ be a group.

1. Show that, for each $g\in G$, the map $C'_g = G\to G, x\mapsto g^{-1}\cdot x\cdot g$ is in $\mathrm{Sym}_G$.
2. Is the map $C' = G\to \mathrm{Sym}(G), g\mapsto C'_g$ permutation representation?

3. Can you make a permutation representation with $M : G \to G, x \mapsto g \cdot x \cdot g$ ?

**Exercise 8.7.4.** Let $X = \{x \in \mathbb{R}^4 | x_1 + x_2 + x_3 + x_4 = 0\}$. Define a permutation representation of $\mathrm{Sym}_4$ on $X$ by setting, for $g \in \mathrm{Sym}_4$, $g(x) = \left(x_{g^{-1}(1), g^{-1}(2), g^{-1}(3), g^{-1}(4)}\right)$.

1. Prove that this indeed a permutation representation.
2. Show that each $g$ in $\mathrm{Sym}_4$ acts as a linear transformation on $X$ and even as an orthogonal transformation.
3. Show that $(1,2)$ acts as a reflection and $(1,2,3,4)$ as a product of a reflection and a rotation.

**Exercise 8.7.5.** Describe the left regular representation $L = G \to \mathrm{Sym}(H/G)$ in each of the following cases.

1. $G = \mathrm{Sym}_3$ and $H = \langle [(1,2)] \rangle_{\mathrm{Sym}_3}$.
2. $G = \mathrm{Sym}_4$ and $H$ is a subgroup of order 4. (There are two different subgroups of order 4 !)

### 8.7.2 Orbits

**Exercise 8.7.6.** Suppose $G$ is a subgroup of $\mathrm{Sym}(X)$ for some set $X$. Then being in an orbit of $G$ defines an equivalence relation. This does not hold when $G$ is a monoid and not a group as will be clear from the following. Let $X$ be the set $\mathbb{Z}$ and $M$ the monoid $[\mathbb{N}, +, 0]$. Define

$f = M \to \mathrm{Sym}(\mathbb{Z})$ by $f(n) = k \mapsto k + 3 \cdot n$.

1. Show that $f$ is a morphism of monoids.
2. Define the relation ˜ on $\mathbb{Z}$ by $x$   $y$ if and only if there exists an $n$ in $\mathbb{N}$ with $f(n)(x) = y$. Show that this relation is not symmetric.

**Exercise 8.7.7.** A square matrix $A$ is called orthogonal if the product $A \cdot A^T$ is the identity matrix. The group $\mathrm{O}(n, R)$ of all orthogonal $n$ by $n$ matrices acts on $\mathbb{R}^n$ by left multiplication. Show that an orbit consists of all vectors of $\mathbb{R}^n$ with a fixed length. So there infinitely many orbits.

**Exercise 8.7.8.** Matrices in $GL(2, R)$ transform lines through the origin in $\mathbb{R}^2$ into lines through the origin. Determine the stabilizer $H$ of the $x$-axis. Determine also the stabilizer in $H$ of the $y$-axis. What is the kernel of the action on the lines?

**Exercise 8.7.9.** Consider the permutation representation of $\mathrm{Sym}_3$ on the left cosets in $\langle [(1,2,3)] \rangle / \mathrm{Sym}_3$. What is the kernel and what is the image of this permutation representation?

**Exercise 8.7.10.** Let $G$ be a group.

1. Determine the conjugacy class of the unit element $e$.

2. In this part, $G$ = $\mathrm{Sym}_4$. Determine the conjugacy classes of each of the following elements: $(1,2,3),(1,2,3,4),(1,2)\cdot(3,4)$.

3. Show that each conjugacy class consists of a single element if $G$ is commutative. Prove also the converse: if each conjugacy class consists of exactly one element, then $G$ is commutative.

4. Prove that all elements from the same conjugacy class have the same order.

**Exercise 8.7.11.** The center, denoted by $\mathbf{Z}(G)$, of a group $G$ is the set $\mathbf{Z}(G) = \{x \in G | \forall g.\,(g \in G) \Rightarrow (x\cdot g = g\cdot x)\}$.

1. Show that $\mathbf{Z}(G)$ is a normal subgroup of $G$.

2. Determine the center of the group $\mathrm{Sym}_3$.

3. What is the center of a commutative group?

4. Determine the conjugacy class of an element from the center.

5. If $a$ is an element of $G$ of order 2 and $\langle[a]\rangle_G$ is a normal subgroup of $G$, then $a$ is an element of $\mathbf{Z}(G)$.

6. Show that scalar multiplication by a nonzero $a$ is contained in the center of $GL(2,R)$.

7. Prove that the center of $GL(2,R)$ consists of scalar matrices only.

**Exercise 8.7.12.** Let $G$ be the group $C_2 \times C_2$. Call its nontrivial elements $a,b,c$ and, as usual, let $e$ be the unit element.

1. Describe the left regular representation of $G$.

2. Describe the action by left multiplication on the set $X$ of subsets of $G$ consisting of two elements. Is the action transitive?

**Exercise 8.7.13.** Consider $\mathrm{Sym}_n$ and let $X$ be the set of all subsets of $\{1,...,n\}$. There is an obvious permutation representation $f = G \to \mathrm{Sym}(X)$ defined by $f(g)(\{a_1,a_2,...,a_n\}) = \{g(a_1),g(a_2),...,g(a_n)\}$, where $\{a_1,a_2,...,a_n\}$ is a subset of $X$. Determine the orbits of $\mathrm{Sym}_n$. Do the same for $\mathrm{Alt}_n$ acting on $X$. (Watch out for $n = 2$.)

**Exercise 8.7.14.** Define a map $\mathrm{Sym}_2 \to \mathrm{Sym}(\mathbb{R}^2) =$ by $f(g)((x_1,x_2)) = \left(x_{g^{-1}(1),g^{-1}(2)}\right)$ for $g$ in $\mathrm{Sym}_2$.

1. Show that $f$ is a permutation representation.

2. What are the orbits of $\mathrm{Sym}_2$ on $\mathbb{R}^2$?

3. Of which vectors in $\mathbb{R}^2$ is the stabilizer equal to $\mathrm{Sym}_2$?

4. What is the geometric significance of the action of $(1,2)$ ?

**Exercise 8.7.15.** Report on similar questions as in Exercise 26 for the action of $\mathrm{Sym}_3$ on $\mathbb{R}^3$ via $f = \mathrm{Sym}_3 \to \mathrm{Sym}(\mathbb{R}^3)$ with $f(g)((x_1,x_2,x_3)) = \left(x_{g^{-1}(1),g^{-1}(2),g^{-1}(3)}\right)$ for $g$ in $\mathrm{Sym}_3$. Determine also the vertices for which the stabilizer is equal to $\langle[(1,2)]\rangle_{\mathrm{Sym}_4}$. Which permutations act as a rotation on $\mathbb{R}^3$ (determine the axis and angle of rotation)? Which act as reflections?

**Exercise 8.7.16.** Consider $G = \mathbb{Z}\times\mathbb{Z}$. Define $f = G \to \mathrm{Sym}(\mathbb{C})$ by $f((m,n))(z) = z\cdot i^{m+n}$.

1. Show that $f$ is a permutation representation.
2. Determine the kernel of $f$.
3. Determine the orbits of $g$ on $\mathbb{C}$.
4. Same questions as in part 1, but with $i$ replaced by $\mathrm{e}^{\frac{2\cdot i}{5}}$.

**Exercise 8.7.17.** Let $V = \{x_i | i \in \{1,...,6\}\}$ in $\{x \in \mathbb{Z}/2\mathbb{Z}^6 | x_1 + x_2 + ... + x_6 = 0\}$. Define a permutation representation $f = \mathrm{Sym}_6 \to \mathrm{Sym}(V)$ by $g(x) = \left(x_{g^{-1}(1)}, x_{g^{-1}(2)}, ..., x_{g^{-1}(6)}\right)$.

1. Show that this is indeed a permutation representation.
2. Show that the stabilizer of $(1,1,0,0,0,0)$ is isomorphic to $\mathrm{Sym}_2 \times \mathrm{Sym}_4$.
3. Determine the orbit of $(1,1,0,0,0,0)$.
4. Generalise to the case where the group $\mathrm{Sym}_n$, with $n$ even, acts on the set $\{x \in \mathbb{Z}/2\mathbb{Z}^6 | x_1 + x_2 + ... + x_n = 0\}$. Deduce from this, by studying the orbits, that
   $2^{n-1} = 1 + \frac{n!}{2!\cdot(n-2)!} + \frac{n!}{4!\cdot(n-4)!} + \frac{n!}{(n-2)!\cdot 2!} + 1$

**Exercise 8.7.18.** Let $S^1 = \{z \in \mathbb{C} | |z| = 1\}$, the circle of radius 1 around 0 in $\mathbb{C}$.

1. Define a map $f = \mathbb{Z} \to \mathrm{Sym}(S^1)$ by $f(n)(z) = i^n\cdot z$ (in short, $n\cdot z = i^n\cdot z$). Show that $f$ is a permutation representation and determine its kernel.
2. Show that the vertices on $S^1$ of a square form an invariant set.
3. Describe the action of the subgroup $2\cdot\mathbb{Z}$ of $\mathbb{Z}$ on $S^1$. What are the invariant subsets of $S^1$ under the action of $2\cdot\mathbb{Z}$ ?

**Exercise 8.7.19.** Let $G = GL(3,\mathbb{R})$ and let $X$ be the set of all pairs $\{u,v\}$ such that $u$ and $v$ are independent vectors in $\mathbb{R}^3$.

1. Show that, for each $A \in GL(2,\mathbb{R})$, the map $X \to X$, $\{u,v\} \mapsto \{A(u), A(v)\}$ is a bijection.
2. Define a permutation representation of $G$ on $X$ as suggested by the previous part. Is it injective? Is it transitive?
3. Same questions as before for $X$ consisting of triples of independent vectors in $\mathbb{R}^3$.

**Exercise 8.7.20.** Suppose $G$ is a group. Let $g$ be an element of $G$. The centralizer $C_G(g)$ of $g$ is the set of elements in $G$ commuting with $g$, that is, $(d \in C_G(g)) \Leftrightarrow ((d \in G) \wedge ((d,g) = (g,d)))$.

1. Show that $C_G(g)$ is a subgroup of $G$ containing $\langle g \rangle$.
2. What is the centraliser of $g$ if $G$ is commutative?
3. When do we have $|C_G(g)| = 1$?
4. Compute the centraliser of $(1,2)$ in $\mathrm{Sym}_4$.
5. Prove that the number of elements in the conjugacy class of $g$ is equal to $\frac{|G|}{|C_G(g)|}$. Conclude that this number is a divisor of $|G|$.

## 8.7.3 Order

**Exercise 8.7.21.** Determine a basis for the automorphism group of the square. Determine the order of the automorphism group of a square. Describe also the action of this group on the two diagonals of the square.

**Exercise 8.7.22.** Prove that $\langle[(1,2),(1,2,3,4)]\rangle_{\mathrm{Sym}_4} = \langle[(2,3),(1,2,3,4)]\rangle_{\mathrm{Sym}_4} = \mathrm{Sym}_4$. What is the order of the subgroup $H = \langle[(2,4),(1,2,3,4)]\rangle_{\mathrm{Sym}_4}$ of $\mathrm{Sym}_4$ ? Provide an isomorphism from the group $H$ to the group $D_4$ of automorphisms of the square.

**Exercise 8.7.23.** Show that a group of order $n$ cannot act transitively on a set with more than $n$ elements.

**Exercise 8.7.24.** Consider a transitive permutation group $G$ on a set $X$. Show that, for each $x$ and $y$ from $X$, the stabilizers $G_x$ and $G_y$ are conjugate, that is, there is $g \in G$ with $g \cdot G_x \cdot g^{-1} = G_y$.

**Exercise 8.7.25.** Let $G$ be a permutation group on the set $X$. The group $G$ is called $t$ - transitive, where $t$ in $N$, if it is transitive on the ordered $t$ - tuples from $X$.

1. Prove that $G$ is $t$-transitive if and only if the stabilizer of each $s$-tuple of elements from $X$ (where $s < t$ ) is transitive on the remaining elements of $X$.
2. Show that

   $\mathrm{Sym}_n$ is $n$ - transitive and that $\mathrm{Alt}_n$ is $n-2$ - transitive on $\{1,...,n\}$.

**Exercise 8.7.26.** Suppose that $G$ is a 2 - transitive permutation group on $\{1,...,n\}$ with $n > 1$.

1. Show that $G = \mathrm{Sym}_n$ if $G$ contains a transposition.
2. Show that $G = \mathrm{Alt}_n$ or $G = \mathrm{Sym}_n$ if $G$ contains a 3-cycle.

**Exercise 8.7.27.** Let $n > 2$. The group $GL(n,R)$ can be viewed as a permutation group on the set $X$ of 1 - dimensional subspaces of $\mathbb{R}^n$. If $g \in G$ and $x \in X$, then $g(X) = \{g(v) | v \in X\}$.

1. What is the kernel of this permutation representation?
2. Show that if $n = 2$, the group $G$ acts 3 - transitively on $X$.
3. For $n > 2$, the group $G$ is 2 - transitive but not 3 - transitive. Prove this.

**Exercise 8.7.28.** Label the vertices as in the following figure and consider the game in which you are allowed to rotate each of the 4 small triangles. Prove that these moves generate the subgroup $\mathrm{Sym}_6$ of $\mathrm{Sym}_6$.

**Exercise 8.7.29.** Label the vertices of a 2 by 2 by 2 cube with the integers $1,2,3,4,5,6,7,8$ as shown in the figure. Consider the following game: each single move consists of turning a face of the cube over 90 degrees (clockwise or counter clockwise). How many different positions can be obtained by applying such moves?

### 8.7.4 Automorphisms

**Exercise 8.7.30.** Show that the group $D_n$ of symmetries of a regular $n$-gon contains $n$ rotations and $n$ reflections. Determine a basis for $D_n$. What is the order of $D_n$?

**Exercise 8.7.31.** Let $G$ be the automorphism group of the tetrahedron. Determine a basis for $G$ and use it to find the order of $G$. Same question for the cube. Describe also the action of the automorphisms on the 4 diagonals of the cube.

**Exercise 8.7.32.** Let $a$ be the positive real fourth root of 2 so that $a$ is a root of $X^4 - 2$. Determine all automorphisms of $\mathbb{Q} \cdot a$.

**Exercise 8.7.33.** Consider the quotient group $\mathbb{Q}/\mathbb{Z}$.

1. Show that each element of the group has finite order.
2. Establish that the group itself has infinite order.
3. What is the order of the element $\frac{28}{16} + \mathbb{Z}$ ?
4. What is the order of an arbitrary element $\frac{a}{b} + \mathbb{Z}$ ?

**Exercise 8.7.34.** The quotient group $\mathrm{Sym}_4/K$, where $K = \langle [(1,2)\,(3,4)\,,(1,3)\,(2,4)] \rangle_{\mathrm{Sym}_4}$ is Klein‘s Vierergroup, is isomorphic with a group of order 6. Which one? $\mathrm{Sym}_3$ or $C_6$ ?

**Exercise 8.7.35.** Let $\mathbb{C}^\times$ be the multiplicative group of the complex numbers distinct from 0.

1. Show that $H = \{x \in \mathbb{C} \mid |x| = 1\}$ is a subgroup of $\mathbb{C}^\times$.
2. Show that the map $f = R \to H, t \mapsto \exp(2 \cdot i \cdot t)$ is a surjective morphism.
3. Prove that $\mathbb{R}/\mathbb{Z}$ is isomorphic to $H$.

**Exercise 8.7.36.** Consider the automorphism group $G$ of a regular octahedron.

1. Show that the automorphism group acts transitively on the set of vertices.
2. Show that the stabilizer of each vertex has order 8. What is the connection with the automorphisms of a square? What is the order of $G$ ?
3. Describe the action of $G$ on the three diagonals of the octahedron. Is the morphism $G \to \mathrm{Sym}(D)$, where $D$ is the set of diagonals, surjective?
4. Is the action of $G$ on the centers of gravity of the 8 faces of the octahedron an injective permutation representation? Do you spot a connection with the cube?
5. Does $G$ act transitively on the set of all unordered pairs of vertices?

**Exercise 8.7.37.** In this exercise we determine the automorphisms of the field $\mathbb{Q}(a)$, where $a = i + \sqrt{2}$.

1. Show that $a^2 - 2 \cdot i \cdot a = 3$. Deduce from this that $i \in \mathbb{Q}(a)$.
2. Prove that $\sqrt{2}$ also belongs to $\mathbb{Q}(a)$.
3. Conclude that $\mathbb{Q}(a) = \mathbb{C}\left(i, \sqrt{2}\right)$.
4. Determine a polynomial $f \in \mathbb{Q}$ of degree 4 having $a$ as a root.
5. What are the zeros of $f$ in $\mathbb{C}$ ?
6. Determine all automorphisms of $\mathbb{Q}(a)$ ; describe such an automorphism by its image on $a$.
7. Construct the multiplication table of this group. Is it a cyclic group? Indicate the images of $i$ and $\sqrt{2}$ under each automorphism.

**Exercise 8.7.38.** Let $z = \mathrm{e}^{\frac{2 \cdot i}{5}}$.

1. Show that $z$ is a root of the polynomial $X^5 - 1$ in $\mathbb{Q}$. What are the roots of this polynomial in $\mathbb{C}$ ?
2. Determine a polynomial $f \in \mathbb{Q}$ of degree 4 having root $z$.
3. Determine the automorphism group of $\mathbb{Q}(z)$ and show that this group is cyclic.

## 8.7.5 Quotient groups

**Exercise 8.7.39.** Prove the following equivalence for a subgroup $N$ of $G$ : $g \cdot N = N \cdot g$ for all $g$ in $G$ iff $g^{-1} \cdot n \cdot g$ in $N$ for all $g$ in $G$ and $n$ in $N$.

**Exercise 8.7.40.** Let $f = G \to H$ be a morphism of groups. Show, by means of an example, that the image $f(G)$ need not be a normal subgroup of $H$.

**Exercise 8.7.41.** The subgroup $\langle[(1,2)(3,4),(1,3)(2,4)]\rangle$ of $\mathrm{Sym}_4$ is called Klein‘s Viergroup.

1. Establish that $K$ has order 4 and is a normal subgroup of $\mathrm{Sym}_4$ as well as $\mathrm{Alt}_4$.
2. Verify that $K$ is isomorphic to $C_2 \times C_2$.
3. Give a non-normal subgroup of $\mathrm{Sym}_3$ that is also isomorphic to $C_2 \times C_2$.

**Exercise 8.7.42.** Let $H$ be a subgroup of the group $G$.

1. Show that each normal subgroup $N$ of $G$ contained in $H$ is also contained in the kernel of the morphism $L = G \rightarrow \mathrm{Sym}(G/H)$.
2. Show that $H$ is a normal subgroup if $G$ has order 9.

**Exercise 8.7.43.** In which of the following cases is the group $N$ a normal subgroup of the group $G$?

1. $G = \mathrm{Sym}_4$ and $N = \langle[(2,3)]\rangle_{\mathrm{Sym}_4}$.
2. $G = \mathrm{Sym}_4$ and $N = \langle[(1,2,3,4)]\rangle_{\mathrm{Sym}_4}$.
3. $N$ is the subgroup of all rotations in the automorphism group $G$ of a regular 5 - gon.

**Exercise 8.7.44.** Determine all normal subgroups of $\mathrm{Sym}_3$.

**Exercise 8.7.45.** Put $G = GL(2,\mathbb{R})$.

1. Show that the diagonal matrices $D$ form a subgroup which is not a normal subgroup of $G$.
2. Prove that the diagonal matrices of the form $a \cdot I$ with nonzero $a$ do form a normal subgroup of $G$.
3. Is the set of upper triangular matrices a normal subgroup of $G$ ?

**Exercise 8.7.46.** Let $G$ be a group.

1. Prove: if $N$ and $M$ are normal subgroups of $G$, then so is $N \cdot M$.
2. Prove: if $N$ is a normal subgroup of $G$ and $H$ a subgroup of $G$, then $N \cdot H$ is a normal subgroup of $H$.
3. Show, by means of the groups $G = \mathrm{Sym}_4$, $H = \mathrm{Alt}_4$, and a suitable subgroup $H$ of $G$, that $N \cdot H$ need not be a normal subgroup of $N$.
4. Show, by means of an example, that the following assertion does not hold in general: If $N$ is a normal subgroup of $H$ and $H$ a normal subgroup of $G$, then $N$ is a normal subgroup of $G$.

5. Show: If $H$ is a subgroup of $G$ and $g \in G$, then $g^{-1} \cdot H \cdot g$ is also a subgroup of $G$.
6. If, moreover, $H$ is the only subgroup of $G$ of order $n$, then $H$ is a normal subgroup of $G$.

**Exercise 8.7.47.** Let $G$ be a finite group, generated by the set $B$ and suppose $H$ is a subgroup of $G$ generated by $A$. Show that $H$ is a normal subgroup of $G$ if and only if $b^{-1} \cdot a \cdot b$ in $H$ for all $b \in B$ and all $a \in A$.

**Exercise 8.7.48.** Suppose $f = G \to H$ is a morphism of groups.

1. Prove: if $N$ is a normal subgroup of $H$, then $f^{-1} \cdot N$ is a normal subgroup of $G$.
2. If $f$ is surjective and $N$ is normal in $G$, then $f \cdot N$ is normal in $H$. Show, by means of an example, that the surjectivity condition cannot be removed.

**Exercise 8.7.49.** If $G$ is a group and $H$ a subgroup of $G$ of index 2, then $H$ is normal in $G$. Prove this in each of the following two ways:

1. By comparing left cosets and right cosets of $H$ in $G$,
2. By use of the left regular representation $G \to \mathrm{Sym}(H/G)$.

Establish also that, for each $g, h \in G$ the intersection $\{g, h, g \cdot h\} \cap H$ is not empty.

**Exercise 8.7.50.** Let $G$ be a group and $X$ a subset of $G$. The normaliser $N_G(X)$ of $X$ in $G$ is the set of elements $g$ of $G$ with $g \cdot X \cdot g^{-1} = X$. Notice that $N_G(X)$ is a subgroup of $G$. Show that $\langle X \rangle_G$ is a normal subgroup of

$N_X(G)$.

**Exercise 8.7.51.** Determine all normal subgroups of $\mathrm{Sym}_4$.

**Exercise 8.7.52.** Prove in each of the following cases that $N$ is a normal subgroup of $G$, and that $H$ is isomorphic to $G/N$.

1. $G = \mathbb{C}^\times$, $N = \{z \in \mathbb{C} \mid |z| = 1\}$, and $H = \{z \in \mathbb{R} | z > 0\}$, with the operation multiplication.
2. $G = \mathbb{R}^\times$, $N = \{-1, 1\}$, and $H = \{z \in \mathbb{R} | z > 0\}$, with the operation multiplication.
3. $G = \mathbb{C}^\times$,

   $N = \{z \in \mathbb{C} \mid |z| = 1\}$ and $H = \{z \in \mathbb{R} | z > 0\}$.
4. $G = \mathbb{Z} \times \mathbb{Z}$, $N = m \cdot \mathbb{Z} \times n \cdot \mathbb{Z}$, and $H = C_m \times C_n$.
5. $G = Q_8$, the quaternion group, $N = \{1, -1\}$, and $H = C_2 \times C_2$.
6. $G$ is the set of all invertible $2 \times 2$ matrices with entries from $\mathbb{Z}/7\mathbb{Z}$; $N$ is the subgroup of those matrices having determinant in $\{1, -1\}$, and $H = C_3$.

## 8.7.6 Structure theorems

**Exercise 8.7.53.** Let $\mathbb{C}^\times$ be the multiplicative group of the complex numbers distinct from 0.

1. Show that $H = \{x \in \mathbb{C} \mid |x| = 1\}$ is a subgroup of $\mathbb{C}^\times$.
2. Show that the map $f = R \to H, t \mapsto \exp(2 \cdot i \cdot t)$ is a surjective morphism.
3. Prove that $\mathbb{R}/\mathbb{Z}$ is isomorphic to $H$.

**Exercise 8.7.54.** Use the table of groups of order at most 10 as given in Section 8. 6 when answering the following questions.

1. Which groups from the table are commutative?
2. Let $G$ be a group of order 8 generated by two elements $a$ and $b$ of order 2 with $(a \cdot b)^4 =$ e. With which group of order 8 from the table is $G$ isomorphic?
3. Which groups of order 8 are (isomorphic to) subgroups of $\mathrm{Sym}_4$ ?

**Exercise 8.7.55.** Determine all groups of order 15 up to isomorphism.

**Exercise 8.7.56.** Let $G$ be a group of order $2 \cdot p$ where $p$ is an odd prime number.

1. Show that $G$ contains a normal subgroup $H$ of order $p$.
2. Prove that $G$ contains an element, $g$ say, of order 2.
3. If $h \in H$ is not 1, and $g \cdot h = h \cdot g$, then $g \cdot h$ is an element of order $2 \cdot p$. Give a proof of this assertion and conclude that in this case $G$ is cyclic and hence isomorphic to $C_{2 \cdot p}$.
4. Let $h \in H$. Prove: If $g \cdot h \neq h \cdot g$, then $g \cdot h' \neq h' \cdot g$ for all $\grave{h} \in H$ with $h \neq 1_G$.
5. Show that for all $h \in H$, the element $g \cdot h \cdot g$ belongs to $H$ ; derive from this that $g \cdot h \cdot g \cdot h = h \cdot g \cdot h \cdot g$.
6. Let $f = g \cdot h \cdot g \cdot h$. Prove that $g \cdot f = f \cdot g$.
7. Verify that, for all $h \in H$ : If $g \cdot h \neq h \cdot g$, then $g \cdot h \cdot g = h^{-1}$.
8. Show that $G$ is isomorphic to $D_{2 \cdot . \mathrm{p}}$, the automorphism group of a regular $p$-gon, if $G$ is not cyclic.