Galois Theory

Lecture Notes

Priv.-Doz. Dr. Michael H. Mertens

Spring 2021

Contents

Preface

For over two millennia it has been the main objective of Algebra to solve equations and study their solutions. This has naturally led to the study of abstract structures such as groups, rings, or fields.

The guiding question of this entire course is whether, and if so how, a polynomial equation

$$f(X) = 0$$

for some polynomial f is solvable, i.e. whether there exists some α such that $f(\alpha) = 0$. Such an α is called a *root* of the polynomial f. The answer to this question naturally depends on the domain where the polynomial is considered: Over the ring of integers $\mathbb{Z}$, the equation $3X - 5 = 0$ has no solution, but over the field of rational numbers $\mathbb{Q}$, it does, namely $\alpha = 5/3$. Similarly one usually learns in school that the equation $X^2 - 2 = 0$ has no solution over $\mathbb{Q}$, i.e. that the real number $\sqrt{2}$ is irrational, but it is solvable over the real numbers $\mathbb{R}$, namely by said number $\sqrt{2} = 1.414213562...$ and its negative $-\sqrt{2}$. There are still polynomial equations over $\mathbb{R}$ which have no solution in $\mathbb{R}$, e.g. $X^2 + 1 = 0$, but by considering the complex numbers $\mathbb{C}$, i.e. numbers of the form $a + b\,\mathrm{i}$ where a and b are real numbers and i is *defined* as one solution of the equation $X^2 + 1 = 0$, i.e. $\mathrm{i}^2 = -1$, one finds that absolutely every polynomial equation has a complex solution. This is known as the *Fundamental Theorem of Algebra* (see Appendix A).

It is a very common problem in Mathematics that one can show that something exists, but it may be much harder to actually construct it explicitly. It is like this also with roots of polynomials. Already in ancient times, people new formulas to solve quadratic equations, i.e. polynomial equations of degree 2. With the invention of (almost) modern algebraic notation it was possible for the Italian mathematicians del Ferro, Tartaglia, Cardano, and Ferrari in the 16th and 17th century to find solution formulae for general cubic (degree 3) and quartic (degree 4) polynomial equations. This lead to the widespread belief that it should be possible to find formulae for the solutions of any polynomial equation, but no one was able to find such a formula for general polynomials of degree larger than 4. As it turns out, such a formula cannot possibly exist in general, which was first discovered independently by Ruffini and Abel in the early 19th century. Roughly speaking they were able to show that the solutions of polynomial equations have certain symmetries and that the existence of a solution formula means that these symmetries must have a specific shape. Any such symmetry that occurs for polynomials of degree ≤ 4 has this shape, but in higher degrees, there are polynomials with more complicated symmetries, so the desired formula cannot exist. It is one of the main goals of this course to understand this result in detail. The general theory behind it was developed by Évariste Galois wherefore it now bears his name, *Galois Theory*.

It is also not uncommon in Mathematics that the resolution of a problem may come from a seemingly unrelated area: In ancient Greece mathematicians were interested in the question whether certain geometric constructions are possible with compass and straightedge (i.e. an unmarked ruler) in finitely many steps. For example, it is not too difficult to divide a given line segment into two equal pieces just using these given tools, or to dissect a given angle. It is more complicated, but still possible to construct a regular pentagon or a pentadecagon (15-gon) with a given side length (or equivalently into a given circle). The ancient Greeks however had four big construction problems which they weren't able to solve:

1. Construction of the regular n-gon for n other than $2^m \cdot 3, 2^m \cdot 4, 2^m \cdot 5, 2^m \cdot 15$ (if a regular n-gon is given for some n, it is easy to construct a regular $2n$-gon just by finding the midpoints of all the edges).

2. Doubling the cube: From a given cube construct another cube that has precisely twice the original cube's volume.

3. Trisecting an angle: Given an angle θ, construct the angle $\theta/3$.

4. Squaring the circle: Construct a square that has the same area as a given circle.

In the late 18th century Gauß was the first to make any meaningful progress in these questions since antiquity: He was able to classify exactly which regular n-gons are constructible by compass and straightedge by turning this geometric question into an algebraic one about polynomial equations. He also provided an explicit method to construct the regular heptadecagon (17-gon). Pierre Wantzel used Gauß's ideas to show in 1837 that doubling the cube as well as trisecting an angle are impossible (in general). The last remaining ancient constructibility problem of squaring the circle wasn't resolved until Ferdinand von Lindemann showed in 1882 that the number $\pi = 3.1415926...$ is *transcendental*, i.e. not the root of any polynomial over $\mathbb{Q}$. Another main point of this course is to illustrate the details of this connection between geometry and algebra as one of the many important applications of Galois Theory.

Galois Theory is ubiquitous throughout many areas of mathematics, such as Topology, Number Theory, Algebraic Geometry, Representation Theory, Differential Equations, and many others. Unfortunately there probably won't be time to discuss these applications in detail in this course.

Prerequisites Students will be assumed to be familiar with the basic concepts of Linear Algebra (abstract vector spaces, endomorphisms, minimal and characteristic polynomials), and Abstract Algebra (e.g. the basic theory of groups, rings, and fields, as well as some single specific results such as the Main Theorem on finitely generated Abelian groups). Much (but not all!) of the necessary material will however be recalled briefly during the course.

Literature It should be sufficient to rely on these lecture notes, which are based very loosely on lecture notes by Prof. Miles Reid from the University of Warwick, which are freely available at `https://homepages.warwick.ac.uk/~masda/MA3D5/Galois.pdf`. Nevertheless there is a large number of alternative texts for additional reading (in no particular order):

- Emil Artin, *Galois Theory*, Dover Publications, 1998 (reprint of the 2nd edition published by Univ. Notre Dame Press, 1944): This is a very compact, but also thorough treatment of the subject, including the basics on Linear Algebra as well as advanced aspects of Galois Theory, written by one of the towering figures of 20th century Algebra.

- Michael Artin, *Algebra*, Pearson, 2nd edition, 2013: One of the (many) standard volumes on abstract algebra, covering also much background on for example Group Theory and Ring Theory.

- Peter Pesic, *Abel's Proof: An Essay on the Sources and Meaning of Mathematical Unsolvability*, MIT Press, 2004: This is recommended for the historical background of Abel's (and Ruffini's) proof of the unsolvability of the general quintic by radicals. The book is written for a general audience, so the focus is on history rather than the actual mathematics, although the author tries to elucidate some of the basic ideas, but not necessarily in a way that is thorough enough for us.

- Serge Lang, *Algebra*, Springer, Graduate Texts in Mathematics 211, 3rd edition, 2005: One of the over 50 textbooks by Serge Lang, and a classic reference for the topic of algebra.

These notes These notes are intentionally a little broader than is perhaps usual. Chapter 2 in particular should contain a lot of material which is already known from other courses, but it wouldn't go amiss to recall this material here. Some readers with a strong background in Abstract Algebra may want to skip some of the sections covering material they know already. To make this easier to judge each section has a short itemized list of topics and results covered in it.

In Chapter 1, we discuss some properties of polynomials and their roots, focussing on polynomials with rational, real, or complex coefficients, which is probably most familiar to most readers. Apart from discussing general relations between polynomials and their roots, we also discuss the known solution formulae for cubic and quartic polynomials. Chapter 2 covers some basic facts on the general theory of rings and fields, recalling (or introducing) important concepts such as ring homomorphisms, ideals, and certain important subclasses of rings. In Chapter 3, we take a close look at the main subject of this course, fields and their extensions. Next to introducing some necessary vocabulary and tools for the later chapters, we discuss one of the main results in this course in this chapter, namely constructibility problems. Chapter 4 then introduces and recalls some basic concepts of group theory and in particular contains a discussion on an important class of groups, the soluble groups, as well as the important Theorem of Jordan-Hölder. In Chapter 5 we put all the previous material together to formulate and prove the key result in this course, the Main Theorem of Galois Theory, and as one application we show that there can't be a solution formula for the general quintic as there was for quadratics, cubics, or quartics. In the final Chapter 6, we take a closer look at the notorious Axiom of Choice and some of its applications.

Some section titles in these notes are marked with an asterisk *. These sections contain additional material which may, but doesn't have to be a little more advanced than the rest. The material in those sections might appear on the homework sheets, but it will not be relevant on any of the quizzes or exams. It is nevertheless recommended to study those as well as the material in there will be useful in later studies, especially in the area of Algebra.

Chapter 1

Polynomial Equations

In this chapter we discuss the known formulae for solutions of a polynomial equation $f(X) = 0$ as well as the very important relations between the roots of a polynomial and its coefficients.

1.1 Roots of polynomials

Topics

- polynomials over $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$
- roots of polynomials
- division with remainder of polynomials
- relation between roots and coefficients

Let us begin by recalling some definitions. To be very concrete we focus on the familiar fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ of rational, real, and complex numbers for this chapter and write K to mean either one of them. Many of the things mentioned here work for general fields in much the same way, although sometimes care must be taken. In later chapters we will address some of these issues and discuss more general fields.

Definition 1.1.1. For a formal variable X and finitely many numbers $a_0, a_1..., a_n \in K$, we call an expression

$$f = a_n X^n + ... + a_1 X + a_0 = \sum_{j=0}^{n} a_j X^j$$

a *polynomial* over K. The numbers $a_0, ..., a_n$ are called the *coefficients* of f. The largest $m \leq n$ such that $a_m \neq 0$ is called the *degree* of the polynomial, we write

$\deg f = m$. If all the coefficients are 0, in which case we also write $f = 0$, then we formally set the degree of f to be $-\infty$. The set of all polynomials over K in the variable X is denoted by $K[X]$.

Often we assume implicitly that the coefficient a_n in Definition 1.1.1 is non-zero, wherefore the degree of the polynomial written as in the definition is n. In fact we often normalise our polynomials so that the leading coefficient $a_n = 1$. In this case we call the polynomial f *monic*.

Remark 1.1.2. *It is easy to see that we can add two polynomials* $f = a_nX^n + ... + a_0, g = b_nX^n + .. + b_0 \in K[X]$ *by adding their respecitve coefficients,*

$$f + g = (a_n + b_n)X^n + ... + (a_0 + b_0) \in K[X].$$

We then have $\deg(f + g) \leq \max\{\deg f, \deg g\}$.

We can also define the product of two polynomials $f, g \in K[X]$ *to get back a polynomial in such a way that* $\deg(f \cdot g) = \deg f + \deg g$ *(Exercise).*

This entire course is about understanding the object of the following definition.

Definition 1.1.3. Let $f = a_nX^n + ... + a_1X + a_0 \in K[X]$ be a polynomial of degree n. For a complex number $\alpha \in \mathbb{C}$ we call

$$f(\alpha) := a_n\alpha^n + ... + a_1\alpha + a_0 \in \mathbb{C}$$

the *value* of f at α. The number $\alpha \in \mathbb{C}$ is called a *root* of f if we have $f(\alpha) = 0$.

We illustrate this with a very familiar example.

Example 1.1.4. As everyone has learned in school one can write down a formula for the roots of a degree 2, i.e. quadratic, polynomial. If $f = aX^2 + bX + c \in \mathbb{R}[X]$ and $a \neq 0$, then the numbers

$$\alpha_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad \text{and} \qquad \alpha_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Here the quantity $\Delta = \Delta(f) = b^2 - 4ac$, called the *discriminant* of f, decides about the quality of these roots: if $\Delta > 0$, then f has two distinct real roots, if $\Delta = 0$, the f has a double root, and if $\Delta < 0$, then f has a pair of two complex conjugate roots.

It is also possible to (almost) reconstruct the coefficients of a polynomial from its roots: We have, as one checks easily,

$$\alpha_1 + \alpha_2 = -\frac{b}{a} \qquad \text{and} \qquad \alpha_1\alpha_2 = \frac{c}{a},$$

so that we can write

$$f = a(X - \alpha_1)(X - \alpha_2).$$

It was Euclid who first showed that the integers are what is now called a *Euclidean domain*, i.e. for two given integers $m, n \neq 0$ we can always write $n = qm + r$, where q and r are again integers satisfying $0 \leq r < |m|$, and this representation is unique. The same is true for polynomials as we shall see now.

Theorem 1.1.5. *Let $f, g \in K[X]$ be non-zero polynomials. Then there exist unique polynomials $q, r \in K[X]$ such that $\deg r < \deg g$ satisfying*

$$f = qg + r.$$

Proof. We have to prove two statements, first existence and then uniqueness of the given representation. We first deal with the question of existence: Suppose that $\deg f \geq \deg g$, otherwise we can choose $q = 0$ and $r = f$ and we are done. We prove the statement by induction on $\deg f$. For $\deg f = 0$, there is nothing to show (polynomials of degree 0 are just constants), so assume the statement is true for polynomials of degree at most n for some $n \geq 0$. We need to show that it is true for polynomials of degree $n+1$. So suppose $f = a_{n+1}X^{n+1} + ... + a_1X + a_0$ and $g = b_mX^m + ... + b_1X + b_0$ and $m \leq n+1$. Let $q_0 = \frac{a_{n+1}}{b_m}X^{n+1-m}$. Then by construction $f - q_0g$ has degree $\leq n$, so that by induction hypothesis we have polynomials q_1, r_1 with $\deg r_1 < \deg g$ such that $f - q_0g = q_1g + r_1$. Rearranging this yields $f = (q_0 + q_1)g + r_1$, so the claim of existence follows by induction.

Now for the uniqueness: Suppose $f = q_1g + r_1 = q_2g + r_2$ with $\deg r_1, \deg r_2 < \deg g$. Then it follows that

$$(q_1 - q_2)g = r_2 - r_1.$$

If $q_1 - q_2 \neq 0$, then the left-hand side has degree at least $\deg g$, while the $\deg(r_2 - r_1) < \deg g$ assumption and the first part of Remark 1.1.2. This can clearly not be true, so we must have $q_1 = q_2$ and consequently also $r_1 = r_2$, therefore uniqueness.

q.e.d.

The inductive proof of Theorem 1.1.5 is effective in that it gives rise to an algorithm to find the polynomials q and r explicitly. This is often referred to as polynomial division. The method is very much akin to the familiar long division algorithm.

Example 1.1.6. Let $f = 3X^4 + 6X^3 - 3X^2 + 4, g = 2X^2 + 3X - 1 \in \mathbb{Q}[X]$. We want to find polynomials $q, r \in \mathbb{Q}[X]$ as in Theorem 1.1.5 such that $f = qg + r$. In the first step we just divide the leading term $3X^4$ of f by the leading term $2X^2$ of g to obtain the leading term $\frac{3}{2}X^2$ of q:

$$3X^4 + 6X^3 \; - 3X^2 \qquad\qquad + 4 = \left(2X^2 + 3X - 1\right)\left(\tfrac{3}{2}X^2 \qquad\qquad\right) \qquad .$$

Now, as in the proof of Theorem 1.1.5, we multiply this leading term of q by g and subtract this from f and look at the resulting leading term:

$$\begin{array}{rrrrrl}
3X^4 & +6X^3 & -3X^2 & & +4 & = \left(2X^2+3X-1\right)\left(\tfrac{3}{2}X^2 \qquad\qquad\qquad\right) \quad . \\
-3X^4 & -\tfrac{9}{2}X^3 & +\tfrac{3}{2}X^2 & & & \\
\cline{1-3}
 & \tfrac{3}{2}X^3 & -\tfrac{3}{2}X^2 & & &
\end{array}$$

Then we repeat the above, now with a polynomial of smaller degree, until we arrive at a polynomial that has smaller degree than g.

$$\begin{array}{rrrrrl}
3X^4 & +6X^3 & -3X^2 & & +4 & = \left(2X^2+3X-1\right)\left(\tfrac{3}{2}X^2+\tfrac{3}{4}X-\tfrac{15}{8}\right) \quad . \\
-3X^4 & -\tfrac{9}{2}X^3 & +\tfrac{3}{2}X^2 & & & \\
\cline{1-3}
 & \tfrac{3}{2}X^3 & -\tfrac{3}{2}X^2 & & & \\
 & -\tfrac{3}{2}X^3 & -\tfrac{9}{4}X^2 & +\tfrac{3}{4}X & & \\
\cline{2-4}
 & & -\tfrac{15}{4}X^2 & +\tfrac{3}{4}X & +4 & \\
 & & \tfrac{15}{4}X^2 & +\tfrac{45}{8}X & -\tfrac{15}{8} & \\
\cline{3-5}
 & & & \tfrac{51}{8}X & +\tfrac{17}{8} &
\end{array}$$

This smaller degree polynomial, $\frac{51}{8}X+\frac{17}{8}$ in this example, is exactly the remainder r, so that the total computation looks like

$$\begin{array}{rrrrrl}
3X^4 & +6X^3 & -3X^2 & & +4 & = \left(2X^2+3X-1\right)\left(\tfrac{3}{2}X^2+\tfrac{3}{4}X-\tfrac{15}{8}\right)+\tfrac{51}{8}X+\tfrac{17}{8}. \\
-3X^4 & -\tfrac{9}{2}X^3 & +\tfrac{3}{2}X^2 & & & \\
\cline{1-3}
 & \tfrac{3}{2}X^3 & -\tfrac{3}{2}X^2 & & & \\
 & -\tfrac{3}{2}X^3 & -\tfrac{9}{4}X^2 & +\tfrac{3}{4}X & & \\
\cline{2-4}
 & & -\tfrac{15}{4}X^2 & +\tfrac{3}{4}X & +4 & \\
 & & \tfrac{15}{4}X^2 & +\tfrac{45}{8}X & -\tfrac{15}{8} & \\
\cline{3-5}
 & & & \tfrac{51}{8}X & +\tfrac{17}{8} &
\end{array}$$

We now relate division of polynomials to roots.

Definition 1.1.7. Let $f, g \in K[X]$ be polynomials. If there is a polynomial $q \in K[X]$ such that $f = q \cdot g$, we say that g *divides* f and we write $g \mid f$. The polynomial g is then also called a *factor* or *divisor* of f. If f has no non-trivial factors, i.e. constants or constant multiples of itself, we say that f is *irreducible*.

Remark 1.1.8. *Whether or not a polynomial is irreducible depends heavily on the chosen field K. For example, the polynomial $X^2 - 2 \in \mathbb{Q}[X]$ is irreducible, but the same polynomial in $\mathbb{R}[X]$ is reducible since $X^2 - 2 = (X - \sqrt{2})(X + \sqrt{2}) \in \mathbb{R}[X]$.*

In the following corollary we relate the roots of a polynomial to its factors.

Corollary 1.1.9. *Let $f \in K[X]$ be a polynomial of degree n and $\alpha \in K$. Then the following are true:*

1. *α is a root of f if and only if $(X - \alpha)$ divides f.*

2. *f has at most n not necessarily distinct roots.*

Proof. Exercise.

q.e.d.

Remark 1.1.10. *If $K = \mathbb{C}$ in Corollary 1.1.9, then the Fundamental Theorem of Algebra, a proof of which can be found in Appendix A, guarantees that $f = a_n X^n + ... + a_0$ has exactly n roots, counted with multiplicities, i.e. there are n (not necessarily distinct) numbers $\alpha_1, ..., \alpha_n \in \mathbb{C}$ such that*

$$f = a_n \prod_{j=1}^{n} (X - \alpha_j).$$

We saw in Example 1.1.4 that there is a relation between the roots and the coefficients of a quadratic polynomial in that we can express the coefficients in terms of the roots and vice versa. We now want to generalise the former property to polynomials of arbitrary degree. In order to do this we need the notion of symmetric polynomials. Recall first the definition of the *symmetric group* on n letters, denoted by S_n, which we may identify with the group of bijective maps

$$\pi : \{1, ..., n\} \to \{1, ..., n\}.$$

Definition 1.1.11. 1. A polynomial in n variables $P \in K[X_1, ..., X_n]$ is called *symmetric* if for any permutation $\pi \in S_n$ we have

$$P(X_{\pi(1)}, ..., X_{\pi(n)}) = P(X_1, ..., X_n).$$

2. For $0 \leq k \leq n$ we let σ_k denote the k^{th} *elementary symmetric polynomial*, which is defined as

$$\sigma_k(X_1, ..., X_n) := \sum_{\substack{M \subseteq \{1,...,n\} \\ \#M = k}} \prod_{j \in M} X_j = \sum_{1 \leq i_1 < i_2 < ... < i_k \leq n} \prod_{j=1}^{k} X_{i_j}.$$

Explicitly the elementary symmetric polynomials look as follows:

$$\begin{aligned}
\sigma_1 &= X_1 + X_2 + ... + X_n \\
\sigma_2 &= X_1X_2 + X_1X_3 + ... + X_{n-1}X_n \\
&\vdots \\
\sigma_{n-1} &= X_1X_2 \cdots X_{n-1} + X_1 \cdots X_{n-2}X_n + ... + X_2X_3 \cdots X_n \\
\sigma_n &= X_1X_2 \cdots X_n.
\end{aligned}$$

We formally set $\sigma_0 = 1$.

Remark 1.1.12. *As the name suggests, the elementary symmetric polynomials are indeed symmetric polynomials.*

Proof. Perhaps easiest way to see this is to introduce an auxiliary variable Y and to consider the polynomial

$$F = \prod_{j=1}^{n}(Y + X_j). \tag{1.1}$$

It follows directly from the distributive law that

$$F = \sum_{k=0}^{n} \sigma_{n-k}(X_1, ..., X_n)Y^k. \tag{1.2}$$

Now for any permutation $\pi \in S_n$ we have

$$\sum_{k=0}^{n} \sigma_{n-k}(X_1, ..., X_n)Y^k = \prod_{j=1}^{n}(Y + X_j) = \prod_{j=1}^{n}(Y + X_{\pi(j)}) = \sum_{k=0}^{n} \sigma_{n-k}(X_{\pi(1)}, ..., X_{\pi(n)})Y^k, \tag{1.3}$$

since permuting the X_j only permutes the order of the factors in the product, but not the product itself. Comparing the coefficients of Y^k on both extremes of (1.3), we see that indeed

$$\sigma_k(X_{\pi(1)}, ..., X_{\pi(n)}) = \sigma_k(X_1, ..., X_n).$$

q.e.d.

From the above proof, we obtain directly our desired relation between roots and coefficients of a polynomial.

Corollary 1.1.13. *Let $f = X^n + a_{n-1}X^{n-1} + ... + a_0 \in K[X]$ be a monic polynomial of degree n with (not necessarily distinct) roots $\alpha_1, ..., \alpha_n$. Then we have for all $k \in \{0, ..., n\}$ that*

$$a_k = (-1)^{n-k}\sigma_{n-k}(\alpha_1, ..., \alpha_n).$$

Proof. By Corollary 1.1.9 we can write $f = \prod_{j=1}^n (X - \alpha_j)$, which by (1.1) and (1.2) can be rewritten as

$$f = \sum_{k=0}^n \sigma_{n-k}(-\alpha_1, ..., -\alpha_n)X^k = \sum_{k=0}^n (-1)^{n-k}\sigma_{n-k}(\alpha_1, ..., \alpha_n)X^k.$$

Comparing coefficients yields the claim.

q.e.d.

1.1.1 Symmetric polynomials*

Before we conclude this section, we take a look at symmetric polynomials in general.

Example 1.1.14. Consider the polynomial $P = X_1^2 + ... + X_n^2$. This is clearly a symmetric polynomial. We comapare this to

$$\sigma_1^2 = \sum_{i_1+...+i_n=2} \binom{2}{i_1, ..., i_{n-1}} X_1^{i_1} \cdots X_n^{i_n} = (X_1^2 + ... + X_n^2) + 2\sum_{i<j} X_i X_j,$$

where we used the so-called *multinomial theorem* in the first step, which states that

$$(X_1 + ... + X_n)^k = \sum_{i_1+...+i_n=k} \binom{k}{i_1, ..., i_{n-1}} X_1^{i_1} \cdots X_n^{i_n}$$

with

$$\binom{k}{i_1, ..., i_{n-1}} = \frac{k!}{i_1! \cdots i_{n-1}! i_n!}$$

and which we will prove in the exercises. We recognise the second summand on the right-hand side to be $2\sigma_2$, so that we find that we can express P in terms of elementary symmetric polynomials,

$$P = \sigma_1^2 - 2\sigma_2.$$

More generally we can express *any* symmetric polynomial in terms of the elementary symmetric polynomials.

Theorem 1.1.15. *Let $P \in K[X_1, ..., X_n]$ be a symmetric polynomial. Then there is a polynomial $Q_P \in K[Y_1, ..., Y_n]$ such that*

$$P = Q_P(\sigma_1, ..., \sigma_n).$$

Proof. Every polynomial is a sum of monomials, i.e. expressions of the form $X_1^{i_1} \cdots X_n^{i_n}$, $i_j \geq 0$. We introduce an order on these monomials, called the *lexicographical order*: We say that $X_n \prec X_{n-1} \prec ... \prec X_1$ and for two monomials $m = X_1^{i_1} \cdots X_n^{i_n}$ and $m' = X_1^{i'_1} \cdots X_n^{i'_n}$ we say that $m \prec m'$ if $i_k < i'_k$ for some $k \leq n$ and $i_j = i'_j$ for all $0 \leq j < k$. If such a k doesn't exist, then the monomials are equal. So the "largest" monomials are those with the largest power of X_1, among those the ones with the largest power of X_2 are largest and so forth.

Now consider the leading monomial, i.e. the one highest with respect to the ordering $\prec$ among those occuring in P, say $X_1^{i_1} \cdots X_n^{i_n}$, of our given polynomial P. Since P is a symmetric polynomial, the exponents are in decreasing order, $i_1 \geq i_2 \geq ... \geq i_n$.

The leading term of the elementary symmetric polynomial σ_k is clearly given by $X_1 \cdots X_k$, wherefore the leading term of a product of the form $\sigma_1^{b_1}\sigma_2^{b_2} \cdots \sigma_n^{b_n}$ for some $b_j \geq 0$ is given by

$$X_1^{b_1+...+b_n} X_2^{b_2+...+b_n} \cdots X_n^{b_n}.$$

By choosing $b_j = i_j - i_{j+1}$ for $j = 1, ..., n-1$ and $b_n = i_n$ we match the leading monomial of P. If a denotes the coefficient of the leading monomial in P, then we see that

$$P_1 = P - a\sigma_1^{i_1-i_2}\sigma_2^{i_2-i_3} \cdots \sigma_n^{i_n}$$

is a symmetric polynomial whose leading monomial is strictly smaller than that of P. This means that at least one of the exponents in the leading monomial of P_1 is smaller than the corresponding one in P, so this process can only be repeated a finite number of times before we reach zero.

q.e.d.

Theorem 1.1.15 is one of the first results one shows in a course on e.g. Invariant Theory. In that language symmetric polynomials are exactly the so-called *invariant ring* of the group S_n.

1.2 Roots of unity

Topics

- roots of unity
- cyclotomic polynomials
- action of the cyclic group

In this short section we discuss a very important class of polynomials and their roots.

The polynomials in question are

$$X^n - 1, \qquad n \in \mathbb{N}.$$

Definition 1.2.1. The complex roots of the polynomial $X^n - 1$ are called the nth *roots of unity.*

An nth root of unity ζ is called *primitive* if $\zeta^m \neq 1$ for all $m < n$.

It is well known that the complex roots of unity are given by ζ_n^k, $k = 0, ..., n-1$, with

$$\zeta_n := e^{2\pi\,\mathrm{i}/n} = \cos(2\pi/n) + \mathrm{i}\sin(2\pi/n).$$

Lemma 1.2.2. *The primitive nth roots of unity are given by ζ_n^k, $k = 0, ..., n-1$, $\gcd(k, n) = 1$. In particular, there are exactly $\varphi(n)$ primitive nth roots of unity, where $\varphi(n) = \#\{k \in \{0, ..., n-1\} : \gcd(n, k) = 1\} = \#(\mathbb{Z}/n\mathbb{Z})^\times$ denotes Euler's totient function.*

Proof. Any root of unity has the form ζ_n^k. Let $d = \gcd(n, k)$. Then we have

$$(\zeta_n^k)^{n/d} = \zeta_n^{n \cdot k/d} = (\zeta_n^n)^{k/d} = 1,$$

since both n/d and k/d are integers. Thus if $d > 1$ then there is a number $m = n/d < n$ (in fact $m \mid n$) such that $(\zeta_n^k)^m = 1$. On the other hand if $d = 1$, then there is $k' \in \{0, ..., n-1\}$ such that $kk' = 1 + \ell n$ for some $\ell \in \mathbb{Z}$, wherefore

$$(\zeta_n^k)^{k'} = \zeta_n^{1+\ell n} = \zeta_n \cdot (\zeta_n^n)^\ell = \zeta_k.$$

Thus if there were some $m < n$ such that $(\zeta_n^k)^m = 1$, then we would also have $\zeta_n^m = 1$, which is absurd.

q.e.d.

The nth *cyclotomic polynomial* is now definied by

$$\Phi_n = \prod_{\substack{k=0 \\ \gcd(n,k)=1}}^{n-1} (X - \zeta_n^k) \in \mathbb{C}[X].$$

One can compute the first few of these polynomials explicitly to find for instance

$$\begin{aligned}
\Phi_1 &= X - 1 \\
\Phi_2 &= X + 1 \\
\Phi_3 &= X^2 + X + 1 \\
\Phi_4 &= X^2 + 1 \\
\Phi_5 &= X^4 + X^3 + X^2 + X + 1 \\
\Phi_6 &= X^2 - X + 1 \\
\Phi_7 &= X^7 + X^6 + X^5 + X^4 + X^3 + X^2 + X + 1 \\
\Phi_8 &= X^4 + 1 \\
\Phi_9 &= X^6 + X^3 + 1.
\end{aligned}$$

This suggests that all these polynomials are indeed defined over $\mathbb{Q}$, which, as we shall see later, together with the fact that they are also irreducible over $\mathbb{Q}$, is indeed true. Another fact that these computations seems to suggest is that the coefficients of Φ_n are always 0 or ± 1, which is actually false: The first counterexample is the coefficient of X^{41} in Φ_{105} which is -2. In fact it can be shown that the coefficients of cyclotomic polynomials become arbitrarily large.

Remark 1.2.3. *Consider the polyomial $X^n - a$ for some $a \in K$ and let $\alpha = \sqrt[n]{a}$ denote the (or better an) nth root of a. Then all complex roots of the polynomial $X^n - a$ are given by $\zeta_n^k \sqrt[n]{a}$, $k = 0, ..., n-1$.*

It should be noted that it is in general not so simple to find all complex roots of a given polynomial from just one.

One important feature of roots of unity is that they realize the cyclic group as a subgroup of $\mathbb{C}^\times$. The action of the cyclic group C_n can be reformulated in a very handy way in terms of roots of unity:

Remark 1.2.4. *Suppose we have an action of $C_n = \langle g \rangle$ on n objects $\alpha_1, ..., \alpha_n$, i.e. g maps*

$$\alpha_1 \mapsto \alpha_2 \mapsto ... \mapsto \alpha_n \mapsto \alpha_1.$$

Now consider the formal expressions

$$\begin{aligned} d_0 &= \alpha_1 + \alpha_2 + \ldots + \alpha_n \\ d_1 &= \alpha_1 + \zeta_n^{-1}\alpha_2 + \ldots + \zeta_n^{-(n-1)}\alpha_n, \\ d_2 &= \alpha_1 + \zeta_n^{-2}\alpha_2 + \ldots + \zeta_n^{-(n-2)}\alpha_n, \\ &\vdots \end{aligned}$$

The action of g then corresponds to multiplying d_j by ζ_n^j. It is also possible to recover $\alpha_1, \ldots, \alpha_n$ from $d_0, \ldots, d_{n-1}$:

$$\alpha_j = \frac{1}{n}\sum_{k=0}^{n-1} \zeta_n^{k(j-1)} d_k.$$

1.3 Cubic and quartic equations

Topics

- Cardano's formula for cubic equations
- Ferrari's formula for quartic equations

In Example 1.1.4 we recalled the well-known way to obtain the roots of a quadratic polynomial directly from its coefficients, using only basic arithmetic operations and square-roots. This has been known since antiquity. In this section, we sketch how to obtain similar, but much more complicated, formulas for the solutions of polynomials of degree 3 or 4. These formulas were found in the 16th and 17th century by various Italian mathematicians (Cardano, Tartaglia, del Ferro, Ferrari).

We first consider monic cubic polynomials

$$X^3 + aX^2 + bX + c \tag{1.4}$$

and begin with a small remark that allows us to simplify the question.

Remark 1.3.1. *Substituting X by $X - a/3$ in (1.4) yields*

$$X^3 + \left(b - \frac{a^2}{3}\right)X + \left(\frac{2}{27}a^3 - \frac{1}{3}ab + c\right),$$

so we may assume that our cubic takes the form

$$X^3 + 3pX + 2q, \tag{1.5}$$

the additional factors in front of p, q being there to make the later formulas look nicer.

We have the following formula, which was first found by Tartaglia and independently by del Ferro and later published (against Tartaglia's strict instructions) by Cardano, which is why it is still referred to as *Cardano's formula.*

Theorem 1.3.2. *The roots of the polynomial $f = X^3 + 3pX + 2q$ are given by*

$$\sqrt[3]{-q + \sqrt{p^3 + q^2}} + \sqrt[3]{-q - \sqrt{p^3 + q^2}},$$

where the cube roots must be chosen so that their product is $-p$.

Before we prove this, let us consider an example.

Example 1.3.3. Consider the polynomial $X^3 + 9X - 26$. In the notation of Theorem 1.3.2, this means that $p = 3$ and $q = -13$, thus $p^3 + q^2 = 196 = 14^2$. The expressions under the cube-roots are therefore $13 + 14 = 27$ and $13 - 14 = -1$. Taking the product of the real cube-roots indeed yields $-p$, so we find one root to be $3 + (-1) = 2$. The other two roots are given by $3\zeta_3 - \zeta_3^2 = -1 + 2i\sqrt{3}$ and $3\zeta_3^2 - \zeta_3 = -1 - 2i\sqrt{3}$.

Proof of Theorem 1.3.2. Let α, β, γ denote the 3 complex roots of the given polynomial f. By Corollary 1.1.13 we find that

$$\alpha + \beta + \gamma = 0 \tag{1.6}$$

$$\alpha\beta + \alpha\gamma + \beta\gamma = 3p \tag{1.7}$$

$$\alpha\beta\gamma = -2q. \tag{1.8}$$

The cyclic group of order 3 acts on the three roots and as well on the set $\{x, y, z\}$, where x, y, z are some formal quantities, so that by the reasoning in Remark 1.2.4 we may write

$$\begin{aligned}\alpha &= x + y + z\\ \beta &= x + \zeta_3^{-1}y + \zeta_3 z\\ \gamma &= x + \zeta_3 y + \zeta_3^{-1}z.\end{aligned}$$

Plugging these expressions into (1.6),(1.7), and (1.8) gives the equations

$$x = 0, \qquad yz = -p, \qquad \text{and} \qquad y^3 + z^3 = -2q.$$

But this means that y^3 and z^3 are the roots of the quadratic polynomial $g = Y^2 + 2qY - p^3$ (cf. 1.1.4), so that we find

$$y^3 = -q + \sqrt{q^2 + p^3} \qquad \text{and} \qquad z^3 = -q - \sqrt{q^2 + p^3}.$$

The additional requirement $yz = -p$ then yields the claim of the theorem.

q.e.d.

As we see in the proof the computation of roots of a quadratic polynomial is required to solve a cubic polynomial. As we shall see now, this pattern continues. In order to find the roots of a quartic polynomial one has to be able to find the roots of a cubic polynomial. The proof of the formula is in spirit very similar to that of Theorem 1.3.2.

Remark 1.3.4. *Let $f = X^4 + aX^3 + bX^2 + cX + d \in K[X]$ be a monic polynomial of degree 4. Then we have $f(X - a/4) = X^4 + rX^2 + sX + t$ for some suitable $r, s, t \in K$, so we may restrict our analysis to polynomials of this shape.*

The following formula for the solution of the quartic was found by Ferrari in the 17th century.

Theorem 1.3.5. *The complex roots of the polynomial $f = X^4 + rX^2 + sX + t \in K[X]$, where not all r, s, t can equal 0, are given by*

$$\frac{1}{2}(u+v+w),\ \frac{1}{2}(u-v-w),\ \frac{1}{2}(-u+v-w),\ \frac{1}{2}(-u-v+w),$$

where u^2, v^2, w^2 are the roots of the resultant cubic

$$Y^3 + 2rY^2 + (r^2 - 4t)Y - s^2$$

and u, v, w have to be chosen such that $uvw = -s$.

Proof. Let $\alpha_1, ..., \alpha_4$ denote the complex roots of f. We have the following relation between the roots and the coefficients,

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0 \tag{1.9}$$
$$\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_1\alpha_4 + \alpha_2\alpha_3 + \alpha_2\alpha_4 + \alpha_3\alpha_4 = r \tag{1.10}$$
$$\alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_2\alpha_4 + \alpha_1\alpha_3\alpha_4 + \alpha_2\alpha_3\alpha_4 = -s \tag{1.11}$$
$$\alpha_1\alpha_2\alpha_3\alpha_4 = t. \tag{1.12}$$

As in the proof of the Cardano formula Theorem 1.3.2 we consider symmetries of the roots, but not with respect to a cyclic group this time. Why will become clear later. We define quantities u, v, w by setting

$$2\alpha_1 = u + v + w \tag{1.13}$$
$$2\alpha_2 = u - v - w \tag{1.14}$$
$$2\alpha_3 = -u + v - w \tag{1.15}$$
$$2\alpha_4 = -u - v + w, \tag{1.16}$$

the factors of 2 on the left-hand side again being more of a cosmetic nature. By substituting (1.13)–(1.16) into (1.9)–(1.12) we obtain, after a somewhat tedious calculation, which we skip, that

$$\begin{aligned} u^2+v^2+w^2 &= -2r, \\ uvw &= -s, \\ u^2v^2+u^2w^2+v^2w^2 &= r^2-4t. \end{aligned} \tag{1.17}$$

Note that the first and third lines in (1.17) are actually the first two elementary symmetric polynomials in u^2, v^2, w^2, and squaring the second line, we get the third elementary symmetric polynomial, so that u^2, v^2, w^2 are precisely the roots of the cubic polynomial

$$Y^3+2rY^2+(r^2-4t)Y-s^2,$$

from where the theorem follows.

q.e.d.

Remark 1.3.6. *In the proofs of Theorem 1.3.2 and Theorem 1.3.5 we exploited certain symmetries of the roots of the polynomials. This is the key idea of Galois theory and will lead to the proof that a formula for the roots of a general quintic, sextic, or higher degree polynomial in terms of the coefficients —using only elementary arithmetic and nth roots— is impossible.*

Chapter 2

Rings and Fields

In this chapter we recall some basic concepts (rings and fields) which should be familiar from a course in abstract algebra. The aim will be to establish some important irreducibility criteria. Note that from here on, we explicitly do not restrict our attention to fields like $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, but consider completely general fields.

2.1 Basic definitions

2.1.1 Integral domains and fraction fields

Topics

- Ring axioms
- Basic properties
- Units
- Integral domains
- Fields of fractions

Definition 2.1.1. A *ring* (with unity) is a set R together with two binary operations $+ : \ R \times R \to R, \ (a, b) \mapsto a + b$ and $\cdot : \ R \times R \to R, \ (a, b) \mapsto a \cdot b$ satisfying the following properties, where a, b, c are arbitrary elements of R:

1. $a \overset{+}{\cdot} b = b \overset{+}{\cdot} a$ (*commutativity*)
2. $(a \overset{+}{\cdot} b) \overset{+}{\cdot} c = a \overset{+}{\cdot} (b \overset{+}{\cdot} c)$ (*associativity*)
3. There exists an element $0 \in R$ such that $a + 0 = a$ for all $a \in R$ (*existence of zero*)

4. For all $a \in R$ there is $b \in R$, called an *additive inverse* of a and usually denoted by $-a$, such that $a + b = 0$ (*existence of inverses*).

5. There exists an element $1 \in R$ such that $a \cdot 1 = a$ for all $a \in R$ (*existence of one*).

6. $a \cdot (b + c) = a \cdot b + a \cdot c$ (*distributivity*).

We don't bother going through all the consequences of this definition, e.g. that inverses, ones and zeros are unique etc. which are easy to derive from the definition.

We collect some names for rings with special properties in the following definition.

Most of the time in this course we work with special rings.

Definition 2.1.2. 1. We call a ring R an *(integral) domain* if it doesn't have any zero-divisors, i.e. we have the implication $a \cdot b = 0 \Rightarrow a = 0$ or $b = 0$.

2. A ring K is called a *field* if $1 \neq 0$ and for every $a \in R \setminus \{0\}$ there exists $b \in R$, called a *multiplicative inverse* of a and usually denoted by a^{-1}, such that $a \cdot b = 1$.

Since many rings one usually works with are integral domains, it is important to remember that this is definitely not true for all rings (think for instance about the rings $\mathbb{Z}/n\mathbb{Z}$ with addition and multiplication defined modulo n for composite n or $\mathbb{Q} \times \mathbb{Q}$, where we define addition and multiplication componentwise).

Definition 2.1.3. Let R be any ring and $u \in R$. We call u a *unit* in R if there exists $v \in R$ such that $u \cdot v = 1$. The set of units is denoted by $R^\times$.

Therfore we think of fields as rings where each non-zero element is a unit. Note also that products of units are again units and the inverse of a unit is again a unit, wherefore $R^\times$ is a *group.*

Definition 2.1.4. Let R, S be rings and $\varphi : R \to S$ be a map. If we have $\varphi(a +_R b) = \varphi(a) +_S \varphi(b)$, $\varphi(a \cdot_R b) = \varphi(a) \cdot_S \varphi(b)$, and $\varphi(1_R) = 1_S$, we call φ a *ring homomorphism.* If φ is injective, we call it an *embedding* (or sometimes *ring monomorphism*), if it is bijective, we call it a *ring isomorphism.*

Integral domains and fields are very closely related, as the following proposition shows.

Proposition 2.1.5. *1. Let R be an integral domain. Then there is a field $K = \operatorname{Frac} R$, called the* field of fractions *of R such that there is a canonical embedding $R \hookrightarrow K$. If we have any field E and an embedding $\iota : R \hookrightarrow E$, then we can extend ι to an embedding $\hat{\iota} : K \hookrightarrow E$.*

2. *Let R be any ring and K a field such that there is an embedding $\iota : R \hookrightarrow K$, then R is an integral domain.*

Proof.

1. For $a, b \in R$, $b \neq 0$, we define the *fraction* $\frac{a}{b}$ as a formal expressions. Two fractions $\frac{a}{b}$ and $\frac{c}{d}$ are considered equal if $a \cdot_R d = b \cdot_R c$ in R, where we indicate the operations in R by a subscript R throughout this proof. Let $K = \operatorname{Frac} R$ denote the set of all such fractions. We define addition of two fractions in the familiar way,
$$\frac{a}{b} +_K \frac{c}{d} := \frac{a \cdot_R d +_R b \cdot_R c}{b \cdot_R d} \quad \text{and} \quad \frac{a}{b} \cdot_K \frac{c}{d} := \frac{a \cdot_R c}{b \cdot_R d}.$$
We note that since R is an integral domain and $b, d \neq 0$ by assumption, we also have $b \cdot_R d \neq 0$, so the addition and multiplication introduced above are indeed well-defined. We can embed R into K in a canonical way through $a \mapsto \frac{a}{1}$. It is straightforward to check that K with the addition and multiplication above is a ring, where the neutral elements are given by $0_K = \frac{0}{1}$ and $1_K = \frac{1}{1}$, which we will just write as 0 and 1 from here on. To see that K is a field we see easily that for a fraction $\frac{a}{b} \neq 0$ (i.e. $a, b \neq 0$) we have
$$\frac{a}{b} \cdot_K \frac{b}{a} = \frac{a \cdot_R b}{b \cdot_R a} = \frac{a \cdot_R b}{a \cdot_R b} = \frac{1}{1} = 1,$$
therfore multiplicative inverses for all non-zero elements, making K a field, as we claimed.

   Now let E be field such that there is an embedding $\iota : R \hookrightarrow E$. Then we can define the map
$$\hat{\iota} : K \to E, \ \frac{a}{b} \mapsto \iota(a) \cdot_E \iota(b)^{-1}.$$
Since $b \neq 0$, we also have $\iota(b) \neq 0$ because ι is injective, so there is a (in fact unique) inverse of $\iota(b)$ in E. One the checks in a straightforward manner that this map is indeed a ring homomorphism. Since all fields are in particular integral domains (exercise) and a ring homomorphism $\varphi : R \to S$ is injective if and only if $\operatorname{Ker} \varphi := \{r \in R \ : \ \varphi(r) = 0\} = \{0\}$, we find that
$$\hat{\iota}\left(\frac{a}{b}\right) = 0 \quad \Leftrightarrow \quad \iota(a) = 0 \text{ or } \iota(b)^{-1} = 0,$$
but $\iota(b)^{-1}$ cannot equal 0, so we must have $a = 0$, whence $\frac{a}{b} = 0$, so that $\hat{\iota}$ is indeed injective.

2. Suppose there are $a, b \in R \setminus \{0\}$ such that $a \cdot b = 0$. Then $\iota(a), \iota(b) \neq 0$ since ι is injective, so that we have
$$0 = \iota(0) = \iota(a \cdot b) = \iota(a) \cdot \iota(b),$$
so that K would have zero-divisors, which is a contradiction to K being a field. Therefore, R cannot have had zero-divisors in the first place, wherefore it is an integral domain by definition.

q.e.d.

Example 2.1.6. The field of fractions for the ring of integers $\mathbb{Z}$ is of course the field of rational numbers $\mathbb{Q}$.

For a polynomial ring $K[X]$ over a field K, its field of fractions is the so-called *field of rational functions*,
$$K(X) := \left\{ \frac{f}{g} \; : \; f, g \in K[X], g \neq 0 \right\}.$$

2.1.2 Ideals and homomorphisms

Topics

- Ideals
- Ring homomorphisms
- Factor rings
- Prime and maximal ideals

We now consider important substructures of rings.

Definition 2.1.7. Let R be a ring and $I \subseteq R$ a subset. We call I an *ideal* of R if

1. we have $0 \in I$ and for $a, b \in I$ we have $a + b \in I$,
2. for $r \in R$ and $a \in I$ we have $r \cdot a \in I$.

We also write $I \trianglelefteq R$.

Ideals are closely connected to ring homomorphisms.

Proposition 2.1.8. *1. Let $\varphi : R \to S$ be a ring homomorphism. Then the* kernel *of φ,*

$$\mathrm{Ker}(\varphi) := \{r \in R \, : \, \varphi(r) = 0\}$$

is an ideal in R.

2. For every ideal $I \trianglelefteq R$ in a ring R there exists a ring S and a ring homomorphism $\varphi : R \to S$ such that $I = \mathrm{Ker}(\varphi)$.

Proof.

1. Exercise.

2. We choose S to be the *factor ring* $R/I := \{a + I \, : \, a \in R\}$ consisting of all residue classes of R modulo I. This means $a, b \in R$ are considered equal in R/I if $a - b \in I$. Defining addition via $(a + I) + (b + I) := (a + b) + I$ and multiplication via $(a+I) \cdot (b+I) := (a \cdot b) + I$ is well-defined and satisfies the ring axioms, so that R/I is indeed a ring. The desired homomorphism φ is then what is sometimes referred to as the *canonical epimorphism* (meaning it is surjective) defined via $a \mapsto a + I$, which is a ring homomorphism by definition of the addition and multiplication in R/I and whose kernel is clearly the ideal I we started with.

q.e.d.

We now define two important classes of ideals.

Definition 2.1.9. Let R be a ring and $I \trianglelefteq R$ an ideal with $I \neq R$.

1. If for all $a, b \in R$ we have he implication

$$ab \in I \Rightarrow a \in I \text{ or } b \in I,$$

we call I a *prime ideal.*

2. I is called a *maximal ideal* if any ideal $J \trianglelefteq R$ containing I is either I or R.

Remark 2.1.10. *The definition of a prime ideal is inspired by a property of (integer) prime numbers, sometimes referred to as Euclid's Lemma: If a prime number p divides a product ab (of integers in other words ab is in the ideal $(p) = p\mathbb{Z}$ generated by p), it has to divide one of the factors.*

It is usually easier to decide whether a given ideal is prime resp. maximal or not by using the following result.

Proposition 2.1.11. *Let R be a ring and $R \neq I \trianglelefteq R$ and ideal.*

1. *I is a prime ideal if and only if R/I is an integral domain.*
2. *I is a maximal ideal if and only if R/I is a field.*

In particular, any maximal ideal is prime.

Proof.

1. Let $a, b \in R$. If $ab \in I$ then $(ab) + I = 0 + I \in R/I$. If I is a prime ideal, this implies that either $a \in I$ or $b \in I$ or equivalently $a + I = 0 + I \in R/I$ or $b + I = 0 + I \in R/I$, so that R/I is an integral domain.

   On the other hand if R/I is an integral domain then we know that for $a, b \in R$ with $(a + I)(b + I) = (ab) + I = 0 + I$, we must have $a + I = 0$ or $b + I = 0$. In other words we have that if $ab \in I$, then either $a \in I$ or $b \in I$, wherefore I is prime.

2. Let $a \in R \setminus I$. If I is maximal then the ideal generated by I and a actually equals R, wherefore in particular there must be some $a' \in R$ and $j \in I$ such that $aa' + j = 1$. But this means precisely that in R/I we have $(a+I)(a'+I) = (aa' + I) = 1 - j + I = 1 + I$, so that $a + I$ is invertible in R/I, wherefore R/I is a field.

   If on the other hand R/I is a field, then for each $a \notin I$, there must be $a' \in R$ such that $(a + I)(a' + I) = 1 + I$ in R/I. Equivalently, this means that there is some $\tilde{j} \in I$ such that $aa' = 1 + \tilde{j}$. It follows that $1 = aa' - \tilde{j}$ lies in the ideal J generated by a and I, so that we must have $J = R$. Therefore I must be maximal and the claim follows.

q.e.d.

2.1.3 Unique factorisation domains, principal ideal domains, and Euclidean domains

Topics

- Unique factorisation domains
- Principal ideal domains
- Euclidean domains

- Prime and irreducible elements
- Greatest common divisor
- Characteristic of a ring

Let us now talk about some more special classes of rings. Before doing so, we need to distinguish two notions which we are used to considering as equivalent from our experience with the integers.

Definition 2.1.12. Let R be ring, $a, b \in R$, and $p \in R$ not a unit.

1. If we have the implication

   $$p = a \cdot b \ \Rightarrow \ a \in R^\times \text{ or } b \in R^\times$$

   we call p *irreducible.*

2. If we have the implication

   $$p \mid (a \cdot b) \ \Rightarrow \ p \mid a \text{ or } p \mid b,$$

   where $a \mid b$ means that there exists some $q \in R$ such that $b = a \cdot q$, then p is called a *prime.*

Remark 2.1.13. *As already mentioned, the concepts of primality and irreducibility coincide for example for the integers $\mathbb{Z}$. Technically, the usual definition of a prime number is that of an irreducible number and the fact that they are indeed prime in this more general sense is known as Euclid's Lemma (see also Remark 2.1.10). However this is not the case for all rings. We will discuss one example in the exercises.*

Definition 2.1.14. Let R be an integral domain.

1. If any $a \in R \setminus (R^\times \cup \{0\})$ has unique (finite) factorisation into irreducibles, i.e. we can write $a = p_1 \cdots p_n$ for irreducible elements $p_1, ..., p_n$ and if we have two such factorisations $a = p_1 \cdots p_n = q_1 \cdots q_m$, then $n = m$ and there is some permutation $\pi : \{1, ..., n\} \to \{1, ..., n\}$ and units $u_1, ..., u_n \in R^\times$ such that $p_j = u_j q_{\pi(j)}$ for all $j \in \{1, ..., n\}$, we call R a *unique factorisation domain* or *factorial domain.*

2. If for any ideal $I \trianglelefteq R$ there exists $a \in R$ such that $I = (a) = aR := \{a \cdot r : r \in R\}$, i.e. each ideal is generated by a single element, and R is an integral domain, we call R a *principal ideal domain.*

3. If there exists a function $N : R \to \mathbb{Z}$ (called a *norm*) such that $N(a) \geq 0$ for all $a \in R$, $N(0) = 0$, and for all $a, b \in R$, $b \neq 0$, there exist $q, r \in R$ such that
$$a = qb + r, \qquad N(r) < N(b),$$
then we call R a *Euclidean domain.*

Example 2.1.15. Euclidean domains are those that admit a division with remainder.

1. We know that the integers admit division with remainder, the norm is then the usual absolute value.

2. In Theorem 1.1.5 we showed that the polynomial ring $K[X]$, where K is $\mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$ (or indeed any field, since we never used any specific properties of the field in the proof), is a Euclidean domain, where the norm function is essentially the degree function, but not quite, as we defined $\deg 0 = -\infty$, which is not a non-negative integer. So to be formally correct, the norm function could be for instance defined by $N(f) = 2^{\deg f}$, where we define (consistently with what analysis would tell us) $2^{-\infty} := 0$.

Note that in a unique factorisation domain, we can always write $a = u \cdot p_1^{m_1} \cdots p_r^{m_r}$ for distinct irreducible elements $p_1, ..., p_r \in R$, a unit $u \in R^\times$, and non-negative integers $m_1, ..., m_r$. This exponents have to be the same up to reordering for any factorisation of a into irreducibles, which motivates the following definition.

Definition 2.1.16. Let R be a unique factorisation domain and suppose we have a fixed set of representatives of primes up to units (e.g. only the positive prime numbers in $\mathbb{Z}$).

1. For $a \in R$ and an irreducible element $p \in R$ we define
$$v_p(a) := \max\{m \in \mathbb{N}_0 \; : \; p^m \mid a\}$$
to be the *p-adic valuation* of a. For a unit $u \in R^\times$ we set $v_p(u) = 0$ for all primes, and also $v_p(0) := \infty$.

2. For $a, b \in R$, not both 0, we can define the *greatest common divisor* of a and b as
$$\gcd(a, b) := \prod_{p \text{ prime}} p^{\min(v_p(a), v_p(b))}. \tag{2.1}$$

Remark 2.1.17. *1. Note that for all but finitely many prime elements, the valuation $v_p(a)$ is zero for any $a \neq 0$, so the product defining the* gcd *in* (2.1) *is always finite and therefore well-defined.*

2. *Using the above definition inductively, one can also define the* gcd *of finitely many elements in a unique factorisation domain.*

These classes of rings are all subsets of one another. To show this we need some preparatory lemmata.

Lemma 2.1.18. *Let R be a principal ideal domain. Then R is* Noetherian, *i.e. for every incresing chain of ideals*

$$I_1 \trianglelefteq I_2 \trianglelefteq I_3 \trianglelefteq ...$$

there is $n \in \mathbb{N}$ such that $I_n = I_m$ for all $m \geq n$.

Proof. Consider the set $I = \bigcup_{j=1}^{\infty} I_j$. Then we claim that I is an ideal in R. For $a, b \in I$, there exist $i, j \in \mathbb{N}$ such that $a \in I_i$ and $b \in I_j$, hence for each $k \geq \max\{i, j\}$ we have $a, b \in I_k$, which is an ideal in R, so that $a + b \in I_k \subseteq I$. For $r \in R$ we have $ra \in I_i \subseteq I$, so that I is indeed an ideal in R. Since R is a principal ideal domain, we have $I = (c)$ for some $c \in I$. But this means that there must be some $n \in \mathbb{N}$ such that $c \in I_n$, whence $I_n = (c) = I$ and $I_m = (c)$ for all $m \geq n$.

q.e.d.

Lemma 2.1.19. *Let R be a principal ideal domain. Then an ideal I is maximal if and only if it is generated by an irreducible element p, $I = (p)$.*

Proof. Let I be a maximal ideal. Since R is a principal ideal domain, there exists $p \in I$ such that $I = (p)$. If we have $p = ab$ for some $a, b \in R$ then we have $I \subseteq (a)$ and $I \subseteq (b)$. But since I is maximal this means that (a) is either equal to R, in which case a would be a unit, or we have $I = (a)$, so that $a = cp$ for some $c \in R$. But this implies $p = pcb$, wherefore b and c are units.

On the other hand let $I = (p)$ for some irreducible element p and $J \trianglelefteq R$ an ideal containing I. Then there is some $c \in J$ with $J = (c)$, so in particular $p = rc$ for some $r \in R$. But since p is irreducible this means that either r or c is a unit, wherefore we have either $J = I$ or $J = R$ and thus I is maximal.

q.e.d.

Lemma 2.1.20. *Let R be a principal ideal domain and $p \in R$ irreducible. Then p is prime.*

Proof. Suppose $p \mid ab$, so $ab \in (p)$. But by Lemma 2.1.19, (p) is a maximal ideal and therefore (see 2.1.11) in particular a prime ideal. By definition this implies that we have $a \in (p)$ or $b \in (p)$ or equivalently $p \mid a$ or $p \mid b$. Thus p is indeed prime.

q.e.d.

Theorem 2.1.21. *1. Any principal ideal domain is a unique factorisation domain.*

2. Any Euclidean domain is a principal ideal domain.

Proof.

1. Let R be a principal ideal domain. We first show that each element in R has an irreducible factor. For this let $a \in R$. If a is itself irreducible, we are done. Otherwise there exist $a_1, b_1 \in R \setminus R^\times$ such that $a = a_1 b_1$, wherefore $a \in (a_1)$ and hence $(a) \subset (a_1)$. This inclusion is strict, because if we had $(a) = (a_1)$, we could write $a_1 = ca$ for some $c \in R$ and thus $a = acb_1$, which implies, since R is an integral domain, that b_1 must be a unit, contrary to our assumption. Now if a_1 is irreducible, we have found an irreducible factor, otherwise we can factor $a_1 = a_2 b_2$ for non-units a_2, b_2, yielding the strict inclusion $(a) \subset (a_1) \subset (a_2)$. Proceeding yields an ascending chain of ideals which by Lemma 2.1.18 has to terminate after finitely many steps. Any generator of the last ideal is then an irreducible factor of a.

   Now we show that a has a factorisation into irreducibles. If a itself is irreducible, there is again nothing to show. If not, from what we showed above, we may write $a = p_1 c_1$ for an irreducible element $p_1 \in R$ and some $c_1 \in R$. Therefore we have $(a) \subset (c_1)$. Now either c_1 is irreducible and we are finished or we can write $c_1 = p_2 c_2$ for some irreducible element p_2 and some element $c_2 \in R$, yielding the inclusion $(a) \subset (c_1) \subset (c_2)$. Again by Lemma 2.1.18, the resulting chain of ideals has to terminate after n steps, say, yielding $a = p_1 p_2 \cdots p_{n-1} c_n$, where c_n has to be irreducible, so we have found our desired factorisation.

   It remains to show that this factorisation is unique up to reordering and units. If we have $a = p_1 \cdots p_n = q_1 \cdots q_m$, then we have $p_1 \mid q_1 \cdots q_m$. Since p is prime by Lemma 2.1.20, this implies that p_1 must divide q_j for some $j \in \{1, ..., m\}$. But since q_j is irreducible, this means that there is a unit $u_1 \in R^\times$ such that $p_1 = u_1 q_j$. Applying the same argument successively to a/p_1, $a/(p_1 p_2), ...$ yields the uniqueness.

2. Let R be a Euclidean domain with norm function N. We first note that if for $a \in R$ we have $N(a) = 0$, then $a = 0$. Otherwise we could write $b = qa + r$ for an arbitrary $b \in R$ and suitable $q, r \in R$ satisfying $0 \leq N(r) < N(a) = 0$, which is a contradiction.

   Now let $I \trianglelefteq R$ be an ideal. If $I = \{0\}$, then it is generated by 0, so we can focus on the case where I contains non-zero elements. Then we may choose $a \in I$ such that
$$N(a) = \min\{N(b) \ : \ b \in I \setminus \{0\}\}.$$
   Note that this minimum is actually attained since the set is non-empty and the non-negative integers are well-ordered, i.e. every non-empty subset contains a smallest element. Note also from the above that $N(a) > 0$. We claim that $I = (a)$. Then for any $b \in I$ there are $q, r \in R$ such that $b = qa + r$ and $N(r) < N(a)$. But $r = b - qa \in I$, so since a has the minimal norm of all non-zero elements in I and the norm of r is strictly smaller, we must have $r = 0$, so $b = qa \in (a)$.

q.e.d.

Remark 2.1.22. *The proof for the uniqueness of factorisation in principal ideal domains is verbatim the same as that for the (known) uniqueness of the prime factorisation in* $\mathbb{Z}$*.*

Remark 2.1.23. *We want to stress that all the inclusions of classes of rings discussed so far (integral domains $\supset$ unique factorisation domains $\supset$ principal ideal domains $\supset$ Euclidean domains) are strict:*

The ring $\mathbb{Z}[\sqrt{-5}] := \{a + b\sqrt{-5} \ : \ a, b \in \mathbb{Z}\}$ is an integral domain, but doesn't have unique factorisation (exercise), the polynomial ring $\mathbb{Z}[X]$ is a factorial domain (see Theorem 2.2.8) but not a principal ideal domain (the ideal $(2, X)$ is not principal), and the ring $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$ is a principal ideal domain, but not a Euclidean domain (the proof of which is too involved to sketch here).

It is probably true for most people that the ring they are most familiar with is the ring of integers $\mathbb{Z}$. As we know, this ring admits division with remainder and is therefore Euclidean, a principal ideal domain, a unique factorization domain, and has all the nice properties we just derived in much greater generality. In some sense, the integers are a very universal ring.

Proposition 2.1.24. *Let R be any ring, then there exits a unique homomorphism $\varphi : \mathbb{Z} \to R$.*

Proof. By definition we must have $\varphi(1) = 1_R$ and $\varphi(0) = 0_R$. Now φ is uniquely determined by setting

$$\varphi(n) = \begin{cases} \underbrace{1_R + ... + 1_R}_{n \text{ times}} & \text{if } n > 0 \\ -\varphi(-n) & \text{if } n < 0. \end{cases} . \tag{2.2}$$

This homomorphism clearly satisfies $\varphi(m + n) = \varphi(m) + \varphi(n)$ and using the fact that multiplication in $\mathbb{Z}$ is nothing else than repeated addition, it is also straightforward to see that it is multiplicative.

Any other homomorphism $\psi : \mathbb{Z} \to R$ must also satisfy $\psi(1) = 1_R$ and $\psi(0) = 0_R$ and because it is a homomorphism, it must satisfy (2.2), wherefore it must equal φ.

q.e.d.

This homomorphism φ gives rise to the following definition.

Definition 2.1.25. Let R be a ring and $\varphi : \mathbb{Z} \to R$ the homormorphism in Proposition 2.1.24. Then the unique integer $n \geq 0$ such that $\ker \varphi = (n)$ is called the *characteristic* of R, we write $n = \text{char}(R)$.

We note the folowing.

Proposition 2.1.26. *If R is an integral domain then we have either* $\text{char}(R) = 0$, *or* $\text{char}(R) = p$ *for a prime number p, so that R contains either $\mathbb{Z}$ or R contains $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ as a subring.*

If $R = K$ is a field, then K contains either $\mathbb{Q}$ (if $\text{char}(K) = 0$ *or $\mathbb{F}_p$ (if* $\text{char}(K) = p$*) as a subfield. $\mathbb{Q}$ resp. $\mathbb{F}_p$ is called the* prime subfield *of K.*

Proof. If $\text{char}(R) = 0$, there is nothing to show in either case since then φ is an embedding and $\mathbb{Q}$ is the fraction field of $\mathbb{Z}$ and therefore the smallest field containing $\mathbb{Z}$.

If $\text{char}(R) = n > 1$ and $n = ab$ is composite, then $\varphi(a)\varphi(b) = 0_R$, so that R has zero-divisors. Hence if R is an integral domain, we must have prime characteristic.

q.e.d.

2.2 Polynomial rings

Topics

- Polynomials over arbitrary rings

In this short section we discuss polynomials over general rings, with a special emphasis on polynomials over fields.

We have already seen the definition of a polynomial ring over a field like $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$ in Definition 1.1.1. The exact same definition of course makes sense when we let the coefficients be in an arbitrary ring.

Definition 2.2.1. Let R be any ring. For a formal variable X and finitely many $a_0, a_1 ..., a_n \in R$, we call an expression

$$f = a_n X^n + ... + a_1 X + a_0 = \sum_{j=0}^{n} a_j X^j$$

a *polynomial* over R. The $a_0, ..., a_n$ are called the *coefficients* of f. The largest $m \leq n$ such that $a_m \neq 0$ is called the *degree* of the polynomial, we write $\deg f = m$. If all the coefficients are 0, in which case we also write $f = 0$, then we formally set the degree of f to be $-\infty$. The set of all polynomials over R in the variable X is denoted by $R[X]$, called the *polynomial ring* over R.

Remark 2.2.2. *Using the usual setting for addition and multiplication of polynomials, we see that $R[X]$ is again a ring.*

We now discuss some general properties of polynomial rings. We begin by some fairly clear ones.

Proposition 2.2.3. *Let R be an integral domain and $f, g \in R[X]$.*

1. $\deg(fg) = \deg f + \deg g$
2. *$R[X]$ is an integral domain.*
3. *f is a unit it $R[X]$ if and only if $\deg f = 0$ and its 0th coefficient is a unit in R.*
4. *R embeds into $R[X]$ via $a \mapsto aX^0$.*

Proof.

1. Let $f = a_n X^n + ...$ and $g = b_m X^m + ...$ with $a_n, b_m \neq 0$, where the dots indicate lower order terms. Then $fg = a_n b_m X^{n+m} + ...$ and since R is an integral domain, we have $a_n b_m \neq 0$, wherefore $\deg(fg) = n + m = \deg f + \deg g$.

2. Follows directly from 1. If you multiply two non-zero polynomials, the degree cannot decrease, so the product cannot be zero.

3. If we have $fg = 1$ for $f, g \in R[X]$, we must have $\deg f + \deg g = \deg 1 = 0$, so both f and g must have degree 0, wherefore they are constant. The only way two constant polynomials can multiply to 1 is clearly when their 0th coefficients are in $R^\times$.

4. This is obvious.

q.e.d.

We have also already shown the following (see Theorem 1.1.5 and notice that we never relied on the fact that the field is anything specific.

Proposition 2.2.4. *If $R = K$ is a field, then $K[X]$ is Euclidean and therefore in particular by Theorem 2.1.21 a unique factorization domain and a principal ideal domain.*

2.2.1 Gauß's lemma

Topics

- Primitive polynomials
- Gauß's Lemma
- Rational root theorem

As we saw in Proposition 2.2.4, the polynomial ring over a field is a Euclidean domain. We would like to investigate properties of polynomial rings over more general rings.

Definition 2.2.5. Let R be a unique factorisation domain and $f = a_nX^n + ... + a_1X + a_0 \in R[X]$ a polynomial. We call f *primitive* if $\gcd(a_0, ..., a_n) = 1$ or in other words if we can write $f = u \cdot g$ for some $u \in R$ and $g \in R[X]$, then u must be a unit.

The following observation is fairly straightforward, but important.

Lemma 2.2.6. *Let R be a unique factorisation domain with field of fractions K. Then any polynomial $f \in K[X]$ can be written as $\frac{p}{q}g$ for $p, q \in R$ with $\gcd(p, q) = 1$ and a primitive polynomial $g \in R[X]$. Up to multiplication by units, the fraction $\frac{p}{q}$ and the polynomial g are unique.*

Proof. Simply multiply f by the least common multiple of the denominators of its coefficients and factor out the greatest common divisor of the numerators.

q.e.d.

We now come to the Gauß's Lemma which gives this section its name.

Theorem 2.2.7. *(Gauß's Lemma) Let R be a unique factorisation domain and $K = \operatorname{Frac} R$ its field of fractions. Then the following are true.*

1. *If $f, g \in R[X]$ are primitive, then so is $f \cdot g$.*
2. *If $f \in R[X]$ is irreducible, then it is also irreducible in $K[X]$.*

Proof.

1. Let p be any prime in R. Since $f = a_nX^n + ... + a_0$ and $g = b_mX^m + ...b_0$ are both primitive, there must be numbers $r \leq n$ and $s \leq m$ such that $a_0, ..., a_{r-1}$ and $b_0, ..., b_{s-1}$ are all divisible by p, but a_r and b_s are not. If such r and s wouldn't exist, then all the coefficients of one of the polynomials would be divisible by p, contradicting primitivity. Now the coefficient of X^{r+s} in $f \cdot g$ is given by

$$\underbrace{a_0b_{r+s} + ... + a_{r-1}b_{s+1}}_{p|} + a_rb_s + \underbrace{a_{r+1}b_{s-1} + ...a_{r+s}b_0}_{p|},$$

   where for $k > n$ we set $a_k := 0$ and similarly $b_\ell := 0$ for $\ell > m$. Since a_r and b_s are both not divisible by p and R is a unique factorisation domain, p connot divide a_rb_s and therefore it doesn't divide the coefficient of X^{r+s} in $f \cdot g$. Therefore, there is no prime number that could divide all the coefficients of $f \cdot g$, wherefore it must be primitive.

2. Let $f \in R[X]$ and write $f = g \cdot h$ for $g, h \in K[X]$. By Lemma 2.2.6 we may write $f = \frac{p}{q} g_0 \cdot h_0$ for $p, q \in R$ with no common factor and $g_0, h_0 \in R[X]$ primitive. By Item 1, $g_0 \cdot h_0$ is primitive, so q must be a unit. But then we have found a factorisation of f in $R[X]$, so the claim follows by contraposition.

q.e.d.

We can now show the following important result.

Theorem 2.2.8. *Let R be a unique factorisation domain. Then $R[X]$ is a unique factorisation domain as well. In particular, any polynomial ring in finitely many variables over a field is a unique factorisation domain.*

Proof. The proof is similar to that of Gauß's Lemma. Let $f \in R[X]$ be a polynomial, which is neither a unit nor zero and let $K = \operatorname{Frac} R$ be the fraction field of R. Then $K[X]$ is Euclidean, wherefore we can write $f = p_1 \cdots p_r$ for irreducible polynomials $p_1, ..., p_r$, which are unique up to constant multiples (i.e. units in $K[X]$). Therefore by Lemma 2.2.6 we can write $f = \frac{a}{b}\tilde{p}_1 \cdots \tilde{p}_r$ for $a, b \in R$ coprime and primitive polynomials $\tilde{p}_1, ..., \tilde{p}_r \in R[X]$. Their product is again primitive by Gauß's Lemma 2.2.7, Item 1, so that b must be a unit in R. Since a has a unique factorisation into prime elements in R, we obtain a factorisation of f into irreducibles in $R[X]$.

It remains to show uniqueness. If we have any other factorisation in $R[X]$, this would yield a different factorisation in $K[X]$, where we know that it is unique, completing the proof.

q.e.d.

Before moving on, we record the following observation which is useful to determine whether a polynomial over a unique factorisation domain R has a root in $K = \operatorname{Frac} R$.

Theorem 2.2.9. *(Rational root theorem) Let R be a unique factorisation domain and $K = \operatorname{Frac} R$ its field of fractions. Then a polynomial $f = a_nX^n + ... + a_0 \in R[X]$ has a root $\alpha = \frac{p}{q} \in K$, p, q coprime, if and only if $p \mid a_0$ and $q \mid a_n$.*

Proof. We may assume without loss of generality that f is primitive. Then $\alpha \in K$ is a root if and only if we have

$$f = (X - \alpha) \cdot g$$

for a polynomial $g = b_{n-1}X^{n-1} + ... + b_0 \in K[X]$. By Lemma 2.2.6 there are $A, B \in R$ coprime such that

$$f = \frac{A}{qB}(qX - p) \cdot \tilde{g}$$

with $\tilde{g} = \tilde{b}_{n-1}X^{n-1} + ... + \tilde{b}_0 \in R[X]$ primitive. Since by Gauß's Lemma 2.2.7 the product $(qX - p) \cdot \tilde{g}$ is again primitive and so was f, it follows that the prefactor $\frac{A}{qB} =: u$ must be a unit. With this we then find by comparing coefficients that $a_n = uq\tilde{b}_{n-1}$ and $a_0 = upb_0$, so that indeed $p \mid a_0$ and $q \mid a_n$ as claimed.

q.e.d.

2.2.2 Irreducibility criteria

Topics

- Reduction $\pmod p$
- Eisenstein's criterion
- Irreducibility of some cyclotomic polynomials

It is often important to test whether or not a given polynomial in $K[X]$ for some field K is irreducible. A far more challenging problem would be to find a factorisation into irreducibles, which unfortunately is out of the scope of these notes.

For the sake of explicitness, we consider the field $K = \mathbb{Q}$, but almost everything here can be generalised to other fields in a straightforward way.

The first criterion for irreducibility is the following.

Theorem 2.2.10. *Let $f = a_n X^n + \dots + a_0 \in \mathbb{Z}[X]$ be a polynomial and p a prime number satisfying $p \nmid a_n$. By considering every coefficient of f modulo p, we obtain a polynomial*

$$\overline{f} = \overline{a}_n X^n + \dots + \overline{a}_0 \in \mathbb{F}_p[X].$$

If $\overline{f}$ is irreducible in $\mathbb{F}_p[X]$, then f is irreducible over $\mathbb{Q}$.

Proof. We may assume without loss of generality that f is primitive, so that by Gauß's Lemma 2.2.7 it is enough to show irreducibility over $\mathbb{Z}$. Suppose $f = g \cdot h$ is reducible over $\mathbb{Z}$. Then, since f is primitive, we must have $\deg g, \deg h \geq 1$. Reducing modulo p we have $\overline{f} = \overline{g} \cdot \overline{h} \in \mathbb{F}_p[X]$. Since the leading coefficient a_n of f is not divisible by p, we must have $\deg \overline{g} = \deg g \geq 1$ and $\deg \overline{h} = \deg h \geq 1$, so that $\overline{f}$ is reducible in $\mathbb{F}_p[X]$. The claim now follows by contraposition.

q.e.d.

Note that it is quite easy in principle to test whether a polynomial over a finite field is irreducible, since there are only finitely many polynomials that may occur as factors, one can simply try them all. There are however usually more efficient ways to accomplish this. If the degree of the polynomial in question is 3 or less, it is actually sufficient to check for roots modulo p (if a degree 3 polynomial over a field is reducible, then at least one of the factors must have degree 1, and therefore it must have a root).

Example 2.2.11.

1. Consider the polynomial $f = X^3 + 39X^2 - 104X - 2 \in \mathbb{Z}[X]$. Reducing this modulo $p = 13$ yields $\overline{f} = X^3 - 2$, which one can check by inspection has no zero over $\mathbb{F}_{13}$ and is therefore irreducible, since the degree of the polynomial is 3. Therefore the polynomial is irreducible over $\mathbb{Q}$. Note however that if one chooses e.g. $p = 2$, then $\overline{f} = X^3 + X^2 + X =$

$X(X^2+X+1)$ is reducible, so it is important to pick the "correct" prime in order to apply Theorem 2.2.10.

2. It is not always possible to apply Theorem 2.2.10 at all. It can be shown by other means that the polynomial $f = X^4 - 10X^2 + 1$ is irreducible over $\mathbb{Q}$, but its reduction $\overline{f}$ modulo any prime p is reducible (Exercise).

Another irreducibility criterion, which is attributed to Eisenstein, is sometimes easier to apply, but again not universal.

Theorem 2.2.12. *(Eisenstein's criterion) Let R be a ring and $P \trianglelefteq R$ be a prime ideal of R. Let $P^2 := \{p \cdot q \,:\, p, q \in P\}$ be the square of the prime ideal P. Suppose that a polynomial $f = a_nX^n + ... + a_0 \in R[X]$ is an* Eisenstein polynomial*, i.e. its coefficients satisfy:*

(a) $a_n \notin P$,

(b) $a_j \in P$ for $j \in \{0, ..., n-1\}$,

(c) $a_0 \notin P^2$.

Then f cannot be factored as $f = g \cdot h$ with polynomials $g, h \in R[X]$ where $\deg(g), \deg(h) \geq 1$.

Proof. Suppose there are polynomials $g = b_mX^m + ... + b_0, h = c_\ell X^\ell + ... + c_0 \in R[X]$ with $m, \ell \geq 1$ and $b_m, c_\ell \neq 0$ such that $f = g \cdot h$. Then we have $a_0 = b_0c_0 \in P \setminus P^2$, so one of b_0 or c_0 must lie in P (since P is a prime ideal), but they cannot both lie in P, since then their product would lie in P^2, which can't be by assumption. Suppose without loss of generality that $b_0 \in P$.

We claim that then all coefficients $b_0, ..., b_m$ lie in P. This is true for b_0, so suppose it is true for all b_j with for all $j < k$, where $k \leq m$. Then we have

$$a_k = \underbrace{b_0c_k + b_1c_{k-1} + ... + b_{k-1}c_1}_{\in P} + b_kc_0.$$

Notice that $m \lneqq n$, so in particular we have $a_k \in P$ by assumption, but since $c_0 \notin P$, this can only happen if $b_k \in P$.

So by induction we have $b_0, ..., b_m \in P$, wherefore $a_n = b_mc_\ell \in P$, contradicting (a). Thus our original assumption must have been false, proving the theorem.

q.e.d.

In combination with Gauß's Lemma we immediately get the following result.

Corollary 2.2.13. *Let $f \in \mathbb{Z}[X]$ be an Eisenstein polynomial. Then f is irreducible over $\mathbb{Q}$.*

Remark 2.2.14.

1. *In most cases we use the ring $R = \mathbb{Z}$, in which case the only viable prime ideals are going to be the ideals $P = (p)$ for a prime number p and $P^2 = (p^2)$.*

2. *Note that the Eisenstein criterion does not assume any further properties of the underlying ring R, making it very flexible.*

We end this section by discussing two examples for the application of the Eisenstein criterion.

Example 2.2.15. For a prime number p consider the pth cyclotomic polynomial

$$\Phi_p(X) = \frac{X^p - 1}{X - 1} = X^{p-1} + X^{p-2} + ... + X + 1 \in \mathbb{Q}[X]$$

(see also Section 1.2). We claim that this polynomial is irreducible over $\mathbb{Q}$. Clearly Φ_p is not an Eisenstein polynomial, but with a suitable change of variable, we can obtain one. Let $Y = X - 1$. Then we find

$$\Phi_p(Y+1) = \frac{(Y+1)^p - 1}{Y} = Y^{p-1} + pY^{p-2} + \binom{p}{2}Y^{p-3} + ... + \binom{p}{p-2} + p.$$

It is a well-known fact that $p \mid \binom{p}{k}$ for $1 \leq k \leq p-1$[1], so all coefficients of $\Phi_p(Y+1)$ except the leading one lie in (p), and the constant coefficient p does not lie in (p^2). Therefore it is an Eisenstein polynomial and hence irreducible over $\mathbb{Q}$.

Example 2.2.16. It is not too hard to show that the p^2th cyclotomic polynomial for a prime p is given by

$$\Phi_{p^2} = \frac{X^{p^2} - 1}{X^p - 1}.$$

We want to show that this is irreducible as well. Replacing again $Y = X - 1$ we find that

$$(Y+1)^{p^2} - 1 = p^2Y + O(Y^2) \quad \text{and} \quad (Y+1)^p - 1 = pY + O(Y^2),$$

Where $O(Y^2)$ represents terms divisible by Y^2. It follows that $\Phi_{p^2}(Y+1) = p + O(Y)$, so the constant term is divisible by p but not by p^2. To examine the other coefficients consider the equation

$$(Y+1)^{p^2} - 1 = \Phi_{p^2}(Y+1) \cdot ((Y+1)^p - 1).$$

[1] We know that $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ is a non-negative integer and for k as stated, the numerator of the fraction is divisible by p but the denominator is not, so it must be divisible by p.

Since modulo a prime p we have $(a+b)^p \equiv a^p + b^p \pmod{p}$, we can write

$$(Y+1)^{p^2} - 1 = ((Y+1)^p)^p - 1 \equiv (Y^p+1)^p - 1 \equiv Y^{p^2} \pmod{p},$$

while

$$\Phi_{p^2}(Y+1) \cdot ((Y+1)^p - 1) \equiv \Phi_{p^2}(Y+1) \cdot Y^p \pmod{p},$$

so that $\Phi_{p^2}(Y+1) \equiv Y^{p^2-p} \pmod{p}$. This is equivalent to saying that all the coefficients of $\Phi_{p^2}(Y+1)$ except for the leading one are divisible by p, wherefore it is an Eisenstein polynomial and hence irreducible over $\mathbb{Q}$.

Chapter 3

Field Theory

After we have established several important facts about rings in the previous chapters, we turn our attention to the main object of study in this course, fields and their extensions.

3.1 First concepts

3.1.1 Extensions and their degrees

Topics

- Field extensions
- Degree theorem
- Simple extensions

Definition 3.1.1. Let K and E be fields. We call E/K a *field extension* if $K \subseteq E$ and addition and multiplication in K are the restriction of addition and multiplication in E to K. We then call K a *subfield* of E and E an *extension field* of K.

Some fairly straightforward examples of this are of course the field extensions $\mathbb{R}/\mathbb{Q}$ or $\mathbb{C}/\mathbb{R}$.

Remark 3.1.2. *Let E/K be a field extension. Then in particular E is a (commutative) K-algebra, i.e. a K-vector space with a compatible multiplication.*

This motivates the following definition.

Definition 3.1.3. Let E/K be a field extension. Then we call $[E : K] := \dim_K E$ the *degree* of the extension. If this degree is finite, we say that E/K is a *finite extension.*

Example 3.1.4.

1. A very important example of finite extensions can be constructed as follows: Let K be a field and $f \in K[X]$ an irreducible polynomial. Since $K[X]$ is a Euclidean domain the ideal $(f) \trianglelefteq K[X]$ is maximal, wherefore $E := K[X]/(f)$ is a field and we can consider K a subfield of E via the embedding $K \hookrightarrow E, a \mapsto a\overline{1}$. Since $\dim_K E = \deg f =: n$ as a basis is given by $\{\overline{1}, \overline{X}, ..., \overline{X^{n-1}}\}$, we have consequently $[E : K] = n$, so that the extension is finite.

2. Let K be field and consider the field $K(X) := \operatorname{Frac} K[X]$, the field of *rational functions* over K. This is an extension field of K, as K can again be embedded as constants, but it is not finite, since the set of monomials $\{X^n : n \in \mathbb{Z}\}$ is infinite and linearly independent[1], so $\dim_K K(X) = \infty$.

Our first result on field extensions is the following, sometimes called the degree theorem.

Theorem 3.1.5. *(Degree theorem) Let K_2/K_1 and K_3/K_2 be field extensions (so also K_3/K_1 is a field extension). Then we have $[K_3 : K_1] = [K_3 : K_2] \cdot [K_2 : K_1]$, with the understanding that if at least one of the extensions K_2/K_1 and K_3/K_2 is infinite, then so is K_3/K_1.*

Proof. If either K_3/K_2 or K_2/K_1 are infinite extension, then K_3 contains an infinite set of elements which is linearly independent over K_1, so K_3/K_1 is infinite as well.

Now let $[K_2 : K_1] =: m$ and $[K_3 : K_2] =: n$ be finite. Then there exists a K_2-basis of K_3 (as a K_2-vector space) $\{b_1, ..., b_n\}$ and a K_1 basis $\{c_1, ..., c_m\}$ of K_2.

Claim: The set $B = \{b_1c_1, b_1c_2, ..., b_nc_m\}$ is a K_1-basis of K_3.

We first show that B is a generating of K_3 as a K_1-vector space. For this let $\alpha \in K_3$. There exist unique $\alpha_1, ..., \alpha_n \in K_2$ such that $\alpha = \sum_{i=1}^n \alpha_i b_i$. Now for each α_i there exist unique $\beta_{i,1}, ..., \beta_{i,m} \in K_1$ such that $\alpha_i = \sum_{j=1}^m \beta_{i,j} c_j$, so that we can write $\alpha = \sum_{i=1}^n \sum_{j=1}^m \beta_{i,j} b_i c_j$ as a K_1-linear combination of elements of B.

It remains to show that B is linearly independent over K_1. Suppose that we have $\beta_{i,j} \in K_1$ such that $\sum_{i=1}^n \sum_{j=1}^m \beta_{i,j} b_i c_j = 0$. This means that

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{m} \beta_{i,j} c_j \right) b_i = 0,$$

but since the set $\{b_1, ..., b_n\}$ is linearly independent over K_2, we must have $\sum_{j=1}^m \beta_{i,j} c_j = 0$ for all i. But since the set $\{c_1, ..., c_m\}$ is linearly independent over K_1, this can

[1] It is however NOT a basis of $K(X)$, since for example the rational function $\frac{1}{X+1}$ is not a linear combination of these monomials

only be if all $\beta_{i,j}$ are 0. It follows therefore that B is linearly independent and thus a K_1-basis of K_3. Since $\#B = mn$, the theorem follows.

q.e.d.

In order to work with explicit examples of field extensions, we often represent them using generators.

Definition 3.1.6. Let E/K be a field extension.

1. For $a_1, ..., a_n \in E$ we denote by $K(a_1, ..., a_n)$ the smallest subfield of E containing K and $a_1, ..., a_n$. Similarly we denote by $K[a_1, ..., a_n]$ the smallest subring of E (not necessarily a field) containing K and $a_1, ..., a_n$.

2. E/K is called a *simple* extension if there exists some $a \in E$ such that $E = K(a)$. Such an a is called a *primitive element* of E/K.

We shall see later that all "reasonably nice" field extensions are indeed simple (Theorem 3.5.6).

3.1.2 Algebraic extensions

Topics

- Algebraic extensions
- Minimal polynomial

Proposition 3.1.7. *Let $E = K(a)/K$ be a simple extension. The we either have $E = K[a]$ or $E \cong K(X)$. In the former case we say that a is* algebraic *over K, in the latter a is said to be* transcendental *over K.*

Proof. Consider the following homomorphism of K-algebras, $\varphi : K[X] \to K(a)$ defined via $X \mapsto a$ (i.e. we plug a into a polynomial. Then the image $\operatorname{Im}\varphi$ of φ is a subring of $K(a)$ and hence an integral domain by Proposition 2.1.5. By the homomorphy theorem we have $\operatorname{Im}\varphi \cong K[X]/\operatorname{Ker}\varphi$, so that $\operatorname{Ker}\varphi$ must be a prime ideal in $K[X]$ (see Proposition 2.1.11). Since $K[X]$ is a principal ideal domain, there are two possibilities.

1. $\operatorname{Ker}\varphi \neq \{0\}$: In this case there is an irreducible polynomial $m \in K[X]$ such that $\operatorname{Ker}\varphi = (m)$. In this case (m) is a maximal ideal and hence, again by Proposition 2.1.11, $\operatorname{Im}\varphi = K[a]$ is a field, wherefore $K[a] = K(a) = E$.

2. $\operatorname{Ker}\varphi = \{0\}$: In this case φ is injective and $\operatorname{Im}\varphi = K[a] \cong K[X]$, which is not a field. But by definition $E = K(a)$ is the minimal field containing K and a, wherefore $E = \operatorname{Frac} K[a] \cong K(X)$.

q.e.d.

Definition 3.1.8. Let E/K be a field extension and $\alpha \in E$ be algebraic over K. The unique monic polynomial $\mu_{\alpha,K} \in K[X]$ which generates $\operatorname{Ker}\varphi$ with φ as in the proof of Proposition 3.1.7, i.e. which satisfies $\mu_{\alpha,K}(\alpha) = 0$, is called the *minimal polynomial* of a over K. The degree of $\mu_{\alpha,K}$ is called the *degree* of α over K.

Remark 3.1.9. *The more usual definition of algebraicity is that $a \in E$ is algebraic over K if and only if there exists a polyomial $m \in K[X]$ such that $m(a) = 0$. This is clearly equivalent to the fact that $K(a) = K[a]$ by the proof of Proposition 3.1.7.*

Here and throughout most of this course, we will deal with algebraic extensions. In Section 3.7 we briefly discuss some elementary properties of transcendental extensions.

Example 3.1.10.

1. Consider the field extension $\mathbb{Q}(a)/\mathbb{Q}$ for $a = \sqrt{2} + \sqrt{3}$. We can compute the minimal polynomial of a by successively taking powers of a and looking for linear dependencies. We have

$$a^2 = 5 + 2\sqrt{6}, \quad a^3 = 11\sqrt{2} + 9\sqrt{3}, \quad a^4 = 49 + 20\sqrt{6} = 10a^2 - 1.$$

   We have thus found the minimal polynomial of a to be $\mu_{a,\mathbb{Q}} = X^4 - 10X^2 + 1$.

2. The minimal polynomial does depend on the field under consideration. Consider the field extension $\mathbb{C}/\mathbb{Q}$ and $a = \zeta_8 = e^{2\pi\,\mathrm{i}/8} = \frac{1+\mathrm{i}}{\sqrt{2}}$ a primitive 8th root of unity. Then $\mu_{\zeta_8,\mathbb{Q}}(X) = \Phi_8(X) = X^4 + 1$. But if we consider the extension $\mathbb{R}/\mathbb{C}$, then of course ζ_8 is still algebraic over $\mathbb{R}$, but one finds that $\mu_{\zeta_8,\mathbb{R}}(X) = X^2 - \sqrt{2}X + 1$.

In what follows we mainly restrict our attention to so-called algebraic extensions.

Definition 3.1.11. Let E/K be a field extension. If each $a \in E$ is algebraic over K, then we call E/K an *algebraic extension*.

Theorem 3.1.12.

1. *Every finite field extension E/K is algebraic.*
2. *If we have $E = K(a_1, ..., a_n)$ where each a_i is algebraic over K, then E/K is a finite extension.*

Proof.

1. Let $a \in E$. Since E/K is finite, the sequence of powers of a, $(1, a, a^2, ...)$ must become linearly dependent at some point, let's say n, so that there exist $\alpha_1, ..., \alpha_n \in K$ such that $\sum_{i=1}^{n} \alpha_i a^i = 0$. Hence the polynomial $m(X) = \sum_{i=1}^{n} \alpha_i X^i$ satisfies $m(a) = 0$, so that a is algebraic over K by Remark 3.1.9.

2. First consider the case $n = 1$. Then the set $\{a_1, ..., a_1^{m-1}\}$, where $m = \deg \mu_{a_1,K} = [K(a_1) : K]$, forms a K-basis of $K(a_1) = K[a_1]$. Now suppose the claim is proven for some n and consider the extension $E = K(a_1, ..., a_{n+1})/K$. By induction, the extension $K(a_1, ..., a_n)/K$ is finite and the extension $E/K(a_1, ..., a_n)$ is finite by the same argument as before, so E/K is finite by the Degree Theorem (Theorem 3.1.5).

q.e.d.

Slightly more generally we have the following result.

Proposition 3.1.13. *Algebraicity is transitive, i.e. if L/K is an algebraic extension and E/L is an algebraic extension, then E/K is an algebraic extension as well.*

Proof. Let $a \in E$. We want to show that a is algebraic over K. By assumption a is algebraic over L, so there exist $\alpha_0, ..., \alpha_{n-1} \in L$ such that

$$a^n + \sum_{i=0}^{n-1} \alpha_i a^i = 0.$$

Since L/K is algebraic, we known from Theorem 3.1.12 that the extension $K(\alpha_0, ..., \alpha_{n-1})/K$ is finite. Since a is algebraic over $L_0 = K(\alpha_0, ..., \alpha_{n-1})$, the extension

$$K(\alpha_0, ..., \alpha_{n-1}, a) = L_0(a)/L_0$$

is finite, thus by the degree theorem the extension $K(\alpha_0, ..., \alpha_{n-1}, a)/K$ is finite and hence algebraic. In particular a is algebraic over K which is what we wanted to show.

q.e.d.

Remark 3.1.14. *A very slightly different way to characterise algebraic extensions E/K would be to say that each $a \in E$ lies in a finite extension of K. This does however not imply that all algebraic extensions are finite. For example the field $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, ...)$, where all square-roots of prime numbers are adjoint to $\mathbb{Q}$ is clearly algebraic over $\mathbb{Q}$ since every of its generators is, but it is not finite.*

3.2 Splitting fields

Topics

- Rupture field of a polynomial
- Field homomorphisms
- Splitting field of a polynomial
- Algebraically closed fields and algebraic closure

In this section we discuss a fairly universal way to generate field extensions. The first one was already alluded to in Example 3.1.4.

Definition 3.2.1. Let K be field and $f \in K[X]$ a monic polynomial of degree n.

1. An extension field E of K is called a *rupture field* of f over K if there exists $a \in E$ such that $f(a) = 0$.
2. The minimal extension field E of K such that f decomposes into linear factors, i.e. there exist $a_1, ..., a_n \in E$ such that $f = \prod_{i=1}^{n}(X - a_i)$ in $E[X]$ and there is no proper subfield F of E such that this factorisation is possible in $F[X]$, is called the *splitting field* of f over K.

Example 3.2.2.

1. The field $\mathbb{C}$ of complex numbers is a rupture field of the polynomial $f = X^2 + 1 \in \mathbb{R}[X]$. In fact it is also the splitting field of this polynomial since $f = (X - i)(X + i) \in \mathbb{C}[X]$ and $[\mathbb{C} : \mathbb{R}] = 2$, so there cannot be any proper subfield of $\mathbb{C}$ containing $\mathbb{R}$ where f factors completely.
2. Consider the case $K = \mathbb{Q}$ and the polynomial $f = X^4 - 2$. Then the field $L = \mathbb{Q}(\sqrt[4]{2})$ is a rupture field of f, but not its splitting field, because $i\sqrt[4]{2}$ is also a root of f, but not contained in L. This means in particular that $\mathbb{Q}(i\sqrt[4]{2})$ is another rupture field of f. Indeed the field $E = \mathbb{Q}(\sqrt[4]{2}, i)$ is the splitting field of f over $\mathbb{Q}$: Since
$$X^4 - 2 = (X - \sqrt[4]{2})(X + \sqrt[4]{2})(X - i\sqrt[4]{2})(X + i\sqrt[4]{2}) \in E[X],$$
f splits completely in E. Any subfield of E where f splits completely must therefore contain in particular $\sqrt[4]{2}$ and $i\sqrt[4]{2}$ and thus also $i = i\sqrt[4]{2}/\sqrt[4]{2}$. But E is by definition the smallest field containing $\sqrt[4]{2}$ and i, so E is indeed the splitting field of f.

The formulation "*the* splitting field" in the above example already indicates that the splitting field of a polynomial is essentially unique. To make this precise and in particular prove this we need the following concepts.

Definition 3.2.3. 1. Let K, L be fields. Then a ring homomorphism $\varphi : K \to L$ is called a *field homomorphism*. A ring isomorphism $\varphi : K \to K$ is called a *field automorphism*.

2. Let E/K and L/K be field extensions. A homomorphism of K-algebras $\varphi : E \to L$ is called a *field homomorphism over K* or *K-homomorphism*. If there exists a K-algebra isomorphism $E \to L$, we write $E \cong_K L$.

3. For a field extension E/K we call

$$\mathrm{Aut}(E) := \{\varphi : E \to E \ : \ \varphi \text{ is a field automorphism}\}$$

the *automorphism group* of E and

$$\mathrm{Aut}_K(E) := \{\varphi : E \to E \ : \ \varphi \text{ is a field automorphism over } K\}$$

the group of K-automorphisms of E over K.

Example 3.2.4. It is important to distinguish between automorphisms and K-automorphisms of a field: Let for instance $K = \mathbb{R}$ and $E = \mathbb{C}$. Then complex conjugation $\overline{a + ib} := a - ib$ is an automorphism of $\mathbb{C}$, and indeed also an $\mathbb{R}$-automorphism of $\mathbb{C}$.

However if we choose $K = \mathbb{Q}(i)$, the complex conjugation is not a K-automorphism of $\mathbb{C}$, since a K-algebra homomorphism φ of $\mathbb{C}$ has to satisfy $\varphi(\alpha a) = \alpha\varphi(a)$ for all $a \in \mathbb{C}$ and $\alpha \in \mathbb{Q}(i)$. This is clearly not satisfied (pick for instance $a = 1$ and $\alpha = i$).

Remark 3.2.5. *1. Note that any field homomorphism $\varphi : K \to L$ is automatically injective (Exercise).*

2. For any field homomorphism $\varphi : K \to L$ there is a unique ring homomorphism $\tilde{\varphi} : K[X] \to L[X]$ which extends φ and satisfies $\tilde{\varphi}(X) = X$. It is given by $\tilde{\varphi}(\sum_{i=0}^{n} a_i X^i) = \sum_{i=0}^{n} \varphi(a_i) X^i$.

For later application we record the following.

Lemma 3.2.6. *Let K be a field and $\varphi : K \to K$ a field homomorphism (such a homomorphism is also called an* endomorphism*). Then the set*

$$F := \mathrm{Fix}(\varphi) := \{\alpha \in K \ : \ \varphi(\alpha) = \alpha\}$$

is a subfield of K, called the fixed subfield *of φ.*

Proof. Since $\varphi(0) = 0$ and $\varphi(1) = 1$ we have $0, 1 \in F$ and for $\alpha, \beta \in F$ we have

$$\alpha + \beta = \varphi(\alpha) + \varphi(\beta) = \varphi(\alpha + \beta)$$

and similarly $\alpha \cdot \beta = \varphi(\alpha \cdot \beta)$, so that $\alpha + \beta, \alpha \cdot \beta \in F$. If $\alpha \neq 0$ we have $\varphi(\alpha^{-1}) = \varphi(\alpha)^{-1}$, so that for $\alpha \in F$, we also have $\alpha^{-1} \in F$.

q.e.d.

We now return to the discussion of rupture fields and splitting fields.

Theorem 3.2.7. *Let K be a field and $f = \sum_{i=0}^{n} a_i X^i$ a polynomial of degree $n \geq 1$.*

1. *There exists a rupture field of f over K.*
2. *Any minimal rupture field L of f has the form $L = K[\alpha]$, where $f(\alpha) = 0$.*
3. *Assume $f \in K[X]$ is irreducible and $\varphi : K \to L$ a field isomorphism. Set $g = \tilde{\varphi}(f) = \sum_{i=0}^{n} \varphi(a_i) X^i \in L[X]$ with $\tilde{\varphi}$ as in Remark 3.2.5. If $E = K[\alpha]$ is a minimal rupture field of f and $F = L[\beta]$ a minimal rupture field of g, then the map*
$$\varphi^* : E \to F, \sum_{i=0}^{n} c_i \alpha^i \mapsto \sum_{i=0}^{n} \varphi(c_i) \beta^i$$
defines an isomorphism of fields which extends φ.
4. *For an irreducible polynomial, any two minimal rupture fields are isomorphic.*

Proof.

1. Let $p \in K[X]$ be an irreducible polynomial dividing f, write $f = p \cdot g$ for some $g \in K[X]$. Then $E = K[X]/(p)$ is a field. The class $\overline{X} = X + (p)$ satisfies $p(\overline{X}) = 0$ in E, and therefore we have $f(\overline{X}) = p(\overline{X}) \cdot g(\overline{X}) = 0$, so E is a rupture field of K.

2. Let L be a minimal rupture field of f over K. Then L contains some α such that $f(\alpha) = 0$. By definition $K(\alpha) \subseteq L$ is the minimal field containing K and α, so we must have $L = K(\alpha)$ and since α is algebraic over K, we have $K(\alpha) = K[\alpha]$.

3. It is clear from the definition that ϕ^* is a field homomorphism. Therefore it is automatically injective and since φ is an isomorphism we know that for each $b_i \in L$ there is a unique $c_i \in K$ such that $\varphi(c_i) = b_i$, so that $\varphi^*(\sum_{i=0}^{n} c_i \alpha^i = \sum_{i=0}^{n} b_i \beta^i$, showing that φ^* is also surjective, so it is indeed an isomorphism as claimed.

4. Follows directly from 3.

q.e.d.

Regarding splitting fields we can say the following.

Theorem 3.2.8. *Let K be a field and $f \in K[X]$ be a polynomial of degree $n \geq 1$.*

1. *There exists an extension field E of K, such that f decomposes into linear factors in $E[X]$.*

2. *Any two splitting fields of f are isomorphic over K.*

Proof.

1. This follows by iterating Theorem 3.2.7 1. Let K_1 be a rupture field of f, so that there is some $\alpha_1 \in K_1$ such that $f(\alpha_1) = 0$. This means that we can write $f = (X - \alpha_1) \cdot f_1$ for some $f_1 \in K_1[X]$. Now repeat the process to find a rupture field K_2 of f_1 and so on. Since the degree of the polynomial in question decreases in every step, this must reach an end at some point, and the resulting field is one where f splits into linear factors.

2. Let E be a splitting field of f over K and set $[E : K] = m$. We prove the following claim, from which the result follows directly, by induction on m.

   Claim: If $\varphi : K \to K'$ is a field isomorphism and E' is a splitting field of $\tilde{\varphi}(f) \in K'[X]$, then φ can be extended to a field isomorphism $\varphi^\bullet : E \to E'$.

   For $m = 1$, f already decomposes into linear factors over K, so $E = K \cong K' = E'$.

   So let $m > 1$ and assume the claim is true for all $\ell < m$. Then there is an irreducible factor p of f of degree $d \geq 2$. Set $p' = \tilde{\varphi}(p)$. Let $\alpha \in E$ and $\alpha' \in E'$ be such that $p(\alpha) = 0$ and $p'(\alpha') = 0$. Then $L := K[\alpha]$ (resp. $L' := K'[\alpha']$) are minimal rupture fields of p (resp. p'), wherefore φ extends to an isomorphism $\varphi^* : L \to L'$ by Theorem 3.2.7 3. But we have

   $$[E : L] = \frac{[E : K]}{[L : K]} = \frac{[E : K]}{d} \lneq m$$

   by the degree theorem and E is a splitting field of $f \in L[X]$ (and E' is a splitting field of $\tilde{\varphi}^*(f) \in L'[X]$). By the induction hypothesis we can therefore extend φ^* to a field isomorphism $\varphi^\bullet : E \to E'$ which extends φ^*, wherefore it also extends φ.

q.e.d.

One question one may have now is whether for every given field K there is an extension field such that *every* polynomial in $K[X]$ splits into linear factors. For example the complex numbers are an extension field of $\mathbb{Q}$ and every polynomial over $\mathbb{Q}$ splits into linear factors over $\mathbb{C}$ by the Fundamental Theorem of Algebra (for a proof see Appendix A).

Definition 3.2.9. A field K is called *algebraically closed* if every polynomial $f \in K[X]$ of degree ≥ 1 has a root in K.

An example of an algebraically closed field is therefore the complex numbers.

Of course not every field can be embedded into the complex numbers, so one may wonder whether in general there is an algebraically closed field containing a given field K.

Definition 3.2.10. Let K be any field. Then extension field E of K is called an *algebraic closure* of K if E is algebraically closed and the extension E/K is algebraic.

Even though the complex numbers are an algebraically closed field containing $\mathbb{Q}$, they are not an algebraic closure of $\mathbb{Q}$ because $\mathbb{C}$ contains so-called transcendental numbers (see Appendix B). From here on out, we will assume that every field has an algebraic closure (Theorem 6.2.4) which is unique up to isomorphism (Corollary 6.2.6). For a field K, we denote its algebraic closure by $\overline{K}$. For later purposes, we record the following lemma about the algebraic closure at this point.

Lemma 3.2.11. *Let $K \subseteq L \subseteq E \subseteq \overline{K}$ be algebraic extensions. Suppose we have a ring homomorphism $\varphi : L \to \overline{K}$ such that $\varphi(a) = a$ for all $a \in K$. Then there exists a ring homomorphism $\psi : E \to \overline{K}$ such that $\psi(\alpha) = \varphi(\alpha)$ for all $\alpha \in L$, i.e. ψ is an entension of φ from L to E.*

The proofs of all these facts require some preparation, in particular the famous *Zorn's Lemma* 6.1.9. We postpone these proofs until Chapter 6.

3.3 Compass and straightedge constructions

In this section we discuss one of the central results in this course, the possibility and impossibility of certain geometric constructions with straightedge (i.e. and unmarked ruler) and compass. As mentioned in the introduction, many of the problems discussed here go back to the ancient Greeks and those that weren't solved by the Greeks themselves, such as the costruction of the regular heptadecagon (17-gon), which is possible, or proving the impossibility of doubling the cube, had remained open for over 2000 years.

One might say now that such geometric questions are in themselves not very interesting, at least not in modern mathematics, but the moral here should be that mathematics has some unexpected connections among its subfields. As we shall see now, the answer for a (seemingly) geometric question sometimes comes from the theory of field extensions.

3.3.1 From geometry to algebra

Topics

- Constructibility
- Basic constructions with compass and straightedge
- Constructible numbers (examples)

First we need to translate geometry into algebra. For this we interpret the Euclidean plane $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$.

Definition 3.3.1. Let $\mathcal{S} \subset \mathbb{C}$ be a finite set of points in the (Euclidean) plane. For any two distinct points $P, Q \in \mathcal{S}$ we can draw a unique straight line through P and Q, denoted PQ. For any three points M, P, Q, which are not all the same, we can draw a unique circle with centre M and radius $|PQ|$, the distance of the points P and Q.

1. A point $P \in \mathbb{C}$ is *1-step constructible* from $\mathcal{S}$ if P is an intersection point of either two lines, a line and a circle, or two circles obtained from points in $\mathcal{S}$.
2. A point $P \in \mathbb{C}$ is called *constructible* from $\mathcal{S}$ if there exist finitely many points $P_1, ..., P_n = P \in \mathbb{C}$ such that P_{i+1} is 1-step constructible from $\mathcal{S} \cup \{P_1, ..., P_i\}$ for all $0 \leq i \leq n-1$.

Example 3.3.2. Here are some standard constructions which should be familiar. Suppose we have our set of given points $\mathcal{S} = \{P, Q, R\}$.

1. Perpendicular bisector: This is the line perpendicular to the line PQ, which intersects at its exact midpoint M, i.e. the lengths $|PM|$ and $|MQ|$ are equal. One draws two circles of the same radius, which needs to be larger that the distance $|PM|$ (so $|PQ|$ would do), one around P and one around Q and connects the intersection points (see Figure 3.1). So in particular the midpoint of any two given points is constructible.

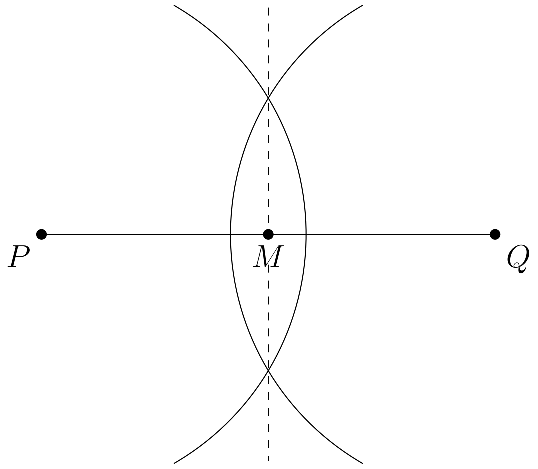


Figure 3.1: Perpendicular bisector

2. Using the perpendicular bisector, we see that we can also construct the line perpendicular to a given line ℓ through a given point P, either on the line or not: Draw a circle around P of radius large enough to intersect the line twice and construct the perpendicular bisector of the two intersection points. The resulting line is perpendicular to the original one and passes through P. By constructing a second perpendicular bisector through P, we can also construct the line parallel to ℓ passing through P.

3. Angle bisector: Given three points $\{P, Q, R\}$ (not all in a straight line), the lines connecting two of them form an angle α, say at P. We can construct a line which exactly cuts this angle in half as follows: Draw a circle around P with radius at most the smaller of the two distances $|PQ|$ and $|PR|$ (or extend the lines PQ and PR sufficiently so that the circle intersects both at points S_1 and S_2. Draw the line S_1S_2 and construct a perpendicular bisector to it. This line goes through P and bisects the angle α exactly (see Figure 3.2).

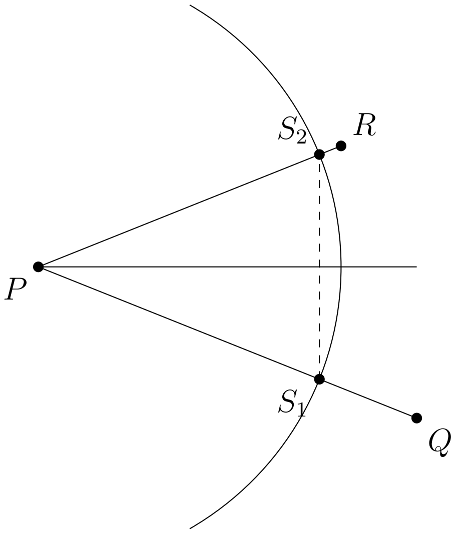


Figure 3.2: Angular bisector

Since we may identify points with numbers, it makes sense to talk about *constructible numbers*, which leads to the question how constructibility behaves with regard to arithmetic.

Theorem 3.3.3. *Suppose we are given a set $\mathcal{S}$ with $\{0,1\} \subseteq \mathcal{S}$. Then the following are true.*

1. *Every rational number is constructible from $\mathcal{S}$.*

2. *The set $K_{\mathcal{S}} := \{a \in \mathbb{C} \, : \, a$ is constructible from $\mathcal{S}\}$ is a subfield of $\mathbb{C}$.*

3. *For $a \in K_{\mathcal{S}}$, both complex roots of $X^2 - a \in K_{\mathcal{S}}[X]$ are in $K_{\mathcal{S}}$, so square-roots of elements in $K_{\mathcal{S}}$ lie in $K_{\mathcal{S}}$.*

4. *We have $a \in K_{\mathcal{S}}$ if and only if $\mathrm{Re}(a) \in K_{\mathcal{S}}$ and $\mathrm{Im}(a) \in K_{\mathcal{S}}$. Also if $a \in K_{\mathcal{S}}$, we have $|a|^2$ and $\overline{a} \in K_{\mathcal{S}}$.*

Proof.

1. We extend the line through 0 and 1, which we identify with the real line, in both directions indefinitely. Then we draw a circle around 1 with radius 1, which yields two intersections, at 0 and 2, so 2 is constructible from $\mathcal{S}$. Continuing this procedure with a circle around 2, 3,... and similarly in the other direction around $0, -1, -2, ...$, we see that all integers are constructible from $\{0, 1\}$.

For any positive real constructible number a, we show that $1/a$ is also constructible (see also Figure 3.3). Construct the line from 0 to a and construct a line perpendicular to $0a$ through a. Then draw a circle of radius 1 around a and obtain an intersection point b which we then connect to 0. We then find 1 on the (extended) line through 0 and a. The perpendicular line to $0a$ through 1 intersects the line $0b$ in a unique point S and since the two triangles $0ab$ and $01S$ are clearly similar, we must have

$$\frac{1}{a} = \frac{|ab|}{|0a|} = \frac{|1S|}{|01|} = |1S|,$$

so we can draw a circle around 0 with radius $1/a$, and its intersection with the real line is the desired reciprocal.

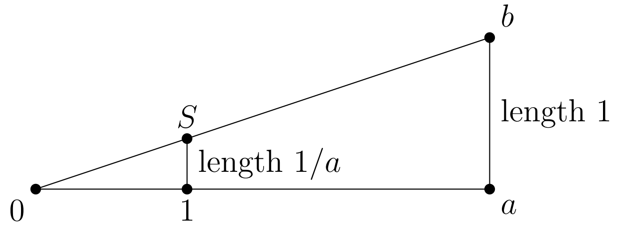


Figure 3.3: Construction of reciprocals

Now if we want to construct a fraction a/b with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$, we can construct $1/b$ as described above and then construct $a \cdot 1/b$ by drawing $|a|$ circles of radius $1/b$ around $0, \pm 1, ...$, depending on the sign of a. Therefore we can construct a/b and thus every rational number.

2. We first show that the sum of two constructible numbers is constructible. Suppose that the three points $0, a, b$ are not all in a straight line. Then we can construct the line parallel to $0a$ through b and the line parallel to $0b$ through a. These two lines intersect in a unique point, which we recognise as $a + b$ from the standard geometric interpretation of the addition of two complex numbers (see Figure 3.4).

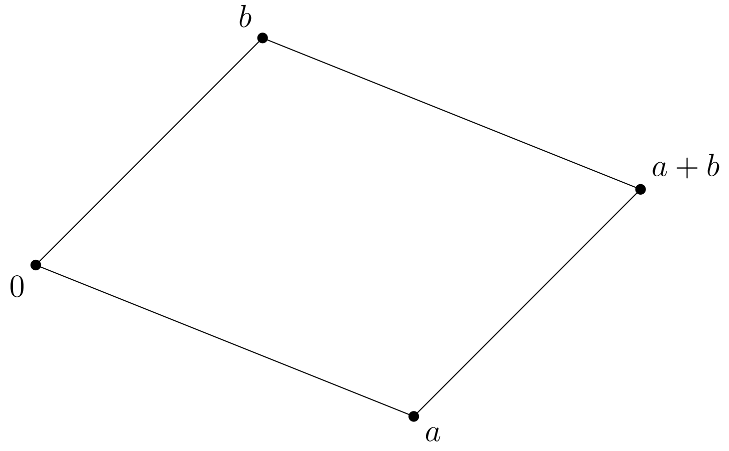


Figure 3.4: Addition of constructible numbers

If $0, a, b$ lie on a straight line, then the circle around a of radius $|0b|$ intersects said line in exactly two points. The point $a+b$ is the the intersection point that farther away from 0 if both a, b lie on the same side of 0, and it is the one closer to 0 if 0 lies between a and b.

Constructing additive inverses is fairly straightforward (exercise).

Next we consider products of constructible numbers. Again we restrict to positive real constructible numbers and leave the general case as an exercise. Draw the line connecting 0 and 1 and extend it to the right. Identify the number b on this line. Next draw a circle of radius $|0a|$ around 0 and connect 0 to any point on the circle, except the intersection points with the (extended) line 01. Call this chosen point A. Draw a line from 1 to A and construct a line parallel to $1A$ through b. This line has a unique intersection point B with the (extended) line $0A$. Since the triangles with vertices $0, 1, A$ and $0, b, B$ are similar, we have

$$\frac{|0b|}{|01|} = \frac{|0B|}{|0A|}.$$

The left-hand side equals b and since $|0A| = a$, it follows that $|0B| = a \cdot b$, as desired (see Figure 3.5).

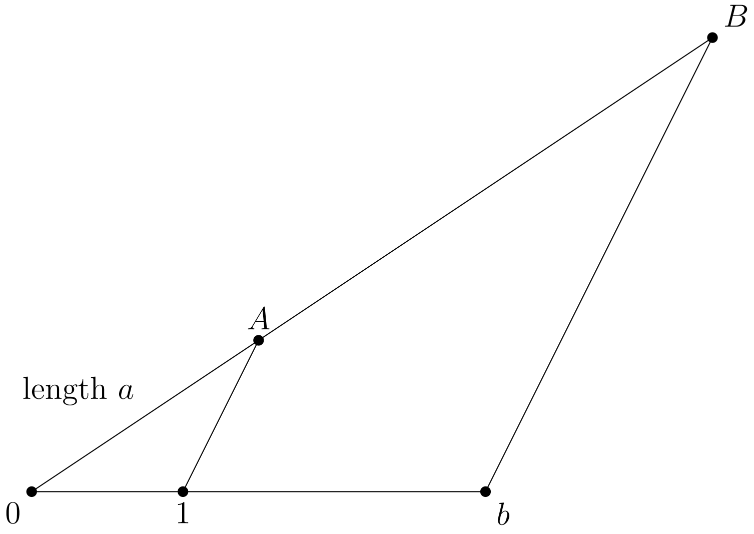


Figure 3.5: Multiplication of constructible numbers

We have already shown above how to construct inverses of positive real numbers and again we leave the general case as an exercise.

3. Again we restrict to the case of positive real constructible numbers and leave the general case as an exercise. Given such a constructible number a, we construct the midpoint M of the line $0(a+1)$ and draw a circle of radius $(a+1)/2$ around M. Next construct the line perpendicular to $0a$ passing through a. This line intersects the circle in a point P. By the Theorem of Thales, the triangle $0, c+1, P$ is a right-angle triangle with right angle at P. Now by the Right Triangle Altitude Theorem, the distance $h = |cP|$ satisfies $h^2 = c \cdot 1$, so $h = \sqrt{c}$ and we are finished (see Figure 3.6).

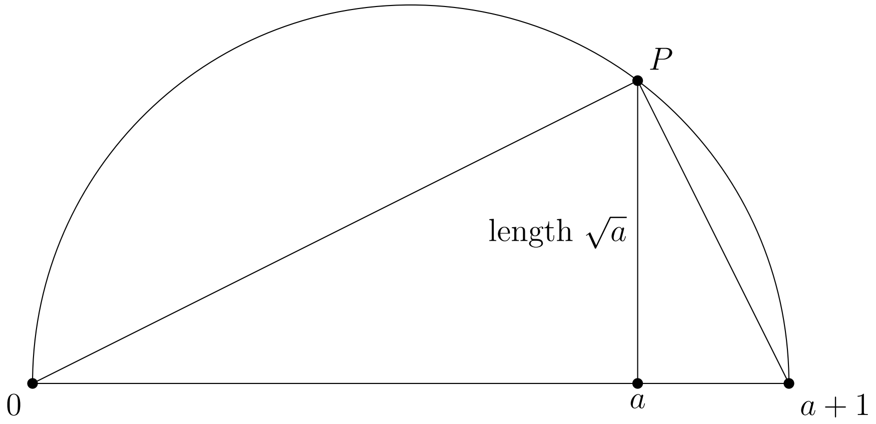


Figure 3.6: Construction of square-roots

4. Similar to the real axis, we can construct the imaginary axis as an orthogonal line to the real one passing through 0. In particular the imaginary unit i is constructible. Therefore if a is constructible, we can construct an orthogonal line to the real line through a to construct $\mathrm{Re}(a)$, and similarly for $\mathrm{Im}(a)$. If both $a, b \in \mathbb{R}$ are constructible, then so is $a + ib$ since constructible numbers form a field. The claim about the (squared) absolute value and the complex conjugate directly follows because we can construct real and imaginary parts.

q.e.d.

3.3.2 Classification of constructible numbers

Topics

- Classification of constructible numbers
- Impossibility of doubling the cube, trisecting an angle, and squaring the circle

In order to classify constructible numbers, we require the following lemma.

Lemma 3.3.4. *Let $L \leq \mathbb{C}$ be a subfield of $\mathbb{C}$ such that* $\mathrm{i} \in L$ *and for each $a \in L$ we also have $\overline{a} \in L$ (denotes the complex conjugate). If $z \in \mathbb{C}$ is 1-step constructible form L, then there exists $w \in \mathbb{C}$ such that $w^2 \in L$ and $z \in L[w]$.*

Proof. We first note that the fact that $\mathrm{i} \in L$ and L contains complex conjugates implies that for each $a \in L$ we have $\mathrm{Re}(a) = \frac{1}{2}(a + \overline{a}) \in L$, $\mathrm{Im}(a) = \frac{1}{2\mathrm{i}}(a - \overline{a}) \in L$, and $|a|^2 = a\overline{a} \in L$.

Now let $z \in \mathbb{C}$ be 1-step constructible from L. There are three cases to distinuish:

1. z is the intersection of two straight lines through points in L:

   In this case there exist four points $\alpha_j = a_j + \mathrm{i}\, b_j\, \mathrm{i} \in L$ and $\beta_j = c_j + \mathrm{i}\, d_j \in L$, $j = 1, 2$ such that the lines through the α_j and the one through β_j intersect in z (in particular, the lines can't be parallel or identical). These lines can be defined through the equations

   $$\ell_1(s) = \alpha_1 + (\alpha_2 - \alpha_1)s \quad \text{and} \quad \ell_2(t) = \beta_1 + (\beta_2 - \beta_1)t,$$

   where s, t are real numbers. So we see that the real and imaginary part of z is the solution to an inhomogeneous linear system of equations with coefficients in L, so the solution, since it is clearly unique, is also defined over L, so we find that $z \in L$.

2. z is an intersection point of a straight line and a circle:

   Let $\alpha_1, \alpha_2 \in L$, so that z lies on the line through α_1 and α_2, i.e. there exists $t \in \mathbb{R}$ such that $z = \alpha_1 + (\alpha_2 - \alpha_1)t$. Furthermore there are points $M, \beta_1, \beta_2 \in L$, such that z lies on the circle of radius $r := |\beta_2 - \beta_1|$ around M, i.e we have $|z - M|^2 = r^2$. Note here that $r^2 \in L$. Therefore we find that

   $$\begin{aligned} r^2 &= \mathrm{Re}(z - M)^2 + \mathrm{Im}(z - M)^2 \\ &= (\mathrm{Re}(\alpha_2 - \alpha_1)t + \mathrm{Re}(\alpha_1 - M))^2 + (\mathrm{Im}(\alpha_2 - \alpha_1)t + \mathrm{Im}(\alpha_1 - M))^2, \end{aligned}$$

   Thus t satisfies a quadratic equation over L, i.e. $[L(z) : L] \leq 2$.

3. z is an intersection point of two circles:

   Let $z = x + \mathrm{i}\, y$, and suppose z lies on the two circles with centre $M_1 = a_1 + \mathrm{i}\, b_1, M_2 = a_2 + \mathrm{i}\, b_2 \in L$ with radii r_1 and r_2. Note again that $r_j^2 \in L$. Therefore x and y satisfy the equations

   $$(x - a_j)^2 + (y - b_j)^2 = r_j^2, \quad j = 1, 2.$$

   Taking the difference of the two equations, we obtain, after some rearranging, the equation

   $$(a_2 - a_1)x + (b_2 - b_1)y = \frac{1}{2}(r_1^2 - r_2^2 - a_1^2 + a_2^2 - b_1^2 + b_2^2),$$

   which is the equation of a line defined over L, so z is really the intersection point of a circle and a line and the claim follows from the previous case.

q.e.d.

With this we can show the key result in this section.

Theorem 3.3.5. *Let $\mathcal{S} \subset \mathbb{C}$ and assume $\{0,1\} \subseteq \mathcal{S}$. For $z \in \mathbb{C}$ the following are equivalent.*

(i) $z \in K_{\mathcal{S}}$.

(ii) There exists a (finite) tower of fields

$$\mathbb{Q}(\mathcal{S} \cup \overline{\mathcal{S}}) =: L_0 \leq L_1 \leq \ldots \leq L_n \leq \mathbb{C}$$

such that $z \in L_n$ and $[L_j : L_{j-1}] = 2$.

Proof. If there is such a tower of fields as in (ii), then each element $a \in L_j$, $j \geq 1$, is constructible from the elements in L_{j-1} because a must satisfy a quadratic equation over L_{j-1} whose solution only involves arithmetic operations and taking square-roots. Since all of these operations are constructible by Theorem 3.3.3, each such a is constructible from L_{j-1}, and hence also from L_0.

If on the other hand z is constructible from $\mathcal{S}$, then there must be a sequence $z_1, \ldots, z_k = z$ and each z_j is 1-step constructible from $\mathcal{S} \cup \{z_1, \ldots, z_{j-1}\}$. Since $\overline{\mathcal{S}}$ is constructible from $\mathcal{S}$, it follows directly from Lemma 3.3.4 that by adjoining z_j to $\mathbb{Q}(\mathcal{S} \cup \overline{\mathcal{S}}, z_1, \ldots, z_{j-1})$ we obtain an extension field of degree at most 2, thus showing the claim.

q.e.d.

We can now discuss a few examples of constructibility problems.

Example 3.3.6. One of the last open construction problems from Ancient Greece was the question whether it is possible to "double the cube", i.e. given a cube with some volume V, can we construct a cube of volume $2V$ using only compass and straightedge?

Let the side length of our given cube be a, so $V = a^3$. If the doubled cube is constructible, then it would have to be possible to construct its side length b from a. But b satisfies the equation $b^3 - 2a^3 = 0$. This polynomial is clearly irreducible over $\mathbb{Q}(a)$, so we have $[\mathbb{Q}(a,b) : \mathbb{Q}(a)] = 3$. But this means that there can't be a tower of fields of consecutive degree 2 over $\mathbb{Q}(a)$ which contains b, so by Theorem 3.3.5 b cannot be constructible.

Example 3.3.7. Another classical problem is whether or not it is possible to trisect a general given angle. Of course some angles can be trisected, for example a right angle is trisectable since it is fairly easy to construct a 30° angle, e.g. by constructing an angular bisector in an equilateral triangle, which is also easy to construct.

But it turns out that e.g. an angle of 30° cannot be trisected. It is clear from the usual interpretation of sine and cosine in a circle, an angle can be constructed if and only if its sine and cosine can be constructed. Now for any angle θ we have the relation

$$\sin(3\theta) = -4\sin^3\theta + 3\sin\theta,$$

which follows easily from the standard angle sum formulas for sine and cosine. Setting $x := \sin(10^\circ)$ we obtain from the special value $\sin(30^\circ) = \frac{1}{2}$ the equation

$$8x^3 - 6x + 1 = 0.$$

By the Rational Root Theorem 2.2.9, the polynomial $8X^3 - 6X + 1 \in \mathbb{Q}[X]$ can only possibly have rational roots at $\alpha \in \{\pm 1, \pm\frac{1}{2}, \pm\frac{1}{4}, \pm\frac{1}{8}\}$. It is easily checked that none of these values are roots of the polynomial, wherefore it is irreducible over $\mathbb{Q}$. But this means that $[\mathbb{Q}[x] : \mathbb{Q}] = 3$, so that x cannot be constructible over $\mathbb{Q} = \mathbb{Q}(\sin(30^\circ))$, again by Theorem 3.3.5.

For the sake of completeness we also mention a third classical constructibility problem, probably the most famous of them all: Squaring the circle, so to construct from a given circle a square with the same area. This has almost become a proverbial expression for an impossible task. It is not hard to see that this problem boils down to the question whether the number $\pi = 3.1415926...$ is constructible. It was first shown by Ferdinand von Lindemann in 1882 that π is indeed transcendental over $\mathbb{Q}$, so in particular π does not lie in any finite extension of $\mathbb{Q}$, let alone in one described in Theorem 3.3.5. Lindemann's proof for the transcendence of π however is too long and involved to include it here.

Remark 3.3.8. *In Ancient Greece, the only tools permitted for geometric constructions were compass and straightedge. In modern times people have investigated how other tools might serve to do exact geometric constructions. For example one can show that using a ruler and* origami, *i.e. one is allowed to fold the plane along a straight line, and it turns out that using this method one can solve general (!) cubic equations geometrically, like compass and straightedge can solve general quadratic equations geometrically.*

3.3.3 Constructibility of regular polygons

Topics

- Fermat primes
- Gauß's theorem on regular polygons

We now discuss one of the results for which Carl Friedrich Gauß became one of the most famed mathematicians of the early 19th century. He classified exactly when the regular n-gon is constructible with compass and straightedge. For this we need the following notion.

Definition 3.3.9. For $n \in \mathbb{N}_0$ we call $F_n := 2^{2^n} + 1$ the nth *Fermat number*. If F_n is prime, we call it a *Fermat prime*.

The first few Fermat numbers are given by

$$F_0 = 3,\ F_1 = 5,\ F_2 = 17,\ F_3 = 257,\ F_4 = 65537,$$

which are all prime. It had been conjectured that F_n might be prime for all n, but Euler was able to show that $F_5 = 641 \cdot 6700417$ is composite. In fact, there is no $n > 4$ known such that F_n is prime, but it hasn't been ruled out that such an n exists.

A further important object which may be familiar from elementary number theory (and Section 1.2) is the following.

Definition 3.3.10. For $n \in \mathbb{N}$ we denote by

$$\varphi(n) := (\mathbb{Z}/n\mathbb{Z})^* = \{k \in \{0, ..., n-1\} \ : \ \gcd(k, n) = 1\}$$

the *Euler totient function.*

We record some elementary properties of this function in the following lemma.

Lemma 3.3.11. *The following are all true:*

1. *For $m, n \in \mathbb{N}$ with $\gcd(m, n) = 1$ we have $\varphi(mn) = \varphi(m)\varphi(n)$.*
2. *For a prime number p and $r \in \mathbb{N}$ we have $\varphi(p^r) = (p-1)p^{r-1}$.*
3. *For $n \in \mathbb{N}$ we have $n = \sum_{d|n} \varphi(d)$.*

Proof. Exercise.

q.e.d.

Now we return to our question about the construction of the regular n-gon. This is of course equivalent to the question whether the primitive nth root of unity $\zeta_n = e^{2\pi i/n}$ is constructible from $\mathbb{Q}$. Gauß showed the following result.

Theorem 3.3.12. *For $n \geq 3$ the following statements are all equivalent.*

(*i*) *The regular n-gon is constructible with compass and straightedge.*

(*ii*) *$\varphi(n)$ is a power of 2.*

(*iii*) *We have $n = 2^r p_1 \cdots p_k$ for some $r \in \mathbb{N}_0$ and pairwise distinct Fermat primes $p_1, ..., p_k$.*

We do not quite have all the necessary tools to prove this result in full. Therefore we defer part of the proof to Section 5.3.

Proof. We leave the equivalence of (*ii*) and (*iii*) as an exercise.

We now show that if the regular n-gon is contructible, then n must have the form $n = 2^r p_1 \cdots p_k$ for pairwise distinct Fermat primes $p_1, ..., p_k$. It is clear that if the regular n-gon is constructible, then so is the regular m-gon for each $m \mid n$. So it suffices to show that for each odd prime p we have $p-1$ is a power of 2 and that the regular p^2-gon is not constructible.

Let $\zeta_p = e^{2\pi i/p}$ be a primitive pth root of unity. We have already seen in Example 2.2.15 that the pth cyclotomic polynomial

$$\Phi_p = \frac{X^p - 1}{X - 1} = X^{p-1} + X^{p-2} + ... + X + 1$$

is irreducible over $\mathbb{Q}$ and has ζ_p as a root, so that the field extension $K = \mathbb{Q}(\zeta_p)/\mathbb{Q}$ has degree $[\mathbb{Q}(\zeta) : \mathbb{Q}] = p - 1$. But since ζ_p is constructible by assumption, we know from Theorem 3.3.5 that ζ_p must lie in some field extension $L/\mathbb{Q}$ of degree $[L : \mathbb{Q}] = 2^\ell$ for some ℓ. Since $\mathbb{Q}(\zeta_p)$ is the smallest field over $\mathbb{Q}$ containing ζ_p, it must be a subfield of L, wherefore by the Degree Theorem 3.1.5 we have

$$p - 1 = [Q(\zeta_p) : \mathbb{Q}] \mid [L : \mathbb{Q}] = 2^\ell,$$

whence $p - 1$ must be a power of 2.

Next let $\zeta_{p^2} = e^{2\pi i/p^2}$ be a primitive p^2th root of unity and consider the p^2th cyclotomic polynomial

$$\Phi_{p^2} = \frac{X^{p^2} - 1}{X^p - 1} = X^{p^2-p} + X^{p^2-2p} + ... + X^p + 1.$$

As we saw in Example 2.2.16, this polynomial is irreducible over $\mathbb{Q}$ and has ζ_{p^2} as a root, wherefore the degree $[\mathbb{Q}(\zeta_{p^2}) : \mathbb{Q}] = p(p-1)$. If ζ_{p^2} were constructible, it would have to lie in a field extension $L/\mathbb{Q}$ whose degree is a power of 2, but this cannot be since the smallest field containint ζ_{p^2} has degree divisible by p. Therefore ζ_{p^2} cannot be constructible.

q.e.d.

Example 3.3.13. To give a flavour of the remainder of the proof of Theorem 3.3.12, namely that for each n of the form given in (iii), the regular n-gon is indeed constructible, we look at the first non-trivial case $n = 5$.

Let $L = \mathbb{Q}(\zeta_5)$ and consider the field $K = L \cap \mathbb{R}$. Then, as discussed above, $[L : \mathbb{Q}] = 4$. We clearly have $L \neq K$, so that $[L : K] \geq 2$.

Now note that $\zeta_5^{-1} = \overline{\zeta}_5$, so that $\phi = \zeta + \zeta^{-1} = \zeta_5 + \zeta_5^4 = 2\cos(2\pi/5)$ is real and hence $\phi \in K$. Therefore ζ_5 is a root of the quadratic polynomial

$$X^2 - \phi X + 1 \in K[X],$$

so indeed we find $[L : K] = 2$. Setting $\phi^* = \zeta_5^2 + \zeta_5^{-2} = \zeta_5^2 + \zeta_5^3$, we see using Theorem 1.1.15 that

$$\phi + \phi^* = \zeta_5 + \zeta_5^2 + \zeta_5^3 + \zeta_5^4 = -1$$

and

$$\phi\phi^* = \zeta_5^3 + \zeta_5^4 + \zeta_5^6 + \zeta_5^7 = -1,$$

so that ϕ and ϕ^* are the roots of the quadratic polynomial

$$X^2 + X - 1 \in \mathbb{Q}[X].$$

Therefore we can represent ζ_5 through the successive solution of quadratic equations, which means that we can construct it. Explicitly we find that

$$\zeta_5 = \frac{\sqrt{5}-1}{4} + \mathrm{i}\,\frac{\sqrt{10+2\sqrt{5}}}{4}. \tag{3.1}$$

3.3.4 Construction of the regular pentagon

Topics

- Construction of the regular pentagon

In this section we demonstrate a geometric construction of the regular pentagon, which, as we have already seen abstractly through Theorem 3.3.12 and Example 3.3.13, must be possible. Indeed we can start with given points 0 and 1, draw the real line through them and first construct the numbers 5 and 6 on there (see Figure 3.7).

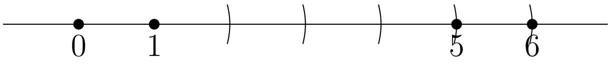


Figure 3.7: Step 1

In the nect step we construct the length $\sqrt{5}$ as described in Theorem 3.3.3: Drawing a circle of radius 3 around 3 and finding one of its intersection point with the line perpedincular to the real line through 5 yields the point $z_0 = 5 + \mathrm{i}\sqrt{5}$, so the distance between 5 and z_1 is exactly $\sqrt{5}$, which we can mark on the real line (see Figure 3.8).

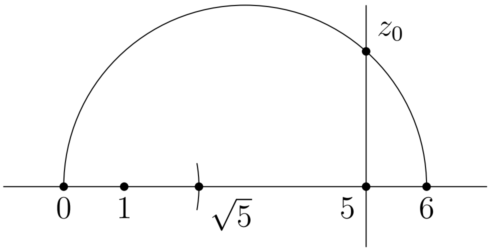


Figure 3.8: Step 2

Now it is straightforward to construct the point $z_1 = -1 + \sqrt{5}$, which is the real part of $4\zeta_5$ by (3.1). Therefore we can draw a circle of radius 4 around 0, so that the line perpendicular to the real axis through z_1 intersects the circle above the real line at $z_2 = 4\zeta_5$ (see Figure 3.9).

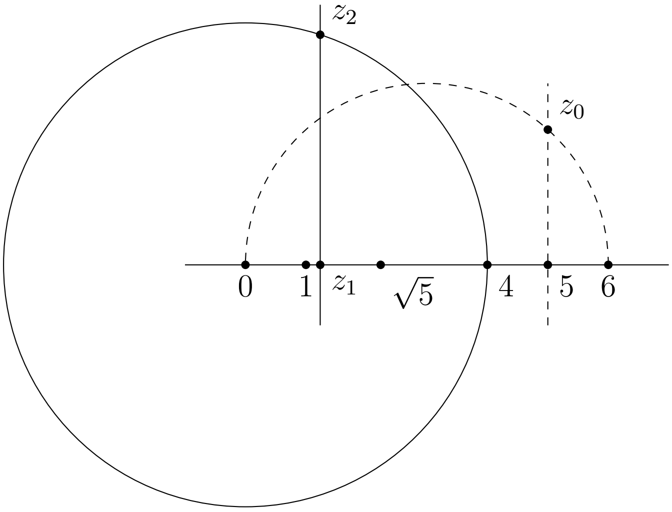


Figure 3.9: Step 3

We now have found two vertices of our regular pentagon, so we can simply construct the remaining three by drawing a circle of radius $\overline{1, z_2}$ around first z_2

to obtain the third vertex z_3, a circle around z_3 with the same radius to find the fourth vertex z_4 and the same again for z_4 to find the remaining fifth vertex z_5. We then connect the vertices and have constructed a regular pentagon with compass and straightedge (see Figure 3.10).

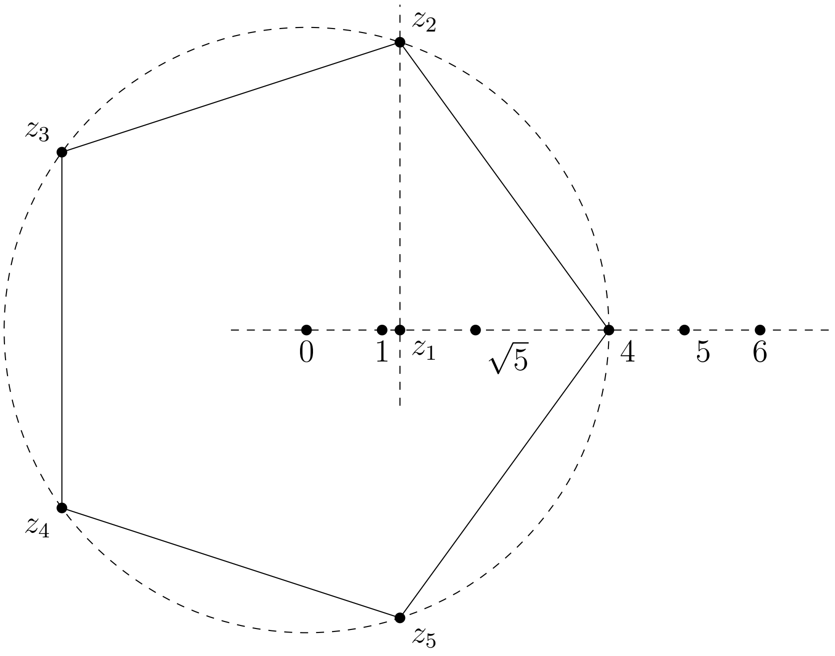


Figure 3.10: Step 4

3.4 Finite fields

Topics

- Frobenius endomorphism
- Fields with positive characteristic
- Formal derivative of a polynomial
- Existence and uniqueness of finite fields

In this section we want to talk briefly about finite fields. Not surprisingly, it is far easier to understand the extensions of finte fields than of infinite fields. This has very practical applications, as for example certain algorithms in cryptography (namely the *Advanced Encryption Standard* AES) rely on the arithmetic in such extensions of finite fields.

We first record the following result.

Proposition 3.4.1. *Let K be any field and $U \leq K^\times$ be a finite subgroup of $K^\times$. Then U is* cyclic*, i.e. there is some $z \in K$ such that every element in U is an integer power of z.*

Proof. Since $K^\times$ is abelian, U is a finitely generated abelian group. If U were not cyclic, the Main Theorem on Finitely Generated Abelian Groups implies that there must be some prime number p and elements $a, b \in U$ such that a and b, as well as all expressions $a^i b^j$ with $0 \leq i, j \leq p-1$ and $i + j > 0$, have order p. But this means that each of the p^2 distinct elements $a^i b^j$ is a root of the polynomial $X^p - 1$, which can have at most p distinct roots in K. Therefore we have a contradiction so that U must be cyclic.

q.e.d.

We now introduce one of the most important maps for fields of positive characteristic.

Lemma 3.4.2. *Let K be a field of characteristic $p > 0$. Then the map*

$$\Phi_p : K \to K, a \mapsto a^p$$

defines an endomorphism of K. If K is finite, the Φ_p is an automorphism of K.

Proof. Φ_p clearly maps 0 to 0 and 1 to 1 and for $a, b \in K$ we have $\Phi_p(ab) = \Phi_p(a)\Phi_p(b)$ and for $a \neq 0$ also $\Phi_p(a^{-1}) = \Phi_p(a)^{-1}$. Note that the binomial coefficient $\binom{p}{k}$ is divisible by p, therefore 0 in K, for each $0 < k < p$, so that we have for $a, b \in K$

$$\Phi_p(a+b) = (a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k} = a^p + b^p = \Phi_p(a) + \Phi_p(b).$$

Therefore Φ_p is an endomorphism.

As a field homomorphism Φ_p is injective and since injective maps from any finite set to itself are automatically also surjective we see that Φ_p is bijective if K is finite.

q.e.d.

Definition 3.4.3. Let K be a field of characteristic $p > 0$. The endomorphism from Lemma 3.4.2 is called the *Frobenius endomorphism* (or just the *Frobenius*) of K.

We now want to classify finite fields completely. In order to do so we need the following general lemma which we also need in the next section.

Lemma 3.4.4. *For any field K and a polynomial $f = a_nX^n + ... + a_0 \in K[X]$ of degree n we denote by $f' := na_nX^{n-1} + ... + a_1$ the (formal)* derivative *of f. Suppose f splits completely into linear factors, i.e. there are numbers $\alpha_1, ..., \alpha_n \in K$ such that $f = a_n \prod_{j=1}^{n}(X - \alpha_j)$. Then the α_j are all distinct if and only if* $\gcd(f, f') = 1$.

Proof. Suppose that two of the α_j are both equal to $\alpha \in K$, so that $(X - \alpha)^2 \mid f$. Thus we may write $f = (X - \alpha)^2 \cdot g$ for some $g \in K[X]$ and we therefore find that $f' = 2(X - \alpha)g + (X - \alpha)^2 g' = (X - \alpha)(2g + (X - \alpha)g')$. Therefore $(X - \alpha) \mid \gcd(f, f')$.

If on the other hand we have $\gcd(f, f') \neq 1$, then there must be some $\alpha_j = \alpha$ such that $(X - \alpha) \mid \gcd(f, f')$. We can therefore write $f = (X - \alpha)h$ for some $h \in K[X]$ and compute $f' = h + (X - \alpha)h'$. But since by asumption $(X - \alpha) \mid f'$, this implies that $(X - \alpha) \mid h$, wherefore we have $(X - \alpha)^2 \mid f$. Thus at least two of the α_j must be equal to α, so they are not all distinct.

q.e.d.

We now come to the announced classification theorem for finite fields.

Theorem 3.4.5. 1. *Let K be a finite field. Then the characteristic of K is a prime number p and $\#K = p^n$ for some $n \geq 1$.*

2. *On the other hand, for each prime power $q = p^n$ there is a field $\mathbb{F}_q$ with q elements. This field $\mathbb{F}_q$ is unique up to isomorphism.*

Proof.

1. We know from Proposition 2.1.26 that the characteristic of K is either 0 or a prime number p. Since K is finite, the characteristic cannot be 0, so it must be some prime p. Therefore K contains the prime field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ as a subfield, so that it is in particular a finite-dimensional vector space over $\mathbb{F}_p$. Therefore we have $\#K = p^n$ with $n = [K : \mathbb{F}_p]$.

2. Let p be a prime number and n be a natural number. We first show existence of a field with p^n elements: Consider the polynomial $f = X^{p^n} - X \in \mathbb{F}_p[X]$ and let K denote its splitting field. Therefore we can factor $f = \prod_{j=1}^{p^n}(X - \alpha_j)$, where $Z := \{\alpha_1, ..., \alpha_{p^n}\} \subseteq K$. Since $f' = -1$ in $K[X]$, we have $\gcd(f, f') = 1$, so that by Lemma 3.4.4 we have $\#Z = p^n$. Now $f(0) = f(1) = 0$ and more generally we find that $a \in K$ is a root of f if and only if $\Phi_p^n(a) = a$. The nth power of the Frobenius is of course again

a field automorphism, so that $Z = \mathrm{Fix}(\Phi_p)$ is a subfield of K containing all the roots of f, so in fact we have $Z = K$ and $\#K = p^n$.

Now we show uniqueness of the field. Let L be any field with $\#L = p^n$. Then the prime field of L is again isomorphic to $\mathbb{F}_p$. Furthermore $L^\times$ is a group with p^n-1 elements, so that we have $a^{p^n-1} = 1$ for all $a \in L^\times = L\backslash\{0\}$ (Exercise). Therefore L contains the p^n roots of the polynomial $X^{p^n} - X \in \mathbb{F}_p[X]$, making L the splitting field of said polynomial. Since the splitting field of a polynomial is unique up to isomorphism by Theorem 3.2.8, we see that K and L are indeed isomorphic as claimed.

q.e.d.

We conclude this section with a few consequences of this classification.

Remark 3.4.6. *1. Every $a \in \mathbb{F}_{p^n}^\times$ which generates $\mathbb{F}_{p^n}^\times$ as a group, satisfies $\mathbb{F}_{p^n} = \mathbb{F}_p[a]$. Such an a is called a* primitive root *of $\mathbb{F}_{p^n}$.*

2. The subfields of $\mathbb{F}_{p^n}$ are precisely $\mathbb{F}_{p^d}$ for $d \mid n$.

3. For $d \mid n$ we have

$$\mathbb{F}_{p^d} = \{a \in \mathbb{F}_{p^n} \ : \ a^{p^d} = a\} = \mathrm{Fix}(\Phi_p^d).$$

Proof. Exercise.

q.e.d.

3.5 Separable extensions

Topics

- Separable extensions and polynomials
- Perfect fields
- Primitive element theorem

In this section we discuss a special class of field extensions, now again for general fields, which will be essential for the formulation of Galois Theory in Chapter 5. We begin with a definition.

Definition 3.5.1. Let K be a field.

1. Let $f \in K[X]$ a polynomial. We call f *separable* if all roots of f in a splitting field of f are distinct. Otherwise we call f *inseparable.*

2. Let E/K be a field extension and $a \in E$. We call a *separable* if its minimal polynomial over K is separable. If every $a \in E$ is separable, we call the extension E/K *separable.*

As we saw in Lemma 3.4.4, it is usually fairly easy to test whether or not a given polynomial is separable or not if it factors completely over the ground field K. By the following remark, this is also true if it only splits in an extension field.

Remark 3.5.2. *Let $f, g \in K[X]$ be polynomials with coefficients in some field K and let $h := \gcd(f, g) \in K[X]$. Then we also have $h = \gcd(f, g) \in E[X]$ for any extension field E/K.*

Proof. Denote by $\tilde{h} \in E[X]$ the greatest common divisor of f and g in $E[X]$.

It follows from the Euclidean algorithm that there exist polynomials $\alpha, \beta \in K[X]$ such that $h = \alpha f + \beta g$. Now let $s \in E[X]$ be a common divisor of f and g in $E[X]$, i.e. there are polynomials $\tilde{f}, \tilde{g} \in E[X]$ such that $f = s\tilde{f}$ and $g = s\tilde{g}$. Thus we find

$$h = \alpha f + \beta g = s(\alpha\tilde{f} + \beta\tilde{g}),$$

so that s is a divisor of h in $E[X]$. In particular we have $\tilde{h} \mid h$ in $E[X]$.

By the same argument applied with interchanged roles of h and $\tilde{h}$ we must also have $h \mid \tilde{h}$, so h and $\tilde{h}$ can only differ by a constant factor.

q.e.d.

By this remark, Lemma 3.4.4 provides a general test for separability of a polynomial. We illustrate this with an example.

Example 3.5.3. 1. Every irreducible polynomial over $\mathbb{Q}$ is separable. Indeed let $f \in \mathbb{Q}[X]$ be irreducible. Then, since $\operatorname{char} \mathbb{Q} = 0$, and $\deg f \geq 1$, we have $f' \neq 0$ and $\deg f' < \deg f$, so we must have $\gcd(f, f') = 1$.

2. Let p be a prime number and let $K = \mathbb{F}_p(t)$ be the rational function field over $\mathbb{F}_p$. Then the polynomial $X^p - t \in K[X]$ is clearly irreducible over K, but we have $f' = pX^{p-1} = 0 \in K[X]$, so that $\gcd(f, f') = f \neq 1$.

As it turns out most extensions we consider in this course are separable.

Definition 3.5.4. A field K is called *perfect* if all its finite extensions are separable.

Theorem 3.5.5. *Let K be a field.*

1. *If* $\operatorname{char} K = 0$*, then* K *is perfect.*

2. *If* K *is a finite field, then* K *is perfect.*

Proof.

1. This follows from the same argument we saw in Example 3.5.3 that every irreducible polynomial over $\mathbb{Q}$ is separable, which is clearly equivalent to $\mathbb{Q}$ being a perfect field.

2. Let $f = \sum_{j=0}^{n} a_n X^n$ be an irreducible polynomial. If $f' \neq 0$, then f is separable by the same argument as above. So assume that $f' = \sum_{j=1}^{n} j a_j X^{j-1} = 0$. Therefore we can have $a_j \neq 0$ only if $j \equiv 0 \pmod p$, so we can write
$$f = \sum_{j=0}^{\lfloor n/p \rfloor} a_{pj} X^{pj}.$$
Since K is finite, we know that the Frobenius map Φ_p is an automorphism of K, so for each a_{pj} there exists a unique b_j such that $\Phi_p(b_j) = b_j^p = a_{pj}$. Therefore the polynomial $g = \sum_{j=0}^{\lfloor n/p \rfloor} b_j X^j$ satisfies
$$g^p = \left(\sum_{j=0}^{\lfloor n/p \rfloor} b_j X^j \right)^p = \sum_{j=0}^{\lfloor n/p \rfloor} b_j^p X^{pj} = \sum_{j=0}^{\lfloor n/p \rfloor} a_{pj} X^{pj} = f.$$
But this is a contradiction to f being irreducible, so that this case cannot occur.

q.e.d.

Note that it is essential in the second part of the proof of Theorem 3.5.5 that the field K be finite. Otherwise we cannot guarantee that the Frobenius is an automorphism, and indeed it does not have to be, as we saw in Example 3.5.3.

We now address an important property of separable extensions. Recall that an extension is called simple if it is generated by a single element, called a *primitive element* (see Definition 3.1.6).

Theorem 3.5.6. *(Primitive element theorem). Let* K *be a field and* $E = K(\beta, \gamma)/K$ *a finite extension, where* γ *is separable over* K*. Then there exists* $\alpha \in E$ *such that* $E = K(\alpha)$*. In particular, every finite, separable extension is simple.*

Proof. If K is finite, then this follows from Theorem 3.4.5 and Remark 3.4.6, so we assume from now on that K is infinite.

Let $\mu_\beta, \mu_\gamma \in K[X]$ be the minimal polynomials of β and γ over K and let L be the splitting field of $\mu_\beta \mu_\gamma$ over K. Thus we can write

$$\mu_\beta = \prod_{i=1}^{n}(X - \beta_i), \quad \mu_\gamma = \prod_{j=1}^{m}(X - \gamma_j) \in L[X],$$

where we assume $\beta = \beta_1$ and $\gamma = \gamma_1$. Since γ is separable over K, we know that $\gamma_j \neq \gamma_{j'}$ for $j \neq j'$. Since we assumed K to be infinite, there exists some $a \in K$ such that

$$\beta_i + a\gamma_j \neq \beta + a\gamma \text{ for all } 1 \leq i \leq n,\ 2 \leq j \leq m.$$

We define $\alpha := \beta + a\gamma$ and claim that $E = K(\alpha)$.

We have $\mu_\beta(\alpha - a\gamma) = \mu_\beta(\beta) = 0$, so that γ is a root of $h := \mu_\beta(\alpha - aX) \in K(\alpha)[X]$. Therefore γ must also be a root of $\gcd(\mu_\gamma, h) \in K(\alpha)[X]$. For each $j \neq 1$ that $h(\gamma_j) = \mu_\beta(\beta + a\gamma - a\gamma_j)$. By the choice of a, we have $\beta + a\gamma - a\gamma_j \neq \beta_i$ for all i, so $h(\gamma_j) \neq 0$. Therefore the only common root of h and μ_γ is γ itself, wherefore we must have $\gcd(h, \mu_\gamma) = (X - \gamma)$, so that $\gamma \in K(\alpha)$ and thus also $\beta = \alpha - a\gamma \in K(\alpha)$, so that we find that $K(\alpha) = E$ as claimed.

q.e.d.

Example 3.5.7. As we saw in Example 3.2.2, the field $K = \mathbb{Q}(\sqrt[4]{2}, \mathrm{i})$ is the splitting field of the polynomial $X^4 - 2 \in \mathbb{Q}[X]$. It follows from the proof of Theorem 3.5.6 that $\alpha = \sqrt[4]{2} + \mathrm{i}$ is a primitive element for K, i.e. $K = \mathbb{Q}(\alpha)$. Indeed one can find explicitly by means of a somewhat tedious computation that

$$\sqrt[4]{2} = \frac{1}{24}(5\alpha^7 + 19\alpha^5 + 5\alpha^3 + 151\alpha) \quad \text{and} \quad \mathrm{i} = -\frac{1}{24}(5\alpha^7 + 19\alpha^5 + 5\alpha^3 + 127\alpha).$$

To conclude this section we add a remark on separable extensions in comparison to the algebraic closure.

Remark 3.5.8. *If $E = K(\alpha) \cong K[X]/(f)$ is a minimal rupture field of an irreducible polynomial $f \in K[X]$ and $\overline{K}$ is a fixed algebraic closure of K, then any K-homomorphism $E \hookrightarrow \overline{K}$ maps a root of f to another root of f and it is uniquely determined by the image of α. Therefore we have*

$$\#\{\varphi :\ E \hookrightarrow \overline{K}\ :\ \varphi\ K\text{-homomorphism}\} = \#\{\beta \in \overline{K}\ :\ f(\beta) = 0\}.$$

It follows therefore by Theorem 3.5.6 that for any finite separable extension E/K of degree n that

$$\#\{\varphi :\ E \hookrightarrow \overline{K}\ :\ \varphi\ K\text{-homomorphism}\} = n,$$

since there exists a primitive element for E and its minimal polynomial has exactly n distinct roots in $\overline{K}$.

3.6 Normal extensions

Topics

- Normal extensions
- Classification as splitting fields of polynomials

We now proceed to study another important class of field extensions. Recall from the discussion at the end of Section 3.2 that we assume that every field K has an *algebraic closure*, which is unique up to isomorphism and which we denote by $\overline{K}$. Also recall the notion of a K-homomorphism (see Definition 3.2.3), i.e. a homomorphism of two extension fields of K which acts as the identity on K.

Definition 3.6.1. Let K be a field and $\overline{K}$ its algebraic closure. We call an algebraic extension $K \subseteq E \subseteq \overline{K}$ *normal* if every K-homomorphism $\varphi : E \to \overline{K}$ satisfies $\varphi(E) = E$.

When it comes to explicit examples, this definition is usually only useful to verify that an extension is not normal.

Example 3.6.2. The extension $E = \mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is *not* a normal extension. For example the map defined by $\sqrt[4]{2} \mapsto \mathrm{i}\sqrt[4]{2}$ yields a $\mathbb{Q}$-homomorphism $\varphi : E \to \overline{\mathbb{Q}}$, but it clearly doesn't map E to itself.

With the following theorem, we find a very useful criterion to test whether or not a given field extension is normal.

Theorem 3.6.3. *Let E/K be a field extension. Then the following statements are equivalent.*

(i) E/K is normal.

(ii) Every irreducible polynomial $f \in K[X]$ which has a root in E decomposes into linear factors in $E[X]$, i.e. all roots of f lie in E.

(iii) The minimal polynomial over K of each element in E decomposes into linear factors in $E[X]$.

(iv) The minimal polynomial over K of each generator of E decomposes into linear factors in $E[X]$.

Proof. We first show the implication $(i) \Rightarrow (ii)$: Let $f \in K[X]$ be irreducible and let $\alpha \in E$ be a root of f. Suppose $\beta \in \overline{K}$ is another root of f. Then the fields $K[\alpha]$ and $K[\beta]$ are isomorphic over K as they can both be identified with $K[X]/(f)$. We call this isomorphism $\varphi : K[\alpha] \to K[\beta]$, where $\varphi(\alpha) = \beta$. By Lemma 3.2.11 (see Chapter 6 for a proof), this isomorphism can be extended to a K-homomorphism $\psi : E \to \overline{K}$. By assumption E is normal, so we have $\psi(E) = E$. But this implies that $\beta = \varphi(\alpha) = \psi(\alpha) \in E$. Since β was an arbitrary root of f, it follows that all roots of f lie in E, as claimed.

The implications $(ii) \Rightarrow (iii) \Rightarrow (iv)$ are obvious. So we are left with showing the implication $(iv) \Rightarrow (i)$. For this let $\varphi : E \to \overline{K}$ be a K-homomorphism and let $\alpha \in E$ be a generator. We define $\beta = \varphi(\alpha)$ and let $\mu_\beta \in K[X]$ be the minimal polynomial of β over K. Our goal now is to show that $\beta \in E$. Since φ is a K-homomorphism, we have

$$0 = \mu_\beta(\beta) = \mu_\beta(\varphi(\alpha)) = \varphi(\mu_\beta(\alpha)),$$

wherefore we must have $\mu_\beta(\alpha) = 0$ since φ is injective. Since $\mu_\beta \in K[X]$ is irreducible, μ_β is also the minimal polynomial of α. By assumption we therefore find that $\mu_\beta \in E[X]$ decomposes into linear factors, so E contains all roots of μ_β and thus in particular $\beta \in E$.

q.e.d.

For finite extensions E/K, which is what we usually deal with, the above result gives the following important classification of normal extensions.

Corollary 3.6.4. *Let E/K be a finite field extension. Then E/K is normal if and only if E is the splitting field of a polynomial $f \in K[X]$.*

Proof. Let E/K be normal. Since E/K is finite by assumption, there exist $\alpha_1, ..., \alpha_r \in E$ such that $E = K[\alpha_1, ..., \alpha_r]$. Every α_i has a minimal polynomial $\mu_{\alpha_i} \in K[X]$, which by Theorem 3.6.3 decomposes into linear factors in E. Therefore E is the splitting field of the polynomial

$$f = \prod_{i=1}^{r} \mu_{\alpha_i} \in K[X].$$

Now suppose E is the splitting field of a polynomial $f \in K[X]$. Then E is generated over K by the roots $\alpha_1, ..., \alpha_r$ of f. Furthermore, the minimal polynomial of each α_i over K must divide f and since f decomposes into linear factors in $E[X]$, so does each of the minimal polynomials. So it follows again by Theorem 3.6.3 that E/K is normal.

q.e.d.

Example 3.6.5. 1. The pth root of unity $\zeta_p = e^{2\pi \mathrm{i}/p}$ for a prime number p generates the field $\mathbb{Q}(\zeta_p)/\mathbb{Q}$. By definition ζ_p is a root of the pth cyclotomic polynomial $\Phi_p = X^{p-1} + ... + X + 1 \in \mathbb{Q}[X]$, which by Example 2.2.15 is irreducible over $\mathbb{Q}$. All roots of Φ_p are given by ζ_p^k, $k = 1, ..., p-1$, according to Lemma 1.2.2, in particular $\mathbb{Q}(\zeta_p)$ is the splitting field of Φ_p over $\mathbb{Q}$, so that the extension $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ is normal.

2. Every extension E/K of degree $[E : K] = 2$ is normal.

3. Normality is not transitive: Let $L = \mathbb{Q}(\sqrt{2})$ and $E = \mathbb{Q}(\sqrt[4]{2})$. Then the extensions $L/\mathbb{Q}$ and E/L have degree 2 and are therefore both normal, but as we saw in Example 3.6.2, the extension $E/\mathbb{Q}$ is not normal.

3.7 Transcendental extensions*

Topics

- Basic properties of transcendental extensions
- Rudimentary Groebner bases

In our considerations up to now we have focussed on algebraic extensions. In this short section, we take brief detour talking a little about *transcendental extensions.* The easiest example of a transcendental extension of a field K is the field of *rational functions* $K(X) = \operatorname{Frac} K[X]$. There are several interesting interesting questions regarding for example transcendental extensions of $\mathbb{Q}$ as subfields of the real or complex numbers. We will take a look at those in Appendix B. Here, we deal with some general questions.

Definition 3.7.1. Let L/K be a field extension. A finite set $A = \{a_1, ..., a_n\} \subset L$ is called *algebraically dependent* over K, if there exists a polynomial $f \in K[X_1, ..., X_n] \backslash \{0\}$, called an *algebraic dependence* of A, such that $f(a_1, ..., a_n) = 0$. If such a polynomial does not exist, then the set is called *algebraically independent.* An arbitrary subset $A \subseteq L$ is called *algebraically independent* if each of its finite subsets is algebraically independent.

When it comes to field extensions, we need the following concept.

Definition 3.7.2. Let L/K be a field extension. An algebraically independent set $B \subset L$ is called a *transcendence basis* of L over K, if $L/K(B)$ is an algebraic extension. If there is a transcendence basis B such that $L = K(B)$, then we call the extension L/K *purely transcendental.*

Example 3.7.3. 1. Consider the field $L = K(X)$ of rational functions over K. Then $B = \{X\}$ is a transcendence basis of L and L is purely transcendental over K. Note however that $B' = \{X^3\}$ is a transcendence basis of L as well, since clearly $[L : K(X^3)] = 3$ is finite and therefore algebraic.

2. Consider the field $L = K(X)[Y]/(Y^2 - X^3 + X)$ over K. Then we could choose $\{X\}$ as a transcendence basis and we clearly have $[L : K(X)] = 2$. We might just as well choose $\{Y\}$ as a transcendence basis with $[L : K(Y)] = 3$.

One may ask whether every field extension admits a transcendence basis. Intuitively, this sounds reasonable (and is indeed true), but the proof relies again on *Zorn's Lemma* 6.1.9, wherefore we postpone it again to Chapter 6. We can however say the following.

Theorem 3.7.4. *Let L/K be a field extension and let $B \subset L$ be a transcendence basis of L. If B is finite, then for all transcendence bases B' of L/K we have $\#B = \#B'$ and one can exchange any element in B with any transcendental element in L and obtain a transcendence basis. If B is infinite, then every transcendence basis is infinite.*

Proof. Suppose first that $B = \{\beta_1, ..., \beta_n\}$ is finite and let $\alpha \in L$ is transcendental over K. Since $L/K(B)$ is algebraic we must have that α is algebraic over $K(B)$, which means that there is a polynomial $0 \neq f \in K[Y, X_1, ..., X_n]$ such that $f(\alpha, \beta_1, ..., \beta_n) = 0$. Since α is transcendental over K, at least one of the variables X_i must occur in f with a positive exponent, and we may assume without loss of generality that this is the case for X_1. It follows that β_1 is algebraic over $K(\alpha, \beta_2, ..., \beta_n)$, wherefore L is algebraic over $K(\alpha, \beta_2, ..., \beta_n)$. We claim that the set $\{\alpha, \beta_2, ..., \beta_n\}$ is algebraically independent over K:

Suppose there is a polynomial $0 \neq g \in K[Y, X_2, ..., X_n]$ satisfying $g(\alpha, \beta_2, ..., \beta_n) = 0$. As in the proof of Theorem 1.1.15, we introduce a lexicographical ordering on $K[Y, X_1, ..., X_n]$ by imposing that $Y \succ X_1 \succ ... \succ X_n$ and for two monomials $m, m' \in K[Y, X_1, ..., X_n]$ we say that $m \succ m'$ if the exponents of all the variables in order agree up to a point and then the exponent of the next smaller variable in m is larger than that in m'. With respect to this ordering, compare the leading terms in the polynomials f from above and g. We can now multiply both f and g by suitable monomials and constants and take their difference to obtain a polynomial f_1, which then has a smaller leading term than the least common multiple of the leading terms of f and g. We can then subtract a suitable multiple of f or g from f_1 to obtain a polynomial f_2 with a smaller leading term than f_1. Continuing this process of constructing a polynomial f_j from $f, g, f_1, ..., f_{j-1}$ with smaller leading term can only be repeated finitely many times and at some point, we must arrive at a non-zero polynomial $F \in K[Y, X_1, ..., X_n]$, whose degree in Y is actually 0, in other words we have

$$F(Y, X_1, ..., X_n) = F(X_1, ..., X_n).$$

Since $f(\alpha, \beta_1, ..., \beta_n) = g(\alpha, \beta_2, ..., \beta_n) = 0$, the same is true for all the polynomials f_j (since they are all of the form $p_j f + q_j g$ for some polynomials $f, g \in K[Y, X_1, ..., X_n]$) and therefore we also have $F(\alpha, \beta_1, ..., \beta_n) = F(\beta_1, ..., \beta_n) = 0$. But this means that the set $\{\beta_1, ..., \beta_n\}$ would be algebraically dependent, which is a contradiction. Therefore, our claim that the set $\{\alpha, \beta_2, ..., \beta_n\}$ is algebraically independent is proven.

Therefore, $\{\alpha, \beta_2, ..., \beta_n\}$ is also a transcendence basis of L/K.

From what we have shown, it follows that we obtain a bijection between any two (finite) transcendence bases by exchanging their elements successively, so they must have the same number of elements.

In particular, if one transcendence basis is finite, all of them have to be, therefore it is also true that if we have an infinite transcendence basis of L/K, then all transcendence bases must be infinite.

q.e.d.

When we used monomial orders in the proof of Theorem 3.7.4 to argue that an algebraic dependence of $\{\alpha, \beta_2, ..., \beta_n\}$ yields an algebraic dependence of $\{\beta_1, ..., \beta_n\}$, we used, in a somewhat vague fashion, the concept of *Groebner bases.* It is very important, e.g. in Algebraic Geometry, to study these bases, but a thorough discussion would take us too far afield for this course. We just illustrate the concept with an example.

Example 3.7.5. Suppose we have the two polynomials $f = Y^2 + YX_1 + X_2^3, g = Y^3 + YX_2 - X_2 \in \mathbb{Q}[Y, X_1, X_2]$ and we want to construct a polynomial $F \in \mathbb{Q}[X_1, X_2]$ of the form $pf + qg$, where $p, q \in \mathbb{Q}[Y, X_1, X_2]$. As described in the proof of Theorem 3.7.4, the leading terms of f and g are Y^2 and Y^3 resp., so we first obtain the polynomial

$$f_1 = Yf_1 - g = Y^2X_1 + YX_2^3 - YX_2 + X_2.$$

This has a larger leading term than f, but

$$f_2 = f_1 - X_1 f = -YX_1^2 + YX_2^3 - YX_2 - X_1X_2^3 + X_2$$

has a strictly smaller leading term $(-YX_1^2)$ than either f or g.

We now want to construct a polynomial with yet a smaller leading term than that of f_2. For this consider first

$$f_3 = Yf_2 + X_1 f_1 = Y^2(X_2^3 - X_2) + Y(-X_1X_2 + X_2) + X_2X_1,$$

whose leading term can be reduced below that of f_2 using a multiple of the polynomial f:

$$f_4 = f_3 - (X_2^3 - X_2)f = Y(-X_1X_2^3 + X_2) + X_1X_2 - X_2^6 + x_2^4.$$

Continuing in this way successively yields

$$\begin{aligned}
f_5 &= X_1f_4 - X_2^3f_2 = Y(X_1X_2 - X_2^6 + X_2^4) + X_1^2X_2 + X_1X_2^4 - X_2^4\\
f_6 &= X_1f_5 + X_2f_2 = Y(-X_1X_2^6 + X_1X_2^4 + X_2^4 - X_2^2) + X_1^3X_2 + X_1^2X_2^4 - 2X_1X_2^4 + X_2^2\\
f_7 &= f_6 + (X_2^5 - X_2^3)f_5 = Y(-X_2^{11} + 2X_2^9 - X_2^7 + X_2^4 - X_2^2) + X_1^3X_2 + X_1^2X_2^6\\
&\qquad + X_2^9X_1 - X_2^7X_1 - 2X_2^4X_1 - X_2^9 + X_2^7 + X_2^2\\
f_8 &= X_1f_7 + (X_2^{10} - 2X_2^8 + X_2^6 + X_2^3 + X_2)f_5\\
&= Y(2X_1X_2^4 - X_2^{16} + 3X_2^{14} - 3X_2^{12} + X_2^{10} - X_2^9 + X_2^5)\\
&\qquad + X_1^4X_2 + X_1^3X_2^6 + X_1^2X_2^{11} - X_1^2X_2^9 - X_1^2X_2^4 + X_1^2X_2^2 + X_1X_2^{14}\\
&\qquad - 2X_1X_2^{12} + X_1X_2^{10} - X_1X_2^9 + 2X_1X_2^7 + X_1X_2^5 + X_1X_2^2\\
&\qquad - X_2^{14} + 2X_2^{12} - X_2^{10} - X_2^7 - X_2^5\\
f_9 &= f_8 - (X_2^5 - X_2^3)f_7 = X_1^4X_2 + X_1^3X_2^4 - 3X_1^2X_2^4 + X_1^2X_2^2 + X_1X_2^9 - 2X_1X_2^7\\
&\qquad + X_1X_2^5 + X_1X_2^2
\end{aligned}$$

So we find a polynomial $F = f_9$ in the ideal generated by f and g which is independent of Y.

By Theorem 3.7.4, the following definition makes sense.

Definition 3.7.6. Let L/K be a field extension and let $B \subset L$ be a finite transcendence basis of L. Then we define the *transcendence degree* of L/K as $\operatorname{trdeg}(L/K) := \#B$.

We note the following corollary from Theorem 3.7.4.

Corollary 3.7.7. *Let E/L and L/K be field extensions of finite transcendence degree. Then we have*

$$\operatorname{trdeg}(E/K) = \operatorname{trdeg}(E/L) + \operatorname{trdeg}(L/K).$$

Proof. Exercise.

q.e.d.

Chapter 4

Group Theory

Groups are among the most fundamental objects in Algebra. In this chapter we establish some basic facts on groups and their actions which we will need in the following chapter to establish Galois Theory. Many of the facts at least in the first section are probably known from previous courses, but for the sake of completeness, we at least sketch most of the proofs.

4.1 Normal subgroups and the homomorphy theorem

We begin by recalling the definition of a group.

Definition 4.1.1. A set $G \neq \emptyset$ with a map $G \times G \to G, (g,h) \mapsto gh$ is called a *group* if the following properties are satisfied:

1. For $g, h, k \in G$ we have $(gh)k = g(hk)$ (Associativity).

2. There exists an element $1 \in G$ such that $1g = g$ for all $g \in G$ (Existence of 1).

3. For all $g \in G$ there is $h \in G$ such that $gh = 1$ (Existence of inverses). We usually write $h = g^{-1}$.

Example 4.1.2.

1. Any ring R forms a group with respect to addition. It is even a commutative or *abelian* group. With respect to multiplication, the units of a ring form an abelian group as well.

2. We have already encountered the *symmetric group* S_n of all permutations of n objects. The multiplication map is the composition of maps. For $n \geq 3$, this group is *not* abelian.

3. For a vector space V over a field K, the linear, bijective maps $V \to V$ form a group under composition of maps, the *general linear group* $\mathrm{GL}(V)$ of V. Choosing a basis of V, this yields the group $\mathrm{GL}_n(K)$ of all invertible $n \times n$ matrices over K, which form a group under matrix multiplication.

One usually encounters groups via their actions. In geometry, such actions correspond to symmetries of an object.

Definition 4.1.3. Let G be a group and M be a set. Then a map $G \times M \to M$, $(g, x) \mapsto g.x$ is called a *(left) group action* if

1. $1.x = x$ for all $x \in M$,
2. For all $g, h \in G$ and $x \in M$ we have $g.(h.x) = (gh).x$.

Example 4.1.4. There are several important actions of G on itself: One (easy) example is via left-multiplication. Somewhat more intricate is the action of G on itself is the action via *conjugation*,

$$(g, x) \mapsto gxg^{-1}.$$

For later reference we define the following concepts regarding group actions.

Definition 4.1.5. Let G be a group and M a set, such that G acts on M from the left.

1. For $x \in M$ we call $G.x := \{g.x \ : \ g \in G\}$ the *orbit* of x under G.
2. If there exists $x \in M$ such that $G.x = M$, we call the action of G on M *transitive*. Note that in this case, every $x \in M$ satisfies $G.x = M$.
3. For $x \in M$ we call the set $\mathrm{Stab}_G(x) := \{g \in G \ : \ g.x = x\}$ the *stabiliser* of x in G.

We now define maps between groups.

Definition 4.1.6. Let G, H be groups.

1. A map $\varphi : G \to H$ is called a *group homomorphism* if for $g_1, g_2 \in G$ we have
$$\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2).$$
If in addition φ is bijective (injective, surjective), we say that φ is a group *isomorphism* (*monomorphism*, *epimorphism*).
2. If there exists an isomorphism $\varphi : G \to H$, we say that G and H are *isomorphic* and write $G \cong H$.

3. The group of maps

$$\mathrm{Aut}(G) := \{\varphi : G \to G \;:\; \varphi \text{ is an isomorphism}\},$$

which is a group under composition of maps, is called the *automorphism group* of G.

Remark 4.1.7. *It is easy to see that each $g \in G$ induces an automorphism $\kappa_g : G \to G$ via $\kappa_g(x) = gxg^{-1}$. Such an automorphism is called an* inner automorphism *and we write* Inn(G) *for the group of inner automorphisms of G. In other words*

$$\kappa : G \to \mathrm{Aut}(G), g \mapsto \kappa_g$$

is a homomorphism of groups.

Definition 4.1.8. Let G be a group.

1. A subset $U \subseteq G$ is called a *subgroup* of G if $1 \in U$ and for $u, v \in U$ we have $uv \in U$ and $u^{-1} \in U$. We write $U \leq G$.

2. We call $N \leq G$ a *normal subgroup* if for all $n \in N$ and $g \in G$ we have $gng^{-1} \in N$. We then write $N \trianglelefteq G$.

3. We call $N \leq G$ a *characteristic subgroup* of G if $\alpha(N) = N$ for all $\alpha \in \mathrm{Aut}(G)$.

4. The *centre* of G is the subgroup

$$Z(G) := \{g \in G \;:\; gh = hg \text{ for all } h \in G\} = \mathrm{Ker}(\kappa).$$

5. For $g, h \in G$ we call $[g, h] := g^{-1}h^{-1}gh$ the *commutator* of g and h. The smallest subgroup of G containing all commutators of elements in G is called the *commutator subgroup* or *derived subgroup* of G,

$$G' = \langle [g, h] \;:\; g, h \in G \rangle.$$

Remark 4.1.9. *1. Let $\varphi : G \to H$ be a homomorphism of groups. Then* $\mathrm{Ker}\,\varphi := \{g \in G \;:\; \varphi(g) = 1\} \trianglelefteq G$ *is a normal subgroup of G and* $\mathrm{Im}(\varphi) := \{h \in H \;:\; h = \varphi(g)$ *for some* $g \in G\} \leq H$ *is a subgroup of H, which is in general not normal.*

2. Characteristic subgroups are in particular normal subgroups.

3. Both $Z(G)$ and G' are characteristic subgroups of G.

4. *The factor group G/G' is abelian. Indeed any normal subgroup $N \trianglelefteq G$ such that G/N is abelian satisfies $G' \trianglelefteq N$.*

5. *If G is a group, $N \trianglelefteq G$ is a normal subgroup and $C \trianglelefteq N$ is a characteristic subgroup of N, then $C \trianglelefteq G$ is a normal subgroup in G. In general this is not true if C is just a normal subgroup of N.*

Proof. Exercise.

q.e.d.

We now turn to an important structural theorem on groups, which in a similar way also exists for rings.

Theorem 4.1.10. *(Homomorphy theorem)*

1. *Let G be a group and $N \trianglelefteq G$ a normal subgroup. Then $G/N := \{gN : g \in G\}$, where we say $gN = hN$ if and only if $gh^{-1} \in N$, is a group with the multiplication defined $(gN)(hN) := (gh)N$, called the* factor group *of G by N. The map $\nu : G \to G/N, g \mapsto gN$ defines an epimorphism, called the* canonical epimorphism.

2. *Let G, H be groups and let $\varphi : G \to H$ be a group homomorphsim. Then for $N := \operatorname{Ker}\varphi$, the map $\overline{\varphi} : G/N \to H, gN \to \varphi(g)$ is injective and we have $\varphi = \overline{\varphi} \circ \nu$, i.e. the following diagram commutes:*

$$\begin{tikzcd}
G \arrow[r, "\varphi"] \arrow[dr, two heads, "\nu"'] & H \\
 & G/\operatorname{Ker}\varphi \arrow[u, hook, "\overline{\varphi}"']
\end{tikzcd}$$

In particular we have that $\operatorname{Im}(\varphi) \cong G/N$.

Proof.

1. This proof is almost exactly the same as the proof of Proposition 2.1.8, except that we have to pay attention to the fact that multiplication in G is not necessarily commutative. First we note that the multiplication defined above is indeed well-defined. For this let $g, \tilde{g}, h, \tilde{h} \in G$ such that $gN = \tilde{g}N$ and $hN = \tilde{h}N$, i.e. there exist $n_1, n_2 \in N$ such that $g = \tilde{g}n_1$ and $h = \tilde{h}n_2$. Then we have

$$gh = (\tilde{g}n_1)(\tilde{h}n_2) = \tilde{g}\tilde{h}\underbrace{\tilde{h}^{-1}n_1\tilde{h}}_{\in N}n_2 = (\tilde{g}\tilde{h})n'$$

for some $n' \in N$, so we have $(gN)(hN) = (\tilde{g}N)(\tilde{h}N)$. The group axioms in G/N follow directly from those in G and so does the claim about the canonical epimorphism ν.

2. First we show again that $\overline{\varphi}$ is well-defined: If we have $g, \tilde{g} \in G$ such that there exists $n \in N = \operatorname{Ker}\varphi$ with $g = \tilde{g}n$, then

$$\overline{\varphi}(gN) = \varphi(g) = \varphi(\tilde{g}n) = \varphi(\tilde{g})\varphi(n) = \varphi(\tilde{g}) = \overline{\varphi}(\tilde{g}N).$$

Furthermore $\overline{\varphi}$ is clearly a homomorphism by the definition of the multiplication in G/N. We now note that a group homomorphism $\psi : G \to H$ is injective if and only if $\operatorname{Ker}\psi = \{1\}$ (Exercise). We have that $\overline{\varphi}(gN) = \varphi(g) = 1$ if and only if $g \in \operatorname{Ker}\varphi$, whence $gN = 1N \in G/N$, therefore $\overline{\varphi}$ is injective.

Next we note that for $g \in G$ we have

$$\overline{\varphi}(\nu(g)) = \overline{\varphi}(gN) = \varphi(g),$$

so that the diagram does commute. Since $\operatorname{Im}(\overline{\varphi}) = \operatorname{Im}(\varphi)$, we also find the claimed isomorphism.

q.e.d.

A useful corollary of the homomorpy theorem is the following "cancellation rule" for normal subgroups.

Corollary 4.1.11. *Let G be a group and $N_1, N_2 \trianglelefteq G$ normal subgroups such that $N_2 \leq N_1$. Then the following are true.*

1. *The factor group N_1/N_2 is a normal subgroup in G/N_2.*
2. *We have that $(G/N_2)/(N_1/N_2) \cong G/N_1$.*

Proof. Exercise.

q.e.d.

In what follows, we will mainly be concerned with finite groups. From now on, all groups G will be finite groups, unless explicitly stated otherwise. For those the following theorem attributed to Lagrange is often important.

Theorem 4.1.12. *(Lagrange's theorem) Let G be a finite group and $U \leq G$ a subgroup. Then $(\#U) \mid (\#G)$.*

Proof. U acts on G via left-multiplication. This defines an equivalence relation on G via $g \sim h$ if and only if there is $u \in U$ such that $ug = h$. The equivalence classes now partition G, i.e. there exist representatives $1 = g_1, ..., g_r \in G$ such that for each $g \in G$ there exists exactly one $j \in \{1, ..., r\}$ such that $g = ug_j$ for some $u \in U$. Clearly all these equivalence classes have the same cardinality $\#U$, so that we obtain that

$$\#G = \sum_{j=1}^{r} \#(Ug_j) = r\#U.$$

q.e.d.

When we discuss soluble groups in Section 4.2 we shall need the following rather special result on so-called p-groups.

Proposition 4.1.13. *Let G be a p-group, i.e. $\#G = p^r$ for some $r \geq 1$ and a prime number p. Then the following are true.*

1. *The centre of G is non-trivial, $Z(G) \neq \{1\}$.*
2. *We have $G \neq G'$.*

Proof.

1. Consider the action of G on itself via conjugation. As in the proof of Theorem 4.1.12, this action defines an equivalence relation, where we say $x \sim y$ for $x, y \in G$ if there is some $g \in G$ such that $x = gyg^{-1}$. The *orbit* of any element $x \in G$, i.e ${}^{G}x := \{gxg^{-1} \ : \ g \in G\}$ has length $\#G/\#C_G(x)$ where $C_G(x) := \{g \in G \ : \ gxg^{-1} = x\}$ denotes the *centraliser* of x in G. Since $C_G(x)$ is a subgroup of G (exercise), its order must be a power of p by Lagrange's Theorem 4.1.12. The orbits again partition the group, i.e. there exist $x_1, ..., x_n \in G$ such that $G = \bigcup_{j=1}^{n} {}^{G}x_j$ and any two of these orbits are either equal or disjoint. Now we have that $g \in Z(G)$ if and only if $C_G(g) = G$ by definition, so if and only if $\#^{G}g = 1$. So we have

   $$p^r = \#G = \sum_{g_j \in Z(G)} \#^{G}g_j + \sum_{g_j \notin Z(G)} \#^{G}g_j = \#Z(G) + \sum_{g_j \notin Z(G)} \#^{G}g_j.$$

   Since $\#^{G}g_j$ is divisible by p for $g_j \notin Z(G)$, we must also have that $\#Z(G)$ is divisible by p, so in particular it cannot be 1.

2. Suppose G is a counterexample of minimal order. Then, since $Z(G)$ is non-trivial and a normal subgroup of G, the factor group $H = G/Z(G)$ is not a counterexample. Therefore we must either have $H = \{1\}$ or $H' \neq H$.

In the first case, G must be abelian and $G' = \{1\} \neq G$, so G is not a counterexample and we have a contradiction. Therefore suppose we are in the latter case, $H \neq H'$. Since for $g, h \in G$ we have $[gZ(G), hZ(G)] = [g, h]Z(G)$ it follows from the fact that $(G/Z(G))' \neq G/Z(G)$ that $G'Z(G) \neq G$, so in particular $G' \neq G$, so again G is not a counterexample and we arrive again at a contradiction.

q.e.d.

4.2 Soluble groups

In this section we introduce the concept of soluble groups.

Definition 4.2.1. 1. For a group G we define its kth *derived subgroup* $G^{(k)}$ inductively by setting

$$G^{(0)} := G, \qquad G^{(k)} = (G^{(k-1)})', \ k > 0.$$

The resulting series of normal subgroups

$$G = G^{(0)} \trianglerighteq G' \trianglerighteq G^{(2)} \trianglerighteq ...$$

is called the *commutator series* of G.

2. The group G is called *soluble* if there exists a $k \in \mathbb{N}$ such that $G^{(k)} = \{1\}$.

Example 4.2.2. 1. Any abelian group is soluble, since G is abelian if and only if $G' = \{1\}$.

2. Any p-group G is soluble, since we know by Proposition 4.1.13 that $G' \neq G$, so the groups in the commutator series must become smaller in each step and thus the series must arrive at $\{1\}$ eventually.

3. Consider the group S_4 of permutations of 4 elements, which we identify with the numbers $\{1, 2, 3, 4\}$. Recall that we can represent these permutations by *cycles*: For example the cycle $(1, 3, 4)$ represents the permutation π with $\pi(1) = 3$, $\pi(3) = 4$, $\pi(4) = 1$ and $\pi(2) = 2$.

   The commutator subgroup of S_4 is given by the *alternating group* $S_4' = A_4$ consisting of all *even* permutations. This can be seen for example by noting that each commutator in S_4 is in A_4 since every commutator gives rise to an even permutation and one can check that every 3-cycle in S_4 can be written as a commutator,

$$(1,2,3) = [(1,3),(2,3)],\ (1,2,4) = [(1,4),(2,4)],\ (2,3,4) = [(2,4),(3,4)]$$

and the 3-cycles generate A_4.

The commutator subgroup of A_4 is known as the *Klein* 4*-group* $V_4 = \langle (1,2)(3,4),(1,3)(2,4)\rangle$ which has order 4 and every element (except 1) has order 2. It can be realised as the symmetry group of a rectangle and it is abelian, so we have $V_4' = \{1\}$. Therefore the commutator series of S_4 is given by

$$S_4 \trianglerighteq A_4 \trianglerighteq V_4 \trianglerighteq \{1\}$$

and we see that S_4 is soluble.

The following theorem gives an alternative description of soluble groups which is sometimes easier to work with.

Theorem 4.2.3. *A group G is soluble if and only if there is a* subnormal series

$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq ... \trianglerighteq G_r = \{1\}$$

(N.B.: we require $G_i \trianglelefteq G_{i-1}$ but not necessarily $G_i \trianglelefteq G$), such that each factor group G_i/G_{i+1}, $i = 0, ...r-1$, is abelian.

Proof. Since for $i \geq 1$ we have $G^{(i)} = (G^{(i-1)})'$, each factor group in the commutator series is abelian, so the commutator series provides a subnormal series with the desired properties.

Suppose on the other hand we have a subnormal series

$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq ... \trianglerighteq G_r = \{1\}$$

such that each factor group G_i/G_{i+1} is abelian. Therefore by Remark 4.1.9 $G_i' \trianglelefteq G_{i+1}$ for all $i = 0, ..., r-1$. It follows by induction that $G^{(i)} \trianglelefteq G_i$ for all i, so in particular $G^{(r)} = \{1\}$, wherefore G is soluble.

q.e.d.

In order to decide whether or not a given group is soluble it is often useful to rely on the following result.

Proposition 4.2.4. *Let G be a group.*

1. *If G is soluble and $U \leq G$ is a subgroup, then U is soluble.*

2. *If G is soluble and $N \trianglelefteq G$ is a normal subgroup, then the factor group G/N is soluble.*

3. *Let $N \trianglelefteq G$ be a normal subgroup. If both N and G/N are soluble, then so is G.*

Proof.

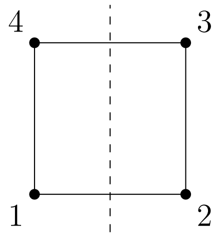
1. Suppose $G^{(k)} = \{1\}$ for some $k \in \mathbb{N}$. If $U \leq G$, it is clear that $U^{(i)} \leq G^{(i)}$ for all i and therefore we have $U^{(k)} = \{1\}$, wherefore U is soluble.

2. We first note that for any homomorphism $\varphi : G \to H$ we have $\varphi(G^{(i)}) = (\varphi(G))^{(i)}$ (Exercise) for every $i \geq 0$. Supposing again that $G^{(k)} = \{1\}$ for some $k \in \mathbb{N}$, this implies for the canonical epimorphism $\nu : G \to G/N$ that
$$(G/N)^{(k)} = (\nu(G))^{(k)} = \nu(G^{(k)}) = \nu(\{1\}) = \{1N\},$$
so that G/N is soluble.

3. Suppose that we have $N^{(k)} = \{1\}$ and $(G/N)^{(n)} = \{1N\}$ for suitable $k, n \in \mathbb{N}$. Consider again the canonical epimorphism $\nu : G \to G/N$. Then we have
$$\{1N\} = (G/N)^{(n)} = \nu(G)^{(n)} = \nu(G^{(n)}),$$
so that we must have $G^{(n)} \leq N$. But then it follows that
$$G^{(n+k)} = (G^{(n)})^{(k)} \leq N^{(k)} = \{1\},$$
wherefore $G^{(n+k)} = \{1\}$ and G is soluble as claimed.

q.e.d.

Example 4.2.5. Consider the so-called *dihedral group* $D_4 \leq S_4$ generated by the 4-cycle $\rho = (1, 2, 3, 4)$ and the double transposition $\tau = (1, 2)(3, 4)$. It can be thought of as the symmetry group of a square: Label the vertices of a square by the numbers $1, 2, 3, 4$:



Then the permutation ρ corresponds to a rotation of the square around the centre by 90°, and τ represents the reflection at the axis marked by the dashed line. It is easy to check that D_4 is not abelian (the two generators don't commute) and that $\#D_4 = 8$. The subgroup $N = \langle \rho \rangle$ has order 4 and thus index 2 in D_4, wherefore we have $N \trianglelefteq D_4$. N is clearly abelian (as a cyclic group) and hence soluble, and the factor group $D/N \cong \langle \tau \rangle$ has order 2 and is therefore also abelian and hence soluble, so that by Proposition 4.2.4, D_4 is soluble.

The same reasoning applies for all dihedral groups $D_n \leq S_n$, $n \geq 3$: It is the symmetry group of the regular n-gon and generated by the n-cycle $(1, 2, ..., n)$ (corresponding to a rotation by $(360/n)°$) and a transposition τ corresponding to a reflection along a symmetry axis containing at most one vertex of the n-gon. It has order $2n$, is not abelian, but soluble.

To conclude this chapter, we consider an important example of a non-soluble group.

Definition 4.2.6. A group G is called *simple* if it has no non-trivial normal subgroups, i.e. if $N \trianglelefteq G$ then either $N = \{1\}$ or $N = G$.

Example 4.2.7. The easiest examples of simple groups are the cyclic groups of prime order C_p: By Lagrange's Theorem 4.1.12, the order of any subgroup of C_p has to divide the order p of C_p, but since p is prime, this means that the order can only be 1 or p, so in fact, C_p doesn't have any non-trivial subgroups, let alone normal ones. In fact all abelian simple groups are isomorphic to C_p for some prime p.

Note that since $G' \trianglelefteq G$, any non-abelian simple group must satisfy $G = G'$ (i.e. G is *perfect*), so that non-abelian simple groups provide examples of non-soluble groups. The rest of this section will be devoted to establishing a whole family of non-abelian simple groups.

For this recall the following definition.

Definition 4.2.8. Let $n \in \mathbb{N}$. Then each permutation $\pi \in S_n$ can be written as a product of transpositions (2-cycles) (ij) with $i, j \in \{1, ..., n\}$ and we call

$$\operatorname{sign} \pi := (-1)^{\#\{\text{transpositions in } \pi\}}$$

the *sign* of π. We call π *even* (resp. *odd*) if $\operatorname{sign}(\pi) = 1$ (resp. $\operatorname{sign}(\pi) = -1$).

The *alternating group* A_n is defined as the group of all even permutations in S_n.

Theorem 4.2.9. *The alternating group A_n is simple for $n \geq 5$. In particular, the symmetric group S_n is not soluble for $n \geq 5$.*

Proof. Claim 1: A_n is generated by 3-cycles: Each permutation $\pi \in A_n$ can be written as product of an even number of 2-cycles, say

$$\pi = (a_1, b_1)(c_1, d_1)...(a_r, b_r)(c_r, d_r),$$

where we may always assume that $a_i \neq b_i$ and $c_i \neq d_i$ and the transpositions (a_i, b_i) and (c_i, d_i) are not identical (otherwise they would cancel and we wouldn't have needed them). As one easily computes, if a, b, c, d are all distinct, we have

$$(a, b)(c, d) = (a, c, b)(a, c, d)$$

and if a, b, c are all distinct, we have

$$(a, b)(b, c) = (a, b, c).$$

Therefore, each pair of transpositions $(a_i, b_i)(c_i, d_i)$ above can be replaced by either one or two 3-cycles. Therefore A_n is generated by 3-cycles.

Claim 2: All 3-cocycles are conjugate in A_n: Recall that the action of S_n on cycles by conjugation is simply given by

$$\pi^{-1}(a_1, ..., a_r)\pi = (\pi(a_1), ..., \pi(a_r)).$$

In particular, for any 3-cycle (a_1, a_2, a_3) in A_n, there is some $\pi \in S_n$ such that $(a_1, a_2, a_3) = \pi^{-1}(1, 2, 3)\pi$. If π is even, then $\pi \in A_n$ and we are done, if not then define $\tau = (4, 5)\pi$. Then τ is even and we have

$$\tau^{-1}(1, 2, 3)\tau = \pi^{-1}(4, 5)(1, 2, 3)(4, 5)\pi = \pi^{-1}(1, 2, 3)\pi = (a_1, a_2, a_3),$$

proving the claim.

Now let $\{1\} \neq N \trianglelefteq A_n$ be a normal subgroup. By Claim 2, it suffices to show that N contains a 3-cycle to show that $N = A_n$ and therefore that A_n is simple.

Case 1: Suppose that N contains an element of the form $\pi = (1, ..., r)\tau$ for some $r \geq 4$ and τ is a product of cycles containing only elements $> r$. For $\delta = (1, 2, 3)$ it follows that $\delta^{-1}\pi\delta \in N$ since $N \trianglelefteq A_n$ and therefore also

$$\pi^{-1}\delta^{-1}\pi\delta = (r, ..., 1)(1, 3, 2)(1, ..., r)(1, 2, 3) = (2, 3, r) \in N.$$

Therefore N contains a 3-cycle and we have $N = A_n$.

Case 2: Suppose N contains an element of the form $\pi = (1, 2, 3)(4, 5, 6)\tau$, where τ is a product of cycles containing only elements > 6. Then set $\delta = (1, 2, 4)$ and observe that by a similar computation as in Case 1 we find that

$$\pi^{-1}\delta^{-1}\pi\delta = (1, 3, 2)(4, 6, 5)(1, 4, 2)(1, 2, 3)(4, 5, 6)(1, 2, 4) = (1, 2, 3, 4, 5, 6) \in N,$$

so N contains a 3-cycle by Case 1.

Case 3: Suppose N contains an element of the form $(1,2,3)\tau$ where τ is a product of disjoint transpositions permuting elements ≥ 4. Then $\pi^2 = (1,3,2) \in N$ and we have found our 3-cycle.

Case 4: Suppose N contains an element of the form $\pi(1,2)(3,4)\tau$, where τ is a product of disjoint transpositions permuting elements ≥ 5. Then let $\delta_1 = (1,2,3)$ and $\delta_2 = (1,2,5)$. Then we have

$$\sigma_1 = \pi^{-1}\delta_1^{-1}\pi\delta_1 = (1,4)(2,3) \in N$$

and

$$\sigma_2 = \delta_2^{-1}\sigma\delta_2 = (1,3)(4,5) \in N.$$

Thus we also have $\sigma_1\sigma_2 = (1,2,3,4,5) \in N$ and we have a 3-cycle in N by Case 1.

Since it is clearly always possible to conjugate any element in A_n into one of those forms, it follows that any non-trivial normal subgroup of A_n contains a 3-cycle and therefore equals A_n, wherefore A_n is simple.

q.e.d.

4.3 Composition series and the Theorem of Jordan-Hölder*

The definition of a soluble group relies on the commutator series (see Definition 4.2.1). In this section we have a closer look at related series which we now define.

Definition 4.3.1. Let G be a group.

1. A collection of subgroups $G_0 = G, G_1, ..., G_r = \{1\}$ satisfying

$$G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq ... \trianglerighteq G_r = \{1\},$$

i.e. $G_i \trianglelefteq G_{i-1}$ for $i = 1, ..., r$ (but not necessarily $G_i \trianglelefteq G$), is called a *subnormal series* of G.

2. A subnormal series

$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq ... \trianglerighteq G_r = \{1\}$$

is called a *composition series* if all its *composition factors* G_{i-1}/G_i, $i = 1, ..., r$, are simple.

Note that we have already encountered this concept in Theorem 4.2.3 and that the commutator series is an example of a subnormal series. It is however not necessarily a composition series.

Example 4.3.2. 1. As we have seen in Example 4.2.2, the commutator series of the group S_4 is given by

$$S_4 \trianglerighteq A_4 \trianglerighteq V_4 \trianglerighteq \{1\}.$$

The factor groups are $S_4/A_4 \cong C_2$, $A_4/V_4 \cong C_3$, and V_4. The first two factors are simple, but V_4 is not simple, so it is not a composition series. We can however refine the subnormal series by introducing a non-trivial normal subgroup of V_4. Note that V_4 is abelian so every subgroup is normal and a non-trivial subgroup is cyclic of order 2. Thus a composition series of S_4 is given by

$$S_4 \trianglerighteq A_4 \trianglerighteq V_4 \trianglerighteq C_2 \trianglerighteq \{1\}$$

With composition factors isomorphic to $C_2,\ C_3,\ C_2,\ C_2$.

2. Let $n \geq 5$. By Theorem 4.2.9, the group A_n is simple, and it is a normal subgroup of index 2 in S_n. The factor group is therefore simple and

$$S_n \trianglerighteq A_n \trianglerighteq \{1\}$$

is a composition series for S_n for $n \geq 5$.

We now want to show that any group has essentially only one composition series. For this we need three preliminary results. The first of these is attributed to Emmy Noether.

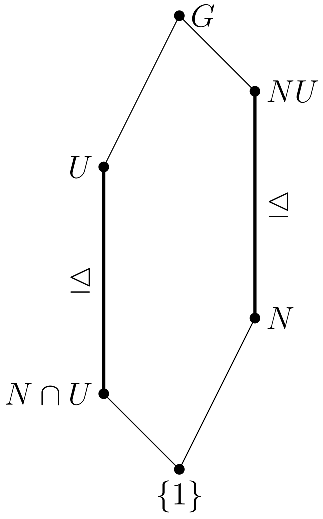
Theorem 4.3.3. *(Noether's Isomorphy Theorem) Let G be a group, $N \trianglelefteq G$ a normal subgroup and $U \leq G$ an arbitrary subgroup. Then the following are true.*

1. *The set $N \cdot U := \{n \cdot u \, : \, n \in N, \ u \in U\}$ is a subgroup of G, $NU \leq G$.*

2. *N is a normal subgroup of NU, $N \trianglelefteq NU$.*

3. *The group $N \cap U$ is a normal subgroup of U, $(N \cap U) \trianglelefteq U$.*

4. *The factor groups NU/N and $U/(N \cap U)$ are isomorphic, $NU/N \cong U/(N \cap U)$.*



Proof.

1. We clearly have $1 \in NU$. Now let $g_1 = n_1u_1, \ g_2 = n_2u_2 \in NU$, i.e. $n_1, n_2 \in N$ and $u_1, u_2 \in U$. It suffices to show that $g_1g_2^{-1} \in NU$ (exercise). We have
$$g_1g_2^{-1} = n_1u_1u_2^{-1}n_2^{-1} = n_1u_1u_2^{-1}n_2^{-1}(u_1u_2^{-1})^{-1}u_1u_2^{-1}.$$
Since N is a normal subgroup in G we have $u_1u_2^{-1}n_2^{-1}(u_1u_2^{-1})^{-1} \in N$ and hence also $n_1u_1u_2^{-1}n_2^{-1}(u_1u_2^{-1})^{-1} \in N$ and since U is a subgroup we have $u_1u_2^{-1} \in U$, so that indeed $g_1g_2^{-1} \in NU$.

2. Since N is a normal subgroup in G, it is also normal in any subgroup of G containing N, so in particular in NU.

3. Let $n \in N \cap U$ and $u \in U$. Then we have $unu^{-1} \in N$ because $n \in N$ and N is normal in G and also $unu^{-1} \in U$ because we assumed $n \in U$. Therefore $unu^{-1} \in N \cap U$, wherefore $N \cap U \trianglelefteq U$ as claimed.

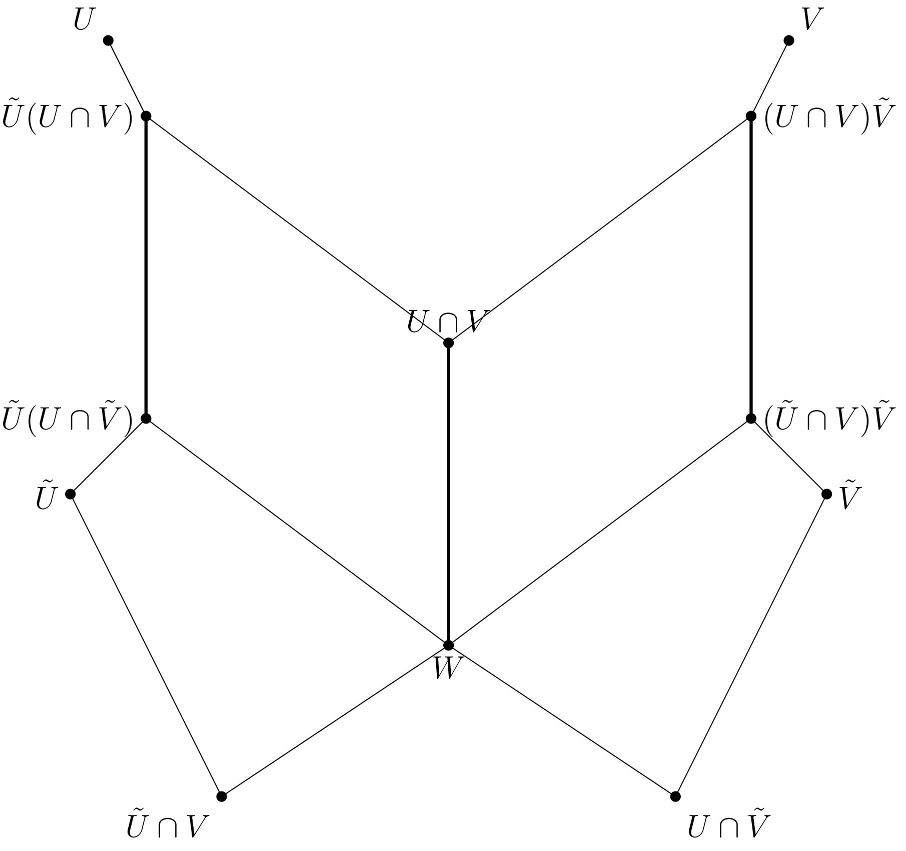
4. Consider the map $\varphi : U \to G/N, \ u \mapsto uN$, the restriction of the canonical epimorphism to U. Then we have $\mathrm{Im}(\varphi) = NU/N$, since for any $nu \in NU$ we have $nuN = uu^{-1}nuN = uN = \varphi(u)$, and clearly $\mathrm{Ker}(\varphi) = N \cap U$. Therefore it follows from the Homomorphy Theorem 4.1.10 that
$$NU/N = \mathrm{Im}(\varphi) \cong U/\mathrm{Ker}(\varphi) = U/(N \cap U).$$

q.e.d.

The next result one was discovered by Zassenhaus.

Theorem 4.3.4. *(Butterfly lemma) Let G be a group and $U, \tilde{U}, V, \tilde{V}$ be subgroups, such that $\tilde{U} \trianglelefteq U$ and $\tilde{V} \trianglelefteq V$. Then we have*

$$\tilde{U}(U\cap V)\big/\tilde{U}(U\cap\tilde{V}) \cong (U\cap V)\tilde{V}\big/(\tilde{U}\cap V)\tilde{V}.$$



$$W = (\tilde{U}\cap V)(U\cap\tilde{V})$$

Proof. By Noether's Isomorphy Theorem 4.3.3 we know that $\tilde{U}\cap V$ and $U\cap\tilde{V}$ are both normal subgroups of $U\cap V$ and that we have isomorphisms

$$\varphi : (U\cap V)/(\tilde{U}\cap V) \to (\tilde{U}(U\cap V))/\tilde{U}, \psi : \ (U\cap V)/(U\cap\tilde{V}) \to ((U\cap V)\tilde{V})/\tilde{V}.$$

We now claim that $W = (\tilde{U}\cap V)(U\cap\tilde{V})$ is normal in $U\cap V$. Let $w = w_1w_2 \in W$ with $w_1 \in \tilde{U}\cap V$ and $w_2 \in U\cap\tilde{V}$ and $g \in U\cap V$. Then we have $gwg^{-1} = (gw_1g^{-1})(gw_2g^{-1})$ and since $\tilde{U}\cap V$ and $U\cap\tilde{V}$ are both normal in $U\cap V$ it follows that $gw_1g^{-1} \in \tilde{U}\cap V$ and $gw_2g^{-1} \in U\cap\tilde{V}$, wherefore $gwg^{-1} \in W$, whence W is indeed normal in $U\cap V$. It follows that

$$\varphi(W/(\tilde{U}\cap V)) = \tilde{U}W/\tilde{U} = \tilde{U}(U\cap\tilde{V})/\tilde{U}$$

and

$$\psi(W/(U \cap \tilde{V})) = W\tilde{V}/\tilde{V} = (\tilde{U} \cap V)\tilde{V}/\tilde{V}.$$

Therefore we have found that φ induces an isomorphism

$$(U \cap V)/W \cong ((U \cap V)/(\tilde{U} \cap V))/(W/(\tilde{U} \cap V)) \cong (\tilde{U}(U \cap V))/(\tilde{U}(U \cap \tilde{V})),$$

where the first isomorphism is justified by Corollary 4.1.11.

Similarly, ψ induces an isomorphism

$$((U \cap V)\tilde{V})/((\tilde{U} \cap V)\tilde{V}) \cong (U \cap V)/W,$$

from where the claim follows.

q.e.d.

We now apply the Butterfly Lemma to show the following important step towards the main theorem of this section.

Theorem 4.3.5. *(Refinement Theorem of Schreier-Zassenhaus) Let G be a group and suppose we have two subnormal series*

$$G = G_0 \trianglerighteq G_1 \trianglerighteq ... \trianglerighteq G_r = \{1\} \quad \textit{and} \quad G = H_0 \trianglerighteq H_1 \trianglerighteq ... \trianglerighteq H_s = \{1\}.$$

Define the groups $G_{i,j} := G_i(G_{i-1} \cap H_j)$ and $H_{i,j} := H_j(H_{j-1} \cap G_i)$. Then we obtain the following refinements of the two subnormal series above,

$$G = G_{1,0} \trianglerighteq G_{1,1} \trianglerighteq ... \trianglerighteq G_{1,s} = G_1 = G_{2,0} \trianglerighteq G_{2,1} \trianglerighteq ... \trianglerighteq G_{r,s} = \{1\}$$

and

$$G = H_{0,1} \trianglerighteq H_{1,1} \trianglerighteq ... \trianglerighteq H_{r,1} = H_1 = H_{0,2} \trianglerighteq H_{1,2} \trianglerighteq ... \trianglerighteq H_{r,s} = \{1\},$$

where we have that

$$G_{i,j-1}/G_{i,j} \cong H_{i-1,j}/H_{i,j}.$$

In particular, all the factor groups obtained from both refined subnormal series are pairwise isomorphic afer reordering.

Proof. As we saw in the proof of the Butterfly Lemma 4.3.4, we do indeed have $G_{i,j} \trianglerighteq G_{i,j+1}$ and $H_{i,j} \trianglerighteq H_{i+1,j}$. Note also that $H_s = G_r = \{1\}$, wherefore we have $G_{i,s} = G_i(G_{i-1} \cap \{1\}) = G_i$ and $G_{i+1,0} = G_{i+1}(G_i \cap H_0) = G_i$ and similarly $H_{r,j} = H_j = H_{0,j+1}$, so the refined series are in fact refinements of the original ones.

Looking at the factor groups we have

$$G_{i,j-1}/G_{i,j} = G_i(G_{i-1} \cap H_{j-1})/(G_i(G_{i-1} \cap H_j))$$

and

$$H_{i-1,j}/H_{i,j} = H_j(G_{i-1} \cap H_{j-1})/(H_j(G_i \cap H_{j-1}))$$

and these groups are isomorphic according to the Butterfly Lemma 4.3.4.

q.e.d.

We are now ready to prove the announced result, the Theorem of Jordan-Hölder, which states that each group has in a sense just one composition series. Recall from Definition 4.3.1 that a composition series of a group is a subnormal series where all the factor groups are simple groups.

Theorem 4.3.6. *(Jordan-Hölder) Let G be a group a with composition series*

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \dots \trianglerighteq G_r = \{1\} \quad \textit{and} \quad G = H_0 \trianglerighteq H_1 \trianglerighteq \dots \trianglerighteq H_s = \{1\}$$

such that $G_i \neq G_{i+1}$ and $H_j \neq H_{j+1}$. Then we have $r = s$ and there is some permutation $\pi \in S_r$ such that the factor groups G_i/G_{i+1} and $H_{\pi(i)}/H_{\pi(i)+1}$ are isomorphic.

Proof. We apply the Refinement Theorem of Schreier-Zassenhaus 4.3.5 to the two composition series. Since all the factor groups G_i/G_{i+1} and H_j/H_{j+1} are simple, all of the groups in the refined series must satisfy $G_{i,j} = G_i$ or $G_{i,j} = G_{i+1}$ (resp. $H_{i,j} = H_j$ or $H_{i,j} = H_{j+1}$ for all i, j. This shows that after elimintaing the duplicate groups in the refined subnormal series, we must have $r = s$. The isomorphy of the composition factors follows from Theorem 4.3.5 again, since the only non-trivial factor groups in the refined series are those which occur in the original composition series.

q.e.d.

Remark 4.3.7. *It is because of the Theorem of Jordan-Hölder that one sometimes makes the comparison of simple groups to primes. In a way, all groups can be built from simple groups in an essentially unique way. The analogy is however not perfect, since it can happen that two non-isomorphic groups have isomorphic composition factors. For instance the groups S_5 and $A_5 \times C_2$ have the same order and isomorphic composition factors (namely A_5 and C_2), but the two groups are not isomorphic, which can be seen for instance by noting that $A_5 \times C_2$ contains an element of order 10 and S_5 does not.*

Chapter 5

Galois Theory

In this chapter we collect the results both on fields and groups we have collected so far to introduce some of the main results of Galois Theory, which aims to understand field extensions by understanding subgroups of the so-called *Galois group*.

5.1 Galois extensions

5.1.1 Galois groups

Recall the following definitions and results from Chapter 3:

A field extension E/K is called *normal* if every K-homomorphism $\varphi : E \to \overline{K}$ for an algebraic closure $\overline{K}$ of K satisfies $\varphi(E) = E$ (see Definition 3.6.1), or equivalently, if the minimal polynomial of any element $\alpha \in E$ decomposes into linear factors in $E[X]$ (Theorem 3.6.3).

A field extension E/K is called *separable* if the minimal polynomial $\mu_\alpha \in K[X]$ of any element $\alpha \in E$ decomposes into distinct linear factors in its splitting field (or in $\overline{K}$) (Definition 3.5.1), or equivalently if $\gcd(\mu_\alpha, \mu_\alpha') = 1$ (Lemma 3.4.4).

We begin by noting the following.

Lemma 5.1.1. *A finite field extension E/K is separable if and only if $[E : K] = \#\{\varphi : E \to \overline{K}\ K\text{-homomorphism}\}$.*

Proof. Let $\alpha \in E$ and $\mu_\alpha \in K[X]$ be its minimal polynomial over K. Then any K-homomorphism $\varphi : K(\alpha) \to \overline{K}$ satisfies $\mu_\alpha(\varphi(\alpha)) = 0$ (which follows essentially from Theorem 3.2.7). On the other hand, if $\beta \in \overline{K}$ is any root of μ_α, then we obtain a K-isomorphism $K(\alpha) \cong_K K(\beta) \subseteq \overline{K}$, so that any root of μ_α. Therefore we have

$$\#\{\varphi : K(\alpha) \to \overline{K}\ K\text{-homomorphism}\} = \#\{\beta \in \overline{K}\ :\ \mu_\alpha(\beta) = 0\}.$$

The right-hand side now equals the degree of μ_α and thus $[K(\alpha):K]$ if and only if μ_α is separable, which is true if and only if $K(\alpha)/K$ is separable.

Iterating this argument with $K(\alpha)$ instead of K yields the claim for E.

q.e.d.

As an immediate corollary we obtain the following (recall the definition of $\mathrm{Aut}_K(E)$ in Definition 3.2.3).

Corollary 5.1.2. *Let E/K be a finite field extension. Then we have $\#\,\mathrm{Aut}_K(E) = [E:K]$ if and only if E/K is both normal and separable.*

Proof. By Lemma 5.1.1, we have $[E:K] = \#\{\varphi : E \to \overline{K}\ K\text{-homomorphism}\}$ if and only if E/K is separable and by definition we have

$$\{\varphi : E \to \overline{K}\ K\text{-homomorphism}\} = \mathrm{Aut}_K(E)$$

if and only if E/K is normal.

q.e.d.

We now come to the central definition of this chapter.

Definition 5.1.3. Let K be a field.

1. For $G \leq \mathrm{Aut}(K)$ we set

$$K^G := \{a \in K \ :\ \sigma(a) = a \text{ for all } \sigma \in G\}$$

the *fixed subfield* of G.

2. A field extension E/K is called *Galois* if there is a finite subgroup $G \leq \mathrm{Aut}(E)$ such that $K = E^G$. In this case we call $G =: \mathrm{Gal}(E/K)$ the *Galois group* of E/K.

The concepts of Galois extensions and Galois groups is named after the 19th century French mathematician Évariste Galois.

Remark 5.1.4. *Note that K^G is indeed a subfield of K by the same argument as in Lemma 3.2.6.*

Example 5.1.5. 1. Let $E = \mathbb{C}$ and $K = \mathbb{R}$. Then $\mathbb{R}$ is the fixed subfield of the automorphism of complex conjugation $z \mapsto \overline{z}$ in $\mathbb{C}$, wherefore $\mathbb{C}/\mathbb{R}$ is a Galois extension with Galois group

$$\mathrm{Gal}(\mathbb{C}/\mathbb{R}) = \langle z \mapsto \overline{z} \rangle \cong C_2.$$

2. Let $E = \mathbb{Q}(\sqrt[3]{2})$ and $K = \mathbb{Q}$. Then any automorphism of E has to fix $\mathbb{Q}$ since $\mathbb{Q}$ is the prime field contained in E. In particular it is uniquely determined by its image of the generator $\sqrt[3]{2}$, which must be a root of the minimal polynomial $X^3 - 2$ in E. But $\sqrt[3]{2}$ is the only root of $X^3 - 2$ in E, so we have $\mathrm{Aut}(E) = \{\mathrm{id}\}$ is trivial. In particular there is no subgroup of $\mathrm{Aut}(E)$ whose fixed subfield is $\mathbb{Q}$, so that $E/\mathbb{Q}$ is *not* Galois. Consider on the other hand the field $L = \mathbb{Q}(\sqrt[3]{2}, \zeta_3)$ and the extension $L/\mathbb{Q}$. Then L is the splitting field of $X^3 - 2$ and admits several automorphsims, which we may define on the generators by $\sigma : \sqrt[3]{2} \mapsto \zeta_3\sqrt[3]{2}$ and $\tau : \zeta_3 \mapsto \zeta_3^{-1}$. Note that τ is just complex conjugation. It turns out that these generate the automorphisms of L and one checks without too much difficulty that $E = L^{\langle\tau\rangle}$ and $\mathbb{Q} = L^{\langle\sigma,\tau\rangle}$, so the extension L/E is Galois with $\mathrm{Gal}(L/E) = \langle\tau\rangle \cong C_2$ and $L/\mathbb{Q}$ is Galois as well with Galois group $\langle\sigma,\tau\rangle \cong S_3$.

3. It follows directly from Theorem 3.4.5 and Remark 3.4.6 that a finite extension E/K of a finite field K is Galois, where the Galois group is generated by the Frobenius automorphism (see Definition 3.4.3).

We first want establish a description of Galois extensions which usually makes it not too hard to recognize them. Before that, we need the following result.

Proposition 5.1.6. *Let E be a field, $G \leq \mathrm{Aut}(E)$ a finite group, and $K := E^G$ the fixed field of G. Then we have*

$$[E : K] = \#G.$$

Proof. It follows exactly as in the proof of Lemma 5.1.1 that $\#G \leq [E : K]$, so we are left with proving that $\#G \geq [E : K]$. For this assume $\#G = n$, $G = \{\sigma_1, ..., \sigma_n\}$ and let $\alpha_1, ..., \alpha_{n+1} \in E$ be arbitrary. We want to show that $\alpha_1, ..., \alpha_{n+1}$ are linearly dependent over K, thus showing that $[E : K] \leq n$. To see this consider the n linear equations over E,

$$\sum_{j=1}^{n+1} \sigma_i(\alpha_j)u_j = 0, \quad i \in \{1, ..., n\}. \tag{5.1}$$

Since this is an underdetermined homogeneous linear systems, there must be a nontrivial solution $(u_1, ..., u_{n+1}) \in E^{n+1}$, where not all u_j are zero. By reordering, we may assume without loss of generality that $u_1, ..., u_r \neq 0$ and $u_{r+1} = ... = u_{n+1} = 0$ and we may choose r minimal with this property. Furthermore we may multiply our solution by any constant in E, so we may further assume that $u_1 = 1$.

Now let $\tau \in G$. It follows that for any $i \in \{1, ..., n\}$ we have

$$0 = \tau\left(\sum_{j=1}^{n+1} \sigma_i(\alpha_j)u_j\right) = \sum_{j=1}^{n+1}(\tau\sigma_i)(\alpha_j)\tau(u_j).$$

Note that τ just permutes the σ_i, so that $(\tau(u_1), ..., \tau(u_{n+1})) \in E^{n+1}$ is also a solution of the linear system in (5.1). But since we assumed $u_1 = 1$, we have $\tau(u_1) = 1$, so that

$$(u_1, ..., u_{n+1}) - (\tau(u_1), ...\tau(u_{n+1})) = (0, u_2 - \tau(u_2), ..., u_{n+1} - \tau(u_{n+1})) \in E^{n+1}$$

is a solution with at most $r - 1$ non-zero entries, so by our assumption on r, it must be the trivial solution. Therefore we find that $\tau(u_j) = u_j$ for all j and $\tau \in G$, which by definition ensures $u_j \in K$.

Therefore we have found our desired linear dependence $\sum_{j=1}^{n+1} u_j \alpha_j = 0$ over K and we find $\dim_K(E) \leq n$, as we wanted to show.

q.e.d.

With this result we can now show the following classification of Galois extensions. In most examples this classification is used to show that an extension is Galois.

Theorem 5.1.7. *A field extension E/K is Galois if and only if it is finite, normal, and separable. In this case we have* $\mathrm{Gal}(E/K) = \mathrm{Aut}_K(E)$.

Proof. Fist let E/K be finite, normal, and separable. Then we know by Corollary 5.1.2 that $G = \mathrm{Aut}_K(E) \leq \mathrm{Aut}(E)$ has order $\#G = [E : K]$. Consider the field $E^G \leq E$. By definition, every $\sigma \in G$ fixes K, so K is a subfield of E^G. On the other hand, each $\sigma \in G$ defines an E^G-automorphism of E, so, since E/E^G is still normal and separable, we have $\#G = [E : E^G] = [E : K]$, so that $E^G = K$, making the extension E/K Galois with Galois group $\mathrm{Gal}(E/K) = \mathrm{Aut}_K(E)$.

Now suppose that E/K is a (finite) Galois extension with Galois group $G \leq \mathrm{Aut}(E)$, i.e. $K = E^G$. We know by Proposition 5.1.6 that $[E : K] = \#G$. Since G is certainly a subgroup of $\mathrm{Aut}_K(E)$ and we know that $\#\mathrm{Aut}_K(E) \leq [E : K]$ by Lemma 5.1.1, we find that $G = \mathrm{Aut}_K(E)$ and $[E : K] = \#G = \#\mathrm{Aut}_K(E)$, whence by Corollary 5.1.2 the extension E/K is both normal and separable.

q.e.d.

5.1.2 The Main Theorem of Galois Theory

We now have all the necessary tools to prove the Main Theorem of Galois Theory.

Definition 5.1.8.

1. For a group G let $\mathcal{U}(G) := \{U \leq G\}$ denote the collection of all subgroups of G.

2. For a field extension E/K let $\mathcal{F}(E/K) := \{L \leq E : K \leq L\}$ the collection of all subfields of E containing K or *intermediate fields* of E/K.

Note that both sets $\mathcal{U}(G)$ and $\mathcal{F}(E/K)$ are *partially ordered* by inclusion.

Remark 5.1.9. *Let E/K be a Galois extension with Galois group G. Then G acts on $\mathcal{U}(G)$ by conjugation, and on $\mathcal{F}(E/K)$ by application, i.e. for any $U \leq G$ and $\sigma \in G$ we have $\sigma U \sigma^{-1} := \{\sigma u \sigma^{-1} : u \in U\} \leq G$ and for $L \in \mathcal{F}(E/K)$ we have $\sigma(L) \in \mathcal{F}(E/K)$.*

Theorem 5.1.10. *(Main Theorem of Galois Theory) Let E/K be a Galois extension with Galois group G. Consider the map*

$$\Phi : \mathcal{U}(G) \to \mathcal{F}(E/K),\ U \mapsto E^U.$$

Then the following are true.

1. *Φ is bijective with inverse map*

$$\Psi : \mathcal{F}(E/K) \to \mathcal{U}(G),\ L \mapsto \mathrm{Aut}_L(E).$$

2. *Φ is inclusion inverting, i.e. if $U \leq V$, then we have $\Phi(V) \leq \Phi(U)$.*

3. *Φ is G-equivariant, i.e. for any $\sigma \in G$ and $U \in \mathcal{U}(G)$ we have $\Phi(\sigma U \sigma^{-1}) = \sigma(\Phi(U))$.*

Proof.

1. We see easily that Φ is a well-defined map, since for a subgroup $U \leq G$, the fixed field E^U is a subfield of E and since $K = E^G$, we have in particular $\sigma(a) = a$ for all $a \in K$ and $\sigma \in U$, so K is a subfield of E^U. Similarly, the map Ψ is well-defined since for $L \in \mathcal{F}(E/K)$ each $\sigma \in \mathrm{Aut}_L(E)$ satisfies by definition $\sigma(\alpha) = \alpha$ for all $\alpha \in L$, so in particular $\sigma(a) = a$ for all $a \in K \leq L$. Hence $\mathrm{Aut}_L(E) \leq \mathrm{Aut}_K(E) = G$.

   We now need to verify that Φ and Ψ are inverses of one another. Let $U \in \mathcal{U}(G)$. Then E/E^U is a Galois extension and by Theorem 5.1.7, we have

$$U = \mathrm{Gal}(E/E^U) = \mathrm{Aut}_{\Phi(U)}(E) = \Psi(\Phi(U)).$$

   Now let $L \in \mathcal{F}(E/K)$ be an intermediate field of E/K. Then the extension E/L is Galois with Galois group $\Psi(L) = \mathrm{Aut}_L(E)$, and therefore

$$F = E^{\Psi(F)} = \Phi(\Psi(F)).$$

   Thus Ψ is an inverse of Φ which therefore is bijective as claimed.

2. Now let $U \leq V \leq G$. It is clear that any element of E that is fixed by V is in particular fixed by U, so that indeed $\Phi(V) = E^V \leq E^U = \Phi(U)$.

3. Let $\sigma \in G$ and $U \in \mathcal{U}(G)$. Then we have $\Phi(\sigma U \sigma^{-1}) = E^{\sigma U \sigma^{-1}}$ and we find that $\alpha \in E^{\sigma U \sigma^{-1}}$ if and only if $\sigma u \sigma^{-1}(\alpha) = \alpha = \sigma(\sigma^{-1}(\alpha))$ for all $u \in U$, which is equivalent to $u(\sigma^{-1}(\alpha)) = \sigma^{-1}(\alpha)$ for all $u \in U$, i.e. $\sigma^{-1}(\alpha) \in E^U$ or in other words $\alpha \in \sigma(E^U)$. Therefore we have indeed
$$\Phi(\sigma U \sigma^{-1}) = \sigma(\Phi(U))$$
as claimed.

q.e.d.

We can say the following about the action of Galois groups on the roots of polynomials.

Proposition 5.1.11. *Let E/K be a Galois extension with Galois group G and choose $\alpha \in E$. Then for every root β of $\mu_{\alpha,K}(X) \in K[X]$ there exists $\sigma \in G$ such that $\sigma(\alpha) = \beta$. Furthermore, we have that the* stabilisor subgroup $\mathrm{Stab}_G(\alpha) := \{\sigma \in G \,:\, \sigma(\alpha) = \alpha\}$ *is the Galois group of $E/K(\alpha)$.*

Proof. Since E/K is Galois and hence in particular normal, all roots of μ_α lie in E. Furthermore we know by the definition of normality that we can extend the K-homomorphism $K(\alpha) \to E$ sending α to a different root β to a K-automorphism of E, i.e. an element in $\mathrm{Gal}(E/K)$. Since clearly $\Phi(\mathrm{Stab}_G(\alpha)) = K(\alpha)$, the second claim follows from the Main Theorem 5.1.10.

q.e.d.

Using the insight from the previous proposition we can also verify the following result.

Proposition 5.1.12. *Let E/K be a Galois extension with Galois group $G =$* $\mathrm{Gal}(E/K)$ *and assume the notation from Theorem 5.1.10. Then the extension L/K for a field $L \in \mathcal{F}(E/K)$ is normal if and only if $\Psi(L) =$* $\mathrm{Aut}_L(E)$ *is a normal subgroup of G. In this case we have* $\mathrm{Gal}(L/K) \cong G/\Psi(L)$.

Proof. Exercise.

q.e.d.

Let us now consider some examples how we can use the Main Theorem to determine all intermediate fields of a given Galois extension.

Example 5.1.13. Let $K = \mathbb{Q}$ and E be the splitting field of the polynomial $X^4 - 2$. As we have seen in Example 3.2.2 we have $E = \mathbb{Q}(\sqrt[4]{2}, \mathrm{i})$ and $[E : K] = 8$.

Since E is the splitting field of a polynomial over a perfect field, E/K is normal and separable and hence Galois with Galois group G. Since G permutes the 4 roots of

$$X^4 - 2 = (X - \sqrt[4]{2})(X - \mathrm{i}\,\sqrt[4]{2})(X + \sqrt[4]{2})(X + \mathrm{i}\,\sqrt[4]{2}) \in E[X],$$

G must be a subgroup of S_4 of order 8. The two endomorphisms of E defined by

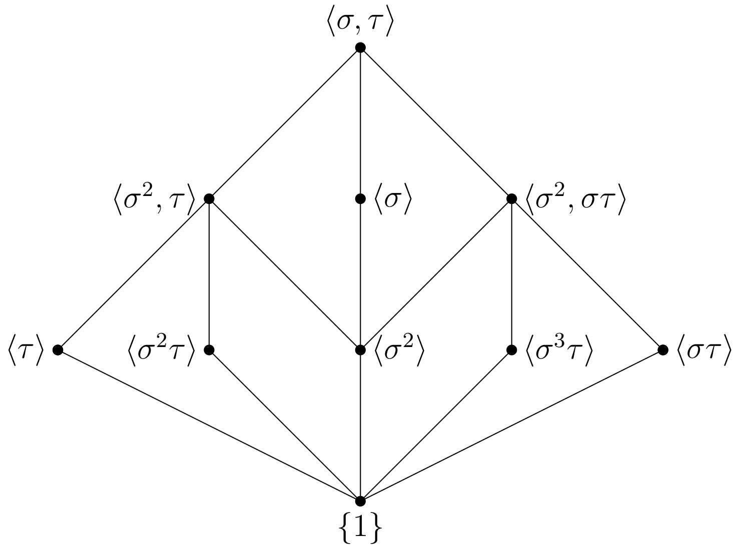
$$\sigma : \begin{cases} \sqrt[4]{2} \mapsto \mathrm{i}\,\sqrt[4]{2} \\ \mathrm{i} \mapsto \mathrm{i} \end{cases} \qquad \text{and} \qquad \tau : \begin{cases} \sqrt[4]{2} \mapsto \sqrt[4]{2} \\ \mathrm{i} \mapsto -\,\mathrm{i} \end{cases}$$

clearly fix $\mathbb{Q}$, so we have $\sigma, \tau \in G$. We now label the roots of $X^4 - 2$ by the numbers $1, 2, 3, 4$ in the order listed above and realise σ and τ as permutations represented by cycles. We find that

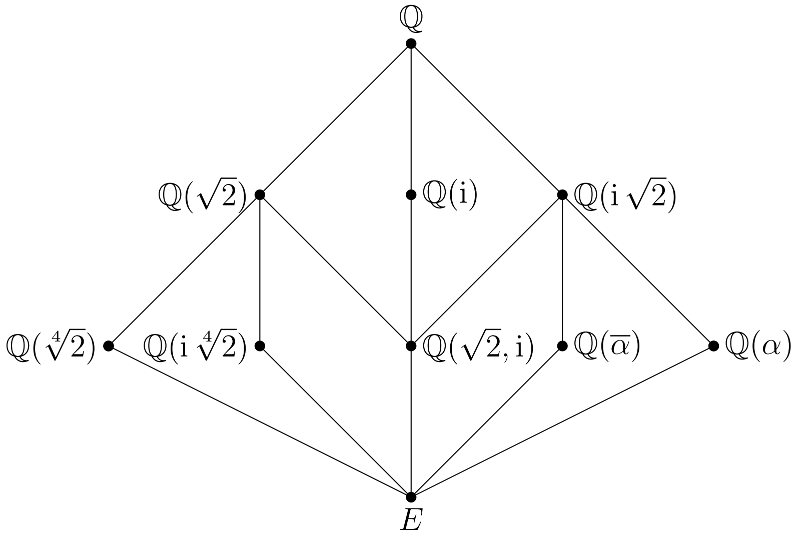
$$\sigma \longleftrightarrow (1, 2, 3, 4) \qquad \text{and} \qquad \tau \longleftrightarrow (2, 4).$$

As one sees without difficulty, these two permutations generate the group D_4 introduced in Example 4.2.5, which has order 8, so we have $G \cong D_4$.

The subgroups of D_4 are given in the following diagram, where larger groups are towards the top. The relative indices of each subgroup in the next larger is always 2.


$\langle \sigma, \tau \rangle$
$\langle \sigma^2, \tau \rangle$
$\langle \sigma \rangle$
$\langle \sigma^2, \sigma\tau \rangle$
$\langle \tau \rangle$
$\langle \sigma^2\tau \rangle$
$\langle \sigma^2 \rangle$
$\langle \sigma^3\tau \rangle$
$\langle \sigma\tau \rangle$
$\{1\}$

By the Main Theorem, we can find the intermediate fields of $E/\mathbb{Q}$ as the fixed subfields of each of the subgroups, where the inclusions are reversed. We therefore obtain the following diagram, where now the larger fields are at the bottom.


$\mathbb{Q}$
$\mathbb{Q}(\sqrt{2})$
$\mathbb{Q}(\mathrm{i})$
$\mathbb{Q}(\mathrm{i}\sqrt{2})$
$\mathbb{Q}(\sqrt[4]{2})$
$\mathbb{Q}(\mathrm{i}\sqrt[4]{2})$
$\mathbb{Q}(\sqrt{2},\mathrm{i})$
$\mathbb{Q}(\overline{\alpha})$
$\mathbb{Q}(\alpha)$
E

In the above diagram, we have used the shorthands $\alpha = (1+\mathrm{i})\sqrt[4]{2}$ and an overline to denote complex conjugation.

We see that quite a few properties of a Galois extension depend on properties of the associated Galois group. This motivates the following definition.

Definition 5.1.14. Let E/K be a Galois extension with Galois group $G := \mathrm{Gal}(E/K)$. We say that E/K is *abelian* resp. *cyclic* if G has that property.

5.2 The Normal Basis Theorem

Recall the Primitive Element Theorem 3.5.6: This implies that for any finite, separable field extension E/K there is an element $\alpha \in E$ such that $E = K(\alpha)$, or in other words if $[E : K] = n$, then the set $\{1, \alpha, \alpha^2, ..., \alpha^{n-1}\}$ forms a basis of E as a K-vector space. In this section, we want to show a related theorem for Galois extensions, where instead of powers of a single element we take Galois conjugates of a single element. A basis obtained in this way is called a *normal basis*.

As for the primitive element theorem we treat finite fields separately. Recall from Example 5.1.5, that any extension of finite fields is a Galois extension where the Galois group is cyclic and generated by the Frobenius automorphism.

Theorem 5.2.1. *(Normal basis Theorem (finite fields))*
Let $q = p^n$ be a prime power, $K = \mathbb{F}_q$ be a finite field, and $E = \mathbb{F}_{q^m}$ a finite extension of K. Further let $\Phi = \Phi_q : E \to E$, $a \mapsto a^q$ denote the Frobenius automorphism of E. Then there exists an element $\beta \in E$ such that the set

$$\{\Phi^j(\beta) \,:\, 0 \leq j < m\} = \{\beta, \beta^q, \beta^{q^2}, ..., \beta^{q^{m-1}}\}$$

forms a K-basis of E.

Proof. Each map $\Phi^j : E \to E$ may be viewed as a group homomorphism of $E^\times$ to itself, a so-called *character*. An important theorem of Artin (see Theorem 5.4.2) implies that the maps $\Phi^0 = \mathrm{id}, \Phi, ..., \Phi^{m-1} : E \to E$ are linearly independent over E and hence also over K as K-linear endomorphisms of E as a K-vector space. Since $\Phi^m = \mathrm{id}$, it follows that the minimal polynomial of Φ as a K-vector space endomorphism if $X^m - 1 \in K[X]$.

Denote by $K[G]$ the K-vector space spanned by the K-linear maps $\Phi^j : E \to E$. This naturally becomes a module over the polynomial ring $K[X]$ by defining the multiplication $X.\Phi^j := \Phi^{j+1}$. Since the minimal polynomial of Φ is $X^m - 1$, we therefore see that

$$K[G] \cong K[X]/(X^m - 1)$$

as $K[X]$-modules and hence as K-vector spaces.

Similarly, E becomes a $K[X]$-module via $X.\alpha = \Phi(\alpha)$. Since any finite $K[X]$-module is isomorphic to a direct sum $\bigoplus_i K[X]/(f_i)$ for some (unique) polynomials f_i of positive degree such that $f_i \mid f_{i+1}$ (known as the Elementary Divisors Theorem) and E is clearly annihilated by the minimal polynomial of Φ, it follows for reasons of dimension that $E \cong K[X]/(X^m - 1)$ as $K[X]$-modules. Note that this is **not** an isomorphism of rings or K-algebras. It follows therefore that

$$K[G] \cong K[X]/(X^m - 1) \cong E$$

as $K[X]$-modules. Under this isomorphism the K-basis $\{1, X, ..., X^{m-1}\}$ of the middle module is mapped to a K-basis of E with the desired property.

q.e.d.

Even though the above proof does offer a way to find a normal basis for a finite field, it is a bit implicit. In some cases however it is not so hard to produce such a basis anyway.

Example 5.2.2. Consider the field $\mathbb{F}_8 = \mathbb{F}_2(\alpha)$ over $\mathbb{F}_2$, where $\alpha^3+\alpha+1 = 0$. As a first attempt to find a normal basis of $\mathbb{F}_8$, one may consider $\{\alpha, \Phi(\alpha) = \alpha^2, \Phi^2(\alpha)\}$, but since $\Phi^2(\alpha) = \alpha^4 = \alpha^2 + \alpha$, this is linearly dependent and hence not a basis. However one finds that

$$\{\alpha + 1, \Phi(\alpha + 1) = \alpha^2 + 1, \Phi^2(\alpha) = \alpha^4 + 1 = \alpha^2 + \alpha + 1\}$$

is linearly independent and thus forms a normal basis.

For infinite fields, it is also possible to find such a basis and the proof is slightly more explicit.

Theorem 5.2.3. *(Normal Basis Theorem (infinite fields))*
Let K be an infinite field and E/K be a finite Galois extension of degree $[E : K] = n$ with Galois group G. Then there exists an element $\beta \in E$ such that the set $\{\sigma(\beta) : \sigma \in G\}$ forms a K-basis of E.

Proof. Since E/K is a Galois extension, it is in particular a separable extension by Theorem 5.1.7, so by the Primitive Element Theorem 3.5.6 there exists some $\alpha \in E$, such that $\{1, \alpha, ..., \alpha^{n-1}\}$ forms a K-basis of E. Let $\mu_\alpha \in K[X]$ denote the minimal polynomial of α over K and write

$$\mu_\alpha = \prod_{i=1}^{n}(X - \alpha_i) \in E[X],$$

where $\alpha_1 = \alpha$ and $\sigma_i \in G$ is determined by $\sigma_i(\alpha) = \alpha_i$. So in particular $\sigma_1 = 1 \in G$. Define the polynomials

$$g = \prod_{i=2}^{n} \frac{X - \alpha_i}{\alpha - \alpha_i} \in E[X]$$

and

$$g_j = \sigma_j(g) = \prod_{\substack{i=1 \\ i \neq j}}^{n} \frac{X - \alpha_i}{\alpha_j - \alpha_i} \in E[X].$$

and consider the matrix $G \in E[X]^{n \times n}$ with

$$G_{ij} = \sigma_i(\sigma_j(g)).$$

Then clearly $G_{ij} = g_k$ where k is determined by $\sigma_k = \sigma_i\sigma_j \in G$. Since by construction we have

$$g_i(\alpha_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

we find that $G(\alpha) \in E^{n \times n}$ is a permutation matrix, i.e. an invertible matrix such that each column has exactly one entry equal to 1 and all others equal 0. Such a matrix has determinant ± 1. This implies that the determinant $D \in E[X]$ of the polynomial matrix G is not the zero polynomial. Since K is infinite, there must be some element $a \in K$ such that $D(a) \neq 0$. Define $\beta := g(a)$ and $\beta_i = g_i(a) = \sigma_i(\beta)$ (so $\beta = \beta_1$). We claim that the set $B = \{\beta_1, ..., \beta_n\}$ forms a K-basis of E, which would be what we want to show. It suffices of course to show that B is linearly independent over K, so suppose there are $a_1, ..., a_n \in K$ such that

$$\sum_{j=1}^{n} a_j\beta_j = 0.$$

It follows that for any $\sigma_i \in G$ we have

$$0 = \sigma_i \left(\sum_{j=1}^{n} a_j \beta_j \right) = \sum_{j=1}^{n} a_j \sigma_i(\sigma_j(g(a)))$$

for all i. In other words we see that

$$G(a) \cdot (a_1, ..., a_n)^{tr} = 0.$$

But since $\det G(a) = D(a) \neq 0$, this is only possible if $a_1 = ... = a_n = 0$, so the claim follows.

q.e.d.

Example 5.2.4. The extension $E = \mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}$ is Galois, since E is easily seen to be the splitting field of $f = X^4 - 10X^2 + 1 \in \mathbb{Q}[X]$, which factors as

$$(X - (\sqrt{2} + \sqrt{3}))(X - (\sqrt{2} - \sqrt{3}))(X - (-\sqrt{2} + \sqrt{3}))(X - (-\sqrt{2} - \sqrt{3})) \in E[X].$$

It is also not hard to see that $\alpha = \sqrt{2} + \sqrt{3}$ is a primitive element of E and f is the minimal polynomial of α. A little computation yields, in the notation of the proof of Theorem 5.2.3,

$$g = \frac{1}{96} \left[(5\alpha^3 - 49\alpha)X^3 + (\alpha^2 - 5)X^2 + (-49\alpha^3 + 485\alpha)X + (-5\alpha^2 + 49) \right].$$

The matrix G is given by

$$G = \begin{pmatrix} g_1 & g_2 & g_3 & g_4 \\ g_2 & g_1 & g_4 & g_3 \\ g_3 & g_4 & g_1 & g_2 \\ g_4 & g_3 & g_2 & g_1 \end{pmatrix},$$

where g_2 is obtained by replacing α in g by $\alpha_2 = \alpha^3 - 10\alpha = \sqrt{2} - \sqrt{3}$, g_3 by $\alpha \mapsto \alpha_3 = -\alpha_2$, and g_4 via $\alpha \mapsto \alpha_4 = -\alpha$. The resulting determinant D is given by

$$D = -\frac{1}{48} \left[X^8 - 25X^6 + 199X^4 - 495X^2 \right].$$

Since $D(1) = 20/3 \neq 0$ we find that the Galois conjugates of

$$\beta = g(1) = -\frac{1}{24} \left[11\alpha^3 + \alpha^2 - 109\alpha - 11 \right] = -\frac{1}{12} \left[6\sqrt{2} - 5\sqrt{3} + \sqrt{6} - 3 \right]$$

form a normal basis of E/K.

5.3 Cyclotomic extensions

Apart from some relatively elementary preliminaries we now have everything in place to complete the proof of Theorem 3.3.12, Gauß's famous classification of those regular n-gons constructible with compass and straightedge. Recall that we have already shown (see Section 3.3.3) that if the regular n-gon is constructible, then n must be of the form $n = 2^r p_1 \cdot ... \cdot p_\ell$, where r is some non-negative integer and $p_1, ..., p_\ell$ are pairwise distinct Fermat primes.

Also recall the definition of the nth cyclotomic polynomial

$$\Phi_n = \prod_{\substack{k=0 \\ \gcd(k,n)=1}}^{n-1} (X - \zeta_n^k),$$

with $\zeta_n = e^{2\pi\,\mathrm{i}/n}$. Recall that it has degree $\varphi(n)$, where φ denotes the Euler totient function (see Lemma 1.2.2).

We begin with the following result.

Lemma 5.3.1. *Let $n \in \mathbb{N}$ be arbitrary. We have the identity*

$$\prod_{d|n} \Phi_d = X^n - 1,$$

where the product runs over all positive divisors of n, and $\Phi_n \in \mathbb{Z}[X]$.

Proof. Each root of $X^n - 1$ in $\mathbb{C}$ is a primitive dth root of unity for some divisor $d \mid n$, so it coincides with one of the roots of Φ_d. Therefore $(X^n - 1) \mid \prod_{d|n} \Phi_d$. On the other hand we know from Lemma 3.3.11 that $\sum_{d|n} \varphi(d) = n$, so that $\deg \prod_{d|n} \Phi_d = \sum_{d|n} \varphi(d) = n$, hence it follows that the left- and right-hand side must be equal up to a constant. Since each of the Φ_d is clearly monic, so is their product, and hence that constant is 1 and the first claim follows.

The second follows by induction: We have $\Phi_1 = X - 1 \in \mathbb{Z}[X]$. Assume that for some $n > 1$ $\Phi_m \in \mathbb{Z}[X]$ for all $m < n$, then it follows from the identity just proven that

$$\Phi_n = \frac{X^n - 1}{\prod_{d|n, d<n} \Phi_d}.$$

By induction hypothesis, each Φ_d in the denominator is a monic polynomial with integer coefficients, and their product divides $X^n - 1$, so that the quotient again must be a monic polynomial with integer coefficients, as claimed.

q.e.d.

We have shown in Example 2.2.15 and Example 2.2.16 using the Eisenstein criterion that for a prime p, the cyclotomic polynomials Φ_p and Φ_{p^2} are irreducible over $\mathbb{Q}$. We now extend this to the following result.

Theorem 5.3.2. *Let $n \in \mathbb{N}$.*

1. *The cyclotomic polynomial $\Phi_n \in \mathbb{Q}[X]$ is irreducible.*

2. *The extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is Galois of degree $\varphi(n)$ with Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$. In particular, the extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is abelian.*

Proof.

1. By Gauß's Lemma 2.2.7, it is enough to show that Φ_n is irreducible in $\mathbb{Z}[X]$. Assume that $g \in \mathbb{Z}[X]$ is a monic, irreducible divisor of Φ_n such that $g(\zeta) = 0$ for some primitive nth root of unity ζ and let p be any prime number such that $p \nmid n$. We claim that $g(\zeta^p) = 0$. By repeatedly applying this result (note that any k with $\gcd(n, k) = 1$ is a product of primes p that don't divide n), it follows that $g = \Phi_n$, wherefore Φ_n is irreducble.

   We now prove the claim. Write $\Phi_n = g \cdot h$ with $h \in \mathbb{Z}[X]$ monic and assume that $g(\zeta^p) \neq 0$. Then we must have $h(\zeta^p) = 0$. Therefore ζ is a root of the polynomial $h(X^p)$. Since g is a product of factors of the form $X - \zeta'$ and the above reasoning is true for any ζ' it follows that $h(X^p) = g \cdot f$ for some monic $f \in \mathbb{Z}[X]$. We now reduce modulo p and find that $\overline{h(X)^p} \equiv \overline{h(X^p)} = \overline{g \cdot f}$. Therfore the reduced polynomials $\overline{h}, \overline{g} \in \mathbb{F}_p[X]$ cannot be coprime, so that $\overline{\Phi_n} \in \mathbb{F}_p[X]$ is not separable. But since $(X^n - 1)' = nX^{n-1} \neq 0 \in \mathbb{F}_p[X]$, we see that $\gcd(X^n - 1, (X^n - 1)') = 1 \in \mathbb{F}_p[X]$, i.e. $X^n - 1 \in \mathbb{F}_p[X]$, which is a multiple of $\overline{\Phi}_n \in \mathbb{F}_p[X]$, is separable by Lemma 3.4.4. This is clearly a contradiction, so that the claim follows.

2. It is clear that $\mathbb{Q}(\zeta_n)$ is the splitting field of Φ_n, which has degree $\varphi(n)$ and by Part 1. is irreducible over $\mathbb{Q}$. Therefore the extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is normal and also separable since $\mathbb{Q}$ is perfect, and thus Galois. Now let $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) =: G$. Then $\sigma(\zeta_n) = \zeta_n^k$ for some $k = k(\sigma) \in \{0, ..., n-1\}$ with $\gcd(n, k) = 1$ and σ is uniquely determined by this image, so that there is an injective map
   $$G \to (\mathbb{Z}/n\mathbb{Z})^\times, \ \sigma \mapsto k(\sigma).$$
   Since $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is Galois we have
   $$\#G = [\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times,$$
   so that this map must be bijective. Since we have for any $\sigma, \tau \in G$ that
   $$\zeta_n^{k(\sigma\tau)} = \sigma(\tau(\zeta_n)) = \sigma(\zeta_n^{k(\sigma)}) = \zeta_n^{k(\tau)k(\sigma)}$$
   we find that this map is an isomorphism of groups.

q.e.d.

Now it is easy to complete the proof of Gauß's theorem.

Proof of Theorem 3.3.12, Part II Let $n = 2^r p_1 \cdots p_\ell$ where $p_1, ..., p_\ell$ are pairwise distinct Fermat primes. Then the extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ has degree $\varphi(n) = 2^m$ for some m and is Galois with Galois group isomorphic to $(\mathbb{Z}/n\mathbb{Z})^\times$ by Theorem 5.3.2. By the main theorem on finitely generated abelian groups, an abelian group of order N contains a subgroup of order d for any divisor d of N, so in particular in our case $(\mathbb{Z}/n\mathbb{Z})^\times$ contains a subgroup U_a of order 2^a for any $a \leq m$ which we can additionally choose so that $U_a \leq U_{a+1}$. It follows by the Main Theorem of Galois Theory 5.1.10 that there exists a tower of fields $\mathbb{Q} = K_0 \leq K_1 \leq ... \leq K_m = \mathbb{Q}(\zeta_n)$ with $K_j = \mathbb{Q}(\zeta_n)^{U_{m-j}}$ with $[K_j : K_{j+1}] = 2$ for all $j = 0, ..., m-1$. Therfore we find by Theorem 3.3.5 that ζ_n and therefore also the regular n-gon, is constructible with compass and straightedge.

q.e.d.

To conclude this section we mention a famous theorem by Kronecker and Weber classifying all abelian extensions of $\mathbb{Q}$. This is a deep result in Algebraic Number Theory and goes far beyond what we can show in this course.

Theorem 5.3.3. *(Kronecker-Weber) Let $K/\mathbb{Q}$ be a (finite) abelian extension of $\mathbb{Q}$. Then there exists some $n \in \mathbb{N}$ such that K is a subfield of $\mathbb{Q}(\zeta_n)$.*

5.4 Cyclic extensions

In this short section we take a closer look at cyclic Galois extensions.

We begin by discussing an important example of cyclic extensions. Note that we have already encountered it in a different formulation in Remark 1.2.3, at least in part.

Proposition 5.4.1. *Let K be a field of characteristic 0 and suppose that K contains primitive nth roots of unity. Further let $a \in K^\times$ and $\alpha \in \overline{K}$ a root of the polynomial $X^n - a \in K[X]$. Then the extension $K(\alpha)/K$ is cyclic — thus in particular Galois— and its Galois group has order d for some $d \mid n$. Furthermore we have $\alpha^d \in K$.*

Proof. Let $\zeta \in K$ be a primitive nth root of unity. Then all roots of $X^n - a$ are given by $\alpha, \zeta\alpha, ..., \zeta^{n-1}\alpha \in K(\alpha)$, so $K(\alpha)$ is the splitting field of that polynomial, so $K(\alpha)/K$ is a normal extension. Since it is a finite extension in characteristic

0 it is also separable and hence Galois. Now let $\sigma \in G = \mathrm{Gal}(K(\alpha)/K)$ be a K-automorphism of $K(\alpha)$. Then $\sigma(\alpha)$ must be a root of $X^n - a$, so we must have $\sigma(\alpha) = \zeta_\sigma \alpha$ for some $\zeta_\sigma \in \{1, \zeta, ..., \zeta^{n-1}\} = \langle \zeta \rangle \cong C_n$. The map

$$G \to \langle \zeta \rangle, \ \sigma \to \zeta_\sigma$$

is clearly a group homomorphism and also injective, so G is isomorphic to a subgroup of C_n, which is then cyclic of order d for some $d \mid n$. In particular each ζ_σ is a (not necessarily primitive) dth root of unity.

It follows that for any $\sigma \in G$ we have

$$\sigma(\alpha^d) = (\sigma(\alpha))^d = (\zeta_\sigma \alpha)^d = \zeta_\sigma^d \alpha^d = \alpha^d$$

since $\zeta_\sigma^d = 1$. Thus $\alpha^d \in K(\alpha)^G = K$ as we claimed.

q.e.d.

We want to show that, at least for fields of characteristic 0 with sufficiently many roots of unity, all cyclic extensions have this particular shape. For this we need two preliminary results which are in themselves quite famous. The first one is due to Emil Artin.

Theorem 5.4.2. *(Artin) Let K be a field and G a group. Then for $n \leq \#G$, any collection of n distinct* characters *of G over K, i.e. homomorphisms $\sigma : G \to K^\times$, is linearly independent as elements of the vector space of all maps $G \to K$.*

Proof. We use induction over n. Since σ cannot be the zero-map, the case $n = 1$ is clear. Now suppose the claim has been shown for n characters for some $n \geq 1$ and consider the $n + 1$ characters $\sigma_1, ..., \sigma_{n+1}$. Assume that these are linearly dependent, so there exist $a_1, ..., a_{n+1} \in K$ not all 0 such that we have

$$a_1\sigma_1(g) + ... + a_{n+1}\sigma_{n+1}(g) = 0 \quad \text{for all } g \in G. \tag{5.2}$$

Since any n of the $n + 1$ characters are linearly independent by the induction hypothesis, we conclude that all $a_j \neq 0$. We now choose any $g_0 \in G$. Plugging in $g_0 g$ in (5.2) yields

$$a_1\sigma(g_0)\sigma_1(g) + ... + a_{n+1}\sigma_{n+1}(g_0)\sigma_{n+1}(g) = 0 \quad \text{for all } g \in G, \tag{5.3}$$

and by multiplying (5.2) by $\sigma_1(g_0) \neq 0$ we obtain

$$a_1\sigma_1(g_0)\sigma_1(g) + ... + a_{n+1}\sigma_1(g_0)\sigma_{n+1}(g_0)\sigma_{n+1}(g) = 0 \quad \text{for all } g \in G. \tag{5.4}$$

Subtracting (5.4) from (5.3) the yields

$$a_2(\sigma_2(g_0) - \sigma_1(g_0))\sigma_2(g) + ... + a_{n+1}(\sigma_{n+1}(g_0) - \sigma_1(g_0))\sigma_{n+1}(g) = 0 \quad \text{for all } g \in G.$$

But again by the induction hypothesis the n characters $\sigma_2, ..., \sigma_{n+1}$ are linearly independent, so this implies that

$$a_j(\sigma_j(g_0) - \sigma_1(g_0)) = 0 \quad \text{for all } j = 2, ..., n+1$$

and since $a_j \neq 0$, it follows that $\sigma_j(g_0) = \sigma_1(g_0)$ for all j. But since g_0 has been chosen arbitrarily this means that $\sigma_j = \sigma_1$ for all j which is a contradiction, since the characters are assumed to be distinct.

q.e.d.

For the next result we need the following concept.

Definition 5.4.3. Let E/K be a separable extension of degree n and let $\sigma_1, ..., \sigma_n : E \hookrightarrow \overline{K}$ denote the n distinct embeddings of E into the algebraic closure of K (see Remark 3.5.8). For $\alpha \in E$ we define its *norm* over K by

$$\mathrm{Nm}_{E/K}(\alpha) := \prod_{i=1}^{n} \sigma_i(\alpha)$$

and its *trace* as

$$\mathrm{Tr}_{E/K}(\alpha) := \sum_{i=1}^{n} \sigma_i(\alpha).$$

Remark 5.4.4. 1. *For a general field extension E/K, one can define the norm and trace of an element α as the determinant and trace resp. of the K-linear map induced by multiplication by α on E as a K-vector space. It is not very hard to show that for separable extensions, these two definitions coincide.*

2. *The norm map is multiplicative, i.e.* $\mathrm{Nm}_{E/K}(\alpha\beta) = \mathrm{Nm}(\alpha)\,\mathrm{Nm}(\beta)$.

We now use Artin's Theorem to prove another famous theorem which often just referred to as *Hilbert's Theorem 90*. The name is due to the fact that it is the 90th theorem in Hilbert's influential *Zahlbericht*. It is however not due to David Hilbert, but already occurs in earlier works by Ernst Eduard Kummer.

Theorem 5.4.5. *(Hilbert's Theorem 90) Let E/K be a cyclic extension with Galois group $G = \mathrm{Gal}(E/K) = \langle \sigma \rangle$ of order n. Then for every $\alpha \in E$ with $\mathrm{Nm}_{E/K}(\alpha) = 1$ there exists some $\beta \in E$ such that $\alpha = \frac{\beta}{\sigma(\beta)}$.*

Proof. To avoid cluttered notation, we write Nm instead of $\mathrm{Nm}_{E/K}$ in this proof.

First we note that since σ extends to an embedding into $\overline{K}$ by Lemma 3.2.11, it follows that $\mathrm{Nm}(\sigma(\beta)) = \mathrm{Nm}(\beta)$ for any $\beta \in E$, so that indeed $\mathrm{Nm}(\beta/\sigma(\beta)) = 1$ as desired.

Now we show the existence of $\beta \in E$ such that $\alpha = \beta/\sigma(\beta)$. By Artin's Theorem 5.4.2 the maps $\mathrm{id} = \sigma^0, \sigma, ..., \sigma^{n-1}$ are linearly independent over E. Therefore the map

$$\tau := \mathrm{id} + \alpha\sigma + \alpha\sigma(\alpha)\sigma^2 + ... + \alpha\sigma(\alpha)\cdots\sigma^{n-2}(\alpha)\sigma^{n-1}$$

is not identically zero, so there must exist some $\gamma \in E$ such that $\tau(\gamma) \neq 0$. For such γ, define $\beta := \tau(\gamma)$. Then we compute

$$\begin{aligned}\sigma(\beta) &= \sigma\left[\gamma + \alpha\sigma(\gamma) + \alpha\sigma(\alpha)\sigma^2(\gamma) + ... + \alpha\sigma(\alpha)\cdots\sigma^{n-2}(\alpha)\sigma^{n-1}(\gamma)\right] \\ &= \sigma(\gamma) + \sigma(\alpha)\sigma^2(\gamma) + \sigma(\alpha)\sigma^2(\alpha)\sigma^3(\gamma) + ... + \sigma(\alpha)\sigma^2(\alpha)\cdots\sigma^{n-1}(\alpha)\sigma^n(\gamma).\end{aligned}$$

With this we find using that $\sigma^n(\gamma) = \gamma$ and the definition of $\mathrm{Nm}(\alpha)$

$$\begin{aligned}\alpha\sigma(\beta) &= \alpha\sigma(\gamma) + \alpha\sigma(\alpha)\sigma^2(\gamma) + \alpha\sigma(\alpha)\sigma^2(\alpha)\sigma^3(\gamma) + ... + \underbrace{\alpha\sigma(\alpha)\sigma^2(\alpha)\cdots\sigma^{n-1}(\alpha)}_{=\mathrm{Nm}(\alpha)=1}\gamma \\ &= \gamma + \alpha\sigma(\gamma) + \alpha\sigma(\alpha)\sigma^2(\gamma) + \alpha\sigma(\alpha)\sigma^2(\alpha)\sigma^3(\gamma) + ... + \alpha\sigma(\alpha)\cdots\sigma^{n-2}(\alpha) \\ &= \tau(\gamma) \\ &= \beta,\end{aligned}$$

wherefore we have $\alpha = \beta/\sigma(\beta)$ as claimed.

q.e.d.

We can now describe cyclic extensions of fields of characteristic 0.

Theorem 5.4.6. *Let K be a field of characteristic 0 containing a primitive nth root of unity. Furthermore let E/K be a cyclic extension of degree n. Then there exists some $\alpha \in E$ such that $\alpha^n \in K$ and $E = K(\alpha)$.*

Proof. Let $\zeta \in K$ be a primitive nth root of unity and let σ be a generator of the Galois group $\mathrm{Gal}(E/K) = G = \langle\sigma\rangle$. As before we write $\mathrm{Nm} = \mathrm{Nm}_{E/K}$.

We clearly have $\mathrm{Nm}(\zeta^{-1}) = 1$, so that by Hilbert's Theorem 90 (Theorem 5.4.5) there is some $\alpha \in E$ such that $\zeta^{-1} = \alpha/\sigma(\alpha)$ or equivalently $\sigma(\alpha) = \zeta\alpha$. It follows that $\sigma^j(\alpha) = \zeta^j\alpha$, so in particular, all images $\sigma^j(\alpha)$ are pairwise distinct. Therefore we have $[K(\alpha) : K] \geq n = [E : K]$, so that $E = K(\alpha)$.

By direct computation we find

$$\sigma(\alpha^n) = (\zeta\alpha)^n = \alpha^n,$$

so that $\alpha^n \in E^G = K$ as we claimed.

q.e.d.

5.5 Solvability of polynomial equations

In this final section of this chapter we return to the original motivating question of this whole course. As we have mentioned several times, there are solution formulas for the roots of polynomials of degrees 2, 3, and 4 in terms of their coefficients using only basic arithmetic and n-th roots, but it is not possible to find such a formula in any higher degrees.

5.5.1 Solutions by radicals and soluble extensions

We begin with the following definition.

Definition 5.5.1. Let K be field of characteristic 0.

1. We call an extension L/K *solvable by radicals* if there exist subfields
$$K = L_0 \leq L_1 \leq ... \leq L_k = L,$$
such that there exists some $\alpha_j \in L_j$ with $L_j = L_{j-1}(\alpha_j)$ and $\alpha_j^{n_j} \in L_{j-1}$ for some $n_j \in \mathbb{N}$.
2. We call an extension L/K *soluble* if there exists a Galois extension E/K such that $L \leq E$ and $\mathrm{Gal}(E/K)$ is soluble.

In more informal terms, an extension that is solvable by radicals is obtained by successively adjoining n_jth roots.

Example 5.5.2. Consider a generic cubic polynomial $f = X^3 + 3pX + 2q \in \mathbb{Q}[X]$ and assume that f is irreducible. Let L be the splitting field of f . Define $D = p^3 + q^2$ and consider the following tower of fields,

$$\begin{aligned}\mathbb{Q} \leq \mathbb{Q}(\mathrm{i}\sqrt{3}) = L_1 \leq L_1(\sqrt{D}) = L_2 \leq L_2\left((\sqrt[3]{-q+\sqrt{D}}\right) &= L_3 \\ \leq L_3\left(\sqrt[3]{-q-\sqrt{D}}\right) &= L_4.\end{aligned}$$

By the Cardano formula (Theorem 1.3.2) we see that any root α is certainly contained in L_4, so that $L/\mathbb{Q}$ is solvable by radicals.

Before coming to the main result of this subsection we need the following two Lemmata.

Lemma 5.5.3. *Let $L = K(\alpha)/K$ be solvable by radicals and let E be the* Galois closure *of L/K, i.e. splitting field of the minimal polynomial of α over K. Then E/K is solvable by radicals as well.*

Proof. Since L/K is solvable by radicals, there is a tower of subfields

$$K = L_0 \leq L_1 \leq ... \leq L_k = L,$$

such that there exists some $\alpha_i \in L_i$ with $L_i = L_{i-1}(\alpha_i)$ and $\alpha_i^{n_i} \in L_{i-1}$ for some $n_i \in \mathbb{N}$. Denote by $\mu_i = \mu_{\alpha_i}$ the minimal polynomial of α_i over K. Then E must be the splitting field of $f = \prod_j \mu_j \in K[X]$ since E must certainly contain all the roots of the μ_j and is minimal with that property. Write

$$\mu_i = \prod_{j=1}^{n_i} (X - \beta_{ij}) \in E[X]$$

and define $E_0 = K$, $E_1 = K(\beta_{1,1}, ..., \beta_{1,n_1})$, and $E_{i+1} = E_i(\beta_{i+1,1}, ..., \beta_{i+1,n_{i+1}})$. Since E/K is normal, for every j there is a K-homomorphism $\sigma : E \to E$ such that $\sigma(\alpha_i) = \beta_{ij}$, whence $\beta_{ij}^{n_i} = \sigma_(\alpha_i)^{n_i} \in \sigma(L_{i-1}) \leq \sigma(E_{i-1})$. But since E_{i-1} is again a splitting field of a polynomial over K, we have that E_{i-1}/K is normal and hence $\sigma(E_{i-1}) = E_{i-1}$. Therefore each extension E_i/E_{i-1} is solvable by radicals and thus also the extension E/K.

q.e.d.

Lemma 5.5.4. *Let $n \in \mathbb{N}$ and E/K be a Galois extension where K has characteristic 0. Let F/E resp. L/K denote the splitting field of $X^n - 1$ over E and K resp. Then the extension F/K is Galois. Furthermore we have that $\operatorname{Gal}(F/K)$ is soluble if and only if $\operatorname{Gal}(E/K)$ is soluble if and only if $\operatorname{Gal}(F/L)$ is soluble.*

Proof. Since E/K is Galois and hence in particular normal, it follows from Corollary 3.6.4 that there is a polynomial $f \in K[X]$ such that E is the spitting field of f. Therefore, F is the splitting field of the polynomial $(X^n - 1) \cdot f \in K[X]$, so that F/K is Galois.

Since both E/K and L/K are normal extensions and contained in F/K, it follows from Proposition 5.1.12 that $\operatorname{Gal}(F/E) \trianglelefteq \operatorname{Gal}(F/K)$ and $\operatorname{Gal}(F/L) \trianglelefteq \operatorname{Gal}(F/K)$ and

$$\operatorname{Gal}(E/K) \cong \operatorname{Gal}(F/K)/\operatorname{Gal}(F/E).$$

Since $\operatorname{Gal}(F/E)$ is abelian and hence soluble, it follows from Proposition 4.2.4 that $\operatorname{Gal}(E/K)$ is soluble if and only if $\operatorname{Gal}(F/K)$ is soluble. Since also $\operatorname{Gal}(L/K) \cong \operatorname{Gal}(F/K)/\operatorname{Gal}(F/L)$ is abelian, it follows again from Proposition 4.2.4 that $\operatorname{Gal}(F/K)$ is soluble if and only if $\operatorname{Gal}(F/L)$ is soluble.

q.e.d.

We now show the key result about the solvability of polynomials.

Theorem 5.5.5. *Let K be a field of characteristic 0 and L/K a finite extension. Then L/K is solvable by radicals if and only if L/K is soluble.*

Proof. "$\Rightarrow$": Assume that L/K is solvable by radicals, i.e. we have a tower of fields

$$K = L_0 \leq L_1 \leq \ldots \leq L_k = L,$$

such that there exists some $\alpha_j \in L_j$ with $L_j = L_{j-1}(\alpha_j)$ and $\alpha_j^{n_j} \in L_{j-1}$ for some $n_j \in \mathbb{N}$. Then Lemma 5.5.3 yields that also the Galois closure of L/K, which we call E, is solvable by radicals, so there is a similar tower of fields leading to E and we call the subfields E_j. Let $n := [E : K]$ and consider the extension $E(\zeta)/K$, where ζ is a primitive nth root of unity. By Lemma 5.5.4, this extension is again Galois with Galois group G. Now consider the tower

$$K(\zeta) \leq E_1(\zeta) \leq \ldots \leq E_\ell(\zeta) = E(\zeta).$$

By Theorem 5.4.6, each one of the extensions $E_i(\zeta)/E_{i-1}(\zeta)$ is cyclic, so by the Main Theorem of Galois Theory 5.1.10 together with Proposition 5.1.12 it follows that there is a subnormal series

$$G \trianglerighteq G_1 \trianglerighteq \ldots \trianglerighteq G_\ell = \{1\}$$

of G, where each factor group G_i/G_{i-1} is cyclic, so in particular abelian. Therefore Theorem 4.2.3 tells us that $\mathrm{Gal}(E(\zeta)/K(\zeta))$ is soluble and thus by Lemma 5.5.4 also $\mathrm{Gal}(E/K)$ is soluble.

"$\Leftarrow$": Assume that L/K is soluble, i.e. there exisits some extension field E/L such that the extension E/K is Galois with soluble Galois group G. By Lemma 5.5.4 we may assume without loss of generality that E contains sufficiently many roots of unity. It follows from Theorem 4.2.3 that there exists a subnormal series

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \ldots \trianglerighteq G_k = \{1\}$$

where each factor group G_i/G_{i+1} is abelian. Indeed by refining the subnormal series, we may (and do) assume that each of these factors is cyclic. By Proposition 5.1.12, there exists therefore a tower of cyclic extensions

$$K = E_0 \leq \ldots \leq E_k = E.$$

By our classification of cyclic extensions in Theorem 5.4.6 we see that each of these fields E_j is obtained by adjoining some n_jth root, so that L/K is solvable by radicals as we claimed.

q.e.d.

This implies abstractly, independently from the formulas of Cardano and Ferrari, that solution formulas for polynomials of degree 3 and 4 must exist.

Corollary 5.5.6. *Any extension $K/\mathbb{Q}$ of degree ≤ 4 is solvable by radicals. In other words there is a solution formula for roots of polynomials of degree ≤ 4 in terms of radicals.*

Proof. A field $K/\mathbb{Q}$ of degree n is a subfield of a finite Galois extension $E/\mathbb{Q}$. Since any Galois automorhism of E permutes the roots of the minimal polynomial of an element of K in E, it follows that the Galois group embeds into the symmetric group S_n. Since S_n is soluble for $n \leq 4$, so is any of its subgroups and the claim follows from Theorem 5.5.5.

q.e.d.

We conclude this subsection with an example.

Example 5.5.7. Since cyclotomic extensions are abelian and thus in particular soluble, it follows that any primitive root of unity $\zeta_n = e^{2\pi\,\mathrm{i}/n}$ must be expressible only in terms of radicals. As an illustration, we list the expressions for $\mathrm{Re}(\zeta_n) = \cos(2\pi/n)$ for the first few n, since the imaginary part is easily obtained from it using the relation $\cos^2\theta + \sin^2\theta = 1$.

$$\begin{aligned}
\cos(2\pi/3) &= -\frac{1}{2}\\
\cos(2\pi/4) &= 0\\
\cos(2\pi/5) &= \frac{-1+\sqrt{5}}{4}\\
\cos(2\pi/6) &= \frac{1}{2}\\
\cos(2\pi/7) &= \frac{-1+\sqrt[3]{14+21\frac{-1+\sqrt{-3}}{2}}+\sqrt[3]{14-21\frac{-1+\sqrt{-3}}{2}}}{3}\\
\cos(2\pi/8) &= \frac{1}{\sqrt{2}}
\end{aligned}$$

Note that some care must be taken in applying the above formula for $\cos(2\pi/7)$ because the third root of a complex number (or even a negative real number) is not uniquely determined. Even though we are expressing real numbers by radicals, the complex numbers in the above expressions cannot be avoided.

5.5.2 Insolubility of the quintic

Since the group S_5 is not soluble, as it contains the non-abelian simple group A_5 as a normal subgroup, it immediately follows from the above discussion that there cannot be a solution formula for a polynomial f of degree 5 whose splitting field over $\mathbb{Q}$ has Galois group S_5, so we need to show that such a polynomial exists.

The remainder of this section is devoted to showing a little more (except for the proof of Proposition 5.5.18, which requires some background in commutative algebra which would take too long to develop), namely that for any n, there exists an irreducible polynomial of degree n such that its splitting field over $\mathbb{Q}$ has Galois group S_n. It follows directly that there can't be a solution formula for polynomials of *any* degree ≥ 5. The proof we present here is an adaptation of the proof given in the book by van der Waerden mentioned in the introduction, as posted by user Mako Kato on Math Stackexchange[1].

We begin by making the following observations.

Lemma 5.5.8. *Let K be field, $f \in K[X]$ a separable polynomial and E/K be the splitting field of f. Then $G := \mathrm{Gal}(E/K)$ acts transitively on the roots of f if and only if f is irreducible.*

Proof. Suppose f is irreducible and that $\alpha \in E$ is a fixed root of f. For any root $\beta \in E$ of f, we have seen that the rupture fields $K(\alpha)$ and $K(\beta)$ are isomorphic, so in other words the K-homomorphism φ defined by $\varphi(\alpha) = \beta$ defines an embedding of $K(\alpha)$ into $\overline{K}$. By Lemma 3.2.11, this extends to an embedding $\sigma : E \hookrightarrow \overline{K}$, and since E is normal, this extension is an endomorphism of E. Therefore we have found $\sigma \in \mathrm{Aut}_K(E) = \mathrm{Gal}(E/K)$ such that $\sigma(\alpha) = \beta$, and therefore G acts transitively on the roots of f.

If $f = gh \in K[X]$, where $g \in K[X]$ irreducible and $\deg h \geq 1$, then G acts on the roots of g in E, and thus cannot act transitively on the roots of f.

q.e.d.

This immediately implies the following.

Corollary 5.5.9. *Let K be a field, $f \in K[X]$ be a separable polynomial over K and E/K be the splitting field of f. Write $f = f_1 \cdots f_r$ for distinct irreducible polynomials $f_j \in K[X]$. Denote by S the set of roots of f in E and similarly let S_j denote the set of roots of f_j. Then $S = \bigcup_j S_j$ and each S_j is a $\mathrm{Gal}(E/K)$-orbit.*

We now have to make a quick detour to finite fields. Recall from Example 5.1.5 that for a finite field K any finite extension E/K is Galois, where the Galois group is cyclic and generated by the Frobenius automorphism of E/K. This leads to the following remark.

[1] `https://math.stackexchange.com/questions/165675/constructing-a-galois-extension-field-with-galois-group-s-n`

Remark 5.5.10. *If K is a finite field and E/K has degree n and $f \in K[X]$ is an irreducible polynomial such that E is its splitting field, then the Frobenius endomorphism viewed as a permutation on the roots of f defines an n-cycle.*

Together with Corollary 5.5.9 this implies the following.

Lemma 5.5.11. *Let K be a finite field and $f \in K[X]$ a separable polynomial. Write $f = f_1 \cdots f_r$ for pairwise distinct irreducible polynomials $f_j \in K[X]$ of degree n_j and let E/K be the splitting field of f. Then the Frobenius automorphism of E/K viewed as permutation on the roots of f is a product of disjoint cycles of lengths $n_1, ..., n_r$.*

We illustrate this with an example.

Example 5.5.12. Let $K = \mathbb{F}_5$, $f = (X^2 - 2)(X^3 + X + 1) \in \mathbb{F}_5[X]$ and E/K be the splitting field of f. Then there is some $\alpha, \beta \in E$ such that $\alpha^2 = 2$ and $\beta^3 + \beta + 1 = 0$. The roots of the quadratic factor of f are then clearly given by $\alpha = \alpha_1$ and $4\alpha = \alpha_2$, while the roots of the cubic factor can be checked to be given by $\beta = \beta_1$, $4\beta^2 + \beta + 1 = \beta_2$, $\beta^2 + 3\beta + 4 = \beta_3$.

Under the Frobenius map $\Phi_5 : E \to E$, $a \mapsto a^5$ we see that

$$\Phi_5(\alpha) = \alpha^5 = \alpha \cdot \alpha^4 = 4\alpha = \alpha_2,$$

and

$$\Phi_5(\beta) = \beta^5 = 4\beta^2 + \beta + 1 = \beta_2,\ \Phi_5(\beta_2) = \beta^{25} = \beta^2 + 3\beta + 4 = \beta_3,\ \Phi_5(\beta_3) = \beta.$$

Therefore if we number the roots $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$ by $1, 2, 3, 4, 5$, then Φ_5 induces the permutation

$$(1, 2)(3, 4, 5).$$

Now we discuss two lemmas on the symmetric group S_n. Recall that S_n is generated by transpositions (i, j), $1 \leq i < j \leq n$, that makes $\binom{n}{2} = \frac{n(n-1)}{2}$ generators. Indeed we can reduce this set of generators by quite a bit to only $n - 1$ generators.

Lemma 5.5.13. *The group S_n is generated by transpositions (k, n) for $1 \leq k \leq n - 1$.*

Proof. Let (a, b), $1 \leq 1 < b \leq n$ be a transposition. If $b = n$, then it is already in our alleged set of generators. If $b \neq n$ we can write $(a, b) = (a, n)(b, n)(a, n)$, so any known generator can be expressed in terms of the new ones, and the claim follows.

q.e.d.

Using this we can show the following.

Lemma 5.5.14. *Let G be a finite permutation group on a finite set M with $\#M = n$. Suppose that*

1. *G acts transitively on M,*
2. *G contains a transposition,*
3. *G contains an $(n-1)$ cycle.*

Then G is isomorphic to S_n.

Proof. Me may assume without loss of generality that $M = \{1, ..., n\}$ and that G contains the cycle $c = (1, ..., n-1)$ as well as the transposition (i, j). Since G acts transitively on M, there is some $g \in G$ such that $g.j = n$. Set $h = g.i$. Then G contains the transpose $g(i,j)g^-1 = (g.i, g.j) = (h, n)$. Conjugating (h, n) by powers of c we see that G contains every transposition (k, n) with $1 \leq k < n$. Therefore, by Lemma 5.5.13, G contains a full set of generators of S_n and therefore, since it is a subgroup of S_n, it must be the full symmetric group.

q.e.d.

This is essentially enough to give an example of a polynomial over $\mathbb{Q}$ of degree 5 with Galois group S_5.

Corollary 5.5.15. *Let G be a subgroup of S_5 acting transitively on $\{1, ..., 5\}$ and containing a transposition (i, j). Then $G = S_5$.*

Proof. By Lemma 5.5.14 it is enough to show that G contains a 4-cycle. Assume without loss of generality that $(i, j) = (1, 2)$. Since G acts transitively on $\{1, 2, 3, 4, 5\}$, for every $k \in \{1, 2, 3, 4, 5\}$, there is some $g \in G$ such that $g.2 = k$. Setting $j = g.1$, we see that the transposition $(j, k) = g(1, 2)g^{-1} \in G$. Furthermore at least one $k \in \{1, 2, 3, 4, 5\}$ occurs in at least two such transpositions, because there are clearly at least 3 transpositions in G, and since they can contain only the numbers $1, 2, 3, 4, 5$, there must be at least two transpositions in G that overlap. The product of these overlapping transposition gives us a 3-cycle $(a, b, c) \in G$. By renumbering, we may assume that $(a, b, c) = (1, 2, 3)$.

Now we claim that there is a transposition $(j, 4) \in G$ with $j \in \{1, 2, 3\}$, so that $(1, 2, 3)(j, 4) \in G$ defines a 4-cycle in G, so we are finished by Lemma 5.5.14. To show the claim we notice that there must be two transpositions containing 4, since by trasitivity, there is $g \in G$ such that $g.1 = 4$ and since $(1, 2, 3)(1, 2)(1, 3, 2) = (2, 3) \in G$, the same is true for $g' = g(2, 3)$. We thus obtain $g(1, 2)g^{-1} = (g.2, 4)$ and $g'(1, 2)g'^{-1} = (g.3, 4)$. Since $g.2, g.3 \neq 4$, at least one of them must lie in $\{1, 2, 3\}$ and the claim follows, completing the proof.

q.e.d.

With this, we can give an example of a quintic polynomial over $\mathbb{Q}$ with Galois group S_5.

Example 5.5.16. Consider the polynomial $f = X^5 - 6X + 3 \in \mathbb{Q}[X]$. By Eisenstein's criterion (see Theorem 2.2.12), this polynomial is irreducible over $\mathbb{Q}$. Let $E/\mathbb{Q}$ be its splitting field. Looking at the graph below (see Figure 5.1), we see that f has exactly three real roots (approximately $-1.6709...$, $0.50550...$, 1.4016) and therefore a pair of complex conjugate roots (approximately $-0.1181...\pm 1.5874...\,\mathrm{i}$).

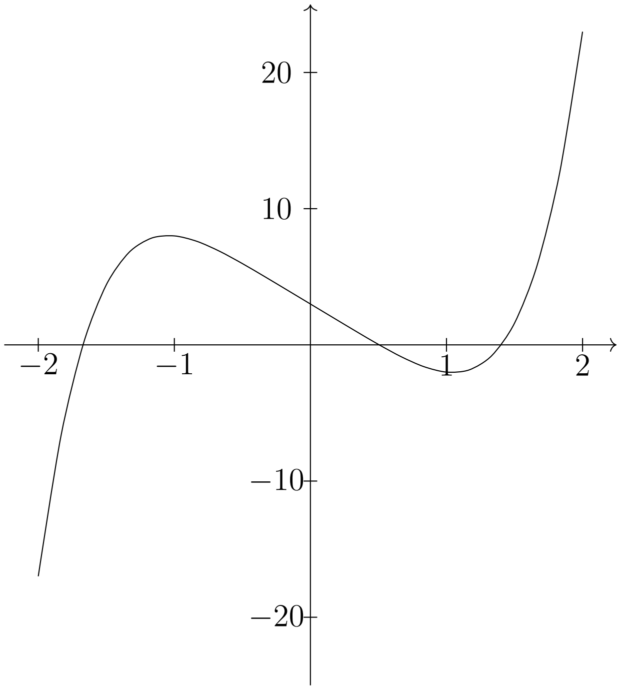


Figure 5.1: Graph of the polynomial function $x \mapsto x^5 - 6x + 3$

Complex conjugation, restricted to E, induces an automorphism of E, flipping the two complex roots and leaving the real ones invariant. Therefore the Galois group $\mathrm{Gal}(E/\mathbb{Q})$ acts transitively on the five roots of f and contains a transposition. Thus we know from Corollary 5.5.15 that $\mathrm{Gal}(E/\mathbb{Q}) \cong S_5$, which is not soluble. Therefore, by Theorem 5.5.5, there can't be a formula for the roots of f in terms of radicals, so the general quintic is insoluble.

Indeed this enough to show that there cannot be solution formulae in terms of radicals for polynomials of any degree ≥ 5, since if there were such a formula in degree n, then it would express the roots of $X^{n-5}f$, and hence those of f in terms of radicals.

Having solved the solubility problem for the quintic in particular, we return to the main objective of this section, constructing Galois extensions of $\mathbb{Q}$ with Galois group S_n. We continue with the following easy observation.

Remark 5.5.17. *Let $f \in \mathbb{Z}[X]$ be a monic polynomial. Suppose the reduction $\overline{f}$ of f modulo a prime number p is separable. Then f is separable over $\mathbb{Q}$.*

Proof. Suppose $f \in \mathbb{Q}[X]$ is not separable. Since $\mathbb{Q}$ has characteristic 0, this implies that there must be some non-constant monic polynomial $g \in \mathbb{Z}[X]$ such that $g^2 \mid f$. But this implies that $\overline{f}$ is divisible by $\overline{g}^2$ in $\mathbb{F}_p[X]$ and is therefore not separable.

q.e.d.

The following proposition will turn out to be the key to the theorem we wish to prove here.

Proposition 5.5.18. *Let $f \in \mathbb{Z}[X]$ be a separable, monic polynomial and p a prime number such that the reduction $\overline{f}$ modulo p is again separable. Write $\overline{f} = \overline{f}_1 \cdots \overline{f}_r$ for pairwise distinct irreducible polynomials overline$f_j \in \mathbb{F}_p[X]$ of degree n_j. Let $E/\mathbb{Q}$ be the splitting field of f, $G := \mathrm{Gal}(E/\mathbb{Q})$ its Galois group, and $S \subset E$ be the set of roots of f in E. Further let $\overline{E}/\mathbb{F}_p$ denote the splitting field of $\overline{f}$ with Galois group $\overline{G} := \mathrm{Gal}(\overline{E}/\mathbb{F}_p)$ and denote the set of roots of $\overline{f}$ in $\overline{E}$ by $\overline{S}$. Then G contains an element which is a product of disjoint cycles of lengths n_j.*

Even though there is a fairly elementary proof for this (actually for a more general statement), it requires some basic results and concepts in commutative algebra which we cannot develop here, so we take this for granted.

With this result, we come to the main result of this section.

Theorem 5.5.19. *For every $n \geq 1$ there exists a Galois extension $E/\mathbb{Q}$ such that $\mathrm{Gal}(E/\mathbb{Q}) \cong S_n$.*

Proof. The claim is of course obvious for $n \leq 2$, so we assume $n \geq 3$.

Let $f_2 \in \mathbb{F}_2[X]$ be an irreducible polynomial of degree n (the minimal polynomial of any generator of $\mathbb{F}_{2^n}^\times$ works, according to Remark 3.4.6).

Let $g_0 \in \mathbb{F}_3[X]$ be a monic polynomial of degree 1 and $g_1 \in \mathbb{F}_3[X]$ a monic, irreducible polynomial of degree $n-1$ (if $n = 2$, pick $g_1 \neq g_0$) and set $f_3 = g_0 g_1 \in \mathbb{F}_3[X]$, which is separable by construction.

Let $h_0 \in \mathbb{F}_5[X]$ be a monic, irreducible polynomial of degree 2. If $n-2$ is odd, then let $h_1 \in \mathbb{F}_5[X]$ be a monic irreducible polynomial of degree $n-2$ and set $f_5 = h_0 h_1 \in \mathbb{F}_5[X]$. If $n-2$ is even, then choose $h_1 \in \mathbb{F}_5[X]$ be a monic polynomial of degree 1 and h_2 and irreducible, monic polynomial of degree $n-3$ (if $n-3 = 1$, then choose $h_2 \neq h_1$) and set $f_5 = h_0 h_1 h_2 \in \mathbb{F}_5[X]$, which is again separable by construction in both cases.

Now lift the polynomials f_2, f_3, f_5 to monic polynomials in $\mathbb{Z}[X]$, which we also call f_2, f_3, f_5, and define $f = -15f_2 + 10f_3 + 6f_5$. Note that $f \in \mathbb{Z}[X]$ is again monic of degree n. Clearly we have the congruences $f \equiv f_p \pmod{p}$ for $p \in \{2, 3, 5\}$.

Since f_2 is irreducible it follows that f is irreducible over $\mathbb{Q}$ (see Theorem 2.2.10 and Gauß's Lemma 2.2.7). Let $E/\mathbb{Q}$ be the splitting field of f, $G := \mathrm{Gal}(E/\mathbb{Q})$ its Galois group, and $S \subset E$ the set of roots of f in E.

Since f is irreducible, G acts transitively on S by Lemma 5.5.8.

Since $f \equiv f_3 \pmod{3}$, it follows from Proposition 5.5.18 that G contains an $(n-1)$-cycle.

Since $f \equiv f_5 \pmod{5}$, it follows by the same proposition that G contains a permutation which is a product of disjoint cycles of lengths $2, n-2$ if n is odd and of lengths $2, 1, n-3$ if n is even, call it g. If n is odd, it follows that g^{n-2} is a transposition, if n is even, then g^{n-3} is a transposition, so in either case, G contains a transposition.

Therefore we find by Lemma 5.5.14 that $G \cong S_n$ as we claimed.

q.e.d.

Example 5.5.20. The polynomial $f = X^5 - 6X + 3$ from Example 5.5.16 is an example for a polynomial with Galois group S_5, but, as one easily checks, it does not arise from the construction in the proof of Theorem 5.5.19.

It is however not too hard to find using trial and error that

$$f_2 = X^5 + X^2 + 1 \in \mathbb{F}_2[X]$$

is irreducible (if it were reducible, it would have to have an irreducible factor of degree 1 or 2 and there is only one irreducible polynomial of degree 2 over $\mathbb{F}_2$). Similarly,

$$f_3 = X(X^4 + X^2 + X + 1) = X^5 + X^3 + X^2 + X \in \mathbb{F}_3[X]$$

and

$$f_5 = (X^2 + X + 1)(X^3 + X + 1) = X^5 + X^4 + 2X^3 + 2X^2 + 2X + 1 \in \mathbb{F}_5[X]$$

are polynomials satisfying the conditions in the proof. Therefore another degree 5 polynomial with Galois group S_5 over $\mathbb{Q}$ is given by

$$f = -15f_2 + 10f_3 + 6f_5 = X^5 + 6X^4 + 22X^3 + 7X^2 + 22X - 9 \in \mathbb{Q}[X].$$

We conclude this section with a few remarks.

Remark 5.5.21. *If one orders monic polynomials in $\mathbb{Z}[X]$ of degree $n \geq 2$ by their largest coefficient in absolute value, the famous Irreducibility Theorem of Hilbert, a result in Analytic Number Theory, implies that asymptotically* 100% *of them are irreducible and have Galois group S_n.*

This means that if one picks a monic, integer polynomial "at random", one should expect that it is irreducible and has the largest possible Galois group and it is essentially impossible to find other polynomials purely by chance (although of course they do of course exist).

Remark 5.5.22. *Theorem 5.5.19 solves the so-called* inverse Galois problem *for the groups S_n and the field $\mathbb{Q}$: Given a finite group G and a field K, is there a Galois extension E/K such that $\mathrm{Gal}(E/K) \cong G$?. For $K = \mathbb{Q}$, this has been studied extensively. It is known for instance that every soluble group arises as a Galois group over $\mathbb{Q}$ (Shafarevich), so does every alternating group A_n (Hilbert), as well as several others, for instance the famous* Monster group, *the largest of the so-called sporadic simple groups (Thompson). It is however not known whether or not* every *finite group is realized as a Galois group over $\mathbb{Q}$. At the time of writing the smallest order of a group not known*[2] *to occur as a Galois group of a finite extension over $\mathbb{Q}$ is* 8160, *the group being a semidirect product of the simple group* $\mathrm{PSL}_2(\mathbb{F}_{16})$ *and C_2.*

Remark 5.5.23. *Regarding solution formulas for the quintic and higher degree polynomials, it should be pointed out that such formulas* do *exist, but they involve more complicated functions than nth roots. For example, it is possible to transform a general quintic polynomial to a polynomial of the form*

$$X^5 - X + t.$$

Hermite was the first to realise that one could use so-called elliptic functions *to express the solutions of this equation. In a sense this is analogous to, but quite a bit more involved than, a classical solution formula for the cubic in terms of trigonometric functions.*

[2] This information was retrieved from Jürgen Klüners's database of number fields at `http://galoisdb.math.upb.de/`.

Chapter 6

Algebraic closure*

In this chapter we want to take one more look at the algebraic closure of a field. In particular, we would like to show that every field has an algebraic closure and that it is essentially unique. However, this fact actually depends on the so-called *Axiom of Choice* in Set Theory which we will discuss first.

6.1 The Axiom of Choice*

In Mathematics an *axiom* is a fundamental fact that is “self-evident” and is therefore to be accepted. This terminology goes back to Euclid who in his treatise *The Elements* derived essentially all facts in geometry known at the time from only five such axioms:

1. There is exactly one straight line through any two given points in the plane.

2. Any finite straight line can be extended to a straight line of arbitrary length.

3. For any given point P and given radius r, there exists exactly one circle with midpoint P and radius r.

4. All right angles are equal to one another.

5. That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

The fifth axiom is clearly much more complicated than the others and not as self-evident. An informal way to state it is that two non-parallel lines always intersect. This axiom has been a point of debate for centuries and it has been shown in the early 19th century by Bolyai and independently Lobachevsky, that without the fifth axiom, one still arrives at valid geometries (spherical and hyperbolic geometry). It has also been attempted multiple times to derive the fifth axiom from the others, but this turned out to be impossible, since the fifth axiom is logically independent from the others.

A similar, but in the details much more complicated, situation arises when one tries to axiomatise Set Theory. There are several "actual" axioms, that one can regard as self-evident. These include for instance that two sets are equal if and only if they contain the same elements, one can form the union of two sets, any set has a power set (the set of all its subsets), and some others. The axiom which is much less self-evident (and is therefore not universally accepted) is the *Axiom of Choice*.

Axiom 6.1.1. *(Axiom of choice) Let $\Lambda \neq \emptyset$ be any set and suppose we have a set $X_\lambda \neq \emptyset$ for each $\lambda \in \Lambda$. Then the* Cartesian product

$$\prod_{\lambda\in\Lambda} X_\lambda := \{(x_\lambda)_{\lambda\in\Lambda} \ : \ x_\lambda \in X_\lambda\}$$

is not empty.

Remark 6.1.2. *Another, less formal, way to formulate the Axiom of Choice is that it is possible to choose $x_\lambda \in X_\lambda$ for each $\lambda \in \Lambda$ simultaneously.*

It has been shown that this axiom is also logically independent of the standard Zermelo-Frenkel axioms of Set Theory, meaning that one can assume that the Axiom of Choice is true or false without generating any logical inconsistencies.

There are two other important variants of this axiom, which look completely unrelated. To state them, we require the following definition.

Definition 6.1.3. Let X be a set. A relation $\preceq$ on X is called an *ordering* or *order* if it satisfies the following properties:

(i) It is *reflexive*, i.e. we have $x \preceq x$ for all $x \in X$.

(ii) It is *antisymmetric*, i.e. if $x \preceq y$ and $y \preceq x$ for some $x, y \in X$, then $x = y$.

(iii) It is *transitive*, i.e. if $x \preceq y$ and $y \preceq z$ for $x, y, z \in X$, then we also have $x \leq z$.

The pair $(X, \preceq)$ is then called an *ordered set.*

The definition of an ordering is of course an abstraction from the familiar $\leq$-relation on the real numbers. In this generality however, there are some special properties that one wishes to distinguish.

Definition 6.1.4. Let $(X, \preceq)$ be an ordered set.

1. We call $(X, \preceq)$ *totally ordered*, if for any $x, y \in X$ we have $a \preceq b$ or $b \preceq a$.
2. The ordered set $(X, \preceq)$ is called *well-ordered* if every non-empty subset $\emptyset \neq Y \subseteq X$ contains a smallest element, i.e. there exists some $y_0 \in Y$ such that $y_0 \preceq y$ for all $y \in Y$.
3. The ordered set $(X, \preceq)$ is *ordered inductively* if any totally ordered subset $Y \subseteq X$ has an *upper bound* in X, i.e. there exists some $x \in X$ (possibly depending on Y) such that $y \preceq x$ for all $y \in Y$.
4. An element $m \in X$ such that we know for every $x \in X$ that if $m \preceq x$ then $x = m$, is called a *maximal element* of X.
5. A subset $S \subseteq X$ is called a *segment* if for any $x \in X$ we have that if there exists some $s \in S$ such that $x \preceq s$, then $x \in S$.

Example 6.1.5.

1. The set of natural numbers $\mathbb{N}$ is totally ordered and well-ordered with respect to the usual $\leq$ ordering. It is however not ordered inductively, as any infinite subset of $\mathbb{N}$ is not bounded above by any natural number.
2. The ordered set $(\mathbb{R}, \leq)$ is totally ordered, but not well-ordered, as for instance the interval $(0, 1]$ has no smallest element.
3. The ordered set $((0, 1], \leq)$ is totally ordered and ordered inductively (any subset is bounded above by 1), but again not well-ordered.
4. Let X be any set and $\mathfrak{P}(X) := \{Y \subseteq X\}$ its power set. Then $(\mathfrak{P}(X), \subseteq)$ is an ordered set, but as soon as X has more than one element, it is not totally ordered.
5. Any interval of the form $(-\infty, a]$ is a segment in $(\mathbb{R}, \leq)$.

Remark 6.1.6. *Let X be a set.*

1. *A well-ordering on X is also a total ordering on X, since for any $x, y \in X$, the set $\{x, y\}$ has a smallest element.*
2. *Let $\preceq$ be an ordering on X. Then for any $x \in X$ the set*

$$X_{\leq x} := \{y \in X \ : \ y \preceq x\}$$

is a segment.

3. If $(X, \preceq)$ is totally ordered and $S \subset X$, so gilt $S \subset X_{\leq x}$ for any $x \in X \setminus S$.

The first variant of the axiom of choice is the so-called *Well-ordering Theorem.*

Axiom 6.1.7. *(Well-ordering Theorem) For any set X there is an ordering $\preceq$ on X such that $(X, \preceq)$ is well-ordered.*

Remark 6.1.8. *For some sets it is possible to construct such an ordering explicitly (for instance the ordering $\leq$ on the natural numbers works). But for other sets, such as (probably) the real numbers, finding this well-ordering explicitly is impossible, it is just known to exist through the Well-ordering Theorem (if we assume it to be true).*

The second variant is known as *Zorn's Lemma.* This is the version of the axiom of choice that is most widely used in the context of algebra.

Axiom 6.1.9. *(Zorn's Lemma) Let $(X, \preceq)$ be a non-empty ordered set. If $(X, \preceq)$ is ordered inductively, then $(X, \preceq)$ contains maximal elements. So if every totally ordered subset, a* chain*, in M has an upper bound in M, then there exists some $m \in X$ such that we have for every $x \in X$ that if $m \preceq x$ then $x = m$.*

We now formulate the main theorem of this section.

Theorem 6.1.10. *The three axioms 6.1.1, 6.1.7, and 6.1.9 are equivalent.*

As already mentioned, it has been shown that the Axiom of Choice is independent of the other axioms of Set Theory, so one has to assume it, and therefore also the Well-ordering Theorem and Zorn's Lemma, as an additional axiom if one wants to use them.

Before proceeding to the proof of Theorem 6.1.10, we give two immediate applications.

Theorem 6.1.11. *Assuming the Axiom of Choice, every vector space V over a field K has a basis.*

Proof. Let V be a vector space over a field K and define

$$\mathcal{B} := \{B \subseteq V \ : \ B \text{ is linearly independent}\}.$$

Then the usual subset relation $\subseteq$ defines an ordering on $\mathcal{B}$. Now let $\mathcal{C} \subseteq \mathcal{B}$ be any chain in $\mathcal{B}$, i.e. a totally ordered subset of $\mathcal{B}$ and consider the set

$$C := \bigcup_{B \in \mathcal{C}} B.$$

Then $C \subseteq V$ is a linearly independent set, for if it were not, there would be *finitely many* $v_1, ..., v_n \in C$ and $\alpha_1, ..., \alpha_n \in K$ not all 0 such that $\sum_{i=1}^{n} \alpha_i v_i = 0$. But since $\mathcal{C}$ is totally ordered, there must be some $B \in \mathcal{C}$ such that $v_1, ..., v_n \in B$, so that B is linearly dependent, which is a contradiction.

Therefore we have $C \in \mathcal{B}$ and it is clearly an upper bound for the chain $\mathcal{C}$. By Zorn's Lemma 6.1.9 $\mathcal{B}$ contains maximal elements, i.e. maximal linearly independent subsets $X \subseteq V$. It remains to show that any such X generates V. If we have $v \in V \setminus \langle X \rangle$, then the set $X \cup \{v\}$ is a proper superset of X and is therefore linearly dependent, so, since X itself is linearly independent, $v \in \langle X \rangle$, which is again a contradiction. Thus X is a linearly independent generating set for V, therfore by definition a basis.

q.e.d.

Remark 6.1.12. *It is in general not possible to find such a basis, e.g. for the* $\mathbb{R}$*-vector space of continuous functions* $f : [0, 1] \to \mathbb{R}$ *or the space of formal power series* $K[\![X]\!]$ *over a field* K*.*

Theorem 6.1.13. *Let* R *be a ring. Then, assuming the Axiom of Choice, every proper ideal* $I \trianglelefteq R$*,* $I \neq R$*, is contained in a maximal ideal.*

Proof. Let $I \trianglelefteq R$, $I \neq R$, be a proper ideal. Consider the set $\mathcal{I}$ of all ideal $J \trianglelefteq R$, $J \neq R$, containing I. This set is ordered inductively, since the union of a chain of ideals is again an ideal (see the proof of Lemma 2.1.18). By Zorn's Lemma 6.1.9 $\mathcal{I}$ contains maximal elements and any such maximal element is clearly a maximal ideal in R containing I.

q.e.d.

Remark 6.1.14. *Since the Axiom of Choice or equivalently Zorn's Lemma is essential in Algebra to prove results such as Theorems 6.1.11 and 6.1.13, we will assume the Axiom of Choice to be true from now on without stating it explicitly.*

It remains to prove Theorem 6.1.10. For this we require the following two lemmata.

Lemma 6.1.15. *Let* $(X_\lambda : \lambda \in \Lambda)$ *a family of subsets of a given set* M *and for each* $\lambda \in \Lambda$ *let* $\preceq_\lambda$ *be a well-ordering on* X_λ *such that for* $X_\mu \subseteq X_\lambda$ *we have that* X_μ *is a segment in* X_λ *and the ordering* $\preceq_\mu$ *is the restriction of* $\preceq_\lambda$ *to* X_μ*.*

If for $\lambda, \mu \in \Lambda$ *we always have either* $X_\mu \subseteq X_\lambda$ *or* $X_\lambda \subseteq X_\mu$*, then the orderings* $\preceq_\lambda$*,* $\lambda \in \Lambda$*, induce a well-ordering* $\preceq$ *on* $X := \bigcup_{\lambda \in \Lambda} X_\lambda$*.*

Proof. Let $x, y \in X$. Then there are $\lambda, \mu \in \Lambda$ such that $x \in X_\lambda$ and $y \in X_\mu$. Since we have $X_\mu \subseteq X_\lambda$ or $X_\lambda \subseteq X_\mu$, we may assume without loss of generality that x and y are both in X_μ. We then set $x \preceq y$ if and only if $x \preceq_\mu y$. Thus $\preceq$ defines an ordering on X, such that the restriction to a subset X_λ is $\preceq_\lambda$.

We need to show that $\preceq$ defines a well-ordering on X. For this let $\emptyset \neq Y \subseteq X$. Then there must be some $\lambda \in \Lambda$ such that $Y_\lambda := Y \cap X_\lambda \neq \emptyset$. Let y_0 be a smallest element of Y_λ, which exists because $\preceq_\lambda$ is a well-ordering on X_λ. Now take an arbitrary $y \in Y$. If $y \in X_\lambda$, then we have $y_0 \preceq y$ by construction. Otherwise we have $y \in X_\mu$ for some $\mu \neq \lambda$. Since we have $y \notin X_\lambda$ we cannot have $X_\mu \subseteq X_\lambda$, wherefore it follows from the assumption that $X_\lambda \subseteq X_\mu$ and X_λ is a segment in X_mu. By Remark 6.1.6 this implies that X_λ consists only of elements $\preceq y$, so in particular we have $y_0 \preceq y$. It therefore follows that y_0 is a smallest element in Y, wherefore $\preceq$ defines a well-ordering on X.

q.e.d.

Lemma 6.1.16. *(Bourbaki's Fundamental Lemma) Let M be a set and $\mathcal{T} \subseteq \mathfrak{P}(M)$ a set of subsets of M such that $\emptyset \in \mathcal{T}$. Further suppose there is a map $p : \mathcal{T} \to M$ with $p(T) \notin T$ for all $T \in \mathcal{T}$. Then there is a subset $X \subseteq M$ and a well-ordering $\preceq$ on X such that the following properties are satisfied:*

1. *For every $x \in X$ we have $X_{\preceq x} \in \mathcal{T}$ (see Remark 6.1.6 for the notation) and $p(X_{\preceq x}) = x$.*
2. *$X \notin \mathcal{T}$.*

Proof. Let $\mathcal{F}$ denote the family of all ordered sets $(X_\lambda, \preceq_\lambda)$, $\lambda \in \Lambda$, satisfying the following properties.

(a) $X_\lambda \in \mathcal{T}$.

(b) $(X_\lambda, \preceq_\lambda)$ is well-ordered.

(c) For every $x \in X_\lambda$ we have $(X_\lambda)_{\preceq_\lambda x} \in \mathcal{T}$ and $p((X_\lambda)_{\preceq_\lambda x}) = x$.

Note that $\mathcal{F}$ trivially contains the empty set.

Furthermore we claim that $\mathcal{F}$ satisfies the conditions of Lemma 6.1.15:

For $\lambda, \mu \in \Lambda$ let

$$V := \left\{x \in X_\lambda \cap \mathfrak{X}_\mu \ : \ ((X_\lambda)_{\preceq_\lambda x}, \preceq_\lambda) = ((X_\mu)_{\preceq_\mu x}, \preceq_\mu)\right\}.$$

Then V is a segment both in $(X_\lambda, \preceq_\lambda)$ and $(X_\mu, \preceq_\mu)$. For this let $x \in X_\lambda$ such that there exists some $v \in V$ such that $x \preceq_\lambda v$. By definition of V this means that we also have $x \preceq_\mu v$ and in particular $x \in X_\lambda \cap X_\mu$. Now any $x' \in X_\lambda$ with $x' \preceq_\lambda x$ also satisfies $x' \preceq_\lambda v$. But again by the definition of V this implies $x' \preceq_\mu v$ and since $(X_\lambda)_{\preceq_\lambda v}$ and $(X_\mu)_{\preceq_\mu v}$ are equal as ordered sets, we must have $x' \preceq_\mu x$. It therefore follows that $(X_\lambda)_{\preceq_\lambda x} = (X_\mu)_{\preceq_\mu x}$ as ordered sets, so $x \in V$, wherefore V is a segment in X_λ and for symmetry reasons also in X_μ.

Indeed much more is true. We have $V = X_\lambda$ or $V = X_\mu$. If not, there would be smallest elements $x_\lambda \in X_\lambda \setminus V$ and $x_\mu \in X_\mu \setminus V$. Then we have $V = (X_\mu)_{\preceq_\mu x_\mu} = (X_\lambda)_{\preceq_\lambda x_\lambda}$ and thus because of condition (c) above

$$x_\mu = p(V) = x_\lambda \in X_\lambda \cap X_\mu$$

and thus $x_\lambda = x_\mu \in V$, which is a contradiction.

Therefore the conditions of Lemma 6.1.15 are satisfied and we obtain a well-ordering $\preceq$ on $X = \bigcup_{X_\lambda \in \mathcal{F}} X_\lambda$ induced by the well-orderings $\preceq_\lambda$, $\lambda \in \Lambda$. For every $x \in X$ there is some $\lambda \in \Lambda$ such that $x \in X_\lambda$ and we have

$$X_{\preceq x} = (X_\lambda)_{\preceq_\lambda x} \in \mathcal{T}$$

by condition (c) above, and by the same condition we have $p(X_{\preceq x}) = x$. The set X therefore satisfies point 1. of the lemma.

If we have $X \notin \mathcal{T}$, then X also satisfies point 2. and we are done. Otherwise we have $X \in \mathcal{T}$ and $x_0 := p(X) \notin X$. Consider the set $X_0 := X \cup \{x_0\}$ with the ordering $\preceq_0$ defined by $x \preceq x_0$ for all $x \in X$ and $\preceq_0 = \preceq$ on X. Then the segments of X_0 are either segments of X or X_0, so that X_0 again satisfies point 1. of the lemma. Furthermore we have $X_0 \notin \mathcal{T}$, otherwise X_0 would be one of the sets X_λ since $(X_0, \preceq_0)$ is well-ordered and therefore a subset of X which is a contradiction.

q.e.d.

We now conclude this section with the proof of Theorem 6.1.10.

Proof of Theorem 6.1.10 We start by showing that the Axiom of Choice 6.1.1 implies Zorn's Lemma 6.1.9. Let $(M, \preceq)$ be an inductively ordered set and set

$$\mathcal{T} := \{T \subseteq M \ : \ T \text{ has an upper bound } t' \in M \setminus T\}.$$

For $T \in \mathcal{T}$ let

$$X_T := \{t' \in M \setminus T \ : \ t' \text{ is an upper bound for } T\}.$$

By the Axiom of Choice there is a map

$$p : \mathcal{T} \to M, \ p(T) \in X_T \text{ for all } T \in \mathcal{T}.$$

By Lemma 6.1.16 there is a subset $X \subseteq M$ with a well-ordering $\preceq_X$ such that

1. For $x \in X$ we have $X_{\preceq_X x} \in \mathcal{T}$ and $p(X_{\preceq_X x}) = x$,
2. $X \notin \mathcal{T}$.

Note that the well-ordering $\preceq_X$ on X a priori has nothing to do with the ordering $\preceq$ on M. We show now that $(X, \preceq)$ (with the ordering from M) is totally ordered. Let $x_1, x_2 \in X$. Since $(X, \preceq_X)$ is well-ordered and thus by Remark 6.1.6 in partiular totally ordered, we may assume that $x_1 \preceq_X x_2$, so $x_1 \in X_{\preceq_X x_2}$. Since $X_{\preceq_X x_2} \in \mathcal{T}$ and $p(X_{\preceq_X x_2}) = x_2$, we find that x_2 is an upper bound for $X_{\preceq_X x_2}$ with respect to the ordering $\preceq$. In particular we have $x_1 \preceq x_2$, so $(X, \preceq)$ is totally ordered.

Since $(M, \preceq)$ is ordered inductively, the totally ordered set $(X, \preceq)$ has an upper bound $m_X \in M$. Since we have $X \notin \mathcal{T}$, X cannot have an upper bound in $M \setminus X$, wherefore $m_X \in X$. Then m_X is a maximal element of $(M, \preceq)$ because every $y \in M$ satisfying $m_X \preceq y$ would also be an upper bound for $(X, \preceq)$, thus contained in X and thus $\preceq m_X$.

We now show that Zorn's Lemma 6.1.9 implies the Well-ordering Theorem 6.1.7.

Let $M \neq \emptyset$ and consider

$$\mathcal{M} := \{(T, \preceq_T) \ : \ T \subseteq M, \ \preceq \ \text{is a well-ordering on } T\}.$$

We define the following ordering on $\mathcal{M}$,

$$(S, \preceq_S) \preceq (T, \preceq_T) \ \Leftrightarrow \begin{cases} S \subseteq T & \text{and} \\ \preceq_S = (\preceq_T)|_S & \text{and} \\ s \preceq_T t & \text{for all } s \in S, t \in T \setminus S. \end{cases}$$

With respect to this ordering $\mathcal{M}$ is ordered inductively, so by Zorn's Lemma there is a maximal element $(X, \preceq_X) \in \mathcal{M}$.

We claim that $X = M$ as a set. Otherwise there exists $m \in M \setminus X$ and we can extend the ordering $\preceq_X$ to and ordering $\preceq_0$ on $X_0 := X \cup \{m\}$ by setting $x \preceq_0 m$ for all $x \in X$. But then $(X_0, \preceq_0) \in \mathcal{M}$ is strictly larger than $(X, \preceq_X)$ with respect to $\preceq$, which is a contradiction to the maximality of $(X, \prec_X)$. Thus $\preceq_X$ defines a well-ordering on M.

Lastly we give the comparibly simple proof that the Well-ordering Theorem 6.1.7 implies the Axiom of Choice 6.1.1.

Let Λ be a set and $(X_\lambda \ : \ \lambda \in \Lambda)$ a family of non-empty sets. By the Well-ordering Theorem, the set $\bigcup_{\lambda \in \Lambda} X_\lambda$ has a well-ordering, and in particular each set X_λ contains a smallest element x_λ with respect to said ordering. But this means the cartesian product of the X_λ contains at least the element $(x_\lambda \ : \ \lambda \in \Lambda)$ and is therefore not empty.

q.e.d.

6.2 The algebraic closure*

In this section we study the algebraic closure of an arbitrary field and in particular provide a proof for Lemma 3.2.11, which we have taken for granted at several occasions throughout this course, for example when we showed that the Galois group of the splitting field of an irreducible separable polynomial acts transitively on the roots of said polynomial (see Lemma 5.5.8) or in the proof of Hilbert's Theorem 90 (see Theorem 5.4.5).

Before moving on, recall the definition of an algebraically closed field (Definition 3.2.9) and the definition of the algebraic closure of a field (Definition 3.2.10)

Definition 6.2.1. Let E/K be a field extension. Then we call

$$\mathrm{Alg}_E(K) := \{\alpha \in E \,:\, \alpha \text{ is algebraic over } K\}$$

the *algebraic closure* of K in E.

Remark 6.2.2. *It is straightforward to see that* $\mathrm{Alg}_E(K)$ *is a subfield of* E*, namely the largest extension field of* K *in* E *which is algebraic over* K*.*

The following remark perhaps seems clear intuitively.

Remark 6.2.3. *Let* E/K *be a field extension such that* E *is algebraically closed. Then* $\mathrm{Alg}_E(K)$ *is an algebraic closure of* K*.*

Proof. Let $f =\in \mathrm{Alg}_E(K)[X]$ be an irreducible polynomial. Since E is algebraically closed, we know that f has a root $\alpha \in E$. We have to show that $\alpha \in \mathrm{Alg}_E(K)$, i.e. that α is algebraic over K. Since all the coefficients of f are algebraic over K, there is a finite extension L/K containing all of them, so $f \in L[X]$. Thus the extensions $L(\alpha)/L$ and L/K are finite, wherefore $L(\alpha)/K$ is finite and hence algebraic. Since f is irreducible over $\mathrm{Alg}_E(K)$ and it has a root in $\mathrm{Alg}_E(K)$, it must have degree 1, hence $\mathrm{Alg}_E(K)$ is algebraically closed and algebraic over K by construction.

q.e.d.

Since the field $\mathbb{C}$ of complex numbers is algebraically closed by the Fundamental Theorem of Algebra (Theorem A.1.1), we obtain immediately that for instance any algebraic extension of $\mathbb{Q}$ and in particular $\mathbb{Q}$ itself has an algebraic closure. For a general field it is much less clear that there is an algebraically closed field containing it. This is however the case.

Theorem 6.2.4. *Let K be an arbitrary field. Then there exists an algebraic closure of K.*

Proof. Let $\mathcal{P} := \{f \in K[X] \setminus K \ : \ f \text{ is monic and irreducible}\}$ and

$$\mathcal{X} := \{\xi_1^{(f)}, ..., \xi_{\deg f}^{(f)} \ : \ f \in \mathcal{P}\}$$

be an (infinite) set of variables, which we will now link to the roots of each $f \in \mathcal{P}$: Recall from Corollary 1.1.13 that the coefficients of a monic polynomial are given by the elementary symmetric polynomials in its roots. With this in mind we define for $f = X^n - a_1 X^{n-1} + a_2 X^{n-2} - ... + (-1)^n a_n \in \mathcal{P}$ the set

$$R_f := \{ \sum_{i_1 < ... < i_k} \xi_{i_1}^{(f)} \cdots \xi_{i_k}^{(f)} - a - k \ : \ k = 1, ..., n\}$$

and let

$$I := \langle \bigcup_{f \in \mathcal{P}} R_f \rangle \trianglelefteq K[\mathcal{X}]$$

be the ideal in the polynomial ring $K[\mathcal{X}]$ (which has infinitely many variables) generated by all the sets R_f for $f \in \mathcal{P}$. Assuming for the moment that $I \neq K[\mathcal{X}]$, we have seen in Theorem 6.1.13, as a consequence of Zorn's Lemma 6.1.9, that I is contained in a maximal ideal $J \trianglelefteq K[\mathcal{X}]$. The factor ring $E := K[\mathcal{X}]/J$ is then a field (see Proposition 2.1.11) which contains K via the embedding $K \cong 1 \cdot K + J \subseteq K[\mathcal{X}]/J$.

Now every element in E is the root of an irreducible polynomial over K by construction (or a polynomial therein), so that E/K is algebraic. Furthermore, every irreducible polynomial $f \in \mathcal{P}$ of degree $n \geq 1$ splits into linear factors

$$f = \prod_{i=1}^{n} (X - \overline{\xi}_i^{(f)}) \in E[X].$$

This implies that E is also algebraically closed, since if α is algebraic over E, it is also algebraic over K and thus a root of a polynomial in $K[X]$. Since this polynomial splits into linear factors in $E[X]$ we have $\alpha \in E$.

It remains to show that $I \neq K[\mathcal{X}]$. If we had $I = K[\mathcal{X}]$, then we would have $1 \in I$, so we could write $1 = \sum_i a_i x_i$ for finitely many $a_i \in K$ and x_i finite products of expressions of the form $\sum_{i_1 < ... < i_k} \xi_{i_1}^{(f)} \cdots \xi_{i_k}^{(f)} - a_k$. Replacing the occuring $\xi_i^{(f)}$ to roots of f, all of these expressions become 0 and since only finitely many polynomials, say $f_1, ..., f_k$ can occur, all of these roots lie in a finite field extension L/K, the splitting field of $\prod_{i=1}^k f_i$. Therefore we have

$$1 = \sum_i a_i x_i = 0 \in L$$

which is a contradiction.

This completes the proof.

q.e.d.

We now come to the proof of Lemma 3.2.11 which we restate here.

Lemma 6.2.5. *Let $K \subseteq L \subseteq E \subseteq \overline{K}$ be algebraic extensions. Suppose we have a ring homomorphism $\varphi : L \to \overline{K}$ such that $\varphi(a) = a$ for all $a \in K$. Then there exists a ring homomorphism $\psi : E \to \overline{K}$ such that $\psi(\alpha) = \varphi(\alpha)$ for all $\alpha \in L$, i.e. ψ is an entension of φ from L to E.*

Proof. Consider the set

$$mathcalA := \{(E', \varphi') \ : \ L \leq E' \leq E, \varphi'|_L = \varphi\}.$$

Then $\mathcal{A}$ is ordered with respect to the ordering $\preceq$ defined by

$$(E_1, \varphi_1) \preceq (E_2, \varphi_2) \text{ if and only if } E_1 \leq E_2 \text{ and } \varphi_2|_{E_1} = \varphi_2.$$

This ordering is in fact inductive, i.e. every chain has an upper bound in $\mathcal{A}$: Let $\{(E_i, \varphi_i) \ : \ i \in I\}$ be some totally ordered subset of $\mathcal{A}$, then we have the upper bound $(\tilde{E}, \tilde{\varphi}) \in \mathcal{A}$ with $\tilde{E} = \bigcup_{i \in I} E_i$ and $\tilde{\varphi}|_{E_i} = \varphi_i$. By Zorn's Lemma 6.1.9 there exists a maximal element $(E', \varphi') \in \mathcal{A}$. We claim that $E' = E$. If not, then choose $\alpha \in E \setminus E'$ and extend φ' to $E'(\alpha)$ in the following way: By assumption α is algebraic over E', so it has a minimal polynomial $\mu_\alpha \in E'[X]$. This minimal polynomial splits into linear factors in $\overline{K}[X]$, so we can set $\varphi'(\alpha) = \beta$ for any $\beta \in \overline{K}$ satisfying $\mu_\alpha(\beta) = 0$. But this contradicts the maximality of (E', φ'), so the claim follows.

q.e.d.

Corollary 6.2.6. *Let $\overline{K}$ and $\widetilde{K}$ be algebraic closures of a field K. Then $\overline{K}$ and $\widetilde{K}$ are isomorphic over K.*

Proof. This follows immediately from Lemma 3.2.11 by choosing $L = K$, $E = \widetilde{K}$ and φ the identity map.

q.e.d.

Remark 6.2.7. *It is important to note that even though the formulation* “the *algebraic closure” of a field K is justified through Corollary 6.2.6, two algebraic closures of the same field are not neceessarily physically identical. This is why it is formally necessary to fix an algebraic closure $\overline{K}$ of a field K and an embedding $K \hookrightarrow \overline{K}$, which we have done implicitly throughout theses notes whenever we talked about it.*

Appendix A

The Fundamental Theorem of Algebra*

A.1 Some history

Throughout especially Chapter 1, we have assumed the famous Fundamental Theorem of Algebra.

Theorem A.1.1. *(Fundamental Theorem of Algebra) The field of complex numbers $\mathbb{C}$ is algebraically closed, i.e. every polynomial $f \in \mathbb{C}[X]$ has a root in $\mathbb{C}$.*

There are at least 100 known proofs of this result. The first attempted proof is due to d'Alembert (1746), a different approach was due to Gauß (1798). It was later noticed that both these proofs had gaps (the gap in Gauß's proof was closed in 1920 by Ostrowski). The first rigorous proof was given by Argand in 1806 and two further (essentially complete) proofs were presented by Gauß in 1816. With the study of complex analysis in the 19th century, it was realised that the Fundamental Theorem of Algebra can be derived in various very easy ways from basic principles in complex analysis, as we have seen in the exercises.

A.2 The proof

Here, we give a relatively elementary proof of the Fundamental Theorem, based on that by d'Alembert.

We begin by recording some basic facts which should be known from a basic course in Calculus or Analysis.

Lemma A.2.1. *1. Let $f \in \mathbb{C}[X]$ be a polynomial. Then the function $f : \mathbb{C} \to \mathbb{C}$, $z \mapsto f(z)$ is continuous.*

2. *For every $z \in \mathbb{C}$ and $m \in \mathbb{N}$, there exists $w \in \mathbb{C}$ such that $w^m = z$.*

3. *(Minimum principle) Every continuous, real-valued function on a compact set $S \subset \mathbb{C}$ attains its minimum in S.*

The key to the proof lies in the following result, which is sometimes referred to as *d'Alembert's Lemma* or *Argand's inequality.*

Lemma A.2.2. *Let $f = \sum_{j=0}^{n} c_j X^j \in \mathbb{C}[X]$ be a polynomial of degree $n \geq 1$ and $a \in \mathbb{C}$. If $f(a) \neq 0$, then every open disk $B_r(a) := \{z \in \mathbb{C} : |z - a| < r\}$ around a contains a point b such that $|f(b)| < |f(a)|$.*

Proof. Let $r > 0$ be arbitrary. Then every point in $B_r(a)$ can be written as $a + w$ for some $w \in \mathbb{C}$ with $|w| < r$. Then we have

$$f(a+w) = \sum_{j=0}^{n} c_j (a+w)^j = \sum_{j=0}^{n} c_j \sum_{k=0}^{j} \binom{j}{k} a^{k-j} w^j = \sum_{k=0}^{n} \left(\sum_{j=k}^{n} \binom{j}{k} c_j a^{k-j} \right) w^k.$$

Extracting the coefficient of w^0, we see that we can write

$$f(a+w) = f(a) + sum_{k=1}^{n} \left(\sum_{j=k}^{n} \binom{j}{k} c_j a^{k-j} \right) w^k f(a) + Cw^m(1 + g(w))$$

for a suitable constant $C \in \mathbb{C} \setminus \{0\}$, $1 \leq m \leq n$, and a polynomial $g \in \mathbb{C}[X]$ of degree $n - m$ satisfying $g(0) = 0$.

We now want to find upper bounds for $|Cw^m|$ and $|g(w)|$. For $|w| < \rho_1 := \sqrt[m]{|f(a)|/C}$. Therefore we have $|Cw^m| < |f(a)|$. Since $g(0) = 0$ and g is continuous, there must be some $\rho_2 > 0$ such that $|g(w)| < 1$ for $|w| < \rho_2$. If we now set $\rho := \min\{\rho_1, \rho_2\}$, we have for all w with $|w| < \rho$ the inequalities

$$|Cw^m| < f(a) \qquad \text{and} \qquad |g(w)| < 1. \tag{A.1}$$

Now let $\zeta \in \mathbb{C}$ be an mth root of $-\frac{f(a)/C}{|f(a)/C|}$ and $0 < \varepsilon < \min\{r, \rho\}$. We claim that $b := a + w_0$ with $w_0 := \varepsilon\zeta$ lies in the disk $B_r(a)$ and satisfies $|f(b)| < |f(a)|$. First note that $|\zeta| = 1$ and thus $|w_0| = \varepsilon < r$, so that indeed $b \in B_r(a)$.

We have

$$|f(b)| = |f(a + w_0)| = |f(a) + Cw_0^m(1 + g(w_0))|.$$

Defining $\delta := \frac{\varepsilon^m}{|f(a)/C|}$, we find by construction that

$$Cw_0^m = C\varepsilon^m\zeta^m = -\frac{eps^m}{|f(a)/C|} f(a) = -\delta f(a).$$

Notice that since $\varepsilon < \rho$, we obtain from (A.1) that $0 < \delta < 1$. We thus obtain using the triangle inequality

$$\begin{aligned} |f(b)| &= |f(a) + Cw_0^m(1 + g(w_0))| = |f(a) - \delta f(a)(1 + g(w_0))| \\ &= |(1 - \delta)f(a) - \delta f(a)g(w_0)| \\ &\leq (1 - \delta)|f(a)| + \delta|f(a)||g(w_0)|. \end{aligned}$$

By (A.1) we have $|g(w_0)| < 1$, so we obtain

$$|f(b)| \lneqq (1 - \delta)|f(a)| + \delta|f(a)| = |f(a)|,$$

as we claimed.

q.e.d.

We are now ready to prove the Fundamental Theorem of Algebra A.1.1.

Proof of Theorem A.1.1. Since we clearly have $\lim_{|z|\to\infty} f(z)z^{-n} = c_n$, it must be true that $|f(z)| \to \infty$ as $|z| \to \infty$. Therefore there must be some $R > 0$ such that $|f(z)| > |f(0)|$ for all $z \in \mathbb{C}$ with $|z| = R$. Now it follows from the minimum principle in Lemma A.2.1 that the continuous, real-valued function $z \mapsto |f(z)|$ attains its minimum in some point z_0 in the compact set $\overline{B_R(0)} = \{z \in \mathbb{C} \, : \, |z| \leq R\}$. Since $|f(z)| > |f(0)| \geq |f(z_0)|$, so that z_0 cannot lie on the boundary of the disk, i.e. $z_0 \in B_R(0)$. If $f(z_0)$ were not zero, then there would be an $r > 0$ such that $B_r(z_0) \subset B_R(0)$ and by d'Alembert's Lemma A.2.2 we could find $z_1 \in B_r(z_0)$ satisfying $|f(z_1)| < |f(z_0)|$, contradicting the choice of z_0. Therefore we must have $f(z_0) = 0$ and the theorem is proven.

q.e.d.

Appendix B

Existence of transcendental numbers*

In Section 3.7 we have considered general properties of transcendental field extensions. In Number Theory, it is an important and generally very hard question, whether a given real (or complex) number is algebraic or transcendental over $\mathbb{Q}$. The two most famous transcendental numbers are very probably $\pi = 3.1415926...$, the ratio of a circle's circumference to its diameter, and Euler's number $e = 2.718281828...$, the base of the natural logarithm. It is however not known whether for instance whether or not $e + \pi$ is transcendental.

The full proof that these numbers are transcendental was first found by Hermite (1873) for e and by von Lindemann (1882) for π, but they are both far too involved to sketch here. We will give two different proofs that transcendental numbers *exist.*

B.1 An abstract counting argument*

It is a well-known fact that the set of rational numbers $\mathbb{Q}$ is *countably infinite*, i.e. there exists a bijective map $\mathbb{N} \to \mathbb{Q}$, and that the real numbers $\mathbb{R}$ are *uncountable.* Both of these facts were first noticed by Cantor. We now show that transcendental numbers exist by showing that there are "more" transcendental numbers than algebraic ones. For this we need a small preparatory lemma.

Lemma B.1.1. *Let $(M_i)_{i \in I}$ be a family of sets, where I is some index set. If all M_i are at most countable and I is countable, then $\bigcup_{i \in I} M_i$ is countable.*

Proof. We may assume without loss of generality that the sets M_i are pairwise disjoint, i.e. $M_i \cap M_j = \emptyset$ if $i \neq j$. Since each M_i is countable, we can enumerate all elements in M_i by $(a_{i1}, a_{i2}, ...)$. Since I is also countable, we may assume without loss of generality that $I = \mathbb{N}$. We now arrange the elements in the sets M_i in a grid

$$\begin{array}{cccc} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}.$$

Since all the a_{ij} are distinct by assumption, we obtain a well-defined map

$$\phi : M = \bigcup_{i \in \mathbb{N}} M_i = \{a_{ij}\} \to \mathbb{N} \times \mathbb{N},\ a_{ij} \mapsto (i, j).$$

Clearly, ϕ is bijective and since $\mathbb{N} \times \mathbb{N}$ is countable, it follows that M is countable.

q.e.d.

Proposition B.1.2. *The set $\overline{Q}$ of algebraic numbers is countably infinite.*

Proof. Every algebraic number $\alpha \in \overline{\mathbb{Q}}$ is the root of a non-zero polynomial $f \in \mathbb{Q}[X]$. Therefore we may write

$$\overline{\mathbb{Q}} = \bigcup_{\substack{f \in \mathbb{Q}[X] \\ \deg f \geq 1}} \{\alpha \in \mathbb{C} \ : \ f(\alpha) = 0\}.$$

Since each of the sets $\{\alpha \in \mathbb{C} \ : \ f(\alpha) = 0\}$ is finite and therefore countable, it suffices to show that $\mathbb{Q}[X]$ is countable. We can write

$$\mathbb{Q}[X] = \{0\} \cup \bigcup_{n \in \mathbb{N}_0} \{f \in \mathbb{Q}[X] \ : \ \deg f = n\}.$$

Now each set $\{f \in \mathbb{Q}[X] \ : \ \deg f = n\}$ can be mapped bijectively to the finite cartesian product $\mathbb{Q}^\times \times \mathbb{Q}^{n-1}$. Since $\mathbb{Q}^\times$ and $\mathbb{Q}^{n-1}$ are both countable, since Cartesian products of countable sets are countable, we find that each of the sets $\{f \in \mathbb{Q}[X] \ : \ \deg f = n\}$ is countable. Therefore $\mathbb{Q}[X]$ is a countable union of countable sets and thus itself countable. This completes the proof.

q.e.d.

As an immediate corrollary we obtain the following.

Corollary B.1.3. *There exist uncountably many transcendental numbers in $\mathbb{R}$.*

B.2 Liouville's construction*

The arguments for the existence of transcendental numbers given in the previous section may be a bit unsatisfactory, since they cannot be used to produce an explicit example of a transcendental number. In this section, we want to give a construction of transcendetal numbers due to Liouville, which produces the first real number which was ever shown to be transcendental. It all relies on the following important result.

Theorem B.2.1. *(Liouville's approximation theorem) Let* $\alpha \in \mathbb{C}$ *be an algebraic number of degree* n*. Then there exists an effective constant* $c = c(\alpha) > 0$ *such that for all* $p, q \in \mathbb{Z}$*,* $q > 00$*, with* $\alpha \neq \frac{p}{q}$ *we have the inequality*

$$\left|\alpha - \frac{p}{q}\right| < c(\alpha)q^{-n}.$$

Proof. By assumption the minimal polynomial of α (over $\mathbb{Q}$) has degree n. By multiplying by the common denominator, we obtain a polynomial $f = a_nX^n + ... + a_0 \in \mathbb{Z}[X]$ of degree n satisfying $f(\alpha) = 0$. Since f is essentially the minimal polynomial of α, it is irreducible over $\mathbb{Q}$, so in particular we have $f(\frac{p}{q}) \neq 0$. Therefore we have that

$$q^n f(\frac{p}{q}) = a_n p^n + ... + a_0 q^n \in \mathbb{Z} \setminus \{0\}$$

is at least 1 in absolute value.

From the binomial formula we obtain

$$\begin{aligned} -f(\frac{p}{q}) = f(\alpha) - f(\frac{p}{q}) &= \sum_{j=1}^{n} a_j \left(\alpha^j - (\frac{p}{q})^j\right) \\ &= \left(\alpha - \frac{p}{q}\right) \sum_{j=1}^{n} a_j \left(\alpha^{j-1} + \alpha^{j-2}\frac{p}{q} + ... + \alpha(\frac{p}{q})^{j-2} + (\frac{p}{q})^{j-1}\right). \end{aligned}$$

Therefore, if $0 < |\alpha - \frac{p}{q}| < 1$, so that in particular $|\frac{p}{q}| < 1 + |\alpha|$ by the triangle inequality, we find that

$$q^{-n} \leq |-f(\frac{p}{q})| \leq \left|\alpha - \frac{p}{q}\right| \sum_{j=1}^{n} |a_j| \sum_{i=0}^{j-1} |\alpha|^i (1 + |\alpha|)^{j-1-i}.$$

Defining $1/c(\alpha) := \sum_{j=1}^{n} |a_j| \sum_{i=0}^{j-1} |\alpha|^i (1 + |\alpha|)^{j-1-i}$, which is permitted, since the sum is a real number ≥ 1. Thus all those p, q satisfying $0 < |\alpha - \frac{p}{q}| < 1$ satisfy the inequality

$$c(\alpha)q^{-n} \leq |\alpha - \frac{p}{q}|,$$

while all those with $|\alpha - \frac{p}{q}| \geq 1$ satisfy the stronger inequality $|\alpha - \frac{p}{q}| \geq q^{-n}$. Note that in either case we have $c(\alpha) \leq 1$.

q.e.d.

More informally speaking, Liouville's approximation theorem says that irrational algebraic numbers (in particular real algebraic numbers) cannot be approximated very well by rational numbers. Liouville himself used this to show the following.

Theorem B.2.2. *Let $h : \mathbb{N} \to \mathbb{N}$ be strictly increasing and satisfying the following gap condition,*

$$\limsup_{n\to\infty} \frac{h(n+1)}{h(n)} = \infty.$$

Then for any choice of digits $c_n \in \{0, ..., 9\}$, $n = 1, 2, 3, ...$, such that infinitely many are non-zero, the number $\sum_{n=1}^{\infty} c_n 10^{-h(n)}$ is transcendental. In particular this is true for the number

$$\lambda = \sum_{n=1}^{\infty} 10^{-n!} = 0.110001000000000000000000010....$$

Proof. Let $\alpha \in \mathbb{R}$ be a number as described in the theorem and suppose that α is algebraic of degree d. By truncating the series defining α at a given point N, we obtain integers

$$p_N = \sum_{n=1}^{N} c_n 10^{h(N)-h(n)}, \qquad q_N = 10^{h(N)},$$

whose quotient clearly approaches α as $N \to \infty$. We have

$$0 < \alpha - \frac{p_N}{q_N} = \sum_{n=N+1}^{\infty} c_n 10^{-h(n)} \leq \sum_{n=N+1}^{\infty} \frac{9}{10^{h(n)}} \leq \frac{9}{10^{h(N+1}} \sum_{n=0}^{\infty} 10^{-}n = 10^{1-h(N+1)}$$

by the geometric series. By Liouville's Theorem B.2.1 however we find that for a suitable constant $c(\alpha)$ we have

$$c(\alpha)10^{-dh(N)} \leq \alpha - \frac{p_N}{q_N}.$$

It follows that we must have $c(\alpha)10^{-dh(N)} \leq 10^{1-h(N+1)}$, which can only be satisfied for all sufficiently large N if

$$\limsup_{N\to\infty} \frac{h(N+1)}{h(N)} \leq d,$$

contradicting the hypothesis that this limit superior should be ∞. Therefore α must be transcendental.

Since $\frac{(n+1)!}{n} = n + 1$, which goes to infinity as n goes to infinity, $h(n) = n!$ satisfies the gap condition, and λ is indeed transcendental.

q.e.d.

Remark B.2.3. *1. There is nothing special about using base* 10 *here, one can do the same proof for any basis.*

2. It is not hard to see that there are in fact uncountably many transcendental numbers of the shape described in Theorem B.2.2. However in a sense "most" transcendental numbers are not *of this shape, meaning that the transcendental numbers from Liouville's Theorem form a set of Lebesgue-measure* 0 *in the interval* $[0, 1]$*. To see this however requires some very deep insights in Analytic Number Theory.*