Algebra

Moritz Müller

February 2, 2026

Menschen, die von der Algebra nichts wissen, können sich auch nicht die wunderbaren Dinge vorstellen, zu denen man mit Hilfe der genannten Wissenschaft gelangen kann.

Gottfried Wilhelm Leibniz

They that are ignorant of Algebra cannot imagine the wonders in this kind are to be done by it.

John Locke

Contents

Preface

History

This course covers classical algebra as the theory of solving polynomial equations, a theory that found its completion in the 19th century. Here, we give a brief historical overview.

An early and famous example is $X^2 = 2$ for the length $\sqrt{2}$ of the diagonal in the unit square. A Babylonian clay tablet dating c.1700 BCE gives a 6 digit approximation of $\sqrt{2}$. The discovery that $\sqrt{2}$ is not rational is sometimes but uncertainly credited to Hippasus (c.530- c.450 BCE). This shattered the numerological esoterics of pythagorean sects, and legend has it that Hippasus was punished by the gods (read: pythagoreans) to drown.

Irrational numbers were hard to conceptualize by the Greeks, dubbed *alogos*, and later 'irrational' in Euclid's *Elements* (c.300 BCE). This may be the most influential textbook ever written. It established the axiomatic method itself and carried it out for geometry. It also contains a basic development of number theory and proves the infinitude of primes.

Diophantus (c.250 BCE) was the first to accept positive rationals as numbers. His book *Arithmetica* introduced the notation of variables and showed how to solve 130 specific quadratic equations in several variables over the rationals or the integers – the latter being known today as *Diophantine equations*. He rejected, however, negative numbers. These first appeared in the Chinese anonymous book *Nine Chapters on the Mathematical Art* (c.100 CE) and Liu Hui (c.350 CE) explained how to compute with them.

The first systematic approach treating equations as the objects of study was *al-Kitabal-Mukhtasar fi Hisab al-Jabr wal-Muqbalah* (The Compendious Book on Calculation by Completion and Balancing) by the polymath al-Khwarizmi (c.780-c.850). Even the word 'algebra' stems from this book – al-Jabr, originally meaning bone-setting. Its topic is how to find positive solutions to quadratic equations. General formulas had been found already by the Indian mathematician Brahmagupta (c.598-c.668) in the book *Brahma-sphuta-siddhanta* (Correctly Established Doctrine of Brahma), notably using negative numbers which were absent from al-Khwarizmi's work. But already his successor, "the Egyptian calculator" Abu Kamil (c.850-c.930) became the first mathematician to accept both negative and irrational numbers. The first proof of the fundamental theorem of number theory (prime factorization) also stems from the Golden Age of Islam, and is due to al-Farisi (1267-1319) in his book *Tadhkira al-ahbab fi bayan al-tahabb* (Memorandum for Friends on the Proof of Amicability), completing the steps taken in Euclid's *Elements*.

The Indo-Arabic decimal notation was introduced to Europe, replacing the cumber-

some Roman notation, by Fibonacci (c.1170-c.1245) in his book *Liber Abaci* (The book of Calculations), following Abu-Kamil. The book described the growth of rabbit populations by what became known as the *Fibbonaci sequence* – known in India since the 6th century.

In Europe, negative numbers were first used in the book *Ars magna* of the Italian polymath Cardano (1501-1576). He was a provocative, irascible gambler, chronically short of money, who said to despise religion and ended up jailed by the inquisition for heresy. *Ars magna* contained formulas solving cubic and quartic equations, namely expressions built from the coefficients by arithmetical operations, divisions and roots. Formulas for the quartic are due to his scholar Ferrari (1522-1565). Formulas for (a special case of) the cubic are due to del Ferro (1465-1526) who kept them secret in order to maintain advantage in public challenges lecturers posed each other at the time in order to win over or defend their positions. On his deathbed he passed them to his scholar Fiore who went on to challenge Tartaglia (*the stammerer*, c.1500-1557). Tartaglia won, figuring out the formulas himself. He revealed them in the form of a cryptic poem to Cardano who had pressured him with insults and sworn an oath of secrecy. After finding dead del Ferro's notebook, Cardano decided to publish them – and enraged Tartaglia.[1]

The formulas required computations with complex numbers, for Cardano a "mental torture". The imaginary unit i got its derogatively meant name from Descartes (1637). Still in 1702, Leibniz called i a "feine und wunderbare Zuflucht des menschlichen Geistes, beinahe ein Zwitterwesen zwischen Sein und Nichtsein." Why the trouble? The first to give calculation rules for complex numbers was Bombelli (1526-1572), crucially considering them as neither negative nor positive – they cannot be organized on an line compatibly with the arithmetical operations. Now, this is a conceptual leap: numbers are often vaguely thought to represent 'magnitudes' or 'quantities'. E.g., the opening sentence of Euler's book *Algebra* (1770) reads "Endlich wird alles dasjenige eine Größe genennt, welches einer Vermehrung oder Verminderung fähig ist, oder wozu sich noch etwas hinzusetzen oder davon wegnehmen läßt". Whatever this means, if anything at all, it seems to imply a linear order since for any two distinct 'quantities' one is smaller than the other.

Number theory gained unlikely momentum from a lawyer, namely Fermat (1607-1665). He did, however, not prove his insights – famous is *Fermat's Last Theorem*, annotated around 1637 at the margin of Diophantus' *Arithmetica*, and finally proved 1994 by Wiles. Fermat's work was continued by Euler (1707-1783). The first rigorous textbook on number theory was *Disquisitiones Arithmeticae* published 1801 by Gauß (1777-1855). This work also contains Gauß' law of quadratic reciprocity, a suprising result that keeps facinating mathematicians through the centuries – by today more than 300 proofs appeared.[2]

Another lawyer, d'Alembert, stated the *fundamental theorem of algebra* in in 1746: all reasonable polynomial equations in one variable have complex solutions. Gauß corrected an error and gave an own proof in 1799, also incomplete, later giving several others. The first complete proof appeared 1813 by a mysterious amateur, named Argand – mysterious

[1] R. W. Feldmann Jr., The Cardano-Tartaglia dispute. The Mathematics Teacher 54 (3): 160-163, 1961.
[2] https://www.mathi.uni-heidelberg.de/~flemmermeyer/qrg_proofs.html

because not much is known about him besides his surname.[3]

A central question that eluded all efforts for centuries was to find formulas a la Cardano to solve quintic equations. Gauß conjectured, in the *Disquisitiones*, that such formulas might not exist. 1799 appeared Ruffini's book *Teoria Generale delle Equazioni, in cui si dimostra impossibile la soluzione algebraica delle equazioni generali di grado superiore al quarto*. The proof was largely ignored at the time (Cauchy being an exception), probably due to its length (> 500 pages) and weird methods - today named group theory. Ruffini's proof was incomplete. A complete proof stems from a shy, modest teenager, named Abel (1802-1829) and later also the "Mozart of mathematics" (Klein). He published *Mémoire sur les équations algébriques ou on démontre l'impossibilité de la résolution de l'équation générale du cinquiéme degré* 1824 at own costs, condensed to 6 pages to save money. Gauß received a copy – and ignored it like everybody else. Abel died at age 26 from tuberculosis, in abject poverty, buried under the debts of his alcoholic parents.

The Abel-Ruffini theorem states that there are no formulas a la Cardano that work generally for all quintic equations. Might every quintic equation be solvable by formulas specially designed for it? No. Concrete counterexamples came from another teenager, namely Galois (1811-1832), not shy at all but a french republican revolutionary who died at the age of 20 in a mysterious duel. His paper *Mémoire sur les conditions de résolubilité des équations par radicaux* revolutionized algebra. It was first rejected as "incomprehensible" (Poisson), published posthumously 1843 by Liouville, and only slowly understood by the community. The revolutionary insight was that the existence of solving formulas is determined by a property of the *Galois group*: those permutations of the roots of the given polynomial that preserve all polynomial equations satisfied by these roots. By lack of fantasy, this group property is called *solvability*.

People started to develop group theory. Poincaré's statement "Les mathématiques ne sont qu'une histoire des groupes" from 1881 underlines the enthusiam continuing all through the 20th century. Feit and Thompson proved 1963 that every finite group of odd order is solvable; the paper has more than 250 pages. The *enormous theorem* classifies all finite so-called *simple* groups - in some sense, the building blocks of all finite groups. The proof is ≈ 15000 pages long and scattered in hundreds of papers by about 100 authors, mainly from the 2nd half of the 20th century. The "proof has never been written down in its entirety, may never be written down, and as presently envisaged would not be comprehensible to any single individual." (E. B. Davies)

The 19th century also found answers to ancient Greek questions on ruler and compass constructions. Wantzel (1814-1848) – also french and dying young and unrecognized – proved 1837 the impossibility of trisecting angles and doubling cubes, working through the nights "faisant alternativement abus de café et d'opium" (Saint-Venan). In 1882, Lindemann proved that π is not a solution of any rational polynomial equation and, thus, squaring the circle is impossible.

At the beginning of the 19th century new mysterious objects entered mathematical practice with Newton and Leibniz' calculus: *infinitessimals*, supposed to behave like pos-

[3] https://mathshistory.st-andrews.ac.uk/Biographies/Argand/

itive reals but being smaller than all $1/2, 1/3, ...$ Their geometric siblings *indivisibles* had been used already by Archimedes (c.287-212 BCE) to calculate volumes. The Jesuate Cavalieri (1598-1647) and others elaborated this method until the Jesuites, for some reason, deemed indivisibles dangerous and banned talk about them in 1632.[4] Dangerous or not, infinitessimals lacked a definition. E.g., Abel said 1826 "the most important parts of mathematics stand without foundation. It is true that most of it is valid, but that is very surprising." Cauchy and Weierstraß re-built calculus without infinitessimals, as it is still taught today. Only much later in the 1960s, A. Robinson defined the *hyperreals* and developed calculus rigorously just as originally envisioned.

This required basic methods of mathematical logic whose development was pushed by Hilbert (1862-1943) during the so-called *foundational crisis of mathematics*. At the time mathematics gained new levels of abstraction following Galois' algebra and Cantor's (1845-1918) set theory. Paradoxes appeared and called for more rigor. Modern logic had already been set up in an unreadable notation called *Begriffsschrift* 1879 by Frege (1848-1925). He wanted to know what the natural numbers are and why $1 + 1 = 2$ but nobody paid attention at his time. Contrarily, the businessman and influential algebraist Kronecker (1823-1891) said "Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk" and opposed Cantor's infinities. He was sided by "the living brain of the rational sciences", the polymath Poincaré (1854-1912) who wrote in 1908 "*Il n'y a pas d'infinie actuel*; les Cantoriens l'ont oublié, et ils sont tombés dans la contradiction." Cantor suffered from biploar disorder, severed ties with his disinterested or openly hostile mathematical contemporaries and, himself deeply religious, wasted his time with worries of catholic clerics that his discoveries might be a vessel for pantheism.[5]

Such were the foreshocks of the foundational crisis. It was mainly fought between Hilbert's *formalism* and Brower's (1881-1966)) *intuitionism* which, in particular, rejected non-constructive existence proofs – at the time a wide-spread source of uneasiness. E.g., Lindemann, Hilbert's thesis advisor, found his (non-constructive) proof of the *Basis Theorem* "unheimlich", and Gordan said first "Das ist keine Mathematik; das ist Theologie" but later conceded "Ich habe mich davon überzeugt, daß die Theologie auch nützlich sein kann." In Hilbert's Göttingen, the "mecca of mathematics", the new abstract, axiomatic approach to algebra was developed by Gordan's student Noether (1882-1935), the first women joining a faculty with the support of Hilbert against sexist norms: "Eine Fakultät ist doch keine Badeanstalt." In 1921, E. Artin (1898-1961) nicknamed "Ma" (for *Ma*thematik), arrived in Göttingen. He and Noether are considered the founders of modern algebra.

The tone during the crisis sharpened 1921 with an article of Hilbert's student Weyl (1885-1955) who had temporarily changed camps. In Hilbert's words the intuitionists want "eine Verbotsdiktatur à la Kronecker errichten. Dies heißt aber unsere Wissenschaft zerstückeln und verstümmeln" and sought a provably consistent axiomatization, like Euclid's of geometry, but of the whole of mathematics including Cantor's infinities: "Aus dem

[4]Slava Gerovitch, Infinitesimal : How a Dangerous Mathematical Theory Shaped the Modern World: A Book Review, in Notices of the AMS 63 (5): 571–574, 2016.

[5]J. W. Dauben, Georg Cantor and Pope Leo XIII: Mathematics, Theology, and the Infinite Dauben, Journal of the History of Ideas 38 (1): 85–108,1977.

Paradies, das Cantor uns geschaffen hat, soll uns niemand vertreiben können". Hilbert's ambitious program was proved impossible 1931 by Gödel (1906-1978), chronically sick, paranoid and according to his friend Einstein "the greatest logician since Ariostotle".

Intuitionism found a degenerate heir in the national socialist "Deutsche Mathematik" of Bieberach, Teichmüller and other nazis. At the university in Berlin, Bieberach was nicknamed "Großinquisitor" for his activity in antisemite persecutions, and, in Göttingen, Teichmüller organized nazi boycotts of Landau's analysis courses. As part of the "Säuberung der Hochschulen von Gelehrten" Noether, a pacifist and former USPD member, got suspended "bis zur endgültigen Entscheidung" (Arbeiterzeitung 26.4.1933) and left to the US in 1933. Artin, then in Hamburg, fled to the US 1937 – while "arisch" his wife was a "Mischling ersten Grades" and he had made no secret about his distaste for the nazis. Asked by Kultusminister Rust (considered an idiot even among his nazi comrades) whether his institute suffered "unter dem Weggang der Juden und Judenfreunde", Hilbert replied "Jelitten? Dat hat nich jelitten, Herr Minister. Dat jibt es doch janich mehr!".

Luckily, Hilbert won the long run. Based on lectures of Noether and Artin "the most influential text of algebra of the twentieth century" was van der Waerden's *Moderne Algebra* (1930). It established for good the abstract, axiomatic approach to algebra via groups, rings and fields and "dramatically changed the way algebra is now taught" (Mac Lane).

On an even larger scale, 20th century mathematics was heavily influenced by Hilbert's 23 problems posed 1900 at the ICM in Paris – some unsolved to date. The 10th asked – more than 2 millenia after *Arithmetica* – whether there exists an algorithm that decides the solvability of Diophantine equations. The word 'algorithm' is coined after (the latin version of) al-Khwarizmi's name and was informal at the time. Church and Turing formalized the concept in 1936, thereby establishing computer science. Matiyasevich answered "no" to Hilbert's 10th in 1975 (building on the work of Davis, Putnam and J. Robinson).

This course

This course covers most of the material mentioned in the historical survey above and is written for readers with basic knowledge of linear algebra and analysis, like a typical 3rd semester student. The material follows Fischer's *Lehrbuch der Algebra*, a canon for german Staatsexamen students, but gives an extra emphasis on number theory. These lecture notes are much shorter mainly due to a concise writing style, and not lack of content.

The abstract axiomatic approach to algebra is now standard and this course makes no exception. In Weyl's words "We cannot help the feeling that certain mathematical structures which have evolved through the combined efforts of the mathematical community bear the stamp of a necessity not affected by the accidents of their historical birth."

This course tries, however, a less standard presentation. We explain first why and then how. Weyl still warned in 1939:

> *Important though the general concepts and propositions may be with which the modern industrious passion for axiomatizing and generalizing has presented us, in algebra perhaps more than anywhere else, nevertheless I am convinced that*

> *the special problems in all their complexity constitute the stock and core of mathematics; and to master their difficulties requires on the whole the harder labor.'*

This is the motto of this course. Historically, the abstract concepts evolved from concrete problems as means to understand them. We claim that this is how they should be taught. In slogan form: the art of mathematics is abstraction, not deduction. As an illustrating example, after the 1st world war a secretive, elitist group of french mathematicians under the alias *Bourbaki* set out, in the words of its member Cartier (1997), "to submit all mathematics to the scheme of Hilbert; what van der Waerden had done for algebra would have to be done for the rest of mathematics." He said "The misunderstanding was that many people thought that it should be taught the way it was written in the books.[...] If you consider it as a textbook, it's a disaster."

Schiller wrote to Goethe (1796) "wo es die Sache leidet, halte ich es immer für besser, nicht mit dem Anfang anzufangen, der immer das schwerste und das leerste ist."

Many courses take a merciless deductive top-down approach in the groups-rings-fields structure. This is "leer" in that typical students lack non-trivial examples of these structures and can only see in the very end how the generality pays off. Good such books counterbalance the general nonsense by extensive motivational discussions and/or arrays of examples. However, Schiller suggests a development bottom-up, starting from familiar structures and stepwise abstracting. This poses a didactical dilemma because it reverses the logical structure. We thus have conflicting goals and leave it to the reader to judge how well we manoeuvered the contradiction.

Exemplary models are M. Artin's *Algebra*, or Borcherd's online courses on elementary number theory. We aim, however, at different material – visible from the fine-structured table of contents. We make some brief comments concerning how we tried a bottom-up approach. Often results are proved twice, with a second "more abstract proof" presented once the appropriate portion of theory is at hand. We hope this helps appreciating that and how the more abstract theory is indeed useful.

Chapter 1 gives set-theoretic constructions of $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$. This is non-canonical and in this sense optional material that we think is valuable especially for ongoing school teachers. The proper material starts in Chapter 2 with $\mathbb{Z}$ – as concrete as it can get. We give special emphasis to applications in computer science: we explain central cryptographic algorithms, namely RSA encryption, the Digital Signature Algorithm and also mention the Diffie-Hellman protocol for key exchange and ElGama-encryption (Section 2.6.1 and 2.7.1). More thoroughly, Sections 2.7.2 and 2.9.1 detail two famous algorithmically efficient probabilistic primality tests, namely the Miller-Rabin test and the Solovay-Strassen test.

Polynomial rings are familiar from linear algebra. Chapter 3 gives material accessible via direct combinatorial arguments and without ring theory. It puts some elementary knowledge (e.g. polynomial division or discriminants) into the conceptually wider context of algebraic field extensions. This aims to reveal the need for more theory. The chapter also functions as a teaser by proving the fundamental theorem of algebra with only some polynomial combinatorics *assuming* the existence of splitting fields (Chapter 6).

Thus, abstract algebra starts quite late in Chapter 4 on rings. Here, our presentation deviates most significantly from a more standard one – by turning it upside down. Having seen quadratic number fields in Chapter 3, it starts with their rings of integers as first examples of less well-behaved rings. This motivates the definition of factorial rings as enjoying a good portion of the familiar divisibility theory. Principal ideal domains appear as examples, easy in that arguments are just abstract versions of known ones. This makes ideal theory well-motivated at the *end* of the chapter (often done in the beginning).

Chapter 5 on groups starts with symmetry groups in the plane, a concrete example accessible with basic linear algebra. We then treat permutation groups as another concrete example, also familiar from linear algera. We continue with three classes of groups that are easily described: cyclic groups are accessible with divisibility theory in $\mathbb{Z}$, finitely generated free abelian groups are accessible with techniques familiar from linear algebra, and finitely presentable groups are handled by direct computations. We define the latter semantically instead of syntactically, thereby avoiding the usual definition based on normal hulls. Normal subgroups are introduced afterwards, starting abstract group theory.

The crown of typical introductory algebra courses is Galois theory treated in Chapter 6 on fields. It starts with ruler and compass constructions as a motivating example. The first theoretical steps elaborate preliminary material from Chapter 3 on algebraic extensions and give quick and easy impossibility results for ruler and compass. We then pay back our debt inherited from Chapter 3 and construct splitting fields and algebraic closures. Back to more concrete structures we treat finite fields and describe Reed-Solomon error correcting codes as an application in computer science (Section 6.5.1). The rest is devoted to Galois theory, the most abstract part of this course. We treat many examples. This is not easy and showcases our motto, Weyl's warning.

Chapter 1

Logical foundations

In this chapter we recall some basic algebraic notions, characterize the natural numbers $\mathbb{N}$ axiomatically and then construct $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$.

Set-theoretic notations

For sets X, Y we let $X \times Y$ denote the set of pairs (x, y) with $x \in X, y \in Y$. For $n \in \mathbb{N}$ we write $X^n := X \times \cdots \times X$ (n times) for the set of n-tuples $(x_1, ..., x_n)$ with $x_1, ..., x_n \in X$. A *(binary) relation* $f \subseteq X \times Y$ is a *function* or *map* if for all $x \in X$ there is at most one $y \in Y$ such that $(x, y) \in f$. Its *domain* is $\{x \in X \mid (x, y) \in f \text{ for some } y \in Y\}$.

A function f is said to be *from* its domain D and *into* Y, symbolically $f : D \to Y$. For $x \in D$ we write $f(x)$ for the unique $y \in Y$ with $(x, y) \in f$ and say f is *defined on* x. If D is clear from context we often refer to f by $x \mapsto f(x)$. E.g., the usual addition $(x, y) \mapsto x + y$ on the naturals $\mathbb{N}$ is the function $+ : \mathbb{N}^2 \to \mathbb{N}$ which, as a set of pairs, equals $\{((x, y), x + y) \mid x, y \in \mathbb{N}\}$; here we use infix notation and write $x + y$ instead $+(x, y)$.

For $X \subseteq D, Y_0 \subseteq Y$ we write

$$f(X) := \{f(x) \mid x \in X\}, \quad f^{-1}(Y_0) := \{x \in D \mid f(x) \in Y_0\}.$$

The *image of* f is $f(D)$. Similarly, for a binary function $\circ : X \times X \to Y$ in infix notation (i.e., $x \circ y := \circ(x, y)$) and for $Z, Z' \subseteq X$ we write

$$Z \circ Z' := \{z \circ z' \mid z \in Z, z' \in Z'\}.$$

For $X \subseteq D$, the *restriction of* f *to* X is $g := f{\restriction}X := f \cap (X \times Y) : X \to Y$. Then f *extends* g. $f : X \to Y$ is *injective* if $f(x) \neq f(x')$ for all $x, x' \in X$ with $x \neq x'$. It is said to be *surjective* or *onto* Y if its image is Y. It is *bijective* if it is both injective and surjective. Then $f^{-1} := \{(y, x) \mid (x, y) \in f\} : Y \to X$ is the *inverse of* f. A bijection from X onto X is a *permutation of* X. E.g., the *identity on* X, namely $x \mapsto x$, i.e.,

$$\mathrm{id}_X := \{(x, x) \mid x \in X\}.$$

If $f : X \to Y$ and $g : Y \to Z$, then the *composition* $g \circ f : X \to Z$ is the function $x \mapsto g(f(x))$, i.e., $g \circ f := \{(x, g(f(x))) \mid x \in X\}$.

Exercise. $f : X \to Y$ is injective if and only if there exists $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$. $f : X \to Y$ is surjective if and only if there exists $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$.

$R \subseteq X^2$ is *reflexive* if $(x, x) \in R$ for all $x \in X$, and *irreflexive* if $(x, x) \notin R$ for all $x \in X$. It is *transitive* if $(x, y), (y, z) \in R$ implies $(x, z) \in R$ for all $x, y, z \in X$. R is an *equivalence relation on* X if it is reflexive, transitive and *symmetric*: $(x, y) \in R$ implies $(y, x) \in R$ for all $x, y \in R$. Then the *equivalence class of* x is $\{y \in X \mid (x, y) \in R\}$. The set of equivalence classes *partitions* X, i.e., the classes are nonempty, pairwise disjoint and their union is X.

R is a *partial order on* X if it is irreflexive and transitive. It is a *linear order* if additionally $(x, y) \in R$ or $(y, x) \in R$ or $x = y$ for all $x, y \in X$. Typically linear orders are denoted $<$ with infix notation, i.e., we write $x < y$ or $y > x$ instead $(x, y) \in <$; then $x \leqslant y$ means $x < y$ or $x = y$.

The 2nd statement in the exercise is (a version of) the *axiom of choice* in set theory. This axiom was first formulated by Zermelo (1871-1953) who considered it unproblematic, even 'logically true'. Nevertheless this axiom was a bone of contention during the foundational crisis of mathematics – conceptually because it embodies a non-constructive existence claim, and technically because it does have some counterintuitive consequences. It is today commonly accepted. We shall occasionally use an equivalent statement, namely *Zorn's lemma* whose statement we recall here.

Let X be a set and $R \subseteq X^2$ a partial order on X. A *chain* is a linearly ordered subset, i.e., a $Y \subseteq X$ such that $R \cap Y^2$ is a linear order on Y. We call R *inductive (on X)* if every chain $Y \subseteq X$ has an *upper bound*, i.e., an $x \in X$ such that $(y, x) \in R$ or $y = x$ for all $y \in Y$. Call $x \in X$ *maximal* if $(x, x') \notin R$ for all $x' \in X$.

Zorn's Lemma. *Inductive partial orders have maximal elements.*

1.1 Basic algebraic notions

1.1.1 Monoids and groups

Definition 1.1.1. A *monoid* is a pair $(G, \circ)$ of a set G and a function $\circ : G^2 \to G$ that is associative (i.e., $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$) and such that there exists a *neutral element* $e \in G$ satisfying $x = x \circ e = e \circ x$. It is *commutative* if $x \circ y = y \circ x$ for all $x, y \in G$.

$(G, \circ)$ is a *group* if every $x \in G$ is *invertible*, i.e., x has an *inverse* $x^{-1} \in G$, i.e., $x \circ x^{-1} = x^{-1} \circ x = e$. Commutative groups are also called *abelian*.

Remark 1.1.2. Let $(G, \circ)$ be a monoid and $x, y \in G$.

1. There cannot be another neutral element $e' \in G$ (since then $e = e \circ e' = e'$).
2. The notation x^{-1} is justified because if y is another inverse of x then $y = y \circ e = y \circ (x \circ x^{-1}) = (y \circ x) \circ x^{-1} = e \circ x^{-1} = x^{-1}$.
3. In particular, $(x^{-1})^{-1} = x$ for all $x \in G$ because x is an inverse of x^{-1}.

4. If for every $x \in G$ there is $y \in G$ with $yx = e$ (omitting $\circ$), then y is an inverse of x: $yx = e$ implies $(yx)y = y$; multiply from the left with z such that $zy = e$ gives $xy = e$.
5. $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$, and similarly $(y^{-1}x^{-1})(xy) = e$, because

$$(xy)(y^{-1}x^{-1}) = x(y(y^{-1}x^{-1})) = x((yy^{-1})x^{-1}) = x(ex^{-1}) = xx^{-1} = e.$$

Exercise 1.1.3. A group $(G, \circ)$ has *cancellation*: if $x \circ z = y \circ z$ or $z \circ x = z \circ y$, then $x = y$. For all $z \in G$ the *left* and *right translation* $x \mapsto z \circ x$ and $x \mapsto x \circ z$ are permutations of G.

Example 1.1.4. In this chapter we are going to explain what $\mathbb{N}, \mathbb{Q}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$ are and how $+, \cdot$ are defined on them. This gives commutative monoids in all cases. With $+$, the sets $\mathbb{Z}, \mathbb{R}, \mathbb{C}$ are abelian groups but not with $\cdot$ because 0 does not have a multiplicative inverse (and in $\mathbb{Z}$ only ± 1 have); $(\mathbb{Q} \setminus \{0\}, \cdot), (\mathbb{R} \setminus \{0\}, \cdot), (\mathbb{C} \setminus \{0\}, \cdot)$ are abelian groups.

Example 1.1.5. For $n > 0$ let $\mathbb{R}^{n \times n}$ be the set of $n \times n$ matrices with entries in $\mathbb{R}$. With matrix addition $+$ it is an abelian group, with matrix multiplication $\cdot$ it is a non-commutative monoid and not a group if $n > 1$. The set of invertible matrices $\mathrm{GL}(n, \mathbb{R}) \subseteq \mathbb{R}^{n \times n}$ is a group with $\cdot$: the *general linear group over* $\mathbb{R}$; it is non-abelian if $n > 1$.

Definition 1.1.6. Let $(G, \circ)$ be a group. A set $\varnothing \neq U \subseteq G$ is a *subgroup of* G if $e \in G$ and for all $x, y \in U$: $x \circ y \in U$ and $x^{-1} \in U$.

Remark 1.1.7. Equivalently, $x \circ y^{-1} \in U$ for all $x, y \in U$.

Clearly, this is implied by U being a subgroup; conversely, $e = x \circ x^{-1} \in U$ and $x, y \in U$ implies $y^{-1} = e \circ y^{-1} \in U$, so also $x \circ y = x \circ (y^{-1})^{-1} \in U$.

Note that then $(U, \circ)$ is a group with the same neutral element – more, precisely we should use the restriction $\circ \restriction U \times U$.

Definition 1.1.8. Let $(G, \circ), (H, *)$ be monoids (groups) with neutral elements e_G, e_H. A *monoid (group) homomorphism from* $(G, \circ)$ *to* $(H, *)$ is a function $\varphi : G \to H$ that *preserves* $\circ$ *and* e, i.e., for all $x, y \in G$:

$$\varphi(x \circ y) = \varphi(x) * \varphi(y), \quad \varphi(e_G) = e_H.$$

The *kernel of* φ is

$$\ker(\varphi) := \{x \in G \mid \varphi(x) = e_H\}.$$

If φ is injective (surjective, bijective), then it is a *monomorphism (epimorphism, isomorphism).* $(G, \circ)$ and $(H, *)$ are *isomorphic*, symbolically $(G, \circ) \cong (H, *)$ if there exists an isomorphism $\varphi : G \to H$; in case, they are isomorphic *via* φ.

An *endomorphism* of $(G, \circ)$ is a homomorphism from $(G, \circ)$ to itself. $\mathrm{Aut}(G, \circ)$ is the set of *automorphisms* of $(G, \circ)$, i.e., bijective endomorphisms.

Remark 1.1.9.

1. If ψ is a another monoid (group) homomorphism from $(H, *)$ to $(H', *')$, then $\psi \circ \varphi$ is a homomorphism from $(G, \circ)$ to $(H', *')$.

2. If $(H, *)$ is a group, preservation of e is automatic: multiply both sides of $\varphi(e_G) = \varphi(e_G \circ e_G) = \varphi(e_G) * \varphi(e_G)$ by $\varphi(e_G)^{-1}$ (inverse in H), and get $e_H = \varphi(e_G)$;
3. φ preserves $\cdot^{-1}$ if defined: assume $x \in G$ has an inverse x^{-1}. Then $\varphi(x^{-1})$ is an inverse of $\varphi(x)$ in $(H, *)$ because $e_H = \varphi(e_G) = \varphi(x \circ x^{-1}) = \varphi(x) * \varphi(x^{-1})$; similarly, $e_H = \varphi(x^{-1}) * \varphi(x)$.
4. If $(G, \circ), (H.*)$ are groups, then φ is injective if and only if $\ker(\varphi) = \{e_G\}$.

   Indeed, $\Rightarrow$ is clear; $\Leftarrow$: if $\varphi(x) = \varphi(y)$, then $e_H = \varphi(x) * \varphi(y)^{-1} = \varphi(x \circ y^{-1})$, so $x \circ y^{-1} \in \ker(\varphi)$, so $x \circ y^{-1} = e_G$, so $x = y$.
5. If $(G, \circ), (H.*)$ are groups and U a subgroup of G, then $\varphi(U)$ is a subgroup of H.

   Indeed, let $y, y' \in \varphi(U)$, say, $\varphi(x) = x', \varphi(x') = y'$ with $x, x' \in U$; then $x' \circ x^{-1} \in U$ and $y' * y^{-1} = \varphi(x') * \varphi(x)^{-1} = \varphi(x' \circ x^{-1}) \in \varphi(U)$.
6. If $(G, \circ), (H.*)$ are groups and V a subgroup of $(H, *)$, then $\varphi^{-1}(V)$ is a subgroup of $(G, \circ)$. In particular, $\ker(\varphi) = \varphi^{-1}(\{e_H\})$ is a subgroup of $(G, \circ)$.

   Indeed, if $x, x' \in U := \varphi^{-1}(V)$, then $x' \circ x^{-1} \in U$ because $\varphi(x' \circ x^{-1}) = \varphi(x') * \varphi(x)^{-1} \in V$ as a subgroup of H.

Exercise 1.1.10 (Subgroup correspondence)**.** Let $(G, \circ), (H, *)$ be groups and $\varphi : G \to H$ a (group) epimorphism. Then $U \mapsto \varphi(U)$ is a bijection from the set of subgroups U of G with $\ker(\varphi) \subseteq U$ onto the set of subgroups V of H; its inverse is $V \mapsto \varphi^{-1}(V)$.

Exercise 1.1.11. If $\varphi : G \to H$ is a group epimorphism and G is abelian, then so is H.

Examples 1.1.12.

1. The map $x \mapsto e^x$ is a group isomorphism from $(\mathbb{R}, +)$ onto $(\mathbb{R}^+, \cdot)$; here, $\mathbb{R}^+$ is the set of positive reals.
2. Preservation of e is not automatic in monoids: the map $x \mapsto \left[\begin{smallmatrix} x & 0 \\ 0 & 0 \end{smallmatrix}\right]$ from $\mathbb{R}$ into $\mathbb{R}^{2\times 2}$ preserves $\cdot$ but not 1.
3. The determinant for invertible matrices satisfies $\det(AB) = \det(A)\det(B)$. This means that $\det : \mathrm{GL}(n, \mathbb{R}) \to (\mathbb{R} \setminus \{0\}, \cdot)$ is a group homomorphism. Since inverses are preserved, it follows that $\det(A^{-1}) = 1/\det(A)$. We have $\ker(\det) = \mathrm{SL}(n, \mathbb{R}) := \{A \in \mathbb{R}^{n\times n} \mid \det(A) = 1\}$; this is the *special linear group*, a subgroup of $(\mathrm{GL}(n, \mathbb{R}), \cdot)$.
4. $z \mapsto |z|$ is a homomorphism from $(\mathbb{C} \setminus \{0\}, \cdot)$ into $(\mathbb{R}^+, \cdot)$; its kernel is the *circle group* $S^1 := \{z \in \mathbb{C} \mid |z| = 1\}$, a subgroup of $(\mathbb{C} \setminus \{0\}, \cdot)$.

Notation: we usually write groups G *multiplicatively* or *additively*, namely $\cdot$ or $+$ for $\circ$. In the former case we write 1 for e, and often omit $\cdot$, i.e., write xy instead $x \cdot y$. In the latter case we write $-x$ instead x^{-1} and 0 for e; we also write $x - y$ instead $x + (-y)$.

1.1.2 Rings and fields

Definition 1.1.13. A *(unitary) ring* is a triple $(R, +, \cdot)$ where $+, \cdot : R^2 \to R$ are called *addition* and *multiplication (of R)* and are such that $(R, +)$ is an abelian group (with neutral element 0), and $(R, \cdot)$ is a monoid (with neutral element 1) and for all $x, y, z \in R$

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z), \quad (x + y) \cdot z = (x \cdot z) + (y \cdot z)$$

It is *commutative* if $x \cdot y = y \cdot x$ for all $x, y \in R$.

A commutative ring $(R, +, \cdot)$ is an *integral domain* if $0 \neq 1$ and there is no *zero-divisor* in R, i.e., an $x \in R \setminus \{0\}$ such that $x \cdot y = 0$ for some $y \in R \setminus \{0\}$.

An integral domain $(R, +, \cdot)$ is a *field* if every $x \in R \setminus \{0\}$ has a multiplicative inverse.

Notation: we omit parentheses and $\cdot$ as usual: e.g., $xy + xz$ is understood as $(x \cdot y) + (x \cdot z)$. To emphasize the ring (group or field) we sometimes write indices $+_R, \cdot_R, 0_R, 1_R$. But we usually omit listing $+, \cdot$ and simply say R is a ring (group, field).

Remark 1.1.14. Let R be a ring. For all $x, y \in R$:

1. A zero-divisor as defined above is often called a *left* zero-divisor; but we shall be interested in this concept only in commutative rings.
2. $0x = 0$ because $0x = (0 + 0)x = 0x + 0x$; similarly, $x0 = 0$.
3. $-(-x) = x$ by Remark 1.1.2 (2).
4. $(-x)y = -(xy)$ because $0 = 0y = (x + (-x))y = xy + (-x)y$. Similarly, $x(-y) = -(xy)$.
5. $(-x)(-y) = xy$ because $(-x)(-y) \overset{(3)}{=} -(x(-y)) \overset{(3)}{=} -(-(xy)) \overset{(2)}{=} xy$.

Lemma 1.1.15. *Let R be a commutative ring with $0 \neq 1$. Then R is an integral domain if and only if it has* cancellation *for $\cdot$, i.e., for all $x, y, z \in R, z \neq 0$: $xz = yz$ implies $x = y$.*

Proof. $\Rightarrow$: if $xz = yz$, then $z(x - y) = 0$; as $z \neq 0$ is not a zero-divisor, $x - y = 0$, i.e., $x = y$.
$\Leftarrow$: if $xz = 0$ for $z \neq 0$, then $xz = 0 \cdot z$, so $x = 0$ by cancellation. □

Definition 1.1.16. Let R be a ring. Then $x \in R$ is a *unit* if it has a multiplicative inverse, i.e., there is $x^{-1} \in R$ such that $x \cdot x^{-1} = x^{-1} \cdot x = 1$. The set of units of R is denoted $R^\times$.

Examples 1.1.17.

1. If R is a ring with $1 = 0$, then $1x = 0x = 0$ for all $x \in R$, so $R = \{0\}$. This is the *trivial* ring. It is commutative and not a field.
2. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields. $\mathbb{Z}$ is an integral domain (hence the name) with $\mathbb{Z}^\times = \{\pm 1\}$.
3. The set $C(\mathbb{R})$ of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is a commutative ring with $+, \cdot$ defined pointwise, i.e., for $f, g \in C(\mathbb{R})$ the sum $f + g$ is defined as the function $x \mapsto f(x) + g(x)$; the product $f \cdot g$ is defined analogously.

4. Let $n > 1$ and V be an n-dimensional vector space over $\mathbb{R}$. The set of endomorphisms of V (linear functions from V to V) is a non-commutative ring with addition defined pointwise and $\circ$ as multiplication. It is isomorphic (see Definition 1.1.21 below) to $\mathbb{R}^{n\times n}$ with the usual matrix operations $+,\cdot$; its units are $(\mathbb{R}^{n\times n})^\times = \mathrm{GL}(n,\mathbb{R})$.
5. $\mathbb{F}_2 := \{0,1\}$ is a field with $\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}$ and $\begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$; note $1 = -1$.

Remark 1.1.18. Let R be a ring.

1. $(R^\times,\cdot)$ is a group.

   Indeed: $1 \in R^\times$ is clear, and $x,y \in R^\times$ implies $xy \in R^\times$ and $x^{-1} \in R^\times$: $(xy)(y^{-1}x^{-1}) = (y^{-1}x^{-1})xy = 1$, so $xy \in R^\times$; and $x^{-1} \in R^\times$ because it has inverse x.
2. If $x \in R^\times$, then $-x \in R^\times$ and $(-x)^{-1} = -x^{-1}$ (by Remark 1.1.14 (4)).
3. No zero-divisor is a unit: if $xy = 0$ and $x \in R^\times$, then $x^{-1}xy = y = 0$.
4. If R is commutative and finite, every $x \in R \setminus \{0\}$ is a zero-divisor or a unit (exercise).
5. Thus, every finite integral domain is a field.
6. R is a field if and only if $(R \setminus \{0\},\cdot)$ is an abelian group (exercise).

Definition 1.1.19. Let R be a ring (field). A subset $S \subseteq R$ is a *subring (subfield) of* R if $0,1 \in S$ and for all $x,y \in S$ we have $x+y \in S, -x \in S, xy \in S$ (and $x^{-1} \in S$ if $x \neq 0$).

Remark 1.1.20. Then S is a ring (field) with $\cdot, +$ the operations of R (restricted to S^2).

Definition 1.1.21. Let R,S be rings (fields). A *ring (field) homomorphism from* R *to* S is a function $\varphi : R \to S$ that *preserves* $+,\cdot$ *and* 1, i.e., for all $x,y \in R$

$$\varphi(x +_R y) = \varphi(x) +_S \varphi(y), \quad \varphi(x \cdot_R y) = \varphi(x) \cdot_S \varphi(y), \quad \varphi(1_R) = 1_S.$$

The *kernel of* φ is

$$\ker(\varphi) := \{x \in R \mid \varphi(x) = 0_S\}.$$

Injective, surjective, bijective homomorphisms are again called *mono-, epi-, isomorphisms*, and R,S are *isomorphic*, symbolically $R \cong S$, if there exists an isomorphism $\varphi : R \to S$.

For $R = S$, homo-, isomorphisms are *endo-, automorphisms* of R; the set of automorphisms of R is denoted $\mathrm{Aut}(R)$.

Remark 1.1.22.

1. φ preserves 0, and $-$ and $\cdot^{-1}$ if defined (Remark 1.1.9). Hence, $\varphi(R^\times) \subseteq S^\times$.
2. If R is a field and S not trivial, then φ is injective. In particular, field homomorphisms are injective.

   Indeed: if $\varphi(x) = \varphi(y)$, then $0_S = \varphi(x) -_S \varphi(y) = \varphi(z)$ for $z := x -_R u$. Hence $\varphi(z) \notin S^\times$ (here we use $0_S \neq 1_S$), so $z \notin R^\times = R \setminus \{0_R\}$, i.e., $z = 0_R$, so $x = y$.

Exercise 1.1.23. Let $\varphi : R \to S$ be a (ring, field) homomorphism.

1. If U is a subring (-field) of R, then $\varphi(U)$ is a subring (-field) of S.
2. If V is a subring (-field) of S, then $\varphi^{-1}(V)$ is a subring (-field) of R.
3. φ is injective if and only if $\ker(\varphi) = \{0_R\}$.
4. If ψ is homomorphism from S to S', then $\psi \circ \varphi$ is one from R to S'.

Exercise 1.1.24. Let R be a group, ring or field. The set of endomorphisms of R together with composition $(\varphi, \psi) \mapsto \varphi \circ \psi$ is a monoid with neutral element id_R. The invertible elements are exactly the automorphisms of R. $(\mathrm{Aut}(R), \circ)$ is a group.

Definition 1.1.25. An *ordered field* is a tuple $(K, +, \cdot, <)$ such that $(K, +, \cdot)$ is a field and $<$ a linear order on K such that for all $x, y, z \in K$:

1. *compatible with $+$:* if $x < y$, then $x + z < x + z$;
2. *compatible with $\cdot$:* if $x < y$ and $0 < z$, then $x \cdot z < y \cdot z$.

K is *archimedian* if for all $x \in K$ there is $n \in \mathbb{N}$ such that $x < \underline{n}$; here, for $n \in \mathbb{N}$ we set $\underline{0} := 0_K$ and $\underline{n+1} := \underline{n} +_K 1_K$.

Exercise 1.1.26. Let K be an ordered field and $x, y \in K$.

1. If $x \neq 0$, then $0 < x^2$. Hence $0 < 1$ and $x^2 \neq -1$. Further, $0 < 1 < \underline{2} < \underline{3} < \ldots$.
2. If $0 < x < y$, then $0 < y^{-1} < x^{-1}$.
3. $<$ is *dense*: for all $x, y \in K$ with $x < y$ there is $z \in K$ such that $x < z < y$.

Example 1.1.27. $\mathbb{Q}, \mathbb{R}$ with the usual order $<$ are archimedian odered fields. By Exercise 1.1.26 (1) there is no linear order $<$ on $\mathbb{C}$ that would make $\mathbb{C}$ an ordered field.

Remark 1.1.28 (Nonstandard analysis)**.** Basic methods of mathematical logic allow to extend $\mathbb{R}$ to a non-archimedian ordered field $\mathbb{R}^*$ that is in a precise sense very similar to $\mathbb{R}$. In $\mathbb{R}^*$ there are *infinitessimals*, elements Δ such that $0 < \Delta < 1/n$ for all $n \in \mathbb{N}$.

1.2 Naturals

We assume that there is a triple $(\mathbb{N}, s, 0)$ with $\mathbb{N}$ a set, $0 \in \mathbb{N}$ and $s : \mathbb{N} \to \mathbb{N}$ a function such that the *Peano axioms* are satisfied:

(P0) s is injective;

(P1) $s(n) \neq 0$ for all $n \in \mathbb{N}$;

(P2) *(induction)* every $X \subseteq \mathbb{N}$ that contains 0 and is s-closed, equals $\mathbb{N}$.

That X is *s-closed* means $s(n) \in X$ for all $n \in X$. We write $1 := s(0), 2 := s(1), \ldots$

Remark 1.2.1. The elements of $\mathbb{N}$ are *natural numbers*. A proof of the existence of such a triple is given in any introductory course of set theory. This can be reasonably explained only in an axiomatic framework and is thus omitted here. To spark some curiosity:

$$0 := \{\},\ 1 := \{0\} = \{\{\}\},\ 2 := \{0,1\} = \{\{\},\{\{\}\}\},\ 3 := \{0,1,2\} = \{\{\},\{\{\}\},\{\{\},\{\{\}\}\}\},\dots$$

We are not interested in what the triple exactly is because they all look the same:

Theorem 1.2.2 (Categoricity)**.** *If $(N, 0', s')$ satisfies the Peano axioms, then there is a unique $\pi : \mathbb{N} \to N$ that* preserves 0 and s*, that is, $\pi(0) = 0'$ and $\pi(s(n)) = s'(\pi(n))$ for all $n \in \mathbb{N}$; this π is bijective.*

Proof. Uniqueness: if π' is another such map, consider $X := \{n \in \mathbb{N} \mid \pi(n) = \pi'(n)\}$. Then $0 \in X$ and if $n \in X$ then $\pi(s(n) = s'(\pi(n)) = s'(\pi'(n)) = \pi'(s(n))$, so $s(n) \in X$. By induction, $X = \mathbb{N}$. Thus $\pi = \pi'$.

Existence: call a set $F \subseteq \mathbb{N} \times N$ *good* if $(0, 0') \in F$ and for all $n \in \mathbb{N}$: if $(n, a) \in F$, then $(s(n), s'(a)) \in F$. E.g., $\mathbb{N} \times N$ is good. The intersection π of all good sets, is good - so π is the smallest good set.

π is a function on $\mathbb{N}$: let X be the set of $n \in \mathbb{N}$ such that $(n, a) \in \pi$ for exactly one $a \in N$. We prove $X = \mathbb{N}$ be induction. Note $0 \in X$: if $(0, a) \in \pi$ for $a \neq 0'$, then deleting $(0, a)$ from π is good – contradiction to π being smallest. We show X is s-closed: let $n \in X$, say $(n, a) \in \pi$; then $(s(n), s'(a)) \in \pi$ being good; as before $(s(n), b) \notin \pi$ for $b \neq s'(a)$.

π is as desired: for $n \in \mathbb{N}$ we have $(n, \pi(n)) \in \pi$, so $(s(n), s'(\pi(n))) \in \pi$ as π is good; this means $\pi(s(n)) = s'(\pi(n))$.

π is injective: let X be the set of $n \in \mathbb{N}$ such that $\pi(n) \neq \pi(m)$ for all $m \in \mathbb{N} \setminus \{n\}$. We claim $0 \in X$. To see this, let $Y := \{m \in \mathbb{N} \mid \pi(0) \neq \pi(m)\} \cup \{0\}$. Then $0 \in Y$ and if $m \in Y$, then $\pi(s(m)) = s'(\pi(m)) \neq 0' = \pi(0)$ (with $\neq$ by (P2)), so $s(m) \in Y$. By induction, $Y = \mathbb{N}$. Thus, $0 \in X$. By induction, it suffices to show $s(n) \in X$ for $n \in X$. Let

$$Z := \{m \in \mathbb{N} \mid \pi(m) \neq \pi(s(n))\} \cup \{s(n)\}.$$

We prove $Z = \mathbb{N}$ by induction. Note $0 \in Z$ because $\pi(0) = 0' \neq s'(\pi(n)) = \pi(s(n))$ (with $\neq$ by (P2)). Assume $m \in Z$. If $s(m) = s(n)$, then $s(m) \in Z$. Otherwise $n \neq m$, so $\pi(n) \neq \pi(m)$ as $n \in X$; then $\pi(s(m)) = s'(\pi(m)) \neq s'(\pi(n)) = \pi(s(n))$ (with $\neq$ by (P0)), so $s(m) \in Z$.

π is surjective: the image of π contains $0'$. If it contains n', say $\pi(n) = n'$, then also $s'(n')$ because $\pi(s(n)) = s'(\pi(n)) = s'(n')$. By induction (in N'), the image equals N'. □

Exercise 1.2.3. For every $n \in \mathbb{N} \setminus \{0\}$ there is a unique $m \in \mathbb{N}$ with $s(m) = n$.

Lemma 1.2.4. *There is a unique function $+ : \mathbb{N}^2 \to \mathbb{N}$ such that $n + 0 = n$ and $n + s(m) = s(n + m)$ for all $n, m \in \mathbb{N}$. It is called* addition (on $\mathbb{N}$).

Proof. It suffices to show for every $n \in \mathbb{N}$ that there is a unique function $f_n : \mathbb{N} \to \mathbb{N}$ with $f_n(0) = n$ and $f_n(s(m)) = s(f_n(m))$ for all $n \in \mathbb{N}$. Then the union of the f_n's is a function $+$ as desired. Any $+'$ as desired equals f_n when the first argument is fixed to n, so $+' = +$.

Let $n \in \mathbb{N}$. Existence is proved as in Theorem 1.2.2. Uniqueness: if f' is another such function, then $f_n(0) = f'(0)$ and, if $f(m) = f'(m)$, then $f_n(s(m)) = s(f_n(m)) = s(f'(m)) = f'(s(m))$; hence, $f = f'$ by induction . □

Proposition 1.2.5 (Properties of addition)**.** $(\mathbb{N}, +)$ *is a commutative monoid with neutral element* 0*. Further, for all* $n, m, k \in \mathbb{N}$*:*

(cancelling) $n + k = m + k$ *implies* $n = m$*.*

Proof. Associativity: let X be the set of k such that $n + (m + k) = (n + m) + k$ for all $n, m \in \mathbb{N}$. We show $X = \mathbb{N}$ by induction. Clearly, $0 \in X$. If $k \in X$, then $n + (m + s(k)) = n + s(m + k) = s(n + (m + k)) = s((n + m) + k) = (n + m) + s(k)$,so $s(k) \in X$.

Commutativity: let X be the set of $n \in \mathbb{N}$ such that $n + m = m + n$ for all $m \in \mathbb{N}$. To see $0 \in X$ is suffices to show $Y := \{m \in \mathbb{N} \mid 0 + m = m\}$ equals $\mathbb{N}$. But $0 \in Y$ and, if $m \in Y$, then $0 + s(m) = s(0 + m) = s(m)$, so $s(m) \in Y$. We are left to show $s(n) \in X$ for $n \in X$, i.e., $Z := \{m \in \mathbb{N} \mid s(n) + m = m + s(n)\}$ equals $\mathbb{N}$. But $0 \in Z$ because $0 \in X$. If $m \in Z$, then

$$\begin{aligned} s(n) + s(m) &= s(s(n) + m) \overset{m \in Z}{=} s(m + s(n)) = s(s(m + n)) \overset{n \in X}{=} s(s(n + m)) \\ &= s(n + s(m)) \overset{n \in X}{=} s(s(m) + n) = s(m) + s(n). \end{aligned}$$

0 is neutral because $0 + n = n + 0$ by commutativity and $n + 0 = n$ by definition of $+$.

Cancelling: let X be the set of $k \in \mathbb{N}$ such that the implication holds for all $n, m \in \mathbb{N}$. Clearly, $0 \in X$. Assume $k \in X$ and $n + s(k) = m + s(k)$. Then $s(n + k) = s(m + k)$. Since s is injective, $n + k = m + k$. Since $k \in X$ this implies $n = m$. □

Similarly as in Lemma 1.2.4 one verifies:

Lemma 1.2.6. *There is a unique function* $\cdot : \mathbb{N}^2 \to \mathbb{N}$ *such that* $n \cdot 0 = 0$ *and* $n \cdot s(m) = n \cdot m + n$ *for all* $n, m \in \mathbb{N}$*. It is called* multiplication (on $\mathbb{N}$).

Example 1.2.7. $2 + 2 = 2 + s(s(0)) = s(2 + s(0)) = s(s(2 + 0)) = s(s(2)) = s(3) = 4$,
$2 \cdot 2 = 2 \cdot s(s(0)) = 2 \cdot s(0) + 2 = (2 \cdot 0 + 2) + 2 = (0 + 2) + 2 = 2 + 2 = 4$.

Proposition 1.2.8 (Properties of multiplication)**.** $(\mathbb{N}, \cdot)$ *is a commutative monoid with neutral element* 1*. Further, for all* $n, m, k \in \mathbb{N}$ *we have*

1. (cancellation) $n \cdot k = m \cdot k$ *and* $k \neq 0$ *implies* $n = m$*;*
2. (distributivity) $k \cdot (n + m) = k \cdot n + k \cdot m$.

Proof. (Left-)distributivity: let X be the set of $m \in \mathbb{N}$ such that $k(n + m) = kn + km$ for all $k, n \in \mathbb{N}$. Clearly, $0 \in X$. Assume $m \in X$. Then $k(n + s(m)) = ks(n + m) = k(n + m) + k = (kn + km) + k = kn + (km + k) = kn + ks(m)$.

Associativity: let X be the set of $k \in \mathbb{N}$ such that $(nm)k = n(mk)$. Clearly, $0 \in X$. Assume $k \in X$. Then, using distributivity,

$$(nm)s(k) = (nm)k + nm \overset{k \in X}{=} n(mk) + nm = n(mk + m) = n(ms(k)).$$

1 is neutral: note $ns(0) = n0 + n = 0 + n = n$. Let X be the set of $n \in \mathbb{N}$ such that $1n = n$. Then $0 \in X$. Assume $n \in X$. Then $1 \cdot s(n) = 1 \cdot n + 1 \overset{n \in X}{=} n + s(0) = s(n)$.

Right-distributivity: we show $(n+m)k = nk+mk$ for all $n, m, k \in \mathbb{N}$.Let X be the set of k such that this holds for all n, m. Assume $k \in X$. Then $ms(k) = mk + m \overset{k\in X}{=} km + m \overset{k\in X}{=} (k+1)m = s(k)m$. Then $(n+m)s(k) = (n+m)k + (n+m) \overset{k\in X}{=} (nk+mk) + (n+m)$. Since $+$ is commutative and associative, this equals $(nk+n) + (mk+m) = ns(k) + ms(k)$.

Commutativity: let X be the set of all k such that $nk = kn$ for all $n \in \mathbb{N}$. To see $0 \in X$ note $0 \cdot 0 = 0$ and, assuming $0 \cdot \ell = 0$ and using $\ell + 1 = s(\ell + 0) = s(\ell)$, we get $0 \cdot s(\ell) = 0(\ell+1) = 0 \cdot \ell + 0 \cdot 1 = 0 + (0 \cdot 0 + 0) = 0$.

Now, if $k \in X$, then $ns(k) = nk + n = kn + n = (k+1)n = s(k)n$ using right-distributivity.

We leave the verification of cancellation as a (non-trivial) exercise. □

Proposition 1.2.9 (Properties of order)**.** *For $n, m \in \mathbb{N}$ let $n < m$ if $n + k = m$ for some $k \in \mathbb{N} \setminus \{0\}$. Then $<$ is a linear order on $\mathbb{N}$ that is compatible with $+, \cdot$ (see Definition 1.1.25) and has minimal element 0.*

Proof. That $0 < n$ for all $n \neq 0$ is trivial. Irreflexivity: let X be the set of $n \in \mathbb{N}$ such that $n + k \neq n$ for all $k \neq 0$. We show $X = \mathbb{N}$. Clearly, $0 \in X$. Let $n \in X$ and $k \neq 0$. Then $s(n) + k = k + s(n) = s(k + n) = s(n + k) \neq s(n)$ since $n + k \neq n$ by $n \in X$ and s is injective.

Transitivity: if $n+k = m$ and $m+k' = \ell$ for $k, k' \in \mathbb{N} \setminus \{0\}$, then $n+(k+k') = (n+k)+k' = m + k' = \ell$. By Exercise 1.2.3, $k' = s(k'')$ for some $k'' \in \mathbb{N}$, so $k + k' = s(k + k'') \neq 0$ by (P2).

Linearity: by irreflexivity and transitivity at most one of $n < m, n = m, m < n$ holds. Let X be the set of n such that at least one holds for all m. Then $0 \in X$ being minimal. Assume $n \in X$. We show $s(n) \in X$ by induction on m. Let Y be the set of $m \in \mathbb{N}$ such that $s(n) < m, s(n) = m$ or $m < s(n)$. Then $0 \in Y$ because $0 < s(n)$. Assume $m \in Y$. Note $n < m, n = m$ or $m < n$, since $n \in X$ These cases imply respectively $s(n) < s(m), s(n) = s(m), s(m) < s(m)$; e.g., if $n < m$, say $n + k = m$ with $k \neq 0$, then $s(n) + k = k + s(n) = s(k + n) = s(n + k) = s(m)$. Hence $s(m) \in Y$.

Compatibility with $+$: let X be the set of $\ell \in \mathbb{N}$ such that $\ell = 0$ or $m + \ell < n + \ell$ holds for all $n < m$ and $\ell \neq 0$. Then $0 \in X$. If $\ell \in X$, then $n + s(\ell) = s(n + \ell) < s(m + \ell) = m + s(\ell)$ is implied by $n + \ell < m + \ell$ as above. Thus, $<$ is compatible with $+$.

We leave compatibility with $\cdot$ as a (now easy) exercise. □

Exercise 1.2.10. For all $n, m \in \mathbb{N}$ wer have: $m < s(n)$ if and only if $m = n$ or $m < n$.

Theorem 1.2.11 (Least number principle)**.** *Every nonempty $X \subseteq \mathbb{N}$ has a minimal element $n \in X$, i.e., $n < m$ for all $m \in X \setminus \{n\}$.*

Proof. Let Y be the set of $n \in \mathbb{N}$ such that $m \notin X$ for all $m < n$. Then $0 \in Y$ by (4) above. As $X \neq \emptyset$, $Y \neq \mathbb{N}$. By induction, Y is not s-closed, i.e., there is $n \in Y$ with $s(n) \notin Y$. Then there exists $m < s(n)$ in X. Then $m = n$ or $m < n$ by Exercise 1.2.10. The latter is impossible as $n \in Y$. Hence $m = n \in X$. Then n is a minimal element of X: if $m \in X \setminus \{n\}$, then $m < n$ or $n < m$; the former is impossible, as $n \in Y$. □

Remark 1.2.12. This section exemplifies what one does in *mathematical logic*: reducing a given body of mathematics to fundamental reasoning principles. Here, the principle is induction, or equivalently, the least number principle. As seen, the reduction often amounts to tedious and somewhat boring work. One interest of mathematical logic is to compare the relative logical strength of the various thereby identified reasoning principles.

1.3 Integers

Idea: $\mathbb{Z}$ enlarges $\mathbb{N}$ by providing solutions of $n + X = 0$. We know $z \in \mathbb{Z}$ equals $n - m$ for some $(n, m) \in \mathbb{N}^2$. This pair is not unique, so z corresponds to the set of such pairs (n', m') with the same "difference", i.e., $n' - m' = n - m$. We can write this as $n' + m = n + m'$, avoiding the not yet explained $-$.

Lemma 1.3.1. *Define* $\sim \subseteq (\mathbb{N}^2)^2$ *setting for* $n, m, n', m' \in \mathbb{N}$*:*

$$(n, m) \sim (n', m') \iff n + m' = n' + m.$$

Then $\sim$ *is an equivalence relation on* $\mathbb{N}^2$*.*

Proof. Reflexivity and symmetry are trivial. For transitivity let $(n_0, m_0) \sim (n_1, m_1) \sim (n_2, m_2)$, i.e., $n_0 + m_1 = n_1 + m_0$ and $n_1 + m_2 = n_2 + m_1$. Then $n_0 + m_1 + n_1 + m_2 = n_1 + m_0 + n_2 + m_1$. Cancellation (see Proposition 1.2.5) $(n_1 + m_1)$ gives $n_0 + m_2 = n_2 + m_0$, i.e., $(n_0, m_0) \sim (n_1, m_1)$. □

Definition 1.3.2. $\mathbb{Z}$ is the set of equivalence classes $[n, m]$ of $(n, m) \in \mathbb{N}^2$ under $\sim$. Its elements are *called integers.*

Definition 1.3.3. For $x, y \in \mathbb{Z}$, say $x = [k, \ell], y = [n, m]$ with $(k, \ell), (n, m) \in \mathbb{N}$, set

$$x + y := [k + n, \ell + m].$$

Remark 1.3.4. This is well-defined: assume $[k, \ell] = [k', \ell']$ and $[n, m] = [n', m']$. Then $k + \ell' = k' + \ell$ and $n + m' = n' + m$. Then $(k + n) + (m' + \ell') = (m + \ell) + (k' + n')$, so $[k + n, m + \ell] = [k' + n', m' + \ell']$.

Proposition 1.3.5. $(\mathbb{Z}, +)$ *is an abelian group with neutral element* $[0, 0]$*.*

Proof. Commutativity: $[k, \ell] + [n, m] = [k + n, \ell + m] = [n + k, m + \ell] = [n, m] + [k, \ell]$. Associativity is similar. $[0, 0]$ is neutral because $[0, 0] + [n, m] = [n + 0, m + 0] = [n, m] = [0+n, 0+m] = [0, 0]+[m, n]$. The inverse of $[n, m]$ is $[m, n]$ because $[n+m, m+n] = [0, 0]$. □

Multiplication is straightforwardly defined: if we already knew what it is, then we could note $(k - \ell)(n - m) = (kn + \ell m) - (km + \ell n)$. We use this equation as a definition:

Definition 1.3.6. For $x, y \in \mathbb{Z}$, say $x = [k, \ell], y = [n, m]$ with $(k, \ell), (n, m) \in \mathbb{N}$, set

$$x \cdot y := [kn + \ell m, km + \ell n].$$

Remark 1.3.7. This is well-defined: let $x = [k, \ell] = [k', \ell']$, $y = [n, m] = [n', m']$, i.e.,

$$k + \ell' = k' + \ell, \quad n + m' = n' + m.$$

We have to show $[kn + \ell m, km + \ell n] = [k'n' + \ell'm', k'm' + \ell'n']$, i.e.,

$$kn + \ell m + k'm' + \ell'n' = k'n' + \ell'm' + km + \ell n.$$

This follows by cancellation (see Proposition 1.2.5) from a simple tricky calculation:

$$\begin{aligned}&(kn+\ell m+k'm'+\ell'n')+(\ell'n+k'm+k'n+\ell'm)\\&=(k+\ell')n+(\ell+k')m+k'(m'+n)+\ell'(n'+m)\\&=(\ell+k')n+(k+\ell')m+k'(n'+m)+\ell'(n+m')\\&=(k'n'+\ell'm'+km+\ell n)+(\ell'n+k'm+k'n+\ell'm).\end{aligned}$$

Theorem 1.3.8. $(\mathbb{Z},+,\cdot)$ *is an integral domain.*

Proof. We first show $(\mathbb{Z},\cdot)$ is a commutative monoid with neutral element $[1,0]$. For neutrality, note $[1,0]\cdot[k,\ell]=[1\cdot k+0\cdot\ell,0\cdot k+1\cdot\ell]=[k,\ell]$. Commutativity: $[k,\ell]\cdot[n,m]=[kn+\ell m,km+\ell n]=[nk+m\ell,mk+n\ell]=[n,m]\cdot[k,\ell]$. Associativity follows similarly from properties of $\mathbb{N}$. Distributivity:

$$\begin{aligned}&([k,\ell]+[n,m])\cdot[r,s]=[k+n,\ell+m]\cdot[r,s]=[rk+rn+s\ell+sm,r\ell+rn+sk+sn]\\&\quad=[rk+s\ell,r\ell+sk]+[rn+sm,rn+sn]=[k,\ell]\cdot[r,s]+[n,m]\cdot[r,s].\end{aligned}$$

By Proposition 1.3.5, $(\mathbb{Z},+,\cdot)$ is a commutative ring. We are left to show that $x\cdot y=[0,0]$ implies $x=[0,0]$ or $y=[0,0]$. Say, $x=[k,\ell],y=[n,m]$, and assume $y\neq[0,0]$, i.e., $n\neq m$, and $[0,0]=[kn+\ell m,km+\ell n]$, i.e., $kn+\ell m=km+\ell n$.

By Proposition 1.2.9, $n<m$ or $m<n$. We assume $n<m$ (the other case is analogous) and write $n+r=m$ for some $r\neq 0$. Plugging this in our assumption above gives $kn+\ell(n+r)=k(n+r)+\ell n$. Using commutativity and cancellation for $+$ (in $\mathbb{N}$), this implies $\ell r=kr$. Cancellation for $\cdot$ (in $\mathbb{N}$) gives $\ell=k$. Thus, $x=[k,\ell]=[0,0]$. □

Remark 1.3.9.

1. Consider the injection $n\mapsto[n,0]$ from $\mathbb{N}$ into $\mathbb{Z}$. Then $+$ (on the image of this map) as defined in $\mathbb{Z}$ extends $+$ as defined in $\mathbb{N}$. E.g., $[4,0]+[5,0]=[4+5,0+0]=[9,0]$.
2. More precisely, $n\mapsto[n,0]$ is a monoid monomorphism both from $(\mathbb{N},+)$ to $(\mathbb{Z},+)$, and from $(\mathbb{N},\cdot)$ to $(\mathbb{Z},\cdot)$.

   Indeed: the map is clearly injective. As $0\mapsto[0,0]$ and $1\mapsto[1,0]$ the neutral elements are preserved. We leave preservation of $+$ to the reader. Preservation of $\cdot$: $[k,0]\cdot[n,0]=[k\cdot n+0\cdot 0,0\cdot n+k\cdot 0]=[k\cdot n,0]$.
3. We "identify" n with $[n,0]$ and, somewhat sloppily, view $\mathbb{N}$ as a subset of $\mathbb{Z}$.

Compatibility of order is explained as in Definition 1.1.25:

Proposition 1.3.10. *For $x,y\in\mathbb{Z}$ let $x<y$ if and only if there is $n\in\mathbb{N}\setminus\{0\}$ such that $x+n=y$ (i.e., $a+[n,0]=b$). Then $<$ is a linear order on $\mathbb{Z}$ that is compatible with $+,\cdot$.*

Proof. We only show compatibility. Assume $x+n=y$ for $n\in\mathbb{N}\setminus\{0\}$. Then $y+z=x+z+n$, so $x+z<y+z$. Assume $0<z$, i.e., $z\in\mathbb{N}\setminus\{0\}$. Then $yz=xz+nz$, so $xz<yz$ as $nz\in\mathbb{N}\setminus\{0\}$. □

Remark 1.3.11. Clearly, $<$ as defined in $\mathbb{Z}$ extends $<$ as defined in $\mathbb{N}$. More precisely, the monomorphism $n\mapsto[n,0]$ *preserves* $<$: $n<m$ in $\mathbb{N}\iff[n,0]<[m,0]$ in $\mathbb{Z}$.

1.4 Rationals

Idea: we want to enlarge $\mathbb{Z}$ to add solutions of $aX = 1$ for $a \neq 0$. We know every $x \in \mathbb{Q}$ equals a/b for some $(a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$. This pair is not unique, so z corresponds to the set of such pairs (a', b') with the same "fraction", i.e., $a/b = a'/b'$. We can write this as $ab' = a'b$, avoiding the not yet explained $/$.

Lemma 1.4.1. *Define* $\sim \subseteq (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))^2$ *setting:*

$$(a, b) \sim (a', b') \iff ab' = a'b.$$

Then $\sim$ *is an equivalence relation.*

Proof. Reflexivity and symmetry are trivial. For transitivity, assume $(a_0, b_0) \sim (a_1, b_1) \sim (a_2, b_2)$, i.e., $a_0b_1 = a_1b_0$ and $a_1b_2 = a_2b_1$.

Then $a_0b_1a_1b_2 = a_1b_0a_2b_1$, so $(a_0b_2 - b_0a_2) \cdot (a_1b_1) = 0$. Assume first that $a_1 \neq 0$. As $b_1 \neq 0$ we have $a_1b_1 \neq 0$ since $\mathbb{Z}$ is an integral domain. Then $(a_0b_2 - b_0a_2) = 0$, i.e., $(a_0, b_0) \sim (a_2, b_2)$.

Now assume $a_1 = 0$. Then $a_0b_1 = 0 = a_2b_1$. Since $\mathbb{Z}$ is an integral domain and $b_1, b_2 \neq 0$ we get $a_0 = a_2 = 0$. Hence $a_0b_2 = a_2b_0 (= 0)$, so $(a_0, b_0) \sim (a_2, b_2)$. □

Definition 1.4.2. $\mathbb{Q}$ is the set of equivalence classes a/b of $(a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ under $\sim$. Its elements are called *rationals*.

Addition and multiplication are straightforwardly defined: if we already knew the rationals we could note $a/b + c/d = (ad + cb)/bd$ and $a/b \cdot c/d = ac/bd$ (we read e.g. ac/bd as $(ac)/(bd)$). We use these equations as definitions:

Definition 1.4.3. For $x, y \in \mathbb{Q}$, say $x = a/b, y = c/d$ with $(a, b), (c, d) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, set

$$x + y := (ad + cb)/bd, \quad x \cdot y := ac/bd.$$

Further, set $x < y$ if and only if $x + z = y$ for some *positive* $z \in \mathbb{Q}$, i.e., $z = e/f$ for some $0 < e, f \in \mathbb{Z}$.

Remark 1.4.4. These are well-defined. Assume $x = a/b = a'/b'$ and $c/d = c'/d'$, i.e., $y = ab' = a'b$ and $cd' = c'd$.

For $+$ we have to show $(ad + cb)/bd = (a'd' + c'b')/b'd'$. This is true:

$$(ad + cb)b'd' = ab'dd' + bb'cd' = a'bdd' + bb'c'd = (a'd' + b'c')bd.$$

For $\cdot$ we have to show $ac/bd = a'c'/b'd'$. This is true: $acb'd' = ab'cd' = a'bc'd$.

For $<$ assume $x + e/f = y$ with $0 \leqslant e, f$. Then $a/b + e/f = (af + eb)/bf = c/d$, i.e., $afd + ebd = cbf$. We claim $a'/b' + e/f = c'/d'$, i.e., $a'fd' + eb'd' = c'b'f$. Argue

$$\begin{aligned} afd \cdot a'b'c'd' + ebd \cdot a'b'c'd' &= cbf \cdot a'b'c'd' \\ a'fd' \cdot ab'c'd + eb'd' \cdot a'bc'd &= c'b'f \cdot a'bcd' \\ a'fd \cdot a'bc'd' + eb'd' \cdot a'bc'd &= c'b'f \cdot a'bc'd. \end{aligned}$$

The 2nd is only re-arranging the 1st, the 3rd uses $ab' = a'b$ on the left, and $cd' = c'd$ on the right. It implies our claim by distributivity and cancellation in $\mathbb{Z}$.

Remark 1.4.5. $x \in \mathbb{Q}$ is positive if and only if $0 < x$ in $\mathbb{Q}$.

Theorem 1.4.6. $(\mathbb{Q}, +, \cdot, <)$ *is an archimedian ordered field.*

Proof. It is easily checked that $(\mathbb{Q}, +)$ and $(\mathbb{Q} \setminus \{0/1\}, \cdot)$ are associative and commutative. The neutral elements are $0/1$ and $1/1$ respectively: $a/b + 0/1 = (a \cdot 1 + 0 \cdot b)/b \cdot 1 = a/b$ and $a/b \cdot 1/1 = a \cdot 1/b \cdot 1 = a/b$. The additive inverse $-(a/b)$ of a/b is $(-a)/b$ because $a/b + (-a)/b = (ab + -(ab))/b = 0/b = 0/1$. The multiplicative inverse $(a/b)^{-1}$ of $a/b \neq 0/1$ is b/a – note $a \neq 0$ because otherwise $a/b = 0/b = 0/1$. Indeed, $a/b \cdot b/a = ab/ba = 1/1$.

For distributivity we use $a'/b' = a'c'/b'c'$ for $c' \neq 0$:

$$(a/b + c/d) \cdot e/f = (ad + cb)/bd \cdot e/f = (ade + cbe)/bdf = (ade\ bdf + cbe\ bdf)/(bdf\ bdf)$$
$$= ade/bdf + cbe/bdf = ae/bf + ce/df = a/b \cdot e/f + c/d \cdot e/f.$$

Concerning $<$ we only verify compatibiliy: if $x < y$ and $z \in \mathbb{Q}$, then $(y + z) - (x + z) = y - x$ is positive, so $y + z < y + z$. Note a product of positive rationals is positive. Hence, if $0 < z$, then $yz - xz = (y - x)z$ is positive, so $xz < yz$.

Archimedian: if x is not positive, then $x < 1$. Otherwise $x = a/b$ with $a, b \in \mathbb{Z}$ and $a, b > 0$. Then $x < a/1$ because $a/1 - a/b = (ab - a)/ab$ with $a(b-1), ab > 0$ in $\mathbb{Z}$. □

Proposition 1.4.7. *The map $a \mapsto a/1$ is a ring monomorphism from $\mathbb{Z}$ into $\mathbb{Q}$. Moreover, it* preserves $<$ *in the sense that for all $a, b \in \mathbb{Z}$:*

$$a < b \text{ in } \mathbb{Z} \iff a/1 < b/1 \text{ in } \mathbb{Q}.$$

Proof. The map is clearly injective. It preserves 1 because $1 \mapsto 1/1$. It preserves $+$ because $a/1 + b/1 = (a \cdot 1 + b \cdot 1)/1 \cdot 1 = (a + b)/1$. It preserves $\cdot$ because $a/1 \cdot b/1 = ab/1 \cdot 1 = ab/1$. Finally, it preserves $<$ because

$$b/1 - a/1 = (b \cdot 1 + (-a) \cdot 1)/1 \cdot 1 \text{ is positive} \iff 0 < b - a \text{ in } \mathbb{Z} \iff a < b \text{ in } \mathbb{Z}. \qquad \square$$

Notation: From now on we "identify" a with $a/1$ and, somewhat sloppily, view $\mathbb{Z}$ as a subset of $\mathbb{Q}$. By the above then $\mathbb{Z}$ is a subring of $\mathbb{Q}$.

Definition 1.4.8. The *absolute value of* $x \in \mathbb{Q}$ is $|x| := \begin{cases} x & \text{if } x \geqslant 0 \\ -x & \text{else.} \end{cases}$

Remark 1.4.9. For all $x, y \in \mathbb{Q}$ we have $|xy| = |x| \cdot |y|$, the *triangle inequality* $|x + y| \leqslant |x| + |y|$, and, $|x| = 0$ if and only if $x = 0$.

1.5 Reals

Write $\mathbb{Q}^+$ for the positive rationals and recall that a sequence $(q_n)_{n \in \mathbb{N}}$ of rationals has *limit* $q \in \mathbb{Q}$ if and only if for all $\epsilon \in \mathbb{Q}^+$ there is $n_0 \in \mathbb{N}$ such that $|q_n - q| < \epsilon$ for all $n > n_0$. Each sequence has at most one limit. If it exists, the sequence is *Cauchy*: for all $\epsilon \in \mathbb{Q}^+$ there is $n_0 \in \mathbb{N}$ such that $|q_n - q_m| < \epsilon$ for all $n, m > n_0$.

Idea: e.g. $X^2 = 2$ has no solution in $\mathbb{Q}$ but can be "approximated" by rationals. We know every real $x \in \mathbb{R}$ is the limit of a Cauchy sequence of rationals. The sequence is not unique, so every real corresponds to a set of Cauchy sequences, namely those with the same limit. We can say two Cauchy sequences have the same limit, without referring to this yet undefined limit, by stating that their difference has limit 0.

Lemma 1.5.1. *Let $\mathcal{C}$ be the set of rational Cauchy sequences. For $(q_n)_n, (p_n)_n \in \mathcal{C}$ set*

$$(q_n)_n \sim (p_n)_n \iff (q_n - p_n)_n \text{ has limit } 0.$$

Then $\sim$ is an equivalence relation on $\mathcal{C}$.

Proof. Reflexivity and symmetry are trivial. For transitivity, assume $(p_n)_n \sim (q_n)_n \sim (r_n)_n$, i.e., $(p_n - q_n)_n$ and $(q_n - r_n)_n$ have limit 0. Then $((p_n - q_n) - (q_n - r_n))_n = (p_n - r_n)_n$ has limit 0. □

Definition 1.5.2. $\mathbb{R}$ is the set of equivalence classes $\overline{(q_n)_n}$ of $(q_n)_n \in \mathcal{C}$ under $\sim$. Its elements are called *reals*. For reals $x = \overline{(q_n)_n}$ and $y = \overline{(p_n)_n}$ set

$$x + y := \overline{(q_n + p_n)_n}, \quad x \cdot y := \overline{(q_n \cdot p_n)_n}.$$

Further, define $x < y$ if and only if there are $\epsilon \in \mathbb{Q}^+$ and $n_0 \in \mathbb{N}$ such that $p_n - q_n > \epsilon$ (in $\mathbb{Q}$) for all $n > n_0$.

Remark 1.5.3. These are well-defined. We leave it as an exercise (in elementary analysis) to verify this for $+$ and $\cdot$.

For $<$, assume $x = \overline{(q_n)_n} = \overline{(q'_n)_n} < y = \overline{(p_n)_n} = \overline{(p'_n)_n}$. Choose $\epsilon \in \mathbb{Q}^+, n_0 \in \mathbb{N}$ such that $p_n - q_n > \epsilon$ for all $n > n_0$. We claim there is $n_1 \in \mathbb{N}$ such that $p'_n - q'_n > \epsilon/2$ for all $n > n_1$. Namely, choose $n_0 < n_1 \in \mathbb{N}$ such that $|q_n - q'_n| < \epsilon/4$ and $|p_n - p'_n| < \epsilon/4$ for all $n > n_1$. Then $p'_n - q'_n = (p'_n - p_n) + (p_n - q_n) + (q_n - q'_n) > -\epsilon/4 + \epsilon - \epsilon/4 = \epsilon/2$ for all $n > n_1$.

Theorem 1.5.4. $(\mathbb{R}, +, \cdot, <)$ *is an archimedian ordered field.*

Proof. It is clear that $+, \cdot$ are associative and commutative with neutral elements $\overline{(0, 0, ...)}$ and $\overline{(1, 1, ...)}$. Distributivity is also clear. The additive inverse of $\overline{(q_n)_n}$ is $\overline{(-q_n)_n}$. For the multiplicative inverse of $\overline{(q_n)_n} \neq \overline{(0, 0, ...)}$, note $(q_n)_n$ does not have limit 0. Being Cauchy, there are $\epsilon \in \mathbb{Q}^+$ and $n_0 \in \mathbb{N}$ such that $|q_n| > \epsilon$ for all $n > n_0$; let $(p_n)_n$ be defined by $p_n := 1$ if $n \leqslant n_0$, and $p_n := q_n^{-1}$ for $n > n_0$. Then $\overline{(q_n)_n} \cdot \overline{(p_n)_n} = \overline{(q_0, ..., q_{n_0}, 1, 1, ...)} = \overline{(1, 1, ...)}$.

We only show compatibility of $<$ with $\cdot$. Let $x = \overline{(q_n)_n} < y = \overline{(p_n)_n}$ and $z = \overline{(r_n)_n}$. Choose $\epsilon \in \mathbb{Q}^+, n_0 \in \mathbb{N}$ such that $p_n - q_n > \epsilon$ for all $n > n_0$. Then $p_n + r_n - q_n - r_n > \epsilon$ for all $n > n_0$, so $x + z < y + z$. Assume $z > 0$, so there are $\delta \in \mathbb{Q}^+, n_1 \in \mathbb{N}$ such that $r_n > \delta$ for all $n > n_1$. Then $p_n r_n - q_n r_n \geqslant \delta(p_n - q_n) > \delta\epsilon$ for all $n > n_1$. Hence, $xz < yz$.

Archimedian: let $x = \overline{(q_n)_n}$. Being Cauchy, there are $q \in \mathbb{Q}, n_0 \in \mathbb{N}$ such that $q_n < q$ for all $n > n_0$. By Theorem 1.4.6 there is $m \in \mathbb{N}$ such that $q < m$ in $\mathbb{Q}$. Then $x < \overline{(m, m, ...)}$. □

Remark 1.5.5.

1. $q \mapsto \overline{(q,q,...)}$ is a field monomorphism from $\mathbb{Q}$ to $\mathbb{R}$; moreover, it preserves $<$. We "identify" q with $\overline{(q,q,...)}$ and, somewhat sloppily, view $\mathbb{Q}$ as a subset of $\mathbb{R}$. Then $\mathbb{Q}$ is a subfield of $\mathbb{R}$.

   Indeed, preservation of $+,\cdot$ is clear, e.g., $\overline{(q,q,...)}\cdot\overline{(p,p,...)} = \overline{(qp,qp,...)}$. It is injective by Remark 1.1.22. It preserves $<$ because

   $$\overline{(q,q,...)} < \overline{(p,p,...)} \iff p-q \in \mathbb{Q}^+ \iff q < p \text{ in } \mathbb{Q}.$$

2. $|x|$ is defined for $x \in \mathbb{R}$ as in Definition 1.4.8; it has the properties in Remark 1.4.9.

For sequences $(x_n)_n$ in $\mathbb{R}$ we define limits in $\mathbb{R}$ and being Cauchy exactly as for $\mathbb{Q}$ (recall the beginning of this section).

Theorem 1.5.6. $\mathbb{R}$ *is* complete*: every Cauchy sequence in* $\mathbb{R}$ *has a limit in* $\mathbb{R}$.

Proof. Let $(x_n)_n$ be Cauchy (in $\mathbb{R}$). For $n \in \mathbb{N}$ write $x_n = \overline{(q_{nk})_k}$ for $(q_{nk})_k \in \mathcal{C}$. For $n \in \mathbb{N}$ choose $k_n \in \mathbb{N}$ such that $|x_n - q_{nk_n}| < 1/n$ for all $k \geqslant k_n$. Then $(q_{nk_n})_n \in \mathcal{C}$ and $(x_n)_n$ has limit $x := \overline{(q_{nk_n})_n}$: given $\epsilon \in \mathbb{Q}^+$, choose $n_0 \in \mathbb{N}$ such that $n_0 > 2/\epsilon$ (archimedian), so $1/n < \epsilon/2$ for all $n > n_0$. Choose $n_0 < n_1 \in \mathbb{N}$ such that $|q_{nk_n} - x| < \epsilon/2$ for all $n > n_1$. Then for all $n \geqslant n_1$:

$$|x_n - x| \leqslant |x_n - q_{nk_n}| + |q_{nk_n} - x| < 1/n + \epsilon/2 < \epsilon. \qquad \square$$

We now show that the order $<$ on $\mathbb{R}$ is determined by its field structure:

Lemma 1.5.7. *Every positive real is a square. Hence, for all* $x, y \in \mathbb{R}$, $x < y$ *if and only if* $x + z^2 = y$ *for some* $z \in \mathbb{R} \setminus \{0\}$.

Proof. (Sketch) Given $x > 0$ define a sequence $(q_n, p_n)_n$ of pairs of rationals such that $p_n^2 \leqslant x \leqslant q_n^2$ and $p_n - q_n < 2^{-n}$. Then $\overline{(q_n)_n}^2 = x$.

If $x < y$, then $x + z^2 = y$ where $z \in \mathbb{R} \setminus \{0\}$ is such that $z^2 = y - x > 0$. Conversely, $y - x = z^2 > 0$ by by Exercise 1.1.26 (1), so $x < y$. $\square$

Corollary 1.5.8. *The only automorphism of the field* $\mathbb{R}$ *is the identity* $\mathrm{id}_{\mathbb{R}}$.

Proof. Let φ be an automorphism. Then $\varphi(1) = 1, \varphi(2) = \varphi(1) + \varphi(1) = 1 + 1 = 2...$, so $\varphi(n) = n$ for all $n \in \mathbb{N}$, so also $\varphi(-n) = -n$ and $\varphi(1/n) = 1/n$. As every $q \in \mathbb{Q}$ equals $\pm n/m$ for some $n, m \in \mathbb{N}$ we have $\varphi(q) = q$ for all $q \in \mathbb{Q}$.

It should be clear that φ preserves $|\cdot|$ (i.e., $|x| = |\varphi(x)|$ for all $x \in \mathbb{R}$). By Lemma 1.5.7, φ preserves $<$: if $x + z^2 = y$ with $z \neq 0$, then $\varphi(x) + \varphi(z)^2 = \varphi(y)$ and $\varphi(z) \neq 0$.

Let $x \in \mathbb{R}$, say $x = \overline{(q_n)_n}$ and set $y := \varphi(x)$. Given $\epsilon \in \mathbb{Q}^+$ it suffices to show $|x - y| < \epsilon$ (then $|x - y| = 0$, so $x - y = 0$). Choose $n \in \mathbb{N}$ such that $|x - q_n| < \epsilon$. Then

$$\epsilon = \varphi(\epsilon) > \varphi(|x - q_n|) = |\varphi(x) - \varphi(q_n)| = |y - q_n|. \qquad \square$$

Remark 1.5.9. We showed our construction of the reals has a series of fundamental properties we expect the reals to have. The reader might still worry whether the "real" reals are "the same" as our $\mathbb{R}$. They are: every complete archimedian ordered field is isomorphic to $\mathbb{R}$ (think about why, the argument is similar to Corollary 1.5.8). "Complete archimedian" can be equivalently replaced by the statement that every non-empty upward bounded subset has a supremum. This statement plus the axioms of ordered fields thus constitute an elegant categorical axiomatization of the reals (as we had for $\mathbb{N}$).

1.6 Complex numbers

Idea: we intend to enlarge the field $\mathbb{R}$ to another $\mathbb{C}$ so that $X^2 = -1$ has a solution (recall Exercise 1.1.26), denoted i, so $i^2 = -1$. Then, for $x, x', y, y' \in \mathbb{R}$:

$$(x+iy)+(x'+iy') = (x+x')+i(y+y'),$$
$$(x+iy)\cdot(x'+iy') = (xx'+ixy'+iyx'+i^2yy') = (xx'-yy')+i(xy'+x'y).$$

To find such a field $\mathbb{C}$ we use these equations as a definition:

Definition 1.6.1. The set of *complex numbers* is $\mathbb{C} := \mathbb{R}^2$. For $z = (x,y), z' = (x',y')$ set

$$z+z' := (x+x', y+y'), \quad z\cdot z' := (xx'-yy', xy'+x'y).$$

For $z = (x,y)$ we call $\mathrm{Re}(z) := x$ and $\mathrm{Im}(z) := y$ the *real* and *imaginary part* of z. The *(complex) conjugation* is

$$z = (x,y) \mapsto \bar{z} := (x,-y).$$

The *absolute value of* $z = (x,y)$ is $|z| := \sqrt{z\bar{z}} = \sqrt{x^2+y^2}$.

Remark 1.6.2. We write $i := (0,1)$. Then $i^2 = (0,1)\cdot(0,1) = (0\cdot 0 - 1\cdot 1, 0\cdot 1 + 0\cdot 1) = (-1,0)$. This is $-1 \in \mathbb{R}$ after our "identification" of $x \in \mathbb{R}$ with $(x,0)$ to be explained below.

Theorem 1.6.3. $(\mathbb{C}, +, \cdot)$ *is a field and conjugation is an automorphism.*

Proof. Associativity, commutativity and distributivity are clear. The additive and multiplicative neutral elements are $(0,0)$ and $(1,0)$. The additive inverse of $z = (x,y)$ is $-z = (-x,-y)$. If $z \neq (0,0)$, then $x^2+y^2 \neq 0$, and its multiplicative inverse is

$$z^{-1} = (x/|z|^2, -y/|z|^2).$$

Indeed: $(x,y)\cdot(x/|z|^2, -y/|z|^2) = \big(xx/|z|^2 - y(-y)/|z|^2, x(-y)/|z|^2 + yx/|z|^2\big) = (1,0)$. We leave the verification that conjugation is an automorphism to the reader. □

Remark 1.6.4. $x \mapsto (x,0)$ is a field monomorphism from $\mathbb{R}$ to $\mathbb{C}$. Indeed: recall Remark 1.1.22 (3); the map preserves 1 and $+, \cdot$ because for all $x, x' \in \mathbb{R}$:

$$(x,0)+(x',0) = (x+x',0), \quad (x,0)\cdot(x',0) = (x\cdot x' - 0\cdot 0, x\cdot 0 + 0\cdot x') = (x\cdot x', 0).$$

We "identify" $x \in \mathbb{R}$ with $(x,0)$ and, somewhat sloppily, view $\mathbb{R}$ as a subset of $\mathbb{R}$. Then $\mathbb{R}$ is a subfield of $\mathbb{C}$. For $z = (x,y)$ we have $z = x + iy$, or, more precisely,

$$z = (x,y) = (x,0) + (0,1)\cdot(y,0).$$

Remark 1.6.5. The absolute value $|z|$ satisfies the properties in Remark 1.4.9.

For sequences $(z_n)_{n\in\mathbb{N}}$ in $\mathbb{C}$ having a *limit* and being *Cauchy* is explained as for rationals and reals (see beginning of Section 1.5). It has limit z if and only if both $(\mathrm{Re}(z_n))_n$ has limit $\mathrm{Re}(z)$ and $(\mathrm{Im}(z_n))_n$ has limit $\mathrm{Im}(z)$. Further, $(z_n)_{n\in\mathbb{N}}$ is Cauchy if and only if both $(\mathrm{Re}(z_n))_n$ and $(\mathrm{Im}(z_n))_n$ are. This implies:

Theorem 1.6.6. *Every Cauchy sequence in $\mathbb{C}$ has a limit in $\mathbb{C}$.*

We define exponentiation, sine and cosine on $\mathbb{C}$ via their power series. Convergence and basic properties are verified as in $\mathbb{R}$, known from calculus.

$$\begin{array}{rcll} e^z & := & 1+z+z^2/2!+z^3/3!+\cdots & = \sum_{k=0}^{\infty} z^k/k!, \\ \sin(z) & := & z-z^3/3!+z^5/5!-\cdots & = \sum_{k=0}^{\infty}(-1)^k z^{2k+1}/(2k+1)!, \\ \cos(z) & := & 1-z^2/2!+z^4/4!-\cdots & = \sum_{k=0}^{\infty}(-1)^k z^{2k}/(2k)!. \end{array}$$

Remark 1.6.7.

1. *Euler's identity* $e^{\pi i}=-1$ follows noting for $\alpha\in\mathbb{R}$:

$$e^{i\alpha}=1+i\alpha-\alpha^2/2!-i\alpha^3/3!+\alpha^4/4!+i\alpha^5/5!-\cdots=\cos(\alpha)+i\sin(\alpha).$$

   In particular, $|e^{i\alpha}|=\sqrt{\cos(\alpha)^2+\sin(\alpha)^2}=1$, i.e., all $e^{i\alpha}$ lie on the unit circle S^1.
2. (*Polar coordinates*) Every $z\in\mathbb{C}$ equals $|z|(\cos(\alpha)+i\sin(\alpha))$ for some unique $\alpha\in[0,2\pi)$, the *argument of* z. This becomes $z=|z|e^{i\alpha}$. For $z'=|z'|e^{i\alpha'}$, complex multiplication becomes $z\cdot z'=|z|\cdot|z'|\cdot e^{i(\alpha+\alpha')}$.
3. (*De Moivre*) For $n>1$ and any $z\in\mathbb{C}$ with argument α we have $w^n=z$ for n distinct

$$w \;=\; |z|^{1/n}e^{i\alpha/n},\;\; |z|^{1/n}e^{i\alpha/n}e^{2\pi i/n},\;\; |z|^{1/n}e^{i\alpha/n}e^{2\cdot 2\pi i/n},\ldots,|z|^{1/n}e^{i\alpha/n}e^{(n-1)2\pi i/n}.$$

Definition 1.6.8 (Roots of unity)**.** Let $n>1$ the *n-th roots of unity* are

$$C_n:=\left\{\zeta_n^0,\ldots,\zeta_n^{n-1}\right\}\quad\text{where }\zeta_n:=e^{2\pi i/n}.$$

Remark 1.6.9. C_n is a subgroup of the circle group S^1 (Example 1.1.12). Further,

$$1+\zeta_n+\zeta_n^2+\cdots+\zeta_n^{n-1}=(\zeta_n^n-1)/(\zeta_n-1)=0.$$

Example 1.6.10. $\zeta_2,\ldots,\zeta_6$ are $-1,\ (-1+\sqrt{-3})/2,\ i,\ \cos(2\pi/5)+i\sin(2\pi/5),\ (1+\sqrt{-3})/2$. E.g., one easily computes $\zeta_6^0,\zeta_6^1,\zeta_6^2,\zeta_6^3,\zeta_6^4,\zeta_6^5$ as

$$1,\ (1+i\sqrt{3})/2,\ (-1+i\sqrt{3})/2,\ -1,\ (-1-i\sqrt{3})/2,\ (1-i\sqrt{3})/2.$$

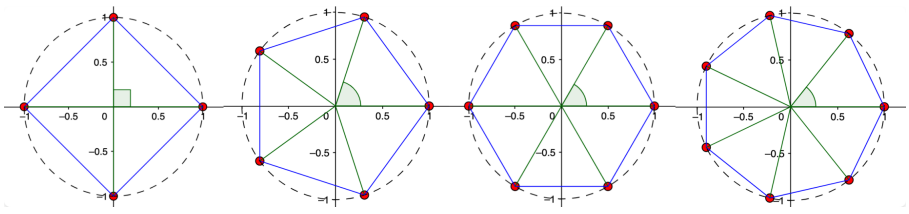


Figure 1.1: For $n=4,5,6,7$, C_n is marked red, the regular n-gon blue.

Chapter 2

Number theory

2.1 The Euclidian algorithm

Definition 2.1.1. Let $x, y \in \mathbb{Z}$. Then x is a *divisor* or *factor of* y, and y a *multiple of* x, symbolically $x \mid y$, if $x \cdot z = y$ for some $z \in \mathbb{Z}$. The negation is denoted $x \nmid y$.

Remark 2.1.2. For all $x, x', y, y', z, u, u' \in \mathbb{Z}$:

1. $x \mid 0, 1 \mid x, -1 \mid x, x \mid x, -x \mid x, x \mid -x$.
2. $\mid$ is transitive: if $x \mid y$ and $y \mid z$, then $x \mid z$.
3. If $x \mid y$ and $x \mid y'$, then $x \mid uy + u'y'$.
4. If $x \mid y$ and $x' \mid y'$, then $xx' \mid yy'$;
5. If $y \neq 0$ and $x \mid y$, then $1 \leqslant |x| \leqslant |y|$;

   Indeed: $y = xu$ implies $|y| = |x| \cdot |u|$ and $y \neq 0$ implies $x \neq 0$, so $|x| \neq 0$, so $|x| \geqslant 1$.
6. The divisors of 1 are ± 1, i.e., $\mathbb{Z}^\times = \{\pm 1\}$ (if $x \mid 1$, then $1 \leqslant |x| \leqslant 1$ by (5), so $|x| = 1$).
7. $x \mid y$ and $y \mid x$, if and only if, $x = y$ or $x = -y$.

   Indeed: $\Leftarrow$: by (1). $\Rightarrow$: if $y = 0$, then $x = 0$ by $y \mid x$; if $y \neq 0$, then $x \neq 0$ by $x \mid y$; by (5), $|x| \leqslant |y| \leqslant |x|$, so $|x| = |y|$.

Theorem 2.1.3 (Euclidian division)**.** *Let $x, y \in \mathbb{Z}$ and $y \neq 0$. Then there is a unique pair $(q, r) \in \mathbb{Z}^2$ with $x = qy + r$ and $0 \leqslant r < |y|$. Moreover, if $x \geqslant 0$ and $y > 0$, then $q \geqslant 0$.*

q is called the quotient *and r the* remainder *of (x, y); we agree both are 0 if $y = 0$.*

Proof. Uniqueness: assume $qy + r = q'y + r'$ with $0 \leqslant r, r' < |y|$. Then $(q - q')y = r' - r$, so $y \mid (r' - r)$. If $r \neq r'$, then by Remark 2.1.2 (5) $|y| \leqslant |r' - r| \leqslant \max\{r, r'\}$ (since $r, r' \geqslant 0$), a contradiction. So $r = r'$. Then $qy = q'y$ and $q = q'$ follow.

Existence: assume first $y > 0$. Let R be the set of $r \in \mathbb{N}$ such that $r = x - qy$ for some $q \in \mathbb{Z}$. Then $R \neq \varnothing$: if $x \geqslant 0$, then $x = x - 0y$, so $x \in R$; if $x < 0$, then $x - xy \in R$.

Let r be a minimal element of R. Then $r \geqslant 0$ and $x = qy + r$ for some $q \in \mathbb{Z}$. Further, $r < |y| = y$ because otherwise $r - y = x - (q + 1)y \in R$, contradicting the minimality of r.

Now assume $y < 0$. Then there are $q, r \in \mathbb{Z}$ such that $x = q'(-y) + r'$ and $0 \leqslant r' < |y| = -y$ and we can set $q := -q', r = r'$.

Moreover: if $y > 0 > q$, then $qy \leqslant -y$, so $x = qy + r \leqslant -y + r < 0$ (as $0 \leqslant r < y$). □

Remark 2.1.4. Call $x \in \mathbb{Z}$ *even* if $2 \mid x$ and otherwise *odd*. Since every $x \in \mathbb{Z}$ equals either $2q + 1$ or $2q$ for some $q \in \mathbb{Z}$, we see x is odd if and only if $x = y + 1$ for some even y.

Lemma 2.1.5. *Every subgroup of* $(\mathbb{Z}, +)$ *equals* $n\mathbb{Z} := \{nz \mid z \in \mathbb{Z}\}$ *for some* $n \in \mathbb{N}$.

Proof. Let U be a subgroup. If $U = \{0\}$, then $U = 0\mathbb{Z}$. Otherwise there is $x \in U \setminus \{0\}$. As $-x \in U$, $U \cap \mathbb{N} \setminus \{0\} \neq \varnothing$. Let n be the minimal element of $U \cap (\mathbb{N} \setminus \{0\})$.

We claim $U = n\mathbb{Z}$. $\supseteq$ is clear. $\subseteq$: let $x \in U$. Write $x = qn + r$ using Euclidian division, so $0 \leqslant r < n$. Then $qn \in U$, so $r = x - qn \in U$. Then $r = 0$ by choice of n, so $x = qn \in n\mathbb{Z}$. □

Definition 2.1.6. Let $n > 0$ and $x_1, ..., x_n \in \mathbb{Z}$ not all 0. Then $x \in \mathbb{Z}$ is a *common divisor* of $x_1, ..., x_n$ if $x \mid x_i$ for all $1 \leqslant i \leqslant n$. The largest of them is the *greatest common divisor of* $x_1, ..., x_n$, denoted $\gcd(x_1, ..., x_n)$. If it equals 1, then $x_1, ..., x_n$ are *coprime*.

Remark 2.1.7. This is well-defined: the set of common divisors contains 1, so is non-empty; it is finite because, if say $x_i \neq 0$, then all common divisors x satisfy $1 \leqslant |x| \leqslant |x_i|$ by Remark 2.1.2 (5).

Lemma 2.1.8 (Bézout)**.** *Let* $n > 0$ *and* $x_1, ..., x_n \in \mathbb{Z}$ *not all 0. Then there are* $c_1, ..., c_n \in \mathbb{Z}$ *such that*

$$\gcd(x_1, ..., x_n) = c_1x_1 + \cdots + c_nx_n.$$

Proof. Let U be the set of all $c_1x_1 + \cdots + c_nx_n$ with $c_i \in \mathbb{Z}$. Then U is a subgroup of $(\mathbb{Z}, +)$, so $U = d\mathbb{Z}$ for some $d \in \mathbb{N}$ by Lemma 2.1.5. As $d \in U$ it suffices to show $d = \gcd(x_1, ..., x_n)$. As all $x_i \in U = d\mathbb{Z}$, d is a common divisor. As not all $x_i = 0$, $U \neq \{0\}$ and $d \neq 0$. By Remark 2.1.2 (3), any common divisor c divides all elements of U, so also d. By Remark 2.1.2 (5), $c \leqslant |c| \leqslant |d| = d$. □

Remark 2.1.9. Let $n > 0$, $x_1, ..., x_n \in \mathbb{Z}$ not all 0, $d := \gcd(x_1, ..., x_n)$ and $x, y, z \in \mathbb{Z} \setminus \{0\}$.

1. Every common divisor of $x_1, ..., x_n$ divides d (by Lemma 2.1.8 and Remark 2.1.2 (3)).
2. $x_1, ..., x_n$ are coprime if and only if $c_1x_1 + ... + c_nx_n = 1$ for certain $c_i \in \mathbb{Z}$;

   Indeed: $\Rightarrow$ by Lemma 2.1.8, $\Leftarrow$: $d \mid 1$ by 2.1.2 (3), so $d = 1$ by Remark 2.1.2 (6).
3. $x_1/d, ..., x_n/d$ are coprime (divide the equation of Lemma 2.1.8 by d and apply (2)).
4. $\gcd(x, y, z) = \gcd(x, \gcd(y, z))$ (by (1), x, y, z and $x, \gcd(y, z)$ have the same divisors).
5. If $x \mid yz$ and x, y are coprime, then $x \mid z$.

   Indeed: by Lemma 2.1.8 write $1 = cx + c'y$ for certain $c, c' \in \mathbb{Z}$, so $z = zcx + zc'y$, so $x \mid z$ (by $x \mid yz$ and Remark 2.1.2 (3)).
6. If $x \mid z$ and $y \mid z$ and x, y are coprime, then $xy \mid z$.

   Indeed: write $z = cx$, so $y \mid xc$, so $y \mid c$ by (5), say $c = c'y$, then $z = c'yx$.

Exercise 2.1.10. Let $x, y, z \in \mathbb{Z} \smallsetminus \{0\}$. Show that $\mathrm{ggT}(x, yz) \mid \mathrm{ggT}(x, y) \cdot \mathrm{ggT}(x, z)$. For coprime y, z we have equality.

By (4), an algorithm for computing the gcd of two integers can be iterated to compute the gcd of any finite number of integers.

Theorem 2.1.11 (Euclidian algorithm)**.** *For $x, y \in \mathbb{Z} \smallsetminus \{0\}$ with $y \nmid x$ let $r_0, r_1, \ldots$ be the sequence with $r_0 := x, r_1 := y$ and, for $i > 0$,*

$$r_{i+1} := \begin{cases} \text{the remainder of } (r_{i-1}, r_i) & \text{if } r_i \neq 0 \\ 0 & \text{else.} \end{cases}$$

Then $r_{n+1} = 0$ for some $0 < n < |y|$ and $r_n = \gcd(x, y)$ for the minimal such n.

Moreover, for this n let $s_0, \ldots, s_n$ and $t_0, \ldots, t_n$ be the sequences with $s_0 := 1, s_1 := 0$ and $t_0 := 0, t_1 := 1$ and for $0 < i < n$, letting q_i be the quotient of (r_{i-1}, r_i),

$$s_{i+1} := s_{i-1} - q_i s_i, \quad t_{i+1} := t_{i-1} - q_i t_i.$$

Then $\gcd(x, y) = s_n x + t_n y$.

Proof. Note $|y| > r_2 > 0$ as $y \nmid x$, and $r_2 > r_3 > \cdots$ are all $\geqslant 0$, so n as claimed exists. Note

$$x = q_1 y + r_2, \quad y = q_2 r_2 + r_3, \quad r_2 = q_3 r_3 + r_4, \quad \ldots \quad r_{n-2} = q_n r_{n-1} + r_n, \quad r_{n-1} = q_n r_n + 0.$$

Work the equations backwards: $r_{n+1} = 0$, so $r_n \mid r_{n-1}$, so $r_n \mid r_{n-2}$ by Remark 2.1.2 (3), etc., so $r_n \mid r_1 = y$ and $r_n \mid r_0 = x$. Hence r_n is a common divisor of x, y.

To see it is the largest, c be a common divisor of x, y. Work the equations forwards: as $r_2 = x - q_1 y$ we have $c \mid r_2$ by Remark 2.1.2 (3); as $r_3 = y - q_2 r_2$ we have $c \mid r_3$, etc., so $c \mid r_n$.

Finally, we claim $r_i = s_i x + t_i y$ for all $i \leqslant n$. This is true for $i = 0, 1$. Inductively,

$$r_{i+1} = r_{i-1} - q_i r_i = (s_{i-1} x + t_{i-1} y) - q_i (s_i x + t_i y) = s_{i+1} x + t_{i+1} y. \qquad \square$$

Example 2.1.12. We compute $\gcd(122, 16) = 2 = r_5$ with r_2, r_3, r_4, r_5, r_6 being $10, 6, 4, 2, 0$:

$$122 = 7 \cdot 16 + 10, \quad 16 = 1 \cdot 10 + 6, \quad 10 = 1 \cdot 6 + 4, \quad 6 = 1 \cdot 4 + 2, \quad 4 = 2 \cdot 2 + 0.$$

Note q_0, q_1, q_2, q_3, q_4 are $7, 1, 1, 1, 2$. Thus,

$$\begin{array}{llll} s_2 = 1 - 7 \cdot 0 & s_3 = 0 - 1 \cdot 1 & s_4 = 1 - 1 \cdot (-1) & s_5 = -1 - 1 \cdot 2 \\ t_2 = 0 - 7 \cdot 1 & t_3 = 1 - 1 \cdot (-7) & t_4 = -7 - 1 \cdot 8 & t_5 = 8 - 1 \cdot (-15). \end{array}$$

and $r_n = 2 = -3 \cdot 122 + 23 \cdot 16$. Or look at the equations and substitute from right to left: 4 by $10 - 6$, then 6 by $16 - 10$, then 10 by $122 - 7 \cdot 16$:

$$\begin{aligned} 2 &= 6 - 4 = 6 - (10 - 6) = 2 \cdot 6 - 10 \\ &= 2 \cdot (16 - 10) - 10 = 2 \cdot 16 - 3 \cdot 10 \\ &= 2 \cdot 16 - 3 \cdot (122 - 7 \cdot 16) = -3 \cdot 122 + 23 \cdot 16. \end{aligned}$$

Exercise 2.1.13. Recall Bézout's lemma. Consider an equation $a_1X_1 + \cdots a_nX_n = b$ where $a_1, ..., a_n, b \in \mathbb{Z}$ and $X_1, ..., X_n$ are variables ranging over $\mathbb{Z}$. How do you decide whether there is a solution and, in case, compute one?

Definition 2.1.14. Let $n > 0$. Then $y \in \mathbb{Z}$ is a *common multiple of* $x_1, ..., x_n \in \mathbb{Z} \setminus \{0\}$ if $x_i \mid y$ for all i. The minimal (least) common multiple in $\mathbb{N} \setminus \{0\}$ is denoted $\operatorname{lcm}(x_1, ..., x_n)$.

Exercise 2.1.15.

1. Show this is well-defined.
2. Show y is a common multiple of $x_1, ..., x_n$ if and only if $\operatorname{lcm}(x_1, ..., x_m) \mid y$.
3. For $x, y \in \mathbb{Z} \setminus \{0\}$ show $|xy| = \gcd(x, y) \cdot \operatorname{lcm}(x, y)$.

2.2 The fundamental theorem of number theory

Definition 2.2.1. $n > 1$ is *prime* if only 1 and n are factors of n; otherwise it is *composite*.

Lemma 2.2.2. *Every integer $x \in \mathbb{Z}$ with $|x| > 1$ has a prime factor.*

Proof. As $|x| \mid x$ there exist a natural $n > 1$ dividing x. The smallest such n is prime. □

Exercise 2.2.3. Every composite natural n has a prime factor $\leqslant \sqrt{n}$.

Theorem 2.2.4 (Euclid)**.** *There are infinitely many primes.*

Proof. Let $p_1, ..., p_n$ be finitely many primes ($n > 0$). Then $z := p_1 \cdots p_n + 1 > 1$. By the lemma, z has a prime factor p. Then p is distinct from all p_i: otherwise, $p \mid z - p_1 \cdots p_n = 1$ by Remark 2.1.2 (3), a contradiction. □

Exercise 2.2.5. Let $n > 1$.

1. Gaps: there are n consecutive naturals that are not prime.
2. There is a prime p with $n < p \leqslant n!$.
3. There are infinitely many primes of the form $4x + 3$ with $x \in \mathbb{N}$.

   Hint: given finitely many primes $3 < p_1, ..., p_n$ of the form $4x + 3$, consider $4p_1 \cdots p_n + 3$. Can all its prime factors have the form $4x + 1$?

Theorem 2.2.6 (Wilson)**.** *A natural $n > 1$ is prime if and only if $n \mid (n-1)! + 1$.*

We defer the proof to Section 2.5. The theorem was stated already by the medieval polymath Ibn al-Haytham (c.965 – c.1040), sometimes called "the first true scientist".

Lemma 2.2.7 (Euclid's lemma)**.** *$p \in \mathbb{N}$ is prime if and only if for all $x, y \in \mathbb{Z}$:*

$$p \mid xy \text{ implies } p \mid x \text{ or } p \mid y.$$

Proof. $\Rightarrow$: if $p \mid xy$ is prime and $p \nmid x$, then $\gcd(p,x) = 1$, so $p \mid y$ by Remark 2.1.9 (5). $\Leftarrow$: if p is not prime, then $p = nm$ for certain $1 \leqslant n, m < p$; then $p \mid nm$ but $p \nmid n$ and $p \nmid m$. □

Lemma 2.2.8. *Let $k, \ell, n > 0$ and p be prime.*

1. *If $x_1, ..., x_n \in \mathbb{Z}$ and $p \mid x_1 \cdots x_n$, then $p \mid x_i$ for some i.*
2. *If $x_1, ..., x_n$ are prime and $p \mid x_1 \cdots x_n$, then $p = x_i$ for some i.*
3. *If $x, y \in \mathbb{Z}$ and $p^k x = p^\ell y$ and $p \nmid x$ and $p \nmid y$, then $k = \ell$ and $x = y$.*

Proof. (1): for $n = 1$ there is nothing to show. Assume $n > 1$ and argue inductively: if $p \mid (x_1 \cdots x_{n-1})x_n$, then $p \mid x_n$ or $p \mid x_1 \cdots x_{n-1}$ by Lemma 2.2.8. In the 1st case we are done, and in the 2nd too by induction.

(2) follows from (1) because $p \mid x_i$ implies $p = 1$ or $p = x_i$ if x_i is prime.

(3): by cancellation it suffices to show $k = \ell$. Otherwise assume $k < \ell$ (the case $\ell < k$ is analogous). Then $p \mid p^{\ell-k} y = x$ as $\ell - k > 0$, a contradiction. □

Theorem 2.2.9 (Fundamental theorem of number theory)**.** *For every natural $n > 1$ there are $r \in \mathbb{N}$ and primes $p_1 < \cdots < p_r$ and naturals $k_1, ..., k_n > 0$ such that*

$$n = p_1^{k_1} \cdots p_r^{k_r}.$$

The numbers $r, p_1, ..., p_r, k_1, ..., k_r$ are unique and called the prime factorization of n.

Proof. Existence: otherwise there is a minimal natural $n > 1$ that is not a product of primes. Then n is not prime, say $n = xy$ with $1 < x, y < n$. By minimality of n, both x and y are products of primes. Hence so is n, a contradiction.

Uniqueness: assume $n = q_1^{\ell_1} \cdots q_s^{\ell_s}$ for $s, \ell_j > 0$ and primes $q_1 < \cdots < q_s$. We show by induction that $r = s$ and $k_i = \ell_i$ and $p_i = q_i$. For $n = 2$ this is clear.

Assume $n > 2$. By Lemma 2.2.8 (2), $p_1 = q_j$ and $q_1 = p_i$ for some i, j. Then $p_1 \leqslant p_i = q_1 \leqslant q_j = p_1$, so $i = j = 1$ and $p_1 = q_1$. Let $x := p_2^{k_2} \cdots p_r^{k_r}$ and $y := q_2^{\ell_2} \cdots q_s^{\ell_s}$ – we agree the empty product is 1. By Lemma 2.2.8 (2), $p_1 \nmid x$ and $p_1 \nmid y$. Since $n = p_1^{k_1} x = q_1^{\ell_1} y$, Lemma 2.2.8 (3) gives $k_1 = \ell_1$ and $x = y$. Now, if $x = 1$, then $r = s = 1$ and we are done. If $x > 1$, then $r, s \geqslant 2$ and $p_2^{k_2} \cdots p_r^{k_r} = q_2^{\ell_2} \cdots q_s^{\ell_s} > 1$. By induction, $r = s$ and $\ell_i = k_i$ and $p_i = q_i$. □

We introduce notation for the k_i's:

Definition 2.2.10. Let p be prime. The *p-adic valuation* $\nu_p : \mathbb{N} \to \mathbb{N}$ is given by

$$\nu_p(n) := \max\{k \in \mathbb{N} \mid p^k \mid n\}.$$

Remark 2.2.11. The fundamental theorem states $n = \prod_p p^{\nu_p(n)}$ where p runs over all primes. Note the product is finite in the sense that all but finitely many factors are 1.

Exercise 2.2.12. Let $n, m \in \mathbb{N}$.

1. $n \mid m$ if and only if $\nu_p(n) \leqslant \nu_p(m)$ for all primes p.

2. How many divisors does n have?
3. $\gcd(n,m) = \prod_p p^{\min\{\nu_p(n),\nu_p(m)\}}$, $\quad \operatorname{lcm}(n,m) = \prod_p p^{\max\{\nu_p(n),\nu_p(m)\}}$.

Notation: for $x \in \mathbb{R}$ let $\lfloor x \rfloor$ denote the largest integer $\leqslant x$.

Note the sum below is finite in that only finitely many terms are $\neq 0$:

Lemma 2.2.13 (Legendre's formula)**.** *Let $n, p \in \mathbb{N}$, p prime. Then*

$$\nu_p(n!) = \textstyle\sum_{k>0} \lfloor n/p^k \rfloor.$$

Proof. Count occurrences of p as a prime factor of a number $\leqslant n$. One per multiple of p: $\lfloor n/p \rfloor$ many. One additional occurrence per multiple of p^2: $\lfloor n/p^2 \rfloor$ many. And so on. □

2.3 Chebychev's prime number theorem

How many primes are there? We want to know the growth rate of:

Definition 2.3.1. For $n \in \mathbb{N}$ let $\pi(n)$ be the number of primes $p \leqslant n$.

Landau notation: Let $f, g : \mathbb{N} \to \mathbb{R}$. Then both $f \leqslant O(g)$ and $g \geqslant \Omega(f)$ mean: there is $c \in \mathbb{N}$ such that $f(n) \leqslant cg(n) + c$ for all $n \in \mathbb{N}$; if additionally $g \leqslant O(f)$, then $f = \Theta(g)$.

Remark 2.3.2. We need some simple bounds on binomial coefficients. For $0 < k \leqslant n$:

$$(n/k)^k \leqslant \binom{n}{k} = \frac{n}{k}\frac{n-1}{k-1}\cdots\frac{n-k+1}{1} \leqslant n^k/k! < (e \cdot n/k)^k.$$

The final $<$ is equivalent to $k^k/k! < e^k$ which is clear from the power series of e^k.

Theorem 2.3.3 (Chebychev 1850)**.** $\pi(n) = \Theta(n/\ln n)$.

Proof. [1] Upper bound: let $n > 0$; a prime $n < p \leqslant 2n$ divides $(2n)!$ but not $n!n!$. Consider the product $\prod_{n<p\leqslant 2n} p$, where p ranges over primes. It divides the binomial coefficient $\binom{2n}{n} = \frac{(2n)!}{n!\cdot n!} \leqslant (2e)^n$ using Remark 2.3.2. It is $\geqslant n^{\pi(2n)-\pi(n)}$. Thus, $n^{\pi(2n)-\pi(n)} \leqslant (2e)^n$ and

$$\pi(2n) - \pi(n) \leqslant 2n/\ln n,$$

Given$n > 2$, choose $\ell \in \mathbb{N}$ minimal such that $2^\ell \geqslant n$. Note $2^\ell \leqslant 2n$ and $\ell \geqslant \ln n/\ln 2$. By the above, $\pi(2^k) - \pi(2^{k-1}) \leqslant 2^k/\ln(2^{k-1}) = 2^k/((k-1)\ln 2)$ for all $k > 0$. Thus,

$$\begin{aligned}\pi(n) &\leqslant \pi(2^\ell) = \pi(2^\ell) - \pi(2^{\ell-1}) + \pi(2^{\ell-1}) - \pi(2^{\ell-2}) + \pi(2^{\ell-2}) - \cdots + \pi(2^2) - \pi(2^1) + \pi(2^1)\\ &\leqslant \left(2^\ell/(\ell-1) + 2^{\ell-1}/(\ell-2) + \cdots + 2^2/1 + 1\right)/\ln 2 \leqslant (12 \cdot 2^\ell/\ell)/\ln 2 \leqslant 24 \cdot n/\ln n,\end{aligned}$$

where the penultimate step is an easy induction. Hence, $\pi(n) \leqslant O(n/\ln n)$.

[1] I learned this proof from S. Glock.

Lower bound: for a prime p we have $\nu_p(\binom{2n}{n}) = \nu_p((2n)!) - 2\cdot\nu_p(n!)$. By Legendre,

$$\nu_p(\tbinom{2n}{n}) = \textstyle\sum_{k>0}\left(\lfloor 2n/p^k\rfloor - 2\cdot\lfloor n/p^k\rfloor\right) \leqslant \max\{k \mid p^k \leqslant 2n\},$$

where the inequality follows from $\lfloor 2x\rfloor - 2\lfloor x\rfloor \in \{0,1\}$ for all $x\in\mathbb{R}$. Thus,

$$2^n \leqslant \tbinom{2n}{n} = \textstyle\prod_{p \text{ prime}} p^{\nu_p(\binom{2n}{n})} \leqslant (2n)^{\pi(2n)}.$$

Thus, $\pi(2n) \geqslant \ln 2\cdot n/\ln(2n)$. This implies $\pi(n) \geqslant \Omega(n/\ln n)$. □

Remark 2.3.4. Chebychev gave estimates close to 1 for the constants in Θ. The *Prime Number Theorem* states that indeed $\pi(n)\ln n/n \to_n 1$. It was conjectured by Legendre and 16 year old Gauß in the 1790s, and proved in 1896 independently by Hadamard and De La Vallée-Poussin following Riemann's seminal work from 1859.

Remark 2.3.5. More about primes.

1. A *Fermat prime* is a prime of the form 2^n+1. Then n is a power of 2: otherwise write $n = 2^k m$ for odd $m>1$ and factor $2^n+1 = (x+1)\cdot(x^{m-1}-x^{m-2}+\cdots-x+1)$ with $x := 2^{2^k}$. Fermat conjectured that all $2^{2^k}+1$ are prime, a surprising error revealed by Euler: $2^{2^5}+1 = 641\cdot 6700417$. It is conjectured that

   $$2^{2^0}+1 = 3,\ 2^{2^1}+1 = 5,\ 2^{2^2}+1 = 17,\ 2^{2^3}+1 = 257,\ 2^{2^4} = 65537$$

   lists all Fermat primes. The largest $2^{2^k}+1$ known to be composite is for $k := 18233954$ with prime factor $7\cdot 2^{18233956}+1$.
2. A *Mersenne prime* is a prime of the form 2^n-1. Then n is prime because $2^{k\ell}-1 = (2^k-1)(1+2^k+2^{2k}\cdots+2^{(\ell-1)k})$. Cole refuted 1903, after "three years of sundays", Mersenne's two centuries old conjecture showing $2^{67}-1 = 193707721\cdot 761838257287$ is not prime. It is conjectured that there are infinitely many such non-primes. It is also conjectured that there are infinitely many Mersenne primes. The largest known is also the largest known prime: $2^{136279841}-1$; it has 41024320 digits.
3. It is unknown whether there are infinitely many *twin primes*: a prime p such that $p+2$ is also prime. The largest currently known one has 388342 digits. Brun showed 1919 that there are much less twin primes than primes: if $\pi_2(n)$ is the number of twin primes $\leqslant n$, then $\pi_2(n)/\pi(n) \to_n 0$.
4. *Goldbach's conjecture* (1742) is also open: is every even $n>2$ the sum of two primes? Computers verified this for all $n \leqslant 4\cdot 10^{18}$.
5. *Dirichlet's theorem* (1837) states that for all coprime $a,b\in\mathbb{N}$ there are infinitely many primes of the form $ax+b$ where $x\in\mathbb{N}$. Theorem 6.9.15 proves this for $b=1$.
6. Green and Tao proved 2004 that there are arbitrarily long *arithmetic progressions of primes*: for every $\ell>0$ there are primes $p_1<\cdots<p_\ell$ such that $p_2-p_1 = \cdots = p_\ell - p_{\ell-1}$.

Exercise 2.3.6. Let $p>3$ be a twin prime and $q := p+2$ (prime). Show $12 \mid p+q$.

2.4 The Chinese remainder theorem

Definition 2.4.1. For $n \in \mathbb{Z}$ call $x, y \in \mathbb{Z}$ *congruent modulo* n, symbolically $x \equiv y \mod n$, if $n \mid x - y$. The equivalence class of $x \in \mathbb{Z}$ is the *residue class of* x *modulo* n and denoted $[x]_n$, or $\bar{x}$ if n is clear from context. The set of equivalence classes is denoted $\mathbb{Z}_n$.

Remark 2.4.2. This is indeed an equivalence relation. It is reflexive because $n \mid x - x = 0$, symmetric because $n \mid x - y$ if and only if $n \mid -(x-y) = y - x$, and transitive because $n \mid x - y$ and $n \mid y - z$ implies $n \mid (x-y) + (y-z) = x - z$.

We usually take $n \in \mathbb{N}$; note n and $-n$ define the same relation. Any two numbers are congruent modulo $n := 1$; only identical numbers are congruent modulo $n := 0$. If $m \mid n$, then congruence modulo n refines congruence modulo m: $x \equiv y \mod n$ implies $x \equiv y \mod m$.

Remark 2.4.3. Let $n > 1$ and $x, x', y, y', z \in \mathbb{Z}$.

1. $[x]_n = x + n\mathbb{Z} := \{x + nz \mid z \in \mathbb{Z}\}$.

   Indeed: $y \in [x]_n \Leftrightarrow n \mid x - y \Leftrightarrow nz = x - y$ for some $z \in \mathbb{Z} \Leftrightarrow y \in x + n\mathbb{Z}$.
2. There are n residue classes modulo n,namely $[0]_n, ..., [n-1]_n$.

   Indeed: clearly, the listed classes are pairwise distinct; the list is complete: for $x \in \mathbb{Z}$ write $x = qn + r$ by Euclidian division, so $0 \leqslant r \leqslant n - 1$ and $r = x - qn \equiv x \mod n$.
3. If $x \equiv x' \mod n$ and $y \equiv y' \mod n$, then

$$x + y \equiv x' + y' \mod n, \quad x - y \equiv x' - y' \mod n, \quad xy \equiv x'y' \mod n.$$

4. If $x \equiv y \mod n$ and $k \in \mathbb{N}$, then $x^k \equiv y^k \mod n$.
5. If $z \neq 0$ and $xz \equiv yz \mod n$, then $x \equiv y \mod n/d$ for $d := \gcd(z, n)$.

   Indeed: $n \mid (xz - yz) = z(x-y)$ implies $n/d \mid (z/d)(x-y)$. By Remark 2.1.9 (3), $n/d, z/d$ are coprime, so $n/d \mid (x-y)$ by Remark 2.1.9 (5).

   E.g., $1 \cdot 2 \equiv 4 \cdot 2 \mod 6$ and $1 \not\equiv 4 \mod 6$ but $1 \equiv 4 \mod 6/2$.

Example 2.4.4. $[0]_5 = [2025]_5, [1]_5 = [66]_5, [2]_5 = [-33]_5, [3]_5 = [-102]_5, [4]_5 = [-1]_5$.

We sample some applications of reasoning with congruences.

Example 2.4.5. What is the last digit of $n := 13^{14}$?

Solution: that the digits are $a_\ell a_{\ell-1} \cdots a_0$ with $a_i \leqslant 9$ means $n = a_\ell 10^\ell + \cdots + a_1 10 + a_0$. Hence, a_0 is the unique natural $\leqslant 9$ with $n \equiv a_0 \mod 10$. We use the *Russian peasant method*, i.e., repeated squaring: modulo 10 we have $13 \equiv 3$, so $13^2 \equiv 9 \equiv -1$, so $13^4 \equiv (-1)^2 \equiv 1$, so $13^8 \equiv 1$; then $13^6 \equiv 13^4 \cdot 13^2 \equiv -1$ and $13^{14} \equiv 13^8 \cdot 13^6 \equiv -1 \equiv 9$. Hence, $a_0 = 9$.

Exercise 2.4.6. A natural n is divisible by 9 (or 3) if and only if so is the sum of its digits.

Exercise 2.4.7. Consider an equation $a_1 X_1 + \cdots + a_r X_r \equiv b \mod n$ where $a_1, ..., a_r, b \in \mathbb{Z}$, $n > 0$ and $X_1, ..., X_r$ are variables ranging over $\mathbb{Z}$. Show it has a solution if and only if $\gcd(a_1, ..., a_r, n) \mid b$. In case, how do you compute one?

Example 2.4.8 (ISBN)**.** Books are assigned an ISBN, a sequence of 11 digits (until 2007) determining 10 naturals $a_1, ..., a_9 \leqslant 9$ and $a_{10} \leqslant 10$. The numbers $a_1, ..., a_9$ code information about the book (editor etc.), and a_{10} is such that $\sum_{i=1}^{10} i \cdot a_i \equiv 0 \mod 11$.

Assume one of the first 9 numbers is corrupted, say a_{i_0}. Then one can still compute $b := \sum_{i \neq i_0} i \cdot a_i$. Then $i_0 \cdot a_{i_0} + b \equiv 0 \mod 11$. Then a_{i_0} is a solution of $i_0 X \equiv -b \mod 11$ and the unique one in $\{0, ..., 9\}$. Thus a_{i_0} can be computed as in the previous exercise.

Theorem 2.4.9 (Chinese remainder theorem)**.** *Let $r > 0$, $a_1, ..., a_r \in \mathbb{Z}$ and assume that $m_1, ..., m_r > 1$ are pairwise coprime. Set $m := m_1 \cdots m_r$.*

1. *The following system of equations has a solution in $\mathbb{Z}$:*
$$X \equiv a_1 \mod m_1, \quad ... \quad X \equiv a_r \mod m_r$$
2. *For all $1 \leqslant i \leqslant r$ there is $b_i \in \mathbb{Z}$ with $\frac{m}{m_i} \cdot b_i \equiv 1 \mod m_i$. Then a solution is*
$$x := a_1 \frac{m}{m_1} b_1 + \cdots + a_r \frac{m}{m_i} b_r.$$
3. *If x_0 is a solution, then the set of solutions is $[x_0]_m$.*

Proof. (2) implies (1). (2): we claim m_i and m/m_i are coprime for all i. Otherwise there is a common prime factor p of m_i and m/m_i. Then $p \mid m_1 \cdots m_{i-1} m_{i+1} \cdots m_r$ and Euclid's lemma gives $j \neq i$ with $p \mid m_j$. As also $p \mid m_i$, this contradicts m_i, m_j being coprime.

Bézout gives $y, z \in \mathbb{Z}$ with $1 = y m_i + z \frac{m}{m_i}$. Then $1 \equiv z m/m_i \mod m_i$ so $b_i := z$ is as claimed. Let x be as in (2) and $1 \leqslant i \leqslant r$. Note $m_i \mid \frac{m}{m_j}$ for $j \neq i$, so $a_j \frac{m}{m_i} b_j \equiv 0 \mod m_i$. Hence $x \equiv a_i \frac{m}{m_i} b_i \equiv a_i \cdot 1 \mod m_i$.

(3): let S be set of solutions and $x_0 \in S$. We show $S = [x_0]_m$. If $x \in S$, then $x \equiv x_0 \equiv a_j \mod m_j$ for all j, i.e., $m_j \mid x_0 - x$. Since m_1, m_2 are coprime, Remark 2.1.9 (6) gives $m_1 m_2 \mid x_0 - x$. But also $m_1 m_2$ and m_3 are coprime (see the argument above), so $m_1 m_2 m_3 \mid x_0 - x$, etc. Hence $m \mid x_0 - x$, i.e., $x_0 \equiv x \mod m$. Conversely, if $m \mid x_0 - x$, so $m_i \mid x_0 - x$ for all i, so $x \equiv x_0 \equiv a_i \mod m_i$ for all i, i.e., $x \in S$. □

Example 2.4.10. We compute all solutions (in $\mathbb{Z}$) of
$$X \equiv 1 \mod 2, \quad X \equiv 2 \mod 3, \quad X \equiv 4 \mod 5.$$
We copy the notation: a_1, a_2, a_3 are $1, 2, 4$ and m_1, m_2, m_3 are $2, 3, 5$ and $m = 30$.

- we want $b_1 \frac{m}{m_1} \equiv 1 \mod m_1$, i.e., $15 \cdot b_1 \equiv 1 \mod 2$ and take $b_1 := 1$;
- we want $b_2 \frac{m}{m_2} \equiv 1 \mod m_2$, i.e., $10 \cdot b_2 \equiv 1 \mod 3$ and take $b_2 := 1$;
- we want $b_3 \frac{m}{m_3} \equiv 1 \mod m_3$, i.e., $6 \cdot b_3 \equiv 1 \mod 5$ and take $b_3 := 1$;
- we get a solution $x_0 := a_1 \frac{m}{m_1} b_1 + a_2 \frac{m}{m_2} b_2 + a_3 \frac{m}{m_3} b_3 = 1 \cdot 15 \cdot 1 + 2 \cdot 10 \cdot 1 + 4 \cdot 6 \cdot 1 = 59$;
- the set of solutions is $[59]_{30} = [-1]_{30}$.

Exercise 2.4.11. Find the smallest $n \in \mathbb{N}$ such that for all $2 \leqslant m \leqslant 7$ the remainder of (n, m) is $m - 1$.

2.5 Residue class rings

Definition 2.5.1. Let $n > 1$. For $x, y \in \mathbb{Z}$ set

$$[x]_n + [y]_n := [x+y]_n, \qquad [x]_n \cdot [y]_n := [x \cdot y]_n.$$

Remark 2.5.2. These are well-defined by Remark 2.4.3 (2) and make $(\mathbb{Z}_n, +, \cdot)$ a commutative ring (exercise). Further, $x \mapsto [x]_n$ is a ring epimorphism from $\mathbb{Z}$ onto $\mathbb{Z}_n$.

Examples 2.5.3. Here are the tables for $\mathbb{Z}_3$ and $\mathbb{Z}_4$:

$\mathbb{Z}_3$:

$+$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{0}$
$\bar{2}$	$\bar{2}$	$\bar{0}$	$\bar{1}$

$\cdot$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{1}$

$\mathbb{Z}_4$:

$+$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{0}$
$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{0}$	$\bar{1}$
$\bar{3}$	$\bar{3}$	$\bar{0}$	$\bar{1}$	$\bar{2}$

$\cdot$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{0}$	$\bar{2}$
$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

In $\mathbb{Z}_3$ we have $\bar{2}^{-1} = \bar{2}$, so every $\bar{x} \neq \bar{0}$ has a multiplicative inverse and $\mathbb{Z}_3$ is a field.

In $\mathbb{Z}_4$ we have $\bar{2} \neq \bar{0}$ and $\bar{2} \cdot \bar{2} = \bar{0}$, so $\bar{2}$ is a zero-divisor and $\mathbb{Z}_4$ is not an integral domain; its units are $\mathbb{Z}_4^\times = \{\bar{1}, \bar{3}\}$ with $\bar{3}^{-1} = \bar{3}$.

Exercise 2.5.4. Let $n > 1$. Give an isomorphism from $(\mathbb{Z}_n, +, \cdot)$ onto $(\{0, ..., n-1\}, +', \cdot')$ where $i +' j, i \cdot' j$ are defined as the remainders of $(i + j, n), (i \cdot j, n)$.

Exercise 2.5.5. Recall Definition 1.6.8. Show $(\mathbb{Z}_n, +) \cong (C_n, \cdot)$.

Proposition 2.5.6. *Let $n > 1$. Then $\mathbb{Z}_n$ is a field if and only if n is prime.*

Proof. $\Leftarrow$: assume n is prime. We have to show that for every $\bar{x} \in \mathbb{Z}_n \setminus \{\bar{0}\}$ there is $\bar{y} \in \mathbb{Z}_n$ such that $\bar{x} \cdot \bar{y} = \bar{1}$. Then $n \nmid x$, so $\gcd(x, n) = 1$ as n is prime. By Bézout's lemma there are $y, z \in \mathbb{Z}$ such that $1 = yx + zn$. Then $\bar{1} = \bar{y} \cdot \bar{x} + \bar{z} \cdot \bar{n}$. But $\bar{n} = \bar{0}$, so $\bar{1} = \bar{y} \cdot \bar{x}$..

$\Rightarrow$: assume n is not prime, so there are $1 < x, y < n$ with $n = xy$. Then $\bar{0} = \bar{x} \cdot \bar{y}$ and $\bar{x} \neq \bar{0}$ and $\bar{y} \neq \bar{0}$. Hence, $\mathbb{Z}_n$ is not an integral domain, so not a field. □

Notation: for a prime p the field $\mathbb{Z}_p$ is denoted $\mathbb{F}_p$.

Using the field structure it is easy to prove:

Theorem 2.5.7 (Wilson)**.** *$n > 1$ is prime if and only if $(n-1)! \equiv -1 \mod n$. Moreover, $(n-1)! \equiv 0 \mod n$ if n is not prime and $n \neq 4$.*

Proof. Clear for $n \leqslant 3$. Assume $n > 3$. $\Leftarrow$: if $n > 1$ is not a prime, let $1 < q < n$ be a prime factor. If $n/q \neq q$, then $n = q \cdot n/q \mid (n-1)!$, so $(n-1)! \equiv 0 \mod n$. If $n/q = q$, then $n = q^2$ and we have two subcases. If $q = 2$, then $n = 4$ and $(4-1)! \equiv 2 \mod 4$; if $q > 2$, then $2q < q^2 = n$ and $2n = q \cdot 2q \mid (n-1)!$, so $(n-1)! \equiv 0 \mod n$.

$\Rightarrow$: if $n > 1$ is prime, $\mathbb{Z}_n$ is a field. Let f map $0 < i < n$ to the unique $0 < j < n$ such that $[i]_n \cdot [j]_n = [1]_n$; note that then $f(j) = i$. We have $f(1) = 1$ and $f(n-1) = n-1$ but no other fixed points: if $1 < i < n-1$, then $i^2 - 1 = (i-1)(i+1) \not\equiv 0 \mod n$ because $i-1, i+1 \not\equiv 0 \mod n$ (and $\mathbb{Z}_n$ is an integral domain). Hence $2 \cdots (n-2) \equiv 1 \mod n$ – it is a product of $(n-3)/2$ many 1 modulo n. Then $(n-1)! \equiv 1 \cdot (n-1) \equiv -1 \mod n$. □

Another easy application is:

Theorem 2.5.8 (Fermat's little theorem)**.** *If p is prime and $x \in \mathbb{Z}$ with $p \nmid x$, then*

$$x^{p-1} \equiv 1 \mod p.$$

Proof. $\bar{y} \mapsto \bar{x} \cdot \bar{y}$ permutes $\mathbb{Z}_p \setminus \{\bar{0}\}$, so $\bar{2} \cdots \overline{p-1} = \overline{x2} \cdots \overline{x(p-1)} = \bar{x}^{p-1} \cdot \bar{2} \cdots \overline{p-1}$, so $\bar{x}^{p-1} = \bar{1}$. □

Remark 2.5.9. There are infinitely many *Carmichael numbers*: composite n such that $x^{n-1} \equiv 1 \mod n$ for all x coprime to n. Below 10^8 there are only 255 of them, the first three are $561 = 3 \cdot 11 \cdot 7$, $1105 = 5 \cdot 13 \cdot 17$ and $1729 = 7 \cdot 13 \cdot 19$.

What does the Chinese remainder theorem tell us about the ring $\mathbb{Z}_n$?

Lemma 2.5.10. *Let $r > 0$ and $(R_1, +_1, \cdot_1), \ldots, (R_r, +_r, \cdot_r)$ be (commutative) rings. Then their* direct product *$R_1 \times \cdots \times R_k$ is a (commutative) ring with $+, \cdot$ defined for all $(x_1, \ldots, x_r)$, $(y_1, \ldots, y_r) \in R_1 \times \cdots \times R_r$ as follows:*

$$\begin{aligned} (x_1, \ldots, x_r) + (y_1, \ldots, y_r) &:= (x_1 +_1 y_1, \ldots, x_r +_r y_r), \\ (x_1, \ldots, x_r) \cdot (y_1, \ldots, y_r) &:= (x_1 \cdot_1 y_1, \ldots, x_r \cdot_r y_r). \end{aligned}$$

Further, $(R_1 \times \cdots \times R_r)^\times = R_1^\times \times \cdots \times R_r^\times$.

Proof. Associativity (commutativity) and distributivity of $+, \cdot$ are clear. The neutral elements are $(0_1, \ldots, 0_r), (1_1, \ldots, 1_r)$ with $0_i, 1_i$ denoting the neutral elements of R_i. The additive inverse of $(x_1, \ldots, x_r)$ is $(-_1 x_1, \ldots, -_r x_r)$ with $-_i$ denoting additive inverse in R_i.

Further: $(x_1, \ldots, x_r) \cdot (y_1, \ldots, y_r) = (1_1, \ldots, 1_r)$ if and only if $x_i \cdot_i y_i = 1_i$ for all i. □

Corollary 2.5.11. *Let $r > 0$ and $m_1, \ldots, m_r > 1$ be pairwise coprime, and $m := m_1 \cdots m_r$. Then $\mathbb{Z}_m \cong \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_r}$ via*

$$[x]_m \mapsto ([x]_{m_1}, \ldots, [x]_{m_r})$$

Proof. The map is well-defined: if $[x]_m = [x']_m$, then $x' \equiv x \mod m_i$ for all i (as $m_i \mid m$), i.e., $([x]_{m_1}, \ldots, [x]_{m_r}) = ([x']_{m_1}, \ldots, [x']_{m_r})$. It is clear that the map preserves $+, \cdot$ and 1. It is surjective by Theorem 2.4.9 (1). Hence it is bijective (both rings have size m). □

Example 2.5.12. $\mathbb{Z}_4 \not\cong \mathbb{Z}_2 \times \mathbb{Z}_2$: note $x + x = 0$ for all x on the right, but not on the left. The additive group of $\mathbb{Z}_2 \times \mathbb{Z}_2$ is the *Klein four-group* and denoted

$$K_4.$$

More generally, one easily sees that (the additive subgroups of) $\mathbb{Z}_n \times \mathbb{Z}_m$ and $\mathbb{Z}_{nm}$ are not isomorphic if n, m are not coprime (exercise).

2.6 Euler's totient

By Proposition 2.5.6, if n is prime and $x, y \in \mathbb{Z}_n \setminus \{0\}$, the equation $\bar{x} \cdot X = \bar{y}$ has exactly 1 solution in $\mathbb{Z}_n$, namely $\bar{x}^{-1}\bar{y}$. What about non-prime $n > 1$? We show there are either 0 or "$\gcd(\bar{x}, n)$" many solutions. We avoid this notation but it makes sense – the reader might verify as an exercise: if $\bar{x} = \bar{x}'$, then $\gcd(x, n) = \gcd(x', n)$.

Lemma 2.6.1. *Let $n > 1$ and $\bar{x}, \bar{y} \in \mathbb{Z}_n$. Then $\bar{x} \cdot X = \bar{y}$ has a solution in $\mathbb{Z}_n$ if and only if*

$$d := \gcd(x, n) \mid y.$$

Moreover, if $\bar{z} \in \mathbb{Z}_n$ is a solution, then there are exactly d solutions, namely

$$\bar{z},\ \overline{z + n/d}, \ldots, \overline{z + (d-1)n/d}.$$

Proof. Assume there is $z \in \mathbb{Z}$ with $\bar{x} \cdot \bar{z} = \bar{y}$, i.e., $\overline{xz} = \bar{y}$; then $n \mid y - xz$, so $y = un + xz$ for some $u \in \mathbb{Z}$; then $d \mid y$. Conversely, assume $d \mid y$, say $y = ud$ for some $u \in \mathbb{Z}$; by Bézout, $y = uu_0 x + uu_1 n$ for some $u_0, u_1 \in \mathbb{Z}$, so $\bar{y} = \overline{uu_0} \cdot \bar{x} + \bar{0}$ and $\overline{uu_0}$ is a solution.

Moreover, the listed classes are solutions: $\bar{x} \cdot \overline{z + in/d} = \overline{xz} + \overline{(x/d)in} = \bar{y} + \bar{0}$. They are pairwise distinct: if $i, j < d$ and $\overline{z + in/d} = \overline{z_0 + jn/d}$, then $n \mid (z + jn/d) - (z + in/d) = (j - i)n/d$, i.e., $un = (j - i)n/d$ for some $u \in \mathbb{Z}$, so $ud = (j - i)$. But $i, j < d$, so $j = i$.

No other solutions: assume $\bar{x} \cdot \bar{z}' = \bar{y}$, so $xz' \equiv xz \equiv y \mod n$. Then $z' \equiv z \mod n/d$ by Remark 2.4.3 (4). Then $z' = z + u \cdot n/d$ for some $u \in \mathbb{Z}$. Write $u = qd + r$ by Euclidian division, so $0 \leqslant r < d$. Then $z' = z + qn + rn/d$, so $\bar{z}' = \overline{z + rn/d}$. □

We leave the verification of the following as an easy and recommended exercise.

Corollary 2.6.2. *Let $n > 1$, and $\bar{x} \in \mathbb{Z}_n \setminus \{\bar{0}\}$. The following are equivalent.*

1. *$\bar{x} \in \mathbb{Z}_n^\times$.*
2. *$\gcd(x, n) = 1$.*
3. *For all $\bar{y} \in \mathbb{Z}_n$ there is $\bar{z} \in \mathbb{Z}_n$ such that $\bar{x} \cdot \bar{z} = \bar{y}$.*
4. *For all $\bar{y} \in \mathbb{Z}_n$ there is exactly one $\bar{z} \in \mathbb{Z}_n$ such that $\bar{x} \cdot \bar{z} = \bar{y}$.*
5. *$\bar{x}$ is not a zero divisor in $\mathbb{Z}_n$.*

Example 2.6.3. Is $[109]_{341} \in \mathbb{Z}_{341}^\times$? If so, find $0 < k < 341$ with $[k]_{341} = [109]_{341}^{-1}$.

Solution. following Euclid's algorithm write:

$$341 = 3 \cdot 109 + 14, \quad 109 = 7 \cdot 14 + 11, \quad 14 = 1 \cdot 11 + 3, \quad 11 = 3 \cdot 3 + 2, \quad 3 = 1 \cdot 2 + 1.$$

So $\gcd(109, 341) = 1$ and the answer is yes. Then plugging in backwards:

$$\begin{aligned} 1 &= 3 - 2 = 3 - (11 - 3 \cdot 3) = 4 \cdot 3 - 11 \\ &= 4 \cdot (14 - 11) - 11 = 4 \cdot 14 - 5 \cdot 11 \\ &= 4 \cdot 14 - 5 \cdot (109 - 7 \cdot 14) = 39 \cdot 14 - 5 \cdot 109 \\ &= 39(341 - 3 \cdot 109) - 5 \cdot 109 = 39 \cdot 341 - 122 \cdot 109, \end{aligned}$$

so $109 \cdot (-122) \equiv 1 \mod 341$ and $[109]_{341}^{-1} = [-122]_{341} = [219]_{341}$. □

Remark 2.6.4. By Remark 1.1.18, $\mathbb{Z}_n^\times$ is an abelian group (with $\cdot$).

Exercise 2.6.5 (Another view of $\mathbb{Z}_n^\times$)**.** Let $n > 1$. If $\bar{x} \in \mathbb{Z}_n^\times$, then $\varphi_{\bar{x}}(\bar{y}) := \bar{x} \cdot \bar{y}$ defines an automorphism of the group $(\mathbb{Z}_n, +)$. Any automorphism has this form. Conclude $\bar{x} \mapsto \varphi_{\bar{x}}$ is a group isomorphism from $(\mathbb{Z}_n^\times, \cdot)$ onto $\mathrm{Aut}(\mathbb{Z}_n, +)$.

Definition 2.6.6. *Euler's totient* $\varphi : \mathbb{N} \setminus \{0\} \to \mathbb{N}$ is defined by $\varphi(1) = 1$ and for $n > 1$:

$$\varphi(n) := |\mathbb{Z}_n^\times|.$$

Remark 2.6.7. By the previous corollary, for $n > 1$, $\varphi(n)$ is the number of $1 \leqslant x < n$ with $\gcd(x, n) = 1$. In particular, $\varphi(p) = p - 1$ for a prime p. For composites, e.g.,

n	4	6	8	9	10	12	14	15	16	18	20	21	22	24	25	26	27	28	30
$\varphi(n)$	2	2	4	6	4	4	6	8	8	6	8	12	10	8	20	12	18	12	8

We now learn how to compute φ. First, we generalize Fermat's little theorem:

Theorem 2.6.8 (Euler)**.** *Let $n > 1$ and $x \in \mathbb{Z}$ be coprime to n. Then*

$$x^{\varphi(n)} \equiv 1 \mod n.$$

Proof. We have $\bar{x} \in \mathbb{Z}_n^\times$. Let $\bar{x}_1, \ldots, \bar{x}_{\varphi(n)}$ list $\mathbb{Z}_n^\times$. Since $\mathbb{Z}_n^\times$ is a group, $\bar{y} \mapsto \bar{x} \cdot \bar{y}$ permutes $\mathbb{Z}_n^\times$ (cf. Exercise 1.1.3). Thus, $\bar{x}^{\varphi(n)} = 1$ follows by cancellation from

$$\bar{x}_1 \cdots \bar{x}_{\varphi(n)} = (\bar{x}\bar{x}_1) \cdots (\bar{x}\bar{x}_{\varphi(n)}) = \bar{x}^{\varphi(n)} \cdot \bar{x}_1 \cdots \bar{x}_{\varphi(n)}. \qquad \square$$

Lemma 2.6.9 (Multiplicativity)**.** $\varphi(nm) = \varphi(n)\varphi(m)$ *for all coprime* $n, m > 0$.

Proof. We can assume $n, m > 1$. By Corollary 2.5.11, $\mathbb{Z}_{nm} \cong \mathbb{Z}_n \times \mathbb{Z}_m$. The isomorphism maps units onto units, i.e., $\mathbb{Z}_{nm}^\times$ onto $(\mathbb{Z}_n \times \mathbb{Z}_m)^\times$. By Lemma 2.5.10, the latter is $\mathbb{Z}_n^\times \times \mathbb{Z}_m^\times$. $\square$

Theorem 2.6.10. *Let $n > 1$ have prime factorization $n = p_1^{k_1} \cdots p_r^{k_r}$. Then*

$$\varphi(n) \;=\; p^{k_1-1}(p_1 - 1) \cdots p^{k_r-1}(p_r - 1) \;=\; n \cdot (1 - 1/p_1) \cdots (1 - 1/p_r).$$

Proof. By Lemma 2.6.9 it suffices to show $\varphi(p^k) = (p^k - p^{k-1})$ for primes p and $k > 0$. For this is suffices to show there are exactly p^{k-1} numbers $1 < x < p^k$ which are *not* coprime to p^k. But $\gcd(x, p^k) > 1$ means $p \mid x$, so $x = pm$ for some $1 \leqslant m \leqslant p^{k-1}$. $\square$

Exercise 2.6.11. How many $0 < n < 3000$ are coprime to 3000? Show $2237^{800} - 1$ is divisible by 3000. Is $2^{46} - 1$ prime?

Theorem 2.6.12 (Totient sum formula)**.** *Let $n > 0$. Then* $\sum_{d \mid n} \varphi(d) = n$.

Proof. We first treat the case $n = p^k$ for a prime p and $k > 0$. Then our sum is

$$\varphi(1) + \varphi(p) + \cdots + \varphi(p^k) = 1 + (p - 1) + p(p - 1) + \cdots + p^{k-1}(p - 1) = p^k.$$

Now proceed by induction. For $n = 1$ our claim is trivial. Given $n > 1$ let write $n = p^k m$ with p prime and $p \nmid m$. The divisors of n are $d, dp, \ldots, dp^k$ where $d \mid m$. Hence our sum is

$$\textstyle\sum_{d \mid m} \varphi(d) + \sum_{d \mid m} \varphi(dp) + \cdots + \sum_{d \mid m} \varphi(dp^k) = (1 + \varphi(p) + \cdots + \varphi(p^k)) \cdot \sum_{d \mid m} \varphi(d).$$

The 1st factor is p^k as seen above, the 2nd is m by induction. $\square$

We shall later see a more abstract proof (cf. Corollary 5.3.24).

2.6.1 RSA encryption

Rivest, Shamir and Adleman suggested 1977 the following protocol of secret communication that is widely in use today. The basic idea is as follows.

Alice wants to send a secret message to Bob, say of 100 bits, that is, she wants to send a number $m < 2^{100}$. She sends a ciphertext instead the message, a number c instead m. Bob should be able to recover m from c but nobody else, so Bob should know something nobody else does, not even Alice. Here is how it is done.

Bob chooses two primes $q > p > 2^{100}$ and e(ncryption) coprime to $\varphi(pq) = (q-1)(p-1)$. The primes p, q are Bob's secret. He sends $n := pq$ and e to Alice. Alice sends to Bob $c :=$ the remainder of (m^e, n). Bob computes d(ecryption) such that $de \equiv 1 \mod \varphi(n)$ (say, by the Euclidian algorithm). He decrypts the message m as the remainder of (c^d, n).

This works: for some $k \in \mathbb{N}$ this remainder is $\equiv c^d = m^{ed} = m^{1+k\varphi(n)} \equiv m \mod n$ by Euler ($m < p < q$ is coprime to n); both the remainder and m are $< n$, so equal.

Is this encryption secure? Can Eve, seeing c, n, e, compute m? The hope is that Eve needs $\varphi(n)$ to compute d and for that she needs p, q, i.e., to factor n. Many believe (or hope) that factoring large integers cannot be done in reasonable time. There are, however, no results in computational complexity theory that would support this belief.

2.7 Primitive roots

By Euler's Theorem 2.6.8 the following is well-defined:

Definition 2.7.1. Let $n > 1$ and $0 < x < n$ be coprime to n. The *order of x modulo n* is the minimal $k > 0$ such that $x^k \equiv 1 \mod n$. If this is $\varphi(n)$, then x is a *primitive root of n*.

Remark 2.7.2. Let $n > 1$ and x coprime to n of order k modulo n.

1. For all $\ell > 0$, $x^\ell \equiv 1 \mod n$ if and only if $k \mid \ell$.

   $\Leftarrow$: if $\ell = km$, then $x^\ell \equiv (x^k)^m \equiv 1^m \equiv 1 \mod n$.
   $\Rightarrow$: write $\ell = qk + r$ with $0 \leqslant r < k$ and $q \geqslant 0$; then $1 \equiv (x^\ell)^q \cdot x^r \equiv x^r \mod n$, so $r = 0$.
2. $k \mid \varphi(n)$ (by Euler's theorem).
3. For all ℓ, ℓ', $x^\ell \equiv x^{\ell'} \mod n$ if and only if $k \mid (\ell - \ell')$.

   Indeed: say, $\ell \leqslant \ell'$; then $x^\ell \equiv x^{\ell'} \mod n$, $x^{\ell-\ell'} \equiv 1 \mod n$, $k \mid (\ell - \ell')$ are equivalent.
4. For all $\ell \in \mathbb{N}$, x^ℓ has order $k/\gcd(k, \ell)$ modulo n.

   Indeed: $(x^\ell)^j \equiv 1 \mod n \overset{(1)}{\Leftrightarrow} k \mid \ell j \Leftrightarrow k/\gcd(k,\ell) \mid j \cdot \ell/\gcd(k,\ell) \Leftrightarrow k/\gcd(k,\ell) \mid j$ (by Remark 2.1.9 (3), (5)). The minimal such $j > 0$ is $k/\gcd(k, \ell)$.

Examples 2.7.3. $2, 3, 4$ have primitive roots $1, 2, 3$. $\mathbb{Z}_6^\times = \{\bar{1}, \bar{5}\}$ and 5 is a primitive root of 6. The tables below show 5 has primitive roots $2, 3$ and 7 has primitive roots $3, 5$.

8 does not have primitive roots: $\mathbb{Z}_8^\times = \{\bar{1}, \bar{3}, \bar{5}, \bar{7}\}$ and $\bar{3}^2 = \bar{5}^2 = \bar{7}^2 = \bar{1}$. Seeing Corollary 2.7.5 below, note there are > 2 many $1 \leqslant x < 8$ with $x^2 \equiv 1 \mod 8$.

$\mathbb{Z}_5^\times$	x^2	x^3	x^4	order
$\bar{2}$	$\bar{4}$	$\bar{3}$	$\bar{1}$	4
$\bar{3}$	$\bar{4}$	$\bar{2}$	$\bar{1}$	4
$\bar{4}$	$\bar{1}$			2

$\mathbb{Z}_7^\times$	x^2	x^3	x^4	x^5	x^6	order
$\bar{2}$	$\bar{4}$	$\bar{1}$				3
$\bar{3}$	$\bar{2}$	$\bar{6}$	$\bar{4}$	$\bar{3}$	$\bar{1}$	6
$\bar{4}$	$\bar{2}$	$\bar{1}$				3
$\bar{5}$	$\bar{4}$	$\bar{6}$	$\bar{2}$	$\bar{3}$	$\bar{1}$	6
$\bar{6}$	$\bar{1}$					2

Lemma 2.7.4. *If $n > 1$ has a primitive root, then $(\mathbb{Z}_{\varphi(n)}, +) \cong (\mathbb{Z}_n^\times, \cdot)$.*

Proof. If x is a primitive root of $n > 1$, then $[1]_n, [x]_n, [x^2]_n, ..., [x^{\varphi(n)-1}]_n$ lists $\mathbb{Z}_n^\times$ and the map $[i]_{\varphi(n)} \mapsto [x^i]_n$ for $0 \leqslant i < \varphi(n)$ is an isomorphism. □

This lemma turns multiplicative problems into additive ones. E.g.:

Corollary 2.7.5. *Assume there exists a primitive root of $n > 1$ and let $y \in \mathbb{Z}$ be coprime with n. Then there are 0 or 2 many $1 \leqslant x < n$ such that $x^2 \equiv y \mod n$.*

Proof. Note no such x can be a zero-divisor in the ring $\mathbb{Z}_n$, so $x \in \mathbb{Z}_n^\times$. Write $m := \varphi(n)$. By the isomorphism the number of such x is the same as the number of $[z]_m \in \mathbb{Z}_m$ such that $[z]_m + [z]_m = [2]_m \cdot [z]_m = [u]_m$ where $[u]_m \in \mathbb{Z}_m$ corresponds to $[y]_n \in \mathbb{Z}_n^\times$ under the isomorphism. By Lemma 2.6.1, there are 0 or 2 of them. □

We saw $2, 3$ have $1 = \varphi(2-1) = \varphi(3-1)$ primitive roots, and $5, 7$ have $2 = \varphi(5-1) = \varphi(7-1)$ many. This is generally so:

Theorem 2.7.6. *If p is prime and $d \mid p-1$, then there are exactly $\varphi(d)$ many naturals below p of order d modulo p. In particular, there are exactly $\varphi(p-1)$ primitive roots of p.*

Proof. For $d \mid \varphi(p) = p-1$ let $\psi(d)$ be the number of $0 < x < p$ that have order d modulo p. Then $\sum_{d \mid p-1} \psi(d) = p-1$. By Theorem 2.6.12 it suffices to show $\psi(d) \leqslant \varphi(d)$ for all $d \mid p-1$.

Analogously to the corollary, one sees that there are $\leqslant d$ many $0 < y < p$ such that $y^d \equiv 1 \mod p$ (alternatively use Corollary 3.3.3 in the field $\mathbb{Z}_p = \mathbb{F}_p$).

Assume $\psi(d) > 0$, so there exists x of order d modulo p. Choose $0 < x_0, ..., x_{d-1} < p$ with $x_0 = 1, x_1 = x, [x_2]_p = [x^2]_p, ..., [x_{d-1}]_p = [x^{d-1}]_p$. As the x_i are pairwise distinct, all y of order d modulo p appear. By Remark 2.7.2 (4), x_i has order d modulo p if and only if $\gcd(i, d) = 1$. These are $\varphi(d)$ many. Thus, $\psi(d) \leqslant \varphi(d)$. □

Exercise 2.7.7. Assume $n > 1$ has a primitive root and let $d \in \mathbb{N}$. How many $0 < x < n$ coprime to n have order d modulo n?

8 is the first natural without a primitive root. What's so special about 8?

Theorem 2.7.8. *Let $n > 1$. Then there exist primitive roots of n if and only if n equals 4 or p^k or $2p^k$ for a prime $p > 2$ and $k \in \mathbb{N}$.*

Proof. $\Rightarrow$: We first show 2^k does not have a primitive root modulo for $k > 2$. Clearly, such a root would be an odd x, say $x = 2y + 1$. Then $x^2 = 4y^2 + 4y + 1 = 4y(y+1) + 1 = 8z + 1$ for some $z \in \mathbb{N}$. And $x^4 = 8^2z^2 + 16z + 1 = 16u + 1$ for some $u \in \mathbb{N}$. And so on. By induction, $x^{2^{k-2}} = 2^k v + 1$ for some $v \in \mathbb{N}$. Hence, $x^{2^{k-2}} \equiv 1 \mod 2^k$, so x has order $\leqslant 2^{k-2} < 2^{k-1} = \varphi(2^k)$.

Assume there is a primitive root of $n > 1$. By the above, it suffices to show that n cannot be written $n = n_0 n_1$ with coprime $n_0, n_1 > 2$. Otherwise, by Corollary 2.5.11, $\mathbb{Z}_n \cong \mathbb{Z}_{n_0} \times \mathbb{Z}_{n_1}$ as rings. By Corollary 2.7.5, $\mathbb{Z}_{n_0} \times \mathbb{Z}_{n_1}$ has $\leqslant 2$ elements x with $x^2 = 1$. But it has $\geqslant 4$, namely $(\pm 1, \pm 1)$– we have $1 \neq -1$ in both $\mathbb{Z}_{n_0}$ and $\mathbb{Z}_{n_1}$ since $n_0, n_1 > 2$.

$\Leftarrow$: we show p^k has a primitive root. We can assume $k > 1$. Let x be a primitive root of p. Set $y := x + p$ and note $y^{p-1} = x^{p-1} + (p-1)x^{p-2}p + mp^2$ for some $m \in \mathbb{Z}$; note $p-2$ makes sense as $p > 2$. By Fermat $x^{p-1} = 1 + m'p$ for some $m' \in \mathbb{Z}$, so $y^{p-1} = 1 + (m' + (p-1)x^{p-2})p + mp^2$. Note $p \nmid (p-1)x^{p-2}$. We can assume that $p \nmid (m' + (p-1)x^{p-2})$: indeed, one easily checks that otherwise this condition is ensured for $y := x + 2p$.

Thus, $y^{p-1} = 1 + m_0 p$ and for some $m_0 \in \mathbb{Z}$ with $p \nmid m_0$. Now $(y^{p-1})^p = 1 + m_0 p^2 + m'' p^3 = 1 + m_1 p^2$ for some $m'', m_1 \in \mathbb{Z}$ with $p \nmid m_1$. Then $(y^{p(p-1)})^p = 1 + m_1 p^3 + m''' p^4 = 1 + m_2 p^3$ for some $m''', m_2 \in \mathbb{Z}$ with $p \nmid m_2$. Continuing, we get for all $\ell \in \mathbb{N}$ some $m_\ell \in \mathbb{Z}$ with $p \nmid m_\ell$ and $y^{p^\ell(p-1)} = 1 + m_\ell p^{\ell+1}$.

The order e of y modulo p^k divides $\varphi(p^k) = p^{k-1}(p-1)$, so $e = p^\ell d$ for some $\ell \leqslant k-1$ and $d \mid (p-1)$. Then $e \mid p^\ell(p-1)$, so $y^{p^\ell(p-1)} \equiv 1 \mod p^k$, so $1 + m_\ell p^{\ell+1} \equiv 1 \mod p^k$, so $\ell + 1 \geqslant k$. Thus $\ell = k-1$ and $e = p^{k-1}d$.

By Fermat, $y^{p^\ell} \equiv y \equiv x \mod p$, so y^{p^ℓ} has order $p-1$ modulo p. Since $(y^{p^\ell})^d = y^e \equiv 1 \mod p$ we have $p - 1 \mid d$. Hence, $p - 1 = d$. Thus, $e = \varphi(p^k)$ and y is a primitive root of p^k.

Finally, we show $2p^k$ has a primitive root: we want an element x of the ring $\mathbb{Z}_{2p^k}$ whose powers are $\mathbb{Z}^\times_{2p^k}$. By Corollary 2.5.11, we want a pair (x, y) whose powers in $\mathbb{Z}_2 \times \mathbb{Z}_{p^k}$ are $\mathbb{Z}^\times_2 \times \mathbb{Z}^\times_{p^k}$. Easy: set $x := 1$ and $y :=$ a primitive root of p^k. □

Remark 2.7.9. No efficient algorithm is known for determining primitive roots.

Exercise 2.7.10. Carmichael numbers do not have primitive roots (cf. Remark 2.5.9).

Exercise 2.7.11. For every prime p there is $n \in \mathbb{N}$ such that $np + 1$ is prime.

Hint: Let q be a prime divisor of $2^p - 1$ and consider the order of 2 modulo q.

2.7.1 Digital Signature Algorithm

Alice wants to sign a message m to Bob. A signature should be a small amount of information allowing Bob to efficiently verify that indeed Alice was the sender. The public information is p, q, g where $p > q$ are primes such that $p \equiv 1 \mod q$ and $g < p$ has order q modulo p (by Theorem 2.7.6 there are $\varphi(q) = q - 1 > 0$ such gs).

Alice chooses the *private key* $0 < x < q$ at random and keeps it secret; she publishes $y := g^x \mod p$, the *public key*; here, we write $z \mod p$ for the remainder of (z, p).

Each time Alice wants to sign a message $m < q$, she chooses $0 < z < q$ at random and computes $0 < z^{-1} < q$ with $z \cdot z^{-1} \equiv 1 \mod q$; she sends the signature (a, b) where

$$a := (g^z \mod p) \mod q, \quad b := z^{-1}(m + xa) \mod q.$$

Bob, upon receiving a pair (a, b), checks that a equals

$$a' := (g^{mb^{-1}} \cdot y^{ab^{-1}} \mod p) \mod q.$$

This check clears for (a, b) sent by Alice: since $z \equiv (m + xa)b^{-1} \mod q$ we have

$$g^z \equiv g^{(m+xa)b^{-1}} \equiv g^{mb^{-1}} \cdot y^{ab^{-1}} \mod p.$$

Can Eve fake Alice's signature? One hopes, the only way to find (a, b) passing Bob's check is to compute x from p, q, g, y, i.e., to solve $y = g^X \mod p$ – this is known as the *discrete log problem* and believed or hoped to be computationally hard.

To sign large messages $m > q$ one replaces m above by a so-called *cryptographic hash of* m, a number $< q$. One chooses large p for security, and small q to get short signatures. E.g., the U.S. Department of Commerce and National Institute of Standards and Technology (1994) officially recommends 1024 bits for p, and 160 bits for q.

Remark 2.7.12 (Diffie-Hellman key exchange 1976)**.** Alice and Bob can share a secret as follows: Bob sets a private key $\tilde{x}$ and a public key $\tilde{y} := g^{\tilde{x}}$ like Alice. The shared secret is $y^{\tilde{x}} \mod p = \tilde{y}^x \mod p$ computable by both but hopefully not by Eve: one hopes computing $g^{x\tilde{x}} \mod p$ from $g^x, g^{\tilde{x}}$ requires $x, \tilde{x}$, so solving instances of the discrete log problem.

Exercise 2.7.13. Why should Alice avoid using the same z signing different messages?

Exercise 2.7.14 (ElGamal encryption)**.** Bob can encrypt a message $m < p$ to Alice as follows. He chooses secretly a random $z < p$ and sends the *ciphertext* (c_0, c_1) where

$$c_0 := g^z \mod p, \quad c_1 := my^z \mod p.$$

Show Alice (knowing x) can decode by computing $c_0^{p-x-1} c_1 \mod p$.

2.7.2 The Miller-Rabin primality test

How to decide whether a given number n is prime? A good "witness for compositionality" of n is a prime factor of n. By Exercise 4.2.8, a composite n has a prime factor $\leqslant \sqrt{n}$. We can check whether one exists by first computing all primes $\leqslant \sqrt{n}$.

The *Sieve of Erathostenes* does this as follows: start with a list of all numbers $\leqslant \sqrt{n}$. Mark 1 and all multiples of 2 except 2; mark all multiples of 3 except 3;...; choose the first unmarked number p not yet considered and mark all its multiples except p. The finally unmarked numbers are the primes $\leqslant \sqrt{n}$. This method is unfeasible: for an input n of only 160 digits, the size of the list is about the number of atoms in the observable universe.

Another way to make the point: to decide whether an input n is prime, you run a program that checks for each $m \leqslant \sqrt{n}$ whether $m \mid n$; say n has 40 digits and your computer does 1 billion checks per sec; then, if n is prime, you wait > 1000 years, and 55 digits make you wait more than the age of the universe.

We now design a fast *probabilistic* algorithm that your PC executes in only a split second on much larger inputs – but has a 0.001% chance of error. The key is a concept of "witness" with the property that most numbers $< n$ are witnesses if n is composite.

Definition 2.7.15. Let $n > 1$ be odd and write $n - 1 = 2^t m$ for odd m and $t > 0$. An *RM-witness for* n is a number $1 \leqslant x \leqslant n$ such that

1. $x^{n-1} \not\equiv 1 \mod n$, or,
2. there is $j < t$ such that $x^{2^j m} \not\equiv \pm 1$ and $x^{2^{j+1}m} \equiv 1 \mod n$.

Proposition 2.7.16. *If $n > 1$ is prime, then there is no RM-witness for n.*

Proof. $x^{n-1} \equiv 1 \mod n$ by Fermat. If $x^{2^j m} \not\equiv \pm 1$ and $x^{2^{j+1}m} \equiv 1 \mod n$, then ± 1 and the remainder of $(x^{2^j m}, n)$ are 3 solutions of $X^2 \equiv 1 \mod n$, contradicting Corollary 2.7.5. □

Exercise 2.7.17. A proper subgroup H of a finite group G has size $|H| \leqslant |G|/2$.

Hint: for $x \in G \smallsetminus H$ consider the set of $xh, h \in H$.

Theorem 2.7.18. *Every odd composite $n > 1$ has $> n/2$ many RM-witnesses.*

Proof. Write $n - 1 = 2^t m$ as above. Call $x \in \mathbb{Z}$ *bad* if it is not an RM-witness. Then $\bar{x} \in \mathbb{Z}_n^\times$ since $x \cdot x^{n-2} \equiv 1 \mod n$. By the exercise, it suffices to show that there is a proper subgroup B of $\mathbb{Z}_n^\times$ that contains $\bar{x}$ for every bad x. This is clear in case there exists $\bar{x} \in \mathbb{Z}_n^\times$ such that $x^{n-1} \not\equiv 1 \mod n$: take B to be the set of $\bar{y} \in \mathbb{Z}_n^\times$ such that $y^{n-1} \equiv 1 \mod n$.

So assume $x^{n-1} \equiv 1 \mod n$ for all $\bar{x} \in \mathbb{Z}_n^\times$, i.e., n is Carmichael. By Exercise 2.7.10 (and Theorem 2.7.8), n is not a prime power, so $n = n_0 n_1$ for odd coprime $n_0, n_1 > 1$.

Let $j \leqslant t$ be maximal such that $x^{2^j m} \equiv -1 \mod n$ for some $x \in \mathbb{Z}$ – such j exist, e.g., $(-1)^{2^0 m} \equiv -1 \mod n$. Set

$$B := \left\{ \bar{y} \in \mathbb{Z}_n^\times \mid y^{2^j m} \equiv \pm 1 \mod n \right\}.$$

Clearly, B is a subgroup and contains $\bar{y}$ for every bad y: either all $y^{2^i m}$ are $\equiv 1 \mod n$, or there is $i < t$ such that $y^{2^i m} \not\equiv 1 \mod n$ and $y^{2^{i+1}m} \equiv 1 \mod n$; as y is bad, $y^{2^i m} \equiv -1 \mod n$; by choice of j we have $i \leqslant j$, and hence $\bar{y} \in B$. We are left to show $B \neq \mathbb{Z}_n^\times$.

The Chinese remainder theorem gives z with $z \equiv x \mod n_0$ and $z \equiv 1 \mod n_1$. Then

$$z^{2^j m} \equiv x^{2^j m} \equiv -1 \mod n_0, \quad z^{2^j m} \equiv 1 \mod n_1.$$

This means, the isomorphism of Corollary 2.5.11 maps $\bar{z}^{2^j m}$ to $(-\bar{1}, \bar{1})$. Note $-\bar{1} \neq \bar{1}$ as $n_0 > 2$. Thus $\bar{z}^{2^j m} \neq \pm 1$ in the ring $\mathbb{Z}_n$; hence, $\bar{z} \notin B$. □

The Miller-Rabin test The algorithm takes as input $n > 1$ odd and $k \geqslant 1$ and works as follows: choose $1 \leqslant x_1, \ldots, x_k \leqslant n - 1$ independently and uniformly at random and test for each of them whether it is an RM-witness for n. If there is one, then answer "composite", otherwise answer "prime".

If n is prime, the algorithm answers "prime" for sure (Proposition 2.7.16). If it is composite, it errs and answers "prime" only with probability $< 2^{-k}$ (Theorem 2.7.18).

Some basic algorithmics give efficient RM-witness-checks.

2.8 The law of quadratic reciprocity

This section studies the question whether $X^2 \equiv a \mod n$ has a solution in $\mathbb{Z}$. The Legendre symbol is a cumbersome but established notation for the answer with a special treatment of the case $a \equiv 0 \mod p$:

Definition 2.8.1. Let $n > 1$ and $x \in \mathbb{Z}$. Then x is a *quadratic residue modulo* n if $y^2 \equiv x \mod n$ for some $y \in \mathbb{Z}$; otherwise it is a *quadratic nonresidue modulo* n.

For a prime $p > 2$, the *Legendre symbol* is

$$\left(\frac{x}{p}\right) := \begin{cases} 1 & \text{if } p \nmid x \text{ and } x \text{ is a quadratic residue modulo } p, \\ -1 & \text{if } p \nmid x \text{ and } x \text{ is a quadratic nonresidue modulo } p, \\ 0 & \text{if } p \mid x. \end{cases}$$

Exercise 2.8.2. Let $p > 2$ be prime, $a, b, c \in \mathbb{Z}$ with a coprime to p. Then $aX^2 + bX + c \equiv 0 \mod p$ has a solution if and only if $(b^2 - 4ac)$ is a quadratic residue modulo p.

Examples 2.8.3. Modulo 11 and 13 and 15 we have quadratic residues $1, 3, 4, 5, 9$ and $1, 3, 4, 9, 10, 12$ and $1, 4, 6, 9, 10$:

in $\mathbb{Z}_{11}$

x	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{7}$	$\bar{8}$	$\bar{9}$	$\bar{10}$
x^2	$\bar{1}$	$\bar{4}$	$\bar{9}$	$\bar{5}$	$\bar{3}$	$\bar{3}$	$\bar{5}$	$\bar{9}$	$\bar{4}$	$\bar{1}$

in $\mathbb{Z}_{13}$

x	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{7}$	$\bar{8}$	$\bar{9}$	$\bar{10}$	$\bar{11}$	$\bar{12}$
x^2	$\bar{1}$	$\bar{4}$	$\bar{9}$	$\bar{3}$	$\bar{12}$	$\bar{10}$	$\bar{10}$	$\bar{12}$	$\bar{3}$	$\bar{9}$	$\bar{4}$	$\bar{1}$

in $\mathbb{Z}_{15}$

x	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{7}$	$\bar{8}$	$\bar{9}$	$\bar{10}$	$\bar{11}$	$\bar{12}$	$\bar{13}$	$\bar{14}$
x^2	$\bar{1}$	$\bar{4}$	$\bar{9}$	$\bar{1}$	$\bar{10}$	$\bar{6}$	$\bar{4}$	$\bar{4}$	$\bar{6}$	$\bar{10}$	$\bar{1}$	$\bar{9}$	$\bar{4}$	$\bar{1}$

Proposition 2.8.4. *Let $p > 2$ be prime. Then exactly half of $\{1, \ldots, p-1\}$ are quadratic residues modulo p.*

Proof. By Theorem 2.7.8, p has a primitive root. Map $1 \leqslant x \leqslant p-1$ to the $1 \leqslant y \leqslant p-1$ such that $x^2 \equiv y \mod p$. By Corollary 2.7.5, each value has exactly 2 preimages. □

Theorem 2.8.5 (Euler)**.** *Let $p > 2$ be prime and $x \in \mathbb{Z}$. Then* $\left(\frac{x}{p}\right) \equiv x^{\frac{p-1}{2}} \mod p$.

First proof. We can assume $x \not\equiv 0 \mod p$ and write $x \equiv g^k \mod p$ where $k \in \mathbb{N}$ and g is a primitive root of p. Then $g^{(p-1)/2} \equiv -1 \mod p$ because it is not 1 and its square is 1 by Fermat. Clearly, x is a quadratic residue modulo p if and only if k is even. But

$$x^{(p-1)/2} \equiv (g^{(p-1)/2})^k \equiv (-1)^k \mod p.$$ □

Second proof. If $\left(\frac{x}{p}\right) = 0$, then $x^{(p-1)/2} \equiv 0 \mod p$. If $\left(\frac{x}{p}\right) = 1$, say $y^2 \equiv x \mod p$, then $x^{(p-1)/2} \equiv y^{p-1} \equiv 1 \mod p$ by Fermat (and $y \not\equiv 0 \mod p$). If $\left(\frac{x}{p}\right) = -1$, we argue as Wilson: pair each $1 \leqslant y \leqslant p-1$ with the $1 \leqslant z \leqslant p-1$ such that $yz \equiv x \mod p$; then z is paired with y, and $y \neq z$ as x is a quadratic nonresidue. Thus $(p-1)! \equiv x^{(p-1)/2} \mod p$ and $(p-1)! \equiv -1 \mod p$ by Wilson. □

Corollary 2.8.6. *Let $p > 2$ be prime and $x, y \in \mathbb{Z}$.*

1. $\left(\frac{xy}{p}\right) = \left(\frac{x}{p}\right) \cdot \left(\frac{y}{p}\right)$.
   The product of two quadratic residues or nonresidues is a quadratic residue, and the product of a quadratic residue with a quadratic nonresidue is a quadratic nonresidue.
2. $\left(\frac{x^2}{p}\right) = 1, \quad \left(\frac{x^2 y}{p}\right) = \left(\frac{y}{p}\right), \quad \left(\frac{1}{p}\right) = 1, \quad \left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$.

Lemma 2.8.7 (Gauß)**.** *Let $p > 2$ be prime and x coprime to p. Let k be the number of remainders of $(x,p), (2x,p), ..., (\frac{p-1}{2}x, p)$ that are $> p/2$. Then*

$$\left(\frac{x}{p}\right) = (-1)^k.$$

Proof. Let $r_1, ..., r_k$ list the remainders $> p/2$ and $s_1, ..., s_\ell$ the remainders $< p/2$. Then $k + \ell = (p-1)/2$. Further, the numbers $p - r_i$ are $< p/2$ and pairwise distinct, and distinct from the s_j: assume $p - r_i = s_j$, and, say, r_i is the remainder for $1 \leqslant i^* \leqslant p - 1/2$ and s_j for $1 \leqslant j^* \leqslant p - 1$; then $p - r_i \equiv i^* x \mod p$ and $s_j \equiv j^* x \mod p$, so $i^* x \equiv j^* x \mod p$, so $p \mid x(i^* + j^*)$; since $p \nmid x$, this implies $p \mid i^* + j^* < p/2 + p/2$, a contradiction.

It follows that the $p - r_i$ and s_j list $1, ..., (p-1)/2$. Thus

$$\frac{p-1}{2}! \equiv (-1)^k r_1 \cdots r_n s_1 \cdots s_\ell \equiv (-1)^k x \cdot 2x \cdots \frac{p-1}{2}x \mod p,$$

so $(-1)^k x^{(p-1)/2} \equiv 1 \mod p$. Now apply the previous theorem. □

Lemma 2.8.8. *Let $p > 2$ be prime and x be odd and coprime to p. Then*

$$\left(\frac{x}{p}\right) = (-1)^t \quad \text{where } t := \sum_{i=1}^{(p-1)/2} \lfloor ix/p \rfloor.$$

Proof. Define $r_1, ..., r_k, s_1, ..., s_\ell$ as in the previous proof. E.g., if r_i is the remainder of (y,p), then $y = p \cdot \lfloor y/p \rfloor + r_i$. Observe

$$\begin{aligned}
\sum_{i=1}^{(p-1)/2} ix &= \sum_{i=1}^{(p-1)/2} p \cdot \lfloor ix/p \rfloor + \sum_{i=1}^{k} r_i + \sum_{j=1}^{\ell} s_j, \\
\sum_{i=1}^{(p-1)/2} i &= \sum_{i=1}^{k} (p - r_i) + \sum_{i=1}^{\ell} s_j = kp - \sum_{i=1}^{k} r_i + \sum_{i=1}^{\ell} s_j, \\
(x-1) \sum_{i=1}^{(p-1)/2} i &= p \cdot \Big(\sum_{i=1}^{(p-1)/2} \lfloor ix/p \rfloor - k \Big) + 2 \sum_{i=1}^{k} r_i,
\end{aligned}$$

where the 3rd follows by subtracting the 2nd from the 1st. Its l.h.s. is even as x is odd, so $\sum_{i=1}^{(p-1)/2} \lfloor ix/p \rfloor - k \equiv 0 \mod 2$. The lemma follows. □

Corollary 2.8.9. *Let $p > 2$ be prime. Then* $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$.

Proof. Note $\sum_{i=1}^{(p-1)/2} i = (p-1)/2\cdot((p-1)+2)/2\cdot 1/2 = (p^2-1)/8$. The last displayed formula above gives $(p^2-1)/8 \equiv \sum_{i=1}^{(p-1)/2}\lfloor i2/p\rfloor - k \mod 2$ and all $\lfloor i2/p\rfloor = 0$. □

Exercise 2.8.10. Let $p > 2$ be prime. Then $X^2 \equiv 2 \mod p$ has a solution if and only if $p \equiv \pm 1 \mod 8$. And $X^2 \equiv -1 \mod p$ has a solution if and only if $p \equiv 1 \mod 4$.

The following result is very surprising. Gauß proved it 1801 and referred to it as the "theorema aureum".

Theorem 2.8.11 (Quadratic reciprocity)**.** *If $p, q > 2$ are distinct primes, then*

$$\left(\frac{p}{q}\right)\cdot\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Proof. Let S be the set of pairs (x, y) with $1 \leqslant x \leqslant p-1$ and $1 \leqslant q \leqslant q-1$; it has size $(p-1)(q-1)/4$. Partition S into S_0, S_1 where S_0 contains the (x, y) with $qx > py$ and S_1 the (x, y) with $qx < py$; note $qx = py$ is impossible. Note $(x, y) \in S_0$ if and only if $1 \leqslant x \leqslant p-1$ and $1 \leqslant y < qx/p$, so $|S_0| = \sum_{x=1}^{(p-1)/2}\lfloor qx/p\rfloor$. Similarly, $|S_1| = \sum_{y=1}^{(q-1)/2}\lfloor py/q\rfloor$.

Now Lemma 2.8.8 implies the theorem:

$$\sum_{x=1}^{(p-1)/2}\lfloor qx/p\rfloor + \sum_{y=1}^{(q-1)/2}\lfloor py/q\rfloor = \frac{p-1}{2}\cdot\frac{q-1}{2}.$$ □

Remark 2.8.12. Note $(p-1)/2$ is even if and only if $p \equiv 1 \mod 4$. Hence another way to phrase the reciprocity law is:

$$\left(\tfrac{p}{q}\right) = \left(\tfrac{q}{p}\right) \quad \text{if } p \text{ or } q \text{ is } \equiv 1 \mod 4,$$
$$\left(\tfrac{p}{q}\right) = -\left(\tfrac{q}{p}\right) \quad \text{if both } p \text{ and } q \text{ are } \equiv 3 \mod 4.$$

The reciprocity law enables very quick computations – if one has prime factorizations:

Example 2.8.13. Is 1001 a quadratic residue modulo the prime 99991? Write $1001 = 7\cdot 11\cdot 13$ and $\left(\frac{1001}{9991}\right) = \left(\frac{7}{99991}\right)\left(\frac{11}{9991}\right)\left(\frac{13}{99991}\right)$ and consider each factor in turn.

- $\left(\frac{7}{99991}\right) = -\left(\frac{99991}{7}\right)$ since both numbers are $3 \mod 4$, $= -\left(\frac{3}{7}\right)$ as $99991 \equiv 3 \mod 7$, $= \left(\frac{7}{3}\right)$ since both numbers are $3 \mod 4$, $= \left(\frac{1}{3}\right) = 1$ since $7 \equiv 1 \mod 3$.
- $\left(\frac{11}{99991}\right) = -\left(\frac{99991}{11}\right) = -\left(\frac{1}{11}\right) = -1$ since $99991 \equiv 1 \mod 11$.
- $\left(\frac{13}{99991}\right) = \left(\frac{99991}{13}\right)$ since $13 \equiv 1 \mod 4$, $= \left(\frac{2^2\cdot 2}{13}\right)$ since $99991 \equiv 8 \mod 13$, $= \left(\frac{2}{13}\right)$ by Corollary 2.8.6 (2), $= (-1)^{(13^2-1)/8} = (-1)^{21} = -1$ by Corollary 2.8.9.

Thus the answer is yes. Indeed, one can verify $38521^2 \equiv 1001 \mod 99991$.

Example 2.8.14. Here is why the prime 773 is a quadratic residue modulo the prime 1373: since $773 \equiv 1 \mod 4$, we have $\left(\frac{773}{1373}\right) = \left(\frac{1373}{773}\right) = \left(\frac{600}{773}\right) = \left(\frac{2^3\cdot 5^2\cdot 3}{773}\right) = \left(\frac{2}{773}\right)\cdot\left(\frac{3}{773}\right)$ by Corollary 2.8.6 (2). Both factors are -1: $\left(\frac{2}{773}\right) = (-1)^{(773^2-1)/8} = (-1)^{74691} = -1$ by Corollary 2.8.9; and $\left(\frac{3}{773}\right) = \left(\frac{773}{3}\right) = \left(\frac{2}{3}\right)$ as $773 \equiv 2 \mod 3$, $= (-1)^{(3^2-1)/8} = -1$.

Exercise 2.8.15. 9907 is prime. Compute $\left(\frac{1001}{9907}\right)$.

2.9 The Jacobi symbol

Definition 2.9.1. The *Jacobi symbol* $\left(\frac{x}{n}\right)$ extends the Legendre symbol to odd naturals $n > 1$. If $n = p_1 \cdots p_\ell$ for odd primes p_i, then, using the Legendre symbol on the r.h.s.,

$$\left(\frac{x}{n}\right) := \left(\frac{x}{p_1}\right) \cdots \left(\frac{x}{p_\ell}\right).$$

Lemma 2.9.2. *Let $x, y \in \mathbb{Z}$ and $n, m > 1$ be odd.*

1. $\left(\frac{xy}{n}\right) = \left(\frac{x}{n}\right) \cdot \left(\frac{y}{n}\right), \quad \left(\frac{x}{nm}\right) = \left(\frac{x}{n}\right) \cdot \left(\frac{x}{m}\right)$.
2. *If* $\gcd(x, n) > 1$, *then* $\left(\frac{x}{n}\right) = 0$.
3. *If* $\gcd(x, n) = 1$, *then* $\left(\frac{x^2}{n}\right) = \left(\frac{x}{n^2}\right) = 1$.
4. *If* $x \equiv y \mod n$, *then* $\left(\frac{x}{n}\right) = \left(\frac{y}{n}\right)$.
5. $\left(\frac{-1}{n}\right) = (-1)^{(n-1)/2}, \quad \left(\frac{2}{n}\right) = (-1)^{(n^2-1)/8}$.
6. *If* $\gcd(n, m) = 1$, *then* $\left(\frac{n}{m}\right) \cdot \left(\frac{m}{n}\right) = (-1)^{\frac{n-1}{2} \cdot \frac{m-1}{2}}$.

Proof. (1)-(4) are easy. For (5), let $n = p_1^{k_1} \cdots p_r^{k_r}$ be the prime factorization. For the first statement, recall Corollary 2.8.6 (2) and note

$$\left(\frac{-1}{n}\right) = \prod_{i=1}^{r} \left(\frac{-1}{p_i}\right)^{k_i} = \prod_{i=1}^{r} ((-1)^{(p_i-1)/2})^{k_i} = (-1)^k, \text{ where } k := \sum_{i=1}^{r} k_i (p_i - 1)/2.$$

We claim $(n-1)/2 \equiv k \mod 2$. Since $p_i - 1$ is even, the binomial formula implies

$$p_i^{k_i} = (1 + (p_i - 1))^{k_i} \equiv 1 + k_i(p_i - 1) \mod 4.$$

Noting $(p_i - 1)(p_j - 1) \equiv 0 \mod 4$ we get

$$n \equiv \prod_{i=1}^{r} (1 + k_i(p_i - 1) \equiv 1 + \sum_{i=1}^{r} k_i(p_i - 1) \equiv 1 + 2k \mod 4.$$

This implies $(n-1)/2 \equiv k \mod 2$ (by Remark 2.4.3 (4)), and thus our claim.

The 2nd statement of (5) is proved similarly using Corollary 2.8.9 and $8 \mid n^2 - 1$. For (6), let $m = q_1^{\ell_1} \cdots q_s^{\ell_s}$ be the prime factorization and use the law of quadratic reciprocity:

$$\left(\frac{n}{m}\right) \cdot \left(\frac{m}{n}\right) = \prod_{ij} \left(\left(\frac{p_i}{q_j}\right) \cdot \left(\frac{q_j}{p_i}\right)\right)^{k_i \ell_j} = (-1)^{k \cdot \ell}, \text{ where } \begin{array}{l} k := \sum_i k_i (p_i - 1)/2 \\ \ell := \sum_j \ell_j (q_j - 1)/2. \end{array}$$

Note $p_i \neq q_j$ as n, m are coprime. As seen above, $k\ell \equiv (n-1)/2 \cdot (m-1)/2 \mod 2$. □

Proposition 2.9.3. *Let $n > 1$ be odd and $x \in \mathbb{Z}$ coprime to n. If x is a quadratic residue modulo n, then* $\left(\frac{x}{n}\right) = 1$.

Proof. Let $n = p_1 \cdots p_\ell$ for odd primes p_i. If $y^2 \equiv x \mod n$, then $y^2 = x \mod p_i$. Since $p_i \nmid x$, $\left(\frac{x}{p_i}\right) = 1$. Hence, $\left(\frac{x}{n}\right) = \left(\frac{x}{p_1}\right) \cdots \left(\frac{x}{p_\ell}\right) = 1$. □

Example 2.9.4. The converse is false: $\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = (-1)(-1) = 1$ and $\left(\frac{8}{15}\right) = \left(\frac{2}{15}\right)^3 = 1$ but $2, 8$ are quadratic nonresidues modulo 15 (Example 2.8.3).

Remark 2.9.5. One can show, using the lemma and some basic algorithmics, that the Jacobi symbol can be efficiently computed. In contrast, computing the Legendre symbol seems to require prime factorizations for which no efficient algorithm is known.

Example 2.9.6. We compute $\left(\frac{1001}{9907}\right)$ from Exercise 2.8.15 using the Jacobi instead of the Legendre symbol: $\left(\frac{1001}{9907}\right) = \left(\frac{9907}{1001}\right)$ since $(1001-1)/2$ is even (hence $(-1)^{(1001-1)/2 \cdot (9907-1)/2} = 1$), so $= \left(\frac{898}{1001}\right)$ as $9907 \equiv 898 \mod 1001$, so $= \left(\frac{2}{1001}\right)\left(\frac{449}{1001}\right) = \left(\frac{449}{1001}\right)$ since $(-1)^{(1001^2-1)/8} = 1$, so $= \left(\frac{1001}{449}\right)$ as $(1001-1)/2$ is even, so $= \left(\frac{103}{449}\right)$ as $1001 \equiv 103 \mod 449$, so $= \left(\frac{449}{103}\right)$ as $(449-1)/2$ is even, so $= \left(\frac{37}{103}\right)$ as $449 \equiv 37 \mod 103$, so $= \left(\frac{103}{37}\right)$ as $(37-1)/2$ is even, so $= \left(\frac{29}{37}\right)$ as $103 \equiv 29 \mod 37$, so $= \left(\frac{37}{29}\right) = \left(\frac{8}{29}\right) = \left(\frac{2}{29}\right)^3 = (-1)^{3(29^2-1)/8} = (-1)^{305} = -1$.

Exercise 2.9.7. $511 = 7 \cdot 73$ is not prime. Compute $\left(\frac{163}{511}\right)$.

2.9.1 The Solovay-Strassen primality test

Definition 2.9.8. Let $n > 1$ be odd. An *Euler-Jacobi-* or *EJ-witness for* n is a number $1 \leqslant x \leqslant n-1$ coprime to n such that $\left(\frac{x}{n}\right) \not\equiv x^{(n-1)/2} \mod n$.

Theorem 2.9.9. *An odd $n > 1$ is composite if and only if there is an EJ-witness for n.*

Proof. $\Leftarrow$ is clear by Euler's theorem 2.8.5. $\Rightarrow$: assume n is composite.

Case 1: $n = p_1 \cdots p_r$ for distinct primes $p_i > 2$ and $r > 1$. Choose $1 \leqslant y \leqslant p_1 - 1$ with $\left(\frac{y}{p_1}\right) = -1$ (Proposition 2.8.4). By the Chinese remainder theorem choose $1 \leqslant x \leqslant n-1$ such that $x \equiv y \mod p_1$ and $x \equiv 1 \mod p_i$ for $i \neq 1$. Then no p_j divides x, so x is coprime to n. Then $\left(\frac{x}{p_1}\right) = \left(\frac{x}{p_1}\right) = -1$ and $\left(\frac{x}{p_i}\right) = \left(\frac{1}{p_i}\right) = 1$ for $i \neq 1$, so $\left(\frac{x}{n}\right) = \prod_j \left(\frac{x}{p_j}\right) = -1$. But $-1 \not\equiv x^{(n-1)/2} \mod n$. Otherwise, as $p_2 \mid n$, also $-1 \equiv x^{(n-1)/2} \equiv 1 \mod p_2$, a contradiction as $p_2 > 2$.

Case 2: $n = p^k m$ for some prime $p > 2, k > 1$ and $m \in \mathbb{N}$ with $p \nmid m$. The Chinese remainder theorem gives $1 \leqslant x \leqslant n-1$ such that $x \equiv (1+p) \mod p^2$ and $x \equiv 1 \mod m$. Then $\gcd(x, n) = 1$. Indeed: $p \nmid x$ (otherwise $p \nmid x - (1+p)$ but $p^2 \mid x - (1+p)$) and no divisor q of m divides x (otherwise $q \nmid x - 1$ but $m \mid x - 1$). Assume x is not an EJ-witness of n. Then

$$x^{n-1} = (x^{(n-1)/2})^2 \equiv \left(\frac{x}{n}\right)^2 = 1 \mod n,$$

Since $p^2 \mid n$ we have $1 \equiv x^{n-1} \equiv (1+p)^{n-1} \mod p^2$ and this is $\equiv 1 + (n-1)p \mod p^2$ by the binomial theorem. Hence, $0 \equiv (n-1)p \mod p^2$. Then $p \mid n-1$, contradicting $p \mid n$. □

Lemma 2.9.10. *If $n > 1$ is odd and composite, then at least half of all $1 \leqslant m \leqslant n-1$ coprime to n are EJ-witnesses for n.*

Proof. Call $m \in \mathbb{Z}$ *bad* if it is coprime to n and $\left(\frac{m}{n}\right) \equiv m^{(n-1)/2} \mod n$. Clearly, 1 is bad and, by Lemma 2.9.2 (1), if $m, \tilde{m}$ are bad, then so is $m\tilde{m}$. Thus, the residue classes of bad $m \in \mathbb{Z}$ form a subgroup of $\mathbb{Z}_n^\times$. It is proper by the previous theorem. By Exercise 2.7.17, its size is $\leqslant |\mathbb{Z}_n^\times|/2$. This is our claim. □

The Solovay-Strassen test The algorithm takes as input $n > 1$ odd and $k \geqslant 1$ and works as follows: choose $1 \leqslant x_1, ..., x_k \leqslant n-1$ independently and uniformly at random and check for each of them whether $\gcd(n, x_i) > 1$ or $\left(\frac{x_i}{n}\right) \not\equiv x_i^{(n-1)/2} \mod n$. If there is one, then answer "composite", otherwise answer "prime".

If n is prime, then this algorithm answers "prime" for sure (Theorem 2.9.9). If n is composite, then each choice of x_i has probability $\geqslant 1/2$ to either have a nontrivial divisor with n (and then n is not prime) or to satisfy $\left(\frac{x_i}{n}\right) \not\equiv x_i^{(n-1)/2} \mod n$. Hence the algorithm errs and answers "prime" only with probability $< 2^{-k}$.

Each check is efficient: we already remarked that the Jacobi symbol can be efficiently computed, and $\gcd(n, m_i)$ can also be efficiently computed using Euclid's algorithm.

Remark 2.9.11 (Computational complexity theory)**.** The *Cobham-Edmonds Thesis* states that a property P of (one or a tuple of) natural numbers is "efficiently decidable" (informal concept) if and only if P is decidable in polynomial time – a formal concept. *Decidability* means that there is an algorithm that given any input $n \in \mathbb{N}$ performs at most a finite sequence of basic computational steps and halts with output 1 or 0 according to whether $n \in P$ or not. It is a matter of no consequence how one defines "basic computational step". Being *polynomial time* means that on an input of length ℓ the number of steps is $O(\ell^c)$ for some constant $c \in \mathbb{N}$ (i.e., independent of the input). Here, the *length* of an input $n \in \mathbb{N}$ is the length of the binary representation of n, so $\ell = \lceil \log(n+1) \rceil$.

E.g., primality is decided by an algorithm that on input n checks for all $m \leqslant \sqrt{n}$ whether $m \mid n$. This can take more than $\sqrt{n} \approx 2^{\ell/2}$ steps, so is *exponential time*. In contrast, the property of (x, n) that x is an RM-witness for n, or the property of (x, n) that $\gcd(x, n) > 1$ or x is an EJ-witness of n are polynomial time decidable. The input length is $\approx \log x + \log n$, so this means the property is decidable in time $O((\log x + \log n)^c)$ for some constant $c \in \mathbb{N}$.

Probabilistic algorithms toss fair coins during their computations, so their output becomes a random variable. In the terminology of computational complexity theory, both the Rabin-Miller test and the Solovay-Strassen test give probabilistic polynomial time algorithms deciding primality with one-sided error (*one-sided* because they do not err when the input is prime). In 2002, Agrawal, Kayal and Saxena found a deterministic polynomial time algorithm for primality. This is outside the scope of this course.

Chapter 3

Polynomials

3.1 Univariate polynomials

In this section, let R, S be commutative rings. A polynomial over R is an expression of the form $a_n X^n + \cdots + a_1 X + a_0$ with $n \in \mathbb{N}, a_i \in R$. But what is an "expression"? In what sense is e.g. $3X^3 + 2X + 1$ the same as $0 \cdot X^2 + 1 + 2X + 3X^3$?

We are used to judge equality of polynomials by comparing coefficients. The idea is thus to *define* a polynomial as a sequence of coefficients, one for each power of X and only finitely many $\neq 0$. E.g. the two "expressions" above become $(1, 2, 0, 3, 0, 0, ...)$.

Definition 3.1.1. Let $R[X]$ be the set of *(univariate) polynomials (over R)*: sequences $(a_k)_{k\in\mathbb{N}}$ with $a_k \in R$ for all $k \in \mathbb{N}$ and $a_k \neq 0$ for only finitely many $k \in \mathbb{N}$. Given another polynomial $(b_k)_{k\in\mathbb{N}}$ define $+$ componentwise and $\cdot$ as the *Cauchy product*:

$$(a_k)_k + (b_k)_k := (a_k + b_k)_k, \quad (a_k)_k \cdot (b_k)_k := \Big(\sum_{i+j=k} a_i \cdot b_j \Big)_k.$$

Remark 3.1.2.

1. $(R[X], +, \cdot)$ is a commutative ring with additive and multiplicative neutral elements $(0, 0, ...)$, the *zero polynomial*, and $(1, 0, ...)$. The additive inverse of $(a_k)_k \in R[X]$ is $-(a_k)_k = (-a_k)_k$ (additive inverse in R on the r.h.s.).
2. $a \mapsto (a, 0, 0, \cdots)$ is a ring monomorphism from R into $R[X]$. We "identify" $a \in R$ with $(a, 0, 0, \cdots)$ and view R as a subring of $R[X]$.
3. Write $X := (0, 1, 0, 0, \cdots)$ and note $X^k = (0, ..., 0, 1, 0, ...)$ for $k \in \mathbb{N}$ (where the 1 is at position k). Given $(a_k)_k \in R[X]$ choose n with $a_k = 0$ for all $k > n$; then

   $$(a_k)_k = a_n X^n + \cdots + a_1 X + a_0.$$

   Formally, e.g., $a_1 X$ stands for $(a_1, 0, 0, ...) \cdot (0, 1, 0, ...)$.

Definition 3.1.3. Let $f = a_n X^n + \cdots + a_0 \in R[X]$ with $a_n \neq 0$. The a_i are *coefficients of f*, a_n is the *lead coefficient*, and a_0 the *constant coefficient*. The *degree of f* is $\deg(f) := n$. The zero polynomial has degree $-\infty$. Polynomials with lead coefficient 1 are *monic*.

Remark 3.1.4. Understanding $-\infty < n$ and $-\infty = -\infty + n = -\infty + -\infty$ for all $n \in \mathbb{N}$, we have for $f, g \in R[X]$:

1. $\deg(f+g) \leqslant \max\{\deg(f), \deg(g)\}$;
2. $\deg(fg) \leqslant \deg(f) + \deg(g)$;
3. if R is an integral domain, then $\deg(fg) = \deg(f) + \deg(g)$.

   Indeed: if $f = a_n X^n + \cdots + a_0, g = b_m X^m + \cdots + b_0$ with $a_n, b_m \neq 0$, then $a_n b_m \neq 0$ as R is an integral domain, and fg has lead coefficient $\sum_{i+j=m+n} a_i b_j = a_n b_m$.

Example 3.1.5. For $f = \bar{2}X + \bar{1} \in \mathbb{Z}_4[X]$ we have $\deg(f) = 1$ and $\deg(f \cdot f) = 0$ since $f^2 = \bar{4}X^2 + \bar{2}X + \bar{2}X + \bar{1} = \bar{1}$.

Lemma 3.1.6. *If R is an integral domain, then so is $R[X]$ and $R[X]^\times = R^\times$.*

Proof. If $f, g \in R[X] \setminus \{0\}$, then their degree is $\geqslant 0$, so $\deg(fg) \geqslant 0$ by the above, so $fg \neq 0$.

To show $R[X]^\times = R^\times$, let $f \in R[X]^\times$. Then $0 = \deg(1) = \deg(ff^{-1}) = \deg(f) + \deg(f^{-1})$ by the above, so $\deg(f) = \deg(f^{-1}) = 0$, i.e., $f, f^{-1} \in R$, so $f \in R^\times$. $\supseteq$ is clear. □

Theorem 3.1.7 (Universal property)**.** *Let $\varphi : R \to S$ a (ring) homomorphism, and $x \in S$. Then there is a unique homomorphism $\Phi : R[X] \to S$ that extends φ and satisfies $\Phi(X) = x$.*

Proof. For $f = a_n X^n + \cdots + a_0 \in R[X]$ with $a_n \neq 0$ define

$$\Phi(f) = \varphi(a_n)x^n + \cdots + \varphi(a_0).$$

Then $\Phi(X) = x$ and Φ extends φ. Uniqueness is clear as every such homomorphism satisfies the equation above. We leave it to the reader to verify that Φ preserves $+, 1$ and treat $\cdot$. Let $g := b_m X^m + \cdots + b_0 \in R[X]$. Then by definition of $\cdot$ and Φ:

$$\Phi(f \cdot g) \;=\; \Phi\Big(\sum_{k=0}^{m+n} \Big(\sum_{i+j=k} a_i b_j \Big) \cdot X^k \Big) = \sum_{k=0}^{m+n} \varphi\Big(\sum_{i+j=k} a_i b_j \Big) \cdot x^k.$$

Note $\sum_{k=0}^{m+n}$ denotes first a sum in $R[X]$ and then in S; $\sum_{i+j=k}$ is a sum in R. Then

$$\Phi(f \cdot g) = \sum_{k=0}^{m+n} \Big(\sum_{i+j=k} \varphi(a_i)\varphi(b_j) \Big) \cdot x^k = \Big(\sum_{i=0}^{n} \varphi(a_i)x^i \Big) \cdot \Big(\sum_{j=0}^{m} \varphi(b_j)x^j \Big) = \Phi(f) \cdot \Phi(g). \qquad \square$$

Remark 3.1.8. There is a unique homomorphism $\Phi : R[X] \to S[X]$ that extends φ and maps X to itself (view $\varphi : R \to S[X]$ and apply the theorem with $x := X$).

Abusing notation we write again φ for Φ. It just replaces coefficients by the φ-values; more precisely, for $f = a_n X^n + \cdots + a_0 \in R[X]$ we have

$$\varphi(f) = \varphi(a_n)X^n + \cdots + \varphi(a_0) \in S[X].$$

Of course polynomials determine functions: we can plug for X a value from R and get another value from R.

Corollary 3.1.9 (Evaluation homomorphism)**.**

1. *Let $a \in R$. Then there is a unique ring homomorphism $\varphi_a : R[X] \to R$ that is the identity on R (i.e., $\varphi_g{\restriction}R = \mathrm{id}_R$) and satisfies $\varphi_a(X) = a$.*

   We write $f(a) := \varphi_a(f)$ for $f \in R[X]$.
2. *Let $g \in R[X]$. Then there is a unique ring homomorphism $\varphi_g : R[X] \to R[X]$ that is the identity on R and satisfies $\varphi_g(X) = g$.*

   We write $f(g) := \varphi_g(f)$ for $f \in R[X]$.

It is important to distinguish a polynomial f from the function $a \mapsto f(a)$:

Example 3.1.10. $f := X^2 + X, g := \bar{0} \in \mathbb{Z}_2[X]$ both determine the function constantly $\bar{0}$.

Definition 3.1.11. Assume R is a subring of S and $A \subseteq S$. The *ring generated by A over R (in S)* is the intersection of all subrings of S containing $R \cup A$; it is denoted $R[A]$. One says, $R[A]$ *results from R (in S) by adjunction of A.*

If $n \in \mathbb{N}$ and $A = \{a_1, ..., a_n\}$ one writes $R[a_1, ..., a_n] := R[A]$.

Remark 3.1.12.

1. This is well-defined because the intersection of a non-empty set of subrings is a subring.
2. $R[A]$ is the smallest subring of S containing $R \cup A$, i.e., $R[A]$ is contained in any other such subring.
3. If $a \in S$, then $R[a] = \{f(a) \in S \mid f \in R[X]\} = \{r_n a^n + \cdots + r_0 \mid n \in \mathbb{N}, r_0, ..., r_n \in R\}$.

   Indeed: $\subseteq$: the r.h.s. is a subring (Exercise 1.1.23) containing $R \cup \{a\}$. $\supseteq$: any subring containing $R \cup \{a\}$ also contains any element of the form $r_n a^n + \cdots + r_0$.

Example 3.1.13. The subring $\mathbb{Z}[1/2, 1/3]$ of $\mathbb{Q}$ contains precisely the rationals $x2^{-n}3^{-m}$ for $x \in \mathbb{Z}$ and $n, m \in \mathbb{N}$.

Proof. The set of these numbers is easily seen to be a subring containing $\mathbb{Z} \cup \{1/2, 1/3\}$ and hence contains $\mathbb{Z}[1/2, 1/3]$; conversely, these numbers are contained in any subring containing $\mathbb{Z} \cup \{1/2, 1/3\}$. □

Proposition 3.1.14. *Every subring R of $\mathbb{Q}$ equals $\mathbb{Z}[\{1/p \mid p \in P\}]$ for $P \subseteq \mathbb{N}$ a set of primes.*

Proof. Clearly, R contains $\mathbb{Z}$, so $R \cup \mathbb{Z}$, so $\mathbb{Z}[R] \subseteq R$; since $R \subseteq \mathbb{Z}[R]$, we have $\mathbb{Z}[R] = R$. We claim $R = \mathbb{Z}[\{1/p \in R \mid p \text{ prime}\}]$. $\supseteq$ is clear. $\subseteq$: if $a/b \in R$, assume a, b coprime and $b > 0$; by Bézout $xa + yb = 1$ for certain $x, y \in \mathbb{Z}$, so $xa/b + y = 1/b \in R$; if $b = p_1 \cdots p_n$ for primes p_i, then $\mathbb{Z}[\{1/p \in R \mid p \text{ prime}\}]$ contains $1/b = 1/p_1 \cdots 1/p_n$, and hence $a/b = a \cdot 1/b$. □

Exercise 3.1.15 (Rational root theorem)**.** Let $f = a_n X^n + \cdots + a_0 \in \mathbb{Z}[X] \subseteq \mathbb{Q}[X]$ with $a_n \neq 0$. Assume $f(a/b) = 0$ for $a/b \in \mathbb{Q}$ with coprime $a, b \in \mathbb{Z}$. Show $a \mid a_0$ and $b \mid a_n$. In particular, $a/b \in \mathbb{Z}$ if f is monic. (Later we generalize this to other rings than $\mathbb{Z}$.)

Exercise 3.1.16. Let K be a field and let $g \in K[X]$. Show $f \mapsto f(g)$ is an automorphism of $K[X]$ if and only if g has degree 1.

3.1.1 Formal power series

A good way to understand a definition is to slightly change it and see what happens. Here, we drop the finiteness condition in the definition of $R[X]$, so allow infinite sums of monomials. The resulting *formal power series* cannot be evaluated like polynomials, so do not determine functions on R. But we can manipulate them much like polynomials.

Definition 3.1.17. The set $R[[X]]$ of *formal power series (over R)* is the set of sequences $(a_k)_{k\in\mathbb{N}}$ with $a_k \in R$. Addition $+$ and multiplication $\cdot$ are defined as for $R[X]$. We write

$$\textstyle\sum_k a_k X^k := (a_k)_k.$$

The *order* of $(a_k)_k \neq (0,0,...)$ is $n := \min\{k \in \mathbb{N} \mid a_k \neq 0\}$.

Remark 3.1.18.

1. $R[[X]]$ is a commutative ring with subring $R[X]$.
2. For $f = (a_k)_k \in R[[X]]$, $X \cdot f = (0, a_0, a_1, ...), X^2 \cdot f = (0, 0, a_0, a_1, ...)$ etc..
3. $(1-X)^{-1} = 1 + X + X^2 + \cdots, \quad (1+X)^{-1} = 1 - X + X^2 - X^3 + \cdots$ (exercise).
4. If R is an integral domain, then so is $R[[X]]$.

   Indeed: let $f := (a_k)_k, g := (b_k)_k \in R[[X]] \setminus \{0\}$ have orders n, m; write $fg = (c_k)_k$; then $c_{n+m} = a_n b_m \neq 0$ as R is an integral domain.

Many polynomials have inverses in $R[[X]]$:

Lemma 3.1.19. *Let R be a commutative ring and $f = \sum_k a_k X^k \in R[[X]]$. Then*

$$f \in R[[X]]^\times \iff a_0 \in R^\times.$$

In case, $f^{-1} = \sum_k b_k X^k$ where $b_0 = a_0^{-1}$ and for all $k > 0$:

$$b_k = -a_0^{-1} \textstyle\sum_{i=1}^k a_i b_{k-i}.$$

Proof. $\Rightarrow$: if $f \cdot \sum_k b_k X^k = (1, 0, ...)$, then $a_0 b_0 = 1$. $\Leftarrow$: assume the b_k's satisfy the recursive equations and write $f \cdot \sum_k b_k X^k = \sum_k c_k X^k$. Then $c_0 = a_0 b_0 = 1$. For $k > 0$,

$$c_k = \textstyle\sum_{i=0}^k a_i b_{k-i} = a_0 b_k + \sum_{i=1}^k a_i b_{k-i} = a_0 b_k + (-a_0) b_k = 0.$$ □

For a field K already a slight extension of $K[[X]]$ is a field:

Definition 3.1.20. Let R be a commutative ring. $R((X))$ is the set of *formal Laurent series*: sequences $(a_k)_{k\in\mathbb{Z}}$ in R such that the are only finitely many $k \in \mathbb{N}$ such that $a_{-k} \neq 0$. For another such sequence $(b_k)_k$ set, as before (note the sums are finite sums in R),

$$(a_k)_k + (b_k)_k := (a_k + b_k)_k, \quad (a_k)_k \cdot (b_k)_k := \big(\textstyle\sum_{i+j=k} a_i b_j\big)_k.$$

Example 3.1.21. Write X for the sequence $(c_k)_k$ with $c_1 = 1$ and $c_k = 0$ for $k \neq 1$. For $\ell \in \mathbb{Z}$, then $X^\ell = (c'_k)_k$ where $c'_\ell = 1$ and $c'_k = 0$ for all $k \neq \ell$. E.g., underlining entry 0 of a sequence, multiplying with X^{-2} or X^2 looks as follows:

$$X^{-2} \cdot (a_k)_k = (..., 0, 1, 0, \underline{0}, 0, ...) \cdot (..., a_{-1}, \underline{a_0}, a_1, ...) = (..., a_0, a_1, \underline{a_2}, a_3, ...),$$
$$X^2 \cdot (a_k)_k = (..., 0, \underline{0}, 0, 1, 0, ...) \cdot (..., a_{-1}, \underline{a_0}, a_1, ...) = (..., a_{-1}, \underline{a_{-2}}, a_{-1}, a_0, ...).$$

Remark 3.1.22. $R((X))$ is a commutative ring and there is a ring monomorphism from $R[[X]]$ into $R((X))$: it maps $f = (a_0, a_1, ...) \in R[[X]]$ to $f^* = (..., 0, 0, a_0, a_1, ...) \in K((X))$. We identify f and f^* and view $R[[X]]$ as a subring of $R((X))$.

Proposition 3.1.23. *If K is a field, then $K((X))$ is a field.*

Proof. Let $f = (a_k)_k \in K((X)) \setminus \{0\}$. The *order of f* is

$$k_0 := \min\{k \in \mathbb{Z} \mid a_k \neq 0\}.$$

Then $g := (c_k)_k := X^{-k_0} \cdot f$ is in $K[[X]]$ with $c_0 \neq 0$. By the lemma, g has an inverse $g^{-1} \in K[[X]]$. Then $X^{-k_0} \cdot g^{-1}$ is an inverse of f in $K((X))$. □

3.2 Polynomial division

We all learned how to do polynomial division in school. Here we prove that it is possible.

Theorem 3.2.1. *Let R be a commutative ring and $f, g \in R[X]$ with $g \neq 0$. Let $b \in R$ be the leading coefficient of g. Then there are $q, r \in R[X]$ and $k \in \mathbb{N}$ such that*

$$b^k f = qg + r \quad \textit{and} \quad \deg(r) < \deg(g).$$

Proof. If $\deg(g) > \deg(f)$, set $q := 0, r := f, k := 0$. So we assume $n := \deg(f) \geqslant \deg(g) \geqslant 0$ and proceed by induction on n.

If $n = 0$, then $f, g \in R$ and $g = b$ and we set $q := f, r := 0, k := 1$. Assume $n > 0$ and write $f = a_n X^n + \cdots + a_0$ and $g = b_m X^m + \cdots + b_0$ with $b = b_m \neq 0$ for some $m \leqslant n$. Set $h := bf - a_n X^{n-m} g$ (understanding $X^0 = 1$). Then $\deg(h) < n$. By induction, there are $q', r' \in R[X]$ and $k \in \mathbb{N}$ such that $b^k h = q'g + r$ with $\deg(r) < m$. Then

$$b^{k+1} f = b^k h + b^k a_d X^{n-m} g = (q' + b^k a_d X^{n-m})g + r.$$ □

Example 3.2.2. As is often the case, the above inductive proof describes a recursive algorithm. Given $f := 3X^3 + X + 1, g := 2X + 1 \in \mathbb{Z}[X]$ compute:

$$h_1 := 2f - 3X^2 g = (6X^3 + 2X + 2) - (6X^3 + 3X^2) = -3X^2 + 2X + 2$$
$$h_2 := 2h_1 + 3X^1 g = (-6X^2 + 4X + 4) + (6X^2 + 3X) = 7X + 4$$
$$h_3 := 2h_2 - 7X^0 g = (14X + 8) - (14X + 7) = 1$$
$$2h_2 = h_3 + 7g = 1 + 7g$$
$$4h_1 = 2h_2 - 6Xg = 1 + (7 - 6X)g$$
$$8f = 4h_1 + 12X^2 g = 1 + (7 - 6X + 12X^2)g$$

Corollary 3.2.3 (Polynomial division)**.** *Let K be a field and $f, g \in K[X]$ with $g \neq 0$. Then there are unique $q, r \in K[X]$ such that $f = qg + r$ and $\deg(r) < \deg(g)$.*

q and r are the quotient *and* remainder *of (f, g).*

Proof. Existence follows from the theorem: divide by $b^k \neq 0$. Uniqueness: assume $f = qg + r = q'g + r'$ where r, r' have degree $< \deg(g)$. Then $(q - q')g = r' - r$ and

$$\deg(g) > \deg(r' - r) = \deg(q - q') + \deg(g),$$

so $\deg(q' - q) = -\infty$, so $q' = q$. This implies $r = r'$. □

Exercise 3.2.4 (Remainder theorem)**.** Let K be a field, $f \in K[X]$ and $a \in K$. Show $f(a)$ is the remainder of $(f, X - a)$. For $K = \mathbb{R}$, assume 5 is the remainder of $(f, X - 1)$, and -1 is the remainder of $(f, X + 2)$. What is the remainder of $(f, (X - 1)(X + 2))$?

We next check the Euclidian algorithm works in $K[X]$ the same way as in $\mathbb{Z}$ – with $\deg(\cdot)$ playing the role of $|\cdot|$. These functions are *Euclidian valuations* (cf. Definition 4.6.8).

Definition 3.2.5. Let K be a field and $f, g \in K[X]$. g is a *divisor of* f, symbolically $g \mid f$, if $f = gh$ for some $h \in K[X]$. For $n > 0$ and $f_1, ..., f_n \in K[X]$ not all zero, a *common divisor of* $f_1, ..., f_n$ is some $g \in K[X]$ such that $g \mid f_i$ for all i.

A *greatest common divisor of* $f_1, ..., f_n$ is a common divisor g that is monic and such that every common divisor of $f_1, ..., f_n$ is a divisor of g. We write $g = \gcd(f_1, ..., f_n)$. If $g = 1$, $f_1, ..., f_n$ are *coprime*.

Remark 3.2.6. There is at most one greatest common divisor g. If g' is another, then $g \mid g'$ and $g' \mid g$. Then g, g' have the same degree, so $g = ag'$ for some $a \in K$. But g is monic, so $a = 1$ and $g = g'$. The reason for requiring monic is just to ensure this uniqueness here.

We now verify existence, thereby justifying the functional notation $\gcd(f_1, ..., f_n)$.

Theorem 3.2.7 (Euclidian algorithm for polynomials)**.** *Let K be a field and $f, g \in K[X]$ with $\deg(f) \geqslant \deg(g) \geqslant 0$ and $g \nmid f$. Let $r_0, r_1, ...$ be given by $r_0 := f, r_1 := g$ and, for $i > 0$,*

$$r_{i+1} := \begin{cases} \text{the remainder of } (r_{i-1}, r_i) & \text{if } r_i \neq 0 \\ 0 & \text{else.} \end{cases}$$

Then $r_{n+1} = 0$ for some $0 < n < \deg(g)$ and $a^{-1}r_n = \gcd(f, g)$ for the minimal such n and a the lead coefficient of r_n.

Moreover, for this n let $s_0, ..., s_n$ and $t_0, ..., t_n$ be the sequences with $s_0 := 1, s_1 := 0$ and $t_0 := 0, t_1 := 1$ and for $0 < i < n$, letting q_i be the quotient of (r_{i-1}, r_i),

$$s_{i+1} := s_{i-1} - q_i s_i, \quad t_{i+1} := t_{i-1} - q_i t_i.$$

Then $r_n = s_n f + t_n g$.

Proof. Note $\deg(g) > r_2 > 0$ as $g \nmid f$, and $r_2 > r_3 > \cdots$ are $\geqslant 0$, so n as claimed exists. Note

$$f = q_1 g + r_2, \quad g = q_2 r_2 + r_3, \quad r_2 = q_3 r_3 + r_4, \quad \ldots \quad r_{n-2} = q_n r_{n-1} + r_n, \quad r_{n-1} = q_n r_n + 0.$$

Work the equations backwards: $r_{n+1} = 0$, so $r_n \mid r_{n-1}$, so $r_n \mid r_{n-2}$, etc., so $r_n \mid r_1 = y$ and $r_n \mid r_0 = x$. Hence r_n and $a^{-1} r_n$ are common divisors of f, g.

Let h be a common divisor of f, g. Work the equations forwards: as $r_2 = f - q_1 g$ we have $h \mid r_2$; as $r_3 = g - q_2 r_2$ we have $h \mid r_3$, etc., so $h \mid r_n$ and $h \mid a^{-1} r_n$.

Finally, we claim $r_i = s_i f + t_i g$ for all $i \leqslant n$. This is true for $i = 0, 1$. Inductively,

$$r_{i+1} = r_{i-1} - q_i r_i = (s_{i-1} f + t_{i-1} g) - q_i (s_i f + t_i g) = s_{i+1} f + t_{i+1} g. \qquad \square$$

Example 3.2.8. In $\mathbb{Q}[X]$, we compute $\gcd(X^6 + 1, X^4 - 1) = X^2 + 1$: polynomial division gives first $X^6 + 1 = X^2 \cdot (X^4 - 1) + (X^2 + 1)$, then $X^4 - 1 = (X^2 - 1)(X^2 + 1)$.

Exercise 3.2.9. For $f := X^3 - X + 1, g := 2X^2 - 3X + 2 \in \mathbb{Q}[X]$ find $s, t \in \mathbb{Q}[X]$ such that $1 = sf + tg$.

Definition 3.2.5 is given with respect to a field K. We justify that the dependence on K is not reflected in the notations, e.g., we just write $g \mid f$ and not, say, $g \mid_K f$.

Remark 3.2.10. Let $L \mid K$ be a field extension and $f, g \in K[X]$. Slightly informally:

1. Since $K[X]$ is a subring in $L[X]$, polynomial equations are true in $K[X]$ if and only if they are true in $L[X]$.
2. Division by remainder done in $K[X]$ gives the same answer as done in $L[X]$;

   Indeed: if $f = qg + r$ by Euclidian division in $K[X]$, then this equation holds also in $L[X]$; by uniqueness, Euclidian division in $L[X]$ also yields $f = qg + r$.
3. $g \mid f$ in $K[X]$ if and only if $g \mid f$ in $L[X]$ (this is just (2) for the case that $r = 0$).
4. The greatest common divisor of f, g is the same, computed in $K[X]$ or $L[X]$ (by (2) and the previous theorem).

The following construction is analogous to $\mathbb{Z}_n$. Section 4.8 gives a general construction.

Exercise 3.2.11 (Polynomial residue classes)**.** Let K be a field and let $g \in K[X], g \neq 0$.

1. On $K[X]$ define an equivalence relation setting $f \sim_g f' \Leftrightarrow g \mid f - f'$. Show that every $f \in K[X]$ is equivalent to exactly one polynomial of degree $< \deg(g)$, namely the remainder of (f, g).
2. Let $K[X]/(g)$ be the set of equivalence classes $[f]_g, f \in K[X]$. Show

   $$[f]_g + [h]_g := [f + h]_g, \quad [f]_g \cdot [h]_g := [fh]_g$$

   are well-defined and make $K[X]/(g)$ a ring. Think about whether it has zero divisors.
3. Let $K_0 \subseteq K$ be a subfield and $K_0[X]/(g)$ be the set of $[f]_g$ with $f \in K_0[X]$. Show this is a subring of $K[X]/(g)$. For $a \in K$ show $K_0[a] \cong K_0[X]/(X - a)$.

3.3 Roots

As in $\mathbb{Z}$ also here Euclidian division is highly consequential.

Definition 3.3.1. Let R be a commutative ring. $a \in R$ is a *root of* $f \in R[X]$ if $f(a) = 0$.

Corollary 3.3.2 (Factor theorem)**.** *Let R be a commutative ring, and $f \in R[X]$. Then $a \in R$ is a root of f if and only if $f = (X-a)g$ for some $g \in R[X]$.*

Proof. $\Leftarrow$ is clear. $\Rightarrow$: by Theorem 3.2.1 write $1^k \cdot f = q(X-a) + r$; then $\deg(r) < 1$ and $0 = f(a) = r(a)$, so $r = 0$ (zero polynomial). □

Corollary 3.3.3. *Let R be an integral domain and $f \in R[X] \setminus \{0\}$. Then f has $\leqslant \deg(f)$ many roots.*

Proof. Induction on $n := \deg(f)$. If $n = 0$, $f \in R \setminus \{0\}$ has 0 roots. Assume $n > 0$. If f has 0 roots, we are done. Otherwise, let $a \in R$ be a root of f and write $f = (X-a)g$ by Corollary 3.3.2. As R is an integral domain, $\deg(g) = n-1$ by Remark 3.1.4 (3). By induction it suffices to show that every root $b \neq a$ of f is a root of g. But $0 = f(b) = (b-a)g(b)$ implies $g(b) = 0$ as $b - a \neq 0$ and R is an integral domain. □

We observed in Example 3.1.10 that distinct polynomials can determine the same function. We now get a better understanding. Recall $\mathbb{F}_p = \mathbb{Z}_p$ denotes the p-element field.

Proposition 3.3.4. *Let $p \in \mathbb{N}$ be a prime number and let $f, g \in \mathbb{F}_p[X]$. Then $f(a) = g(a)$ for all $a \in \mathbb{F}_p$ if and only if $(X^p - X) \mid f - g$ in $\mathbb{F}_p[X]$.*

Proof. $\Leftarrow$ by Fermat's little theorem. $\Rightarrow$: $f - g = (X-\bar{0})\cdots(X-\overline{(p-1)})h$ for some $h \in \mathbb{F}_p[X]$ (Corollary 3.3.2). But $(X-\bar{0})\cdots(X-\overline{(p-1)})$ equals $X^p - X$ because their difference has degree $< p$ and p many roots, so equals 0. □

Example 3.3.5. We saw in Example 2.7.3 (4) that $X^2 - 1$ has roots $\bar{1}, \bar{3}, \bar{5}, \bar{7}$, i.e., $\pm\bar{1}, \pm\bar{3}$ in $\mathbb{Z}_8$. In $\mathbb{Z}_8[X]$ we factor $X^2 - 1 = (X-\bar{1})(X+\bar{1}) = (X-\bar{3})(X+\bar{3})$.

Exercise 3.3.6. If R is an infinite integral domain, then distinct polynomials determine distinct functions.

Corollary 3.3.7 (Interpolation)**.** *Let K be a field, $n > 0$ and $a_1, \ldots, a_n \in K$ pairwise distinct and $b_1, \ldots, b_n \in K$. There is a unique $f \in K[X]$ of degree $< n$ such that $f(a_i) = b_i$ for all i.*

First proof by Lagrange interpolation. Uniqueness: if f' is another such polynomial, then $f - f'$ has degree $< n$ but n roots $a_1, \ldots, a_n$, so $f - f' = 0$.

Existence: set $f := L_1 b_1 + \cdots + L_n b_n$ where $L_j \in K[X]$ satisfies $L_i(a_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$ E.g.,

$$L_i := \frac{(X-a_1)\cdots(X-a_{i-1})(X-a_{i+1})\cdots(X-a_n)}{(a_i-a_1)\cdots(a_i-a_{i-1})(a_i-a_{i+1})\cdots(a_i-a_n)}.$$ □

Second proof by linear algebra. The coefficients $x := (x_1, \ldots, x_n) \in K^n$ of the wanted degree $\leqslant n-1$ polynomial must satisfy $Ax = b$ where A is the $n \times n$ matrix over K with i-th row $a_i^0, a_i^1, \ldots, a_i^{n-1}$ and $b := (b_1, \ldots, b_n) \in K^n$ a column vector. This is a Vandermonde matrix, so invertible, and hence $Ax = b$ for a unique x (namely, $x = A^{-1}b$). □

3.3.1 Multiple roots

Definition 3.3.8. Let K be a field. A $a \in K$ a root of $f \in K[X] \setminus K$ is *multiple* if $(X-a)^k \mid f$ for some $k > 1$; the maximal such $k \geqslant 1$ is the *multiplicity of a wrt f*.

Exercise 3.3.9. Show the multiplicity is well-defined, and k is the multiplicity if and only if $f = (X-a)^k g$ for some $g \in K[X]$ with $g(a) \neq 0$.

Definition 3.3.10. Let K be a field. The *formal derivative of* $f = a_n X^n + \cdots + a_0 \in K[X]$ is

$$f' := \underline{n} a_n X^{n-1} + \cdots + \underline{2} a_2 X + a_1;$$

recall $\underline{n} = 1_K + \cdots + 1_K$ (n times). We define $f^{(0)} := f$ and $f^{(k+1)} := (f^{(k)})'$.

Lemma 3.3.11. *Let K be a field and $f, g \in K[X]$ and $a, b \in K$.*

1. *$(af + bg)' = af' + bg'$ and $(fg)' = f'g + fg'$.*
2. *If $\mathrm{char}(K) = 0$, then $f' = 0$ if and only if $f \in K$.*
3. *If $p := \mathrm{char}(K) > 0$, then $f' = 0$ if and only if $f = g(X^p)$ for some $g \in K[X]$.*

Proof. (1): Linearity is easy to check. For $i, j > 0$ note

$$(X^i X^j)' = (X^{i+j})' = \underline{(i+j)} X^{i+j-1} = \underline{i} X^{i-1} X^j + X^i \underline{j} X^{j-1} = (X^i)' X^j + X^i (X^j)'.$$

Let $f = \sum_{i=0}^n a_i X^i, g = \sum_{j=0}^m b_j X^j$. Then by linearity

$$(fg)' = \textstyle\sum_{ij} a_i b_j (X^i X^j)' = \sum_{ij} a_i b_j (X^i)' X^j + \sum_{ij} a_i b_j X^i (X^j)' = f'g + fg'.$$

(2) and (3): $f' = 0$ if and only if $\underline{n} a_n = \underline{(n-1)} a_{n-1} = \cdots = \underline{2} a_2 = a_1 = 0$. If $\mathrm{char}(K) = 0$, this means $a_n = \cdots = a_1 = 0$, i.e., $f \in K$. If $p = \mathrm{char}(K) > 0$, this means $a_i = 0$ for all $i \leqslant n$ with $p \nmid i$; this means $f = a_{\ell p} X^{\ell p} + a_{(\ell-1)p} X^{(\ell-1)p} + \cdots + a_p X^p + a_0$ where ℓ is maximal with $\ell p \leqslant n$; set $g := a_{\ell p} X^\ell + a_{(\ell-1)p} X^{\ell-1} + \cdots + a_p X + a_0$. □

Example 3.3.12. In $\mathbb{F}_5[X]$, e.g., $(\bar{4} X^{15} + \bar{3} X^{10} + \bar{2} X^5 + \bar{1})' = 0$.

Lemma 3.3.13. *Let K be a field, $f \in K[X] \setminus K$ and $a \in K$ a root of f.*

1. *a is a multiple root of f if and only if $f'(a) = 0$.*
2. *If $\mathrm{char}(K) = 0$, the multiplicity of a wrt f is the maximal $k > 0$ such that*

$$f^{(0)}(a) = f^{(1)}(a) = \cdots = f^{(k-1)}(a) = 0.$$

Proof. (1): write $f = (X-a)g$ for some $g \in K[X]$ (Corollary 3.3.2). Then $f' = g + (X-a)g'$, so $f'(a) = g(a)$. Thus, $f'(a) = 0$ implies $(X-a) \mid g$, so $(X-a)^2 \mid f$ and a is multiple.

Conversely, if a is multiple, then $f = (X-a)((X-a)h)$ for some $h \in K[X]$; by the product rule, $f' = (X-a)h + (X-a)((X-a)h)'$, so $f'(a) = 0$.

(2): let k be the multiplicity of a wrt f. Note $\deg(f) \geqslant k$, so $f^{(j)} \neq 0$ for $j \leqslant k$ – here we use $\mathrm{char}(K) = 0$. We prove that the multiplicity of a wrt $f^{(j)}$ is $k-j$.

If $j = 0$, this is trivial. Assume the claim for $j < k$, and write $f^{(j)} = (X-a)^{k-j}g$ with $g(a) \neq 0$ by Exercise 3.3.9. An easy induction shows $((X-a)^\ell)' = \underline{\ell}(X-a)^{\ell-1}$. Thus,

$$f^{(j+1)} = \underline{(k-j)}(X-a)^{k-j-1}g + (X-a)^{k-j}g' = (X-a)^{k-(j+1)} \cdot \big(\underline{(k-j)}g + (X-a)g'\big).$$

Evaluating the r.h.s. factor on a gives $\underline{(k-j)}g(a)$. This is $\neq 0$ since $\underline{k-j} \neq 0$ and $g(a) \neq 0$ – here we use $\mathrm{char}(K) = 0$ again. This implies the multiplicity of a wrt $f^{(j+1)}$ is $k-(j+1)$. □

3.4 Quotient fields

Where are the roots? Consider the tower

$$\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

The polynomial $2X - 1 \in \mathbb{Z}[X]$ has a root in $\mathbb{Q}$ but not in $\mathbb{Z}$; the polynomial $X^2 - 2 \in \mathbb{Z}[X]$ has a root in $\mathbb{R}$ but not in $\mathbb{Q}$, and $X^2 + 1 \in \mathbb{Z}[X]$ has one in $\mathbb{C}$ but not in $\mathbb{R}$.

Smallest subrings of $\mathbb{C}$ where these polynomials have a root are $\mathbb{Z}[1/2], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[i]$. What are the smallest fields?

Definition 3.4.1. $L \mid K$ *is a field extension* if L is a field and K a subfield of L. Let $A \subseteq L$. The *field generated by A over K (in L)* is the intersection of all subfields of L containing $K \cup A$; it is denoted $K(A)$. One says, $K(A)$ *results from K (in L) by adjunction of A.*

If $n \in \mathbb{N}$ and $A = \{a_1, ..., a_n\}$ one writes $K(a_1, ..., a_n) := K(A)$.

Remark 3.4.2. This is well-defined because the intersection of a set of subfields is a subfield. Clearly, $K(A)$ is the smallest subfield of L containing $K \cup A$, i.e., $K(A)$ is contained in any other such subfield.

We generalize the construction of $\mathbb{Q}$ from $\mathbb{Z}$ in Section 1.4.

Definition 3.4.3. Let R be an integral domain. $\mathrm{Quot}(R)$ is the set of equivalence classes a/b of $(a, b) \in R \times (R \setminus \{0\})$ under the equivalence relation $\sim \subseteq (R \times (R \setminus \{0\}))^2$ defined by

$$(a, b) \sim (a', b') \iff ab' = a'b.$$

$\mathrm{Quot}(R)$ is *quotient field of R* with $+, \cdot$ defined for $x = a/b, y = c/d \in \mathrm{Quot}(R)$ by

$$x + y := (ad + cb)/bd, \quad x \cdot y := ac/bd.$$

Remark 3.4.4. Exactly as done in Section 1.4 for $R = \mathbb{Z}$ one verifies:

1. $\sim$ is an equivalence relation, $+, \cdot$ are well-defined and $\mathrm{Quot}(R)$ is a field.
2. $a \mapsto a/1$ is a ring monomorphism from R into $\mathrm{Quot}(R)$. If R is a field, it is an isomorphism (surjective because $a/b = ab^{-1}/1$).

3. We "identify" a with $a/1$ and view R as a subring of $\mathrm{Quot}(R)$.

$\mathrm{Quot}(R)$ is the smallest field extending R:

Theorem 3.4.5 (Universal property)**.** *Let R be an integral domain, K a field and $\varphi : R \to K$ a ring monomorphism. Then there is a unique field monomorphism $\Phi : \mathrm{Quot}(R) \to K$ that extends φ; for $a, b \in R$ with $b \neq 0$ we have*

$$\Phi(a/b) = \varphi(a) \cdot \varphi(b)^{-1}.$$

Proof. The equality follows from $\Phi(a/b) = \Phi(a/1)\Phi(1/b) = \varphi(a)\varphi(b)^{-1}$ since $b^{-1} = 1/b$ in $\mathrm{Quot}(R)$; note $\varphi(b) \neq 0$ as $b \neq 0$. Thus, uniqueness is clear. Well-defined: if $a/b = a'/b'$ where $a', b' \in R, b' \neq 0$, i.e., $ab' = a'b$, then $\varphi(a)\varphi(b') = \varphi(a')\varphi(b)$, so $\varphi(a')\varphi(b)^{-1} = \varphi(a')\varphi(b')^{-1}$, i.e., $\Phi(a/b) = \Phi(a'/b')$. Clearly, $\Phi(1/1) = 1$.

Φ preserves $+$: $\Phi(a/b + c/d) = \varphi(ad + cb)\varphi(bd)^{-1} = \varphi(a)\varphi(d)\varphi(bd)^{-1} + \varphi(c)\varphi(b)\varphi(bd)^{-1}$. But $\varphi(bd)^{-1} = \varphi(b)^{-1}\varphi(d)^{-1}$, so $= \varphi(a)\varphi(b)^{-1} + \varphi(c)\varphi(d)^{-1} = \Phi(a/b) + \Phi(c/d)$.

Φ preserves $\cdot$: $\Phi(a/b \cdot c/d) = \varphi(ac)\varphi(bd)^{-1} = \varphi(a)\varphi(c)\varphi(b)^{-1}\varphi(d)^{-1} = \Phi(a/b) \cdot \Phi(c/d)$. □

Exercise 3.4.6. Let K be a field. Then $K((X)) \cong \mathrm{Quot}(K[[X]])$.

Definition 3.4.7. Assume R is a subring of the field K. The *quotient field of R in K* is

$$\{ab^{-1} \mid a, b \in R, b \neq 0\},$$

the image of the unique homomorphism $\Phi : \mathrm{Quot}(R) \to K$ extending $\mathrm{id}_R : R \to K$.

Remark 3.4.8. It is not hard to see that this is the smallest subfield of K that contains R.

In particular, if $L \mid K$ is a field extension and $A \subseteq L$, then $K(A)$ equals the quotient field of the subring $K[A]$ of L.

The quotient fields of $\mathbb{Z}, \mathbb{Z}[1/2], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[i]$ in $\mathbb{C}$ are $\mathbb{Q}, \mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{Q}(i)$. The subrings $\mathbb{Z}[\sqrt{2}], \mathbb{Z}[i]$ of $\mathbb{Q}(\sqrt{2}), \mathbb{Q}(i)$ play a similar role as the subring $\mathbb{Z}$ of $\mathbb{Q}$ – see Section 4.1.

By Lemma 3.1.6, with R also $R[X]$ is an integral domain. Hence we can define:

Definition 3.4.9. Let R be an integral domain. The field of *rational functions over R* is

$$R(X) := \mathrm{Quot}(R[X]).$$

Remark 3.4.10. Intuitively, a rational function over R is a fraction f/g of polynomials. Despite the wording it does in general not determine a function on R, i.e., we do not have an evaluation homomorphism (e.g., what should $\varphi_0(1/X)$ be?).

Exercise 3.4.11. $\mathrm{Quot}(R[X]) \cong \mathrm{Quot}(\mathrm{Quot}(R)[X])$.

3.4.1 Prime fields

Definition 3.4.12. Let K be a field. The smallest $k > 0$ such that $\underline{k} = 0_K$ is the *characteristic* $\mathrm{char}(K)$ *of* K; if no such $k > 0$ exists, we set $\mathrm{char}(K) := 0$.

Remark 3.4.13. Let K be a field. Recall $\underline{n} = 1_K + \cdots + 1_K$ for $n \in \mathbb{N}$ (Definition 1.1.25) and extend this notation to $\mathbb{Z}$ setting $\underline{-n} := -\underline{n}$ (additive inverse in K in the r.h.s.).

1. $\mathrm{char}(K)$ is 0 or prime.

   Indeed: $0 < \mathrm{char}(K) =: n = k\ell$ with $1 \leqslant k, \ell < n$ implies $\underline{n} = \underline{k}\cdot\underline{\ell} = 0_K$ but both $\underline{k}, \underline{\ell} \neq 0_K$, so are zero-divisors, a contradiction.
2. If $\mathrm{char}(K) = 0$, then $a \mapsto \underline{a}$ is a ring monomorphism from $\mathbb{Z}$ to K.

   Indeed, it is clearly a homomorphism. It is injective: if $\underline{a} = \underline{b}$, then $\underline{a-b} = 0_K$, so $a - b = 0$ and $a = b$.
3. If $p := \mathrm{char}(K) > 0$, then $[a]_p \mapsto \underline{a}$ is a ring monomorphism from $\mathbb{Z}_p$ to K.

   Indeed, this is well-defined: if $[a]_p = [b]_p$, then $a - b = pc$ for some $c \in \mathbb{Z}$, so $\underline{a} - \underline{b} = \underline{pc} = 0_K$, so $\underline{a} = \underline{b}$. It is injective since these implications can be reversed.

Theorem 3.4.14. *A field K has a smallest subfield, the* prime field of K. *If* $\mathrm{char}(K) = 0$, *it is isomorphic to* $\mathbb{Q}$. *If* $p := \mathrm{char}(K) > 0$, *it is isomorphic to* $\mathbb{F}_p$.

Proof. Every subring of K contains the elements $\underline{a}, a \in \mathbb{Z}$. They form a subring R of K (being the image of a ring homomorphism) isomorphic to $\mathbb{Z}$ (resp. to $\mathbb{Z}_p = \mathbb{F}_p$). The quotient field of R in K is the smallest subfield of K. It is isomorphic to $\mathrm{Quot}(\mathbb{Z})$ and thus to $\mathbb{Q}$ (resp. equals $R \cong \mathbb{F}_p$ being a field). □

Exercise 3.4.15. Let K be a field. Every endomorphism φ of K is the identity on its prime field $P \subseteq K$, i.e., $\varphi{\restriction}P = \mathrm{id}_P$.

Lemma 3.4.16 (Frobenius)**.** *Let K be a field of characteristic $p > 0$. Then $x \mapsto x^p$ is an endomorphism of K, the* Frobenius endomorphism (of K).

Proof. Clearly, the map preserves $\cdot$. For $+$ note $(x+y)^p = \sum_{i=0}^{p} \binom{p}{i} x^{p-i}y^i$ and for $0 < i < p$, p divides $\binom{p}{i} = p(p-1)\cdots(p-i+1)/i!$ because no factor in $i!$ can cancel p; hence, $\underline{\binom{p}{i}} x^{p-i}y^i = 0_K$. Thus $(x+y)^p = x^p + y^p$ (the so-called *freshman's dream*). □

It is natural to ask when the Frobenius endomorphism is surjective, and hence an automorphism (Remark 1.1.22 (2)). This is going to play a role in the last chapter and therefore earns a definition:

Definition 3.4.17. A field K is *perfect* either if $\mathrm{char}(K) = 0$ or, if $\mathrm{char}(K) > 0$ and the Frobenius endomorphism of K is surjective.

Examples 3.4.18.

1. Finite fields are perfect.

   Indeed, the characteristic is positive, and the Frobenius homomorphism injective (Remark 1.1.22 (3)), so surjective by finiteness.

2. Let p be prime. Then $\mathbb{F}_p(X)$, the field of rational functions over $\mathbb{F}_p$, is not perfect.

   Indeed: if $(f/g)^p = X$ for some $f, g \in \mathbb{F}_p[X]$ with $g \neq 0$, then $f^p = g^p X$, so $p\deg(f) = p\deg(g) + 1$. But this is nonsense because $\deg(f), \deg(g) \geqslant 0$.

3.5 Algebraicity

Let $L \mid K$ be a field extension. For $a \in L$, clearly $K[a] \subseteq K(a)$ and, $K[a] = K(a)$ if and only if $K[a]$ is a field. When does this happen?

Example 3.5.1. $\mathbb{R}[i] = \mathbb{R}(i) = \mathbb{C}$ because $\mathbb{R}[i] \subseteq \mathbb{R}(i) \subseteq \mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\} \subseteq \mathbb{R}[i]$. Using the notation below, we have $m_i^{\mathbb{Q}} = m_i^{\mathbb{R}} = X^2 + 1$.

Definition 3.5.2. $a \in L$ is *algebraic over* K if it is a root of some $f \in K[X]$; otherwise a is *transcendental over* K. For $L = \mathbb{C}, K = \mathbb{Q}$ we omit "over $\mathbb{Q}$".

If $a \in L$ is algebraic over K, then the *minimal polynomial* m_a^K *of* a *over* K is a monic $f \in K[X] \smallsetminus K$ of minimal degree with root a.

Remark 3.5.3. This is well-defined: let $f \in K[X] \smallsetminus K$ be of minimal degree, say n, with root a; we can take f to be monic by dividing by the lead coefficient. If $g \in K[X]$ is another such polynomial, then $f - g$ has root a and degree $< n$; hence $f - g = 0$, i.e, $f = g$.

Remark 3.5.4 (Cantor)**.** "Most" reals are transcendental since the set of algebraic numbers is countable.

Definition 3.5.5. $f \in K[X]$ is *reducible* if $f \in K$ or f is the product of two polynomials in $K[X]$ of positive degree; otherwise f is *irreducible*.

Lemma 3.5.6. *Let $a \in L$ be algebraic over K and $f \in K[X]$ monic. The following are equivalent.*

1. *$f = m_a^K$.*
2. *a is a root of f and f is irreducible.*
3. *a is a root of f and $f \mid g$ for every $g \in K[X]$ with root a.*

Proof. $1 \Rightarrow 3$: assume $g(a) = 0$ and write $g = qf + r$ with $\deg(r) < \deg(f)$; then $0 = g(a) = h(a)f(a) + r(a) = r(a)$; then $r = 0$, so $f \mid g$.

$3 \Rightarrow 1$: $f \mid g$ implies $\deg(f) \leqslant \deg(g)$, so f has minimal degree among the polynomials with root a; as f is monic, $f = m_a^K$.

$1 \Rightarrow 2$: clearly, $\deg(m_a^K) > 0$; if $m_a^K = gh$ for $g, h \in K[X]$ of positive degree, then g, h have degree $< \deg(m_a^K)$ by Remark 3.1.4 (3). Then $0 = f(a) = g(a)h(a)$, so at least one

of $g(a), h(a) \neq 0$. Divide by the lead coefficient to get a monic polyomial with root a and degree $< \deg(m_a^K)$, a contradiction.

$2 \Rightarrow 1$: assume $f(a) = 0$ and f is irreducible. By $1 \Rightarrow 3$, $f = m_a^K h$ for some $h \in K[X]$; by irreducibility, $\deg(h) = 0$, so $h \in K$; as f and m_a^K are monic, $h = 1$ and $f = m_a^K$. □

Corollary 3.5.7. *Let $f, g \in K[X]$. If f is irreducible over K and f, g have a common root in L, then $f \mid g$.*

Theorem 3.5.8. *Let $a \in L$. Then $K[a]$ is a field if and only if a is algebraic over K.*

Proof. $\Rightarrow$: $a^{-1} = f(a)$ for some $f \in K[X]$, so a is a root of $Xf - 1 \in K[X]$.

$\Leftarrow$: given $0 \neq b \in K[a]$ we look for an inverse in $K[a]$. Write $b = f(a)$ for $f \in K[X]$. By irreducibility, $\gcd(f, m_a^K)$ is either 1 or m_a^K. But it cannot be m_a^K because $m_a^K \nmid f$ as $f(a) \neq 0$. Thus, the Euclidian algorithm for polynomials gives $1 = gf + hm_a^K$ for certain $g, h \in K[X]$. Then $g(a) \in K[a]$ is an inverse of b:

$$1 = g(a)f(a) + h(a)m_a^K(a) = g(a)b.$$ □

Exercise 3.5.9. If $f \in K[X]$ has degree 2 or 3, then f is irreducible if and only if f does not have a root in K.

Exercise 3.5.10. Let $a \in \mathbb{C}$ be a root of $f := X^3 - X + 1 \in \mathbb{Q}[X]$. Show $f = m_a^{\mathbb{Q}}$. Show $b := 2a^2 - 3a + 2 \neq 0$ and find $h \in \mathbb{Q}[X]$ with $h(a) = 1/b$ in $\mathbb{Q}(a)$. (*Hint:* Exercise 3.2.9.)

Irreducibility for higher degrees is difficult to understand and studied in the next chapter. Algebraicity is a central concept of algebra but we do not yet have the theoretical means to understand it and defer a more serious study to the last chapter.

3.5.1 Quadratic and cubic equations

A description of the quadratic case is in reach of our currently available methods, in fact, we learned in school how to solve quadratic equations. Landau said, however, "Bitte vergessen Sie alles, was Sie in der Schule gelernt haben, denn Sie haben es nicht gelernt."

Recall, in a ring or field K we write $\underline{2} := 1_K + 1_K, \underline{3} := \underline{2} + 1_K$ etc.

Lemma 3.5.11. *Let K be a field with* $\mathrm{char}(K) \neq 2$ *and $f = aX^2 + bX + c \in K[X]$ with $a \neq 0$. The* discriminant *of f is*

$$D_f := b^2 - \underline{4}ac \in K.$$

1. *f is reducible if and only if D_f is a square in K.*
2. *If $L \mid K$ is a field extension with $\alpha \in L$ a root of f, then $K(\alpha) = K(\delta)$ for some $\delta \in L$ with $\delta^2 = D_f$; we write $K(\sqrt{D_f}) := K(\delta)$.*
3. *Then $K(\delta) = K + K\delta := \{x + y\delta \mid y, y \in K\}$.*

Proof. (2): $a\alpha^2+b\alpha+c=0$ is equivalent to $\underline{4}a^2\alpha^2+\underline{4}ab\alpha=-\underline{4}ac$ (note $\underline{4}ac\neq 0$ as $\mathrm{char}(K)\neq 2$), hence to $(\underline{2}a\alpha)^2+\underline{2}\cdot\underline{2}a\alpha b+b^2=b^2-\underline{4}ac$, so to $(\underline{2}a\alpha+b)^2=D_f$. Set $\delta:=\underline{2}a\alpha+b$.

Clearly, $K(\alpha)=K(\delta)$. The notation $K(\sqrt{D_f})$ is justified because it does not matter which square root δ we choose: the other one is $-\delta$, and $K(\delta)=K(-\delta)$.

(1): if f is reducible, it has a root $\alpha\in K$, so, by the above, D_f is a square in K. Conversely, if $\delta^2=D_f$ with $\delta\in K$, the above shows $(\delta-b)/(\underline{2}a)\in K$ is a root of f (*Mitternachtsformel*), so f is reducible. Note division by $\underline{2}$ requires $\mathrm{char}(K)\neq 2$.

(3) is trivial if $\delta\in K$. Assume $\delta\notin K$. $\supseteq$ is trivial. For $\subseteq$ it suffices to show $K+K\delta$ is a subfield of L. It clearly contains $0,1$ and with α,β also $\alpha+\beta,\alpha\cdot\beta,-\beta$ and, if $\alpha=x+y\delta\neq 0$, then $\alpha^{-1}=(x-y\delta)/(x^2-y^2D_f)\in K+K\delta$: note $x^2-y^2\delta\neq 0$ as $\delta\notin K$. □

For a real $r>0$ we understand $\sqrt{r}$ to be positive; further, $\sqrt{r}=i\sqrt{|r|}\in\mathbb{C}$ for $r<0$.

Corollary 3.5.12. *Let $\alpha\in\mathbb{C}\smallsetminus\mathbb{Q}$ be a root of a quadratic polynomial over $\mathbb{Q}$. Then $\mathbb{Q}(\alpha)$ is a* quadratic number field*, i.e.,*

$$\mathbb{Q}(\alpha)=\mathbb{Q}(\sqrt{d})=\mathbb{Q}+\mathbb{Q}\sqrt{d}$$

for some $d\in\mathbb{Z}$ satisfying $d=-1$, or, $|d|>1$ is square-free*: $m^2\nmid|d|$ for all $m>1$.*

Proof. Let $f\in\mathbb{Q}[X]$ be quadratic and $\alpha\in\mathbb{C}$ a root. We can assume $f\in\mathbb{Z}[X]$ (otherwise multiply with some large integer). Then $\mathbb{Q}(\alpha)=\mathbb{Q}(\sqrt{D_f})$ and, as $\alpha\notin\mathbb{Q}$, $D_f\in\mathbb{Z}$ is not a square in $\mathbb{Q}$. If $|D_f|>1$ is not square-free, say, $D_f=m^2D'$ for $D'\in\mathbb{Z}$ and $m>1$, then $\sqrt{D_f}=m\sqrt{D'}$ and $\mathbb{Q}(\sqrt{D_f})=\mathbb{Q}(\sqrt{D'})$. Then $|D'|<|D_f|$, so continuing like this for finitely many steps we get $\mathbb{Q}(\alpha)=\mathbb{Q}(\sqrt{d})$ for a square-free $|d|$. □

Remark 3.5.13. Read $K(\alpha)=K(\sqrt{D_f})$ as a statement about solvability of the equation $f=0$: that a solution α is in $K(\sqrt{D_f})$ means that we can construct it by finitely many applications of addition, subtraction, multiplication and division plus a single use of taking some square root. Of course, we already know the formula from school.

As mentioned in the introduction, allowing more uses of taking (possibly non-square) roots allows to construct roots of polynomials of degree 3 and 4. In the last chapter we show this is not generally possible for degree 5. Here, we treat the cubic case.

Remark 3.5.14 (Tschirnhausen transformation)**.** We want to solve $a_3X^3+a_2X^2+a_1X+a_0=0$ where $a_i\in\mathbb{C}$ and $a_3\neq 0$. We can assume $a_3=1$ (otherwise divide by a_3). Substitution of $X+a_2/3$ for X yields $X^3+aX+b=0$ for certain $a,b\in\mathbb{C}$.

Recall, $\zeta_3=e^{2\pi i/3}=(-1+\sqrt{-3})/2$.

Proposition 3.5.15 (Cardano's formulas)**.** *Let $f:=X^3+aX+b\in\mathbb{C}[X]$. Let*

$$D_f=-4a^3-27b^2$$

be the discriminant *of f. Let $\delta,x,y\in\mathbb{C}$ be such that*

$$\delta^2=-D_f/27,\ x^3=(\delta-b)/2,\ y^3=(\delta+b)/2.$$

Then the roots of f in $\mathbb{C}$ are $x-y$, $\zeta_3x-\zeta_3^2y$, $\zeta_3^2x-\zeta_3y$.

Proof. Since $(x-y)^3 = x^3 - 3x^2y + 3xy^2 - y^3$ we have $(x-y)^3 + 3xy(x-y) - (x^3 - y^3) = 0$. Then $x - y$ is a root if x, y are *good*: $xy = a/3, x^3 - y^3 = -b$. We solve for x, y.

Being good implies $4x^3y^3 = 4a^3/27$ and $x^6 - 2x^3y^3 + y^6 = b^2$. Adding these gives

$$(x^3 + y^3)^2 = b^2 + 4a^3/27$$

The r.h.s. is $-D_f/27$. Hence, $x^3 + y^3 = \delta$. Using $x^3 - y^3 = -b$ gives the equations listed. The roots of f are those choices of 3rd roots that give good x, y. □

3.6 Multivariate polynomials

Multivariate polynomials are "expressions" like $X^2YZ^3 + 2XY^2 - 3Y$ and it should be clear how to sum and multiply them. Formally, we proceed similarly as in Definition 3.1.1. There we wrote sequences $(a_k)_k$ assigning coefficients to powers X^k; now we assign coefficients to *primitive monomials*, above X^2YZ^3, XY^2 and Y.

Let R be a commutative ring and $I \neq \varnothing$ a set – intuitively, we want variables $X_i, i \in I$.

Definition 3.6.1. Let M be the set of *primitive monomials:* functions $m : I \to \mathbb{N}$ such that there are only finitely many $i \in I$ with $m(i) > 0$.

Given $m, m' \in M$, define $m \odot m' \in M$ setting for all $i \in I$:

$$(m \odot m')(i) := m(i) + m'(i).$$

$R[X_i, i \in I]$ is the set of *(multivariate) polynomials (over R in variables $(X_i)_{i \in I}$)*: functions $f : M \to R$ such that there are are only finitely many $m \in M$ with $f(m) \neq 0$.

Given polynomials f, g define polynomials $f + g$ and $f \cdot g$ setting for all $m \in M$:

$$(f+g)(m) := f(m) + g(m), \quad (f \cdot g)(m) := \sum_{m_0 \odot m_1 = m} f(m_0)g(m_1).$$

Observe that the sum above is a finite sum in R. As in the univariate case we now introduce the familiar notation and do suitable identifications.

Remark 3.6.2.

1. $(M, \odot)$ is a Monoid with neutral element $1_M :=$ the function constantly 0. For $m \in M$ let $i_1, ..., i_k$ list the $i \in I$ with $m(i) > 0$; we write
$$m = X_{i_1}^{m(i_1)} \cdots X_{i_k}^{m(i_k)}.$$
2. For $m \in M$ let $g_m \in R[X_i, i \in I]$ map m to 1 and $m' \neq m$ to 0. Then $m \mapsto g_m$ is a monoid monomorphism from $(M, \odot)$ to $(R[X_i, i \in I], \cdot)$.

   We "identify" m with g_m and thereby view M as a subset of $R[X_i, i \in I]$.
3. For m as above and $i \in I, k \in \mathbb{N}$ we have in $R[X_i, i \in I]$:
$$m = X_{i_1}^{m(i_1)} \cdot ... \cdot X_{i_k}^{m(i_k)} \quad \text{and} \quad X_i^k = X_i \cdot ... \cdot X_i \ (k \text{ times}).$$

4. For $a \in R$ let $f_a \in R[X_i, i \in I]$ map 1_M to a and $m \neq 1_M$ to 0. Then $a \mapsto f_a$ is a ring monomorphism from R into $R[X_i, i \in I]$.

   We "identify" a with f_a and view R as a subring of $R[X_i, i \in I]$.

   Indeed, $R[X_i, i \in I]$ is a commutative ring with neutral elements $0_{R[X_i, i\in I]} := f_0$ (*zero polyomial*) and $1_{R[X_i, i\in I]} := f_1$; additive inverses are given by $(-f)(m) = -f(m)$ (inverse in R on the r.h.s.).
5. Given $f \in R[X_i, i \in I] \setminus \{0\}$ let $m_0, ..., m_s \in M$ for $s \in \mathbb{N}$ list the $m \in M$ with $f(m) \neq 0$. Writing $a_j := f(m_j)$ for $j \leqslant s$ we have

$$f = a_0 m_0 + \cdots + a_s m_s.$$

Definition 3.6.3. A *monomial* is a polynomial in $R[X_i, i \in I]$ of the form am where $a \in R \setminus \{0\}, m \in M$; we say X_i *occurs in* am if $m(i) > 0$. A polynomial $f \neq 0$ as in (5) above has *(total) degree*

$$\deg(f) := \max_{j \leqslant s} \sum_{i \in I} m(i);$$

note the sums are finite; the zero polynomial has degree $-\infty$. Writing

$$f = f(X_{i_1}, ..., X_{i_r})$$

means that the variables that occur in the $a_j m_j$ are among $X_{i_1}, ..., X_{i_r}$.

Notation: We write e.g. $R[X, Y, Z]$ instead of $R[X_i, i \in \{1, 2, 3\}]$.

Exercise 3.6.4. Show $R[X, Y] \cong R[X][Y]$. If R is an integral domain, so is $R[X_i, i \in I]$. If R is an integral domain, and $f, g \in R[X_i, i \in I]$, then $\deg(fg) = \deg(f) + \deg(g)$.

Infer that $R[X_i, i \in I]^\times = R^\times$.

Hint: for the 2nd statement, assume $f, g \in R[X_1, ..., X_n]$. Order monomials according to the *graded lexicographic order* of exponent tuples in $\mathbb{N}^n$: $(e_1, ..., e_n) <_{lex} (d_1, ..., d_n)$ if $\neq$ and either $\sum_i e_i < \sum_i d_i$, or $\sum_i e_i = \sum_i d_i$ and $e_j < d_j$ for the first j with $e_j \neq d_j$.

Definition 3.6.5. Let R be an integral domain. $R(X_i, i \in I) := \mathrm{Quot}(R[X_i, i \in I])$ is the *field of multivariate rational functions over R in variables $X_i, i \in I$.*

Of course, we write $R(X, Y, Z)$ for $\mathrm{Quot}(R[X, Y, Z])$ and the like. The following is analogous to Theorem 3.1.7.

Theorem 3.6.6 (Universal property)**.** *Let I be a nonempty set, R, S be commutative rings, $\varphi : R \to S$ a ring homomorphism, and $x_i \in S$ for every $i \in I$. Then there exists a unique ring homomorphism $\Phi : R[X_i, i \in I] \to S$ that extends φ and satisfies $\Phi(X_i) = x_i$ for all $i \in I$.*

Corollary 3.6.7 (Evaluation homomorphism)**.**

1. *For $i \in I$ let $g_i \in R[X_i, i \in I]$. There is a unique ring homomorphism $\varphi_{(g_i)_i} : R[X_i, i \in I] \to R[X_i, i \in I]$ that is the identity on R and satisfies $\varphi_{(g_i)_i}(X_i) = g_i$ for all $i \in I$.*

   If $f = f(X_{i_1}, ..., X_{i_r})$, we write $f(g_{i_1}, ..., g_{i_r}) := \varphi_{(g_i)_i}(f)$.

2. *For $i \in I$ let $a_i \in R$. There is a unique ring homomorphism $\varphi_{(a_i)_i} : R[X_i, i \in I] \to R$ that is the identity on R and satisfies $\varphi_{(a_i)_i}(X_i) = a_i$ for all $i \in I$.*

   If $f = f(X_{i_1}, ..., X_{i_r})$, we write $f(a_{i_1}, ..., a_{i_r}) := \varphi_{(a_i)_i}(f)$.

Exercise 3.6.8. Let $f \in R[X, Y]$ and $a, b \in R$. Formalize and prove: plugging a, b for X, Y in f is the same as plugging first a for X and then b for Y.

The following generalizes Remark 3.1.12 (3). We leave the proof as an exercise.

Lemma 3.6.9. *Let S be a commutative ring with subring R and $A = \{a_i \mid i \in I\} \subseteq S$. Then $R[A]$ is the image of $\varphi_{(a_i)_i}$, i.e.,*

$$R[A] = \big\{ f(a_{i_1}, ..., a_{i_n}) \mid n \in \mathbb{N}, i_1, ..., i_n \in I, f(X_{i_1}, ..., X_{i_n}) \in R[X_i, i \in I] \big\}.$$

In particular, for $n > 0$ and $a_1, ..., a_n \in S$ we have

$$R[a_1, ..., a_n] = \big\{ f(a_1, ..., a_n) \mid f(X_1, ..., X_n) \in R[X_1, ..., X_n] \big\}.$$

Lemma 3.6.10.

1. *Let S, S' be commutative rings and R be a subring of both S and S', $A \subseteq S$ and $\varphi : S \to S'$ a ring homomorphism with $\varphi \restriction R = \mathrm{id}_R$. Then*

$$\varphi(R[A]) = R[\varphi(A)].$$

   Moreover, if a ring homomorphism $\psi : R \to S'$ agrees with φ on $R \cup A$, then also on $R[A]$.

2. *Let $L \mid K$ and $L' \mid K$ be field extensions, $A \subseteq L$ and $\varphi : L \to L'$ a field homomorphism with $\varphi \restriction K = \mathrm{id}_K$. Then*

$$\varphi(K(A)) = K(\varphi(A)).$$

   Moreover, if a field homomorphism $\psi : L \to L'$ agrees with φ on $K \cup A$, then also on $K(A)$.

Proof. We can assume $A \neq \varnothing$. (1) $\supseteq$: $\varphi(R[A])$ is a subring of S' that contains $R \cup \varphi(A)$. $\subseteq$: write $A = \{a_i \mid i \in I\}$ for a suitable set I and let $x \in R[A]$, say $x = f(a_{i_1}, ..., a_{i_n})$ as in the previous lemma; then $\varphi(x) = f(\varphi(a_{i_1}), ..., \varphi(a_{i_n})) \in R[\varphi(A)]$.

Moreover: $\psi(x) = f(\psi(a_{i_1}), ..., \psi(a_{i_n})) = f(\varphi(a_{i_1}), ..., \varphi(a_{i_n})) = \varphi(x)$.

(2) $\supseteq$: $\varphi(K[A])$ is a subfield of L' that contains $K \cup \varphi(A)$. $\subseteq$: let $x \in K(A)$, say $x = y/z$ for $y, z \in K[A]$ with $z \neq 0$; then $\varphi(x) = \varphi(y)/\varphi(z) \in K(\varphi(A))$ by (1).

Moreover: $\psi(x) = \psi(y)/\psi(z) = \varphi(y)/\varphi(z) = \varphi(x)$ since φ, ψ agree on $K[A]$ by (1). □

Definition 3.6.11. Let $n > 0$ and $f \in R[X_1, ..., X_n]$. Then $\bar{a} = (a_1, ..., a_n) \in R^n$ is a *root of f* if and only if $f(\bar{a}) = 0$.

Exercise 3.6.12. ...if and only if there are $g_1, ..., g_n \in R[X_1, ..., X_n]$ such that

$$f = g_1(X_1 - a_1) + \cdots + g_n(X_n - a_n).$$

Exercise 3.6.13. Let $d, n > 0$ and $f, g \in R[X_1, ..., X_n]$. Then $f = f(X_1, ..., X_n)$ is *homogeneous* if all its monomials have the same degree. Equivalently, writing $d := \deg(f)$,

$$f(YX_1, ..., YX_n) = Y^d f(X_1, ..., X_n),$$

in $R[X_1, ..., X_n, Y]$. Further, fg is homogeneous if and only if both f and g are.

3.7 Symmetric polynomials

The polynomials $X^2Y^2 + 3XY + 2X + 2Y$ and $X^2 + Y^2$ are *symmetric* in the sense that they do not change when X, Y are interchanged. E.g., $X^2Y + XY$ is not symmetric.

Definition 3.7.1. Let R be a commutative ring, $n > 0$ and $f = f(X_1, ..., X_n) \in R[X_1, ..., X_n]$. Then f is *symmetric* if $f = f^\sigma$ for all permutations $\sigma : \{1, ..., n\} \to \{1, ..., n\}$; here,

$$f^\sigma := f(X_{\sigma(1)}, ..., X_{\sigma(n)}).$$

For $1 \leqslant k \leqslant n$, the *k-th elementary symmetric polynomial in n variables* is

$$s_{n,k} := \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} X_{i_1} \cdots X_{i_k}.$$

Example 3.7.2. $s_{3,1} = X_1 + X_2 + X_3$, $s_{3,2} = X_1X_2 + X_1X_3 + X_2X_3$, $s_{3,3} = X_1X_2X_3$.

Remark 3.7.3.

1. All $s_{n,k}$ are symmetric. In $R[X_1, ..., X_n][X]$, we have *Vieta's formula*

   $$(X - X_1) \cdots (X - X_n) = X^n - s_{n,1}X^{n-1} + s_{n,2}X^{n-1} - \cdots + (-1)^n s_{n,n}.$$

   The coefficients of a monic degree n polynomial with n roots (not necessarily distinct) are values of symmetric polynomials of these roots.
2. $(f+g)^\sigma = f^\sigma + g^\sigma$ and $(f \cdot g)^\sigma = f^\sigma \cdot g^\sigma$, so $f \mapsto f^\sigma$ is an automorphism of $R[X_1, ..., X_n]$. The set of symmetric polynomials is the fixed ring (see below).
3. Hence, $g(s_{n,1}, ..., s_{n,n})$ is symmetric for every $g \in R[X_1, ..., X_n]$.

Definition 3.7.4. Let R be a group, ring or field, and $\Phi \subseteq \mathrm{Aut}(R)$. The *fixed group (ring, field) of* Φ is

$$R^\Phi := \{x \in R \mid \varphi(x) = x \text{ for all } \varphi \in \Phi\}.$$

Remark 3.7.5. R^Φ is a subgroup (subring, subfield) of R.

Example 3.7.6. $X_1^2X_2^2 + 3X_1X_2 + 2X_1 + 2X_2 = s_{2,2}^2 + 3s_{2,2} + 2s_{2,1}$ and $X_1^2 + X_2^2 = s_{2,1}^2 - 2s_{2,2}$.

Theorem 3.7.7. *Let R be a commutative ring and $n > 0$. For every symmetric $f \in R[X_1, ..., X_n]$ there is a unique $g \in R[X_1, ..., X_n]$ such that $f = g(s_{n,1}, ..., s_{n,n})$.*

Hence, the evaluation homomorphism mapping X_k to $s_{n,k}$, i.e., $\varphi_{(s_{n,k})_k}$, is an isomorphism from $R[X_1, ..., X_n]$ onto the subring of symmetric polynomials.

Proof. The *weighted degree* of a monomial $aX_1^{r_1}\cdots X_n^{r_n}$ with $a \neq 0$ is $1\cdot r_1 + \cdots + n\cdot r_n$. The *weighted degree* $\mathrm{wdeg}(g)$ of $g \in R[X_1,...,X_n]\setminus\{0\}$ is the maximum weighted degree of its monomials; $\mathrm{wdeg}(0) := -\infty$. One easily checks $\deg(\varphi(g)) = \mathrm{wdeg}(g)$ where $\varphi := \varphi_{(s_{n,k})_k}$.

We proceed by induction on n. For $n = 1$ our claim is trivial: $X_1 = s_{1,1}$, φ is the identity and all polynomials in $R[X_1]$ are symmetric. Let $n > 1$ and assume our claim for $n-1$.

We first show φ is surjective. Otherwise, choose f symmetric of minimal degree d outside the image of φ. For a polynomial $g(X_1,...,X_n)$ let $\tilde{g} := g(X_1,...,X_{n-1},0)$. Then $\tilde{f}$ is symmetric with $n-1$ variables. Thus, $\tilde{f} = g(s_{n-1,1},...,s_{n-1,n-1})$ for some $g \in R[X_1,...,X_{n-1}]$.

Set $h := f - g(s_{n,1},...,s_{n,n-1})$. Note h is symmetric of degree $\leqslant d$: indeed, $d \geqslant \deg(\tilde{f}) = \mathrm{wdeg}(g) = \deg(g(s_{n,1},...,s_{n,n-1}))$. Further, noting $s_{n-1,k} = \tilde{s}_{n,k}$ for $k < n$,

$$\tilde{h} = \tilde{f} - g(\tilde{s}_{n,1},...,\tilde{s}_{n,n-1}) = \tilde{f} - g(s_{n-1,1},...,s_{n-1,n}) = 0$$

It follows that X_n occurs (with positive exponent) in all monomials of h. As h is symmetric, this holds for all X_k, and we can write $h = s_{n,n}h'$ for some h'. This h' is symmetric since h is: $s_{n,n}h' = h = h^\sigma = s_{n,n}^\sigma h'^\sigma$ implies $h' = h'^\sigma$ since $s_{n,n}^\sigma = s_{n,n}$. But $\deg(h') < d$, so $h' = g'(s_{n,1},...,s_{n,n})$ for some g'. But then f is in the image of φ, a contradiction:

$$f = h + g(s_{n,1},...,s_{n,n-1}) = s_{n,n}g'(s_{n,1},...,s_{n,n}) + g(s_{n,1},...,s_{n,n-1}).$$

For injectivity, we show $\ker(\varphi) = \{0\}$. Otherwise choose f of minimal degree $d > 0$ with $\varphi(f) = 0$. Then, noting $\tilde{s}_{n,n} = 0$,

$$0 = f(\tilde{s}_{n,1},...,\tilde{s}_{n,n-1},\tilde{s}_{n,n}) = f(s_{n-1,1},...,s_{n-1,n-1},0)$$

By induction, $f(X_1,...,X_{n-1},0) = 0$. Thus, $f = X_n f'$ for some $f' \neq 0$. Then $0 = \varphi(f) = s_{n,n}\varphi(f')$ implies $\varphi(f') = 0$. But $\deg(f') < \deg(f)$, a contradiction. □

Exercise 3.7.8. The ring of symmetric polynomials equals the subring $R[s_{n,1},...,s_{n,n}]$ of $R[X_1,...,X_n]$. (Notation check.)

Here is an important consequence:

Corollary 3.7.9. *Let K be a field, R a subring and let $f \in R[X]$ be monic of degree $n > 0$. Assume $f = (X-\alpha_1)\cdots(X-\alpha_n)$ with $\alpha_i \in K$. Let $g(X_1,...,X_n) \in R[X_1...,X_n]$ be symmetric.*

Then $g(\alpha_1,...,\alpha_n) \in R$.

Proof. Write $\bar{X}, \bar{\alpha}$ for the n-tuples of the X_i, α_i and $f = X^n - a_1X^{n-1} + a_2X^{n-2} - \cdots + (-1)^n a_n$ with $a_i \in R$. By the theorem, $g(\bar{X}) = G(s_{n,1},...,s_{n,n})$ for some $G \in R[\bar{X}]$. By Vieta's formula, $s_{n,i}(\bar{\alpha}) = a_i$. Then

$$g(\bar{\alpha}) = G(s_{n,1}(\bar{\alpha}),...,s_{n,n}(\bar{\alpha})) = G(a_1,...,a_n) \in R.$$ □

Exercise 3.7.10. Let K be a field, R a subring, $f \in R[X]$, $a \in R\setminus\{0\}$, $n > 0$ and assume that $f = a(X-\alpha_1)\cdots(X-\alpha_n) \in R[X]$ with $\alpha_i \in K$. Let $g \in R[X_1,...,X_n]$ be symmetric of (total) degree $\leqslant n$ such that a^n divides all coefficients of g. Then $g(\bar{\alpha}) \in R$.

Hint: find a monic $f^* \in R[X]$ with roots $a\alpha_i$.

Example 3.7.11 (Higher discriminants)**.** Let $R, f, K, \bar{\alpha}$ be as above. Then

$$D(\bar{X}) := \prod_{i<j}(X_i - X_j)^2$$

is symmetric, so equals $\Delta_n(s_{n,1}, ..., s_{n,n})$ for some $\Delta_n \in R[\bar{X}]$. By Vieta's formula,

$$D(\bar{\alpha}) = \Delta_n(a_1, ..., a_n) \in R,$$

where $f = X^n - a_1X^{n-1} + a_2X^{n-2} - \cdots + (-1)^n a_n$ with $a_i \in R$. This is called the *discriminant of* f and denoted D_f. It can be computed without knowing K and roots $\bar{\alpha}$. It is 0 if and only if f has a multiple root in K – in particular, this does not depend on the choice of K.

The above generalizes the familiar formula for $n = 2$:

$$f = X^2 - s_{2,1}(\bar{\alpha})X + s_{2,2}(\bar{\alpha}), \quad D_f = (X_1 - X_2)^2 = s_{2,1}^2 - \underline{4}s_{2,2}, \quad \Delta_2 = X_1^2 - \underline{4}X_2.$$

Unfortunately, Δ_n is complicated for $n > 2$. E.g.,

$$\Delta_3 = -\underline{4}X_1^3X_3 + X_1^2X_2^2 + \underline{18}X_1X_2X_3 - \underline{4}X_2^3 - \underline{27}X_3^2.$$

This matches the definition of D_f in Proposition 3.5.15.

Exercise 3.7.12. $D_f = (-1)^{\binom{n}{2}} \prod_{i \neq j}(\alpha_i - \alpha_j) = (-1)^{\binom{n}{2}} \cdot f'(\alpha_1) \cdots f'(\alpha_n)$.

Exercise 3.7.13. For $R = \mathbb{R}, K = \mathbb{C}, n = 3$ above, show $D_f = 0, < 0, > 0$ if and only if f has a multiple root, one real and two complex roots, resp., three real roots.

Remark 3.7.14. We defined D_f for monic f, i.e., with lead coefficient $a_0 = 1$. Definitions for $a_0 \neq 1$ add a normalizing factor, e.g., often but not always a_0^{2n-2} is used. We omit this in order not to spoil the elegance.

3.8 The fundamental theorem of algebra

In the last chapter we shall prove (Theorem 6.3.5):

Theorem 3.8.1. *Let K be a field and let $f \in K[X]$ have degree $n > 0$. Then there exists a field extension $L \mid K$ and $a \in K$ and $a_1, ..., a_n \in L$ such that $f = a(X - a_1) \cdots (X - a_n)$.*

Already in 1795 Laplace sketched a proof of the fundamental theorem of algebra that in Lagrange's words "ne laisse rien désirer comme simple démonstracion". It is based on the above, the theorem on symmetric polynomials, and the following basic facts:

Remark 3.8.2. Recall conjugation $z \mapsto \bar{z}$ is an automorphism of $\mathbb{C}$.

1. For $f = a_nX^n + \cdots + a_0 \in \mathbb{C}[X]$ let $\bar{f} := \bar{a}_nX^n + \cdots + \bar{a}_0$; then $\overline{f(z)} = \bar{f}(\bar{z})$ for all $z \in \mathbb{C}$.
2. For Theorem 3.8.3 it suffices to show that every $g \in \mathbb{R}[X] \smallsetminus \mathbb{R}$ has a root in $\mathbb{C}$.

   Given $f \in \mathbb{C}[X]$, note $\overline{f\bar{f}} = \bar{f}f$, so $f\bar{f} \in \mathbb{R}[X]$; a root $a \in \mathbb{C}$ of $f\bar{f}$ is a root of f or of $\bar{f}$; in the 2nd case $\bar{a}$ is a root of f.

3. Every $f \in \mathbb{R}[X]$ of odd degree has a root in $\mathbb{R}$.

   Assume f is monic; then $f(-n) < 0$ and $f(n) > 0$ for large enough $n \in \mathbb{N}$; apply the mean value theorem.
4. Every quadratic $f \in \mathbb{C}[X]$ has a root in $\mathbb{C}$ (exercise).

Theorem 3.8.3 (Fundamental theorem of algebra)**.** *Every $f \in \mathbb{C}[X] \setminus \mathbb{C}$ has a root in $\mathbb{C}$.*

Proof. Given $f \in \mathbb{R}[X] \setminus \mathbb{R}$ we look for a root of f in $\mathbb{C}$ (Remark 3.8.2 (2)). We can assume f is monic, say of degree $n > 0$. Write $n = 2^k m$ for $k, m \in \mathbb{N}$ and m odd. We proceed by induction on k. For $k = 0$ apply Remark 3.8.2 (3). Assume $k > 0$.

By Theorem 3.8.1, $f = (X-\alpha_1)\cdots(X-\alpha_n)$ where the α_i are in some field extension $L \mid \mathbb{C}$. For $t \in \mathbb{R}$ set

$$f_t := \prod_{1 \leqslant i < j \leqslant n} (X - (\alpha_i + \alpha_j + t\alpha_i\alpha_j)).$$

By Vieta's formula the coefficients of f_t are $s_{n(n-1)/2,k}(\ldots\beta_{ij}\ldots)$ where $1 \leqslant k \leqslant n(n-1)/2$ and $\beta_{ij} := \alpha_i + \alpha_j + t\alpha_i\alpha_j$ for $1 \leqslant i < j \leqslant n$. But these are symmetric polynomials in α_i for $1 \leqslant i \leqslant n$. By Corollary 3.7.9, $f_t \in \mathbb{R}[X]$.

We have $\deg(f_t) = n(n-1)/2 = 2^{k-1}m(2^k m - 1) = 2^{k-1}m'$ for m' odd. By induction, f_t has a root in $\mathbb{C}$. This root equals $\alpha_i + \alpha_j + t\alpha_i\alpha_j$ for some $i < j$. This way each t is mapped to a pair $i < j$ and there are $t \neq s$ mapped to the same $i < j$. Then both $\alpha_i + \alpha_j + t\alpha_i\alpha_j, \alpha_i + \alpha_j + s\alpha_i\alpha_j \in \mathbb{C}$. As $s \neq t$, $z_0 := \alpha_1\alpha_2 \in \mathbb{C}$ and $z_1 := \alpha_i + \alpha_j \in \mathbb{C}$ and $(X - \alpha_i)(X - \alpha_j) = X^2 - z_1 X + z_0 \in \mathbb{C}[X]$. Remark 3.8.2 (4) gives a root $z \in \mathbb{C}$ of this polynomial, so $z = \alpha_i$ or $z = \alpha_j$, so z is a root of f. □

Exercise 3.8.4. For every $f \in \mathbb{C}[X]$ of degree $n \in \mathbb{N}$ there are $a_0, \ldots, a_n \in \mathbb{C}$ such that $f = a_0(X-a_1)\cdots(X-a_n)$. Infer that every $f \in \mathbb{R}[X] \setminus \mathbb{R}$ is a product of linear and quadratic polynomials in $\mathbb{R}[X]$.

3.9 Transcendence of π

Hermite showed 1873 that e is transcendental. He said "I shall risk nothing on an attempt to prove the transcendence of π. If others undertake this enterprise, no one will be happier than I in their success. But believe me, it will not fail to cost them some effort." We give Niven's "relatively simple proof" (1939) based on "an ingenious device" of Hurwitz.

Theorem 3.9.1 (Lindemann 1882)**.** *π is transcendental.*

Proof. For contradiction, assume π is a root of $g \in \mathbb{Q}[X]$. Then $i\pi$ is a root of $g(iX)g(-iX)$ and it is straightforward to check $g(iX)g(-iX) \in \mathbb{Q}[X]$. Hence, $i\pi$ is a root of some monic $f \in \mathbb{Q}[X]$. By Exercise 3.8.4, $f = (X - \alpha_1)\cdots(X - \alpha_n)$ for some $n \in \mathbb{N}$ and $\alpha_i \in \mathbb{C}$ and, say, $\alpha_1 = i\pi$. By Euler's equality $e^{i\pi} = -1$, so

$$0 = (e^{\alpha_1} + 1)\cdots(e^{\alpha_n} + 1) = e^{\beta_1} + \cdots + e^{\beta_{2^n}} = e^{\beta_1} + \cdots + e^{\beta_r} + k.$$

Here, the β_j enumerate the sums $\epsilon_1\alpha_1+\cdots+\epsilon_n\alpha_n$ with $\epsilon_1,\ldots,\epsilon_n \in \{0,1\}$. For the 2nd equality we assume $\beta_1,\ldots,\beta_r$ list the $\beta_j \neq 0$ and set $k := 2^n - r \in \mathbb{N}$.

By Vieta's formula, $\prod_{j=1}^{2^n}(X-\beta_j) = X^{2^n-r}\prod_{j=1}^{r}(X-\beta_j)$ is a symmetric polynomial whose coefficients are $s_{2^n,\ell}(\beta_1,\ldots,\beta_{2^r})$ for $1 \leqslant \ell \leqslant 2^n$. But this is a symmetric polynomial in $\alpha_1,\ldots,\alpha_n$, so the coefficients are in $\mathbb{Q}$ by Corollary 3.7.9. Thus, for a suitable $a \in \mathbb{Z}$,

$$g := a(X-\beta_1)\cdots(X-\beta_r) \in \mathbb{Z}[X].$$

For a prime p to be chosen later, set $s := rp-1$ and define $h \in \mathbb{Z}[X]$ of degree $s+p$ by

$$h(X) := a^s X^{p-1} g(X)^p.$$

Define 'Hurwitz's device' $H \in \mathbb{Z}[X]$ using formal derivatives

$$H := h + h^{(1)} + \cdots + h^{(s+p)}.$$

Note $H' = h^{(1)} + \cdots + h^{(s+p)}$. By evaluation, H, h determine functions from $\mathbb{C}$ to $\mathbb{C}$, namely, $x \mapsto H(x)$ and $x \mapsto h(x)$. We denote these functions by $H(x), h(x)$ for a complex variable x. Using derivatives in the sense of calculus, we have $(e^{-x}H(x))' = -e^{-x}h(x)$ and hence

$$e^{-x}H(x) - e^0 H(0) = \int_0^x -e^{-z}h(z)dz.$$

Substituting yx for z gives

$$H(x) - e^x H(0) = -x\int_0^1 e^{(1-y)x}h(yx)dy.$$

Plug the β_i's for x and sum the equations:

$$\sum_{i=1}^r H(\beta_i) + kH(0) = -\sum_{i=1}^r \beta_i \int_0^1 e^{(1-y)\beta_i}h(y\beta_i)dy.$$

Consider the r.h.s. as a function of the (not displayed) prime p. It is easy to check, that the r.h.s. has absolute value $\leqslant c^p$ for some $c \in \mathbb{N}$. Dividing the r.h.s. by $p!$ gives a number with absolute value < 1 for all sufficiently large primes p. We get the desired contradiction by showing that the l.h.s. is an integer $\neq 0$ for all sufficiently large primes p.

We first examine $\sum_{i=1}^r H(\beta_i)$. Let $1 \leqslant j \leqslant r$. As β_j is a root of g, we have $h^{(t)}(\beta_j) = 0$ for all $t < p$. Let $t \geqslant p$. Since products of any p consecutive integers is divisible by $p!$, the coefficients of $h^{(t)}$ are divisible by $a^s \cdot p!$; also note $\deg(h^{(t)}) \leqslant s$. Thus $x_t := \sum_{j=1}^r h^{(t)}(\beta_j) = h_t(\beta_1,\ldots,\beta_r)$ for $h_t \in \mathbb{Z}[X_1,\ldots,X_r]$ symmetric of degree $\leqslant s$ with coefficients divisible by $p!a^s$. By Exercise 3.7.10, $x_t \in p!\mathbb{Z}$. We thus have

$$\sum_{j=1}^r H(\beta_j) \in p!\mathbb{Z}.$$

Obviously, $kH(0) \in \mathbb{Z}$, so it suffices to show $p \nmid kH(0)$ for sufficiently large primes p. Note $h^{(t)}(0) = 0$ for $t < p-1$ and $h^{(t)}(0) \in p!\mathbb{Z}$ for $t \geqslant p$. Further,

$$h^{(p-1)}(0) = a^s \cdot (p-1)! \cdot g(0)^p = a^s \cdot (p-1)! \cdot a a_0^p,$$

where $a_0 \in \mathbb{Z}$ is the constant coefficient of g. But no prime $p > k, a, a_0$ divides $kh^{(p-1)}(0)$. $\square$

Chapter 4

Ring theory

4.1 Quadratic integer rings

Let $d \in \mathbb{Z}$ be such that either $d = -1$, or, $|d| > 1$ is square free. Recall the quadratic number field $\mathbb{Q}(\sqrt{d}) = \mathbb{Q} + \mathbb{Q}\sqrt{d}$ from Corollary 3.5.12.

Definition 4.1.1. For $\alpha = x + y\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ with $x, y \in \mathbb{Q}$ let

$$\bar{\alpha} := x - y\sqrt{d}$$

be the *conjugation of* α. The *norm* and *trace* of α are

$$N(\alpha) := \alpha \cdot \bar{\alpha} = x^2 - y^2 d, \quad T(\alpha) := \alpha + \bar{\alpha} = 2x.$$

Remark 4.1.2.

1. Conjugation is an automorphism of $\mathbb{Q}(\sqrt{d})$ that fixes $\mathbb{Q}$. E.g., $\cdot$ is preserved:

$$\overline{(x + y\sqrt{d})(u + v\sqrt{d})} = xu + yvd - xv\sqrt{d} - yu\sqrt{d} = (x - y\sqrt{d})(u - v\sqrt{d}).$$

2. $N(\alpha \cdot \beta) = \alpha\beta\bar{\alpha}\bar{\beta} = N(\alpha) \cdot N(\beta)$ and $T(\alpha + \beta) = T(\alpha) + T(\beta)$.
3. $N(\alpha) = 0$ if and only if $\alpha = 0$.

   Indeed: $N(\alpha) = x^2 - y^2 d = 0$ clearly implies $x, y = 0$ in case $d < 0$. If $d > 0$, it suffices to show $y = 0$: otherwise $(x/y)^2 = d$, so $\sqrt{d} \in \mathbb{Q}$, a contradiction.

Exercise 4.1.3. $\mathbb{Q}(\sqrt{d})$ is a vector space over $\mathbb{Q}$ with basis $1, \sqrt{d}$. For $\alpha \in \mathbb{Q}(\sqrt{d})$ let $A_\alpha \in \mathbb{Q}^{2\times 2}$ be the matrix representing the linear map $\beta \mapsto \alpha \cdot \beta$ on this vector space.

Show $N(\alpha) = \det(A_\alpha)$, and $T(\alpha)$ equals the trace of A_α in the sense of linear algebra.

Definition 4.1.4. Call $\alpha \in \mathbb{Q}(\sqrt{d})$ *integral* if $N(\alpha), T(\alpha) \in \mathbb{Z}$. The set of integral elements is $\mathcal{O}_d$. It is called a *quadratic integer ring*, and *imaginary* if $d < 0$ and *real* if $d > 0$.

$\mathcal{O}_{-1}$ are also called the *Gaussian integers*, and $\mathcal{O}_{-3}$ the *Eisenstein integers*.

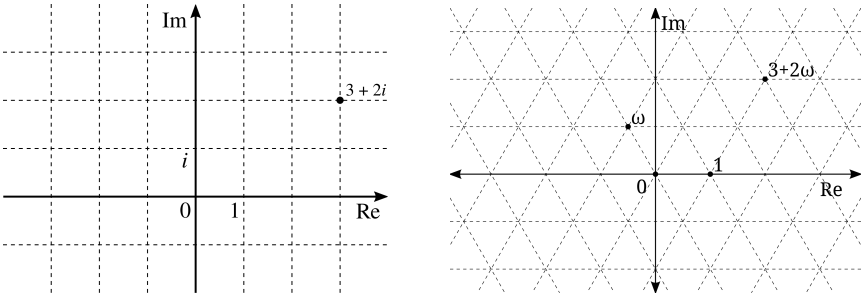


Figure 4.1: The Gaussian and Eisenstein integers

Lemma 4.1.5.

1. *If* $d \equiv 2,3 \mod 4$, *then* $\mathcal{O}_d = \mathbb{Z} + \mathbb{Z}\sqrt{d} := \left\{x + y\sqrt{d} \mid x, y \in \mathbb{Z}\right\}$.
2. *If* $d \equiv 1 \mod 4$, *then for* $\omega := (1+\sqrt{d})/2$

$$\mathcal{O}_d = \left\{(x + y\sqrt{d})/2 \mid x, y \in \mathbb{Z}, 2 \mid x - y\right\} = \mathbb{Z} + \mathbb{Z}\omega.$$

Proof. The 2nd equality in (2) is straightforward. $\supseteq$ is trivial in (1); in (2) let $x, y \in \mathbb{Z}$ with $x - y$ even and $\alpha = (x + y\sqrt{d})/2$. We show $N(\alpha) = (x^2 - y^2 d)/4 \in \mathbb{Z}$: since also $x + y$ is even, $4 \mid (x + y)(x - y) = x^2 - y^2$; then $4 \mid x^2 - y^2 - (d-1)y^2 = x^2 - y^2 d$.

$\subseteq$: let $x, y \in \mathbb{Q}$ and assume $\alpha := x + y\sqrt{d}$ is integral, so $T(\alpha) = 2x \in \mathbb{Z}$ and $N(\alpha) = x^2 - y^2 d \in \mathbb{Z}$. Say, $x = r/2$ for $r \in \mathbb{Z}$. Then $4N(\alpha) = (2x)^2 - d(2y)^2 \in \mathbb{Z}$ and $(2x)^2 \in \mathbb{Z}$, so $d(2y)^2 \in \mathbb{Z}$. Since d is square free, this implies $2y \in \mathbb{Z}$. Say, $y = s/2$ for $s \in \mathbb{Z}$. Then $N(\alpha) = r^2/4 - ds^2/4 \in \mathbb{Z}$, so

$$r^2 - ds^2 \equiv 0 \mod 4.$$

Note all squares are 0 or 1 modulo 4. Hence, in (1) we get $r^2 \equiv s^2 \equiv 0 \mod 4$. Then r, s are even and $x, y \in \mathbb{Z}$. In (2), write $r^2 - s^2 = (r+s)(r-s) \equiv 0 \mod 4$. Then $r \equiv s \mod 2$, say $r = s + 2t$ with $t \in \mathbb{Z}$. Then $\alpha = (r/2) + (s/2)\sqrt{d} = t + s\omega$. □

Corollary 4.1.6. *$\mathcal{O}_d$ is a subring of $\mathbb{Q}(\sqrt{d})$ and, in particular, an integral domain.*

Proof. Clearly, $1, 0 \in \mathcal{O}_d$. Let $\alpha, \beta \in \mathcal{O}_d$. By the lemma, $\alpha + \beta, -\alpha \in \mathcal{O}_d$ and we have to show $\alpha \cdot \beta \in \mathcal{O}_d$. This is obvious if $d \equiv 2,3 \mod 4$. If $d \equiv 1 \mod 4$, say $d = 4r + 1$ for $r \in \mathbb{Z}$, note $\alpha \cdot \beta \in \mathbb{Z} + \mathbb{Z}\omega + \mathbb{Z}\omega^2$ and $\omega^2 = (1/4 + \sqrt{d}/2 + d/4) = \omega + r$, so $\mathbb{Z}\omega^2 \subseteq \mathbb{Z} + \mathbb{Z}\omega$. □

Lemma 4.1.7 (Pell equalities)**.**

1. $\mathcal{O}_d^\times = \left\{\alpha \in \mathcal{O}_d \mid N(\alpha) = \pm 1\right\}$.
2. *If* $d \equiv 2,3 \mod 4$, *then for all* $x, y \in \mathbb{Z}$: $x + y\sqrt{d} \in \mathcal{O}_d^\times \Leftrightarrow x^2 - y^2 d = \pm 1$.
3. *If* $d \equiv 1 \mod 4$, *then for all* $x, y \in \mathbb{Z}$: $(x + y\sqrt{d})/2 \in \mathcal{O}_d^\times \Leftrightarrow x^2 - y^2 d = \pm 4$.

Proof. (1) $\subseteq$: $1 = N(\alpha\alpha^{-1}) = N(\alpha)N(\alpha^{-1})$ and $N(\alpha) \in \mathbb{Z}^\times = \{\pm 1\}$. $\supseteq$: if $\pm 1 = N(\alpha) = \alpha\bar{\alpha}$, then $\pm\bar{\alpha}$ is an inverse of α.

(2) and (3) $\Rightarrow$ follows from (1). (3)$\Leftarrow$: $x^2 - y^2 d \equiv (x-y)(x+y) \equiv 0 \mod 4$ implies $x - y$ is even, so $(x + y\sqrt{d})/2 \in \mathcal{O}_d$. □

This allows to determine the units in imaginary quadratic number rings. Recall C_n denotes the group of n-th roots of unity (Definition 1.6.8).

Corollary 4.1.8. $\mathcal{O}_d^\times = \begin{cases} C_2 = \{\pm 1\} & \text{if } d = -2 \text{ or } d < -3 \\ C_4 = \{\pm 1, \pm i\} & \text{if } d = -1 \\ C_6 = \{\pm 1, (\pm 1 \pm i\sqrt{3})/2\} & \text{if } d = -3. \end{cases}$

Proof. For $d \equiv 2, 3 \mod 4$ and $d \neq -1$, the Pell equation $x^2 + y^2|d| = 1$ has only 'trivial' solutions$(x, y) = (\pm 1, 0)$; for $d = -1$ we additionally have $(x, y) = (0, \pm 1)$, so get $\pm i \in \mathcal{O}_{-1}^\times$.

For $d \equiv 1 \mod 4$ and $d < -3$ we have $d \leqslant -7$, and the Pell equation $x^2 + y^2|d| = 4$ has only trivial solutions. For $d = -3$, we have additionally $(\pm 1, \pm 1)$, so get $(\pm 1 \pm i\sqrt{3})/2 \in \mathcal{O}_{-3}^\times$. □

Units in real quadratic number rings are more difficult to determine.

Example 4.1.9. $\mathcal{O}_2^\times$ is infinite: $\epsilon := 1 + \sqrt{2}$ has norm -1 so is in $\mathcal{O}_2^\times$. Also $-\epsilon = -1 - \sqrt{2}, \epsilon^{-1} = -1 + \sqrt{2}, -\epsilon^{-1} = 1 - \sqrt{2} \in \mathcal{O}_2^\times$. The powers $(1 + \sqrt{2})^k$ have norm $(-1)^k$, so are in $\mathcal{O}_2^\times$ and pairwise distinct (their absolute values in $\mathbb{R}$ grow).

Similarly, $\mathcal{O}_5^\times$ is infinite: it contains $\epsilon := 1/2 + \sqrt{5}/2, -\epsilon, \epsilon^{-1}, -\epsilon^{-1}$ and powers.

Remark 4.1.10 (Integer rings)**.** The following explains the wording "integral" and the rationale behind its definition. (3) is used to define the *integer ring of* $\mathbb{Q}(\alpha)$ for any algebraic $\alpha \in \mathbb{C}$. Recall Definition 3.5.2.

Theorem 4.1.11. *For $\alpha \in \mathbb{Q}(\sqrt{d})$ the following are equivalent.*

1. *α is integral.*
2. *$m_\alpha^\mathbb{Q} \in \mathbb{Z}[X]$.*
3. *α is a root of a monic polynomial in $\mathbb{Z}[X]$.*

Yet incomplete proof. $1 \Leftrightarrow 2$: if $\alpha \in \mathbb{Q}$, it has minimal polynomial $X - \alpha$ and, by Lemma 4.1.5, being integral means $\alpha \in \mathbb{Z}$. So assume $\alpha \notin \mathbb{Q}$. Then $\deg(m_\alpha^\mathbb{Q}) > 1$. Hence,

$$m_\alpha^\mathbb{Q} = (X - \alpha)(X - \bar{\alpha}) = X^2 - T(\alpha)X + N(\alpha).$$

$2 \Rightarrow 3$ is trivial. For $3 \Rightarrow 2$ choose a monic *integer* polynomial of minimal degree that has α as a root. Is this the minimal polynomial? Or is it possible to find a monic *rational* polynomial of smaller degree? We need a better understanding of divisibility in rings.

We shall complete the proof in Section 4.4 as follows. Assume α is a root of a monic $f \in \mathbb{Z}[X]$. Then $f = m_\alpha^\mathbb{Q} \cdot g$ for some $g \in \mathbb{Q}[X]$ by Lemma 3.5.6. But $f, m_\alpha^\mathbb{Q}$ are monic, and we shall prove in Lemma 4.4.15 that this implies $m_\alpha^\mathbb{Q} \in \mathbb{Z}[X]$. □

4.2 Irreducible and prime elements

In this section we make the discomforting observation that there exist natural rings where divisibility behaves quite differently from what we are used to from the integers or polynomials, in particular, we face rings that violate Euclid's lemma 2.2.7.

Let R be a commutative ring.

Definition 4.2.1. Let $x, y \in R$. Then x is a *divisor of y (in R)*, and y a *multiple of x (in R)*, symbolically $x \mid y$, if $x \cdot z = y$ for some $z \in R$. The set of multiples of x is

$$xR := \{xz \mid z \in R\}.$$

We say, x and y are *associate (in R)*, symbolically $x \sim y$, if both $x \mid y$ and $y \mid x$.

Remark 4.2.2. For all $x, x', y, y', z, u, u' \in R$:

1. $x \mid 0, 1 \mid x, -1 \mid x, x \mid x, -x \mid x, x \mid -x$ (recall Lemma 1.1.14 (3)).
2. $\mid$ is transitive: if $x \mid y$ and $y \mid z$, then $x \mid z$.
3. $\sim$ is an equivalence relation on R.
4. If $x \mid y$ and $x \mid y'$, then $x \mid uy + u'y'$.
5. If $x \mid y$ and $x' \mid y'$, then $xx' \mid yy'$.
6. If $x \mid y$, then $x\varepsilon \mid y\varepsilon'$ for all $\varepsilon, \varepsilon' \in R^\times$. Indeed: if $xz = y$, then $x\varepsilon(\varepsilon^{-1}z\varepsilon') = y\varepsilon'$.
7. If $x\varepsilon = y$ for some $\varepsilon \in R^\times$, then $x = y\varepsilon^{-1}$, so $x \sim y$.
8. $x \mid y \iff x \in yR \iff yR \subseteq xR$; in particular, $x \sim y \iff xR = yR$.
9. $x \in R^\times \iff x \mid 1 \iff xR = R$.

Definition 4.2.3. $x \in R$ is *irreducible (in R)* if $x \neq 0, x \notin R^\times$ and x is not the product of two non-units, i.e., all divisors of x are units or $\sim x$. Otherwise x is *reducible (in R)*.

Lemma 4.2.4. *Let R be an integral domain and $x, y \in R$.*

1. *$x \sim y$ if and only if $x\varepsilon = y$ for some $\varepsilon \in R^\times$.*
2. *If x is irreducible and associate to y, then y is irreducible.*
3. *x is irreducible if and only if $x \neq 0$, $xR \neq R$ and for all $y \in R$: $xR \subsetneq yR$ implies $yR = R$.*

Proof. (1) $\Rightarrow$ is Remark 4.2.2 (7). (1) $\Leftarrow$: assume $xz = y, yz' = x$; if $x = 0$, then $y = 0$ and $x = 1y$; if $x \neq 0$, then $xzz' = x$, so $zz' = 1$ as R is an integral domain; thus, $z, z' \in R^\times$.

(2): if $x = \varepsilon y$ for some $\varepsilon \in R^\times$, then $y \neq 0$ and $y \notin R^\times$ (otherwise $x = 0$ or $x \in R^\times$); by Remark 4.2.2 (6), x, y have the same divisors; hence, y is irreducible.

(3) $\Rightarrow$: $xR \neq R$ as $x \notin R^\times$ by Remark 4.2.2 (9); if $xR \subseteq yR$, then $y \mid x$ by Remark 4.2.2 (8), so y is a unit or associate to x. Then $yR = R$ or $yR = xR$ by Remark 4.2.2 (9), (8).

(3) $\Leftarrow$: $x \notin R^\times$ by Remark 4.2.2 (9); assume $y \mid x$; then $xR \subseteq yR$ by Remark 4.2.2 (8), so $xR = yR$ or $yR = R$; this implies $x \sim y$ or $y \in R^\times$ by Remark 4.2.2 (9), (8). □

Examples 4.2.5.

1. Recall $\mathbb{Z}^\times = \{\pm 1\}$. Remark 2.1.2 (6) becomes (1) above. By definition, the irreducible elements of $\mathbb{Z}$ are $\pm p$ for prime numbers $p \in \mathbb{N}$.
2. Let K be a field. Then $f \in K[X] \setminus \{0\}$ is irreducible by the above definition if and only if it is irreducible by Definition 3.5.5 (recall $K[X]^\times = K^\times$ by Lemma 3.1.6).
3. $2X \in \mathbb{Z}[X]$ is not the product of two polynomials of positive degree but reducible as $2, X \notin \mathbb{Z}[X]^\times = \mathbb{Z}^\times = \{\pm 1\}$.
4. $2 = (1+i)\cdot(1-i)$ is reducible in $\mathcal{O}_{-1}$ – note $2, 1 \pm i$ are not units by Corollary 4.1.8.
5. $2 = \sqrt{2}\cdot\sqrt{2}$ is reducible in $\mathcal{O}_2$ – note $2, \sqrt{2}$ are not units by Lemma 4.1.7 (1).
6. $2, 3, 1 \pm i\sqrt{5}, 2 \pm i\sqrt{5}$ are irreducible in $\mathcal{O}_{-5}$.

   Indeed: assume α is one of these and $\alpha = \beta\gamma$. Then $N(\alpha) = N(\beta)N(\gamma) \in \{4, 6, 9\}$. Assume β, γ are not units, so have norm $\neq \pm 1$. Then $N(\beta) \in \{\pm 2, \pm 3\}$. But $\pm 2, \pm 3$ are not of the form $x^2 + y^2 5$ for $x, y \in \mathbb{Z}$.

Exercise 4.2.6. Let $\mathcal{O}_d$ be a quadratic integer ring.

1. If $\alpha \in \mathcal{O}_d$ and $N(\alpha)$ is prime, then α is irreducible in $\mathcal{O}_d$.
2. A prime $p \in \mathbb{N} \subseteq \mathcal{O}_d$ is reducible if and only if there exists $\alpha \in \mathcal{O}_d$ with $N(\alpha) = p$.

Euclid's lemma 2.2.7 states that in $\mathbb{Z}$ irreducible elements are *prime*:

Definition 4.2.7. $p \in R$ is *prime (in R)* if $p \neq 0, p \notin R^\times$ and for all $x, y \in R$:

$$p \mid xy \text{ implies } p \mid x \text{ or } p \mid y.$$

Exercise 4.2.8. Let R be an integral domain, $p \in R$ be prime and $x, y, q, x_1, \ldots, x_n \in R$.

1. If $p \mid x_1 \cdots x_n$, then $p \mid x_i$ for some i (by induction on n).
2. If $q \sim p$, then q is prime.
3. If R is an integral domain and q is prime and $p \mid q$, then $p \sim q$.

Lemma 4.2.9. *Let R be an integral domain. If $p \in R$ is prime, then p is irreducible.*

Proof. If $x \mid p$, say $p = xy$ for some $y \in R$, then $p \mid x$ or $p \mid y$. In the 1st case, $p \sim x$. In the 2nd, $xpz = p$ for some $z \in R$, so $xz = 1$ (as R is an integral domain), so x is a unit. □

Example 4.2.10. $\bar{3}$ is prime in $\mathbb{Z}_6$ but not irreducible as $\bar{3} = \bar{3}\cdot\bar{3}$.

Recall, in $\mathbb{Z}$, the converse is stated as Euclid's lemma 2.2.7. Recall also that, in $K[X]$, we copied Euclid's algorithm. We can also copy the proof of Euclid's lemma:

Lemma 4.2.11. *Let K be a field. In $K[X]$ irreducibles are prime.*

Proof. Assume $f, g, h \in K[X]$, f is irreducible, $f \mid gh$. Let $d = \gcd(f, g)$. Since f is irreducible, $d \sim f$ or $d \in K[X]^\times = K \setminus \{0\}$. In the 1st case, $f \mid g$. In the 2nd, $d = 1$. By Theorem 3.2.7, $1 = sf + tg$ for certain $s, t \in K[X]$. Then $h = sfh + tgh$, so $f \mid h$. □

Example 4.2.12. Euclid's lemma fails in some integral domains:

1. Let $R \subseteq \mathbb{R}[X]$ contain the polynomials $a_0 + a_1X + \cdots$ with $a_0 \in \mathbb{Q}$. It is easy to check that R is a subring. In R, X is irreducible (having degree 1) but not prime: $X \mid (\sqrt{2}X)^2$ but $X \nmid \sqrt{2}X$ since $\sqrt{2} \notin R$.
2. 2 is irreducible in $\mathcal{O}_{-5}$ (by Example 4.2.5) but not prime: $2 \mid 6 = (1 + i\sqrt{5})(1 - i\sqrt{5})$ but $2 \nmid 1 \pm i\sqrt{5}$: if $1 \pm i\sqrt{5} = 2\alpha$, then $6 = 4N(\alpha)$, contradicting $N(\alpha) \in \mathbb{Z}$.

Exercise 4.2.13. Let K be a field and $R \subseteq K[X]$ be the set of $a_nX^n + \cdots + a_0$ with $n \in \mathbb{N}, a_i \in K$ and $a_1 = 0$. Show R is a subring. Show X^3 is irreducible and not prime in R.

Exercise 4.2.14. Let $\mathcal{O}_d$ be a quadratic integer ring and π be prime in $\mathcal{O}_d$. Then there is a unique prime number $p \in \mathbb{N}$ such that $\pi \mid p$; moreover, $|N(\pi)| \in \{p, p^2\}$. If $|N(\pi)| = p^2$, then $\pi \sim p$ (in $\mathcal{O}_d$).

Lemma 4.2.15. *Let R be an integral domain and $x, y \in R$.*

1. *$x \mid y$ in R if and only if $x \mid y$ in $R[X]$.*
2. *$x \sim y$ in R if and only if $x \sim y$ in $R[X]$.*
3. *x is irreducible in R if and only if x is irreducible in $R[X]$.*
4. *x is prime in R if and only if x is prime in $R[X]$.*

Proof. We leave (1)-(3) as an exercise. (4) $\Leftarrow$: assume $x \mid yz$ for $y, z \in R$, so $x \mid yz$ in $R[X]$. Then $x \mid y$ or $x \mid z$ in $R[X]$ (x being prime in $R[X]$), so $x \mid y$ or $x \mid z$ in R by (1).

(4) $\Rightarrow$: assume x is not prime in $R[X]$. Choose $f, g \in R[X]$ such that $x \mid fg$, $x \nmid f$ and $x \nmid g$ in $R[X]$. We claim x is not prime in R. Write $f = a_nX^n + \cdots + a_0$ and $g = b_mX^m + \cdots + b_0$.

Let k be the minimal $i \leqslant n$ such that $x \nmid a_i$, and ℓ be the minimal $j \leqslant m$ such that $x \nmid b_j$. Write $f = f_0 + f_1$ with $f_0 := a_{k-1}X^{k-1} + \cdots + a_0$ (the empty sum is 0) and $f_1 = a_nX^n + \cdots + a_kX^k$. Analogously write $g = g_0 + g_1$. Then x divides f_0g_0, f_0g_1, f_1, g_0 and $fg = f_0g_0 + f_1g_0 + f_0g_1 + f_1g_1$, so $x \mid f_1g_1$. Say $xh = f_1g_1$ and let h have coefficients c_i. Then $xc_{k+\ell} = a_kb_\ell$. Hence $x \mid a_kb_\ell, x \nmid a_k, x \nmid b_\ell$ and x is not prime. □

4.3 Factorial rings

In this section we shall see that reasoning about divisibility follows familiar lines in rings that allow prime factorizations:

Definition 4.3.1. A ring is *factorial* if it is an integral domain such that every nonzero non-unit is a (finite) product of prime elements.

A more honest generalization of prime factorization in $\mathbb{Z}$ is (2) below.

Theorem 4.3.2. *Let R be an integral domain. The following are equivalent.*

1. *R is factorial.*

2. *Every nonzero non-unit of R is a product of irreducible elements that is* essentially unique*: if $n, m > 1$ and $q_1 \cdots q_n = q'_1 \cdots q'_m$ for irreducible $q_1, ..., q_n, q'_1, ..., q'_m \in R$, then $n = m$ and, after a possible re-enumeration, $q_i \sim q'_i$ for all $1 \leqslant i \leqslant n$.*
3. *Irreducibles are prime and every nonzero non-unit of R is a product of irreducibles.*

Proof. $1 \Rightarrow 3$: by Lemma 4.2.9, a decomposition into primes is one into irreducibles. Every irreducible $q \in R$ is prime: write $q = p_1 ... p_n$ for primes $p_i \in R$; then $n = 1$ and $q = p_1$.

$3 \Rightarrow 2$: it suffices to verify essential uniqueness for prime elements q_i, q'_j. If $q_1 \cdots q_n = q'_1 \cdots q'_m$, then $q_1 \sim q'_j$ for some j by Exercise 4.2.8 (1) and (3). We can assume $i = 1$. Then $q_1 = \varepsilon q'_1$ by Lemma 4.2.4 (1) and $\varepsilon q'_1 q_2 \cdots q_n = q'_1 \cdots q'_m$. Since R is an integral domain, $\tilde{q}_2 q_3 \cdots q_n = q'_2 \cdots q_m$ for $\tilde{q}_2 := \varepsilon q_2$, prime by Exercise 4.2.8 (2). Continuing yields the claim.

$2 \Rightarrow 1$: it suffices to show every irreducible $q \in R$ is prime. Let $q \mid xy$, say $qz = xy$. Write $x = q_1 \cdots q_n, y = q'_1 \cdots q'_m, z = q''_1 \cdots q''_\ell$ for irreducibles q_i, q'_j, q''_k. Then $q q''_1 \cdots q''_\ell = q_1 \cdots q_n q'_1 \cdots q'_m$. By essential uniqueness, q is associate to some q_i or some q'_k. Then $q = \varepsilon q_i$ or $q = \varepsilon q''_k$ for some $\varepsilon \in R^\times$ by Lemma 4.2.4 (1). As $q_i \mid x, q''_k \mid y$ we get $q \mid x$ or $q \mid y$ by Remark 4.2.2 (6). □

Remark 4.3.3. Let R be a factorial ring and let $P \subseteq R$ *represent the primes in* R: it contains exactly one element of every $\sim$-equivalence class of a prime element of R. Then for every nonzero non-unit $x \in R$ there are unique $n, e_1, ..., e_n > 0, p_1, ..., p_n \in P$ and $\varepsilon \in R^\times$ such that

$$x = \varepsilon p_1^{e_1} \cdots p_n^{e_n}.$$

For $R = \mathbb{Z}$ and P the set of prime numbers, this is the fundamental theorem.

Example 4.3.4. $\mathcal{O}_{-5}$ is not factorial by Example 4.2.12. By Example 4.2.5 (6) we have decompositions into pairwise not associate irreducibles (recall $\mathcal{O}_{-5}^\times = \{\pm 1\}$ by Corollary 4.1.8):

$$6 = 2 \cdot 3 = (1 + i\sqrt{5})(1 - i\sqrt{5}), \quad 9 = 3 \cdot 3 = (2 + i\sqrt{5})(2 - i\sqrt{5}).$$

Exercise 4.3.5. The subring $\mathbb{Z} + \mathbb{Z}\sqrt{-3} \subseteq \mathcal{O}_{-3}$ is not factorial.

Lacking a notion of size like $|\cdot|$, it is not obvious how to generalize gcd and lcm to an integral domain R. But we can use Remark 2.1.9 (1) and Exercise 2.1.15 (2):

Definition 4.3.6. Let R be an integral domain, $n > 0$ and $x_1, ..., x_n \in R$ not all zero. A *common divisor of* $x_1, ..., x_n$ is an $x \in R$ such that $x \mid x_i$ for all i; it is *greatest* if additionally all common divisors divide x. If 1 is a greatest common divisor, then $x_1, ..., x_n$ are *coprime*.

A *common multiple of* $x_1, ..., x_n$ is an $x \in R$ such that $x_i \mid x$ for all i; it is *least* if additionally x divides all common multiples.

Remark 4.3.7. In case greatest common divisors or least common multiples exist at all, we avoid the functional notations gcd, lcm because we only have 'almost' uniqueness: any two greatest common divisors are associate, any two least common multiples are associate.

Proposition 4.3.8. *Let R be a factorial ring and $n > 0$ and $x_1, ..., x_n \in R \setminus \{0\}$. Let P represent the primes in R. Let $m > 0$ and $p_1, ..., p_m \in P$ and $e_{ij} \in \mathbb{N}$ for $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m$, and $\varepsilon_1 ..., \varepsilon_n \in R^\times$ such that for all $1 \leqslant i \leqslant n$:*

$$x_i = \varepsilon_i p_1^{e_{i1}} \cdots p_m^{e_{im}}.$$

Then $x_1, ..., x_n$ have greatest common divisor and least common multiple

$$p_1^{\min_i e_{i1}} \cdots p_m^{\min_i e_{im}} \quad \text{and} \quad p_1^{\max_i e_{i1}} \cdots p_m^{\max_i e_{im}}.$$

Proof. We treat gcd (lcm is similar). It is clear that $d := p_1^{\min_i e_{i1}} \cdots p_m^{\min_i e_{im}}$ is a common divisor. Let c be a common divisor and write $c = \varepsilon q_1^{e_1} \cdots q_\ell^{e_\ell}$ according Remark 4.3.3, in particular with $e_j > 0$. Fix $1 \leqslant i \leqslant n$. Then every q_j divides x_i and hence some $p_i^{e_{ik}}$, so $q_j = p_i$ by Exercise 4.2.8 (3). Thus $\{q_1, ..., q_\ell\} \subseteq \{p_1, ..., p_n\}$ and $c = \varepsilon p_1^{e'_1} \cdots p_m^{e'_m}$ for certain $e'_j \in \mathbb{N}$. Write $cy = x_i$ for some $y \in R$. As $y \mid x_i$ we similarly write $y = \varepsilon' p_1^{e''_1} \cdots p_m^{e''_m}$ for certain $e''_j \in \mathbb{N}$. Then $x_i = \varepsilon\varepsilon' p_1^{e'_1 + e''_1} \cdots p_m^{e'_m + e''_m}$. It follows $e_{ij} = e'_j + e''_j$ by uniqueness Remark 4.3.3. Hence $e'_j \leqslant e_{ij}$ for all i, j. As $1 \leqslant i \leqslant n$ was arbitrary, $e'_j \leqslant \min_i e_{ij}$. It follows that $c \mid d$. □

Exercise 4.3.9. Let R be a factorial ring, $y, z \in R$, $n > 0$ and $x_1, ..., x_n \in R$ not all zero.

1. $x_1, ..., x_n$ are coprime if and only if there does not exists a prime element $p \in R$ that divides all x_i.
2. If y is a greatest common divisor of $x_1, ..., x_n$ and $y_1 y = x_1, ..., y_n y = x_n$, then $y_1, ..., y_n$ are coprime.
3. If y is a greatest common divisor of $x_1, ..., x_n$, then zy is a greatest common divisor of $zx_1, ..., zx_n$.
4. If x, y are coprime and $x \mid yz$, then $x \mid z$.

Lemma 4.3.10. *Let R be a factorial ring.*

1. *For all $x \in R, y \in R \setminus \{0\}$ there are coprime $x' \in R, y' \in R \setminus \{0\}$ such that $x/y = x'/y'$ in $\mathrm{Quot}(R)$; such x', y' are unique up to $\sim$.*
2. *For all $n > 0$ and $x_1, ..., x_n, y_1, ..., y_n \in R \setminus \{0\}$ there exists $z \in R$ such that $zx_1/y_1, ..., zx_n/y_n \in R$ and have the same greatest common divisors as $x_1, ..., x_n$.*

Proof. (1): write $x = zx', y = zy'$ for a greatest common divisor z of x, y. Then $x/y = x'/y'$ in $\mathrm{Quot}(R)$ and x', y' are coprime by Exercise 4.3.9 (2).

(2): write $x_i = \varepsilon_i p_1^{e_{i1}} \cdots p_m^{e_{im}}$ and $y_i = \delta_i q_1^{d_{i1}} \cdots q_m^{d_{im}}$ according Proposition 4.3.8 for suitable primes p_i, q_i and units ε_i, δ_i. We can assume $\{p_1, ..., p_m\} \cap \{q_1, ..., q_m\} = \varnothing$.

Let $z := q_1^{\max_i d_{i1}} \cdots q_m^{\max_i d_{im}}$ be a least common multiple of the y_i. Then

$$zx_j/y_j = \delta_j^{-1} \varepsilon_j q_1^{\max_i d_{i1} - d_{j1}} \cdots q_m^{\max_i q_{im} - d_{jm}} p_1^{e_{j1}} \cdots p_m^{e_{jm}}$$

for all $1 \leqslant j \leqslant n$. By Proposition 4.3.8 a greatest common divisor of the zx_j/y_j is

$$q_1^{\min_j(\max_i d_{i1} - d_{j1})} \cdots q_m^{\min_j(\max_i d_{im} - d_{jm})} p_1^{\min_i e_{i1}} \cdots p_m^{\min_i e_{im}} = p_1^{\min_i e_{i1}} \cdots p_m^{\min_i e_{im}}.$$

This is a greatest common divisor of the x_j's. This is enough by Remark 4.3.7. □

4.4 Polynomial factorization

Recall our main interest are polynomial rings. This section gives good news showing they are well-behaved:

Theorem 4.4.1 (Gauß)**.** *If R is a factorial ring, then so is $R[X]$.*

$\mathbb{Z}[X_1]$ is factorial. Then $\mathbb{Z}[X_1, X_2] \cong \mathbb{Z}[X_1][X_2]$ is factorial. And so on. Thus:

Corollary 4.4.2. *Let $n > 0$ and K be a field. Then $\mathbb{Z}[X_1, ..., X_n]$ and $K[X_1, ..., X_n]$ are factorial.*

We aim to verify Theorem 4.3.2 (3) and first ask for decompositions into irreducibles. We explore the straightforward idea to repeatedly replace reducible factors by products.

Definition 4.4.3. Let R be an integral domain. A *proper divisor chain* in R is a sequence $(x_n)_{n\in\mathbb{N}}$ of elements of R such that $x_{n+1} \mid x_n$ and $x_{n+1} \nmid x_n$ for all $n \in \mathbb{N}$.

Example 4.4.4. $\mathbb{Z}$ does not have proper divisor chains. Indeed: if $y \mid x$, then $|y| \leqslant |x|$ and $|y| = |x|$ means $y \sim x$. Hence, a proper divisor chain satisfies $|x_0| > |x_1| > \cdots$, impossible.

Lemma 4.4.5. *A factorial ring R does not have proper divisor chains.*

Proof. Assume $(x_n)_n$ is a proper divisor chain. Let P represent primes in R and write $x_0 = \varepsilon_0 p_1^{e_1} \cdots p_m^{e_m}$ according Remark 4.3.3. Then $x_1 = \varepsilon_1 p_1^{e_1'} \cdots p_m^{e_m'}$ with $\varepsilon_2 \in R^\times$ and $e_i' \leqslant e_i$ for all i and $e_i' < e_i$ for at least one i. Continuing gives $x_n \in R^\times$ for $n := \sum_i e_i$. Then $x_{n+1} \in R^\times$ and $x_{n+1} \sim x_n$, a contradiction. □

Lemma 4.4.6. *In an integral domain R without proper divisor chains, every nonzero non-unit is a product of irreducibles.*

Proof. Assume $x_0 \in R$ is a nonzero non-unit that is *not* a product of irreducibles. Then x_0 is reducible, so $x_0 = y_1 y_2$ where y_1, y_2 are nonzero non-units. Then $y_1, y_2 \nmid x_0$: if, say, $y_1 \sim x$, i.e., $y_1 = \varepsilon x$ for $\varepsilon \in R^\times$, then $x = \varepsilon x y_2$, so $1 = \varepsilon y_2$ (R is an integral domain) and $y_2 \in R^\times$, contradiction. At least one of y_1, y_2 is *not* a product of irreducibles. Call it x_1. Continuing gives a proper divisor chain. □

Lemma 4.4.7. *If R is an integral domain without proper divisor chains, so is $R[X]$.*

Proof. Let $(f_n)_n$ be a proper divisor chain in $R[X]$. Then all f_n are nonzero, so have degree $\geqslant 0$. Then $\deg(f_0) \geqslant \deg(f_1) \geqslant \cdots$ so there are $d, n_0 \in \mathbb{N}$ such that $\deg(f_n) = d$ for all $n \geqslant n_0$. Let a_n be the lead coefficient of f_n. For $n \geqslant n_0$ write $f_{n+1} g_n = f_n$ and note $d = \deg(f_n) = \deg(f_{n+1}) + \deg(g_n) = d + \deg(g_n)$, so $\deg(g_n) = 0$, so $g_n \in R$. Then $a_n = a_{n+1} g_n$, so $a_{n+1} \mid a_n$ in $R[X]$ and hence in R by Lemma 4.2.15 (1). Since R does not have a proper divisor chain, there is $m \geqslant n_0$ such that $a_{m+1} = \varepsilon a_m$ for some $\varepsilon \in R^\times$. But $a_m = a_{m+1} g_m$, so $g_m \in R^\times = R[X]^\times$ by Lemma 3.1.6. Thus, $f_{m+1} \sim f_m$ in $R[X]$. □

Corollary 4.4.8. *If R is a factorial ring, then every nonzero non-unit in $R[X]$ is a finite product of irreducibles.*

Corollary 4.4.9. *Let K be a field. Then $K[X]$ is factorial.*

Proof. K is factorial, so in $K[X]$ nonzero non-units are finite products of irreducibles. But irreducibles are prime by Lemma 4.2.11. By Theorem 4.3.2, $K[X]$ is factorial. □

We shall re-prove this in the next section by more abstract means.

Exercise 4.4.10. In $\mathbb{F}_5[X]$ we have $\bar{3}X^2 + \bar{4}X + \bar{3} = (\bar{3}X + \bar{2})(X + \bar{4}) = (\bar{4}X + \bar{1})(\bar{2}X + \bar{3})$. Why does this not contradict the uniqueness of factorizations? (*Hint:* $\bar{2} \cdot \bar{3} = \bar{1}$)

We next aim to show that irreducibels are prime in $\mathbb{Z}[X_1, ..., X_n]$ and $K[X_1, ..., X_n]$.

Definition 4.4.11. Let R be a factorial ring, and $f \in R[X] \setminus \{0\}$. A *content of f* is a greatest common divisor of the coefficients of f. f is *primitive* if 1 is a content of f.

Example 4.4.12. In $\mathbb{Z}[X]$, the contents of $12X^3 + 16X + 8$ are ± 4. E.g., $49X^5 + 10X$ and $3X + 4$ are primitive.

Remark 4.4.13. Let R be a factorial ring and $f \in R[X] \setminus \{0\}$ with content $a \in R$.

1. If $b \in R$ is a content of f, then $a \sim b$ (by Remark 4.3.7).
2. There is a primitive $g \in R[X]$ such that $f = ag$.

   Indeed: define g from f replacing every coefficient c by some b with $ba = c$; these b are coprime by Exercise 4.3.9 (2).
3. If $b \in R \setminus \{0\}$, then bf has content ba (by Exercise 4.3.9 (3)).

Lemma 4.4.14 (Gauß)**.** *Let R be a factorial ring. If $f, g \in R[X]$ have contents a, b, then fg has content ab.*

Proof. By Remark 4.4.13 (2) write $f = af', g = bg'$ with primitive f', g'. By Remark 4.4.13 (3), a content of fg is abc with c a content of $f'g'$. We claim that we can take $c = 1$, i.e., that $f'g'$ is primitive. Write $f' = a_nX^n + \cdots + a_0, g' = b_mX^m + \cdots + b_0$ and $f'g' = c_{m+n}X^{n+m} + \cdots + c_0$.

Assume $f'g'$ is not primitive, say p is a prime divisor of the c_j's. Let r be the minimal $i \leqslant n$ such that $p \nmid a_i$. Let s be the minimal $j \leqslant m$ such that $p \nmid b_j$. Note $c_{r+s} = \sum_{i+j=r+s} a_ib_j$. If $(i, j) \neq (r, s)$, then $i < r$ or $j < s$, so $p \mid a_ib_j$. Thus, p divides all $a_ib_j \neq a_rb_s$. As $p \mid c_{r+s}$, we get $p \mid a_rb_s$. As p is prime, $p \mid a_r$ or $p \mid b_s$, a contradiction. □

Lemma 4.4.15. *Let R be a factorial ring, $f \in R[X] \setminus \{0\}$, $g, h \in \mathrm{Quot}(R)[X]$ and $f = gh$.*

1. *There are $\hat{g}, \hat{h} \in R[X]$ of the same degree as g, h such that $f = \hat{g}\hat{h}$.*
2. *If $g \in R[X]$ is primitive, then $h \in R[X]$.*
3. *If f, g are monic, then $g, h \in R[X]$.*

Proof. (1): by Lemma 4.3.10 (2) there are $a, b \in R$ such that $ag, bh \in R[X]$. By Remark 4.4.13 (3) there are $a', b' \in R$ and primitive $g', h' \in R[X]$ such that $ag = a'g', bh = b'h'$. Then $abf = a'b'g'h'$. Let c be a content of f. By Remark 4.4.13 (4), abc is a content of abf. By Gauß' Lemma, $a'b' \cdot 1 \cdot 1$ is also a content of abf, so $abc\varepsilon = a'b'$ for some $\varepsilon \in R^\times$ by Remark 4.4.13 (1). Hence, $abf = abc\varepsilon \cdot g'h'$, so $f = c\varepsilon g'h'$. Set $\hat{g} := g', \hat{h} := c\varepsilon h'$.

(2): if $g \in R[X]$ is primitive, we can choose $a = a' = 1$ above. Then $g = g' = \hat{g}$, so $gh = f = \hat{g}\hat{h} = g\hat{h}$ and hence $h = \hat{h} \in R[X]$.

(3): if f, g are monic, so is h. By Lemma 4.3.10 (2), ag, bh are primitive (their coefficients' gcd equals that of nominators in g, h). Hence we can choose $a' = b' = 1$ and $c = 1$. Then $abc\varepsilon = a'b'$ gives $ab \in R^\times$. Since $abg, abh \in R[X]$, also $g, h \in R[X]$. □

Exercise 4.4.16 (General rational root theorem)**.** Let R be a factorial ring. If $f \in R[X]$ is monic and $x \in \mathrm{Quot}(R)$ a root, then $x \in R$.

Theorem 4.4.17. *Let R be a factorial ring and $f \in R[X] \setminus R$. Then f is irreducible in $R[X]$ if and only if f is primitive and irreducible in* $\mathrm{Quot}(R)[X]$.

Proof. $\Rightarrow$: if f is not primitive, then $f = pg$ for some prime $p \in R$ (Exercise 4.3.9 (1)) and $g \in R[X] \setminus R$. Then $p, g \notin R[X]^\times = R^\times$ (Lemma 3.1.6), so f is reducible.

If f is reducible in $\mathrm{Quot}(R)[X]$, then $f = gh$ for certain $g, h \in \mathrm{Quot}(R)[X]$ with positive degree. But then $\hat{g}, \hat{h}$ from the lemma have positive degree, so f is reducible in $R[X]$.

$\Leftarrow$: assume $f = gh$ for $g, h \in R[X] \setminus R[X]^\times$. Then $g, h \notin R$ as f is primitive, so g, h have positive degree, so $g, h \notin \mathrm{Quot}(R)[X]^\times$, so f is reducible in $\mathrm{Quot}(R)[X]$. □

Proof of Gauß' theorem 4.4.1. By Corollary 4.4.8 and Theorem 4.3.2 (3) we are left to show that every irreducible $f \in R[X]$ is prime.

Case $f \in R$. Then f is irreducible in R by Lemma 4.2.15 (3), so prime in R by Theorem 4.3.2, so prime in $R[X]$ by Lemma 4.2.15 (4).

Case $f \notin R$. Then $\deg(f) > 0$, so, by Theorem 4.4.17, f is primitive and irreducible in $\mathrm{Quot}(R)[X]$. Now, $\mathrm{Quot}(R)[X]$ is factorial by Lemma 4.2.11, so, by Theorem 4.3.2, f is prime in $\mathrm{Quot}(R)[X]$. Assume $f \mid gh$ for $g, h \in R[X]$. Then $f \mid g$ or $f \mid h$ in $\mathrm{Quot}(R)[X]$. As f is primitive, Lemma 4.4.15 (2) implies $f \mid g$ or $f \mid h$ in $R[X]$. □

4.5 Eisenstein's irreducibility criterion

No efficient algorithm to test irreducibility of a given polynomial in $\mathbb{Z}[X]$ is known, examples are treated by ad hoc arguments and tricks. In slogan form: irreducibility is an art, not a technique. In this section we learn some tricks. We start generalizing Exercise 3.5.9:

Corollary 4.5.1. *Let R be a factorial ring and $f \in R[X]$ have degree 2 or 3. Then f is irreducible if and only if f is primitive and has no root in* $\mathrm{Quot}(R)$.

If additionally f is monic, then f is irreducible if and only if it has no root in R.

Proof. f has a root $a \in \mathrm{Quot}(R)$ if and only if $(X-a) \mid f$ in $\mathrm{Quot}(R)$ (by Corollary 3.3.2), equivalently f is reducible in $\mathrm{Quot}(R)[X]$. Now apply Theorem 4.4.17. The 2nd statement follows from Exercise 4.4.16. □

Example 4.5.2. $f := X^2 + 3X + 1$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$.

Proof. For $a \in \mathbb{Z}$ with $|a| \geqslant 3$ we have $|f(a)| \geqslant |a|^2 - 3|a| + 1 \geqslant |a|(|a| - 3) + 1 > 0$, so $f(a) \neq 0$. Further, none of $\pm 2, \pm 1, 0$ is a root. □

Example 4.5.3. $f := X^2 + Y^2 + 1$ is irreducible in $\mathbb{Z}[X, Y]$.

Proof. We view $f = X^2 + a \in \mathbb{Z}[Y][X]$ with $a := Y^2 + 1$ and show f has no root in $\mathbb{Z}[Y]$. Otherwise $g^2 = -Y^2 - 1$ for some $g \in \mathbb{Z}[Y]$. Then -1 is the square of the constant term of g, contradiction. □

Exercise 4.5.4. To show $f := X^4 - 2X^3 + X + 1 \in \mathbb{Z}[X]$ is irreducible, first note it has no linear factors, then show $f = (X^2 + aX + b)(X^2 + cX + d)$ for $a, b, c, d \in \mathbb{Z}$ is impossible by comparing coefficients.

Example 4.5.5. $f := X^4 - 10X^2 + 1$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$.

1st proof. In $\mathbb{Z}[X]$, f does not have a linear factor: such a factor would have the form $X - a$ (as f is monic) but f does not have a root in $\mathbb{Z}$. Since f is primitive it suffices to show f is irreducible in $\mathbb{Z}[X]$. Assume otherwise. Writing f as a product of irreducibles thus reads $f = g_0 g_1$ with g_0, g_1 of degree 2; say $g_0 := X^2 + bX + c$. One easily checks $X^2 + c \nmid f$ for all c, so $b \neq 0$. The evaluation homomorphism mapping X to $-X$ is an automorphism of $\mathbb{Z}[X]$. Hence, $g_0(-X) \mid f(-X) = f$. Since $g_0(-X) \neq g_0$, we have $g_0(-X) = g_1 = X^2 - bX + c$. We get $c^2 = 1, b^2 - 2c = 10$. But there are no such $b, c \in \mathbb{Z}$. □

2nd proof. Assume $f = gh$ for $g, h \in \mathbb{Z}[X]$ of degree 2. Note $|f(a)|$ is 1 or a prime in $\mathbb{N}$ for $a = 0, \pm 2, \pm 4, \pm 6, \pm 8$ which are 9 values. For each of them $g(a)$ or $h(a)$ is ± 1. But g, h take each value $\leqslant 2$ times (by interpolation), a contradiction. □

We shall see a third proof in Section 6.2.

Exercise 4.5.6. If $f \in \mathbb{Z}[X]$ and $|f(a)|$ is 1 or a prime in $\mathbb{N}$ for $> 2\deg(f)$ many $a \in \mathbb{Z}$, then f is irreducible in $\mathbb{Z}[X]$.

Theorem 4.5.7 (Eisenstein)**.** *Let R be a factorial ring, $n > 0$ and $f = a_n X^n + \cdots + a_0 \in R[X]$ be primitive with $a_n \neq 0$. Then f is irreducible in both $R[X]$ and $\mathrm{Quot}(R)[X]$ if there exists a prime $p \in R$ such that:*

$$p \mid a_0, \ldots, p \mid a_{n-1},\ p \nmid a_n \text{ and } p^2 \nmid a_0.$$

Proof. Assume there is such a p. By Theorem 4.4.17 it suffices to show f is irreducible in $R[X]$. Assume $f = gh$ for $f, g \in R[X]$, say $g = b_k X^k + \cdots + b_0, h = c_\ell X^\ell + \cdots + c_0$ with $b_k, c_\ell \neq 0$ and $\ell + k = n$. We claim $g \in R^\times$ or $h \in R^\times$.

Since $p \mid a_0 = b_0c_0$ and $p^2 \nmid a_0$, p divides exactly one of b_0, c_0. Say, $p \mid b_0, p \nmid c_0$. As $p \nmid a_n = b_kc_\ell$ there is $0 < i \leqslant k$ such that $p \mid b_0, \ldots, p \mid b_{i-1}, p \nmid b_i$. Then $p \nmid b_ic_0$ and hence

$$p \nmid a_i = b_ic_0 + b_{i-1}c_1 + \cdots + b_0c_i,$$

setting $c_j := 0$ for $j > \ell$. Thus $i = n = k$, so $\ell = 0$ and $h \in R$. As f is primitive, $h \in R^\times$. □

Example 4.5.8. $f := 16X^5 - 9X^4 + 3X^2 + 6X - 21 \in \mathbb{Z}[X]$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$ (Eisenstein with $p = 3$).

Examples 4.5.9. Let $n > 0$ and $p \in \mathbb{N}$ a prime number.

1. $f := X^n - p$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$.
2. $f := X^{p-1} + \cdots + X + 1$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$.
3. $f := X^n + Y^n - 1$ is irreducible in $\mathbb{Z}[X, Y]$.

Proof. (1): by Eisenstein. (2): let φ be the evaluation homomorphism mapping X to $X+1$ (Corollary 3.1.9 (1)). This is an automorphism of $\mathbb{Z}[X]$. It thus suffices to show $\varphi(f)$ is irreducible. Since $f \cdot (X-1) = X^p - 1$, we have $\varphi(f) \cdot X = (X+1)^p - 1$, so

$$\varphi(f) = X^{p-1} + \tbinom{p}{p-1}X^{p-2} + \cdots + \tbinom{p}{1}.$$

Apply Eisenstein: primitive and $\binom{p}{1} = p$ and $p \mid \binom{p}{i} = p \cdot \binom{p-1}{i}$ for all $1 \leqslant i \leqslant p-1$.

(3): view f in $\mathbb{Z}[Y][X]$ and write $f = X^n + a$ with $a := Y^n - 1 \in \mathbb{Z}[Y]$. Then $a = (Y-1)g$ for $g := (Y^{n-1} + \cdots + Y + 1)$. Note $Y - 1$ is prime in $\mathbb{Z}[Y]$ and $Y - 1 \nmid g$ since $g(1) \neq 0$, so $(Y-1)^2 \nmid a$. Since f is primitive, Eisenstein applies. □

Exercise 4.5.10.

1. $X^9 + 10X^5 + 15X^3 + a$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$ for infinitely many $a \in \mathbb{Z}$.
2. $X^3 + 3X + 2$ and $X^4 + 1$ are irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$ (*Hint:* plug $X + 1$ for X).
3. $Y^2 + XY + X$ is irreducible in $\mathbb{Z}[X, Y]$.
4. $Z^2 + Y^2 - XY + X$ is irreducible in $\mathbb{Q}(X, Y)[Z]$.

4.6 Principal ideal domains

Definition 4.6.1. Let R be a commutative ring. $I \subseteq R$ is an *ideal (of R)* if I is a subgroup of $(R, +)$ and $xr \in I$ for all $x \in I, r \in R$. It is *proper* if $I \neq R$. It is *trivial* if $I = \{0\}$ or $I = R$. It is *principal* if $I = xR$ for some $x \in R$.

R is a *principal ideal domain* if R is an integral domain such that all ideals of R are principal, i.e., for every ideal I of R there is $x \in R$ such that $I = xR$.

Remark 4.6.2. Let $R \neq \{0\}$ be a commutative ring.

1. An ideal I of R is proper if and only if $I \cap R^\times = \varnothing$.

   Indeed: if $\varepsilon \in I \cap R^\times$ and $x \in R$, then $x = \varepsilon(\varepsilon^{-1}x) \in I$.
2. R is a field if and only if all ideals are trivial.

   $\Rightarrow$ by (1). $\Leftarrow$: if $0 \neq x \in R$, then $xR \neq \{0\}$, so $xR = R \ni 1$, so $x \in R^\times$.

Examples 4.6.3.

1. $\mathbb{Z}$ is a principal ideal domain by Lemma 2.1.5.
2. $\mathbb{Z}[X]$ is not a principal ideal domain.

   Indeed: let I be the set of polynomials in $\mathbb{Z}[X]$ with an even constant term. This is an ideal but not principal. Otherwise $I = f\mathbb{Z}[X]$ for some $f \in \mathbb{Z}[X]$. Then there are $g, h \in \mathbb{Z}[X]$ with $2 = fg$ and $X = fh$. The former implies $f, g \in \mathbb{Z}$. Then the latter implies $h = yX$ for some $y \in \mathbb{Z}$ with $yf = 1$, so $f = \pm 1$. Then $f\mathbb{Z}[X] = \mathbb{Z}[X]$, so $I = \mathbb{Z}[X]$. But $1 \notin I$, contradiction.
3. For a field K, $K[X]$ is a principal ideal domain (see Example 4.6.10 and Lemma 4.6.9 below) but $K[X, Y]$ is not (exercise).

Exercise 4.6.4. For a field K, $K[[X]]$ is a principal ideal domain.

Lemma 4.6.5. *In a principal ideal domain, irreducibles are prime.*

Proof. Let R be a principal ideal domain, $q, x, y \in R$ and assume $q \mid xy$ and q is irreducible. Consider $J := xR + qR = \{xr + qs \mid r, s \in R\}$. This is an ideal, so principal, say $J = zR$. Then $qR \subseteq J = zR$ and, since q is irreducible, Lemma 4.2.4 (2) gives $qR = zR$ or $zR = R$.

In case $qR = zR$, we have $xR \subseteq qR$, so $q \mid x$ by Remark 4.2.2 (8). In case $J = zR = R$, choose $r_0, s_0 \in R$ with $xr_0 + qs_0 = 1$, and note $q \mid xr_0y + qs_0y = y$ by Remark 4.2.2 (4). □

Theorem 4.6.6. *Principal ideal domains are factorial.*

Proof. By Theorem 4.3.2 (3) and Lemmas 4.4.6, 4.6.5 it suffices to show that a principal ideal domain R does not have proper divisor chains. Assume $(x_n)_n$ is one. Then

$$x_0R \subsetneq x_1R \subsetneq x_2R \subsetneq \cdots,$$

by Remark 4.2.2 (8). Let $I := \bigcup_n x_nR$ and observe I is an ideal. Choose $x \in R$ with $I = xR$. As $x \in I$, there is $n \in \mathbb{N}$ such that $x \in x_nR$. Then $x_{n+1}R \subseteq I = xR \subseteq x_nR$, contradiction. □

Exercise 4.6.7. Principal ideal domains R satisfy Bézout's lemma: for $x_1, \ldots, x_n \in R$ not all zero, there are $r_1, \ldots, r_n \in R$ such that $r_1x_1 + \cdots + r_nx_n$ is a gcd of $x_1, \ldots, x_n$.

4.6.1 Euclidian domains

Which integral domains are principal ideal domains? Beginning number theory, our main tool was Euclidian division:

Definition 4.6.8. R is an *Euclidian domain* if R is an integral domain and there is a *Euclidian valuation*

$$\delta : R \setminus \{0\} \to \mathbb{N},$$

i.e., for all $x, y \in R \setminus \{0\}$ there are $q, r \in R$ such that $x = qy + r$ and, $r = 0$ or $\delta(r) < \delta(x)$.

Lemma 4.6.9. *Euclidian domains are principal ideal domains.*

Proof. Let R be a Euclidian domain with valuation δ, and $I \subseteq R$ an ideal. If $I = \{0\}$, then $I = 0R$ is principal. Assume $I \neq \{0\}$. Choose $x \in I$ with minimal δ-value. We claim $I = xR$.

$\supseteq$ is clear. $\subseteq$: given $y \in I$, write $y = qx + r$ with $q, r \in R$ and, $r = 0$ or $\delta(r) < \delta(x)$; then $r = y - qx \in I$, so $r = 0$ by choice of x; then $y = qx \in xR$. □

Examples 4.6.10.

1. $\mathbb{Z}$ has Euclidian valuation $x \mapsto |x|$ (by Euclidian division).
2. A field K has Euclidian valuation $x \mapsto 0$.
3. $K[X]$ has Euclidian valuation $f \mapsto \deg(f)$ (by Theorem 3.2.1).
4. The Gaussian integers $\mathcal{O}_{-1} = \mathbb{Z} + \mathbb{Z}i$ have Euclidian valuation $\alpha \mapsto N(\alpha)$.

   Indeed: note $N(x+iy) = x^2 + y^2 = |x+iy|^2$ extends to $\mathbb{Q}(i)$. We want for all $\alpha, \beta \in \mathcal{O}_{-1}$ some $q, r \in \mathcal{O}_{-1}$ such that $\alpha = \beta q + r$ with $r = 0$ or $N(r) < N(\beta)$. Working in $\mathbb{Q}(i)$, this means $N(r/\beta) = N(\alpha/\beta - q) < 1$. So we ask for $q \in \mathbb{Z} + \mathbb{Z}i$ with $|\cdot|$-distance < 1 to a given point (namely α/β). The maximal distance to grid points is realized by the midpoints of the squares of the grid. This is $|(1+i)/2| = 1/\sqrt{2} < 1$.
5. Similarly, $\mathcal{O}_{-2} = \mathbb{Z} + \mathbb{Z}\sqrt{2}$ has Euclidian valuation $\alpha \mapsto N(\alpha)$.

More abstract proof of Corollary 4.4.9. By (3) above, $K[X]$ is a Euclidian domain, so a principal ideal domain by Lemma 4.6.9, so factorial by Theorem 4.6.6. □

Remark 4.6.11. It is known that $\alpha \mapsto |N(\alpha)|$ is a Euclidian valuation in $\mathcal{O}_d$ for exactly 21 values of d, namely $-1, -2, -3, -7, -11, 2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 41, 57, 73$. It is conjectured that there are infinitely many $d > 0$ such that $\mathcal{O}_d$ is Euclidian. E.g., $\mathcal{O}_{14}$ is known to be Euclidian but not by its norm. Heegner proved 1952 that there are exactly 9 values of $d < 0$ such that $\mathcal{O}_d$ is a principal ideal domain, namely $-1, -2, -3, -7, -11, -19, -43, -67, -163$. E.g., $\mathcal{O}_{-19}$ is known to be a principal ideal domain that is not Euclidian.

Remark 4.6.12. What if we require (q, r) in Definition 4.6.8 to be unique? It was observed by Jodeit (1967) that then $K := R^\times \cup \{0\}$ is a field and, $R = K$ or $R \cong K[X]$.

Exercise 4.6.13. Give a version of the Euclidian algorithm for a Euclidian ring.

Exercise 4.6.14. Are subrings of Euclidian rings also Euclidian?

Exercise 4.6.15 (Euclidian valuations)**.** Let R be a Euclidian domain.

1. R has a Euclidian valuation δ satisfying $\delta(x) \leqslant \delta(xy)$ for all $x, y \in R \setminus \{0\}$.
   Hint: given an arbitrary valuation δ', set $\delta(x) := \min\{\delta'(xy) \mid y \in R \setminus \{0\}\}$.
2. For all $x \in R \setminus \{0\}, \varepsilon \in R^\times, y \notin R^\times \cup \{0\}$:
$$\delta(1) \leqslant \delta(x), \quad \delta(\varepsilon x) = \delta(x), \quad \delta(x) < \delta(xy).$$
3. $x \in R \setminus \{0\}$ is a unit if and only if $\delta(x) = \delta(1)$.

4.7 Ideals

Let R, S be a commutative rings. Ideals are 'ideal numbers' and we compute with them:

Lemma 4.7.1. *Let $\mathcal{I}$ be the set of ideals of R. Then $(\mathcal{I}, +)$ and $(\mathcal{I}, \cdot)$ are commutative monoids with neutral elements $\{0\}$ and R. Here, for $I, J \in \mathcal{I}$:*

$$\begin{aligned} I + J &:= \big\{x + y \mid x \in I, y \in J\big\}, \\ I \cdot J &:= \big\{x_1 y_1 + \cdots + x_n y_n \mid n \in \mathbb{N}, x_1, \ldots, x_n \in I, y_1, \ldots, y_n \in J\big\}. \end{aligned}$$

Moreover, $I + I = I$ and $(I + I') \cdot J = I \cdot J + I' \cdot J$ for all $I, I', J \in \mathcal{I}$.

Proof. We only verify distributivity: if $z \in (I + I') \cdot J$, then $z = (x_1 + x_1')y_1 + \cdots + (x_n + x_n')y_n$ for some $n \in \mathbb{N}, x_i \in I, x_i' \in I', y_i \in J$. Then $z = (x_1 y_1 + \cdots + x_n y_n) + (x_1' y_1 + \cdots x_n' y_n) \in I \cdot J + I' \cdot J$.
Conversely, $(I + I')J$ contains IJ and $I'J$ and is closed under $+$. □

Remark 4.7.2. Let I, J be ideals of R. Then

$$I \cdot J \subseteq I \cap J \subseteq I \cup J \subseteq I + J.$$

Clearly, $I + J$ is the smallest ideal containing both I and J (every such ideal contains $I + J$). $I \cap J$ is an ideal, and clearly the largest one contained in both I and J. $I \cup J$ is not always an ideal, e.g. in $\mathbb{Z}$, $21 = 6 + 15 \notin 6\mathbb{Z} \cup 15\mathbb{Z}$.

Definition 4.7.3. The ideal *generated by* $X \subseteq R$ is $(X) := \bigcap_{I \text{ ideal}, X \subseteq I} I$.

Remark 4.7.4. Intersections of nonempty sets of ideals are ideals; in particular, (X) is an ideal. It is the smallest ideal that contains X. For $x_1, \ldots, x_n \in R$ we have

$$(x_1, \ldots, x_n) := (\{x_1, \ldots, x_n\}) = x_1 R + \cdots + x_n R.$$

For $x, y \in R$, note $xR = (x)$ and $(\varnothing) = (0) = \{0\}$, and $(x) \cdot (y) = (xy)$. Using distributivity,

$$(x_1, x_2, x_3) \cdot (y) = ((x_1) + (x_2) + (x_3)) \cdot (y) = (x_1 y) + (x_2 y) + (x_3 y) = (x_1 y, x_2 y, x_3 y).$$

Exercise 4.7.5. Let R be a principal ideal domain, $n > 0$ and $x_1, ..., x_n \in R$ not all zero. Show $x \in R$ is a greatest common divisor of $x_1, ..., x_n$ if and only if $(x) = (x_1, ..., x_n)$. Show $x \in R$ is a least common multiple of $x_1, ..., x_n$ if and only if $\bigcap_{i=1}^{n}(x_i) = (x)$.

Example 4.7.6. In $\mathbb{Z}$, $(6) \cdot (15) = (90) \subsetneq (30) = (6) \cap (15) \subsetneq (6) \cup (15) \subsetneq (6) + (15) = (3)$.

Why are the finite sums needed in the definition of the ideal product?

Example 4.7.7. In $\mathbb{Z}[X]$, consider the ideals $I := (2, X), J := (3, X)$ and note

$$I \cdot J = ((2) + (X)) \cdot ((3) + (X)) = (2 \cdot 3) + (2X) + (3X) + (X^2) = (6) + (X).$$

The last equality follows from $I' := (2X) + (3X) + (X^2) = (X)$; indeed, $\subseteq$ as X divides all elements of I', and $\supseteq$ as $X = 2X \cdot (-1) + 3X \cdot 1 + X^2 \cdot 0 \in I'$. Less formally, note I, J and $I \cdot J$ are the ideals of polynomials with constant term divisible by $2, 3$ and 6.

But $X \in I \cdot J$ cannot be written $X = fg$ with $f \in I, g \in J$ because X is irreducible and I, J (are non-trivial, so) do not contain units ± 1.

Exercise 4.7.8. $\{f(X, Y) \in \mathbb{R}[X, Y] \mid f(x, x^2) = 0 \text{ for all } x \in \mathbb{R}\} \subseteq \mathbb{R}[X, Y]$ is the ideal of real polynomials vanishing on the parabola. Show it equals $(Y - X^2)$.

The ideal of real polynomials vanishing on $(a, b) \in \mathbb{R}^2$ is $(X - a, Y - b)$. Verify $(X - 2, Y - 4) \supseteq (Y - X^2)$ algebraically.

Definition 4.7.9. Let I, J be ideals of R.

1. I, J are *coprime* if $I + J = R$.
2. I is *prime* if I is proper and for all $x, y \in R$: $xy \in I$ implies $x \in I$ or $y \in I$.
3. I is *maximal* if I is proper and for every ideal J: $I \subseteq J$ implies $I = J$ or $J = R$.

Exercise 4.7.10. Maximal ideals are prime.

Remark 4.7.11. Assume R is a principal ideal domain and let $x, y, z \in R \setminus \{0\}$.

1. The ideals xR, yR are coprime if and only if x, y are coprime (in R, cf. Exercise 4.3.9).

   $\Rightarrow$: if $xR + yR = R$ choose $r, s \in R$ such that $xr + ys = 1$. Let $d \neq 0$ be a common divisor of x, y. Then $d \mid xr + sy = 1$. Hence, 1 is a greatest common divisor.
   $\Leftarrow$: Exercise 4.6.7 gives $1 \in xR + yR$, so $xR + yR = R$.
2. xR is a prime ideal if and only if x is prime. (Holds in any commutative ring.)

   $\Rightarrow$: if $x \mid yz$, then $yz \in xR$, so $y \in xR$ or $z \in xR$, so $x \mid y$ or $x \mid z$.
   $\Leftarrow$: if $yz \in xR$, then $x \mid yz$, so $x \mid y$ or $x \mid z$, so $y \in xR$ or $z \in xR$.
3. xR is a maximal ideal if and only if x is irreducible (by Lemma 4.2.4).
4. An ideal $\neq \{0\}$ of R is maximal if and only if it is prime (Exercise 4.7.10, Lemma 4.2.9).

Example 4.7.12. The maximal ideals of $\mathbb{Z}$ are $p\mathbb{Z}$ for a prime number p. The prime ideals of $\mathbb{Z}$ are these plus $0\mathbb{Z} = \{0\}$.

Example 4.7.13. Let K be a field. In $K[X,Y]$, the ideal (X) is prime because X is prime in $K[X,Y]$ (Remark 4.7.11 (3)). (X) is not maximal as $(X) \subsetneq (X,Y)$. (X,Y) is maximal: the elements of $K[X,Y] \setminus (X,Y)$ are the polynomials with non-zero constant term; any such f can be written $f = g + a$ with $g \in (X,Y)$ and $a \in K^\times$, so $1 \in (X,Y,f)$.

Example 4.7.14. Let $L \mid K$ be a field extension, and $a \in L$ be algebraic over K. The set I_a of $f \in K[X]$ with $f(a) = 0$ is an ideal. That a is algebraic over K means that $I_a \neq \{0\}$.

By Lemma 3.5.6 ($1 \Rightarrow 3$), $I_a = m_a^K K[X]$. By Lemma 3.5.6 ($1 \Rightarrow 2$) , m_a^K is irreducible, so I_a is maximal.

Conversely, $I_a = gK[X]$ for some $g \in K[X]$ (Example 4.6.3). Then $g \neq 0$, so we can choose g monic. Then $f \in I_a$ if and only if $g \mid f$ and Lemma 3.5.6 ($3 \Rightarrow 1$) shows $g = m_a^K$.

Theorem 4.7.15. *Every proper ideal of R is contained in a maximal one.*

Proof. Let $\mathcal{I}$ be the set of proper ideals, partially ordered by $\subsetneq$. Let $\mathcal{C}$ be a chain. Then $I := \bigcup \mathcal{C} \in \mathcal{I}$. Indeed, I is clearly an ideal and it is proper: otherwise $1 \in I$, so $1 \in J$ for some $J \in \mathcal{C}$, so $J = R \notin \mathcal{I}$, contradiction. Hence, the partial order is inductive. By Zorn's lemma it contains a maximal element. This is a maximal ideal. □

Remark 4.7.16. Let $\varphi : R \to S$ be a ring homomorphism.

1. If J is an ideal of S, then $\varphi^{-1}(J) = \{x \in R \mid \varphi(x) \in J\}$ is an ideal of R that contains $\ker(\varphi)$. In particular, $\ker(\varphi) = \varphi^{-1}(\{0\})$ is an ideal of R.

   Indeed: $\varphi^{-1}(J) \supseteq \ker(\varphi)$ because $0 \in J$. We show $\varphi^{-1}(J)$ is an ideal. Let $x, y \in \varphi^{-1}(J)$ and $r \in R$; then $\varphi(x-y) = \varphi(x) - \varphi(y) \in J$, so $x - y \in \varphi^{-1}(J)$; further, $\varphi(xr) = \varphi(x)\varphi(r) \in J$, so $xr \in \varphi^{-1}(J)$.
2. If J is a prime ideal of S, then $\varphi^{-1}(J)$ is a prime ideal of R.

   Indeed: $I := \varphi^{-1}(J)$ is an ideal by (1) and proper as $1 \notin I$ since $\varphi(1) = 1 \notin J$. If $xy \in I$, then $\varphi(xy) = \varphi(x)\varphi(y) \in J$. As J is prime, $\varphi(x) \in J$ or $\varphi(y) \in J$, so $x \in I$ or $y \in I$.
3. If S is an integral domain, then $\ker(\varphi)$ is a prime ideal of R.

   Indeed: if $xy \in \ker(\varphi)$, then $\varphi(x)\varphi(y) = 0$. Since S is an integral domain, $\varphi(x) = 0$ or $\varphi(y) = 0$, i.e., $x \in \ker(\varphi)$ or $y \in \ker(\varphi)$.
4. If φ is surjective and I is an ideal of R, then $\varphi(I)$ is an ideal of S.

   Indeed: let $y, y' \in \varphi(I)$, say $\varphi(x) = y, \varphi(x) = y'$ for $x, x' \in I$; then $y - y' = \varphi(x) - \varphi(x') = \varphi(x - x') \in \varphi(I)$. Given $s \in S$, choose $r \in R$ with $\varphi(r) = s$ by surjectivity; then $ys = \varphi(x)\varphi(r) = \varphi(xr) \in \varphi(I)$.
5. In (4) surjectivity cannot be omitted: e.g., the identity $\mathrm{id}_{\mathbb{Z}}$ is a ring monomorphism from $\mathbb{Z}$ into $\mathbb{Q}$ and $\mathrm{id}_{\mathbb{Z}}(2\mathbb{Z}) = 2\mathbb{Z}$ is not an ideal of $\mathbb{Q}$.

Proposition 4.7.17 (Ideal correspondence)**.** *Let $\varphi : R \to S$ be a ring epimorphism. Then $I \mapsto \varphi(I)$ is a bijection from the set of ideals I of R with $\ker(\varphi) \subseteq I$ onto the set of all ideals of S; its inverse is $J \mapsto \varphi^{-1}(J)$ for ideals J of S.*

Proof. We show $I^* := \varphi^{-1}(\varphi(I)) = I$ for every ideal I of R with $\ker(\varphi) \subseteq I$. $\supseteq$ is clear. $\subseteq$: let $x^* \in I^*$, so $\varphi(x^*) \in \varphi(I)$, so $\varphi(x^*) = \varphi(x)$ for some $x \in I$. Then $\varphi(x^* - x) = 0$, so $x^* - x \in \ker(\varphi) \subseteq I$. Then $x^* = (x^* - x) + x \in I$.

We show $J^* := \varphi(\varphi^{-1}(J)) = J$ for every ideal J of S. $\subseteq$ is clear. $\supseteq$: let $y \in J$; by surjectivity, there is $x \in R$ such that $\varphi(x) = y$; then $x \in \varphi^{-1}(J)$, so $y \in J^*$. □

4.7.1 Noetherian rings

We saw $\mathbb{Z}$ and $K[X]$ are principal ideal domains, but $\mathbb{Z}[X]$ and $K[X][Y]$ are not. A weaker property is preserved moving to the polynomial rings:

Definition 4.7.18. R is *noetherian* if every ideal I of R is *finitely generated*, i.e., $I = (X)$ for some finite $X \subseteq R$.

Proposition 4.7.19. *The following are equivalent.*

1. *R is noetherian.*
2. Ascending chain condition: *if $I_0 \subseteq I_1 \subseteq \cdots$ are ideals of R, then there is $n \in \mathbb{N}$ such hat $I_n = I_m$ for all $m \geqslant n$.*
3. Noetherian recursion: *every nonempty set $\mathcal{I}$ of ideals of R contains a $\subseteq$-maximal element $I \in \mathcal{I}$, i.e., for all $J \in \mathcal{I}$: $I \subseteq J$ implies $I = J$.*

Proof. $1 \Rightarrow 2$: given $I_0 \subseteq I_1 \subseteq \cdots$, set $I := \bigcup_n I_n$ and note this is an ideal. Choose $r > 0$ and $x_1, ..., x_r \in R$ such that $I = (x_1, ..., x_r)$. Choose $n \in \mathbb{N}$ such that $x_1, ..., x_r \in I_n$. Then $I_n \subseteq I_m \subseteq I \subseteq I_n$, so $I_n = I_m$ for all $m \geqslant n$.

$2 \Rightarrow 3$: assume $\mathcal{I} \neq \varnothing$ does not have a maximal element. Choose $I_0 \in \mathcal{I}$. As I_0 is not maximal, there is $I_1 \in \mathcal{I}$ with $I_0 \subsetneq I_1$. Continuing gives a chain violating (2).

$3 \Rightarrow 1$: given an ideal I of R let $\mathcal{I}$ be the set of ideals (X) for $X \subseteq I$ finite. Then a maximal $(X) \in \mathcal{I}$ equals I: if $x \in I \setminus (X)$, then $(X) \subsetneq (X \cup \{x\}) \in \mathcal{I}$ is not maximal. □

Exercise 4.7.20. The ascending chain condition restricted to principal ideals is equivalent to not having proper divisor chains.

Theorem 4.7.21 (Hilbert's basis theorem)**.** *If R is noetherian, then so is $R[X]$.*

Proof. Assume I is an ideal of $R[X]$ that is not finitely generated. Clearly, $I \neq \{0\}$. Choose $f_0 \in I \setminus \{0\}$ of minimal degree n_0. As $I \neq (f_0)$, choose $f_1 \in I \setminus (f_0)$ of minimal degree n_1. As $I \neq (f_0, f_1)$, choose $f_2 \in I \setminus (f_0, f_1)$ of minimal degree n_2. And so on.

Then $n_0 \leqslant n_1 \leqslant n_2 \leqslant \cdots$. Let a_n be the lead coefficient of f_n. Then $(a_0) \subseteq (a_0, a_1) \subseteq \cdots$ is a chain of ideals in R. As R is noetherian, $(a_0, ..., a_k) = (a_0, ..., a_{k+1})$ for some $k \in \mathbb{N}$. Then $a_{k+1} = r_0 a_0 + \cdots + r_k a_k$ for certain $r_i \in R$. Set

$$g := \textstyle\sum_{i \leqslant k} r_i X^{n_{k+1} - n_i} f_i.$$

Then $g \in (f_0, ..., f_k)$ has degree n_{k+1} and lead coefficient a_{k+1}. Then $f_{k+1} - g \in I \setminus (f_0, ..., f_k)$. But $f_{k+1} - g$ has degree $< n_{k+1}$, in contradiction to the choice of f_{k+1}. □

Corollary 4.7.22. *If K is a field and $n > 0$, then $K[X_1, ..., X_n]$ is noetherian.*

4.8 Residue class rings

We defined $\mathbb{Z}_n$ from $\mathbb{Z}$ by identifying two integers that differ only by multiples of n, i.e., an element of the principal ideal $n\mathbb{Z}$. In the notation below this becomes $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/(n)$. In Exercise 3.2.11 we defined $K[X]/(g)$ by identifying two polynomials that differ only by multiples of g, i.e., an element of the principal ideal (g).

We now give a general definition. Let R, S be commutative rings.

Definition 4.8.1. Let I be an ideal of R. Call $x, y \in R$ *congruent modulo* I, symbolically $x \equiv y \mod I$, if $x - y \in I$. The equivalence class of x is $x + I := \{x + y \mid y \in I\}$.

The set of equivalence classes is R/I. Define $+, \cdot$ on R/I setting

$$(x + I) + (y + I) := (x + y) + I, \quad (x + I) \cdot (y + I) := xy + I,$$

for all $x, y \in R$. For $X \subseteq R$ we let $X/I := \{x + I \mid x \in X\}$.

Example 4.8.2. The set C of rational Cauchy sequences are a ring with componentwise addition and multiplication. Those with limit 0 form an ideal N. We defined $\mathbb{R}$ as C/N.

Theorem 4.8.3. *Let I be an ideal in R. Then $(R/I, +, \cdot)$ is a commutative ring, the* residue class ring modulo I. *The* canonical projection π_I *given by*

$$\pi_I(x) := x + I$$

for $x \in R$ is a ring epimorphism from R onto R/I with kernel I.

Proof. It is easy to check that congruence is an equivalence relation with classes $x + I$. We show $+, \cdot$ are well-defined: assume $u := x - x', v := y - y' \in I$. Then $(x+y)+I = (u+x'+v+y')+I = (x + y) + I$ as $u + v \in I$, and $xy + I = (u + x')(v + y') + I = x'y' + (uv + uy' + vx') + I = x'y' + I$ as $(uv + uy' + vx') \in I$.

It is easy to see that $+, \cdot$ are associative and commutative with neutral elements $0 + I = I$ and $1 + I$. The additive inverse of $x + I$ is $(-x) + I$. Distributivity is also clear.

The canonical projection is obviously surjective, preserves 1 and also $+, \cdot$ by definition of $+, \cdot$ in R/I. For the kernel, note $x + I = 0 + I$ if and only if $x \in I$. □

Exercise 4.8.4. Let $x \in R$ and I be an ideal of R. Then $x + I \in (R/I)^\times$ if and only if xR is coprime to I.

Ideal correspondence for the epimorphism $\pi_I : R \to R/I$ gives:

Corollary 4.8.5. *Let I be an ideal in R. The ideals of R/I are exactly the sets J/I for J an ideal of R with $I \subseteq J$.*

Exercise 4.8.6. Iterating factoring does not yield anything new: let I, J be ideals of R with $I \subseteq J$. Then $(R/I)/(J/I) \cong R/J$ via $(x + I) + J/I \mapsto x + J$.

Exercise 4.8.7. Let I, J be ideals of R. Then the map $x + (I \cap J) \mapsto x + J$ is a bijection from $I/(I \cap J)$ onto $(I + J)/J$ that preserves $+$ and $\cdot$.

Exercise 4.8.8. Show $\mathbb{Z}_{10}/([5]_{10}) \cong \mathbb{Z}_5$. For ideals of $\mathbb{Z}$ show $(3)/(15)$ is isomorphic to $((3)+(5))/(5)$ and $\mathbb{Z}_5$.

Lemma 4.8.9. *Let I be an ideal of R.*

1. *I is a prime ideal if and only if R/I is an integral domain.*
2. *I is maximal if and only if R/I is a field.*

Proof. (1) $\Rightarrow$: since I is proper, R/I satisfies $0 \neq 1$. If $(x+I)(y+I) = 0+I$, then $xy \in I$, so $x \in I$ or $y \in I$, so $x+I = I$ or $y+I = I$.

(1) $\Leftarrow$: since R/I satisfies $0 \neq 1$, I is proper. If $xy \in I$, then $I = xy + I = (x+I)(y+I)$, so $x+I = I$ or $y+I = I$, i.e., $x \in I$ or $y \in I$.

(2) $\Rightarrow$: as I is proper, R/I satisfies $0 \neq 1$. If R/I is not a field, it has a non-trivial ideal by Remark 4.6.2. By Corollary 4.8.5 we can write it as J/I for some ideal J of R with $I \subseteq J$. Being nontrivial means $I \subsetneq J \subsetneq R$. Hence, I is not maximal.

(2) $\Leftarrow$: as R/I satisfies $0 \neq 1$, I is proper. If I is not maximal, there is an ideal $I \subsetneq J \subsetneq R$. By Corollary 4.8.5, J/I is a nontrivial ideal of R/I, so R/I is not a field. □

More abstract proof of Lemma 4.6.5. If $x \in R$ is irreducible, then xR is maximal by Remark 4.7.11 (3). By (2) above, R/xR is a field, hence an integral domain, so xR is a prime ideal by (3) above, so x is prime by Remark 4.7.11 (2). □

Remark 4.8.10. Proposition 2.5.6 observed $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ is a field if and only if p is a prime number. By Example 4.7.12, these $p\mathbb{Z}$ are precisely the maximal ideals of $\mathbb{Z}$.

Theorem 4.8.11 (Universal property)**.** *Let I be an ideal in R and $\varphi : R \to S$ a ring homomorphism with $I \subseteq \ker(\varphi)$. There is a unique ring homomorphism $\bar{\varphi} : R/I \to S$ with*

$$\varphi = \bar{\varphi} \circ \pi_I.$$

Proof. $\varphi = \bar{\varphi} \circ \pi_I$ forces the definition $\bar{\varphi}(x+I) := \varphi(x)$, so uniqueness is clear. Well-defined: if $x+I = x'+I$, then $x - x' \in I \subseteq \ker(\varphi)$, so $0 = \varphi(x-x') = \varphi(x) - \varphi(x')$, so $\varphi(x) = \varphi(x')$.

It is clear that $\bar{\varphi}$ is a homomorphism. E.g., it preserves $\cdot$:

$$\bar{\varphi}((x+I)\cdot(y+I)) = \bar{\varphi}(xy+I) = \varphi(xy) = \varphi(x)\varphi(y) = \bar{\varphi}(x+I)\cdot\bar{\varphi}(y+I). \qquad \square$$

Corollary 4.8.12 (Isomorphism theorem for rings)**.** *Let $\varphi : R \to S$ be a ring epimorphism. Then $\bar{\varphi} : R/\ker(\varphi) \cong S$.*

Proof. Write $I := \ker(\varphi)$. $\bar{\varphi}$ is surjective: if $y \in S$, say $\varphi(x) = y$, then $\bar{\varphi}(x+I) = y$. Injective: $\bar{\varphi}(x+I) = 0$ implies $\varphi(x) = \bar{\varphi}(\pi_I(x)) = 0$, so $x \in I$, so $x+I = I$ is the 0 of R/I. □

Exercise 4.8.13. Recall the ring $C(\mathbb{R})$ from Example 1.1.17 (5). For $a \in \mathbb{R}$ show that $I_a := \{f \in C(\mathbb{R}) \mid f(a) = 0\}$ is an ideal of $C(\mathbb{R})$ and $C(\mathbb{R})/I_a \cong \mathbb{R}$.

Example 4.8.14. Let K be a field, $n > 0$ and $a_1, \ldots, a_n \in K$. Then $(X_1 - a_1, \ldots, X_n - a_n)$ is a maximal ideal of $K[X_1, \ldots, X_n]$.

Proof. Let $\varphi_{\bar{a}}$ be the evaluation homomorphism (cf. Corollary 3.6.7 (2)) that maps X_i to a_i. It is onto K. Then $K[X_1,...,X_n]/\ker(\varphi_{\bar{a}}) \cong K$ is a field, so $\ker(\varphi_{\bar{a}})$ is a maximal ideal by Lemma 4.8.9 (2). But $\ker(\varphi_{\bar{a}}) = (X_1 - a_1, ..., X_n - a_n)$. This has been shown "by hand" in Exercise 3.6.12. More abstractly we can now argue as follows: let I denote the r.h.s. ideal and argue: $X_i \equiv a_i \mod I$ for all i, so for any $f(\bar{X}) \in K[\bar{X}]$ we have $f(\bar{X}) \equiv f(\bar{a}) \mod I$; hence, $f(\bar{a}) = 0$ if and only if $f(\bar{X}) \equiv 0 \mod I$, i.e., $f(\bar{X}) \in I$. □

Corollary 4.8.15. *Let $L \mid K$ be a field extension and $a \in L$. Then*

1. *If a is algebraic over K, then $K(a) = K[a] \cong K[X]/(m_a^K)$ via an isomorphism that maps $x \in K$ to $x + (m_a^K)$ and a to $X + (m_a^K)$.*
2. *If a is transcendent over K, then $K(a) \cong K(X)$ via an isomorphism that fixes each $x \in K$ (maps it to itself) and maps a to $X + (m_a^K)$*

Proof. (1): $=$ by Theorem 3.5.8 and $\cong$ by $\ker(\varphi_a) = I_a = (m_a^K)$ as noted in Example 4.7.14.
(2): $\ker(\varphi_a) = \{0\}$, so $K[a] \cong K[X]$, so $K(a) \cong K(X)$. □

Remark 4.8.16.

1. Recall Exercise 3.2.11 showed $K[a] \cong K[X]/(X - a)$ (as defined there). The advantage of the description $K[X]/(m_a^K)$ is that it uses only data from K.
2. In Theorem 3.5.8 we showed, assuming a is algebraic over K, that $K[a]$ is a field using the Euclidian algorithm for polynomials. More abstractly we can now argue as follows: m_a^K is irreducible (Lemma 3.5.6), so (m_a^K) is a maximal ideal in $K[X]$, so $K[X]/(m_a^K)$ is a field, hence the isomorphic $K[a]$ is a field.

Examples 4.8.17. $\mathbb{R}[X]/(X^2+1) \cong \mathbb{R}[i] = \mathbb{C}$. Similarly, for a quadratic number field, $\mathbb{Q}(\sqrt{d}) \cong \mathbb{Q}[X]/(X^2 - d)$. By Lindemann, $\mathbb{Q}(\pi) \cong \mathbb{Q}(X)$.

Theorem 4.8.18 (General Chinese remainder theorem)**.** *Let $n > 1$ and $I_1, ..., I_n$ pairwise coprime ideals of R. Then*

$$R/I_1 \cdots I_n \cong R/I_1 \times \cdots \times R/I_n.$$

Proof. We first show for every $1 \leqslant i \leqslant n$ that I_i is coprime to $I_i^* := \bigcap_{i \neq j} I_j$. For each $j \neq i$ choose $x_j \in I_i$ and $y_j \in I_j$ such that $1 = x_j + y_j$. Then $1 = \prod_{i \neq j}(x_j + y_j) = \prod_{i \neq j} y_j + z$ for some $z \in I_i$ and $\prod_{i \neq j} y_j \in I_i^*$. Hence, $1 \in I_i^* + I_i$.

Second we show $\bigcap_{i=1}^k I_i = \prod_{i=1}^k I_i$ for all $2 \leqslant k \leqslant n$. $k = 2$: $\supseteq$ is clear; $\subseteq$: let $z \in I_1 \cap I_2$; since I_1, I_2 are coprime, there are $x \in I_1, y \in I_2$ such that $x + y = 1$; then $z = zx + zy \in I_1 \cdot I_2$.

Assuming inductively $J := \bigcap_{i=1}^k I_i = \prod_{i=1}^k I_i$, we show $J \cap I_{k+1} = J \cdot I_{k+1}$. This follows as in the case $k = 2$ since J, I_{k+1} are coprime ($J \supseteq I_{k+1}^*$ and I_{k+1}, I_{k+1}^* are coprime).

The map $x \mapsto (x + I_1, ..., x + I_n)$ is clearly a homomorphism from R to the product. It has kernel $\bigcap_{i=1}^n I_i = I_1 \cdots I_n$. By the isomorphism theorem we are left to show surjectivity.

Given $(z_1, ..., z_n) \in R^n$ we want $z \in R$ with $z \equiv z_i \mod I_i$ for all i. For every i choose $x_i \in I_i, y_i \in I_i^*$ with $x_i + y_i = 1$. Then $y_i \equiv 1$ and $y_j \equiv 0 \mod I_i$ for every $j \neq i$. Thus,

$$z := \textstyle\sum_{j=1}^n z_j y_j \equiv z_i y_i \equiv z_i \mod I_i.$$

□

4.8.1 An irreducibility criterion

Exercise 4.8.19 (Generalized Eisenstein criterion)**.** Let R be a factorial ring, $n > 0$ and $f = a_n X^n + \cdots + a_0 \in R[X]$ be primitive with $a_n \neq 0$. Let P be a prime ideal of R such that $a_n \notin P$, $a_i \in P$ for all $i \neq n$, $a_0 \notin P \cdot P$. Then f is irreducible in $R[X]$ and $\mathrm{Quot}(R)[X]$.

Recall the notation from Remark 3.1.8.

Theorem 4.8.20. *Let R be a factorial ring, and $f \in R[X] \setminus R$ primitive with lead coefficient $a \in R$. Let P be a prime ideal of R with $a \notin P$.*

If $\pi_P(f)$ is irreducible in $(R/P)[X]$, then f is irreducible in $R[X]$ and $\mathrm{Quot}(R)[X]$.

Proof. Write $\bar{x} := \pi_P(x), \bar{g} := \pi_P(g)$ for $x \in R, g \in R[X]$. By Theorem 4.4.17 it suffices to show f is irreducible in $R[X]$. Otherwise $f = gh$ for non-units $g, h \in R[X]$. As f is primitive, $g, h \notin R$, so have positive degree.

We have $\bar{f} = \bar{g}\bar{h}$ and show $\bar{g}, \bar{h}$ are not units. But $(R/P)[X]^\times = (R/P)^\times$ by Lemma 3.1.6 because R/P is an integral domain by Lemma 4.8.9. It thus suffices to show $\bar{g}, \bar{h}$ have positive degree. Then $0 \neq \bar{a} = \bar{b}\bar{c}$ in R/P where $b, c \in R$ are the lead coefficients of g, h. Then both $\bar{b}, \bar{c} \neq 0$, so $\bar{g}, \bar{h}$ have the same degrees as g, h. □

Let us spell out what this means for $R = \mathbb{Z}$. Recall, the nontrivial prime ideals of $\mathbb{Z}$ are $p\mathbb{Z}$ for prime $p \in \mathbb{N}$ (Example 4.7.12) and $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$.

Corollary 4.8.21. *Let $f \in \mathbb{Z}[X] \setminus \mathbb{Z}$ be primitive with lead coefficient a and $p \in \mathbb{N}$ prime with $p \nmid a$. Let $\bar{f} \in \mathbb{F}_p[X]$ be obtained from f by replacing each coefficient b of f by $\bar{b} := [b]_p \in \mathbb{F}_p$.*

If $\bar{f}$ is irreducible in $\mathbb{F}_p[X]$, then f is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$.

Example 4.8.22. $f := 5X^3 + 6X^2 - 7X + 3$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$. Note Eisenstein's criterion does not apply.

Proof. f is primitive and for $p = 2$ we have $\bar{f} = X^3 - X + \bar{1} \in \mathbb{F}_2[X]$ is irreducible in $\mathbb{F}_2[X]$: it has degree 3 and no root in $\mathbb{F}_2$. □

Example 4.8.23. The converse fails: $f := X^2 + 1$ is primitive and irreducible in $\mathbb{Z}[X]$ but for $p = 2$ we have $\bar{f} = X^2 + \bar{1} = (X + \bar{1})^2$ in $\mathbb{F}_2[X]$.

Example 4.8.24. The converse fails *badly*: $f := X^4 + 1$ is primitive and irreducible in $\mathbb{Z}[X]$ (Exercise 4.5.10) but $\bar{f}$ is reducible in $\mathbb{F}_p[X]$ for *every* prime $p \in \mathbb{N}$.

Proof. If $-\bar{1}$ is a square in $\mathbb{F}_p$ (e.g., for $p = 2$), say $\bar{a}^2 = -\bar{1}$, then $X^4 + \bar{1} = (X^2 + \bar{a})(X^2 - \bar{a})$ is reducible in $\mathbb{F}_p[X]$. If $p > 2$ and $\bar{2}$ is a square in $\mathbb{F}_p$, say $\bar{2} = \bar{b}^2$, then factor

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

Assume $p > 2$ and $-\bar{1}, \bar{2}$ are not squares in $\mathbb{F}_p$. Then $-\bar{2}$ is a square in $\mathbb{F}_p$ by Corollary 2.8.6 (1). More directly: let x be a primitive root of p and write $-\bar{1} = \bar{x}^r, \bar{2} = \bar{x}^s$ for $r, s \in \mathbb{N}$; then r, s are odd, so $-\bar{2} = \bar{c}^2$ for $\bar{c} := \bar{x}^{(r+s)/2}$ in $\mathbb{F}_p$. Then factor

$$X^4 + \bar{1} = (X^2 - \bar{c}X - \bar{1})(X^2 + \bar{c}X - \bar{1}).$$ □

Exercise 4.8.25. For $R = \mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, consider $f := X^3 + X^2 + X + \underline{2} \in R[X]$ (where $\underline{2} := 1_R + 1_R$). For which of these R is f irreducible?

Chapter 5

Group theory

5.1 Isometries

We recall some linear algebra. For $n > 0$ consider the vector space $\mathbb{R}^n$ with the *standard basis* $e_1, ..., e_n$ where $e_i = (0, ..., 0, 1, 0, ..., 0)$ with 1 at the i-th component. The *inner product* is $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ where $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n$. If $\langle x, y\rangle = 0$, then x, y are *orthogonal*. The *(Euclidian) norm* or *length of* x is $\|x\| := \sqrt{\langle x, x\rangle}$.

For a matrix $A \in \mathbb{R}^{n\times n}$ we have $\langle Ax, y\rangle = \langle x, A^\top y\rangle$ where $\cdot^\top$ denotes matrix transpose and x, y are viewed as column vectors. The matrix $A \in \mathbb{R}^{n\times n}$ *of* a linear map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ has i-th column $\varphi(e_i)$; then $\varphi(x) = Ax$. We call φ *orthogonal* if its matrix A is *orthogonal*, i.e., $A^\top A = I_n$ where $I_n \in \mathbb{R}^n$ is the matrix of the identity $\mathrm{id}_{\mathbb{R}^n}$.

Exercise 5.1.1. For all $x, y \in \mathbb{R}^n$: if $\langle x, y\rangle = \langle x, x\rangle = \langle y, y\rangle$, then $x = y$.

Remark 5.1.2. Let $n > 0$.

1. Every orthogonal $A \in \mathbb{R}^{n\times n}$ has determinant $\det(A) = \pm 1$.

   Indeed: $1 = \det(I_n) = \det(A^\top A) = \det(A)\det(A^\top) = \det(A)^2$.
2. The set of orthogonal matrices $\mathrm{O}(n, \mathbb{R})$ is a subgroup of the general linear group $\mathrm{GL}(n, \mathbb{R})$, namely the *orthogonal group*.

   Indeed: $\mathrm{O}(n, \mathbb{R}) \subseteq \mathrm{GL}(n, \mathbb{R})$ is clear, and if $A, B \in \mathrm{O}(n, \mathbb{R})$, then $AB^{-1} \in \mathrm{O}(n, \mathbb{R})$:

   $$AB^{-1}(AB^{-1})^\top = AB^\top(AB^\top)^\top = AB^\top B^{\top\top}A^\top = AB^{-1}BA^\top = AA^\top = I_n.$$
3. The $A \in \mathrm{O}(n, \mathbb{R})$ with $\det(A) = 1$ form a subgroup of $\mathrm{O}(n, \mathbb{R})$, the *special orthogonal group* $\mathrm{SO}(n, \mathbb{R})$. Indeed: if $A, B \in \mathrm{SO}(n, \mathbb{R})$, then $\det(AB^{-1}) = \det(A)/\det(B) = 1$.

Orthogonal matrices are important for group theory – in the next section we prove:

Theorem 5.1.3. *Every finite group of order $n \in \mathbb{N}$ is isomorphic to a subgroup of* $\mathrm{O}(n, \mathbb{R})$.

Example 5.1.4 (Orthogonal group of the plane)**.** Basic linear algebra shows that

$$R_\alpha := \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \quad \text{and} \quad S_\alpha := \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{bmatrix}$$

for $\alpha \in \mathbb{R}$ have determinants 1 and -1, respectively, and every $A \in \mathrm{O}(2,\mathbb{R})$ has one of these forms. R_α is the matrix of a rotation $\rho_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$, counterclockwise with angle α. S_α is the matrix of a reflection $\sigma_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ of the plane about the line with angle $\alpha/2$ with the e_1-axis. In particular,

$$S_0 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

describes the reflection σ_0 about the e_1-axis and $S_\alpha = R_\alpha S_0$, i.e., $\sigma_\alpha = \rho_\alpha \circ \sigma_0$.

Proposition 5.1.5. *Let $n > 0$. A linear map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal if and only if it* preserves lengths*, i.e., $\|\varphi(x)\| = \|x\|$ for all $x \in \mathbb{R}^n$.*

Proof. Let A be the matrix of φ. $\Rightarrow$: if A is orthogonal, then $\|\varphi(x)\|^2 = \langle Ax, Ax\rangle = \langle x, A^\top Ax\rangle = \langle x, I_n x\rangle = \|x\|^2$. $\Leftarrow$: assume $\|Ax\| = \|x\|$ for all $x \in \mathbb{R}^n$. Then $\langle Ax, Ay\rangle = \langle x, y\rangle$ for all $x, y \in \mathbb{R}^n$. Indeed: by $\|x+y\| = \|A(x+y)\|$ we have $\|x^2\| + 2\langle x, y\rangle + \|y\|^2 = \langle x+y, x+y\rangle = \langle Ax + Ay, Ax + Ay\rangle = \|Ax\|^2 + 2\langle Ax, Ay\rangle + \|Ay\|^2 = \|x\|^2 + 2\langle Ax, Ay\rangle + \|y\|^2$.

Thus, $\langle Ax, Ay\rangle = x^\top A^\top Ay = x^\top y$, so $x^\top(AA^\top - I_n)y = 0$ for all $x, y \in \mathbb{R}^n$. This implies $A^\top A = I_n$, i.e., A is orthogonal. Indeed: if $B \in \mathbb{R}^{n\times n}$ satisfies $x^\top By = 0$ for all $x, y \in \mathbb{R}^n$, then all entries of B are 0 (the ij-th entry is $e_i^\top B e_j$). □

Definition 5.1.6. Let $n > 0$. A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is an *isometry of* $\mathbb{R}^n$ if it is *distance preserving*: $\|f(x) - f(y)\| = \|x - y\|$ for all $x, y \in \mathbb{R}^n$. The set of isometries of $\mathbb{R}^n$ is $\mathrm{I}(n,\mathbb{R})$.

Example 5.1.7. Orthogonal linear maps φ are isometries: $\|\varphi(x) - \varphi(y)\| = \|\varphi(x - y)\| = \|x - y\|$. For $a \in \mathbb{R}^n$ the *translation*

$$t_a(x) := x + a$$

is an isometry which is not linear (unless $a = 0$). Note $\varphi \circ t_a = t_{\varphi(a)} \circ \varphi$.

Lemma 5.1.8. *Let $n > 0$ and $f : \mathbb{R}^n \to \mathbb{R}^n$. The following are equivalent.*

1. *f is an isometry with $f(0) = 0$.*
2. *f* preserves the inner product*, i.e., $\langle f(x), f(y)\rangle = \langle x, y\rangle$ for all $x, y \in \mathbb{R}^n$.*
3. *f is an orthogonal linear map.*

Proof. $3 \Rightarrow 1$ is noted above. Write $x' := f(x)$. $1 \Rightarrow 2$: let f accord (1). Then $\langle x' - y', x' - y'\rangle = \langle x - y, x - y\rangle$. For $y = 0$, noting $0' = 0$, we get $\langle x', x'\rangle = \langle x, x\rangle$. Similarly, $\langle y', y'\rangle = \langle y, y\rangle$. Then $\langle x', x'\rangle - 2\langle x', y'\rangle + \langle y', y'\rangle = \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle$ implies $\langle x', y'\rangle = \langle x, y\rangle$.

$2 \Rightarrow 3$: let f accord (2). By Proposition 5.1.5, it suffices to show f is linear, i.e., $x' + y' = z'$ where $z := x + y$ and $(rx)' = rx'$ for $r \in \mathbb{R}$. The latter being analogous, we show the former. By Exercise 5.1.1, it suffices to show $\langle z', z'\rangle = \langle x' + y', x' + y'\rangle = \langle z', x' + y'\rangle$, i.e.,

$$\langle z', z'\rangle = \langle x', x'\rangle + 2\langle x', y'\rangle + \langle y', y'\rangle = \langle z', x'\rangle + \langle z', y'\rangle.$$

We drop primes using assumption (2), so this is true by $z = x + y$. □

Theorem 5.1.9. *Let $n > 0$. For every isometry f of $\mathbb{R}^n$ there is a unique orthogonal linear map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ such that $f = t_a \circ \varphi$ for $a := f(0)$.*

Proof. We have to show that $\varphi := t_{-a} \circ f$ is a linear orthogonal map. But as a composition of isometries, it is an isometry and clearly $\varphi(0) = 0$. Apply the lemma. □

Remark 5.1.10. Here is how isometries compose. Let $a, b \in \mathbb{R}^n$ and φ, ψ be orthogonal linear maps. Then $(t_a \circ \varphi) \circ (t_b \circ \psi) = t_a \circ t_{\varphi(b)} \circ \varphi \circ \psi = t_{a+\varphi(b)} \circ (\varphi \circ \psi)$.

Corollary 5.1.11. *Let $n > 0$. $\mathrm{I}(n, \mathbb{R})$ is a group (with composition $\circ$).*

Proof. The composition of isometries is an isometry and the neutral element is $\mathrm{id}_{\mathbb{R}^n}$. The inverse of $t_a \circ \varphi$ is $\varphi^{-1} \circ t_{-a}$ – an isometry. □

Example 5.1.12 (Plane isometries)**.** Every isometry of $\mathbb{R}^2$ is of the form $t_a \circ \rho_\alpha$ or $t_a \circ \rho_\alpha \circ \sigma_0$ where $\alpha \in \mathbb{R}$ and $a \in \mathbb{R}^2$. Compositions are computed by the rules (omitting $\circ$):

$$\rho_\alpha t_a = t_{\rho_\alpha(a)} \rho_\alpha, \quad \sigma_0 t_a = t_{\sigma_0(a)} \sigma_0, \quad \sigma_0 \rho_\alpha = \rho_{-\alpha} \sigma_0, \quad t_a t_b = t_{a+b}, \quad \rho_\alpha \rho_\beta = \rho_{\alpha+\beta}.$$

E.g., $(t_{(1,1)} \rho_{\pi/2}) \circ (t_{(1,0)} \pi_{\pi/2} \sigma_0) = t_{(1,1)} t_{(0,1)} \rho_{\pi/2} \rho_{\pi/2} \sigma_0 = t_{(1,2)} \rho_\pi \sigma_0$ is $(x, y) \mapsto (-x + 1, y + 2)$.

One can show that every plane isometry is a translation, or a rotation about some point, or a reflection about some line, or a glide reflection (reflection followed by a translation).

Definition 5.1.13 (Symmetries of figures)**.** Let $n > 0$ and $F \subseteq \mathbb{R}^n$ (read "figure"). A *symmetry of F* is an isometry f of $\mathbb{R}^n$ with $f(F) = F$.

Exercise 5.1.14. Show the set of symmetries of F is a subgroup of $\mathrm{I}(n, \mathbb{R})$. Let $F := \{(x, y) \in \mathbb{R}^2 \mid x = \pm 1 \text{ or } y = \pm 1\}$ be the square centered at $(0, 0)$ with side length 2. List the 8 symmetries of F. Show they are 'generated' by $\rho_{\pi/2}$ and σ_0.

Exercise 5.1.15. The symmetries of the unit circle are $\mathrm{O}(2, \mathbb{R})$. $\mathrm{SO}(2, \mathbb{R})$ is isomorphic to the circle group S^1 (Example 1.1.12).

5.1.1 Dihedral groups

Lemma 5.1.16. *Let G be a nontrivial subgroup of $(\mathbb{R}, +)$ that is* discrete*, i.e., there is $\epsilon > 0$ such that $|g| > \epsilon$ for all $g \in G$. Then $G = \mathbb{Z}a$ for some real $a > 0$.*

Proof. Let $g_0 \neq 0$ in G. We can assume $g_0 > 0$ (otherwise use $-g_0 \in G$). If $g, g' \in G$ are distinct, they have distance $|g - g'| > \epsilon$. Hence, $G \cap (0, g_0]$ is nonempty and finite, so contains a minimal element a. Then $\mathbb{Z}a \subseteq G$. Conversely, let $g \in G$ and write $g = (z + \delta)a$ for $z \in \mathbb{Z}$ and $0 \leqslant \delta < 1$. Then $g' := g - za = \delta a \in G$ satisfies $0 \leqslant g' < a$, so $g' = 0$, so $\delta = 0$, so $g \in \mathbb{Z}a$. □

Theorem 5.1.17. *For every finite subgroup G of $\mathrm{O}(2, \mathbb{R})$ there is $n > 0$ such that either the elements of G are*

$$R_{2\pi/n}, R_{2 \cdot 2\pi/n}, \ldots, R_{n \cdot 2\pi/n} = I_2,$$

or G is isomorphic to the dihedral group D_n *whose elements are*

$$R_{2\pi/n}, R_{2 \cdot 2\pi/n}, \ldots, R_{n \cdot 2\pi/n} = I_2, \quad R_{n \cdot 2\pi/n} S_0, R_{2 \cdot 2\pi/n} S_0, \ldots, R_{n \cdot 2\pi/n} S_0 = S_0.$$

Proof. First case: G contains only rotations R_α. Then $G' := \{\alpha \mid R_\alpha \in G\}$ is a discrete subgroup of $(\mathbb{R}, +)$ (as G is finite). Choose $a > 0$ with $G' = \mathbb{Z}a$. As $I_2 = R_{2\pi} \in G$, G' contains 2π, so $a = 2\pi/n$ for some $n \in \mathbb{N}$. Then

$$G = \{R_{z\cdot 2\pi/n} \mid z \in \mathbb{Z}\} = \{R_{2\pi/n}, ..., R_{n\cdot 2\pi/n}\}.$$

Second case: G contains some reflection S_α. By linear algebra one sees that S_α has orthogonal eigenvectors with eigenvalues 1 and -1, so there is $C \in \mathrm{O}(2, \mathbb{R})$ such that $C^{-1}S_\alpha C = S_0$. Replacing G by the isomorphic subgroup $C^{-1}GC = \{C^{-1}AC \mid A \in G\}$ (see below), we can assume $S_0 \in G$. Let H be the subgroup of G containing all rotations in G. Choose $R_{2\pi/n}$ for H according to the 1st case. Then $D_n \subseteq G$. We claim $=$. Let $A \in G$. If $A \in H$, then $A \in D_n$. Otherwise, $A = R_\alpha S_0$ describes a reflection. Then G contains $AS_0 = R_\alpha S_0^2 = R_\alpha$, so $R_\alpha \in H$ is a power of $R_{2\pi/n}$. Again, $A \in D_n$. □

The argument concerning $C^{-1}GC$ is a general one:

Definition 5.1.18. Let $(G, \cdot)$ be a group and $g \in G$. The map $x \mapsto gxg^{-1}$ is *conjugation by g*. Such maps are *inner automorphisms of G*.

Exercise 5.1.19. Check that inner automorphisms are automorphisms. If α is an automorphism of G and $H \subseteq G$ a subgroup, then $\alpha(H)$ is a subgroup of G isomorphic to H.

Exercise 5.1.20. $D_2 \cong K_4$, the Klein four-group (cf. Example 2.5.12).

Example 5.1.21. Let $n > 2$ and $P_n \subseteq \mathbb{R}^2$ be the *regular n-gon*: it contains

$$v_k := \big(\cos(k2\pi/n), \sin(k2\pi/n)\big)$$

for $k = 0, ..., n-1$ and line segments connecting consecutive ones. Then D_n is isomorphic to the symmetries of P_n.

Proof. Symmetries of P_n permute the v_k's (since they preserve lengths) and this permutation determines the map on all other points – hence there are finitely many such symmetries. We claim they are represented by matrices in $\mathrm{O}(2, \mathbb{R})$. Then we are done by the theorem: the symmetries contain σ_0 and $\rho_{2\pi/n}$ but no rotation ρ_α with $0 < \alpha < 2\pi/n$.

To prove the claim, write a symmetry f as $x \mapsto Ax + a$ with $A \in \mathrm{O}(2, \mathbb{R})$ and $a \in \mathbb{R}^2$. We have to show $a = 0$. Recalling Definition 1.6.8, the v_k's are the n-th roots of unity (viewing $\mathbb{R}^2$ as $\mathbb{C}$). By Remark 1.6.9, we have $\sum_{k=0}^{n-1} v_k = 0$ in $\mathbb{R}^2$. As f permutes the v_k's,

$$0 = \textstyle\sum_{k=0}^{n-1}(Av_k + a) = A(\sum_{k=1}^{n-1} v_k) + na = na.$$ □

5.2 Permutations

Definition 5.2.1. For a set X let $\mathrm{Sym}(X)$ be the set of permutations of X. The *symmetric group over X* is $(\mathrm{Sym}(X), \circ)$ where $\circ$ denotes composition. We often write $\sigma\tau$ instead $\sigma \circ \tau$.

For $n \in \mathbb{N}$ write

$$S_n := \mathrm{Sym}(\{1, ..., n\}).$$

Remark 5.2.2.

1. The neutral element of $\mathrm{Sym}(X)$ is the identity id_X on X; the group inverse of $\sigma \in \mathrm{Sym}(X)$ is the inverse function σ^{-1}.
2. The set X plays no role: if Y is a set and $B : X \to Y$ a bijection, then $\mathrm{Sym}(X) \cong \mathrm{Sym}(Y)$ via $\sigma \mapsto B \circ \sigma \circ B^{-1}$ (exercise).
3. S_n has order $|S_n| = n!$. If $n > 2$, then S_n is not abelian.

Proposition 5.2.3 (Cayley)**.** *Every group G is isomorphic to a subgroup of* $\mathrm{Sym}(G)$.

Proof. It suffices to define a group monomorphism from G into $\mathrm{Sym}(G)$. For $x \in G$ let λ_x be the map $y \mapsto xy$. By Exercise 1.1.3, $\lambda_x \in \mathrm{Sym}(G)$. Define $\varphi : G \to \mathrm{Sym}(G)$ by

$$\varphi(x) := \lambda_x.$$

Homomorphism: we have to show that $\varphi(xy) = \varphi(x) \circ \varphi(y)$, i.e., $\lambda_{xy} = \lambda_x \circ \lambda_y$. But for all $z \in G$, $\lambda_{xy}(z) = xyz = \lambda_x(yz) = \lambda_x(\lambda_y(z))$.

Injective: if $\varphi(x) = \varphi(y)$, then $\lambda_x = \lambda_y$, so $xz = yz$ for all $z \in G$, so $x = y$. □

Proof of Theorem 5.1.3. It suffices to find a monomorphism ψ from S_n into $\mathrm{O}(n, \mathbb{R})$. Let ψ map $\sigma \in S_n$ to the permutation matrix P_σ with i-th column $e_{\sigma(i)}$ (from the standard basis). Then ψ is clearly injective and a homomorphism: $P_{\sigma\tau} = P_\sigma P_\tau$ because for all i:

$$P_\sigma P_\tau e_i = P_\sigma e_{\tau(i)} = e_{\sigma(\tau(i))} = P_{\sigma\tau} e_i.$$ □

Definition 5.2.4. Let $n > 1$. $\sigma \in S_n$ is a *cycle* if it is a k-cycle for some $k > 1$, that is, if there are pairwise distinct $i_1, ..., i_k \in \{1, ..., n\}$ such that $\sigma(i_1) = i_2, \sigma(i_2) = i_3, ..., \sigma(i_{k-1}) = i_k, \sigma(i_k) = i_1$ and $\sigma(i) = i$ for all $i \in \{1, ..., n\} \setminus \{i_1, ..., i_k\}$. We write

$$\sigma = (i_1 i_2 \cdots i_k).$$

Two cycles $(i_1 \cdots i_k)$ and $(j_1 \cdots j_\ell)$ are *disjoint* if $\{i_1, ..., i_k\} \cap \{j_1, ..., j_\ell\} = \varnothing$. 2-cycles (ij) (with $|i - j| = 1$) are *(neighbor) transpositions*.

Remark 5.2.5. Let $k > 1$ and consider S_n.

1. The inverse of a k-cycle is a k-cycle: $(i_1 i_2 \cdots i_k)^{-1} = (i_k i_{k-1} \cdots i_1)$.
2. $(i_1 i_2 \cdots i_k) = (i_k i_1 i_2 \cdots i_{k-1}) = (i_{k-1} i_k i_1 i_2 \cdots i_{k-1}) = \cdots$.
3. Write $\sigma^2 = \sigma\sigma, \sigma^3 = \sigma\sigma^2, ...$. For a k-cycle σ we have $\sigma^k = \mathrm{id}_{\{1,...,n\}}$ and $\sigma^\ell \neq \sigma^{\ell'}$ for all $\ell < \ell' < k$ (exercise). E.g., $(123)((123)(123)) = (123)(132) = \mathrm{id}_{\{1,2,3\}}$.
4. Disjoint cycles σ, τ commute: $\sigma\tau = \tau\sigma$.

Example 5.2.6. Consider the symmetries of the regular 3-gon P_3 (the triangle). There is exactly one symmetry for every permutation of the 3 vertices. So $D_3 \cong S_3$.

In more detail, the elements of $D_3 \subseteq \mathrm{O}(2, \mathbb{R})$ are the identity I_2, reflections $S_0, S_{2\pi/3}, S_{4\pi/3}$ and rotations $R_{2\pi/3}, R_{4\pi/3}$. S_3 has $3! = 6$ elements, namely the identity $\mathrm{id}_{\{1,2,3\}}$, transpositions $\tau_1 := (23), \tau_2 := (12), \tau_3 := (13)$, and 3-cycles $\sigma_1 = (123), \sigma_2 := (132)$.

The order of the lists show the isomorphism from D_3 onto S_3, namely: $I_2 \mapsto \mathrm{id}_{\{1,2,3\}}, S_0 \mapsto \tau_1, S_{2\pi/3} \mapsto \tau_2$ etc.. The vertices of P_3 are $1, e^{2\pi i/3}, e^{4\pi i/3}$. E.g., S_0 is the reflection about the x-axis; it behaves like τ_1 in that it fixes vertex 1 and swaps vertices 2 and 3.

Theorem 5.2.7 (Cycle decomposition)**.** *Let $n > 1$. Every $\sigma \in S_n$ is a product of disjoint cycles. The set of these cycles is unique.*

Proof. By convention we understand the identity $\mathrm{id}_{\{1,\dots,n\}}$ equals the empty product.

Existence: let $\mathrm{id}_{\{1,\dots,n\}} \neq \sigma \in S_n$. Consider 'fixed' and 'moved' points: $F := \{i \mid \sigma(i) = i\}$ and $M_1 := \{1,\dots,n\} \setminus F \neq \varnothing$. Choose $i_1 \in M_1$ and consider $\sigma(i_1), \sigma^2(i_1), \dots$. Choose $d_1 < d_2$ with $\sigma^{d_1}(i_1) = \sigma^{d_2}(i_1)$. Then $\sigma^{d_2-d_1}(i_1) = i_1$. Hence there is a minimal k_1 with $\sigma^{k_1+1}(i_1) = i_1$. Then we have a k_1-cycle $\sigma_1 := (i_1\ \sigma(i_1) \cdots \sigma^{k_1}(i_1))$ in M_1. Note $k_1 > 0$ as $i_1 \notin F$.

Let $M_2 := M_1 \setminus \{i_1, \sigma(i_1), \dots, \sigma^{k_1}(i_1)\}$. If $M_2 = \varnothing$, then $\sigma = \sigma_1$ and we are done. Otherwise, choose $i_2 \in M_2$ and find a cycle $\sigma_2 = (i_2\ \sigma(i_1) \cdots \sigma^{k_2}(i_2))$ as before. We observe that σ_1, σ_2 are disjoint: every element of $C_1 := \{i_1, \sigma(i_1), \dots, \sigma^{k_1}(i_1)\}$ equals $\sigma(x)$ for another $x \in C_1$; as $i_2 \notin C_1$, by injectivity $\sigma(i_2) \notin C_1$; for the same reason $\sigma^2(i_2) \notin C_1$ and so on.

Continue.

Uniqueness: assume $\sigma = \sigma_1 \cdots \sigma_k = \tau_1 \cdots \tau_\ell$ for disjoint cycles σ_i and disjoint cycles τ_j. Let $i \notin F$. Since disjoint cycles commute, we can assume i appears in σ_1 and τ_1. Let $d > 1$ be minimal with $\sigma^d(i) = i$. Then $\sigma_1 = \tau_1 = (i\ \sigma(i) \cdots \sigma^{d-1}(i))$. Hence $\sigma_2 \cdots \sigma_k = \tau_2 \cdots \tau_\ell$.

Continue. □

Corollary 5.2.8. *Let $n > 1$. Every $\sigma \in S_n$ is a product of neighbor transpositions.*

Proof. A cycle is a product of transpositions: $(i_1 i_2 \cdots i_k) = (i_1 i_k) \cdots (i_1 i_3)(i_1 i_2)$. A transposition (ij),with $i < j$, is a product of neighbor transpositions: move j stepwise to position i, and then i from position $i+1$ stepwise to position j:

$$(j-1\ j)\ (j-1\ j-2) \cdots (i+1\ i+2)\ \ (i\ i+1)\ (i+1\ i+2) \cdots (j-2\ j-1)\ (j-1\ j). \qquad \square$$

Example 5.2.9. In S_4: $(14) = (34)(23)(12)(23)(34)$.

Below we understand $\{\pm 1\}$ as a group with multiplication, i.e., C_2 (Definition 1.6.8).

Theorem 5.2.10. *Let $n > 1$. There exists a group homomorphism*

$$\mathrm{sign} : S_n \to \{\pm 1\}$$

mapping transpositions to -1. In particular, it maps $\sigma \in S_n$ to 1 if and only if σ is a product of an even number of transpositions.

Proof. Define $\mathrm{sign} := \det \circ \psi$ where ψ is the homomorphism from Theorem 5.1.3, i.e.,

$$\mathrm{sign}(\sigma) := \det(P_\sigma).$$

Note $P_{(i\ i+1)}$ results from I_n by swapping row i and row $i+1$. Hence, neighbor transpositions have sign -1. We just saw that (ij) with $i < j$ is a product of $2(j-i)-1$ neighbor transpositions, so $\mathrm{sign}((ij)) = (-1)^{2(j-1)-1} = -1$. The 2nd sentence follows. □

Definition 5.2.11. Let $n > 1$. A permutation $\sigma \in S_n$ is *even* if $\mathrm{sign}(\sigma) = 1$, and otherwise *odd*. The set of even permutations is denoted A_n and called *alternating group*.

Remark 5.2.12. A_n is a subgroup of S_n of order $n!/2$

Proof. Subgroup: $\sigma, \tau \in A_n$, then $\sigma\tau^{-1} \in A_n$ since $\operatorname{sign}(\sigma\tau^{-1}) = \operatorname{sign}(\sigma)\operatorname{sign}(\tau)^{-1} = 1$. For the order, one easily checks that $\sigma \mapsto (12)\sigma$ is a bijection from A_n onto $S_n \setminus A_n$. □

Exercise 5.2.13. Let $n > 1$ and $1 \leqslant k \leqslant n$. k-cycles in S_n have sign $(-1)^{k-1}$.

Example 5.2.14. Writing $\sigma \in S_4$ as a product of disjoint cycles, we see σ is a 4-cycle, a 3-cycle or a product of at most 2 transpositions. If σ is even, it is a 3-cycle or a product of 2 disjoint transpositions (not disjoint would give a 3-cyle or $1 = \mathrm{id}_{\{1,2,3,4\}}$). Hence, the $4!/2 = 12$ elements of A_4 are

$$(123), (132), (234), (243), (134), (143), (142), (124)$$
$$(12)(34), (13)(24), (14)(23), 1.$$

The second row forms a subgroup K_4' of A_4 consisting of all x with $x^2 = 1$. It is isomorphic to Klein's four-group K_4 (write down an isomorphism or apply Proposition 5.3.19).

5.3 Cyclic groups

Notation: if not stated otherwise we write groups G in multiplicative notation. The neutral element is denoted 1 or 1_G. Given $x \in G$ we write $x^0 := 1$ and for $k \in \mathbb{N} \setminus \{0\}$:

$$x^k := x \cdots x \ (k \text{ times}), \quad x^{-k} := x^{-1} \cdots x^{-1} \ (k \text{ times})$$

In additive notation $(G, +)$, the neutral element is 0 and $x^{\pm k}$ becomes $\pm kx$.

Definition 5.3.1. For a group G and $X \subseteq G$ the *subgroup of G generated by X* is

$$\langle X \rangle := \left\{ x_1^{\epsilon_1} \cdots x_n^{\epsilon_n} \mid n \in \mathbb{N} \setminus \{0\}, x_i \in X, \epsilon_i \in \{\pm 1\} \text{ for all } 1 \leqslant i \leqslant n \right\}.$$

Exercise 5.3.2. $\langle X \rangle$ is a subgroup of G that contains X and it is the smallest such subgroup (i.e., is contained in any other such subgroup).

Definition 5.3.3. Let G be a group and $X \subseteq G$. If $G = \langle X \rangle$, then G is *generated by X*. G is *finitely generated* if G is generated by some finite $X \subseteq G$. We write

$$\langle x_1, \ldots, x_n \rangle := \langle X \rangle,$$

if $X = \{x_1, \ldots, x_n\}$ for some $n \in \mathbb{N}$. G is *cyclic* if $G = \langle x \rangle$ for some *generator* $x \in G$.

Remark 5.3.4. Let G be a group and $x \in G$.

1. $\langle x \rangle = \{x^k \mid k \in \mathbb{Z}\}$. In particular, $x \mapsto x^k$ is an epimorphism from $(\mathbb{Z}, +)$ onto $\langle x \rangle$.
2. If G is abelian and $x_1, \ldots, x_n \in G$, then $\langle x_1, \ldots, x_n \rangle = \{x_1^{k_1} \cdots x_n^{k_n} \mid k_1, \ldots, k_n \in \mathbb{Z}\}$. In particular, $(k_1, \ldots, k_n) \mapsto x_1^{k_1} \cdots x_n^{k_n}$ is an epimorphism from $(\mathbb{Z}^n, +)$ onto $\langle x_1, \ldots, x_n \rangle$; here, $(\mathbb{Z}^n, +)$ is the additive group of the ring $\mathbb{Z}^n$.

3. Cyclic groups are abelian. Indeed, $x^k x^{k'} = x^{k+k'} = x^{k'} x^k$.

Example 5.3.5. Let $n > 1$.

1. $(\mathbb{Z}, +)$ and $(\mathbb{Z}_n, +)$ are cyclic, $(\mathbb{Z}^n, +)$ is not.
2. In $\mathrm{O}(2, \mathbb{R})$, $\langle R_{2\pi/n}, S_0 \rangle = D_n$.
3. By Theorems 5.2.7, 5.2.8: $S_n = \langle \text{cycles} \rangle = \langle \text{transpositions} \rangle = \langle \text{neighbor transpositions} \rangle$.
4. $S_n = \langle (12\cdots n), (12) \rangle$. This follows from Corollary 5.2.8 noting for $\sigma := (12\cdots n)$

$$(23) = \sigma(12)\sigma^{-1}, (34) = \sigma(23)\sigma^{-1}, \ldots, (n-1\ n) = \sigma(n-2\ n-1)\sigma^{-1}.$$

5. $(\mathbb{Q}, +)$ is not finitely generated.

   Indeed: given finitely many $a_1/b_1, \ldots, a_n/b_n$ with $a_i, b_i \in \mathbb{Z}$ choose a prime $p \nmid b_1 \cdots b_n$; then $1/p \neq \sum_i c_i \cdot a_i/b_i$ for all $c_i \in \mathbb{Z}$, so $1/p \notin \langle a_1/b_1, \ldots, a_n/b_n \rangle$.
6. Similarly, the additive group $(\mathbb{Q}/\mathbb{Z}, +)$ of the ring $\mathbb{Q}/\mathbb{Z}$ is not finitely generated.

Exercise 5.3.6. Let $1 \leqslant i < j \leqslant n$ and $\gcd(j-i, n) = 1$. Then $S_n = \langle (12\cdots n), (ij) \rangle$.

Example 5.3.7. Let $n > 2$. The alternating group A_n is generated by the 3-cycles.

Proof. Clearly, 3-cycles are in A_n (Exercise 5.2.13). By Theorem 5.2.10, every $\sigma \in A_n$ is a product of an even number of transpositions. It thus suffices to write a product of 2 transpositions $(ij), (k\ell)$ as a product of 3-cycles.

If $(ij), (k\ell)$ are disjoint, then $(ij)(k\ell) = (kji)(k\ell i)$. Otherwise we can assume $i = k$. If $j = \ell$, then $(ij)(k\ell)$ is the identity, so assume $j \neq \ell$. Then $(ij)(k\ell) = (ij)(i\ell) = (i\ell j)$. □

Definition 5.3.8. Let G be a group, and $x \in G$. The *order* $\mathrm{ord}(x)$ *of* x *(in* G*)* is the order (i.e., cardinality) of $\langle x \rangle$; we write $\mathrm{ord}(x) := \infty$, if $\langle x \rangle$ is infinite.

The following lemma gathers all you have to know about orders.

Lemma 5.3.9. *Let* G *be a group and* $x \in G$ *have finite order* n*.*

1. $E := \{k > 0 \mid x^k = 1\} \neq \varnothing$,
2. $n = \min E$,
3. $\langle x \rangle = \{1, x, \ldots, x^{n-1}\}$,
4. *for every* $k \in \mathbb{Z}$*:* $x^k = 1 \iff n \mid k$.
5. *for every* $k \in \mathbb{Z}$*:* $\mathrm{ord}(x^k) = n/\gcd(k, n)$.

Proof. (1): as $\langle x \rangle$ is finite and $x, x^2, x^3, \ldots \in \langle x \rangle$ there are $0 < k < \ell$ with $x^k = x^\ell$, so $x^{\ell-k} = 1$, so $k - \ell \in E \neq \varnothing$. Let $t := \min E$.

(2,3): the elements $1, x, \ldots, x^{t-1}$ are pairwise distinct (otherwise there are $0 < k < \ell < t$ with $\ell - k \in E$, a contradiction to the choice of t). We claim these elements list $\langle x \rangle$. They are clearly contained in $\langle x \rangle$. We show they contain x^k for all $k \in \mathbb{Z}$. Write $k = qt + r$ by Euclidian division, so $0 \leqslant r < t$. Then $x^k = (x^t)^q x^r = x^r$.

(4): if $n \mid k$, say $n\ell = k$, then $x^k = (x^n)^\ell = 1$. Conversely, assume $x^k = 1$ and write $k = qn + r$ with $0 \leqslant r < n$ by Euclidian division; then $1 = x^k = (x^n)^q x^r = x^r$, so $r = 0$ by (2).

(5): argue as in Remark 2.7.2 (4): $(x^k)^j = 1 \Leftrightarrow n \mid kj \Leftrightarrow n/\gcd(k,n) \mid kj/\gcd(k,n)$ $\Leftrightarrow n/\gcd(k,n) \mid j$ (by Remark 2.1.9 (3), (5)). The minimal such $j > 0$ is $n/\gcd(k,n)$. □

Definition 5.3.10. The *exponent* $\exp(G)$ *of* a group G is the minimal $n > 0$ such that $x^n = 1$ for all $x \in G$; if no such n exists, then $\exp(G) := \infty$.

Remark 5.3.11. If $\exp(G) \neq \infty$, then $\operatorname{ord}(x) \mid \exp(G)$ for all $x \in G$ (by the lemma (4)).

Exercise 5.3.12. Groups with exponent 2 are abelian.

Remark 5.3.13 (Boolean rings)**.** One easily (but tediously) checks that $(P(\mathbb{N}), +)$ is an uncountable group where $X+Y := (X \smallsetminus Y) \cup (Y \smallsetminus X)$ is the symmetric difference of $X, Y \subseteq \mathbb{N}$. The neutral element is $0 := \varnothing$ and the inverse is $-X = X$. Then $X + X = 0$, so the exponent is 2. If we set $X \cdot Y := X \cap Y$ we get a commutative ring with $1 := \mathbb{N}$ satisfying $X^2 = X$ for all $X \subseteq \mathbb{N}$. Such rings are called *Boolean* and are "the same" as so-called *Boolean algebras* studied in mathematical logic.

Examples 5.3.14. Let $n > 1$.

1. In $(\mathbb{Z}_6, +)$, list 'powers' $1x = x, 2x = x + x, 3x = x + x + x, \ldots$:

$1x$	$2x$	$3x$	$4x$	$5x$	$6x$	$\operatorname{ord}(x)$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	1
$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	6
$\bar{2}$	$\bar{4}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	3
$\bar{3}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	2
$\bar{4}$	$\bar{2}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	3
$\bar{5}$	$\bar{4}$	$\bar{3}$	$\bar{2}$	$\bar{1}$	$\bar{0}$	6

2. In $(\mathbb{Z}_n^\times, \cdot)$, the order of $[x]_n$ is the order of x modulo n as defined in 2.7.1. $\bar{x}$ is a generator if and only if x is a primitive root modulo n. Thus, Theorem 2.7.8 determines the n for which $\mathbb{Z}_n^\times$ is cyclic.
3. In $\mathrm{O}(2, \mathbb{R})$, $R_{2\pi/n}$ has order n, $R_{k2\pi/n}$ has order $n/\gcd(nk, n)$ and $R_{k2\pi/n}S_0$ has order 2 (because it is a reflection – or compute $R_{k\cdot 2\pi/n}S_0R_{k\cdot 2\pi/n}S_0 = R_{k\cdot 2\pi/n}S_0S_0R_{-k\cdot 2\pi/n} = I_2$).
4. In $\sigma \in S_n$, if $\sigma = \tau_1 \cdots \tau_r$ for disjoint k_i-cycles τ_i, then σ has order $\operatorname{lcm}(k_1, \ldots, k_r)$.

   Indeed, $\sigma^\ell = \tau_1^\ell \cdots \tau_r^\ell$ is the identity if and only if $k_i \mid \ell$ for all i (Remark 5.2.5 (3), (4)).

Note that all orders of elements above divide the group order (if finite). This always happens and follows from Lagrange's Theorem 5.6.4 proved in the next section.

Lemma 5.3.15. *Let G be a finite group and $x \in G$. Then* $\operatorname{ord}(x) \mid |G|$.

Exercise 5.3.16. Prove this for abelian groups (*Hint:* cf. Theorem 2.6.8).

Theorem 5.3.17 (Classification of cyclic groups)**.** *A group is cyclic if and only if it is isomorphic to either* $(\mathbb{Z},+)$ *or* $(\mathbb{Z}_n,+)$ *for some* $n>0$.

Proof. Let G be cyclic, say with generator x, so $G=\{x^k \mid k\in\mathbb{Z}\}$. Define $\varphi:\mathbb{Z}\to G$ by $\varphi(k):=x^k$. Note $\varphi(k+\ell)=\varphi(k)\varphi(\ell)$, so φ is a group homomorphism from $(\mathbb{Z},+)$ to $(G,\cdot)$. It is clearly surjective. If it is injective, we are done. Otherwise, say $x^k=x^\ell$ for $k<\ell$ implies $x^{\ell-k}=1$, so $n:=\mathrm{ord}(x)$ is finite and $G=\{1,x,...,x^{n-1}\}$. We claim $\psi(\bar{k}):=x^k$ defines an isomorphism from $(\mathbb{Z}_n,+)$ onto $(G,\cdot)$.

Well-defined: if $\bar{k}=\bar{\ell}$, write $k=\ell+nm$ for some $m\in\mathbb{Z}$ and note

$$\psi(\bar{k})=x^k=x^\ell(x^n)^m=x^\ell=\psi(\bar{\ell}).$$

That ψ is a homomorphism and surjective is clear. Injective: if $\psi(\bar{k})=\psi(\bar{\ell})$ with $0\leqslant k<\ell<n$, then $x^{\ell-k}=1$ and $\ell-k<n$, contradicting $\mathrm{ord}(x)=n$. □

Recall $\mathbb{Z}_n\cong C_n$, a group in multiplicative notation (Exercise 2.5.5).

Proposition 5.3.18. *Every group of prime order* p *is isomorphic to* C_p.

Proof. Let G be a group of order p and $x\in G\smallsetminus\{1\}$. Then $1<\mathrm{ord}(x)\mid p$ by Lemma 5.3.15, so $\mathrm{ord}(x)=p$, so $G=\langle x\rangle\cong C_p$. □

Recall the Klein four-group K_4 from Example 2.5.12.

Proposition 5.3.19. *Every group of order 4 is isomorphic to either* C_4 *or* K_4.

Proof. Let G be a group order 4. If there is an element of order 4, then $G\cong C_4$. Otherwise all elements $\neq 1$ have order 2 by Lemma 5.3.15. Let $1,x,y,z$ list the elements. Then $z=xy$ because: $xy\neq x$ as otherwise $y=1$; $xy\neq y$ as otherwise $x=1$; $xy\neq 1$ as otherwise $x=x^2y=y$ (since $x^2=1$). Similarly, $yx=z$. So we know the left partial table:

$\cdot$	1	x	y	z
1	1	x	y	z
x	x	1	z	
y	y	z	1	
z	z			1

$\cdot$	1	x	y	z
1	1	x	y	z
x	x	1	z	y
y	y	z	1	x
z	z	y	x	1

$+$	$(\bar{0},\bar{0})$	$(\bar{0},\bar{1})$	$(\bar{1},\bar{0})$	$(\bar{1},\bar{1})$
$(\bar{0},\bar{0})$	$(\bar{0},\bar{0})$	$(\bar{0},\bar{1})$	$(\bar{1},\bar{0})$	$(\bar{1},\bar{1})$
$(\bar{0},\bar{1})$	$(\bar{0},\bar{1})$	$(\bar{0},\bar{0})$	$(\bar{1},\bar{1})$	$(\bar{1},\bar{0})$
$(\bar{1},\bar{0})$	$(\bar{1},\bar{0})$	$(\bar{1},\bar{1})$	$(\bar{0},\bar{0})$	$(\bar{0},\bar{1})$
$(\bar{1},\bar{1})$	$(\bar{1},\bar{1})$	$(\bar{1},\bar{0})$	$(\bar{0},\bar{1})$	$(\bar{0},\bar{0})$

This determines G as the 2nd table because every column and every row must list G. The 3rd table is K_4 and we see $G\cong K_4$ via $x,y,z\mapsto(\bar{0},\bar{1}),(\bar{1},\bar{0}),(\bar{1},\bar{1})$. □

What subgroups can cyclic groups have?

Example 5.3.20. $(\mathbb{Z}_6,+)$ has subgroups $\langle\bar{0}\rangle=\{\bar{0}\}$ and $\langle\bar{2}\rangle=\langle\bar{4}\rangle=\{\bar{4},\bar{2},\bar{0}\}$ and $\langle\bar{3}\rangle=\{\bar{3},\bar{0}\}$ and $\langle\bar{1}\rangle=\langle\bar{5}\rangle=\mathbb{Z}_6$. By the corollary below these are all subgroups.

Theorem 5.3.21. *Let* $n,k>0$ *and* G *be a finite cyclic group of order* n *with generator* x.

1. *If* $k\mid n$, *then* $\langle x^{n/k}\rangle$ *is a subgroup of* G *of order* k.

2. If U is a subgroup of G of order k, then $k \mid n$ and $U = \langle x^{n/k} \rangle$.

Proof. (1): by Lemma 5.3.9 (5), $\operatorname{ord}(x^{n/k}) = n/\gcd(n/k, n) = k$. The proof of (2) is tricky: the case $k = 1$ is trivial, so assume $k > 1$. Write $U = \{x^{e_1}, ..., x^{e_k}\}$ for $e_i \in \mathbb{Z}$. Let $d := \gcd(e_1, ..., e_k)$ and write $d = c_1 e_1 + \cdots + c_k e_k$ with $c_i \in \mathbb{Z}$ by Bézout. Then $\langle x^d \rangle \subseteq U$ because $x^d = (x^{e_1})^{c_1} \cdots (x^{e_k})^{c_k} \in U$. Clearly, $x^{e_i} = (x^d)^{e_i/d} \in \langle x^d \rangle$, so $U = \langle x^d \rangle$. Then $k = |U| = \operatorname{ord}(x^d) = n/\gcd(d, n)$ by Lemma 5.3.9 (5). Hence, $n/k \mid d$, so $U = \langle x^d \rangle \subseteq \langle x^{n/k} \rangle$. By (1), $\operatorname{ord}(x^{n/k}) = k = |U|$, so $U = \langle x^{n/k} \rangle$. □

Exercise 5.3.22. If a group has an element of finite order, then also one of prime order.

Corollary 5.3.23. *Subgroups of cyclic groups are cyclic.*

Proof. For finite cyclic groups the claim follows from the theorem. Infinite cyclic groups are isomorphic to $(\mathbb{Z}, +)$ by Theorem 5.3.17. Their subgroups are isomorphic to $\mathbb{Z}$ or $n\mathbb{Z}$ for some $n \in \mathbb{N}$ by Lemma 2.1.5, so cyclic. □

The following gives a more abstract proof of Theorem 2.6.12 on Euler's totient.

Corollary 5.3.24. *Let $n > 1$ and $d \mid n$. In $(\mathbb{Z}_n, +)$ there are exactly $\varphi(d)$ many elements of order d. In particular, $n = \sum_{d|n} \varphi(d)$.*

Proof. In $\mathbb{Z}_n$, $\bar{k} = k \cdot \bar{1}$ has order n, i.e., generates $\mathbb{Z}_n$, if and only if $\gcd(n, k) = 1$: this follows from Lemma 5.3.9 (5) or Corollary 2.6.2. Thus, $\mathbb{Z}_n$ has $\varphi(n)$ many generators. For $d \mid n$, there is exactly one subgroup U_d of $\mathbb{Z}_n$ of order d. The elements of order d in $\mathbb{Z}_n$ are the generators of U_d. Since $U_d \cong \mathbb{Z}_d$, there are exactly $\varphi(d)$ many. 2nd statement: the sets $\{x \in \mathbb{Z}_n \mid \operatorname{ord}(x) = d\}$ with $d \mid n$ partition $\mathbb{Z}_n$ (Lemma 5.3.15). □

We gain an important insight:

Corollary 5.3.25. *Let K be a field and G be a finite subgroup of its multiplicative group. Then G is cyclic. In fact, for every $d \mid |G|$ there are exactly $\varphi(d)$ elements of order d.*

Proof. For $d \mid |G| =: n$ let $\psi(d)$ be the number of elements of order d in G. Then

$$\textstyle\sum_{d|n} \psi(d) = n = \sum_{d|n} \varphi(n).$$

Thus it suffices to show $\psi(d) \leqslant \varphi(d)$ for all $d \mid n$. Assume $\psi(d) \neq 0$ and choose $x \in G$ of order d. Every $y \in \langle x \rangle$ satisfies $y^d = 1$. Hence, $\langle x \rangle$ contains d many such y and thus all such y: the polynomial $Y^d - 1$ has $\leqslant d$ roots in K. Thus, the elements of order d are the generators of $\langle x \rangle$. As $\langle x \rangle \cong \mathbb{Z}_d$ it has $\varphi(d)$ many generators. Thus, $\psi(d) = \varphi(d)$. □

Exercise 5.3.26. List all subgroups of $(\mathbb{Z}_{15}, +)$.

Exercise 5.3.27. Let G be abelian.

1. If $x, y \in G$ have orders $n, m \in \mathbb{N}$ and n, m are coprime, then xy has order nm.
2. If $x, y \in G$ have orders $n, m \in \mathbb{N}$, then there exists $z \in G$ of order $\operatorname{lcm}(n, m)$.

   Hint: Induction on n. For $n > 1$ write $n = p^c n', m = p^d m'$ for a suitable prime p.
3. If G is finite, then $\exp(G) = \max_{x \in G} \operatorname{ord}(x)$.

5.4 Finitely generated free abelian groups

In this section we write groups G additively, with neutral element 0 or 0_G. We use the notation ax for $x \in G$ and $a \in \mathbb{Z}$ from the previous section. We also write $\mathbb{Z}x := \{ax \mid x \in \mathbb{Z}\}$.

Remark 5.4.1. This notation allows to view an abelian group G as a so-called $\mathbb{Z}$-*module*: the map $(a, x) \mapsto ax$ from $\mathbb{Z}\times G$ to G satisfies the vectorspace axioms of scalar multiplication (but $\mathbb{Z}$ is not a field), namely, for all $a, b \in \mathbb{Z}$ and $x, y \in G$:

$$1x = x, \quad (ab)x = a(bx), \quad (a+b)x = ax + bx, \quad a(x+y) = ax + by.$$

Exercise 5.4.2 (Divisible torsion-free abelian groups)**.** An abelian group G is *divisible* if for every $x \in G$ and $n > 0$ there is $y \in G$ with $ny = x$. It is *torsion-free* if $nx \neq 0$ for all $x \in G \setminus \{0_G\}$ and $n > 0$. Torsion-free divisible abelian groups "are" $\mathbb{Q}$-vectorspaces. Why?

Recall Remark 5.3.4 (2) in additive notation: if G is abelian and $x_1, ..., x_n \in G$, then

$$\langle x_1, ..., x_n \rangle = \{a_1x_1 + \cdots + a_nx_n \mid a_1, ..., a_n \in \mathbb{Z}\}.$$

Definition 5.4.3. A finitely generated abelian group G is *free* if it has a $\mathbb{Z}$-*basis*: a tuple $\bar{x} = (x_1, ..., x_r) \in G^r$ for some $r \in \mathbb{N}$ such that for all $y \in G$ there is a unique $(a_1, ..., a_r) \in \mathbb{Z}^r$ such that $y = a_1x_1 + \cdots + a_rx_r$. We call $(a_1, ..., a_n)$ the *coordinates of* y *wrt* $\bar{x}$, and

$$a_1x_1 + \cdots + a_nx_n \mapsto (a_1, ..., a_r)$$

the *coordinate map wrt* $\bar{x}$.

Remark 5.4.4. The coordinate map is a group isomorphism from G onto $\mathbb{Z}^r$.

Example 5.4.5.

1. $\{0\}$ has the empty tuple as $\mathbb{Z}$-basis: by convention, the empty sum is 0.
2. $\mathbb{Z}$ has the $\mathbb{Z}$-bases (1) and (-1); for $n > 0$, $n\mathbb{Z}$ has $\mathbb{Z}$-bases (n) and $(-n)$.
3. Let $r > 0$. Then $(\mathbb{Z}^r, +)$ has the *standard* $\mathbb{Z}$-basis $(e_1, ..., e_r)$ where e_i has 1 at the i-th component and 0 elsewhere.
4. Abelian groups with an element $x \neq 0$ of finite order do not have a $\mathbb{Z}$-basis.

   This follows from Remark 5.4.4. More concretely, given $\bar{x} = (x_1, ..., x_r) \in G^r$ and $\mathrm{ord}(x) = n \neq \infty$ write $x = a_1x_1 + \cdots + a_rx_r$ for certain $a_i \in \mathbb{Z}$, not all 0. Then $0_G = 0x_1 + \cdots + 0x_r = nx = na_1x_r + \cdots + na_rx_r$ and the na_i are not all 0.

Exercise 5.4.6. Let G be an abelian group. Show $(x_1, ..., x_r) \in G^r$ is a $\mathbb{Z}$-basis of G if and only if for every $y \in G$ there is $(a_1, ..., a_r) \in \mathbb{Z}^r$ with $y = a_1x_1 + \cdots + a_rx_r$, and, for all $(a_1, ..., a_r) \in \mathbb{Z}^r$: $a_1x_1 + \cdots + a_rx_r = 0_G$ implies $a_i = 0$ for all i.

In case, for every $a \in \mathbb{Z}$ and $i \neq 1$ also $(x_1 + ax_i, x_2, ..., x_r)$ is a $\mathbb{Z}$-basis.

Exercise 5.4.7 (Universal Property)**.** Let G be an abelian group with $\mathbb{Z}$-basis $\bar{x} = (x_1, ..., x_r)$. Let H be an arbitrary abelian group and $y_1, ..., y_r \in H$. Then the map $x_i \mapsto y_i$ has a unique extension to a homomorphism from G into H.

We have an analogue of the notion of vector space dimension:

Theorem 5.4.8. *All $\mathbb{Z}$-bases of a finitely generated free abelian group G have the same length, called the* rank *of G.*

Proof. Let $\bar{x} := (x_1, ..., x_r)$ and $\bar{y} := (y_1, ..., y_s)$ be $\mathbb{Z}$-bases of G, and φ the coordinate map wrt $\bar{x}$. We have to show $r = s$. Note $\varphi(x_1) = e_1, ..., \varphi(x_r) = e_r$ is the standard basis of the vector space $\mathbb{Q}^r$. It suffices to show that the $\varphi(y_i)$'s form a basis of $\mathbb{Q}^r$.

Linearly independent: assume $0 = (a_1/b_1)\varphi(y_1) + \cdots + (a_s/b_s)\varphi(y_s)$ for certain $a_i, b_i \in \mathbb{Z}$. Set $b := b_1 \cdots b_s$. Then, since φ is a homomorphism, we have with $c_i := ba_i/b_i \in \mathbb{Z}$:

$$0 = c_1\varphi(y_1) + \cdots + c_s\varphi(y_s) = \varphi(c_1y_1 + \cdots + c_sy_s)$$

But $\ker(\varphi) = \{0_G\}$, so $c_1y_1 + \cdots + c_sy_s = 0_G$. Since $(y_1, ..., y_s)$ is a $\mathbb{Z}$-basis, all $c_i = 0$. But then all $a_i/b_i = 0$.

Generating: given $v := (a_1/b_1, ..., a_r/b_r) \in \mathbb{Q}^r$ with $a_i, b_i \in \mathbb{Z}$, set $b := b_1 \cdots b_r$ and note $bv = (c_1, ..., c_r) \in \mathbb{Z}^r$ where $c_i := ba_i/b_i \in \mathbb{Z}$. As φ is surjective, there is $x \in G$ with $\varphi(x) = bv$. Since $\bar{y}$ is a $\mathbb{Z}$-basis, there are $d_1, ..., d_s \in \mathbb{Z}$ such that $x = d_1y_1 + \cdots + d_sy_s$. Then $bv = \varphi(x) = d_1\varphi(y_1) + \cdots + d_s\varphi(y_s)$. Hence, v is a linear combination of the $\varphi(y_i)$'s. □

Theorem 5.4.9. *Let G be a finitely generated free abelian group of rank r and U a subgroup of G. Then U is a finitely generated free abelian group of some rank $s \leqslant r$.*

Moreover, there exists a $\mathbb{Z}$-basis $(x_1, ..., x_r)$ of G and $d_1, ..., d_s > 0$ such that $(d_1x_1, ..., d_sx_s)$ is a $\mathbb{Z}$-basis of U and

$$d_1 \mid d_2, \quad d_2 \mid d_3, \quad \ldots \quad d_{s-1} \mid d_s.$$

Proof. Induction on r. If $r \leqslant 1$, the claim follows easily from Example 5.4.5. Let $r > 1$. If $U = \{0_G\}$, the claim is trivial ($s := 0$), so assume $U \neq \{0_G\}$.

Wrt to some $\mathbb{Z}$-basis any $u \in U \setminus \{0_G\}$ has at least one coordinate $\neq 0$, so u or $-u$ has at least one positive coordinate. Let $d_1 \in \mathbb{N} \setminus \{0\}$ be minimal such that there exist a $\mathbb{Z}$-basis $\bar{x} := (x_1, ..., x_r)$, an $u_1 \in U$ and $1 \leqslant i \leqslant n$ such that u_1 has i-th coordinate d_1 wrt $\bar{x}$. We can assume $i = 1$ (otherwise re-index). Let u_1 have coordinates $(d_1, a_2, ..., a_r)$ wrt $\bar{x}$.

We claim $d_1 \mid a_i$ for all $2 \leqslant i \leqslant n$. Write $a_i = q_id_1 + r_i$ with $0 \leqslant r_i < d_1$ by Euclidian division. By Exercise 5.4.6 $(x_1 + q_ix_i, x_2, ..., x_r)$ is a $\mathbb{Z}$-basis. The i-th coordinate of u_1 wrt this basis is r_i. Then $r_i = 0$ by choice of d_1.

Thus, u_1 has coordinates $(d_1, q_2d_1, ..., q_rd_1)$ wrt $\bar{x}$, so $u_1 = d_1y_1$ for

$$y_1 := x_1 + q_2x_2 + \cdots + q_rx_r.$$

Then $\bar{x}' := (y_1, x_2, ..., x_r)$ is a $\mathbb{Z}$-basis (apply $r-1$ times Exercise 5.4.6). Since $u_1 = d_1y_1$, the coordinates of u_1 wrt $\bar{x}'$ are $(d_1, 0, ..., 0)$. Let G_0 be the subgroup of G with 1st coordinate

0 wrt $\bar{x}'$. Let U_0 be the subgroup of G_0 of elements u_0 obtained from $u \in U$ by changing the 1st coordinate of u wrt $\bar{x}'$ to 0.

Since G_0 has rank $r-1$, induction gives a $\mathbb{Z}$-basis $(y_2, ..., y_r)$ of G_0 and $s \leqslant r$ and positive $d_2, ..., d_s \in \mathbb{N}$ such that $(d_2y_2, ..., d_sy_s)$ is a $\mathbb{Z}$-basis of U_0 and $d_2 \mid d_3 \mid \cdots \mid d_s$.

Every $x \in G$ can be uniquely written $x = ay_1 + y_0$ with $y_0 \in G_0, a \in \mathbb{Z}$. Hence, $\bar{y} := (y_1, y_2, ..., y_r)$ is a $\mathbb{Z}$-basis of G. We claim that $(d_1y_1, d_2y_2, ..., d_sy_s)$ is a $\mathbb{Z}$-basis of U.

We have to show for every $u \in U$ that d_1 divides the 1st coordinate b_1 of u wrt $\bar{y}$. Write $u = b_1y_1 + u_0$ for $u_0 \in U_0$ and write $b_1 = qd_1 + r$ with $0 \leqslant r < d_1$.Then $u - qu_1 = ry_1 + u_0$ has 1st coordinate r wrt $\bar{y}$; hence, $r = 0$ by choice of d_1.

In case, U_0 has the empty tuple as basis, then all $u \in U$ have wrt $\bar{y}$ all coordinates 0 except possibly the 1st which is divided by d_1. Hence (d_1y_1) is a basis of U.

Otherwise we are left to show that $d_1 \mid d_2$: write $d_2 = qd_1 + r$ with $0 \leqslant r < d_1$; then $\bar{y}' := (y_1 + qy_2, y_2, ..., y_r)$ is a $\mathbb{Z}$-basis of G. Let $u_2 := d_2y_2 \in U$ and consider $u_1 + u_2 \in U$. Note $u_1 + u_2 = d_1(y_1 + qy_2) + ry_2$ has 2nd coordinate r wrt $\bar{y}'$. Hence, $r = 0$ by choice of d_1. □

Exercise 5.4.10. Let G, H be finitely generated free abelian groups with ranks r, s. Make precise and prove: homomorphisms from G to H are given by $r \times s$-matrices over $\mathbb{Z}$.

Corollary 5.4.11. *Subgroups of finitely generated abelian groups are finitely generated.*

Proof. Let G be an abelian group generated by $x_1, ..., x_r \in G$. Let H be a subgroup of G. By Remark 5.3.4 (2) there is an epimorphism $\varphi : \mathbb{Z}^r \to G$. Then $\varphi^{-1}(H)$ is a subgroup of $\mathbb{Z}^r$. By the theorem it is generated by a finite set $Y \subseteq \mathbb{Z}^r$. Then $\varphi(Y)$ generates H. □

Remark 5.4.12. There are non-abelian finitely generated groups with subgroups that are not finitely generated. We shall see an example in Exercise 5.5.12.

5.5 Finitely presentable groups

Recall, the dihedral group $G := D_n$ is generated in $\mathrm{O}(2, \mathbb{R})$ by $R_{2\pi/n}, S_0$. Forget everything about G except that it is generated by elements r, s satisfying the relations

$$r^n = 1, \quad s^2 = 1, \quad sr = r^{-1}s.$$

Write elements as a words like $srsrr^{-1}rs^{-1}rrr$. The relations allow computations, e.g.:

$$srsrr^{-1}rs^{-1}rrr = r^{-1}ssrr^{-1}rs^{-1}r^3 = r^{-1}rs^{-1}r^3 = s^{-1}r^3 = sr^3 = (r^{-1})^3s = (r^{n-1})^3s = r^{n-3}s.$$

The 1st = follows using $sr = r^{-1}s$ on the first 2 letters; the 2nd uses $rr^{-1} = 1$ and $s^2 = 1$; the 3rd uses $r^{-1}r = 1$; the 4th uses $s^{-1} = s$ implied by $s^2 = 1$; the 5th uses 3 times $sr = r^{-1}s$; the 6th uses $r^{-1} = r^{n-1}$ implied by $r^n = 1$; the 7th uses $r^n = 1$ twice.

It is clear that such computations allow to re-write every word into one of $1, r, ..., r^{n-1}$, $s, rs, ..., r^{n-1}s$. Easy computations show these $2n$ elements have D_n's truth table. If no more equalities hold, besides the implied ones, we see $G \cong D_n$.

The above is highly informal, politely said. E.g., the stated equalities are plainly false because the words are distinct – what is equal is their 'meaning'. Also the words 'imply' and 'computation' are unclear. General definitions are given in mathematical logic. Here, we only give some simplified ad hoc definitions.

Definition 5.5.1. An *alphabet* is a non-empty set of *letters*. For a set $X \neq \varnothing$ we consider the alphabet $A(X)$ that has two letters x, x^{-1} for every $x \in X$. A *word (over $A(X)$)* is a finite tuple of letters. The set of words is $A(X)^*$. We write $w = x_1 \cdots x_n$ for $w = (x_1, ..., x_n) \in A(X)^n$ (where $n \in \mathbb{N}$). If $v = y_1 \cdots y_m \in A(X)^m$ then

$$wv := x_1 \cdots x_n y_1 \cdots y_m \in A(X)^{n+m}.$$

We write 1 for the *empty word*, the unique word of length 0. Abbreviations:

$$(x^{-1})^{-1} := x \text{ for } x \in X \text{ and } \quad w^{-1} := x_n^{-1} \cdots x_1^{-1},\ w^2 = ww,\ w^3 = w^2 w, ...$$

Remark 5.5.2. $A(X)^*$ with $(w, v) \mapsto wv$ is a monoid with neutral element 1.

Definition 5.5.3. Let $X \neq \varnothing$ and G a group. A *valuation (of X) in G* is a map $V : X \to G$. We extend V to $V^* : A(X)^* \to G$ setting first $V^*(x^{-1}) := V(x)^{-1}$ for $x \in X$ and then for $w = x_1 \cdots x_n \in A(X)^n$:

$$V^*(w) := V^*(x_1) \cdots V^*(x_n);$$

the r.h.s. is a product in G; as usual, we agree the empty product is 1_G, so $V^*(1) = 1_G$. We abuse notation and just write V instead of V^*.

A *relation (over X)* is a pair $(w, v) \in (A(X)^*)^2$. It is *true under V* if $V(w) = V(v)$. A set of relations R *(logically) implies* (w, v) if for all groups G and all valuations V in G:

$$\text{if every relation in } R \text{ is true under } V, \text{ then } (w, v) \text{ is true under } V.$$

We leave it as an exercise to verify:

Remark 5.5.4. Let $u, v, w \in A(X)^*$.

1. $\varnothing$ implies $(w, w), (w^{-1}w, 1), (ww^{-1}, 1)$.
2. $\{(u, v)\}$ implies $(v, u), (uw, vw), (wu, wv)$.
3. $\{(u, v), (v, w)\}$ implies (u, w).
4. By these rules alone, each of $(u, v), (v^{-1}u, 1), (uv^{-1}, 1)$ implies every other.

Intuitively, (w, v) stands for the assertion $w = v$. By (2), $\{(u, v)\}$ implies (w, w') where w is obtained from w' by substituting a subword u by v. We refrain from a formal definition, but such substitutions are what a 'computation with R' does; it consists in a sequence of 'equalities' $w = v$ with (w, v) implied by R.

Example 5.5.5. For $X = \{r, s\}$ our 'computations' used the relations (s, s^{-1}) and (r^{-1}, r^{n-1}). These are implied by the given $R = \{(r^n, 1), (s^2, 1), (sr, r^{-1}s)\}$. Our 'computations' show that for every word w there is $k < n$ such that (w, r^k) or $(w, r^k s)$ is implied by R.

Definition 5.5.6. Let $X \neq \varnothing$ be a set and R a set of relations over X. Consider the equivalence relation $\sim_R$ on $A(X)^*$ given by:

$$w \sim_R w' \iff R \text{ implies } (w, w').$$

$\langle X \mid R\rangle$ is the set of equivalence classes $[w]_R := \{w' \mid w \sim_R w'\}$ for $w \in A(X)^*$; we often omit index R. For $w, v \in A(X)^*$ set

$$[w]_R \cdot [v]_R := [wv]_R.$$

For $X = \{x, y, ...\}, R = \{(v, w), (v', w'), ...\}$ write $\langle x, y, ... \mid v = w, v' = w', ...\rangle$ for $\langle X \mid R\rangle$.

Lemma 5.5.7. *Let X, R be as above. Then $\langle X \mid R\rangle$ is a group with neutral element $[1]_R$ and inverses $[w]_R^{-1} = [w^{-1}]_R$ for all $w \in A(X)^*$.*

Proof. $\cdot$ is well-defined: given $[w] = [w']$ and $[v] = [v']$ we have to show $[wv] = [w'v']$. Let G be a group and V a valuation of X in G such that every relation in R is true under V. Then $V(w) = V(w')$ and $V(v) = V(v')$. Thus, $V(wv) = V(w)V(v) = V(w')V(v') = V(w'v')$.

$[1]$ is neutral (recall 1 is the empty word): $[w] \cdot [1] = [w1] = [w] = [1w] = [1] \cdot [w]$.

Inverse: $[w] \cdot [w^{-1}] = [w^{-1}] \cdot [w] = [1]$ means $[w^{-1}w] = [ww^{-1}] = [1]$ and holds by Remark 5.5.4 (1). □

Definition 5.5.8. A group G is *finitely presentable* if $G \cong \langle X \mid R\rangle$ for some finite set $X \neq \varnothing$ and a finite set of relations R over X.

For a set $X \neq \varnothing$ the *free group generated by* X is $F(X) := \langle X \mid \varnothing\rangle$.

Finitely generated free abelian groups are finitely presentable (recall Remark 5.4.4). They are called "free" because, intuitively, they are obtained from the free group by requiring commutativity and nothing else:

Example 5.5.9. $\mathbb{Z} \cong \langle x \mid \varnothing\rangle = F(\{x\})$ and, for $n > 1$,

$$\mathbb{Z}^n \cong \left\langle x_1, ..., x_n \mid \{(x_i x_j, x_j x_i) \mid 1 \leqslant i < j \leqslant n\}\right\rangle.$$

Proof. The 1st statement is clear. For the 2nd, assume $n = 2$ for notational simplicity and write x, y for x_1, x_2. Let $w \in A(\{x, y\})^*$. Then $R := \{(xy, yx)\}$ implies the relation $(w, x^{a_w} y^{b_w})$ where a_w is the number of occurrences x in w minus the number of occurrences of x^{-1} in w, and b_w is similarly defined. Hence, R implies $(wv, x^{a_w + a_v} y^{b_w + b_v})$ for any word v.

Define $\varphi : \langle \{x, y\} \mid R\rangle \to \mathbb{Z}^2$ by $\varphi([w]) := (a_w, b_w)$ (omitting index R). This is well-defined: assume $[w] = [v]$; then R implies $(x^{a_w} y^{b_w}, x^{a_v} y^{b_v})$; consider the valuation V of $\{x, y\}$ in $\mathbb{Z}^2$ mapping x, y to $(1, 0), (0, 1)$; as R is true under V, so is $(x^{a_w} y^{b_w}, x^{a_v} y^{b_v})$, that is, $(a_w, b_w) = V(x^{a_w} y^{b_w}) = V(x^{a_v} y^{b_v}) = (a_v, b_v)$ in $\mathbb{Z}^2$.

Surjectivity is trivial and injectivity is easy: if $\varphi([w]) = \varphi([v])$, then $[w] = [x^{a_w} y^{b_w}] = [x^{a_v} y^{b_v}] = [v]$. To see φ is an homomorphism, recall we observed above that $(a_{wv}, b_{wv}) = (a_w + a_v, b_w + b_v)$; the l.h.s. is $\varphi([wv]) = \varphi([w][v])$, the r.h.s. is $\varphi([w])\varphi([v])$. □

Exercise 5.5.10. Let $n > 1$. Then $C_n \cong \langle x \mid x^n = 1\rangle$.

Finite groups are finitely presentable:

Exercise 5.5.11. Show $G \cong \langle G \mid R\rangle$ where R 'is' the computation table of G.

Exercise 5.5.12. In $F(\{0,1\})$, the subgroup generated by $[0^n10^n]_\varnothing, n \in \mathbb{N}$, is isomorphic to $F(\mathbb{N})$ and hence not finitely generated.

Remark 5.5.13 (Word problem for groups)**.** In 1911 Dehn, a student of Hilbert, introduced the *word problem* for a finitely presented group $\langle X \mid R\rangle$, namely to decide for a given word $w \in A(X)^*$ whether R implies $(w, 1)$. In 1955 the soviet mathematician Novikov found a finitely presented group with undecidable word problem. Beyond the iron curtain, Boone gave 1958 a different proof.

Theorem 5.5.14 (Universal property)**.** *Let $X \neq \varnothing$, R a set of relations over X. Further, let V be a valuation of X in a group G such that every relation in R is true under V.*

Then there is exactly one homomorphism $\varphi_V : \langle X \mid R\rangle \to G$ with $\varphi_V([x]_R) = V(x)$ for all $x \in X$. Moreover, a relation (w', w) is true under V if and only if $[w^{-1}w']_R \in \ker(\varphi_V)$.

Proof. Assume φ_V is such an homomorphism. Then $\varphi_V([x^{-1}]) = \varphi_V([x]^{-1}) = \varphi_V([x])^{-1} = V(x)^{-1}$, so $\varphi([y]) = V(y)$ for all $y \in A(X)$. Then for all $w = x_1 \cdots x_n \in A(X)^n$:

$$\varphi_V([w]) = \varphi_V([x_1]\cdots[x_n]) = \varphi_V([x_1])\cdots\varphi_V([x_n]) = V(x_1)\cdots V(x_n) = V(w).$$

Thus, the only possibility is to define $\varphi_V([w]) := V(w)$. It is easy to check that this is a well-defined homomorphism.

For the moreover-part note the equivalences: (w', w) is true under V, $V(w') = V(w)$, $\varphi_V([w']) = \varphi_V([w])$, $\varphi_V([w])^{-1}\varphi_V([w']) = 1_G$, $\varphi_V([w^{-1}w']) = 1_G$. □

Exercise 5.5.15 (Universality of free groups)**.** A group G generated by $X \neq \varnothing$ is isomorphic to $F(X)/N$ for some N. Make precise: every at most countable group is a factor of $F(\mathbb{N})$.

In the terminology of mathematical logic, the question of finite presentability is the same as the question for the finite axiomatizability of a certain equational theory:

Theorem 5.5.16. *A group G is finitely presentable if and only if G is generated by a finite set $\varnothing \neq X \subseteq G$ such that there exists a finite set R of relations over X that implies all relations over X that are true under the valuation id_X in G.*

Proof. $\Rightarrow$: assume $\varphi : \langle X \mid R\rangle \cong G$ for finite X, R, write $x' := \varphi([x]_R)$ for $x \in X$ and set $X' := \{x' \mid x \in X\}$. Clearly, $G = \langle X'\rangle$. For a word w over X let w' be obtained by priming all letters, and set $R' := \{(u', v') \mid (u, v) \in R\}$. We claim R' implies all relations over X' true under $\mathrm{id}_{X'}$: if (u', v') is true under $\mathrm{id}_{X'}$, then (u, v) is true under the valuation $x \mapsto [x]_R$ in $\langle X \mid R\rangle$; this means $[u]_R = [v]_R$, i.e., (u, v) is implied by R; then (u', v') is implied by R'.

$\Leftarrow$: choose $\varphi_V : \langle X \mid R\rangle \to G$ for the valuation $V := \mathrm{id}_X$ in G according to Theorem 5.5.14. Then φ_V is surjective because its image contains X. We show it is injective: if $[w]_R \in \ker(\varphi_V)$, then $(w, 1)$ is true under $V = \mathrm{id}_X$, so implied by R, hence $[w]_R = [1]_R$. □

Remark 5.5.17. The so-called *lamplighter group* is an example of a finitely generated group that is not finitely presentable (Baumslag 1961).

Example 5.5.18. Let $n > 1$. $D_n \cong \langle r, s \mid r^n = s^2 = 1, sr = r^{-1}s \rangle$.

Proof. Theorem 5.5.14 with the valuation $r, s \mapsto R_{2\pi/n}, S_0$ of $\{r, s\}$ in D_n gives an homomorphism from the r.h.s. into D_n. It is surjective because its image contains $R_{2\pi/n}, S_0$ generating D_n. It is injective because the r.h.s. has order $\leqslant 2n$ by Remark 5.5.5.

More concretely: 'computations' show that every word has the form $r^k, r^k s$ up to equivalence and verify the intended multiplication table. The words are pairwise not equivalent because there exists a valuation giving them distinct values (the above one in D_n). □

Example 5.5.19 (Quaternion group)**.** Hamilton 1844:

> *"And here there dawned on me the notion that we must admit, in some sense, a fourth dimension of space for the purpose of calculating with triples ... An electric circuit seemed to close, and a spark flashed forth."*

In $\mathrm{GL}(2, \mathbb{C})$ consider the matrices

$$\mathbf{e} := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{i} := \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \mathbf{j} := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{k} := \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

Then $Q_8 := \{ \pm \mathbf{e}, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k} \}$ is a subgroup of $\mathrm{GL}(2, \mathbb{C})$. The *Hamilton rules* are:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -\mathbf{e}.$$

These relations determine Q_8:

$$Q_8 \;\cong\; \langle \bar{e}, i, j, k \mid \bar{e}^2 = 1, i^2 = j^2 = k^2 = ijk = \bar{e} \rangle \;\cong\; \langle x, y \mid x^4 = 1, x^2 = y^2, yx = x^{-1}y \rangle.$$

Proof. The $\bar{e}ijk$-relations R are true under the obvious valuation in Q_8. Theorem 5.5.14 gives an homomorphism onto Q_8. It is surjective because its image generates Q_8. For injectivity, we show $\langle \bar{e}, i, j, k \mid R \rangle$ has order $\leqslant 8$. It suffices to verify the intended multiplication table for $1, \bar{e}, i, j, k, \bar{e}i, \bar{e}j, \bar{e}k$. To see R implies e.g. (ij, k) note it implies $(ijk, \bar{e}), (ijk^2, \bar{e}k)$, $(ijk^2, k\bar{e}), (ij\bar{e}, k\bar{e}), (ij, k)$. We omit the rest.

Let S denote the xy-relations. Theorem 5.5.14 gives an epimorphism from $\langle \bar{e}, i, j, k \mid R \rangle$ onto $\langle x, y \mid S \rangle$ once we show the relations in R are true under the valuation $\bar{e}, i, j, k \mapsto x^2, x, y, xy$. E.g. $V(ijk) = V(\bar{e})$ means $[xyxy]_S = [x^2]_S$ – 'compute' $xyxy = xx^{-1}yy = y^2 = x^2$. Thus $\langle x, y \mid S \rangle$ has order $\leqslant 8$. We are left to find an epimorphism onto Q_8: verify that the relations in S are true under the valuation $x, y \mapsto \mathbf{i}, \mathbf{j}$. □

5.6 Normal subgroups

Recall, the ring $\mathbb{Z}_n$ is obtained by identifying elements $x, y \in \mathbb{Z}$ that differ only by a multiple of n. More generally, given a ring R, the ring R/I is obtained by identifying elements x, y that differ only by an element of the ideal I.

We proceed similarly with a group G and identify elements $x, y \in G$ that differ only by an element of a subgroup U. If G is non-abelian, this has two meanings: for $x \in G, u \in U$ identify x and xu, or, identify x and ux.

Definition 5.6.1. Let G be a group, $x \in G$ and U a subgroup. Then

$$xU := \{xu \mid u \in U\}, \quad Ux := \{ux \mid u \in U\}$$

is the *left*, resp., *right coset of x wrt U*. We set

$$G/U := \{xU \mid x \in G\}, \quad U\backslash G := \{Ux \mid x \in G\}.$$

The *index of U in G* is $[G : U] := |G/U|$. We write $[G : U] = \infty$ if this is infinite.

In an additively written group the notation xU becomes $x + U$.

We leave the following as an easy exercise.

Remark 5.6.2. G/U is a partition of G, namely the set of equivalence classes of the equivalence relation $\sim_U$ on G given by

$$x \sim_U y \iff x^{-1}y \in U.$$

We have $xU = yU \Leftrightarrow x^{-1}y \in U \Leftrightarrow x \in yU$ and, in particular, $xU = U \Leftrightarrow x \in U$.

Exercise 5.6.3. Picture the multiplicative group $\mathbb{C}^\times$ of $\mathbb{C}$ as the plane without the origin.

1. $\mathbb{C}^\times/S^1$ is the partition of $\mathbb{C}^\times$ into circles around the origin, one for each radius $r > 0$.
2. $\mathbb{C}^\times/\mathbb{R}^\times$ is the partition of $\mathbb{C}^\times$ into lines through the origin (without the origin), one for each angle $0 \leqslant \alpha < \pi$ with the x-axis.
3. Find a subgroup $U \subseteq \mathbb{C}^\times$ such that $\mathbb{C}^\times/U$ is the partition of $\mathbb{C}^\times$ into orthogonal crosses centered at the origin.

Theorem 5.6.4 (Lagrange)**.** *Let G be a finite group and U a subgroup. Then*

$$|G| = |U| \cdot [G : U].$$

Proof. By the remark, it suffices to show that every left coset xU has cardinality $|U|$. But $u \mapsto xu$ is a bijection from U onto xU. □

Remark 5.6.5. The same proof works if we defined the index as $|U\backslash G|$. Hence

$$|G/U| = |U\backslash G| = |G|/|U|.$$

Can we make G/U into a group, as we did for $\mathbb{Z}_n$? We want $xU \cdot yU := xyU$. Is this well-defined? Assume $xU = x'U, yU = y'U$, i.e., $x = x'u$ and $y = y'v$ for certain $u, v \in U$. We want $xyU = x'y'U$, i.e., $xy = x'y'w$ for some $w \in U$. We know $xy = x'uy'v$, so only need $uy' = y'u'$ for some $u' \in U$ (then set $w := u'v$). This means we want $Uy' \subseteq Uy'$ for all $y' \in G$.

This condition can be equivalently formulated as follows.

Exercise 5.6.6. Let G be a group and U a subgroup. The following are equivalent:

1. $xUx^{-1} = U$ for all $x \in G$.
2. $xUx^{-1} \subseteq U$ for all $x \in G$.
3. $Ux = xU$ for all $x \in G$.
4. $Ux \subseteq xU$ for all $x \in G$.
5. $xU \subseteq Ux$ for all $x \in G$.

Remark 5.6.7. Recalling Definition 5.1.18, (1) states $\alpha(U) = U$ for all inner automorphisms α of G, i.e., conjugation by any $x \in G$ permutes U.

Definition 5.6.8. Let G be a group. A subgroup U is *normal (in G)*, symbolically

$$U \lhd G,$$

if $xUx^{-1} = U$ for all $x \in G$. If this is the case, then the *factor group of G modulo U* is $(G/U, \cdot)$ where for $x, y \in G$:

$$xU \cdot yU := xyU.$$

Remark 5.6.9. Let $U \lhd G$.

1. We already checked that $\cdot$ on G/U is well-defined.
2. $(G/U, \cdot)$ is a group: the neutral element is $1_{G/U} = U$; the inverse of xU is $x^{-1}U$.
3. If G is abelian, then all subgroups V are normal and G/V is abelian.

   Indeed, for $x, y \in G$: $xVx^{-1} = \{xvx^{-1} \mid v \in V\} = \{xx^{-1}v \mid v \in V\} = V$ and $xV \cdot yV = xyV = yxV = yV \cdot xV$.
4. $\{1\}, G$ are the *trivial* normal subgroups of G. Indeed: $x\{1\}x^{-1} = \{1\}$, $xGx^{-1} = G$.
5. Subgroups of index 2 are normal.

   Indeed: let V be such a subgroup and $x \in G$. If $x \in V$, then $xV = V = Vx$. If $x \notin V$, then both xV, Vx equal $G \setminus V$, the unique coset $\neq V$ in G/V or $V\backslash G$.

Exercise 5.6.10. Let G be a group, U a subgroup and $N \lhd G$. Then

$$NU := \{nu \mid n \in N, u \in U\}$$

is a subgroup of G. More generally, if V is a subgroup of G such that $UV = VU$, then UV is a subgroup of G.

Exercise 5.6.11. If G is a cyclic group and U a subgroup, then G/U is a cyclic group.

Example 5.6.12.

1. Let $n > 0$. As $(\mathbb{Z}, +)$ is abelian, its subgroup $n\mathbb{Z}$ is normal, and $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ by definition; the index is $[\mathbb{Z} : n\mathbb{Z}] = n$.
2. The subgroup $U := \{1, (12)\}$ is not normal in S_3 because $(123)(12)(321) = (23) \notin U$.
3. A_n is normal in S_n since it has index 2 by Remark 5.2.12 (or argue as below).
4. The special orthogonal group $U := \mathrm{SO}(n, \mathbb{R})$ is normal in $G := \mathrm{O}(n, \mathbb{R})$.

   We verify index 2 showing $AU = G \setminus U$ for all $A \in G \setminus U$. $\subseteq$: if $B \in U$, then $\det(AB) = -1$, so $AB \in G \setminus U$. $\supseteq$: if $B \in G \setminus U$, then $A^{-1}B \in U$ and $B = A(A^{-1}B) \in AU$.
5. Recall Example 5.2.14: $K_4' \lhd A_4$.

   Use Remark 5.6.7: note K_4' is the set of elements of order 2 in A_4; hence it is permuted by all (inner) automorphisms of A_4.

As a first application of factor groups we show:

Proposition 5.6.13. *Let G be a finite abelian group and p be a prime divisor of $|G|$. Then G has a subgroup of order p.*

Proof. We show by induction on $|G|$, that G has an element of order p. If $|G| = p$, use Proposition 5.3.18. Assume $|G| > p$. Let $x \in G$ have prime order q (Exercise 5.3.22) and assume $p \neq q$. Since G is abelian, $N := \langle x \rangle \lhd G$ and G/N is abelian. G/N has order $|G|/q$. As $p \mid |G|$ and $p \neq q$, also $p \mid |G|/q$. Induction gives an element $yN \in G/N$ of order p. Let n be the order of y in G. Then $(yN)^n = 1_{G/N}$, so $p \mid n$. If $p = n$ we are done, so assume $p \neq n$. By Lemma 5.3.9 (5), $y^{n/p}$ has order $n/\gcd(n/p, n) = p$. □

Remark 5.6.14. We shall later see either one of the assumptions abelian or primality (of p) can be omitted (Theorem 5.12.1 and Corollary 5.11.5). It is not generally true that a finite group G contains for every divisor d of $|G|$ a subgroup of order d (Example 5.7.13).

Example 5.6.15. There are groups $U \lhd V \lhd G$ with $U \not\lhd G$.

Proof. Consider the unit square (regular 4-gon) with corners numbered counterclockwise 1-4 starting at $(1, 0)$. Its symmetry group D_4 (Exercise 5.1.14) permutes the corners and is isomorphic to the following subgroup of S_4, writing $1 = \mathrm{id}_{\{1,2,3,4\}}$:

$$D_4' := \big\{1, (13), (24), (14)(23), (12)(34), (13)(24), (1234), (1432)\big\}.$$

D_4' has the subgroup K_4' from Example 5.2.14. Let $C_2' := \{1, (14)(23)\} \subseteq K_4'$. Then $C_2' \lhd K_4'$ and $K_4' \lhd D_4'$ because both have index 2 (Remark 5.6.9 (5)). But $C_2' \not\lhd D_4'$ because for $\sigma := (13) \in D_4', \tau := (14)(23) \in C_2'$:

$$\sigma^{-1}\tau\sigma = (31)\ (14)(23)\ (13) = (12)(34) \notin C_2'.$$ □

Example 5.6.16. The quaternion group Q_8 from Example 5.5.19 is not abelian (e.g. $\mathbf{ij} = \mathbf{k}$ and $\mathbf{ji} = -\mathbf{k}$) but every subgroup of Q_8 is normal.

Proof. Let U be a subgroup of Q_8. By Lagrange, $|U|$ devides $|Q_8| = 8$, so $|U| \in \{1, 2, 4, 8\}$. If $|U| = 1$ or $|U| = 8$, then $U = \{\mathbf{e}\}$ or $|U| = Q_8$, so U is normal. If $|U| = 4$, then $[Q_8 : U] = 2$ and U is normal by Remark 5.6.9 (5).

Finally, if $|U| = 2$, then $U = \{\mathbf{e}, x\}$ for some $x \neq \mathbf{e}$ with $x^2 = \mathbf{e}$. Hence, $x = -\mathbf{e}$. But $A(\pm\mathbf{e})A^{-1} = \pm\mathbf{e}$ for all $A \in Q_8$ (in fact, $\mathrm{GL}(2, \mathbb{C})$), so $U = \{\pm\mathbf{e}\}$ is normal. □

Exercise 5.6.17. For $n > 1$, show both $\mathrm{O}(n, \mathbb{R})/\mathrm{SO}(n, \mathbb{R})$ and S_n/A_n are isomorphic to C_2. Further, show $Q_8/\{\pm\mathbf{e}\} \cong K_4$. (Note the factor groups are well-defined.)

Exercise 5.6.18. Every abelian group of order 6 is isomorphic to C_6.

Example 5.6.19. Every non-abelian group of order 8 is isomorphic to D_4 or to Q_8.

Proof. Let G be non-abelian of order 8. Possible orders of elements are $1, 2, 4, 8$. Order 8 is impossible since G is not cyclic. By Exercise 5.3.12, not all elements have order 2. Hence there is an element $a \in G$ of order 4. Then $\langle a \rangle$ has index 2 in G, that is, for $b \in G \setminus \langle a \rangle$

$$G = \{1, a, a^2, a^3\} \cup \{b, ba, ba^2, ba^3\}.$$

What is $c := bab^{-1}$? Note $c \in \langle a \rangle \lhd G$. But $c \neq 1$ (else $a = 1$), $c \neq a$ (else $ba = ab$ and G is abelian), $c \neq a^2$ (else $1 = c^2 = ba^2b^{-1}$, so $a^2 = 1$). Hence, $c = a^3 = a^{-1}$, so $ba = a^{-1}b$.

What is b^2? Since $G/\langle a \rangle \cong C_2$, we have $b^2\langle a \rangle = (b\langle a \rangle)^2 = 1_{G/\langle a \rangle} = \langle a \rangle$, so $b^2 \in \langle a \rangle$. But $b^2 \neq a$ and $b^2 \neq a^{-1}$ as otherwise $\mathrm{ord}(b) = 8$. Thus we have two cases:

$$a^4 = 1, \quad ba = a^{-1}b, \quad b^2 = 1 \qquad \text{or} \qquad a^4 = 1, \quad ba = a^{-1}b, \quad b^2 = a^2.$$

By Examples 5.5.18 and 5.5.19 these relations R are those for D_4 and for Q_8, i.e., $\langle a, b \mid R \rangle$ is isomorphic to D_4 or Q_8. For the valuation $V := \mathrm{id}_{\{a,b\}}$ in G, Theorem 5.5.14 gives an homomorphism $\varphi_V : \langle a, b \mid R \rangle \to G$ containing a, b in its image. But a, b generate G, so φ_V is surjective. Since we know $\langle a, b \mid R \rangle$ has order 8, φ_V is injective, so an isomorphism. □

How many abelian groups of order 8 are there? You might want to check all possible 8^{8^2} multiplication tables with a computer. But this number dwarfs the number of atoms in the solar system. Some theory is needed. We shall see that there are exactly 3 abelian groups of order 8 up to isomorphism, and exactly 49 of order 100000 (Examples 5.11.18, 5.11.17).

5.6.1 Normal hull

We show that our definition of $\langle X \mid R \rangle$ is equivalent to a more standard one using normal hulls. It is surprising that the concept of normality is linked to logical implication.

Definition 5.6.20. Let G be a group and $Y \subseteq G$. The *normal hull of Y (in G)* is

$$\langle\langle Y \rangle\rangle := \left\langle \{ xyx^{-1} \mid y \in Y, x \in G \} \right\rangle.$$

Exercise 5.6.21. Show $\langle\langle Y \rangle\rangle$ is the smallest normal subgroup of G containing Y. That is: $\langle\langle Y \rangle\rangle \lhd G$ and $\langle\langle Y \rangle\rangle \subseteq N$ for all $N \lhd G$ with $Y \subseteq N$.

Lemma 5.6.22. *Let $X \neq \varnothing$ be a set, (u, v) a relation over X and R a set of relations over $X \neq \varnothing$. The following are equivalent.*

1. *R implies (u, v).*
2. *$[v^{-1}u]_\varnothing$ is in the normal hull of $\{[w^{-1}w']_\varnothing \mid (w', w) \in R\}$ in $F(X)$, denoted $\langle\langle R \rangle\rangle$.*

Proof. $\Rightarrow$: consider the valuation $V(x) := [x]\langle\langle R \rangle\rangle$ of X in $F(X)/\langle\langle R \rangle\rangle$. Then

$$V(u) = V(v) \iff [u]\langle\langle R \rangle\rangle = [v]\langle\langle R \rangle\rangle \iff [u^{-1}v] \in \langle\langle R \rangle\rangle,$$

for every relation (u, v) over X. In particular, all relations in R are true under V. Hence, any implied relation (u, v) is also true under V. Hence, $[v^{-1}u] \in \langle\langle R \rangle\rangle$.

$\Leftarrow$: assume (u, v) is not implied by R. Then there is a group G and a valuation V in G such that all relations in R are true under V but $V(u) \neq V(v)$. Choose φ_V according to Theorem 5.5.14 and set $N := \ker(\varphi_V)$. Then $[w^{-1}w'] \in N$ for all $(w', w) \in R$. By Exercise 5.6.21, $\langle\langle R \rangle\rangle \subseteq N$. Since (u, v) is not true under V, again Theorem 5.5.14 gives $[v^{-1}u] \notin N$. Hence, $[v^{-1}u] \notin \langle\langle R \rangle\rangle$. □

Corollary 5.6.23. *Let $X \neq \varnothing$ be a set and R a set of relations over X. Then*

$$\langle X \mid R \rangle \cong F(X)/\langle\langle R \rangle\rangle.$$

Proof. The map $[w]_R \mapsto [w]_\varnothing\langle\langle R \rangle\rangle$ is well-defined and bijective – use the lemma:

$$u \sim_R v \iff R \text{ implies } (u, v) \iff [v^{-1}u]_\varnothing \in \langle\langle R \rangle\rangle \iff [u]_\varnothing\langle\langle R \rangle\rangle = [v]_\varnothing\langle\langle R \rangle\rangle.$$

It is easily checked to be a homomorphism, hence an isomorphism. □

5.6.2 Simple groups

Definition 5.6.24. A group G is *simple* if $G \neq \{1\}$ and its only normal subgroups are $\{1\}$ and G.

Examples 5.6.25.

1. Groups of prime order are simple (by Lagrange). An abelian group is simple if and only if it has prime order (Proposition 5.6.13).
2. For $n > 2$, S_n is not simple as $A_n \lhd S_n$ (Example 5.6.12 (3)).
3. For $n > 1$, D_n is not simple: $N := \{R_{k \cdot 2\pi/n} \mid k < n\}$ has index 2 in D_n, so is normal.

Example 5.6.26 (Galois)**.** A_5 is simple.

Proof. Assume $\{1\} \neq N \lhd A_5$. We have to show $N = A_5$. We first claim that every 3-cycle (ijk) is conjugate to (123) in A_5, i.e., $\sigma(ijk)\sigma^{-1} = (123)$ for some $\sigma \in A_5$. Note $\sigma(ijk)\sigma^{-1} = (\sigma(i)\sigma(j)\sigma(k))$, so we clearly find such $\sigma \in S_n$. If $\sigma \notin A_5$, note $(45)\sigma \in A_5$ and

$$(45)\sigma(ijk)\sigma^{-1}(45) = (45)(123)(45) = (123).$$

It thus suffices to show that N contains some 3-cycle – then it contains all of them as N is normal, and we are done by Example 5.3.7. Let $1 \neq \sigma \in N$. If σ is not a 3-cycle its cycle decomposition is $(ij)(k\ell)$ or $(ijk\ell m)$. But

$$(ijm)(ij)(k\ell)(ijm)^{-1}\ (ij)(k\ell) = (imj), \quad (ijk)(ijk\ell m)(ijk)^{-1}(ijk\ell m)^{-1} = (ij\ell). \qquad \square$$

Theorem 5.6.27 (Jordan 1870)**.** *Let $n > 1$. Then A_n is simple if and only if $n \notin \{2, 4\}$.*

Proof. A_2 is trivial, A_3 has prime order $3!/2 = 3$. A_4 has normal subgroup K_4' from Example 5.6.12 (5). We know A_5 is simple. Let $n > 5$ and $N \lhd A_n$.

Case 1: N contains some $\sigma \neq 1$ with a fixed-point, i.e., $\sigma(i) = i$ for some i. Let $G_i = \{\sigma \in A_n \mid \sigma(i) = i\}$. Clearly, $G_i \cong A_{n-1}$ and $\{1\} \neq N \cap G_i \lhd G_i$. By induction (and $n > 5$), $N \cap G_i = G_i$, i.e., $G_i \subseteq N$. But for every j and $\sigma \in G_j$ we have $\tau\sigma\tau^{-1} \in G_i \subseteq N$ for any $\tau \in A_n$ with $\tau(j) = i$, hence $\sigma = \tau^{-1}\tau\sigma\tau^{-1}\tau \in N$. Thus, N contains all $G_1, \ldots, G_n$. Every element of A_n is a product of pairs of transpositions. As $n \geqslant 5$ such a pair is in some G_j. Thus, $N = A_n$.

Case 2: every permutation in $N \smallsetminus \{1\}$ moves all numbers. Then distinct $\sigma, \sigma' \in N$ disagree on all $1 \leqslant i \leqslant n$: if $\sigma(i) = \sigma'(i)$, then $\sigma'\sigma^{-1}(i) = i$, so $\sigma'\sigma^{-1} = 1$, so $\sigma = \sigma'$.

Given $\sigma \in N$, we show $\sigma = 1$. Write σ as a product of disjoint cycles. Assume a k-cycle $(a_1 \cdots a_k)$ with $k \geqslant 3$ appears. As $n \geqslant 5$ there is a 3-cycle τ that contains a_3 but not a_1, a_2. Then $\tau(a_1 \cdots a_k)\tau^{-1} = (\tau(a_1) \cdots \tau(a_k)) = (a_1 a_2 \tau(a_3) \cdots \tau(a_k))$. Hence, $\tau\sigma\tau^{-1} \in N$ agrees with σ on a_1 but not on a_2, a contradiction.

Thus, σ is a product of disjoint transpositions. We claim the product is empty. Otherwise, $\geqslant 3$ transpositions appear (here we use $n > 5$). Say, $\sigma = (a_1 a_2)(a_3 a_4)(a_5 a_6) \cdots$. Let $\tau := (a_1 a_2)(a_3 a_5)$. Then $\tau\sigma\tau^{-1} = (a_1 a_2)(a_5 a_4)(a_3 a_6) \cdots$. But this agrees with σ on a_1, a_2 but not on a_3, contradiction. $\square$

5.7 Noether's isomorphism theorems

Normal subgroups are best identified as kernels:

Lemma 5.7.1. *Let $\varphi : G \to G'$ be a group homomorphism.*

1. *If $N' \lhd G'$, then $\varphi^{-1}(N') \lhd G$; in particular, $\ker(\varphi) = \varphi^{-1}(\{1_{G'}\}) \lhd G$.*
2. *If $N \lhd G$, then $\varphi(N) \lhd \varphi(G)$.*

Proof. (1): by Remark 1.1.9 (6), $\varphi^{-1}(N')$ is a subgroup of G; let $x \in G$ and $n \in \varphi^{-1}(N')$; then $\varphi(xnx^{-1}) = \varphi(x)\varphi(n)\varphi(x)^{-1}$ is in N' because $\varphi(n) \in N' \lhd G'$; hence, $xnx^{-1} \in \varphi^{-1}(N')$.

(2): by Remark 1.1.9 (5), $\varphi(N)$ is a subgroup of $\varphi(G)$, a subgroup of G'; let $n' \in \varphi(N)$ and $x' \in \varphi(G)$, say $\varphi(n) = n', \varphi(x) = x'$ for $n \in N, x \in G$. Then $x'n'x'^{-1} = \varphi(x)\varphi(n)\varphi(x)^{-1} = \varphi(xnx^{-1}) \in \varphi(N)$ because $xnx^{-1} \in N \lhd G$. $\square$

Remark 5.7.2. In (2), $\varphi(N) \lhd G'$ can fail. E.g., if G is a subgroup of G' that is not normal and $\varphi = \mathrm{id}_G$, then $N := G \lhd G$ but $\varphi(N) = N \not\lhd G'$.

Examples 5.7.3. Let $n > 1$. A_n is the kernel of $\mathrm{sign} : S_n \to \{\pm 1\}$. $\mathrm{SO}(n, \mathbb{R})$ is the kernel of $\det : \mathrm{O}(n, \mathbb{R}) \to \{\pm 1\}$. $\mathrm{SL}(n, \mathbb{R})$ is the kernel of $\det : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^\times$.

All normal subgroups are kernels:

Proposition 5.7.4. *Let G be a group and $N \lhd G$. The* canonical projection π_N *given by $\pi_N(x) := xN$ is an epimorphism from G onto G/N with kernel N.*

Proof. π_N is clearly surjective. It is a group homomorphism by definition of G/N. For the kernel note $\pi_N(x) = 1_{G/N} \iff xN = N \iff x \in N$. □

Theorem 5.7.5 (Universal property)**.** *Let $\varphi : G \to G'$ be a group homomorphism and $N \lhd G$ with $N \subseteq \ker(\varphi)$. Then there is exactly one group homomorphism $\bar{\varphi} : G/N \to G'$ with $\varphi = \bar{\varphi} \circ \pi_N$.*

Exercise 5.7.6. Prove this.

Exercise 5.7.7 (Correspondence theorem)**.** Let G be a group and $N \lhd G$. Then

$$N' \mapsto \pi_N(N') = N'/N$$

is a bijection from the set of $N' \lhd G$ with $N \subseteq N'$ onto the set of $\tilde{N} \lhd G/N$; the inverse is

$$\tilde{N} \mapsto \pi_N^{-1}(\tilde{N}).$$

Theorem 5.7.8 (1st isomorphism theorem)**.** *Let $\varphi : G \to G'$ be a group epimorphism. Then*

$$G/\ker(\varphi) \cong G'.$$

In fact, there is exactly one isomorphism $\bar{\varphi} : G/\ker(\varphi) \cong G'$ with $\varphi = \bar{\varphi} \circ \pi_{\ker(\varphi)}$.

Proof. Take the homomorphism $\bar{\varphi}$ from the universal property with $N := \ker(\varphi)$. It is surjective: let $x' \in G'$ and choose $x \in G$ with $\varphi(x) = x'$; then $\bar{\varphi}(xN) = \varphi(x) = x'$.

It is injective: if $xN \in \ker(\bar{\varphi})$, then $\varphi(x) = 1$, so $x \in N$, so $xN = N = 1_{G/N}$. □

More abstract proof of Theorem 5.3.17. If G is a cyclic group, there is an epimorphism $\varphi : \mathbb{Z} \to G$ (Remark 5.3.4 (1)). Choose $n \in \mathbb{N}$ such that $\ker(\varphi) = n\mathbb{Z}$ (Lemma 2.1.5). By the theorem, $\mathbb{Z}/n\mathbb{Z} \cong G$. But $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}$ for $n = 0$, and $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$ for $n > 0$. □

Example 5.7.9. $\mathrm{SO}(2, \mathbb{R}) \cong \mathbb{R}/\mathbb{Z}$ for the additive groups $\mathbb{R}$ and $\mathbb{Z}$.

Proof. The map $\alpha \mapsto e^{2\pi i \cdot \alpha}$ is an epimorphism from $(\mathbb{R}, +)$ onto the circle group $(S^1, \cdot)$. The kernel is $\mathbb{Z}$. By the theorem, $\mathbb{R}/\mathbb{Z} \cong S^1$. But $S^1 \cong \mathrm{SO}(2, \mathbb{R})$ by Exercise 5.1.15. □

Exercise 5.7.10. For the additive groups $\mathbb{C}, \mathbb{Z}$ show $\mathbb{C}/\mathbb{Z} \cong \mathbb{C}^\times$.

Theorem 5.7.11 (2nd isomorphism theorem)**.** *Let G be a group, U a subgroup and $N \lhd G$. Then $N \cap U \lhd U$ and $N \lhd NU$ and*

$$U/(N \cap U) \cong (NU)/N.$$

In particular, if G is finite, then $|NU| = |N| \cdot |U|/|N \cap U|$.

Proof. Clearly, $N \cap U$ is a subgroup of U. It is normal: let $u \in U, n \in N \cap U$; then $unu^{-1} \in U$ since U is a subgroup, and $unu^{-1} \in N$ since $N \lhd G$, so $unu^{-1} \in N \cap U$.

NU is a subgroup of G by Exercise 5.6.10. As $N = N1 \subseteq NU$, N is a subgroup of NU; since $N \lhd G$, also $N \lhd NU$.

We define the isomorphism φ setting for $u \in U$:

$$\varphi(u(N \cap U)) := uN.$$

Well-defined: if $u(N \cap U) = u'(N \cap U)$ for $u, u' \in U$, then $u^{-1}u' \in N \cap U$, so $\in N$, so $uN = u'N$, i.e., $\varphi(u(N \cap U)) = \varphi(u'(N \cap U))$. φ is clearly a homomorphism.

Injective: if $u \in U$ and $u(N \cap U) \in \ker(\varphi)$, then $\varphi(u(N \cap U)) = N$, so $uN = N$, so $u \in N$, so $u \in N \cap U$, so $u(N \cap U) = N \cap U = 1_{U/(N \cap U)}$.

Surjective: let $nuN \in NU/N$ where $n \in N, u \in U$; then $nuN = nNu = Nu = uN$ as $N \lhd G$; hence, $\varphi(u(N \cap U)) = nuN$.

In particular, $|U|/|N \cap U| = |U/(N \cap U)| = |NU/N| = |NU|/|N|$. □

Exercise 5.7.12. The 2nd statement holds also for N not normal.

Example 5.7.13. A_4 has order 12 but no subgroup of order 6.

Proof. Assume U is a subgroup of A_4 of order 6. Then $[A_4 : U] = 2$ by Lagrange, so $U \lhd A_4$ by Remark 5.6.9 (5). Recall $K_4' = \{1, (12)(34), (14)(23), (13)(24)\} \lhd A_4$ from Theorem 5.6.27. Then $K_4' \not\subseteq U$ since $|K_4'| \nmid |U|$. Let $\tau \in K_4' \setminus U$. By index 2, $A_4 = U \cup \tau U$. Hence, $A_4 = K_4'U$. By the theorem, $12 = |A_4| = 4 \cdot 6/|U \cap K_4'|$, so $|U \cap K_4'| = 2$. Thus, U contains exactly one of $(12)(34), (14)(23), (13)(24)$. Whichever it is, U is not closed under conjugation by $\sigma := (123)$ – a contradiction to $U \lhd A_4$:

$$\sigma(12)(34)\sigma^{-1} = (23)(14), \quad \sigma(14)(23)\sigma^{-1} = (24)(31), \quad \sigma(13)(24)\sigma^{-1} = (12)(34). \quad \square$$

Example 5.7.14. Let $n > 0$, $G := \mathrm{GL}(n, \mathbb{R}), U := \mathrm{SL}(n, R)$ and

$$N := \mathrm{D}(n, \mathbb{R}) := \{\alpha I_n \mid \alpha \in \mathbb{R}^\times\}.$$

Then $N \lhd G$ and $NU = G$. Further, $N \cap U = \{I_n\}$. Thus $U/(N \cap U) \cong U = \mathrm{SL}(n, R)$. By the theorem, $\mathrm{SL}(n, R) \cong NU/N = \mathrm{GL}(n, \mathbb{R})/\mathrm{D}(n, \mathbb{R})$.

Theorem 5.7.15 (3rd isomorphism theorem)**.** *Let G be a group, $N, N' \lhd G$ and $N \subseteq N'$. Then $N'/N \lhd G/N$ and*

$$(G/N)/(N'/N) \cong G/N'.$$

In particular, if G is finite, $[G : N] = [G : N'] \cdot [N' : N]$.

Proof. $N'/N = \pi_N(N') \lhd \pi_N(G) = G/N$ by Lemma 5.7.1 (2). For $x \in G$ set

$$\psi(xN) := xN'.$$

It is easy to check that this is a well-defined epimorphism from G/N onto G/N'. We have $xN \in \ker(\psi)$ if and only if $xN' = 1_{G/N'} = N'$, i.e., $x \in N'$. Hence $\ker(\psi) = N'/N$. Now apply the 1st isomorphism theorem. □

Exercise 5.7.16. The 2nd statement also holds for N' not normal.

5.8 Solvable groups

Recalling the preface, solvable groups get their name from Galois theory treated in the next chapter. It is a beautiful fact that they can be independently motivated starting from the following question. Can we make a group G abelian by somehow minimally mod-ing out non-commuting elements, that is, find a minimal $N \lhd G$ such that G/N is abelian?

Definition 5.8.1. Let G be a group. The *commutator of* $x, y \in G$ is

$$[x, y] := xyx^{-1}y^{-1}.$$

The *commutator group of* G is the subgroup

$$[G, G] := \langle [x, y] \mid x, y \in G \rangle.$$

Remark 5.8.2. Let G be a group and $x, y \in G$.

1. $xy = yxx^{-1}y^{-1}xy = yx[x^{-1}, y^{-1}]$, $[y, x]xy = yx$, $[x, y]^{-1} = [y, x]$.
2. Hence, $[G, G]$ equals the set of products of commutators.
3. G is abelian if and only if $[G, G] = \{1\}$.

Remark 5.8.3. It is not so easy to find groups G such that $[G, G]$ contains non-commutators. A known example is the free group with 2 generators (see Definition 5.5.8); the smallest example has order 96.

Theorem 5.8.4. *Let G be a group and $N \lhd G$. Then $[G, G] \lhd G$ and, G/N is abelian if and only if $[G, G] \subseteq N$. In particular, the* abelianization $G/[G, G]$ *is abelian.*

Proof. To see $[G, G] \lhd G$, let $x, y, z \in G$. Then

$$z[x, y]z^{-1} = zxyx^{-1}y^{-1}z^{-1} = zxz^{-1}zyz^{-1}\ zx^{-1}z^{-1}\ zy^{-1}z^{-1}\ zz^{-1} = [zxz^{-1}, zyz^{-1}] \in [G, G].$$

By the remark, every $x \in [G, G]$ is a finite product $x = x_1 \cdots x_n$ of commutators x_i. By the above, $zxz^{-1} = zx_1z^{-1}zx_2z^{-1}z \cdots z^{-1}zx_nz^{-1}$ is a product of commutators, so $zxz^{-1} \in [G, G]$.

Assume G/N is abelian. Then $\pi_N([x, y]) = \pi_N(x)\pi_N(y)\pi_N(x)^{-1}\pi_N(y)^{-1} = 1_{G/N}$. Hence, $[x, y] \in N$ for all $x, y \in G$, so $[G, G] \subseteq N$.

Assume $[G, G] \subseteq N$. Then $xNyN = xyN = yx[x^{-1}, y^{-1}]N = yxN = yNxN$. □

Exercise 5.8.5 (Universal property)**.** Let $\varphi : G \to G'$ be a group homomorphism where G' is abelian. Then there is exactly one homomorphism $\bar{\varphi} : G/[G,G] \to G'$ with $\varphi = \bar{\varphi} \circ \pi_{[G,G]}$.

Example 5.8.6. Let $n > 1$. Then $[S_n, S_n] = A_n$.

Proof. As $A_n \lhd S_n$ has index 2, $S_n/A_n \cong C_2$ is abelian, so $[S_n, S_n] \subseteq A_n$ by the theorem.

For $n = 2$, $A_2 = \{1\} \subseteq [S_2, S_2]$. For $n > 2$, $A_n \subseteq [S_n, S_n]$ because A_n is generated by the 3-cycles (ijk) (Example 5.3.7) and $(ijk) = (jk)(ij)(jk)(ij) = [(jk), (ij)] \in [S_n, S_n]$. □

Example 5.8.7. $[A_2, A_2] = [A_3, A_3] = \{1\}$, $[A_4, A_4] \cong K_4$, and $[A_n, A_n] = A_n$ for $n > 4$.

Proof. A_2 is trivial, A_3 has order $3!/2 = 3$, so is isomorphic to C_3 (Proposition 5.3.18), hence abelian. For $n = 4$, consider $K_4' \lhd A_4$ from Example 5.6.12 (5); it has index $(4!/2)/4 = 3$ in A_4, so $A_4/K_4' \cong C_3$ is abelian; by the theorem, $[A_4, A_4] \subseteq K_4'$. Conversely, note K_4' contains (the identity and) products of two disjoint transpositions $(ij)(k\ell)$ and

$$(ij)(k\ell) = (ijk)(ij\ell)(kji)(\ell ji) = [(ijk), (ij\ell)] \in [A_4, A_4].$$

For $n > 4$, it suffices to show every 3-cycle is a commutator of 3-cycles (Example 5.3.7):

$$(123) = (124)(135)(421)(531) = [(124), (135)]. \qquad \square$$

Remark 5.8.8. Let G be a group. Then

$$G^{(0)} := G \ \rhd \ G^{(1)} := [G^{(0)}, G^{(0)}] \ \rhd \ G^{(2)} := [G^{(1)}, G^{(1)}] \ \rhd \ G^{(3)} := [G^{(2)}, G^{(2)}] \ \rhd \cdots$$

and all $G^{(k)}/G^{(k+1)}$ are abelian. If G is finite, this series eventually stabilizes, i.e., $G^{(k)=G^{(k+1)}} = \cdots$ for some k. Does it stabilize with the trivial $\{1\}$ or something bigger?

Definition 5.8.9. A group G is *solvable* if $G^{(k)} = \{1\}$ for some $k \in \mathbb{N}$.

Example 5.8.10. S_2, S_3, S_4 are solvable, and S_n is not solvable for $n > 4$. Same for A_n. In fact, for $G = S_2, S_3, S_4$ and $G = S_n$ with $n > 4$ the sequences $G^{(0)} \rhd G^{(1)} \rhd G^{(2)} \rhd \cdots$ read

$$S_2 \rhd \{1\} \qquad S_3 \rhd A_3 \rhd \{1\} \qquad S_4 \rhd A_4 \rhd K_4' \rhd \{1\} \qquad S_n \rhd A_n \rhd A_n \rhd \cdots$$

Remark 5.8.11.

1. Abelian groups G are solvable ($G^{(1)} = \{1\}$). In particular, groups of prime order are solvable (Proposition 5.3.18). Later we find ourselves able to prove that also groups of order a prime power are solvable (Corollary 5.13.13).
2. Non-abelian simple groups G are not solvable. Using Example 5.6.27, we see again that A_n is not solvable for $n > 4$.

   Indeed, $[G, G] \lhd G$ and $[G, G] \neq \{1\}$ implies $[G, G] = G$, so $G^{(k)} = G$ for all $k \in \mathbb{N}$.
3. Subgroups U of solvable groups G are solvable.

   Say, $G^{(k)} = \{1\}$. By Exercise 5.8.12 below with $\varphi = \mathrm{id}_U : U \to G$ we have $U^{(k)} \subseteq G^{(k)}$, so $U^{(k)} = \{1\}$.

Exercise 5.8.12. Let $\varphi : G \to H$ be a group homomorphism. Then for all $k \in \mathbb{N}$:

$$\varphi(G^{(k)}) = \varphi(G)^{(k)} \subseteq H^{(k)}.$$

Lemma 5.8.13. *Let G be a group and $N \lhd G$. Then G is solvable if and only if both N and G/N are solvable.*

Proof. $\Rightarrow$: assume G is solvable, say $G^{(k)} = \{1\}$. We just remarked, N is solvable. To see G/N is solvable use Exercise 5.8.12 with $\varphi := \pi_N$:

$$(G/N)^{(k)} = \pi_N(G)^{(k)} = \pi_N(G^{(k)}) = \pi_N(\{1_G\}) = \{1_{G/N}\}.$$

$\Leftarrow$: say, $(G/N)^{(k)} = \{1_{G/N}\} = \{N\}$. By Exercise 5.8.12, $\pi_N(G^{(k)}) = \pi_N(G)^{(k)} = (G/N)^{(k)} = \{N\}$ and hence $G^{(k)} \subseteq N$. By Remark 5.8.11 (3), $G^{(k)}$ is solvable. Thus, there is $\ell \in \mathbb{N}$ such that $\{1\} = (G^{(k)})^{(\ell)} = G^{(k+\ell)}$. □

Example 5.8.14. Let $n > 0$. The dihedral group D_n is solvable.

Proof. D_1, D_2 are abelian. For $n > 2$, we have $R \lhd D_n$ for the abelian subgroup of rotations R (of index 2). Both R and D_n/R (order 2) are abelian, so solvable. □

The $G^{(k)}$ of a solvable group G from a so-called *subnormal series*; "sub" because possibly some $G^{(k)} \not\lhd G$. The following is sometimes taken as a definition of solvability.

Theorem 5.8.15. *Let G be a group. Then G is solvable if and only if there exist $\ell \in \mathbb{N}$ and a* subnormal series of G (of length ℓ)

$$\{1\} = N_0 \lhd N_1 \lhd \cdots \lhd N_\ell = G$$

with abelian factors N_{k+1}/N_k *for $k < \ell$.*

Proof. $\Rightarrow$ is trivial. $\Leftarrow$: it suffices to show that $G^{(k)} \subseteq N_{\ell-k}$ for all $k \leqslant \ell$. This is trivial for $k = 0$. Inductively assume $k < \ell$ and $G^{(k)} \subseteq N_{\ell-k}$. Then

$$G^{(k+1)} = [G^{(k)}, G^{(k)}] \subseteq [N_{\ell-k}, N_{\ell-k}] \subseteq N_{\ell-(k+1)},$$

where the last $\subseteq$ follows from Theorem 5.8.4 because $N_{\ell-k}/N_{\ell-(k+1)}$ is abelian. □

Example 5.8.16. We saw the subnormal series $\{1\} \lhd K_4' \lhd A_4$ of A_4 with factors of order 4 and $(4!/2)/4 = 3$. This can be "refined" to $\{1\} \lhd \{1, (12)(34)\} \lhd K_4' \lhd A_4$ with factors of prime orders 2, 2 and 3. Recall, prime order implies abelian (Proposition 5.3.18).

Theorem 5.8.17. *A finite solvable group has a subnormal series with factors of prime order.*

Proof. Let G be a finite solvable group. Then it has a subnormal series of some length ℓ as in the previous theorem. We can assume $N_k \neq N_{k+1}$ for all $k < \ell$. Assume there is $0 < k \leqslant \ell$ such that N_k/N_{k-1} does not have prime order. Lets us assume $k = \ell$ and write $N := N_{\ell-1}$. We claim that there exists $N \lhd N' \lhd G$ such that N'/N has prime order.

The theorem follows: note N'/N is abelian (prime order) and $G/N' \cong (G/N)/(N'/N)$ (3rd isomorphism theorem) is abelian as a factor of an abelian group. Hence, we can just repeat 'inserting' such N' until all factors have prime order. To see this eventually succeeds, note the index drops: $[G : N'] = [G : N]/[N' : N] < [G : N]$.

We are left to prove the claim: as $G \neq N$ we have $[G : N] > 1$. Let p be a prime divisor of $[G : N']$. By Proposition 5.6.13, G/N has a subgroup $\tilde{U}$ of order p. Then $\tilde{U} \lhd G/N$ as G/N is abelian. Let $N' := \pi_N^{-1}(\tilde{U})$. Then $N' \lhd G$ by Lemma 5.7.1 and $\ker(\pi_N) = N \subseteq N'$. Then $N \lhd N'$ and $N'/N \cong \tilde{U}$ by the 1st isomorphism theorem. □

Example 5.8.18. The subnormal series $\{I_2\} \lhd \{\pm I_2\} \lhd \mathrm{SL}(2,\mathbb{R}) \lhd \mathrm{GL}(2,\mathbb{R})$ has a non-abelian factor and cannot be refined: one can show that $\{\pm I_2\}$ is the only non-trivial normal subgroup of $\mathrm{SL}(2,\mathbb{R})$. Instead of $\mathbb{R}$ this holds for every field $\mathbb{F}_q$ of size $q > 3$.

As an exercise, show $\mathrm{GL}(2,\mathbb{F}_q)$ has size $(q^3 - q)(q - 1)$ and $\mathrm{SL}(2,\mathbb{F}_q)$ has size $q^3 - q$.

5.9 Direct products

Products of groups are defined just like products of rings (cf. Lemma 2.5.10):

Definition 5.9.1. Let $r > 1$ and $G_1, ..., G_r$ be groups. Then their *direct product* $G_1 \times \cdots \times G_r$ is the group with $\cdot$ defined for $(x_1, ..., x_r)$, $(y_1, ..., y_r) \in G_1 \times \cdots \times G_r$ as follows:

$$(x_1, ..., x_r) \cdot (y_1, ..., y_r) \quad := \quad (x_1 \cdot y_1, ..., x_r \cdot y_r).$$

If all G_i are the same group G we write G^r for the direct product. For additively written G_i the direct product is written $G_1 \oplus \cdots \oplus G_r$ and called *direct sum*.

Remark 5.9.2.

1. Above $x_i \cdot y_i$ refers to the group operation of G_i. $(G_1 \times \cdots \times G_r, \cdot)$ is a group with neutral element $(1_{G_1}, ..., 1_{G_r})$ and inverses $(x_1, ..., x_r)^{-1} = (x_1^{-1}, ..., x_r^{-1})$.
2. $\times$ is associative and commutative in the sense that $G_0 \times (G_1 \times G_2) \cong (G_0 \times G_1) \times G_2$ and $G_1 \times G_2 \cong G_2 \times G_1$. Sloppily, we shall not notationally distinguish products $G_1 \times \cdots \times G_r$ parenthesized in various ways. E.g., we do not distinguish G^5 and $G^2 \times G^3$.
3. For every $1 \leqslant i \leqslant r$, the *projection* π_i given by $\pi_i(x_1, ..., x_r) := x_i$ is an epimorphism from $G_1 \times \cdots \times G_r$ onto G_i. The kernel is

$$\ker(\pi_i) = \{(x_1, ..., x_r) \mid x_i = 1_{G_i}\} \lhd G_1 \times \cdots \times G_r,$$

and $G_1 \times \cdots \times G_k / \ker(\pi_i) \cong G_i$ (1st isomorphism theorem).

4. For every $1 \leqslant i \leqslant r$, we have a monomorphism from G_i into $G_1 \times \cdots \times G_r$:

$$x \mapsto (1_{G_1}, ..., 1_{G_{i-1}}, x, 1_{G_{i+1}}, ..., 1_{G_r}).$$

The image of this map is a normal subgroup of $G_1 \times \cdots \times G_r$.

5. If $\varphi_i : G_i \cong G_i'$ for all i, then $G_1 \times \cdots \times G_r \cong G_1' \times \cdots \times G_r'$ via

$$(x_1, ..., x_r) \mapsto (\varphi_1(x_1), ..., \varphi_r(x_r)).$$

6. $G_1 \times \cdots \times G_r$ is abelian if and only if every G_i is abelian.

Examples 5.9.3. Our definitions click: $\mathbb{Z}^r$ defined as the additive group of the ring $\mathbb{Z}^r$ equals $\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$ (r times) defined above. In particular, $K_4 = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \cong C_2 \times C_2$.

Also note that Corollary 2.5.11 states $\mathbb{Z}_{nm} \cong \mathbb{Z}_n \oplus \mathbb{Z}_m$ for coprime n, m.

Exercise 5.9.4. Let G_1, G_2 be finite groups of coprime orders. Show every subgroup H of $G_1 \times G_2$ has the form $H_1 \times H_2$ for subgroups H_i of G_i.

Show the coprimality assumption cannot be omitted.

Exercise 5.9.5. Let G_1, G_2 be cyclic groups. When is $G_1 \times G_2$ cyclic?

Exercise 5.9.6. Let $G_1, ..., G_r$ be groups and $N_i \lhd G_i$ for $i = 1, ..., r$. Then

$$N' := N_1 \times \cdots \times N_r \lhd G_1 \times \cdots \times G_r =: G' \quad \text{and} \quad G'/N' \cong G_1/N_1 \times \cdots \times G_r/N_r.$$

Theorem 5.9.7 (Universal property)**.** *Let G_1, G_2 be groups. Then $G_1 \times G_2$ is up to isomorphism the unique group satisfying:*

1. *there are epimorphisms $\varphi_1 : G_1 \times G_2 \to G_1$ and $\varphi_2 : G_1 \times G_2 \to G_2$.*
2. *if G' is a group and $\psi_1 : G' \to G_1$ and $\psi_2 : G' \to G_2$ are homomorphisms, then there is exactly one homomorphism $\Phi : G' \to G_1 \times G_2$ such that $\psi_1 = \varphi_1 \circ \Phi$ and $\psi_2 = \varphi_2 \circ \Phi$.*

Proof. $G_1 \times G_2$ has properties (1) and (2): for (1) take $\varphi_i := \pi_i$. For (2), given G', ψ_1, ψ_2 the equations force to set $\Phi(x) := (\psi_1(x), \psi_2(x))$ for $x \in G'$. This is indeed a homomorphism from G' to $G_1 \times G_2$.

Let H be a group satisfying (1) and (2). Apply (2) for H with $G' := G_1 \times G_2$ and $\pi_i : G' \to G_i$. This gives a homomorphism $\Phi : G' \to H$ with $\pi_i = \varphi_i \circ \Phi$.

We show Φ is bijective. Since $G_1 \times G_2$ satisfies (2) we get $\hat{\Phi} : H \to G_1 \times G_2$ with $\varphi_i = \pi_i \circ \hat{\Phi}$. Then

$$\pi_i \circ \hat{\Phi} \circ \Phi = \varphi_i \circ \Phi = \pi_i$$

It follows that $\hat{\Phi} \circ \Phi = \mathrm{id}_{G_1 \times G_2}$. Hence Φ is surjective. To see injectivity, assume $\Phi(x_1, x_2) = 1_H$ for $(x_1, x_2) \in G_1 \times G_2$; then $x_i = \pi_i(x_1, x_2) = \varphi_i(\Phi(x_1, x_2)) = \varphi_i(1_H) = 1_{G_i}$. Hence, (x_1, x_2) is the neutral element of $G_1 \times G_2$. □

When is G isomorphic to $U_1 \times U_2$ for subgroups U_1, U_2 of G?

Lemma 5.9.8. *Let G be a group and U, V be subgroups satisfying $uv = vu$ for all $u \in U, v \in V$. Then UV is a subgroup of G and the following are equivalent:*

1. *$(u, v) \mapsto uv$ is an isomorphism from $U \times V$ onto UV;*
2. *for every $x \in UV$ there is a unique pair $(u, v) \in U \times V$ with $uv = x$;*
3. *$U \cap V = \{1\}$.*

Proof. UV is a subgroup by Exercise 5.6.10; directly: let $u_0v_0, u_1v_1 \in UV$ with $u_0, u_1 \in U$ and $v_0, v_1 \in V$; then $u_0v_0(u_1v_1)^{-1} = u_0v_0v_1^{-1}u_1^{-1} = u_0v_0u_1^{-1}v_1^{-1} = u_0u_1^{-1}v_0v_1^{-1} \in UV$.

Let φ denote the map in (a); it is a homomorphism:

$$\varphi((u_0, v_0)(u_1, v_1)) = \varphi(u_0u_1, v_0v_1) = u_0u_1v_0v_1 = u_0v_0u_1v_1 = \varphi(u_0, v_0)\varphi(u_1, v_1).$$

φ is obviously surjective. (2) states injectivity, so (1) and (2) are equivalent. We show injectivity, i.e., $\ker(\varphi) = \{(1, 1)\}$ is equivalent to (3).

$\Leftarrow$: let $(u, v) \in \ker(\varphi)$, i.e., $uv = 1$; then $u = v^{-1}$, so both $u \in V$ and $v^{-1} \in U$, so $u, v \in U \cap V = \{1\}$, so $(u, v) = (1, 1)$.

$\Rightarrow$: if $x \in U \cap V$, then $(x, x^{-1}) \in \ker(\varphi) = \{(1, 1)\}$, so $x = 1$. □

Example 5.9.9. S_3 has the subgroups $U := \{1, (12)\}$ and $V := \{1, (13)\}$. Then $UV = \{1, (12), (13), (132)\}$ is not a subgroup of S_3 as $(13)(12) = (123) \notin UV$.

Theorem 5.9.10. *Let G be a group and let $N, N' \lhd G$ satisfy $NN' = G$ and $N \cap N = \{1\}$. Then $(n, n') \mapsto nn'$ is an isomorphism from $N \times N'$ onto G.*

We say G is the inner direct product of N and N'.

Proof. We check the assumption of the lemma: for $n \in N, m \in N'$ we have $nmn^{-1}m^{-1}$ is in $nN = N$ because $mn^{-1}m^{-1} \in N$ by $N \lhd G$. But also $nmn^{-1}m^{-1} \in N'm^{-1} = N'$ because $nmn^{-1} \in N'$ by $N' \lhd G$. Hence, $nmn^{-1}m^{-1} \in N \cap N' = \{1\}$, so $nm = mn$. □

Exercise 5.9.11. $G_1 \times G_2$ is the inner direct product of $G_1' := G_1 \times \{1_{G_2}\}$ and $G_2' := \{1_{G_1}\} \times G_2$. Every group G is the inner product of its trivial subgroups $\{1_G\}$ and G.

Remark 5.9.12. Let $r > 1$. A group G is the *inner direct product of* $N_1, ..., N_r \lhd G$ if $(n_1, ..., n_r) \mapsto n_1 \cdots n_r$ is an isomorphism from $N_1 \times \cdots \times N_r$ onto G.

Similarly to the theorem, one can show that this is the case if and only if $N_1 \cdots N_r = G$, and $N_i \cap N_1 \cdots N_{i-1}N_{i+1} \cdots N_r = \{1\}$ for all $1 \leqslant i \leqslant r$.

Example 5.9.13. That an abelian group $(G, +)$ has $\mathbb{Z}$-basis $(x_1, ..., x_r)$ means that G is the inner direct sum of $\langle x_1 \rangle, ..., \langle x_r \rangle$.

5.10 Semidirect products

Given a group G to analyze, it is good information finding out that every $x \in G$ can be uniquely written $x = u_1u_2$ for elements u_1, u_2 of proper subgroups U_1, U_2. If U_1, U_2 are both normal, Theorem 5.9.10 tells us that G decomposes as $U_1 \times U_2$. What can we say about G if only one of U_1, U_2 is normal? This is what actually happens:

Example 5.10.1. Recall Theorem 5.1.17 and consider D_n for $n > 2$. The subgroup $N := \{R_{k2\pi/n} \mid 1 \leqslant k \leqslant n\}$ is normal and the subgroup $U := \{I_2, S_0\}$ is not (e.g., $R_{2\pi/3} S_0 R_{2\pi/3}^{-1} = R_{4\pi/3} S_0 \notin U$). We saw $NU = D_n$ and $N \cap U = \{I_2\}$.

But D_n is not the inner product of N and U – e.g. use Remark 5.9.2 (4): D_n is not abelian while both N and U are. Or note in an inner product we would have for $k, k' < n$, and $b, b' < 1$ that $R_{k\cdot 2\pi/n} S_0^b \cdot R_{k'\cdot 2\pi/n} S_0^{b'} = R_{(k+k')\cdot 2\pi/n} \cdot S_0^b S_0^{b'}$ which is not true.

Example 5.10.2. Let $n > 0$ and recall Corollary 5.1.11. The group of isometries $\mathrm{I}(n, \mathbb{R})$ has as subgroups the set of translations $N := \mathrm{T}(n, \mathbb{R}) := \{t_a \mid a \in \mathbb{R}^n\}$ and the set of orthogonal linear maps $U := \mathrm{O}(n, R)$ (using the same symbol as for the representing matrices). Then Theorem 5.1.9 states that $\mathrm{I}(n, \mathbb{R}) = NU$ and, obviously, $N \cap U = \{\mathrm{id}_{\mathbb{R}^n}\}$.

N is easily checked to be normal nut U is not: $(t_a)^{-1} \varphi t_a = t_{-a+\varphi(a)} \varphi \notin U$ unless $a = \varphi(a)$. Clearly, $\mathrm{I}(n, \mathbb{R})$ is not the inner product of N and U. This would mean that $(t_a \varphi)(t_b \psi) = t_a t_b \varphi \psi$ – nonsense.

Remark 5.10.3. Assume $G = NU, N \cap U = \{1\}$ for subgroups $N, U \subseteq G$, i.e., every $x \in G$ can be uniquely written as $x = nu$ with $n \in N, u \in U$. If both $N, U \lhd G$ then G is the inner product of N, U and we have

$$nu \cdot n'u' = nn' \cdot uu'.$$

In the examples above only N is normal. We can, however, always write

$$nu \cdot n'u' = n \cdot un'u^{-1} \cdot uu'.$$

Here, n' gets conjugated with u, i.e., moved by the inner automorphism determined by u (cf. Definition 5.1.18). Let's turn this into a definition.

Recall, $(\mathrm{Aut}(G), \circ)$ is the *automorphism group of* a group G (Exercise 1.1.24).

Definition 5.10.4. Let G_1, G_2 be groups, and $\Phi : G_2 \to \mathrm{Aut}(G_1)$ an homomorphism. The *semidirect product of G_1 and G_2 wrt Φ*, denoted

$$G_1 \rtimes_\Phi G_2,$$

is the set $G_1 \times G_2$ together with the operation $\cdot$ given by

$$(x_1, x_2) \cdot (y_1, y_2) := \big(x_1 \Phi_{x_2}(y_1), x_2 y_2\big)$$

for all $(x_1, x_2), (y_1, y_2) \in G_1 \times G_2$; here, we write $\Phi_{x_2} := \Phi(x_2)$.

Remark 5.10.5. If Φ is constantly $\mathrm{id}_{G_1} \in \mathrm{Aut}(G_1)$, then $G_1 \rtimes_\Phi G_2 = G_1 \times G_2$.

Lemma 5.10.6. *Let G_1, G_2 be groups, and $\Phi : G_2 \to \mathrm{Aut}(G_1)$ an homomorphism. Then $G_1 \rtimes_\Phi G_2$ is a group with neutral element $(1_{G_1}, 1_{G_2})$ and for all $(x_1, x_2) \in G_1 \times G_2$:*

$$(x_1, x_2)^{-1} = \big(\Phi_{x_2^{-1}}(x_1^{-1}), x_2^{-1}\big).$$

Proof. For the neutral element note $\Phi_{1_{G_2}} = \mathrm{id}_{G_1}$ and $\Phi_x(1_{G_1}) = 1_{G_1}$ for all $x \in G_2$; hence,

$$(x_1, x_2) \cdot (1_{G_1}, 1_{G_2}) = \big(x_1\Phi_{x_2}(1_{G_1}), x_2 1_{G_2}\big) = (x_1, x_2),$$
$$(1_{G_1}, 1_{G_2}) \cdot (x_1, x_2) = \big(1_{G_1}\Phi_{1_{G_2}}(x_1), 1_{G_2}x_2\big) = (x_1, x_2).$$

For the inverse note $\Phi_x \circ \Phi_{x^{-1}} = \mathrm{id}_{G_1}$ for all $x \in G_2$ and $\Phi_x(y)\Phi_x(y^{-1}) = \Phi_x(yy^{-1}) = \Phi_x(1_{G_1}) = 1_{G_1}$ for all $y \in G_1$; hence,

$$(x_1, x_2) \cdot \big(\Phi_{x_2^{-1}}(x_1^{-1}), x_2^{-1}\big) = \big(x_1\Phi_{x_2}\big(\Phi_{x_2^{-1}}(x_1^{-1})\big), x_2x_2^{-1}\big) = (x_1x_1^{-1}, x_2x_2^{-1}) = (1_{G_1}, 1_{G_2}),$$
$$\big(\Phi_{x_2^{-1}}(x_1^{-1}), x_2^{-1}\big) \cdot (x_1, x_2) = \big(\Phi_{x_2^{-1}}(x_1^{-1})\Phi_{x_2^{-1}}(x_1), x_2^{-1}x_2\big) = (1_{G_1}, 1_{G_2})$$

To verify associativity let $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in G_1 \times G_2$. Then

$$\big((x_1, x_2)(y_1, y_2)\big)(z_1, z_2) = \big(x_1\Phi_{x_2}(y_1), x_2y_2\big)(z_1, z_2) = \big(x_1\Phi_{x_2}(y_1)\Phi_{x_2y_2}(z_1), x_2y_2z_2\big),$$
$$(x_1, x_2)\big((y_1, y_2)(z_1, z_2)\big) = (x_1, x_2)\big(y_1\Phi_{y_2}(z_1), y_2z_2\big) = \big(x_1\Phi_{x_2}(y_1\Phi_{y_2}(z_1)), x_2y_2z_2\big).$$

But $\Phi_{x_2}(y_1)\Phi_{x_2y_2}(z_1) = \Phi_{x_2}(y_1)\Phi_{x_2}(\Phi_{y_2}(z_1)) = \Phi_{x_2}(y_1\Phi_{y_2}(z_1))$. □

Proposition 5.10.7. *Let G_1, G_2 be groups, and $\Phi : G_2 \to \mathrm{Aut}(G_1)$ an homomorphism. Then $G_1 \rtimes_\Phi G_2$ is abelian if and only if both G_1 and G_2 are abelian and Φ is constantly id_{G_1}.*

Proof. $\Rightarrow$: assume $G_1 \rtimes_\Phi G_2$ is abelian. Clearly, G_2 is abelian. For G_1, let $x, y \in G_1$. Then $(x, 1_{G_2}) \cdot (y, 1_{G_2}) = (x\Phi_{1_{G_2}}(y), 1_{G_2}1_{G_2}) = (xy, 1_{G_2})$ since $\Phi_{1_{G_2}} = \mathrm{id}_{G_1}$. By assumption, this equals $(y, 1_{G_2}) \cdot (x, 1_{G_2}) = (y\Phi_{1_{G_2}}(x), 1_{G_2}1_{G_2}) = (yx, 1_{G_2})$. Thus, $xy = yx$.

We show $\Phi_y(x) = x$ for all $y \in G_2, x \in G_1$. Note $(1_{G_1}, y) \cdot (x, 1_{G_2}) = (\Phi_y(x), y)$. Since $G_1 \rtimes_\Phi G_2$ is abelian, this equals $(x, 1_{G_2}) \cdot (1_{G_1}, y) = (x\Phi_{1_{G_2}}(1_{G_1}), y) = (x, y)$.

$\Leftarrow$: by Remarks 5.10.5 and 5.9.2 (4). □

Example 5.10.8. There exists a non-abelian group of order 2025.

Proof. $\mathbb{Z}_{45} \rtimes_\Phi \mathbb{Z}_{45}$ is non-abelian for a non-constant homomorphism $\Phi : \mathbb{Z}_{45} \to \mathrm{Aut}(\mathbb{Z}_{45})$. Does such Φ exist? By Exercise 2.6.5, $\mathrm{Aut}(\mathbb{Z}_{45}) \cong \mathbb{Z}_{45}^\times$ has order $\varphi(3^2 \cdot 5) = 3(3-1)(5-1)$ (Theorem 2.6.10). Proposition 5.6.13 gives $\psi \in \mathrm{Aut}(\mathbb{Z}_{45})$ of order 3 (indeed, $\overline{16}$ has order 3 in $\mathbb{Z}_{45}^\times$). Since $3 \mid 45$ there is a (unique) Φ mapping $\bar{1}$ to ψ. □

Exercise 5.10.9. Construct a non-abelian group of order 2015.

Exercise 5.10.10. Let G_1, G_2 be groups, and $\Phi : G_2 \to \mathrm{Aut}(G_1)$ an homomorphism. Show $N := G_1 \times \{1_{G_2}\}, U := \{1_{G_1}\} \times G_2$ are subgroups of $G_1 \rtimes_\Phi G_2$, N is normal, and $NU = G_1 \rtimes_\Phi G_2$.

Exercise 5.10.11. Let G_1, G_2 be groups, and $\Phi, \Psi : G_2 \to \mathrm{Aut}(G_1)$ be monomorphisms with the same image. Then $G_1 \rtimes_\Phi G_2 \cong G_1 \rtimes_\Psi G_2$.

Exercise 5.10.12. Assume G_1, G_2 be groups, $\Phi : G_2 \to \mathrm{Aut}(G_1)$ an homomorphism, and $\varphi_1 : G_1 \cong G_1', \varphi_2 : G_2 \cong G_2'$. Determine Ψ, ψ with $\psi : G_1 \rtimes_\Phi G_2 \cong G_1' \rtimes_\Psi G_2'$.

Theorem 5.10.13. *Let G be a group, $N \lhd G$ and U a subgroup and assume $NU = G$ and $N \cap U = \{1_G\}$. Then $(n,u) \mapsto nu$ is an isomorphism from $N \rtimes U$ onto G.*

Here, $\rtimes$ stands for $\rtimes_\Phi$ where $\Phi : U \to \mathrm{Aut}(N)$ maps $u \in U$ to conjugation by u, i.e., $\Phi_u : N \to N$ is $n \mapsto unu^{-1}$. We say G is the inner semidirect product of N and U.

Proof. Note Φ_u permutes N because $N \lhd G$, so $\Phi_u \in \mathrm{Aut}(N)$. Let Ψ denote $(n,u) \mapsto nu$. We verify it is a homomorphism: for $(n_1,u_1),(n_2,u_2) \in N \times U$,

$$\Psi\big((n_1,u_1)(n_2,u_2)\big) = \Psi\big(n_1\Phi_{u_1}(n_2), u_1u_2\big) = \Psi\big(n_1u_1n_2u_1^{-1}, u_1u_2\big)$$
$$= n_1u_1n_2u_1^{-1}u_1u_2 = n_1u_1n_2u_2 = \Psi(n_1,u_1)\Psi(n_2,u_2).$$

Ψ is surjective by $NU = G$. Injective: if $\Psi(n,u) = nu = 1_G$, then $n = u^{-1} \in U$, so $n \in N \cap U = \{1_G\}$. Hence, $n = 1_G$ and $u = nu = 1_G$, so $(n,u) = (1_G, 1_G)$. □

Example 5.10.14. Recall Example 5.10.2 and let $n > 0$. Then $\mathrm{I}(n,\mathbb{R})$ is the inner semidirect product of $\mathrm{T}(n,\mathbb{R})$ and $\mathrm{O}(n,\mathbb{R})$.

Example 5.10.15. Recall Example 5.10.1 and let $n > 1$. D_n is the inner semidirect product of $N := \{R_{k2\pi/n} \mid k = 1,...,n\}$ and $U := \{I_2, S_0\}$. Since $N \cong C_n$ and $U \cong C_2$ we get

$$D_n \cong C_n \rtimes_\Phi C_2$$

for a certain $\Phi : C_2 \to \mathrm{Aut}(C_n)$ by Exercise 5.10.12. Concretely: note $S_0R_{k2\pi/n}S_0^{-1} = R_{k2\pi/n}^{-1}$, i.e., conjugation with S_0 gives the inverse in N. Since $N \cong C_n$ via $R_{k2\pi/n} \mapsto \zeta_k$ and $U \cong C_2$ via $I_2, S_0 \mapsto 1, -1$ we can define Φ setting $\Phi_1 := \mathrm{id}_{C_n}$ and Φ_{-1} to be $x \mapsto x^{-1}$.

If $n = 2$, then $x = x^{-1}$, so Φ is constant and $D_2 \cong C_2 \times C_2 \cong K_4$ (Remark 5.10.5).

Exercise 5.10.16. For $n > 1$, S_n is the inner semidirect product of A_n and $\{1, (12)\}$.

Remark 5.10.17. To express that G is the inner (semi)direct product of N, U it is unfortunately quite common to use the notations $G = N \rtimes U$ and $G = N \times U$. This notation creates a confusing ambiguity of the equality symbol: the same statement also expresses that G equals a set of pairs, namely the product of N, U according to Definitions 5.9.1, 5.10.4. For distinction it is common to express the latter saying G is the *outer* (semi)direct product of N, U. But thereby one only escalates the confusion because only one (semi)direct product operation is defined. We avoid all these notations and modes of speech.

5.11 Finitely generated abelian groups

In this section we use additive notation for groups unless stated otherwise.

Definition 5.11.1. Let G be an abelian group. The *torsion subgroup of G* is

$$T(G) := \big\{x \in G \mid \mathrm{ord}(x) \neq \infty\big\}.$$

G is a *torsion group* if $T(G) = G$. It is *torsion-free* if $T(G) = \{0\}$ (cf. Exercise 5.4.2).

Remark 5.11.2. $T(G)$ is a subgroup of G and $G/T(G)$ is torsion-free.

Proof. 1st statement: let $x, y \in T(G)$ have orders $n, m \in \mathbb{N}$, i.e., $nx = 0, my = 0$. Then $m(-y) = 0$ and $nm(x-y) = 0$, so $x - y \in T(G)$.

2nd statement: assume $x + T(G) \in G/T(G)$ has order $n \in \mathbb{N}$; then $T(G) = n(x + T(G)) = nx + T(G)$, so $nx \in T(G)$, say $mnx = 0$, so $x \in T(G)$, so $x + T(G) = T(G)$. □

Example 5.11.3.

1. Above it is crucial that G is abelian: in $\mathrm{I}(2, \mathbb{R})$ (multiplicative notation), s_0 is the reflection at line $y = 0$, and $s_1 := t_{(0,1)} s_0 t_{(0,1)}^{-1}$ the reflection at line $y = 1$. Then s_0, s_1 have order 2 and $s_0 s_1 = t_{-(0,2)}$ has infinite order.
2. Let $n > 1$. $\mathbb{Z}^n$ is torsion-free, $\mathbb{Z}_n$ is a torsion group.
3. Let $r, s \in \mathbb{N}$ and $n_1, ..., n_s > 1$. Then (forgetting about parentheses)
$$T(\mathbb{Z}^r \oplus \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_s}) = \{0\}^r \oplus \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_s} \cong \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_s}.$$
4. $(\mathbb{Q}/\mathbb{Z}, +)$ is a torsion group with infinite exponent.

   Indeed: $b \cdot (a/b + \mathbb{Z}) = 0 + \mathbb{Z}$ for $a, b \in \mathbb{Z}, b > 0$, so $\mathrm{ord}(a/b + \mathbb{Z}) \leqslant b$, and $\mathrm{ord}(1/b + \mathbb{Z}) = b$.

Kronecker classified finite abelian groups 1870, generalizing a result from Gauß' *Disquisitiones Arithmeticae* 1801. The following is due to Poincaré 1900:

Theorem 5.11.4 (Classification of finitely generated abelian groups)**.** *Let G be a finitely generated abelian group. Then there exists unique naturals $r, s \in \mathbb{N}$ and* invariant factors *$d_1, ..., d_s > 1$ with $d_i \mid d_{i+1}$ for all $i < s$ such that*
$$G \cong \mathbb{Z}^r \oplus \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}.$$
Moreover, d_s is the exponent of $T(G)$ and r is the rank of the finitely generated free abelian group $G/T(G)$, called rank *of G.*

Proof. If G is generated by a set of cardinality t, then there is a surjective homomorphism from $\mathbb{Z}^t$ onto G (Remark 5.3.4 (2)). Let U be its kernel, so $G \cong \mathbb{Z}^t/U$ by the 1st isomorphism theorem. By Theorem 5.4.9, there is a $\mathbb{Z}$-basis $\bar{x} = (x_1, ..., x_t)$ of $\mathbb{Z}^t$ and $s \leqslant t$ and positive $d_1 \mid \cdots \mid d_s$ such that $(d_1 x_{t-s+1}, ..., d_s x_t)$ is a $\mathbb{Z}$-basis of U.

The coordinate map wrt to $\bar{x}$ (Remark 5.4.4) maps U onto $V := \{0\}^r \oplus d_1\mathbb{Z} \oplus \cdots \oplus d_s\mathbb{Z}$ where $r := t - s$. Then $G \cong \mathbb{Z}^t/V$, so by Exercise 5.9.6
$$G \cong \mathbb{Z}^r \oplus \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s} =: G'.$$
We can assume $d_1 > 1$. Then $T(G') = \{0\}^r \oplus \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}$. Thus, $G/T(G) \cong G'/T(G') \cong \mathbb{Z}^r$ (Exercise 5.9.6) is free of rank r. Since $T(G) \cong \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}$ and $d_1 \mid \cdots \mid d_s$, we have $d_s x = 0$ for all $x \in T(G)$ satisfy $d_s x = 0$. Since $(0, ..., 0, 1)$ has order d_s, we have $\exp(T(G)) = d_s$.

We show that r, s and the d_i are unique. Assume $G \cong \mathbb{Z}^{r'} \oplus \mathbb{Z}_{e_1} \oplus \cdots \oplus \mathbb{Z}_{e_{s'}}$ where $e_1 \mid \cdots \mid e_{s'}$ are > 1. Then $r = r' =$ rank of $G/T(G)$. Hence $\mathbb{Z}_{e_1} \oplus \cdots \oplus \mathbb{Z}_{e_{s'}} \cong \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}$. Then $e_{s'} = d_s = \exp(T(G))$ and the sums without the last factor are isomorphic; they have exponent d_{s-1} and $e_{s'-1}$, so $d_{s-1} = e_{s'-1}$. Continuing like this gives $s = s'$ and $d_i = e_i$. □

Corollary 5.11.5. *Let $n, d > 1$ and G an abelian group of order n. If $d \mid n$, then G has a subgroup of order d.*

Proof. We can assume $G = \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}$ where $d_1 \mid \cdots \mid d_s$. Then we can write $d = e_1 \cdots e_s$ with $e_i \mid d_i$. By Theorem 5.3.21 (1), there is a subgroup U_i of $\mathbb{Z}_{d_i}$ of order e_i. Then $U_1 \oplus \cdots \oplus U_s$ is a subgroup of G of order d. □

Example 5.11.6. Above, the subgroup is not necessarily unique: $\{0\} \oplus \mathbb{Z}_2$ and $\mathbb{Z}_2 \oplus \{0\}$ are distinct subgroups of $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

2nd proof of Corollary 5.3.25. In the classification for G we have $r = 0$ as G is finite. Then $d_s = \exp(T(G)) = \exp(G)$. Clearly, $x^{d_s} = 1$ for all $x \in G$. In other words, every $x \in G$ is a root of $X^{d_s} - 1 \in K[X]$. Since there are $\leqslant d_s$ roots, $|G| \leqslant d_s$. By Exercise 5.3.27, G contains an element of order d_s, so G is cyclic. □

Exercise 5.11.7. Let G be a finitely generated abelian group. Then G is free if and only if it is torsion-free. G is finite if and only if it is a torsion group.

Exercise 5.11.8. Subgroups of finitely generated abelian groups are finitely generated.

Remark 5.11.9 (Burnside problem)**.** One defines non-abelian torsion groups analogously. Burnside asked 1902 whether all finitely generated torsion groups are finite. In 1964 Golod and Shafarevich finally answered "no". Even finitely generated groups with finite exponent can be infinite (Novikov, Adian 1968), and so-called *Tarski monsters* are striking examples (Olshanskii 1979). Which finite exponents can appear is not fully understood.

5.11.1 Finite abelian groups

Finite abelian groups of a given order are products of certain $\mathbb{Z}_q$'s. Which ones? The answer is not much more than notational warfare:

Definition 5.11.10. A *partition of* $k \in \mathbb{N}$ is a tuple $\bar{\ell} = (\ell_1, ..., \ell_s) \in \mathbb{N}^r$ for some $r > 0$ such that $0 < \ell_1 \leqslant ... \leqslant \ell_s$ and $k = \ell_1 + \cdots + \ell_s$. A *partition of* $(k_1, ..., k_r) \in \mathbb{N}^r$ for some $r > 0$ is a tuple $\pi := (\bar{\ell}_1, ..., \bar{\ell}_r)$ such that $\bar{\ell}_i$ is a partition of k_i.

For $n > 1$ with prime factorization $p_1^{k_1} \cdots p_r^{k_r}$ let $\mathcal{P}(n)$ be the set of partitions of $(k_1, ..., k_r)$. For $\pi = (\bar{\ell}_1, ..., \bar{\ell}_r) \in \mathcal{P}(n)$ and writing $\bar{\ell}_i = (\ell_{i1}, ..., \ell_{is_i})$ define the group

$$G_\pi := \mathbb{Z}_{p_1^{\ell_{11}}} \oplus \cdots \oplus \mathbb{Z}_{p_1^{\ell_{1s_1}}} \oplus \ \cdots \cdots \ \oplus \mathbb{Z}_{p_r^{\ell_{r1}}} \oplus \cdots \oplus \mathbb{Z}_{p_r^{\ell_{rs_r}}}.$$

Theorem 5.11.11 (Classification of finite abelian groups)**.** *Let $n > 1$. Every abelian group of order n is isomorphic to G_π for exactly one $\pi \in \mathcal{P}(n)$.*

Proof. Let $n = p_1^{k_1} \cdots p_r^{k_r}$ be the prime factorization. We first show that the G_π are pairwise non-isomorphic. Assume $G_\pi \cong G_{\pi'}$ for $\pi, \pi' \in \mathcal{P}(n)$. Write $\pi = (\bar{\ell}_1, ..., \bar{\ell}_s), \pi' = (\bar{\ell}'_1, ..., \bar{\ell}'_{s'})$ and $\bar{\ell}_i = (\ell_{i1}, ..., \ell_{is_i})$ and $\bar{\ell}'_i = (\ell'_{i1}, ..., \ell'_{is'_i})$. Abbreviate

$$q_{ij} := p_i^{\ell_{ij}}, \quad q'_{ij} := p_i^{\ell'_{ij}}.$$

Let $1 \leqslant i \leqslant r$. We have to show $s_i = s'_i$ and $\bar{\ell}_i = \bar{\ell}'_i$. The elements of G_π whose order is a power of p_i is a subgroup (use Lemma 5.3.9 (4)) isomorphic to $G_i := \mathbb{Z}_{q_{i1}} \oplus \cdots \oplus \mathbb{Z}_{q_{is_i}}$. The isomorphism from G_π onto $G_{\pi'}$ maps this subgroup onto the elements of $G_{\pi'}$ whose order is a power of p_i. This is isomorphic to $G'_i := \mathbb{Z}_{q'_{i1}} \oplus \cdots \oplus \mathbb{Z}_{q'_{is'_i}}$. Since $\ell_{ij}, j \leqslant s_i$, is nondecreasing, $\exp(G_i) = q_{is_i}$. Similarly, $\exp(G'_i) = q'_{is'_i}$. By isomorphism, $\ell_{is_i} = \ell'_{is'_i}$. Thus, deleting the last factors, the groups stay isomorphic. Continuing gives $s_i = s'_i$ and $\bar{\ell}_i = \bar{\ell}'_i$.

We now show every abelian group G of order n is isomorphic to G_π for some $\pi \in \mathcal{P}(n)$. We know $G \cong \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}$ where the d_j are the invariant factors. Write $d_j = p_1^{\ell_{1j}} \cdots p_r^{\ell_{rj}}$ for certain $\ell_{1j}, \ldots, \ell_{rj} \geqslant 0$ and primes $p_1, \ldots, p_r$. Set

$$F(i,j) := \begin{cases} \mathbb{Z}_{p_i^{\ell_{ij}}} & \text{if } \ell_{ij} \neq 0, \\ \{0\} & \text{if } \ell_{ij} = 0. \end{cases}$$

By Corollary 2.5.11, $\mathbb{Z}_{d_j} \cong F(1,j) \oplus \cdots \oplus F(r,j)$. Hence G is isomorphic to the direct sum of the direct sums of the columns of

$$\begin{matrix} F(1,1) & \cdots & F(1,s) \\ \vdots & \ddots & \vdots \\ F(r,1) & \cdots & F(r,s) \end{matrix}$$

Shuffling factors, G is isomorphic to the direct sum of the direct sums of the rows. If we omit factors $\{0\}$ (namely $F(i,j)$ with $\ell_{ij} = 0$) this product equals G_π for some $\pi \in \mathcal{P}(n)$. Indeed: write $\bar{\ell}_i := (\ell_{i1}, \ldots, \ell_{is})$; since $n = d_1 \cdots d_s$ we have $k_i = \ell_{i1} + \cdots + \ell_{is}$; further, $\ell_{i1} \leqslant \cdots \leqslant \ell_{is}$ because $d_1 \mid \cdots \mid d_s$. Take π to be $(\bar{\ell}_1, \ldots, \bar{\ell}_r)$ with all $\ell_{ij} = 0$ deleted. □

Exercise 5.11.12 (Interpretation of the s_i)**.** Let G be an abelian group, additively written, and $p, n \in \mathbb{N}$ with p prime. Let $\mathbb{F}_p := \mathbb{Z}_p$ denote the p-element field. Let $pG := \{px \mid x \in G\}$.

1. G/pG is in a natural way a vector space over $\mathbb{F}_p$.
2. If G is finite and $p \nmid |G|$, then G/pG is trivial.
3. If $p \mid n$, then $\mathbb{Z}_n/p\mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_p$.
4. If $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_s}$ for certain $n_i > 0$, then the dimension of the vector space G/pG equals the number of n_i that are divisible by p.

Corollary 5.11.13 (Primary decomposition)**.** *Let G be a finitely generated abelian group. Then*

$$G \cong \mathbb{Z}^r \oplus \mathbb{Z}_{q_1} \oplus \cdots \oplus \mathbb{Z}_{q_t}$$

for unique naturals $r, t \in \mathbb{N}$ and some sequence $(q_1, \ldots, q_t)$ of (not necessarily distinct) prime powers; the sequence is unique up to re-indexing.

Proof. Replace in Theorem 5.11.4 the factor $\mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_s}$ by the isomorphic G_π. This has the required form. Any presentation of the desired form has $\mathbb{Z}_{q_1} \oplus \cdots \oplus \mathbb{Z}_{q_t}$ isomorphic to G_π, so the factors $\mathbb{Z}_{q_i}$ are those of G_π. □

Exercise 5.11.14. Every simple finite abelian group is isomorphic to $\mathbb{Z}_p$ for a prime p.

Corollary 5.11.15. *Let $n > 1$ have prime factorization $n = p_1^{k_1} \cdots p_r^{k_r}$. For $k \in \mathbb{N}$ let $p(k)$ be the number of partitions of k. Then there are exactly $p(k_1) \cdots p(k_r)$ many abelian groups of order n up to isomorphism.*

Remark 5.11.16. It is known that the *partition function* $p(k)$ satisfies

k	1	2	3	4	5	6	7	8	9	10	15	20	25	30	100
$p(k)$	1	2	3	5	7	11	15	22	30	42	176	627	1958	5604	190569292

$p(1000)$ has 32 digits. Asymptotically we have $p(k) \sim \dfrac{1}{4k\sqrt{3}} \cdot e^{\pi\sqrt{2k/3}}$.

Example 5.11.17. There are exactly $p(3)p(5) = 3 \cdot 7 = 21$ pairwise non-isomorphic abelian groups of order $25000 = 2^3 \cdot 5^5$ and $p(5)p(5) = 49$ of order $100000 = 2^5 \cdot 5^5$.

Recall Example 5.6.19. We now complete the list of all 5 groups of order 8.

Example 5.11.18. List all abelian groups of order 8 (up to isomorphism).

Solution: $8 = 2^3$. Partitions of 3: $(3), (1,2), (1,1,1)$. This gives the groups

$$\mathbb{Z}_{2^3}, \quad \mathbb{Z}_{2^1} \oplus \mathbb{Z}_{2^2}, \quad \mathbb{Z}_{2^1} \oplus \mathbb{Z}_{2^1} \oplus \mathbb{Z}_{2^1}.$$

Example 5.11.19. List all abelian groups of order 100 (up to isomorphism).

Solution: $100 = 2^2 \cdot 5^2$. Partitions of 2: $(2), (1,1)$. This gives the groups

$$\mathbb{Z}_{2^2} \oplus \mathbb{Z}_{5^2}, \quad \mathbb{Z}_{2^2} \oplus \mathbb{Z}_{5^1} \oplus \mathbb{Z}_{5^1}, \quad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_{5^2}, \quad \mathbb{Z}_{2^1} \oplus \mathbb{Z}_{2^1} \oplus \mathbb{Z}_{5^1} \oplus \mathbb{Z}_{5^1},$$

that is, using Corollary 2.5.11: $\mathbb{Z}_{100}$, $\mathbb{Z}_5 \oplus \mathbb{Z}_{20}$, $\mathbb{Z}_2 \oplus \mathbb{Z}_{50}$, $\mathbb{Z}_{10} \oplus \mathbb{Z}_{10}$.

Exercise 5.11.20. Find the smallest $n \in \mathbb{N}$ such that there are exactly 6 abelian groups of order n.

5.12 Group actions

Understanding a group requires understanding its subgroups. Given a finite group G, it only has subgroups of order dividing $|G|$ (Lagrange). For cyclic groups we find a unique subgroup of order d for every $d \mid n$ (Theorem 5.3.21). For abelian groups we find a not necessarily unique one (Corollary 5.11.5, Example 5.11.6). In general, subgroups of order d might not exist (Example 5.7.13). The following strengthens Proposition 5.6.13.

Theorem 5.12.1 (Cauchy)**.** *Let G be a finite group and p a prime divisor of $|G|$. Then G contains a subgroup of order p.*

The proof is based on a new perspective on groups:

Definition 5.12.2. Let G be a group and $X \neq \varnothing$ a set.

1. An *action of G on X* is a map $\tau : G \times X \to X$ such that $\tau_{gh}(x) = \tau_g(\tau_h(x))$ and $\tau_1(x) = x$ for all $g, h \in G, x \in X$; here, $\tau_g(x) := \tau(g, x)$. We say G *acts on X by τ*.
2. For $x \in X$, the *stabilizer of $x \in X$ (under τ)* is $G_x := \{g \in G \mid \tau_g(x) = x\}$.
3. The *orbit of $x \in X$ (under τ)* is $G(x) := \{\tau_g(x) \in X \mid g \in G\}$.
4. The set of *fixed-points (of τ)* is $X^G := \{x \in X \mid \tau_g(x) = x \text{ for all } g \in G\}$.
5. The action is *faithful* if for all $g, h \in G$: if $\tau_g(x) = \tau_h(x)$ for all $x \in X$, then $g = h$.
6. The action is *transitive* if for all $x, y \in X$ there is $g \in G$ with $\tau_g(x) = y$.

Remark 5.12.3. Let τ an action of G on X.

1. For every $g \in G$, τ_g is a permutation of X because $\tau_{g^{-1}} \circ \tau_g = \tau_g \circ \tau_{g^{-1}} = \mathrm{id}_X$; e.g., $\tau_g(\tau_{g^{-1}}(x)) = \tau_{gg^{-1}}(x) = \tau_1(x) = x$.
2. The definition means that $g \mapsto \tau_g$ is a homomorphism from G into $\mathrm{Sym}(X)$. Being faithful means it is injective.
3. For every $x \in X$, the stabilizer G_x is a subgroup of G.

   Indeed: if $g, h \in G_x$ then $\tau_{gh^{-1}}(x) = \tau_{gh^{-1}}(\tau_h(x))) = \tau_g(\tau_{h^{-1}}(\tau_h(x))) = \tau_g(x) = x$.

Examples 5.12.4. Let $n > 0$.

1. $\mathrm{GL}(n, \mathbb{R})$ acts on $\mathbb{R}^n$ via $(A, x) \mapsto Ax$. This action is faithful (Ae_i is the i-th column of A) and not transitive (as 0 is a fixed-point).
2. $\mathrm{I}(n, \mathbb{R})$ acts on the power set of $\mathbb{R}^n$ by $(f, F) \mapsto f(F)$ (where f is an isometry and $F \subseteq \mathbb{R}^n$). The stabilizer of F is the set of symmetries of F (Definition 5.1.13).
3. Let $n > 2$. Each symmetry of the regular n-gon permutes the n vertices and is determined by this permutation (cf. Example 5.1.21). Numbering the vertices $1, \ldots, n$ we get a faithful action of D_n on $\{1, \ldots, n\}$. Remark 5.12.3 (2) shows D_n is isomorphic to a subgroup of S_n. Examples 5.2.6 and 5.6.15 spelled this out for D_3 and D_4.
4. Let K be a field. Then $(K[X], +)$ acts on K via $(f, x) \mapsto f(x)$. If K is infinite, the action is faithful (Exercise 3.3.5). If $K = \mathbb{F}_p$ for a prime p, the action is not faithful (Proposition 3.3.4).
5. Let K a field. The symmetric group S_n acts faithfully on $K[X_1, \ldots, X_n]$ by $(\sigma, f) \mapsto f^\sigma$ as of Definition 3.7.1. The fixed-points are the symmetric polynomials.
6. Let K be a group, ring or field. A subgroup Φ of $\mathrm{Aut}(K)$ faithfully acts on K by $(\varphi, x) \mapsto \varphi(x)$. The fixed-points are K^Φ as of Definition 3.7.4.
7. A group G acts on G by *left translation* $(g, x) \mapsto gx$ and also by *right translation* $(g, x) \mapsto xg^{-1}$. Both actions are faithful and transitive. Remark 5.12.3 (2) implies G is isomorphic to a subgroup of $\mathrm{Sym}(G)$ – this is Proposition 5.2.3.

Exercise 5.12.5. C_2 and $\mathbb{R}_{>0}$ act on $\mathbb{C}$ via $\begin{matrix}(1, z) \mapsto z \\ (-1, z) \mapsto \bar{z}\end{matrix}$ and $(r, z) \mapsto rz$. Are these faithful? What are the fixed-points? Find $T \subseteq \mathbb{C}$ containing exactly one element per orbit.

Exercise 5.12.6. Let τ be an action of G on X. Define $\sim_\tau \subseteq X^2$ by

$$x \sim_\tau y \iff \tau_g(x) = y \text{ for some } g \in G.$$

Show this is an equivalence relation and the equivalence classes are the orbits.

Lemma 5.12.7 (Orbit-stabilizer lemma)**.** *Let τ be an action of G on X and assume G is finite. Then for every $x \in X$:*

$$|G(x)| = [G : G_x].$$

Proof. Define $f : G(x) \to G/G_x$ setting $f(y) := gG_x$ where $g \in G$ is such that $\tau_g(x) = y$.

Well-defined: if $\tau_g(x) = \tau_h(x)$, then $g^{-1}h \in G_x$, so $gG_x = hG_x$.

Injective: assume $f(y) = f(z)$ for $y, z \in G_x$, say, $f(y) = gG_x, f(z) = hG_x$ where $\tau_g(x) = y$ and $\tau_h(x) = z$; then $g^{-1}h \in G_x$, i.e., $\tau_{g^{-1}h}(x) = x$, so $\tau_g(x) = \tau_h(x)$, i.e., $y = z$.

Surjective: if $g \in G$, then $f(\tau_g(x)) = gG_x$. □

Lemma 5.12.8. *Let τ be an action of G on X and assume G, X are finite. Let T contain exactly one element from each orbit.*

1. (Bahngleichung) $|X| = \sum_{x \in T} [G : G_x]$.
2. (Burnside's lemma) $|T| = \sum_{g \in G} |\{x \in X \mid \tau_g(x) = x\}|/|G|$.
3. (Fixed-point lemma) *If $|G| > 1$ is a power of a prime p, then $|X| \equiv |X^G| \mod p$.*

Proof. (1): by the exercise the orbits partition X, so $|X| = \sum_{x \in T} |G(x)|$. Then apply the previous lemma. (2): note

$$\sum_{g \in G} |\{x \in X \mid \tau_g(x) = x\}| = |\{(g, x) \mid \tau_g(x) = x\}| = \sum_{x \in X} |G_x|.$$

By the previous lemma and Lagrange, $|G(x)| = |G|/|G_x|$, so $\sum_{x \in X} |G_x| = |G| \sum_{x \in X} 1/|G(x)|$. Since $G(x), x \in T$, partitions X, this sum equals

$$|G| \sum_{x \in T} \sum_{y \in G(x)} 1/|G(x)| = |G| \sum_{x \in T} 1 = |G| \cdot |T|.$$

(3): Let $x \in T^* := T \setminus X^G$. By the orbit-stabilizer lemma, $|G(x)| = [G : G_x] > 1$ and by Lagrange $[G : G_x] \mid |G|$, a power of p. Thus, $p \mid [G : G_x]$.

For $x \in T \setminus T^*$ we have $G_x = G$, so $[G : G_x] = 1$. Thus $p \mid |X| - |X^G|$ because, by (1), $|X| = |X^G| + \sum_{x \in T^*} [G : G_x]$. □

Remark 5.12.9. Burnsides lemma states that the the number of orbits equals the expected value of the number of points fixed by $g \in G$ chosen uniformly at random.

Exercise 5.12.10. How many length 6 necklaces can you design with 10 beads? Cutting the necklace gives a string, e.g., 122183, which is "the same" as 831221 (cut at a different place). Formalize "the same" by an action of C_6 on $\{1..., 10\}^6$ and use Burnside's lemma.

Exercise 5.12.11. If G acts on $X, |G| = 55, |X| = 19$, then there are at least 3 fixed points.

Definition 5.12.12. G acts on G by *conjugation* $(g,x) \mapsto gxg^{-1}$. The orbit of $x \in G$ is called *conjugacy class of* x, and the stabilizer of x is called *centralizer of* x and denoted $\mathrm{Z}(x)$.

The *center of* G is

$$\mathrm{Z}(G) := \bigcap_{x\in G} \mathrm{Z}(x) = \{g \in G \mid gx = xg \text{ for all } x \in G\}.$$

Theorem 5.12.13. *Let G act on G by conjugation.*

1. *$\mathrm{Z}(G)$ is the set of fixed-points.*
2. *$\mathrm{Z}(G)$ is an abelian normal subgroup of G.*
3. (Class equation) *If G is finite and T contains exactly one element from every orbit of size $\geqslant 2$, then*
$$|G| = |\mathrm{Z}(G)| + \textstyle\sum_{x\in T}[G : \mathrm{Z}(x)].$$

Proof. (1): $g \in \mathrm{Z}(G)$ means $g = xgx^{-1}$ for all $x \in G$. (2): $\mathrm{Z}(G)$ is a subgroup as an intersection of subgroups. If $g \in \mathrm{Z}(G)$ and $x \in G$, then $xgx^{-1} = g \in \mathrm{Z}(G)$, so $\mathrm{Z}(G)$ is normal. If $g,h \in \mathrm{Z}(G)$, then $hgh^{-1} = g$, so $hg = gh$; hence, $\mathrm{Z}(G)$ is abelian.

(3) follows from (1) and the Bahngleichung. □

Proof of Cauchy's Theorem 5.12.1. Induction on $|G|$. Choose T as in the class equation. By Lagrange, $|G| = |\mathrm{Z}(x)|\cdot[G : \mathrm{Z}(x)]$ for all $x \in T$; since p is prime, $p \mid |\mathrm{Z}(x)|$ or $p \mid [G : \mathrm{Z}(x)]$.

Assume $p \mid [G : \mathrm{Z}(x)]$ for all $x \in T$. By the class equation, $p \mid |\mathrm{Z}(G)|$. As $\mathrm{Z}(G)$ is abelian, Proposition 5.6.13 gives an order p subgroup of $\mathrm{Z}(G)$, hence of G.

Assume $p \mid |\mathrm{Z}(x)|$ for some $x \in T$. By the orbit-stabilizer lemma, $[G : \mathrm{Z}(x)] = |G(x)| \geqslant 2$. By Lagrange, $|\mathrm{Z}(x)| < |G|$. By induction, $\mathrm{Z}(x)$ and hence G has a subgroup of order p. □

Definition 5.12.14. G acts on the set of its subgroups $\mathcal{U}$ by *conjugation*: $(x,U) \mapsto xUx^{-1}$. $U,V \in \mathcal{U}$ are *conjugate* if they are in the same orbit, i.e., $gUg^{-1} = V$ for some $g \in G$.

The stabilizer of U is called *normalizer of* U and denoted $\mathrm{N}(U)$, i.e.,

$$\mathrm{N}(U) = \{g \in G \mid gUg^{-1} = U\}.$$

Proposition 5.12.15. *Let U be a subgroup of G. $\mathrm{N}(U)$ is the largest subgroup V of G such that $U \lhd V$, i.e., it contains all such V. In particular, $U \lhd G$ if and only if $\mathrm{N}(U) = G$.*

Proof. $\mathrm{N}(U)$ is a subgroup of G by Remark 5.12.3 (4), and $U \lhd \mathrm{N}(U)$ by definition. Let V be a subgroup of G with $U \lhd V$. Then $gUg^{-1} = U$ for all $g \in V$, so $V \subseteq \mathrm{N}(U)$. □

Here is a more clever use of some group operation:

Proposition 5.12.16. *Let G be a finite group, p be the smallest prime divisor of $|G|$ and U a subgroup of index p. Then U is normal.*

Proof. Let G act on G/U via $(g, xU) \mapsto gxU$. Remark 5.12.3 (1) gives an homomorphism $\varphi : G \to \mathrm{Sym}(G/U)$, namely $\varphi(g)$ is the map $xU \mapsto gxU$. Then $N := \ker(\varphi) \subseteq U$: if $g \notin U$, then $gU \neq U$, so $\varphi(g) \neq \mathrm{id}_{G/U}$. By Lemma 5.7.1 (1), $N \lhd G$. By Exercise 5.7.16,

$$k := [G : N] = [G : U] \cdot [U : N] = p \cdot [U : N],$$

By the 1st isomorphism theorem, $G/N \cong \varphi(G)$, so $|\varphi(G)| = k \geqslant p$. But $\varphi(G)$ is a subgroup of $\mathrm{Sym}(G/U) \cong S_p$, so $k \mid p!$ by Lagrange. As $k \mid |G|$ we have $p = k$ by assumption on p. Then $[U : N] = 1$, so $U = N$. Thus $U \lhd G$. □

Exercise 5.12.17. Let τ be a transitive operation of G on X. Assume G, X are finite, $|X| > 1$ and $\{1\} \neq N \lhd G$.

(a) There is $g \in G$ such that τ_g has no fixed-points.

(b) The orbits under (the restriction of the operation to) N have all the same size.

Assume $p := |X|$ is prime and τ is faithful.

(c) G has an element of order p. Further, N operates transitively and faithfully.

(d) If G is solvable and $N_0 = \{1\} \lhd N_1 \lhd \cdots \lhd N_k = G$ a subnormal series with factors of prime order (cf. Theorem 5.8.17), then $N_1 \cong C_p$.

Exercise 5.12.18. Let G be a simple, non-abelian group with a subgroup of index $n \geqslant 2$. Show G is isomorphic to a subgroup of A_n and $n \geqslant 5$.

5.13 Sylow's theorems

Definition 5.13.1. Let p be a prime. A group G is a *p-group* if for every $x \in G$ there is $n \in \mathbb{N}$ such that $\mathrm{ord}(x) = p^n$.

Lemma 5.13.2. *Let p be prime. A finite group is a p-group if and only if its order is a power of p.*

Proof. $\Leftarrow$ is clear by Lagrange. $\Rightarrow$: if $|G|$ is not a power of p, then $|G|$ has a prime divisor $q \neq p$. Then Cauchy's theorem 5.12.1 gives a subgroup of order q. By Proposition 5.3.18 it is isomorphic to C_q. Thus, G contains an element of order q, so is not a p-group. □

Example 5.13.3. Let p be prime. From the primary decomposition we see that a finite abelian group G is a p-group if and only if $G \cong \mathbb{Z}_{p^{k_1}} \oplus \cdots \oplus \mathbb{Z}_{p^{k_s}}$ for some $s \in \mathbb{N}$ and $k_i \in \mathbb{N}$.

Example 5.13.4 (Prüfer p-group)**.** Let p be prime. Let $\mathbb{Z}(p^\infty) := \bigcup_{n \in \mathbb{N}} C_{p^n}$.

1. $\mathbb{Z}(p^\infty)$ is a subgroup of the circle group $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$.
2. $\mathbb{Z}(p^\infty)$ is a p-group with $\exp(\mathbb{Z}(p^\infty)) = \infty$.
3. Every proper subgroup U of $\mathbb{Z}(p^\infty)$ equals C_{p^n} for some $n \in \mathbb{N}$.

4. $\mathbb{Z}(p^\infty)$ is divisible (cf. Exercise 5.4.2).

Proof. (1) is clear. (2): the exponent is ∞ because ζ_{p^n} has order p^n. Given $x \in \mathbb{Z}(p^\infty)$ choose $n \in \mathbb{N}$ minimal with $x \in C_{p^n}$; write $x = \zeta_{p^n}^k$ for some $k \in \mathbb{N}$. Then $p \nmid k$, so $\gcd(k, p^n) = 1$. By Lemma 5.3.9 (5), x has order p^n.

(3): as just seen, for every $x \in U$ we have $\langle x \rangle = C_{p^{n_x}} \subseteq U$ for some $n_x \in \mathbb{N}$. Since U is proper, there is a maximal $m \in \mathbb{N}$ among the n_x's. Then $U = C_{p^m}$.

(4): it suffices to find for every $x \in G$ and every prime q some $y \in G$ with $y^q = x$. Choose $n, k \in \mathbb{N}$ such that $x = \zeta_{p^n}^k$. If $q = p$, set $y := \zeta_{p^{n+1}}^k$. If $q \neq p$, choose $a, b \in \mathbb{Z}$ with $ap^n + bq = 1$ by Bézout; then $x = x^{ap^n} x^{bq} = x^{bq}$ and we take $y := x^b$. □

Lemma 5.13.5. *Let p be prime. Finite non-trivial p-groups have non-trivial center.*

Proof. Let G be a finite p-group. By Lemma 5.13.2, $|G| = p^n$ for some $n \in \mathbb{N}$. Then $n > 0$ since $G \neq \{1\}$. Write $p^n = |\mathrm{Z}(G)| + \sum_{x \in T} [G : \mathrm{Z}(x)]$ by the class equation (Theorem 5.12.13 (3)). For $x \in T$, $[G : \mathrm{Z}(x)] \geqslant 2$ divides $|G| = p^n$ by Lagrange, so $p \mid [G : \mathrm{Z}(x)]$. Hence, $p \mid |\mathrm{Z}(G)|$, so $\mathrm{Z}(G) \neq \{1\}$. □

As an application we generalize Proposition 5.3.19 and classify the groups of order p^2. In the end of this section we classify the groups of order pq with distinct primes p, q.

Example 5.13.6 (Groups of order p^2)**.** Let p be prime and G a group of order p^2. Then G is isomorphic to $\mathbb{Z}_p \oplus \mathbb{Z}_p$ or to $\mathbb{Z}_{p^2}$.

Proof. By the classification of finite abelian groups it suffices to show G is abelian. By the lemma, $\mathrm{Z}(G)$ has order > 1 and devides p^2 by Lagrange. If $|\mathrm{Z}(G)| = p^2$, then G is abelian. Hence, it suffices to show $|\mathrm{Z}(G)| \neq p$. Otherwise, choose $x \in G \setminus \mathrm{Z}(G)$. Then $\mathrm{Z}(x)$ contains both $\mathrm{Z}(G)$ and x, so is larger than $\mathrm{Z}(G)$. Since $|\mathrm{Z}(x)|$ divides $|G| = p^2$, we have $|\mathrm{Z}(x)| = p^2$, so $\mathrm{Z}(x) = G$ and $x \in \mathrm{Z}(G)$, a contradiction. □

Definition 5.13.7. Let p be prime and G a finite group of order $p^\ell m$ with $p \nmid m$ where $\ell, m \in \mathbb{N}$. A *p-subgroup (of G)* is a subgroup that is a p-group.

A *p-Sylow subgroup* is a p-subgroup of order p^ℓ.

Remark 5.13.8. Let G, p, ℓ, m be as above.

1. If $\ell = 0$, then $\{1\}$ is a p-Sylow subgroup.
2. By Lemma 5.13.2, p-subgroups have order p^k for some k and $k \leqslant \ell$ (Lagrange).
3. For every $x \in G$, if U is a p-(Sylow) subgroup, then so is xUx^{-1} (conjugation is an automorphism).
4. If $q \neq p$ is prime, P a p-subgroup P and Q a q-subgroup, then $P \cap Q = \{1\}$.

   Indeed: if $x \in P \cap Q$, then $\mathrm{ord}(x)$ divides $|P|$, a power of p, and $|Q|$, a power of q, so $\mathrm{ord}(x) = 1$. (Lemmas 5.3.15, 5.13.2).

5. If $\ell = 1$, then any p-Sylow subgroups $P \neq P'$ have trivial intersection $P \cap P' = \{1\}$.

   Indeed: $P \cap P'$ is a subgroup of P that is proper because $|P| = |P'|$. Its order divides $|P| = p$ (Lagrange), so equals 1.

Exercise 5.13.9. The case of abelian G is easy. Let p be a prime divisor of $|G|$. Show there is exactly one p-Sylow subgroup of G, namely the *p-torsion subgroup*:

$$T_p(G) := \{x \in G \mid \text{ord}(x) \text{ is a power of } p\}.$$

Show G is the inner direct sum of its p-torsion subgroups.

Exercise 5.13.10. Let $p, n, m \in \mathbb{N}$ with $n > 0, m > 1$ and p prime. Let $\mathbb{F}_q$ be a finite field of size $q := p^n$. Show $|\mathrm{GL}(m, \mathbb{F}_q)| = q^{m(m-1)/2} \prod_{i=1}^{m} (q^i - 1)$. Let G be the set of upper triangular matrices over $\mathbb{F}_q$ with 1's on the diagonal. Show G is p-Sylow subgroup of $\mathrm{GL}(m, \mathbb{F}_q)$.

Recall the normalizer $\mathrm{N}(U)$ of a subgroup from Proposition 5.12.15.

Lemma 5.13.11. *Let p be prime, G a finite group and U a non-trivial p-subgroup. Then*

$$[\mathrm{N}(U) : U] \equiv [G : U] \mod p.$$

Proof. Let U act on G/U by left-translation, i.e., $(u, gU) \mapsto ugU$. By Lemma 5.13.2, $|U|$ is a power of p. By the fixed-point lemma $[G : U] \equiv |F| \mod p$ where F is the set of fixed-points of the action. What does it mean to be a fixed-point? $ugU = gU$ for all $u \in U$ is equivalent to $g^{-1}ugU = U$ for all $u \in U$, hence to $g^{-1}Ug = U$, hence to $g \in \mathrm{N}(U)$. Thus, $gU \in F$ if and only if $g \in \mathrm{N}(U)$. Thus $|F| = [\mathrm{N}(U) : U]$. □

Theorem 5.13.12 (1st Sylow theorem)**.** *Let p be prime and G a finite group of order $p^\ell m$ with $p \nmid m$ where $\ell, m \in \mathbb{N}$. Then p-Sylow subgroups exist. In fact,*

1. *For all $k \leqslant \ell$ there exist subgroups of G of order p^k.*
2. *Every subgroup of order p^k with $k < \ell$ is normal in some subgroup of order p^{k+1}.*

Proof. (1) is proved by induction on k. For $k = 0$ take $\{1\}$. Assume $k < \ell$ and there is a subgroup U of order p^k. Then $p \mid [G : U] = p^{\ell-k}m$. By the lemma, $p \mid [\mathrm{N}(U) : U]$. By Cauchy's theorem, $\mathrm{N}(U)/U$ has a subgroup $\tilde{U}$ of order p.

Let $\pi_U : \mathrm{N}(U) \to \mathrm{N}(U)/U$ be the canonical projection. Then $U' := \pi_U^{-1}(\tilde{U})$ is a subgroup of $\mathrm{N}(U)$ and hence of G. Its order is $|\tilde{U}| \cdot |U| = p^{k+1}$.

(2): let U be a subgroup of order p^k with $k < \ell$. Define U' as above. Then $U \subseteq U' \subseteq \mathrm{N}(U)$ and $U \lhd \mathrm{N}(U)$. This implies $U \lhd U'$. □

Corollary 5.13.13. *Let p be prime. Finite p-groups are solvable.*

Proof. A finite p-group G has order p^ℓ for some $\ell \in \mathbb{N}$ by Lemma 5.13.2. Let $N_0 := \{1\}$ and N_1 be a subgroup of order p. Then choose a N_2 of order p^2 with $N_1 \lhd N_2$. Continuing ℓ times, gives a subnormal series of G. Then N_{k+1}/N_k has order p, so is isomorphic to C_p (Proposition 5.3.18), so abelian. □

Theorem 5.13.14 (2nd Sylow theorem)**.** *Let p be prime and G a finite group.*

1. *For every p-subgroup U and every p-Sylow subgroup P there is $g \in G$ such that $gUg^{-1} \subseteq P$.*
2. *Any two p-Sylow subgroups are conjugate.*
3. *A p-Sylow subgroup is normal if and only if it is the only p-Sylow subgroup.*

Proof. (1): we can assume U is non-trivial. Let U act on G/P be left translation, i.e., $(u, gP) \mapsto ugP$. By Lemma 5.13.2, $|U|$ is a power of p. By the fixed-point lemma, $[G : P] \equiv |F| \mod p$ where F is the set of fixed-points. Since P is p-Sylow, $[G : P] = |G|/|P| \not\equiv 0 \mod p$. Hence $F \neq \varnothing$. Let $gP \in F$. Then $g^{-1}ugP = P$ for all $u \in U$, so $g^{-1}U(g^{-1})^{-1} \subseteq P$.

(2) follows from (1) because all p-Sylow subgroups have the same order.

(3): let P be a p-Sylow subgroup. Recall $P \lhd G$ means $gPg^{-1} = P$ for all $g \in G$. But the sets $gPg^{-1}, g \in G$, are precisely the p-Sylow subgroups (by (2) and Remark 5.13.8 (2)). □

Exercise 5.13.15. Let G be a finite group, U a subgroup, p prime and P a p-Sylow subgroup of G. Show that for every p-Sylow subgroup Q of U there are a p-Sylow subgroup P of G and $x \in G$ such that $Q = U \cap xPx^{-1}$.

Theorem 5.13.16 (3rd Sylow theorem)**.** *Let p be prime and G a finite group of order $p^\ell m$ with $p \nmid m$ where $\ell, m \in \mathbb{N}$. Let s_p denote the number of p-Sylow subgroups of G.*

Then $s_p \mid m$ and $s_p \equiv 1 \mod p$.

Proof. Let G operate on the set of p-Sylow subgroups X by conjugation, i.e., $(g, P) \mapsto gPg^{-1}$. Note $X \neq \varnothing$ by the 1st Sylow theorem, and the action is transitive by the 2nd Sylow theorem, i.e., X equals the orbit $G(\tilde{P})$ – we fix $\tilde{P} \in X$ arbitrarily.

By definition, the stabilizer $G_{\tilde{P}}$ is the normalizer $\mathrm{N}(\tilde{P})$. By the orbit-stabilizer lemma,

$$s_p = |X| = |G(\tilde{P})| = [G : \mathrm{N}(\tilde{P})].$$

Recalling $|\tilde{P}| = p^\ell$ we get $s_p \mid m$ using Lagrange as follows:

$$p^\ell m = |G| = [G : \mathrm{N}(\tilde{P})] \cdot |\mathrm{N}(\tilde{P})| = [G : \mathrm{N}(\tilde{P})] \cdot [\mathrm{N}(\tilde{P}) : \tilde{P}] \cdot |\tilde{P}| = s_p \cdot [\mathrm{N}(\tilde{P}) : \tilde{P}] \cdot p^\ell.$$

We now show $s_p \equiv 1 \mod p$. Let $\tilde{P}$ act on X by conjugation. Let $T \subseteq X$ contain exactly one $P \in X$ of each orbit. For the stabilizers $\tilde{P}_P$ of $P \in X$ the Bahngleichung reads

$$s_p = \textstyle\sum_{P \in T} [\tilde{P} : \tilde{P}_P].$$

Since $[\tilde{P} : \tilde{P}_P] \mid |\tilde{P}| = p^\ell$ by Lagrange, we can write $[\tilde{P} : \tilde{P}_P] = p^{k_P}$ for some $k_P \leqslant \ell$. It now suffices to show

$$k_P = 0 \iff P = \tilde{P}.$$

$\Leftarrow$ is clear because $\tilde{P}_{\tilde{P}} = \tilde{P}$, so $[\tilde{P} : \tilde{P}_{\tilde{P}}] = 1$. $\Rightarrow$: if $k_P = 0$, then $[\tilde{P} : \tilde{P}_P] = 1$, so $\tilde{P} = \tilde{P}_P$, so $gPg^{-1} \subseteq P$ for all $g \in \tilde{P}$. By definition, $\tilde{P} \subseteq \mathrm{N}(P)$. By Proposition 5.12.15, $P \lhd \mathrm{N}(P)$. Then both $\tilde{P}$ and P are p-Sylow subgroups of $\mathrm{N}(P)$. By the 2nd Sylow theorem, they are conjugate in $\mathrm{N}(P)$, i.e., $\tilde{P} = gPg^{-1}$ for some $g \in \mathrm{N}(P)$. But $gPg^{-1} = P$ by normality. □

The Sylow theorems are our main tool for analyzing finite groups.

Example 5.13.17. There are no simple groups of order 10, 20, 30, 40, 50, 70, 80 or 90 (by Example 5.6.26, A_5 is simple of order 60).

Proof. Let G be a group. Case $|G| = 20 = 2^2 \cdot 5$: by the 3rd Sylow theorem $s_5 \mid 4$ and $s_5 \equiv 1 \mod 5$, so $s_5 = 1$. By the 2nd Sylow theorem, the unique 5-Sylow subgrup is a normal subgroup of G of order 5. Hence, G is not simple. Cases $|G| = 10, 40, 50, 70$ are similar.

Case $|G| = 30 = 2 \cdot 3 \cdot 5$: 3rd Sylow gives $s_5 \mid 6$ and $s_5 \equiv 1 \mod 5$, so $s_5 \in \{1, 6\}$. Further, $s_3 \mid 10$ and $s_3 \equiv 1 \mod 3$, so $s_3 \in \{1, 10\}$. If $s_5 = 1$ or $s_3 = 1$, we are done as before. So assume $s_5 = 6$ and $s_3 = 10$. Remark 5.13.8 (5) gives $10 \cdot (3-1) = 20$ elements of order 3 and $6 \cdot (5-1) = 24$ elements of order 5, contradicting $|G| = 30$.

The case $|G| = 80$ is similar (exercise). The case $|G| = 90 = 2 \cdot 3^2 \cdot 5$ is more complicated: $s_3 \mid 10$ and $s_3 \equiv 1 \mod 3$ implies $s_3 \in \{1, 10\}$. $s_5 \mid 18$ and $s_5 \equiv 1 \mod 5$ implies $s_5 \in \{1, 6\}$. If one is $= 1$ we are done. So assume $s_3 = 10, s_5 = 6$.

Then there are $6 \cdot (5-1) = 24$ elements of order 5, disjoint from all 3-Sylow subgroups (Remark 5.13.8 (4), (5)). If the 3-Sylow subgroups intersect trivially, they comprise another $1 + 10 \cdot (9-1) = 81$ elements – too many.

Hence, there are distinct 3-Sylow subgroups P, Q with $|P \cap Q| > 1$, so $|P \cap Q| = 3$ by Lagrange. By Exercise 5.7.12, $|PQ| = |P| \cdot |Q| / |P \cap Q| = 27$.

By Proposition 5.12.16, $P \cap Q$ is normal in both P and Q. Hence, $P, Q \subseteq N := \mathrm{N}(P \cap Q)$, the normalizer of $P \cap Q$ in G. Thus, $|N| \geqslant 27$. Further, $|N| \mid 90$ and $9 \mid |N|$ by Lagrange (P is a subgroup of N). Thus, $|N| \in \{45, 90\}$. If $|N| = 45$, then $N \lhd G$ (index 2) and we are done. If $|N| = 90$, then $N = G$, so $P \cap Q \lhd G$ and we are done. □

Exercise 5.13.18. In case $|G| = 30$ above, show G has a subgroup of order 15 and infer that $s_3 = s_5 = 1$.

Example 5.13.19 (Groups of order pq)**.** Let $p < q$ be primes. There is a homomorphism $\Phi : C_p \to \mathrm{Aut}(C_q)$ such that every group of order pq is isomorphic to either C_{pq} or $C_q \rtimes_\Phi C_p$.

Moreover, the 2nd case happes only if $p \mid q - 1$.

Proof. Let G have order pq. 3rd Sylow gives $k, \ell \in \mathbb{N}$ such that $s_p \mid q$ and $s_p = 1 + kp$, and $s_q \mid p$ and $s_q = 1 + \ell q$. Then $s_q = 1$: otherwise $s_q = p$, so $p - 1 = \ell q$, so $\ell \neq 0$ and $q < p$, a contradiction. Let P, Q be p-, resp., q-Sylow subgroups. Then $Q \lhd G$ by 2nd Sylow. Moreover, $P \cap Q = \{1\}$ by Remark 5.13.8 (4). By the 2nd isomorphism theorem, $|PQ| = |P||Q|/|P \cap Q| = pq = |G|$. Thus, $G = PQ$.

By Theorem 5.10.13, $G \cong Q \rtimes P$. As $Q \cong C_q$ and $P \cong C_p$, we have $Q \rtimes P \cong C_q \rtimes_\Phi C_p$ for some homomorphism $\Phi : C_p \to \mathrm{Aut}(C_q)$ (Exercise 5.10.12).

As $\ker(\Phi)$ is a subgroup of C_p, its order divides p by Langrange. If this order is p, then Φ is constant and $G \cong C_q \times C_p \cong C_{pq}$ (Remark 5.10.5).

Assume $\ker(\Phi)$ has order 1. Then Φ is injective. By Exercise 5.10.11 it suffices to show such Φ have only one possible image. By Exercise 2.6.5, $\mathrm{Aut}(C_q) \cong \mathbb{Z}_q^\times$. By Example 5.3.14 (2), $\mathbb{Z}_q^\times$ is cyclic, so has at most one subgroup of order p (Theorem 5.3.21).

As $|\mathbb{Z}_q^\times| = q - 1$ (Remark 2.6.7), the 2nd case implies $p \mid q - 1$ by Lagrange. □

Example 5.13.20. The 2nd case above happens for the dihedral group D_q for a prime $q > 2$. It has order $2q$ and is not cyclic (Example 5.3.14 (4)). In fact, we already saw in Example 5.10.15 that $D_q \cong C_q \rtimes_\Phi C_2$ for a certain $\Phi : C_2 \to \mathrm{Aut}(C_q)$.

Exercise 5.13.21. Let $p \neq q$ be prime. Groups of order p^2q are inner semidirect products of proper subgroups.

Exercise 5.13.22. Groups of order 922 or 2022 are solvable.

Exercise 5.13.23. Let p be prime and $p \leqslant n < p^2$. Every p-Sylow subgroup of S_n is abelian.

Example 5.13.24. We list all groups (up to $\cong$) of order $\leqslant 10$ using Proposition 5.3.19 for order 4, the above for 6 and 10, Examples 5.11.18, 5.6.19 for 8, and Example 5.13.6 for 9:

$$\{1\},\ C_2,\ C_3,\ \underbrace{C_4,\ K_4}_{\text{order } 4=2^2},\ C_5,\ \underbrace{C_6,\ D_3}_{\text{order } 6},\ C_7,\ \underbrace{C_8,\ C_2\times C_4,\ C_2^3,\ D_4,\ Q_8}_{\text{order } 8=2^3},\ \underbrace{C_9,\ C_3^2}_{\text{order } 9=3^2},\ \underbrace{C_{10},\ D_5}_{\text{order } 10}.$$

The numbers of groups of order 2^n for $n = 4, ..., 9$ are $14, 51, 267, 2328, 56092, 10494213$. The numbers of groups of order 3^n for $n = 3, ..., 7$ are $5, 15, 67, 504, 9310$.

For fixed prime p, no somehow explicit formula is known for the number of groups of order p^n. The *Higman-Sims formula* states the upper bound $p^{2n^3/27+O(n^{8/3})}$.

Chapter 6

Field theory

6.1 Ruler and compass constructions

Definition 6.1.1. We refer to the elements of $\mathbb{R}^2$ as *points*. A point p is *constructible* if there are $n \in \mathbb{N}$ and a sequence $p_1, ..., p_n$ of points with $p_n = p$ such that for all $1 \leqslant i \leqslant n$ there are $1 \leqslant i_1, i_2, i_3, i_4, i_5, i_6 < i$ such that one of the following holds.

1. $p_i \in \{(0,0), (1,0)\}$,
2. p_i is the intersection of the line through $p_{i_1} \neq p_{i_2}$ and the line through $p_{i_3} \neq p_{i_4}$,
3. p_i is an intersection of the line through $p_{i_1} \neq p_{i_2}$ and the circle of radius $\|p_{i_3} - p_{i_4}\|$ around p_{i_5},
4. p_i is an intersection of the circle of radius $\|p_{i_1} - p_{i_2}\|$ around p_3 and the circle with radius $\|p_{i_4} - p_{i_5}\|$ around p_{i_6}.

The set of constructible points is denoted Con.

Notation: as usual we view the set $\mathbb{R}^2$ as the set of complex numbers $\mathbb{C}$ and thereby $\mathrm{Con} \subseteq \mathbb{C}$. Accordingly, we write e.g. $0, 1, i, e^{i\alpha}$ instead $(0,0), (1,0), (0,1), (\sin(\alpha), \cos(\alpha))$ and call a real $r \in \mathbb{R}$ *constructible* if $(r, 0)$ is constructible. It is easy to see that $z \in \mathbb{C}$ is constructible if and only if both reals $\mathrm{Re}(z), \mathrm{Im}(z)$ are constructible.

Remark 6.1.2. Every constructible point is determined by a sequence as above, so by a number $k \in \mathbb{N}$ and for each $1 \leqslant i \leqslant k$: a number 1-4 determining the rule, numbers $i_1, ..., i_6 < i$ and a bit determining which of the $\leqslant 2$ intersections is taken. This way one sees that Con is countable, so "most" points are not constructible.

Remark 6.1.3 (Classical Greek problems[1])**.**

1. *Delian problem:* is $\sqrt[3]{2}$ constructible?

   Given a cube can you construct another of double volume?

[1] https://mathshistory.st-andrews.ac.uk/HistTopics/category-greeks/

"Eratosthenes, in his work entitled Platonicus relates that, when the god proclaimed to the Delians through the oracle that, in order to get rid of a plague, they should construct an altar double that of the existing one, their craftsmen fell into great perplexity in their efforts to discover how a solid could be made the double of a similar solid; they therefore went to ask Plato about it, and he replied that the oracle meant, not that the god wanted an altar of double the size, but that he wished, in setting them the task, to shame the Greeks for their neglect of mathematics and their contempt of geometry." (Theon of Smyrna)

2. *Angle trisection:* is $e^{i\alpha/3}$ constructible from $\alpha \in [0, 2\pi)$? *from* means that $e^{i\alpha}$ is allowed aside $0, 1$ in in Rule 1 of the definition above.

   Given an angle α can you construct an angle $\alpha/3$?

3. *Squaring the circle:* is $\sqrt{\pi}$ constructible?

   Given a circle, can you construct a square of the same area?

   The egyptian Rhind papyrus (c.1850 BCE) gives a geometric approximation of π by 3.1605. The first mathematician on record trying to square the circle was Anaxagoras c.450 BCE while in prison for heresy, namely for "teaching that the sun was a red-hot stone and the moon was earth." (Russell) The problem became popular in ancient Greece. Remarkably, unlike in modern times, erroneous 'proofs' were scarce.

4. *Construction of regular n-gons:* for which $n > 2$ is $\zeta_n = e^{2\pi i/n}$ constructible?

   Given a circle, can you inscribe a regular n-gon?

Lemma 6.1.4. Con *is a subfield of* $\mathbb{C}$*. It is closed under conjugation and square roots, i.e., if* $z \in$ Con*, then* $\bar{z} \in$ Con *and* $\pm\sqrt{z} \in \mathbb{C}$*.*

Sketch of proof. It suffices to construct $0, 1, z + z', -z, z \cdot z', z^{-1}, \bar{z}, \pm\sqrt{z}$ from $z, z' \in \mathbb{C}$. This is done by school geometry. We only explain how to construct $\sqrt{z} = \sqrt{r}e^{i\alpha/2}$ from $z = re^{i\alpha}$ where $r, \alpha > 0$. Bisect the angle of the x-axis and the line through z and 0 (the origin) and intersect with a circle around 0 of radius $\sqrt{r}$. How to construct $\sqrt{r}$ from r? Construct $-r$ (i.e., $(-r, 0)$), the midpoint between $-r$ and 1 on the x-axis, and then a circle around it that intersects the x-axis in $-r$ and 1. Construct a line perpendicular to the x-axis and passing through 0. Its intersection u with the circle gives a right triangle $-r, 1, u$ by Thales' theorem. Its height $\mathrm{Im}(u)$ is $\sqrt{1 \cdot r}$ by Euclid's right triangle altitude theorem. □

Theorem 6.1.5. $z \in \mathbb{C}$ *is constructible if and only if there are* $n \in \mathbb{N}$ *and fields*

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n \subseteq \mathbb{C}$$

with $z \in K_n$ *and for all* $i < n$*, the extension* $K_{i+1} \mid K_i$ results by adjunction of a square root*, i.e.,* $K_{i+1} = K_i(a)$ *for some* $a \in \mathbb{C}$ *with* $a^2 \in K_i$*.*

Proof. $\Leftarrow$: we show $K_i \subseteq$ Con by induction on i. For $i = 0$, note $K_0 = \mathbb{Q}$ is the prime field of Con. Assume $K_i \subseteq$ Con and $K_{i+1} = K_i(a)$ with $a^2 \in K_i$ and $a \in \mathbb{C}$. Then $K_i \cup \{a\} \subseteq$ Con since Con is closed under square roots. Then $K_{i+1} = K_i(a) \subseteq$ Con since Con is a subfield of $\mathbb{C}$.

$\Rightarrow$: let $z_1, ..., z_n = z$ witness that z is constructible. We claim that for all $i \leqslant n$ there are $K_0 := \mathbb{Q}(i) \subseteq K_1 \subseteq \cdots \subseteq K_i$ such that $z_i \in K_i$ and for all $j \leqslant i$ we have $K_{j+1} = K_j(a)$ for some root $a \in \mathbb{C}$ of some $f \in K_j[X]$ with $\deg(f) \leqslant 2$. This suffices by Lemma 3.5.11 (2).

Lemma 3.5.11 (3) implies $K_{j+1} = K_j + K_j \cdot \sqrt{D_f}$, so K_{j+1} is closed under complex conjugation if K_j is. By induction, all K_j are closed under conjugation.

Observe: if $z := x + iy \in K_j$ then $x = (z + \bar{z})/2 \in K_j$ and $y = (z - \bar{z})/(2i) \in K_j$ (as $i \in K_j$); hence, if $z, z' \in K_j$, then $\|z - z'\|^2 \in K_j$.

We proceed by induction on i. Our claim is trivial for $i = 0$. Assume it for $i < n$ and distinguish cases on how z_{i+1} is obtained.

Case: $z_{i+1} \in \{0, 1\}$. Set $f := X - 1$ (and $K_{i+1} = K_i$).

Case: z_{i+1} is an intersection of two lines. A line through $x_0 + iy_0, x_1 + iy_1 \in K_i$ is the set of $(x, y) \in \mathbb{R}^2$ satisfying $(x_1 - x_0)(y - y_0) = (y_1 - y_0)(x - x_0)$. Here, $x_0, y_0, x_1, y_1 \in K_i \cap \mathbb{R}$. Hence the intersection of two lines means solving a system of linear equations with coefficients in K_i. The solution is in K_i, so we can again take $f := X - 1$.

Case: z_{i+1} is the intersection of a circle and a line. A circle with radius r around $x_0 + iy_0 \in K_i$ is the set of (x, y) satisfying $(x - x_0)^2 + (y - y_0)^2 = r^2$. Here, $x_0, y_0, r^2 \in K_i \cap \mathbb{R}$. To find the intersection with a line, eliminate one variable given the linear equation for the line and solve a quadratic equation in the other: this equation is f.

The field extension that contains a root of f contains one coordinate of z_{i+1}, hence also the other (due to the linear equation), and hence z_{i+1} (since it contains i).

Case: z_{i+1} is the intersection of two circles. Given two circles $(x - x_0)^2 + (y - y_0)^2 = r^2$ and $(x - x_1)^2 + (y - y_1)^2 = s^2$ with $r^2, s^2, x_0, x_1, y_0, y_1 \in K_i \cap \mathbb{R}$, subtract the equations and get a linear equation. Eliminate one variable and proceed as before. □

Informally, $z \in \mathbb{C}$ is constructible if and only if z equals some expression built from the field operations, fractions and square roots. We illustrate this by a famous example:

Example 6.1.6 (Regular 17-gon)**.** To construct ζ_{17} it suffices to construct $\cos(2\pi/17)$. This is possible as noted by 18 year old Gauß:

$$\cos(2\pi/17) = -\frac{1}{16} + \frac{\sqrt{17}}{16} + \frac{1}{16}\sqrt{34 - 2\sqrt{17}} + \frac{1}{8}\sqrt{17 + 3\sqrt{17} - \sqrt{34 - 2\sqrt{17}} - 2\sqrt{34 + 2\sqrt{17}}}.$$

We have $a := 34 - 2\sqrt{17}, b := 34 + 2\sqrt{17} \in \mathbb{Q}(\sqrt{17})$, so $c := 17 + 3\sqrt{17} - \sqrt{a} - 2\sqrt{b} \in \mathbb{Q}(17, \sqrt{a}, \sqrt{b})$. Thus, $\cos(2\pi/17)$ is captured by successively adjoining square roots, e.g. via

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{17}) \subseteq \mathbb{Q}(\sqrt{17}, \sqrt{a}) \subseteq \mathbb{Q}(\sqrt{17}, \sqrt{a}, \sqrt{b}) \subseteq \mathbb{Q}(\sqrt{17}, \sqrt{a}, \sqrt{b}, \sqrt{c}).$$

This field tower is poorly motivated and all but unique. We need some theory.

6.2 Algebraic extensions

Let $L \mid K$ be a field extension. By *L as a K-vector space* we mean: the vectors are L with $+_L$, the scalar field is K and the scalar multiplication is $(a, b) \mapsto a \cdot_L b$ for $a \in K, b \in L$.

Examples 6.2.1. We know for $\mathbb{C}$ and quadratic number fields $\mathbb{Q}(\sqrt{d})$ (cf. Example 4.8.17):

$$\mathbb{R}[X]/(X^2+1) \cong \mathbb{R}(i) = \mathbb{C} = \{x+yi \mid x,y \in \mathbb{R}\},$$
$$\mathbb{Q}[X]/(X^2-d) \cong \mathbb{Q}(\sqrt{d}) = \{x+y\sqrt{d} \mid x,y \in \mathbb{Q}\}.$$

$\mathbb{C}$ as an $\mathbb{R}$-vector space has basis $1, i$, and, $\mathbb{Q}(\sqrt{d})$ as a $\mathbb{Q}$-vector space has basis $1, \sqrt{d}$.

Definition 6.2.2.

1. $L \mid K$ is *finitely generated* if $L = K(A)$ for some finite $A \subseteq L$.
2. $L \mid K$ is *simple* if $L = K(a)$ for some $a \in L$, a *primitive element* of $L \mid K$.
3. The *degree* $[L:K]$ of $L \mid K$ is the dimension of L as a K-vector space; we write $[L:K] = \infty$ if the degree is infinite.
4. $L \mid K$ is *finite* if $[L:K]$ is finite.
5. $L \mid K$ is *algebraic* if every $a \in L$ is algebraic over K.

Remark 6.2.3.

1. If $L \mid K$ is finite, then it is finitely generated.

   Indeed: if $x_1, ..., x_n$ is a basis of L as a K-vector space, then $L = K(x_1, ..., x_n)$.
2. $[L:K] = 1$ if and only if $L = K$.

   Indeed: if $L = K$, then 1 is a basis of L as a K-vector space; in particular, K is a 1-dimensional subspace of L; hence $[L:K] = 1$ implies $L = K$.
3. If $L \mid K$ is finite and $a \in L$, then a is a root of some $f \in K[X]$ with $\deg(f) \leqslant [L:K]$. In particular, a is algebraic over K.

   Indeed: let $n := [L:K]$; then $1, a, a^2, ..., a^n$ are linearly dependent vectors, so $x_0 + x_1 a + \cdots + x_n a^n = 0$ for certain $x_i \in K$; then a is a root of $x_n X^n + \cdots + x_0 \in K[X]$.

Examples 6.2.4. $\mathbb{Q}(\pi) \mid \mathbb{Q}$ is simple of infinite degree (by Remark 6.2.3 (3) and Lindemann's theorem). $\mathbb{R} \mid \mathbb{Q}$ is not finitely generated and hence $[\mathbb{R}:\mathbb{Q}] = \infty$.

Indeed, assume $\mathbb{R} = \mathbb{Q}(A)$ for $A \subseteq \mathbb{R}$ of size $n \in \mathbb{N}$; then there is a surjection from $\mathbb{Q}[X_1, ..., X_n]$ onto $\mathbb{R}$; but $\mathbb{Q}[X_1, ..., X_n]$ is countable.

Recall minimal polynomials from Definition 3.5.2. The $\cong$ below is Corollary 4.8.15.

Theorem 6.2.5. *$a \in L$ is algebraic over K if and only if $K(a) \mid K$ is finite. In case, $n := \deg(m_a^K) = [K(a):K]$ and $1, a, ..., a^{n-1}$ is a basis of $K(a)$ as a K-vector space.*

In particular, then

$$K[X]/(m_a^K) \cong K(a) = \{x_0 + x_1 a + \cdots + x_{n-1} a^{n-1} \mid x_0, ..., x_{n-1} \in K\}.$$

Proof. $\Leftarrow$ follows from Remark 6.2.3 (3). $\Rightarrow$: to show $1, a, ..., a^{n-1}$ generate $K(a)$ as a K-vector space, we verify the displayed equality. By Theorem 3.5.8, $K(a) = K[a]$ and we have to show that every $f(a)$ with $f \in K[X]$ equals $r(a)$ for some $g \in K[X]$ with $\deg(r) < n$. Write $f = q \cdot m_a^K + r$ with $\deg(r) < n$ by polynomial division; then $f(a) = r(a)$.

Independent: assume $x_0 + x_1 a + \cdots x_{n-1} a^{n-1} = 0$ with $x_i \in K$ not all 0; then a is a root of $x_{n-1}X^{n-1} + x_{n-2}X^{n-2} + \cdots + x_0 \in K[X] \setminus \{0\}$ of degree $< n$; we then also find a monic such f, contradicting the definition of m_a^K. □

Exercise 6.2.6. Recall Exercise 3.5.10. Let $a \in \mathbb{C}$ be a root of $X^3 - X + 1 \in \mathbb{Q}[X]$. The basis $1, a, a^2$ determines a vector space isomorphism $\varphi : \mathbb{Q}(a) \cong \mathbb{Q}^3$. Compute $\varphi(a^4), \varphi(a^5), \varphi(a^6)$.

$x \mapsto ax$ in $\mathbb{Q}(a)$ corresponds to an endomorphism of $\mathbb{Q}^3$ – what is its matrix?

Exercise 6.2.7. Let K be a field and $f \in K[X] \setminus \{0\}$. View $K[X]/(f)$ naturally as a K-vector space and show its dimension is $\deg(f)$.

Example 6.2.8. $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3}) = \left\{x_1 + x_2\sqrt{2} + x_3\sqrt{3} + x_4\sqrt{6} \mid x_1, ..., x_4 \in \mathbb{Q}\right\}$.

Proof. Compute powers of $\alpha := \sqrt{2} + \sqrt{3}$:

$$\alpha^2 = 5 + 2\sqrt{6},\ \alpha^3 = 11\sqrt{2} + 9\sqrt{3},\ \alpha^4 = 49 + 20\sqrt{6}.$$

We see, α is a root of $f := X^4 - 10X^2 + 1 \in \mathbb{Q}[X]$. Further, $\sqrt{2} = (\alpha^3 - 9\alpha)/2, \sqrt{3} = -(\alpha^3 - 11\alpha)/2 \in \mathbb{Q}(\alpha)$. Thus, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\alpha)$.

By Example 4.5.5, f is irreducible, so $f = m_\alpha^{\mathbb{Q}}$. By the theorem, $1, \alpha, \alpha^2, \alpha^3$ is a basis of $\mathbb{Q}(\alpha)$ as a $\mathbb{Q}$-vector space. But then also $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$ is a basis. □

Such ad hoc argument won't take us far. We need more theory.

Proposition 6.2.9. *Assume $K \neq L$ and $\mathrm{char}(K) \neq 2$. Then $[L : K] = 2$ if and only if L is obtained from K by adjunction of a square root.*

Proof. $\Rightarrow$: there is $a \in L \setminus K$ such that $1, a$ is a basis of L as a K-vector space. Then $L = K(a)$ and $[K(a) : K] = \deg(m_a^K) = 2$. By Lemma 3.5.11, $K(a) = K(\sqrt{D})$ for the discriminant $D \in K$ of m_a^K.

$\Leftarrow$: assume $L = K(a)$ with $a^2 \in K$. Then $a \notin K$ as $L \neq K$. Then $X^2 - a^2 \in K[X]$ is irreducible, so equals m_a^K. Then $[K(a) : K] = \deg(m_a^K) = 2$. □

Example 6.2.10. Let $\mathbb{C} \mid K \mid \mathbb{Q}$ be field extensions. By Corollary 3.5.12, if $[K : \mathbb{Q}] = 2$, then K is a quadratic number field.

Definition 6.2.11. M is a *(proper) intermediate field* of $L \mid K$ if $M \mid K$ and $L \mid M$ are field extensions (and $L \neq M \neq K$).

We agree that $n < \infty = n \cdot \infty = \infty \cdot n = \infty \cdot \infty$ for all $n \in \mathbb{N}, n > 0$.

Theorem 6.2.12 (Degree formula)**.** *If M is an intermediate field of $L \mid K$, then*

$$[L : K] = [L : M] \cdot [M : K].$$

Proof. If $[L:M]=\infty$ or $[M:K]=\infty$, then $[L:K]=\infty$. Assume $[L:M]=n, [M:K]=m$ for $n,m\in\mathbb{N}$. Let $x_1,\dots,x_n$ be a basis of L as an M-vector space, and $y_1,\dots,y_m$ a basis of M as a K-vector space. We claim the x_iy_j are a basis of L as a K-vector space.

Generating: let $x\in L$; write $x=a_1x_1+\cdots+a_nx_n$ with $a_i\in M$ and $a_i=b_{i1}y_1+\cdots+b_{im}y_m$ with $b_{ij}\in K$. Then $x=\sum_{ij}b_{ij}x_iy_j$.

Independent: assume $\sum_{ij}a_{ij}x_iy_j=0$ with $a_{ij}\in K$; then $b_1x_1+\cdots+b_nx_n=0$ for $b_i:=\sum_j a_{ij}y_j\in M$. Then $b_i=0$ for all i because $x_1,\dots,x_n$ are independent. Then $a_{ij}=0$ for all j because $y_1,\dots,y_m$ are independent. □

Remark 6.2.13. Together with Remark 6.2.3 (2) we see: field extensions of prime degree, like $\mathbb{C}\mid\mathbb{R}$, do not have proper intermediate fields.

3rd proof of Example 4.5.5. By Example 6.2.8, $f:=X^4-10X^2+1\in\mathbb{Q}[X]$ has root $\alpha:=\sqrt{2}+\sqrt{3}\in\mathbb{R}$. It suffices to show $[\mathbb{Q}(\alpha):\mathbb{Q}]=4$ – then $f=m_\alpha^{\mathbb{Q}}$ is irreducible.

We know $\mathbb{Q}(\alpha)=\mathbb{Q}(\sqrt{2},\sqrt{3})$. By the degree formula

$$[\mathbb{Q}(\alpha):\mathbb{Q}]=[\mathbb{Q}(\sqrt{2},\sqrt{3}):\mathbb{Q}(\sqrt{2})]\cdot[\mathbb{Q}(\sqrt{2}):\mathbb{Q}].$$

Clearly, both factors are $\leqslant 2$. But then they are $=2$ because $\sqrt{2}\notin\mathbb{Q}$ and $\sqrt{3}\notin\mathbb{Q}(\sqrt{2})$. Indeed: assume $\sqrt{3}\in\mathbb{Q}(\sqrt{2})$; then $\sqrt{3}=x+y\sqrt{2}$ for certain $x,y\in\mathbb{Q}$; then $y\neq 0$ as $\sqrt{3}\notin\mathbb{Q}$ and $x\neq 0$ as $\sqrt{3/2}\notin\mathbb{Q}$; then $3=x^2+2xy\sqrt{2}+2y^2$ implies $\sqrt{2}\in\mathbb{Q}$ – contradiction. □

Theorem 6.2.14. *For $a\in L$ the following are equivalent.*

1. *a is transcendental over K.*
2. *$K[a]\subsetneq K(a)$.*
3. *$f\mapsto f(a)$ is an isomorphism from $K[X]$ onto $K[a]$.*
4. *$[K(a):K]=\infty$.*
5. *There is an infinite sequence of fields $K(a)\supsetneq K_1\supsetneq K_2\supsetneq\cdots\supsetneq K$.*

Proof. $1\Leftrightarrow 2$ follows from Theorem 3.5.8, and $1\Leftrightarrow 4$ from Theorem 6.2.5.

$1\Leftrightarrow 3$: clearly, $f\mapsto f(a)$ is an epimorphism, so (3) states it is injective, i.e., its kernel is $\{0\}$; but this means a is transcendental over K.

$1\Rightarrow 5$: clearly, with a also a^2 is transcendental over K. Further, $K(a^2)\subsetneq K(a)$: otherwise $a\in K(a^2)$, so $a=f(a^2)/g(a^2)$ for $f,g\in K[X]$; then a is a root of $Xg(X^2)-f(X^2)$, a contradiction. This gives a chain $K(a)\supsetneq K(a^2)\supsetneq K(a^{2^2})\supsetneq K(a^{2^3})\supsetneq\cdots$.

$5\Rightarrow 4$: since $[K_i:K_{i+1}]\geqslant 2$ (Remark 6.2.3 (2)), Theorem 6.2.12 gives for all $n>0$:

$$[K(a):K]\geqslant[K(a):K_n]=[K(a):K_1]\cdots[K_{n-1}:K_n]\geqslant 2^n.$$ □

Exercise 6.2.15. Every $f\in K(X)\smallsetminus K$ is transcendental over K.

Corollary 6.2.16.

1. *Let $n>1$ and $a_1,\dots,a_n\in L$ such that a_i is algebraic over $K(a_1,\dots,a_{i-1})$ for all $1\leqslant i\leqslant n$. Then $K(a_1,\dots,a_n)\mid K$ is finite.*

2. *$K \mid L$ is finite if and only if it is finitely generated and algebraic.*
3. *Let M be an intermediate field. Then $L \mid K$ is algebraic if and only both $L \mid M$ and $M \mid K$ are algebraic.*

Proof. 1: write $K_i := K(a_1, ..., a_i)$, so $K_{i+1} = K_i(a_{i+1})$ and $K_{i+1} \mid K_i$ is finite by Theorem 6.2.5. By the degree formula, $K_n \mid K$ is finite:

$$[K_n : K] = [K_n : K_{n-1}] \cdots [K_2 : K_1] \cdot [K_1 : K].$$

$(2 \Rightarrow)$ by Remark 6.2.3 (1) and (3) and $(2 \Leftarrow)$ by (1).

$(3 \Rightarrow)$ is trivial. $(3 \Leftarrow)$: let $a \in L$; say a is a root of $b_n X^n + \cdots + b_0 \in M[X]$. Then a is algebraic over $\tilde{M} := K(b_0, ..., b_n)$ and by assumption, the b_i are algebraic over K. By (1), $\tilde{M}(a) \mid K$ is finite. By Remark 6.2.3 (3), a is algebraic over K. □

Exercise 6.2.17. Let $a_0, a_1 \in L$ algebraic over K and

$$n_0 := [K(a_0) : K], \quad n_1 := [K(a_1) : K] = n_1, \quad n := [K(a_0, a_1) : K].$$

Then $n \leqslant n_0 n_1$ and n_0, n_1 both divide n; in particular, $n = n_0 n_1$ if n_0, n_1 are coprime. The same holds more generally for finite tuples $\bar{a}_0, \bar{a}_1$ of algebraic elements.

Examples 6.2.18.

1. The complex roots of $X^3 - 2 \in \mathbb{Q}[X]$ are $\alpha_1 := \sqrt[3]{2}, \alpha_2 := \zeta_3 \alpha_1, \alpha_3 := \zeta_3^2 \alpha_1$ where $\zeta_3 = e^{2\pi i/3}$. Then $[\mathbb{Q}(\alpha_i) : \mathbb{Q}] = 3$ and $[\mathbb{Q}(\alpha_1, \alpha_2) : \mathbb{Q}] = 6 < 3 \cdot 3$.

   Indeed, note $\mathbb{Q}(\alpha_1, \alpha_2) = \mathbb{Q}(\alpha_1, \zeta_3)$ and ζ_3 is a root of $X^2 + X + 1 \in \mathbb{Q}[X]$ (Remark 1.6.9). Hence, $[\mathbb{Q}(\zeta_3) : \mathbb{Q}] = 2$, so $[\mathbb{Q}(\alpha_1, \zeta_3) : \mathbb{Q}] = 6$ by the exercise.
2. $[\mathbb{Q}(i) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ and $[\mathbb{Q}(i, \sqrt{2}) : \mathbb{Q}] = 2 \cdot 2$.

   Indeed: $i \notin \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$, so $[\mathbb{Q}(i, \sqrt{2}) : \mathbb{Q}(\sqrt{2})] = 2$.

Exercise 6.2.19. Let $a \in \mathbb{C}$ be a root of $X^4 + 2X + 2 \in \mathbb{Q}[X]$. Show $\sqrt[3]{2} \notin \mathbb{Q}(a)$.

6.2.1 Relative algebraic closure

Theorem 6.2.20. *The set $\overline{K}^L$ of $a \in L$ that are algebraic over K is an intermediate field of $L \mid K$ called the* relative algebraic closure of K in L.

It contains every $a \in L$ that is algebraic over $\overline{K}^L$.

Proof. Given $a, b \in \overline{K}^L$, say, roots of $f, g \in K[X]$, we have to find polynomials having $a - b$ and ab^{-1} (if $b \neq 0$) as roots. It is difficult to construct such polynomials from f, g. Arguing abstractly is easier: $K(a, b) \mid K$ is finite, hence algebraic by Corollary 6.2.16; thus, $a - b, ab^{-1} \in K(a, b)$ are algebraic over K.

2nd statement: if $a \in L$ is algebraic over $\overline{K}^L$, then $\overline{K}^L(a) \mid \overline{K}^L$ is finite (Theorem 6.2.5), hence algebraic (Corollary 6.2.16 (2)). Since also $\overline{K}^L \mid K$ is algebraic, $\overline{K}^L(a) \mid K$ is algebraic by Corollary 6.2.16 (3); hence, a is algebraic over K, i.e., $a \in \overline{K}^L$. □

Corollary 6.2.21. *If $A \subseteq \overline{K}^L$, then $K(A) \mid K$ is algebraic.*

Proof. Since $A \subseteq \overline{K}^L$ and $\overline{K}^L$ is a field, we have $K(A) \subseteq \overline{K}^L$. □

Example 6.2.22. $\overline{\mathbb{Q}}^{\mathbb{C}} \mid \mathbb{Q}$ is not finitely generated.

Proof. $\sqrt[n]{2}$ has minimal polynomial $X^n - 2$ over $\mathbb{Q}$ (irreducible by Eisenstein). Hence, $[\overline{\mathbb{Q}}^{\mathbb{C}} : \mathbb{Q}] \geqslant [\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}] = n$ (Theorem 6.2.5), so $[\overline{\mathbb{Q}}^{\mathbb{C}} : \mathbb{Q}] = \infty$. As $\overline{\mathbb{Q}}^{\mathbb{C}} \mid \mathbb{Q}$ is algebraic, Corollary 6.2.16 (2) gives the claim. □

6.2.2 Impossibility for ruler and compass

Corollary 6.2.23. *If $z \in \mathbb{C}$ is constructible, then z is algebraic (over $\mathbb{Q}$) and $[\mathbb{Q}(z) : \mathbb{Q}]$ is a power of 2.*

Proof. Choose $\mathbb{Q} \subsetneq K_1 \subsetneq \cdots \subsetneq K_n \ni z$ according Theorem 6.1.5. Then $[K_{i+1} : K_i] = 2$. Hence, $[K_n : \mathbb{Q}] = 2^n$. Since $2^n = [K_n : K(z)] \cdot [K(z) : \mathbb{Q}]$, also $[K(z) : \mathbb{Q}]$ is a power of 2. □

Example 6.2.24. The converse is false: it is known that $f := X^4 - 4X + 2 \in \mathbb{Q}[X]$ has a non-constructible root $z \in \mathbb{C}$; but $[\mathbb{Q}(z) : \mathbb{Q}] = 4$ since f is irreducible by Eisenstein.

Example 6.2.25 (Classical Greek problems, again)**.**

1. The Delian problem is unsolvable: $\sqrt[3]{2}$ is not constructible because $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ is not a power of 2.
2. Squaring the circle is impossible: Lindemann's theorem implies $\sqrt{\pi}$ is not algebraic.
3. Trisecting angles is not always impossible: $e^{\pi i/3}$ is constructible but $e^{\pi i/9}$ is not.

   Indeed: let $\alpha := \pi/9$ and $z := e^{i\alpha} + e^{-i\alpha} = 2\cos(\pi/9)$ and $e^{3i\alpha} + e^{-3i\alpha} = 2\cos(\pi/3) = 1$; hence $z^3 = (e^{i\alpha} + e^{-i\alpha})^3 = e^{3i\alpha} + 3e^{i\alpha} + 3e^{-i\alpha} + e^{-3i\alpha} = 1 + 3z$. Hence, z is a root of $X^3 - 3X - 1 \in \mathbb{Q}[X]$; this is irreducible because it has no root in $\mathbb{Z}$ (Exercise 3.1.15); hence, $[\mathbb{Q}(z) : \mathbb{Q}] = 3$; hence neither z nor $\cos(\alpha) = \mathrm{Re}(e^{i\alpha})$ nor $e^{i\alpha}$ is constructible.

We do not have yet the theoretical means to understand the constructibility of regular n-gons and delay an answer to Section 6.9.1.

6.3 Splitting fields

Lemma 6.3.1 (Kronecker)**.** *Let K be a field and $f \in K[X] \setminus K$. Then there exists a field extension $L \mid K$ such that f has a root in L. Moreover, $[L : K] \leqslant \deg(f)$.*

Proof. Let g be a monic irreducible factor of f and write $g = X^n + a_{n-1}X^{n-1} + \cdots + a_0$. By Example 4.6.3 (3), $K[X]$ is a principal ideal domain and by Remark 4.7.11 (3), (g) is a maximal ideal in $K[X]$. By Lemma 4.8.9 (2), $L := K[X]/(g)$ is a field.

Since the canonical projection $\pi_{(g)}$ is a homomorphism,

$$0_L = \pi_{(g)}(g) = \bar{X}^n + \bar{a}_{n-1}\bar{X}^{n-1} + \cdots + \bar{a}_0$$

where we write $\bar{x} := \pi_{(g)}(x)$. Note $\pi_{(g)}\restriction K : K \to L$ is injective as a field homomorphism. Identifying $a \in K$ with $\bar{a}$ we sloppily view K as a subfield of L. Then $\pi_{(g)}(g) = g(\bar{X})$. Thus $\alpha := \bar{X} \in L$ is a root of g, hence also of f. Moreover: since g is irreducible and monic, $g = m_\alpha^K$ by Lemma 3.5.6. Then $[K(\alpha) : K] = \deg(g) \leqslant \deg(f)$ (in fact, $K(\alpha) = L$). □

Exercise 6.3.2. Let K be a field. Then $f, g \in K[X]$ are not coprime if and only if f, g have a common root in some field extension $L \mid K$. In particular, f and f' are not coprime if and only if f has a multiple root in some field extension $L \mid K$.

Definition 6.3.3. Let K be a field and $f \in K[X]$ of degree $n > 0$. L is a *splitting field of f over K* if $L \mid K$ is a field extension and

1. *f splits in L*: there are $a_1, ..., a_n \in L$ and $b \in K$ such that $f = b(X - a_1)\cdots(X - a_n)$;
2. if M is an intermediate field of $L \mid K$ such that f splits in M, then $M = L$

Remark 6.3.4. Then:

1. $L = K(a_1, ..., a_n)$; in particular, $L \mid K$ is finite (Corollary 6.2.16 (1)).
2. $a \in K$ is the lead coefficient of f, and $\{a_1, ..., a_n\}$ the set of roots of f in L.
3. Since $L[X]$ is factorial: if $g \mid f$ with $g \in L[X]$, then $g = b(X - a_{i_1})\cdots(X - a_{i_r})$ for certain $1 \leqslant i_1, ..., i_r \leqslant n$ and $b \in K$; in particular, g splits in L.

Theorem 6.3.5 (Existence of splitting fields)**.** *Let K be a field. For every $f \in K[X]$ of degree $n > 0$ there exists a splitting field L of f over K with $[L : K] \leqslant n!$.*

Proof. We can assume f is monic. Choose $L_1 \mid K$ of degree $\leqslant n$ with a root $a_1 \in L_1$ of f. In $L_1[X]$ write $f = (X - a_1)f_1$ with $f_1 \in L_1[X]$ and $\deg(f_1) = n - 1$ (Corollary 3.3.2). Choose $L_2 \mid L_1$ of degree $\leqslant n-1$ with a root $a_2 \in L_2$ of f_1. In $L_2[X]$ write $f = (X-a_1)(X-a_2)f_2$ with $f_2 \in L_2[X]$ and $\deg(f_2) = n - 2$. Continue and get $L_n \mid K$ such that $f = (X - a_1)\cdots(X - a_n)$. Note $[L_n : K] = [L_n : L_{n-1}]\cdots[L_2 : L_1][L_1 : K] \leqslant n!$. Set $L := K(a_1, ..., a_n)$. As this is an intermediate field of $L_n \mid K$, its degree divides $[L_n : K]$, so is $\leqslant n!$.

Let M be an intermediate field of $L \mid K$ such that f splits in M, say $f = (X-b_1)\cdots(X-b_n)$ with $b_i \in M$. For every a_i we have $0 = f(a) = (a_i - b_1)\cdots(a_i - b_n)$, so a_i equals some $b_j \in M$. Thus, $L = K(a_1, ..., a_n) \subseteq M \subseteq L$, so $L = M$. □

Example 6.3.6. Recall Example 6.2.18 (1). $\mathbb{Q}(\sqrt[3]{2}, \zeta_3)$ is a splitting field of $X^3 - 2 \in \mathbb{Q}[X]$ over $\mathbb{Q}$ and $[\mathbb{Q}(\sqrt[3]{2}, \zeta_3) : \mathbb{Q}] = 6 = 3!$.

Example 6.3.7. $\mathbb{Q}(\sqrt[8]{2}, i)$ is a splitting field of $X^8 - 2$ over $\mathbb{Q}$, and $[\mathbb{Q}(\sqrt[8]{2}, i) : \mathbb{Q}] = 16 < 8!$.

Proof. The complex roots are $\alpha := \sqrt[8]{2}, \zeta_8\alpha, ..., \zeta_8^7\alpha$. The splitting field contains $\zeta_8 = \alpha^7 \cdot \zeta_8\alpha/2$, so equals $\mathbb{Q}(\alpha, \zeta_8)$. Note $X^8 - 2$ is irreducible (Eisenstein), so $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 8$.

Further, $\zeta_8 \notin \mathbb{Q}(\alpha) \subseteq \mathbb{R}$, so $[\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha)] > 1$. Compute

$$(\zeta_8 + \zeta_8^{-1})^2 = \zeta_8^2 + 2 + \zeta_8^{-2} = i + 2 - i = 2.$$

Hence, $\zeta_8 + \zeta_8^{-1}$ equals one of the square roots $\pm\alpha^4$ of 2 in $\mathbb{Q}(\alpha)$; then $\zeta_8^2 + 1 = \pm\alpha^4 \cdot \zeta_8$, so ζ_8 is a root of a quadratic polynomial in $\mathbb{Q}(\alpha)[X]$. Hence $[\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha)] = 2$. Thus, $[\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}] = [\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha)] \cdot [\mathbb{Q}(\alpha) : \mathbb{Q}] = 2 \cdot 8$.

Finally, $[\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha, i)] = 1$ because $\mathbb{Q}(\alpha, i) \subseteq \mathbb{Q}(\alpha, \zeta_8)$ and

$$2 = [\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha)] = [\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha, i)][\mathbb{Q}(\alpha, i) : \mathbb{Q}(\alpha)] = [\mathbb{Q}(\alpha, \zeta_8) : \mathbb{Q}(\alpha, i)] \cdot 2. \quad \square$$

Exercise 6.3.8. The splitting field of $X^4 - 4X^2 + 2$ over $\mathbb{Q}$ is $\mathbb{Q}(\alpha)$ where $\alpha := \sqrt{2 - \sqrt{2}}$. What is $[\mathbb{Q}(\alpha) : \mathbb{Q}]$?

We now aim to show that splitting fields are unique in some sense.

Remark 6.3.9. Recall Remark 3.1.8: for commutative rings R, S, a homomorphism $\varphi : R \to S$ and $f = a_n X^n + \cdots + a_0 \in R[X]$ we write

$$\varphi(f) := \varphi(a_n)X^n + \cdots + \varphi(a_0).$$

If $a \in R$ is a root of f, then $\varphi(a) \in S$ is a root of $\varphi(f)$ because

$$\varphi(f)(\varphi(a)) = \varphi(a_n)\varphi(a)^n + \cdots + \varphi(a_0) = \varphi(a_n a^n + \cdots + a_0) = \varphi(f(a)) = \varphi(0_R) = 0_S.$$

Theorem 6.3.10. *Let $L \mid K$ and $L' \mid K'$ be field extensions and $\varphi : K \to K'$ a homomorphism. Let $f \in K[X]$ be irreducible and $a \in L$ a root of f.*

1. *For every root $a' \in L'$ of $\varphi(f)$ there is exactly one homomorphism $\psi : K(a) \to L'$ that extends φ and maps a to a'.*
2. *If $\varphi(f)$ has n roots in L', then there are exactly n homomorphisms $\psi : K(a) \to L'$ that extend φ.*

Proof. (2) follows from (1) and the previous remark. (1): uniqueness is easy. Assume ψ, ψ' are homomorphism as stated. We have $K(a) = K[a]$ (Theorem 3.5.8). Let $g(a) \in K[a]$ with $g \in K[X]$. Then $\psi(g(a)) = \varphi(g)(\psi(a)) = \varphi(g)(\psi'(a)) = \psi'(g(a))$.

For existence, let $\chi : K[X] \to L'$ extend φ and map X to a' (Theorem 3.1.7). Then $\chi(f) = \varphi(f)(a') = 0_{L'}$, so $(f) \subseteq \ker(\chi)$. But $\ker(\chi) \neq K[X]$ and (f) is a maximal ideal in $K[X]$ (Remark 4.7.11), so $(f) = \ker(\chi)$. The isomorphism theorem for rings gives a monomorphism $\chi' : K[X]/(f) \to L'$ with $\chi' \circ \pi_{(f)} = \chi$. Then $\chi'(X + (f)) = \chi(X) = a'$ and $\chi'(b + (f)) = \chi(b) = \varphi(b)$ for $b \in K$. By Corollary 4.8.15 (and $(f) = (m_a^K)$), there is $\chi'' : K(a) \cong K[X]/(f)$ mapping a to $X + (f)$ and $b \in K$ to $b + (f)$. Set $\psi := \chi' \circ \chi''$. $\square$

Definition 6.3.11. Let $L \mid K$ and $L' \mid K$ be field extensions. A *K-homomorphism (K-isomorphism)* is a homomorphism (isomorphism) $\varphi : L \to L'$ that fixes K, i.e., $\varphi{\restriction}K = \mathrm{id}_K$.

If $L = L'$, we speak of a *K-endomorphism (K-automorphism) of L*.

Remark 6.3.12. Let $L \mid K$ and $L' \mid K$ be field extensions and $f \in K[X] \setminus K$.

1. If f is irreducible, and $a \in L, a' \in L'$ are roots of f, then there is exactly one K-isomorphism from $K(a)$ onto $K(a')$ with $\varphi(a) = a'$.

   Indeed: apply the theorem with $\varphi = \mathrm{id}_K$; clearly, φ is onto $K(a')$.
2. A K-homomorphism $\varphi : L \to L'$ maps roots of f in L to roots of f in L' (Remark 6.3.9).
3. A K-automorphism φ of L permutes the roots of f in L.

   Indeed by (2), φ maps roots to roots; there are $\leqslant \deg(f)$ many (Corollary 3.3.3); as a field homomorphism φ is injective (Remark 1.1.22 (2)).
4. Let $A \subseteq L$ and $\varphi, \psi : K(A) \to L'$ are K-homomorphisms that agree on A, then $\varphi = \psi$ (Lemma 3.6.10).

Corollary 6.3.13. *Let $L \mid K$ and $L' \mid K$ be field extensions, and assume $L \mid K$ is finite. Then there are $\leqslant [L : K]$ many K-homomorphisms from L to L'.*

Proof. Let $n := [L : K]$ and choose $a_1, \ldots, a_r \in L$ with $K(a_1, \ldots, a_r) = L$. Let $K_i := K(a_1, \ldots, a_i)$ (and $K_0 = K$). By the degree formula $n = n_0 n_1 \cdots n_r$ where $n_0 := 1$ and $n_i := [K_i : K_{i-1}] = \deg(m_{a_i}^{K_{i-1}})$ for $0 < i \leqslant r$. We proceed by induction on r.

For $r = 0$, our claim is trivial. For $r > 0$, we assume inductively that there are $m \leqslant n_1 \ldots n_{r-1}$ many K-homomorphisms φ from K_{r-1} to L'. By Theorem 6.3.10 (2), every φ has exactly m_r many extensions to a homomorphism from K_r to L' where m_r is the number of roots of $\varphi(m_{a_r}^{K_{r-1}})$ in L'. In total, there are $m \cdot m_r$ homomorphisms from $K_r = L$ to L'. But $m_r \leqslant \deg \varphi(m_{a_r}^{K_{r-1}}) = n_r$. Our claim follows. □

Theorem 6.3.14 (Uniqueness of splitting fields)**.** *Let K be a field and $f \in K[X] \setminus K$. Then the splitting field of f over K is unique up to K-isomorphism.*

In fact, if L, L' are splitting fields of f over K, and $a \in L, a' \in L'$ roots of some irreducible factor g of f in $K[X]$, then there is a K-isomorphism from L onto L' that maps a to a'.

Proof. Let L, L' be two splitting fields of f over K. Let $a_1, \ldots, a_n \in L$ and $a'_1, \ldots, a'_n$ list the (not necessarily distinct) roots of f in L, L'. Then $L = K(a_1, \ldots, a_n), L' = K(a'_1, \ldots, a'_n)$. Since roots of g are roots of f, we can assume $a = a_1$. Since $g \mid f$, it has some a'_j as root in L' (Remark 6.3.4), say, $a'_j = a'_1$. By Remark 6.3.12, there is a K-isomorphism $\varphi_1 : K(a_1) \to K(a'_1)$ that maps a_1 to a'_1.

Write $f = (X - a_1) f_1$ where $f_1 \in K(a_1)[X]$. Its roots in L are $a_2, \ldots, a_n$. Let $g_2 \in K(a_1)$ be an irreducible factor of f_1. Some a_j for $j > 1$ is a root of g_2 in L (Remark 6.3.4 (3)). We can assume $j = 2$. We have $f = (X - a'_1)\varphi_1(f_1)$ where $\varphi_1(f_1) \in K(a'_1)$ has roots $a'_2, \ldots, a'_n$ in L', $\varphi_1(g_2) \in K(a'_1)$ is an irreducible factor of $\varphi(f_1)$ and some a'_k for $k > 1$ is a root of $\varphi(g_2)$ in L'. We can assume $k = 2$.

By Theorem 6.3.10, φ_1 extends to a homomorphism $\varphi_2 : K(a_1, a_2) \to L'$ mapping a_2 to a_2'; clearly, φ_2 is a K-isomorphism from $K(a_1, a_2)$ onto $K(a_1', a_2')$. Continue. □

Exercise 6.3.15. Let $L \mid K$ be finite with $[L : K] = n$ and $a \in L$. Assume K-automorphisms of L map a to $\geqslant n$ many values. Show a is primitive, i.e., $K(a) = L$.

Example 6.3.16. Let $L := \mathbb{Q}(\sqrt[4]{2}, i), K := \mathbb{Q}$. Determine $\mathrm{Aut}(L)$ and infer $L = \mathbb{Q}(\sqrt[4]{2} + i)$.

Solution. Write $\alpha := \sqrt[4]{2}$. We have $m_\alpha^{\mathbb{Q}} = X^4 - 2$ (irreducible by Eisenstein) and $m_i^{\mathbb{Q}(\alpha)} = X^2 + 1$ (irreducible as it has no root in $\mathbb{Q}(\alpha) \subseteq \mathbb{R}$). The degree formula gives $[L : K] = 8$. Recall, automorphisms are $\mathbb{Q}$-automorphisms (Exercise 3.4.15), so there are $\leqslant 8$ automorphisms.

Theorem 6.3.10 gives $\psi_1, \psi_2, \psi_3, \psi_4 \in \mathrm{Aut}(\mathbb{Q}(\alpha))$ mapping α to one of the 4 roots $\pm\alpha, \pm i\alpha$ of $m_\alpha^{\mathbb{Q}}$. By the corollary above each ψ_i extends to two automorphisms of L mapping i to $\pm i$. This gives $4 \cdot 2$ automorphisms, so all of them.

Hence, $\alpha + i$ is primitive by the previous exercise: 8 values $\pm\alpha \pm i, \pm i\alpha \pm i$. □

Exercise 6.3.17. Determine $[\mathbb{Q}(\sqrt{3}, \sqrt[3]{3}) : \mathbb{Q}]$ and find all homomorphisms from $\mathbb{Q}(\sqrt{3}, \sqrt[3]{3})$ into $\mathbb{C}$. Which ones are automorphisms of $\mathbb{Q}(\sqrt{3}, \sqrt[3]{3})$?

Exercise 6.3.18. Let $p_1 < \cdots < p_n$ be prime and $K := \mathbb{Q}(\sqrt{p_1}, ..., \sqrt{p_n})$. Show $\sqrt{q} \notin K$ for q a product of other primes (induction on n). Infer $[K : \mathbb{Q}] = 2^n$ and $K = \mathbb{Q}(\sqrt{p_1} + \cdots + \sqrt{p_n})$.

6.3.1 Normal extensions

Splitting fields have the following important property with a nonsensical name:

Lemma 6.3.19. *Let L be the splitting field of $f \in K[X] \setminus K$ over K. Then the field extension $L \mid K$ is* normal*: it is algebraic and for every irreducible $g \in K[X]$, if g has a root in L, then g splits in L.*

Proof. Let $a_1, ..., a_n$ be the roots of f in L. Then $L = K(a_1, ..., a_n)$, so $L \mid K$ is algebraic by Corollary 6.2.16 (1). Let $g \in K[X]$ be irreducible and $b \in L$ a root. Let L' be the splitting field of g over L. We have to show that every root $b' \in L'$ of g is in L.

By the theorem there is a K-automorphism φ of L' that maps b to b'. It suffices to show $\varphi(L) \subseteq L$. But for every $1 \leqslant i \leqslant n$, $\varphi(a_i)$ is a root of $\varphi(f) = f$, so $\varphi(a_i) = a_j \in L$ for some j. Then $\varphi(L) = \varphi(K(a_1, ..., a_n)) = K(\varphi(\{a_1, ..., a_n\})) \subseteq L$ by Lemma 3.6.10 (2). □

Exercise 6.3.20. A partial converse: if $L \mid K$ is a finite and normal field extension with $L \neq K$, then L is a splitting field over K of some $f \in K[X] \setminus K$.

Hint: f is the product of minimal polynomials of generators.

Remark 6.3.21.

1. If $L \mid M \mid K$ are field extensions and $L \mid K$ is normal, then so is $L \mid M$.

   Indeed: let $f \in M[X]$ be irreducible with root $a \in L$. We have to show f splits in L. We can assume f is monic. Then $f = m_a^M$. By Lemma 3.5.6, $f \mid m_a^K$ in $M[X]$ (note a is algebraic over K). As m_a^K splits in L, so does f (Remark 6.3.4).

2. There exist field extensions $L \mid M \mid K$ with $L \mid K$ normal and $M \mid K$ not normal.

   E.g., for $\alpha := \sqrt[3]{2}$ we have $\mathbb{Q}(\alpha, \zeta_3) \mid \mathbb{Q}(\alpha) \mid \mathbb{Q}$ (Example 6.3.6). $\mathbb{Q}(\alpha) \mid \mathbb{Q}$ is not normal: $X^3 - 2$ does not split since $\mathbb{Q}(\alpha)$ it does not contain the other two roots.

   In fact, $X^3 - 2$ factors $(X - \alpha) \cdot (X^2 + \alpha X + \alpha^2)$ in $\mathbb{Q}(\alpha)$.
3. There exist $L \mid M \mid K$ with both $M \mid K$ and $L \mid M$ normal, and $L \mid K$ not normal.

   E.g., $\mathbb{Q}(\sqrt[4]{2}) \mid \mathbb{Q}(\sqrt{2}) \mid \mathbb{Q}$: note that adjoining square roots gives normal extensions and $\mathbb{Q}(\sqrt[4]{2}) \mid \mathbb{Q}$ is not normal, e.g., it does not contain the root $i\sqrt[4]{2}$ of $X^4 - 2$.

   In fact, $X^4 - 2$ factors $(X - \sqrt[4]{2})(X + \sqrt[4]{2})(X^2 + \sqrt{2})$ in $\mathbb{Q}(\sqrt[4]{2})$.

The reason why normality is important is the following:

Lemma 6.3.22. *Let $L \mid K$ be an algebraic field extension. Then $L \mid K$ is normal if and only if $\varphi(L) = L$ for all field extensions $L' \mid L$ and K-homomorphisms $\varphi : L \to L'$.*

Proof. $\Rightarrow$: assume $L \mid K$ is normal and let L', φ be given. We show $\varphi(L) = L$.

For $a \in L$, normality gives $m_a^K = (X - a_1) \cdots (X - a_n)$ for certain $a_i \in L$. Note $m_a^K = \varphi(m_a^K) = (X - \varphi(a_1)) \cdots (X - \varphi(a_n))$. As $\varphi(a)$ is a root of m_a^K we have $\varphi(a) = a_i$ for some i. Conversely, a is a root of $\varphi(m_a^K)$ so equals $\varphi(a_i)$ for some i.

$\Leftarrow$: assume the r.h.s.. We have to show that every $f \in K[X] \setminus K$ with a root $a \in L$ splits in L. We can assume f is irreducible. Let L' be a splitting field of f over L and let $a' \in L'$ a root of f. We have to show $a' \in L$. By Theorem 6.3.10, there is a K-homomorphism $\varphi : K(a) \to L'$ that maps a to a'. By assumption, $\varphi(L) \subseteq L$, so $a' = \varphi(a) \in L$. □

Remark 6.3.23. The same proof works also with the variant of the r.h.s. that has $\varphi(L) \subseteq L$ instead $\varphi(L) = L$.

In the next section we observe that one can make field extensions normal by making them minimally larger:

Theorem 6.3.24 (Normal hull)**.** *Every algebraic field extension $L \mid K$ has a* normal hull *N: $N \mid L$ is a field extension and $N \mid K$ is normal and, if N' is an intermediate field of $N \mid L$ such that $N' \mid K$ is normal, then $N = N'$. Moreover, N is unique up to L-isomorphism.*

6.4 Algebraic closure

Definition 6.4.1. Let K be a field. K is *algebraically closed* if every $f \in K[X] \setminus K$ has a root in K. A field $\overline{K}$ is an *algebraic closure of K* if $\overline{K}$ is algebraically closed and $\overline{K} \mid K$ is an algebraic field extension.

Remark 6.4.2. Let K be a field.

1. K is algebraically closed if and only if the irreducible polynomials in $K[X]$ are precisely the linear ones.

   $\Rightarrow$: if $f \in K[X]$ has degree > 1, then it has a root $a \in K$ and we can write $f = (X-a)g$; then $\deg(g) = \deg(f) - 1 > 0$ and f is reducible.
   $\Leftarrow$: an irreducible factor of a given $f \in K[X] \setminus K$ is linear, say it equals $aX + b$ for $a, b \in K, a \neq 0$; then $-b/a \in K$ is a root of f.
2. If K is algebraically closed if and only if every $f \in K[X] \setminus K$ splits in K.

   $\Leftarrow$ is clear and $\Rightarrow$ follows from (1), since $K[X]$ is factorial (this solves Exercise 3.8.4).
3. K is algebraically closed if and only if $L = K$ for every algebraic field extension $L \mid K$.

   $\Rightarrow$: if $a \in L$, then the irreducible m_a^K is linear and hence $a \in K$.
   $\Leftarrow$: if $f \in K[X] \setminus K$ does not have a root in K, then the splitting field of f is a proper algebraic extension.
4. Let $L \mid K$ be a field extension and L algebraically closed. Then $\overline{K}^L$ is an algebraic closure of K (by Theorem 6.2.20).
5. If $\overline{K}$ is an algebraic closure of K, then $\overline{K} \mid K$ is normal. If $f \in K[X]$ and $a_1, ..., a_n$ lists its roots in $\overline{K}$, then $K(a_1, ..., a_n)$ is the splitting field of f over K

Examples 6.4.3. The fundamental theorem of algebra (Theorem 3.8.3) states that $\mathbb{C}$ is algebraically closed, so $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$.

By (4) above, $\overline{\mathbb{Q}}^{\mathbb{C}}$ is an algebraic closure of $\mathbb{Q}$.

Theorem 6.4.4 (Embedding theorem)**.** *Let $L \mid K$ be an algebraic field extension, M an intermediate field, $\overline{K}$ an algebraic closure of K and $\varphi : M \to \overline{K}$ a K-homomorphism.*

Then there exists a K-homomorphism $\psi : L \to \overline{K}$ that extends φ. In particular, there exists a K-homomorphism $\varphi : L \to \overline{K}$.

Proof. Consider the set H of K-homomorphisms $\chi : N \to \overline{K}$ whose domain N is an intermediate field of $L \mid M$ and that extend φ.

View each χ as a set of ordered pairs. Then H is partially ordered by $\subsetneq$. Further, $(H, \subsetneq)$ is inductive: every chain $C \subseteq H$ has upper bound $\chi^* := \bigcup C = \bigcup_{\chi \in C} \chi \in H$.

By Zorn's lemma, H contains a maximal element ψ. We claim $N = L$ for its domain N. Otherwise choose $a \in L \setminus N$. Then $L \mid N$ is algebraic (Corollary 6.2.16) and $m_a^N \in N[X]$ irreducible. Choose a root $a' \in \overline{K}$ of $\psi(m_a^N) \in \overline{K}[X]$. Theorem 6.3.10 gives a K-homomorphism $\chi : N(a) \to \overline{K}$ (with $\chi(a) = a'$) that extends ψ, contradicting maximality. □

Theorem 6.4.5. *Every field K has an algebraic closure $\overline{K}$, unique up to K-isomorphism.*

Proof. Uniqueness: for another algebraic closure $\tilde{K}$, the embedding theorem gives a K-homomorphism $\varphi : \tilde{K} \to \overline{K}$. Then $\varphi(\tilde{K})$ is an intermediate field of $\overline{K} \mid K$ and algebraically closed. By Corollary 6.2.16 (3), $\overline{K} \mid \varphi(\tilde{K})$ is algebraic. By Remark 6.4.2 (3), $\varphi(\tilde{K}) = \overline{K}$.

Existence: write $P := K[X] \setminus K$. For $f \in P$ let X_f be a variable and set $R := K[X_f, f \in P]$; note $f(X_f) \in R$ for all $f \in P$. Let

$$I := \big(\{f(X_f) \mid f \in P\}\big).$$

We claim I is a proper ideal of R. Otherwise $1 \in I$, so $1 = g_1 f_1(X_{f_1}) + \cdots + g_n f_n(X_{f_n})$ for some $n \in N$ and $f_i \in P$ and $g_i \in R$. Let L be the splitting field of $f := f_1 \cdots f_n$ and $a_i \in L$ a root of $f_i(X_{f_i})$. Plug a_i for X_{f_i} and 0_L for all other variables X_f (evaluation homomorphism). Then $1 = 0$ in L, a contradiction.

By Theorem 4.7.15, I is contained in a maximal ideal I^* of R. Then R/I^* is a field (Lemma 4.8.9). Identifying $x \in K$ with $x + I^*$ we view K as a subfield of R/I^*. Every $f \in P$ has root $x_f := X_f + I^* \in R/I^*$ because $f(X_f + I^*) = f(X_f) + I^* = I^* = 0_{R/I^*}$. Set

$$K_1 := K\big(\{x_f \mid f \in P\}\big).$$

By Corollary 6.2.21, $K_1 \mid K$ is algebraic. Repeat the construction with K_1 in place of K and get an algebraic field extension $K_2 \mid K_1$ such that every $f \in K_1[X] \setminus K_1$ has a root in K_2. And so on: $K_0 := K \subseteq K_1 \subseteq K_2 \subseteq \cdots$. Define

$$\overline{K} := \bigcup_{n \in \mathbb{N}} K_n.$$

It is easy to see that $\overline{K}$ is a field. The field extension $\overline{K} \mid K$ is algebraic: if $x \in \overline{K}$, then $x \in K_n$ for some n; but $K_n \mid K$ is algebraic by Corollary 6.2.16 (3).

$\overline{K}$ is algebraically closed: let $f \in \overline{K}[X] \setminus \overline{K}$; choose $n \in \mathbb{N}$ such that the finitely many coefficients of f are in K_n, i.e., $f \in K_n[X]$; then f has a root in K_{n+1}, hence in $\overline{K}$. □

Remark 6.4.6. The existence of algebraically closed field extensions has deeper reasons. They exemplify so-called *existentially closed* structures constructed in mathematical logic.

Lemma 6.4.7 (Homogeneity)**.** *Let K be a field and $a, b \in \overline{K}$ be roots of an irreducible $f \in K[X]$. Then there exists a K-automorphism φ of $\overline{K}$ with $\varphi(a) = b$.*

Proof. By Theorem 6.3.10 there exists a K-homomorphism $\psi : K(a) \to \overline{K}$ with $\psi(a) = b$. By the embedding theorem there is a K-homomorphism $\varphi : \overline{K} \to \overline{K}$ extending ψ. Then $\varphi(\overline{K})$ is an intermediate field of $\overline{K} \mid K$ and algebraically closed. By Remark 6.4.2 (3), $\varphi(\overline{K}) = \overline{K}$, so φ is a K-automorphism of $\overline{K}$. □

We now revisit the concept of normality.

Lemma 6.4.8. *Let L be an intermediate field of $\overline{K} \mid K$. Then the following are equivalent.*

1. *$L \mid K$ is normal.*
2. *For every irreducible $f \in K[X]$, L contains either all or none of the roots of f in $\overline{K}$.*
3. *$\varphi(L) = L$ for every K-automorphisms φ of $\overline{K}$.*

Proof. By Corollary 6.2.16 (3), $L \mid K$ is algebraic. $1 \Rightarrow 3$ follows from Lemma 6.3.22.

$3 \Rightarrow 2$: assume $f \in K[X]$ is irreducible, $a, b \in \overline{K}$ are roots and $a \in L$ and $b \notin L$. By homogeneity there is a K-automorphism φ of $\overline{K}$ with $\varphi(a) = b$. Then $\varphi(L) \neq L$.

$2 \Rightarrow 1$: let $f \in K[X]$ be irreducible with a root in L. Write $f = b(X - a_1) \cdots (X - a_n)$ for certain $a_i \in \overline{K}, b \in K$ (Remark 6.4.2 (2)). By (2), all $a_i \in L$, so f splits in L. □

Proof of Theorem 6.3.24. Existence: let $A \subseteq \overline{L}$ be the set of all roots of all irreducible $g \in K[X]$ that have a root in L. Then $N := L(A)$ is a normal hull of $L \mid K$.

Uniqueness: let $\tilde{N}$ be a normal hull of $L \mid K$. By the embedding theorem, there is an L-homomorphism $\varphi : \tilde{N} \to \overline{L}$. Then $\tilde{N}$ is L-isomorphic to $\varphi(\tilde{N})$, so $\varphi(\tilde{N})$ is a normal hull of $L \mid K$. Thus, $\varphi(\tilde{N}) = N$. Indeed, as $\varphi(\tilde{N}) \mid K$ is normal, $A \subseteq \varphi(\tilde{N})$, so $N \subseteq \varphi(\tilde{N})$ and N is an intermediate field of $\varphi(\tilde{N}) \mid L$; since $\varphi(\tilde{N})$ is a normal hull, $N = \varphi(\tilde{N})$. □

6.5 Finite fields

Proposition 6.5.1. *The cardinality of a finite field K is a power of* $\mathrm{char}(K)$.

Proof. $p := \mathrm{char}(K) > 0$ because K is finite. By Theorem 3.4.14, we can assume K has prime field $\mathbb{F}_p$. As K is finite, $n := [K : \mathbb{F}_p]$ is finite. Then K as an $\mathbb{F}_p$-vector space is isomorphic to $\mathbb{F}_p^n$, so has cardinality p^n. □

Especially important for computer science is the case $\mathrm{char}(K) = 2$ and $[K : \mathbb{F}_2] = n$: this endows the set of binary strings of length n with the structure of a field or an $\mathbb{F}_2$-vector space. We now show such fields exist for all $n > 0$ and, in fact, classify all finite fields. The proof is done by connecting the dots of many things we learned so far. It also exemplifies how group theory is useful for field theory.

Theorem 6.5.2 (Classification of finite fields)**.** *Let $p, n, k > 0$ with p prime and let $q := p^n$.*

1. *There exists an up to isomorphism unique field $\mathbb{F}_q$ with q elements, namely the splitting field of $X^q - X$ over $\mathbb{F}_p$; it consists of the roots of this polynomial.*
2. *There exist irreducible $f \in \mathbb{F}_p[X]$ of degree n. For any such f we have $\mathbb{F}_q \cong \mathbb{F}_p[X]/(f)$ and every root of f in $\mathbb{F}_q$ is a primitive element of $\mathbb{F}_q \mid \mathbb{F}_p$.*
3. *$\mathbb{F}_q$ has a subfield of cardinality p^k, i.e., isomorphic to $\mathbb{F}_{p^k}$, if and only if $k \mid n$.*

Proof. (1): let L be the splitting field of $X^q - X$ over $\mathbb{F}_p$ and $a_1, \ldots, a_q \in L$ list the roots.

We claim the a_i are pairwise distinct. Otherwise some a_i is a multiple root of $X^q - X$. By Lemma 3.3.13, a_i is a root of $(X^q - X)' = qX - 1$. But $qa_i = 0$ in L since $\mathrm{char}(L) = p \mid q$.

We next claim $\{a_1, \ldots, a_q\}$ is a subfield of L – then $L = \{a_1, \ldots, a_q\}$ is a field of cardinality q. We have to show: if $a, b \in L$ are roots of $X^q - X$, then so are $ab, a^{-1}, -1$ and $a + b$. That ab, a^{-1} are roots is clear. $-1^q = -1$ if q is odd; if q is even, $p = 2$ and $1 = -1$. For $a + b$ we have to show $(a + b)^q = a + b$, i.e., $= a^q + b^q$. By Lemma 3.4.16, this is true for $q = p$, i.e., $n = 1$. For $n > 1$ it follows by induction.

Uniqueness: if K is a field with q elements, its characteristic is p by the proposition, so we can assume K extends $\mathbb{F}_p$. Since $K^\times$ has order $q - 1$, every $a \in K$ is a root of $X^q - X$; hence, K is also a splitting field, so $K \cong L$ (Theorem 6.3.14).

(2): by Corollary 5.3.25, $\mathbb{F}_q^\times$ is cyclic, say, generated by $a \in \mathbb{F}_q$. Clearly, $\mathbb{F}_q = \mathbb{F}_p(a)$. By Theorem 6.2.5, $n = [\mathbb{F}_q : \mathbb{F}_p] = \deg(m_a^{\mathbb{F}_p})$, so $m_a^{\mathbb{F}_p}$ is irreducible of degree n.

Let $f \in \mathbb{F}_p[X]$ be such and $b \in \mathbb{F}_q$ a root. Then $\mathbb{F}_p[X]/(f) \cong \mathbb{F}_p(b)$ and $[\mathbb{F}_p(b) : \mathbb{F}_p] = n$, so $|\mathbb{F}_p(b)| = q$, so $\mathbb{F}_p(b) = \mathbb{F}_q$.

(3): let K be a subfield of $\mathbb{F}_q$. Then $K \mid \mathbb{F}_p$, so $|K| = p^k$ for some $k \leqslant n$ by the proposition. Then $k = [K : \mathbb{F}_p]$ divides $n = [\mathbb{F}_q : \mathbb{F}_p]$ by the degree formula.

Conversely, assume $k \mid n$, say $n = k\ell$. Let $q' := p^k$. Then $q' - 1 \mid q - 1$ because

$$(p^k)^\ell - 1 = (p^k - 1) \cdot ((p^k)^{\ell-1} + \cdots + p^k + 1).$$

As $\mathbb{F}_q^\times$ is cyclic of order $q-1$, Theorem 5.3.21 gives a subgroup $U \subseteq \mathbb{F}_q^\times$ of order $q'-1$. Every $x \in U$ has order dividing $|U| = q' - 1$ by Lagrange, so $x^{q'} = x$. Hence, $U \cup \{0\}$ is a set of q' many roots of $X^{q'} - X$ in $\mathbb{F}_q$. We saw above that this set is a subfield. □

Exercise 6.5.3. Let $p, k, n > 0$, p prime and $k \mid n$. Show $X^{p^k} - X \mid X^{p^n} - X$ and infer from this that $\mathbb{F}_{p^n}$ has a subfield $\cong \mathbb{F}_{p^k}$.

Corollary 6.5.4. *Let $p, n, k > 0$ with p prime and $q := p^n$. Let $f \in \mathbb{F}_p[X]$ be irreducible of degree k. Then f is a factor of $X^q - X$ if and only if $k \mid n$.*

Proof. $\Rightarrow$: by Remark 6.3.4, f splits in $\mathbb{F}_q$, so has root $a \in \mathbb{F}_q$. Then $\mathbb{F}_p(a) \subseteq \mathbb{F}_q$ and $[\mathbb{F}_p(a) : \mathbb{F}_p] = \deg(f) = k$ divides $[\mathbb{F}_q : \mathbb{F}_p] = n$ by the degree formula.

$\Leftarrow$: let $a \in \overline{F}_p$ be a root of f. Then $[\mathbb{F}_p(a) : \mathbb{F}_p] = k$ and $\mathbb{F}_q$ contains a field isomorphic to $\mathbb{F}_p(a)$. Then f has a root $b \in \mathbb{F}_q$. As b is a root of $X^q - X$, $f \mid X^q - X$ by Corollary 3.5.7. □

Theorem 6.5.5. *Let p be a prime. Recursively, let $F_0 := \mathbb{F}_p$ and choose F_{n+1} as an extension of F_n and isomorphic to $\mathbb{F}_{p^{(n+1)!}}$. Then $\overline{\mathbb{F}}_p \cong \bigcup_n F_n$.*

Proof. The F_n are well-defined by (3) of the theorem. It is easy to see that $\bigcup_n F_n$ is a field. By Theorem 6.4.5 it suffices to show $\bigcup_n F_n$ is an algebraic closure of $\mathbb{F}_p$.

To see $\bigcup_n F_n \mid \mathbb{F}_p$ is algebraic, let $x \in \bigcup_n F_n$ and choose $n \in \mathbb{N}$ with $x \in F_n$. As $F_n \mid \mathbb{F}_p$ is finite, it is algebraic, so x is algebraic over $\mathbb{F}_p$.

To see $\bigcup_n F_n$ is algebraically closed, are looking for a root of a given $f \in \bigcup_n F_n[X] \setminus \bigcup_n F_n$. Choose $n \in \mathbb{N}$ with $f \in F_n[X]$. We can assume f is irreducible in $F_n[X]$. By Kronecker's lemma there is a finite field extension $L \mid F_n$ where f has a root. By the theorem, $L \cong \mathbb{F}_{p^k}$ for some k (with $n! \mid k$). By Corollary 3.5.7, $f \mid X^{p^k} - X$ in $F_n[X]$. Now, L is isomorphic to a subfield F of $F_k \cong \mathbb{F}_{p^{k!}}$. The isomorphism maps F_n onto F_n (as the set of roots of $X^{p^{n!}} - X$), so F extends F_n. As F is the splitting field of $X^{p^k} - X$ over $\mathbb{F}_p$ and $f \mid X^{p^k} - X$ in $F_n[X]$, also f splits in F. □

Remark 6.5.6. The sequence F_n above has a poorly motivated ad hoc definition; more generically, one can define $\overline{\mathbb{F}}_p$ as a so-called *direct limit* of the fields $\mathbb{F}_{p^n}$.

Example 6.5.7. $\mathbb{F}_{2^3}$ is the set of roots of $X^8 - X$ which factors over $\mathbb{F}_2[X]$ as

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

Then $\mathbb{F}_{2^3} \cong \mathbb{F}_2[X]/(f)$ for f one of the degree 3 factors, say, $f := X^3 + X + 1$. Let $x \in \overline{\mathbb{F}}_2$ be a root of f, so $\mathbb{F}_{2^3} \cong \mathbb{F}_2(x)$. The elements of $\mathbb{F}_{2^3}$ are the $\mathbb{F}_2$-linear combinations of $1, x, x^2$. Here is a vector space isomorphism from $\mathbb{F}_2(x)$ onto $\mathbb{F}_2^3$:

0	1	x	x^2	$x^3 = 1 + x$	$x^4 = x + x^2$	$x^5 = 1 + x + x^2$	$x^6 = 1 + x^2$
$(0,0,0)$	$(1,0,0)$	$(0,1,0)$	$(0,0,1)$	$(1,1,0)$	$(1,0,1)$	$(0,1,1)$	$(1,1,1)$

Calculations use the rules $1+1=0$ and $x^3=1+x$ (note $x=-x$). E.g., $(1+x)(x+x^2)=x+x^2+x^2+x^3=x+x^3=x+(1+x)=1$, so $(1+x)^{-1}=x+x^2$. Such calculations verify the stated equalities for the powers of x. We see x generates the group $\mathbb{F}_2(x)^\times$.

Example 6.5.8. Both X^2+1 and X^2+X-1 are irreducible over $\mathbb{F}_3=\{0,\pm1\}$. For respective roots $x,y\in\overline{\mathbb{F}}_3$ we have $\mathbb{F}_{3^2}\cong\mathbb{F}_3(x)\cong\mathbb{F}_3(y)$. Calculations use the rules $1+1=-1$ and $x^2=-1$, resp., $y^2=1-y$. E.g., $y^3=y\cdot y^2=y-y^2=y-1+y=-1-y$.

Then $x^4=1$ in $\mathbb{F}_3(x)^\times$ while y generates $\mathbb{F}_3(y)^\times$:

1	y	$y^2=1-y$	$y^3=-1-y$	$y^4=-1$	$y^5=-y$	$y^6=-1+y$	$y^7=1+y$
$(1,0)$	$(0,1)$	$(1,-1)$	$(-1,-1)$	$(-1,0)$	$(0,-1)$	$(-1,1)$	$(1,1)$

Example 6.5.9. $\mathbb{F}_{2^4}$ is the set of roots of $X^{16}-X$ which factors over $\mathbb{F}_2[X]$ as

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

Then $\mathbb{F}_{2^4}\cong\mathbb{F}_2(x)\cong\mathbb{F}_2(y)$ for roots $x,y\in\overline{\mathbb{F}}_2$ of X^4+X+1 and X^4+X^3+1. Calculations use the rules $x^4=1+x$, resp., $y^4=1+y^3$. Both x and y generate $\mathbb{F}_2(x)^\times$, resp., $\mathbb{F}_2(y)^\times$:

$x^4=1+x$	$x^5=x+x^2$	$x^6=x^2+x^3$	$x^7=1+x+x^3$	$x^8=1+x^2$	...
$(1,1,0,0)$	$(0,1,1,0)$	$(0,0,1,1)$	$(1,1,0,1)$	$(1,0,1,0)$	

$y^4=1+y^3$	$y^5=1+y+y^3$	$y^6=1+y+y^2+y^3$	$y^7=1+y+y^2$	$y^8=y+y^2+y^3$	...
$(1,0,0,1)$	$(1,1,0,1)$	$(1,1,1,1)$	$(1,1,1,0)$	$(0,1,1,1)$	

Exercise 6.5.10. Let $0,1,a,b$ list $\mathbb{F}_4$. Determine the tables for $+,\cdot$. Verify $f:=X^4+X+1=(X^2+X+a)(X^2+X+b)$ in $\mathbb{F}_4[X]$ and show f is irreducible in $\mathbb{F}_2[X]$. Determine the degree of the splitting field of f over $\mathbb{F}_4$.

6.5.1 Reed-Solomon codes

Reed-Solomon codes are the key component of how data are stored on CDs or DVDs. Instead of the given data x one stores an *error-correcting code* $C(x)$ of x such that x can be retrieved even after considerable parts of the code $C(x)$ are damaged.

Let $q\geqslant m>n$ with q a power of a prime p and consider the data as a vector $x=(x_0,...,x_{n-1})\in\mathbb{F}_q^n$, that is, a length n word over an alphabet of size q. Fix pairwise distinct $a_1,...,a_m\in\mathbb{F}_q$ – in practice, one often uses powers of a primitive element of $\mathbb{F}_q\mid\mathbb{F}_p$. Define

$$C(x):=(p_x(a_1),...,p_x(a_m))\in\mathbb{F}_q^m,\qquad\text{where } p_x:=x_{n-1}X^{n-1}+\cdots+x_0\in\mathbb{F}_q[X].$$

Note $C:\mathbb{F}_q^n\to\mathbb{F}_q^m$ is linear: $C(x)=Ax$ where A is the $(m\times n)$ Vandermonde matrix with i-th row $1,a_i,\cdots,a_i^{n-1}$. The key observation is this: for distinct $x,y\in\mathbb{F}_q^k$, the codes $C(x),C(y)$ have the same i-th entry if and only if a_i is a root of p_x-p_y, a polynomial of degree $<n$. Hence, $C(x)$ and $C(y)$ differ in $\geqslant m-n+1$ entries.

This means that $\leqslant(m-n)/2$ errors can be corrected: x is uniquely determined by any tuple obtained by changing $\leqslant(m-n)/2$ entries of $C(x)$.

Remark 6.5.11. We cheated by using many letters – if only bits are used, one additionally codes the letters in $\mathbb{F}_q$ by bit strings.

Sipser and Spielman's *expander codes* (1996) code n bits by $m \leqslant O(n)$ bits, correct a constant fraction of errors and have very efficient encoding and decoding algorithms.

6.6 Separable extensions

Definition 6.6.1. Let K be a field. $f \in K[X] \setminus K$ is *inseparable (over K)* if there exists an irreducible factor $g \in K[X]$ of f that has a multiple root in $\overline{K}$; otherwise, f is *separable*.

Lemma 6.6.2. *Let K be a field and $f \in K[X]$ irreducible. The following are equivalent.*

1. *f is inseparable.*
2. *f has $< \deg(f)$ many roots in $\overline{K}$.*
3. *f has $< \deg(f)$ many roots in some $L \mid K$ such that f splits in L.*
4. *f has a multiple root in some field extension $L \mid K$.*
5. *f has formal derivative $f' = 0$.*

Proof. $1 \Rightarrow 2$: the unique factorization of f in $\overline{K}[X]$ reads $f = b(X - a_1)\cdots(X - a_n)$ where for $n := \deg(f) > 0$, $b \in K$ and $a_i \in \overline{K}$ (Remark 6.4.2 (1)). The a_i are all roots of f in $\overline{K}$. If $a \in \overline{K}$ is a multiple root of f, then $a = a_i = a_j$ for some $i \neq j$.

$2 \Rightarrow 3 \Rightarrow 4$ are trivial and $5 \Rightarrow 1$ follows from Lemma 3.3.13 (1). $4 \Rightarrow 5$: by Exercise 6.3.2, f, f' have a common divisor g of positive degree. Since f is irreducible, f, g are associate, so $\deg(f) = \deg(g)$. As $\deg(f') < \deg(f)$, $g \mid f'$ implies $f' = 0$. □

Remark 6.6.3. For monic f we defined the discriminant D_f in Example 3.7.11 and noted that $D_f = 0$ is equivalent to (4).

Example 6.6.4. For an example of an irreducible inseparable f we need an infinite field of positive characteristic p. Set $K := \mathbb{F}_p(T)$. By Eisenstein, $f := X^p - T$ is irreducible in $K[X]$ and $\mathrm{Quot}(\mathbb{F}_p[T])[X] = K[X]$. f is inseparable since $f' = pX^{p-1} = 0$.

In fact, let $a \in \overline{K}$ be a root of f; then a is a multiple root of f by Lemma 3.4.16:

$$(X - a)^p = X^p - a^p = X^p - T.$$

This just plugs X^p for X in the irreducible separable $X - T$. All examples are like this:

Proposition 6.6.5. *Let K be a field with $p := \mathrm{char}(K) > 0$ and $f \in K[X]$ be irreducible. Then f is inseparable if and only if $f = g(X^{p^n})$ for some $n > 0$ and some irreducible separable $g \in K[X]$.*

Proof. $\Leftarrow$: clearly $f' = 0$. $\Rightarrow$: by Lemma 3.3.11 (3), $f = g_1(X^p)$ for some $g_1 \in K[X]$. With f also g_1 is irreducible. If g_1 is separable, we are done. Otherwise, repeat and write $g_1 = g_2(X^p)$ for some $g_2 \in K[X]$. And so on. As the degrees of the g_is decrease this process stops with a separable g_n. Then $f = g_1(X^p) = g_2((X^p)^p) = \cdots = g_n(X^{p^n})$. □

We now explain the mysterious Definition 3.4.17.

Proposition 6.6.6. *A field K is perfect if and only if every $f \in K[X] \setminus K$ is separable.*

Proof. Equivalently, in the r.h.s., require separability only for irreducible f. $\Rightarrow$: if $\mathrm{char}(K) = 0$, then Lemma 6.6.2 (5) is false by Lemma 3.3.11 (2). Assume $\mathrm{char}(K) =: p > 0$ and the Frobenius endomorphism is surjective. If f is inseparable, then $f' = 0$, so $f = g(X^p)$ for some $g = a_m X^m + \cdots + a_0 \in K[X]$ by Lemma 3.3.11 (3); but $a_i = b_i^p$ for certain $b_i \in K$; then f is reducible: $f = (b_m X)^{pm} + (b_{m-1}X)^{p(m-1)} + \cdots + b_0^p = (b_m X^m + \cdots + b_0)^p$.

$\Leftarrow$: if K is not perfect, then $p := \mathrm{char}(K) > 0$ and there is $x \in K \setminus K^p$. Choose a root $a \in \overline{K}$ of $X^p - x$. Then $m_a^K \mid X^p - x$ by Lemma 3.5.6. Then $m_a^K = (X-a)^d$ for some $d \in \mathbb{N}$ because $X^p - x = (X-a)^p$. As $a \notin K$, Theorem 6.2.5 implies $d = \deg(m_a^K) = [K(a) : K] > 1$. Thus m_a^K is irreducible and inseparable. □

Hence, separability is automatic in most cases of interest. In other words, it is annoying because for a general theory we need to pay attention.

Definition 6.6.7. A field extension $L \mid K$ is *separable* if all $a \in L$ are *separable over* K, i.e., a is algebraic over K and m_a^K is separable.

Remark 6.6.8. Let M be an intermediate field of $L \mid K$. If $a \in L$ is separable over K, then a is also separable over M: a multiple root of m_a^M would also be one of m_a^K because $m_a^M \mid m_a^K$ by Lemma 3.5.6.

Exercise 6.6.9. Let $L \mid K =: K_0$ be a field extension, $r \in \mathbb{N}$ and $a_1, ..., a_r \in L$ such that for all $1 \leqslant i \leqslant r$, a_i is separable over $K_{i-1} := K(a_1, ..., a_{i-1})$. Let $n_i := [K_i : K_{i-1}]$. Then there are exactly $n_1 \cdots n_r$ many K-homomorphisms from K_r to $\overline{K}$.

Hint: follow the proof of Corollary 6.3.13.

Theorem 6.6.10. *Let $L \mid K$ be a finite field extension and $n := [L : K]$. Then $L \mid K$ is separable if and only if there are exactly n many K-homomorphisms from L to $\overline{K}$.*

Proof. $\Rightarrow$: write $L = K(a_1, ..., a_r)$ by Corollary 6.2.16 (2). Each a_i is separable over K, so also over $K(a_1, ..., a_{i-1})$ (Remark 6.6.8). The exercise gives $n = n_1 \cdots n_r$ many K-homomorphisms. But $n = [L : K]$ by the degree formula.

$\Leftarrow$: assume $a \in L$ is not separable over K. Let $d := \deg(m_a^K) = [K(a) : K]$ and note $n = [L : K(a)] \cdot d$ by the degree formula. By Lemma 6.6.2 (2), m_a^K has $< d$ many roots in $\overline{K}$. By Theorem 6.3.10 (2) there are $< d$ many K-homomorphisms $\varphi : K(a) \to \overline{K}$. It suffices to show that every such φ has $\leqslant [L : K(a)]$ extensions to an homomorphism from L to $\overline{K}$. Identifying $\varphi(K(a)) \subseteq \overline{K}$ with $K(a)$, we count $K(a)$-homomorphisms from L to $\overline{K}$. Corollary 6.3.13 states there are $\leqslant [L : K(a)]$ many. □

The exercise thus implies:

Corollary 6.6.11. *Let $L \mid K$ be a field extension, $r \in \mathbb{N}$ and $a_1, ..., a_r \in L$ such that for all $1 \leqslant i \leqslant r$, a_i is separable over $K(a_1, ..., a_{i-1})$. Then $K(a_1, ..., a_n) \mid K$ is separable.*

Corollary 6.6.12. *Let $L \mid K$ be a field extension and M an intermediate field. Then $L \mid K$ is separable if and only if both $L \mid M$ and $M \mid K$ are separable.*

Proof. $\Rightarrow$: by Corollary 6.2.16 (3) and Remark 6.6.8. $\Leftarrow$: by Corollary 6.2.16 (3), $L \mid K$ is algebraic. Let $a \in L$. Since $L \mid M$ is separable, m_a^M is separable over M. Let $a_1, ..., a_n \in M$ be the coefficients of m_a^M, so $m_a^M \in M'[X]$ for $M' := K(a_1, ..., a_n)$. Then a is separable over M'. As $M \mid K$ is separable, all a_i are separable over K. By Remark 6.6.8, a_2 is separable over $K(a_1)$, a_3 over $K(a_1, a_2)$ and so on. By the previous corollary, $K(a_1, ..., a_n, a) \mid K$ is separable, so a is separable over K. □

Exercise 6.6.13. Let $L \mid K$ be a field extension.

1. The set $L_{\mathrm{sep},K}$ of $a \in L$ that are separable over K is an intermediate field.
2. If $a \in L$ is separable over $L_{\mathrm{sep},K}$, then $a \in L_{\mathrm{sep},K}$.
3. $K(A) \mid K$ is separable for $A \subseteq L_{\mathrm{sep},K}$.

Remark 6.6.14. Some modes of speech that we shall not employ: $L_{\mathrm{sep},K}$ is called the *relative separable closure of K in L*. For $L := \overline{K}$ it is the *separable closure of K*. $[L_{\mathrm{sep},K} : K]$ is the *separable degree of $L \mid K$*; one can prove the degree formula for it.

6.6.1 Primitive element theorem

Theorem 6.6.15 (Steinitz 1910)**.** *Every finite, separable field extension is simple.*

Proof. Let $L \mid K$ be a finite, separable field extension. If K is finite, then so is L and a primitive element is a generator of $L^\times$ (Corollary 5.3.25). Assume K is infinite.

Write $L = K(a_1, ..., a_n)$ for $n \in \mathbb{N}$ and $a_i \in L$. If $n < 2$ there is nothing to show. Inductively, $L = K(a_1, ..., a_{n-1})(a_n) = K(a)(a_n)$ for some $a \in L$ – hence, we are left with the case $n = 2$, i.e., $L = K(a_1, a_2)$. We can assume $a_1, a_2 \notin K$, so $m_{a_1}^K, m_{a_2}^K$ have degree > 1.

We are looking for $b \in L$ such that $L = K(b)$. For $c \in K$ we try $b := a_1 + ca_2$. Call c *bad* if $K(b) \subsetneq K(a_1, a_2)$. It suffices to show that there are only finitely many bad c.

Assume c is bad. Then $a_2 \notin K(b)$ as otherwise also $a_1 \in K(b)$ and $K(b) = K(a_1, a_2)$. Hence, $m_{a_2}^{K(b)}$ has degree > 1. By Remark 6.6.8, a_2 is separable over $K(b)$. Hence, $\overline{L}$ contains a root $a_2' \neq a_2$ of $m_{a_2}^{K(b)}$. By Theorem 6.3.10, there is a $K(b)$-homomorphism $\varphi : K(b, a_2) \to \overline{L}$ with $\varphi(a_2) = a_2'$. Then $b = \varphi(b)$, i.e., $a_1 + ca_2 = \varphi(a_1 + ca_2) = \varphi(a_1) + ca_2'$, so $c = (\varphi(a_1) - a_1)/(a_2 - a_2')$. Since $\varphi(a_1), a_2'$ are roots of $m_{a_1}^{K(b)}, m_{a_2}^{K(b)}$, they are also roots of $m_{a_1}^K, m_{a_2}^K$ (Lemma 3.5.6).

We see that every bad c is of the form $(a_1' - a_1)/(a_2 - a_2')$ for roots a_1', a_2' of $m_{a_1}^K, m_{a_2}^K$. Thus, there are only finitely many bad c. □

Corollary 6.6.16. *A separable field extension $L \mid K$ is finite if and only if there is $n \in \mathbb{N}$ such that for all $a \in L$ we have $[K(a) : K] \leqslant n$.*

Proof. $\Rightarrow$: by the degree formula, $[K(a):K] \mid n := [L:K]$.

$\Leftarrow$: assume $[L:K] = \infty$ and let $n > 0$; choose $a_1 \in L \smallsetminus K_0, K_0 := K$, $a_2 \in L \smallsetminus K_1, K_1 := K(a_1)$, and so on. This is possible because $L = K(a_1, ..., a_n)$ implies $[L:K] < \infty$ by Corollary 6.2.16 (1). Then $[K_{i+1}:K_i] \geqslant 2$, so $[K_n:K] = [K_n:K_{n-1}] \cdots [K_1:K_0] \geqslant 2^n > n$. Each a_{i+1} is separable over K, hence K_i (Corollary 6.6.12), so $K_n \mid K$ is separable (Corollary 6.6.11) By the theorem, $K_n = K(a)$ for some $a \in L$. □

Example 6.6.17. We look for primitive elements of $\mathbb{Q}(\alpha, \zeta_3) \mid \mathbb{Q}$ from Example 6.2.18 where $\alpha := \sqrt[3]{2}$. We have $m_\alpha^{\mathbb{Q}} = X^3 - 2$ with complex roots $\alpha, \alpha\zeta_3, \alpha\zeta_3^2$, and $m_{\zeta_3}^{\mathbb{Q}} = X^2 + X + 1$ (Example 4.5.9 (2)) with roots ζ_3, ζ_3^2. We want a primitive element of the form $\alpha + c\zeta_3$ and see $c \neq 0$ is bad in the sense of Steinitz' theorem only for the values

$$\frac{\alpha\zeta_3 - \alpha}{\zeta_3 - \zeta_3^2} = -\alpha\zeta_3^2, \quad \frac{\alpha\zeta_3^2 - \alpha}{\zeta_3 - \zeta_3^2} = -\alpha(1 + \zeta_3^2).$$

So no rational $c \neq 0$ is bad; e.g., $\mathbb{Q}(\alpha, \zeta_3) = \mathbb{Q}(\alpha + \zeta_3)$.

Example 6.6.18. We saw $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$ in Example 6.2.8 using some ad hoc tricks. We can now argue as follows: note $\pm\sqrt{3}$ are the roots of $m_{\sqrt{3}}^{\mathbb{Q}} = X^2 - 3$ and $\pm\sqrt{2}$ are the roots of $m_{\sqrt{2}}^{\mathbb{Q}} = X^2 - 2$; then 1 is not bad in the sense of Steinitz' theorem because $1 \notin \{(\pm\sqrt{2} - \sqrt{2})/(\sqrt{3} + \sqrt{3})\}$.

Example 6.6.19. For $\alpha := \sqrt[4]{2}$ we saw $\mathbb{Q}(\alpha, i) = \mathbb{Q}(\alpha + i)$ in Example 6.3.16 using some ad hoc tricks. We can now argue as follows: note $\pm\alpha, \pm i\alpha$ are the roots of $m_\alpha^{\mathbb{Q}} = X^4 - 2$ and $\pm i$ are the roots of $m_i^{\mathbb{Q}} = X^2 + 1$; then 1 is not bad in the sense of Steinitz' theorem because $1 \notin \{(-\alpha - \alpha/(2i), \pm i\alpha - \alpha)/(2i)\}$.

The assumption of separability cannot be omitted:

Example 6.6.20. Let p be prime, $K := \mathbb{F}_p(X, Y)$ and $L := K(\sqrt[p]{X}, \sqrt[p]{Y}) \mid K$. Then $L \mid K$ is finite and not simple.

Proof. To see $L \mid K$ is finite, note L is the splitting field of $f := (Z^p - X)(Z^p - Y)$. This is because $f = (Z - \sqrt[p]{X})^p \cdot (Z - \sqrt[p]{X})^p$ by Frobenius.

We claim $a^p \in K$ for all $a \in L$. We have $L = K(\sqrt[p]{X})(\sqrt[p]{Y}) = K(\sqrt[p]{X})[\sqrt[p]{Y}] = K[\sqrt[p]{X}, \sqrt[p]{Y}]$ by Lemma 3.5.8, so $a = g(\sqrt[p]{X}, \sqrt[p]{X})$ for some $g \in K[Z_0, Z_1]$. With the Frobenius endomorphism φ of K we have $a^p = \varphi(g)(X, Y) \in K$.

We show $L \neq K(a)$ for every $a \in L$. Note $m_a^K(Z) \mid Z^p - a^p \in K[Z]$ by the claim, so $[K(a):K] = \deg(m_a^K) \leqslant p$. But $[L:K] = [L:K(\sqrt[p]{X})] \cdot [K(\sqrt[p]{X}):K] > p$ because $[L:K(\sqrt[p]{X})] > 1$ as $\sqrt[p]{Y} \notin K(\sqrt[p]{X})$ and $[K(\sqrt[p]{X}):K] = \deg(m_{\sqrt[p]{X}}^K) = \deg(Z^p - X) = p$ since $Z^p - X$ is irreducible in $(\mathbb{F}_p[X, Y])[Z]$ and $K(Z)$ by Eisenstein (with prime X). □

6.7 Galois extensions

Definition 6.7.1. A field extension $L \mid K$ is *Galois* if it is separable and normal. The *Galois group* $G(L \mid K)$ *of* $L \mid K$ is the set of K-automorphisms of L.

Remark 6.7.2.

1. If K is the prime field of G, then $G(L \mid K) = \mathrm{Aut}(L)$ (Exercise 3.4.15).
2. If $K = L$, then $G(L \mid K) = \{\mathrm{id}_L\}$.
3. If $L \mid M \mid M' \mid K$ are field extensions, then $G(L \mid M)$ is a subgroup of $G(L \mid M')$.
4. If $L \mid K$ is finite, then $|G(L \mid K)| \leqslant [L : K]$ (Corollary 6.3.13).

Example 6.7.3. $\mathbb{C} \mid \mathbb{R}$ is Galois, $G(\mathbb{C} \mid \mathbb{R})$ contains the identity and conjugation, so is isomorphic to C_2. Similarly for quadratic number fields (Definition 4.1.1).

Lemma 6.7.4. *Let $L \mid K$ be Galois and M an intermediate field. Then $L \mid M$ is Galois, and $M \mid K$ is Galois if and only if $\varphi(M) = M$ for all $\varphi \in G(L \mid K)$.*

Proof. $L \mid M$ is separable by Corollary 6.6.12 and normal by Remark 6.3.21 (1), hence Galois. $M \mid K$ is separable by Corollary 6.6.12. We have to show $M \mid K$ is normal if and only if $\varphi(M) = M$ for all $\varphi \in G(L \mid K)$. $\Rightarrow$ follows from Lemma 6.3.22. $\Leftarrow$: by the embedding theorem we can assume that L is a subfield of $\overline{K}$. By Lemma 6.4.8 we have to show $\psi(M) = M$ for all K-automorphisms ψ of $\overline{K}$. But since $L \mid K$ is normal, this lemma implies $\psi(L) = L$, so $\psi{\restriction}L \in G(L \mid K)$. Hence $\psi(M) = (\psi{\restriction}L)(M) = M$ by assumption. □

Recall Example 5.12.4 (4) and Definition 3.7.4: for $G \subseteq \mathrm{Aut}(L)$, evaluation $(\varphi, x) \mapsto \varphi(x)$ defines an action of G on L and the fixed field is

$$L^G := \{x \in L \mid \varphi(x) = x \text{ for all } \varphi \in G\}.$$

Lemma 6.7.5. *Let L be a field, G a subgroup of $\mathrm{Aut}(L)$. Then $a \in L$ is algebraic over L^G if and only if the orbit $G(a) = \{\varphi(a) \mid \varphi \in G\}$ is finite; in case,*

$$m_a^{L^G} = \textstyle\prod_{b \in G(a)} (X - b)$$

Proof. $\Rightarrow$: if a is a root of $f \in L^G[X]$, then every $\varphi \in G$ maps a to a root of f in L (Remark 6.3.12 (2)); there are $\leqslant \deg(f)$ many.

$\Leftarrow$: let $b_1, \ldots, b_n$ list $G(a)$. By Vieta's formula, the coefficients of $f := \prod_{b \in G(a)} (X - b)$ are $\pm s_{n,k}(b_1, \ldots, b_n)$ for $1 \leqslant k \leqslant n$. These are in L^G: for $\varphi \in G$ we have $\varphi(s_{n,k}(b_1, \ldots, b_n)) = s_{n,k}(\varphi(b_1), \ldots, \varphi(b_n)) = s_{n,k}(b_1, \ldots, b_n)$ because φ permutes $G(a)$ and $s_{n,k}$ is symmetric.

To show $m_a^{L^G} = f$ we verify Lemma 3.5.6 (3): let $g \in L^G[X]$ have a as a root. Let $b \in G(a)$ and choose $\varphi \in G$ with $\varphi(a) = b$. Then b is a root of $\varphi(g) = g$, so $(X - b) \mid g$. Since this holds for all $b \in G(a)$ we get $f \mid g$. □

Theorem 6.7.6. *An algebraic field extension $L \mid K$ is Galois if and only if $L^{G(L|K)} = K$.*

Proof. Write $G := G(L \mid K)$. By the embedding theorem we can assume $L \subseteq \overline{K}$.

$\Rightarrow$: assume there exists $a \in L^G \setminus K$. Then m_a^K has degree $\geqslant 2$; as it is separable, there exists a root $b \in \overline{K}$ with $b \neq a$; by homogeneity (Lemma 6.4.7) there is a K-automorphism φ of $\overline{K}$ with $\varphi(a) = b$. But $L \mid K$ is normal, so Lemma 6.4.8 implies $\varphi(L) = L$. Then $\varphi{\restriction}L \in G$ and $(\varphi{\restriction}L)(a) = b \neq a$, contradicting $a \in L^G$.

$\Leftarrow$: as every $a \in L$ is algebraic over $K = L^G$, the lemma shows m_a^K is separable. For normality, let $f \in K[X]$ be irreducible with a root $a \in L$; we have to show f splits in L. We can assume f is monic. By the lemma, $f = m_a^K$ splits in L. □

Theorem 6.7.7. *Let L be a field and G a finite subgroup of $\mathrm{Aut}(L)$. Then $L \mid L^G$ is a finite Galois extension of degree $[L : L^G] = |G|$ with Galois group G.*

Proof. Let $G' := G(L \mid L^G)$; then $G \subseteq G'$, so $L^{G'} \subseteq L^G$. The converse is trivial, so $L^{G'} = L^G$. Then $L \mid L^G$ is Galois by the previous theorem.

For $a \in L$ we have $\deg(m_a^{L^G}) = |G(a)|$ by the lemma. But $|G| = [G : G_a]|G_a| = |G(a)||G_a|$ by the orbit-stabilizer lemma, so $\deg(m_a^{L^G}) \mid |G|$. By Corollary 6.6.16, $L \mid L^G$ is finite.

By the primitive element theorem, $L = L^G(a)$ for some $a \in L$. Every $\varphi \in G$ that fixes a equals id_L (Remark 6.3.12 (4)), so $G_a = \{\mathrm{id}_L\}$. Then

$$|G| = |G(a)| = \deg(m_a^{L^G}) = [L : L^G].$$

$G = G'$ follows from $G \subseteq G'$ and $[L : L^G] = |G| \leqslant |G'| \leqslant [L : L^G]$ (Remark 6.7.2 (4)). □

Below, note (3) implies that $L \mid K$ is finite.

Theorem 6.7.8 (Artin's characterization)**.** *Let $L \mid K$ be a finite field extension and $G := G(L \mid K)$. The following are equivalent.*

1. *$L \mid K$ is Galois.*
2. *$|G| = [L : K]$.*
3. *L is the splitting field of some separable $f \in K[X] \setminus K$.*

Proof. Note $L \mid K$ is algebraic by Corollary 6.2.16. Further, $G \leqslant [L : K]$ is finite (Corollary 6.3.13), so $[L : L^G] = |G|$ by Theorem 6.7.7. Then $[L : K] = |G| \cdot [L^G : K]$ by the degree formula. Thus, $[L : K] = |G|$ if and only if $[L^G : K] = 1$, if and only if $L^G = K$ (Remark 6.2.3 (2)). Hence, $2 \Leftrightarrow 1$ by Theorem 6.7.6.

$1 \Rightarrow 3$: as $L \mid K$ is finite and normal, by Exercise 6.3.20, L is the splitting field of some $f \in K[X] \setminus K$ which is a product of certain minimal polynomials over K; these are separable since $L \mid K$ is separable; hence f is separable (by definition).

$3 \Rightarrow 1$: $L \mid K$ is normal by Lemma 6.3.19. Write $L = K(a_1, \ldots, a_n)$ for $a_i \in L$ roots of f. Each a_i is a root of an irreducible factor of f, so separable over K and hence over $K(a_1, \ldots, a_{i-1})$ (Remark 6.6.8). Thus, $L \mid K$ is separable by Corollary 6.6.11. □

Definition 6.7.9. If $f \in K[X] \setminus K$ and L is the splitting field of f over K, then

$$G(f, K) := G(L \mid K).$$

Remark 6.7.10 (Galois' idea)**.** Let $f \in K[X] \setminus K$ and A the set of roots of f in L, the splitting field of f over K; let $a_1, ..., a_n$ list A.

1. $(\varphi, x) \mapsto \varphi(x)$ is a faithful action of $G(f, K)$ on A (Remark 6.3.12 (2), (4)).
2. Restriction $\varphi \mapsto \varphi{\restriction}A$ is a group monomorphism from $G(f, K)$ into $\mathrm{Sym}(A) \cong S_n$ (Remark 5.12.3 (2)). Hence, $|G(f, K)| \mid |S_n| = n!$ (by Lagrange).
3. More concretely, for $\varphi \in G(f, K)$ define $\varphi^* \in S_n$ setting

$$\varphi^*(i) = j \Leftrightarrow \varphi(a_i) = a_j$$

   for all $i, j \in \{1, ..., n\}$. Then $\varphi \mapsto \varphi^*$ is a group monomorphism from $G(f, k)$ into S_n.
4. If f is irreducible, then the action is transitive (Theorem 6.3.14) and $n \mid |G(f, K)|$.

   Indeed: write $G := G(f, K)$ and let $a \in A$; then $|G| = [G : G_a]|G_a| = |G(a)||G_a|$ by the orbit-stabilizer lemma; by transitivity, $G(a) = A$ has size n.

Galois' revolutionary insight is that the finite group $G(f, K)^* \subseteq S_n$ contains important information about the polynomial equation $f = 0$.

Example 6.7.11. The splitting field of $f := X^3 - 2$ over $\mathbb{Q}$ is $\mathbb{Q}(\alpha, \zeta_3)$ for $\alpha := \sqrt[3]{2}$ (Example 6.3.6). By Artin's characterization, $\mathbb{Q}(\alpha, \zeta_3) \mid \mathbb{Q}$ is Galois and the Galois group $G := G(f, \mathbb{Q})$ has order $[\mathbb{Q}(\alpha, \zeta_3) : \mathbb{Q}] = 6$ (Examples 6.2.18 (1)). By Remark 6.7.10 (2), G is isomorphic to a subgroup of S_3. As $|S_3| = 3! = 6$ we see $G \cong S_3$.

Alternatively, use Corollary 6.7.17 below: $G \cong S_3$ because $D_f = -27 \cdot 2^2$ is not a square in $\mathbb{Q}$. Thus, G is non-abelian and solvable (Remark 5.8.11 (3)).

Example 6.7.12. The splitting field of $f := X^4 - 5X^2 + 6 = (X^2 - 2)(X^2 - 3) \in \mathbb{Q}[X]$ is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. By Artin's characterization, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) \mid \mathbb{Q}$ is Galois with Galois group $G := G(f, \mathbb{Q})$ of order $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$ (Example 6.2.8). Hence, $G \cong C_4$ or $G \cong K_4$ (Proposition 5.3.19). G containes the identity and $\varphi_1, \varphi_2, \varphi_3$ determined by

$$\sqrt{2}, \sqrt{3} \mapsto -\sqrt{2}, \sqrt{3}, \quad \sqrt{2}, \sqrt{3} \mapsto \sqrt{2}, -\sqrt{3}, \quad \sqrt{2}, \sqrt{3} \mapsto -\sqrt{2}, -\sqrt{3}.$$

To see e.g. φ_3 exists, Remark 6.3.12 gives an ($\mathbb{Q}$-)automorphism ψ of $\mathbb{Q}(\sqrt{2})$ with $\psi(\sqrt{2}) = -\sqrt{2}$; then Theorem 6.3.10 gives φ_3 as an extension to $\mathbb{Q}(\sqrt{2})(\sqrt{3})$.

Each of them has order 2 in G because φ_i^2 fixes all roots $\pm\sqrt{2}, \pm\sqrt{3}$, so $\varphi^2 = \mathrm{id}_{\mathbb{Q}(\sqrt{2},\sqrt{3})}$ (Remark 6.7.10 (1)). Hence, $G \cong K_4$ is abelian and not cyclic.

Exercise 6.7.13. Let $f := X^4 - 5 \in \mathbb{Q}[X]$. Describe the elements of $G(f, \mathbb{Q})$. Show $G(f, \mathbb{Q}(i)) \cong C_4$ and $G(f, \mathbb{Q}(\sqrt{5})) \cong K_4$.

Exercise 6.7.14. Let $f := X^4 - 4X^2 + 2$ from Exercise 6.3.8. Its roots are $\pm\alpha, \pm\beta$ for $\alpha = \sqrt{2 - \sqrt{2}}, \beta = \sqrt{2 + \sqrt{2}}$, and its splitting field over $\mathbb{Q}$ is $\mathbb{Q}(\alpha)$. Show there is a unique $\sigma \in G(f, \mathbb{Q})$ with $\sigma(\alpha) = \beta$ and compute $\sigma(\sqrt{2})$ and $\sigma(\beta)$. Infer that $G(f, \mathbb{Q})$ is cyclic.

Exercise 6.7.15. In Exercise 6.3.18 show $G(K \mid \mathbb{Q}) \cong \mathbb{Z}_2^n$.

We continue with some more abstract examples, and determine Galois groups in some familiar settings. Recall the discriminant D_f of a monic polynomial f from Example 3.7.11.

Proposition 6.7.16. *Let K be a field and $f \in K[X]$ be monic, separable of degree $n > 1$. Then $G(f,K)^* \subseteq A_n$ if and only if $\sqrt{D_f} \in K$.*

Proof. Let $a_1, ..., a_n$ list the roots of f in L, the splitting field of f over K. Then

$$\sqrt{D_f} = \prod_{1 \leqslant i < j \leqslant n} (a_i - a_j).$$

This implies $\varphi(\sqrt{D_f}) = \operatorname{sign}(\varphi^*)\sqrt{D_f}$ for every $\varphi \in G(L \mid K)$. Hence, $G(L \mid K)^* \subseteq A_n$ if and only if every $\varphi \in G(L \mid K)$ fixes $\sqrt{D_f}$. This is equivalent to $\sqrt{D_f} \in K$ by Theorem 6.7.6 (note $L \mid K$ is Galois by Artin). □

Corollary 6.7.17 (Galois theory of the cubic)**.** *Let K be a field and $f \in K[X]$ be monic, cubic, separable and irreducible. If $\sqrt{D_f} \in K$, then $G(f,K) \cong A_3$; otherwise, $G(f,K) \cong S_3$.*

Proof. By Remark 6.7.10 (4), the order of $G := G(f,K)^* \subseteq S_3$ is divisible by 3. Since $|G| \mid |S_3| = 6$, we have $|G| \in \{3, 6\}$. Then, $G = A_3 = \{1, (123), (321)\}$ or $G = S_3$. Apply the proposition. □

Example 6.7.18. Consider $f^{\pm} := X^3 \pm 3X + 1 \in \mathbb{Q}[X]$. Using the formula in Example 3.7.11 compute $D_{f^+} = -63$ and $D_{f^-} = 9$. Hence, $G(f^+, \mathbb{Q}) \cong S_3$ and $G(f^-, \mathbb{Q}) \cong A_3$.

Definition 6.7.19. A finite Galois extension $L \mid K$ is *cyclic, abelian* or *solvable* if so is its Galois group $G(L \mid K)$.

Proposition 6.7.20. *Let $p, n \in \mathbb{N}$, $n > 0$ and p prime. Then $\mathbb{F}_{p^n} \mid \mathbb{F}_p$ is a finite cyclic Galois extension; $G(\mathbb{F}_{p^n} \mid \mathbb{F}_p)$ is generated by the Frobenius endomorphism.*

Proof. $\mathbb{F}_{p^n} \mid \mathbb{F}_p$ is obviously finite, and Galois by Artin because $\mathbb{F}_{p^n}$ is the splitting field of $X^{p^n} - X$ over $\mathbb{F}_p$ (Theorem 6.5.2) and $\mathbb{F}_p$ is perfect (Example 3.4.18). Then $G := G(\mathbb{F}_{p^n} \mid \mathbb{F}_p)$ has order $[\mathbb{F}_{p^n} : \mathbb{F}_p] = n$ by Artin. As $\mathbb{F}_{p^n}$ is perfect, the Frobenius endomorphism φ is an automorphism of $\mathbb{F}_{p^n}$. It fixes the prime field $\mathbb{F}_p$, so $\varphi \in G$.

Let x generate $\mathbb{F}_{p^n}^\times$ (Corollary 5.3.25), i.e., x has order $p^n - 1$ in $\mathbb{F}_{p^n}^\times$. Note $\varphi^k(x) = x^{p^k}$, so the minimal k for which $\varphi^k(x) = x$ is n. As $\mathbb{F}_p(x) = \mathbb{F}_{p^n}$, this implies $\varphi^n = \mathrm{id}_{\mathbb{F}_{p^n}}$. Hence, φ has order n in G, so generates G. □

Exercise 6.7.21. Generalize to the case where p is a prime power.

Recall Exercise 3.7.8: $K[s_{n,1}, ...]$ is the subring of symmetric polynomials in $K[\bar{X}]$.

Proposition 6.7.22. *Let K be a field, $n > 0$ and $M := K(s_{n,1}, ..., s_{n,n})$. Then*

$$G(f,M) \cong S_n$$

where $f := (X - X_1) \cdots (X - X_n) \in M[X]$.

Proof. $f \in M[X]$ by Vieta's formula, and $L := M(\bar{X}) = K(\bar{X})$ is the splitting field of f over M. As f is separable, $L \mid M$ is Galois by Artin. By Remark 6.7.10, $G(f,K)^* \subseteq S_n$. It thus suffices to show $|G(f,K)| \geqslant n!$.

For $\sigma \in S_n$ let $\hat{\sigma} : K[\bar{X}] \to K[\bar{X}]$ be the ring automorphism $f \mapsto f^\sigma$ (permute the variables X_i by σ). Note $\hat{\sigma}$ fixes the subring $K[s_{n,1}, ..., s_{n,n}]$. It uniquely extends to an automorphism of the quotient field $K(\bar{X}) = L$ via $f/g \mapsto \hat{\sigma}(f)/\hat{\sigma}(g)$ – again denoted $\hat{\sigma}$. Then $\hat{\sigma}$ fixes M, so $\hat{\sigma} \in G(L \mid M) = G(f, M)$. It is clear that $\sigma \mapsto \hat{\sigma}$ is injective. □

6.8 Galois theory

The application of group theoretic terminology to field extensions might appear strange. The theorem below is maybe the most beautiful result of this course. Roughly speaking, it equates the structure of a Galois extension with the structure of its Galois group.

Theorem 6.8.1 (Main theorem of Galois theory)**.** *Let $L \mid K$ be a finite Galois extension with Galois group $G := G(L \mid K)$.*

1. *$U \mapsto L^U$ is a bijection from the set $\mathcal{U}$ of subgroups of G onto the set $\mathcal{M}$ of intermediate fields of $L \mid K$; its inverse is $M \mapsto G(L \mid M)$; both maps reverse inclusions.*
2. *For all $M \in \mathcal{M}$: $L \mid M$ is a Galois extension.*
3. *For all $U \in \mathcal{U}$: $L \mid L^U$ is a Galois extension with Galois group U.*
4. *For all $M \in \mathcal{M}$ and all $\varphi \in G$: $G(L \mid \varphi(M)) = \varphi G(L \mid M)\varphi^{-1}$.*
5. *For all $M \in \mathcal{M}$: $[M : K] = [G : G(L \mid M)]$ and, $M \mid K$ is a Galois extension if and only if $G(L \mid M) \lhd G$; in case, $G(M \mid K) \cong G/G(L \mid M)$.*

$$\begin{array}{ccccc} \{1\} & \xrightarrow{n/k} & U & \xrightarrow{k} & G \\ L & \xleftarrow{n/k} & M & \xleftarrow{k} & K \end{array}$$

Figure 6.1: corresponding U, M with $[M : K] = [G : U] = k$, $[L : K] = |G| = n$

Proof. Note G is finite because $|G| = [L : K]$ (by Artin). (2) follows from Lemma 6.7.4 and (3) from Theorem 6.7.7.

(1): let Φ denote $U \mapsto L^U$ and Ψ denote $M \mapsto G(L \mid M)$. Then $\Psi(\Phi(U)) = G(L \mid L^U) = U$ by Theorem 6.7.7. Further, $\Phi(\Psi(M)) = L^{G(L|M)}$. By (2), $L \mid M$ is Galois, so $L^{G(L|M)} = M$ by Theorem 6.7.6. Thus, Φ is bijective and $\Phi^{-1} = \Psi$.

Clearly, if $U, U' \in \mathcal{U}$, $M, M' \in \mathcal{M}$ with $U \subseteq U', M \subseteq M'$, then $\Phi(U) = L^U \supseteq L^{U'} = \Phi(U')$ and $\Psi(M) = G(L \mid M) \supseteq G(L \mid M') = \Psi(M')$.

(4) $\supseteq$: let $\psi \in G(L \mid M)$ and $x \in \varphi(M)$, say $x = \varphi(y)$ with $y \in M$. Note $\psi(y) = y$. Then (omitting $\circ$): $\varphi\psi\varphi^{-1}(x) = \varphi\psi(y) = \varphi(y) = x$. Hence, $\varphi\psi\varphi^{-1} \in G(L \mid \varphi(M))$.

(4) $\subseteq$: using the inclusion already proved,

$$\varphi G(L \mid M)\varphi^{-1} = \varphi G(L \mid \varphi^{-1}\varphi(M))\varphi^{-1} \supseteq \varphi\varphi^{-1}G(L \mid \varphi(M))\varphi\varphi^{-1} = G(L \mid \varphi(M)).$$

(5): by the degree formula and Artin

$$[M : K] = [L : K]/[L : M] = |G|/|G(L \mid M)| = [G : G(L \mid M)].$$

Further, we have the equivalences:

$M \mid K$ is Galois		
$\iff$	$M = \varphi(M)$ for all $\varphi \in G$	by Lemma 6.7.4
$\iff$	$G(L \mid M) = G(L \mid \varphi(M))$ for all $\varphi \in G$	since Ψ is injective by (1)
$\iff$	$G(L \mid M) = \varphi G(L \mid M)\varphi^{-1}$ for all $\varphi \in G$	by (4)
$\iff$	$G(L \mid M) \lhd G$.	

Assume $M \mid K$ is Galois. For $\varphi \in G$ we have $\varphi(M) = M$ by Lemma 6.7.4, so $\varphi \mapsto \varphi{\restriction}M$ is a group homomorphism from G to $G(M \mid K)$. Its kernel is $G(L \mid M)$. By Noether's 1st isomorphism theorem we are left to show that the map is surjective. Let $\psi \in G(M \mid K)$. The embedding theorem allows to assume $L \subseteq \overline{K}$ and gives an extension $\varphi : L \to \overline{K}$. Since $L \mid K$ is normal, Lemma 6.4.8 yields $\varphi(L) = L$. Then $\varphi \in G$ and $\varphi{\restriction}M = \psi$. □

A first application:

More abstract proof of the fundamental theorem 3.8.3 after Artin. By Remark 3.8.2 (2) it suffices to show every $f \in \mathbb{R}[X] \smallsetminus \mathbb{R}$ has a root in $\mathbb{C}$. Let L be the splitting field of $(X^2+1)f$ over $\mathbb{R}$. Then $L \mid \mathbb{C} \mid \mathbb{R}$ and we claim $L = \mathbb{C}$. By Artin $L \mid \mathbb{R}$ is Galois and $|G| = [L : \mathbb{R}]$ for $G := G(L \mid \mathbb{R})$. As $[\mathbb{C} : \mathbb{R}] = 2$ we have $2 \mid [L : \mathbb{R}]$ be the degree formula. By the 1st Sylow theorem, G has a 2-Sylow subgroup U. Set $M := L^U$. Then $[M : \mathbb{R}] = [G : U]$ is odd. By the primitive element theorem, $M = \mathbb{R}(a)$ for some $a \in \mathbb{C}$. Then $\deg(m_a^{\mathbb{R}}) = [M : \mathbb{R}]$ is odd, so $m_a^{\mathbb{R}}$ has a root in $\mathbb{R}$ (Remark 3.8.2 (3)). Thus, $\deg(m_a^{\mathbb{R}}) = 1$ and $a \in \mathbb{R}$, so $M = \mathbb{R}$. Then $G = U$, so $|G| = 2^k$ for some $k > 0$. Then $[L \mid \mathbb{C}] = 2^{k-1}$ by the degree formula, so $G' := G(L \mid \mathbb{C})$ has order 2^{k-1}. We claim $k = 1$: then $[L : \mathbb{C}] = 1$, so $L = \mathbb{C}$ and we are done.

Otherwise, by the 1st Sylow theorem, G' has a subgroup U' of order 2^{k-2}. As $L \mid \mathbb{C}$ is Galois with Galois group G', $[L^{U'} : \mathbb{C}] = [G' : U'] = 2$ – impossible by Remark 3.8.2 (4). □

Which intermediate fields M does $L \mid K$ have?

Corollary 6.8.2. *Let $L \mid K$ be a finite Galois extension.*

1. *If $L \mid K$ is cyclic, then there is for every $d \mid [L : K]$ exactly one intermediate field M with $[M : K] = d$; then $M \mid K$ is Galois and cyclic.*
2. *If $L \mid K$ is abelian, then there is for every $d \mid [L : K]$ at least one intermediate field M with $[M : K] = d$ and $M \mid K$ is Galois and abelian.*
3. *If $L \mid K$ is solvable and $L \neq K$, then there are $\ell > 0$ and intermediate fields $K = M_0 \subsetneq M_1 \subsetneq M_2 \cdots \subsetneq M_\ell = L$ such that for all $i < \ell$, $M_{i+1} \mid M_i$ is Galois of prime degree.*

   For every intermediate field M of $L \mid K$: the field extension $L \mid M$ is Galois and solvable and, if $M \mid K$ is Galois, then $M \mid K$ is solvable.

Proof. Write $G := G(L \mid K)$. (1): the intermediate fields M with $[M : K] = d$ correspond to subgroups U of G of index d. there is exactly one by Theorem 5.3.21. $M \mid K$ is Galois because all subgroups are normal since G is abelian. (2) is similar using Corollary 5.11.5.

(3): by Theorem 5.8.17 there is a subnormal series $N_\ell = \{\mathrm{id}_L\} \lhd N_{\ell-1} \lhd \cdots \lhd N_0 = G$ with (abelian) factors N_i/N_{i+1} of prime order. For $M_i := L^{N_i}$ we have $M_0 = L^G = K$ by Theorem 6.7.6, and $M_0 \subseteq \cdots \subseteq M_\ell = L$. Note $N_i = G(L \mid M_i)$.

Since $L \mid M_i$ is finite Galois we can use (5) of the main theorem with M_i in the role of K: $M_{i+1} \mid M_i$ is Galois because $G(L \mid M_{i+1}) = N_{i+1} \lhd N_i = G(L \mid M_i)$; further, $[M_{i+1} : M_i] = [G(L \mid M_i) : G(L \mid M_{i+1})] = [N_i : N_{i+1}]$ is prime.

For M an intermediate field, note $L \mid M$ is Galois and $G(L \mid M)$ is solvable as a subgroup of $G(L \mid K)$ (Remark 5.8.11 (3)). If $M \mid K$ is Galois, then $G(L \mid M) \lhd G$ and $G(M \mid K) \cong G/G(L \mid M)$ is solvable by Lemma 5.8.13. □

Exercise 6.8.3. Finite separable field extensions have finitely many intermediate fields.

Corollary 6.8.4. *Let $z \in \mathbb{C}$ be algebraic and $L \subseteq \mathbb{C}$ the splitting field of $m_z^{\mathbb{Q}}$ over $\mathbb{Q}$. If $[L : \mathbb{Q}]$ is a power of 2, then z is constructible.*

Proof. Say, $[L : \mathbb{Q}] = 2^n$. By Artin, $L \mid \mathbb{Q}$ is Galois and $|G(L \mid \mathbb{Q})| = 2^n$. By Corollary 5.13.13, $G(L \mid \mathbb{Q})$ is solvable. Choose intermediate fields M_i as in (3) above. As $[M_{i+1} : M_i] \mid 2^n$ is prime, $[M_{i+1} : M_i] = 2$. By Proposition 6.2.9, M_{i+1} results from M_i by adjunction of a square root. By Theorem 6.1.5, z is constructible. □

Example 6.8.5. We continue Example 6.7.12 of the Galois extension $L := \mathbb{Q}(\sqrt{2}, \sqrt{3}) \mid \mathbb{Q}$ with Galois group $G := \{\mathrm{id}_L, \varphi_1, \varphi_2, \varphi_2\}$. The proper nontrivial subgroups are $\langle \varphi_i \rangle$ of index 2. Thus their fixed fields have degree 2. E.g., φ_1 fixes $\sqrt{3}$, so $\mathbb{Q}(\sqrt{3}) \subseteq L^{\langle \varphi_1 \rangle}$; since both field extensions of $\mathbb{Q}$ have degree 2 we have $\mathbb{Q}(\sqrt{3}) = L^{\langle \varphi_1 \rangle}$. Analogously, $\mathbb{Q}(\sqrt{2}) = L^{\langle \varphi_2 \rangle}$. Finally note $\varphi_3(\sqrt{6}) = \varphi_3(\sqrt{2})\varphi(\sqrt{3}) = (-\sqrt{2}) \cdot (-\sqrt{3}) = \sqrt{6}$, so $\mathbb{Q}(\sqrt{6}) = L^{\langle \varphi_3 \rangle}$.

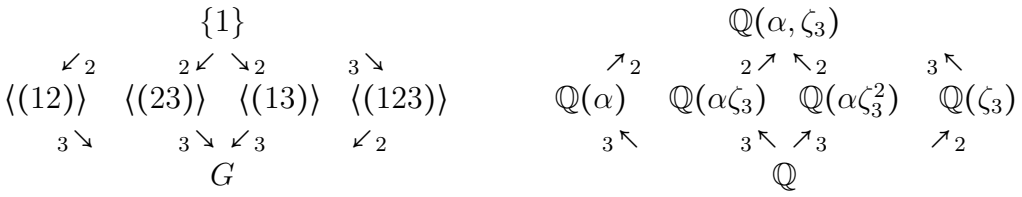
We display non-trivial subgroups and proper subfields fields with arrows indicating subgroups of index 2, resp., field extensions of degree 2.

$$\begin{array}{ccccc} \langle \varphi_1 \rangle & \langle \varphi_2 \rangle & \langle \varphi_3 \rangle & \qquad & \mathbb{Q}(\sqrt{2}) \quad \mathbb{Q}(\sqrt{3}) \quad \mathbb{Q}(\sqrt{6}) \\ \searrow & \downarrow & \swarrow & & \nwarrow \quad \uparrow \quad \nearrow \\ & G & & & \mathbb{Q} \end{array}$$

Example 6.8.6. We continue Example 6.7.11 of the Galois extension $L := \mathbb{Q}(\alpha, \zeta_3) \mid \mathbb{Q}$ (where $\alpha = \sqrt[3]{2}$) with Galois group $G := G(X^3 - 2, \mathbb{Q}) \cong S_3$. Number the roots $a_1 := \alpha, a_2 := \alpha\zeta_3, a_2 := \alpha\zeta_3^2$. Let us abuse our notation for S_3 to denote elements of G, e.g., (12) denotes the automorphism that swaps a_1, a_2 and fixes a_3. Then the nontrivial proper subgroups are $A_3 \cong \langle (123) \rangle$ of index 2, and $\langle (12) \rangle, \langle (13) \rangle, \langle (23) \rangle$ of index 3.

As (12) fixes a_3, we have $\mathbb{Q}(a_3) \subseteq L^{\langle (12) \rangle}$; since both fields are degree 3 extensions of $\mathbb{Q}$ we have $\mathbb{Q}(a_3) = L^{\langle (12) \rangle}$. Similarly, $\mathbb{Q}(a_2) = L^{\langle (13) \rangle}$ and $\mathbb{Q}(a_1) = L^{\langle (23) \rangle}$. For $L^{\langle (123) \rangle}$ we know it is the unique degree 2 extension, and we saw $[\mathbb{Q}(\zeta_3) : \mathbb{Q}] = 2$; hence $\mathbb{Q}(\zeta_3) = L^{\langle (123) \rangle}$.

We display subgroups and intermediate fields with arrows indexed by the index of the subgroup, resp., the degree of the field extension.



Exercise 6.8.7. Continuing Exercises 6.3.8, 6.7.14, show $\mathbb{Q}(\sqrt{2-\sqrt{2}}) \mid \mathbb{Q}$ has exactly one proper intermediate field, namely $\mathbb{Q}(\sqrt{2})$.

A more complicated example:

Example 6.8.8. One can show by elementary means that $f := X^4 - X^2 - 1 \in \mathbb{Q}[X]$ is irreducible with roots $\pm\sqrt{\phi}, \pm i/\sqrt{\phi} \in \mathbb{C}$ where $\phi := (1+\sqrt{5})/2$ is the golden ratio. The roots of f are the vertices of a square inscribed in a circle of radius $\sqrt{\phi}$ and we now show the automorphisms of its splitting field act on them like the symmetry group of this square.

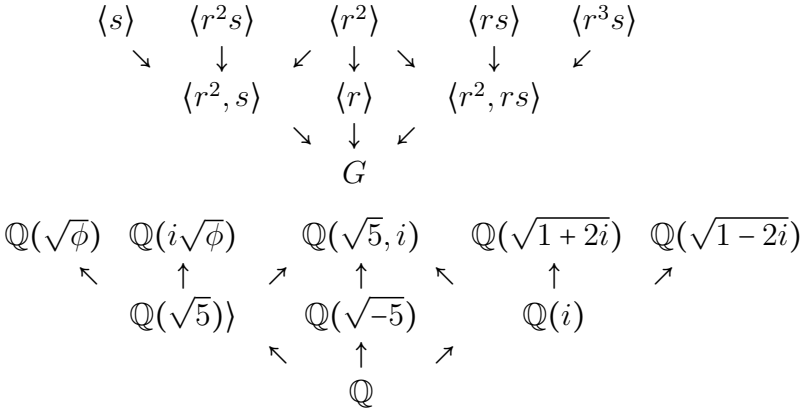
The splitting field of f over $\mathbb{Q}$ is $L := \mathbb{Q}(\sqrt{\phi}, i)$, so $L \mid \mathbb{Q}$ is Galois with Galois group $G := G(f, \mathbb{Q})$. Then $|G| = [L : \mathbb{Q}] = 8$ because and $[L : \mathbb{Q}(\sqrt{\phi})] = 2$ (as $i \notin \mathbb{Q}(\sqrt{\phi}) \subseteq \mathbb{R}$) and $[\mathbb{Q}(\sqrt{\phi}) : \mathbb{Q}] = \deg(f) = 4$. Since $i/\sqrt{\phi} \notin \mathbb{Q}(\sqrt{\phi}) \subseteq \mathbb{R}$ we see $\mathbb{Q}(\sqrt{\phi}) \mid \mathbb{Q}$ is not normal, hence not Galois, so $G(L \mid \mathbb{Q}(\sqrt{\phi})) \not\trianglelefteq G$. The only group of order 8 with a non-normal subgroup is D_4 (cf. Examples 5.6.19 and 5.5.19). Hence, $G \cong D_4$.

We determine an isomorphism: every $\varphi \in G$ is determined by $\leqslant 4$ choices $\pm\sqrt{\phi}, \pm i/\sqrt{\phi}$ for $\varphi(\sqrt{\phi})$ and $\leqslant 2$ choices $\pm i$ for $\varphi(i)$, so all choices are possible:

$\varphi(\sqrt{\phi})$	$\sqrt{\phi}$	$-\sqrt{\phi}$	$i/\sqrt{\phi}$	$-i/\sqrt{\phi}$	$\sqrt{\phi}$	$-\sqrt{\phi}$	$i/\sqrt{\phi}$	$-i/\sqrt{\phi}$
$\varphi(i)$	i	i	i	i	$-i$	$-i$	$-i$	$-i$
φ	id	r^2	rs	r^3s	s	r^2s	r	r^3

where we introduce names r, s: mapping them to $R_{2\pi/4}, S_0$ determines an isomorphism onto D_4 (cf. Theorem 5.1.17). Note the isomorphism only equates the action on the roots, not on all of L, e.g., $r(1) = 1$ is not a rotation by $\pi/2$.

We display the nontrivial subgroups and the corresponding fixed fields. The arrows indicate subgroups of index 2, resp., field extensions of degree 2:



To see e.g. $\mathbb{Q}(i) = L^{\langle r^2, rs\rangle}$ note $\subseteq$ follows because both r^2 and rs fix i; for $\supseteq$ note $[\mathbb{Q}(i) : \mathbb{Q}]$ has the required degree $[L^{\langle r^2,rs\rangle} : \mathbb{Q}] = [G : \langle r^2, rs\rangle] = 2$.

How did we find, e.g., $b = \sqrt{1+2i}$ to see $\mathbb{Q}(b) = L^{\langle rs\rangle}$? We want $b \in L$ fixed by rs but not by r^2. Since rs has order 2, a natural guess is $b := \sqrt{\phi} + (rs)(\sqrt{\phi}) = \sqrt{\phi} + i/\sqrt{\phi}$; indeed, $r^2(b) = -\sqrt{\phi} - i\sqrt{\phi} = -b \neq b$ and a direct computation shows $b^2 = 1 + 2i$.

Exercise 6.8.9. Let K be a field and $f \in K[X]$ irreducible and separable. Show: if $G(f, K)$ is abelian, then it has order $\deg(f)$.

6.9 Cyclotomic fields

Definition 6.9.1. Let K be a field and $n > 0$. An *nth roots of unity (over K)* is a root of $X^n - 1$ in $\overline{K}$; their set is denoted $C_n^K \subseteq \overline{K}$. A *primitive* nth root of unity is one that generates C_n^K as a subgroup of $\overline{K}^\times$.

The *nth cyclotomic field over K* is $K(C_n^K)$, the splitting field of $X^n - 1 \in K[X]$ over K.

Remark 6.9.2.

1. $C_n^{\mathbb{Q}} = C_n = \{1, \zeta_n, \zeta_n^2, ..., \zeta_n^{n-1}\} \subseteq \mathbb{C}$ as of Definition 1.6.8 with $\mathbb{Q}(\zeta_n)$ as the nth cyclotomic field over $\mathbb{Q}$. ζ_n is a primitive nth root of unity over $\mathbb{Q}$.
2. C_n^K is indeed a subgroup of $\overline{K}^\times$, and hence cyclic (Corollary 5.3.25).

   Indeed: if $x, y \in C_n^K$, then $xy^{-1} \in C_n^K$ since $(xy^{-1})^n = x^n(y^n)^{-1} = 1$.
3. If $\zeta \in C_n^K$ is primitive, then the nth cyclotomic field over K is $K(\zeta)$.
4. For $\mathrm{char}(K) = p > 0$, write $n = p^k m$ with $p \nmid m$; then in $K[X]$,

$$X^n - 1 = (X^m - 1)^{p^k},$$

   so the nth roots of unity (over K) are the mth roots of unity.

Theorem 6.9.3. *Let $n > 0$, L be the nth cyclotomic field over K, and $\mathrm{char}(K) \nmid n$. Then C_n^K has order n and $\varphi(n)$ primitive elements (Euler's totient). Moreover, $L \mid K$ is a finite Galois extension and $G(L \mid K)$ is isomorphic to a subgroup of $\mathbb{Z}_n^\times$.*

Proof. $X^n - 1$ has derivative $\underline{n}X^{n-1}$; for $x \in C_n^K$ we have $\underline{n}x^{n-1} \neq 0$, so no root is multiple (Lemma 3.3.13 (1)); hence, $|C_n^K| = n$. If $\zeta \in C_n^K$ is primitive, then $C_n^K = \{1, \zeta, ..., \zeta^{n-1}\}$ has order n. Moreover, ζ^k is primitive if and only if $\gcd(k, n) = 1$ (Lemma 5.3.9 (5)).

Moreover: $L \mid K$ is clearly finite (Remark 6.3.4), and by the above, $X^n - 1$ is separable, so $L \mid K$ is Galois by Artin. $\varphi{\restriction}C_n^K$ permutes C_n^K for $\varphi \in G(L \mid K)$ by Remark 6.3.12 (3). Hence, $\varphi \mapsto \varphi{\restriction}C_n^K$ is a group homomorphism from $G(L \mid K)$ into $\mathrm{Aut}(C_n^K)$. It is injective by Remark 6.3.12 (4). But $\mathrm{Aut}(C_n^K) \cong \mathbb{Z}_n^\times$ by Exercise 2.6.5 – $C_n^K \cong \mathbb{Z}_n$ via $\zeta^k \mapsto \bar{k}$. □

Examples 6.9.4. For prime p we factor $X^p - 1 = (X - 1)(X^{p-1} + \cdots + X + 1)$ (Example 4.5.9 (2)), so the pth cyclotomic field over $\mathbb{Q}$ is a degree $p - 1$ extension.

Recalling Example 1.6.10, we have $\mathbb{Q}(\zeta_3) = \mathbb{Q}(\sqrt{-3}), \mathbb{Q}(\zeta_4) = \mathbb{Q}(i), \mathbb{Q}(\zeta_6) = \mathbb{Q}(\sqrt{-3})$ are extensions of degree $2 = \varphi(3) = \varphi(4) = \varphi(6)$. We have $\zeta_5 = \cos(2\pi/5) + i\sin(2\pi/5)$ and $\cos(2\pi/5) = (\sqrt{5}-1)/4, \sin(2\pi/5) = \sqrt{(\sqrt{5}+5)/8}$, so $\mathbb{Q}(\zeta_5) = \mathbb{Q}(\sqrt{5})(\sqrt{-a})$ where $a = (\sqrt{5}+5)/8$ has degree $4 = \varphi(6)$. We prove below that $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$ for all $n > 0$.

Definition 6.9.5. Let $n > 0$ and K a field with $\mathrm{char}(K) \nmid n$ and let $a_1, ..., a_{\varphi(n)} \in \overline{K}$ list the primitive nth roots of unity. Then the *nth cyclotomic polynomial over* K is

$$\Phi_n^K := \prod_{i=1}^{\varphi(n)} (X - a_i).$$

Lemma 6.9.6. *Let $n > 0$ and K a field with $\mathrm{char}(K) \nmid n$.*

1. $X^n - 1 = \prod_{d|n} \Phi_d^K$.
2. *If $\mathrm{char}(K) = 0$ and K extends $\mathbb{Q}$, then $\Phi_n^K \in \mathbb{Z}[X]$.*
3. *If $\mathrm{char}(K) = p > 0$ and K extends $\mathbb{F}_p$, then $\Phi_n^K \in \mathbb{F}_p[X]$.*

Proof. (1): as $|C_n^K| = n$ we can write $X^n - 1 = \prod_{a \in C_n^K} (X - a)$ in $\overline{K}[X]$; as $C_n^K \cong \mathbb{Z}_n$ (additive) there are exactly $\varphi(d)$ elements of order $d \mid n$ (Corollary 5.3.24); these are the primitive dth roots of unity; thus, (1) just groups the factors $X - a$ according orders.

(2) is proved by induction on n. For $n = 1$, $\Phi_1^K = X - 1 \in \mathbb{Z}[X]$. For $n > 1$, we have $X^n - 1 = \Phi_n^K \cdot f$ with $f := \prod_{d|n, d<n} \Phi_d^K$; by induction, $f \in \mathbb{Z}[X]$; hence, polynomial divison of $X^n - 1$ by f yields Φ_n when done in $\overline{K}[X]$. But $X^n - 1, f \in \mathbb{Z}[X]$ and f is monic, so polynomial division runs in $\mathbb{Z}[X]$ (Theorem 3.2.1). Hence, $\Phi_n^K \in \mathbb{Z}[X]$.

The proof of (3) is analogous. □

Remark 6.9.7. This gives a recursive procedure to compute $\Phi_n^{\mathbb{Q}}$:

$$\begin{aligned}
\Phi_1^{\mathbb{Q}} &= X - 1 \\
\Phi_2^{\mathbb{Q}} &= (X^2 - 1)/(X - 1) = X + 1 \\
\Phi_3^{\mathbb{Q}} &= (X^3 - 1)/(X - 1) = X^2 + X + 1 \\
\Phi_4^{\mathbb{Q}} &= (X^4 - 1)/((X - 1)(X + 1)) = (X^2 + 1)(X^2 - 1)/(X^2 - 1) = X^2 + 1 \\
\Phi_5^{\mathbb{Q}} &= (X^5 - 1)/(X - 1) = X^4 + X^4 + X^3 + X^2 + X + 1 \\
\Phi_6^{\mathbb{Q}} &= (X^6 - 1)/((X - 1)(X + 1)(X^2 + X + 1)) = X^2 - X + 1 \\
\Phi_{30}^{\mathbb{Q}} &= X^8 + X^7 - X^5 - X^4 - X^3 + X + 1.
\end{aligned}$$

The first with a coefficient $\neq 0, 1$ is $\Phi_{105}^{\mathbb{Q}}$ which has degree 48.

Theorem 6.9.8 (Gauß)**.** *Let $n > 0$. Then $\Phi_n^{\mathbb{Q}} = m_{\zeta_n}^{\mathbb{Q}}$. In particular, $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$ and $G(\mathbb{Q}(\zeta_n) \mid \mathbb{Q}) \cong \mathbb{Z}_n^\times$.*

Proof. The 2nd sentence is clear by Theorem 6.9.3. We claim that for every primitive $\zeta \in C_n^{\mathbb{Q}}$ and prime $p \nmid n$ we have $m_\zeta^{\mathbb{Q}} = m_{\zeta^p}^{\mathbb{Q}}$.

Assume $\neq$. Both $f := m_\zeta^{\mathbb{Q}}, g := m_{\zeta^p}^{\mathbb{Q}} \mid X^n - 1$. Since f, g are monic, Lemma 4.4.15 shows $f, g \in \mathbb{Z}[X]$. Since f, g are coprime, $fg \mid X^n - 1$, say $fgh = X^n - 1$. As fg is monic, $h \in \mathbb{Z}[X]$.

Note ζ is a root of $g(X^p)$, so $g(X^p) = fh'$ by Lemma 3.5.6, and again $h' \in \mathbb{Z}[X]$. Let $\bar{f}$ be obtained by replacing coefficients $x \in \mathbb{Z}$ by $\bar{x} = [x]_p$; i.e., this is the extension of the canonical homomorphism from $\mathbb{Z}$ onto $\mathbb{Z}_p = \mathbb{F}_p$ to $\mathbb{Z}[X]$. Then $\bar{f} \cdot \bar{h}' = \overline{g(X^p)} = \bar{g}^p$ by the Frobenius homomorphism. Let $\tilde{f}$ be an irreducible factor of $\bar{f}$ in $\mathbb{F}_p[X]$. Then $\tilde{f} \mid \bar{g}$, so $\tilde{f}^2 \mid \overline{X^n - 1} = X^n - 1$. But then a root of $\tilde{f}$ (in $\overline{\mathbb{F}_p}$) is a multiple root of $X^n - 1$. This contradicts $|C_n^{\mathbb{F}_p}| = n$ by Theorem 6.9.3. This proves the claim.

Let $\zeta \in C_n^{\mathbb{Q}}$ be primitive. Then $\zeta = \zeta_n^k$ with k, n coprime. Write $k = p_1 \cdots p_\ell$ for primes $p_i \nmid n$. Then $m_{\zeta_n}^{\mathbb{Q}} = m_{\zeta_n^{p_1}}^{\mathbb{Q}}$ by the claim. As $\zeta_n^{p_1}$ is also a primitive nth root of unity, the claim implies $m_{\zeta_n^{p_1 p_2}}^{\mathbb{Q}} = m_{\zeta_n^{p_1}}^{\mathbb{Q}} = m_{\zeta_n}^{\mathbb{Q}}$ and so on. Hence, $m_{\zeta}^{\mathbb{Q}} = m_{\zeta_n}^{\mathbb{Q}}$.

This implies that every primitive nth root of unity is a root of $m_{\zeta_n}^{\mathbb{Q}}$ and therefore $\Phi_n^{\mathbb{Q}} \mid m_{\zeta_n}^{\mathbb{Q}}$. Since also $m_{\zeta_n}^{\mathbb{Q}} \mid \Phi_n^{\mathbb{Q}}$ (Lemma 3.5.6) and both are monic, $m_{\zeta_n}^{\mathbb{Q}} = \Phi_n^{\mathbb{Q}}$. □

Example 6.9.9. Consider 7th roots of unity over $\mathbb{F}_2$. We have the factorization

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

All roots of unity $\neq 1$ are primitive; 3 have the first cubic factor as minimal polynomial and the other 3 the second; $\Phi_7^{\mathbb{F}_2}$ is reducible; the 7th cyclotomic field over $\mathbb{F}_2$ is an extension of degree $3 \neq \varphi(7) = 6$. Note 3 is the order of 2 modulo 7. We see in the proof below that this is no coincidence.

Recall Definition 2.7.1 and Theorem 2.7.8.

Proposition 6.9.10. *Let $n > 0$ and $p \nmid n$ prime. $\Phi_n^{\mathbb{F}_p}$ is irreducible in $\mathbb{F}_p[X]$ if and only if p is a primitive root of n.*

Proof. For $\zeta \in C_n^{\mathbb{F}_p}$ primitive, $m_{\zeta}^{\mathbb{F}_p} \mid \Phi_n^{\mathbb{F}_p}$. Hence, $\Phi_n^{\mathbb{F}_p}$ is irreducible in $\mathbb{F}_p[X]$ if and only if $m_{\zeta}^{\mathbb{F}_p} = \Phi_{\zeta}^{\mathbb{F}_p}$, if and only if $d := \deg(m_{\zeta}^{\mathbb{F}_p}) = \varphi(n)$, if and only if (by Artin) $d = [\mathbb{F}_p(\zeta) : \mathbb{F}_p] = \varphi(n)$. Now $\mathbb{F}_p(\zeta) \cong \mathbb{F}_{p^d}$, and we can assume $=$. By Proposition 6.7.20, d equals the order of the Frobenius endomorphism ψ in $G := G(\mathbb{F}_{p^d} \mid \mathbb{F}_p)$. The group monomorphism of G into $\mathbb{Z}_n^\times$ from Theorem 6.9.3 (recall the proof) maps ψ to $\bar{p} \in \mathbb{Z}_n^\times$. Thus, d is the order of $\bar{p}$ in $\mathbb{Z}_n^\times$. This means that $d = \varphi(n)$ if and only if p is a primitive root of n. □

Exercise 6.9.11. Show $f := X^{12} - 729 \in \mathbb{Q}[X]$ has splitting field $\mathbb{Q}(\zeta_{12})$ over $\mathbb{Q}$ (use $729 = 3^6$ and $\zeta_{12} = (\sqrt{3} + i)/2$). Determine all intermediate fields of $\mathbb{Q}(\zeta_{12}) \mid \mathbb{Q}$.

Exercise 6.9.12. $n > 0$ is prime if and only if $\Phi_n^{\mathbb{Q}} = X^{n-1} + \cdots + X + 1$.

6.9.1 Constructibility of regular n-gons

We are now equipped to address the last of the classical Greek problems on ruler and compass constructions, the construction of regular n-gons.

Recall, *Fermat primes* are primes of the form $2^k + 1$ for $k \in \mathbb{N}$ (Remark 2.3.5 (1)).

Theorem 6.9.13 (Gauß, Wantzel)**.** *Let $n > 2$. Then ζ_n is constructible if and only if $n = 2^m p_1 \cdots p_\ell$ for some $\ell, m \in \mathbb{N}$ and pairwise distinct Fermat primes p_i.*

Proof. We claim ζ_n is constructible if and only if $\varphi(n)$ is a power of 2. $\Rightarrow$: if ζ_n is constructible, then $[\mathbb{Q}(\zeta_n) : \mathbb{Q}]$ is a power of 2 by Corollary 6.2.23; but $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$ by Gauß' Theorem 6.9.8. $\Leftarrow$ follows from Corollary 6.8.4 because $\mathbb{Q}(\zeta_n)$ is the splitting field of $m_{\zeta_n}^{\mathbb{Q}}$ over $\mathbb{Q}$ (Remark 6.9.2 (3)).

Let $n = p_1^{k_1} \cdots p_\ell^{k_\ell}$ be the prime factorization of n. By Theorem 2.6.10,

$$\varphi(n) = p_1^{k_1 - 1} \cdots p_\ell^{k_\ell - 1} (p_1 - 1) \cdots (p_\ell - 1).$$

This is a power of 2 if and only if for all $p_i \neq 2$ we have $k_i = 1$ and $p_i - 1$ is a power of 2. □

Remark 6.9.14. The ancient Greeks constructed regular n-gons for $n = 2^k, 2^k 3, 2^k 5$ where $k \in \mathbb{N}$. Gauß' 17-gon marked the first progress after 2 milennia (Example 6.1.6).

Since there are 5 known Fermat primes (see Remark 2.3.5 (1)), for odd n we know $2^5 - 1 = 33$ constructible regular n-gons; the largest is $n = 4294967295$. Isn't it remarkable that we know this regular n-gon is constructible without ever possibly see a construction?

An explicit construction of a regular $2^{16} + 1 = 65537$-gon was worked out in more than 221 pages and 10 years by Hermes (1894); he believed "Geduld ist die Pforte der Freude."

6.9.2 Dirichlet's theorem

As a further application we prove a special case of Dirichlet's theorem (cf. Remark 2.3.5 (5)).

Theorem 6.9.15 (Dirichlet)**.** *For every $n > 1$ there are infinitely many primes p with*

$$p \equiv 1 \mod n.$$

Proof. Let $s \in \mathbb{N}$ and $p_1, \ldots, p_s$ be primes with $p_i \equiv 1 \mod n$. We have to find another such prime. Let $x := np_1 \cdots p_s > 1$, understanding $x = n$ if $s = 0$. Note $|\Phi_n^{\mathbb{Q}}(x)| = \prod_j |x - \zeta_n^j| > 1$ where j ranges over $1 \leqslant j \leqslant n$ coprime to n; hence, $\Phi_n^{\mathbb{Q}}(x) \notin \{0, \pm 1\}$. As $\Phi_n^{\mathbb{Q}} \in \mathbb{Z}[X]$, we have $\Phi_n^{\mathbb{Q}}(x) \in \mathbb{Z}$. Let p be a prime divisor of $\Phi_n^{\mathbb{Q}}(x)$.

By $\Phi_n^{\mathbb{Q}}(x) \mid x^n - 1$ we have $x^n \equiv 1 \mod p$, so $p \nmid x$, so $p \neq p_i$. We are left to show $p \equiv 1 \mod n$. As $p \nmid x$ we have $\bar{x} \in \mathbb{Z}_p^\times$; let k be its order. As $x^n \equiv 1 \mod p$ we have $k \mid n$. Since $x^{p-1} \equiv 1 \mod p$ by Fermat, also $k \mid p - 1$. It suffices to show $k = n$.

Otherwise, $n = k\ell$ for some $\ell > 1$. Now, the roots of

$$(X^n - 1)/(X^k - 1) = (X^k)^{\ell - 1} + \cdots + X^k + 1 =: f$$

in $\mathbb{C}$ are the nth roots of unity that are not kth roots of unity, in particular, they contain the primitive nth roots of unity. Thus, $\Phi_n^{\mathbb{Q}} \mid f$ in $\mathbb{C}[X]$, hence in $\mathbb{Q}[X]$. Since both are monic, $\Phi_n^{\mathbb{Q}} \mid f$ in $\mathbb{Z}[X]$ (Lemma 4.4.15). Thus,

$$\Phi_n^{\mathbb{Q}}(x) \mid f(x) \equiv \ell \mod p.$$

As $p \mid \Phi_n^{\mathbb{Q}}$ we get $p \mid \ell$; as $\ell \mid n \mid x$ we get $p \mid x$; but $p \nmid x$ was observed above. □

Remark 6.9.16 (Inverse Galois problem)**.** It is unknown whether every finite group H is the Galois group of some Galois extension of $\mathbb{Q}$. Shafarevich (1958) verified this for all solvable H. For illustration, we show:

Proposition 6.9.17. *For every $n > 1$ there exists a Galois field extension $L \mid \mathbb{Q}$ with $G(L \mid \mathbb{Q}) \cong \mathbb{Z}_n$.*

Proof. Dirichlet's theorem gives a prime p with $p \equiv 1 \mod n$. By Theorem 6.9.8, $\mathbb{Q}(\zeta_p) \mid \mathbb{Q}$ is Galois of degree $\varphi(p) = p-1$ with Galois group $G \cong \mathbb{Z}_p^\times$, cyclic of order $p-1$ (Theorem 2.7.6). As $n \mid p-1$, G has a cyclic subgroup U of order $(p-1)/n$ (Theorem 5.3.21), so $[G : U] = n$. As G is abelian, U is normal. By (5) of the main theorem, $L := \mathbb{Q}(\zeta_p)^U \mid \mathbb{Q}$ is Galois with $G(L \mid \mathbb{Q}) \cong G/U$. But G/U is cyclic of order $[G : U] = n$, so $\cong \mathbb{Z}_n$. □

Example 6.9.18. $G\big(\mathbb{Q}(\cos(2\pi/7)) \mid \mathbb{Q}\big) \cong \mathbb{Z}_3$.

Proof. Follow the proof above with $n = 3$ and $p = 7$. Then $G(\mathbb{Q}(\zeta_7) \mid \mathbb{Q}) \cong \mathbb{Z}_6^\times$. Indeed, it contains the automorphisms φ_k with $\varphi_k(\zeta_7) = \zeta_7^k$ for $k < 7$. Note $\varphi_6^2 = \mathrm{id}$ as $\varphi_6(\varphi_6(\zeta_7)) = \zeta^{36} = \zeta_7$. Hence $U := \{\mathrm{id}, \varphi_6\}$ is a subgroup of order $2 = 6/3$. Then $\mathbb{Q}(\zeta_7)^U \mid \mathbb{Q}$ is Galois with group $G(\mathbb{Q}(\zeta_7)^U \mid \mathbb{Q}) \cong G(\mathbb{Q}(\zeta_7) \mid \mathbb{Q})/U \cong \mathbb{Z}_3$.

To determine $\mathbb{Q}(\zeta_7)^U$ it suffices to find $\alpha \in \mathbb{Q}(\zeta_7)^U \smallsetminus \mathbb{Q}$: then $[\mathbb{Q}(\alpha) : \mathbb{Q}] > 1$ and $3 = [\mathbb{Q}(\zeta_7)^U : \mathbb{Q}] = [\mathbb{Q}(\zeta_7)^U : \mathbb{Q}(\alpha)] \cdot [\mathbb{Q}(\alpha) : \mathbb{Q}]$ imply $[\mathbb{Q}(\zeta_7)^U : \mathbb{Q}(\alpha)] = 1$, so $\mathbb{Q}(\zeta_7)^U = \mathbb{Q}(\alpha)$.

But $\varphi_6(\zeta_7 + \zeta_7^6) = \zeta_7^6 + \zeta_7^{36} = \zeta_7^6 + \zeta_7 \in L^U$ and $\zeta_7^6 = \bar{\zeta}_7$ (complex conjugation), so $\zeta_7 + \zeta_7^6 = 2\mathrm{Re}(\zeta_7) = 2\cos(2\pi/7) =: \alpha$. It is known that $\alpha \notin \mathbb{Q}$. □

Exercise 6.9.19. For every finite group H there are field extensions $L \mid M \mid \mathbb{Q}$ with $L \mid M$ Galois and $G(L \mid M) \cong H$. (*Hint:* Proposition 6.7.22.)

6.10 Adjunctions of roots

Definition 6.10.1. Let $n > 0$. A field extension $L \mid K$ *results by adjunction of an nth root* if and only if $L = K(b)$ for some $b \in L$ with $b^n \in K$.

$L \mid K$ *results by adjunction of a root* if this happens for some $n > 0$.

Example 6.10.2. $\mathbb{Q}(\sqrt[3]{2}) \mid \mathbb{Q}$ and $\mathbb{Q}(\sqrt[3]{2}, \zeta_3) \mid \mathbb{Q}(\zeta_3)$ and $\mathbb{Q}(\zeta_3) \mid \mathbb{Q}$ result by adjunction of a root. The 1st is not Galois, the others are.

Theorem 6.10.3. *Let $L \mid K$ a field extension of degree $n > 1$ and assume K contains a primitive nth root of unity and $\mathrm{char}(K) \nmid n$. Then $L \mid K$ results by adjoining an nth root if and only if $L \mid K$ is a cyclic Galois extension.*

Lemma 6.10.4 (Pure equations)**.** *Let $n > 1$ and K be a field with $\mathrm{char}(K) \nmid n$. Let $a \in K^\times$, and $L \subseteq \overline{K}$ be the splitting field of $X^n - a$ over K.*

1. *L contains a primitive nth root of unity ζ.*
2. *If $b \in L$ is a root of $X^n - a$, then $L = K(b, \zeta)$.*

3. $L \mid K$ *and* $L \mid K(\zeta)$ *are Galois extensions.*
4. $G(L \mid K(\zeta))$ *is isomorphic to a subgroup of* $\mathbb{Z}_n$*; it is isomorphic to* $\mathbb{Z}_n$ *if* $X^n - a$ *is irreducible in* $K(\zeta)[X]$*.*

Proof. (1): no root of $X^n - a$ is a root of $(X^n - a)' = \underline{n}X^{n-1}$, so has n pairwise distinct roots $b_1, \dots, b_n \in L$ (Lemma 3.3.13). Then b_i/b_1 are n pairwise distinct roots of unity. Thus, L contains the nth cyclotomic field over K.

(2): $b, \zeta b, \zeta^2 b, \dots, \zeta^{n-1} b$ are pairwise distinct roots of $X^n - a$, so equal $b_1, \dots, b_r$ above. Then $L = K(b_1, \dots, b_n) = K(b, \zeta)$.

(3): $X^n - a$ is separabel (having n roots), so $L \mid K$ (clearly, finite) is Galois by Artin. Then also $L \mid K(\zeta)$ is Galois by Lemma 6.7.4.

(4): every $\varphi \in G(L \mid K(\zeta))$ maps b to $\zeta^{k_\varphi} b$ for some $k_\varphi < n$; as $L = K(\zeta)(b)$ this determines φ (Remark 6.3.12). Thus, $\varphi \mapsto \bar{k}_\varphi$ defines a group monomorphism into $\mathbb{Z}_n$. If $X^n - a$ is irreducible in $K(\zeta)[X]$, the action of $G(L \mid K(\zeta))$ on the roots is transitive (Remark 6.7.10 (3)), i.e., this monomorphism is surjective. □

Proof of $\Rightarrow$ *in Theorem 6.10.3.* Assume $L = K(b)$ with $a := b^n \in K$ and $\zeta \in K$ for an n-th root of unity ζ. Apply Lemma 6.10.4: by (2), L is the splitting field of $X^n - a$, $L \mid K$ is Galois by (3), and the Galois group is cyclic by (4) (and Corollary 5.3.23). □

For the case $K = \mathbb{Q}$, we can additionally describe $G(L \mid K)$. Recall Exercise 2.6.5.

Proposition 6.10.5. *Let* $n > 1, a \in \mathbb{Q}$ *and assume* $f := X^n - a \in \mathbb{Q}[X]$ *is irreducible in* $\mathbb{Q}(\zeta_n)[X]$*. Let* $\Phi : \mathbb{Z}_n^\times \cong \mathrm{Aut}(\mathbb{Z}_n)$ *be given by* $\Phi(\bar{k})(\bar{x}) := \overline{kx}$ *for all* $\bar{k} \in \mathbb{Z}_n^\times, \bar{x} \in \mathbb{Z}_n$*. Then*

$$G(f, \mathbb{Q}) \cong \mathbb{Z}_n \rtimes_\Phi \mathbb{Z}_n^\times.$$

Proof. Let $L \subseteq \mathbb{C}$ be the splitting field of f over $\mathbb{Q}$. Then $\zeta_n \in L$. The roots are $b, \zeta_n b, \dots, \zeta_n^{n-1} b$ for some $b \in L$. Let $\varphi \in G(f, \mathbb{Q}) = G(L \mid \mathbb{Q})$. Then $\varphi(\zeta_n)$ is a primitive nth root of unity, $\varphi(\zeta_n) = \zeta_n^{\ell_\varphi}$ for some $\ell_\varphi < n$, coprime to n. Using the notation from the previous proof, we define the desired isomorphism Ψ by

$$\Psi(\varphi) := \left(\bar{k}_\varphi, \bar{\ell}_\varphi\right).$$

Then Ψ is injective: $(k_\varphi, \ell\text{–}\varphi)$ determine the values of φ on the roots. Surjectivity follows using Artin, the degree formula and the lemma plus Theorem 6.9.8:

$$|G(f, \mathbb{Q})| = [L : \mathbb{Q}] = [L : \mathbb{Q}(\zeta_n)] \cdot [\mathbb{Q}(\zeta_n) : \mathbb{Q}] = |\mathbb{Z}_n| \cdot |\mathbb{Z}_n^\times|.$$

To show Ψ is a homomorphism, let $\varphi, \psi \in G(L \mid \mathbb{Q})$; then

$$\begin{aligned} \varphi(\psi(b)) &= \varphi(\zeta_n^{k_\psi} b) = \varphi(\zeta_n)^{k_\psi} \varphi(b) = \zeta_n^{k_\psi \ell_\varphi} \zeta_n^{k_\varphi} b = \zeta_n^{k_\psi \ell_\varphi + k_\varphi} b, \\ \varphi(\psi(\zeta_n)) &= \varphi(\zeta_n^{\ell_\psi}) = \zeta_n^{\ell_\varphi \ell_\psi}. \end{aligned}$$

Thus, $\Psi(\varphi \circ \psi) = \left(\overline{k_\psi \ell_\varphi + k_\varphi}, \overline{\ell_\varphi \ell_\psi}\right) = \left(\bar{k}_\varphi + \Phi(\bar{\ell}_\varphi)(\bar{k}_\psi), \bar{\ell}_\varphi \cdot \bar{\ell}_\psi\right) = \Psi(\varphi) \cdot \Psi(\psi)$. □

Remark 6.10.6. Let K be a field, $a \in K$ and p prime. Then $X^p - a \in K[X]$ is irreducible if and only if it has no root in K.

Proof. $\Rightarrow$ is clear. $\Leftarrow$: assume $g \mid X^p - a$ for some monic $g \in K[X]$ of degree $0 < d < p$. Write $X^p - a = (X - \alpha_1)\cdots(X - \alpha_p)$ with $\alpha_i \in \overline{K}$ and assume $g = (X - \alpha_1)\cdots(X - \alpha_d)$. As g has coefficients in K we have $b := \alpha_1 \cdots \alpha_d \in K$. As $\alpha_i^p = a$ we have $b^p = a^d$. By Bézout, $1 = xd + yp$ for certain $x, y \in \mathbb{Z}$. Then $a = a^{xd}a^{yp} = b^{px}a^{yp}$, so $X^p - a$ has root $b^x a^y \in K$. □

The following states that so-called *characters of G in K* are linearly independent.

Lemma 6.10.7 (Artin)**.** *Let $n > 0$, K a field, G a group, $\chi_1, \ldots, \chi_n : G \to K^\times$ pairwise distinct group homomorphisms, and $a_1, \ldots, a_n \in K$ not all 0. Then there is $g \in G$ such that*

$$a_1\chi_1(g) + \cdots + a_n\chi_n(g) \neq 0.$$

Proof. Otherwise there is a minimal $n > 0$ such that this fails. Then $n \neq 1$ since $a_1\chi_1(1_G) = a_1 \neq 0$. Let $g \in G$. Choose $g' \in G$ with $\chi_1(g') \neq \chi_n(g')$. Then $\sum_i a_i\chi_i(g) = 0$ implies $\sum_i a_i\chi_n(g')\chi_i(g) = 0$. Further, $\sum_{i=1}^n a_i\chi_i(g'g) = \sum_i a_i\chi_i(g')\chi_i(g) = 0$. Subtracting gives

$$\textstyle\sum_{i=1}^n a_i(\chi_i(g') - \chi_n(g'))\chi_i(g) = \sum_{i=1}^{n-1} a_i(\chi_i(g') - \chi_n(g'))\chi_i(g).$$

As g is arbitrary, and n minimal, all $a_i(\chi_i(g') - \chi_n(g')) = 0$. But $(\chi_1(g') - \chi_n(g')) \neq 0$, so $a_1 = 0$. But then $\sum_{i=2}^n a_i\chi_i(g) = 0$ for all $g \in G$, so again by minimality of n, all $a_i = 0$. □

Proof of $\Leftarrow$ in Theorem 6.10.3. Let $\zeta \in K$ be a primitive nth root of unity. Let φ generate $G := G(L \mid K)$. Since $|G| = n$ by Artin, $\mathrm{id}_L, \varphi, \ldots, \varphi^{n-1}$ lists G. Note $\chi_k := \varphi^k{\restriction}L^\times$ for $k < n$ are pairwise distinct group homomorphisms from $L^\times$ into $L^\times$. By Artin's lemma above, there is $x \in L^\times$ such that $b := \sum_{k=0}^{n-1} \zeta^k\chi_k(x) \neq 0$. Then, as ζ is fixed by φ and $\varphi^n(x) = x$,

$$\varphi(b) = \varphi(x) + \zeta\varphi^2(x) + \cdots + \zeta^{n-2}\varphi^{n-1}(x) + \zeta^{n-1}\varphi^n(x) = \zeta^{-1}b.$$

Thus $\varphi(b^n) = \zeta^{-n}b^n = b^n$, so $a := b^n$ is fixed by $G(L \mid K)$. Thus, $a \in K$ by Theorem 6.7.6.

We are left to show $L = K(b)$. By the degree formula, it suffices to show $[K(b) : K]$ has degree $\geqslant n$. By Corollary 6.8.2 (1), $K(b) \mid K$ is Galois, so $[K(b) : K] = |G(K(b) : K)|$ by Artin. Hence, it suffices to find n pairwise distinct K-automorphisms of $K(b)$.

Easy: for $k < n$ we have $\varphi^k(b) = \zeta^{-k}b$ and, as $b \neq 0$, these values are pairwise distinct; hence, $\varphi^k{\restriction}K(b)$ are pairwise distinct; they take values in $K(b)$ by Lemma 6.7.4. □

6.11 Radical extensions

Definition 6.11.1. A field extension $L \mid K$ is a *radical extension* if there are $r > 0$ and $L_0 = K \subseteq L_1 \subseteq \cdots \subseteq L_r = L$ such that $L_{i+1} \mid L_i$ results by adjunction of a root for all $i < r$.

$f \in K[X]$ is *solvable with radicals (over K)* if f splits in some radical extension of K.

Intuitively, this means the roots of f can be computed from the coefficients using field operations and n-th roots for various $n > 1$.

Examples 6.11.2.

1. By Theorem 6.1.5, every constructible $z \in \mathbb{C}$ is contained in a radical extension of $\mathbb{Q}$, namely one obtained by successive adjunctions of square roots.
2. Lemma 3.5.11 shows, if $\mathrm{char}(K) \neq 2$, then a quadratic $f \in K[X]$ splits in $K(\sqrt{D_f})$.
3. Cardano's formulas (Proposition 3.5.15 and Remark 3.5.14) show that every cubic $f \in \mathbb{Q}[X]$ is solvable by radicals. In particular, $f = X^3 + aX^2 + b \in \mathbb{Q}[X]$ splits in L_4:

$$L_0 := \mathbb{Q} \subseteq L_1 := L_0(\zeta_3) \subseteq L_2 := L_1(\delta) \subseteq L_3 := L_2(x) \subseteq L_4 := L_3(y),$$

where $\zeta_3^3 = 1 \in L_0$, $\delta^2 = -D_f/27 \in L_1$, $x^3 = (\delta - b)/2 \in L_2$, $y^3 = (\delta + b)/2 \in L_3$.

The following is the reason why solvable groups are called 'solvable'.

Theorem 6.11.3. *Let K be a field with $\mathrm{char}(K) = 0$ and $f \in K[X]$. Then f is solvable with radicals if and only if $G(f, K)$ is solvable.*

Of course, we intend to apply Galois theory. A difficulty is that radical extensions can be non-Galois, even if all steps $L_{i+1} \mid L_i$ are Galois.

Example 6.11.4. $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[4]{2})$ is a radical extension with Galois steps but $\mathbb{Q}(\sqrt[4]{2}) \mid \mathbb{Q}$ is not Galois as it does not contain the complex roots $\pm i\sqrt[4]{2}$ of $m^{\mathbb{Q}}_{\sqrt[4]{2}} = X^4 - 2$.

Lemma 6.11.5. *Assume $L_0 = K \subseteq L_1 \subseteq \cdots \subseteq L_r = L$ are fields with $\mathrm{char}(K) = 0$ such that $L_{i+1} \mid L_i$ results by adjunction of a root for all $i < r$. Then there exist $m \in \mathbb{N}$ such that for all multiples $n \in \mathbb{N}$ of m there exist $s \in \mathbb{N}$ and fields*

$$K \subseteq L'_0 \subseteq L'_1 \subseteq \cdots \subseteq L'_s$$

such that $L \subseteq L'_s$ and $L'_s \mid K$ is finite Galois, $L'_0 = K(\zeta)$ for a primitive nth root of unity $\zeta \in \overline{K}$, $L'_0 \mid K$ and $L'_{i+1} \mid L'_i$ for all $i < s$ are cyclic Galois extensions.

Proof. Induction on r. For $r = 0$ set $m := 1$ and let $n > 0$ be given. For a primitive n-th root of unity $\zeta \in \overline{K}$, and $K(\zeta) \mid K$ is a cyclic Galois extension (Theorem 6.9.3).

Let $r > 0$ and L_is be given. Let $L_r = L_{r-1}(b)$ with $b^k \in L_{r-1}$. For the first $r-1$ many L_is choose m' by induction and set $m := km'$. Let n be a multiple of m. Then m is a multiple of m' so induction gives $L'_0, \ldots, L'_{s'}$, in particular, $L'_{s'} \mid K$ is Galois and contains b^k.

Write $G := G(L'_{s'} \mid K)$ and define

$$f := \textstyle\prod_{\varphi \in G}(X^k - \varphi(b^k)).$$

Then $f \in L'_{s'}[X]$. But $\varphi(f) = f$ for all $\varphi \in G$, so $f \in (L'_{s'})^G[X] = K[X]$ (Theorem 6.7.6). By Artin, $L'_{s'}$ is the splitting field of some $g \in K[X]$. Let L' be the splitting field of fg over K, so $L' \mid K$ is finite Galois by Artin. Then g splits over L', so $L'_{s'} \subseteq L'$. As $f(b) = 0$, also $b \in L'$ and hence $L_r = L_{r-1}(b) \subseteq L'_{s'}(b) \subseteq L'$.

But L' is a radical extension of $L'_{s'}$: successively adjoin the roots of f. These are kth roots of certain $\varphi(b^k) \in L'_{s'}$. Since $L'_0 \subseteq L'_{s'}$ contains a primitive nth root of unity ζ, it contains also a primitive kth root of unity, namely $\zeta^{n/k}$ (note $k \mid n$). Thus, each such adjoinment produces a cyclic Galois extension by Theorem 6.10.3. □

Proof of Theorem 6.11.3. $\Rightarrow$: given a radical extension $L \mid K$ where f splits, choose $K \subseteq L_0' \subseteq \cdots \subseteq L_s' =: L'$ as in the lemma. By the main theorem,

$$\{\mathrm{id}_{L'}\} = G(L' \mid L_s') \lhd G(L' \mid L_{s-1}') \lhd \cdots \lhd G(L' \mid L_0') \lhd G(L' \mid K).$$

Moreover, for $i < s$, $G(L' \mid L_i')/G(L' \mid L_{i+1}') \cong G(L_{i+1}' \mid L_i')$ is cyclic, so abelian. Also $G(L' \mid K)/G(L' \mid L_0') \cong G(L_0' \mid K) = G(K(\zeta) \mid K)$ is abelian by Theorem 6.9.3.

Thus, $G(L' \mid K)$ is solvable (Theorem 5.8.15). Let $\tilde{L} \subseteq L'$ be the splitting field of f over K. Then $\tilde{L} \mid K$ is Galois by Artin. By the main theorem $G(L' \mid \tilde{L}) \lhd G(L' \mid K)$ and

$$G(\tilde{L} \mid K) \cong G(L' \mid K)/G(L' \mid \tilde{L}).$$

But factors of solvable groups are solvable (Lemma 5.8.13).

$\Leftarrow$: let $G(f, K) = G(L \mid K)$ for $L \subseteq \overline{K}$ the splitting field of f. Choose intermediate fields $K = M_0, M_1, \ldots, M_\ell$ of $L \mid K$ according to Corollary 6.8.2 (3).

Let $n := [L : K]$ and ζ be a primitive nth root of unity. By Artin, M_1 is the splitting field of some $f \in K[X]$ over K. Then $M_1(\zeta)$ is the splitting field of $(X^n - 1)f$ over $M_0 = K$, so $M_1(\zeta) \mid K$ is Galois, so $M_1(\zeta) \mid K(\zeta)$ is Galois (Lemma 6.7.4).

Further, $\varphi(M_1) = M_1$ for every $\varphi \in G(M_1(\zeta) \mid K(\zeta)) \subseteq G(M_1(\zeta) \mid K)$ (Lemma 6.7.4), i.e., $\varphi{\restriction}M_1 \in G(M_1 \mid K)$. Thus, $\varphi \mapsto \varphi{\restriction}M_1$ is a group homomorphism from $G(M_1(\zeta) \mid K(\zeta))$ into $G(M_1 \mid K)$. It is injective: if φ, ψ agree on M_1, then also on $M_1(\zeta)$ (both fix ζ).

Thus, $G(M_1(\zeta) \mid K(\zeta))$ is isomorphic to a subgroup of $G(M_1 \mid K)$. But $p_1 := [M_1 : K] = |G(M_1 \mid K)|$ is prime, so $|G(M_1(\zeta) \mid K(\zeta))| = [M_1(\zeta) : K(\zeta)]$ is p_1 or 1.

Repeating this argument,

$$M_0 = K \subseteq K(\zeta) \subseteq M_1(\zeta) \subseteq \cdots \subseteq M_\ell(\zeta) = L(\zeta)$$

with $M_i(\zeta) \mid M_{i-1}(\zeta)$ Galois of prime degree $p_i := [M_i : M_{i-1}]$ or 1. Consider an i with $M_i(\zeta) \neq M_{i-1}(\zeta)$. By Artin, $G(M_i(\zeta) \mid M_{i-1}(\zeta))$ has order p_i, so is cyclic. As $p_i \mid n$, $M_{i-1}(\zeta)$ contains a primitive p_ith root of unity (namely, ζ^{n/p_i}). By Theorem 6.10.3, $M_i(\zeta) \mid M_{i-1}(\zeta)$ results by adjoining a p_ith root.

Since also the first step $K(\zeta) \mid K$ results by adjoining a root, $L(\zeta) \mid K$ is a radical extension. Since f splits in L, it splits in $L(\zeta)$. □

6.11.1 The Abel-Ruffini theorem

Recall Examples 6.11.2. The *Mitternachtsformel* is general in the sense that it is a single formula where we plug the coefficients of a given quadratic polynomial to find roots. Similarly, we saw general formulas for degree 3 and Cardano's school also found general formulas for degree 4. Other degrees cannot be handled this way:

Theorem 6.11.6 (Abel-Ruffini)**.** *Let $n > 1$ and $S_1, \ldots, S_n$ be variables and K be a field of characteristic* 0*. The* general degree n polynomial over K

$$f := X^n - S_1 X^{n-1} + S_2 X^{n-1} + \cdots + (-1)^n S_n \in K(S_1, \ldots, S_n)[X]$$

is solvable with radicals over K if and only if $n < 5$.

Proof. Let L be the splitting field of f over $K(S_1, ..., S_n)$. Then $L \mid L_0 := K(S_1, ..., S_n)$ is Galois by Artin (note f is separable since $\operatorname{char}(K(S_1, ..., S_n)) = 0$). Let $x_1, ..., x_n \in L$ be the roots of f. Then $L = K(x_1, ..., x_n)$ because $S_i = s_{n,i}(x_1, ..., x_n)$ by Vieta's formula.

By Theorem 6.11.3 and Example 5.8.10 it suffices to show $G(L \mid L_0) \cong S_n$. By Proposition 6.7.22 is suffices to show $G(L \mid L_0) \cong G(M \mid M_0)$ where $M := K(X_1, ..., X_n)$ and $M_0 := K(s_{n,1}, ..., s_{n,n})$. By Theorem 3.7.7, $L_0 \cong M_0$ via the K-homomorphism φ that maps S_i to $s_{n,i}$. We are left to show that φ^{-1} extends to an isomorphism from M onto L.

We verify this for the homomorphism $\psi : M \to L$ determined by $X_i \mapsto x_i$. Since

$$\psi(s_{n,i}) = s_{n,i}(x_1, ..., x_n) = S_i = \varphi^{-1}(s_{n,i}),$$

ψ extends φ^{-1}. For surjectivity, note M is the splitting field of

$$(X - X_1)\cdots(X - X_n) = X^n - s_{n,1}X^{n-1} + \cdots + (-1)^n s_{n,n} = \varphi(f).$$

over M_0. Thus, $\psi(M) \subseteq L$ is a splitting field of $\psi(\varphi(f)) = f$ over L_0. Thus, $\psi(M) = L$. □

Can one find 'special' formulas? For each degree 5 equation an own one? No:

Corollary 6.11.7. *Let $f \in \mathbb{Q}[X]$ be irreducible of degree 5 and assume f has exactly 3 roots in $\mathbb{R}$. Then f is not solvable with radicals over $\mathbb{Q}$.*

Proof. By Remark 6.7.10, $\varphi \mapsto \varphi^*$ is a group monomorphism from $G := G(f, \mathbb{Q})$ into S_5. We claim it is surjective. We show that the φ^* generate S_5.

By Remark 6.7.10 (4), $5 \mid |G|$. By Cauchy's theorem 5.12.1, G contains an element φ of order 5. Then $\varphi^* \in S_5$ has order 5, so must be a 5-cycle: write φ^* as a product of disjoint cycles by Theorem 5.2.7 and use Example 5.3.14 (4). We can assume $\varphi^* = (12345)$, enumerating the 5 complex roots of f accordingly.

Letting $\psi \in G$ be complex conjugation, ψ^* is a transposition, namely (ij) if the 2 roots in $\mathbb{C} \setminus \mathbb{R}$ are the ith and the jth. By Exercise 5.3.6, $\langle \varphi^*, \psi^* \rangle = S_5$. □

Example 6.11.8. $f := X^5 - 4X + 2 \in \mathbb{Q}[X]$ cannot be solved with radicals over $\mathbb{Q}$.

Indeed, f is irreducible by Eisenstein and has exactly 3 real roots.