Integer and Polynomial Arithmetic
First Steps towards Abstract Algebra

Arjeh M. Cohen
Hans Cuypers
Hans Sterk

October 8, 2007

Contents

Preface

Algebra Interactive grew out of algebra lectures given at the Eindhoven University of Technology, The Netherlands, over the past few years. It was delevoped as IDA: Interactive Document on Algebra. Its aim is to bring elementary algebra to life through modern means, and provide students with a sophisticated learning environment with emphasis on computational and algorithmic aspects. New technology enriches the material in that

- many examples allow for experimenting, dynamic illustrations occur throughout the text,
- buttons enable focusing on specific aspects (proofs, examples, excercises, illustrations)
- on-line multiple choice tests are offered,
- on-line calculators pertaining to the subjects covered are available,
- various ways of cross referencing are supported.

Chapter 1

Integer arithmetic

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.

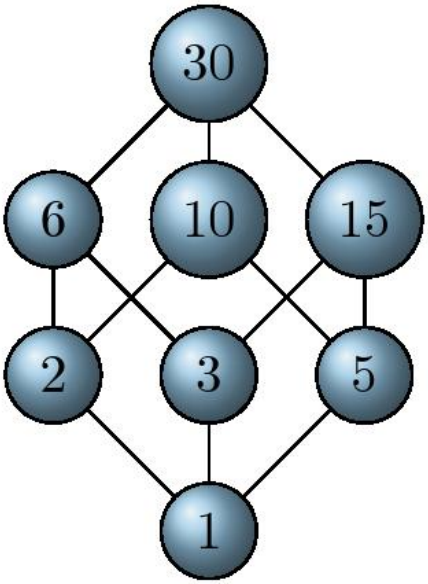

Figure 1.1: A schematic representation of all positive divisors of 30.

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 $\text{quot}(a,b)$.

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

$$n^m-1{=}(n-1){\cdot}(n^{m-1}+n^{m-2}+\cdots+1).$$

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 thata, b, and c are integers.*

(a) *If a divides b, and b divides c, then a divides c.*

(b) *If a divides b and c, then a divides $x{\cdot}b+y{\cdot}c$ for all integers x and y.*

(c) *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 we 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.

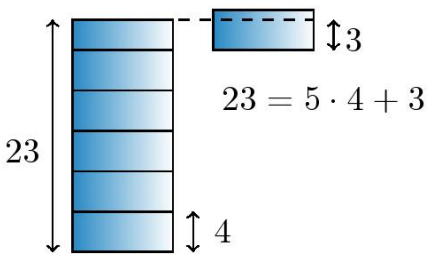

Figure 1.2: A division with remainder.

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 $r{=}r'$.
□

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 $\mathrm{rem}(a,b)$.

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 Division and Remainder Theorem 1.1.6 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)·|b|≤|a| do
    | x := q , q := x + 1
r := |a| − q·|b|
if a≥0∧b > 0
    | then
    |     | return
    |     |     | [q, r]
    | else
    |     | if a≥0∧b < 0
    |     |     | then
    |     |     |     | return
    |     |     |     |     | [−q, r]
    |     |     | else
    |     |     |     | if a < 0∧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.

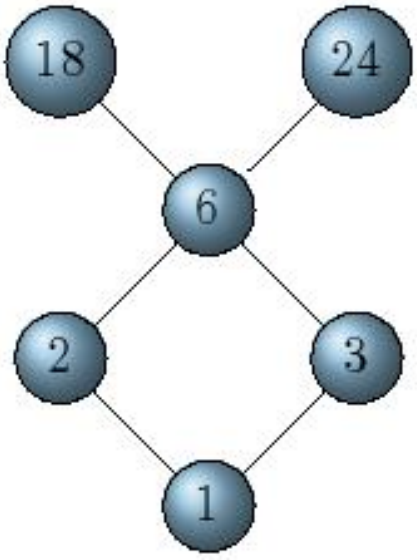

Figure 1.3: Positive common divisors of 18 and 24.

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*.

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 Lemma 1.1.5. 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.

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 we have a=q·b, then also a=$(-q)$·$(-b)$. 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 $\gcd(a, b)$=$\gcd(|a|, |b|)$.

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

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.

For any two nonzero integers a and b there exists a positive common multiple, namely $|a$·$b|$. 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.
We denote the least common multiple (lcm) of a and b by $\mathrm{lcm}(a, b)$.

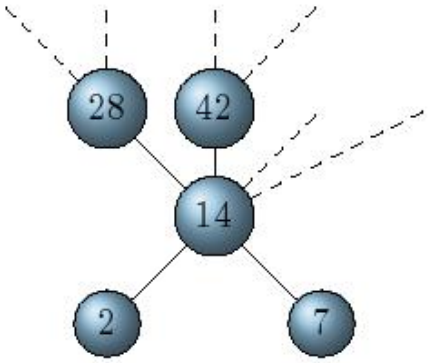

Figure 1.4: Some positive common multiples of 2 and 7.

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, $\mathrm{lcm}(13,5)$=65.

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

Theorem 1.1.19. *Let a and b be positive integers. Then*

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

Proof. Our strategy is to apply division with remainder to $a{\cdot}b$ and $\mathrm{lcm}(a,b)$, 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 $a{\cdot}b{=}q{\cdot}\mathrm{lcm}(a,b)+r$ as

$$r{=}a{\cdot}b-q{\cdot}\mathrm{lcm}(a,b).$$

Since both $a{\cdot}b$ and $\mathrm{lcm}(a,b)$ 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 $r < \text{lcm}(a, b)$ by the Division and Remainder Theorem 1.1.6, so r=0. Consequently, $a{\cdot}b{=}q{\cdot}\text{lcm}(a, b)$.
Next, we claim that q divides a and b. To see this, first let u be such that $\text{lcm}(a, b){=}u{\cdot}b$. 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 $a{\cdot}b{=}q{\cdot}\text{lcm}(a, b)$.

□

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, $\text{lcm}(a, b)$=120 and $a{\cdot}b$=360. We see that 3·120=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,

$$\text{lcm}(n + 1, n - 1){=}\frac{(n + 1){\cdot}(n - 1)}{2}$$

if n is odd, and

$$\text{lcm}(n + 1, n - 1){=}(n + 1){\cdot}(n - 1)$$

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 Lemma 1.1.5, where $q{=}\text{quot}(a, b)$ and $r{=}\text{rem}(a, b)$.
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 Remark 1.1.16.

Figure 1.5: 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 Lemma 1.1.5.

$$\gcd(a, b)=\gcd(b, a),$$

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

(for every integer k), and

$$\gcd(a, 0)=a.$$

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 $(a,b)^\top$ is replaced by $(b,a-q{\cdot}b)^\top$. We can also write this as the product of the matrix M and the vector $(a,b)^\top$, 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 1.2.1, 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 1.2.1 how to express the intermediate results in the input integers.

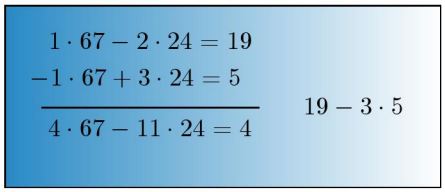

Figure 1.6: One step in the Extended Euclidean Algorithm applied to 67 and 24. Using the expressions for the intermediate results 19 and 5, the next occurring integer, 4, can also be expressed in the input values.

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**
  $a_1 := a\ ,\ b_1 := b$
  $u_1 := u\ ,\ v_1 := v\ ,\ x_1 := x\ ,\ y_1 := y$
  $a := b_1\ ,\ b := \mathrm{rem}(a_1,b_1)$
  $x := u_1\ ,\ y := v_1$
  $u := x_1 - \mathrm{quot}(a_1,b_1){\cdot}u_1\ ,\ v := y_1 - \mathrm{quot}(a_1,b_1){\cdot}v_1$
return
  $[a,x,y]$

Proof.
Correctness

Find the gcd of a and b using Euclid's algorithm 1.2.1. 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 1.2.1 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.
Each row of the following table represents a step in the 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

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 1.2.5 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.*

(a) $\gcd(a,b)=d$.

(b) *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.*

(c) *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 1.2.5 we have d=x·$a + y$·b for some integers x and y. If c is any common divisor of a and b, then it also divides x·$a + y$·b=d, see Lemma 1.1.5. 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·$a + y$·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·$a + y$·b implies that d divides e (see Lemma 1.1.5). So d≤e. Moreover, as a result of the Extended Euclidean Algorithm 1.2.5, d itself can also be written as an integral linear combination of a and b. So d≥e by the defining property of e. Hence e must be equal to d. This proves the equivalence.
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 1.2.9 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·$a + y$·b=1.*

Proof. Apply the previous theorem 1.2.9 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

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

A first application of the Characterization of the gcd 1.2.9 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, Corollary 1.2.10 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 $a|(x \cdot a \cdot c + y \cdot b \cdot c)$ which equals c, which proves the proposition.

□

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 1.2.5 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.
We first discuss a special case, the *homogeneous equation*, i.e., the case where c equals 0.

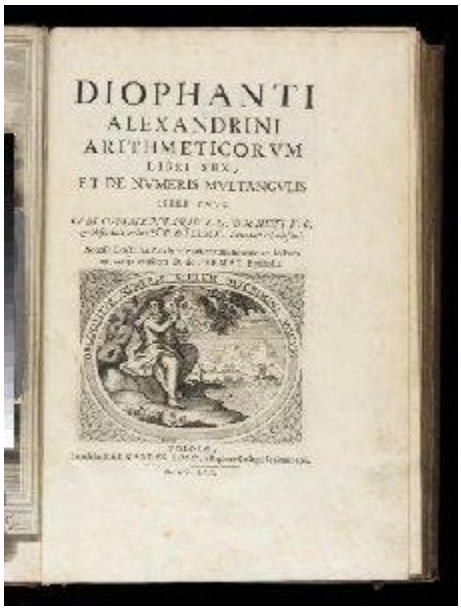

Figure 1.7: Diophantus' book on Arithmetic. Diophantus' work inspired Fermat to write in the margin of this book his famous last theorem: for $n > 2$ there are no nonzero integers x, y and z, such that $x^n + y^n = z^n$.

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 Proposition 1.2.13. 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 1.3.1 we conclude the following.

Theorem 1.3.2. *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 1.2.5 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 1.2.10.
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 1.3.1. Our equation then reads $x{\cdot}\frac{a}{d} + y{\cdot}\frac{b}{d}{=}0$, where $\gcd(\frac{a}{d}, \frac{b}{d}){=}1$. Lemma 1.3.1 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 Euclidean Algorithm 1.2.1 as in Example 1.2.3, we find

$$\gcd(24, 15){=}3.$$

By Theorem 1.3.2,

$$x{=}\frac{15{\cdot}n}{3}{=}5{\cdot}n$$

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)**.**

- *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
    $e := \text{Extgcd}(a, b)$
    $g := e[1]$
    $x_0 := e[2]$
    $y_0 := e[3]$
if $g|c$
    then
        return
            $\left\{\left(\frac{x_0 \cdot c - n \cdot b}{g}, \frac{y_0 \cdot c + n \cdot a}{g}\right) \mid n \in \mathbb{Z}\right\}$
    else
        return
            $\emptyset$

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 Theorem 1.3.2 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.

□

Note the structure of the solutions: $\left(\frac{x_0 \cdot c}{\gcd(a,b)}, \frac{y_0 \cdot c}{\gcd(a,b)}\right)$ 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.

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 1.3.4, we use the Extended Euclidean Algorithm 1.2.5 to compute the gcd of 24 and 15 and express it as a linear combination of these numbers. We find

$$\begin{aligned} \gcd(24, 15) &= \\ 3 &= \\ 2 \cdot 24 - 3 \cdot 15. & \end{aligned}$$

As 3 divides 63, there are solutions.
By the Linear Diophantine Equation Solving Algorithm 1.3.4 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.

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.

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.

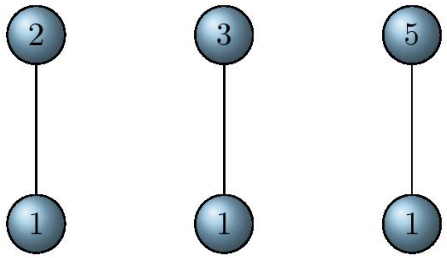

Figure 1.8: A prime has only 'trivial' divisors.

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 $p_1, \ldots, p_n$, 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 $i \in \{1, \ldots, n\}$, the integer m is clearly not divisible by p_i. Hence, the smallest divisor p of m greater than 1 is distinct from $p_1, \ldots, p_n$.
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 $p_1, \ldots, p_n$. This contradicts the assumption that $p_1, \ldots, p_n$ 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

Example 1.4.6. Although there are infinitely many prime numbers, see Euclid's Theorem 1.4.4, 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 $n \in \{2, \ldots, 101\}$, is n.

Example 1.4.7. Suppose that L is a finite list of primes, for example

$$L = [2, 3, 5, 7, 11, 13, 17]\,.$$

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).

Figure 1.9: 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.

If M is a list of integers and m is an integer, we shall write $M \sqcup m$ 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.*

Figure 1.10: Eratosthenes (about 276 BC-194 BC).

Sieve := **procedure**(n)
local variables
    $L := \{2, \ldots, n\}$
    $M := [\,]$
    m
while $L \neq [\,]$ **do**
    $m := L[1]$, $L := L \setminus m \cdot \{1, \ldots, n\}$, $M := M \sqcup m$
return
    M

Proof. 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.
By construction, the output list M consists of all numbers in $\{2, \ldots, n\}$ 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 1.4.9. 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	

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 1.4.9 equals the number of primes in the interval $\{1, \ldots, n\}$. 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 1.4.9 can also be used as a prime test. However, to avoid problems of big memory use as indicated in Remark 1.4.11, 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 $\{1, \ldots, n\}$. 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 $\{1, \ldots, n\}$. Then we have*

$$\lim_{n \uparrow \infty} \left(\frac{primes(n)}{\frac{n}{\ln(n)}} \right) = 1.$$

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.

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 1.4.13, the probability of having picked a prime is roughly $\frac{1}{\ln(10^{100})}$. Hence we expect to find a prime after at most $\ln(10^{100}) < 300$ picks.
Using a fast prime test (which do exist!), this can be easily done by a computer.

Application 1.4.15. 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

Figure 1.11: Jacques Hadamard (1865-1963).

asked the mathematics department to come up with a quick and very secure way of checking. The mathematicians' response is the following.

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=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.

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 rem(x, p)=rem(y, p). This means that the conclusion $x \neq y$ is justified if rem$(x, p) \neq$rem(y, p). So suppose that rem(x, p)=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 $\frac{b\cdot n}{2^n}$.
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**

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 Proposition 1.2.13 to find that p divides c.

Only if

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.

The Prime Characterization Theorem 1.4.16 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 Theorem 1.4.16 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.

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.

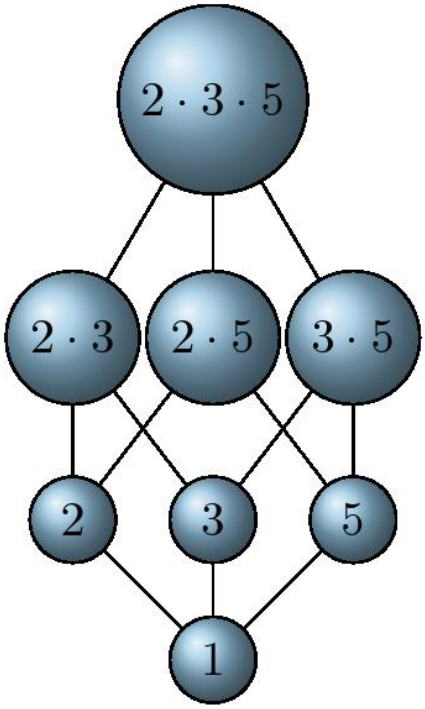

Figure 1.12: Building integers from primes.

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·$(\frac{a}{b})$, 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 Corollary 1.4.18 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.

□

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

$$a=\prod_{p\in\mathbb{P}} p^{\mathrm{ord}(p,a)}.$$

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.

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:

$$\begin{aligned}
&114381625757888867669235779976146612010218296721242362562561842935706935245733897830597123563958705058989075147599290026879543541 =\\
&\quad 349052951084765094914784961990389813341776463849338784399082057 7\cdot\\
&\quad 327691329932667095499619881908344614131776429679929425397982885 33.
\end{aligned}$$

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 1.4.9, 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.

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

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	$2^2 \cdot 5^1$

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 $(\frac{a}{b})^2{=}2$. 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 $(\frac{a}{b})^2{=}2$ 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 $\frac{a}{b}{=}\sqrt[n]{q}$, then $(\frac{a}{b})^n{=}q$. 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: $6{=}3{\cdot}2{=}(1 + \sqrt{-5}){\cdot}(1 - \sqrt{-5})$. The system R is an example of a ring, an algebraic structure with properties similar to those of $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$.

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*

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

and

$$\mathrm{lcm}(a, b){=}\prod_{p\in\mathbb{P}} p^{\max(\mathrm{ord}(p,a),\mathrm{ord}(p,b))}.$$

In particular we have $a{\cdot}b{=}\gcd(a, b){\cdot}\mathrm{lcm}(a, b)$.

Proof. We prove the first equality.
For each prime p we certainly have: $\min(\mathrm{ord}(p, a), \mathrm{ord}(p, b)){\leq}\mathrm{ord}(p, a)$ and $\min(\mathrm{ord}(p, a), \mathrm{ord}(p, b)){\leq}\mathrm{ord}(p, b)$. Hence the right-hand side of the equality

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

is a common divisor of a and b. In particular, by the Characterization of the gcd 1.2.9, we find that the right-hand side divides $\gcd(a, b)$.

On the other hand, if for some prime p we have $\text{ord}(p, \gcd(a, b))=m$, then p^m divides both a and b. Therefore, $m \leq \text{ord}(p, a)$ and $m \leq \text{ord}(p, b)$.
Hence the left-hand side of the equation

$$\gcd(a, b)=\prod_{p \in \mathbb{P}} p^{\min(\text{ord}(p,a),\text{ord}(p,b))}$$

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 Theorem 1.1.19 another proof of this statement is given.

□

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.

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 $n=\text{ord}(p, a)$. Then the binomial coefficient $\binom{a}{p^n}$ is not divisible by p.
Indeed, the binomial coefficient $\binom{a}{p^n}$ 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 $\text{ord}(p, b)$ equals $\text{ord}(p, a - b)$. 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

$$\operatorname{lcm}(a,b)=2\cdot3\cdot5\cdot23\cdot41.$$

Each of the factors in the above products is prime. You can check this with the prime test of Eratosthenes 1.4.12.

1.6 Number systems

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. 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 $a=[a_k,\ldots,a_0]_b$.*

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 $a_0, \ldots, a_k$ 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 $a{=}[a_k, \ldots, a_0, r]_b$.

Uniqueness of the representation.

Suppose that $a{=}[a_k, \ldots, a_0]_b$ and also $a{=}[c_l, \ldots, c_0]_b$ 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 $[c_l, \ldots, c_1]_b$ and $[a_k, \ldots, a_1]_b$ in the b-ary number system. So, by the induction hypothesis, $k{=}l$ and $a_i{=}c_i$ for all $i{\in}\{1, \ldots, k\}$. 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 1.6.3 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 $\text{rem}(a, b)$ at the beginning of the string and replace a by $\text{quot}(a, b)$. 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=quot(a, b)	rem(a, 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.

(a) 480, 175;

(b) 5621, 192;

(c) 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=$-b$.

Hint.
Use the definition of divisor.

Solution.
Since a divides b and b divides a, there exist integers p and q such that b=$p{\cdot}a$ and a=$q{\cdot}b$. Then a=$q{\cdot}p{\cdot}a$. But $a\neq 0$, so 1=$q{\cdot}p$. Since p and q are integers, we conclude that p and q are both equal to 1 or are both equal to -1. In the first case a=b and in the second case a=$-b$.

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

Hint.
Use the definition of divisor.

Solution.
There exist integers p and q such that $b{=}p{\cdot}a$ and $d{=}q{\cdot}c$. Hence $b{\cdot}d{=}p{\cdot}q{\cdot}a{\cdot}c$. Since the product $p{\cdot}q$ is an integer, the definition of divisor shows that $a{\cdot}c$ is a divisor of $b{\cdot}d$.

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

Hint.
In the induction step use $3^{4\cdot(n+1)}-1{=}(3^{4\cdot n}-1){\cdot}3^4+3^4-1$.

Solution.
For $n{=}1$, the statement reads 10 divides 80, which is clearly true. Suppose that the statement is true for n. To show that the statement also holds for $n+1$, we rewrite $3^{4\cdot(n+1)}-1$ as $(3^{4\cdot n}-1){\cdot}3^4+3^4-1$. By induction, the term $(3^{4\cdot n}-1){\cdot}3^4$ is divisible by 10. Since 3^4-1 is divisible by 10, application of Lemma 1.1.5 finishes the proof.

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.

Hint.
In the induction step use $a^{n+1}-b^{n+1}{=}a{\cdot}(a^n-b^n)+(a-b){\cdot}b^n$.

Solution.
For $n{=}1$ the statement is evidently true since $a-b$ divides $a-b$. Suppose that the statement holds for n. In order to show that the statement also holds for $n+1$, we rewrite $a^{n+1}-b^{n+1}$ as $a{\cdot}(a^n-b^n)+(a-b){\cdot}b^n$, a sum of two terms. The induction hypothesis implies that the first term is divisible by $a-b$. The second term is clearly divisible by $a-b$. Lemma 1.1.5 implies that the sum is then also divisible by $a-b$.

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

(a) 48, 15;

(b) 21, 19;

(c) 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.

Solution.
If d is a positive common divisor of c and b, then d divides c and b, so d divides a and b. But then d must be less than or equal to the greatest common divisor of a and b, which is 1. So d must be 1. In particular, the greatest common divisor of c and b is 1

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

(a)
$$\gcd(a,b)=\gcd(b,a),$$

(b)
$$\gcd(a,b)=\gcd(a,b-k\cdot a),$$

(c)
$$\gcd(a,0)=|a|.$$

Hint.
Use the Definition of the gcd 1.1.12.

Solution.
If both a and b are zero, then, in all three cases, the left-hand side and right-hand side are both equal to zero. Suppose therefore that they are not both zero. Then, by Definition 1.1.12, $\gcd(a,b)$ is the largest common divisor of a and b.
The set of divisors of a and b coincides with the set of divisors of b and a. Hence, their largest members are equal. This settles the first statement.
The set of divisors of a and b coincides with the set of divisors of a and $\gcd(a,b-k\cdot a)$. Hence, their largest members are equal. This settles the second statement.
Since every integer divides 0, the set of divisors of a and 0 coincides with the set of divisors of a. The largest member of this set is clearly $|a|$. This settles the third statement.

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

Hint.
For all integers a the number a^3-1 equals the product $(a-1)\cdot(a^2+a+1)$

Solution.
For all integers a the number a^3-1 is equal to the product $(a-1)\cdot(a^2+a+1)$. With a=10^n this reads

$$10^{3\cdot n}=(10^n-1)\cdot((10^n)^2+10^n+1)+1.$$

So the quotient of division of $10^{3 \cdot n}$ by $10^n - 1$ is equal to $(10^n)^2 + 10^n + 1$, and the remainder is equal to 1.

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.

Solution.
If a is a positive integer that divides n, then clearly a divides n^2. If a also divides n^2+3, then a must be a divisor of the difference of these two numbers $(n^2+3)-n^2$, i.e., 3. So a must be 1 or 3.
Here are examples of the two cases. For n=2, the integers 2 and 2^2+3 are relatively prime. For n=3, the gcd of 3 and $3^2 + 3$ is 3.

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?

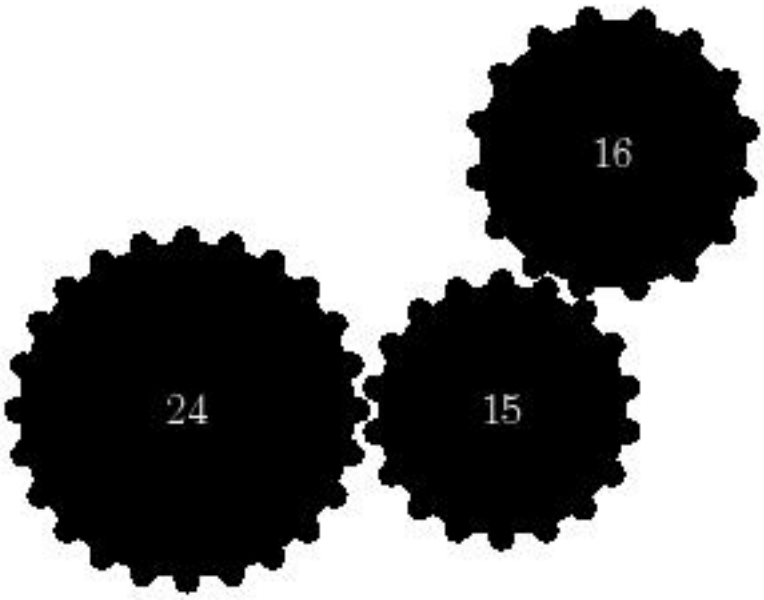

Figure 1.13: Three cogs

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?

Solution.
Let m be an odd integer. Write $m{=}2{\cdot}n+1$ and square both sides: $m^2{=}4{\cdot}n^2+4{\cdot}n+1$. Clearly this number is odd and leaves remainder 1 upon division by 4.
Since either n or $n+1$ is divisible by 2, the part $4{\cdot}n^2+4{\cdot}n{=}4{\cdot}n{\cdot}(n+1)$ is divisible by 8. So, m^2 leaves remainder 1 upon division by 8.
Taking $n{=}1$, we see that $m^2{=}9$ does not have remainder 1 upon division by 16, so the statement no longer holds if we replace 8 by 16.

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

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

Hint.
What can you say about the remainder of the division of c by $\mathrm{lcm}(a,b)$? Notice that this remainder is divisible by both a and b.

Solution.
Dividing c by $\mathrm{lcm}(a,b)$ yields

$$c{=}q{\cdot}\mathrm{lcm}(a,b)+r,$$

where r is a nonnegative integer less than $\mathrm{lcm}(a,b)$. Since a divides c and $\mathrm{lcm}(a,b)$, it also divides r. Similarly, b divides r. We conclude that r is a common multiple of a and b. But, as r is nonnegative and less than $\mathrm{lcm}(a,b)$, the definition of the lcm implies that r is 0. In turn, this implies that c is a multiple of $\mathrm{lcm}(a,b)$.

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:

(a) 480, 175;

(b) 5621, 192;

(c) 983675, 105120.

Hint.
Use the Extended Euclidean Algorithm 1.2.5.

Solution.
The following answers have been found using the Extended Euclidean Algorithm 1.2.5.

(a) The greatest common divisor of a=480 and b=175 equals 5. It can be written as $5{=}{-}4{\cdot}a + 11{\cdot}b$.

(b) The greatest common divisor of a=5621 and b=192 equals 1. It can be written as $1{=}29{\cdot}a + {-}849{\cdot}b$.

(c) The greatest common divisor of a=983675 and b=105120 equals 365. It can be written as $365{=}{-}137{\cdot}a + 1282{\cdot}b$.

Exercise 1.7.16. Show that, for all positive integers x and y, and nonnegative z, we have

$$\gcd(z{\cdot}x, z{\cdot}y){=}z{\cdot}\gcd(x, y).$$

Hint.
Use the Characterization of the gcd 1.2.9 to show that $z{\cdot}\gcd(x, y)$ is indeed the gcd of $z{\cdot}x$ and $z{\cdot}y$.

Solution.
If z=0, then the statement is evidently true. So we may assume that x, y, and z are all positive. Clearly, $z{\cdot}\gcd(x, y)$ is a positive divisor of both $z{\cdot}x$ and $z{\cdot}y$. Hence, it is less than or equal to $\gcd(z{\cdot}x, z{\cdot}y)$. Moreover, it can be written as a linear combination of $z{\cdot}x$ and $z{\cdot}y$ with integer coefficients. Indeed, since there exist integers a and b with $\gcd(x, y)$ equal to $a{\cdot}x + b{\cdot}y$, we can write $z{\cdot}\gcd(x, y)$ as $a{\cdot}z{\cdot}x + b{\cdot}z{\cdot}y$. The Characterization of the gcd 1.2.9 then implies that

$$\gcd(z{\cdot}x, z{\cdot}y){=}z{\cdot}\gcd(x, y).$$

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.

Hint.
Use the Characterization of Relatively Prime Numbers 1.2.10.

Solution.
By the Extended Euclidean Algorithm 1.2.5 there exist integers x and y such that $x{\cdot}a + y{\cdot}b{=}d$. Dividing both sides by d we obtain the relation $x{\cdot}a/d + y{\cdot}b/d{=}1$. By the Characterization of Relatively Prime Numbers 1.2.10 the integers 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

$$\gcd(a, b, c)=x\cdot a + y\cdot b + z\cdot c.$$

Hint.
Use the fact, stated in Exercise 1.7.18, that

$$\gcd(a, b, c)=\gcd(\gcd(a, b), c)$$

and, of course, the Extended Euclidean Algorithm 1.2.5.

Solution.
Using the Extended Euclidean Algorithm 1.2.5, we can find pairs of integers p, q and r, s, respectively, satisfying the equations

$$\gcd(a, b)=p\cdot a + q\cdot b$$

and

$$\gcd(\gcd(a, b), c)=r\cdot\gcd(a, b) + s\cdot c.$$

Using the fact that $\gcd(a, b, c)=\gcd(\gcd(a, b), c)$, we can express the gcd of a, b, and c as

$$\gcd(a, b, c)=x\cdot a + y\cdot b + z\cdot c,$$

where $x=r\cdot p$, $y=r\cdot q$ and $z=s$.

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.

Hint.
Notice that 18 and 25 are relatively prime. Use this to express a as an integral linear combination of $18\cdot a$ and $25\cdot a$.

Solution.
Notice that 18 and 25 are relatively prime. With the help of the Extended Euclidean Algorithm 1.2.5 we find that

$$1=7\cdot 18 - 5\cdot 25.$$

Multiplying both sides of the equation by a yields

$$a=7\cdot 18\cdot a - 5\cdot 25\cdot a.$$

As $18\cdot a$ and $25\cdot a$ are integers, the right-hand side is an integer, and so a 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:

(a) 5672, 234;

(b) 5311, 121;

(c) 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.

(a) $22{\cdot}x + 32{\cdot}y{=}12$;

(b) $12{\cdot}x + 25{\cdot}y{=}11$;

(c) $24{\cdot}x + 36{\cdot}y{=}18$.

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?

Solution.
The first question comes down to finding the nonnegative solutions of the Diophantine equation $5{\cdot}x + 20{\cdot}y{=}50$, where x denotes the number of 5 eurocent coins and y denotes the number of 20 eurocent coins used. Solutions of the equation are $x{=}10 + 4{\cdot}t$ and $y{=}{-}t$, where t is an integer. Nonnegative solutions are therefore $(10, 0)$, $(6, 1)$ and $(2, 2)$. There is exactly one solution where precisely 7 coins are used: $(6, 1)$.

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.

Hint.
First eliminate x from the two equations and solve the resulting equation in two variables.

Solution.
Subtract the first equation from the second. The integer solutions of the resulting equation $y + 2 \cdot z = 10$ are given by $y = 10 + 2 \cdot t$ and $z = -t$, where t is an arbitrary integer. Now substitute this result in the first equation: $x + 10 + 2 \cdot t - 3 \cdot t = 19$. So $x = 9 + t$. The integer solutions to the system of two equations are therefore given by $x = 9 + t$, $y = 10 + 2 \cdot t$, and $z = -t$, where t runs through the integers.
The positive solutions are obtained by taking $t = -1$, $t = -2$, $t = -3$ and $t = -4$, respectively. The corresponding solutions are (denoted as triples): $(8, 8, 1)$, $(7, 6, 2)$, $(6, 4, 3)$, and $(5, 2, 4)$.

1.7.4 Prime numbers

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

Hint.
Write $n^2 - 4$ as $(n + 2) \cdot (n - 2)$.

Solution.
Let p be a prime of the form $n^2 - 4$. Without loss of generality we assume n to be positive. Since $n^2 - 4$ equals $(n + 2) \cdot (n - 2)$, both $n - 2$ and $n + 2$ are divisors of p. But p has only trivial divisors, so $n - 2$, being the smaller of the two divisors found, must be equal to 1. Consequently, $n = 3$ and $p = 5$.

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

Hint.
Notice that p divides $7 \cdot q$.

Solution.
Notice that p divides $p \cdot q - 4 \cdot p$, which is equal to $7 \cdot q$. As p is prime, it divides 7 or q by the Prime Characterization Theorem 1.4.16. In the first case p equals 7 and, after division by 7, the original equation gives $q = 4 + q$, a contradiction. Therefore, $p | q$. But q is also a prime, so $p = q$. The original equation now reads $p^2 = 11 \cdot p$. Division by p gives $p = 11$. In other words, p and q are both equal to 11.

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

Hint.
Imitate the proof of Euclid's Theorem 1.4.4 stating that there are infinitely many primes.

Solution.
Suppose that there are only a finite number of primes of the form $4{\cdot}n+3$, say p_1, ... , p_r.
Consider the product

$$m{=}\prod_{i=1}^{r} p_i.$$

Clearly, m is odd. If $\mathrm{rem}(m,4){=}3$, put $M{=}m+4$; if $\mathrm{rem}(m,4){=}1$, put $M{=}m+2$. Then M is of the form $4{\cdot}n+3$. Since none of the primes p_j divides 2 or 4, there is no p_j dividing M. So, all prime divisors of M are of the form $4{\cdot}n+1$. But as the product of any number of elements of the form $4{\cdot}n+1$, in particular M, is again a number of the form $4{\cdot}n+1$, we encounter a contradiction to the fact that M is of the form $4{\cdot}n+3$.

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 $\gcd(p,a){=}p$.

Solution.
If p is prime, than 1 and p are the only positive divisors of p. Since $\gcd(p,a)$ is a positive divisor of p for every integer a, $\gcd(p,a)$ itself can only be 1 or p.
To prove the converse, we show that if p is not prime, then there exists an integer a such that $\gcd(p,a)$ is not equal to 1 or p. Suppose therefore that p is not a prime, so that it can be written as a product $a{\cdot}b$, where a and b are both greater than 1 and less than p. Then $\gcd(p,a){=}a$, Since $a{\neq}1$ and $a{\neq}p$, the proof is finished.

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 $\gcd(p,a/p^n){=}1$.

Solution.
Let n be the maximal element of the set of positive integers k such that p^k divides a. (This set is nonempty, since 1 belongs to it, and finite, since only finitely many powers of p are less than a. So this set has indeed a maximal element.) Since n belongs to the set, the quotient a/p^n is an integer. The maximality of n implies that p^{n+1} does not divide a. Now if p were a divisor of the quotient a/p^n, then it immediately follows that p^{n+1} divides a, which is impossible by what we just noted. Since 1 and p are the only positive divisors of p, the integer 1 is the only possibility left for the gcd of p and a/p^n.

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.

Solution.
$111=3\cdot 37$;
$143=11\cdot 13$;
$724=2^2\cdot 181$;
$1011=3\cdot 337$.

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

Solution.
Suppose that a and b are two integers with $b\neq 0$ and $\sqrt[3]{17}=a/b$. Without loss of generality we may assume a and b to be relatively prime (for the fraction is equal to the fraction with numerator a/g and denominator b/g, where $g=\gcd(a,b)$). Taking the third power of both sides and rewriting the above equation, we find $a^3=17\cdot b^3$. So 17 divides a^3. As 17 is a prime, the Prime Characterization Theorem 1.4.16 gives that 17 divides a. But then 17^3 divides $17\cdot b^3$ and hence 17 must divide b as well. This contradicts the condition that a and b are relatively prime. Hence 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.

Hint.
Write $3\cdot p+1=n^2$, use the identity $n^2-1=(n+1)\cdot(n-1)$ and apply the Unique Factorization Theorem 1.5.1.

Solution.
Suppose that $3\cdot p+1$ is equal to a square, say n^2. Then $3\cdot p+1=n^2$, which we rewrite as $3\cdot p=(n+1)\cdot(n-1)$. By the Unique Factorization Theorem 1.5.1 we must be in one of the following three cases: $3\cdot p=n+1$ and $1=n-1$, or $3=n+1$ and $p=n-1$, or $p=n+1$ and $3=n-1$. It is easy to see that the last case is the only possibility. In that case $n=4$ and $p=5$.

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.

Solution.
Suppose that a octaves higher corresponds to b fifths. Then the frequency obtained is both 2^a and $(3/2)^b$ times the original frequency. This leads to the equation $2^a{=}(3/2)^b$. Ridding ourselves of the denominator, we find $3^b{=}2^{a+b}$. By the Unique Factorization Theorem 1.5.1 this forces a and b both to be equal to 0. In other words, the only frequency occurring in both series is the original one.

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 $a{\cdot}b$.

Solution.
By the Unique Factorization Theorem 1.5.1 there are primes p_i such that $n{=}p_1{\cdot}\ldots{\cdot}p_k$. As a and b divide n, each of them is a product of some of these p_i's. Since a and b are coprime, no prime occurring in the factorization of a occurs in the factorization of b and vice versa. So we may assume, after reshuffling the indices, that $a{=}p_1{\cdot}\ldots{\cdot}p_l$ and $b{=}p_{l+1}{\cdot}\ldots{\cdot}p_m$ for certain numbers l and m. But then $a{\cdot}b{=}p_1{\cdot}\ldots{\cdot}p_m$, so $a{\cdot}b$ is a divisor of n, as required.

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

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

Exercise 1.7.43. Determine $\gcd(2^4{\cdot}3^2{\cdot}5{\cdot}7^6{\cdot}11, 2^2{\cdot}3^2{\cdot}5^3{\cdot}11)$.

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

Hint.
Use the factorization into primes.

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

(a) $2^3{\cdot}5^7{\cdot}11$ and $2^2{\cdot}3^4{\cdot}5^2{\cdot}11^4$;

(b) $2^1{\cdot}3^3{\cdot}5^2$ and $2^2{\cdot}3^4{\cdot}5{\cdot}11$;

(c) $3^2{\cdot}4^5{\cdot}7^2$ and $2^3{\cdot}3^2{\cdot}6^5{\cdot}7^2$.

Hint.
Use the factorization into primes.

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

Solution.
If a or b is 0, the statement is easy. So we restrict ourselves to the case where a and b are nonzero. For every prime p and every nonzero integer c we have $\mathrm{ord}(p, c^2)=2\cdot\mathrm{ord}(p, c)$. But that implies that

$$\begin{aligned}
\gcd(a^2, b^2) &= \\
\prod_{p\in\mathbb{P}} p^{\min(\mathrm{ord}(p,a^2),\mathrm{ord}(p,b^2))} &= \\
\prod_{p\in\mathbb{P}} p^{\min(2\cdot\mathrm{ord}(p,a),2\cdot\mathrm{ord}(p,b))} &= \\
\prod_{p\in\mathbb{P}} p^{2\cdot\min(\mathrm{ord}(p,a),\mathrm{ord}(p,b))} &= \\
(\prod_{p\in\mathbb{P}} p^{\min(\mathrm{ord}(p,a),\mathrm{ord}(p,b))})^2 &= \\
(\gcd(a, b))^2.
\end{aligned}$$

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?

Hint.
Try b=3.

Solution.
Suppose that we use the four weights 1, b, b^2, and b^3. To be able to identify the weight on a balance, we need to be able to write it as a sum

$$a_0 + a_1\cdot b + a_2\cdot b^2 + a_3\cdot b^3$$

with each $a_i \in \{-1, 0, 1\}$.
The only way to achieve the weight 2 will be to write it as $b - 1$, which implies $b=3$. Using that $2 \cdot 3^i = 3^{i+1} - 3^i$, we can indeed identify each weight less than or equal to $1 + 3 + 3^2 + 3^3 = 40$.

Exercise 1.7.51. The decimal representation of an integer n is $[\text{abcabc}]_{10}$, where a, b and c are elements from $\{0, \ldots, 9\}$.
Prove that 7, 11, and 13 are divisors of n.

Hint.
$[\text{abcabc}]_{10} = [\text{abc}]_{10} \cdot 1001$.

Solution.
Since n equals $[\text{abcabc}]_{10} = [\text{abc}]_{10} \cdot 1001$, it is divisible by all divisors of $1001 = 7 \cdot 11 \cdot 13$.

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.

1.8 Summary

This chapter is about multiplication and division of integers. We consider the following topics.

- Division with remainder: determine the quotient and remainder.
- Euclid's algorithm for finding the gcd.
- The extended Euclidean algorithm for expressing the gcd of a and b in the form $x \cdot a + y \cdot b$.
- Prime numbers: the building blocks of integers. We discuss:
  - There are infinitely many prime numbers.
  - Using Eratosthenes' sieve you can determine all the primes less than a given number.
  - Prime factorization: Any number is the product of primes; this factorization is unique (up to the order).

Chapter 2

Modular integer arithmetic

2.1 Arithmetic modulo an integer

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.

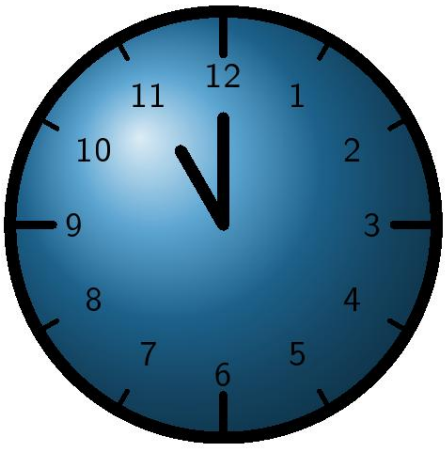

Figure 2.1: Clock arithmetic

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:

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

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.

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 $a \equiv c \pmod{n}$.

□

Here the set $k + n\cdot\mathbb{Z}$ consists of all integers of the form $k + n\cdot m$ where m is an integer.
The class containing the integer k (i.e., $k + n\cdot\mathbb{Z}$) will also be denoted by $k \pmod n$. The integer k is a representative of this equivalence class. If no confusion arises, we will also denote the class $k \pmod n$ by k itself.

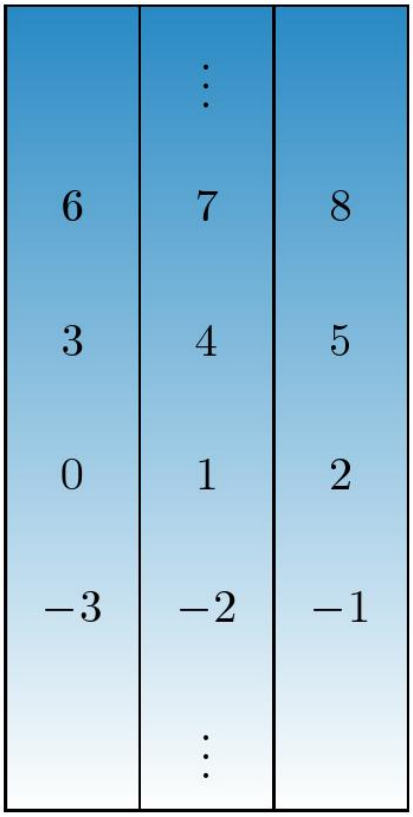

Figure 2.2: Congruence modulo 3 splits the integers in three disjoint subsets. These subsets are represented by columns. Integers in the same subset differ by a multiple of 3.

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 Arithmetic **??**, 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.

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.

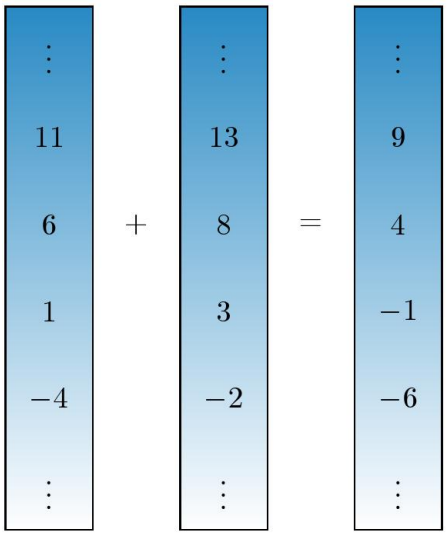

Figure 2.3: Addition of congruence classes is defined in terms of representatives. For instance, to add the two congruence classes modulo 5 above take any representatives in each of these classes, say 6 in the first and 3 in the second. Then their sum, 9, is a representative of the sum of the two classes.

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:* $a \pmod n + b \pmod n = a + b \pmod n$.
- *Multiplication:* $a \pmod n \cdot b \pmod n = a \cdot b \pmod n$.

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

$$\begin{aligned} (x + y) - (x' + y') &= \\ x - x' + y - y' &= \\ n \cdot a + n \cdot b &= \\ n \cdot (a + b). & \end{aligned}$$

Hence $x + y \equiv x' + y' \pmod n$.

Multiplication

By the above we find

$$\begin{aligned} x \cdot y - x' \cdot y' &= \\ x \cdot (y - y') + (x - x') \cdot y' &= \\ n \cdot b \cdot x + n \cdot a \cdot y' &= \\ n \cdot (b \cdot x + a \cdot y'). & \end{aligned}$$

Hence $x \cdot y \equiv x' \cdot y' \pmod n$.

□

Example 2.1.7 (Tables for $\mathbb{Z}/17\mathbb{Z}$)**.** 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

Below is the multiplication 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	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

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.

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

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

Proof. The laws hold for integers. For instance, in the case of commutativity, we have $a + b = b + a$. Now apply the Addition and Multiplication Theorem 2.1.6 to both sides. The commutativity for $\mathbb{Z}/n\mathbb{Z}$ follows. The proofs of the other equalities are similar. □

A *neutral* element for the addition is $0 \pmod n$. Indeed, $a \pmod n + 0 = a \pmod n$ and $0 + a \pmod n = a \pmod n$. The *opposite* of $a \pmod n \in \mathbb{Z}/n\mathbb{Z}$ is $-a \pmod n$, the unique element b such that $a \pmod n + b \pmod n = 0$.
A *neutral* element for the multiplication is $1 \pmod n$, as $a \pmod n \cdot 1 = a \pmod n$ and $1 \cdot a \pmod n = a \pmod n$.
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.

Example 2.1.9. Using the common rules of addition and multiplication we find the following applications of modular arithmetic.

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, $(0 \pmod 4)^2 = 0 \pmod 4$, $(1 \pmod 4)^2 = 1 \pmod 4$, $(2 \pmod 4)^2 = 0 \pmod 4$ and $(3 \pmod 4)^2 = 1 \pmod 4$. Now if m and n are integral, then $m^2 + n^2 \pmod 4 = m^2 \pmod 4 + n^2 \pmod 4$, and, by the above, this sum can only take the values $0 \pmod 4$, $1 \pmod 4$, or $2 \pmod 4$. 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.

The nine test

Suppose that $a = [a_k, \ldots, a_0]_{10}$ is the usual digital representation 1.6.1 of a. The well-known nine test

$$9|a \Leftrightarrow 9|(a_k + \ldots + a_0)$$

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

$$[a_k, \ldots, a_0]_{10} = a_k \cdot 10^k + \ldots + a_0 \cdot 10^0$$

reduction modulo 9 implies that $a \equiv a_k + \ldots + a_0 \pmod 9$. Thus $9|a$ if and only if $9|(a_k + \ldots + a_0)$.

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!

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 $19|(10^9 + 1)$.

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

Definition 2.1.10. An element $a \pmod n \in \mathbb{Z}/n\mathbb{Z}$ is called *invertible* if there is an element $b \pmod n$, called *inverse* of $b \pmod n$, such that $a \pmod n \cdot b \pmod n = 1$.

If $a \pmod n$ is invertible, its inverse (which is unique, as follows from Remark 2.1.12) will be denoted by $(a \pmod n)^{-1}$.

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 $\mathbb{Z}/n\mathbb{Z}$.

An integer a will be called *invertible modulo* n if its class $a \pmod n$ is invertible.

Example 2.1.11. In $\mathbb{Z}/18\mathbb{Z}$ the element $5 \pmod{18}$ is invertible. Indeed, since $2 \cdot 18 - 7 \cdot 5 = 1$, the inverse of $5 \pmod{18}$ is $-7 \pmod{18}$. The element $6 \pmod{18}$ is not invertible, since any multiple of 6 is either congruent to 0, 6, or 12 modulo 18.

Remark 2.1.12. Multiplicative inverses are unique, i.e., every invertible element has exactly one inverse. For, if

$$\begin{aligned} (a \pmod n) \cdot (b \pmod n) &= \\ (a \pmod n) \cdot (c \pmod n) &= \\ 1, & \end{aligned}$$

then

$$\begin{aligned} b \pmod n &= \\ (b \pmod n) \cdot (a \pmod n) \cdot (c \pmod n) &= \\ (a \pmod n) \cdot (b \pmod n) \cdot (c \pmod n) &= \\ c \pmod n. & \end{aligned}$$

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

(a) *The class* $a \pmod n$ *in* $\mathbb{Z}/n\mathbb{Z}$ *has a multiplicative inverse if and only if* $\gcd(a, n)=1$*.*

(b) *If* a *and* n *are relatively prime, then the inverse of* $a \pmod n$ *is the class* $\text{Extgcd}(a, n)_2 \pmod n$*.*

(c) *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 1.2.5, 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 $(a \pmod n) \cdot (x \pmod n) + 0 = 1$. In particular, $x \pmod n$ is the inverse of $a \pmod n$.
Notice that x indeed coincides with $\text{Extgcd}(a, n)_2$ modulo n, which proves the second statement.

Only if

If $a \pmod n$ has an inverse $b \pmod n$ 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 1.2.9, we find $\gcd(a, n)=1$.

□

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 $\mathbb{C}$.

Example 2.1.14. The invertible elements in $\mathbb{Z}/2^n\mathbb{Z}$ are the classes $x \pmod{2^n}$ for which x is an odd integer.
Indeed, the gcd of x and 2^n equals 1 if and only if x is odd.

Suppose that n and a are integers with $n > 1$ and $\gcd(a, n)=1$. The Characterization of Modular Invertibility 2.1.13 not only gives the existence of the inverse of $a \pmod n$ in $\mathbb{Z}/n\mathbb{Z}$, but also a way to compute this inverse.

Algorithm 2.1.15 (Modular Inverse)**.**

- *Input: integers $n > 1$ and a*
- *Output: the inverse of the class $a \pmod n$ of a in $\mathbb{Z}/n\mathbb{Z}$ if it exists, and 0 otherwise*

Inverse := **procedure**(a, n)
local variables
  $E := \text{Extgcd}(a, n)$
if E_1=1
  then
    return
      $E_2 \pmod n$
  else
    return
      0

Proof.
Termination

By the absence of loops this is obvious.

Correctness

Obvious by part (b) of the Characterization of Modular Invertibility 2.1.13. □

Example 2.1.16. Consider a=24 and n=35. Then a and n are relative prime. So $a \pmod n$ has an inverse. To find the inverse of $a \pmod n$, we apply the Extended Euclidean Algorithm. This gives the following expression of 1 as a linear combination of a and n.

$$1=35\cdot 11 - 24\cdot 16.$$

We deduce that the inverse of $a \pmod n$ equals $-16 \pmod n$.

Besides invertible elements in $\mathbb{Z}/n\mathbb{Z}$, which can be viewed as divisors of 1, see 2.1.10, one can also consider the divisors of 0.

Definition 2.1.17. An element $a \pmod n \in \mathbb{Z}/n\mathbb{Z}$ not equal to 0 is called a *zero divisor* if there is a nonzero element $b \pmod n$ such that $a \pmod n \cdot b \pmod n = 0$.

Example 2.1.18. 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 (mod 24), 4 (mod 24), 6 (mod 24), 8 (mod 24), 9 (mod 24), 10 (mod 24), 12 (mod 24), 14 (mod 24), 15 (mod 24), 16 (mod 24), 18 (mod 24), 20 (mod 24), 21 (mod 24), and 22 (mod 24).

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

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

(a) *The class $a \pmod n$ in $\mathbb{Z}/n\mathbb{Z}$ is a zero divisor if and only if $\gcd(a, n) > 1$ and $a \pmod n$ is nonzero.*

(b) *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 $b \pmod n$ of b is nonzero, but $a{\cdot}b$ is a multiple of n and so $a{\cdot}b \pmod n{=}0$. This translates to $a \pmod n{\cdot}b \pmod n{=}0$ in $\mathbb{Z}/n\mathbb{Z}$. In particular, $a \pmod n$ is a zero divisor.

Only if

If $a \pmod n$ is a zero divisor, then it is nonzero and there is a nonzero element $b \pmod n$ in $\mathbb{Z}/n\mathbb{Z}$ with $a \pmod n{\cdot}b \pmod n{=}0$. So, for the representative b_0 of $b \pmod n$ 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 Theorem 1.1.19, relating the gcd with the lcm, implies that $\gcd(a, n) > 1$.

□

Since an element $a \pmod n$ of $\mathbb{Z}/n\mathbb{Z}$ is either 0, a zero divisor, or invertible, the Modular Inverse Algorithm 2.1.15 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.

Example 2.1.20. Below you find the multiplication table of $(\mathbb{Z}/17\mathbb{Z})\backslash\{0\}$. 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.

*	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

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 2.1.21. *Let a, b, and n be integers with $n > 1$.*

(a) *If $a \pmod n$ and $b \pmod n$ 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 $(a \pmod n) \cdot (b \pmod n)$ is given by $(b \pmod n)^{-1} \cdot (a \pmod n)^{-1}$.*

(b) *If $a \pmod n$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$ and $b \pmod n$ an arbitrary element, then $a \pmod n \cdot b \pmod n$ is either 0 or a zero divisor.*

Proof. As

$$\begin{aligned}((a \pmod n)\cdot(b \pmod n))\cdot((b \pmod n)^{-1}\cdot(a \pmod n)^{-1}) &= \\ (a \pmod n)\cdot(a \pmod n)^{-1} &= \\ 1,&\end{aligned}$$

the inverse of $a \pmod n \cdot b \pmod n$ is $(b \pmod n)^{-1}\cdot(a \pmod n)^{-1}$. This establishes the second assertion. The first assertion is a direct consequence. If $a \pmod n$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$, then there is a nonzero element $c \pmod n$ with $a \pmod n \cdot c \pmod n$ equal to 0. But then $a \pmod n \cdot b \pmod n \cdot c \pmod n$ is also equal to 0. So $a \pmod n \cdot b \pmod n$ is 0 or a zero divisor.

□

Example 2.1.22. 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

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 $5^2 \pmod n = 1 \pmod n$. So indeed, the set of invertible elements is closed under multiplication.

2.2 Linear congruences

In addition to the linear equation $a\cdot x = b$ with integer coefficients in the single unknown x, we study the related equation $a\cdot x \equiv b \pmod n$ in the unknown x, which is called a *linear congruence*. It is closely related to the equation $a \pmod n \cdot x \pmod n = b \pmod n$ where $a \pmod n$ and $b \pmod n$ are elements of $\mathbb{Z}/n\mathbb{Z}$ and the unknown $x \pmod n$ 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 $a{\cdot}x + n{\cdot}y{=}b$ in the unknown x and y, a linear Diophantine equation **??**. The results on linear Diophantine equations **??** 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
  $E := \mathrm{Extgcd}(a, n)$
  $g := E_1$
  $z := E_2$
if $g|b$
  then
    return
      $\left\{ z{\cdot}\frac{b}{g} + k{\cdot}\frac{n}{g} \pmod{n} \mid k \in \mathbb{Z}/n\mathbb{Z} \right\}$
  else
    return
      $\emptyset$

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 Algorithm 1.3.4 for solving linear Diophantine equations.

□

Remark 2.2.2. In the terminology of Algorithm 2.2.1, the solutions of the related equation $a \pmod n \cdot x \pmod n = b \pmod n$ over $\mathbb{Z}/n\mathbb{Z}$ are the elements of the set

$$\left\{ z \cdot \frac{b}{g} + k \cdot \frac{n}{g} \pmod n \mid k \in \mathbb{Z}/n\mathbb{Z} \right\}.$$

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 1.2.5 we find

$$\begin{aligned} \gcd(24, 15) &= \\ 3 &= \\ 2 \cdot 24 - 3 \cdot 15. & \end{aligned}$$

Now 3 divides 12, so the solution set is $\{(2 \cdot 12 + k \cdot 15)/3 \mid k \in \mathbb{Z}\}$.

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 $x = 8 \pmod{15}$. Other solutions can be found by adding multiples of $15/3 \pmod{15}$ to this particular solution.

So, the complete set of solutions for x consists of the classes $3 \pmod{15}$, $8 \pmod{15}$, and $13 \pmod{15}$.

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 $x \equiv a_i \pmod{n_i}$ with $i \in \{1, \ldots, k\}$ has a solution. Indeed, the integer*

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

where for each i we have $y_i = \mathrm{Extgcd}(\frac{n}{n_i}, n_i)_3$, 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 1.2.5 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.

□

Here is another way of making the last statement of Theorem 2.2.4: If x is a solution, then the set of all solutions is the set $x \pmod{\prod_{i=1}^{k} n_i}$.

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

$$13, 27, 41, 55, 69, 83, 97, 111, 125, 139, 153, 167, 181, 195, 209, 223, 237.$$

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

$$5, 22, 39, 56, 73, 90, 107, 124, 141, 158, 175, 192, 209, 226.$$

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

The Chinese Remainder Theorem 2.2.4 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**
    $y_i := \text{Extgcd}(\frac{n}{n_i}, n_i)_3$

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 2.2.4.

□

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.

Figure 2.4: Pierre de Fermat (1601-1665)

Theorem 2.3.1 (Fermat's Little Theorem)**.** *Let p be a prime. For every integer a we have $a^p \equiv a \pmod{p}$. In particular, if a is not in $0 \pmod{p}$ then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Equivalently, for all elements $a \pmod{p}$ in $\mathbb{Z}/p\mathbb{Z}$ we have $(a \pmod{p})^p = a \pmod{p}$. For nonzero elements $a \pmod{p}$ we have

$$(a \pmod{p})^{p-1} = 1 \pmod{p}.$$

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

$$\sum_{i=0}^{p} \binom{p}{i} \cdot a^i.$$

Recall that the binomial coefficient is determined by

$$\binom{p}{i} = \frac{p!}{(p-i)! \cdot i!}.$$

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 $\binom{p}{i}$ 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 $(a+1)^p \equiv a+1 \pmod{p}$. 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 $a \pmod{p}$ in $\mathbb{Z}/p\mathbb{Z}$ we have $(a \pmod{p})^p = a \pmod{p}$.

For $a \pmod{p}$ equal to 0 the statements are trivial. Thus assume that $a \pmod{p}$ 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

$$M_{a \pmod{p}} : (\mathbb{Z}/p\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times, \quad b \pmod{p} \mapsto a \pmod{p} \cdot b \pmod{p},$$

that is, multiplication by $a \pmod{p}$. As $\mathbb{Z}/p\mathbb{Z}$ contains no zero divisors, see Theorem 2.1.13, the map is well defined. Moreover, this map is bijective. Indeed, its inverse is $M_{a \pmod{p}^{-1}}$, multiplication by $(a \pmod{p})^{-1}$. 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

$$\prod_{z\in(\mathbb{Z}/p\mathbb{Z})^\times} M_{a \pmod p}(z).$$

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

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

By Theorem 2.1.13 the product

$$\prod_{z\in(\mathbb{Z}/p\mathbb{Z})^\times} z$$

is nonzero and hence invertible, see Lemma 2.1.21. Therefore, $(a \pmod p)^{p-1} = 1 \pmod p$. Multiplying both sides of the equation by $a \pmod p$ proves the assertion.
The other statements in Fermat's Little Theorem 2.3.1 follow easily from the above assertions.

□

Example 2.3.2. As 7 is prime, Fermat's Little Theorem 2.3.1 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 2.3.1 we have

$$2^6 \equiv 1 \pmod 7,$$

so

$$1234^{1234} \equiv 2^{1234} \equiv 2^{6\cdot 205+4} \equiv 2^4 \equiv 2 \pmod 7.$$

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 **??**. 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 2.3.1 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. The Euler totient function $\Phi : \mathbb{N} \to \mathbb{N}$ 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

Theorem 2.3.7 (Euler Totient)**.** *The Euler Totient Function 2.3.5 satisfies the following properties.*

(a) *Suppose that n and m are positive integers. If* $\gcd(n,m)=1$*, then* $\Phi(n{\cdot}m)=\Phi(n){\cdot}\Phi(m)$*.*

(b) *If p is a prime and n a positive integer, then* $\Phi(p^n)=p^n-p^{n-1}$*.*

(c) *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

$$\Phi(a)=\prod_{i=1}^{s}((p_i)^{n_i}-(p_i)^{n_i-1}).$$

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

$$F\colon(\mathbb{Z}/n{\cdot}m\mathbb{Z})^\times\to(\mathbb{Z}/n\mathbb{Z})^\times\times(\mathbb{Z}/m\mathbb{Z})^\times$$

defined by

$$F(a \pmod{n{\cdot}m})=(a \pmod{n}, a \pmod{m})$$

is well defined.
The Chinese Remainder Theorem 2.2.4 implies that for each pair $(b \pmod{n}, c \pmod{m})$ in $(\mathbb{Z}/n\mathbb{Z})^\times\times(\mathbb{Z}/m\mathbb{Z})^\times$ there is one and only one class $a \pmod{n{\cdot}m}$ which is mapped onto the pair $(b \pmod{n}, c \pmod{m})$ 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 $\{1,\ldots,p^n\}$, we find $\Phi(p^n)=p^n-p^{n-1}$.

Part (c).

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

□

Example 2.3.8. By Theorem 2.3.7,

$$\begin{aligned}\Phi(100) &= \\ \Phi(2^2\cdot 5^2) &= \\ \Phi(2^2)\cdot\Phi(5^2) &= \\ (2^2-2)\cdot(5^2-5) &= \\ 40.&\end{aligned}$$

Let n be a prime. Then $\Phi(n)=n-1$. So, by Fermat's Little Theorem 2.3.1 we have $(a \pmod n)^{\Phi(n)}=1 \pmod n$. This statement can be generalized to arbitrary n.

Figure 2.5: Leonard Euler

Theorem 2.3.9 (Euler's Theorem)**.** *Suppose n is an integer with $n\geq 2$. Let $a \pmod n$ be an element of $(\mathbb{Z}/n\mathbb{Z})^\times$. Then*

$$(a \pmod n)^{\Phi(n)}=1 \pmod n.$$

Proof. The proof of the theorem almost literally follows the second proof of Fermat's little Theorem 2.3.1.
Suppose $a \pmod n$ in $(\mathbb{Z}/n\mathbb{Z})^\times$. Consider the map

$$M_{a \pmod n} \colon (\mathbb{Z}/n\mathbb{Z})^\times \to (\mathbb{Z}/n\mathbb{Z})^\times, \quad z \mapsto a \pmod n \cdot z.$$

In other words, $M_{a \pmod n}$ is multiplication by $a \pmod n$. By Lemma 2.1.21, this map is well defined. Moreover, the map is bijective. Indeed, its inverse is given by $M_{(a \pmod n)^{-1}}$, multiplication by $(a \pmod n)^{-1}$. 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

$$\prod_{z \in (\mathbb{Z}/n\mathbb{Z})^\times} M_{a \pmod n}(z).$$

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

$$\prod_{z \in (\mathbb{Z}/n\mathbb{Z})^\times} a \pmod n \cdot z = (a \pmod n)^{\Phi(p)} \cdot \prod_{z \in (\mathbb{Z}/n\mathbb{Z})^\times} z.$$

By Lemma 2.1.21 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 $(a \pmod n)^{\Phi(n)} = 1 \pmod n$. This proves the theorem.

□

Example 2.3.10. The set $(\mathbb{Z}/15\mathbb{Z})^\times$ contains 8 elements, one of them being $7 \pmod{15}$. For this element we have

$$7^8 \equiv 49^4 \equiv 4^4 \equiv 1^2 \equiv 1 \pmod{15}.$$

This in accordance with Euler's Theorem 2.3.9.

Let n be an integer. The *order* of an element $a \pmod n$ in $(\mathbb{Z}/n\mathbb{Z})^\times$ is the smallest positive integer m such that $(a \pmod n)^m = 1$. By Euler's Theorem 2.3.9 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 2.3.11. *Let n be an integer greater than 1.*

(a) *If $a \pmod n \in \mathbb{Z}/n\mathbb{Z}$ satisfies $(a \pmod n)^m = 1$ for some positive integer m, then $a \pmod n$ is invertible and its order divides m.*

(b) *For all elements $a \pmod n$ in $(\mathbb{Z}/n\mathbb{Z})^\times$ the order of $a \pmod n$ is a divisor of $\Phi(n)$.*

(c) *If $\mathbb{Z}/n\mathbb{Z}$ contains an element $a \pmod n$ of order $n-1$, then n is prime.*

Proof.
Part (a)

Suppose $a \pmod n \in \mathbb{Z}/n\mathbb{Z}$ satisfies $(a \pmod n)^m = 1$ for some integer m. Then, since $(a \pmod n) \cdot (a \pmod n)^{m-1} = 1$, the element $a \pmod n$ is invertible with inverse $(a \pmod n)^{m-1}$.
Let k be the order of $a \pmod n$, and set $q = \mathrm{quot}(m,k)$ and $r = \mathrm{rem}(m,k)$. Then $(a \pmod n)^r$ equals $(a \pmod n)^{m-q\cdot k} = (a \pmod n)^m \cdot ((a \pmod n)^k)^{-q}$, 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 2.3.9.

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 2.3.12. The element $7 \pmod{15}$ of $\mathbb{Z}/15\mathbb{Z}$ satisfies

$$7^4 \equiv 49^2 \equiv 4^2 \equiv 1 \pmod{15}.$$

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

Remark 2.3.13. Fermat's Little Theorem 2.3.1 and Theorem 2.3.11 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 $a \pmod{n}$ is not invertible, so $\gcd(a, n)\neq 1$.) The only Carmichael number less than 1000 is 561.

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 *data* into $E(data)$. The customer then sends the encrypted message $E(data)$ 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 $E(data)$. This means that E and D have the property that $D(E(data))=data$. The idea is that, in case one does not know D, it is hard (or almost impossible) to discover *data* from the encrypted message $E(data)$. Only the trusted shop can find the personal information in *data* by applying D to $E(data)$.
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 2.3.9. 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.

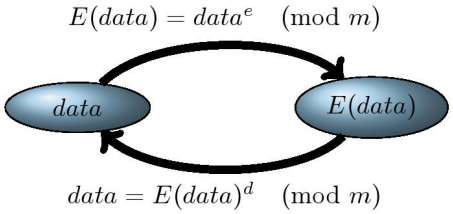

Figure 2.6:

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 2.3.9 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 Factorization 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 $a \equiv b \pmod{n}$.

Hint.
Use the definition.

Solution.
If $a{=}q_1{\cdot}n + r$ and $b{=}q_2{\cdot}n + r$, then $a - b{=}(q_1 - q_2){\cdot}n$, so n divides $a - b$. Hence, a and b are congruent modulo 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.

Solution.
Since m divides $a - b$, the integer m also divides $(a - b){\cdot}(a + b)$, i.e., $a^2 - b^2$.
Here is an easy counterexample to the last statement in the exercise: 1 is congruent to 4 modulo 3, but 1^2 is not congruent to 4^2 modulo 3^2.

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?

Hint.
Use the rules for addition and multiplication modulo 5.

Solution.
Computing modulo 5, we find that a^3 is congruent to 2^3, which is congruent to 3. Likewise, $-3{\cdot}a$ is congruent to -6, and hence to 4. So $a^3 - 3{\cdot}a + 1$ is congruent to $3 - 4 + 1$, which is equal to 0.

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.

Solution.
The product $a{\cdot}b$ is congruent to 5 modulo 7 by the rules for modular arithmetic. But then $a{\cdot}b$ is of the form $7{\cdot}q + 5$ for some positive integer q. Since 5 is less than 7, this shows that 5 is the remainder 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.

Hint.
Work modulo 4.

Solution.
Let x be an integer whose representation in the digital system is $x{=}[x_k, \ldots, x_1, x_0]_{10}$. Then $r{=}\mathrm{rem}(x, 100)$ has representation $r{=}[x_1, x_0]_{10}$.
Since $x = 100{\cdot}\mathrm{quot}(x, 100) + r = 4{\cdot}25{\cdot}\mathrm{quot}(x, 100) + r$, we find x to be divisible by 4 if and only if r is divisible by 4.

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?

Hint.
Note that 1000 is divisible by 8.

Solution.
Let x be an integer, whose representation in the digital system is $x{=}[x_k, \ldots, x_1, x_0]_{10}$. Then $r{=}\mathrm{rem}(x, 1000)$ has representation $r{=}[x_2, x_1, x_0]_{10}$.
Since $x = 1000{\cdot}\mathrm{quot}(x, 1000) + r = 8{\cdot}125{\cdot}\mathrm{quot}(x, 100) + r$, we find x to be divisible by 8 if and only if r is divisible by 8.

To test divisibility of r by 8, we notice that $r = 100{\cdot}x_2+10{\cdot}x_1+1{\cdot}x_0$, which modulo 8 equals $4{\cdot}x_2 + 2{\cdot}x_1 + x_0$. So x to be divisible by 8 if and only if $4{\cdot}x_2 + 2{\cdot}x_1 + x_0$ is divisible by 8.

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

Hint.
Notice that $a^n \equiv 1 \pmod{a-1}$.

Solution.
Suppose $x{=}[x_k, \ldots, x_0]_a$. Then

$$(a-1)|x$$

if and only if

$$(a-1)|(x_k + \ldots + x_0).$$

The proof of this claim runs like the proof of the nine test in Example 2.1.9.
Since $a \equiv 1 \pmod{a-1}$, we find $a^n \equiv 1 \pmod{a-1}$ for all integers n. As $[x_k, \ldots, x_0]_a{=}x_k{\cdot}a^k + \ldots + x_0{\cdot}a^0$ reduction modulo $a - 1$ implies that $x \equiv x_k + \ldots + x_0 \pmod{a-1}$. Thus $(a-1)|x$ if and only if $(a-1)|(x_k + \ldots + x_0)$.

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

Hint.
Consider the different possibilities modulo 3.

Solution.
If $n \equiv 0 \pmod 3$, then $n^4 + n^2 + 1 \equiv 0 + 0 + 1 \pmod 3$. Thus $n^4 + n^2 + 1$ is not divisible by 3.
If $n \equiv 1 \pmod 3$ or $n \equiv -1 \pmod 3$, then

$$n^4 + n^2 + 1 \equiv 1 + 1 + 1 \equiv 0 \pmod 3.$$

Therefore, $n^4 + n^2 + 1$ is divisible by 3.

Exercise 2.5.9. Prove the following statements:

(a) $13|(10^6 - 1)$.

(b) $17|(10^8 + 1)$.

(c) If $n \not\equiv 0 \pmod 5$, then $n^4 + 64$ is not prime.

(d) The number $2^{1000} + 5$ is divisible by 3.

(e) For every $n > 0$ we find that 3 is a divisor of $2^{2{\cdot}n} - 1$.

Hint.
Modular arithmetic.

Solution.

(a) We have $10^6 \equiv -3^6 \pmod{13}$. Furthermore $3^3 \equiv 1 \pmod{13}$, so $10^6 \equiv 1^2 \pmod{13}$ and the result follows.

(b) As 6·17=102, we have $10^8 + 1 \equiv 100^4 + 1 \equiv (-2)^4 + 1 \equiv 17 \equiv 0 \pmod{17}$

(c) $n^4+64 \equiv (n^4-1)+65 \equiv (n^4-1) \pmod 5$. But $(n^4-1)=(n^2-1)\cdot(n^2+1)$. Since a nonzero square is 1 or $-1 \pmod 5$ we find $(n^4-1) \equiv 0 \pmod{}$. Therefore, n^4+64 is divisible by 5.

(d) $2^{1000} + 5 \equiv (-1)^{1000} + 2 \equiv 1 + 2 \equiv 0 \pmod 3$.

(e) $2^{2\cdot n} - 1 \equiv (2^n - 1)\cdot(2^n + 1) \equiv 0 \pmod 3$.

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

(a) $3 \in \mathbb{Z}/37\mathbb{Z}$;

(b) $4 \in \mathbb{Z}/14\mathbb{Z}$.

Hint.
Extended Euclidean Algorithm 1.2.5.

Solution.
By the Characterization of Modular Invertibility 2.1.13, modulo 37 the inverse of 3 is 25. The inverse of 4 modulo 14 does not exist.

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 $641 | (2^{2^5} + 1)$.

(a) $641=2^9 + 2^7 + 1$ and so $2^7\cdot 5 \equiv 2^7\cdot(2^2 + 1) \equiv -1 \pmod{641}$.

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

Hint.
Modular arithmetic.

Solution.
We have to show that 641 divides $2^{2^5} + 1$, or in other words that $2^{2^5} + 1 \equiv 0 \pmod{641}$.
We compute

$$2^{32}+1 \equiv (2^7)^4\cdot 2^4+1 \equiv (2^7)^4\cdot(-5^4)+1 \equiv -(2^7\cdot 5)^4+1 \equiv -(-1)^4+1 \equiv 0 \pmod{641}.$$

Exercise 2.5.12. The binomial coefficient $\binom{p}{k}$ (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 $\binom{p}{k}$ 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.

Hint.
Use Newton's Binomium.

Solution.
If k is positive but less than p, there is no factor p in the denominator. So, for k different from 0 and p, this binomial coefficient is divisible by p. Thus, after expanding $(x+y)^p$ with the help of Newton's Binomium and after reducing the result modulo p we are left with $x^p + y^p$.

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

Hint.
An element x of $\mathbb{Z}/n\mathbb{Z}$ has an inverse if and only if $\gcd(x, n){=}1$.

Solution.
$(\mathbb{Z}/4\mathbb{Z})^\times{=}\{1, 3\}$, $(\mathbb{Z}/6\mathbb{Z})^\times{=}\{1, 5\}$ and $(\mathbb{Z}/12\mathbb{Z})^\times{=}\{1, 5, 7, 11\}$.

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

Hint.
An element x has an inverse if and only if $\gcd(x, n){=}1$.

Solution.
The invertible elements of of $\mathbb{Z}/p^2\mathbb{Z}$ are the elements in $\{1, 2, \ldots, p^2 - 1\}$ that are not a multiple of p. Clearly there are precisely $p^2 - p$ such elements.

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.

Solution.
Let n be equal to 12. Then 7 (mod 12) and 5 (mod 12) are invertible, but their sum is 12. Then 0 (mod 12) is not.
The class 7 (mod 12) is invertible but its product with 2 (mod 12) is 2 (mod 12) which is not invertible.
3 (mod 12) and 4 (mod 12) are zero divisors, but their sum 7 (mod 12) is not.

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:

(a) $2{\cdot}x \equiv 37 \pmod{21}$

(b) $5{\cdot}x \equiv 15 \pmod{25}$

(c) $3{\cdot}x \equiv 7 \pmod{18}$

Hint.
Use the Extended Euclidean Algorithm 1.2.5.

Solution.

(a) $x \equiv 8 \pmod{21}$.

(b) $x \equiv 3 \pmod{5}$.

(c) No solution.

Exercise 2.5.20. Solve the following system of linear congruences:

$$2{\cdot}x \equiv 37 \pmod{5}$$

and

$$3{\cdot}x \equiv 48 \pmod{7}.$$

Hint.
Use the Chinese Remainder Theorem 2.2.4.

Solution.
Multiplying the first equation by 3 yields $x \equiv 1 \pmod{5}$. Multiplying the second equation by -2 yields $x \equiv 2 \pmod{7}$. So $x \equiv 16 \pmod{35}$.

Exercise 2.5.21. Solve the following system of linear congruences:

$$x + y \equiv 6 \pmod{11}$$

and

$$2{\cdot}x - y \equiv 8 \pmod{11}.$$

Hint.
Reduce the problem to an equation with one variable.

Solution.
Add the two equations and find $3{\cdot}x \equiv 3 \pmod{11}$. As 3 is invertible modulo 11, we derive $x \equiv 1 \pmod{11}$. But then $y{=}5 \pmod{11}$.

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 2.3.1,

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

also true?

Hint.
Can you find an inverse for x?

Solution.
If for all nonzero x we have $x^{p-1} \equiv 1 \pmod{p}$, then every nonzero element $x \pmod{p}$ of $\mathbb{Z}/p\mathbb{Z}$ has an inverse, namely $x^{p-2} \pmod{p}$. So, indeed, p is a prime.

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.

Solution.
We show that $2^{341} - 2$ is divisible by 11 and 31, which implies that it is divisible by 341.
Modulo 11 we have

$$2^{341} - 2 \equiv 2^{34 \cdot 10+1} - 2 \equiv 2^1 - 2 \equiv 0 \pmod{11}.$$

Modulo 31 we have

$$2^{341} - 2 \equiv 2^{30 \cdot 11+11} - 2 \equiv 2^{11} - 2 \equiv 2^{5 \cdot 2+1} - 2 \equiv 0 \pmod{31}.$$

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 $\mathbb{Z}/n\mathbb{Z}$?

Solution.
$n - \Phi(n)$.

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

Solution.
Since the order of x is equal to $\Phi(n)$, we find the elements x, x^2, ..., $x^{\Phi(n)}$ to be different. Moreover, as all these elements are different, they form the set of $\Phi(n)$ invertible elements of $\mathbb{Z}/n\mathbb{Z}$.

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.

Solution.
The integer 2623 factors into the primes 43 and 61. So, with p=43 and q=61 we find $(p-1)\cdot(q-1)$=2520.
Next we look for a decoding number d satisfying $d \cdot e \equiv 1 \pmod{2520}$. A solution to this equation is d=613. To find the first two letters we compute $\mathrm{rem}(0249^{613}, 2623)$.

Here you might want to use a computer. The remainder equals 0405. This implies that the first two letters of the decoded text are a D and a E.
Going on like this, we find the encoded text:

DECODING SUCCEEDED.

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'?

2.6 Summary

In $\mathbb{Z}/n\mathbb{Z}$, the quotient ring modulo n, one can

- add, subtract, and multiply,
- divide 1 by a provided $\gcd(a, n)$=1; the result is called the inverse of a.

For these new rings we have looked at

- linear congruences,
- Chinese Remainder Theorem,
- Fermat's Little Theorem, which is useful for:
- cryptography (RSA).

Chapter 3

Polynomial arithmetic

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

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

$$a_0 + a_1 \cdot X + \ldots + a_n \cdot X^n = \sum_{k=0}^{n} a_k \cdot X^k.$$

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

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 $a_0, \ldots, a_n$ 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,

$$a_0 + a_1 \cdot X + \ldots + a_n \cdot X^n = a_n \cdot X^n + \ldots + a_1 \cdot X + a_0.$$

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.

Let $a = a_0 + a_1 \cdot X + \ldots + a_n \cdot X^n$ and $b = b_0 + b_1 \cdot X + \ldots + b_m \cdot X^m$ 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

$$a \cdot b = c_0 + c_1 \cdot X + \ldots + c_{2 \cdot m} \cdot X^{2 \cdot m},$$

  where

$$c_k = a_0 \cdot b_k + a_1 \cdot b_{k-1} + \ldots + a_k \cdot b_0.$$

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);
- $(a \cdot b) \cdot c = a(b \cdot c)$ (associativity of multiplication);
- $a(b + c) = a \cdot b + a \cdot c$ (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

$$a_0 \cdot b_k + a_1 \cdot b_{k-1} + \ldots + a_k \cdot b_0 = b_0 \cdot a_k + b_1 \cdot a_{k-1} + \ldots + b_k \cdot a_0$$

(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.

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

$$a \cdot b = X^5 + X^3 + X - 1.$$

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.

$$(1+X)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot X^k.$$

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.

$$X + 2 \cdot X^2 + 3 \cdot X^3 - 4 \cdot X^2 = X + -2 \cdot X^2 + 3 \cdot X^3.$$

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 $a = a_0 + a_1 \cdot X + \ldots + a_n \cdot X^n$ 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.*

(a) $\text{degree}(a{\cdot}b){=}\text{degree}(a)+\text{degree}(b)$.

(b) $\text{degree}(a+b){\leq}\max(\text{degree}(a),\text{degree}(b))$.

(c) *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 $-\infty+m{=}m$ 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.
If $b{\neq}0$ divides a, the polynomial q with $a{=}q{\cdot}b$ is unique and is called the *quotient* of a and b; it is denoted by $\frac{a}{b}$ or a/b.

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.

Example 3.2.5. The polynomial X^2-1 is a divisor of X^6-1, for

$$X^6-1{=}(X^2-1){\cdot}(X^4+X^2+1).$$

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.

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* $\deg(r) < \deg(b)$.
The polynomials q and r are uniquely determined. They are called the quotient *and* remainder *of a divided by b.*

Proof. (Compare this proof with the proof of Division with Remainder 1.1.5 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 degree(r) < degree(b) and also $a{=}q'{\cdot}b + r'$ with degree(r') < 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 3.2.3, the degree of $r' - r$ is less than the degree of a, so, by Part 1 of the Degree Formulas 3.2.3, both sides of the equality must be equal to 0. In particular, $r' - r{=}0$ and, as $a{\neq}0$, also $q{=}q'$. □

The quotient q is denoted by $\mathrm{quot}(a, b)$ and the remainder r is denoted by $\mathrm{rem}(a, b)$, just like for integers.

Remark 3.2.8. At various places in the proof of Division with Remainder Theorem 3.2.7 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:

$$\begin{aligned} a - 2{\cdot}X^2{\cdot}b &= \\ 2{\cdot}X^4 + X - 2{\cdot}X^2{\cdot}(X^2 + 1) &= \\ -2{\cdot}X^2 + X. \end{aligned}$$

  From this step we conclude that $2{\cdot}X^2$ is a term of the quotient q. We now have

$$a{=}2{\cdot}X^2{\cdot}b + -2{\cdot}X^2 + X.$$

  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

$$2{\cdot}X^2 + X + 2{\cdot}(X^2 + 1){=}X + 2.$$

  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.,

$$2{\cdot}X^4 + X{=}(2{\cdot}X^2 - 2){\cdot}(X^2 + 1) + X + 2.$$

The Division and Remainder theorem 3.2.7 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**
  $q := q + \frac{\text{lc}(r)}{\text{lc}(b)}{\cdot}X^{\text{degree}(r)-\text{degree}(b)}$
  $r := r - \frac{\text{lc}(r)}{\text{lc}(b)}{\cdot}X^{\text{degree}(r)-\text{degree}(b)}{\cdot}b$, $n :=$ 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.

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.

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 the Existence and Uniqueness of the gcd Theorem 3.2.14.

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 the gcd Theorem 3.2.14.

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 can be found among the polynomials of the form $x \cdot a + y \cdot b$, where x and y are 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. Since 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 3.2.3 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**
  $c := a\ ,\ a := b\ ,\ b := \mathrm{rem}(c,b)$
return
  $\frac{a}{\mathrm{lc}(a)}$

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

$$\begin{aligned}\gcd(X^4-1,X^6-1) &= \\ \gcd(X^6-1,X^4-1) &= \\ \gcd(X^4-1,X^2-1) &= \\ \gcd(X^2-1,0) &= \\ X^2-1.&\end{aligned}$$

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

Proof. As degree(b) goes strictly down at each step, termination is guaranteed.
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 3.2.16, 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 $\mathrm{lc}(a)$.
Although we do not use the equality involving u and v, it is worth notingthat, 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=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 2×3 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

$$X^2-1=(-X^2)\cdot(X^4-1)+1\cdot(X^6-1).$$

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 3.2.18 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 3.2.20, 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 are both instances of *Euclidean rings*, algebraic structures for which there exists a measure with comparable properties.

The Extended Euclidean Algorithm 3.2.18 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.*

(a) $\gcd(a,b){=}d$.

(b) *The polynomial d is a monic common divisor of a and b of maximal degree.*

(c) *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 Theorem **??**.

The third statement is equivalent to the first.

Let d=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 degree(d)≤degree(e). Moreover, as a result of the Extended Euclidean Algorithm **??**, d itself can also be written as a combination of a and b. So degree(e)≤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.

□

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.

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 3.2.23.

$$(1 + X - X^2) \cdot (X^5 + 1) + (-1 + X - X^2 - X^3 + X^4) \cdot (X^3 - 1) = 2.$$

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

$$a \colon R \to R, \quad r \mapsto a(r),$$

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 $\mathbb{R}[X]$.

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 3.3.5 (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 be a polynomial in $R[X]$ of degree at most $n-1$. Write

$$f(X) = f_0 + f_1 \cdot X + \ldots + f_{n-1} \cdot X^{n-1}$$

and substitute the given values. This transforms the problem into that of solving the system of linear equations:

$$f(x_i) = f_0 + f_1 \cdot x_i + \ldots + f_{n-1} \cdot (x_i)^{n-1}$$

where $i \in \{1, \ldots, n\}$.
This system can be rewritten in matrix form as $M \cdot f = a$, where f is the vector $(f_0, \ldots, f_{n-1})^\top$, a is the vector $(a_1, \ldots, a_n)^\top$, and M the matrix

$$\begin{pmatrix} 1 & x_1 & (x_1)^2 & \ldots & (x_1)^{n-1} \\ 1 & x_2 & (x_2)^2 & \ldots & (x_2)^{n-1} \\ & & & \ldots & \\ 1 & x_n & (x_n)^2 & \ldots & (x_n)^{n-1} \end{pmatrix}.$$

The coefficients of the polynomial f (or, equivalently, the coordinates of the vector f) are the n unknowns of the system of linear equations.
Now M is a so-called *Vandermonde matrix*. It has the special property that its determinant is given by the formula

$$\det(M)=\prod_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,n\},i>j}(x_i-x_j).$$

Since the x_i are chosen to be distinct, the determinant is nonzero. So, M is invertible. This means that the system of linear equations has exactly one solution. In other words, there is a unique polynomial f over R of degree at most $n-1$ satisfying the interpolation conditions of the theorem.

□

Example 3.3.6. 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 Theorem 3.3.5, applied for $n=2$.

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

Theorem 3.3.7 (Characterization of the Zeros of a Polynomial)**.** *Let R be a field and $f \in R[X]$.*

(a) *An element $x \in R$ is a zero of f if and only if $X - x$ divides f.*

(b) *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 3.2.3, every divisor of it has degree at most n, so the claim implies that f has at most n different zeros. □

Remark 3.3.8. Another proof of the second statement of the theorem (and the claim used in the proof) will follow from Unique Factorization 3.4.6.

Example 3.3.9. 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,

$$X^6 = (X^2 - X + 1) \cdot (X^2 + X + 1) \cdot (X + 1) \cdot (X - 1),$$

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.

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 3.3.10 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 3.3.7.
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

$$\begin{aligned} f(\overline{x}) &= \\ \overline{f(x)} &= \\ \overline{0} &= \\ 0. & \end{aligned}$$

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. A polynomial $f \in R[X]$ is called *irreducible* if $\text{degree}(f) > 0$ 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*.

We shall study factorizations of a polynomial, that is, ways to write the polynomial as a product of polynomials of smaller degree.

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 $X^3 + X^2 + 1$.

With the help of the Fundamental Theorem of Algebra 3.3.10, we can determine 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 **??**, 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 **??**, 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

$$a{\cdot}(X - \frac{-b + \sqrt{b^2 - 4{\cdot}a{\cdot}c}}{2{\cdot}a}){\cdot}(X - \frac{-b - \sqrt{b^2 - 4{\cdot}a{\cdot}c}}{2{\cdot}a}).$$

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

$$a{\cdot}(X - \frac{-b + i{\cdot}\sqrt{4{\cdot}a{\cdot}c - b^2}}{2{\cdot}a}){\cdot}(X - \frac{-b - i{\cdot}\sqrt{4{\cdot}a{\cdot}c - b^2}}{2{\cdot}a}).$$

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 3.2.18.

□

Compare the next theorem with similar results on integers **??**.

Lemma 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 3.2.18, 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 Characterization of Relative Prime Polynomials 3.4.7 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 $f{=}(\frac{f}{g}){\cdot}g$, 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 Theorem **??** 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 $(X^2+1){\cdot}(X+1){\cdot}(X-1)$.
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 $(X^2-X-1)\cdot(X^2+X-1)$. 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

$$(a_0+a_1\cdot X+\ldots+a_{n-1}\cdot X^{n-1}+a_n\cdot X^n)^p = a_0+a_1\cdot X^p+\ldots+a_{n-1}\cdot X^{p\cdot(n-1)}+a_n\cdot X^{p\cdot n}.$$

Hint.
Use Fermat's Little Theorem 2.3.1.

Solution.
Let p be a prime. We will prove the claim using induction on the degree of the polynomial.
If the degree is 0, the result follows directly from Fermat's Little Theorem 2.3.1. So suppose the claim is valid for all f in $(\mathbb{Z}/p\mathbb{Z})[X]$ of degree less than n. We write $f=a\cdot X^n+g$, where g is a polynomial of degree less than n. It suffices to prove that

$$(a\cdot X^n+g)^p = a\cdot X^{n\cdot p}+g^p.$$

For then the rest follows by the induction hypothesis. Expanding the power on the left-hand side using Newton's Binomium, we find that all but (possibly) the first and the last term vanish. Indeed, for $1<k<p$, the binomial coefficient $\binom{p}{k}$ is equal to 0 modulo p. Moreover, by Fermat's Little Theorem 2.3.1 the coefficient of the term $X^{n\cdot p}$ equals a. This proves the claim.

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 $\mathbb{Z}/3\mathbb{Z}$.

Hint.
Use Euclid's Algorithm for Polynomials 3.2.16.

Solution.
The gcd of X^2+1 and X^3+1 over $\mathbb{Q}$. Euclid's Algorithm for Polynomials 3.2.16 gives

$$X^3+1 = X\cdot(X^2+1)+1-X,$$
$$X^2+1 = (-1-X)\cdot(1-X)+2.$$

So,

$$\begin{aligned}\gcd(X^3+1, X^2+1) &= \\ \gcd(X^2+1, 1-X) &= \\ \gcd(1-X, 2) &= \\ 2.\end{aligned}$$

Notice that the gcd of polynomials is determined up to a scalar multiple and that we agreed to take the gcd to be monic, so that here the gcd is 1.
In order to express the gcd as a combination of the starting polynomials, we need to keep track at every step how the remainder can be expressed in such a way. So in this example we obtain:

$$\begin{aligned}2 &= \\ X^2+1+(X+1)\cdot(1-X) &= \\ X^2+1+(X+1)\cdot(X^3+1-X\cdot(X^2+1)) &= \\ (1-X-X^2)\cdot(X^2+1)+(X+1)\cdot(X^3+1).\end{aligned}$$

Since we claimed above that we could equally well say that 1 is the gcd we should be able to express it as a combination of the polynomials. Well, here it is:

$$1 = \frac{1}{2}\cdot(X+1)\cdot(X^3+1)+\frac{1}{2}\cdot(-X^2-X+1)\cdot(X^2+1).$$

We compute the gcd of X^2+1 and X^3+1 over $\mathbb{Z}/2\mathbb{Z}$ in a similar manner.

Euclid's Algorithm for Polynomials 3.2.16 gives

$$\gcd(X^3+1, X^2+1) = X+1.$$

We can express it as

$$X^3+1+(X^2+1)\cdot X = X+1.$$

Finally, application of Euclid's Algorithm for Polynomials 3.2.16 to the polynomials X^2-X+1 and X^3+X+2 over $\mathbb{Z}/3\mathbb{Z}$ gives:

$$\gcd(X^2-X+1+1, X^3+X+1) = X+1.$$

This gcd can be expressed as

$$X^3+X+2-(X+1)\cdot(X^2-X+1) = X+1.$$

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 $\mathbb{Z}[X]$.

Hint.
First prove that the leading coefficient of q is an integer. Then proceed by induction on the degree of q.

Solution.
These polynomials satisfy

$$a = q\cdot b + r.$$

It is given that a and b have integer coefficients, moreover b is monic. We will prove that both q and r have integer coefficients. First, observe that the leading coefficient of q is an integer.
Since the degree of r is strictly less than the degree of b we have:

$$\mathrm{lc}(a) = \mathrm{lc}(q)\cdot\mathrm{lc}(b).$$

As the leading coefficient of b equals 1 and the leading coefficient of a is an integer, the leading coefficient of q must also be an integer. We will now prove that all coefficients of q are integers by using induction on the degree of q.
If q is of degree 0, then it only has one coefficient, which by the above remark must be an integer. Now suppose that q is of degree n and that we have proven the claim for the cases where the degree of q is less than n. We know that the leading coefficient of q is an integer, so we can express q as $c\cdot X^n + q'$, for some integer c. We can now rewrite the equation $a = q\cdot b + r$ as

$$(a - c\cdot X^n)\cdot b = q'\cdot b + r,$$

where $q' = q - c\cdot X^n$. Since the degree of q' is less than n we already know, by our induction hypothesis, that it has integer coefficients. So the same holds for q.
Finally, we have to prove that r has integer coefficients. The expression $a - q\cdot b$ has only integer coefficients, and it is equal to r. This finishes the proof.

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 $p{\cdot}a + q{\cdot}b + r{\cdot}c{=}1$.

Hint.

- Use the definition to unravel both sides; note the convention about gcd's having leading coefficient 1.
- Use the previous item and the Extended Euclidean Algorithm 3.2.18.

Solution.

- Let d be the gcd on the left-hand side and d' the gcd on the right-hand side. Then d' is a common divisor of $\gcd(b, c)$ and a. So d' is a common divisor of a, b, c and therefore divides d.

  Since d is a common divisor of a, b, c, it is certainly a common divisor of b and c and therefore a divisor of $\gcd(b, c)$. But then d is a common divisor of $\gcd(b, c)$ and a, and we conclude that it is a divisor of d'. As d and d' have leading coefficient 1, they must be equal.
- Suppose a, b, c are relatively prime. The previous item implies that $\gcd(b, c)$ and a are relatively prime. By the Extended Euclidean Algorithm 3.2.18 there exist polynomials x and y such that
$$x{\cdot}\gcd(b, c) + y{\cdot}a{=}1.$$
  Again by the Extended Euclidean Algorithm 3.2.18 there exist polynomials u and v such that
$$\gcd(b, c){=}u{\cdot}b + v{\cdot}c.$$
  By substitution we then find
$$x(u{\cdot}b + v{\cdot}c) + y{\cdot}a{=}1.$$
  So, with $p{=}y$, $q{=}x{\cdot}u$ and $r{=}x{\cdot}v$ we have found
$$p{\cdot}a + q{\cdot}b + r{\cdot}c{=}\gcd(a, b, c).$$

Conversely, suppose there exist polynomials p, q, r such that

$$p{\cdot}a + q{\cdot}b + r{\cdot}c{=}1.$$

If d is a common divisor of a, b, c, then d is also a divisor of the combination $p{\cdot}a + q{\cdot}b + r{\cdot}c$. Since this combination equals 1, we conclude that $\gcd(a, b, c)$ equals 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.

Solution.
Since a divides b and c, there are polynomials e and f such that $b{=}a{\cdot}e$ and $c{=}a{\cdot}f$. But then $b + c{\cdot}d{=}a{\cdot}(e + f{\cdot}d)$, which proves that a divides $b + c{\cdot}d$.

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.

(a) $a{=}X^4 + 3{\cdot}X^2 + X + 1$ and $b{=}X^2 + X + 1$ in $\mathbb{Q}[X]$;

(b) $a{=}X^4 + 3{\cdot}X^2 + X + 1$ and $b{=}X^2 + X + 1$ in $\mathbb{Z}/2\mathbb{Z}[X]$;

(c) $a{=}X^4 + 3{\cdot}X^2 + X + 1$ 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.

Solution.
Every polynomial f that can be written as $p{\cdot}a + q{\cdot}b$ with p and q polynomials over R is divisible by any common divisor of a and b. Since f itself is a common divisor of a and b, it is a greatest common divisor.

Exercise 3.5.10. Determine polynomials a and b in $\mathbb{Q}[X]$ such that

$$a{\cdot}(X^2 + 1) + b{\cdot}(X^3 - X + 1){=}X - 1.$$

Exercise 3.5.11. Determine polynomials a and b in $\mathbb{Z}/2\mathbb{Z}[X]$ such that

$$a{\cdot}(X^2 + 1) + b{\cdot}(X^3 - X + 1){=}X - 1.$$

3.5.3 Polynomial functions

Exercise 3.5.12. Find all zeros of each of the following polynomials

(a) $X^2 + 2{\cdot}X + 2$ in $(\mathbb{Z}/5\mathbb{Z})[X]$;

(b) $X^2 + X + 1$ in $(\mathbb{Z}/24\mathbb{Z})[X]$;

(c) $X{\cdot}(X + 1){\cdot}(X + 2)$ in $(\mathbb{Z}/12\mathbb{Z})[X]$;

(d) $2{\cdot}X^2 + 13{\cdot}X + 9$ in $(\mathbb{Z}/33\mathbb{Z})[X]$.

Hint.

(a) Split off a square: $(X + 1)^2 + 1$ and rewrite 1 as -2^2.

(b) Rewrite as $X{\cdot}(X+1)+1$ and show that substituting an integer always leads to an odd number modulo 24.

(c) Be careful: don't draw the conclusion that a zero should be a zero of at least one of the factors. Instead of trying each of the numbers $0, 1, \ldots, 11$ consider the question: when is the product of three consecutive integers divisible by 12?

(d) Replace the coefficients by even representatives (modulo 33) and get rid of the leading coefficient 2. Then split off a square. Beware: 33 is not a prime.

Solution.

(a) Of course, since 5 is such a small number, one can easily check case by case. One finds the zeros 1 and 2. More elegantly, split off a square $(X+1)^2+1$ and note that $1 \equiv -2^2 \pmod 5$, so that $(X+1)^2 - 2^2$ factors as $(X+3){\cdot}(X-1)$. We conclude that 1 and 3 are the zeros.

(b) Rewrite the polynomial as $X{\cdot}(X + 1) + 1$. If you substitute an integer in $X{\cdot}(X + 1)$, the result is always even. Calculating modulo 24, the result is always even modulo 24 and so will never be equal to -1 modulo 24. Therefore there are no solutions to this equation.

(c) Whenever you substitute an integer you get the product of three consecutive integers. Being 0 (mod 12) means that the product is divisible by 12=4·3. Among three consecutive integers there is always one that is divisible by 3. If the first and third integer are even, then the product is divisible by 4·3. If only the second integer is even, then the product is divisible by 12 if this even number is divisible by 4. Otherwise the product is not divisible by 12. In the range $0, 1, \ldots, 11$ this happens precisely for the products $1{\cdot}2{\cdot}3, 5{\cdot}6{\cdot}7$, and $9{\cdot}10{\cdot}11$. So the 'nonzeros' are $1, 5, 9$ and the zeros are $0, 2, 3, 4, 6, 7, 8, 10, 11$.

(d) Since 2 and 33 are relatively prime, 2 is invertible in $\mathbb{Z}/33\mathbb{Z}$. Using $13 \equiv -20 \pmod{33}$ and $9 \equiv 42 \pmod{33}$, and dividing the polynomial by 2, we can rewrite the equation $2{\cdot}X^2 + 13{\cdot}X + 9{=}0$ as $X^2 - 10{\cdot}X + 21{=}0$. Next we split off a square: $(X-5)^2 - 4{=}0$. So we must solve $(X-5)^2 - 2^2{=}0$ or, equivalently, $(X-5-2){\cdot}(X-5+2){=}0$, i.e., $(X-7){\cdot}(X-3){=}0$. Two zeros are immediate: 3 and 7. But more zeros could arise since the product on the left-hand side is also zero modulo 33 if one of the factors is zero modulo 3 and the other zero modulo 11. This leads to the zeros 18 and 25.

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 $n{\in}\mathbb{Z}$.

Hint.
If $f(n){=}p$ is a prime, consider $f(n + k{\cdot}p)$.

Solution.
Suppose that $f(n){=}p$ is a prime. We claim that p is a divisor of $f(n + k{\cdot}p)$, for any integer k.
We look at the difference $f(n) - f(n + k{\cdot}p)$; it is a sum of expressions of the form $a_i{\cdot}n^i - a_i{\cdot}(n + k{\cdot}p)^i$. Applying Newton's Binomium, we see that the term $a_i{\cdot}n^i$ drops out and that the remainder is divisible by p. The only way the numbers $f(n + k{\cdot}p)$ can all be prime is when they are all equal to p. But in that case the polynomial would be constant, contradicting the assumption that the degree of f is at least 1.

Exercise 3.5.14. Find all polynomials $p{\in}\mathbb{Q}[X]$ that satisfy $p(x){=}p(-x)$ for any x in $\mathbb{Q}$.

Hint.
Notice that there are polynomials a and b such that we can write the polynomial p as $p(X){=}a(X^2) + b(X^2){\cdot}X$.

Solution.
There exist polynomials a and b such that $p(X){=}a(X^2) + b(X^2){\cdot}X$. The equation $p(x){=}p(-x)$ yields:

$$a(x^2) + b(x^2){\cdot}x{=}a(x^2) - b(x^2){\cdot}x,$$

for all x in $\mathbb{Q}$. From this it follows that $b(x^2)(2{\cdot}x){=}0$ for all x so that b has infinitely many zeros and therefore equals 0. We conclude that p is a polynomial in X with only terms of even degree.
On the other hand, if p is a polynomial with only terms of even degree, then it satisfies the condition $p(x){=}p(-x)$ for all x.

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 $\mathbb{Z}/6\mathbb{Z}$?

Solution.

If we are working over $\mathbb{Z}/2\mathbb{Z}$, then always $x{=}{-}x$ and hence also $p(x){=}p(-x)$ for any polynomial p.
Now suppose that we are working over $\mathbb{Z}/6\mathbb{Z}$. Then we follow the strategy of the previous exercise. We first write $p(X){=}a(X^2) + b(X^2){\cdot}X$. Again, the condition $p(x){=}p(-x)$ translates to $b(x^2)(2{\cdot}x){=}0$. This is equivalent to requiring that $b(1)$, $b(2)$, $b(4)$, and $b(5)$ be 3 modulo 6.

3.5.4 Factorization

Exercise 3.5.16. Consider the polynomial $a{=}a_0 + a_1{\cdot}X + \ldots + a_{n-1}{\cdot}X^{n-1} + a_n{\cdot}X^n$ in $\mathbb{Z}[X]$, with $a_n{\neq}0$.

(a) Prove: If $r{\in}\mathbb{Z}$ is a zero of a, then r is a divisor of a_0.

(b) 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.

(c) Find all rational roots of the polynomial $15 - 32{\cdot}X + 3{\cdot}X^2 + 2{\cdot}X^3$.

Hint.
Substitute r for x in a. Now look which terms are divisible by r and which terms are not. What can you conclude from that?

Solution.
Let r be root of $a_0 + a_1{\cdot}X + \ldots + a_{n-1}{\cdot}X^{n-1} + a_n{\cdot}X^n$. Substitute r for x and rewrite to get:

$$-a_0{=}a_1{\cdot}r + \ldots + a_{n-1}{\cdot}r^{n-1} + a_n{\cdot}r^n.$$

Since r divides the right-hand side, it must also divide the left-hand side. This proves the first assertion.
Now suppose that $\frac{r}{s}$ is a root of a, with r and s relatively prime. Of course, s cannot be zero. Again we substitute this in a, and multiply by s^n. We get:

$$0{=}a_0{\cdot}s^n + a_1{\cdot}s^{n-1}{\cdot}r + \ldots + a_{n-1}{\cdot}s{\cdot}r^{n-1} + a_n{\cdot}r^n.$$

From this equation, we infer that the term $a_n{\cdot}r^n$ must be divisible by s, since all other terms are. As r and s are relatively prime, we conclude that s must divide a_n. In a similar way, the equation shows that r must divide the term $a_0{\cdot}s^n$. But then r must divide a_0, again since r and s are relatively prime.
If r/s is a rational root of $15 - 32{\cdot}X + 3{\cdot}X^2 + 2{\cdot}X^3$, then r must divide 15 and s must divide 2. Therefore there are only 32 cases to consider. Trying them all, we find that there are three rational roots: $3, \frac{1}{2}, -5$.

Exercise 3.5.17. Consider the ring $(\mathbb{Z}/3\mathbb{Z})[X]$ of polynomials in X with integer coefficients modulo 3.

(a) How many polynomials of degree n are there in $(\mathbb{Z}/3\mathbb{Z})[X]$?

(b) Determine all irreducible polynomials in $(\mathbb{Z}/3\mathbb{Z})[X]$ of degrees 2 and 3.

Hint.
Recall that a reducible polynomial of degree 2 or 3 always has a zero.

Solution.
A polynomial in $(\mathbb{Z}/3\mathbb{Z})[X]$ of degree n has the form $a_0+\ldots+a_{n-1}\cdot X^{n-1}+a_n\cdot X^n$. The coefficient a_n can be 1 or 2, all the other coefficients can be $0, 1$ or 2. In total there are $2\cdot 3^n$ choices.
If a polynomial of degree 2 or 3 is reducible, it must have a factor of degree 1; hence it must have a root. So to check whether a given polynomial (of degree 2 or 3) is irreducible, it suffices to check if $0, 1$ or 2 is a zero. We will only do the degree 2 case in detail. First, note that we may assume that $a_2{=}1$ (if a is an irreducible with $a_2{=}2$, then $-a$ is also irreducible). Next, we may also assume that a_0 is not 0, (since in that case we will have 0 as a root). The list of remaining polynomials is: $X^2{+}0\cdot X{+}1$, irreducible; $X^2+0\cdot X+2, x{=}1$ is a zero; $X^2+1\cdot X+1, x{=}1$ is a zero; $X^2+1\cdot X+2$, irreducible; $X^2+2\cdot X+1, x{=}2$ is a zero; $X^2+2\cdot X+2$, irreducible. The above three polynomials and minus these three polynomials together are all six irreducible polynomials of degree 2.
The same strategy (only with some more work) yields all irreducible polynomials of degree 3.
The result of that analysis is that the irreducible polynomials of degree 3 with leading coefficient 1 are X^3-X+1, X^3-X^2+1, X^3+X^2-1, X^3+X^2+X-1, X^3-X^2+X+1, X^3+X^2-X+1 and X^3-X^2-X-1.

Exercise 3.5.18. Verify the identity of polynomials

$$(X^2-1)^2+(2\cdot X)^2{=}(X^2+1)^2.$$

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 $(X^2-1)^2+(2\cdot X)^2{=}(X^2+1)^2$.

Hint.
Expand the left-hand side of the identity, and evaluate it in p/q.

Solution.
Expanding the left-hand side of the identity yields

$$X^4 - 2{\cdot}X^2 + 1 + 4{\cdot}X^2 = X^4 + 2{\cdot}X^2 + 1,$$

which equals the right-hand side. Substituting $\frac{p}{q}$ in the identity and multiplying through by q^4 produces the identity of integers $(p^2 - q^2)^2 + (2{\cdot}p{\cdot}q)^2 = (p^2 + q^2)^2$. For example, if you take $p=3$ and $q=2$, this yields $5^2 + 12^2 = 13^2$.

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 $d(X + a)$.

Hint.
If $h(X)$ divides $f(X)$, then $h(X + a)$ divides $f(X + a)$.

Solution.
By the Extended Euclidean Algorithm 3.2.18, there exist polynomials $r(X)$ and $s(X)$ such that

$$r(X){\cdot}f(X) + s(X){\cdot}g(X) = d(X).$$

This implies the relation

$$r(X + a){\cdot}f(X + a) + s(X + a){\cdot}g(X + a) = d(X + a).$$

Since $d(X)$ divides both $f(X)$ and $g(X)$, we find that $d(X + a)$ divides both $f(X + a)$ and $g(X + a)$. So $d(X + a)$ is a common divisor of $f(X + a)$ and $g(X + a)$. Every common divisor $h(X)$ of $f(X + a)$ and $g(X + a)$ divides

$$r(X + a){\cdot}f(X + a) + s(X + a){\cdot}g(X + a) = d(X + a).$$

So $h(X)$ is a divisor of $d(X + a)$. Since the leading coefficient of $d(X + a)$ is equal to the leading coefficient of $d(X + a)$, the latter is the greatest common divisor of $f(X + a)$ and $g(X + a)$.

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 3.2.18 to find constants a, b, c such that $\frac{3}{X^3-1} = \frac{a}{X-1} + \frac{b{\cdot}X+c}{X^2+X+1}$.

Hint.
Find the gcd of $X - 1$ and $X^2 + X + 1$ and write it as a combination of these two polynomials. Then divide both sides by $X^3 - 1$.

Solution.

The Extended Euclidean Algorithm 3.2.18 yields the relation

$$X^2 + X + 1 + (-X - 2)\cdot(X - 1) = 3.$$

Now divide both sides by $X^3 - 1$.

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 $R[X]$.

Hint.
Think of the proof of the theorem stating that there are infinitely many primes 1.4.4.

Solution.
Suppose $p_1, \ldots, p_n$ are distinct irreducible polynomials in $R[X]$ and consider the polynomial $p_1 \cdot \ldots \cdot p_n + 1$. None of the n polynomials $p_1, \ldots, p_n$ divides $p_1 \cdot \ldots \cdot p_n + 1$, so $p_1 \cdot \ldots \cdot p_n + 1$ has an irreducible factor different from $p_1, \ldots, p_n$. This argument shows that there is no bound on the number of irreducible polynomials. Note that this argument also shows that there are infinitely many irreducible polynomials with leading coefficient 1.

Exercise 3.5.22. Determine all irreducible polynomials p and q in $\mathbb{Z}[X]$ that satisfy the equation $(X^2 + 1)\cdot p + (X + 2)\cdot q = p \cdot q$.

Hint.
Move one term from the left-hand side to the right-hand side, then rewrite the new right-hand side as a product of two factors and compare with the factors on the left-hand side.

Solution.
Rewrite $(X^2 + 1)\cdot p + (X + 2)\cdot q = p \cdot q$ as $(X^2 + 1)\cdot p = q \cdot (p - X - 2)$. Since p divides the left-hand side it divides the right-hand side. As p is irreducible, it divides q or $p - X - 2$; in the latter case it divides $X + 2$.

(a) p divides q. Then $p = a \cdot q$ for some rational number a. Substituting in the equation we find $q = X^2 + 1 + \frac{1}{a}\cdot(X + 2)$. It follows that $a = -1$ or 1 because q has integer coefficients. This leads to two candidate solutions:

- For $a = 1$ we find $p = q = X^2 + X + 3$.
- For $a = -1$ we find $p = -X^2 + X + 1$ and $q = X^2 - X - 1$.

It is easily verified that these solutions are irreducible polynomials since their zeros are nonrational, and that they satisfy the given equation.

(b) p divides $X + 2$. Then $p = b \cdot (X + 2)$ for some integer b. Substituting in the equation we find $q = \frac{b}{b-1}\cdot(X^2 + 1)$. But this polynomial has integer coefficients only if $b = 2$. In conclusion, we find $p = 2 \cdot X + 4$ and $q = 2 \cdot X^2 + 2$. Since both are irreducible and since they satisfy the equation we have found another solution.

3.6 Summary

Polynomials (with coefficients from an arithmetic system like $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}/n\mathbb{Z}$) form again an arithmetical system (a polynomial ring) in which one can add, subtract, and multiply. To describe a polynomial we introduced concepts like coefficient, term, and monomial. From polynomials we can make a polynomial function (with the coefficient ring for the domain and codomain). For polynomials with coefficients from a field we discussed (similarly to the integer case)

- division with remainder and a division algorithm
- Euclid's Algorithm for determining the gcd
- the Extended Euclidean Algorithm for writing the gcd of two polynomials as a linear combination of the two polynomials
- irreducible polynomials (analogous to prime numbers)
- factorization (analogous to integer factorization)

Chapter 4

Modular polynomial arithmetic

4.1 Congruence modulo a polynomial

In Modular Arithmetic **??**, computation modulo a fixed integer n is discussed. 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 Theorem 3.3.7 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 Theorem 3.3.7 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 congruence classes are in bijective correspondence with $\mathbb{Q}$, the set of possible values of the polynomial function of d, see Definition 3.3.1.

We introduce some notation for the equivalence classes of congruence modulo d.

Definition 4.1.5. By $d \cdot R[X]$ we denote the set

$$\{f{\in}R[X] \mid \exists g \,.\, f{=}g{\cdot}d\}\,.$$

The equivalence class $\{f \in R[X] \mid \exists g \,.\, f = a + g \cdot d\}$, containing the polynomial a, is called the *residue class* modulo d of a and is denoted by $a + d \cdot R[X]$. The set of residue classes modulo d is denoted by $R[X]/d \cdot R[X]$. This set is called the *residue class ring* or *quotient ring* modulo d.

Other notations for the residue class modulo d containing the polynomial a are:

- a, when it is clear we mean the residue class,
- or $a + (d)$.

In these notations, naturally, a is the most obvious representative from the residue class $a + d \cdot R[X]$, 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 $R[X]/d \cdot R[X]$ is similar to the notation $\mathbb{Z}/n\mathbb{Z}$ introduced in Proposition 2.1.3.

Example 4.1.6. In $\mathbb{Q}[X]$, the polynomials X^6 and 1 represent the same residue class modulo $X^2 - X + 1$.

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 $a + d \cdot R[X]$ be the class of a modulo d. The proof is divided into two parts. Together they imply the theorem.

There exists a representative of $a + d \cdot R[X]$ 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 Theorem 4.1.7, 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.

$$0, 1, 2, X, 1 + X, 2 + X, 2{\cdot}X, 1 + 2{\cdot}X, 2 + 2{\cdot}X$$

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 $1+X+(X^2+1)\cdot(\mathbb{Z}/3\mathbb{Z})[X]$ 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 $R[X]/d \cdot R[X]$.
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 $a + d \cdot R[X]$ and $b + d \cdot R[X]$ in $R[X]/d \cdot R[X]$ are defined as follows.

- Sum: $(a + d \cdot R[X]) + (b + d \cdot R[X]){=}(a + b) + d \cdot R[X]$;
- Product: $(a + d \cdot R[X]){\cdot}(b + d \cdot R[X]){=}(a{\cdot}b) + d \cdot R[X]$.

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 4.2.2. 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 $\mathbb{Q}[X]/d\cdot\mathbb{Q}[X]$ we find

$$\begin{aligned}(a+d\cdot\mathbb{Q}[X])+(b+d\cdot\mathbb{Q}[X]) &= \\ (a+b)+d\cdot\mathbb{Q}[X] &= \\ X^3+3{\cdot}X^2+1+X^2+2{\cdot}X-1+d\cdot\mathbb{Q}[X] &= \\ X^3+4{\cdot}X^2+2{\cdot}X+d\cdot\mathbb{Q}[X] &= \\ -2{\cdot}X-3+d\cdot\mathbb{Q}[X]\end{aligned}$$

The product modulo d equals

$$\begin{aligned}(a+d\cdot\mathbb{Q}[X])\cdot(b+d\cdot\mathbb{Q}[X]) &= \\ (a{\cdot}b)+d\cdot\mathbb{Q}[X] &= \\ (X^3+3{\cdot}X^2+1)\cdot(X^2+2{\cdot}X-1)+d\cdot\mathbb{Q}[X] &= \\ -1+2{\cdot}X-2{\cdot}X^2+5{\cdot}X^3+5{\cdot}X^4+X^5+d\cdot\mathbb{Q}[X] &= \\ 5+8{\cdot}X+d\cdot\mathbb{Q}[X]\end{aligned}$$

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 $0+d\cdot R[X]$ is called the *zero element* of $R[X]/d\cdot R[X]$ and
- the element $1+d\cdot R[X]$ is called the *unity* or *unit element*.

We often simply denote these elements by 0 and 1, respectively.

Theorem 4.2.3 (Arithmetical Rules)**.** *For arbitrary* $a \in R[X]/d \cdot R[X]$ *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* $b \in R[X]/d \cdot R[X]$ *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 $a \cdot (0 + d \cdot R) = a' \cdot 0 + d \cdot R$ according to the definition of multiplication. The multiplication in R yields $a' \cdot 0 = 0$, so that we find $a' \cdot 0 + d \cdot R = 0 + d \cdot R = d \cdot R = 0$. 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 4.2.4. Let $R{=}\mathbb{Z}/2\mathbb{Z}$ and $d{=}X^3+X+1$. Then the residue class a of X in $R[X]/d\cdot R[X]$ satisfies $a^7{=}1$. Indeed, $X^7-1{=}d{\cdot}(X^4+X^2+X+1)$ in $R[X]$, so

$$\begin{aligned} a^7-1 &= \\ 0{\cdot}(a^4+a^2+a+1) &= \\ 0 & \end{aligned}$$

.

Some more rules are given in the theorem below.

Theorem 4.2.5 (General Arithmetical Rules)**.** *For all a, b, and c in $R[X]/d\cdot R[X]$ 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 4.2.6. 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 $a{\in}\mathbb{R}[X]/(X^2+1)\cdot\mathbb{R}[X]$ be the equivalence class containing the element $(X^3+1)^{27}{\cdot}(X^2+X+1)^{35}$ 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

$$\begin{aligned}(x^3+1)^{27}\cdot(x^2+x+1)^{35} &= \\ (-x+1)^{27}\cdot(-1+x+1)^{35} &= \\ (-x+1)\cdot((-x+1)^2)^{13}\cdot x^{35} &= \\ (-x+1)\cdot(-2\cdot x)^{13}\cdot x^{35} &= \\ (-x+1)\cdot(-2)^{13}\cdot x^{48} &= \\ 2^{13}\cdot x - 2^{13}.\end{aligned}$$

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. The restriction of the residue class map to R is the map

$$j\colon R\to R[X]/d\cdot R[X] : a\mapsto a + d\cdot R[X].$$

Lemma 4.2.7. *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 4.1.7 implies $c=0$.

□

The injectivity of j tells us that within $R[X]/d\cdot R[X]$ 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.

Example 4.2.8. Let $R=\mathbb{R}$, the real numbers, and take $d=X^2+1$. Then the residue class ring $\mathbb{R}[X]/d\cdot\mathbb{R}[X]$ is a description of the complex numbers $\mathbb{C}$, with the role of the complex number i being played by $X+d\cdot\mathbb{R}[X]$. Indeed,

$$\begin{aligned}(X+d\cdot R[X])^2 &= \\ X^2+d\cdot R[X] &= \\ -1+d\cdot R[X].\end{aligned}$$

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 4.2.9. 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.

Let R be a field and d a polynomial of degree $n>0$ in $R[X]$. The residue class ring $R[X]/d\cdot R[X]$ carries a vector space structure as follows.

Theorem 4.2.10. *The residue class ring $S{=}R[X]/d\cdot R[X]$ 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 3.2.7 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 3.2.3, degree(f)$\geq$degree(d), a contradiction as degree(d)=n. This proves that the vectors are linearly independent.

□

Example 4.2.11. Given is the residue class ring $S=(\mathbb{Z}/2\mathbb{Z})[X]/d\cdot(\mathbb{Z}/2\mathbb{Z})[X]$, 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 $R[X]/d\cdot R[X]$ form a subspace $R[X]_{<n}$ of the vector space $R[X]$. A complement is formed by the multiples of d:

Theorem 4.2.12. *The ring $R[X]$ has the following vector space decomposition:*

$$R[X]=R[X]_{<n}+d\cdot R[X].$$

Furthermore, the map

$$R[X]\to R[X]_{<n},\quad f\mapsto \mathrm{rem}(f,d)$$

is the linear projection onto $R[X]_{<n}$ with kernel $d\cdot R[X]$.

Proof. Division with Remainder 3.2.7 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 $f \mapsto \mathrm{rem}(f,d)$ 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 4.2.13. Let $R = \mathbb{Z}/2\mathbb{Z}$ and $d = X^2 + X + 1 \in R[X]$. The matrix of the map

$$R[X]_{<5} \to R[X]/d \cdot R[X], \quad f \mapsto f + d \cdot R[X]$$

with respect to the basis $1, X, X^2, X^3, X^4$ of $R[X]_{<5}$ and the basis $1 + d \cdot R[X], X + d \cdot R[X]$ of $R[X]/d \cdot R[X]$ 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 $f \mapsto \mathrm{rem}(f, X^{n+1})$ 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 $f(x) = a(x) + O(x^{n+1})$ for $x \to 0$.
Recall from Analysis or Calculus that this big Oh notation means that there are positive real constants C and ε such that

$$|f(x) - a(x)| \leq C \cdot |x^{n+1}|$$

for all x with $|x| < \varepsilon$.
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 $\mathrm{rem}(F{\cdot}G, X^{n+1})$ and $\mathrm{rem}(F+G, X^{n+1})$ 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 $x{\mapsto}F(x)$ 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 $g(x){=}G(x) + O(x^{n+1})$ for x going to 0.
Then we have

$$\begin{aligned}
(f{\cdot}g - F{\cdot}G)(x) &= \\
(f{\cdot}g - F{\cdot}g)(x) + (F{\cdot}g - F{\cdot}G)(x) &= \\
(f(x) - F(x)){\cdot}g(x) + F(x){\cdot}(g(x) - G(x)) &= \\
O(x^{n+1}){\cdot}g(x) + F(x){\cdot}O(x^{n+1}) &= \\
O(x^{n+1}) &
\end{aligned}$$

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 $x{\mapsto}e^x$ around 0 is the function $x{\mapsto}1 + x + x^2/2$.
The second-order approximation of the function $x{\mapsto}\sin(x)$ is the function $x{\mapsto}x$.

But then the second order approximation of the product function $x \mapsto \sin(x) \cdot e^x$ equals the function $x \mapsto x + x^2$, which is the remainder of the division of $x \cdot (1 + x + x^2/2)$ by x^2.

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 $\{0, 1, \ldots, n-1\}$. This is similar for polynomials over $\mathbb{Z}/n\mathbb{Z}$. The relation is clarified by the following map.
$I \colon \mathbb{Z}[X]/n \cdot \mathbb{Z}[X] \to (\mathbb{Z}/n\mathbb{Z})[X], \quad a_0 + a_1 \cdot X + \ldots + a_m \cdot X^m + n \cdot \mathbb{Z}[X] \mapsto a_0 + a_1 \cdot X + \ldots + a_m \cdot X^m$.
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 $a = a_0 + a_1 \cdot X + \ldots + a_m \cdot X^m$ and $b = b_0 + b_1 \cdot X + \ldots + b_m \cdot X^m$ 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 $i = 0, 1, \ldots, m$. This implies that $a_i = \mathrm{rem}(b_i, n)$ for $i = 0, 1, \ldots, m$. So our definition does not depend on the representative a or b that we have chosen.

I respects addition.

Suppose that $a{=}a_0 + a_1{\cdot}X + \ldots + a_m{\cdot}X^m$ and $b{=}b_0 + b_1{\cdot}X + \ldots + b_k{\cdot}X^k$ are elements of $\mathbb{Z}[X]$. Then, adding some powers of X, we can assume that $k{=}m$. Now $I(a + b + n \cdot \mathbb{Z}[X])$ equals $(a_0{+}b_0){+}(a_1{+}b_1){\cdot}X{+}\ldots{+}(a_m{+}b_m){\cdot}X^m$ in $(\mathbb{Z}/n\mathbb{Z})[X]$, which is equal to $a_0{+}a_1{\cdot}X{+}\ldots{+}a_m{\cdot}X^m{+}b_0{+}b_1{\cdot}X{+}\ldots{+}b_m{\cdot}X^m$. But the latter is equal to $I(a + n \cdot \mathbb{Z}[X]) + I(b + n \cdot \mathbb{Z}[X])$.

I respects zeros.

Indeed, $I(0 + n \cdot \mathbb{Z}[X]){=}0$.

I respects multiplication.

The proof is similar to the proof of the fact that I respects addition.

I respects units.

Indeed, $I(1 + n \cdot \mathbb{Z}[X]){=}1$.

I is a bijection.

Suppose that a and b are in $\mathbb{Z}[X]$ and satisfy $I(a){=}I(b)$. As I respects addition and scalar multiplication, $I(a - b){=}0$. But then it is straightforward to check that $a - b{=}0$ and hence $a{=}b$.

□

The conclusion of the above result is that the arithmetic in $\mathbb{Z}[X]/n \cdot \mathbb{Z}[X]$ is nothing but the arithmetic in $(\mathbb{Z}/n\mathbb{Z})[X]$. In mathematical jargon: The two arithmetical structures are isomorphic (i.e., equal of form).

Example 4.3.4. The image of $3{+}6{\cdot}X{+}8{\cdot}X^2{+}2{\cdot}X^3{-}88{\cdot}X^4{\in}\mathbb{Z}[X]/5 \cdot \mathbb{Z}[X]$ under the map I of the theorem is $3 + X + 3{\cdot}X^2 + 2{\cdot}X^3 + 2{\cdot}X^4{\in}(\mathbb{Z}/5\mathbb{Z})[X]$.

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 $R[X]$. In the newly constructed arithmetical system $R[X]/d \cdot R[X]$ we have not yet

considered division, since it comes with various complications.

Definition 4.4.1. Suppose that d is a nonconstant polynomial in $R[X]$. Then $f \in R[X]/d \cdot R[X]$ is called *invertible* with respect to multiplication if there exists a $g \in R[X]/d \cdot R[X]$ 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 $R[X]/d \cdot R[X]$ and both g and h are inverses of f. Then

$$\begin{aligned} g &= \\ g \cdot 1 &= \\ g \cdot (f \cdot h) &= \\ (g \cdot f) \cdot h &= \\ 1 \cdot h &= \\ h. & \end{aligned}$$

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 $R[X]$ of positive degree.
The following characterization of the invertible elements in $R[X]/d \cdot R[X]$ yields also a way of computing inverses with the help of the Extended Euclidean Algorithm 3.2.18.

Theorem 4.4.3 (Characterization of Invertibility in Residue Class Rings)**.** *Let a be a polynomial in $R[X]$. Then the residue class $a + d \cdot R[X]$ in $R[X]/d \cdot R[X]$ has an inverse if and only if* $\gcd(a, d) = 1$.

Proof.
If.

If the residue class $a + d \cdot R[X]$ has inverse $b + d \cdot R[X]$, then $a \cdot b = 1 + d \cdot R[X]$.
Hence there is a polynomial p with $a \cdot b + p \cdot d = 1$.
According to the Characterization of Relative Prime Polynomials 3.4.7, $\gcd(a, d) = 1$.

Only if.

If $\gcd(a, d)$=1, then the Extended Euclidean Algorithm 3.2.18 produces polynomials b and p such that $a \cdot b + p \cdot d$=1. But then b represents an inverse of the residue class $a + d \cdot R[X]$.

□

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 $R[X]/d \cdot R[X]$ if and only if the constant term of a differs from 0.

Theorem 4.4.3 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=$R[X]/d \cdot R[X]$ 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 Theorem 4.4.3, 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 $R[X]/d \cdot R[X]$ contains the following four elements: 0, 1, a and a+1, where a=$X + d \cdot R[X]$. The multiplication table for the four elements from $R[X]/d \cdot R[X]$ is as follows:

$\cdot$	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

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 $(\mathbb{Z}/p\mathbb{Z})[X]/d \cdot (\mathbb{Z}/p\mathbb{Z})[X]$ 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. Nevertheless, we state the main result on finite fields.

Theorem 4.5.1. *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 $(\mathbb{Z}/p\mathbb{Z})[X]/d \cdot (\mathbb{Z}/p\mathbb{Z})[X]$ 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 $S = (\mathbb{Z}/3\mathbb{Z})[X]/d \cdot (\mathbb{Z}/3\mathbb{Z})[X]$ 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 Theorem 4.5.1, 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 $(\mathbb{Z}/p\mathbb{Z})[X]/d \cdot (\mathbb{Z}/p\mathbb{Z})[X]$ is a field with exactly p^n elements.*
Moreover, this field is the unique field with p^n elements.

Proof. According to Corollary 4.4.5, the residue class ring $S{=}(\mathbb{Z}/p\mathbb{Z})[X]/d\cdot(\mathbb{Z}/p\mathbb{Z})[X]$ is a field. On the other hand, S is a vector space over $\mathbb{Z}/p\mathbb{Z}$ of dimension n (see the theorem 4.2.10). 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.

□

Example 4.5.4. Let $f{=}X^3+X+1$ be a polynomial in $(\mathbb{Z}/2\mathbb{Z})[X]$. The residue class ring $(\mathbb{Z}/2\mathbb{Z})[X]/f\cdot(\mathbb{Z}/2\mathbb{Z})[X]$ 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+
1	1	a	$1+a$	a^2	a^2+1	a^2+a	a^2+
a	a	a^2	a^2+a	$1+a$	1	a^2+a+1	a^2+
$1+a$	$1+a$	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^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	$1+$

Notice that in each row (and each column) of the table one finds a 1, implying that each element has an inverse. So, $(\mathbb{Z}/2\mathbb{Z})[X]/f\cdot(\mathbb{Z}/2\mathbb{Z})[X]$ 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 $S{=}(\mathbb{Z}/p\mathbb{Z})[X]/d\cdot(\mathbb{Z}/p\mathbb{Z})[X]$.

Theorem 4.5.5. *Write $q{=}p^n$ for the cardinality of S. Then, for each $a,b{\in}S$,*

(a) $a+a+\ldots+a{=}0$ *(with p terms);*

(b) $(a+b)^p{=}a^p+b^p$*;*

(c) $a^q{=}a$ *(Fermat's Little Theorem).*

Proof. We prove the three parts of the theorem separately.

Part 1. $a + a + \ldots + a = 0$ **(with p terms).**

We have $a + a + \ldots + a = (1 + 1 + \ldots + 1) \cdot a = p \cdot a = 0$.

Part 2. $(a + b)^p = a^p + b^p$**.**

Expand $(a + b)^p$ by means of Newton's Binomium. As each binomial coefficient $\binom{p}{i}$ with i different from 0 and p is zero modulo p (see the proof of Fermat's Little Theorem 2.3), 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 2.3.1.
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

$$M_a \colon S^\times \to S^\times, \quad b \mapsto a \cdot b,$$

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 Theorem 4.5.5 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 2.3.1 for finite fields! (Note that the proof does not use the particular construction of the field S.)

Theorem 4.5.5 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 Chapter 2 we met 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. Below you find a few examples where information is secured, so that we can detect and correct possible errors.

CD and DVD

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 music transmitted to the listener by speakers. 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 music has to be stored on disc in such a way that the player can correct its errors.

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.

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 4.6.3. 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,

$$11011001101000111 0011$$

is transformed into

$$11011000110100011110 0111.$$

If at most one mistake occurs in each substring of eight characters, these errors can be detected, but not corrected.

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 4.6.4. 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 4.6.5. We consider the numbers $0, \ldots, 15$ in their Binary Representation 1.6.1, 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 0010011.

The first 4 coordinates form the binary notation of the number. The remaining 3 positions are filled in the following way:

0	0000000
1	0001011
2	0010101
3	0011110
4	0100110
5	0101101
6	0110011
7	0111000
8	1000111
9	1001100
10	1010010
11	1011001
12	1100001
13	1101010
14	1110100
15	1111111

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.

Now we address the real 'coding' aspects.

Definition 4.6.6. 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 δ as defined indeed satisfies the axioms for a distance function with values in $\mathbb{N}$, viz., $\delta(v, w)=0$ if and only if $v=w$, symmetry: $\delta(v, w)=\delta(w, v)$, and the triangle inequality: $\delta(v, w)+\delta(w, u)\geq\delta(v, u)$, where u, v, and w belong to V.

$\delta(v,w)=0$ **if and only if** $v=w$**.**

Clearly, v and w differ in zero positions if and only if they coincide.

Symmetry: $\delta(v,w)=\delta(w,v)$**.**

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: $\delta(v,w)+\delta(w,u)\geq\delta(v,u)$**.**

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, $\delta(v,u)\leq|S\cup T|$. As $|S\cup T|\leq|S|+|T|$, $|S|=\delta(v,w)$, and $|T|=\delta(w,u)$, this implies the triangle inequality.

□

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.

Example 4.6.7. The code in Example 4.6.5 can also be depicted graphically. Let x be a number in $\{0,\ldots,15\}$. 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 $abcdefg$. 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 $[a,b,c,d]_2$.

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 $S=(\mathbb{Z}/p\mathbb{Z})[X]/(X^n-1)\cdot(\mathbb{Z}/p\mathbb{Z})[X]$. This ring has the structure of a vector space over the field $\mathbb{Z}/p\mathbb{Z}$ with basis $1,\ldots,X^{n-1}$, cf. Theorem 4.2.10. So each element of S can be represented by the vector of coefficients with respect to this basis, and vice versa:

$$a=a_0+a_1\cdot X+\ldots+a_{n-1}\cdot X^{n-1}+(X^n-1)\cdot(\mathbb{Z}/p\mathbb{Z})[X]$$

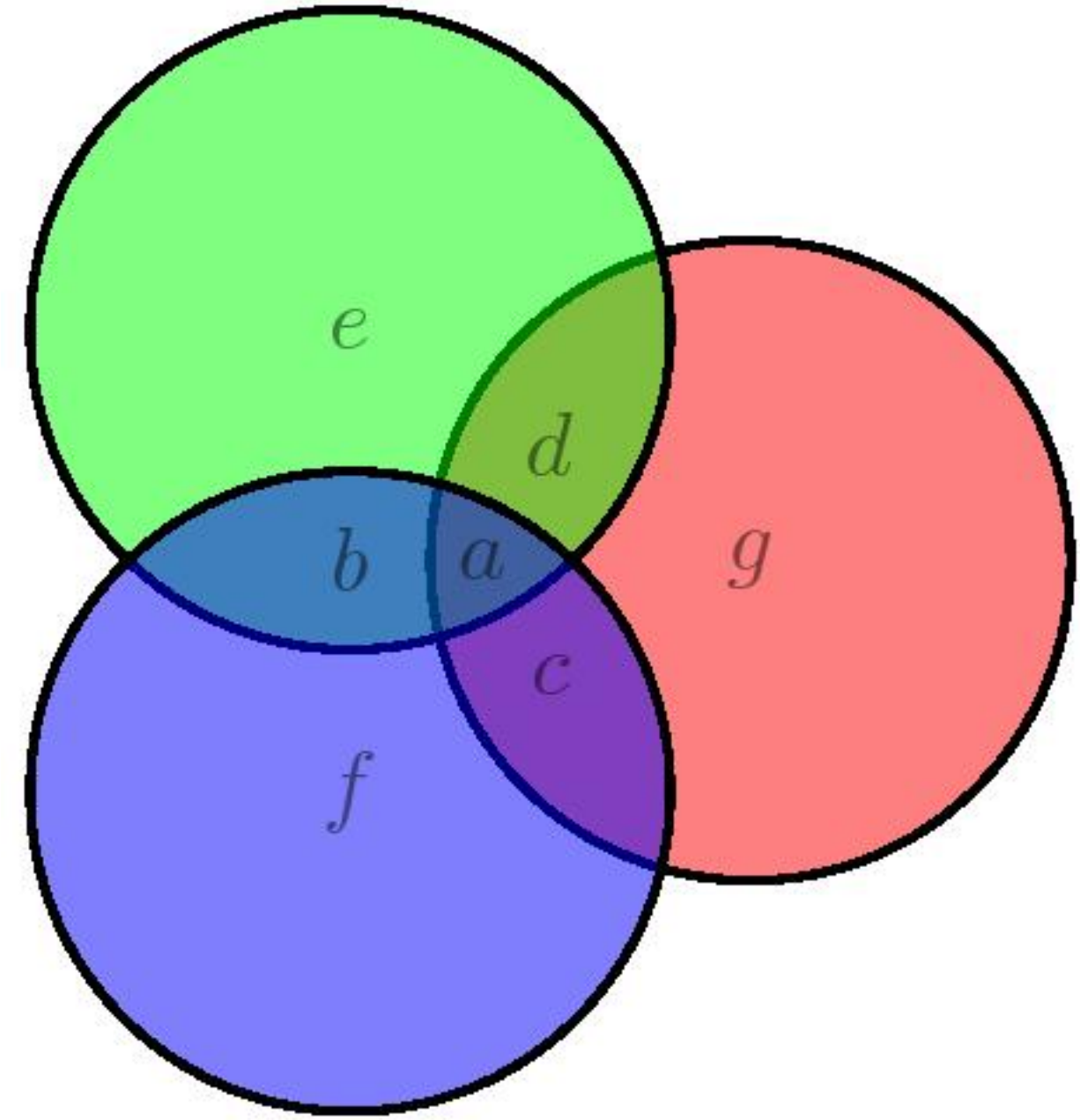

Figure 4.1:

corresponds bijectively to

$$a{=}(a_0, a_1, \ldots, a_{n-1}).$$

The polynomial $X^n - 1$ is reducible for $n > 1$: it is divisible by $X - 1$.

Definition 4.6.8. Let g be a divisor of $X^n - 1$ over $\mathbb{Z}/p\mathbb{Z}$. The image under the linear map $(\mathbb{Z}/p\mathbb{Z})[X]{\to}S, \quad a{\mapsto}a{\cdot}g + (X^n - 1) \cdot (\mathbb{Z}/p\mathbb{Z})[X]$ is called the *cyclic code* of length n generated by g.

Let l be the degree of g and write $k{=}n - l$. The elements $g, \ldots, X^{k-1}{\cdot}g$ form a basis for the image space C of the map from Definition 4.6.8. 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.

Example 4.6.9. 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 4.6.5.

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 4.6.10 (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 $-\mathrm{rem}(c{\cdot}h, X^n)$.*

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 $m{\in}(X^n - 1) \cdot (\mathbb{Z}/p\mathbb{Z})[X]$. By the Degree Formulas 3.2.3, 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:

$$\begin{aligned} c{\cdot}h &= \\ a{\cdot}g{\cdot}h &= \\ a{\cdot}(X^n-1) &= \\ X^n{\cdot}a-a. & \end{aligned}$$

After Division with Remainder 3.2.7, we conclude $-a{=}\mathrm{rem}(c{\cdot}h, X^n)$.

□

Example 4.6.11. 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

$$\begin{aligned} c &= \\ \mathrm{rem}(a{\cdot}g, X^7-1) &= \\ X^5+X^4+X^3+X. & \end{aligned}$$

Since $c{\cdot}h{=}X^8+X$, the polynomial of minimal degree in $c{\cdot}h{+}(X^7)\cdot(\mathbb{Z}/2\mathbb{Z})[X]$ is X, which coincides with a.

Let d and g be polynomials in the polynomial ring $R[X]$. We will consider the residue class ring $S{=}R[X]/d\cdot R[X]$. For an element $s{\in}S$ the substitution of s for X in g gives the element $g(s)$ of S, see the Polynomial Function Definition 3.3.1.
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. Theorem 3.3.7.
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 4.6.12 (BCH bound)**.** *Set $d{=}X^n-1$ and write $S{=}(\mathbb{Z}/p\mathbb{Z})[X]/(X^n-1)\cdot(\mathbb{Z}/p\mathbb{Z})[X]$, 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 $g(a^j){=}0$ contains a sequence of m consecutive integers, then the minimal distance of the (n,k)-code C generated by g is at least $m+1$.

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.

Example 4.6.13. 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 Theorem 4.6.12 can be applied with p=2 and m=2. Note that 3 is also the minimum distance of C.

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.

(a) $a{=}X^3, b{=}1, c{=}X^2+X+1$ as polynomials over $\mathbb{Q}$.

(b) $a{=}X^4+X+2, b{=}X+3, c{=}X+1$ as polynomials over $\mathbb{Z}/5\mathbb{Z}$.

(c) $a{=}(X^3+X+1)^5, b{=}(X^2+2{\cdot}X)^5, c{=}X-1$ over $\mathbb{Q}$.

Hint.
Apply division with remainder to $a-b$ by c.

Solution.

(a) $a{=}X^3, b{=}1, c{=}X^2+X+1$ in $\mathbb{Q}[X]$. Recall that a and b are congruent if their difference is divisible by c. This is the case as $X^3-1{=}(X-1){\cdot}(X^2+X+1)$.

(b) $a{=}X^4+X+2, b{=}X+3, c{=}X+1$ in $(\mathbb{Z}/5\mathbb{Z})[X]$. Now, $a-b$ is divisible by $X+1$ if and only if -1 is a zero of it. Substituting $X{=}-1$, we see that this is indeed the case. Hence a and b are congruent modulo c.

(c) $a{=}(X^3+X+1)^5, b{=}(X^2+2{\cdot}X)^5, c{=}X-1$ in $\mathbb{Q}[X]$. Again it suffices to check that 1 is a zero of $a-b$. Since $3^5-3^5{=}0$, they are congruent modulo c.

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.

(a) $a{=}X^4, d{=}X^2+X+1$ in $\mathbb{Q}[X]$,

(b) $a{=}X^4+X^2+1, d{=}X^2+X+1$ in $(\mathbb{Z}/2\mathbb{Z})[X]$.

Hint.
Divide a by d and determine the remainder.

Solution.
Polynomial division gives:

(a) $a{=}X^4{=}(X^2+X+1){\cdot}(X^2-X)+X$. Hence X is the representative of the residue class of a modulo d of smallest degree.

(b) $a{=}X^4+X^2+1{=}(X^2+X+1)^2{=}d^2$. Hence 0 is the representative of $\mathrm{rem}(a,d)$ of smallest degree.

Exercise 4.7.3. Determine representatives for all congruence classes for each of the following residue class rings.

(a) $(\mathbb{Z}/2\mathbb{Z})[X]/(X^3+1)\cdot(\mathbb{Z}/2\mathbb{Z})[X]$,

(b) $\mathbb{Q}[X]/(X-1)\cdot\mathbb{Q}[X]$,

(c) $\mathbb{R}[X]/2\cdot\mathbb{R}[X]$.

Hint.
Use division with remainder. What are the possible remainders?

Solution.
Representatives are those polynomials of degree less than the one with respect to which the residue class ring is formed. Hence, a complete set of representatives is as follows.

(a) For $(\mathbb{Z}/2\mathbb{Z})[X]/\big(X^3+1\big)\cdot(\mathbb{Z}/2\mathbb{Z})[X]$: All polynomials of degree less than 3, that is $\{0,1,X,X+1,X^2,X^2+1,X^2+X,X^2+X+1\}$.

(b) For $\mathbb{Q}[X]/(X-1)\cdot\mathbb{Q}[X]$: All constants, that is, the set $\mathbb{Q}$.

(c) For $\mathbb{R}[X]/2\cdot\mathbb{R}[X]$: All polynomials of degree less than 0, that is, only the polynomial 0.

4.7.2 The residue class ring

Exercise 4.7.4. Consider the residue class a of X in $S{=}(\mathbb{Z}/2\mathbb{Z})[X]/(X^2+X+1)\cdot(\mathbb{Z}/2\mathbb{Z})[X]$

(a) Describe the elements of S in terms of 'polynomials' in a.

(b) Compose a multiplication table for S.

(c) Show that $a^{17}{=}a+1$.

Hint.
Express all powers of a as linear combinations of a^2, a, and 1.

Solution.

(a) Since S is a vector space over $\mathbb{Z}/2\mathbb{Z}$ of dimension 2, it has 4 elements. The elements $1, a, a^2{=}a+1$ are distinct, so, together with 0, they are all elements of S.

(b) Notice that $a^3 = a^{2+1} = a{\cdot}(a{+}1) = a^2{+}a = 1$. Thus we obtain the following table for the nonzero elements of S.

$\cdot$	1	a	$a+1$
1	1	a	$a+1$
a	a	$a+1$	1
$a+1$	$a+1$	1	a

(c) $a^{17}{=}a^{15+2}{=}a^2{=}a+1$.

Exercise 4.7.5. Let $a{\in}\mathbb{R}$. We define the map $eval\colon \mathbb{R}[X]/(X-a)\cdot\mathbb{R}[X]{\to}\mathbb{R}$ by $eval(f+(X-a)\cdot\mathbb{R}[X]){=}f(a)$.

(a) Show that this map is well defined.

(b) Show that eval is a bijection.

Hint.

(a) Show that the definition does not depend on the representative chosen.

(b) Show that the representative of f of minimal degree is uniquely determined by $f(a)$.

Solution.

(a) In order to show that the map is well defined we prove that the definition of *eval* does not depend on the choice of a representative in a residue class modulo $X-a$. Suppose that f and g are congruent. Then $f-g$ is a multiple of $X-a$ and so $f(a)-g(a)=0$. We conclude that, if f and g are congruent modulo $X-a$, then $f(a)$ is equal to $g(a)$.

(b) A bijection is a mapping that is both injective and surjective. The latter condition is easy since for any element $z\in\mathbb{R}$ we have eval $eval(z+(X-a)\cdot\mathbb{R}[x])=z$. It remains to show that the mapping *eval* is injective. Let f and g be representatives of (possibly different) classes modulo $X-a$ and suppose that *eval* takes the same value on these classes, that is $f(a)=g(a)$. This implies that $f-g$ has a as a root, and hence, by Theorem 3.3.7, is a multiple of $X-a$. So f and g are congruent modulo $X-a$. This shows that the mapping *eval* is indeed injective.

Exercise 4.7.6. We define the two maps f_+ and f_- from $\mathbb{Q}[X]/(X^2-2)\cdot\mathbb{Q}[X]$ to $\mathbb{Q}+\mathbb{Q}\cdot\sqrt{2}$ in the following way. For any residue class $g+(X^2-2)\cdot\mathbb{Q}[X]$ we have

$$f_+(g+\left(X^2-2\right)\cdot\mathbb{Q}[X])=g(\sqrt{2})$$

and

$$f_-(g+\left(X^2-2\right)\cdot\mathbb{Q}[X])=g(-\sqrt{2}).$$

(a) 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.

(b) Show that f_+ and f_- are both injective.

(c) Show that both f_+ and f_- are both surjective.

(d) Show that, for all a, b in $\mathbb{Q}[X]/(X^2-2)\cdot\mathbb{Q}[X]$,

$$f_+(a+b)=f_+(a)+f_+(b),$$

$$f_+(a\cdot b)=f_+(a)\cdot f_+(b),$$

$$f_-(a+b)=f_-(a)+f_-(b),$$

$$f_-(a\cdot b)=f_-(a)\cdot f_-(b).$$

Both maps give a way to associate the residue class ring $\mathbb{Q}[X]/(X^2-2)\cdot\mathbb{Q}[X]$ to $\mathbb{Q}+\mathbb{Q}\cdot\sqrt{2}$.

Hint.
Compare the map with *eval* of Exercise 4.7.5.

Solution.
We will only do the proofs for $+$ since the proofs for $-$ are similar.

(a) We show that $+$ is well defined. Let a and b be two polynomials that are congruent modulo X^2-2. We show that $f_+(a)=f_+(b)$. Since a and b are congruent, their difference is divisible by X^2-2, that is $a-b=(X^2-2)\cdot q$ for some polynomial q. Substituting $\sqrt{2}$ yields

$$\begin{aligned} a(\sqrt{2})-b(\sqrt{2}) &= \\ ((\sqrt{2})^2-2)\cdot q(\sqrt{2}) &= \\ 0. & \end{aligned}$$

So indeed, $f_+(a)=f_+(b)$.

(b) We show that f_+ is injective. Let a and b be polynomials in $\mathbb{Q}[X]$. We are interested in their classes modulo X^2-2. We may assume that both a and b are of degree at most one. In particular, $a=a_0+a_1\cdot X$ and $b=b_0+b_1\cdot X$, for some rational numbers a_0, a_1, b_0, and b_1. Suppose that $f_+(a)=f_+(b)$. Then

$$a_0+a_1\cdot\sqrt{2}=b_0+b_1\cdot\sqrt{2}$$

and so

$$a_0-b_0=\sqrt{2}\cdot(b_1-a_1).$$

But the left hand side is rational and the right hand side is zero or irrational. Hence the only possible way they can be equal is when they are both zero. This implies $a=b$, and injectivity of f_+ follows.

(c) For each pair a_0, a_1 of rational numbers, we have $f_+(a_0+a_1\cdot X+\left(X^2-2\right)\cdot\mathbb{Q}[X])=a_0+a_1\cdot\sqrt{2}$ Therefore, the map f_+ is surjective.

(d) We show the multiplicative law $f_+(a\cdot b)=f_+(a)\cdot f_+(b)$ and leave the remaining parts to the reader.

Let a and b be two polynomials in $\mathbb{Q}[X]$. We are only interested in their classes modulo X^2-2 and so we may assume that both a and b are of degree at most one. In particular, $a=a_0+a_1\cdot X$ and $b=b_0+b_1\cdot X$, for some rational

numbers a_0, a_1, b_0, and b_1. Then

$$\begin{aligned} f_+(a\cdot b) &= \\ f_+((a_0+a_1\cdot X)\cdot(b_0+b_1\cdot X)) &= \\ f_+(a_0\cdot b_0+(a_1\cdot b_0+a_0\cdot b_1)\cdot X+a_1\cdot b_1\cdot X^2) &= \\ a_0\cdot b_0+(a_1\cdot b_0+a_0\cdot b_1)\cdot\sqrt{2}+a_1\cdot b_1\cdot 2 &= \\ (a_0+a_1\cdot\sqrt{2})\cdot(b_0+b_1\cdot\sqrt{2}) &= \\ f_+(a)\cdot f_+(b). \end{aligned}$$

Exercise 4.7.7. Find the representative of degree less than 5 of the residue class of $(1+X)\cdot(1+X^3)\cdot(1+X^4)\cdot(1+X^5)$ in $(\mathbb{Z}/2\mathbb{Z})[X]/(X^5)\cdot(\mathbb{Z}/2\mathbb{Z})[X]$.

Hint.
Rather than expanding or performing division with remainder, you should work in the residue class ring. This means rewriting the expressions using the relation $a{=}0$, where a is the residue class of X.

Solution.
In the residue class ring $(\mathbb{Z}/2\mathbb{Z})[X]/\big(X^5\big)\cdot(\mathbb{Z}/2\mathbb{Z})[X]$, the element X^5 is equal to 0, so the product immediately reduces to $(1+X)\cdot(1+X^3)\cdot(1+X^4)$. Moreover, in the expansion of this product, the terms $X\cdot 1\cdot X^4, 1\cdot X^3\cdot X^4$, and $X\cdot X^3\cdot X^4$ vanish for the same reason. Therefore, the product consists of $2^3{=}8$ terms, 5 of which have degree less than 5. Two terms are equal to X^4 and cancel, leaving the following representative of degree less than five: $1+X+X^3$.

Exercise 4.7.8. The polynomial f in $\mathbb{Q}[X]$ satisfies the relation

$$(X^3+1)\cdot f+a\cdot(X^2+1){=}X^3-1$$

for some polynomial a in $\mathbb{Q}[X]$. Determine the remainder upon division of f by X^2+1.

Hint.
Calculate modulo X^2+1 and find a degree one polynomial which is congruent to f modulo X^2+1.

Solution.
Denote by x the residue class of X modulo X^2+1. Reduction modulo X^2+1 of the given relation yields the equation

$$(-x+1)\cdot f{=}-x-1$$

in $\mathbb{Q}[X]/\big(X^2+1\big)\cdot\mathbb{Q}[X]$.

Multiplying the left hand side with the class of $(x+1)/2$ yields

$$\begin{aligned}\frac{x+1}{2}\cdot(-x+1)\cdot f &= \\ \frac{1}{2}\cdot(1-x^2)\cdot f &= \\ \frac{1}{2}\cdot 2\cdot f.\end{aligned}$$

For the right hand side we obtain

$$\begin{aligned}\frac{x+1}{2}\cdot(-x-1) &= \\ \frac{1}{2}\cdot(x^2+2\cdot x+1) &= \\ -x.\end{aligned}$$

Therefore, f is congruent to $-X$ modulo X^2+1. This implies of course that the remainder of f upon division by X^2+1 equals $-X$.

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.

Hint.
Have a look at the proof of the Chinese Remainder Theorem 2.2.4 for integers.

Solution.
The proof is almost identical to the proof of the Chinese Remainder Theorem 2.2.4 for integers.
Since c and d have gcd equal to 1, there are x and y with $x{\cdot}c + y{\cdot}d{=}1$, cf. Theorem 3.2.23. Multiplying both sides with $b-a$ yields $x{\cdot}c{\cdot}(b-a) + y{\cdot}d{\cdot}(b-a){=}b-a$. Now consider $f = a - x{\cdot}c{\cdot}(a-b) = b - y{\cdot}d{\cdot}(a-b)$. The polynomial f satisfies the relations $f \equiv a \pmod{c}$ and $f \equiv b \pmod{c}$. Replacing f by its remainder upon division by $a{\cdot}b$ yields a polynomial still satisfying the relations and also having a degree smaller than $n{\cdot}m$. This proves the existence.
If both f and g are polynomials of degree less than $n{\cdot}m$ satisfying both relations, then $f-g$ is a polynomial of degree less than $n{\cdot}m$ and equal to 0 modulo c and d, and as c and d are relatively prime, also modulo $c{\cdot}d$, cf. Theorem 3.4.8. However, there is only one such representative of the residue class of rem(0) modulo $c{\cdot}d$, namely 0. Hence $f{=}g$.

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 Exercise 4.7.9.

Hint.
Have a look at Exercise 4.7.9.

Solution.
Use the Extended Euclidean Algorithm 3.2.18 to find x and y in the equation $x{\cdot}c + y{\cdot}d{=}1$.
Define f to be $\mathrm{rem}(a + x{\cdot}c{\cdot}(b - a), c{\cdot}d)$.

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.

(a) $\frac{1}{1+x}$

(b) $\frac{1}{1+x+x^2}$

(c) $\frac{1}{\cos(x)}$

Hint.
Use the Extended Euclidean Algorithm 3.2.18 to find inverses of the denominators modulo x^4. For $\cos(x)$, use the Taylor expansion $1 + \frac{1}{2}{\cdot}x^2$.

Solution.

(a) Using the Extended Euclidean Algorithm 3.2.18, we find that the inverse of $1 + X$ modulo X^4 is $1 - X + X^2 - X^3$. Hence, the Taylor expansion is $1 - x + x^2 - x^3$.

(b) Similarly, the inverse of $\mathrm{rem}(1 + X + X^2, X^4)$ is $1 - X + X^3$, so the Taylor expansion is $1 - x + x^3$.

(c) The Taylor expansion of order 4 of the function cos equals $1 - \frac{1}{2}{\cdot}x^2$. The inverse of that polynomial modulo x^4 equals $1 + \frac{1}{2}{\cdot}x^2$.

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.

(a) $\frac{1}{1-x}$

(b) $\frac{1}{1-x+x^3}$

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 $(\mathbb{Z}/3\mathbb{Z})[X]/(X^2+1)$.
Solve the following equation for z: $a{\cdot}z{=}b$.

Hint.
Compute the inverse of a and multiply both sides of the equation with it.

Solution.
Since $1{+}X$ and $X^2{+}1$ are relatively prime, the element a has an inverse, which we compute using the Extended Euclidean Algorithm 3.2.18: $(X-1){\cdot}(X+1)-1{\cdot}(X^2+1){=}1$. The inverse of a is therefore the class of $X-1$. Multiplying both sides of the equation $a{\cdot}z{=}b$ with the class of $X-1$ and simplifying (modulo X^2+1) yields

$$z \equiv (1+2{\cdot}X){\cdot}(X-1) \equiv 2{\cdot}X^2 - X - 1 \equiv 2{\cdot}X \pmod{X^2+1}.$$

Therefore $z{=}2{\cdot}X + \big(X^2+1\big)$.

Exercise 4.7.14. Consider the element $a{=}X + (X^3+X+1)\cdot\mathbb{Q}[X]$ in $\mathbb{Q}[X]/(X^2+X+1)\cdot$

(a) Show that $X^3{+}X{+}1$ is irreducible in $\mathbb{Q}[X]$. Conclude that $\mathbb{Q}[X]/(X^2+X+1)\cdot\mathbb{Q}[X]$ is a field.

(b) Write $\frac{1}{a}$ as $p+q{\cdot}a+r{\cdot}a^2$ with $p,q,r{\in}\mathbb{Q}$.

(c) Write $\frac{1}{a+2}$ as $p+q{\cdot}a+r{\cdot}a^2$ with $p,q,r{\in}\mathbb{Q}$.

(d) Same question for $\frac{1}{a^2+a+1}$.

Hint.
Use the Extended Euclidean Algorithm 3.2.18 to find inverses.

Solution.

(a) The polynomial $X^3{+}X{+}1$ has no zeros in $\mathbb{Q}$. By Theorem 3.3.7, this implies that it has no linear factors. As its degree is less than 4, it is irreducible.

(b) Since $a{\cdot}(a^2+1){=}{-}1$, the inverse of a is $-1-a^2$.

(c) By the Extended Euclidean Algorithm 3.2.18 we have $\frac{1}{9}{\cdot}(X^2-2{\cdot}X+5){\cdot}(X+2)+\frac{-1}{9}{\cdot}(X^3+X$ This shows that $\frac{5}{9}-\frac{2}{9}{\cdot}a+\frac{1}{9}{\cdot}a^2$ is the inverse of $a+2$.

(d) Similarly to the previous part, the Extended Euclidean Algorithm 3.2.18 gives $3{\cdot}(X^2-2{\cdot}X+2){\cdot}(X^2+X+1)+3{\cdot}(1-X){\cdot}(X^3+X+1){=}1$, and so $3{\cdot}(a^2-2{\cdot}a+2)$ is the inverse of 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:

(a) If $f|d$, then f is invertible in $R[X]/d \cdot R[X]$.

(b) If the degree of d is larger than 1 and $R{=}\mathbb{Z}$, then $R[X]/d \cdot R[X]$ is infinite.

(c) If a and b are elements from $R[X]/d \cdot R[X]$ with $a{\cdot}b{=}0$, but a nor b are equal to 0, then both a and b are not invertible.

(d) If a, b, and c are elements from $R[X]/d \cdot R[X]$ with $a{\cdot}b{=}a{\cdot}c$, then $b{=}c$.

(e) If a, b, and c are elements from $R[X]/d \cdot R[X]$ with $a{\cdot}b{=}a{\cdot}c$ and a is invertible, then $b{=}c$.

(f) If $a^4{=}0$ for some element a in $R[X]/d \cdot R[X]$, then $1 - a$ is invertible.

Hint.
Do all elements have an inverse?

Solution.

(a) False. Take $f{=}X$ and $d{=}X^2$.

(b) True. Every element of $\mathbb{Z}$ represents a different class. This can be shown similarly to the proof of Theorem 4.2.7.

(c) True. Suppose that a has an inverse. Then, multiplying the equation $a{\cdot}b{=}0$ by a^{-1} from the left gives $b{=}0$, contradicting that b is nonzero. Hence a is not invertible. The proof that b is not invertible is similar.

(d) False. Of course, $a{=}0$ (and any choice for the other variables) gives a counterexample. But even if a is not zero it is false; for example, $X + 1$ is not equal to 1 but $X{\cdot}(X + 1){=}X{\cdot}1$ if we compute modulo X^2.

(e) True. Multiply with the inverse of a to prove it.

(f) True. The identity $(1 - a){\cdot}(1 + a + a^2 + a^3){=}1 - a^4{=}1$ shows that $1 - a$ 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 $R{\to}R[X]/d \cdot R[X]$, $\quad a{\mapsto}a + d \cdot R[X]$ is bijective. Prove this.

Hint.
Use division with remainder.

Solution.
Division with remainder shows that the map is surjective. To show that the map is injective, consider two elements a and b of R.
Their images are equal if they differ by a multiple of d. Since the degree of d is 1, this can only happen if this multiple is 0, i.e., if $a{=}b$.

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.

(a) Let $f, g{\in}K[X]$ with f irreducible and let a be the class of X in $K[X]/f \cdot K[X]$. Show that $f|g$ if and only if a is a zero of g, where we view g as a polynomial with coefficients in $K[X]/f \cdot K[X]$.

(b) 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.

Hint.

(a) Translate $g(a){=}0$ into congruences in the polynomial ring $K[X]$. yyy

(b) Substitute a for X in $X^6 - X^3 + 1$ and use $a^2 + a + 1{=}0$.

Solution.

(a) $g(a){=}0$ means that $g(X) \equiv 0 \pmod{f}$. That is just saying that f divides g.

(b) The class a satisfies $a^2 + a + 1{=}0$. Now substitute in $X^6 - X^3 + 1$:

$$\begin{aligned} a^6 - a^3 + 1 &= \\ a^3{\cdot}(a^3 - 1) + 1 &= \\ a^3{\cdot}(a-1){\cdot}(a^2 + a + 1) + 1 &= \\ 1. & \end{aligned}$$

So f does not divide g.

Exercise 4.7.18. Let R be a ring. A polynomial in $R[X]$ is called *monic* if its leading coefficient equals 1.

(a) If d is a monic polynomial in $R[X]$ of positive degree n, then each residue class in $R[X]/d \cdot R[X]$ contains an element of degree smaller than n. Prove this.

(b) Let R be equal to $\mathbb{Z}/4\mathbb{Z}$ and d the polynomial $2{\cdot}X$. Verify that the class of X in $R[X]/d \cdot R[X]$ does not contain an element of degree 0.

Hint.

(a) Use induction on the degree.

(b) Consider elements of the form $2{\cdot}X{\cdot}f$ for some polynomial f.

Solution.

(a) Let f be an arbitrary polynomial over R of degree m. If $m < n$, then it obviously has a representative modulo d of degree less than n. We proceed by induction on m. So suppose that the assertion holds for all polynomials of degree less than $m > n-1$ and let f be a polynomial of degree m. If its leading coefficient is a, then $f - a{\cdot}d{\cdot}X^{m-n}$ is congruent to $\mathrm{rem}(f,d)$ and of degree less than m. Hence, by induction, it has a representative of degree less than n. This then, is also a representative of $\mathrm{rem}(f,d)$.

(b) If X has a representative of degree 0, then there exists a $c{\in}\mathbb{Z}/4\mathbb{Z}$ such that $2{\cdot}X$ divides $X - c$. However all coefficients of a multiple of $2{\cdot}X$ are 0 or 2, while the coefficient of X in $X - c$ is 1. This is a contradiction. Hence X does not have a representative of degree less than 1.

4.7.5 Finite fields

Exercise 4.7.19. Let $d{=}X^4 + X + 1{\in}(\mathbb{Z}/2\mathbb{Z})[X]$ and write $S{=}(\mathbb{Z}/2\mathbb{Z})[X]/d \cdot (\mathbb{Z}/2\mathbb{Z})[X]$.

(a) Prove that d is irreducible.

(b) Determine the addition and multiplication table for the field S.

(c) 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.

Hint.

(a) Consider possible irreducible factors of degrees 1 and 2.

(b) How would you represent each element of S?

(c) Write down all powers of $X + d \cdot (\mathbb{Z}/2\mathbb{Z})[X]$ until you reach 1. Then answer the questions.

Solution.

(a) The polynomial d has no zeros. Thus, if it is reducible, then it is the product of two irreducible polynomials of degree 2. However, the only irreducible polynomial of degree 2 is $p{=}X^2+X+1$. But d does not equal p^2.

(b) The field S is a vector space of dimension 4 over $\mathbb{Z}/2\mathbb{Z}$. Therefore, S has $2^4{=}16$ elements. Let a be the element $X + d\cdot(\mathbb{Z}/2\mathbb{Z})[X]$ of S. We write down some powers of a: a, a^2, a^3, $a^4{=}a+1$, $a^5{=}a^2+a$, $a^6{=}a^3+a^2$, $a^7{=}a^3+a+1$, $a^8{=}a^2+1$, $a^9{=}a^3+a$, $a^{10}{=}a^2+a+1$, $a^{11}{=}a^3+a^2+a$, $a^{12}{=}a^3+a^2+a+1$, $a^{13}{=}a^3+a^2+1$, $a^{14}{=}a^3+1$, $a^{15}{=}1$.

We have found all 15 nonzero elements of S as powers of a. From this it is easy to write down an addition and multiplication table.

(c) A subfield of order 4 contains $0, 1$, and at least one element of the form a^i (with i between 0 and 15). Then it will also contain $a^{2\cdot i}, a^{3\cdot i}$, etc. If the subfield is to have only 4 elements, then we conclude from the table that $i{=}5$ or 10. It is easily checked that $\{0, 1, a^5, a^{10}\}$ is indeed a subfield.

Exercise 4.7.20. Let $K{=}(\mathbb{Z}/2\mathbb{Z})[X]/d\cdot(\mathbb{Z}/2\mathbb{Z})[X]$, where $d{=}X^3+X+1$ and let a be the class of X modulo d.

(a) 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.

(b) Show that $(X^3+X+1)|(X^7+1)$ and that $K{=}\{0, 1, a, a^2, a^3, a^4, a^5, a^6\}$.

(c) The element a is a zero of X^3+X+1 (viewed as polynomial in K). Express all zeros as powers of a.

(d) Find the zeros of X^3+X^2+1.

Hint.

(a) For establishing the irreducibility of X^3+X+1 it suffices to show that this polynomial has no zeros in $\mathbb{Z}/2\mathbb{Z}$. Why?

(b) Note that $a^7{=}1$ and prove that $1, a, a^2, \ldots, a^6$ are distinct.

(c) Square the equality $a^3+a+1{=}0$ or simply try the elements mentioned in the previous item.

(d) Divide X^7+1 by $X+1$ and by X^3+X^2+1.

Solution.

(a) Since X^3+X+1 has degree 3, it suffices to check that X^3+X+1 has no zeros in $\mathbb{Z}/2\mathbb{Z}$ (reducibility of X^3+X+1 is equivalent to having a degree one factor and that is equivalent to having a zero). Substituting 0 and 1 for X in X^3+X+1 leads to the value 1. So X^3+X+1 is an irreducible polynomial. By Theorem 4.2.10, the number of elements in the quotient ring is 2^3=8.

(b) Notice that

$$X^7+1=(X^3+X+1)\cdot(X^4+X^2+X+1).$$

Since $a^3+a+1=0$, this implies that $a^7=1$. It suffices to show that $1, a, a^2, \ldots, a^6$ are all distinct. Suppose $a^i=a^j$ for $0\leq i<j\leq 6$. Cancelling powers of a we reduce to considering the relation $1=a^j$ for $0<j<7$. If $a^j-1=0$, then $(X^3+X+1)|(X^j-1)$. This is clearly impossible if $j\leq 3$. If j is at least 4, then, by multiplying both sides of $1=a^j$ with a^{7-j}, we find again a relation $1=a^i$ with $0<i\leq 3$ (here we use that $a^7=1$); this is impossible.

(c) a^2 is a zero:

$$(a^2)^3+a^2+1=(a^3+a)^2+1=1+1=0.$$

a^4 is a zero:

$$(a^4)^3+a^4+1=(a^3+a)^4+1=1+1=0.$$

(d) The zeros of X^3+X^2+1 are the elements a^3, a^5, a^6 as is easily verified. Note that $1, a, a^2, a^3, a^4, a^5, a^6$ are precisely the zeros of X^7+1 and that

$$X^7+1=(X+1)\cdot(X^3+X^2+1)\cdot(X^3+X+1).$$

Exercise 4.7.21. Let $d=X^3+X+1\in(\mathbb{Z}/2\mathbb{Z})[X]$ and write $S=(\mathbb{Z}/2\mathbb{Z})[X]/d\cdot(\mathbb{Z}/2\mathbb{Z})[X]$.

(a) Prove that d is irreducible and conclude that S is a field.

(b) Show that each nonzero element of S is a power of $X+d\cdot(\mathbb{Z}/2\mathbb{Z})[X]$.

Solution.

(a) Since d has no zero's and is of degree 3, it is irreducible. As a consequence S is a field with 8 elements.

(b) If we denote the element $X+d\cdot(\mathbb{Z}/2\mathbb{Z})[X]$ by a, then we see that each nonzero element of S is a power of a. Indeed, the nonzero elements of S are: a, a^2, $a^3=a+1$, $a^4=a^2+a$, $a^5=a^2+a+1$, $a^6=a^2+1$, and $a^7=1$.

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

$$\frac{2^n}{\sum_{i=0}^{e} \binom{n}{i}}$$

codewords.

Solution.
Since two code words are at distance at least d, each word at distance at most e from a code word c is at distance at least $e + 1$ from any other codeword.
As there are exactly $\sum_{i=0}^{e} \binom{n}{i}$ words at distance at most e from a fixed code word, we see that there are at least $|C| \cdot \sum_{i=0}^{e} \binom{n}{i}$ different words in $(\mathbb{Z}/2\mathbb{Z})^n$. This number is at most the total number of words in $(\mathbb{Z}/2\mathbb{Z})^n$, which is 2^n. The result follows immediately from this inequality.

4.8 Summary

Analogously to the arithmetic modulo an integer n in $\mathbb{Z}$, we have the arithmetic modulo a polynomial d in a polynomial ring $R[X]$. This leads to new arithmetical systems. The topics discussed in this chapter are:

- congruence modulo a polynomial, residue classes
- the construction of the residue class ring $R[X]/d \cdot R[X]$
- addition, subtraction (opposite), multiplication in $R[X]/d \cdot R[X]$
- relation between computation modulo X^n and computing with approximations up to order n in analysis.

When R is a field, further topics are:

- invertibility in $R[X]/d \cdot R[X]$
- the role of Euclid's Algorithm for Polynomials 3.2.16

- the R-vector space structure of $R[X]/d \cdot R[X]$
- multiplication by an element from this ring defines a linear map
- $R[X]/d \cdot R[X]$ is a field if and only if d is an irreducible polynomial
- construction of finite fields

As an application of the construction of finite fields, error-correcting codes are discussed; these codes are used for safe transport of data over (electronic) communication lines.

Chapter 5

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.

We are mainly concerned with bijections of finite set X to itself. Often we work with the set X of integers from 1 to n, thus $X = Y = \{1, \ldots, n\}$. 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.

Example 5.1.2. We give two well known examples:
The exponential function

The function $f\colon \mathbb{R}\to\mathbb{R}$ given by $f(x)=e^x$ is injective.
Namely, if $f(x)=f(y)$, then $e^x=e^y$ and thus $e^{x-y}=1$. 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.

A quadratic function

Let P denote the set of positive real numbers. The function $f\colon P\to P$ 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.

Let X be a set. We introduce permutations of X and describe multiplication of permutations as composition of maps.

Definition 5.1.3.

- 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 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 $X=\{1,\ldots,n\}$, we also write Sym_n instead of $\mathrm{Sym}(X)$. Furthermore, a permutation f of X is often given by $[f(1),f(2),\ldots,f(n)]$.

The product of two permutations in Sym_n is again a permutation and hence an element of Sym_n. (Prove this!)
The identity map $id\colon X\to X$ plays a special role: $g=g\cdot id$ and $g=id\cdot g$, for all $g\in S$ in Sym_n. The inverse of $g\in S$, denoted by g^{-1}, is again a permutation and satisfies $g^{-1}\cdot g=id$ and $g\cdot g^{-1}=id$. We call id the identity element for the product on Sym_n. We often use 1 to denote the identity element. For every positive integer m, we denote by g^m the product of m factors g. Instead of $(g^{-1})^m$ we also write g^{-m}.
We call Sym_n the symmetric group of degree n.

Example 5.1.4. Let g and h be the permutations of $\{1,\ldots,4\}$ 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}=$[3,1,2,4]$.
We will shortly describe notations for permutations that are more convenient for our purposes than the lists we have seen so far: matrices and disjoint cycles.

Remark 5.1.5. Sometimes the product $g{\cdot}h$ 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 $x{\in}X$, and f a map write x^f for the image of x under f. 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 can be described in matrix notation by a 2 by n matrix with the numbers $1,\ldots,n$ in the first row and the images of $1,2,\ldots,n$ (in that order) in the second row. Since there are $n!$ possibilities to fill the second row, the following theorem holds.

Theorem 5.1.6. Sym_n *has exactly* $n!$ *elements.*

The first row of the 2 by n matrix describing a permutation in Sym_n, is always $1,2,\ldots,n$ and hence yields no essential information. Therefore, we often omit the first row; the permutation is then given in list notation. For example, $\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix}$ becomes $[3,1,2]$ in list notation.
Nevertheless, the matrix notation is useful for calculating products and inverses.

- Product: To calculate $g{\cdot}h$ for two permutations g, h in Sym_n, we first look up, for each $i{\in}\{1,\ldots,n\}$, 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 $1, 2, \ldots, n$.

Example 5.1.7. Sym_3 has the following 6 elements: $\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}$, $\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}$, $\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}$, $\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}$, $\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix}$, $\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}$.
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:
$[1,2,3]$, $[1,3,2]$, $[2,1,3]$, $[2,3,1]$, $[3,1,2]$, $[3,2,1]$.

Example 5.1.8.

Definition 5.1.9. The order of a permutation g is the smallest positive integer m such that $g^m{=}e$.

Example 5.1.10.
- The order of the identity is 1.
- The order of the permutation $[2,1,3]$ (in list notation) in Sym_3 is 2.
- The order of the permutation $g{=}[2,3,4,1]$ (in list notation) in 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 Sym_n, then the permutations $g, g^2, g^3, \ldots$ can not all be distinct, because there are only finitely many permutations in 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.

5.2 Cycles

Let g be a permutation of S_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. Notation: $\mathrm{fix}(g,X)$.
- The support of g is the complement in X of $\mathrm{fix}(g,X)$. Notation: $support(g)$.
- 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. We observe the following:

- The fixed points of a permutation g form the set $\mathrm{fix}(g)=\{x\in X \mid g(x)=x\}$. The notation refers to the verb 'to fix'.
- The support of g equals $\{x\in X \mid g(x)\neq x\}$. 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 $support(g)=\{a_1,\ldots,a_m\}$, where the a_i are pairwise distinct. We say g is an m-cycle if $g(a_i)=g(a_{i+1})$ for all $i\in\{1,\ldots,m-1\}$ and $g(a_m)=a_1$. For such a cycle g we also use the cycle notation $(a_1,\ldots,a_m)$).
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 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 $\mathrm{Sym}(aritha.minus(n,1))$ with the part of $\mathrm{Sym}(n)$ consisting of all permutations fixing n : $\mathrm{Sym}(n-1)=\{g\in\mathrm{Sym}(n) \mid g(n)=n\}$.
- 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

c‘ are called disjoint if the intersection of their supports is empty. Two disjoint cycles always commute. (Prove this!) A cycle $(a_1, a_2, \ldots, a_n)$ also commutes with its inverse $(a_n, \ldots, a_2, a_1)$

Every element in $\mathrm{Sym}(n)$ is a product of cycles. Even more is true:

Theorem 5.2.6. *Every permutation 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 $|support(g)| \leq k$ 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 $support(g)$. 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 $c{=}(a_1, a_2, \ldots, a_n)$. Its support is a subset of $support(g)$. So the permutation $h{=}g{\cdot}c^{-1}$ fixes all points of fix g as well as the points a, i,
$i < m + 1$. Indeed, $h(a_i) = g{\cdot}c^{-1}(a_i) = g(a_{i-1}) = a_i$, where we set $a_m{=}a_0$. This implies that the support of h is contained in $support(g)\backslash\{a_1, a_2, \ldots, a_n\}$. By the induction assumption we may write h as a product of disjoint cycles $c_1, c_2, \ldots, c_k$. The support of these cycles is contained in $support(h)$ and therefore disjoint from $\{a_1, a_2, \ldots, a_n\}$. But then $g = h{\cdot}c = c_1{\cdot}c_2{\cdot}\ldots{\cdot}c_k{\cdot}c$ 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 $c_1, c_2, \ldots, c_k$ and at the same time of the disjoint cycles $d_1, d_2, \ldots, d_l$, all of length at least 2. We prove the uniqueness by induction on t. 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 $c_2{\cdot}\ldots{\cdot}c_k = (c_1)^{-1}{\cdot}g = (d_1)^{-1}{\cdot}g = d_2{\cdot}\ldots{\cdot}d_l$. The induction hypothesis yields that $k-1{=}l-1$ and, possibly after renumbering of the indices, $d_i{=}d_i$ for all i from 0 to k. This proves the proposition.

□

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.

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 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 $(1, g{\cdot}1, g^2{\cdot}1, \ldots)$. 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 $(2, g{\cdot}2, g^2{\cdot}2, \ldots)$. 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 $(5, g{\cdot}5, g^2{\cdot}5, \ldots)$, 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.

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 proposition, 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 $h{\cdot}g{\cdot}h^{-1}\mapsto$ 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 $h{\cdot}(a_1,\ldots,a_m)){\cdot}a^{-1}{=}(h(a_1),\ldots,h(a_m)))$.*
- *If $g_1,\ldots,g_k$ are in* $\mathrm{Sym}(n)$, *then $h{\cdot}g_1\cdots g_k{\cdot}h^{-1}{=}h{\cdot}g_1{\cdot}h^{-1}\cdots h\cdot g_k\cdot h^{-1}$. In particular, if $c_1,\ldots,c_k$ are (disjoint) cycles, then $h{\cdot}c_1{\cdot}\ldots{\cdot}c_k{\cdot}h^{-1}$ is the product of the (disjoint) cycles $h{\cdot}c_1{\cdot}h^{-1},\ldots,h{\cdot}c_k{\cdot}h^{-1}$.*

Proof. The proof of both items in the lemma are easy verifications if you take the following approach.

- Part 1 : Conjugation of a cycle Conjugation of a cycleWe compute $h{\cdot}g{\cdot}h^{-1}{\cdot}x$ by distinguishing two cases.
  - $x{=}h{\cdot}a$ for some $1{\leq}i{\leq}m$: $h{\cdot}g{\cdot}h^{-1}{\cdot}h{\cdot}a{=}h{\cdot}g{\cdot}a{=}h{\cdot}a$ (with the convention that $a{=}a$).
  - If x is not in $\{h{\cdot}a,\ldots,h{\cdot}a\}$ then $h^{-1}{\cdot}x$ is not in $\{a,\ldots,a\}$, so that $g{\cdot}h^{-1}{\cdot}x{=}h^{-1}{\cdot}x$ and consequently $h{\cdot}g{\cdot}h^{-1}{\cdot}x{=}x$.

  We conclude that $h{\cdot}g{\cdot}h^{-1}{=}(h(a_1),h(a_2),\ldots,h(a_n))$.
- Part 2 : Conjugation of a product of permutations Conjugation of a product of permutationsThe second item of the lemma follows once you realize that in the product $h{\cdot}g{\cdot}h^{-1}{\cdot}h{\cdot}g{\cdot}h^{-1}{\cdot}\ldots{\cdot}h{\cdot}g{\cdot}h^{-1}$ the pairs $h^{-1}{\cdot}h$ cancel, so that $h{\cdot}g{\cdot}g{\cdot}\ldots{\cdot}g{\cdot}h^{-1}$ is what remains.In particular,

for cycles c, the first item of the lemma then shows that the product $h{\cdot}c{\cdot}h^{-1}{\cdot}h{\cdot}c{\cdot}h^{-1}{\cdot}\ldots{\cdot}h{\cdot}c{\cdot}h^{-1}$ is the product of the cycles $h{\cdot}c{\cdot}h^{-1}$. If cycles have disjoint supports, then their conjugates also have disjoint supports: The support of $h{\cdot}c{\cdot}h^{-1}$ is h (supp c) (see the first item of the lemma), so that the supports of $h{\cdot}c{\cdot}h^{-1}, h{\cdot}c{\cdot}h^{-1}, \ldots, h{\cdot}c{\cdot}h^{-1}$ are the sets h (supp c), h (supp c),. .., h (supp c). Since h is a bijection, these sets are disjoint if the sets supp $c, \ldots,$ supp c are disjoint.

□

Example 5.2.10. Let be an equilateral triangle with vertices $A, B,$ and $\mathbb{C}$. The reflection in the line L through B and the midpoint of the edge $A{\cdot}\mathbb{C}$ induces a permutation of the three vertices: $A{\mapsto}C$, $B{\mapsto}B$, $C{\mapsto}A$. If we name the three vertices $1, 2, 3$ for A, B, C, respectively, then we can describe the reflection by the permutation $(1, 3)$. A rotation through + 120 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 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* Sym(n) *have the same cycle type if and only if there exists a permutation k in* Sym(n) *with $g{=}k{\cdot}h{\cdot}k^{-1}$.*

Proof.

- If IfThis implication follows from the conjugation formulas.
- 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 $s_i{=}s_{i,1}{\cdot}s_{i,2}{\cdot}\ldots{\cdot}s_{i,k_i}$ and $t_i{=}t_{i,1}{\cdot}t_{i,2}{\cdot}\ldots{\cdot}s_{i,k_i}$. Denote by u a permutation with $u(s_{i,j}){=}t_{i,j}$ 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 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 $\{1,\ldots,m-1\}{\cdot}i \mapsto (a(i),a(i+1))$,
and the proof is complete.

□

Example 5.2.14. Let $a{=}\{a,\ldots,a\}$ 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 of the list, compare it with the predecessor a , and switch both elements if a is less than a . 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, then it yields, after applying all the transpositions $(i-1,i)$ where a is less than a a permutation with $j{=}a$ for all $j{\in}\{1,\ldots,n\}$. 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. See the Bubble Sort algorithm at work!

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 $c_1 \cdot \ldots \cdot c_k$ with k even, as the product of transpositions $d_1 \cdot \ldots \cdot d_m$ with m odd. Then $1 = c_1 \cdot \ldots \cdot c_k \cdot d_1^{-1} \cdot \ldots \cdot d_m^{-1}$ 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 $1 = t_1 \cdot \ldots \cdot t_m$ 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 $\{1, \ldots, n\}$, we can shift all transpositions which contain 1 to the front without violating the minimality of m.

There is an index i with $i \in \{2, \ldots, l\}$ such that $t_i = t_1$.
We must have $t_1 \cdot t_2 \cdot \ldots \cdot t_l(1) = 1$. Therefore $2 = t_1(1)$ lies is in the support of $t_2 \cdot \ldots \cdot t_l$, 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 S_n. The sign (signum) of g, denoted by $\text{sgn}(g)$, 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 $\text{sgn}(g)=1$ and odd if $\text{sgn}(g)=-1$.

Example 5.3.3.

The sign is multiplicative.

Theorem 5.3.4. *For all permutations g, h in $\text{Sym}(n)$, we have $\text{sgn}(g\cdot h)=\text{sgn}(g)\cdot\text{sgn}(h)$.*

Proof. Let g and h be elements of $\text{Sym}(n)$.

- 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.

□

We also say that sgn is a multiplicative map from Sym_n to $\{1,-1\}$. (The notion morphism explores this view further in a general context.)

Remark 5.3.5.

- 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 $g \cdot g^{-1}=e$, we find sgn g * sgn g^{-1} = sgn e=1, so that sgn g and sgn g^{-1} must both be 1 or both be - 1.
- Every m-cycle $(a_1, \ldots, a_m)$ can be written as the product of $m-1$ transpositions: $(a_1, \ldots, a_m)=(a_1, a_2)\cdot(a_2, a_3)\cdot \ldots \cdot(a_{m-1}, a_m)$. 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.

The previous theorem implies the following way of determining the sign.

Corollary 5.3.6. *If a permutation g is written as a product of cycles, then* $\mathrm{sgn}(g)=(-1)^w$*, where w is the number of cycles of even length.*

Proof. Since sgn 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.

□

Example 5.3.7.

Application 5.3.8. 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 Sym_n of the products $\mathrm{sgn}(g)\cdot A_{1g(1)}\cdot A_{2g(2)}\cdot \ldots \cdot A_{ng(n)}$, i.e., $\det(A)=\sum_{g\in} \mathrm{Sym}_n \mathrm{sgn}(g)\cdot A_{1g(1)}\cdot A_{2g(2)}\cdot \ldots \cdot A_{ng(n)}$.
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 sgn is multiplicative implies that products and inverses of even permutations are even. This gives rise to the following definition.

Definition 5.3.9. By A_n we denote the set of even permutations in S_n. We call A_n the alternating group on n letters.

There are just as many even permutations as there are odd permutations in S_n.

Example 5.3.10. For n=3, the even permutations are (in cycle notation): $e, (2,3,1)$ and $(3,1,2)$.

Remark 5.3.11. This set is closed with respect to taking products and inverse elements.

There are just as many even as odd permutations in Sym_n.

Theorem 5.3.12. *For $n > 1$, the alternating group Alt_n contains precisely $\frac{n!}{2}$ elements.*

Proof. An element g of Sym_n is even (respectively, odd), if and only if the product $g{\cdot}(1,2)$ is odd (respectively, even). Hence the map $g{\mapsto}g{\cdot}(1,2)$ defines a bijection between the even and the odd elements of Sym_n. But then precisely half of the $n!$ elements of Sym_n are even.

□

Example 5.3.13.

3-cycles are the smallest nontrivial even cycles. They are the building blocks for even permutations:

Theorem 5.3.14. *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 $(a, b) \cdot (c, d) = (a, b) \cdot (b, c) \cdot (b, c) \cdot (c, d) = (a, b, c) \cdot (b, c, d)$.
- Without loss of generality we are left with the case where a, b, d are pairwise distinct and $b = c$. But then $(a, b) \cdot (b, d) = (a, b, d)$.

This proves the theorem.

□

Example 5.3.15.

5.4 Exercises

5.4.1 Symmetric Groups

Exercise 5.4.1. In 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 Sym_n. Show that if $i \in support(g)$, then $g(i) \in support(g)$.

Exercise 5.4.3. How many elements of 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 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=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 $\mathbb{Z}/3\mathbb{Z}[X]/(X^2+1)\cdot\mathbb{Z}/3\mathbb{Z}[X]$ 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 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 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 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 $g^m = \mathrm{id}$ for some positive number m, show that t divides m. In particular, if c is a t-cycle and $c^m = \mathrm{id}$ for some positive number m, then m is divisible by t.
- Prove that if the permutation g has cycle structure $m_1, \ldots, m_r$, then the order of g equals lcm$(m_1, \ldots, m_r)$.

Exercise 5.4.8.

- Prove that for $n > 4$ every permutation in 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 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 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 Sym_n that commutes with all elements of Sym_n.

Exercise 5.4.11. Write all elements of Alt_4. as products of disjoint cycles.

Exercise 5.4.12. Let $a = (1, 2)$ and $b = (2, \ldots, n)$.

- 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 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 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 Sym_3.
- Prove: If g, h are permutations in 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 $\mathrm{sgn}(g) = \det(M)$.

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 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 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 $A^\top$ are equal, i.e., prove $\det A = \det A^\top$.

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 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?

5.5 Summary

The main subject is the bijections of the set $\{1, 2, \ldots, n\}$ to itself, the so-called permutations of $1, 2, \ldots, n$. Permutations may be multiplied (product of permutations) and we may take their inverses (the inverse of a permutation). In both cases the result is again a permutation. The set of all permutations of $1, 2, \ldots, n$ is denoted Sym_n. We deal with the following subjects.

- Presentations of a permutation:
  - as an ordered row;
  - as a $2 \cdot n$ matrix;
  - by a permutation matrix;
  - as a product of disjoint cycles (NB, it is a theorem that this is possible);
  - as a product of transpositions (NB, it is a theorem that this is possible);
  - by a graph.

  For this the following terms are important:
  - m-cycle, 2-cycle or transposition;
  - the fixed points of a permutation;
  - disjoint cycles.
- Calculating with permutations in the different presentations.
- The sign of a permutation with the rules: even * even = even, odd * even = odd, odd * odd = even.
- The subset Alt_n of even permutations in Alt_n.

Index