Contents of MTH1001–Algebra

Sandro Mattarei, University of Lincoln

Lecture notes of Algebra. Week 1

1. Introduction: various types of numbers; notation

- The *natural* numbers $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.
  - They originate from counting, but note that (for us) they start from 0.
  - Can be added (and also multiplied), but not subtracted in general: we can say that $1 - 3$ *should be* the same as $2 - 4$, $3 - 5$, etc., but such a thing does not exist within the natural numbers.
  - *Solution:* we *invent* a symbol for it, namely -2 (the *opposite*).
- The *integer* numbers $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$ (or the *integers*).
  - They can be added and subtracted arbitrarily.
  - They can be arbitrarily multiplied, but not divided in general: even if we exclude dividing by 0, which could never make sense, we can say that $2 : 3$ should be the same as $4 : 6$, $6 : 9$, or even $(-2) : (-3)$, etc., but such a thing does not exist within the integers.
  - *Solution:* we *invent* a symbol for it, namely $2/3$ (the *reciprocal*).
  - Differently from subtraction in $\mathbb{Z}$, we still need pairs of integers to represent the result of an arbitrary division, as the reciprocals alone, such as $1/2$, $1/(-3)$, etc., would not be enough.
- The *rational* numbers $\mathbb{Q}$.
  - They consist of *fractions* of integers, hence of the form a/b, with $a, b \in \mathbb{Z}$ and $b \neq 0$, with the understanding that a/b and c/d represent the same rational number exactly when $ad = bc$.

    (Note that the condition $ad = bc$ makes sense already in $\mathbb{Z}$.)
  - They can be arbitrarily added and subtracted, multiplied and divided, with the only exception that we cannot divide by zero.

    (There is no way to remedy this, as $0 \cdot b = 0$ whatever b is.)
- The *real* numbers $\mathbb{R}$.
  - These are harder to construct from $\mathbb{Q}$, we will not go into this.
  - They include things like π, and like $\sqrt{2}$, but not $\sqrt{-2}$, or $\sqrt{-1}$.
- The *complex* numbers $\mathbb{C}$.
  - Each complex number has the form $a + bi$, for unique $a, b \in \mathbb{R}$.
  - Operations are done 'normally' with the only extra rule that $i^2 = -1$.
  - More info and practice on $\mathbb{C}$ in the Calculus module.

Identifying $\mathbb{Z} \ni 2 = 2/1 \in \mathbb{Q}$ etc. we have $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

2. Divisibility and the greatest common divisor in the integers

2.1. Divisibility in the integers. What does it mean that an integer b divides an integer a? One interpretation is that a/b is an integer as well. There are at least two problems with this possible definition. Firstly, it does not tell you whether 0 divides 0 or not, because $0/0$ does not make sense. Secondly, for generalisations it is good to have a definition which only mentions integer numbers, while a/b might generally be a rational number.

Definition 1 (Divisibility in the integers). Let $a, b \in \mathbb{Z}$. We say that b divides a, and we write $b \mid a$, if there is $c \in \mathbb{Z}$ such that $a = b \cdot c$.

We can express divisibility in various equivalent ways. The expressions

- b divides a,
- b is a divisor of a,
- a is a multiple of b,
- a is divisible by b,

all have the same meaning, which we write symbolically as $b \mid a$.

Do not confuse the symbol $|$, a vertical bar, with the fraction sign $/$, as $b \mid a$ is not related with the fraction b/a, also written $\frac{a}{b}$, but rather with a/b.

In fact, the statement that $b \mid a$ is equivalent with the quotient a/b being an integer, but only for $b \neq 0$. We cannot divide 0 by 0 and so cannot write $0/0$, but $0 \mid 0$ is a true statement, because $0 = 0 \cdot 1$ (or $0 = 0 \cdot 3$, or $0 = 0 \cdot (-22)$, or even $0 = 0 \cdot 0$; uniqueness of c is not required in the above definition).

Hence the integer b divides the integer a when the equation

$$a = bc$$

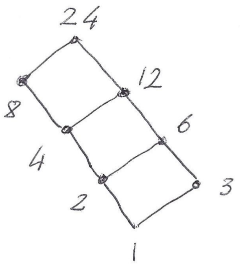
has a solution c in the integers. For example, 2 divides 6 because $6 = 2 \cdot 3$, and so the above equation, for $a = 6$ and $b = 2$, admits the solution $c = 3$. Note that $6 = 2 \cdot 3 = 3 \cdot 2$, hence the same equation tells us that 3 divides 6. In fact, the divisors of a nonzero integer a come in pairs: we can match any divisor b of a to the divisor a/b (which may possibly equal b, in case $a = b^2$).

For $a \in \mathbb{Z}$ we set

$$D(a) = \{x \in \mathbb{Z} : x \mid a\},$$

the set of divisors of a. Note that if $b \in D(a)$ then $-b \in D(a)$ as well, and also $D(-a) = D(a)$ for every a.

Example. $D(24) = \{\pm1, \pm2, \pm3, \pm4, \pm6, \pm8\pm12, \pm24\}$. The divisors are found more systematically starting from the factorisation of 24 into prime factors, $24 = 2 \cdot 2 \cdot 2 \cdot 3 = 2^3 \cdot 3$, and are best arranged in a *Hasse diagram*,



where the ascending lines indicate that the number below divides the one above. For simplicity we have omitted the ± signs, hence omitting the negative divisors, because after all $b \mid -a$ if and only if $b \mid a$, etc. Of course this representation works best if the number has only two prime divisors.

Example. Every integer b divides 0, because $0 = b \cdot 0$. Hence $D(0) = \mathbb{Z}$.

Note that if we had to place 0 somewhere in a (partial) Hasse diagram of $\mathbb{Z}$, we would have to place it 'higher' than any other integer: with respect to divisibility 0 is sort of 'the largest integer,' despite being the smallest in absolute value.

Example. However, if 0 divides a, then $a = 0 \cdot c = 0$ for some c, and so $a = 0$. This means that the only integer $a \in \mathbb{Z}$ such that $0 \in D(a)$ is $a = 0$ itself.

For each a we have $a \mid a$ (*reflexivity*), because $a = a \cdot 1$. Divisibility also enjoys *transitivity*: if $a \mid b$ and $b \mid c$, then $a \mid c$.

Symmetry does not generally hold, as $b \mid a$ does not imply $a \mid b$. In fact, $b \mid a$ and $a \mid b$ can only hold simultaneously when $a = \pm b$. Because this fact is important for the sequel, here is a proof.

Proof. Since the assertions $\pm b \mid \pm a$, with all possible choices of signs, are equivalent to each other, we may assume that $a, b \geq 0$, and so we only need to prove that under this condition $b \mid a$ and $a \mid b$ together imply $a = b$. First, if $a = 0$ then $0 \mid b$ implies $b = 0$, and conversely. Hence we may actually assume $a, b > 0$. Now $b \mid a$ means $b = ac$ for some integer c, which must be positive as well, hence $c \geq 1$, and so $b = ac \geq a \cdot 1 = a$. Similarly we find $a \leq b$, and we conclude $a = b$ as desired. □

2.2. Division with remainder in the integers.

Theorem 2 (Division with remainder in the integers). *Given two integers a, b, with $b > 0$, there exist unique $q, r \in \mathbb{Z}$ such that $a = b \cdot q + r$, with $0 \leq r < b$.*

We will not give a formal proof of Theorem 2, but the division which Theorem 2 describes is the ordinary division with remainder which you know from school, except that here we allow a to be negative.

EXAMPLE. Dividing $a = -13$ by $b = 5$ gives quotient $q = -3$ and remainder 2, because $-13 = 5 \cdot (-3) + 2$, and $0 \leq 2 < 5$.

Theorem 2 actually remains true also for $b < 0$, provided that we replace the second condition with $0 \leq r < |b|$. Of course there is no way to make the theorem work for $b = 0$.

The following variant of integer division allows negative remainders, and aims at making the remainders as small as possible *in absolute value*. This may be convenient in certain situations in order to have simpler calculations.

THEOREM 3 (Variant of division with remainder in the integers). *Given two integers a, b, with $b > 0$, there exist unique $q, r \in \mathbb{Z}$ such that $a = b \cdot q + r$, with $-b < 2r \leq b$.*

The condition on the remainder is equivalent to $-b/2 < r \leq b/2$, but we have preferred the other formulation because it takes place in the integers rather than in the rationals.

EXAMPLE. When $a = 28$ and $b = 5$, using this variant of division we will get $28 = 5 \cdot 6 - 2$ instead of $28 = 5 \cdot 5 + 3$, so the remainder will be $r = -2$ (which is smaller than 3 in absolute value).

When $a = 21$ and $b = 6$, we will have $21 = 6 \cdot 3 + 3$ or $21 = 6 \cdot 4 - 3$, but the condition on the remainder chooses the positive remainder $r = 3$ in this case.

The theorem on division with remainder (in either version, Theorem 2 or Theorem 3) yields the following useful characterization of divisibility.

COROLLARY 4. *Let $a, b \in \mathbb{Z}$ with $b \neq 0$. The following assertions are equivalent:*

(1) *b divides a;*

(2) *the remainder of the division of a by b is zero.*

PROOF. If b divides a then $a = b \cdot c = b \cdot c + 0$ for some c. Because of uniqueness, the remainder must be zero.

Conversely, if $r = 0$, then $a = b \cdot q + r = b \cdot q$, and hence b divides a. □

2.3. The greatest common divisor. Recall the definition of greatest common divisor which is usually learnt in school. Let a, b be positive integers. An integer d is called the greatest common divisor of a and b if

(1) d divides a and b, and

(2) if c is any integer which divides both a and b, then $c \leq d$.

This definition is not good for us because it does not generalise properly to other contexts (polynomials and further). Also, with this definition there would be no greatest common divisor in the special case $a = b = 0$, because the divisors of 0 are *all the integers*, and they do not have a maximum.

A better definition replaces 'max in the natural roder given by $\leq$' with 'max with respect to divisibility'.

Definition 5 (Greatest common divisor). Let $a, b \in \mathbb{Z}$. An integer d is called *a greatest common divisor* of a and b if

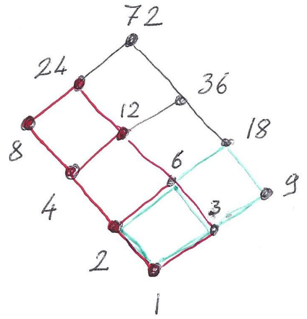
(1) d divides a and b, and
(2) if c is any integer which divides both a and b, then $c \mid d$.

A greatest common divisor of a and b (or *a GCD* in short) is denoted with $\gcd(a, b)$, or more simply with (a, b) (which we will preferably use).

Example. Let $a = 24$ and $b = 18$, and compute their GCD as learnt in school. First we need to factorise a and b as products of prime numbers, and in this case we have

$$24 = 2 \cdot 2 \cdot 2 \cdot 3 = 2^3 \cdot 3 \qquad\qquad 18 = 2 \cdot 3 \cdot 3 = 2 \cdot 3^2$$

The school's rule (which we will recall later together with primes and unique factorisation) tells us that $(24, 18) = 2 \cdot 3 = 6$, and also that the least common multiple (see definition later) of 24 and 18 is $2^3 \cdot 3^2 = 72$. In such a simple case we can actually arrange all divisors of 72 in the following Hasse diagram,



where in red and green we see the Hasse diagrams of the divisors of 24 and 18, respectively. Note that the common divisors of 24 and 18 are precisely the divisors of 6, so $D(24) \cap D(18) = \{\pm 1, \pm 2, \pm 3, \pm 6\}$. Hence 6 is the greatest common divisor of 24 and 18 in the school's sense of being the larger numerically, but also in the stronger sense of our definition.

The school's rule used in the above example to find the greatest common divisor of two (or more) integers is something like *the GCD equals the product of all common prime factors, taken with the lower (or lowest) exponent.* This is correct but is not practical, because factorisation into products of primes is a hard computational problem when the numbers are large, as we will see (while computing a greatest common divisor is much faster, using Euclid's algorithm, see below). Besides, it is theoretically more convenient to deduce the unique factorisation of integers from the existence of the greatest common divisor and its properties (which we will do in a later section), rather than the other way around.

Note that if d is a greatest common divisor of a and b, then $-d$ is also a greatest common divisor. But there cannot be other greatest common divisors of a and b. In fact, if d and d' are both greatest common divisors of a and b, then Definition 5 implies $d \mid d'$ and $d' \mid d$, and we know that $d' = \pm d$ follows. Hence the GCD is not unique, but almost (and we will accept and use the traditional expression 'the GCD' in place of the more correct 'a GCD').

3. The Euclidean algorithm in the integers

To prove that the GCD of any two integers actually exist we will describe an actual algorithm to compute it, called the *Euclidean algorithm*.

3.1. The Euclidean algorithm. Before describing the Euclidean algorithm we start with a numerical example.

Example. We compute the GCD of $a = 78$ and $b = 33$, using the Euclidean algorithm which we will explain below. That consists of a sequence of divisions, in this case

$$\begin{aligned} 78 &= 33 \cdot 2 + 12 \\ 33 &= 12 \cdot 2 + 9 \\ 12 &= 9 \cdot 1 + 3 \\ 9 &= 3 \cdot 1 + 0 \end{aligned}$$

So we have started with dividing one of the given numbers by the other, with remainder: $a = bq + r$. Then we have discarded the first number a, let b and the remainder r take the previous roles of a and b, and we have divided again. This continues until one of the divisions had remainder zero. The remainder of the previous division (hence the last nonzero remainder, if we like) is the greatest common divisor of a and b, so in this case $(78, 33) = 3$. Correct, because $78 = 2 \cdot 3 \cdot 13$ and $33 = 3 \cdot 11$.

It is convenient to think of the list of remainders as including the two original numbers, hence $78, 33, 12, 9, 3, 1, 0$ in this case. Each of them (starting from the third) is determined as the remainder of dividing the previous two in that order. Hence division with remainder gives us a recursive way to construct this sequence (where the first two are given to get started), and the sequence ends as soon as it reaches zero (which it eventually will because it is decreasing).

Now we present a general description of the Euclidean algorithm, starting with a lemma which explains the crucial reason why the algorithm works. We introduce the notation $D(a, b) = D(a) \cap D(b)$ for the set of common divisors of two integers a, b. Then it is easy to see that a greatest common divisor of a and b is any number d such that $D(a,b) = D(d)$. For example, the GCD of $a = 0$ and $b = 0$ is 0, because $D(0,0) = \mathbb{Z} = D(0)$. More generally, the GCD of any a and 0 is a, because $D(a,0) = D(a) \cap D(0) = D(a) \cap \mathbb{Z} = D(a)$.

In the general case, the existence of the GCD can be proved as follows. Because $D(-a) = D(a)$, we can restrict ourselves to the case $a, b \geq 0$. Because $D(a,b) = D(b,a)$, we may also assume $a \geq b$. The crucial step is the following lemma.

Lemma 6. *If $a = bq + c$, then $D(a,b) = D(b,c)$.*

Proof. We have to show that every element of $D(a,b)$ is also an element of $D(b,c)$, and the converse, namely, that every element of $D(b,c)$ is also an element of $D(a,b)$.

If $d \in D(a,b)$, that is, if d is any common divisor of a and b, then d also divides bq, and hence it also divides the difference $a - bq = c$. Because d divided b in the first place, we conclude that it divides both b and c, which means $d \in D(b,c)$.

The converse is similar. If if d is a common divisor of b and c, then d also divides bq, and hence it also divides the sum $bq + c = a$. Therefore, d divides both a and b, which means $d \in D(a,b)$. □

We use Lemma 6 as follows. Because we have seen the case $b = 0$, suppose $b > 0$. Dividing a by b we find

$$a = bq_1 + r_1, \qquad 0 \leq r_1 < b.$$

According to Lemma 6 we have $D(a,b) = D(b,r_1)$, and so for the sake of computing our GCD we may replace a with b and b with r_1, with the advantage that r_1 is smaller than b. Assuming $r_1 > 0$ we divide b by r_1, with remainder $r_2 < r_1$, and continue in the same way by dividing and replacing, until one of the divisions has remainder zero. That is bound to happen because the successive remainders $r_1, r_2, r_3, \ldots$ are strictly decreasing nonnegative integers. In conclusion, we have performed a sequence of divisions

$$\begin{aligned} a &= bq_1 + r_1, & \qquad & 0 < r_1 < b, \\ b &= r_1 q_2 + r_2, & & 0 < r_2 < r_1, \\ r_1 &= r_2 q_3 + r_3, & & 0 < r_3 < r_2, \\ &\vdots & & \vdots \\ r_{i-2} &= r_{i-1} q_i + r_i, & & 0 < r_i < r_{i-1}, \\ r_{i-1} &= r_i q_{i+1}, & & \end{aligned}$$

where the remainder r_{i+1} equals zero. Then the last nonzero remainder, r_i is the desired GCD of a and b, because

$$D(a,b) = D(b,r_1) = D(r_1,r_2) = \cdots = D(r_{i-1},r_i) = D(r_i,0) = D(r_i).$$

For example, let us compute the GCD of 18 and 14. We have $18 = 14 \cdot 1 + 4$, and so $D(18, 14) = D(14, 4)$. Again, $14 = 4 \cdot 3 + 2$, and so $D(14, 4) = D(4, 2)$. And again, $4 = 2 \cdot 2 + 0$, and so $D(4, 2) = D(2, 0) = D(2)$. Hence 2 is the desired GCD. Of course this was also easy to do with the school method but, for example, finding the GCD of 1987 and 2203 by the school method would not be so easy. This method is called *the Euclidean algorithm (of successive divisions).*

3.2. Examples and further comments on the Euclidean algorithm.

Example. Compute the GCD of 59 and 22:

$$\begin{aligned} 59 &= 22 \cdot 2 + 15 \\ 21 &= 15 \cdot 1 + 7 \\ 15 &= 7 \cdot 2 + 1 \end{aligned}$$

Hence $(59, 22) = 1$. Note that when the Euclidean algorithm reaches a remainder 1 there is no need to do or write down the last division $7 = 1 \cdot 7 + 0$, as it is obvious that its remainder will be zero. When two integers have greatest common divisor 1 as in this case we say that they are *relatively prime,* or that they are *coprime.* Do not confuse being coprime with being prime: 59 is actually a prime but 22 is not. Two integers may be coprime without either being prime, for example 4 and 15.

Example. Compute the GCD of 34 and 21:

$$\begin{aligned} 34 &= 21 \cdot 1 + 13 \\ 21 &= 13 \cdot 1 + 8 \\ 13 &= 8 \cdot 1 + 5 \\ 8 &= 5 \cdot 1 + 3 \\ 5 &= 3 \cdot 1 + 2 \\ 3 &= 2 \cdot 1 + 1 \end{aligned}$$

Hence $(34, 21) = 1$. In this case the Euclidean algorithm has been as slow as it can possibly be, because all quotients happened to be 1. This will occur exactly when the starting numbers a and b are consecutive Fibonacci numbers. The *Fibonacci numbers* $F_0, F_1, F_2, \ldots$ are defined by the recurrence relation

$$F_n = F_{n-1} + F_{n-2} \qquad (n \geq 2; \quad F_0 = 1, F_1 = 1),$$

and so their sequence begins with $0, 1, 1, 2, 3, 5, 13, 21, 34, 55, 89, 144, \ldots$

The following example illustrates how computing a greatest common divisor by means of the Euclidean algorithm can be much faster than by the school method of factorising the two numbers.

Example. We use the Euclidean algorithm to compute the greatest common divisor of 391 and 299. The Euclidean algorithm reads

$$391 = 299 \cdot 1 + 92$$
$$299 = 92 \cdot 3 + 23$$
$$92 = 23 \cdot 4.$$

Hence $(391, 299) = 23$.

In particular, we discover that $391 = 17 \cdot 23$ and $299 = 13 \cdot 23$ (the complete factorisations of these two numbers into prime factors), without previously factorising either number. Note that factorising 391 directly, for example, would have taken a while, because the standard procedure would be:

- checking if 391 is divisible by 2: it is not (easy, check if last digit is even or odd);
- checking if it is divisible by 3: it is not (a shortcut is checking if the sum of its digits is divisible by 3);
- checking if it is divisible by 5: it is not (because its last digit is not 0 or 5);
- checking if it is divisible by 7: it is not (not so easy, just perform the division);
- checking if it is divisible by 11: it is not (a shortcut is taking the alternating sum of the digits (that is, with alternating signs, here $3 - 9 + 1$) is divisible by 11);
- checking if it is divisible by 13: it is not (not so easy, just perform the division);
- checking if it is divisible by 17, and finally finding that it is.

Example. Compute the GCD of 2203 and 1987:

$$2203 = 1987 \cdot 1 + 216$$
$$1987 = 216 \cdot 9 + 43$$
$$216 = 43 \cdot 5 + 1$$

Hence $(2203, 1987) = 1$, so these two numbers are coprime (or relatively prime). In this case both 2203 and 1987 are actually prime numbers, and so finding their GCD by factorising them would have taken even longer than in the previous example.

In particular, because $\sqrt{1987}$ is about 44.5, we would have spent a long time trying and dividing it by the primes $2, 3, 5, 7, 11, \ldots, 43$ before discovering that it actually is a prime. At that point it would be enough to check that 1987 does not divide 2203 in order to conclude that $(2203, 1987) = 1$. However, this procedure would have taken a long time (most of it to factorise 1987).

By contrast, the Euclidean algorithm was very fast to give us $(2203, 1987) = 1$, but does not tell us that they are prime. More generally, when the Euclidean algorithm on a and b concludes with $(a, b) = 1$, it gives us no clue about the factorisations of a and b.

The successive divisions in the Euclidean algorithm can also be done in the variant where the remainders are taken as small as possible in absolute value. When possible this may make the Euclidean algorithm conclude a little faster.

Example. We compute the GCD of 29 and 18, on the left in the standard way, and on the right taking negative remainders when convenient:

$$\begin{array}{rlcrl}
29 &= 18 \cdot 1 + 11 & \qquad\qquad & 29 &= 18 \cdot 2 - 7 \\
18 &= 11 \cdot 1 + 7 & & 18 &= 7 \cdot 3 - 3 \\
11 &= 7 \cdot 1 + 4 & & 7 &= 3 \cdot 2 + 1 \\
7 &= 4 \cdot 1 + 3 & & & \\
4 &= 3 \cdot 1 + 1 & & &
\end{array}$$

We see that taking negative remainders whenever those are smaller than the positive ones in absolute value made the algorithm run faster (in fewer steps), and this is generally a good strategy. However, even if when we choose to take negative remainders only sometimes, that may still make our algorithm run a bit faster than with just standard divisions. Here are a couple of examples:

$$\begin{array}{rlcrl}
29 &= 18 \cdot 2 - 7 & \qquad\qquad & 29 &= 18 \cdot 1 + 11 \\
18 &= 7 \cdot 2 + 4 & & 18 &= 11 \cdot 2 - 4 \\
7 &= 4 \cdot 2 - 1 & & 11 &= 4 \cdot 3 - 1
\end{array}$$

Example. Compute the GCD of 34 and 21 as in a previous example (the one on Fibonacci numbers), but allowing negative remainders:

$$\begin{aligned}
34 &= 21 \cdot 2 - 8 \\
21 &= 8 \cdot 3 - 3 \\
8 &= 3 \cdot 3 - 1
\end{aligned}$$

Hence $(34, 21) = 1$. This has taken half as many divisions as doing it in the normal way as before. In fact, note that the successive remainders, in absolute value, have been $34, 21, 8, 3, 1$, so every other Fibonacci number (apart from the start) rather than going through all of them as we did before, 34,21,13,8,5,3,2,1.

Lecture notes of Algebra. Week 2

4. The extended Euclidean algorithm in the integers

4.1. The extended Euclidean algorithm. The calculations done in the Euclidean algorithm on integers a and b can be read backwards and allow one to express their greatest common divisor $d = (a, b)$ *as linear combination of* a *and* b *(with integer coefficients),* meaning finding integers x and y such that $d = ax + by$. The fact that this is always possible is an important fact known as *Bézout's Lemma.*

Lemma 7 (Bézout's Lemma). *Let* $a, b \in \mathbb{Z}$, *and let* $d = (a, b)$ *be their greatest common divisor. Then there exist* $x, y \in \mathbb{Z}$ *such that*

$$ax + by = d.$$

Example. Recall the calculations of a previous example, where we applied the Euclidean algorithm to 391 and 299 to find $(391, 299) = 23$.

$$\begin{aligned} 391 &= 299 \cdot 1 + 92 \\ 299 &= 92 \cdot 3 + 23 \\ 92 &= 23 \cdot 4. \end{aligned}$$

Working backwards through the above calculations we find

$$\begin{aligned} \mathbf{23} &= \mathbf{299} - \mathbf{92} \cdot 3 \\ &= \mathbf{299} - (\mathbf{391} - \mathbf{299} \cdot 1) \cdot 3 = -\mathbf{391} \cdot 3 + \mathbf{299} \cdot 4. \end{aligned}$$

Hence $d = (391, 299) = 391x + 299y$ with $x = -3$ and $y = 4$. In the calculation I have set the remainders in boldface in order to better distinguish them from the various coefficients involved. (It is not necessary to do so if one is tidy, but if you find it helpful in writing by hand you may them underline them, for example.)

Here is what we have done. We have started by writing the GCD 23 as a linear combination of the previous two remainders 299 and 92, using the division where 23 appears as the remainder. Then we have considered the smaller of those two remainders, which is 92, and using the previous division, where 92 appeared as the remainder, we have replaced it with a combination of 391 and 299. Now after simplification the GCD 23 has been expressed as linear combination of the remainders 391 and 299. We are done in this case, but in general we may continue, at each step getting rid of the lower of the two remainders in terms of which the GCD is expressed, at the expense of a the previous (larger) reminder, until we have expressed d as a combination of a and b.

It may be convenient (depending on preference) to arrange the calculations for the extended part of the algorithm next to the main part, but going up from the bottom, as follows,

$$\begin{array}{rl}
391 = 299 \cdot 1 + 92 & \qquad = \mathbf{299} - (\mathbf{391} - \mathbf{299} \cdot 1) \cdot 3 = -\mathbf{391} \cdot 3 + \mathbf{299} \cdot 4 \\
299 = 92 \cdot 3 + 23 & \mathbf{23} = \mathbf{299} - \mathbf{92} \cdot 3 \\
92 = 23 \cdot 4 &
\end{array}$$

In this way each line at the right-hand side uses exactly the result of the division to the left of it, and an extra calculation is done continuing on the same line.

Reading the calculations of the Euclidean algorithm backwards as in the example is sometimes called the *extended Euclidean algorithm* (which really includes the original part of the algorithm). Writing out the procedure in a formal general way (rather than just for specific numbers as in the example) actually provides a proof of Bézout's lemma. This would be an example of a *constructive proof,* as opposed to a *non-constructive proof* which merely proves the existence of x and y without actually giving a procedure (that is, an *algorithm*) for computing them.[1]

Example. Compute the GCD d of 83 and 53, and then express it in the form $d = 83x + 53y$ for some integers x and y (that is, find a solution in the integers of the equation $83x + 53y = d$). Here is the Euclidean algorithm, which shows $d = (83, 53) = 1$, and the extended part to the right of it:

$$\begin{array}{rl}
83 = 53 \cdot 1 + 30 & \qquad = -\mathbf{53} \cdot 13 + (\mathbf{83} - \mathbf{53} \cdot 1) \cdot 23 = \mathbf{83} \cdot 23 - \mathbf{53} \cdot 36 \\
53 = 30 \cdot 1 + 23 & \qquad = \mathbf{30} \cdot 10 - (\mathbf{53} - \mathbf{30} \cdot 1) \cdot 13 = -\mathbf{53} \cdot 13 + \mathbf{30} \cdot 23 \\
30 = 23 \cdot 1 + 7 & \qquad = -\mathbf{23} \cdot 3 + (\mathbf{30} - \mathbf{23} \cdot 1) \cdot 10 = \mathbf{30} \cdot 10 - \mathbf{23} \cdot 13 \\
23 = 7 \cdot 3 + 2 & \qquad = \mathbf{7} - (\mathbf{23} - \mathbf{7} \cdot 3) \cdot 3 = -\mathbf{23} \cdot 3 + \mathbf{7} \cdot 10 \\
7 = 2 \cdot 3 + 1 & \mathbf{1} = \mathbf{7} - \mathbf{2} \cdot 3
\end{array}$$

So we find $1 = 83 \cdot 23 + 53 \cdot (-36)$. Note that, of x and y, one had necessarily to be negative and one positive, if we want the result to be 1. However, which of them will be positive and which negative we do not know until we do Euclid's algorithm, as it depends on the parity of the number of steps.

Remark. One may show that the solution x, y of $ax + by = d$ which is produced by the extended Euclidean algorithm always satisfies $|x| \leq |b/d|$ and $|y| \leq |a/d|$.

[1] Many such non-constructive proof exist in mathematics, and are often shorter and more elegant than corresponding constructive ones. Obviously constructive proofs have an important practical advantage. But for certain theorems only non-constructive proofs are known.

4.2. (Optional) A variant of the extended Euclidean algorithm. We now present a variant of the extended Euclidean algorithm, which is perhaps easier to describe formally, and involves doing some calculations on the side while executing the Euclidean algorithm, rather than working back from the end once the algorithm is finished.[2] We start with writing

$$\begin{array}{rcccc} a &=& a\cdot 1 &+& b\cdot 0 \\ b &=& a\cdot 0 &+& b\cdot 1 \end{array}$$

We divide as in the Euclidean algorithm, $a = bq_1 + r_1$, with $0 \leq r_1 < b$, and then extend the table as follows:

$$\begin{array}{rcccc} a &=& a\cdot 1 &+& b\cdot 0 \\ b &=& a\cdot 0 &+& b\cdot 1 \\ r_1 &=& a\cdot 1 &+& b\cdot(-q_1) \end{array}$$

We divide again: $b = r_1q_2 + r_2$, with $0 \leq r_2 < r_1$. Hence $r_2 = b - r_1q_2$. We use the last two rows of the above table to express r_2 in terms of a and b:

$$\begin{array}{rcccc} a &=& a\cdot 1 &+& b\cdot 0 \\ b &=& a\cdot 0 &+& b\cdot 1 \\ r_1 &=& a\cdot 1 &+& b\cdot(-q_1) \\ r_2 &=& a\cdot u_2 &+& b\cdot v_2 \end{array}$$

Here $u_2 = -q_2$ and $v_2 = 1 + q_1q_2$. But these exact expressions for u_2 and v_2 are not important, the important fact is that they exist, and they can be found by taking a suitable linear combination of the previous two lines of the table. Eventually one of the remainders will be the GCD d of a and b, the table will read

$$\begin{array}{rcccc} a &=& a\cdot 1 &+& b\cdot 0 \\ b &=& a\cdot 0 &+& b\cdot 1 \\ r_1 &=& a\cdot 1 &+& b\cdot(-q_1) \\ r_2 &=& a\cdot u_2 &+& b\cdot v_2 \\ && \vdots && \\ d &=& a\cdot u &+& b\cdot v \end{array}$$

and we will have found the desired linear combination.

[2] This version has practical advantages over the other, which are more apparent when the number of steps of the algorithm is large. For example, when implemented on a computer (or, especially, a smartcard with limited memory) this variant requires very little memory compared to the other, as only two consecutive steps of the calculation need to be kept in the memory at a given time (as opposed to memorising all the calculations done in the Euclidean algorithm in order to read them backwards at the end).

For example, take $a = 24$ and $b = 14$. The successive divisions are

$$\begin{array}{rcrcr} \mathbf{24} & = & \mathbf{14} \cdot 1 & + & \mathbf{10} \\ \mathbf{14} & = & \mathbf{10} \cdot 1 & + & \mathbf{4} \\ \mathbf{10} & = & \mathbf{4} \cdot 2 & + & \mathbf{2} \\ \mathbf{4} & = & \mathbf{2} \cdot 1 & + & 0 \end{array}$$

This shows that the GCD is 2. (We have set remainders in boldface font for clarity.) Now we compute, as explained earlier,

$$\begin{array}{rcccc} \mathbf{24} & = & \mathbf{24} \cdot 1 & + & \mathbf{14} \cdot 0 \\ \mathbf{14} & = & \mathbf{24} \cdot 0 & + & \mathbf{14} \cdot 1 \\ \mathbf{10} & = & \mathbf{24} \cdot 1 & + & \mathbf{14} \cdot (-1) \\ \mathbf{4} & = & \mathbf{24} \cdot (-1) & + & \mathbf{14} \cdot 2 \\ \mathbf{2} & = & \mathbf{24} \cdot 3 & + & \mathbf{14} \cdot (-5) \end{array}$$

5. Some consequences of Bézout's lemma

5.1. Some basic properties of the GCD. The important case when two integers a and b have GCD equal to 1 has a special name: such a and b are called *coprime,* or *relatively prime.* (Do not confuse this with the notion of a *prime.*)

The following four lemmas express important properties of the GCD. For convenience we will refer to them collectively as *the Arithmetical Lemmas.* Lemma B is the most used (and Lemmas C and D are direct consequences of it).

LEMMA 8 (Arithmetical Lemma A). *For two integers a and b, the following properties are equivalent:*

(1) *a and b are coprime;*

(2) *there are $x, y \in \mathbb{Z}$ such that $ax + by = 1$.*

PROOF. That (1) implies (2) is simply Bézout's lemma.

The converse is easier: the greatest common divisor d of a and b divides both ax and by, hence it divides their sum 1, and consequently $d = 1$ (or -1, but we have seen that it makes no difference when taking GCD's). □

However, beware that if $ax + by = d$ for certain integers, and $d > 1$, then we can only conclude that the greatest common divisor (a, b) divides d, but not that $(a, b) = d$.

LEMMA 9 (Arithmetical Lemma B). *Suppose $(a, b) = 1$. If a divides the product $b \cdot c$, then a divides c.*

PROOF. According to Bézout's lemma, there exist $x, y \in \mathbb{Z}$ such that

$$ax + by = 1.$$

Multiplying both sides by c we find

$$a(xc) + (bc)y = c.$$

Because a divides each of the products at the left-hand side, it divides their sum as well, which is c. □

Lemma 10 (Arithmetical Lemma C). *Let a and b be integers, not both zero. Then*

$$\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1.$$

Proof. According to Bézout's lemma, there exist $x, y \in \mathbb{Z}$ such that

$$ax + by = (a, b).$$

After dividing by $(a, b) \neq 0$ we find

$$\frac{a}{(a,b)}x + \frac{b}{(a,b)}y = 1,$$

and hence $\frac{a}{(a,b)}$ and $\frac{b}{(a,b)}$ are coprime. □

Lemma C is relevant to simplifying fractions 'in one go': instead of several steps like $42/60 = 21/30 = 7/10$, because $(42, 60) = 6$ Lemma C tells us that $\frac{42/6}{60/6} = \frac{7}{10}$ cannot be further simplified. Hence simplifying a fraction can always be done in a single step (and the Euclidean algorithm provides the factor to simplify).

Lemma 11 (Arithmetical Lemma D). *Let a and b integers, not both zero. If $a \mid b \cdot c$, then*

$$\frac{a}{(a,b)} \mid c.$$

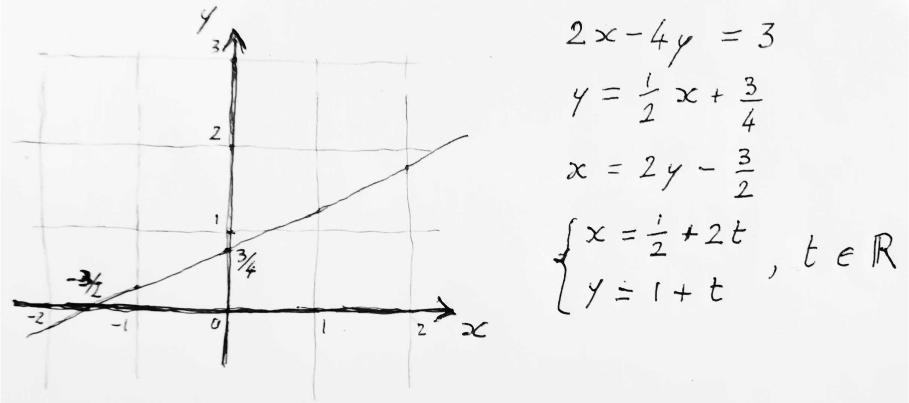
Proof. By assumption there exists x such that $ax = bc$. Dividing both sides by (a, b) (which is not zero) we get

$$\frac{a}{(a,b)} \cdot x = \frac{b}{(a,b)} \cdot c.$$

According to Arithmetical Lemma C, the integers $\frac{a}{(a,b)}$ and $\frac{b}{(a,b)}$ are coprime, and hence Arithmetical Lemma B applies. □

5.2. Solving $ax + by = c$ in $\mathbb{Z}$. Consider the equation $ax + by = c$, where a, b, c are integers. We want to solve the equation *in the integers*. By *solving* we mean finding *all* solutions (which may include saying that there are none, in certain cases), that is, all pairs of integers x, y such that $ax + by = c$, where a, b, c are any given integers.

Example. The equation $2x - 4y = -3$ has infinitely many solutions in the real numbers. In fact, the equation describes a straight line in the plane in Cartesian coordinates, and the solutions of the equation correspond to the points on the line.



Hence all solutions can be described depending on a single arbitrary real parameter, for example as all pairs $(x, y) = (t, \frac{1}{2}t + \frac{3}{4})$ with $t \in \mathbb{R}$, or as all $(x, y) = (s, -2s - \frac{3}{2})$ with $s \in \mathbb{R}$ (or in other ways as well, different *parametrisations* of the line). However, for none of these solutions both x and y take integer values, and so the equation $2x - 4y = -3$ has no solution in the integers. For example, because the left-hand side can only be an even number if x and y are integers, while the right-hand side -3 is odd.

More generally, if an integer d divides both a and b, and the equation $ax + by = c$ has at least one solution (x_0, y_0) in the integers, then d must also divide $c = ax_0 + by_0$. Consequently, if we find a common divisor d of a and b which does not divide c as well, then we know that $ax + by = c$ cannot have any integer solution. But we need not check all common divisors d of a and b, as checking if their GCD $d = (a, b)$ divides c takes care of them all at once.

Therefore, a *necessary* condition for the equation $ax + by = c$ to have integer solutions is that (a, b) divides c. But that is actually also a *sufficient* condition. This is because according to Bézout's lemma we have $a' + by' = (a, b)$ for certain integers x_1 and y_1 (which we can find through the extended Euclidean algorithm), and then $(x_0, y_0) = \big(x_1 \frac{c}{(a,b)}, y_1 \frac{c}{(a,b)}\big)$ is a solution of $ax + by = c$.

Example. Suppose we need to solve $83x + 53y = -2$ in the integers. In a previous example we found by the extended Euclidean algorithm that 83 and 53 are coprime, and that their GCD 1 can be written as $1 = 83 \cdot 23 + 53 \cdot (-36)$. Multiplying both sides by -2 we see that one solution of $83x + 53y = -2$ is given by $x_0 = 23 \cdot (-2) = -46$ and $y_0 = -36 \cdot (-2) = 72$.

At this point we may assume that a and b are not both zero, otherwise the equation becomes $0 = c$, which is easy to discuss (no solutions if $c \neq 0$, and any arbitrary integer values of x and y are solutions if $c = 0$). The discussion so far makes sense even when

$a = b = 0$, but what we are going to do now would not. The cases where just one of a or b is zero would be easy to discuss separately, but there is no need to do that as the following discussion will take care of those cases as well.

We have already seen that in order for $ax + by = c$ to have integer solutions we need that (a, b) divides c. Because a and b are not both zero, $(a, b) \neq 0$, and so we may divide both sides of the equation by (a, b) and obtain the equivalent equation $a'x+b'y = c'$, where $a' = a/(a, b)$, $b' = b/(a, b)$, and $c' = c/(a, b)$. Now $(a', b') = 1$ according to Arithmetical Lemma C.

It remains to see how to solve $a'x + b'y = c'$, but for simplicity let us switch notation and just call a, b, c the new coefficients. So we want to solve the equation $ax + by = c$, with $(a, b) = 1$. We already know how to find one integer solution, so a pair of integers x_0 and y_0 satisfying $ax_0 + by_0 = c$. If k is any integer, then $x_0 - kb$ and $y + ka$ give another solution, simply because

$$a(x_0 - kb) + b(y_0 + ka) = ax_0 + by_0 = c.$$

Now we show that every solution x, y has that form. If x and y form an arbitrary solution, that means $ax + by = c$. Subtracting this equation from $ax_0 + by_0 = c$ side by side and rearranging we find $a(x_0 - x) = b(y - y_0)$. Hence b divides $a(x_0 - x)$, but we have assumed that $(b, a) = 1$, and so b divides $x_0 - x$ according to Arithmetical Lemma B. Hence $x_0 - x = kb$ for some $k \in \mathbb{Z}$. Substituting this into $a(x_0 - x) = b(y - y_0)$ and dividing by b we find $y - y_0 = ka$. (This works only if $b \neq 0$; but if $b = 0$ then $a = \pm 1$ because $(a, b) = 1$, and then the conclusions remain correct, namely $x = x_0$ and $y = k \in \mathbb{Z}$ arbitrary in this case.) Hence $x = x_0 - kb$ and $y = y_0 + ka$ as we wanted. In conclusion, we have shown that under the condition $(a, b) = 1$ all integer solutions of $ax + by = c$ are given by the formulas

$$x = x_0 - kb, \qquad y = y_0 + ka, \qquad \text{for } k \in \mathbb{Z}.$$

Example. We want to find all integer solutions of the equation $54x + 21y = 15$. We start with the extended Euclidean algorithm on 54 and 21.

$$\begin{aligned}
& & 54 &= 54 \cdot 1 + 21 \cdot 0 \\
& & 21 &= 54 \cdot 0 + 21 \cdot 1 \\
54 &= 21 \cdot 3 - 9 \qquad & 9 &= 54 \cdot (-1) + 21 \cdot 3 \\
21 &= 9 \cdot 2 + 3 \qquad & 3 &= 54 \cdot 2 + 21 \cdot (-5)
\end{aligned}$$

The first part of the Euclidean algorithm tells us that $(54, 21) = 3$, and because that divides 15 the equation has solutions in the integers (otherwise we could have stopped right there).

The extended part of the Euclidean algorithm tells us that $54 \cdot 2 + 21 \cdot (-5) = 3$, and so it gives us a solution of the equation $54x + 21y = 3$, which is different from the one we are interested in. But multiplying both sides by 5 we find $54 \cdot 10 + 21 \cdot (-25) = 15$, and so $x_0 = 10$ and $y_0 = -25$ form a solution of our equation $54x + 21y = 15$.

Now, one sees immediately $x = 10 - 21k$ and $y = -25 + 10k$ certainly are other solutions of $54x + 21y = 15$, for any integer k, but these formulas would not give us *all* integer solutions. The problem is that $(54, 21) = 3 \neq 1$, so this would be a wrong way to proceed.

The correct way is replacing $54x+21y = 15$ with the equation $18x+7y = 5$ (obtained by dividing by 3), which is equivalent to the other and so has the same solutions. Now 18 and 7 are coprime, and so our theory tells us that the formulas $x = 10 - 7k$ and $y = -25 + 2k$ do give us all solutions of $18x + 7y = 5$ as k varies over the integers, and so also the solutions of our original equation $54x + 21y = 15$, which is equivalent.

Remark (Optional). There is an interesting variation, useful in some applications, which is solving the equation $ax + by = c$ in the *nonnegative* integers. Let a, b be coprime positive integers. Then for each integer $c \geq (a - 1)(b - 1)$ there exist integers $x, y \geq 0$ such that $ax + by = c$.

We omit the proof of this fact (which is not difficult), but for practice we show that the stated hypothesis on c is *best possible,* because $ax + by = ab - a - b$ has no solution with integers $x, y \geq 0$. In fact, rewriting the equation in the form $a(x+1)+b(y+1) = ab$, because $(a, b) = 1$ we have that $b \mid x + 1$ and $a \mid y + 1$. But if $x, y \geq 0$ then $x + 1$ and $y + 1$ are positive multiples of b and a, hence $x + 1 \geq b$ and $y + 1 \geq a$. But then $a(x + 1) + b(y + 1) \geq 2ab > ab$, a contradiction.

As an example, every integer which is at least $(4 - 1)(3 - 1) = 6$ can be written as $4x + 3y$ with x, y nonnegative integers:

$$\begin{aligned}
6 &= 4 \cdot 0 + 3 \cdot 2 \\
7 &= 4 \cdot 1 + 3 \cdot 1 \\
8 &= 4 \cdot 2 + 3 \cdot 0 \\
9 &= 4 \cdot 0 + 3 \cdot 3 \\
10 &= 4 \cdot 1 + 3 \cdot 2 \\
11 &= 4 \cdot 2 + 3 \cdot 1 \\
12 &= 4 \cdot 3 + 3 \cdot 0 = 4 \cdot 0 + 3 \cdot 4 \\
13 &= 4 \cdot 1 + 3 \cdot 3 \\
14 &= 4 \cdot 2 + 3 \cdot 2 \\
15 &= 4 \cdot 3 + 3 \cdot 1 = 4 \cdot 0 + 3 \cdot 5
\end{aligned}$$

and so on. However, $6-1=5$ cannot be written like that, as one of the coefficients needs to be negative: $5 = 4 \cdot 2 - 3 \cdot 1$.

6. Unique factorisation in the integers

6.1. Primes, and unique factorisation.

DEFINITION 12. An integer $a > 1$ is *composite* if it can be written as $a = bc$, where b and c are positive integers larger than 1 (or, equivalently, b and c are smaller than a; or, equivalently again, where $1 < b < a$); it is *prime* if it is not composite. Equivalently, an integer $a > 1$ is a prime if it has no divisors different from 1 and itself.

We have deliberately excluded 1 from both definitions: it is neither a prime nor a composite. (For several reasons this is much more convenient than including it within the set of primes.) Note that two different primes are certainly coprime (as their only positive common factor is 1), but two integers can be coprime without being primes, for example $6 = 2 \cdot 3$ and $35 = 5 \cdot 7$. This is the reason for the term *coprime*: having no prime factors in common.

LEMMA 13. *Let p be a prime (integer), and let a, b be integers. If p divides the product ab, then p divides either a or b.*

PROOF. Suppose that p does not divide a. Then (p, a), which divides p, is not p, and so it can only be 1. Then Arithmetical Lemma B applies, and hence p divides b, as desired. □

THEOREM 14 (Fundamental Theorem of Arithmetic). *Every integer larger than* 1 *factorises into a product of primes, and in a unique way.*

(OPTIONAL) SKETCH OF PROOF. A proper proof requires induction, but the idea of the proof of existence is easy enough to explain informally. If our number n is not prime, we may factorise it as $n = ab$, with a and b integers smaller than n. If a and b are primes then we are done. Otherwise we may factorise at least one of them, and so on. This process cannot go on forever, because otherwise we would keep producing smaller and smaller positive integers, and so it must eventually stop, and we are left with a factorisation of n as a product of primes.

A proof of uniqueness uses Lemma 13, and goes roughly as follows. Suppose that n can be written in two ways as a product of primes, say

$$n = p_1 p_2 \cdots p_s = q_1 q_2 \cdots q_t,$$

where the p_i and the q_j are (positive) primes, possibly with repetitions. (Note that we initially we do not even know if the two factorisations involve the same number of prime factors.) In particular, p_1 divides $q_1 q_2 \cdots q_t$, but according to Lemma 13 (possibly repeated a number of times) it follows that p_1 divides at least one of the factors q_j. Possibly after renumbering them, say that p_1 divides q_1. But q_1 being a prime itself, its only positive divisors are 1 and q_1. Because $p_1 \neq 1$ we have $p_1 = q_1$. Hence

$$n/p_1 = p_2 \cdots p_s = q_2 \cdots q_t,$$

and we may repeat the same argument on those two factorisations (which have fewer prime factors than the original ones). Eventually (details omitted) we reach the desired conclusion that the two factorisations are the same. □

We have only worked with positive integers in this subsection, but we might as well have allowed any nonzero integers. However, the definition of composite in Definition 12, for an integer $a \neq, 0, \pm 1$, should be modified into the existence of some factorisation $a = bc$, where neither a nor b is ± 1 (that is, a *proper* factorisation). Here ± 1 play a special role because they are the *invertible* integers: those integers x for which $1/x$ is also an integer (called the *inverse* of x, although this is traditionally called the *reciprocal* of x). Then -2,-3,$-5, \ldots$ are also primes, but they are essentially the same as $2, 3, 5, \ldots$, respectively, because they can be obtained from the latter through multiplication by an invertible integer (that is, -1). With this extended meaning, uniqueness of factorisation still holds up to changing sign to some factors (that is, multiplying them by -1), as in $6 = 2 \cdot 3 = (-2) \cdot (-3)$, or $-6 = 2 \cdot (-3) = (-2) \cdot 3$.

6.2. The least common multiple.

Definition 15 (Least common multiple). Let $a, b \in \mathbb{Z}$. An integer m is a *least common multiple (lcm)* of a and b if

(1) a and b divide m, and
(2) if c is any integer divisible by both a and b, then m divides c.

A least common multiple of a and b is sometimes denoted by $[a, b]$ (or $\operatorname{lcm}(a, b)$).

As for the greatest common divisor, this definition easily shows that, if the least common multiple of two integers exists, then it is unique up to a sign. (This is because, again, if we have two least common multiples then each of them divides the other.) However, to prove that the lowest common multiple exists (and also to compute it) we may now rely on what we already know about the greatest common divisor, as follows.

Let c be any common multiple of a and b. Hence $c = b \cdot x$ for some integer x. Because $a \mid b \cdot x$, Arithmetical Lemma D implies $\frac{a}{(a,b)} \mid x$, and so $x = \frac{a}{(a,b)} \cdot y$ for some integer y. Therefore,

$$c = b \cdot x = b \cdot \frac{a}{(a, b)} \cdot y = \frac{ab}{(a, b)} \cdot y.$$

Hence every common multiple of a and b is also a multiple of $m = ab/(a,b)$. Moreover, m itself is a multiple of a and b, because

$$m = \frac{ab}{(a,b)} = a \cdot \frac{b}{(a,b)} = b \cdot \frac{a}{(a,b)},$$

and the two fractions are integers. Hence m is a least common multiple of a and b according to Definition 15. We have discovered the following:

Theorem 16. *$(a,b) \cdot [a,b] = a \cdot b$.*

However, you actually already knew this from school, because of the following way of computing GCD and lcm (whose correctness follows from the theorem on unique factorisation):

Theorem 17.

- *The GCD of two integers is the product of all their* common *prime factors, each taken with the* lower *exponent.*
- *The lcm of two integers is the product of all their prime factors,* common or not*, each taken with the* higher *exponent.*

If this is the formulation you knew from school, note that you may remove the distinction between *common* factors and *common or not* simply by writing both factorisations of a and b using the same prime factors, but possibly with exponent zero.

Example. We compute GCD and lcm of

$$a = 12 = 2^2 \cdot 3^1 \cdot 5^0, \qquad b = 45 = 2^0 \cdot 3^2 \cdot 5^1,$$

where we have included 5^0 and 2^0 in their factorisations in order to use the same set of primes. Taking each prime factor with the lower exponent we get

$$(12,45) = 2^0 3^1 5^0 = 3,$$

and taking each prime factor with the higher exponent we get

$$[12,45] = 2^2 3^2 5^1 = 180.$$

If you look at both GCD and lcm, you see that altogether we have taken all powers of 2, 3, and 5, which appear in the factorisations of 12 and 45, just in a different order. Consequently, we have

$$(24,45) \cdot [24,45] = 2^0 3^1 5^0 \cdot 2^2 3^2 5^1 = 2^2 3^1 5^0 \cdot 2^0 3^2 5^1 = 12 \cdot 45,$$

as predicted by Theorem 16.

Remark. The parentheses and square brackets notation used to denote GCD and lcm can be extended to more than two integers. Theorem 17 easily extends by replacing *lower* and *higher* exponent with *lowest* and *highest* exponent. However, the product of (a,b,c)

and $[a, b, c]$ does not equal the product abc of three integers. The correct generalisation of Theorem 16 to the case of three integers is

$$[a, b, c] = abc \cdot \frac{(a, b, c)}{(a, b) \cdot (a, c) \cdot (b, c)}.$$

6.3. (Optional) Computational complexity. The school method for finding the GCD of two integers a and b (say both positive) starts with factorising a and b into products of prime factors. When a and b are large this is not a good idea, as we informally explain now.

The simplest method for factorising an integer a (called *method of trial divisions*) is dividing it by $2, 3, 4, 5, \ldots$. (One may restrict oneself to dividing by the prime numbers $2, 3, 5, 7, \ldots$, but this is not a huge saving; more on this later.) One can stop at $\lfloor \sqrt{a} \rfloor$. In fact, if a is not a prime, and hence $a = bc$, with $0 < b, c < a$, then we may assume $b \leq c$, hence $b^2 \leq bc = a$, and so $b \leq \sqrt{a}$. If we have not find a factor of a after dividing it by $2, 3, \ldots \lfloor \sqrt{a} \rfloor$ (that is, none of the divisions gave zero remainder), then we must conclude that a is a prime.

Therefore, in order to find a proper factor of a in case it is composite the method of trial divisions requires at most $\lfloor \sqrt{a} \rfloor - 1$ divisions, in the worst case. (And in case a was prime, which we would usually not know in advance, the method produces a (long) proof that a is prime only after exactly $\lfloor \sqrt{a} \rfloor - 1$ divisions.) Here the -1 makes an insignificant difference when a is large, and so does taking the integral part of $\sqrt{a}$. Therefore, we may say that the method of trial divisions requires at most *about* $\sqrt{a}$ divisions, in the worst case, to factorise a.

There is a theory of *computational complexity,* which studies, roughly speaking, how many operations are required (which can be then translated into how long a computer takes to perform them) to complete a certain algorithm. There are a lot more details to this, and variations. For example, we may be interested in how long an algorithm takes to complete *on average* rather than in the worst possible case. Also, for the sake of simplicity, above we just counted the number of divisions required to factorise a, without considering that some single divisions may take longer than others to be done (depending on the sizes of the numbers involved).

We now show that the Euclidean algorithm takes much fewer operations to complete than factorising integers of similar size. In fact, in the general step we divide

$$r_j = r_{j+1} q_{j+2} + r_{j+2},$$

and hence $r_j > r_{j+1} > r_{j+2}$. But because $q_{j+2} \geq 1$ we have

$$r_j = r_{j+1} q_{j+2} + r_{j+2} \geq r_{j+1} + r_{j+2} > 2 r_{j+2}.$$

Hence, although the remainder may not change much in two consecutive divisions, every two divisions it decreases at least by a factor two:

$$r_{j+2} < \frac{1}{2} r_j.$$

Hence if we start with $r_{-1} = a > r_0 = b > 0$, and the last nonzero remainder is r_{2k} or r_{2k+1}, then

$$1 \leq r_{2k} < \frac{r_0}{2^k} = \frac{b}{2^k},$$

whence $2^k < b$, which means $k < \log_2(b)$. Hence the total number of divisions required, which is either $2k+1$ or $2k+2$, is no more than $2k+2 < 2\log_2(b)+2$. Because $\log_2(b) = \log_2(10)\log_{10}(b)$, and $\log_2(10) = \ln(10)/\ln(2)$ equals about $3.32\cdots$, we conclude that the total number of divisions required for the Euclidean algorithm on $a > b$ is, roughly, less than $7\log_{10}(b)$, that is, again roughly, seven times the number of decimal digits of b.

If, for example, b is about 10^{200}, the Euclidean algorithm requires less than about $7 \cdot 200 = 1400$ divisions with remainder. These can be done in a split second by a modern computer. Attempting to factorise b by trial divisions, however, may require up to $\sqrt{b} = 10^{100}$ divisions with remainder. This is a huge number, with about 100 digits. For example, if a computer were able to perform one trillion, that is, 10^{12}, divisions with remainder in a second, it would take $10^{100}/10^{12} = 10^{88}$ seconds to finish. (For comparison, the fastest desktop personal computers in 2015 can process about 200 000 MIPS, that is, only one fifth of a trillion *instructions* per second; and of course each of our trial divisions would require many such instructions.) To gauge how long 10^{88} seconds is, consider that in a year there are $60 * 60 * 24 * 365$ seconds, which is about $3 \cdot 10^7$ seconds. Hence 10^{88} seconds is around $3 \cdot 10^{80}$ years. For comparison, note that the estimated age of the universe is only 13.8 billion years, that is, $13.8 \cdot 10^9$ years!

One may object that, in the method of trial divisions, instead of dividing by the integers $2, 3, 4, 5, \ldots$, it would be enough to divide by the prime numbers $2, 3, 5, 7, 11, 13, 17, \ldots$. Although this seems to make a significant time saving at the beginning (for example, dividing only by 2 and the *odd* integers $3, 5, 7, 9, 11, \ldots$, which certainly include all the remaining primes, already saves about half the time), this saving becomes less and less important as the numbers increase. This is because the prime numbers are certainly sparse, but still *relatively many,* so to say, among the positive integers. In fact, not only there are infinitely many primes, but we have

Theorem 18 (The prime number theorem). *The function*

$$\pi(n) := \sharp\{p\colon\ p \text{ is a prime and } p \leq n\},$$

which counts the number of primes p not exceeding n, satisfies

$$\pi(n) \sim \frac{n}{\log n},$$

meaning that the ratio of the two sides tends to the limit 1 *as* n *tends to infinity.*

For example, of the integers with less than 100 decimal digits (that is, integers less than 10^{100}), about one in 230 is a prime (because $\ln(10^{100}) = 100 \cdot \ln(10)$ is about $230.258 \cdots$). Hence if we were to try factorise a number b, as in the previous example, hence of size around 10^{200}, by only doing trial divisions by the primes less than $\sqrt{b}$, we would save a meager factor 230 in the huge required time.

Lecture notes of Algebra. Week 3

7. Writing numbers in a different base

7.1. Integers, and then real numbers, written in a base b. Fix an integer $b > 1$, called *base.* Then every non-negative integer n can be written in the form

$$n = d_{k-1}b^{k-1} + d_{k-2}b^{k-2} + \cdots + d_1 b + d_0,$$

and will be denoted by $(d_{k-1} \dots d_1 d_0)_b$, where each d_i is a digit in base b, that is, a symbol for one of the integers $0, 1, \dots, b-2, b-1$. For example, if $b \leq 10$ one one can use the ordinary (decimal) $0, 1, \dots, b-1$ digits to represent themselves in base b. However, if $b > 10$ it may be convenient to use extra symbols to denote the digits having values $10, 11, \dots, b-1$. For example, when $b = 16$ (the *hexadecimal* system) it is customary to use the letters from A to F as digits with values 10 to 15.

It is neither necessary nor convenient to assume $d_{k-1} \neq 0$. If that holds then we say that n has (exactly) k digits in base b. Not assuming that allows us to identify writings such as 010 or 0010 for the number 10 in decimal notation. Allowing for that we have that the above expression for n in base b is unique, that is, each digit d_j (including the leading zeroes) depends only on n. [3]

As for base 10, this generalises from positive integers to arbitrary positive real numbers. Every positive real number n can be written as $\sum_{i<k} d_i b^i$ (with i ranging over all integers less than k, hence possibly negative) and denoted by $(d_{k-1}d_{k-2} \dots d_1 d_0, d_{-1}d_{-2} \dots)_b$, where the digits may continue indefinitely to the right. This writing is also unique, but with exceptions, occurring when all d_i starting with a certain d_j equal $b-1$; in that case we may replace d_i for $i \geq j$ with 0 (and possibly omitting it in writing) and increase d_{j-1} by one.

Remark. The number of digits in base b of a non-negative integer n is given by the formula

$$k = \lfloor \log_b n \rfloor + 1 = \left\lfloor \frac{\log n}{\log b} \right\rfloor + 1.$$

This is because $b^{k-1} \leq n < b^k$.

[3] A proof of uniqueness essentially amounts to writing up formally and more generally the familiar way in which we tell which of two different integers in base 10 is larger, by scanning the digits from left to right until we find a different digit (assuming they have the same number of digits). A proof of existence of the expression, for any positive integer n, can be obtained by writing up formally the procedure for writing and integer in an arbitrary base b, which we explain later.

7.2. Converting integers from base b to base 10. For n an integer we conveniently write the conversion as

$$n = (\dots((d_{k-1}b + d_{k-2})b + d_{k-3})b + \dots + d_1)b + d_0.$$

This method (the same as we conveniently use to evaluate a polynomial on some number, in thic case the base b) requires only $k-1$ multiplications by b, and $k-1$ additions. It is also easy to perform on a pocket calculator (as long as that is not too smart, meaning that it should execute operations in the order in which we type them in.) For example,

$$(61405)_7 = ((6 \cdot 7 + 1) \cdot 7 + 4) \cdot 7 \cdot 7 + 5 = 14950.$$

The calculations can be conveniently arranged as follows, a special case of Ruffini's rule for polynomials (or Horner scheme, see a later section on polynomials):

	6	1	4	0	5
7		42	301	2135	14945
	6	43	305	2135	14950

7.3. Converting integers from base 10 to base b. If n is an integer, its last digit in base b, which is d_0, can be obtained as the remainder of dividing n by b, then d_1 is obtained as the remainder of dividing the previous quotient by b, etc., until we get quotient zero. For example converting $(14950)_{10}$ to base 7 can be done as follows:

$$\begin{aligned} 14950 &= 7 \cdot 2135 + 5 \\ 2135 &= 7 \cdot 305 + 0 \\ 305 &= 7 \cdot 43 + 4 \\ 43 &= 7 \cdot 6 + 1 \\ 6 &= 7 \cdot 0 + 6 \end{aligned}$$

We conclude that $(14950)_{10} = (61405)_7$.

This algorithm can be efficiently executed on a pocket calculator (which does not do divisions with remainder) if we repeatedly divide by b starting with n (without subtracting the remainders), and comparing the fractional part of each quotient with a table which we will have prepared in advance, containing $0/b, 1/b, 2/b, \dots, (b-1)/b$. The correct digit d at each stage will be found according to the rule $d/b \leq f < (d+1)/b$, where f is the fractional part of the quotient. In the previous example we will have

$$\begin{array}{l} 1/7{=}0.142\ldots \\ 2/7{=}0.285\ldots \\ 3/7{=}0.428\ldots \\ 4/7{=}0.571\ldots \\ 5/7{=}0.714\ldots \\ 6/7{=}0.857\ldots \end{array}$$

We will find

$$\begin{array}{r} 14950\qquad\quad . \\ 2135.714\ldots \\ 305.102\ldots \\ 43.586\ldots \\ 6.226\ldots \\ 0.889\ldots \end{array}$$

and hence

0.714	5
0.102	0
0.586	4
0.226	1
0.889	6

Note that such a table will need to be sufficiently accurate, especially if there are sequences of consecutive digits 6 in the expansion of n in base b. Also, the divisions will need to be done with a sufficient accuracy, otherwise we may find an incorrect answer, as we exemplify now.

EXAMPLE. Consider the integer 16806, in decimal notation. If we perform the divisions by 7 with a precision limited to three digits after the point, we find

$$\begin{array}{r} 16806\qquad\quad . \\ 2400.857\ldots \\ 342.980\ldots \\ 48.997\ldots \\ 7.000\ldots \\ 1\ldots \\ 0.142\ldots \end{array}$$

and hence

0.857	6
0.980	6
0.997	6
0.0	0
0.0	0
0.142	1

It would appear that $(16806)_{10} = (100666)_7$, but that is wrong, and the correct conclusion is $(16806)_{10} = (66666)_7$. The problem is that

$$\frac{6}{7} + \frac{6}{7^2} + \frac{6}{7^3} + \frac{6}{7^4} = 0.99958307\ldots,$$

which gets approximated to 1 if we only use three digits after the point (In other words, the 7.000 appearing in the penultimate row of the above calculation should really be a little less than 7. Unfortunately, this may produce a large error in the final result: in this case an error of $(1000)_7 = 7^3$.)

Remark. We have used two different methods to convert from base b to base 10, and from base 10 to base b (inverse to each other: *multiplying* in the former case, and *dividing* in the latter). By symmetry, each of those methods would also work for the other task, but it would involve doing the relevant calculations in base b rather than in base 10. Of course if $b = 2$ and the conversions are to be done by a computer working in base 2, rather than by a human, the two algorithms would be exchanged.

Remark. Converting from *binary* (base 2) to *hexadecimal* (base 16, where the customary symbols for the digits are $0, 1, \ldots, 9, A, B, C, D, E, F$), and from hexadecimal to binary, will be much simpler than described above. For example, to convert from binary to hexadecimal it will be sufficient to split the bits into blocks of four starting from the decimal points, and then convert each block into the corresponding hexadecimal digit.

7.4. Converting real numbers from base b to base 10. Consider a positive number written in base b with a finite number h of digits after the point. To convert it to decimal one may use the same procedure as for an integer, but continuing with the digits (in base b) after the point, up to the last digit d_{-h}, and then divide the result by b^h. (This is because ignoring the point amounts to multiplying our number by b^h.)

Example. To convert $(14.22)_5$ to decimal, remove the point (which means multiplying by 5^2), convert $(1422)_5 = 237$, and then divide by 5^2: $(14.22)_5 = 237/25 = 9.48$.

Note that until just before this last step you work with integer numbers, thus avoiding approximation errors. In fact, because of the final division the decimal expansion of the number may have infinitely many digits, even though you started with finitely many digits in base b. This cannot occur if $b = 2$ or 5 or, more generally, if the base b has only 2 and 5 as prime factors. (See a later subsection about periodic numbers.)

Example. We have $(1.22)_3 = (122)_3/3^2 = 17/9 = 5.222\cdots = 5.\dot{2}$.

If the number to be converted has infinitely many digits, one can do the same with an approximation (keeping a few digits after the point).

Example. To convert $(2.\dot{1})_3 = (2.11111\cdots)_3$ into decimal, we may convert the approximation $(2.111)_3 = (2111)_3/27 = 67/27$, which equals $2 + \frac{13}{27} = 2.\dot{4}8\dot{1}$, so roughly 2.48. In this case we can actually convert it exactly, namely $(2.\dot{1})_3 = 2.5$, because $(2.\dot{1})_3 \cdot 2 = (11.\dot{2})_3 = (12)_3 = 5$. More generally, if the number is periodic (written in any base) then it is rational and there is a rule to convert it into a fraction of integers, see a later subsection.

7.5. Converting real numbers from base 10 to base b. If n is not an integer we can deal separately with the integer part and the fractional part. In fact, multiplying the latter by b and taking the integer part of the result we get d_{-1}; then we may repeat this procedure with the fractional part of the result and find d_{-2}, etc.

EXAMPLE. To convert 2.481 to base 3, write it as $2 + 0.481$.

- $0.481 \cdot 3 = 1.443$, so first digit after the point will be 1;
- $0.443 \cdot 3 = 1.329$, so second digit after the point will be 1;
- $0.329 \cdot 3 = 0.987$, so third digit after the point will be 0;
- $0.987 \cdot 3 = 2.961$, so fourth digit after the point will be 2;
- $0.961 \cdot 3 = 2.883$, so fifth digit after the point will be 2; and so on.
- In conclusion, $2.481 = (2.11022\cdots)_3$.

This can be conveniently done on a simple pocket calculator, and we can also avoid subtracting the integral part at each step by using a trick which we have seen before, and requires a preliminary table with (at least approximate) decimal expansions of $1/b, 2/b, \ldots (b-1)/b$. In this case we have $1/3 = 0.\dot{3}$ and $2/3 = 0.\dot{6}$. We proceed as before, but reading each digit off the fractional part (rather than from the integral part) *before* multiplying by b, hence subtracting the integral part becomes unnecessary.

Hence to convert 2.481 to base 3, write it as $2 + 0.481$. Then

- because $\dot{3} \leq 0.481 < \dot{6}$, the first digit after the point will be 1;
- $0.481 \cdot 3 = 1.443$, and as $\dot{3} \leq 0.443 < \dot{6}$ the second digit after the point will be 1;
- $1.443 \cdot 3 = 4.329$, and as $0.329 < \dot{3}$ the third digit after the point will be 0;
- $4.329 \cdot 3 = 12.987$, and as $\dot{6} \leq 0.987$ the fourth digit after the point will be 2;
- $12.987 \cdot 3 = 38.961$, and as $\dot{6} \leq 0.961$ the fifth digit after the point will be 2;
- $38.961 \cdot 3 = 116.883$, and as $\dot{6} \leq 0.883$ the sixth digit after the point will be 2; and so on.
- In conclusion, we find $2.481 = (2.110222\cdots)_3$.

EXAMPLE. Converting the number π to base 2 we will find:

$$(3,1415926\ldots)_{10} = (11,00100100001111110\ldots)_2.$$

Please check that yourself, starting with as many decimal digits of π as they fit on your calculator, and proceeding like in the previous example, comparing the fractional part with $1/2 = 0.5$ after each multiplication by 2.

8. Arithmetic and geometric progressions

A finite sequence $a_1, a_2, \ldots, a_n$ of (real or) complex numbers is called an *arithmetic progression* if the difference $d = a_{k+1} - a_k$ between each two consecutive terms is constant, meaning that it is independent of k. We could also write this condition as $a_{k+1} - a_k = a_k - a_{k-1}$ for all k (to which this applies, that is, $1 < k < n$), or, equivalently, $a_{k+1} + a_{k-1} = 2a_k$.

Hence a sequence is an arithmetic progression precisely when each term is the average, or arithmetic mean, of the preceding and the following term. [4]

For an arithmetic progression we have

$$a_n = a_1 + d(n-1).$$

The sum of an arithmetic progression, also called an *arithmetic series* $a_1 + a_2 + \cdots + a_n$, can be computed as follows:

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n = \frac{(a_1 + a_n)\cdot n}{2}.$$

This can be shown by noting that any two terms which have the same distance from both ends have the same sum $a_k + a_{n-k+1} = \big(a_1 + d(k-1)\big) + \big(a_1 + d(n-k)\big) = 2a_1 + d(n-1) = a_1 + a_n$, and hence

$$\begin{aligned} 2(a_1 + a_2 + \cdots + a_n) = {} & (a_1 + a_2 + \cdots \quad +a_n) \\ & + (a_n + a_{n-1} + \cdots \quad +a_1) \\ = {} & (a_1 + a_n)\cdot n. \end{aligned}$$

Because the indices of an arithmetic progression may cover a different range, for example $a_3, a_4, \ldots, a_9$, the formula for the sum is best remembered as *the sum of the first and last term, times the total number of terms, divided by two.* (Or, equivalently, *the arithmetic mean (or average) of the first and last term, times the total number of terms.*) More generally, an arithmetic progression may have infinite terms. In fact, any finite arithmetic progression with at least two terms can be extended, on either side, to form an infinite arithmetic progression, in a unique way.

Geometric progressions are analogous to arithmetic progressions, except that sums and differences are replaced by products and quotients (ratios). Hence a finite sequence $a_1, a_2, \ldots, a_n$, made of nonzero numbers, is called a *geometric progression* if the ratio $r = a_{k+1}/a_k$ between each two consecutive terms is constant. Note that the ratio r is nonzero, but may possibly be negative, in which case the terms of the progression have alternating signs. Similarly as for arithmetic progressions, a sequence is a geometric progression precisely when $a_{k+1}\cdot a_{k-1} = a_k^2$ for all k which make sense, that is to say, when each term is the geometric mean of the preceding and the following term.

For a geometric progression we have

$$a_n = a_1 \cdot r^{n-1}.$$

[4] A sequence satisfying $a_{k+1} + a_{k-1} \geq 2a_k$ for all k is usually called *convex,* and *strictly convex* if $a_{k+1} + a_{k-1} > 2a_k$ for all k. Similarly, a sequence satisfying $a_{k+1} + a_{k-1} \leq 2a_k$ for all k is usually called *concave.* You may relate this to Calculus notions if you note that $a_{k+1} - 2a_k + a_{k-1}$ is a discrete analogue of the 'second derivative' of a function (and here a_k is a function of the discrete variable k).

In a finite geometric progressions, any two terms which have the same distance from both ends have the same product. Hence, arguing in a similar way as for the sum of an arithmetic progression we see that the product of all terms of a geometric progression, say made of positive real numbers for simplicity, is given by

$$\prod_{k=1}^{n} a_k = a_1 \cdot a_2 \cdots a_n = \sqrt{(a_1 a_n)^n}.$$

This can also be thought of as *the n-th power* of the *geometric mean* $\sqrt{a_1 a_n}$ of the first and last terms.

One may also consider the sum of a geometric progression, also called a *geometric series*. If $a_1, a_2, \ldots, a_n$ is a geometric progression with (common) ratio r, then

$$a_1 + a_2 + a_3 + \cdots + a_n = a_1(1 + r + r^2 + \cdots + r^{n-1}) = a_1 \frac{r^n - 1}{r - 1} = a_1 \frac{1 - r^n}{1 - r}.$$

This can also be used to compute the sum of an infinite geometric progression $a_1, a_2, \ldots$ (continuing indefinitely to the right). The corresponding geometric series converges (see the Calculus module for the meaning of this) exactly when r^n tends to zero as n tends to $+\infty$, which occurs exactly when $|r| < 1$. In that case the sum of the series is given by

$$a_1 + a_2 + a_3 + \cdots = a_1(1 + r + r^2 + r^3 + \cdots) = \frac{a_1}{1 - r}.$$

Remark (Optional: Arithmetic, geometric, and harmonic mean). We mentioned above the *arithmetic mean* $(a+b)/2$ of two numbers, and the *geometric mean* $\sqrt{ab}$ of two *positive real* numbers. Note that the geometric mean never exceeds the arithmetic mean, that is, $\sqrt{ab} \leq (a+b)/2$ for all positive real numbers a, b. In fact, because both sides are positive this inequality is equivalent to $ab \leq (a+b)^2/4$. In turn, this is equivalent to $0 \leq (a+b)^2 - 4ab$, that is, $0 \leq (a-b)^2$, which is certainly true for all positive real numbers a, b. A third type of mean occurs in some applications, the *harmonic mean* $\frac{1}{\frac{1}{2}(\frac{1}{a}+\frac{1}{b})} = \frac{2ab}{a+b}$. The harmonic mean of two positive real numbers a, b never exceeds their geometric mean (similar proof), and so we have

$$\frac{2ab}{a+b} \leq \sqrt{ab} \leq \frac{a+b}{2},$$

that is, [harmonic mean]$\leq$[geometric mean]$\leq$[arithmetic mean]. Note that the product of the arithmetic mean and the harmonic mean equals the square of the geometric mean; this is another (but equivalent) explanation of why the geometric mean takes an intermediate value between the other two means.

9. Periodic numbers (in decimal notation or any other base)

Example. Converting a real number with periodic decimal expansion into a fraction:

$$0.171717\cdots = 0.\overline{17} = 0.17 \cdot 1.\overline{01} = 0.17 \cdot \big(1 + (0.01) + (0.01)^2 + \cdots\big) = \frac{0.17}{1 - 0.01} = \frac{17}{99}.$$

It is easy to discover how to extend this to the most general situation of a *periodic decimal expansion* (also called a *repeated* or *recurring decimal,* for which various notations are in use) with both an *integer part* and a *pre-period*:

$$1234.56789789789\cdots = 1234.56\overline{789} = 1234.56\dot{7}8\dot{9} = \frac{123456789 - 123456}{99900}.$$

The numerator of the fraction equals

$$\text{[integer part|pre-period|period]} \quad \text{minus} \quad \text{[integer part|pre-period]},$$

ignoring the decimal dot; the denominator has as many 9s as the number of digits of the period, followed by as many 0s as the digits of the pre-period.

The procedure explained in the example shows that a real number whose decimal expansion is periodic, is actually a rational number (a fraction of integers). Obviously, a real number whose decimal expansion is finite is also a rational number. (This may actually be viewed as a special case of a periodic expansion where the period is $\dot{0}$, and the procedure for converting it to a fraction still works...) More is true: *a real number has finite or periodic decimal expansion if and only if it is rational.* We have just seen the ‘only if’ implication, that is, the ‘$\Rightarrow$’ implication. To prove the opposite implication, note that when computing the decimal expansion of a rational number m/n (hence with m and n integers), that is, when performing the ordinary school division algorithm (similar to long division for polynomials), at each step at most n remainders r are possible (as $0 \leq r < n$). Once we have finished ‘carrying down’ all the digits from m (so the following ones would all be zeroes, which we usually do not write), sooner or later one of the remainders will have to repeat, and from that point on a whole bunch of steps of the division algorithm will have to repeat periodically. It is easier to see what happens by working out an example than explaining it in words, but it follows that the resulting decimal expansion must be periodic.

The same procedure would work in any base b, just replace the digits 9 used in the rule with the digit $b - 1$. However, beware that those numbers you are writing as numerator and denominator of the fraction will be in base b, so you may then need to convert them to decimal in order to write the fraction in the ordinary way. For example,

$$(3.\dot{2}\dot{1})_7 = \frac{(321)_7 - (3)_7}{(66)_7} = \frac{(315)_7}{(66)_7} = \frac{159}{48}.$$

Note that some fractions of integers may have a finite expansion when written in some base, and a periodic infinite expansion when written in some other base:

$$\frac{1}{2} = 0.5 = (0.1)_2 = (0.\dot{1})_3 = (0.2)_4 = (0.\dot{2})_5 = (0.3)_6 = (0.\dot{3})_7 = \cdots$$

$$\frac{1}{3} = 0.\dot{3} = (0.\dot{0}\dot{1})_2 = (0.1)_3 = (0.\dot{1})_4 = (0.\dot{1}\dot{3})_5 = (0.2)_6 = \cdots$$

10. Polynomials

In this and the following sections we will consider polynomials with coefficients in a *field* F. Some examples of fields are $\mathbb{Q}$ (the field of rational numbers), $\mathbb{R}$ (the field of real numbers), $\mathbb{C}$ (the field of complex numbers). These fields satisfy $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$, but later we will discover some other interesting and unrelated fields. Most of what we do with polynomials in the following section does not really depend on which field we use (unless we state that explicitly), and so you may take F to be some field which is familiar to you, say $\mathbb{Q}$ or $\mathbb{R}$, and worry only later about the proper definition of an arbitrary field. Very roughly, a field is a set of 'numbers' which you can add, subtract, and multiply arbitrarily, and also divide except for dividing by zero, and where those operations satisfy the familiar calculation 'rules,' such as $a - (b - c) = a - b + c$, $ab = ba$, $a(b + c) = ab + ac$, etc.

Example. The set of integers $\mathbb{Z}$ is not a field, because division cannot be done arbitrarily: $2/3$ is not an integer. One can of course consider polynomials with integer coefficients (being special rational numbers), but the theorems which we will see may not be true, and the algorithms may not work, unless we accept to use rational numbers, which are a field.

Example. There are many different fields, and even some with a finite number of elements. The simplest example is the field with two elements, usually denoted by $\mathbb{F}_2$. (Later in the module we will learn about the field $\mathbb{F}_p$ of p elements, where p is any prime.) Its elements are two symbols $\bar{0}$ and $\bar{1}$, which add and multiply exactly as the integers 0 and 1 do, except that $\bar{1} + \bar{1} = \bar{0}$. Here are the full addition and multiplication tables in $\mathbb{F}_2$:

$+$	$\bar{0}$	$\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{1}$
$\bar{1}$	$\bar{1}$	$\bar{0}$

$\cdot$	$\bar{0}$	$\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$

One can verify that $\mathbb{F}_2$ satisfies the definition of a field (which we have not explicitly given but roughly amounts to saying that the usual properties of addition, subtraction, multiplication and division hold).

For the moment, we may think of a polynomial (in the single *indeterminate* x) as an expression of the form $f(x) = a_n x^n + \cdots + a_1 x + a_0$. (Arranging them in the opposite order is just another convention in use.) The notation $f(x)$ is borrowed from Calculus to stress that one may think of the polynomial as a special type of function of the "variable" x, while $a_0, a_1, \ldots, a_n$ are to be thought of as "constants" (even though in some applications they may themselves be expressions depending on parameters, other than x). Strictly speaking, a function of this type should be called *a polynomial function,* rather than *a* *polynomial.* Thinking of polynomials as functions is OK as long as one works with real coefficients, but is not quite the best in view of generalizations.

Thus, a polynomial with coefficients in the field F is an expression of the form $f(x) = a_nx^n + \cdots + a_1x + a_0$ with $a_0, \ldots, a_n \in F$, for some n. If β is any element of F, we may *evaluate $f(x)$ on β, or for $x = \beta$,* and compute the value $f(\beta) = a_n\beta^n + \cdots + a_1\beta + a_0$.

DEFINITION 19. The degree of a non-zero polynomial $f(x)$ is the largest integer n such that $a_n \neq 0$, and is denoted by $n = \deg(f)$.

The coefficient a_n in the definition is called *the leading coefficient*, and if $a_n = 1$ then $f(x)$ is said to be *monic.* We also call a_nx^n *the leading term* of the polynomial, and a_0 *the constant term.* We do not assign a degree to the zero polynomial. (In our notation $f(x) = a_nx^n + \cdots + a_1x + a_0$ we have not assumed that $a_n \neq 0$. This is actually convenient, and all that notation tells us is that $\deg(f(x)) \leq n$, or $f(x)$ might possibly be the zero polynomial.)

The sum of two polynomials is given by

$$(a_nx^n + \cdots + a_1x + a_0) + (b_nx^n + \cdots + b_1x + b_0) = (a_n + b_n)x^n + \cdots + (a_1 + b_1)x + (a_0 + b_0).$$

Note that the two polynomials need not have the same degree, and to make this formula simpler we have taken advantage of the possibility of adding zero coefficients in front of one of them to make both polynomials formally start with the same x^n.

The product of two polynomials can be computed by removing the parentheses using the distributive law, and then collecting like powers of x. Hence

$$\begin{aligned}(a_nx^n + \cdots &+ a_1x + a_0) \cdot (b_mx^m + \cdots + b_1x + b_0)\\ &= a_nb_mx^{n+m} + (a_{n-1}b_m + a_nb_{m-1})x^{n+m-1} + \cdots\\ &\qquad \cdots + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + (a_0b_1 + a_1b_0)x + a_0b_0.\end{aligned}$$

Those two formulas show that the degrees of a sum and of a product of polynomial satisfy

$$\deg\bigl(f(x) + g(x)\bigr) \leq \max\bigl(\deg(f(x)), \deg(g(x))\bigr),$$

and

$$\deg\bigl(f(x) \cdot g(x)\bigr) = \deg\bigl(f(x)\bigr) + \deg\bigl(g(x)\bigr),$$

provided that all degrees make sense, that is, unless one of the polynomials involved is the zero polynomial.

11. Polynomial division with remainder

THEOREM 20 (Division Algorithm for $F[x]$). *Let F be a field and let $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. Then there exist unique polynomials $q(x), r(x) \in F[x]$ such that*

$$f(x) = g(x)q(x) + r(x)$$

where

$$\textit{either}\quad r(x)=0\quad\textit{or}\quad \deg\bigl(r(x)\bigr)<\deg\bigl(g(x)\bigr).$$

The polynomials $q(x)$ and $r(x)$ are called the *quotient* and the *remainder* of the division of $f(x)$ by $g(x)$.

Remark (The degree of the zero polynomial). An alternate convention is to assign a degree to the zero polynomial as well. To make things work properly, however, the zero polynomial should be assigned some negative number, say -1. Doing so, the condition

$$\text{either}\quad r(x)=0\quad\text{or}\quad \deg\bigl(r(x)\bigr)<\deg\bigl(g(x)\bigr)$$

in Theorem 20 can be rephrased, more simply, as

$$\deg\bigl(r(x)\bigr)<\deg\bigl(g(x)\bigr).$$

An even better convention is assigning the symbol $-\infty$ to the zero polynomial, which makes the properties given above on the degrees of a sum and a product remain true even if one of the polynomials involved is the zero polynomial (with some natural interpretations such as $(-\infty)+3=-\infty$, or $(-\infty)+(-\infty)=-\infty$).

We omit the proof of Theorem 20, which is just a bit tedious to write down formally. However, a proof of the *existence* of $q(x)$ and $r(x)$ is just a formal transcription of what the actual algorithm does, as illustrated in the following example.

Example. In $\mathbb{Q}[x]$, divide $f(x)=2x^4+x^2-x+1$ by $g(x)=2x-1$ with remainder.

$$
\begin{array}{r|ccccccccc}
\multicolumn{1}{r}{} & x^3 & + & \frac{1}{2}x^2 & + & \frac{3}{4}x & - & \frac{1}{8} & & \\ \cline{2-10}
2x-1 & 2x^4 & + & 0x^3 & + & x^2 & - & x & + & 1 \\
\multicolumn{1}{r}{} & 2x^4 & - & x^3 & & & & & & \\ \cline{2-4}
\multicolumn{1}{r}{} & & & x^3 & + & x^2 & & & & \\
\multicolumn{1}{r}{} & & & x^3 & - & \frac{1}{2}x^2 & & & & \\ \cline{4-6}
\multicolumn{1}{r}{} & & & & & \frac{3}{2}x^2 & - & x & & \\
\multicolumn{1}{r}{} & & & & & \frac{3}{2}x^2 & - & \frac{3}{4}x & & \\ \cline{6-8}
\multicolumn{1}{r}{} & & & & & & - & \frac{1}{4}x & + & 1 \\
\multicolumn{1}{r}{} & & & & & & - & \frac{1}{4}x & + & \frac{1}{8} \\ \cline{7-10}
\multicolumn{1}{r}{} & & & & & & & & & \frac{7}{8}
\end{array}
$$

Therefore, we have $q(x)=x^3+\frac{1}{2}x^2+\frac{3}{4}x-\frac{1}{8}$ and $r(x)=\frac{7}{8}$. Another perfectly acceptable (and perhaps even preferable) answer is

$$2x^4+x^2-x+1=(2x-1)\cdot\left(x^3+\frac{1}{2}x^2+\frac{3}{4}x-\frac{1}{8}\right)+\frac{7}{8}.$$

Note that the algorithm stops as soon as we obtain a *remainder* which is zero or has degree less than the degree of $g(x)$, as stated in Theorem 20. For example, if we stopped two steps too early (omitting the last four lines) we would obtain that $f(x) = g(x)q_1(x) + r_1(x)$ with $q_1(x) = x^3 + \frac{1}{2}x^2$ and $r_1(x) = \frac{3}{2}x^2 - x + 1$, which is a true equality, but those are not the correct quotient and remainder because $\deg\big(r_1(x)\big) = 2$ is not less than $\deg(2x - 1) = 1$, and $r_1(x)$ is not zero either.

Note that, although the two original polynomials in the above example had integer coefficients, we had to use rational numbers in the course of the calculation, and also to express the final result. That example shows that division with remainder of polynomials in $\mathbb{Z}[x]$ would simply not work, and the reason is that $\mathbb{Z}$ is not a field (as we cannot divide arbitrarily by non-zero integers). Division of polynomials with integer coefficients does work in a restricted situation, namely when the polynomial we are dividing by is monic (that is, it has leading coefficient 1), as in the next example.

Example. In $\mathbb{Q}[x]$, divide $f(x) = 2x^4 - x^3 + 3x^2 + x - 2$ by $g(x) = x^2 - 2x + 2$ with remainder.

$$\begin{array}{rccccccccc}
 & 2x^2 & + & 3x & + & 5 & & & & \\ \cline{2-10}
\multicolumn{1}{r|}{x^2-2x+2} & 2x^4 & - & x^3 & + & 3x^2 & + & x & - & 2 \\
 & 2x^4 & - & 4x^3 & + & 4x^2 & & & & \\ \cline{2-6}
 & & & 3x^3 & - & x^2 & + & x & & \\
 & & & 3x^3 & - & 6x^2 & + & 6x & & \\ \cline{4-8}
 & & & & & 5x^2 & - & 5x & - & 2 \\
 & & & & & 5x^2 & - & 10x & + & 10 \\ \cline{6-10}
 & & & & & & & 5x & - & 12
\end{array}$$

Therefore $q(x) = 2x^2 + 3x + 5$ and $r(x) = 5x - 12$. Or, more explicitly,

$$2x^4 + -x^3 + 3x^2 + x - 2 = (x^2 - 2x + 2) \cdot (2x^2 + 3x + 5) + (5x - 12).$$

A proof of the *uniqueness* of $q(x)$ and $r(x)$ in Theorem 20 is easier to write down formally than a proof of their existence.

Proof of uniqueness of $q(x)$ and $r(x)$ in Theorem 20. Suppose that the division can be done in two ways,

$$f(x) = g(x)q(x) + r(x) \qquad \text{and} \qquad f(x) = g(x)q_1(x) + r_1(x),$$

with both $r(x)$ and $r_1(x)$ satisfying the required condition. Then we claim that $q_1(x) = q(x)$ and $r_1(x) = r(x)$. In fact, putting together the two equalities we obtain

$$g(x)q(x) + r(x) = f(x) = g(x)q_1(x) + r_1(x),$$

and hence

$$g(x)[q_1(x) - q(x)] = r(x) - r_1(x).$$

If the right-hand side were different from zero, then its degree would be less than the degree of $g(x)$. However, the left-hand side, if nonzero, would have degree $\deg(g(x)) + \deg[q_1(x) - q(x)] \geq \deg(g(x))$. This is impossible, and so we have to conclude that each side of the equality is zero. This implies that $r_1(x) = r(x)$, and $q_1(x) = q(x)$ (because $g(x)$ is not the zero polynomial), as we wanted to prove. □

REMARK 21. Although division with remainder for polynomials is conceptually very similar to division with remainder for integers, with polynomials we have no analogue of the variant of division with $-b/2 < r \leq b/2$: there is only one way to do division with remainder for polynomials.

Lecture notes of Algebra. Week 4

12. The Remainder Theorem and the Factor Theorem

Let $f(x) \in F[x]$, and $\alpha \in F$. We say that α is a *root* (or a *zero*) of $f(x)$ if $f(\alpha) = 0$. In other words, if we obtain zero after substituting α for x in $f(x) = a_n x^n + \cdots + a_1 x + a_0$, that is, computing $f(\alpha) = a_n \alpha^n + \cdots + a_1 \alpha + a_0$.

We say that a polynomial $g(x)$ divides $f(x)$ if there is another polynomial $h(x)$ such that $f(x) = g(x) \cdot h(x)$. This occurs exactly when dividing $f(x)$ by $g(x)$ we obtain zero as the remainder.

LEMMA 22 (The Remainder Theorem and the Factor Theorem). *Let F be a field, $0 \neq f(x) \in F[x]$, $\alpha \in F$. Then*

(1) *$f(\alpha)$ equals the remainder of the division of $f(x)$ by $x - \alpha$;*

(2) *α is a root of $f(x)$ if, and only if, $x - \alpha$ divides $f(x)$.*

PROOF. Dividing $f(x)$ by $(x - \alpha)$ we obtain $f(x) = (x - \alpha) \cdot q(x) + r$, where r is either zero or a polynomial of degree less than 1, hence a constant $r \in F$ in both cases. Evaluating on α we find $f(\alpha) = (\alpha - \alpha) \cdot q(\alpha) + r = r$, which proves the remainder theorem.

The Factor Theorem is an immediate consequence: the equality $f(x) = (x - \alpha) \cdot q(x) + r$ shows that $x - \alpha$ divides $f(x)$ (exactly) if, and only if, $r = 0$; but $r = f(\alpha)$ according to the remainder theorem. □

A nice application of the Factor Theorem arises when $f(x) = x^n \pm a^n$, where $a \neq 0$ is a constant. Because $f(a) = a^n \pm a^n$ and $f(-a) = (-1)^n a^n \pm a^n$ the Factor Theorem implies:

- $x^n - a^n$ is always divisible by $x - a$;
- $x^n - a^n$ is divisible by $x + a$ exactly when n is even;
- $x^n + a^n$ is divisible by $x + a$ exactly when n is odd;
- $x^n + a^n$ is never divisible by $x - a$.

For example,

$$\begin{aligned} x^2 - a^2 &= (x - a)(x + a) \\ x^3 - a^3 &= (x - a)(x^2 + ax + a^2) \\ x^3 + a^3 &= (x + a)(x^2 - ax + a^2) \\ x^4 - a^4 &= (x - a)(x^3 + ax^2 + a^2 x + a^3) = (x + a)(x^3 - ax^2 + a^2 x - a^3) \end{aligned}$$

Note, however, that if we had to factorise $x^4 - a^4$ it would be smarter to think of it as $(x^2)^2 - (a^2)^2$, and so

$$x^4 - a^4 = (x^2 - a^2)(x^2 + a^2) = (x - a)(x + a)(x^2 + a^2).$$

Note also that, if $a \in \mathbb{R}$ and we work over the real numbers, $x^2 + a^2$ cannot be further factorised, again because of the Factor Theorem, since it cannot have any real roots: whatever real value we assign to x will make $x^2 + a^2$ a positive number, hence never zero. Over the complex numbers, however, we have

$$x^4 - a^4 = (x - a)(x + a)(x - ai)(x + ai),$$

as ai and $-ai$ are the square roots of $-a^2$. (Every polynomial in $\mathbb{C}[x]$ can be factorised into a product of factors of degree 1, this is called *the Fundamental Theorem of Algebra.*)

Similarly, if we have to factorise $x^6 - a^6$, we best proceed as follows:

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

Over the real numbers the two quadratic factors cannot be further factorised, because they have negative discriminant $a^2 - 4a^2 = -3a^2$. Starting, instead, with

$$\begin{aligned} x^6 - a^6 &= (x^2)^3 - (a^2)^3 \\ &= (x^2 - a^2)(x^4 + a^2x^2 + a^4) \\ &= (x - a)(x + a)(x^4 + a^2x^2 + a^4), \end{aligned}$$

would not have been as good, because none of the above rules applies directly to further factorise the factor of degree four. However, there is another trick which can be very useful in certain situations, namely,

$$\begin{aligned} x^4 + a^2x^2 + a^4 &= (x^4 + 2a^2x^2 + a^4) - a^2x^2 \\ &= (x^2 + a^2)^2 - (ax)^2 \\ &= (x^2 - ax + a^2)(x^2 + ax + a^2), \end{aligned}$$

and so we would recover the complete (or full) factorisation of $x^6 - a^6$ as before. This trick also allows us to factorise $x^6 + a^6$, where writing it as $(x^3)^2 + (a^3)^2$ would have been a dead end:

$$x^6 + a^6 = (x^2 + a^2)(x^4 - a^2x^2 + a^4).$$

One could show that this is a complete factorisation over the rational numbers (if a is rational), but over the real numbers the same trick factorises the factor of degree four:

$$\begin{aligned} x^4 - a^2x^2 + a^4 &= (x^4 + 2a^2x^2 + a^4) - 3a^2x^2 \\ &= (x^2 + a^2)^2 - (\sqrt{3}ax)^2 \\ &= (x^2 - a\sqrt{3}x + a^2)(x^2 + a\sqrt{3}x + a^2). \end{aligned}$$

13. Ruffini's rule

Because of the Factor Theorem, the very special case of polynomial division where we divide by a binomial of the form $x-a$, for some constant a, is important. In this case the ordinary division algorithm is very sparse, and all the numbers involved can be arranged in a more compact notation, which we illustrate by an example: to divide $f(x)=x^4+3x^3-5x-10$ by $x-2$ we write

$$\begin{array}{r|rrrr|r}
 & 1 & 3 & 0 & -5 & -10 \\
2 & & 2 & 10 & 20 & 30 \\
\hline
 & 1 & 5 & 10 & 15 & 20
\end{array}$$

and conclude that $x^4+3x^3-5x-10=(x^3+5x^2+10x+15)(x-2)+20$. This method of performing the division is called *Ruffini's method* (or *Ruffini's rule*). (An extension of it exists, called *synthetic division,* which allows division by any monic polynomial rather than a special binomial $x-a$.)

According to the Factor Theorem, the remainder 20 of the division equals $f(2)$, and so this algorithm can also be used to *evaluate* a polynomial $f(x)$ on a number a (that is, to compute $f(a)$). This algorithm (called *Horner scheme,* but equivalent to Ruffini's rule), which really amounts to rewriting $x^4+3x^3-5x-10$ as

$$((((x+3)x+0)x-5)x-10,$$

is more efficient than the obvious one (computing the various powers of 2, multiplying them by the corresponding coefficients, and then adding up the results), as it requires the same number of addition, but about half as many multiplications. Not a drastic saving, but significant.

Exercise. For a generic polynomial $f(x)$ of degree n (that is, avoiding special cases where some coefficient is zero), exactly how many additions and how many multiplications are required to compute $f(a)$ by the two methods?

Example. Continuing with a previous example, we already know that $f(x)=x^n-a^n$ is divisible by $x-a$. In fact, dividing by means of Ruffini's rule we find remainder zero:

$$\begin{array}{c|ccccc|c}
 & 1 & 0 & 0 & \cdots & 0 & -a^n \\
a & & a & a^2 & \cdots & a^{n-1} & a^n \\
\hline
 & 1 & a & a^2 & \cdots & a^{n-1} & 0
\end{array}$$

But Ruffini's rule also gives us the quotient, and so we find

$$x^n-a^n=(x-a)(x^{n-1}+ax^{n-2}+\cdots+a^{n-2}x+a^{n-1}).$$

Of course this identity could be easily verified directly by executing the multiplication on the RHS, but the point is that Ruffini's rule produces the quotient very quickly. When

n is odd (and only then) a similar identity

$$x^n + a^n = (x+a)(x^{n-1} - ax^{n-2} + \cdots - a^{n-2}x + a^{n-1})$$

can be obtained by applying Ruffini's rule to divide $f(x) = x^n + a^n$ by $x + a$, but it is quicker to deduce the identity by replacing a with $-a$ in the previous identity.

14. Expansion of a polynomial in terms of $x - a$

The following example illustrates how iterating Ruffini's rule we can expand a polynomial in x into powers of $x - a$, that is, if we like, into *a polynomial in $x - a$*. Say we want to expand $f(x) = x^3 + 2x^2 - x - 3$ into powers of $x - 2$. Then we do the following: we divide $f(x)$ by $x - 2$, then we divide the resulting quotient by $x - 2$, then we divide the resulting quotient by $x - 2$, and so on until the quotient is zero.

$$\begin{array}{r|rrrr}
 & 1 & 2 & -1 & -3 \\
2 & & 2 & 8 & 14 \\
\hline
 & 1 & 4 & 7 & \mathbf{11} \\
2 & & 2 & 12 & \\
\hline
 & 1 & 6 & \mathbf{19} & \\
2 & & 2 & & \\
\hline
 & 1 & \mathbf{8} & & \\
2 & & & & \\
\hline
 & \mathbf{1} & & &
\end{array}$$

The final result of this calculation is that

$$x^3 + 2x^2 - x - 3 = 1(x-2)^3 + 8(x-2)^2 + 19(x-2) + 11.$$

The reason why it works is that the term 11 can be obtained as the remainder of dividing the polynomial by $x - 2$, which is done via Ruffini's rule. The quotient $x^2 + 4x + 7$ of this division is eventually going to be written as $1(x-2)^2 + 8(x-2) + 19$, and hence 19 can be obtained as the quotient of dividing it by $x - 2$. And so on.

Of course an alternative way of expanding a polynomial into powers of $x - a$ is substituting $x = y + a$ into it, then expanding the various powers of $y + a$ involved, thus converting it into a polynomial in y after the appropriate simplifications, and finally set $y = x - a$. This procedure, however, involves more operations and hence is computationally less efficient.[5]

[5] The way described of expanding a polynomial in terms of $x - a$ is analogous to the efficient way to convert an integer from decimal to another base b, which was described earlier in the notes.

15. Divisibility, GCD, and the Euclidean algorithm for polynomials

Divisibility, GCD and lcm, and the Euclidean algorithm, work for polynomials with coefficients in a field F very much the same as they do for integers, with only small adjustments. To begin with, divisibility, divisors, etc., are defined in the same way:

Definition 23 (Divisibility for polynomials). Let $f(x)$ and $g(x)$ be polynomials with coefficients in a field F. We say that $g(x)$ divides $f(x)$, and we write $g(x) \mid f(x)$, if there is a polynomial $h(x) \in F[x]$ such that $f(x) = g(x) \cdot h(x)$.

With polynomials, if $f(x) \mid g(x)$ and $g(x) \mid f(x)$, then $g(x) = c \cdot f(x)$ for some nonzero constant c. In fact, the nonzero constants play the same role in $F[x]$ as ± 1 play in $\mathbb{Z}$: they are exactly the elements which are *invertible,* that is, which have an *inverse* (belonging to the same set). This means that the only polynomials $c(x)$ such that there is a polynomial $d(x)$ with $c(x) \cdot d(x) = 1$, are exactly the nonzero constant polynomials $c(x)$ (which, being constants, we may simply write as c). To show this rigorously, taking the degrees in the equality $c(x) \cdot d(x) = 1$ gives us $\deg\bigl(c(x)\bigr) + \deg\bigl(d(x)\bigr) = 0$, which can only happen if $\deg\bigl(c(x)\bigr) = \deg\bigl(d(x)\bigr) = 0$, that is, $c(x)$ is a nonzero constant polynomial (and so is $d(x)$).

The greatest common divisor of two polynomials $f(x)$ and $g(x)$ is defined in the same way as for the integers:

Definition 24 (Greatest common divisor). Let $f(x)$ and $g(x)$ be polynomials with coefficients in a field F (hence $f(x), g(x) \in F[x]$, in a formula). A polynomial $d(x) \in F[x]$ is called *a greatest common divisor* of $f(x)$ and $g(x)$ if

(1) $d(x)$ divides $f(x)$ and $g(x)$, and
(2) if $c(x) \in F[x]$ is any polynomial which divides both $f(x)$ and $g(x)$, then $c(x)$ divides $d(x)$.

Remark. Beware that this is different from the definition given in Part II, Chapter 2, of the recommended book by Childs. The definition given there is more in the style of a ‘school’ definition, but the one we give here has the advantage of being (essentially) the same for integers, for polynomials, and for other contexts: whenever you have a concept of divisibility, there is a concept of *a* greatest common divisor (which need not be precisely unique).

A greatest common divisor of $f(x)$ and $g(x)$, denoted by $\bigl(f(x), g(x)\bigr)$, is only unique up to multiplying it by a nonzero constant. Hence saying that the GCD of two polynomials is $x+3$ is equivalent to saying that it is $2x+6$, or $\frac{1}{3}x+1$, etc. Among all those equivalent GCD’s one usually chooses the one which is monic. (This is similar to choosing the positive greatest common divisors of two integers, rather than its opposite.) As we do for integers, if two polynomials $f(x)$ and $g(x)$ have greatest common divisor 1 then we say that they are *coprime*.

The Euclidean algorithm and the extended Euclidean algorithm work for polynomials in the same way as for the integers. Hence, given two polynomials $f(x)$ and $g(x)$, with $\deg\bigl(f(x)\bigr) \geq \deg\bigl(g(x)\bigr)$ (otherwise we just swap the two polynomials), we start the algorithm by dividing the first polynomials by the second:

$$f(x) = g(x)q_1(x) + r_1(x), \qquad \text{with } \deg\bigl(r_1(x)\bigr) < \deg\bigl(g(x)\bigr).$$

Then we divide $g(x)$ by the first remainder $r_1(x)$,

$$g(x) = r_1(x)q_2(x) + r_2(x), \qquad \text{with } \deg\bigl(r_2(x)\bigr) < \deg\bigl(r_1(x)\bigr),$$

and repeat the procedure until some remainder is zero. The last nonzero remainder is then the GCD of the two polynomials. As in the case of integers, this is justified by noting the following basic fact: if $f(x) = g(x)q(x) + r(x)$ then $\mathrm{GCD}\bigl(f(x), g(x)\bigr) = \mathrm{GCD}\bigl(g(x), r(x)\bigr)$. Hence if $r_i(x)$ is the last nonzero remainder, then $\mathrm{GCD}\bigl(f(x), g(x)\bigr) = \mathrm{GCD}\bigl(g(x), r_1(x)\bigr) = \mathrm{GCD}\bigl(r_1(x), r_2(x)\bigr) = \cdots = \mathrm{GCD}\bigl(r_i(x), 0\bigr) = r_i(x)$.

The number of divisions required by the Euclidean algorithm on polynomials is at most the lower of the degrees of the two polynomials. This is easy to see as the degree of each remainder is less than the degree of the previous remainder.

Example. We compute the GCD of the polynomials $x^3 + 2x^2 + x$ and $x^2 + x - 1$ using the Euclidean algorithm:

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

The remainder of the second division is -1, so there is no point in doing a third division, as dividing by -1 (or by 1, or $2/3$, or any nonzero rational number) would give remainder zero. Hence the last nonzero remainder is -1, and so GCD of $x^3 + 2x^2 + x$ and $x^2 + x - 1$ is 1. In words, those polynomials are coprime.

Example. We compute the GCD of the polynomials $x^{3n} - 1$ and $x^{2n} - 1$, where n is any positive integer. The Euclidean algorithm reads:

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

Hence the last nonzero remainder is $x^n - 1$, and so GCD of $x^{3n} - 1$ and $x^{2n} - 1$ is $x^n - 1$.

We should have known from the start that $x^n - 1$ divides both polynomials. In fact $x^{3n} - 1 = (x^n - 1)(x^{2n} + x^n + 1)$ follow from the general identity $x^3 - a^n = (x - a)(x^2 + ax + a^2)$. Similarly, we should have known that $x^{2n} - 1 = (x^n - 1)(x^n + 1)$. Knowing these factorisations, another way to prove that $x^2 - 1$ is actually the *greatest* common divisor of $x^{3n}-1$ and $x^{2n}-1$ (rather than just *a* common divisor) would then be showing that $x^{2n}+x^n+1$ and x^n+1 are coprime. Of course we can do that by applying the Euclidean algorithm, which in this case consists of a single division: $x^{2n}+x^n+1=(x^n+1)\cdot x^n+1$. This tells us that their GCD is 1, and so the GCD of $x^{3n}-1$ and $x^{2n}-1$ is x^n-1.

Remark. One can actually prove that for any positive integers m and n the greatest common divisor of x^m-1 and x^n-1 is $x^{(m,n)}-1$ (where the exponent (m,n) means the GCD of m and n, as usual).

The conclusion of the extended Euclidean algorithm can be formally stated as Bézout's Lemma for polynomials:

Lemma 25 (Bézout's Lemma for polynomials). *Let $f(x), g(x) \in F[x]$, where F is a field, and let $d(x) = \big(f(x), g(x)\big)$ be their greatest common divisor. Then there exist polynomials $u(x), v(x) \in F[x]$ such that*

$$f(x)\,u(x) + g(x)\,v(x) = d(x).$$

It is not difficult to show that if neither $f(x)$ or $g(x)$ is the zero polynomial then the polynomials $u(x)$ and $v(x)$ produced by the extended Euclidean algorithm satisfy

$$\deg\big(u(x)\big) < \deg\big(g(x)\big) \qquad \text{and} \qquad \deg\big(v(x)\big) < \deg\big(f(x)\big).$$

Example. Reading the divisions in the previous example backwards we find:

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

So we have found find two polynomials $u(x)$ and $v(x)$ such that

$$(x^3+2x^2+x)\cdot u(x) + (x^2+x-1)\cdot v(x) = 1,$$

namely,

$$u(x) = x, \quad \text{and} \quad v(x) = -x^2-x-1.$$

Note that these polynomials satisfy

$$\frac{1}{(x^3+2x^2+x)(x^2+x-1)} = \frac{u(x)}{x^2+x-1} + \frac{v(x)}{x^3+2x^2+x}.$$

Hence the extended Euclidean algorithm has allowed us to write the fraction on the LHS as a sum of the two 'simpler' fractions on the RHS. Note that in each of the two fractions on the RHS the numerator has degree strictly less than the denominator. This is a particular instance of the general fact on the degrees of $u(x)$ and $v(x)$ mentioned before the example.

Example. We compute the *monic* greatest common divisor $d(x)$ of $x^3 - x^2 + x - 6$ and $x^3 + x - 10$:

$$\begin{aligned} x^3 - x^2 + x - 6 &= (x^3 + x - 10) \cdot 1 + (-x^2 + 4) \\ x^3 + x - 10 &= (x^2 - 4) \cdot x + (5x - 10) \\ x^2 - 4 &= (x - 2)(x + 2). \end{aligned}$$

The last nonzero remainder is $5x - 10 = 5(x - 2)$, and so the *monic* GCD is $d(x) = (x^3 - x^2 + x - 6, x^3 + x - 10) = x - 2$. Of course it would also be correct to say that a GCD is $5x - 10$, as much as $\frac{1}{2}x - 2$, or $-\frac{2}{3}x + \frac{4}{3}$, etc., but as a standard way of choosing one we have asked for the *monic* GCD, which is $x - 2$.

Now we carry out the extended part of the Euclidean algorithm, by reading those divisions backwards, and we find

$$\begin{aligned} 5x - 10 &= (x^3 + x - 10) - (x^2 - 4) \cdot x \\ &= (x^3 + x - 10) + [(x^3 - x^2 + x - 6) - (x^3 + x - 10) \cdot 1] \cdot x \\ &= (x^3 + x - 10) \cdot x - (x^3 + x - 10) \cdot (x - 1) \end{aligned}$$

So we have found polynomials $u(x)$ and $v(x)$ whose existence is stated in Bézout's Lemma: $u(x) = \frac{1}{5}x$ and $v(x) = -\frac{1}{5}(x - 1)$ satisfy

$$(x^3 - x^2 + x - 6) \cdot u(x) + (x^3 + x - 10) \cdot v(x) = d(x) = x - 2.$$

Example. When the GCD of two polynomials is not 1, as in the previous example, there is an alternative way to proceed with the extended Euclidean algorithm in order to have simpler calculations: once we have found that $(x^3 - x^2 + x - 6, x^3 + x - 10) = x - 2$, we may divide both polynomials by their GCD $x - 2$ and factorise them as follows:

$$\begin{aligned} x^3 - x^2 + x - 6 &= (x - 2)(x^2 + x + 3) \\ x^3 + x - 10 &= (x - 2)(x^2 + 2x + 5). \end{aligned}$$

We see from these factorisations that there was nothing special about the 5 coefficients in the GCD $5x - 10$ which we found using the Euclidean algorithm on the original polynomial, as there is no trace of that in the factorisations.

Now we carry out the extended Euclidean algorithm using the quotients by $x - 2$ instead of our original polynomials:

$$\begin{aligned} x^2 + x + 3 &= (x^2 + 2x + 5) \cdot 1 + (-x - 2) \\ x^2 + 2x + 5 &= (x + 2) \cdot x + 5. \end{aligned}$$

Hence the GCD of those polynomials is 1 (or 5, or 2/3 if you like, they all mean the same in this polynomial context, because all nonzero constants are invertible in $\mathbb{Q}[x]$). Reading the divisions backwards we find:

$$\begin{aligned}5 &= (x^2+2x+5) - (x+2)\cdot x\\ &= (x^2+2x+5) + [(x^2+x+3) - (x^2+2x+5)\cdot 1]\cdot x\\ &= (x^2+x+3)\cdot x - (x^2+2x+5)\cdot (x-1).\end{aligned}$$

So we have found the same $u(x) = \frac{1}{5}x$ and $v(x) = -\frac{1}{5}(x-1)$ as before. In fact, those polynomials satisfy

$$(x^2+x+3)\cdot u(x) + (x^2+2x+5)\cdot v(x) = 1,$$

and if we multiply both sides of this equality by $x-2$ we recover

$$(x^3-x^2+x-6)\cdot u(x) + (x^3+x-10)\cdot v(x) = x-2,$$

which we obtained the first time.

Note that *it would not be possible* to find polynomials $s(x)$ and $t(x)$ such that

$$(x^3-x^2+x-6)\cdot s(x) + (x^3+x-10)\cdot t(x) = 1,$$

because x^3-x^2+x-6 and x^3+x-10 are not coprime. In fact, each of them is a multiple of $x-2$, so the left-hand side is as well, but the right-hand side 1 is not, so that is impossible.

Lecture notes of Algebra. Week 5

16. Irreducible polynomials, and unique factorisation

The four Arithmetical Lemmas for integers remain true for polynomials, after making the obvious changes in terminology, and can be proved in the same way using Bézout's Lemma. In particular, the most important of them, Arithmetical Lemma B, is as follows: *if the polynomials $f(x)$ and $g(x)$ are coprime, and $f(x)$ divides the product $g(x) \cdot h(x)$, then $f(x)$ divides $h(x)$.*

Prime polynomials are usually rather called *irreducible* polynomials. As was the case for the integers, one has to exclude the zero polynomial and the invertible polynomials (which are the analogues of ± 1 in the integers) from the definition of irreducible. Because the invertible polynomials are the nonzero constants (the polynomials of degree zero), the definition of irreducible will only apply to non-constant polynomials, that is, to polynomials of positive degree.

Definition 26. A non-constant polynomial $f(x) \in F[x]$ is *reducible* in $F[x]$ if it can be written as $f(x) = g(x)\,h(x)$, where $g(x)$ and $h(x)$ are polynomials in $F[x]$ of positive degree (or, equivalently, $\deg(g(x))$ and $\deg(h(x))$ are smaller than $\deg(f(x))$; or, equivalently again, where $0 < \deg(g(x)) < \deg(f(x))$); it is *irreducible* in $F[x]$ if it is not reducible.

Because $\deg\bigl(g(x)\,h(x)\bigr) = \deg\bigl(g(x)\bigr)$, any polynomial of degree 1, hence of the form $ax + b$ with $a \neq 0$, is always irreducible, as 1 cannot be written as a sum of two positive integers.

Note that the notions of reducible and irreducible depend on the field in which we view the coefficients of our polynomial: $x^2 + 1$ is irreducible as a polynomial in $\mathbb{R}[x]$, but not as a polynomial in $\mathbb{C}[x]$, because $x^2 + 1 = (x - i)(x + i)$. To stress which field F is being used, one usually specifies *irreducible in $F[x]$*, or also *irreducible over F*. (Hence $x^2 + 1$ is irreducible over $\mathbb{R}$, but reducible over $\mathbb{C}$.)

Theorem 14 on unique factorisation in the integers has an analogue for polynomials.

Theorem 27 (Unique Factorisation Theorem for polynomials). *Every polynomial of positive degree (which is the same as saying* non-constant*) over a field F factorises into a product of irreducible polynomials (irreducible over the same field F).*

Also, the factorisation is essentially *unique, namely, unique up to permuting the factors, but also to multiplying each irreducible factor by some nonzero constant (that is, by some invertible polynomial).*

For example,

$$\begin{aligned} 2x^2+10x+12 &= 2(x+2)(x+3) = (2x+4)(x+3) = (x+2)(2x+6) \\ &= (3x+6)\Big(\frac{2}{3}x+2\Big), \qquad \text{and so on.} \end{aligned}$$

We call the factorisation into a product of irreducible polynomials the *complete factorisation* of $f(x)$ over F (or in $F[x]$). Note that, once again, with polynomials it is essential to state over which field we are working, because the answer may be different over different fields.

Example. The polynomial $x^4-3x^2+2 \in \mathbb{Q}[x]$ can be written in three essentially different ways

$$x^4-3x^2+2 = (x^2-1)(x^2-4) = (x^2-3x+2)(x^2+3x+2) = (x^2-x-2)(x^2+x-2)$$

as the product of two polynomials of degree 2. However, this does not contradict the above theorem on unique factorisation because none of those quadratic factors is irreducible over $\mathbb{Q}$, in fact $x^4-3x^2+2 = (x-1)(x+1)(x-2)(x+2)$, and the above quadratic factors are obtained by pairing and multiplying together the irreducible factors in different ways.

Example. The polynomial x^3-2 is irreducible in $\mathbb{Q}[x]$. (This will be justified in later sections.) In $\mathbb{R}[x]$ it factorises as

$$x^3-2 = (x-\sqrt[3]{2})(x^2+\sqrt[3]{2}\,x+\sqrt[3]{2}^2).$$

This is a complete factorisation in $\mathbb{R}[x]$ because the quadratic factor is irreducible. In fact, its discriminant is $\sqrt[3]{2}^2 - 4\cdot\sqrt[3]{2}^2 = -3\sqrt[3]{2}^2 < 0$, and so that quadratic polynomial has no real roots. However, in $\mathbb{C}[x]$ the polynomial x^3-2 factorises as

$$x^3-2 = (x-\sqrt[3]{2})(x-\omega\sqrt[3]{2})(x-\bar{\omega}\sqrt[3]{2}),$$

where $\omega = (-1 \pm i\sqrt{3})/2$.

17. Quadratic polynomials

You should know well from school how to find the roots of a quadratic polynomial ax^2+bx+c (hence with $a \neq 0$, otherwise it would not be quadratic), which means the same as finding the solutions of the corresponding equation $ax^2+bx+c=0$. In fact, the trick of *completing the square* ax^2+bx at the left-hand side brings the equation to the equivalent form [6]

$$a\left(x+\frac{b}{2a}\right)^2 = \frac{b^2-4ac}{4a}.$$

[6] To be precise, this works over any field $\mathbb{F}$ where $2 \neq 0$. The meaning of this unfamiliar condition will be clarified in a later algebra course, but note that it is not satisfied when $F = \mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$, the field of 2 elements, where $[2] = [1+1]$ is the same as $[0]$.

The numerator of the fraction at the right-hand side is the *discriminant* of the quadratic equation (or polynomial), and the existence of the solutions depends on whether it is a square in the field F under consideration (which is the same as being nonnegative if $F = \mathbb{R}$, but may be a different condition otherwise). If $b^2 - 4ac$ is not a square of an element of F, then no $x \in F$ can make the above equality true, and so the polynomial has no root in F. If $b^2 - 4ac$ is a square of an element of F, which means that it has a square root in F, denote it by $\sqrt{b^2 - 4ac}$, then the equation becomes

$$\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\sqrt{b^2 - 4ac}}{2a}\right)^2 = 0$$

and, in turn,

$$\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right) = 0.$$

In conclusion, one may distinguish two cases:

- if $b^2 - 4ac$ is not the square of an element of F, and so does not have a square root in F, then the polynomial has no root in F;
- if $b^2 - 4ac$ has a square root in F, then the polynomial has roots given by the familiar formula
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a};$$
of course the two roots coincide when $b^2 - 4ac = 0$ (which is sometimes stated separately as a third case). [7]

These cases can be expressed in terms of reducibility (over a field F containing all coefficients) of the quadratic polynomial $ax^2 + bx + c$ (with $a \neq 0$):

- it is irreducible if $b^2 - 4ac$ does not have a square root in F;
- it is reducible if $b^2 - 4ac$ has a square root in F; in fact in that case we have $ax^2 + bx + c = a(x - \alpha_1)(x - \alpha_2)$, where $\alpha_1, \alpha_2 = (-b \pm \sqrt{b^2 - 4ac})/(2a)$ are its roots (in any order); of course $\alpha_1 = \alpha_2$ when $b^2 - 4ac = 0$.

Example. A quadratic polynomial $ax^2+bx+c \in \mathbb{R}[x]$ (hence with $a \neq 0$) is irreducible exactly when its discriminant b^2-4ac has no square roots in $\mathbb{R}$, hence exactly when b^2-4ac is negative.

Example. Because any complex number has square roots in $\mathbb{C}$, quadratic polynomials in $\mathbb{C}[x]$ are always reducible (and hence factorise completely into products of two polynomials of degree one).

[7] Note that whenever it is convenient to collect a factor 2 from the coefficient b, for example when b is an even integer, or, say, $b = 6\sqrt{5} = 2 \cdot 3\sqrt{5}$, etc., it may be easier to use the slightly simpler formula $x = \left(-B \pm \sqrt{B^2 - ac}\right)/a$ for the roots of the polynomial $ax^2 + 2Bx + c$ (that is, where $b = 2B$).

Example. The polynomial x^2-2 is irreducible over $\mathbb{Q}$, but reducible over $\mathbb{R}$, because $x^2-2=(x-\sqrt{2})(x+\sqrt{2})$, and $\sqrt{2}\notin\mathbb{Q}$, which means that $\sqrt{2}$ is irrational (see below, or as an application of the Rational Root Test later).

(Optional) Proof that $\sqrt{2}$ is irrational. We need to show that $\sqrt{2}$ cannot be equal to any fraction a/b of integers. Suppose it is. (In these cases one says *suppose, for a contradiction,* and then try to show that this assumption does lead to a contradiction, meaning, proving that some fact would have to be both true and false.)

In that case the fraction can certainly be reduced to simplest terms, and so we may assume that $\sqrt{2}=a/b$, with $a,b\in\mathbb{Z}$, clearly with $b\neq 0$, and the further condition $(a,b)=1$. Multiplying by b and squaring we find $2b^2=a^2$. Hence 2 divides $a^2=a\cdot a$, but because 2 is prime it follows that 2 divides a, and hence $a=2c$ for some integer c. Substituting into our equation we find $2b^2=4c^2$, whence $b^2=2c^2$. Hence 2 divides b^2, and because 2 is prime it divides b. So we have found that 2 divides both a and b, and hence 2 divides $(a,b)=1$. But this is certainly false, and we have found the desired contradiction. (We have concluded that 2 divides 1, but at the same time we know that 2 does not divide 1.)

The only way to resolve the contradiction is to admit that we made a false assumption at the beginning, namely, in assuming that $\sqrt{2}$ is equal to some fraction a/b of integers. Hence this is not possible, which means that $\sqrt{2}$ is irrational. □

18. The maximum number of roots of a polynomial

How many roots does a polynomial have? Well, the zero polynomial has all the roots we want (any number is a root), so we look at non-zero polynomials. A polynomial of degree 1 has always exactly one root, namely, $-b/a$ if the polynomial is $ax+b$ (with $a\neq 0$ otherwise it would not have degree 1). For a polynomial of degree 2 the answer may depend on the field where we are viewing the coefficients: it may have two roots, or one root (which we may think of a double root, but we count only once if we meant to ask about how many *distinct* roots), or none. In any case, at most two. Here is a more general result, which we can prove using the Factor Theorem.

Theorem 28. *A polynomial of degree $n\geq 0$ over a field F has at most n roots in F.*

Proof. Let f be a polynomial of degree $n\geq 0$. If f has no roots at all in F, we are done, as $0\leq n$ (so the statement is correct in this case). If f has (at least) a root α_1, then according to the Factor Theorem (see Lemma 22) we have

$$f(x)=(x-\alpha_1)\cdot f_2(x)$$

for some polynomial $f_2(x)$. Now it may happen that $f_2(x)$ has no roots (and then $f(x)$ has exactly one root, so the statement is correct because $1\leq n$). But if $f_2(x)$ does have a root α_2 (possibly equal to α_1), then

$$f_2(x) = (x - \alpha_2) \cdot f_3(x),$$

whence

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdot f_3(x).$$

Continuing in this way, sooner or later we arrive at

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_m) \cdot f_{m+1}(x),$$

where $f_{m+1}(x)$ has no roots in F (possibly because it is a constant). In fact, this must occur for some $m \leq \deg(f) = n$, because taking degrees in the above equality we find $\deg(f) = m + \deg(f_{m+1}) \geq m$.

Now let β be a root of $f(x)$. To complete the proof it will be enough to show that β equals one among $\alpha_1, \ldots, \alpha_m$, of which there are at most m distinct ones (possibly fewer!), and consequently at most n as we have just shown $m \leq n$. In fact, if $f(\beta) = 0$ then

$$0 = f(\beta) = (\beta - \alpha_1) \cdots (\beta - \alpha_m) \cdot f_{m+1}(\beta).$$

We have $f_{m+1}(\beta) \neq 0$ because $f_{m+1}(x)$ has no roots in F, and so at least one other factor in the above product must vanish, say $\beta - \alpha_j$, and hence $\beta = \alpha_j$, as desired. □

In particular, any nonzero polynomial has finitely many roots. This behaviour of polynomial functions is different from other familiar functions, for example the function $f(x) = \sin(x)$ has infinitely many real zeroes, namely,

$$\sin(x) = 0 \qquad \Leftrightarrow \qquad x = 2\pi k \quad \text{for} \quad k \in \mathbb{Z}.$$

Corollary 29. *A polynomial $f(x)$ of degree n is uniquely determined by the values it takes on $n + 1$ distinct elements of F.*

Proof. We are assuming that for $n + 1$ distinct numbers $b_1, \ldots, b_{n+1}$ we know the values

$$f(b_1) = c_1, \quad f(b_2) = c_2, \quad \ldots \quad f(b_{n+1}) = c_{n+1}.$$

Suppose $g(x)$ is any polynomial of degree n which satisfies

$$g(b_1) = c_1, \quad g(b_2) = c_2, \quad \ldots \quad g(b_{n+1}) = c_{n+1}.$$

Then the difference $h(x) = f(x) - g(x)$ is either zero or a nonzero polynomial of degree *at most* n, and it satisfies

$$h(b_1) = 0, \quad h(b_2) = 0, \quad \ldots \quad h(b_{n+1}) = 0.$$

Hence $h(x)$ has at least $n + 1$ roots, while a nonzero polynomial of degree at most n has at most n roots. Consequently, $h(x)$ can only be the zero polynomial, and hence $g(x) = f(x)$. □

An important consequence of the above corollary is that two different polynomials $f(x), g(x) \in \mathbb{R}[x]$ give rise to different functions $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$. Consequently, the apparently simpler alternate definition of a polynomial as a function $f\colon \mathbb{R} \to \mathbb{R}$ of a particular shape (that is, which can be written as $f(x) = a_n x^n + \cdots + a_1 x + a_0$) is actually equivalent to the one we are using (so we can identify polynomials with the functions they gives rise to), but only *if we work over an infinite field,* such as $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$.

Example. Consider the polynomials $f(x) = \bar{1}x = x$ and $g(x) = \bar{1}x^2 = x^2$, where the coefficients belong to the field of two elements $\mathbb{F}_2$. According to our definition of polynomials they are different polynomials (because their coefficients are different; and actually even their degrees) but give rise to the same function $\mathbb{F}_2 \to \mathbb{F}_2$, because

$$f(\bar{0}) = \bar{0} = g(\bar{0}), \qquad \text{and} \qquad f(\bar{1}) = \bar{1} = g(\bar{1}).$$

19. Polynomial interpolation

Given real numbers $\alpha_1, \alpha_2, \beta_1, \beta_2$, with $\alpha_1 \neq \alpha_2$, there is a unique polynomial $f(x)$ of degree at most one (hence a linear polynomial or a constant polynomial), such that

$$f(\alpha_1) = \beta_1, \quad \text{and} \quad f(\alpha_2) = \beta_2.$$

In fact, because $f(x) = ax + b$ for some $a, b \in \mathbb{R}$, the required conditions amount to

$$\begin{cases} a \cdot \alpha_1 + b = \beta_1 \\ a \cdot \alpha_2 + b = \beta_2 \end{cases}$$

Solving this system we would find

$$a = \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1}, \quad \text{and then} \quad b = \beta_1 - \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1},$$

and hence there is a unique solution, which can be written as

$$f(x) = \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1} \cdot (x - \alpha_1) + \beta_1.$$

The following result generalises this fact to an arbitrary number of values. We state it for complex numbers rather than real numbers, but we could as well take for $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ elements of any field (the same field for all of them).

Theorem 30 (Interpolation theorem). *Given n distinct complex numbers $\alpha_1, \ldots, \alpha_n$, and n arbitrary (not necessarily distinct) complex numbers $\beta_1, \ldots, \beta_n$, there is a unique polynomial $f(x)$ of degree less than n such that*

$$f(\alpha_1) = \beta_1, \quad \ldots, \quad f(\alpha_n) = \beta_n.$$

(Optional) Proof. We have already proved the uniqueness of $f(x)$ in Corollary 29. There are several ways to prove the existence of $f(x)$, both direct and indirect (and hence not explicit). We present an explicit proof, based on *Lagrange interpolation.*

The polynomial

$$(x-\alpha_2)(x-\alpha_3)\cdots(x-\alpha_n)=\frac{(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)}{x-\alpha_1}$$

has $\alpha_2, \alpha_3, \ldots, \alpha_n$ as roots, and so it takes the value 0 on $\alpha_2, \alpha_3, \ldots, \alpha_n$. On α_1 it takes the value $(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)\cdots(\alpha_1-\alpha_n)$. Consequently, the polynomial

$$p_1(x)=\frac{(x-\alpha_2)(x-\alpha_3)\cdots(x-\alpha_n)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)\cdots(\alpha_1-\alpha_n)}$$

satisfies

$$p_1(\alpha_1)=1, \quad p_1(\alpha_2)=0, \quad \ldots, \quad p_1(\alpha_n)=0.$$

Similarly, we can construct a polynomial

$$p_1(x)=\frac{(x-\alpha_1)\quad(x-\alpha_3)(x-\alpha_4)\cdots(x-\alpha_n)}{(\alpha_2-\alpha_1)\quad(\alpha_2-\alpha_3)(\alpha_2-\alpha_4)\cdots(\alpha_2-\alpha_n)}$$

(where $x-\alpha_2$ is omitted from the numerator, and $\alpha_2-\alpha_2$ is correspondingly omitted from the denominator), which takes the value 1 on α_2 and vanishes (that means, it takes the value zero) on $\alpha_1, \alpha_3, \alpha_4, \ldots, \alpha_n$. After finding $p_3(x), \ldots, p_n(x)$ with similar properties, we see that the polynomial

$$f(x)=\beta_1\cdot p_1(x)+\cdots+\beta_1\cdot p_1(x)$$

satisfies

$$f(\alpha_1)=\beta_1, \quad \ldots, \quad f(\alpha_n)=\beta_n,$$

as required, and has degree at most $n-1$ because each of $p_1(x), \ldots, p_n(x)$ has degree exactly $n-1$. □

Each polynomial $p_k(x)$ in the above proof is obtained as

$$p_k(x)=g_k(x)/g_k(\alpha_k), \quad \text{where} \quad g_k(x)=\prod_{j\neq k}(x-\alpha_j).$$

Note that it may be quicker to compute $(x-\alpha_1)\cdots(x-\alpha_n)$ first, a polynomial of degree n, and then obtain each $g_k(x)$ by dividing that by $x-\alpha_k$, which can be efficiently done using Ruffini's rule. [8]

[8] A rough way to see why is noting that computing $(x-\alpha_1)\cdots(x-\alpha_n)$ requires $n-1$ polynomial multiplications, and on the way to the conclusion we would have already computed one of the $g_k(x)$, say $g_n(x)=(x-\alpha_1)\cdots(x-\alpha_{n-1})$. Then it remains to obtain the remaining $g_k(x)$ by applying Ruffini's rule $n-1$ times. This total of $2n-2$ multiplications/divisions should be contrasted with the $n(n-1)$ multiplications required to compute each $g_k(x)$ separately. (More precise computational estimates would take into account that polynomial multiplications may take different times depending on the size of the factors, but then we would still observe a similar difference in speed of the two methods.)

Example. Find the unique polynomial $f(x)$ of degree at most three, such that

$$f(-2) = -5, \qquad f(-1) = 3, \qquad f(1) = 1, \qquad f(2) = 3.$$

The polynomial

$$p_1(x) = \frac{(x+1)(x-1)(x-2)}{(-2+1)(-2-1)(-2-2)} = -\frac{1}{12}(x^3 - 2x^2 - x + 2)$$

has -1, 1 and 2 as roots and satisfies $p_1(-2) = 1$. The polynomial

$$p_2(x) = \frac{(x+2)(x-1)(x-2)}{(-1+2)(-1-1)(-1-2)} = \frac{1}{6}(x^3 - x^2 - 4x + 4)$$

has -2, 1 and 2 as roots and satisfies $p_2(-1) = 1$. The polynomial

$$p_3(x) = \frac{(x+2)(x+1)(x-2)}{(1+2)(1+1)(1-2)} = -\frac{1}{6}(x^3 + x^2 - 4x - 4)$$

has -2, -1 and 2 as roots and satisfies $p_3(1) = 1$. The polynomial

$$p_4(x) = \frac{(x+2)(x+1)(x-1)}{(2+2)(2+1)(2-1)} = \frac{1}{12}(x^3 + 2x^2 - x - 2)$$

has -2, -1 and 1 as roots and satisfies $p_4(2) = 1$.

Note that because of the particular symmetry of our problem, $p_3(x)$ and $p_4(x)$ could have also been obtained as $p_3(x) = p_2(-x)$ and $p_4(x) = p_1(-x)$. Of course it will not be so in general.

We conclude that

$$\begin{aligned} f(x) &= -5 \cdot p_1(x) + 3 \cdot p_2(x) + p_3(x) + 3 \cdot p_4(x) \\ &= \frac{5}{12}x^3 - \frac{5}{6}x^2 - \frac{5}{12}x + \frac{5}{6} \\ &\quad + \frac{1}{2}x^3 - \frac{1}{2}x^2 - 2x + 2 \\ &\quad - \frac{1}{6}x^3 - \frac{1}{6}x^2 + \frac{2}{3}x + \frac{2}{3} \\ &\quad + \frac{1}{4}x^3 + \frac{1}{2}x^2 - \frac{1}{4}x - \frac{1}{2} \\ &= x^3 - x^2 - 2x + 3. \end{aligned}$$

Example. Find the unique polynomial $g(x)$ of degree at most three, such that

$$g(-2) = 1, \qquad g(-1) = -1, \qquad g(1) = -1, \qquad g(2) = 1.$$

We can reuse the calculations of the previous example, and find

$$f(x) = p_1(x) - p_2(x) - p_3(x) + p_4(x) = \frac{2}{3}x^2 - \frac{5}{3}.$$

Hence this time the required polynomial has actually degree two. According to the interpolation theorem, Theorem 30, it is the unique polynomial of degree *less than four,* which is the same as *at most three,* which satisfies the given conditions. Of course, had we known that it has degree at most two, it would have been uniquely determined by any three of those four conditions, again according to the interpolation theorem.

20. Irreducibility and roots for quadratic and cubic polynomials

Consider a polynomial $f(x)$, of positive degree, with coefficients in a field F. Recall that, according to the Factor Theorem an element α of F is a root of $f(x)$ (that is, $f(\alpha) = 0$) exactly when $x - \alpha$ is a factor of $f(x)$ (that is, it divides $f(x)$). Hence each time we find a root α of $f(x)$ we have achieved a partial factorisation of $f(x)$ as $f(x) = (x - \alpha)\, g(x)$, for some polynomial $g(x)$ (again with coefficients in F).

In particular, if a polynomial $f(x)$ of degree larger than one has a root in F, then it is reducible in $F[x]$. For polynomials of degree two or three this implication can be inverted.

Proposition 31. *A quadratic or cubic polynomial (that is, of degree two or three) over a field F is irreducible over F exactly when it does not have any root in F.*

Proof. The statement is equivalent to the following: a quadratic or cubic polynomial over a field F is reducible over F exactly when it has some root in F (meaning *at least one root in F*). Now we prove this statement.

If our polynomial, say $f(x)$, has a root α in F, then according to the Factor Theorem we have $f(x) = (x-\alpha)\, g(x)$ for some polynomial $g(x) \in F[x]$, and hence $f(x)$ is reducible over F.

Conversely, if $f(x)$ is reducible over F, then $f(x) = g(x)\, h(x)$ for some polynomials of positive degree $g(x), h(x) \in F[x]$. Because $\deg\bigl(f(x)\bigr) = \deg\bigl(g(x)\bigr) + \deg\bigl(h(x)\bigr)$, and $\deg\bigl(f(x)\bigr)$ equals two or three, then at least one of the factors, say $g(x)$, must have degree one, and hence be of the form $g(x) = ax + b$, with $a, b \in F$ and $a \neq 0$. Then $-b/a$ is a root of $g(x)$, and hence of $f(x)$. In conclusion, $f(g)$ has at least one root in F. □

Of course a polynomial of degree one, hence of the form $ax + b$ with $a, b \in F$ and $a \neq 0$, is irreducible but has always a root in F, namely $-b/a$. This criterion for being irreducible (or reducible) does not work for polynomials of degree four or higher. In fact, it is possible that a polynomial (of degree at least four) has a proper factorisation over F even if it does not have any root in F, as the following examples show.

Example. The polynomial $x^4 + 5x^2 + 4$ factorises over $\mathbb{R}$ as $(x^2 + 1)(x^2 + 4)$, but it has no real roots. In fact, its complex roots are $\pm i$ and $\pm 2i$, but none of them is real. Hence $x^4 + 5x^2 + 4$ is reducible over $\mathbb{R}$ (and, in fact, it is the product of the two irreducible polynomials $x^2 + 1$ and $x^2 + 4$), despite having no roots in $\mathbb{R}$.

Example. The polynomial x^4+1 has no roots in $\mathbb{R}$, but is not irreducible in $\mathbb{R}[x]$. More generally, we have

$$\begin{aligned} x^4+a^4 &= (x^4+2a^2x^2+a^4)-2a^2x^2 \\ &= (x^2+a^2)^2-(\sqrt{2}ax)^2 \\ &= [(x^2+a^2)-\sqrt{2}ax][(x^2+a^2)+\sqrt{2}ax] \\ &= (x^2-\sqrt{2}ax+a^2)(x^2+\sqrt{2}ax+a^2). \end{aligned}$$

Hence if $a \neq 0$ is a real number, then x^4+a^4 has no real roots, but is reducible in $\mathbb{R}[x]$, and the factorisation given here is its complete factorisation in $\mathbb{R}[x]$. For example, $x^4+4=(x^2-2x+2)(x^2+2x+2)$, a factorisation which is even in $\mathbb{Q}[x]$.

21. The Fundamental Theorem of Algebra

We know how to solve equations of degree 1 and 2. In case of quadratic equations the formula involves taking a square root (of the discriminant). More complicated formulas exist for finding the roots of polynomials of degree three and four (which means solving the corresponding equations). Those formulas (due to Del Ferro/Tartaglia/Cardano for cubics, and Ferrari/Cardano for quartics, all in 16th century) involve the usual algebraic operations together with taking square and cube roots.

However, the Norwegian mathematician Abel proved in 1824 (building on previous partial work of Ruffini) that there is no analogous formula expressing the roots of a polynomial of degree five or higher in the general case (the Abel-Ruffini Theorem). This essentially means that there are polynomials $f(x)$ of degree five (for example x^5-x-1) whose roots cannot be described by taking the coefficients of $f(x)$ and manipulating them by the usual algebraic operations together with the operations of taking nth roots (forming radicals), in the way we do for quadratic polynomials (and can be done for cubic and quartic polynomials).

Despite the impossibility of a general formula for degree larger than four, the fundamental Theorem of Algebra (first proof by Argand in 1806, more proofs by Gauss soon later) asserts that at least one complex root exixts for any non-constant polynomial.

Theorem 32 (Fundamental Theorem of Algebra). *Every polynomial in $\mathbb{C}[x]$ of positive degree has at least one root in $\mathbb{C}$.*

Corollary 33. *The irreducible polynomials in $\mathbb{C}[x]$ are those of degree one.*

Many proofs are known but none is really easy, so we will not prove the theorem.

Proof. The polynomials of degree one are always irreducible, so we only need to prove the converse: if $f(x)$ is irreducible in $\mathbb{C}[x]$, then $f(x)$ has degree one.

Because of the Fundamental Theorem of Algebra, $f(x)$ has at least one root α. By the Factor Theorem we have $f(x) = (x-\alpha)\,g(x)$, for some $g(x) \in \mathbb{C}[x]$. This cannot be a proper factorisation of $f(x)$, because $f(x)$ is irreducible, and so $g(x)$ must be a constant. Consequently, $f(x)$ equals $x-\alpha$ times a nonzero constant, and so $f(x)$ has degree one. □

Corollary 34. *Every polynomial of positive degree in $\mathbb{C}[x]$ is a product of polynomials of degree one (also called linear polynomials).*

Proof. We know that if F is any field then every polynomial of positive degree in $F[x]$ is a product of irreducible polynomials. But when $F=\mathbb{C}$ all irreducible polynomials have degree one. □

Example. One can show that the polynomial x^5-x-1 is irreducible in $\mathbb{Q}[x]$. There exists no formula for the roots (using only algebraic operations and radicals), but one can find numerically that one root is approximately 1.167 (and more precisely 1.167303978…). This is a real root, but of course, in particular, it is a complex root. According to the Factor Theorem, $x-1.167$ (approximately) divides x^5-x-1, and Ruffini's rule gives us [9]

$$\begin{array}{r|rrrrr|r}
 & 1 & 0 & 0 & 0 & -1 & -1 \\
1.167 & & 1.167 & 1.362 & 1.590 & 1.856 & -1 \\
\hline
 & 1 & 1.167 & 1.362 & 1.590 & 0.856 & 0
\end{array}$$

and so we find

$$x^5-x-1 \approx (x-1.167)(x^4+1.167x^3+1.362x^2+1.590x+0.856).$$

In turn, the factor of degree 4 has at least one complex root, and continuing in this way we eventually find the complete complex factorisation of x^5-x-1,

$$\approx (x-1.167)(x-0.181+1.083\,i)(x-0.181-1.083\,i)(x+0.764+0.352\,i)(x+0.764-0.352\,i).$$

We see that the non-real roots come in conjugate pairs (see next section), and if we multiply together the corresponding factors we find the complete factorisation of x^5-x-1 in $\mathbb{R}[x]$, which is

$$x^5-x-1 \approx (x-1.167)(x^2-0.362x+1.207)(x^2+1.529x+0.709).$$

We will see the theory of the factorisation over $\mathbb{R}$ in the next section.

[9] Here we have written only three decimal digits after the point of each number, but we have done the calculations with higher precision. In reality we will never find remainder exactly zero with a calculator, as we are using an approximation of the true root. For example, $1.167^5-1.167-1 \approx -0.0025$.

22. Roots and factorisations of a polynomial with real coefficients

Recall that any complex number α can be uniquely written in the form $\alpha = s + it$, where s and t are real numbers. For now we may actually take this as a definition of complex numbers, and define addition and multiplication by treating them like ordinary "expressions" except that every time we encounter i^2 we may replace it with -1 (so i is not a letter like any other but satisfies the "simplification rule" $i^2 = -1$). Hence complex numbers can be added, subtracted, and multiplied as follows:

$$(s + it) \pm (u + iv) = (s \pm u) + i(t \pm v),$$
$$(s + it)(u + iv) = su + i(sv + tu) + i^2 tv = (su - tv) + i(sv + tu).$$

To perform division it is useful to introduce the *conjugate* of a complex number $\alpha = s + it$, which is $\bar{\alpha} = s - it$. Because $\alpha\bar{\alpha} = (s + it)(s - it) = s^2 + t^2 = |\alpha|^2$ (where the *modulus* $|\alpha|$ is the nonnegative real number given by $|s + it| = \sqrt{s^2 + t^2}$), the reciprocal of α can be computed as $\dfrac{1}{\alpha} = \dfrac{\bar{\alpha}}{|\alpha|^2}$, and general division of complex numbers then easily follows.

Now note that complex conjugation has the following properties, which hold for any $\alpha, \beta \in \mathbb{C}$:

$$\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}, \qquad\qquad \overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}.$$

These properties essentially say that conjugation, in the sense of the map $\mathbb{C} \to \mathbb{C}$ taking $\alpha \mapsto \overline{\alpha}$, is an *automorphism* of $\mathbb{C}$. They can be verified by writing $\alpha = s+it$ and $\beta = u+iv$ and checking

$$\overline{(s + it) + (u + iv)} = \overline{(s + u) + i(t + v)} = (s + u) - i(t + v) = (s - it) + (u - iv),$$
$$\overline{(s + it)(u + iv)} = \overline{(su - tv) + i(sv + tu)} = (su - tv) - i(sv + tu) = (s - it)(u - iv).$$

Note also that a complex number α is actually real precisely when $\overline{\alpha} = \alpha$. Other properties follow, such as $\overline{\alpha - \beta} = \overline{\alpha + (-1)\beta} = \overline{\alpha} + \overline{(-1)\beta} = \overline{\alpha} + (-1)\overline{\beta} = \overline{\alpha} - \overline{\beta}$ for subtraction, and a similar one for division, $\overline{\alpha/\beta} = \overline{\alpha}/\overline{\beta}$. Also, $\overline{\alpha^2} = \overline{\alpha}^{\,2}$, and more generally $\overline{\alpha^n} = \overline{\alpha}^{\,n}$.

As an application of these basic properties of conjugation, we now show that if a complex number is a root of a polynomial with real coefficients, then its conjugate is also a root.

LEMMA 35. *If a complex number $\alpha = s + it$ is a root of a polynomial $f(x) \in \mathbb{R}[x]$, then its conjugate $\overline{\alpha} = s - it$ is a root as well.*

Proof. Write the polynomial as $f(x) = a_n x^n + \cdots + a_2 x^2 + a_1 x + a_0$, hence $a_j \in \mathbb{R}$. Then if α is any complex number, not necessarily a root of $f(x)$, we have

$$\begin{aligned} f(\overline{\alpha}) &= a_n \,\overline{\alpha}^{\,n} + \cdots + a_2 \,\overline{\alpha}^{\,2} + a_1 \,\overline{\alpha} + a_0 \\ &= a_n \,\overline{\alpha^n} + \cdots + a_2 \,\overline{\alpha^2} + a_1 \,\overline{\alpha} + a_0 && \text{(because } \overline{\alpha^n} = \overline{\alpha}^{\,n}\text{)} \\ &= \overline{a_n \,\alpha^n} + \cdots + \overline{a_2 \,\alpha^2} + \overline{a_1 \,\alpha} + \overline{a_0} && \text{(because } \overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta} \text{ and } \overline{a_j} = a_j\text{)} \\ &= \overline{a_n \,\alpha^n + \cdots + a_2 \,\alpha^2 + a_1 \,\alpha + a_0} && \text{(because } \overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}\text{)} \\ &= \overline{f(\alpha)}. \end{aligned}$$

In particular, if α is a root of $f(x)$, which means $f(\alpha) = 0$, then $f(\overline{\alpha}) = \overline{f(\alpha)} = 0$, and so $\overline{\alpha}$ is also a root. □

Note that if α is any complex number, then the polynomial

$$(x - \alpha)(x - \overline{\alpha}) = x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}$$

has real coefficients. In fact, if $\alpha = r + it$, then $\alpha + \overline{\alpha} = (r + it) + (r - it) = 2r$, and $|\alpha|^2 = \alpha\overline{\alpha} = (r + it)(r - it) = r^2 + t^2$. That quadratic polynomial has discriminant

$$(\alpha + \overline{\alpha})^2 - 4\alpha\overline{\alpha} = (\alpha - \overline{\alpha})^2 = (2it)^2 = -4t^2,$$

which is zero if α is real (and it must be so because α is a double root in that case), but is negative otherwise (and it must be so because the polynomial has no real roots in that case).

Theorem 36. *The irreducible polynomials in $\mathbb{R}[x]$ are precisely the polynomials of degree 1, and the quadratic polynomials $ax^2 + bx + c$ with $b^2 - 4ac < 0$.*

Proof. A polynomial of degree 1 is always irreducible. We have just seen that a polynomial $ax^2 + bx + c$ with $b^2 - 4ac < 0$ has no real roots. We know that for polynomials of degree two or three this implies that the polynomial is irreducible.

Now we prove the converse. Let $f(x)$ be any irreducible polynomial in $\mathbb{R}[x]$. By definition, it must have positive degree, and so by the Fundamental Theorem of Algebra it has at least one complex root α. By the Factor Theorem (in $\mathbb{C}[x]$) we have $f(x) = (x - \alpha)\, g(x)$, for some $g(x) \in \mathbb{C}[x]$.

If α is real, then $g(x) \in \mathbb{R}[x]$. Because we are assuming $f(x)$ irreducible, it follows that $g(x)$ is constant, and so $f(x)$ has degree 1.

If α is not real, then we have seen that $\overline{\alpha}$ is also a root of $f(x) = (x - \alpha)\, g(x)$, and being different from α it must be a root of $g(x)$, so $g(\overline{\alpha}) = 0$. Hence $x - \overline{\alpha}$ must divide $g(x)$, and so

$$f(x) = (x - \alpha)(x - \overline{\alpha}) \cdot h(x) = [x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}] \cdot h(x)$$

for some $h(x) \in \mathbb{C}[x]$. But we have seen earlier that the quadratic factor $x^2-(\alpha+\overline{\alpha})x+\alpha\overline{\alpha}$ has real coefficients, and so $h(x)$ has real coefficients as well. Because we are assuming $f(x)$ irreducible, it follows that $g(x)$ is constant, and so $f(x)$ has degree 2. Also, it has negative discriminant as claimed in the theorem because its roots α and $\overline{\alpha}$ are not real. □

Corollary 37. *Every polynomial of positive degree in $\mathbb{R}[x]$ is a product of polynomials of degree one and of quadratic polynomials with negative discriminant.*

Consequently, a polynomial of odd degree in $\mathbb{R}[x]$ has always at least one real root.

Example. Consider the polynomial $f(x) = 4x^4+20x^3+30x^2-40x+26$, and suppose that we have somehow found out that $-3+2i$ is a root, meaning that $f(-3+2i) = 0$. Our task is to find the remaining complex roots, and obtain a complete factorisation of $f(x)$ in $\mathbb{C}[x]$.

According to the Factor Theorem, $x+3-2i$ is a factor of $f(x)$. Hence we divide $f(x)$ by $x+3-2i$ using Ruffini's rule:

$$\begin{array}{r|cccc|c} & 4 & 20 & 30 & -40 & 26 \\ -3+2i & & -12+8i & -40-8i & 46+4i & -26 \\ \hline & 4 & 8+8i & -10-8i & 6+4i & 0 \end{array}$$

This confirms that $f(-3+2i)$ is actually a root of $f(x)$ as claimed. It also tells us that

$$f(x) = (x+3-2i) \cdot [4x^3 + (8+8i)x^2 + (-10-8i)x + (6+4i)].$$

Because $-3+2i$ is a root of $f(x)$ we know that its conjugate $-3-2i$ is a root as well, and because that is not a root of its factor $x+3-2i$ it must be a root of the other factor, a cubic polynomial. Dividing that by $x+3+2i$ using Ruffini's rule, we find

$$\begin{array}{r|ccc|c} & 4 & 8+8i & -10-8i & 6+4i \\ -3-2i & & -12-8i & 12+8i & -6-4i \\ \hline & 4 & -4 & 2 & 0 \end{array}$$

Hence $f(x) = (x+3-2i)(x+3+2i)(4x^2-4x+2)$. Finally, the roots of the quadratic factor can easily be found to be $(1 \pm i)/2$, by means of the usual formula, and so the complete factorisation of $f(x)$ in $\mathbb{C}[x]$ is

$$f(x) = (x+3-2i)(x+3+2i)(2x-1-i)(2x-1+i).$$

The complete factorisation of $f(x)$ in $\mathbb{R}[x]$ can be found by multiplying together the pairs of linear factors corresponding to conjugate roots: $f(x) = (x^2+6x+13)(4x^2-4x+2)$.

Lecture notes of Algebra. Week 6

23. Rational roots of a polynomial with integer coefficients

There is a test which allows, with a finite amount of calculations, to find *all* rational roots of a polynomials with rational coefficients, or to conclude that none exists if none is found. It is a rather specialised test, but it is sometimes useful, and its proof is a good illustration of the use of Arithmetical Lemma B on divisibility.

Note that Arithmetical Lemma B can be applied repeatedly to a product of more than two factors, and hence: if an integer divides a product of several integers, and is coprime (separately) with each of the factors except one, then it divides that factor. For example, if a divides a product bcd, and $(a,b)=1$ and $(a,c)=1$, then a must divide d. In fact, because a divides $b(cd)$, and $(a,b)=1$ we have that a divides cd, and then because $(a,c)=1$ we have that a divides d.

Before we apply the test we we may reduce to the case of a polynomial with integer coefficients by multiplying our polynomial by a suitable integer (say the least common multiple of all denominators occurring in the coefficients). Then we may divide by any common factor of the coefficients; strictly speaking, this is not required for the validity of the following test, but it may avoid us lots of superfluous calculations.

Theorem 38 (The Rational Root Test). *Consider a polynomial $f(x)=a_nx^n+\cdots+a_1x+a_0$ with integer coefficients and $a_na_0\neq 0$. If r/s is a rational root of $f(x)$, written as a fraction of integers in lowest terms (that is, with $(r,s)=1$), then r divides the constant term a_0, and s divides the leading coefficient a_n.*

Proof. After expanding $s^n\cdot f(r/s)=0$ we find

$$a_nr^n+a_{n-1}r^{n-1}s+a_{n-2}r^{n-2}s^2+\cdots+a_2r^2s^{n-2}+a_1rs^{n-1}+a_0s^n=0.$$

Because r divides all terms preceding the last one, it must divide the last term as well, that is, $r\mid a_0s^n$. But because $(r,s)=1$, Arithmetical Lemma B implies that r divides a_0.

In a similar way, because s divides all terms following the first one, it must divide the first term a_nr^n as well. Because $(r,s)=1$ it follows that s divides a_n. □

Example. Find all the rational roots of the polynomial $f(x)=2x^3+15x^2+27x+10$, and then factorise it over $\mathbb{Q}$. According to the test, if $r/s\in\mathbb{Q}$ is a root of $f(x)$, with $\gcd(r,s)=1$, then r divides 10 and s divides 2, hence $r\in\{\pm1,\pm2,\pm5,\pm10\}$ and $s\in\{\pm1,\pm2\}$. Consequently, the possibilities for r/s are

$$\pm1,\ \pm2,\ \pm5,\ \pm10,\ \pm\frac{1}{2},\ \pm\frac{5}{2}.$$

However, because the coefficients of the polynomial are all positive, no positive real number can be a root, and so we only have to test the negative ones. Going through the list from left to right, we find $f(-2) = f(-5) = f(-1/2) = 0$, at which point we can stop because $f(x)$ cannot have more than three roots, and we conclude that

$$f(x) = 2(x+2)(x+5)\big(x+\tfrac{1}{2}\big) = (x+2)(x+5)(2x+1),$$

which is the desired complete factorisation of $f(x)$ over $\mathbb{Q}$. The last expression is actually a complete factorisation over $\mathbb{Z}$. (It is a general fact, known as *Gauss' lemma,* that any factorisation over $\mathbb{Q}$ of a polynomial with integer coefficients leads to a corresponding factorisation over $\mathbb{Z}$ by suitably rearranging some scalar factors.) Alternatively, once we have found the first root -2 we may divide $f(x)$ by $x+2$, and then proceed to factorise the resulting quadratic polynomial.

We can use the Rational Root Test to prove that certain radicals represent irrational numbers, as follows.

Example. We prove that $\sqrt{3}$ is irrational. We start with noting that $\sqrt{3}$ is a root of the polynomial x^2-3. By the Rational Root Test, if r/s is a rational root of x^2-3, with $r, s \in \mathbb{Z}$ and $(r, s) = 1$, then $r \mid 3$ and $s \mid 1$, and so $r/s \in \{\pm 1, \pm 3\}$. None of those numbers is a root, hence x^2-3 has no rational root, and so $\sqrt{3}$ is irrational.

Example. We prove that $\sqrt[3]{25/3}$ is irrational. Here $\sqrt[3]{25/3}$ is a root of the polynomial $3x^3-25$. By the Rational Root Test, if r/s is a rational root of $3x^3-25$, with $r, s \in \mathbb{Z}$ and $(r, s) = 1$, then r divides 25 and s divides 3, and so $r/s \in \{\pm 1, \pm 5, \pm 25, \pm 1/3, \pm 5/3, \pm 25/3\}$.

To conclude that $\sqrt[3]{25/3}$ cannot be rational it is enough to check that none of those 12 rational numbers is a root of the polynomial (and so $\sqrt[3]{25/3}$ cannot be equal to any of those rational numbers). However, it may not be necessary to check them all. For example, no negative real number can possibly be a root of $3x^3-25$, and so we only need to check the positive ones. We can do even better if we are able to locate the real roots of the polynomial more precisely. For example, noting that $3 \cdot 2^3 - 25 = -1 < 0$, and $3 \cdot 3^3 - 25 = 56 > 0$, and that the function $x \mapsto x^3$ is increasing, any real root of $3x^3-25$ must be larger than 2 and less than 3. However, none of the candidates which we have found for rational roots is between 2 and 3, and so we conclude that $3x^2-25$ has no rational root, and hence that $\sqrt[3]{25/3}$ is irrational.

24. Some special polynomials: biquadratic polynomials

A *biquadratic polynomial* is a polynomial of degree four where the terms of odd degree are missing, and so has the form ax^4+bx^2+c, with $a \neq 0$. Because it can be viewed as $a(x^2)^2+bx^2+c$, the standard way of finding its roots is setting $y = x^2$, and then solving $ay^2+by+c=0$ (by completing the square, or by the explicit formula). If β_1, β_2 are the roots of this quadratic equation in y, then the roots of the biquadratic polynomial are the solutions of either $x^2 = \beta_1$ or $x^2 = \beta_2$, and so they are the square roots of β_1 and the square roots of β_2.

Example. To find the roots of the biquadratic polynomial $2x^4 + x^2 - 6$ we set $x^2 = y$ and then calculate the roots of the resulting quadratic polynomial $2y^2 + y - 6$, finding $y = 3/2$ or $y = -2$. In terms of x this means $x^2 = 3/2$ or $x^2 = -2$, which leads to $x = \pm\sqrt{3/2} = \pm\sqrt{6}/2$ or $x = \pm i\sqrt{2}$. Hence the full factorisation of the polynomial over $\mathbb{C}$ is

$$\begin{aligned} 2x^4 + x^2 - 6 &= 2(x - \sqrt{6}/2)(x + \sqrt{6}/2)(x - i\sqrt{2})(x + i\sqrt{2}) \\ &= (\sqrt{2}\,x - \sqrt{3})(\sqrt{2}\,x + \sqrt{3})(x - i\sqrt{2})(x + i\sqrt{2}), \end{aligned}$$

whichever form we prefer (as the latter has no denominators, but more radicals). Its complete factorisation over $\mathbb{R}$ is

$$2x^4 + x^2 - 6 = 2(x - \sqrt{6}/2)(x + \sqrt{6}/2)(x^2 + 2),$$

and $x^2 + 2$ is irreducible over $\mathbb{R}$ because it has degree two and has no real roots. Finally, its complete factorisation over $\mathbb{Q}$ is

$$2x^4 + x^2 - 6 = (2x^2 - 3)(x^2 + 2),$$

where again the two quadratic factors are irreducible over $\mathbb{Q}$ because they have no rational roots.

25. (Optional) Double radicals

A *double radical* is an expression of the form $\sqrt{a \pm \sqrt{b}}$. (This is a special case of a *nested radical*, see `https://en.wikipedia.org/wiki/Nested_radical`.) Such an expression may occur, for example, when solving biquadratic equations and quartic self-reciprocal equations, or already when solving quadratic equations if the coefficients involve radicals. A double radical can sometimes be expressed as a sum of difference of *simple radicals*, using the identity

$$\text{(Double Radical Identity)} \qquad \sqrt{a \pm \sqrt{b}} = \sqrt{\frac{a + \sqrt{a^2 - b}}{2}} \pm \sqrt{\frac{a - \sqrt{a^2 - b}}{2}}.$$

Of course the right-hand side is the sum or difference of two simple radicals only when $a^2 - b$ is an exact square, otherwise the identity expresses a double radical as a more complicated expression, a sum of two double radicals.

Example. We have

$$\sqrt{5 \pm 2\sqrt{6}} = \sqrt{5 \pm \sqrt{24}} = \sqrt{\frac{5 + \sqrt{5^2 - 24}}{2}} \pm \sqrt{\frac{5 - \sqrt{5^2 - 24}}{2}} = \sqrt{3} \pm \sqrt{2}.$$

In fact, squaring $\sqrt{3} \pm \sqrt{2}$ we get

$$(\sqrt{3} \pm \sqrt{2})^2 = \sqrt{3}^2 \pm 2\sqrt{3}\sqrt{2} + \sqrt{2}^2 = 3 \pm 2\sqrt{6} + 2 = 5 \pm 2\sqrt{6}.$$

Example. We have

$$\sqrt{6 \pm 2\sqrt{3}} = \sqrt{6 \pm \sqrt{12}} = \sqrt{\frac{6 + \sqrt{6^2 - 12}}{2}} \pm \sqrt{\frac{5 - \sqrt{6^2 - 12}}{2}} = \sqrt{3 + \sqrt{6}} \pm \sqrt{3 - \sqrt{6}}.$$

This is correct but not very useful.

For the Double Radical Identity to really make sense we should put proper limitations on the values allowed for a and b. Convenient limitations for the present exposition are that a, b are real numbers with $a \geq 0$ and $0 \leq b \leq a^2$. This ensures that all the square roots appearing in this formula (on either side, and both the inner ones and the outer ones) have a nonnegative real argument c. Recall here that $\sqrt{c}$ is assigned a unique meaning when c is a nonnegative real number: $\sqrt{c}$ is the unique *non-negative* real number whose square equals c (and so of the two solutions of $x^2 = c$ one *chooses* to denote by $\sqrt{c}$ the non-negative one, for convenience). Under these conditions on a and b one can simply verify the formula by noting (that all the involved square roots are defined in the real numbers, and) that the right-hand side is nonnegative, by squaring both sides and checking that they give the same result after simplification.

Proof of the Double Radical Identity. Writing $R = \sqrt{a^2 - b}$ for brevity, the square of the right-hand side equals

$$\begin{aligned} \left(\sqrt{\frac{a+R}{2}} \pm \sqrt{\frac{a-R}{2}}\right) &= \frac{a+R}{2} + \frac{a-R}{2} \pm 2\sqrt{\frac{a+R}{2}}\sqrt{\frac{a-R}{2}} \\ &= a \pm \sqrt{a^2 - R^2} = a \pm \sqrt{b}. \end{aligned}$$

Hence the square of the right-hand side of the Double Radical Identity equals the square of the left-hand side of the Double Radical Identity. To conclude a proof of the Double Radical Identity we need to make sure that both sides have the same signs (or are both zero). In fact, our assumptions $a \geq 0$ and $0 \leq b \leq a^2$ imply that all radicals involved are real and non-negative (by convention we choose the non-negative root of a non-negative real number). In particular, the left-hand side is non-negative. Also, of the two radicals at the right-hand side the first is not less than the second (because $a + R \geq a - R$), so their difference is non-negative (in case of the minus sign out of ±). □

This procedure gives a correct and perfectly rigorous *proof* of the identity, but has the drawback that it does not tell us where the identity comes from: we apparently have to recall the identity by heart before we can verify it.

We now show how the identity can be derived from scratch. Hence if we happen to forget it, here is how it can be recovered. As a motivation we may preliminarly note that if we square a sum $\sqrt{x} + \sqrt{y}$ we get $x + y + 2\sqrt{xy}$. Now this latter expression is easily recognisable as the square of $\sqrt{x}+\sqrt{y}$ as long as the terms x and y of the sum are written separately, but not anymore once they are added together. Taking them apart is exactly what our identity aims to achieve. So, given a double radical $\sqrt{a \pm \sqrt{b}}$ (with $0 \leq b \leq a^2$) it is natural to try and express it in the form $\sqrt{a \pm \sqrt{b}} = \sqrt{x} \pm \sqrt{y}$, for some x and y to be determined. Squaring both sides we find $a \pm \sqrt{b} = x + y \pm 2\sqrt{xy}$, at which point we are led to impose

$$\begin{cases} x + y = a \\ 4xy = b. \end{cases}$$

Hence the required x and y are the roots of the quadratic polynomial $z^2 - az + b/4$ in the indeterminate z, which of course are $\big(a \pm \sqrt{a^2 - b}\big)/2$, leading to the desired identity after choosing $y \leq x$.

Example. Let us experiment with the Double Radical Identity on a case where the conditions $a \geq 0$ and $a^2 \geq b$ are *not* satisfied. For example, the double radical $\sqrt{2 - \sqrt{5}}$ does not represent a real number, because $2 - \sqrt{5} < 0$. In fact, the condition $a^2 \geq b$ is not satisfied here. If we change sign to what is under the square root we get a real radical $\sqrt{-2 + \sqrt{5}}$, but the condition $a^2 \geq b$ is still not satisfied, and actually $a \geq 0$ is not satisfied either. The Double Radical Identity would give

$$\sqrt{-2 + \sqrt{5}} = \sqrt{\frac{-2 + \sqrt{(-2)^2 - 5}}{2}} + \sqrt{\frac{-2 - \sqrt{(-2)^2 - 5}}{2}} = \sqrt{\frac{-2 + i}{2}} + \sqrt{\frac{-2 - i}{2}},$$

which is not particularly useful as they the simple radicals involve complex numbers. Similarly for the real radical $\sqrt{2 + \sqrt{5}}$ the condition $a \geq 0$ is satisfied, but $a \geq 0$ is not, and the Double Radical Identity would give $\sqrt{+2 + \sqrt{5}} = \sqrt{1 + i/2} + \sqrt{1 - i/2}$. The Double Radical Identity may be less useful when complex numbers are involved, but is still correct if properly interpreted. We explore that in the next section.

26. (Optional) Double radicals in the complex case

Now we take a look at the Double Radical Identity more generally, without the above assumptions on a and b. The identity remains valid for a and b any complex numbers, provided we are careful with the meaning of the square roots. In fact, while $\sqrt{a}$ has, by convention, a unique meaning when a is a nonnegative real number, namely, the only positive root of $x^2 - a$, for arbitrary complex $a \neq 0$ the symbol $\sqrt{a}$ actually takes two opposite values, the roots of $x^2 - a$, as it is not possible to make a consistent choice of one root or the other based on algebraic means. In particular, we usually think of $\sqrt{-1}$ as the imaginary unit i, but we could detect no difference in any calculation if we had set

$\sqrt{-1}$ to be $-i$ (as long as we are consistent). [10] When correctly interpreted, the following identity remains valid for any complex numbers a and b:

$$\sqrt{a+\sqrt{b}} = \sqrt{\frac{a+\sqrt{a^2-b}}{2}} + \sqrt{\frac{a-\sqrt{a^2-b}}{2}}.$$

Here any of the six radicals appearing may take two values, which leads to many possible interpretations, only some of which are correct. However, the same radical $\sqrt{a^2-b}$ appears twice on the right-hand side, and whenever that occurs in a formula there is a convention to use the same value for that in both instances (which one being immaterial as it appears with a + sign in one case and a − sign in the other). [11] So in the end the right-hand side generally takes four values, depending on a choice between two values for each of the outside radicals. The left-hand side also generally takes four possible values, depending on a choice between two values for $\sqrt{b}$, and once that choice has been made, on a choice between two values of $\sqrt{a+\sqrt{b}}$. How to match each of the two opposite pairs of values for the right-hand side to each of the two opposite pairs of values for the left-hand side must be made by looking at the arguments of the complex numbers involved, as both possible matches are algebraically equivalent (that is, there is no algebraic way to choose the appropriate match). [12] With this interpretation, the above identity for complex a, b can be simply verified by squaring both sides as we did for the real case.

One application of the double radical identity for complex numbers is to computing square roots of complex numbers. Consider a complex number written in the usual form $a+ib$, where a, b are real numbers. Together with its complex conjugate it can be thought of as $a \pm ib = a + \sqrt{-b^2}$, and the double radical formula gives

$$\begin{aligned} \sqrt{a \pm \sqrt{-b^2}} &= \sqrt{\frac{a+\sqrt{a^2+b^2}}{2}} \pm \sqrt{\frac{a-\sqrt{a^2+b^2}}{2}} \\ &= \sqrt{\frac{\sqrt{a^2+b^2}+a}{2}} \pm i\sqrt{\frac{\sqrt{a^2+b^2}-a}{2}}. \end{aligned}$$

Note that both $\sqrt{a^2+b^2}+a$ and $\sqrt{a^2+b^2}-a$ are nonnegative real numbers, and so both outer radicals in the final expression are square roots of nonnegative real numbers, where we have a convention in place that they usually represent the nonnegative square root. However, because of how we have obtained this expression they should still be considered as complex radicals, each taking two opposite values. Hence the above formula should be

[10] This issue is a little delicate to be discussed further at this point, but it is related with the *conjugation* map $a+ib \mapsto a-ib$ being an *automorphism* of the complex field $\mathbb{C}$.

[11] A similar situation occurs, for example, in Cardano's formulas for the solutions of cubic equations.

[12] In the double radical identity seen at the beginning of the subsection all radicals took nonnegative real values, which amounts to their arguments being 0, rather than the other possible choice π.

interpreted as

$$\pm\sqrt{a \pm \sqrt{-b^2}} = \pm\sqrt{\frac{\sqrt{a^2+b^2}+a}{2}} \pm i\sqrt{\frac{\sqrt{a^2+b^2}-a}{2}},$$

where all the outer radicals are now read as single-valued real radicals. As discussed above, all four ± signs in the equation are independent (they need not match), and so each side takes four possible values, namely two pairs of opposite values, and the only way to match them correctly is to look at the arguments of the complex numbers involved.

27. Square roots of complex numbers

Solving a quadratic equation with complex coefficients may require taking square roots of complex numbers. Now we look at the problem of computing square roots of complex numbers again, independently of the previous (optional) section on complex double radicals. We will see how computing those square roots in terms of real radicals leads to a biquadratic equation.

Consider a complex number written in the standard form $a + ib$, where $a, b \in \mathbb{R}$, and assume $a + ib \neq 0$ as we may. Of course if the number is written in polar form $a + ib = \rho \cdot (\cos\theta + i\sin\theta)$, where $\rho = \sqrt{a^2+b^2} > 0$ is its modulus and $0 \leq \theta < 2\pi$ is its argument, then we have general formulas for all the nth roots of $a + ib$, namely,

$$\rho^{1/n} \cdot \left(\cos\left(\frac{\theta+2k\pi}{n}\right) + i\sin\left(\frac{\theta+2k\pi}{n}\right)\right), \qquad \text{for } k = 0, \ldots, n-1$$

(or $-n/2 < k \leq n/2$ if we prefer). However, we would like to find the square roots of $a + ib$ algebraically, without using trigonometric functions. More precisely, $a + ib \neq 0$ will have two distinct square roots in the complex numbers, say $x + iy$ and $-x - iy$, with $x, y \in \mathbb{R}$. We would like to compute the real numbers x and y from a and b in an algebraic way, using the four basic operations and possibly radicals.

We have $(x + iy)^2 = a + ib$, that is,

$$x^2 - y^2 + 2ixy = a + ib.$$

Because x, y, a, b are real, this equation is equivalent to the system of equations [13]

$$\begin{cases} x^2 - y^2 = a \\ 2xy = b \end{cases}$$

[13] Note that this is a system of degree $2 \cdot 2 = 4$, and in general we could expect it to have four solutions in the complex numbers. For example, in the special case where $b = 0$ one finds that either $x = 0$ and $y^2 = -a$, or $y = 0$ and $x^2 = a$. However, if $a \neq 0$ then depending on the sign of a exactly one of these two cases will lead to real solutions for both x and y, which are the only acceptable ones for our problem, and produce the two square roots of a in each case.

If we assume $b \neq 0$, which is reasonable because otherwise $a+ib$ is a real number and its square roots are easy to find, then neither x nor y can be zero, and so from the second equation we find $y = b/(2x)$. Substituting this into the first equation we get

$$x^2 - \left(\frac{b}{2x}\right)^2 = a,$$

that is,

$$4x^4 - 4ax^2 - b^2 = 0.$$

This is a biquadratic equation, and solving it (without even performing the usual substitution and back, now that we know how it works) we find

$$x^2 = \frac{2a \pm \sqrt{(2a)^2 + 4b^2}}{4} = \frac{a \pm \sqrt{a^2+b^2}}{2}.$$

Because $a^2 < a^2 + b^2$, the right-hand side will only be nonnegative (and actually positive) when we take the $+$ sign in front of the radical, and because our x needs to be real only that case leads to acceptable solutions. Hence $x^2 = \bigl(a + \sqrt{a^2+b^2}\bigr)/2$ and, consequently, $y^2 = x^2 - a = \bigl(-a + \sqrt{a^2+b^2}\bigr)/2$. In conclusion, the two square roots $\pm(x+iy)$ of $a+ib$ are obtained by taking

$$x = \pm\sqrt{\frac{\sqrt{a^2+b^2}+a}{2}}, \quad \text{and} \quad y = \pm\sqrt{\frac{\sqrt{a^2+b^2}-a}{2}},$$

with matching signs in front of the two main radicals if $b > 0$, and with opposite signs if $b < 0$, as one recognises from the second equation of the system. This can be summarized in the formula

$$\pm\sqrt{a \pm ib} = \pm\sqrt{\frac{\sqrt{a^2+b^2}+a}{2}} \pm i\sqrt{\frac{\sqrt{a^2+b^2}-a}{2}},$$

where the signs have to be appropriately matched as discussed.

Note that there is no need to memorise such complicated formulas: to compute the square roots of a given complex number, just apply the procedure described (see the example below). Nevertheless, let us play with those formulas a bit.

If we write the nonzero complex number $a+ib$ in polar form $a+ib = \rho\cdot(\cos\theta + i\sin\theta)$, where $\rho = \sqrt{a^2+b^2}$ is its modulus and $0 \leq \theta < 2\pi$ is its argument, we can read off the above formulas the halving formulas for sin and cos. In fact, as discussed more generally for nth roots at the beginning of this section, the square roots of $a+ib = \rho(\cos\theta + i\sin\theta)$ will be $\pm\sqrt{\rho}\cdot\bigl(\cos(\theta/2) + i\sin(\theta/2)\bigr)$, and so our formulas for the square roots of $a+ib$ imply

$$\cos\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{1+\cos\theta}{2}}, \quad \text{and} \quad \sin\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{1-\cos\theta}{2}},$$

where the signs have to be taken appropriately. Of course these angle halving formulas can also be obtained more directly by inverting the angle duplication formulas

$$\cos\theta = 2\cos^2\Big(\frac{\theta}{2}\Big) - 1 = 1 - 2\sin^2\Big(\frac{\theta}{2}\Big).$$

Example. Compute the square roots of the complex number $-3 + 4i$, expressing them using only square roots of real numbers (that is, expressing their real and complex part using only algebraic operations on real numbers).

We look for real numbers x and y such that $(x + iy)^2 = -3 + 4i$, that is,

$$x^2 - y^2 + 2ixy = -3 + 4i.$$

Because x, y are real, this equation is equivalent to the system of equations

$$\begin{cases} x^2 - y^2 = -3 \\ 2xy = 4 \end{cases}$$

From the second equation we find $y = 2/x$. Substituting this into the first equation we get $x^2 - (2/x)^2 = -3$, that is, $x^4 + 3x^2 - 4 = 0$, or $(x^2 - 1)(x^2 + 4) = 0$. This has roots ± 1 and $\pm 2i$, but because x must be real for our problem we may only take $x = \pm 1$, and correspondingly $y = \pm 2$. In conclusion, the square roots of $-3+4i$ are $1+2i$ and $-1-2i$.

If we had forgotten that x and y are real, and so we accepted $x = \pm 2i$, and correspondingly $y = \mp i$, the argument would not be quite correct but we would still get the correct square roots, as $2i + i(-i) = 1 + 2i$ and its opposite.

Remark 39. (Optional) It is also possible to compute the square roots of $-3+4i$ by writing it as $-3 + \sqrt{-16}$ and applying the Double Radical Identity:

$$\sqrt{-3+4i} = \sqrt{-3+\sqrt{-16}} = \sqrt{\frac{3+\sqrt{3^2+16}}{2}} + \sqrt{\frac{3-\sqrt{3^2+16}}{2}} = \sqrt{1} + \sqrt{-4}.$$

However, as explained in the previous section using the Double Radical Identity leaves sign ambiguities when working with complex numbers, namely, $\sqrt{1}$ can mean 1 or -1, and $\sqrt{-4}$ can mean $2i$ or -21: we now must decide how to correctly match the signs by looking at arguments of the complex numbers involved (or guessing one possible match and square the result to check if it was the correct choice). In essence, the ambiguity which we have is between the square roots of $-3 + 4i$ and those of $-3 - 4i$, which are their conjugates.

28. Some special polynomials: self-reciprocal polynomials

The reason why biquadratic polynomials are much easier to factorise than arbitrary quartic polynomials is that they satisfy a special symmetry: they satisfy $f(-x) = f(x)$ (this condition is what, more generally, characterises *even* functions). A similar condition, but involving reciprocals instead of opposites, defines *self-reciprocal* polynomials.

A polynomial $f(x)$ of positive degree n is called *self-reciprocal* if $x^n \cdot f(1/x) = f(x)$. If $f(x) = a_n x^n + \cdots + a_1 x + a_0$ then *its reciprocal polynomial* is

$$x^n \cdot f\left(\frac{1}{x}\right) = a_n + a_{n-1}x + a_{n-2}x^2 + \cdots + a_1 x^{n-1} + a_0 x^n,$$

and so $x^n \cdot f(1/x)$ is a polynomial of degree n, whose coefficients are the coefficients of $f(x)$ read in the opposite order. Hence a polynomial $f(x)$ is self-reciprocal if it equals its reciprocal polynomial, and that means that its sequence of coefficients reads the same backwards as forwards: $a_n = a_0$, $a_{n-1} = a_1$, etc.

The definition of self-reciprocal shows that all roots α are nonzero (because $a_0 = a_n \neq 0$), and that whenever α is a root, its *reciprocal* $1/\alpha$ is a root as well. (This is the reason of the name *self-reciprocal.*) If $f(x)$ is self-reciprocal of odd degree, then one sees at once that -1 is a root, hence $f(x)$ is divisible by $x + 1$, and one can show that the quotient is also self-reciprocal, but of course of even degree. Hence it is enough to see how to deal with a self-reciprocal polynomial of even degree n. We could start with the case of quadratic polynomials, but because we already know how to find their roots in general we pass directly to the case of degree four.

A quartic self-reciprocal polynomial will have the form $ax^4+bx^3+cx^2+bx+a$. The idea for finding its roots, that is, for solving the corresponding equations, is to suitably match x and $1/x$, by taking their sum $x+1/x$. We may start with dividing by x^2 the corresponding equation (as we know that 0 cannot be a root), obtaining $ax^2 + bx + c + b/x + a/x^2 = 0$, that is,

$$a\left(x^2 + \frac{1}{x^2}\right) + b\left(x + \frac{1}{x}\right) + c = 0.$$

Now note that we can express $x^2 + 1/x^2$ in terms of $x + 1/x$,

$$x^2 + \frac{1}{x^2} = \left(x + \frac{1}{x}\right)^2 - 2,$$

and so we can write our equation in the equivalent form

$$a\left(x + \frac{1}{x}\right)^2 + b\left(x + \frac{1}{x}\right) + c - 2a = 0.$$

Now we may set $y = x + 1/x$, and solve the quadratic equation $ay^2 + by + c - 2a = 0$. It may not have any solutions in F, and in that case our quartic self-reciprocal equations cannot have any solutions in F either (because if $\alpha \in F$ were a solution, then $\beta = \alpha + 1/\alpha$ would be a solution of the quadratic equation). If it does have solutions, say β_1 and β_2 (which may be equal), then we may try and solve $x + 1/x = \beta_1$ and $x + 1/x = \beta_2$. Each of these may or may not have solutions in F, giving at most four solutions of our quartic equation. [14]

[14] A self-reciprocal equation of degree six may be dealt with in a similar way, and solving it reduces to solving a cubic equation. In this case we would also have to deal with an expression $x^3 + 1/x^3$, which

Example. We use the above method to find all complex roots of the self-reciprocal polynomial $6x^4 + 5x^3 - 38x^2 + 5x + 6$. After equating the polynomial to zero we divide by x^2, rearrange the terms and get

$$6\left(x^2 + \frac{1}{x^2}\right) + 5\left(x + \frac{1}{x}\right) - 38 = 0.$$

Using $(x + 1/x)^2 = x^2 + 2 + 1/x^2$ the equation becomes

$$6\left(x + \frac{1}{x}\right)^2 + 5\left(x + \frac{1}{x}\right) - 50 = 0,$$

which reads $6y^2 + 5y - 50 = 0$ after setting $x + 1/x = y$. This quadratic equation has roots $5/2$ and $-10/3$, and by substituting $y = x + 1/x$ again we find the two equations

$$x + \frac{1}{x} = \frac{5}{2}, \quad \text{and} \quad x + \frac{1}{x} = -\frac{10}{3}.$$

After multiplying by $2x$ or $3x$ they become

$$2x^2 - 5x + 2 = 0, \quad \text{and} \quad 3x^2 + 10x + 3 = 0,$$

whose solutions are 2, $1/2$, and -3, $-1/3$, respectively. Hence these four numbers are the roots of the original polynomial. Because they are all four real, the polynomial factorises into a product of four linear factors already over $\mathbb{R}$, namely,

$$\begin{aligned} 6x^4 + 5x^3 - 38x^2 + 5x + 6 &= 6(x-2)(x-1/2)(x+3)(x+1/3) \\ &= (x-2)(2x-1)(x+3)(3x+1). \end{aligned}$$

Because the roots have turned out to be all rational in this example, we could of course also have found them by the Rational Root Test. But we have used a general method for quartic self-reciprocal polynomials, which would work even if there were no rational roots.

29. (Optional) An application: exact trig values of the angle $2\pi/5$

We will compute the exact values of the trigonometric functions (sin and cos, from which the others follows easily) of multiples of the angle $\pi/5$, that is, 36°. These can be found through geometric arguments. As a general rule, formulas for trigonometric functions are best dealt with by working with the exponential form of complex numbers, and so we set

$$\omega := \exp(2\pi\, i/5) = \cos(2\pi/5) + i\, \sin(2\pi/5).$$

can be done by noting that

$$\left(x + \frac{1}{x}\right)^3 = x^3 + 3x + \frac{3}{x} + \frac{1}{x^3} = \left(x^3 + \frac{1}{x^3}\right) + 3\left(x + \frac{1}{x}\right),$$

whence $x^3 + 1/x^3 = (x+1/x)^3 - 3(x+1/x) = y^3 - 3y$, etc. There is a formula for solving cubic equations, but we will not see that, and so we stop here with this observation.

Now $\omega^5 = \exp(2\pi i/5)^5 = \exp(2\pi i) = 1$, and so ω is a fifth root of unity, and hence a root of the polynomial $x^5 - 1$. Because $x^5 - 1 = (x-1)(x^4 + x^3 + x^2 + x + 1)$ and $\omega \neq 1$, we deduce that ω is a root of $x^4 + x^3 + x^2 + x + 1$. Each of ω^2, $\omega^3 = \omega^{-2}$, and $\omega^4 = \omega^{-1}$ is also a fifth root of unity different from 1, hence these numbers are also roots of $x^4 + x^3 + x^2 + x + 1$. Being all distinct, they must be all complex roots of this polynomial, and so

$$x^4 + x^3 + x^2 + x + 1 = (x - \omega)(x - \omega^2)(x - \omega^{-2})(x - \omega^{-1}).$$

Because this polynomial is self-reciprocal polynomial, we have learnt how to compute its roots. After equating the polynomial to zero we divide by x^2, rearrange the terms and get

$$x^2 + \frac{1}{x^2} + x + \frac{1}{x} + 1 = 0.$$

Now because $(x + 1/x)^2 = x^2 + 2 + 1/x^2$ we can transform the equation into the equivalent equation

$$\left(x + \frac{1}{x}\right)^2 + \left(x + \frac{1}{x}\right) - 1 = 0.$$

After setting $x + 1/x = y$ we solve the resulting equation $y^2 + y - 1 = 0$, and find

$$y = \frac{-1 \pm \sqrt{5}}{2}.$$

Now we substitute $y = x + 1/x$ and solve the two equations

$$x + \frac{1}{x} = \frac{-1 \pm \sqrt{5}}{2},$$

which after multiplication by $2x$ become

$$2x^2 - (-1 \pm \sqrt{5})x + 2 = 0.$$

The equation $2x^2 - (-1 + \sqrt{5})x + 2 = 0$ has solutions

$$\frac{\sqrt{5} - 1 \pm \sqrt{(\sqrt{5} - 1)^2 - 16}}{4} = \frac{\sqrt{5} - 1 \pm \sqrt{-10 - 2\sqrt{5}}}{4} = \frac{\sqrt{5} - 1}{4} \pm i \frac{\sqrt{10 + 2\sqrt{5}}}{4},$$

and so we conclude that

$$\cos(2\pi/5) = \frac{\sqrt{5} - 1}{4}, \quad \text{and} \quad \sin(2\pi/5) = \frac{\sqrt{10 + 2\sqrt{5}}}{4}.$$

Of course $\sin(2\pi/5)$ could also be computed from $\cos(2\pi/5)$ using the relation $\cos^2 \alpha + \sin^2 \alpha = 1$. The double radical $\sqrt{10 + 2\sqrt{5}}$ here cannot be simplified, as $10^2 - (2\sqrt{5})^2 = 80$ is not a perfect square.

Similarly, by solving $2x^2 - (-1 + \sqrt{5})x + 2 = 0$ we find that

$$\cos(4\pi/5) = \frac{-1 - \sqrt{5}}{4}, \quad \text{and} \quad \sin(4\pi/5) = \frac{\sqrt{10 - 2\sqrt{5}}}{4}.$$

From this we obtain

$$\cos(\pi/5) = \frac{1+\sqrt{5}}{4}, \quad \text{and} \quad \sin(\pi/5) = \frac{\sqrt{10-2\sqrt{5}}}{4}.$$

Hence each side of a regular pentagon of radius 1 has length $2\sin(\pi/5) = \bigl(\sqrt{10-2\sqrt{5}}\bigr)/2$.

30. Symmetric functions of the roots of quadratic polynomials

In this and the next section we study an important relation between the coefficients of a polynomial and its roots. We start with discussing the case of quadratic polynomials. If the quadratic polynomial ax^2+bx+c (hence with $a \neq 0$) has at least one root in the field F, then we have seen earlier on that it factorises as the product of two polynomials of degree one. By collecting suitable scalar factors those two factors can be taken to have the form $x-\alpha$ and $x-\beta$, and so we have

$$ax^2+bx+c = a(x-\alpha)(x-\beta) = a\bigl(x^2-(\alpha+\beta)x+\alpha\beta\bigr).$$

Hence $-b/a = \alpha+\beta$ (the sum of the roots), and $c/a = \alpha\beta$ (the product of the roots). This can be used to guess the roots of a quadratic polynomial in simple cases, but also, more usefully, to use a quadratic equation for solving systems of two equations as in the following example.

Example. Solving the system

$$\begin{cases} x+y=s \\ xy=p \end{cases}$$

in the unknowns x and y, means finding all pairs of numbers x, y whose sum equals s and whose product equals p. One could express y in terms of x using the first equation, hence $y = s-x$, substitute for y in the second equation, solve the corresponding quadratic equation in x obtained, etc., but it is more efficient to exploit the symmetry of the system and to proceed as follows.

The desired x and y will be the roots of the polynomial $(z-x)(z-y)$ in the indeterminate z, which can be written as $z^2-(x+y)z+xy$, and hence equals z^2-sz+p. Therefore, its complex roots are given by the formula $\bigl(s \pm \sqrt{s^2-4p}\bigr)/2$. One of the roots will be x, and the other will be y. Of course if the roots are distinct then there are two ways to match x and y to the two roots, and this gives us two solutions (x, y) for our symmetric system. The roots will be equal exactly when $s^2 = 4p$, and in that case the system has a 'double' solution $(x, y) = (s/2, s/2)$.

The polynomials $x+y$ and xy are examples of *symmetric polynomials* in the indeterminates x and y. More generally, a polynomial $f(x, y)$ in x and y is a *symmetric polynomial* if it is unchanged by interchanging the indeterminates x and y, which means

$f(y,x) = f(x,y)$. For example, $x^3 + 2x^2y + 2xy^2 + y^3 + 5xy - 4x - 4y + 7$ is a symmetric polynomial. The special polynomials $x + y$ and xy are called the *elementary symmetric polynomials.*

THEOREM 40. *Every symmetric polynomial $f(x,y)$ (with coefficients in any field F) can be expressed as a polynomial (also with coefficients in F) in the elementary symmetric polynomials $x + y$ and xy.*

This means that if $f(x,y)$ is a symmetric polynomial in x and y (hence with the condition $f(y,x) = f(x,y)$), then $f(x,y) = g(x + y, xy)$, for some polynomial $g(s,p)$ in two indeterminates s and p. More is true: if $f(x,y)$ has integer coefficients, then $g(s,p)$ has integer coefficients as well. We will not prove these statements in general, but just illustrate them with a couple of special cases which can be guessed at once:

$$x^2 + y^2 = (x + y)^2 - 2xy, \qquad x^3 + y^3 = (x + y)^3 - 3xy(x + y).$$

The next case takes a bit more work,

$$\begin{aligned} x^4 + y^4 &= (x + y)^4 - xy(4x^2 + 6xy + 4y^2) \\ &= (x + y)^4 - xy\bigl(4(x + y)^2 - 2xy\bigr) \\ &= (x + y)^4 - 4xy(x + y)^2 + 2(xy)^2. \end{aligned}$$

Hence, in the notation of Theorem 40, the symmetric polynomial $f(x,y) = x^4 + y^4$ can be written as $f(x,y) = g(x + y, xy)$, where $g(s,p) = s^4 - 4s^2p + 2p^2$.

More generally, according to Theorem 40 sums $x^n + y^n$ of higher powers can also be expressed as polynomials in $x+y$ and xy (with integer coefficients): one need to do some work as above and use the analogous expressions for smaller values of n. A different and more efficient way of achieving this same goal is described in a later section.

EXAMPLE. The symmetric polynomial $f(x,y) = x^3+2x^2y+2xy^2+y^3+5xy-4x-4y+7$ can be written as

$$f(x,y) = (x + y)^3 - xy(x + y) + 5xy - 4(x + y) + 7 = g(x + y, xy),$$

where $g(s,p) = s^3 - sp - 4s + 5p + 7$.

31. (Optional) The symmetries involved in biquadratic and self-reciprocal polynomials

The methods which we used to solve biquadratic and quartic self-reciprocal equations can be justified in terms of symmetric polynomials.

Indeed, the fact that a biquadratic polynomial $f(x)$ is unchanged when replacing x with $-x$ (that is, $f(-x) = f(x)$, which more generally characterises the *even* polynomials, as opposed to the *odd* polynomials, satisfying $f(-x) = -f(x)$), suggests expressing it in terms of the ‘elementary symmetric polynomials’ in x and $-x$, which are $x+(-x)=0$ and $x\cdot(-x)=-x^2$. (We have slightly abused language here, as x and $-x$ are not independent indeterminates.) This is, in fact, what we did, thinking of a biquadratic polynomial as a polynomial in x^2 (which makes no practical difference from using $-x^2$ instead).

Similarly, a self-reciprocal polynomial $f(x)$ of degree n satisfies $x^n\cdot f(1/x)=f(x)$, hence it is not quite left unchanged by replacing x with $1/x$, but almost. In fact, our first step in finding the roots of $f(x)$ (and then factorising it), for even n, was dividing $f(x)$ by $x^{n/2}$. Now the rational expression (that is, quotient of two polynomials) $g(x)=f(x)/x^{n/2}$ satisfies $g(1/x)=g(x)$, because

$$g(1/x)=\frac{f(1/x)}{(1/x)^{n/2}}=x^{n/2}\cdot f(1/x)=f(x)/x^{n/2}=g(x).$$

Hence $g(x)$ is left unchanged by replacing x with $1/x$, and as a consequence of Theorem 40 (which extends to quotients of polynomials) it can be expressed in terms of the ‘elementary symmetric polynomials’ in x and $1/x$, which are $x+1/x$ and $x\cdot 1/x=1$. This is precisely what we did when finding the roots of self-reciprocal polynomials: we expressed x^2+1/x^2 as a polynomial in $x+1/x$ (and we would do the same with each x^k+1/x^k if we wished to deal with self-reciprocal polynomials of even degree higher than 4).

Lecture notes of Algebra. Week 8

32. Examples of symmetric systems

A system of equations in x and y is *symmetric* if each of the equations is a symmetric polynomial (or *rational expression,* meaning a quotient of polynomials) in x and y. In that case one expresses both equations in terms of $s = x + y$ and $p = xy$, after which one can try to solve the corresponding system in s and p, and finally recover x and y. Note that if we interchange the values of x and y in any solution (x, y) we get another solution, because a symmetric system does not change if we interchange x and y.

Example. To solve the system

$$\begin{cases} x^4 + y^4 = 17 \\ x + y = 3 \end{cases}$$

in the unknowns x and y, we express the left-hand side of the first equation in terms of $x+y$ and xy as we have just learnt, and then substitute the value for $x+y$ obtained from the second equation, obtaining

$$\begin{cases} 3^4 - 4 \cdot 3^2 \cdot xy + 2(xy)^2 = 17 \\ x + y = 3 \end{cases}$$

which becomes

$$\begin{cases} (xy)^2 - 18(xy) + 32 = 0 \\ x + y = 3 \end{cases}$$

Solving the first equation for xy we see that the system is equivalent to

$$\begin{cases} xy = 2 \\ x + y = 3 \end{cases} \quad \text{or} \quad \begin{cases} xy = 16 \\ x + y = 3 \end{cases}$$

Solving these two systems as we learned earlier we find altogether four different solutions in the complex numbers, namely,

$$(x, y) = (2, 1),\ (1, 2),\ \left(\frac{3}{2} + i\frac{\sqrt{55}}{2}, \frac{3}{2} - i\frac{\sqrt{55}}{2}\right),\ \left(\frac{3}{2} - i\frac{\sqrt{55}}{2}, \frac{3}{2} + i\frac{\sqrt{55}}{2}\right).$$

It is also true that sums $x^n + y^n$ with *negative* exponent n can be expressed in terms of $x + y$ and xy, however not as a polynomial in $x + y$ and xy but a quotient of two polynomials (which is called a *rational expression*). For example,

$$x^{-1} + y^{-1} = \frac{1}{x} + \frac{1}{y} = \frac{x + y}{xy}.$$

More generally,

$$x^{-n} + y^{-n} = \frac{1}{x^n} + \frac{1}{y^n} = \frac{x^n + y^n}{(xy)^n},$$

hence if we know how to express $x^n + y^n$, for a positive n, in terms of $x + y$ and xy, then we also know how to express $x^{-n} + y^{-n}$.

Example. Find all the complex solutions of the following (symmetric) system:

$$\begin{cases} \dfrac{1}{x} + \dfrac{1}{y} = 1 \\ x + y = 2 \end{cases}$$

After rewriting $1/x + 1/y$ as $(x + y)/(xy)$ in the first equation, multiplying both sides by xy, and substituting into it the value of $x + y$ given by the second equation, we find

$$\begin{cases} xy = 2 \\ x + y = 2 \end{cases}$$

Because according to the first equation xy is nonzero, we did not introduce any more solutions when we multiplied by xy, and so this system is equivalent to the original system. Solving it as usual we find the solutions

$$(x, y) = (1 + i, 1 - i),\ (1 - i, 1 + i).$$

Example. Find all the complex solutions of the following symmetric system:

$$\begin{cases} x + y + \dfrac{1}{x} + \dfrac{1}{y} = \dfrac{1}{2} \\ x^2 + y^2 + xy = 3 \end{cases}$$

After multiplying the first equation by xy and expressing everything in terms of the elementary symmetric polynomials $x+y$ and xy, we can write the system in the equivalent form:

$$\begin{cases} 2(x + y)xy + 2(x + y) = xy \\ (x + y)^2 - xy = 3 \end{cases}$$

Note that this is a system of degree $3 \cdot 2 = 6$, because the first equation has degree 3 (the 'combined' degree in x and y, since there is a term x^2y, for example). Obtaining xy from the second equation and then substituting into the first equation we get

$$\begin{cases} 2(x + y)^3 - 4(x + y) = (x + y)^2 - 3 \\ xy = (x + y)^2 - 3 \end{cases}$$

Now the first equation contains only $x + y$, so after setting $z = x + y$ the first equation becomes $2z^3 - z^2 - 4z + 3 = 0$. Using the Rational Root Test it is not hard to find that this polynomial has roots 1, then 1 again, and then $-3/2$ (so the left-hand side of the equation factorises as $2z^3 - z^2 - 4z + 3 = (z - 1)^2(2z + 3)$, with 1 being a double root).

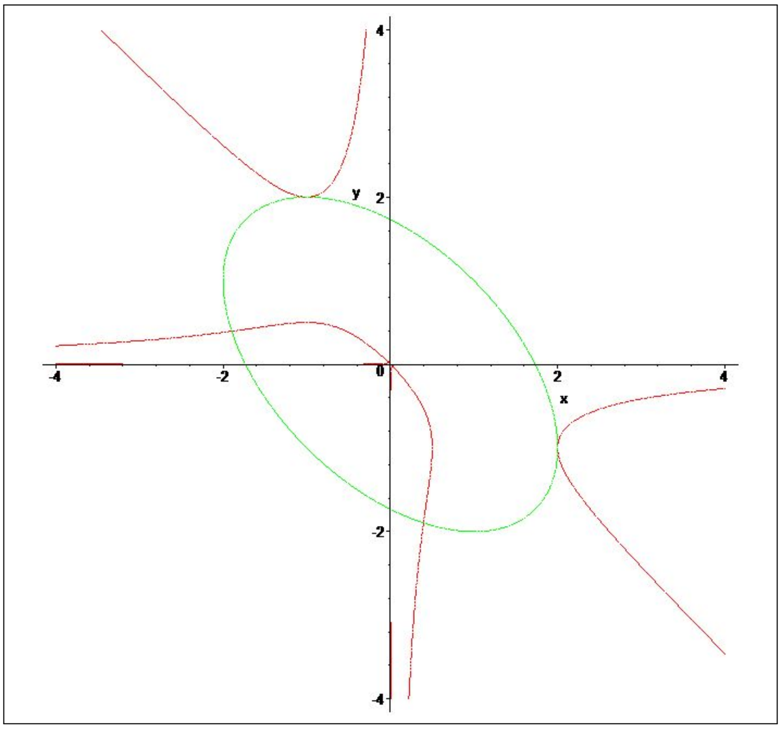
For each of those values of $z = x + y$ one can use the other equation of the system to compute $xy = z^2 - 3$, and so the system is equivalent to

$$\begin{cases} x + y = 1 \\ \quad xy = -2 \end{cases} \quad \text{or} \quad \begin{cases} x + y = -3/2 \\ \quad xy = -3/4 \end{cases}$$

(where each root of the first system should really be counted twice). Solving these two systems as we learned earlier we find altogether four different solutions of the system in the complex numbers, and all four are actually real:

$$(x, y) = (2, -1),\ (-1, 2),\ \Big(\frac{-3+\sqrt{21}}{2}, \frac{-3-\sqrt{21}}{2}\Big),\ \Big(\frac{-3+\sqrt{21}}{2}, \frac{-3+\sqrt{21}}{2}\Big).$$

The first two should actually be counted as *double solutions,* as they came from a double root of the cubic equation, and counting that way we have actually found six solutions of our system of degree six. In the following graph the two equations of our system are represented by the red curve and the green curve, respectively, and we see that the double roots manifest themselves as points where the two curve intersect with the same tangent.



From now on we make a small notational change, using α and β in place of x and y, so we can use x again (rather than z as above) as the indeterminate of our polynomial

$$x^2 - sx + p = (x - \alpha)(x - \beta).$$

Another application of the elementary symmetric polynomials is that they allow us to compute symmetric expressions of the roots of a polynomial (quadratic for now) in terms of the coefficients of the polynomial, without actually computing the roots.

Example. Compute $\alpha^2+\beta^2$, $\alpha^3+\beta^3$, and $\alpha^{-2}+\beta^{-2}$, where α and β are the roots of the polynomial x^2-5x+3. A direct approach based on finding expressions for α and β first, which are $(5\pm\sqrt{13})/2$, would involve complicated calculations with radicals, which would eventually simplify and give rational numbers as final answers (actually, integers in this case). It is much better to use the fact that $\alpha+\beta=5$ and $\alpha\beta=3$, whence

$$\alpha^2+\beta^2=(\alpha+\beta)^2-2\alpha\beta=25-2\cdot 3=19,$$

$$\alpha^3+\beta^3=(\alpha+\beta)^3-3\alpha\beta(\alpha+\beta)=125-3\cdot 3\cdot 5=80,$$

$$\alpha^{-2}+\beta^{-2}=\frac{\alpha^2+\beta^2}{(\alpha\beta)^2}=\frac{(\alpha+\beta)^2-2\alpha\beta}{(\alpha\beta)^2}=\frac{5^2-2\cdot 3}{3^2}=\frac{19}{9}.$$

This last one could also be found by looking at the reciprocal polynomial $3x^2-5x+1$, whose roots are $1/\alpha$, $1/\beta$, $1/\gamma$:

$$\frac{1}{\alpha^2}+\frac{1}{\beta^2}=\Big(\frac{1}{\alpha}+\frac{1}{\beta}\Big)^2-2\frac{1}{\alpha}\frac{1}{\beta}=\Big(\frac{5}{3}\Big)^2-2\cdot\frac{1}{3}=\frac{19}{9}.$$

33. (Optional) Expressing a sum of two equal powers in terms of symmetric polynomials

A systematic and instructive way of expressing $\alpha^n+\beta^n$ in terms of $\alpha+\beta=s$ and $\alpha\beta=p$ is as follows. We start with the fundamental equations

$$\alpha^2-s\alpha+p=0 \quad\text{and}\quad \beta^2-s\beta+p=0.$$

which we can rewrite in the equivalent form

$$\alpha^2=s\alpha-p \quad\text{and}\quad \beta^2=s\beta-p.$$

Adding them together we obtain

$$\alpha^2+\beta^2=s(\alpha+\beta)-2p=s^2-2p.$$

If instead of adding the two equations together we first multiply them by α and β, respectively,

$$\alpha^3=s\alpha^2-p\alpha \quad\text{and}\quad \beta^3=s\beta^2-p\beta,$$

and then add them together, we find

$$\alpha^3+\beta^3=s(\alpha^2+\beta^2)-p(\alpha+\beta)=s(s^2-2p)-ps=s^3-3sp.$$

If, instead, we multiply the resulting two equations again by α and β, respectively, and then add them together, we find

$$\alpha^4+\beta^4=s(\alpha^3+\beta^3)-p(\alpha^2+\beta^2)=s(s^3-3sp)-p(s^2-2p)=s^3-4s^2p+2p^2,$$

and so on.

We can also state what we have found as follows. If we set $c_n = \alpha^n + \beta^n$, the sequence of numbers c_n satisfies the (quadratic) *linear recurrence relation* [15]

$$c_n = s \cdot c_{n-1} - p \cdot c_{n-2},$$

which may also write as $c_n - sc_{n-1} + pc_{n-2} = 0$ if we prefer. (Note that the coefficients are the same as those of our polynomial $x^2 - sx + 1$.) The whole sequence is then completely (and explicitly) determined by this linear recurrence relation together with the *initial conditions*

$$c_1 = \alpha + \beta = s, \quad \text{and} \quad c_2 = \alpha^2 + \beta^2 = s^2 - 2p,$$

or, even better,

$$c_0 = \alpha^0 + \beta^0 = 1 + 1 = 2, \quad \text{and} \quad c_1 = \alpha + \beta = s.$$

The relation can also be used backwards, to compute sums $\alpha^n + \beta^n$ for negative n (as long as $\alpha\beta \neq 0$).

Example. Let α and β be the complex roots of $x^2 + x + 1$. Hence $\alpha + \beta = -1$ and $\alpha\beta = 1$. Proceeding as we have seen above we have

$$\alpha^2 = -\alpha - 1 \quad \text{and} \quad \beta^2 = -\beta - 1.$$

Adding them together we obtain

$$\alpha^2 + \beta^2 = -\alpha - \beta - 2 = -(-1) - 2 = -1.$$

If instead of adding the two equations together we first multiply them by α and β, respectively,

$$\alpha^3 = -\alpha^2 - \alpha \quad \text{and} \quad \beta^3 = -\beta^2 - \beta,$$

and then add them together, we find

$$\alpha^3 + \beta^3 = -(\alpha^2 + \beta^2) - (\alpha + \beta) = -(-1) - (-1) = 2.$$

If, instead, we multiply the two equations again by α and β, respectively, and then add them together, we find

$$\alpha^4 + \beta^4 = -(\alpha^3 + \beta^3) - (\alpha^2 + \beta^2) = -2 - (-1) = -1,$$

and so on. Continuing this way we will find

$$\alpha^5 + \beta^5 = -(\alpha^4 + \beta^4) - (\alpha^3 + \beta^3) = -(-1) - 2 = -1,$$

and

$$\alpha^6 + \beta^6 = -(\alpha^5 + \beta^5) - (\alpha^4 + \beta^4) = -(-1) - (-1) = 2,$$

[15] This is a sort of discrete version of a linear differential equation of the second order, such as $y'' + ay' + b = 0$, and a similar theory can be developed.

Hence $\alpha^k + \beta^k$, for $k = 1, \ldots, 6$, takes the values -1, -1, 2, -1, -1, 2. We may suspect that this would continue periodically, and this is in fact correct.

One way to see that if we set $c_n = \alpha^n + \beta^n$ as done earlier, the sequence of numbers c_n satisfies the (quadratic) linear recurrence relation

$$c_n = -c_{n-1} - c_{n-2},$$

with the initial values

$$c_1 = \alpha + \beta = -1, \quad \text{and} \quad c_2 = \alpha^2 + \beta^2 = -1.$$

We may use the recurrence to compute c_3, c_4, and c_5. At this point we see that $c_4 = c_1 = -1$ and $c_5 = c_2 = -1$, and because each term of the sequence only depends on the previous two we may conclude that the sequence repeats periodically, every three steps.

Another explanation is that our polynomial $x^2 + x + 1$ is a very special polynomial: because $x^2 + x + 1 = (x^3 - 1)/(x - 1)$, its roots $\alpha = (-1 + i\sqrt{3})/2$ and $\beta = (-1 - i\sqrt{3})/2$ satisfy $\alpha^3 = 1$ and $\beta^3 = 1$. (They are the *primitive cubic roots of* 1, and $x^2 + x + 1$ is a *cyclotomic polynomial.*) Also, $\alpha^2 = -\alpha - 1 = \beta$. This explains why the value of $\alpha^k + \beta^k$ repeats periodically every three steps: it does so because the (complex) value of each term, α^k and β^k, repeats periodically every three steps.

34. Symmetric functions of the roots of a cubic polynomial

What we have seen about symmetric functions of the roots of quadratic polynomials generalizes nicely to polynomials of higher degree. We start with the case of a cubic polynomial. After dividing by the leading coefficients we may always assume the polynomial to be monic, and we conveniently write it in the form $x^3 - sx^2 + rx - p$, with alternating signs. If the polynomial factorises as the product of three linear factors over some field we will have

$$\begin{aligned} x^3 - sx^2 + rx - p &= (x - \alpha)(x - \beta)(x - \gamma) \\ &= x^3 - (\alpha + \beta + \gamma)x^2 + (\alpha\beta + \alpha\gamma + \beta\gamma)x - \alpha\beta\gamma, \end{aligned}$$

and so

$$\begin{aligned} \alpha + \beta + \gamma &= s, \\ \alpha\beta + \alpha\gamma + \beta\gamma &= r, \\ \alpha\beta\gamma &= p. \end{aligned}$$

Those expressions at the left-hand sides (hence the sum of the three roots, a new expression in the second equation, and the product of the three roots) are the *elementary symmetric polynomials* in α, β, γ, when those are viewed as indeterminates. More generally, a polynomial $f(\alpha, \beta, \gamma)$ in three (independent) indeterminates α, β, γ is called *symmetric* if it is unchanged after *permuting* (that is, swapping around, rearranging) the three indeterminates, in any of the six possible ways:

$$f(\alpha,\beta,\gamma)=f(\beta,\alpha,\gamma)=f(\alpha,\gamma,\beta)=f(\gamma,\beta,\alpha)=f(\beta,\gamma,\alpha)=f(\gamma,\alpha,\beta).$$

In order to verify that a polynomial is symmetric it is actually sufficient to check that

$$f(\alpha,\beta,\gamma)=f(\beta,\alpha,\gamma)=f(\alpha,\gamma,\beta)$$

(interchanging the first two indeterminates, and interchanging the last two), because the remaining permutations of α, β, and γ, can be realised by appropriately repeating those two simple exchanges (called *transpositions*) in some order. As in the quadratic case one can prove that arbitrary symmetric polynomials can always be expressed in terms of the elementary ones.

Theorem 41. *Every symmetric polynomial $f(\alpha,\beta,\gamma)$ (with coefficients in any field F) can be expressed as a polynomial in the elementary symmetric polynomials $e_1(\alpha,\beta,\gamma)=\alpha+\beta+\gamma$, $e_2(\alpha,\beta,\gamma)=\alpha\beta+\alpha\gamma+\beta\gamma$, and $e_3(\alpha,\beta,\gamma)=\alpha\beta\gamma$.*

Hence if $f(\alpha,\beta,\gamma)$ is a symmetric polynomial in α, β, and γ, then

$$f(\alpha,\beta,\gamma)=g(\alpha+\beta+\gamma,\alpha\beta+\alpha\gamma+\beta\gamma,\alpha\beta\gamma),$$

for some polynomial $g(s,r,p)$ in three indeterminates s, r, and p. As in the quadratic case, more is true: if $f(\alpha,\beta,\gamma)$ has integer coefficients, then $g(s,r,p)$ has integer coefficients as well. For example, we have

$$\alpha^2+\beta^2+\gamma^2=(\alpha+\beta+\gamma)^2-2(\alpha\beta+\alpha\gamma+\beta\gamma)=s^2-2r.$$

Example. Suppose we want to express $\alpha^2\beta+\alpha^2\gamma+\beta^2\gamma+\alpha\beta^2+\alpha\gamma^2+\beta\gamma^2$ in terms of the elementary symmetric polynomials in α,β,γ. First of all, note that this is actually a symmetric polynomial, so Theorem 41 really applies. In fact, if we take one of the terms, say $\alpha^2\beta$, then by applying to it all permutations of α,β,γ we obtain precisely all terms of the sum (and each exactly once in this case). Hence according to Theorem 41 it must be possible to express this polynomial as a polynomial in the elementary symmetric polynomials $s=\alpha+\beta+\gamma$, $r=\alpha\beta+\alpha\gamma+\beta\gamma$, $p=\alpha\beta\gamma$.

Because $\alpha^2\beta$ can be written as the product of $\alpha\beta$ (which is a term of r) and α (which is a term of s), this suggests considering the product rs:

$$\begin{aligned}(\alpha\beta+\alpha\gamma+\beta\gamma)(\alpha+\beta+\gamma)&=\alpha^2\beta+\alpha^2\gamma+\alpha\beta\gamma\\&\quad+\alpha\beta^2+\alpha\beta\gamma+\beta^2\gamma\\&\quad+\alpha\beta\gamma+\alpha\gamma^2+\beta\gamma^2.\end{aligned}$$

Hence we conclude that

$$\alpha^2\beta+\alpha^2\gamma+\beta^2\gamma+\alpha\beta^2+\alpha\gamma^2+\beta\gamma^2=(\alpha\beta+\alpha\gamma+\beta\gamma)(\alpha+\beta+\gamma)-3\alpha\beta\gamma=rs-3p.$$

Example. Let α, β and γ be the complex roots of the polynomial $x^3 + x^2 - 2x - 5$. Compute $\alpha^2 + \beta^2 + \gamma^2$ and $\alpha^{-2} + \beta^{-2} + \gamma^{-2}$.

We express the desired quantities in terms of

$$\alpha + \beta + \gamma = -1, \quad \alpha\beta + \alpha\gamma + \beta\gamma = -2, \quad \text{and} \quad \alpha\beta\gamma = 5.$$

For the former quantity we have

$$\alpha^2 + \beta^2 + \gamma^2 = (\alpha + \beta + \gamma)^2 - 2(\alpha\beta + \alpha\gamma + \beta\gamma) = (-1)^2 - 2 \cdot (-2) = 5.$$

For the latter quantity, where negative powers of the roots appear, we consider the reciprocal polynomial $-5x^3 - 2x^2 + x + 1$, whose roots are α^{-1}, β^{-1}, and γ^{-1}. From the coefficients of the reciprocal polynomial we see that

$$\frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma} = -\frac{2}{5}, \quad \frac{1}{\alpha\beta} + \frac{1}{\alpha\gamma} + \frac{1}{\beta\gamma} = -\frac{1}{5}, \quad \text{and} \quad \frac{1}{\alpha\beta\gamma} = \frac{1}{5}.$$

Consequently, we have

$$\frac{1}{\alpha^2} + \frac{1}{\beta^2} + \frac{1}{\gamma^2} = \left(\frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma}\right)^2 - 2\left(\frac{1}{\alpha\beta} + \frac{1}{\alpha\gamma} + \frac{1}{\beta\gamma}\right) = \left(-\frac{2}{5}\right)^2 - 2 \cdot \left(-\frac{1}{5}\right) = \frac{14}{25}.$$

Note that one may also compute $\alpha^{-2} + \beta^{-2} + \gamma^{-2}$ without passing through the reciprocal polynomial, by directly expressing it in terms of the elementary symmetric polynomials:

$$\begin{aligned}\frac{1}{\alpha^2} + \frac{1}{\beta^2} + \frac{1}{\gamma^2} &= \frac{\alpha^2\beta^2 + \alpha^2\gamma^2 + \beta^2\gamma^2}{(\alpha\beta\gamma)^2} \\ &= \frac{(\alpha\beta + \alpha\gamma + \beta\gamma)^2 - 2(\alpha\beta\gamma)(\alpha + \beta + \gamma)}{(\alpha\beta\gamma)^2} \\ &= \frac{(-2)^2 - 2 \cdot 5 \cdot (-1)}{5^2} = \frac{14}{25}.\end{aligned}$$

However, this procedure appears a little more complicated than using the reciprocal polynomial (despite being equivalent to that).

35. (Optional) Symmetric polynomials in many indeterminates

Symmetric polynomials can be defined similarly for an arbitrary number n of indeterminates. The general definition of elementary symmetric polynomials, and the fundamental theorem of symmetric polynomials, are then as follows.

Definition 42. If $x_1, \dots, x_n$ and x are independent indeterminates, then the *elementary symmetric polynomials* $e_k = e_k(x_1, \dots, x_n)$, for $k = 1, \dots, n$, are defined by the identity

$$(x - x_1)(x - x_2) \cdots (x - x_n) = x^n - e_1 x^{n-1} + e_2 x^{n-2} - e_3 x^{n-3} + \cdots + (-1)^n e_n.$$

Hence e_k is the coefficient of x^{n-k} in the polynomial $(x-x_1)(x-x_2)\cdots(x-x_n)$ (once this is expanded into powers of x, its standard form). One can see that

$$e_k(x_1,\ldots,x_n) = \sum_{1\le j_1<j_2<\cdots<j_k\le n} x_{j_1}x_{j_2}\cdots x_{j_k}.$$

In particular, e_k is a *homogeneous* polynomial of degree k, and $e_1(x_1,\ldots,x_n) = x_1 + \cdots + x_n$, and $e_n(x_1,\ldots,x_n) = x_1\cdots x_n$. (A *homogeneous polynomial* of degree k is a polynomial (in several indeterminates) where each term alone has the same degree k.) More generally, according to the above formula $e_k(x_1,\ldots,x_n)$ is the sum of all distinct products of k distinct indeterminates taken from $x_1,\ldots,x_n$. The number of such distinct products equals the binomial coefficient $\binom{n}{k}$ (see one of the next sections).

Theorem 43 (The fundamental theorem of symmetric polynomials). *Every symmetric polynomial $f(x_1,\ldots,x_n)$ with integer coefficients can be expressed as a polynomial with integer coefficients in the elementary symmetric polynomials $e_k(x_1,\ldots,x_n)$, with $k=1,\ldots,n$.*

Lecture notes of Algebra. Week 9

36. Congruences in the integers

We will introduce congruences in the integers first, but everything we do will work very similarly later for congruences of polynomials with coefficients in a field.

Definition 44. Let $n \in \mathbb{Z}$. Two integers a, b are said to be *congruent modulo* n, and we write $a \equiv b \pmod{n}$, when n divides $a-b$, that is, there is $c \in \mathbb{Z}$ such that $a = b+cn$.

The special cases $n = 0$ and $n = 1$ are not very useful: $a \equiv b \pmod{0}$ is equivalent to $a = b$, and $a \equiv b \pmod{1}$ holds for any integers a, b. Also, because $a \equiv b \pmod{n}$ is always equivalent to $a \equiv b \pmod{(-n)}$, we may as well assume $n \geq 0$.

It is not difficult to see that $a \equiv b \pmod{n}$ is equivalent to each of the following conditions:

- b is in the arithmetic progression
$$\{a + nk\}_{k\in\mathbb{Z}} = \{\dots a - 2n, a - n, a, a + n, a + 2n, \dots\};$$
- the set $\{a + nk : k \in \mathbb{Z}\}$ is the same as the set $\{b + nk : k \in \mathbb{Z}\}$ (even if possibly indexed differently as arithmetic progressions);
- a and b give the same remainder when divided by n (assuming we have fixed one of the correct conventions for the remainders r, say $0 \leq r < n$, or $-n < 2r \leq n$).

It is also easy to see that congruence modulo a fixed integer n is an *equivalence relation*. This means that it is

- *reflexive:* $a \equiv a \pmod{n}$, for all a;
- *symmetric:* if $a \equiv b \pmod{n}$, then $b \equiv a \pmod{n}$;
- *transitive:* if $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$.

For $a \in \mathbb{Z}$, and the given n, we say that $\{a + nk : k \in \mathbb{Z}\}$ is the *congruence class modulo* n to which a belongs (or *the congruence class of* a when n is clear). Hence it is the set of all integers which are congruent to a modulo n. One also uses the shorter notation $[a]_n$ for the set $\{a + nk : k \in \mathbb{Z}\}$, and even just $[a]$ when n is clear. We also say that a is a representative (which really means the same as an element) for the class $[a]_n$. The set of congruence classes modulo n is denoted by $\mathbb{Z}/n\mathbb{Z}$, or also by (the less appropriate but shorter notation) $\mathbb{Z}_n$. Hence $\mathbb{Z}/n\mathbb{Z}$ has n elements (for n a positive integer), which are:

$$[0]_n, [1]_n, \dots, [n-1]_n.$$

However, note that each of them can be written in infinitely many different ways. For example, $[0]_n = [n]_n$, $[2]_5 = [7]_5 = [-38]_5$, etc. It is also important to note that the congruence classes modulo n form a *partition* of $\mathbb{Z}$, that is,

- their union is the whole $\mathbb{Z}$;

- any two distinct of them are disjoint (that is, they have empty intersection).

Another way of saying this is that every integer a belongs to *exactly one* of the congruence classes modulo n (in fact, to its class $[a]_n$). Note that $[a]_n = [b]_n$ is another way of writing $a \equiv b \pmod{n}$.

This is something which happens in general when we have a set S with an equivalence relation $\sim$ on it: the corresponding *equivalence classes* (defined as $[a]_\sim = \{b \in S : b \sim a\}$) form a partition of S. (Conversely, given a partition of S, there is a unique equivalence relation on S whose equivalence classes form that partition; hence *equivalence relations* and *partitions* are really different descriptions for the same concept.) The set of equivalence classes is denoted by $S/\sim$ and is called *the quotient set.*

Here is a geometric example of an equivalence relation: if S is the set of all straight lines which lie on a certain plane π, then the parallelism relation is an equivalence relation on S. Thus, here two straight lines r and r' are considered equivalent, $r \sim r'$, when they are parallel. An equivalence class, that is, the set of all straight lines (lying on π and) parallel to r, is often called a *direction.* Hence the quotient set $A/\sim$ is the set of directions of the plane π.

Back to congruences, in the special case where $n = 2$ we have common names for the congruence classes: the elements of $[0]_2$ being called *even,* and the elements of $[1]_2$ *odd.* Note also that, because of our decimal notation for writing numbers, the congruence classes modulo 10 are easy: two positive integers are congruent modulo 10 (that is, they belong to the same congruence class modulo 10) when they have the same right-most digit. (What about two negative integers, or a positive and a negative one?)

37. Solving the congruence $ax \equiv b \pmod{n}$

Consider the congruence $ax = b \pmod{n}$, where n is a nonzero integer (which we may assume positive if we like), a, b are arbitrary integers, and x is an unknown integer. We need to solve the congruence, which means finding *all* integer values of x which satisfy the congruence (that if finding all *solutions*). If it were a normal equation instead of a congruence, say with real numbers, then the procedure for solving it would be dividing both sides by a, assuming $a \neq 0$. However, this may not work with congruences, because congruences can always be added, subtracted, and multiplied, but not divided in general. We may still take inspiration from this analogy, but we have to proceed more carefully.

By definition, the congruence means $ax = b + kn$ for some $k \in \mathbb{Z}$, or $b = ax - kn$, and we have already learnt how to solve this for the unknowns x and k, using the extended Euclidean algorithm applied to a and n. Let us see how that procedure reads in the present congruence notation. Recall that if the greatest common divisor $d = (a, n)$ does not divide b, then there is no solution (because d would then divide both ax and kn, but not their difference, which is impossible). Hence the congruence $ax = b \pmod{n}$ has no solutions if d does not divide b.

Assume now that $d = (a, n)$ divides b. Then the extended Euclidean algorithm gives us integers s and t such that $d = as + nt$, and after multiplying by b/d we find $b = a(sb/d) + n(tb/d)$. In particular, we find a solution $x_0 = sb/d$ of our congruence $ax = b \pmod{n}$ (and also a corresponding value for k, namely $k_0 = -tb/d$, which we do not really need). To find all solutions, note that if x is any solution of $ax = b \pmod{n}$ then by subtracting $ax_0 = b \pmod{n}$ from this congruence we get $a(x - x_0) = 0 \pmod{n}$, which means $n \mid a(x - x_0)$. According to one of the arithmetical lemmas it follows that n/d divides $x - x_0$. Conversely, any integer x of the form $x = x_0 + h(n/d)$, for some $h \in \mathbb{Z}$, will be a solution, because

$$ax = a\left(x_0 + h\frac{n}{d}\right) = ax_0 + h\frac{a}{d} \cdot n \equiv ax_0 \equiv b \pmod{n}.$$

Hence the solutions of $ax = b \pmod{n}$ are exactly all integers x of the form $x = x_0 + h(n/d)$ for some $h \in \mathbb{Z}$. That is to say, the solutions of $ax = b \pmod{n}$ are described by $x \equiv x_0 \pmod{n/d}$. Hence note that solving $ax = b \pmod{n}$ means transforming it into a congruence of the form $x \equiv \cdots$, but with the possibly smaller modulus n/d. The modulus will remain the same only in the special case where $(a, n) = 1$. (With a terminology to be introduced in the next subsection, this occurs exactly when the class $[a]_n$ is invertible in $\mathbb{Z}/n\mathbb{Z}$.)

In practice, in the case where $d = (a, n)$ divides b it may be convenient to proceed in a different way, which is equivalent in theory but might be easier for calculations. In fact, the congruence $ax = b \pmod{n}$ is equivalent to the congruence $(a/d)x = b/d \pmod{n/d}$, obtained by dividing by d not just both sides, but also the modulus. This is because the former means $ax = b + kn$ for some $k \in \mathbb{Z}$, while the latter means $(a/d)x = b/d + k(n/d)$ for some $k \in \mathbb{Z}$, and those two facts are clearly equivalent. [16] Now $a'x = b' \pmod{n'}$, where $a' = a/d$, $b' = b/d$, and $n' = n/d$, belong to the special case where $(a', n') = (a/d, n/d) = 1$ (according to one of the arithmetical lemmas), which we have learnt how to solve first. Once we have found one solution x_0, the general solution will be $x \equiv x_0 \pmod{(a', b')}$.

Example. Find all solutions of the congruence $9\,x \equiv 4 \pmod{30}$.

The congruence means $9x = 4 + 30k$ for some $k \in \mathbb{Z}$, or $4 = 9x - 30k$. Because $(9, 30) = 3$ does not divide 4, the congruence has no solution.

[16] Note that we can always multiply both sides of a congruence by the same integer: the new congruence will be a consequence of the original congruence, but will generally be weaker than that (that is, it will generally not imply the original congruence). However, we obtain an equivalent congruence if we multiply not just both sides, but also the modulus, by the same integer; what we are using here is just the reverse procedure.

Example. Find all solutions of the congruence $7\,x \equiv 4 \pmod{30}$.

The congruence means $7x = 4 + 30k$ for some $k \in \mathbb{Z}$, or $4 = 7x - 30k$. Because $(7, 30) = 1$ obviously divides 4, we know that the congruence has solutions. In fact, the extended Euclidean algorithm gives $1 = -30 \cdot 3 + 7 \cdot 13$, and hence $7 \cdot 13 \equiv 1 \pmod{30}$. This means that 13 *is an inverse of* 7 *modulo* 30. (Or, if we prefer, it means that [13] *is the inverse of* [7] *in* $\mathbb{Z}/30\mathbb{Z}$, see a later section for this terminology.) Hence the congruence $7x \equiv 4 \pmod{30}$ can be solved by multiplying both sides by 13, and we obtain $x \equiv 13 \cdot 7x \equiv 13 \cdot 4 \equiv 22 \pmod{30}$.

Example. Find all solutions of the congruence $21\,x \equiv 12 \pmod{90}$.

Because $(21, 90) = 3$ divides 12, the congruence has solutions. In fact, the congruence means $21x = 12 + 90k$ for some $k \in \mathbb{Z}$, or $12 = 21x - 90k$. But this is equivalent to $4 = 7x - 30k$ for some $k \in \mathbb{Z}$, which in turn means $7\,x \equiv 4 \pmod{30}$.

So, in short, we may divide both sides and the modulus of the original congruence by 3 and obtain the equivalent congruence $7x \equiv 4 \pmod{30}$. In the previous example we solved this congruence and found $x \equiv 22 \pmod{30}$, so this describes all solutions.

Note that if we wanted to express all solutions modulo 90 rather than modulo 30, we would have three cases:

$$x \equiv 22 \pmod{90}, \quad \text{or} \quad x \equiv 52 \pmod{90}, \quad \text{or} \quad x \equiv 82 \pmod{90},$$

and so it is much simpler to describe the general solution as $x \equiv 22 \pmod{30}$.

38. Solving systems of two congruences

Let us start with a couple of examples.

Example. Suppose we want to solve $\begin{cases} x \equiv 3 \pmod{10} \\ x \equiv 6 \pmod{14} \end{cases}$.

So we want to find all integers x which are congruent to 3 modulo 10 and congruent to 6 modulo 14. But in particular the first congruence tells us that x is odd (as it differs a multiple of 10 from the odd number 3), and the second congruence tells us that x is even. Those are incompatible, and so the system has no solution. To phrase the argument in a more general way, the GCD 2 of 10 and 14 should divide both $x - 3$ and $x - 6$ (because 10 should divide $x - 3$ and 14 divides $x - 6$), and hence 2 should divide their difference, which is $6 - 3 = 3$, false.

Example. Change the previous system to $\begin{cases} x \equiv 3 \pmod{10} \\ x \equiv 11 \pmod{14} \end{cases}$.

Now $(10, 14) = 2$ divides $11 - 3 = 8 = 2 \cdot 4$, and so there is no obstacle as in the previous example. The extended Euclidean algorithm on 14 and 10 tells us that $2 = -14 \cdot 2 + 10 \cdot 3$.

Multiplying both sides by 4 we find $11-3=8=-14\cdot 8+10\cdot 12$. Moving certain terms to the opposite side we find $11+14\cdot 8=3+10\cdot 12$. Hence the common value $x_0=123$ of both sides is a solution of the system.

To find all solutions, note that if x_0 and x are two solutions of the system, which means

$$\begin{cases} x_0 \equiv 3 & \pmod{10} \\ x_0 \equiv 11 & \pmod{14} \end{cases} \quad \text{and} \quad \begin{cases} x \equiv 3 & \pmod{10} \\ x \equiv 11 & \pmod{14} \end{cases},$$

then $\begin{cases} x-x_0 \equiv 0 & \pmod{10} \\ x-x_0 \equiv 0 & \pmod{14} \end{cases}$ by taking differences. Hence $x-x_0$ is both a multiple of 10 and of 14, and so it is a multiple of their lowest common multiple $10\cdot 14/(10,14)=70$. Hence $x=123+70k$ for some integer k. Conversely, any choice of $k\in\mathbb{Z}$ gives us a solution of the system (because each of the two equations remains satisfied by adding to $x_0=123$ a multiple of 70), and so we have found *all* solutions of the system.

The general solution can also be expressed as $x\equiv 123 \pmod{70}$, that is, solving a system of congruences (if it has any solutions) means replacing it with a single congruence (with a different modulus) which is equivalent to it.

Note that the extended Euclidean algorithm gave us a particular solution $x_0=123$, from which we have obtained the general solution. The smallest positive solution of the system is $123-70=53$, and the smallest in absolute value is $53-70=-17$. We could have proceeded in slightly different ways, for example after finding $11-3=8=-14\cdot 8+10\cdot 12$ we could have changed the coefficients of the linear combination of 14 and 10 to $11-3=14\cdot 10-10\cdot 2$ (as we learnt when solving $ax+by=c$ in the integers), and that would have given us the particular solution $x_0'=3-10\cdot 2=11-14\cdot 2=-17$. another way could have been rewriting the system in the equivalent form $\begin{cases} x \equiv 3 & \pmod{10} \\ x \equiv -3 & \pmod{14} \end{cases}$, and because of the smaller numbers in absolute value solving it in the same way would have given us the particular solution $x_0=-17$.

Now we repeat the argument of the above examples in general, and we learn how to solve the system

$$\begin{cases} x \equiv a & \pmod{m} \\ x \equiv b & \pmod{n} \end{cases},$$

for arbitrary integers m,n,a,b. If there is a solution x_0, then for some integers s,t we have,

$$x_0=a+ms, \qquad x_0=b+nt,$$

and hence

$$a+ms=b+nt,$$

which we can rewrite as

$$b - a = ms - nt.$$

Hence, a necessary condition for the existence of a solution is that the greatest common divisor $d = (m, n)$ of m and n, which clearly divides the right-hand side, divides the left-hand side $b - a$as well. In other words, it is necessary that $a \equiv b \pmod{(m, n)}$. If this does not hold, then the system has no solution.

However, if $d \mid b - a$, then because of Bézout's lemma there are u, v such that

$$d = mu - nv,$$

and they can be found using the extended Euclidean algorithm. After multiplying by the integer $(b - a)/d$ we obtain

$$b - a = m\left(u \cdot \frac{b-a}{d}\right) - n\left(v \cdot \frac{b-a}{d}\right).$$

Therefore, we have found at least one solution of the system, namely,

$$x_0 = a + m\left(u \cdot \frac{b-a}{d}\right) = b + n\left(v \cdot \frac{b-a}{d}\right).$$

How do find *all* solutions? If x is any solution, then

$$\begin{cases} x \equiv x_0 & \pmod{m} \\ x \equiv x_0 & \pmod{n}, \end{cases}$$

and hence $x - x_0$ is a common multiple of m and n. Therefore, it is a multiple of their least common multiple $[m, n] = mn/(m, n)$. In other words, if the system has any solutions, then all solutions are the integers x such that

$$x \equiv x_0 \pmod{[m, n]},$$

where x_0 is one particular solution. Hence solving the above system (when it has solutions) is replacing the two congruences with a single congruence (with respect to a different modulus, $[m, n]$).

Example. Consider the following question: find all positive integers x, less than 100, which give remainder 4 when divided by 11, and remainder 12 when divided by 13.

The given conditions on x are

$$\begin{cases} x \equiv 4 & \pmod{11} \\ x \equiv 12 & \pmod{13} \\ 0 < x < 100. \end{cases}$$

The moduli 11 and 13 are coprime, hence the system of congruences certainly has solutions x, but of course we do not know how many, if any, will satisfy $0 < x < 100$. The extended Euclidean algorithm on 13 and 11 gives $1 = -13 \cdot 5 + 11 \cdot 6$. Hence $12 -$

$4 = 8 = -13 \cdot 40 + 11 \cdot 48$, and so one solution of the system of congruences will be $x_0 = 4 + 11 \cdot 48 = 12 + 13 \cdot 40 = 532$. Consequently, the general solution of the system is $x \equiv 532 \pmod{143}$, which is our concise way of writing $x = 532 + 143k$ for $k \in \mathbb{Z}$. So the solutions in descending order starting from 532 are 532, $532 - 143 = 389$, 246, 103, $-40, \ldots$, and we find that no solution x actually satisfies $0 < x < 100$. Hence there are no integers x satisfying the requirements. (The set of solutions to the question is empty.)

Note that a better way to proceed which involves smaller numbers is as follows. Once we get to $12 - 4 = 8 = -13 \cdot 40 + 11 \cdot 48$ we should remember that we may always add to the coefficient 40 some multiple of 11 as long as we subtract from the coefficient 48 the corresponding multiple of 13. Hence we could work instead with $12-4 = 8 = -13 \cdot 7 + 11 \cdot 9$, which gives us the smaller solution $4 + 11 \cdot 9 = 12 + 13 \cdot 7 = 103$ of the system of congruences. Or we could work with $12 - 4 = 8 = 13 \cdot 4 - 11 \cdot 4$, which gives us the solution $4 - 11 \cdot 4 = 12 - 13 \cdot 4 = -40$.

39. The Chinese Remainder Theorem

The special case of what we have discovered on a system of two congruences where m and n are coprime is called *the Chinese Remainder Theorem.*

Theorem 45 (Chinese Remainder Theorem). *Let m and n be integers with $(m, n) = 1$, and let a and b be arbitrary integers. Then the system of congruences*

$$\begin{cases} x \equiv a \pmod{m} \\ x \equiv b \pmod{n} \end{cases}$$

has always solutions, and any two solutions differ by a multiple of mn.

The last part of the statement is the easier one. Hence, in a way, the more important part of Theorem 45 is that the system has at least one solution. We have proved that in a *constructive* way, by giving an algorithm for finding a solution, based on the extended Euclidean algorithm.

40. Solving systems of more than two congruences

We have noted that solving a system of two congruences means replacing the system with a single congruence (with respect to a different modulus) which expresses the solutions of the system. Consequently, we now also know how to solve systems of many congruences (if there are any solutions): we start by solving the system consisting of the first two congruences (or any two, if other choices appear more convenient), which means replacing both with a single congruence, so that the system has now one less congruence; now we repeat the procedure until we are left with just one congruence.

Hence to solve a system $\begin{cases} x \equiv a \pmod{m} \\ x \equiv b \pmod{n} \\ x \equiv c \pmod{r} \end{cases}$ of three congruences, first solve the system made of the first two, for example, hence $\begin{cases} x \equiv a \pmod{m} \\ x \equiv b \pmod{n} \end{cases}$. If this has no solution,, then the full system of three congruences has no solution either. If it does have solutions, they can be written as a single congruence $x \equiv x_0 \pmod{[m,n]}$ as described earlier. Now put this together with the third equation of the original system: the original system is equivalent to $\begin{cases} x \equiv x_0 \pmod{[m,n]} \\ x \equiv c \pmod{r} \end{cases}$. Now solve this system in the same way. It may have no solutions, but if it does we know how to find them, and once we find one particular solution x_1 (which is probably different from x_0), the general solution will be $x = x_1 + k[m,n,r]$ for $k \in \mathbb{Z}$, where $[m,n,r]$ is the lowest common multiple of the three modules, which can also be written as $x \equiv x_1 \pmod{[m,n,r]}$.

Example. To solve the system $\begin{cases} x \equiv 3 \pmod{5} \\ x \equiv -2 \pmod{6} \\ x \equiv 1 \pmod{7} \end{cases}$ we start with solving the system consisting of the first two equations, $\begin{cases} x \equiv 3 \pmod{5} \\ x \equiv -2 \pmod{6} \end{cases}$.

Because the numbers involved are so small, it is easy to guess that -2 is a solution (but in any examination setting you should rather apply the standard method given by the extended Euclidean algorithm), and so all solutions are given by $x \equiv -2 \pmod{30}$. Hence the original system is equivalent to $\begin{cases} x \equiv -2 \pmod{30} \\ x \equiv 1 \pmod{7} \end{cases}$.

The extended Euclidean algorithm yields $1 = -30 \cdot 3 + 7 \cdot 13$, and hence $1 - (-2) = 3 = -30 \cdot 9 + 7 \cdot 39$. We may use this as it is, or make it easier by adding $30 \cdot 7$ to the first term and subtracting $7 \cdot 30$ from the second term, thus obtaining $1 - (-2) = 3 = -30 \cdot 2 + 7 \cdot 9$. Hence one solution of our system is $x \equiv -2 - 30 \cdot 2 = 1 - 7 \cdot 9 = -62$. Because $[5,6,7] = [30,7] = 210$, all solutions of the system are given by $x \equiv -62 \pmod{210}$. Equivalently, they are given by $x \equiv 148 \pmod{210}$.

41. Basic properties of congruences, and computing with classes

Congruences with respect to the same modulus can be added and multiplied: if $a \equiv a' \pmod{n}$, and $b \equiv b' \pmod{n}$, then

- $a + b \equiv a' + b' \pmod{n}$, and

- $ab \equiv a'b' \pmod{n}$.

For example, the latter holds because $ab - a'b' = ab - ab' + ab' - a'b' = a(b-b') + (a-a')b'$ a sum of two multiples of n, hence a multiple of n.

Warning: in general congruences cannot be divided. For example, we have $2 \equiv 6 \pmod{4}$, and $2 \equiv 2 \pmod{4}$, but $2/2 = 1 \not\equiv 6/2 = 3 \pmod{4}$. Note also the related fact that the *cancellation law* does not hold in general (it does only when the modulus is a prime): $2 \cdot 2 \equiv 0 \pmod{4}$, but $2 \not\equiv 0 \pmod{4}$.

The above properties imply that we can define two operations with congruence classes modulo n, namely addition and multiplication, as follows:

- $[a]_n + [b]_n := [a+b]_n$, and
- $[a]_n \cdot [b]_n := [a \cdot b]_n$.

In fact, those properties insure that the result of any of these operations does not depend on which representatives of the two classes (here a and b) we choose to compute the operations. For example, $[2]_{10} \cdot [3]_{10} = [6]_{10}$, and if we rewrite $[2]_{10}$ and $[3]_{10}$ in the equivalent way $[12]_{10}$ and $[-7]_{10}$, say, then our rule for multiplication yields $[12]_{10} \cdot [-7]_{10} = [-84]_{10}$, which is the same as $[6]_{10}$. In fact, in general we may rewrite $[a]_n$ as $[a']_n$, where we mean $[a]_n = [a']_n$, and similarly $[b]_n = [b']_n$. If we choose a and b as representatives of the two classes then our definition of their product yields $[ab]_n$, and if we choose a' and b' as representatives then the same definition yields $[a'b']_n$ as the product of the classes. Fortunately, $[ab]_n = [a'b']_n$ according to the second basic property of congruences seen above. This good behaviour may not be guaranteed if we had defined some operation between classes in some careless way. What we have just done is checking that the operation of multiplication between classes is *well defined.* The same should be done for addition. [17]

This problem of *good definition* may be a little difficult to grasp at first (and will be studied in detail in a later module), but here is an example of how it occurs in a different settings. If we wanted to define *the numerator of a rational number* we would incur into a similar problem. In fact, the numerator of a fraction is clear, but a rational number can be written as a fraction of integers in many different ways (in fact, infinitely many): $2/3 = 4/6 = 6/9 = (-2)/(-3)$, etc. Which of the numbers before the fraction sign should we take as the numerator of *the rational number* (represented by the fraction) $2/3$? Hence such *numerator of a rational number* was not *well defined.* Of course we can make it into a *good definition* if we specify a unique way of choosing the numerator, for example imposing that it must be positive and coprime with the denominator of a fraction.

[17] The existence of the two operations on elements of $\mathbb{Z}/n\mathbb{Z}$, and the basic properties those operations satisfy, can be summarised by saying that $\mathbb{Z}/n\mathbb{Z}$ becomes a *commutative ring* with those operations.

A perhaps more convincing example occurs if we wanted to invent a new, strange operation with rational numbers, call it $\oplus$, say, by setting

$$\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d}.$$

Ignoring the additional problem that the denominator $b+d$ might be zero, we note that $1/2 = 2/4$, but

$$\frac{1}{2} \oplus \frac{2}{3} = \frac{3}{5} \neq \frac{4}{7} = \frac{2}{4} \oplus \frac{2}{3}.$$

Again, the result depends on the way we write the rational number 1/2, and so it is not *well defined*.

Example. The addition and multiplication tables for $\mathbb{Z}/5\mathbb{Z}$ are as follows:

+	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]	[0]
[2]	[2]	[3]	[4]	[0]	[1]
[3]	[3]	[4]	[0]	[1]	[2]
[4]	[4]	[0]	[1]	[2]	[3]

·	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]
[2]	[0]	[2]	[4]	[1]	[3]
[3]	[0]	[3]	[1]	[4]	[2]
[4]	[0]	[4]	[3]	[2]	[1]

One may also write $[-2]$ instead of [3] and $[-1]$ instead of [4] everywhere in these tables, to keep the calculations smaller (especially when dealing with $\mathbb{Z}/n\mathbb{Z}$ for larger n).

Example. The addition and multiplication tables for $\mathbb{Z}/6\mathbb{Z}$ are as follows:

+	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[1]	[2]	[3]	[4]	[5]
[1]	[1]	[2]	[3]	[4]	[5]	[0]
[2]	[2]	[3]	[4]	[5]	[0]	[1]
[3]	[3]	[4]	[5]	[0]	[1]	[2]
[4]	[4]	[5]	[0]	[1]	[2]	[3]
[5]	[5]	[0]	[1]	[2]	[3]	[4]

·	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

The addition table modulo 6 does not really look much different from that modulo 5, and so is rather predictable modulo any n, and not so interesting. The multiplication tables, however, look quite different. We will soon come back to these tables to understand the reason for these differences.

Lecture notes of Algebra. Week 10

42. *Casting out nines,* and divisibility criteria

Here is an immediate application of what we have seen, which shows the convenience of working with congruence classes. Let $n = 9$. Consider a positive integer written in decimal notation $a = (a_{k-1}a_{k-2} \dots a_1 a_0)_{10}$, which means [18]

$$(a_{k-1}a_{k-2} \dots a_1 a_0)_{10} = a_{k-1}10^{k-1} + a_{k-2}10^{k-2} + \cdots + a_1 10 + a_0,$$

where $a_j \in \{0, 1, 2, \dots 9\}$, the allowed digits in base 10.

Now we compute the class of $[a]$ according to the rules we have seen. since n is fixed we may omit it and safely write congruence classes as $[a]$ rather than $[a]_n$. We have

$$\begin{aligned} [a] &= [a_{k-1}10^{k-1} + a_{k-2}10^{k-2} + \cdots + a_1 10 + a_0] \\ &= [a_{k-1}][10]^{k-1} + [a_{k-2}][10]^{k-2} + \cdots + [a_1][10] + [a_0] \\ &= [a_{k-1}] + [a_{k-2}] + \cdots + [a_1] + [a_0] \\ &= [a_{k-1} + a_{k-2} + \cdots + a_1 + a_0]. \end{aligned}$$

Here we have used the fact that $[10] = [1]$. We have found that the congruence class modulo 9 of a positive integer is the same as that of the sum of its digits. This is exactly what is used in the correctness check for arithmetical operations called *casting out nines:*

$$\begin{aligned} [178564] &= [1 + 7 + 8 + 5 + 6 + 4] = [1 + 8] + [7 + 5] + [6 + 4] \\ &= [9] + [12] + [10] = [0] + [1 + 2] + [1 + 0] = [4]. \end{aligned}$$

The same argument also gives us the well-known divisibility criterion by 9 (or by 3, which is analogous because 3 divides 9): a positive integer, written in decimal notation, is divisible by 9 exactly when the sum of its digits is..

If, instead, we consider $n = 11$, then we find

$$\begin{aligned} [a] &= [a_{k-1}10^{k-1} + a_{k-2}10^{k-2} + \cdots + a_1 10 + a_0] \\ &= [a_{k-1}][10]^{k-1} + [a_{k-2}][10]^{k-2} + \cdots + [a_1][10] + [a_0] \\ &= [a_{k-1}][-1]^{k-1} + [a_{k-2}][-1]^{k-2} + \cdots - [a_1] + [a_0] \\ &= [a_{k-1}(-1)^{k-1} + a_{k-2}(-1)^{k-2} + \cdots - a_1 + a_0]. \end{aligned}$$

[18] More generally, we may use a similar notation to denote a number written in any base b (as long as b is an integer with $b > 1$). We use the notation $(\cdot)_b$ not just to remind us that b is the basis used, but also that what is inside the parentheses is not the product of the digits, of course, but a sequence of digits; we may separate the digits with commas as in $a = (a_{k-1}, a_{k-2}, \dots a_1, a_0)_{10}$ for better clarity, but then the notation may become cumbersome. Note that we have named the left-most digit a_{k-1} rather than a_k simply because if $a_{k-1} \neq 0$ then the integer has exactly k digits.

Hence the congruence class modulo 11 of a positive integer is the same as that of the sum of its digits taken *with alternating signs* (starting from the right-most digit taken with the positive sign). Here we have been smart enough to write $[10] = [-1]$.

In a similar way one can devise similar divisibility criteria for other values of n, but they may become more laborious. For example, let $n = 7$. We have

$$\begin{aligned}
(\quad [10]^0 &= [1], \quad) \\
[10]^1 &= [3], \\
[10]^2 &= [3]^2 = [9] = [2], \\
[10]^3 &= [10]^2 \cdot [10] = [2] \cdot [3] = [6] = [-1], \\
[10]^4 &= [10]^3 \cdot [10] = [-1] \cdot [3] = [-3], \\
[10]^5 &= [10]^3 \cdot [10]^2 = [1] \cdot [2] = [-2], \\
[10]^6 &= ([10]^3)^2 = [-1]^2 = [1],
\end{aligned}$$

and so from this point on the powers of [10] will repeat. (We will soon see the theory behind this.) Therefore, with similar calculations as we did for previous values of n, we find

$$\begin{aligned}
[a] = [&a_0 + 3a_1 + 2a_2 - a_3 - 3a_4 - 2a_5 + \\
&a_6 + 3a_7 + 2a_8 + \dots].
\end{aligned}$$

This divisibility criterion by 7 is rather complicated, as it involves a sequence of coefficients repeating every 6 steps. A divisibility criterion by 13 would require again a repetition of 6 steps, and a divisibility criterion by 17 would require again a repetition of 16 steps. We are better off with a criterion of divisibility by 37, which repeats after only three steps:

$$\begin{aligned}
(\quad [10]^0 &= [1], \quad) \\
[10]^1 &= [10], \\
[10]^2 &= [100] = [-11], \\
[10]^3 &= [-11] \cdot [10] = [-110] = [1].
\end{aligned}$$

Hence in $\mathbb{Z}/37\mathbb{Z}$ we have

$$\begin{aligned}
[a] = [&a_0 + 10a_1 - 11a_2 + \\
&a_3 + 10a_4 - 11a_5 + \dots].
\end{aligned}$$

This criterion is short because $37 \mid 10^3 - 1$; in fact, $10^3 - 1 = 3^3 \cdot 37$.

If we only need to know whether an integer is divisible by n, and do not care about its precise reminder in case it is not, then simpler recursive criterions are available. For example, in the case of $n = 7$ we may note that $a = (a_{k-1}a_{k-2}\dots a_1a_0)_{10} = 10b + a_0$, where $b = (a_{k-1}a_{k-2}\dots a_1)_{10}$ is obtained from a by omitting its right-most digit. Now 7 divides a exactly when 7 divides $-2a = -20b - 2a_0 \equiv b - 2a_0 \pmod 7$. (We have chosen to multiply by -2 because that is an inverse of 10 modulo 7.) In other words, we have found that 7 divides $(a_{k-1}a_{k-2}\dots a_1a_0)_{10}$ exactly when 7 divides the difference $(a_{k-1}a_{k-2}\dots a_1)_{10} - 2a_0$, which has one digit less. Now we may apply this step recursively.

43. Invertible elements in $\mathbb{Z}/n\mathbb{Z}$

Definition. A congruence class $[a]_n$ in $\mathbb{Z}/n\mathbb{Z}$ is *invertibile* when there exists a class $[b]_n$ such that $[a]_n \cdot [b]_n = [1]_n$. We call the class $[b]_n$ the *inverse* of $[a]_n$ (or the inverse of $[a]$ in $\mathbb{Z}/n\mathbb{Z}$).

Sometimes one also says that *(the integer) b is an inverse of (the integer) a modulo n.* We also write $[a]_n^{-1} = [b]_n$ (as taking a power with exponent -1 means taking the reciprocal).

For example, the class $[2]_5$ in $\mathbb{Z}/5\mathbb{Z}$ is invertible, because $[2]_5 \cdot [3]_5 = [1]_5$, and so $[3]_5$ is its inverse. Or we say that 3 is an inverse of 2 modulo 5, and 2 is an inverse of 3 modulo 5.

If a class $[a]$ in $\mathbb{Z}/n\mathbb{Z}$ has an inverse (and hence is invertible), then that inverse is unique: if $[b]$ and $[b']$ are both inverses of $[a]$, then

$$[b'] = [b'] \cdot 1 = [b'] \cdot \big([a] \cdot [b]\big) = \big([b'] \cdot [a]\big) \cdot [b] = [1] \cdot [b] = [b].$$

That is why we can talk about *the inverse* rather than *an inverse.*

Example. Look again at the multiplication tables for $\mathbb{Z}/5\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$:

$\cdot$ in $\mathbb{Z}/5\mathbb{Z}$	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]
[2]	[0]	[2]	[4]	[1]	[3]
[3]	[0]	[3]	[1]	[4]	[2]
[4]	[0]	[4]	[3]	[2]	[1]

$\cdot$ in $\mathbb{Z}/6\mathbb{Z}$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

We see that all nonzero classes in $\mathbb{Z}/5\mathbb{Z}$ (that is, all except the zero class [0]) are invertible: in each row corresponding to each of those classes the class [1] appears somewhere, precisely in the column corresponding to the inverse class. However, in the multiplication table for $\mathbb{Z}/6\mathbb{Z}$ this only occurs for the classes [1] and $[5] = [-1]$: those are the only invertible classes in $\mathbb{Z}/6\mathbb{Z}$. Actually, in each of the remaining rows the class [0] appears more than once, meaning that there are nonzero classes which multiplied give the zero class: $[2] \cdot [3] = [0]$ and $[4] \cdot [3] = [0]$.

Proposition 46. *A class $[a]_n$ in $\mathbb{Z}/n\mathbb{Z}$ is invertible precisely when $(a, n) = 1$.*

Proof. By definition $[a]$ in $\mathbb{Z}/n\mathbb{Z}$ is invertible when $[a] \cdot [b]$ for some class $[b] \in \mathbb{Z}/n\mathbb{Z}$. That means $ab \equiv 1 \pmod{n}$, hence $[a]$ is invertible when there exists k such that $ab = 1 + nk$, which we can also write as $ab - nk = 1$.

Suppose such an integer b exists. Then the greatest common divisor of a and n divides the left-hand side of $ab - nk = 1$, but then it must divide the right-hand side 1, and hence $(a, n) = 1$.

Conversely, if $(a, n) = 1$ then the extended Euclidean algorithm allows us to find integers b and k such that $ab - nk = 1$. Then the class $[b]$ is the desired inverse of $[a]$. $\square$

In particular, when $n = p$, a prime number, every nonzero class $[a]$ is invertible, because $(a, p) = 1$ if $a \not\equiv 0 \pmod{p}$. Hence $\mathbb{Z}/p\mathbb{Z}$ is a *field* when p is a prime number, the *field with p elements.* Being a field one also uses the notation $\mathbb{F}_p$ for $\mathbb{Z}/p\mathbb{Z}$ (but not for $\mathbb{Z}/n\mathbb{Z}$ when n is not prime, because in that case we do not have a field). You will learn a proper definition of a field in the second-year module *Algebraic Structures,* but for now just think that a field is roughly a set of 'numbers' which can be added, subtracted, multiplied and divided (except for dividing by zero, which is never possible) and follow the same rules as more familiar numbers such as the rational numbers or the real numbers. (The integers are not a field because they cannot generally divide one by another *exactly,* that is, without a remainder.) So they can be used, for examples as coefficients for polynomials, as we will see in a later section.

Example. The multiplication table of $\mathbb{Z}/8\mathbb{Z}$ is as follows (discarding the square brackets to denote classes, for simplicity).

$\cdot \pmod{8}$	0	1	2	3	4	-3	-2	-1
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	-3	-2	-1
2	0	2	4	-2	0	2	4	-2
3	0	3	-2	1	4	-1	2	-3
4	0	4	0	4	0	4	0	4
-3	0	-3	2	-1	4	1	-2	3
-2	0	-2	4	2	0	-2	4	2
-1	0	-1	-2	-3	4	3	2	1

As we know from the Proposition, the invertible classes are those represented by integers coprime with 8, that is to say by odd integers, hence the invertible classes are $[1]$, $[-1]$, $[3]$ and $[-3]$. From the main diagonal of the multiplication table we see that every invertible class has square equal to $[1]$, meaning that it is the inverse of itself (while the remaining classes have square $[0]$ or $[4]$).

In terms of congruences, this means that the square of every odd integer is congruent to 1 modulo 8. One can also see this as follows. Writing an odd integer as $2k+1$ we have $(2k+1)^2 = 4k^2 + 4k + 1 = 4k(k+1) + 1$. This is clearly congruent to 1 modulo 4, but because either k or $k+1$ is even, it is actually congruent to 1 modulo 8.

Example. The multiplication table of $\mathbb{Z}/15\mathbb{Z}$ is as follows:

$\cdot$ (mod 15)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
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
2	0	2	4	6	8	10	12	14	1	3	5	7	9	11	13
3	0	3	6	9	12	0	3	6	9	12	0	3	6	9	12
4	0	4	8	12	1	5	9	13	2	6	10	14	3	7	11
5	0	5	10	0	5	10	0	5	10	0	5	10	0	5	10
6	0	6	12	3	9	0	6	12	3	9	0	6	12	3	9
7	0	7	14	6	13	5	12	4	11	3	10	2	9	1	8
8	0	8	1	9	2	10	3	11	4	12	5	13	6	14	7
9	0	9	3	12	6	0	9	3	12	6	0	9	3	12	6
10	0	10	5	0	10	5	0	10	5	0	10	5	0	10	5
11	0	11	7	3	14	10	6	2	13	9	5	1	12	8	4
12	0	12	9	6	3	0	12	9	6	3	0	12	9	6	3
13	0	13	11	9	7	5	3	1	14	12	10	8	6	4	2
14	0	14	13	12	11	10	9	8	7	6	5	4	3	2	1

However, it may be more convenient to write the elements of $\mathbb{Z}/15\mathbb{Z}$ as $[0], [1], [2], \ldots [7]$ followed by $[-7], [-6], \ldots, [-2], [-1]$, and the table then reads:

$\cdot$ (mod 15)	0	1	2	3	4	5	6	7	−7	−6	−5	−4	−3	−2	−1
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	−7	−6	−5	−4	−3	−2	−1
2	0	2	4	6	−7	−5	−3	−1	1	3	5	7	−6	−4	−2
3	0	3	6	−6	−3	0	3	6	−6	−3	0	3	6	−6	−3
4	0	4	−7	−3	1	5	−6	−2	2	6	−5	−1	3	7	−4
5	0	5	−5	0	5	−5	0	5	−5	0	5	−5	0	5	−5
6	0	6	−3	3	−6	0	6	−3	3	−6	0	6	−3	3	−6
7	0	7	−1	6	−2	5	−3	4	−4	3	−5	2	−6	1	−7
−7	0	−7	1	−6	2	−5	3	−4	4	−3	5	−2	6	−1	7
−6	0	−6	3	−3	6	0	−6	3	−3	6	0	−6	3	−3	6
−5	0	−5	5	0	−5	5	0	−5	5	0	−5	5	0	−5	5
−4	0	−4	7	3	−1	−5	6	2	−2	−6	5	1	−3	−7	4
−3	0	−3	−6	6	3	0	−3	−6	6	3	0	−3	−6	6	3
−2	0	−2	−4	−6	7	5	3	1	−1	−3	−5	−7	6	4	2
−1	0	−1	−2	−3	−4	−5	−6	−7	7	6	5	4	3	2	1

By locating [1] in each row where it appears, we find the pairs of inverse elements in the corresponding row and column. One thus finds that the invertible elements are

- [1], with inverse [1];
- [2], with inverse [8] = [−7]; this is because $[2]\cdot[-7]=[-14]=[1]$;
- [4], with inverse [4] (itself); this is because $[4]^2=[16]=[1]$;
- [7], with inverse [13] = [−2];

and their opposites

- [−1], with inverse [−1];
- [−2], with inverse [7];
- [−4], with inverse [−4] (itself);
- [−7], with inverse [2].

Note that the search is made faster by a couple of facts:

- if $[a]$ is invertible, with inverse $[b]$, then $[b]$ is also invertible, with inverse $[a]$;
- if $[a]$ is invertible, with inverse $[b]$, then $[-a]$ is invertible, with inverse $[-b]$; this is because if $[a]\cdot[b]=1$, then $[-a]\cdot[-b]=[(-a)(-b)]=[ab]=[a]\cdot[b]=1$.

We may note that the multiplication table can be divided into four quadrants, the top-left one being as follows:

· (mod 15)	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	−7	−5	−3	−1
3	0	3	6	−6	−3	0	3	6
4	0	4	−7	−3	1	5	−6	−2
5	0	5	−5	0	5	−5	0	5
6	0	6	−3	3	−6	0	6	−3
7	0	7	−1	6	−2	5	−3	4

Because each of the four quadrants can be obtained from this by reflections and sign changes, we could answer the question about inverses by looking just at this portion of the multiplication table. For example, the −1 in the last line tells us that $[7]\cdot[2]=[-1]$, whence [7] is invertible, with inverse [−2].

Example. Fix $n=19$, that is, we will work in $\mathbb{F}_{19}=\mathbb{Z}/19\mathbb{Z}$, the field of 19 elements, and denote classes modulo 19 by $[a]$ rather than the more precise $[a]_{19}$. Because 19 is prime every nonzero class is invertible. For example, the inverse of [4] is [5] because $[4]\cdot[5]=[20]=[1]$, and we may write $[4]^{-1}=[5]$.

Because $[6]\cdot[3] = [18] = [-1]$ we also see that $[6]\cdot[-3] = [1]$, and so $[6]^{-1} = [-3] = [16]$. Here we see that it is more convenient to use the representatives $-9, \dots, 9$ for the classes modulo 19, rather than $0, \dots, 18$.

To find the inverse of [7], which is less obvious, use the extended Euclidean algorithm

$$\begin{aligned} 19 &= 7\cdot 3 - 2 & &= 7 + (19 - 7\cdot 3)\cdot 3 = 19\cdot 3 - 7\cdot 8 \\ 7 &= 2\cdot 3 + 1 & 1 &= 7 - 2\cdot 3 \end{aligned}$$

Hence $1 = 19\cdot 3 - 7\cdot 8$, and so $[7]\cdot[-8] = [1]$, meaning that the inverse of [7] is $[-8]$ (or [11] if one prefers).

The value of the fraction $[3]/[7]$ in $\mathbb{F}_{19}$ is $[3]\cdot[7]^{-1} = [3]\cdot[-8] = [-24] = [-5]$. To check the result, compute $[-5]\cdot[7] = [-35] = [3]$. Note that sometimes one simply writes $3/7 = -5$ in $\mathbb{F}_{19}$, abusing notation (which is useful when doing many such calculation, in a fixed finite field).

44. Wilson's theorem

Here is an example of application of the concept of inverse to prove a classical result.

Theorem 47 (Wilson's Theorem). *If p is a prime then* $(p-1)! \equiv -1 \pmod{p}$.

For example, $10! = 3628800$, and we can confirm that $10! \equiv -1 \pmod{11}$ using the divisibility criterion modulo 11, as $0 - 0 + 8 - 8 + 2 - 6 + 3 = -1$.

Proof. In alternate notation, our goal is proving that

$$[1]\cdot[2]\cdot[3]\cdots[p-1] = [-1]$$

in $\mathbb{Z}/p\mathbb{Z}$. The factors on the left-hand side are exactly all nonzero elements of $\mathbb{Z}/p\mathbb{Z}$. Now each of those elements $[a]$ can be paired off with its inverse $[a]^{-1}$, which will also appear as a factor, except for those cases where $[a]$ is its own inverse, which means $[a]\cdot[a] = [1]$. This is equivalent to $a^2 - 1 \equiv 0 \pmod{p}$, or again to $p \mid (a-1)(a+1)$. Because p is a prime this only occurs if p divides at least one of the two factors, which means $a \equiv \pm 1 \pmod{p}$.

In conclusion, after pairing off in the product $[1]\cdot[2]\cdot[3]\cdots[p-1]$ and removing each cancelling pair $[a]\cdot[a]^{-1}$ with $[a] \neq [a]^{-1}$, one is left with $[1]\cdot[p-1] = [-1]$ when $p > 2$, and with $[1] = [-1]$ when $p = 2$. This proves the conclusion in all cases. □

45. Computing powers efficiently in $\mathbb{Z}/n\mathbb{Z}$

Example. Compute $[7]^{-12}$ in $\mathbb{F}_{19} = \mathbb{Z}/19\mathbb{Z}$. Powers with a negative exponent only make sense for invertible classes, but [7] is indeed an invertible class in $\mathbb{Z}/19\mathbb{Z}$.

The usual rules for powers hold like for ordinary numbers, so $[7]^{-12}$ means either $([7]^{-1})^{12}$ or $([7]^{12})^{-1}$, which are the same. We have found in a previous example that $[7]^{-1} = [-8]$, hence $[7]^{-12} = [-8]^{12} = [-1]^{12} \cdot [8]^{12} = [8]^{12}$.

Now it would be generally silly to multiply [8] by itself many times, so if possible one should follow a shorter path, for example using $12 = 2 \cdot 2 \cdot 3$:

$$[8]^{12} = 8^{2\cdot 2\cdot 3} = (([8]^2)^2)^3 = ([64]^2)^3 = ([7]^2)^3 = [49]^3 = [-8]^3 = [-8]^2\cdot[-8] = [7]\cdot[-8] = [1].$$

Hence $[7]^{-12} = [1]$. Here is another way to do the calculation:

$$[8]^{12} = [8]^{3\cdot 2\cdot 2} = (([8]^3)^2)^2 = (([8]^2 \cdot [8])^2)^2 = (([64] \cdot [8])^2)^2 = (([7] \cdot [8])^2)^2 = ([-1]^2)^2 = [1].$$

And yet another way of doing the whole calculation from the start:

$$\begin{aligned}[7]^{-12} = [7]^{3\cdot 2\cdot 2\cdot(-1)} &= (((([7]^3)^2)^2)^{-1} \\ &= (((([7]^2 \cdot [7])^2)^2)^{-1} = ((([-8] \cdot [7])^2)^2)^{-1} = (([1]^2)^2)^{-1} = [1].\end{aligned}$$

Here we were lucky that $[7]^3 = [1]$, which of course was not predictable.

Example. Find the last decimal digit of 13^{999}.

The last decimal digit of a positive integer a tells us to which congruence class modulo 10 the number a belongs, so the question is really computing $[13]^{999}$ in $\mathbb{Z}/10\mathbb{Z}$, which is the same as $[3]^{999}$. Differently from the previous example, factorising the exponent 999 would not be very useful ($999 = 3^4 \cdot 37$). However, if we compute the first few powers of [3], sequentially, we find

$$\begin{aligned}&[3]^1 = [3] \\ &[3]^2 = [3] \cdot [3] = [9] = [-1] \\ &[3]^3 = [3]^2 \cdot [3] = [-1] \cdot [3] = [-3] \\ &[3]^4 = [3]^3 \cdot [3] = [-3] \cdot [3] = [-9] = [1]\end{aligned}$$

Actually, we could have noted that $[3]^4 = [1]$ right after finding $[3]^2 = [-1]$, as $[-1]^2 = [1]$. Because $[1] = [3]^0$ this means that the powers of [3] will repeat every four steps. Hence what we do is divide the exponent 999 by 4 with remainder, $999 = 4 \cdot 249 + 3$, and so

$$[3]^{999} = [3]^{4\cdot 249+3} = ([3]^4)^{249} \cdot [3]^3 = [1]^{249} \cdot [3]^3 = [3]^3 = [-3].$$

Because $[-3] = [7]$, we find that $13^{999} \equiv 7 \pmod{10}$, and so its last decimal digit is 7.

46. Euler's function

For applications it is important to know the number of invertible classes in $\mathbb{Z}/n\mathbb{Z}$. As a function of n, a positive integer, this is called *Euler's phi function* (or *Euler's totient* *function*), denoted by φ. Thus, $\varphi(n)$ is the number of invertible classes in $\mathbb{Z}/n\mathbb{Z}$. However, because of Proposition 46 it can also be computed as

$$\varphi(n) = |\{0 \leq k < n : (k, n) = 1\}|$$

Hence $\varphi(n)$ equals the number of integers k, in the range $0 \leq k < n$, which are coprime with n.

Here are some special cases. When $n = p$ is a prime we have $\varphi(p) = p - 1$.

When $n = p^2$ is the square of a prime p then $\varphi(p^2) = p^2 - p = p(p-1)$. This is because the integers which *are not* coprime with p^2 are precisely those which are multiples of p. The multiples p in the range $0 \leq k < n$ under considerations are those integers k of the form $k = pj$, where $0 \leq j < n/p$. Hence their number is $n/p = p$, and subtracting this from $n = p^2$ gives the result.

An analogous argument shows that when $n = p^r$ is a power of a prime p we have $\varphi(p^r) = p^r - p^{r-1} = p^{r-1}(p-1)$.

Another special case is when $n = pq$ is a product of two distinct primes p and q. In fact,the integers k which *are not* coprime with pq are either multiples of p or multiples of q. Hence in the range $0 \leq k < pq$ there are exactly $p + q - 1$ of them (having subtracted 1 because $k = 0$ was counted twice), and hence [19]

$$\varphi(pq) = pq - (p + q - 1) = (p-1)(q-1).$$

47. Fermat's little theorem

THEOREM 48 (Fermat's little theorem). *If p is a prime, and a is any integer, then $a^p \equiv a \pmod{p}$.*

PROOF (OPTIONAL). If a is a multiple of p then the conclusion is clear, as both sides of the congruence are congruent to 0 modulo p. Hence suppose that a is not a multiple of p. Then we will prove that we actually have $a^{p-1} \equiv 1 \pmod{p}$, from which the conclusion follows after multiplying by a.

In fact, if a is not a multiple of p then the classes $[a], [2a], [3a], \ldots, [(p-1)a]$ in $\mathbb{Z}/p\mathbb{Z}$ are all different from $[0]$, and they are all distinct. In fact, if $[ia] = [ja]$ for some $0 \leq i, j < p$, then $[(i-j)a] = [0]$, which means that $(i-j)a$ is a multiple of p. Because p is a prime and does not divide a, it must divide $i - j$, and it follows that $i = j$. Hence different i, j in the range $0 \leq i, j < p$ give rise to different classes $[ia]$ and $[ja]$.

In conclusion, the classes $[a], [2a], [3a], \ldots, [(p-1)a]$ are $p-1$ distinct nonzero classes, and so they coincide with all $p-1$ nonzero classes in $\mathbb{Z}/p\mathbb{Z}$, which are $[1], [2], [3], \ldots, [p-1]$.

[19] Note that $\varphi(pq)$ equals $\varphi(p) \cdot \varphi(q)$. In fact, we have more generally $\varphi(mn) = \varphi(m) \cdot \varphi(m)$ whenever $(m, n) = 1$. This can be proved using the Chinese remainder theorem.

Hence the product

$$[1] \cdot [2] \cdot [3] \cdots [p-1] = [(p-1)!],$$

which is nonzero, and hence invertible in $\mathbb{Z}/p\mathbb{Z}$, equals the product

$$\begin{aligned} [a] \cdot [2a] \cdot [3a] \cdots [(p-1)a] &= [1] \cdot [a] \quad \cdot [2] \cdot [a] \quad \cdot [3] \cdot [a] \quad \cdots [p-1] \cdot [a] \\ &= \big([1] \cdot [2] \cdot [3] \cdots [p-1]\big) \cdot [a]^{p-1} \\ &= [(p-1)!] \cdot [a]^{p-1}. \end{aligned}$$

Multiplying both sides of the equality $[(p-1)!] \cdot [a]^{p-1} = [(p-1)!]$ by the inverse of the nonzero element $[(p-1)!]$ gives us $[a]^{p-1} = [1]$ in $\mathbb{Z}/p\mathbb{Z}$, as desired. □

Note that in order to carry out the above proof we need not know the actual value of $[(p-1)!]$ (which is given by Wilson's theorem in the previous section, Theorem 47), but only that it is different from the zero class $[0]$, and hence it is invertible because p is a prime. In the course of the proof of Fermat's little theorem we have actually proved the following more precise statement.

Corollary 49. *If p is a prime, and a is any integer not divisible by p, then $a^{p-1} \equiv 1 \pmod{p}$.*

Although we have proved this fact first in order to prove Fermat's little theorem, we have called it a Corollary because of historical reasons, and because it can also be easily deduced from Fermat's little theorem, as we show now. [20]

Proof. According to Fermat's little theorem, the prime p divides the integer $a^p - a = a(a^{p-1}-1)$. However, here we have the additional hypothesis that $p \nmid a$, and hence p divides $a^{p-1}-1$. This is equivalent to the desired conclusion. □

Corollary 49 can be equivalently formulated as a fact in the ring $\mathbb{Z}/p\mathbb{Z}$ (which is actually a field, the field $\mathbb{F}_p$ of p elements): if p is a prime and $[a] \neq [0]$ in $\mathbb{Z}/p\mathbb{Z}$, then $[a]^{p-1} = [1]$. Similarly, Fermat's little theorem can be equivalently formulated as follows: if p is a prime, then $[a]^p = [a]$ for all $[a] \in \mathbb{Z}/p\mathbb{Z}$.

48. Euler's theorem

Corollary 49 (of Fermat's little theorem) admits a generalisation to an arbitrary modulus $n > 1$ in place of the prime number p, involving Euler's phi function.

Theorem 50 (Euler's theorem). *If $n > 1$ is an integer, and a is any integer which is coprime with n, then $a^{\varphi(n)} \equiv 1 \pmod{n}$.*

[20] This shows that the Corollary is essentially equivalent to Fermat's little theorem, meaning that the arguments needed to obtain either one from the other are much shorter or simpler than those needed to prove either fact. The statement in the Corollary actually has a better form for generalisations, as we will see in Section 54.

When n is a prime we have $\varphi(n) = n - 1$, and so the above statement specialises to the corollary of Fermat's little theorem. We will not prove Euler's theorem in this module (although one one can prove it along the lines of our proof of Fermat's little theorem). However, we will prove one special case, where n is a product of two distinct primes p and q, deducing it from Corollary 49. Note that the integers k which are not coprime with pq are either multiples of p or multiples of q. Hence in the range $0 \leq k < pq$ there are exactly $p + q - 1$ of them, and hence

$$\varphi(pq) = pq - (p + q - 1) = (p - 1)(q - 1) \qquad [\, = \varphi(p) \cdot \varphi(q)\,].$$

The special case of Euler's theorem is as follows.

PROPOSITION 51. *If p and q are distinct primes, and a is any integer not divisible by p or q, then $a^{(p-1)(q-1)} \equiv 1 \pmod{pq}$.*

PROOF. We know from Corollary 49 that $a^{p-1} \equiv 1 \pmod{p}$. Taking the $(q-1)$-st power of each side we obtain $a^{(p-1)(q-1)} \equiv 1 \pmod{p}$. Exchanging the roles of p and q we obtain $a^{(p-1)(q-1)} \equiv 1 \pmod{q}$. Hence $a^{(p-1)(q-1)} - 1$ is both a multiple of p and a multiple of q. Because p and q are different primes it follows that $a^{(p-1)(q-1)} - 1$ is a multiple of pq, which is the desired conclusion. □

49. Some applications of Fermat's little theorem

Corollary 49 tells us that the powers of any nonzero class $[a]$ in $\mathbb{Z}/p\mathbb{Z}$ repeat every $p - 1$ steps: multiplying by $[a]$ both sides of $[a]^{p-1} = [1]$ we get $[a]^p = [a]$, which is of course Fermat's little theorem (except that the statement of Fermat's little theorem includes the trivial case of the zero class, $[0]^p = 0$), but if we keep multiplying by $[a]$ we get $[a]^{p+1} = [a]^2$, then $[a]^{p+2} = [a]^3$, etc. Eventually this way we reach $[a]^{2(p-1)} = [1]$, which could be obtained more quickly by squaring both sides of $[a]^{p-1} = [1]$, and then $[a]^{3(p-1)} = [1]$, and so on. Because of this if we need to compute a power $[a]^k$ with large k it is convenient to divide k by $p - 1$, hence write k in the form $k = (p - 1)q + r$, with $0 \leq r < p - 1$, and then we will have

$$[a]^k = [a]^{(p-1)q+r} = ([a]^{p-1})^q \cdot [a]^r = [1]^q \cdot [a]^r = [a]^r.$$

EXAMPLE. Suppose we need to compute $[2]^{12500}$ in $\mathbb{Z}/13\mathbb{Z}$. We have done similar calculations before, but because the exponent is large it would take us quite some time to actually compute that power, for example as

$$[2]^{12500} = [2]^{2^2 \cdot 5^5} = \Big(\Big(\Big(\Big(\Big(([2]^2)^2\Big)^5\Big)^5\Big)^5\Big)^5\Big)^5.$$

Another way is writing the exponent 125000 in binary and proceed as we have seen in an earlier section, turning the calculation of the power into a sequence of squarings and multiplications by [2] (which depend on the bits of the exponent). That method is more general because it does not require us to factorise the exponent 125000. Still, although very fast on a computer, it would take us quite a bit by hand.

However, here the modulus 13 is prime, and the class [2] is invertible, hence because of the corollary to Fermat's little theorem we have $2^{12} \equiv 1 \pmod{13}$, which means $[2]^{12} = [1]$ in $\mathbb{Z}/13\mathbb{Z}$. Consequently, $[2]^{24} = [1]$, $[2]^{36} = [1]$, etc., and we only have to determine the remainder of 12500 when divided by 12. A little more formally, we have

$$[2]^{12500} = [2]^{12\cdot 1041+8} = \left([2]^{12}\right)^{1041} \cdot [2]^8 = [1]^{1041} \cdot [2]^8 = [2]^8 = [256] = [-4].$$

In the last step we have computed $2^8 = 256$ directly because it was easy, but note that a more efficient method for a general situation would be $[2]^8 = [2]^{2\cdot 4} = [4]^{2\cdot 2} = [3]^2 = [-4]$.

Example (The last digit of a fifth power). For any integer a we have $a^5 \equiv a \pmod{10}$. This means that if a is any nonnegative integer written in decimal notation, then a and a^5 have the same right-most digit. Of course you may check this directly, after noting that the right-most digit of a only depends on the right-most digit of a (because carries propagate to the left):

$$0^5 = 0, \qquad 1^5 = 1, \qquad 2^5 = 32, \qquad 3^5 = 243, \qquad 4^5 = 1024, \ldots$$

(Incidentally, note that such a calculation is more efficiently done by directly working with residue classes modulo 10, for example,

$$[7]^5 = [-3]^5 = \left([-3]^2\right)^2 \cdot [-3] = [9]^2 \cdot [-3] = [-1]^2 \cdot [-3] = [1] \cdot [-3] = [-3] = [7],$$

that is, reducing modulo 10 every time we can, rather than actually computing $7^5 = 16807$.) However, it is more instructive to deduce this from Fermat's little theorem, as follows.

Because of Fermat's little theorem, we have $a^5 \equiv a \pmod 5$, and $a^2 \equiv a \pmod 2$. The latter (by iteration) implies that $a^k \equiv a \pmod 2$ for all a, for any positive integer exponent k). (In fact, this last congruence can also easily be checked by distinguishing the cases a odd or even.) Now $a^5 \equiv a \pmod 5$ and $a^5 \equiv a \pmod 2$ are together equivalent to our goal $a^5 \equiv a \pmod{10}$.

What we have shown in this example is that every integer of the form $a^5 - a$ is a multiple of 10. More is actually true, namely, every integer of the form $a^5 - a$ is a multiple of 30. In fact, to show that $a^5 - a$ is a multiple of 3 one may note that according to Corollary 49 to Fermat's little theorem we have $a^2 \equiv 1 \pmod 3$ whenever $3 \nmid a$. Squaring we get $a^4 \equiv 1 \pmod 3$, and multiplying both sides by a we get $a^5 \equiv a \pmod 3$. This last congruence is clearly true also when $3 \mid a$, and hence holds for every integer a.

Example (The last digit of a square). For completeness we note the last digit of a square a^2 (of an integer) can only be $0, 1, 4, 5, 6, 9$ (and never $2, 3, 6, 7$). Of course this can be verified directly, but it is smarter to work separately modulo 2 and modulo 5. In fact, modulo 5 we have

$$[0]^2 = [0], \qquad [\pm 1]^2 = [1], \qquad [\pm 2]^2 = [4] = [-1],$$

and so no square will ever be congruent to 2 or 3 modulo 5. However, squares can be either even or odd. Putting these two facts together (and using the Chinese remainder theorem, Theorem 45) gives the above $3 \cdot 2 = 6$ possibilities for the last digit of a square.

As a consequence, the last digit of a fourth power can only be $0, 1, 5$, or 6.

Example (The last digit of a cube). The last digit of a cube a^3 can be any digit $0, 1, \ldots, 9$ (generally different from the last digit of a, though). Again, this can be simply verified case-by-case, but the following explanation based on Fermat's little theorem is smarter (and can be generalised to other situations).

Because of Corollary 49 to Fermat's little theorem we have $a^4 \equiv 1 \pmod 5$ whenever $5 \nmid a$. But then squaring we get $a^8 \equiv 1 \pmod 5$, and consequently $a^9 \equiv a \pmod 5$. This last congruence is clearly true also when $5 \mid a$, and hence holds for every integer a. Clearly $a^9 \equiv a \pmod 2$ as well, for every integer a. Putting those together we conclude that $a^9 \equiv a \pmod{10}$ for every integer a. We can think of this as the map from $\mathbb{Z}/10\mathbb{Z}$ to itself given by $[a] \mapsto [a]^9$ being the identity.

Now, because $[a]^9 = ([a]^3)^3$, it follows that the *cubing* map from $\mathbb{Z}/10\mathbb{Z}$ to itself given by $[a] \mapsto [a]^3$ gives the identity map when composed with itself. In particular, the cubing map is invertible (and it actually equals its own inverse, but this is not important here), and consequently, it is bijective.

For example, to find an integer a whose cube ends with the digit 3, just cube this and take $a = 3^3 = 27$. Of course we might instead take $a = 7$ (because only the last digit of a matters). In fact, $7^3 = 343$, as desired.

Remark. A slightly more advanced argument, using Euler's theorem rather than just Fermat's little theorem, would show that every odd integer less than 100, and not a multiple of 5, occurs as the last two digits of some cube of an integer. If we included even integers, and also integers which are multiples of 5, we would find that a majority of the 100 possible pairs of digits, namely, exactly 63 of them, occur as the last two digits of the cube of some integer. By contrast, only 22 different pairs of digits occur as the two right-most digits of a square, namely

$$00, 01, 04, 09, 16, 21, 24, 25, 29, 36, 41, 44, 49, 56, 61, 64, 69, 76, 81, 84, 89, 96.$$

Only 12 different pairs of digits occur as the two right-most digits of a fourth power, namely

$$00, 01, 16, 21, 25, 36, 41, 56, 61, 76, 81, 96.$$

Only 15 different pairs of digits occur as the two right-most digits of a fifth power, namely

$$00, 01, 07, 24, 25, 32, 43, 49, 51, 57, 68, 75, 76, 93, 99.$$

Even fewer appear as the two right-most digits of a tenth power, namely,

$$00, 01, 24, 25, 49, 76.$$

But the record holder in this situation is twentieth powers: for any integer a, its twentieth power a^{20} ends with one of the following four pairs of digits:

$$00, 01, 25, 76.$$

In particular, if a is an even integer not ending with 0, then a^{20} ends with the digits 76. (Try this on a pocket calculator, of course with some ingenuity to avoid huge numbers.) This last fact appropriately extends to the last three digits of certain powers (and more digits if we liked, by taking appropriate exponents): for any integer a, its power a^{100} ends with one of the following four pairs of digits:

$$000, 001, 376, 625.$$

All these facts can be nicely explained in the language of congruences, with just a little more theory beyond Euler's theorem.

Lecture notes of Algebra. Week 11

50. Polynomials over the finite field $\mathbb{F}_p$

We have observed earlier that when p is a prime the ring $\mathbb{Z}/p\mathbb{Z}$ is actually a field, which we denote by $\mathbb{F}_p$. Hence the elements of $\mathbb{F}_p$ are the residue classes $[0], [1], \ldots, [p-1]$, but we may prefer to use $[-(p-1)/2], \ldots, [0], \ldots, [(p-1)/2]$ if $p > 2$.

In this section we consider polynomials with coefficients in $\mathbb{F}_p$, which are very useful in modern technological applications of Algebra. To simplify notation when writing polynomials with coefficients in $\mathbb{F}_p$ one sometimes writes the coefficients as integers a, while really meaning their class $[a]$ modulo p. Hence a polynomial in $\mathbb{F}_p[x]$ looks like a polynomial in $\mathbb{Z}[x]$, but we must not confuse the meaning and keep in mind that all calculations with coefficients must be done in $\mathbb{F}_p$ (that is, modulo p). Also, we may take any polynomial with integer coefficients and think of it as a polynomial in $\mathbb{F}_p[x]$ by *viewing its coefficients modulo p,* but its properties such as being irreducible or not will depend on the particular prime p chosen.

Recall that a polynomial $f(x) \in F[x]$ of positive degree is *reducible* in $F[x]$ if it can be written as $f(x) = g(x)\,h(x)$, where $g(x)$ and $h(x)$ are polynomials in $F[x]$ of positive degree; it is *irreducible* in $F[x]$ if it is not reducible.

Example. The polynomial x^2-1 (which really means $f(x) = [1]x^2+[0]x+[-1]$ if we want to be pedantic) is reducible in $\mathbb{F}_p[x]$, whatever the prime p is, because $x^2-1 = (x-1)(x+1)$. So in general if $f(x)$ has a proper factorisation in $\mathbb{Z}[x]$, then that factorisation will remain valid even when the coefficients are viewed modulo any prime p.

However, consider the polynomial x^2+2 can be read as having coefficients in $\mathbb{F}_3$, or in $\mathbb{F}_5$, etc. It is reducible in $\mathbb{F}_[x]$, because it is just another way of writing x^2-1, but it is irreducible in $\mathbb{F}_5[x]$. (See below how to check that.)

Recall the criterion for a polynomial of degree 2 or 3 to be irreducible in $F[x]$: it is irreducible precisely if it does not have any root in F. Because now the field $F = \mathbb{F}_p$ has a finite number of elements, we can check this with a finite amount of calculation: it will be enough to compute $f([0]), f([1]), \ldots, f([p-1])$.

Example. The polynomial $f(x) = x^2+x-1$ is irreducible in $\mathbb{F}_3[x]$, because it has no root in $\mathbb{F}_3$:

$$f([0]) = [-1], \quad f([1]) = [1], \quad f([-1]) = [-1].$$

Example. The polynomial $f(x) = x^3+x^2-x+1$ is irreducible in $\mathbb{F}_3[x]$, because it has no root in $\mathbb{F}_3$:

$$f([0]) = [1], \quad f([1]) = [1+1-1+1] = [-1], \quad f([-1]) = [-1+1+1+1] = [-1].$$

Example. The polynomial $x^2 - 2$ is irreducible in $\mathbb{F}_3[x]$ or in $\mathbb{F}_5[x]$, because there is no element $[a]$ in those fields such that $[a]^2 = [2]$ (check), but it is reducible in $\mathbb{F}_7[x]$, where $[3]^2 = [9] = [2]$, and hence $x^2 - [2] = (x - [3])(x + [3])$, or $x^2 - 2 = (x - 3)(x + 3)$ in $\mathbb{F}_7[x]$ for simpler notation.

Because for polynomials with coefficients in $\mathbb{F}_p$ there is only a finite number of possible choices for each coefficient, there is a finite number of distinct polynomials in $\mathbb{F}_p[x]$ of a given degree, so in principle one could look at all of them and check which are irreducible. In the next example we determine all (monic) irreducible quadratic polynomials in $\mathbb{F}_3[x]$ and $\mathbb{F}_5[x]$ (and learn to find those in $\mathbb{F}_p[x]$, in principle). Recall from the section on quadratic polynomials that we can tell whether a quadratic polynomial over a field F is irreducible by looking at its discriminant. For that to work we actually need that $2 \neq 0$ in the field F (recall that the formula for solving quadratic equations involves dividing by 2, and we would not want to divide by zero). Under that assumption we found that a quadratic polynomial over F has a root in F if and only if its discriminant is a square in F (and in that case one can find the roots by the usual formula). But as we recalled above a quadratic (or cubic) polynomial in $F[x]$ is irreducible precisely when it does not have any root in F. Putting these facts together we see that a quadratic polynomial in $F[x]$ is irreducible if and only if its discriminant *is not* a square in F.

Example 52. Consider $F = \mathbb{F}_p$, for some odd prime p. To apply the considerations recalled above we must exclude the case of $\mathbb{F}_2$, in which case $[2] = [0]$, and so dividing by 2 is not allowed. (Of the four monic quadratic polynomials over $\mathbb{F}_2$ precisely one is irreducible, namely $x^2 + x + 1$, because it has no roots in $\mathbb{F}_2$.) We want to find the quadratic irreducible polynomials in $\mathbb{F}_p[x]$. If a polynomial is not monic we can making it monic by dividing it by its leading coefficient (and irreducibility or not is not affected), so it is enough to look at monic quadratic polynomials.

Hence consider $x^2 + bx + c$, an arbitrary monic quadratic polynomial in $\mathbb{F}_p[x]$. It is irreducible if and only if its discriminant $b^2 - 4c$ is not a square in $\mathbb{F}_p$. We can find which elements of $\mathbb{F}_p$ are squares simply by computing $[0]^2, [\pm 1]^2, [2]^2, \ldots, [(p-1)/2]^2$, and the remaining elements of $\mathbb{F}_p$ will not be squares, by exclusion.

For example, take $p = 3$. Then there are $3^2 = 9$ polynomials of the form $x^2 + bx + c$ in $\mathbb{F}_3[x]$, and we want to find the irreducible ones. The squares in $\mathbb{F}_3$ are $[0]^2 = [0]$ and $[\pm 1]^2 = [1]$, so the only non-square is $[-1]$. Hence $x^2 + bx + c$ is irreducible in $\mathbb{F}_3[x]$ if and only if $b^2 - 4c = -1$ (in $\mathbb{F}_3$, omitting to write bracket for simplicity). Because $[4] = [1]$ this means $c = b^2 + 1$, hence the irreducible polynomials are $x^2 + bx + (b^2 + 1)$, for $b \in \mathbb{F}_3$. So they are

$$x^2 + 1, \quad x^2 + x - 1, \quad x^2 - x - 1.$$

Hence we know without checking that each of the remaining six monic quadratic polynomials in $\mathbb{F}_3[x]$ is reducible. For example $x^2 = x \cdot x$, $x^2 - 1 = (x-1)(x+1)$, $x^2 + x + 1 = (x-1)^2$, etc.

As another example, take $p = 5$. Then there are $5^2 = 25$ polynomials of the form $x^2 + bx + c$ in $\mathbb{F}_5[x]$, and we want to find the irreducible ones. The squares in $\mathbb{F}_5$ are $[0]^2 = [0]$, $[\pm 1]^2 = [1]$, and $[\pm 2]^2 = [-1]$, so the only non-squares are $[2]$ and $[-2]$. Hence $x^2 + bx + c$ is irreducible in $\mathbb{F}_5[x]$ if and only if $b^2 - 4c = \pm 2$. Because $[4] = [-1]$ this means $c = -b^2 \pm 2$ (or rather $c = -b^2 \mp 2$, but the order does not matter here), hence the irreducible polynomials are $x^2 + bx - (b^2 \pm 2)$, for $b \in \mathbb{F}_5$. So there are precisely 10 of them, and they are

$$\begin{array}{llllll} x^2+2, & x^2-2, & x^2+x+2, & x^2+x+1, & x^2+2x-1, & x^2+2x-2, \\ & & x^2-x+2, & x^2-x+1, & x^2-2x-1, & x^2-2x-2. \end{array}$$

Again, any of the other 15 monic quadratic polynomials in $\mathbb{F}_5[x]$ is reducible.

Example. The polynomial $f(x) = x^4 - x^2 + 1$ has no root in $\mathbb{F}_3$,

$$f([0]) = [1], \quad f([1]) = [1], \quad f([-1]) = [1],$$

but this does not tell us whether it is irreducible or reducible in $\mathbb{F}_3[x]$, as our criterion does not apply to polynomials of degree higher than 3. In fact, $f(x)$ is reducible, because

$$x^4 - x^2 + 1 = x^4 + 2x^2 + 1 = (x^2+1)^2 \qquad \text{in } \mathbb{F}_3[x].$$

Example. Similarly, the polynomial $f(x) = x^4 + x^3 + x - 1$ has no root in $\mathbb{F}_3$,

$$f([0]) = [-1], \quad f([1]) = [-1], \quad f([-1]) = [1],$$

but it is reducible in $\mathbb{F}_3[x]$, because

$$x^4 + x^3 + x - 1 = (x^2+1)(x^2+x-1) \qquad \text{in } \mathbb{F}_3[x].$$

Unlike the factorisation of the previous example, this factorisation was not easy to guess, and there is no really easy and fast method to find it.

51. Polynomials with integer coefficients, viewed modulo a prime

We have agreed that for simplicity the elements of a field $\mathbb{F}_p$ can be written as integers (that is, without square brackets), as long as we make clear that they are actually classes in $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$. But we can also take any polynomial with integer coefficients and the view the coefficients modulo a fixed prime p. For example, $2x^4 + 3x^3 - 10x^2 + 7x - 11$ has integer coefficients (so it is in $\mathbb{Z}[x]$), but

- viewing the coefficients modulo 2 we get
$[2]x^4 + [3]x^3 - [10]x^2 + [7]x - [11] = x^3 + x + [1] \in \mathbb{F}_2[x]$,

- viewing the coefficients modulo 3 we get
  $[2]x^4 + [3]x^3 - [10]x^2 + [7]x - [11] = -x^4 - x + [1] \in \mathbb{F}_3[x]$
- viewing the coefficients modulo 5 we get
  $[2]x^4 + [3]x^3 - [10]x^2 + [7]x - [11] = [2]x^4 - [2]x^3 + [2]x - [1] \in \mathbb{F}_5[x]$,

and so on.

The properties of the resulting polynomial can clearly change a lot depending on the prime which we use (even the degree, see example above). We illustrate what can happen on an extended example. (We will now stop writing square brackets to denote classes modulo p, we will just write the coefficients as integers and specify when they are meant to be viewed as elements of $\mathbb{F}_p$.)

Example. Consider the polynomials

$$f(x) = x^3 - 2x^2 + 3x - 6, \quad \text{and} \quad g(x) = x^3 + 3x^2 + x + 3.$$

They have integer coefficients, but as mentioned above we can view them as polynomials in $\mathbb{F}_p[x]$, where p is any prime, by viewing their coefficients modulo p (that is, interpreting 3 as $[3]_p$, etc.). We will study them

(a) as polynomials in $\mathbb{Q}[x]$;
(b) viewed as polynomials in $\mathbb{F}_2[x]$;
(c) viewed as polynomials in $\mathbb{F}_3[x]$;
(d) viewed as polynomials in $\mathbb{F}_5[x]$.

We first compute the greatest common divisors of $f(x)$ and $g(x)$ by applying the Euclidean algorithm in the various cases.

(a) Viewing them as polynomials in $\mathbb{Q}[x]$ we find

$$\begin{aligned} x^3 - 2x^2 + 3x - 6 &= (x^3 + 3x^2 + x + 3) \cdot 1 + (-5x^2 + 2x - 9) \\ x^3 + 3x^2 + x + 3 &= (5x^2 - 2x + 9) \cdot \Big(\frac{1}{5}\,x + \frac{17}{25}\Big) + \Big(\frac{14\,x}{25} - \frac{78}{25}\Big) \\ 5x^2 - 2x + 9 &= (7x - 39)\Big(\frac{5}{7}\,x + \frac{181}{49}\Big) + \frac{7500}{49} \end{aligned}$$

This shows that the two polynomials are coprime over $\mathbb{Q}$. Note that even though $f(x)$ and $g(x)$ have quite small integers as coefficients, the calculations have produced rather large fractions. The phenomenon we are experiencing here is *coefficient explosion,* which may be a problem when working over $\mathbb{Q}$. This cannot occur when working over $\mathbb{F}_p$, as we will see below, because the coefficient can only take p possible values.

(b) The beginning of the Euclidean algorithm over $\mathbb{Q}$ can be read modulo 2 (as long as no denominators appear), but it finishes earlier:

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

Hence the GCD over $\mathbb{F}_2$ is $x^2 + 1$. Alternatively, in this case the two polynomials viewed modulo 2 would have been easy to factorise directly.

(c) Here, too, the two polynomials viewed modulo 3 can be factorised at once, $f(x) = x^3 + x^2 = x^2(x + 1)$ and $g(x) = x^3 + x = x(x^2 + 1)$ in $\mathbb{F}_3[x]$, and so their GCD is x.

(d) Working modulo 5, the first division of the Euclidean algorithm over $\mathbb{Q}$ reads

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

but then it is convenient to divide the remainder by 2, which means multiplying it by its inverse -2, and then the next division will read

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

showing that the GCD over $\mathbb{F}_5$ is $x - 2$.

Now we take another look at this particular example using the fact that the factorisations of $f(x)$ and $g(x)$ in $\mathbb{Q}[x]$ are easy to spot in this case (or can be found by the Rational Root Test), namely,

$$f(x) = (x - 2)(x^2 + 3), \qquad \text{and} \qquad g(x) = (x + 3)(x^2 + 1),$$

where all factors shown are irreducible over $\mathbb{Q}$. It follows at once that the polynomials are coprime over $\mathbb{Q}$, that is, their GCD is 1. Because the factorisations found for $f(x)$ and $g(x)$ are actually over $\mathbb{Z}$, they remain valid when we view them as congruences modulo any prime p, except that

- some factor of $f(x)$ may be congruent to some factor of $g(x)$; in fact, modulo 2 this is the case for the factor $x^2 + 3$ of $f(x)$ and the factor $x^2 + 1$ of $g(x)$;
- the quadratic factors, which are irreducible over $\mathbb{Q}$, may become reducible when viewed modulo p; for example, the factor $x^2 + 3$ of $f(x)$ is reducible modulo 2, and also modulo 3;
- some of the further linear factors obtained for $f(x)$ modulo p may then be a factor of $g(x)$, and conversely.

So we now look at the factorisations of $f(x)$ and $g(x)$ over F_2, $\mathbb{F}_3$, and $\mathbb{F}_5$, to understand their GCD's which we found earlier in the various cases.

(b) Reading modulo 2 the factorisations found above, and possibly factorising further, we find $f(x) = x(x^2+1) = x(x+1)^2$ and $g(x) = (x+1)(x^2+1) = (x+1)^3$ in $\mathbb{F}_2[x]$. Hence their GCD over $\mathbb{F}_2$ is $(x+1)^2 = x^2+1$ (as we found earlier).

(c) Reading modulo 3 the factorisations found above, and possibly factorising further, we find $f(x) = (x+1)x^2$ and $g(x) = x(x^2+1)$, in $\mathbb{F}_3[x]$. Here x^2+1 remains irreducible over $\mathbb{F}_3$, and so the GCD of $f(x)$ and $g(x)$ over $\mathbb{F}_3$ is x.

(d) Reading modulo 5 the factorisations found above, and possibly factorising further, we find $f(x) = (x-2)(x^2-2)$ and $g(x) = (x-2)(x^2-4) = (x-2)^2(x+2)$, in $\mathbb{F}_3[x]$. Here x^2+2 remains irreducible over $\mathbb{F}_5$, and so the GCD of $f(x)$ and $g(x)$ over $\mathbb{F}_5$ is $x-2$.

52. (Optional) Congruences with polynomials

One may also consider congruences with polynomials, with coefficients in a field, and most of what we have seen about congruences with integers in previous sections works in a similar way with polynomials. The definition of congruence with polynomials will be entirely similar. A system of congruences

$$\begin{cases} f(x) \equiv a(x) & \pmod{m(x)} \\ f(x) \equiv b(x) & \pmod{n(x)}, \end{cases}$$

where $f(x)$ is an unknown polynomials, and all the others are given polynomials, will have solutions exactly when the GCD $\big(a(x), b(x)\big)$ divides the difference $b(x) - a(x)$. In that case, one particular solution $f_0(x)$ will be found through the extended Euclidean algorithm, and the general solution will be $f(x) \equiv f_0(x) \pmod{[m(x), n(x)]}$ (or $f(x) \equiv f_0(x) + k(x) \cdot [m(x), n(x)]$ if you prefer, where $k(x)$ is an arbitrary *polynomial*). We have a Chinese remainder theorem for polynomials, and the congruence $a(x)f(x) \equiv b(x) \pmod{n(x)}$ with polynomials can be solved in a similar way as with integers.

Instead of going through all details we present an important example. Consider the congruence classes of polynomials in $\mathbb{R}[x]$ modulo x^2+1. They form the set

$$C = \{[a+bx] : a, b \in \mathbb{R}\},$$

where $[a+bx]$ denotes the congruence class $[a+bx]_{x^2+1} = \{a + bx + k(x) \cdot (x^2+1) : k(x) \in \mathbb{R}[x]\}$. In fact, any congruence class modulo x^2+1 has a *unique* representative of the form $ax+b$, because this is the remainder of dividing *any* representative by x^2+1. Now we can define addition and multiplication of congruence classes as we did to define $\mathbb{Z}/n\mathbb{Z}$. Hence the sum of two classes is given by

$$[a+bx] + [c+dx] = [(a+c) + (b+d)x].$$

Their product is

$$[a+bx]\cdot[c+dx]=[ac+(ad+bc)x+bdx^2]$$

which after taking the remainder of dividing $ac+(ad+bc)x+bdx^2$ by x^2+1 reads

$$[a+bx]\cdot[c+dx]=[ac-bd+(ad+bc)x].$$

This is the multiplication rule for complex numbers, where the role of the imaginary unit i is played by $[x]$, because $[x]^2=[-1]$. This is how the complex numbers are constructed from the real numbers in an algebraic way, and is a model for a much more general, important algebraic construction.

53. (Optional) Rationalisation of denominators

To rationalise a denominator means the following: if we have an expression which is a fraction, and the denominator involves taking some roots (square, or cube, etc.), transform the expression into an equivalent one where no roots appears in the denominator. This is typically done by multiplying denominator and numerator by a suitable expression (which must be the same), as in

$$\frac{1}{3+\sqrt{2}}=\frac{1}{3+\sqrt{2}}\cdot\frac{3-\sqrt{2}}{3-\sqrt{2}}=\frac{3-\sqrt{2}}{(3+\sqrt{2})(3-\sqrt{2})}=\frac{3-\sqrt{2}}{3^2-(\sqrt{2})^2}=\frac{3-\sqrt{2}}{7}.$$

Note also that we may also think of 3 as $\sqrt{9}$ in this calculation, and so the procedure would have been very similar if we had to rationalise the denominator in $1/(\sqrt{5}+\sqrt{2})$, for example. More generally, to rationalise a denominator of the form $\sqrt{a}+\sqrt{b}$, multiply that, as well as the numerator, by $\sqrt{a}-\sqrt{b}$, after which the new denominator will be $(\sqrt{a}+\sqrt{b})(\sqrt{a}-\sqrt{b})=a-b$, without roots signs, which will now instead appear in the numerator. Similarly, to rationalise a denominator of the form $\sqrt{a}-\sqrt{b}$, multiply that, as well as the numerator, by $\sqrt{a}+\sqrt{b}$. And of course the numerator might be something different from 1: calculations could be more complicated but there is no conceptual extra difficulty, so we take the numerator to be 1 here for simplicity.

Note that this procedure is very similar as to when we want to compute the reciprocal of a complex number $a+bi$ (where we multiply by $a-bi$), just think of the complex number as $a+b\sqrt{-1}$, or $a+\sqrt{-b^2}$, or $\sqrt{a^2}+\sqrt{-b^2}$.

If the denominator is a sum of two cube roots, $\sqrt[3]{a}+\sqrt[3]{b}$, then multiply that and the numerator by $\sqrt[3]{a^2}-\sqrt[3]{ab}+\sqrt[3]{b^2}$, and similarly, if the denominator is $\sqrt[3]{a}-\sqrt[3]{b}$, then multiply that and the numerator by $\sqrt[3]{a^2}+\sqrt[3]{ab}+\sqrt[3]{b^2}$.

Example.

$$\begin{aligned}\frac{1}{2\sqrt[3]{3}+\sqrt[3]{5}} &= \frac{1}{2\sqrt[3]{3}+\sqrt[3]{5}}\cdot\frac{4\sqrt[3]{9}-2\sqrt[3]{15}+\sqrt[3]{25}}{4\sqrt[3]{9}-2\sqrt[3]{15}+\sqrt[3]{25}}\\ &= \frac{4\sqrt[3]{9}-2\sqrt[3]{15}+\sqrt[3]{25}}{(2\sqrt[3]{3})^3+(\sqrt[3]{5})^3} = \frac{4\sqrt[3]{9}-2\sqrt[3]{15}+\sqrt[3]{25}}{29}.\end{aligned}$$

But what if the denominator is more complicated, such as $\sqrt[3]{2}^2 - 3\sqrt[3]{2}+3$, which can also be written as $\sqrt[3]{4}-3\sqrt[3]{2}+3$? The idea is to note that $\sqrt[3]{2}$ is a root of the polynomial x^3-2, and then we apply a method based on the extended Euclidean algorithm, which we explain on examples. Before solving this particular example we start with a simpler one. this example.

Example. Rationalise the denominator of $\dfrac{1}{\sqrt[3]{5}^2-\sqrt[3]{5}+1}$.

This case can be easily dealt with directly by multiplying denominator and numerator by $\sqrt[3]{5}+1$, so the new denominator will become $(\sqrt[3]{5}^2-\sqrt[3]{5}+1)(\sqrt[3]{5}+1) = \sqrt[3]{5}^3-1 = 5-1=4$. However, we pretend not to see that and use a more general method.

We are asked to find the multiplicative inverse (that is, the reciprocal) of $\alpha^2-\alpha+1$, where $\alpha=\sqrt[3]{5}$ satisfies $\alpha^3-5=0$. It is very similar in spirit to asking for the inverse of an integer a modulo another integer n, hence *knowing that* $n\equiv 0$. As we know that can be done if $(a,n)=1$, and the method starts with the extended Euclidean algorithm on n and a. So we proceed similarly but working with polynomials instead, and perform the extended Euclidean on the polynomial x^3-5 (of which α is a root) and the polynomial x^2-x+1 (of which we want to compute the reciprocal of the value taken for $x=\alpha$):

$$x^3-5=(x^2-x+1)(x+1)-4,$$

and hence the conclusion of Bézout's Lemma, very simple to obtain in this case, is

$$4=-(x^3-5)+(x^2-x+1)(x+1)$$

(or rather this divided by 4). If we substitute $x=\alpha$ in this equality we have $\alpha^3-5=0$, and hence we are left with

$$4=(\alpha^2-\alpha+1)(\alpha+1).$$

(In practice, this is the same as viewing the conclusion of Bézout's Lemma modulo x^3-5.) But then

$$\frac{1}{\sqrt[3]{5}^2-\sqrt[3]{5}+1}=\frac{1}{\alpha^2-\alpha+1}=\frac{\alpha+1}{4}=\frac{\sqrt[3]{5}+1}{4},$$

which is what we wanted.

Example. Rationalise the denominator of $\dfrac{1}{\sqrt[3]{2}^2-3\sqrt[3]{2}+3}$.

In this case we set $\alpha = \sqrt[3]{2}$. We are asked to find the multiplicative inverse (that is, the reciprocal) of $\alpha^2 - 3\alpha + 3$, where α satisfies $\alpha^3 - 2 = 0$. We perform the extended Euclidean on the polynomial $x^3 - 2$ and the polynomial $x^2 - 3x + 3$:

$$\begin{aligned} x^3 - 2 &= (x^2 - 3x + 3)(x + 3) + (6x - 11) \\ x^2 - 3x + 3 &= (6x - 11)\left(\frac{1}{6}x - \frac{7}{36}\right) + \frac{31}{36} \end{aligned}$$

and then

$$\begin{aligned} 31 &= 36(x^2 - 3x + 3) - (6x - 11)(6x - 7) \\ &= 36(x^2 - 3x + 3) - [(x^3 - 2) - (x^2 - 3x + 3)(x + 3)](6x - 7) \\ &= -(x^3 - 2)(6x - 7) + (x^2 - 3x + 3)[(x + 3)(6x - 7) + 36] \\ &= -(x^3 - 2)(6x - 7) + (x^2 - 3x + 3)(6x^2 + 11x + 15). \end{aligned}$$

If we substitute $x = \alpha$ in this expression we have $\alpha^3 - 2 = 0$, and hence

$$31 = (\alpha^2 - 3\alpha + 3)(6\alpha^2 + 11\alpha + 15).$$

But then

$$\frac{1}{\sqrt[3]{2}^2 - 3\sqrt[3]{2} + 3} = \frac{1}{\alpha^2 - 3\alpha + 3} = \frac{6\alpha^2 + 11\alpha + 15}{31} = \frac{6\sqrt[3]{2}^2 + 11\sqrt[3]{2} + 15}{31},$$

which is precisely what we wanted.

Lecture notes of Algebra. Week 15

54. Cryptography

A *cryptosystem* can generically be described as

$$\mathcal{P} \xrightarrow{f} \mathcal{C} \xrightarrow{f^{-1}} \mathcal{P},$$

where

- $\mathcal{P} =$ the set of all possible *elementary* plain messages,
- $\mathcal{C} =$ set of the possible encrypted messages,
- $f =$ encryption function,
- $f^{-1} =$ its inverse, the decryption function

For example, $\mathcal{P}$ might be the set of letters of the English alphabet, perhaps including punctuation, or a set of numbers corresponding to them, or the elements of a *group* or some other algebraic structure. To use larger sets, $\mathcal{P}$ might be the set of pairs of letters, or the set of blocks of k letters. The way in which numbers are attached to letters or blocks of letters need not be secret: only the function f must be secret, which is usually done by choosing f and f^{-1} within a class of functions depending on a parameter (or a set of parameters) K, the *encryption key* to be kept secret. To f^{-1} there will correspond a *decryption key.*

Example (Caesar's shift). Here is a very old cryptosystem, used since the time of the Roman empire. Let both $\mathcal{P}$ and $\mathcal{C}$ be the set of 26 letters of the English alphabet, listed in the standard order, let f the function which translates (that is, shifts) the letters in some direction by a number of places (the key), wrapping around at either end:

$$\begin{array}{lcccccc} \mathcal{P}: & A & B & C & D & \dots & Z \\ \mathcal{C}: & D & E & F & G & \dots & C \end{array}$$

In order to generalise this example we first describe it in a mathematical way. First we number the letters, starting from 1, or perhaps better from zero. Hence to each letter there corresponds a nonnegative number less than 26,

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

There is nothing secret up to this point. We may generalise this example by taking $\mathcal{P} = \mathcal{C} = \mathbb{Z}/N\mathbb{Z}$ and $f(P) = P + b$, where $P \in \mathcal{P}$ and b is an arbitrary element of $\mathbb{Z}/N\mathbb{Z}$. Here f is bijective and $f^{-1}(C) = C - b$ for all $C \in \mathcal{C}$. As an example, take $N = 26$ and $b = 3$. We want to encrypt the word YES, hence to Y there corresponds the number 24, to E the number 4, and to S the number 18. First we associate to our message YES the numeric string (or sequence) $24, 4, 18$, where those integers are really to be read modulo 26, that is, they must be interpreted as elements of $\mathbb{Z}/26\mathbb{Z}$. Now

$$\begin{aligned} f([24]) &= [24] + [3] = [1] \\ f([4]) &= [4] + [3] = [7] \\ f([18]) &= [18] + [3] = [21] \end{aligned}$$

Hence we send $1, 7, 21$, or possibly its translation into a strings of letters, BHV. The recipient of the message will then apply f^{-1} to recover the plain message:

$$\begin{aligned} f^{-1}([1]) &= [1] - [3] = [24] \\ f^{-1}([7]) &= [7] - [3] = [4] \\ f^{-1}([21]) &= [21] - [3] = [18] \end{aligned}$$

Example (Affine maps). This generalises the previous method. Let $\mathcal{P} = \mathcal{C} = \mathbb{Z}/N\mathbb{Z}$ and let f be defined by $f(x) = [a] \cdot x + [b]$ for $x \in \mathcal{P}$, where $[a], [b] \in \mathbb{Z}/N\mathbb{Z}$, with $(a, N) = 1$ in order for f to be invertible. Of course this can also be written as $f(x) \equiv ax + b \pmod{N}$ for $x \in \{0, 1 \dots, N - 1\}$ if we prefer. Such a map f is called an *invertible affine map* on $\mathbb{Z}/N\mathbb{Z}$. The inverse of f is given by $f^{-1}(y) = [a]^{-1}y - [a]^{-1}[b]$ in $\mathbb{Z}/N\mathbb{Z}$, or $f^{-1}(y) = a^{-1}y - a^{-1}b \pmod{N}$ if we prefer, where a^{-1} denotes any integer which is an inverse of a modulo N (that is, such that $[a^{-1}] = [a]^{-1}$).

Example. Let $N = 26$, $a = 7$, and $b = 16$, which are admissible because $(7, 26) = 1$. Again we want to send the message YES, and we start with associating to it the numeric string $24, 4, 18$. Because $f(x) = 7x + 16 \pmod{26}$ we find

$$\begin{aligned} f([24]) &= [7] \cdot [24] + [16] = [2] \\ f([4]) &= [7] \cdot [4] + [16] = [18] \\ f([18]) &= [7] \cdot [18] + [16] = [12], \end{aligned}$$

and hence we send $2, 18, 12$, or its equivalent string of letters CSM.

To decrypt the message, the recipient uses the decrypting function

$$f^{-1}(y) = a^{-1}y - a^{-1}b = a^{-1}(y - b) = [15]y - [15] \cdot [16] = [15]y - [6],$$

which can also be, perhaps less conveniently, written $f^{-1}(y) = [15]y + [20]$, and finds

$$\begin{aligned} f^{-1}([2]) &= [15] \cdot [2] - [6] = [24] \\ f^{-1}([18]) &= [15] \cdot [18] - [6] = [4] \\ f^{-1}([12]) &= [15] \cdot [12] - [6] = [18] \end{aligned}$$

which corresponds to the plain message YES.

Unfortunately, this cryptosystem based on affine maps is very vulnerable to a type of attack called *frequency analysis*. In *cryptanalysis* one usually assumes that the person who aims at *breaking* the system, that is, to find a way to decrypt messages without being the authorised recipient, knows the general structure of the cryptosystem, and hence that the only information to be discovered is the secret key. Frequency analysis is based on the following idea: if a sufficiently long message has been intercepted, some letters will appear with a higher frequency than others, and one may assume that the relative frequency with which a letter appears in the plain message is close to the relative frequency with which it appears, on average, in the English language (assuming that the plain text is in English). Such frequencies (expressed in percentages) are approximately as follows:

letter	E	T	A	O	I	N	S	R	H	D	L	U	C
frequency	12.02	9.10	8.12	7.68	7.31	6.95	6.28	6.02	5.92	4.32	3.98	2.88	2.71

letter	M	F	Y	W	G	P	B	V	K	X	Q	J	Z
frequency	2.61	2.30	2.11	2.09	2.03	1.82	1.49	1.11	0.69	0.17	0.11	0.10	0.07

One may then assume that the most frequent letter in the plaintext, and hence hopefully in the English language if the text is long enough (say thousands of letters or more), gets encrypted into the letter which appears most frequently in the ciphertext. After this first guess one may make a second guess about the second most frequent letter, etc. After a few guesses the probability of guessing wrong will quickly become too high, but with the cryptosystem just described guessing how two letters are encrypted is (generally) sufficient to discover the key, as we concretely see in the next Example.

Example (Frequency analysis). Suppose we have a long string of ciphertext (say made of thousands of letters), in the same 26-letter alphabet used above. We know that the plain text was enciphered (or encrypted) using an affine transformation $x \mapsto y \equiv ax+b \pmod{26}$ (after translating it into numbers, and then translating the result back into letters). Suppose we count how many times each letter appears in the ciphertext, and we find that the most frequently occurring letter of ciphertext is U, and the second most frequently occurring letter is N. It is then reasonable to assume that these are the encryptions of E and T, respectively, which are likely to be the first and second most frequent letters in a sufficiently long English language text.

So we guess that the encryption affine function $x \mapsto [a]x + [b]$ maps $[4]$, which corresponds to the letter E, to $[20] = [-6]$, which corresponds to the letter U. Similarly, it maps $[19] = [-7]$, which corresponds to the the letter T, to $[13]$, which corresponds to the letter N. Hence we should have

$$\begin{cases} [a] \cdot [4] + [b] = [-6] \\ [a] \cdot [-7] + [b] = [13] \end{cases}$$

This is a system of two linear equations in the indeterminates $[a]$, $[b]$ in $\mathbb{Z}/26/\mathbb{Z}$ or, if we prefer, the system of congruences

$$\begin{cases} 4a + b \equiv -6 \pmod{26} \\ -7a + b \equiv 13 \pmod{26}, \end{cases}$$

the difference being just a matter of notation. By taking the difference of the two equations we find $11a \equiv 7 \pmod{26}$. Luckily (as in some cases this may not work and we may need to guess a third letter), 11 is invertible modulo 26, with inverse -7 (because the extended Euclidean algorithm gives $1 = 26 \cdot 3 - 11 \cdot 7$). Multiplying both sides of the congruence by -7 we get $a \equiv 3 \pmod{26}$. Substituting this into either congruence of the system we find $b \equiv 8 \pmod{26}$. In conclusion, the encrypting transformation is $x \mapsto y \equiv 3x + 8 \pmod{26}$.

To find the decrypting transformation, which is its inverse, we need to solve for x the congruence $y \equiv 3x + 8 \pmod{26}$. Because an inverse of 3 modulo 26 is 9 one finds that the decrypting transformation is $x \equiv 9(y - 8) \equiv 9y + 6 \pmod{26}$. (Of course we could have found the decrypting transformation directly without finding the encrypting one first, if we wanted.)

Example (Power maps). Unlike the affine maps, which are still essentially used as components (but not just the only one) of real-life cryptosystems, the following is a rather academic example, which offers no practical advantage over affine maps, but introduces an idea which will be useful in the next sections.

Let $\mathcal{P} = \mathcal{C} = \mathbb{Z}/N\mathbb{Z}$ and let f be defined by $f(x) = x^e$ for $x \in \mathcal{P}$, where a is some positive integer, chosen in such a way that f is invertible. This may be developed for arbitrary N, but assume here for simplicity that N is prime. Suppose that $(e, N-1) = 1$, let d be a positive inverse of e modulo $N - 1$, and consider the analogous function g, defined by $g(x) = x^d$ for $x \in \mathcal{P}$. Then we claim that $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ is invertible, and $g : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ is its inverse.

To prove that we first note that both $g \circ f$ and $f \circ g$ are given by $(g \circ f)(x) = (x^e)^d = x^{ed}$, and so we only need to show that $x^{ed} = x$ for all $x \in \mathbb{Z}/N\mathbb{Z}$. This is certainly true for $x = [0]$, so we assume $x \neq [0]$. Recall that according to the corollary of Fermat's little theorem (Theorem 48) we have $x^{N-1} = [1]$ for all $x \in \mathbb{Z}/N\mathbb{Z}$ with $x \neq [0]$. Now because e and d are inverses of each other modulo $N - 1$, we have $ed = 1 + k(N - 1)$ for some integer k, which is positive because so are e and d. Therefore, we have

$$x^{ed} = x^{(N-1)k+1} = (x^{N-1})^k \cdot x^1 = [1]^k \cdot x = [1] \cdot x = x,$$

as we wanted to show.

55. Public key cryptography

The cryptosystems described in an earlier subsection are *secret-key cryptosystems.* Both the encryption and the decryption keys must be kept secret, because one can be easily computed from the other.

A *public-key cryptosystem* involves a *public key* (known to everyone) and a *private key* (known to a single person, the intended recipient of the messages). The crucial requirement is that computing f^{-1} from knowing only f is unfeasible because it would take too long.

public key (for encryption) f known to everyone	private key (for decryption) f^{-1} known only to the recipient

One of the advantages of public-key cryptosystems is that it is not necessary that the two parties have a secure communication beforehand in order to exchange keys (or agree on keys). They work as follows (the fictional characters Alice and Bob are traditional in this area):

- Bob wants to send a message to Alice;
- Alice knows both f and f^{-1};
- Alice sends f to Bob, through a reliable but public channel; [21]
- Bob sends a message to Alice, again through a public channel, after encrypting it with f;
- Alice decrypts the message using f^{-1} (and only she can do that).

In order to implement a public-key cryptosystem one needs a *trapdoor one-way function* f: besides f being bijective, and hence having an inverse f^{-1}, it must be very difficult for anyone to discover how to compute f^{-1} knowing only how to compute f, except for the legitimate receiver, who possesses an additional piece of information which allows to do that (the *trapdoor*).

In the RSA public-key cryptosystem which we describe in the next subsection, the trapdoor which allows to recover f^{-1} from f is the knowledge of the factorisation of the product of two very large primes.

56. RSA public-key cryptography

We describe the most well-known public-key cryptosystem, due to Rivest, Shamir, and Adleman (1978), whence the acronym RSA. Here is how the recipient of the messages, Alice, chooses her public key and her private key.

[21] This means that f will become known to everyone. That the channel is reliable is needed to avoid other types of risks; however, if f is modified in the transmission then it will simply not work.

- Alice chooses two distinct very large primes p_A and q_A, say of 300 decimal digits each, which means about 1000 bits each. [22]
- Alice computes $n_A = p_A \cdot q_A$ and $\varphi(n_A) = \varphi(p_A) \cdot \varphi(q_A)$.
- Alice chooses a positive integer e_A, such that $(e_A, \varphi(n_A)) = 1$.
- She computes a (positive) inverse d_A of e_A modulo $\varphi(n_A)$, which means $e_A d_A \equiv 1 \pmod{\varphi(n_A)}$.

Now Alice's public key (for encrypting) is (n_A, e_A), and her private key (for decrypting) is (n_A, d_A). Of course n_A is known to everyone as it is part of the public key, and so the only private part is d_A. Now d_A is very hard to obtain if one only knows the public key (n_A, e_A); in fact, it is believed to be as hard as factorising n_A, see below in this subsection.

Hence the situation is as follows:

public information	private information
n_A, e_A	p_A, q_A, d_A

However, Alice may as well well forget the prime factors p_A and q_A of n_A, which she only used to compute d_A but will not be used for encrypting or decrypting.

Now we explain how Bob can send a message to Alice. Because only Alice's public and private keys are involved, let us simplify the notation and write n for n_A, e for e_A, etc.

- Let P the elementary message unit that Bob intends to send Alice; it must be an element of $\mathbb{Z}/n\mathbb{Z}$, but we may also identify it as a nonnegative integer less than n;
- Bob, who knows the public key (n_A, e_A) of Alice, computes $C \equiv P^e \pmod{n}$;
- Bob sends C to Alice.

Then Alice decrypts Bob's message as follows.

- Alice receives C;
- Alice, who is the only person knowing d, computes $C^d = P^{ed} \equiv P \pmod{n}$ (see below for a proof).

Summarising, the encryption function f_A (relative to Alice, that is, used to send messages to Alice) is

$$f_A(P) = P^{e_A} \pmod{n_A}$$

[22] In modern practical implementations (as of 2015) the product $p_A \cdot q_A$, called an RSA key, has between 1024 and 4096 binary digits, and the current recommendation is that it should have at least 2048 bits. A challenge RSA key of 768 bits, the largest so far, was factorised in 2010 and took 1500 CPU years (two years of real time, on many computers). Although factorisation is so hard, producing *industrial-grade* primes of this size (meaning that there is a negligible probability that they are not primes), or even much larger ones, is not a problem.

and its inverse, the decryption function f_A^{-1}, is

$$f_A^{-1}(C) = C^{d_A} \pmod{n_A}$$

Hence encryption is done by raising to the exponent e_A modulo n_A, and decryption is done by raising to the exponent d_A modulo n_A. Equivalently, we may say that $\mathcal{P} = \mathcal{C} = \mathbb{Z}/n_A\mathbb{Z}$, and that encryption and decryption are taking the powers to the exponents e_A and d_A in the ring $\mathbb{Z}/n_A\mathbb{Z}$.

Now we prove that $P^{ed} \equiv P \pmod{n}$. First of all, $ed \equiv 1 \pmod{\varphi(n)}$, which means $ed = 1 + k \cdot \varphi(n)$ for some integer k, necessarily positive. Under the additional assumption $(P, n) = 1$ then using Euler's theorem (or Proposition 51, a special case of Euler's theorem which we have proved) we have

$$P^{ed} = P^{1+k\cdot\varphi(n)} = P \cdot (P^{\varphi(n)})^k \equiv P \cdot 1^k = P \pmod{n},$$

as desired. Now, our additional assumption $(P, n) = 1$ will be satisfied for most messages (as only $p + q - 1$ of the possible pq messages do not satisfy it, which is a tiny fraction of all messages when p and q are very large). However, even when it is not satisfied one can still prove that $P^{ed} \equiv P \pmod{n}$, and hence any encrypted message will be correctly decrypted. [23]

We claimed that the private key d_A is very hard to compute if one only knows the public key (n_a, e_a). One rather simplistic argument is that d could be computed from e if one also knew $\varphi(n)$, by means of the extended Euclidean algorithm (which is how Alice probably computed it). However, $\varphi(n)$ must be very hard to compute from the public key, because if we knew $n = n_A = pq$ and $\varphi(n) = (p-1)(q-1) = pq-p-q+1 = n+1-(p+q)$ then we would know both the product and the sum of p and q, and hence it would be easy to find them by solving a quadratic equation: p and q are the two roots of the polynomial $x^2 - \big(n + 1 - \varphi(n)\big)x + n$. But factorising n into p times q is believed to be a hard problem. The problem with this simple argument is that in principle there may be ways of computing d_A from the public key (n_A, e_A) without actually computing $\varphi(n_A)$ and hence factorising n_A. However, more convincing arguments for security exist.

Remark. We have seen a greatly simplified description of the RSA cryptosystem. In a real-life implementation one has to avoid many pitfalls which may make it insecure. To

[23] Here is a proof. If $(P, n) > 1$ then P is a multiple of either p or q. If it is a multiple of both then $P \equiv 0 \pmod{n}$ and hence also $P^{ed} \equiv 0 \pmod{n}$. It remains to see the cases where P is a multiple of either p or q, but not both. Suppose that P is a multiple of p but not of q (the other case being similar). Then P^{ed} is also a multiple of p, and hence $P^{ed} \equiv P \pmod{p}$. On the other hand, P is not a multiple of q, hence it is invertible modulo q, and so $P^{q-1} \equiv 1 \pmod{q}$. Now because $ed-1 = k\cdot\varphi(n) = k\cdot(p-1)(q-1)$ is a multiple of $q - 1$ we have $P^{ed-1} \equiv 1 \pmod{q}$, and hence $P^{ed} \equiv P \pmod{q}$. In conclusion, P^{ed} is congruent to P both modulo p and modulo q, which are distinct primes, and hence the congruence holds modulo their product $pq = n$, that is, $P^{ed} \equiv P \pmod{n}$, as desired.

start with, the chosen primes p and q should satisfy several criteria, the easiest of which are the following:

- p and q cannot be too close to each other (for example, one should be at least a couple of digits larger than the other), otherwise pq might be easy to factorise by *Fermat factorisation,* see a later subsection;
- each of $p-1$ and $q-1$ should have at least one large prime divisor (otherwise factorising pq might become easier than average, for different reasons, using the so-called *Pollard* $p-1$ *method*).

The encryption and decryption exponents e and f must also be chosen carefully. In particular, neither of them should be too small (as one may be tempted to choose in order to speed up either encryption or decryption, for example when done by a device of small computational power, such as a *smart card*).

57. (Optional) Digital signatures

One important issue in public-key cryptography is *authentication.* When Bob sends an encrypted message to Alice, how can Alice know that the sender is really Bob? After all, anyone knows Alice's encrypting function f_A, which in the case of the RSA cryptosystem means they know Alice's public key (n_A, e_A), and so anyone can send a message to Alice pretending to be Bob. We are assuming the communication channel is insecure, which is the very reason why we need cryptography. Note that this issue would not arise in classical, secret-key cryptography, because in that case Alice and Bob would have agreed on a secret key (or pair of matching keys) beforehand, which no one else knows.

The solution is that Bob can use his own *private* key (not his public key) (n_B, d_B), that is, his *decryption* function f_B^{-1}, to produce a signature and send it to Alice. This is done by applying f_B^{-1} to something recognisable by Alice, call it S, which may include an identification number, and possibly a time stamp (date and time of the day, say) to make it change every time while remaining recognisable by Alice. Hence Bob computes $f_B^{-1}(S)$ and sends the result to Alice. Alice then applies f_B to it, which is public, and after she recognises $f_B\big(f_B^{-1}(S)\big) = S$ she will know that the sender is really Bob, because only he could have possibly have applied f^{-1} to S. Recall that we assumed that public keys are communicated through a possibly insecure, but reliable channel, so that the identity of who hands out a public key, in this case Bob, is certain.

This description is a little simplified, and in a real communication Bob would also encrypt $f_B^{-1}(S)$ with Alice's public key before sending it, and actually probably append $f_B^{-1}(S)$ to the message he wants to send to Alice before encrypting. If we ignore the message here for simplicity, Bob would compute and send $f_A\big(f_B^{-1}(S)\big)$ to Alice, who would first decrypt it by applying f_A^{-1}, and then check Bob's signature by applying f_B.

Digital signatures are commonly employed in *smart cards,* which are cards containing a small chip, for example ATM cards. In this case the role of Bob is taken by the smart card, and the role of Alice by the ATM. The ATM tells the smart card some message S, the smart card (Bob) returns $f_B^{-1}(S)$, and finally the ATM checks the identity of the smart card by computing $f_B\big(f_B^{-1}(S)\big) = S$.

58. Fermat factorisation

We mentioned earlier that the primes p and q used in the RSA cryptosystem should not be too close to each other, otherwise the product pq may become easy to factorise using the Fermat factorisation method.

In its simplest form, the Fermat factorisation method aims at factorising a positive integer n by discovering integers a and b such that $a^2 \equiv b^2 \pmod{n}$ but $a \not\equiv \pm b \pmod{n}$. In fact, if that is the case then n divides $a^2 - b^2 = (a-b)(a+b)$, but neither factor, and then $(n, a-b)$ provides a proper factor of n.

A simple way to find a and b such that $a^2 \equiv b^2 \pmod{n}$ but $a \not\equiv \pm b \pmod{n}$ is choosing $a \geq \sqrt{n}$ and checking if the remainder of dividing a^2 by n is a square (of some integer b). This works very efficiently if a positive odd integer n is the product of two (not necessarily prime) integers p and q, say with $p < q$, which are close to each other. One checks for a several consecutive increasing integers starting with $\lceil \sqrt{n} \rceil$, computing $a^2 - n$ and increasing a by one at each step until $a^2 - n$ is a square. [24]

Example. If $n = 17399$, we proceed as follows, starting with $a = \lceil \sqrt{n} \rceil = 132$:

$$132^2 - 17399 = 17424 - 17399 = 25 = 5^2,$$

and hence $17399 = (132-5)(132+5) = 127 \cdot 137$. Wow, this was fast!

If $n = 15943$, we start with $a = \lceil \sqrt{n} \rceil = 127$:

$$127^2 - 15943 = 16129 - 15943 = 186, \quad \text{not a square,}$$

$$128^2 - 15943 = 16384 - 15943 = 441 = 21^2,$$

and hence $15943 = (128-21)(128+21) = 107 \cdot 149$.

[24] In fact, the method is successful as soon as a reaches the value $(p+q)/2$, and hence after a number of steps close to $(p + n/p)/2 - \sqrt{n} = (\sqrt{n} - p)^2/2p$, which is small exactly when p and q are very close. However, this method can be very inefficient when applied in the general case. For example, if $n = pq$ and q is close to p^2 the method takes about $n^{2/3}/2$ steps, which is worse than trial division.

Lecture notes of Algebra.
Appendix: binomial coefficients

This Appendix contains optional material on binomial coefficients, which could have been covered early in the module. Much of this will be familiar to you from A-levels, but you may find the extension to negative and fractional exponents new and interesting.

59. (Optional) Binomial coefficients

For any integers n and k with $0 \leq k \leq n$, the symbol $\binom{n}{k}$, read *n choose k,* and called a *binomial coefficient* (we will see later why), denotes the number of different subsets with k elements of a set S with n elements. In a different wording, it counts the number of ways, disregarding order, in which k objects can be chosen from among n distinct objects. This is also sometimes described as the number of *combinations of n objects taken k at a time,* or *combinations of n objects of class k*, or *k-combinations of n objects.* Another way to say this is the number of *unordered selections of k objects out of n objects.* Clearly the nature of the elements of S does not matter for the count, as long as their number is n, and so we may as well take $S = \{1, 2, \ldots, n\}$ as a convenient example. (Note that such a description, even though we have used the number '2' in it, should be interpreted as $S = \{1\}$ if $n = 1$, and $S = \{\} = \emptyset$, the empty set, when $S = 0$.) Because any set has the empty set $\emptyset = \{\}$ and itself as subsets, we have $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$.

Example. The set $S = \{1, 2, 3, 4\}$ has

- 1 subset with 0 elements: the empty set $\emptyset = \{\ \}$;
- 4 subsets with 1 element: $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$;
- 6 subsets with 2 elements: $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, $\{3, 4\}$;
- 4 subsets with 3 elements: $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$, $\{2, 3, 4\}$;
- 1 subsets with 4 elements: $S = \{1, 2, 3, 4\}$.

When listing the subsets of k elements we have used the *lexicographic order,* which not strictly necessary but is convenient in practice (also to make sure that we do not miss any). Hence the numbers of combinations $1, 4, 6, 4, 1$ which we have found are denoted by $\binom{4}{0}$, $\binom{4}{1}$, $\binom{4}{2}$, $\binom{4}{3}$, $\binom{4}{4}$. We will see in the next section that these numbers form a row of *Pascal's triangle.*

In a different terminology, we have written down, and counted, all combinations of 4 objects taken k at a time, for $k = 1, \ldots, 4$. In yet another terminology, we have written down and counted all unordered selections of k objects out of 4 objects, for $k = 1, \ldots, 4$.

If we had chosen to list and count all *ordered selections,* however, we would have found many more. By an *ordered selection of k objects out of n objects* we mean one where the order in which we choose (or select) the elements makes a difference. (We denote such a selection with square brackets here, but this is not a standard notation.) In this example, out of the 4 objects $1, 2, 3, 4$ there are (again in lexicographic order)

- 1 unordered selection of 0 objects: [ ] (selecting no object);
- 4 unordered selections of 1 object: [1], [2], [3], [4];
- 12 unordered selections of 2 objects: $[1, 2]$, $[1, 3]$, $[1, 4]$, $[2, 1]$, $[2, 3]$, $[2, 4]$, $[3, 1]$, $[3, 2]$, $[3, 4]$, $[4, 1]$, $[4, 2]$, $[4, 3]$;
- 24 unordered selections of 3 objects (commas omitted for shortness): [123], [124], [132], [134], [142], [143], [213], [214], [231], [234], [241], [243], [312], [314], [321], [324], [341], [342], [412], [413], [421], [423], [431], [432];
- 24 unordered selections of 4 objects (also called *permutations of* 4 *objects*):
  [1234], [1243], [1324], [1342], [1423], [1432],
  [2134], [2143], [2314], [2341], [2413], [2431],
  [3124], [3142], [3214], [3241], [3412], [3421],
  [4123], [4132], [4213], [4231], [4312], [4321].

The ordered selections of n objects out of n objects are also commonly known as *permutations of n objects*. Hence a permutation of n objects is an arrangement of the objects as a list in a particular order. [25]

Note that the number of ordered selections of $n - 1$ objects out of n objects is the same as the number of permutations of n objects: this is because after selecting $n - 1$ of the n objects in a certain order, there is only one way to select an nth object to form a permutation, by taking the only element left. More generally, ordered selections are easier to count than unordered selections, as follows.

Proposition 53. *For $0 \leq k \leq n$, the number of ordered selections of k objects out of n objects equals $n(n-1)(n-2)\cdots(n-k+1)$.*

The expression $n(n-1)(n-2)\cdots(n-k+1)$ is the product of k consecutive natural numbers, starting from n and descending. When $k = 0$ the notation $n(n-1)(n-2)\cdots(n-k+1)$ should be read as an empty product, and hence given the value 1 (according to an established convention).

Proof. We can select the first object in n different ways. After choosing the first object, there remain precisely $n - 1$ objects to choose from, so the ways to select two objects, in order, are $n(n - 1)$. After two objects have been selected, there are exactly

[25] However, beware that in the *Algebraic Structures* module, and more general in more advanced algebra, the name *permutation* is given to the operation of passing from a certain arrangement of the n objects to another, rather than to the arrangement itself.

$n-2$ ways to select a third one, so the ways of selecting three objects, in order, are $n(n-1)(n-2)$. And so on. [26] □

Proposition 54. *For $0 \le k \le n$, the number of unordered selections of k objects out of n objects, hence the binomial coefficient $\binom{n}{k}$, can be computed by the* explicit formula for binomial coefficients

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!\,(n-k)!}.$$

The second expression for $\binom{n}{k}$, involving factorials (where $n! := n(n-1)\cdots 2\cdot 1$ for n a positive integer, and $0! := 1$) is more compact than the previous expression, but takes longer to compute (because it involves more factors, that are eventually going to cancel out). Also, the latter expression only makes sense for $0 \le k \le n$, while the former makes sense for any integer n, or even complex number n, generalisations which we will study later.

Proof. Given any unordered selections of k objects out of n objects, we can obtain precisely $k!$ ordered selections from it by ordering its objects in all possible ways (the permutations of those k objects involved). Clearly two different unordered selections will produce different ordered selections by doing this, because they will involve different sets of k objects. Consequently, the number $n(n-1)(n-2)\cdots(n-k+1)$ of ordered selections equals $k!$ times the number $\binom{n}{k}$ of ordered selections.

This proves the first expression for $\binom{n}{k}$ given in the proposition. The second expression $n!/(k!\,(n-k)!)$ reduces to the first after simplifying $(n-k)!$, because $n! = n(n-1)\cdots(n-k+1)\cdot(n-k)!$. □

60. (Optional) The binomial theorem and Pascal's triangle

In this section we will develop some basic properties of binomial coefficients. When possible it will be instructive to do so starting from the combinatorial definition, without relying on the explicit formulas found in Proposition 54.

The easiest property which the binomial coefficients satisfy is the *symmetry formula*

$$\text{(Pascal Symmetry)} \qquad \binom{n}{n-k} = \binom{n}{k}.$$

Of course this follows at once from the explicit formula $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$, but here is a combinatorial proof from the definition.

Proof. This is because for each subset R with k elements of S, its complement $S \setminus R = \{x \in S : x \notin R\}$ has $n-k$ elements, and so counting the different subsets of S

[26] The expression *and so on* is mathematically unsatisfactory. It can be made rigorous by using *mathematical induction,* which you will learn in the *Methods of Proof* module.

with k elements will give the same result as counting those with $n-k$ elements, which are precisely their complements. □

It is customary to extend the definition of $\binom{n}{k}$ to any integer k by defining it to be zero when $k<0$ or $k>n$. In fact, this is also quite natural, because a set of n elements cannot have a subset with a negative number of elements, or with more than n elements, and so the number of such subsets is actually zero in those cases.

Binomial coefficients take their name from the following result (usually credited to Isaac Newton (1642–1727), who, however, is responsible for a generalization to n a rational number).

Theorem 55 (The binomial theorem). *For any complex numbers a, b, and for any integer $n \geq 0$, we have*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Proof. When expanding the product

$$(a+b)^n = \underbrace{(a+b)\cdots(a+b)}_{n},$$

the term $a^{n-k}b^k$ will occur in a number of ways, each of them corresponding to selecting b from exactly k of the n factors $(a+b)$, and a from the remaining ones. Hence the number of ways, which will be the coefficient of $a^{n-k}b^k$ in the expansion, equals the binomial coefficient $\binom{n}{k}$. □

Because we have agreed that the binomial coefficient $\binom{n}{k}$ vanishes outside the range $0 \leq k \leq n$, it is sometimes convenient to omit the range in the sum notation, simply writing $\sum_k$ instead of $\sum_{k=0}^{n}$, and think of the summation index k as ranging over the natural numbers, or even over all the integers (but all terms will be zero except for those with $0 \leq k \leq n$). This trick facilitates some polynomial manipulations where keeping track of the correct summation range might be awkward.

An equivalent way of stating the binomial theorem is the polynomial identity

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$$

in $F[x]$. In fact, we can recover the formulation with a and b by substituting $x = b/a$ into the polynomial:

$$(a+b)^n = a^n \cdot \left(1 + \frac{b}{a}\right)^n = a^n \sum_{k=0}^{n} \binom{n}{k} (b/a)^k.$$

The binomial coefficients satisfy various identities, the most important of which is the *addition formula* (or *Pascal's rule,* or *the main recursion formula,...*)

$$\text{(Pascal Addition)} \qquad \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Proof. All subsets A with k elements of the set $S = \{1, 2, \ldots, n\}$ can be divided into two types, namely, those such that $n \in A$, and those such that $n \notin A$. Those of the former type are as many as the subsets with $k-1$ elements of the set $S \setminus \{n\} = \{1, 2, \ldots, n-1\}$, as one sees by removing n from both A and S, and so their number equals $\binom{n-1}{k-1}$. Those of the latter type are actually contained in $\{1, 2, \ldots, n-1\}$, and so their number is the number of subsets of k elements of $\{1, 2, \ldots, n-1\}$, which equals $\binom{n-1}{k}$. □

We can also prove the addition formula (Pascal Addition) starting from the Binomial Theorem, as follows.

Proof. We have

$$\begin{aligned}\sum_k \binom{n}{k} x^k &= (1+x)^n = (1+x)^{n-1} \cdot (1+x) \\ &= \sum_k \binom{n-1}{k} x^k + \sum_k \binom{n-1}{k} x^{k+1} \\ &= \sum_k \binom{n-1}{k} x^k + \sum_k \binom{n-1}{k-1} x^k\end{aligned}$$

The addition formula follows by comparing the coefficient of x^k in the left-hand side and in the right-hand side. □

The addition formula (Pascal Addition) allows us to compute all binomial coefficients recursively, as long as we know some initial values. For example, the traditional way of arranging the binomial coefficients into the rows of *Pascal's triangle*,

$$\begin{array}{llllll}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
1 & 5 & 10 & 10 & 5 & 1 \\
\vdots & \vdots & & & \ddots & \ddots
\end{array}$$

is setting $\binom{n}{0} = \binom{n}{n} = 1$ for all natural numbers n, and then using the addition formula (Pascal Addition) to recover the rest. (Pascal's triangle is sometimes drawn in a left-right symmetric way, but the way we have drawn it here is better, because rows are indexed by n and columns by k, which also works better for generalizations, as we shall see.)

A better choice for the initial values is setting $\binom{n}{0} = 1$ for all integer $n \geq 0$, and $\binom{0}{k} = 0$ for all integer $k > 0$. This amounts to extending each row of Pascal's triangle indefinitely to the right by filling it with zeroes. An even better choice for the initial values is setting $\binom{n}{0} = 1$ for all integer $n \geq 0$, and $\binom{0}{k} = 0$ for all integer $k \neq 0$. This amounts to filling each row of Pascal's triangle with zeroes on both sides, and matches our convention that $\binom{n}{k} = 0$ if $k < 0$ or $k > n$.

It is possible to compute a binomial coefficient just from one nearby (knowing n and k), without having to fill in all previous rows of Pascal's triangle. This is based on one of the following two formulas, which follow directly from the explicit formula for binomial coefficients:

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}, \quad \text{and} \quad \binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1},$$

both valid for any $n \in \mathbb{Z}$, and any positive integer k.

Proof.

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n}{k}\cdot\frac{(n-1)\cdots(n-k+1)}{(k-1)\cdots 1} = \frac{n}{k}\binom{n-1}{k-1}$$

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n-k+1}{k}\cdot\frac{n\cdots(n-k+2)}{(k-1)\cdots 1} = \frac{n-k+1}{k}\binom{n}{k-1}$$

□

The second identity might be slightly harder to remember than the first (although both are easy to work out when needed), but is perhaps more useful because it allows one to compute each coefficient of $(1+x)^n$ from the previous one, without having to compute all previous rows of Pascal's triangle before.

Example. Here is how to compute the binomial expansion of $(a+b)^n$ without having to write down all of Pascal's triangle up to that row:

- Suppose we have computed the coefficient of $a^{n-k+1}\,b^{k-1}$, which of course is $\binom{n}{k-1}$. So we have found the term $\binom{n}{k-1}a^{n-k+1}\,b^{k-1}$ of the binomial expansion.
- Multiply this term by the exponent of a in it, and then decrease the exponent of a by 1: so we get $(n-k+1)\binom{n}{k-1}a^{n-k}\,b^{k-1}$.
- Now increase the exponent of b by 1, and then divide by the (new) exponent of b: we get $\frac{n-k+1}{k}\binom{n}{k-1}a^{n-k}\,b^{k}$, which we have seen is the same as $\binom{n}{k}a^{n-k}\,b^{k}$.

For example, to compute $(a+b)^{10}$, suppose we have already written down $(a+n)^{10} = a^{10} + 10a^9b + 45a^8b^2 + \cdots$. Then to get the next term:

- Take the last term computed: $45a^8b^2$.

- Multiply this term by the exponent of a, which is 8, and then decrease the exponent of a by 1: so we get $8 \cdot 45a7b^2$.
- Now increase the exponent of b by 1, and then divide by the exponent of b: we get $\dfrac{8 \cdot 45}{3}a^7b^3 = 120a^7b^3$.

So we have just found $(a+n)^{10} = a^{10} + 10a^9b + 45a^8b^2 + 120a^7b^3 + \cdots$. Note that we have not computed $8 \cdot 45$ immediately, but we have waited, and in fact at the next step a division by 3 came, and simplification made the calculation slightly easier.

61. (Optional) The binomial series for negative integer exponents

We can also use the addition formula (Pascal Addition) to extend Pascal's triangle upwards, as soon as we know (or define) exactly one entry in each row above the row corresponding to $n = 0$. The right choice here is defining $\binom{n}{0} = 1$ for all integer n, including negative n (even though the symmetry formula $\binom{n}{n-k} = \binom{n}{k}$ does no longer hold for negative n). In this way the definition of binomial coefficient $\binom{n}{k}$ is extended to arbitrary integer values of both n and k: they are the integers obtained by recursively using the addition formula

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \qquad \text{for all integers } n, k,$$

starting from the *initial conditions*

$$\binom{n}{0} = 1 \quad \text{for all integers } n, \qquad \text{and} \qquad \binom{0}{k} = 0 \quad \text{for integer } k \neq 0.$$

This can be done in a unique way, and so this is a valid definition (even if it may seem impractical). The resulting extension of Pascal's triangle looks as follows:

$$\begin{array}{ccccccc}
\vdots & & & & & & \\
1 & \vdots & & & & & \\
1 & -5 & & & & & \\
1 & -4 & 10 & & & & \\
1 & -3 & 6 & -10 & & & \\
1 & -2 & 3 & -4 & 5 & \cdots & \\
1 & -1 & 1 & -1 & 1 & -1 & \cdots \\
1 & & & & & & \\
1 & 1 & & & & & \\
1 & 2 & 1 & & & & \\
1 & 3 & 3 & 1 & & & \\
1 & 4 & 6 & 4 & 1 & & \\
1 & 5 & 10 & 10 & 5 & 1 & \\
\vdots & \vdots & & & & \ddots & \ddots
\end{array}$$

The North-East quadrant of this picture is a sort of replica of Pascal's triangle, apart from the entries of some columns carrying a negative sign. This reflects another identity for binomial coefficients, which we may call *the negation formula* (see below for a proof):

$$\text{(Pascal Negation)} \qquad \binom{-n}{k} = (-1)^k \binom{n+k-1}{k}.$$

This sort of symmetry of the extension of Pascal's triangle is valid for any integers n and k, and hence it is more general than the symmetry expressed by the symmetry formula (Pascal Symmetry), which only holds as long as $n \geq 0$.

For example, the negation formula (Pascal Negation) can be easily deduced from the explicit formula for binomial coefficients, as follows:

$$\begin{aligned}\binom{-n}{k} &= \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} \\ &= (-1)^k \cdot \frac{n(n+1)\cdots(n+k-1)}{k!} = (-1)^k \binom{n+k-1}{k}.\end{aligned}$$

All entries to the left of the column of ones in the above diagram are zero. Hence each row starts, in practice, from $k = 0$. The rows for $n < 0$ continue indefinitely to the right, and so each such row cannot be considered the sequence of coefficients of a polynomial. However, it can be interpreted as the sequence of coefficients of a power series $\sum_{k=0}^{\infty} \binom{n}{k} x^k$, also written $\sum_{k\geq 0} \binom{n}{k} x^k$, or even $\sum_k \binom{n}{k} x^k$ since the coefficients are zero anyway for $k < 0$.

Example. For $n = -1$, using the addition formula recursively (and upwards), or through the explicit formula, we find $\binom{-1}{k} = (-1)^k$ for $k \geq 0$ (we will omit k *integer* for brevity from now on). Hence

$$(1+x)^{-1} = \frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k = \sum_{k=0}^{\infty} (-x)^k.$$

It might be convenient to replace x with $-x$, whence

$$(1-x)^{-1} = \frac{1}{1-x} = \sum_{k=0}^{\infty} x^k.$$

We already know this one from the subsection on arithmetic and geometric progressions, as we used it to sum an infinite geometric series. However, note that the formula for summing a *finite* geometric series can be recovered from it, as follows:

$$\sum_{k=0}^{n-1} x^k = \sum_{k=0}^{\infty} x^k - \sum_{k=n}^{\infty} x^k = \sum_{k=0}^{\infty} x^k - x^n \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} - \frac{x^n}{1-x} = \frac{1-x^n}{1-x}.$$

(When exactly taking the difference of two series in this way makes sense will be seen later in the Calculus module.) Note that this deduction is only valid for $|x| < 1$, but the final conclusion is an equality between polynomials (thre right-hand side is a rational expression, but becomes a polynomial after cancellation), which, consequently, is true for all values of $x \neq 1$ (because of the division by $x - 1$, even though it would eventually cancel). This is because if two polynomials take the same value for infinitely many values of x, for example for all real x with $|x| = 1$, then they are the same polynomial.

EXAMPLE. For $n = -2$, again using the addition formula recursively, or through the explicit formula, we find $\binom{-2}{k} = (-1)^k(k+1)$ for $k \geq 0$. Again, the binomial series will look simpler after replacing x with $-x$, whence

$$(1-x)^{-2} = \frac{1}{(1-x)^2} = \sum_{k=0}^{\infty}(k+1)x^k = 1 + 2x + 3x^2 + 4x^3 + 5x^4 + \cdots .$$

If you prefer, it may look even simpler after multiplying by x,

$$\frac{x}{(1-x)^2} = \sum_{k=1}^{\infty} kx^k = x + 2x^2 + 3x^3 + 4x^4 + \cdots .$$

This can be used to compute the sum of an infinite series such as

$$1 + \frac{2}{2} + \frac{3}{2^2} + \frac{4}{2^3} + \frac{5}{2^4} + \cdots = \frac{1}{\left(1 - \frac{1}{2}\right)^2} = 4$$

(sometimes called an *arithmetico-geometric* series, but this name is not really important).

EXAMPLE. (Optional) As in a previous example we can adapt the formula for $\sum_{k=1}^{\infty} kx^k$ to a finite sum if we like:

$$\begin{aligned}\sum_{k=1}^{n} kx^k &= \sum_{k=1}^{\infty} kx^k - \sum_{k=n+1}^{\infty} kx^k \\ &= \sum_{k=1}^{\infty} kx^k - x^n \sum_{k=1}^{\infty}(k+n)x^k \\ &= \sum_{k=1}^{\infty} kx^k - x^n \sum_{k=1}^{\infty} kx^k - nx^n \sum_{k=1}^{\infty} x^k \\ &= \frac{x}{(1-x)^2} - \frac{x^{n+1}}{(1-x)^2} - \frac{nx^{n+1}}{1-x} \\ &= \frac{x^{n+1}(nx - n - 1) + x}{(x-1)^2}\end{aligned}$$

Again, we have deduced this formula for $|x| < 1$, but as in the previous example this now holds for arbitrary $x \neq 1$. For example, we have

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \frac{4}{2^4} + \cdots + \frac{10}{2^{10}} = \frac{(1/2)^{11}(10/2 - 10 - 1) + 1/2}{(-1/2)^2} = 2 - \frac{12}{2^{10}} = \frac{509}{256},$$

but also

$$1 \cdot 2 + 2 \cdot 2^2 + 3 \cdot 2^3 + 4 \cdot 2^4 + \cdots + 10 \cdot 2^{10} = \frac{2^{11}(10 \cdot 2 - 10 - 1) + 2}{(2-1)^2} = 18434.$$

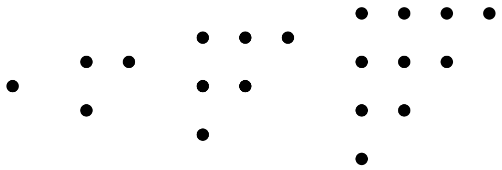
Example. For $n = -3$ we find

$$\binom{-3}{k} = \frac{(-3)(-4)\cdots(-k-1)(k-2)}{k!} = (-1)^k \frac{(k+1)(k+2)}{2} = (-1)^k \binom{k+2}{2}$$

for $k \geq 0$. Replacing x with $-x$ (to avoid the alternating signs) we find

$$(1-x)^{-3} = \frac{1}{(1-x)^3} = \sum_{k=0}^{\infty} \frac{(k+1)(k+2)}{2} x^k = 1 + 3x + 6x^2 + 10x^3 + 15x^4 + \cdots .$$

The coefficients, which also appear in the third column of Pascal's triangle, are sometimes called the *triangular numbers*:



Hence the first few of them can be read off the decimal digits of

$$\frac{1}{0.99^3} = 1.\,03\,06\,10\,15\,21\,28\,36\,45\,55\,66\,78 \cdots .$$

Example. Noting that

$$(k+1)^2 = 2\frac{(k+1)k}{2} + (k+1)$$

and combining the two previous examples we find

$$\frac{1+x}{(1-x)^3} = \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} (k+1)^2 x^k = 1 + 4x + 9x^2 + 16x^3 + 25x^4 + \cdots ,$$

and hence

$$\frac{1.01}{0.99^3} = 1.\,04\,09\,16\,25\,36\,49\,64 \cdots .$$

62. (Optional) The binomial series for arbitrary real or complex exponents

The explicit formula for binomial coefficients, in its expression without factorials, actually still makes sense when n is any number, even real or complex, as long as k remains a natural number. This suggests a more general definition of binomial coefficients,

$$\binom{n}{k} := \frac{n(n-1)\cdots(n-k+1)}{k!}, \qquad \text{for } n \in \mathbb{C} \text{ and } k \text{ a nonnegative integer.}$$

Binomial coefficients defined this way still satisfy the addition formula (Pascal Addition), and so for $n \in \mathbb{Z}$ they coincide with those we constructed earlier by backwards recursion based on the addition formula. More importantly, the following generalised version of the binomial theorem (Theorem 55) holds.

Theorem 56 (Generalised binomial theorem). *For any complex number n, and for any complex number x with $|x| < 1$ we have*

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

That is, the series at the right-hand side converges, and its sum is the number at the left-hand side.

The theorem is really about the Taylor series at $x = 0$ for the function $(1+x)^n$, see the Calculus module.

Example. For $n = 1/2$ we have

$$\binom{1/2}{1} = \frac{1}{2}, \quad \binom{1/2}{2} = \frac{\frac{1}{2}(\frac{1}{2}-1)}{2} = -\frac{1}{8}, \quad \binom{1/2}{3} = \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{6} = \frac{1}{16},$$

and so on, whence

$$\sqrt{1+x} = (1+x)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \frac{7}{256}x^5 - \cdots.$$

In this case it may be convenient to replace x with $4x$ to get rid of the denominators, and we would find

$$\sqrt{1+4x} = (1+4x)^{1/2} = 1 + 2x - 2x^2 + 4x^3 - 10x^4 + 28x^5 - \cdots.$$