One step beyond the solvable equation

The theory of the solvable and unsolvable quintic featuring Galois theory, Klein's icosahedron and elliptic curves

Sander Bessels

11th March 2006

Dedicated to Isabell, the love of my life,
to Hannah Elena Freya, our beautiful daughter
and to the unborn child that will join our family in the near future...

... and to my love of mathematics.

Contents

1 Introduction **5**

2 History **8**

2.1 Mathematics in the ancient civilizations 8

2.2 A new Italian wave of development 8

2.3 A logical barrier 9

2.4 The tragedies of Ruffini, Abel and Galois 10

2.5 The neverending story 11

3 Equations until fourth degree **14**

3.1 Solving Equations 14

3.2 The general cubic equation 16

3.3 The general quartic equation 18

4 Field theory **20**

4.1 Field extensions 20

4.2 Morphisms 27

5 Galois theory **32**

5.1 Normal extensions, Galois extensions 32

5.2 Main theorem 33

5.3 The Galois group of polynomials 37

5.4 Unsolvability of the quintic by radicals 39

6 Reduction of the quintic to normal forms **45**

6.1 Tschirnhaus transformations 45

6.2 Principal equation 46

6.3 Bring-Jerrard normal form 47

6.4 Brioschi normal form 48

7 Solving the quintic by using the icosahedron 52

7.1 The symmetry group of the icosahedron 52

7.2 Projection of the Riemann sphere to the complex plane 53

7.3 Invariants of the symmetry group of the icosahedron 56

7.4 The form problem of the icosahedron 58

8 Solving the quintic by using elliptic curves 64

8.1 Introduction to elliptic curves in the projective plane 64

8.2 The group structure 68

8.3 Torsion 73

8.4 Isomorphism and the j-invariant 80

8.5 Solution of the unsolvable quintic 82

1 Introduction

This paper was written in the first half of 2005 as my graduate thesis in mathematics and deals with the study of polynomial equations in one variable of degree ≤ 5. It is written in such a way that almost all students with some interest in algebra will be able to understand it. Especially while attending one of the introductory algebra courses, this manuscript might prove a welcome addition, because it offers elementary introductions to field theory, Galois theory, projective geometry and elliptic curves, while at the same time presenting interesting applications and illuminating examples.

Mathematically more mature readers on the other hand will be able to go through the material quickly, refreshing their memory on things that are already known, like the unsolvability of the quintic or the associativity of the group law of an elliptic curve and probably learning a few new things, like the method of reducing the quintic to the Brioschi form and the close relations between the icosahedron, various forms of the quintic and the 5-torsion on an elliptic curve.

In chapter 2, we tell some of the history of the subject. The problem of solving algebraic equations in one variable is very old, in fact it is as old as mathematics itself. Therefore the history of this subject reflects the state of mathematics from the Babylonians (300 BC) to the present time. This history is mainly based on biographies from the internet database of the university of St. Andrews[11] and has a personal human touch to it.

In chapter 3, we give the solutions of the general equations up to degree four (the quadratic, cubic and quartic equation), clearing the way for the study of the quintic.

In chapter 4, we will develop field theory, mainly because the solution of a polynomial equation by radicals can be described by a property of a field extension of the coefficient field. We develop it from scratch, so no prior knowledge of field theory is presupposed. We will see that algebraic field extensions can be seen as vector spaces, but also as polynomial rings modulo a certain ideal. We will explain how to construct root fields and splitting fields, while at the same time presenting many clear examples of them. Afterwards, we uncover the properties of field extensions, by a study of their morphisms, leading eventually to the study of the group of automorphisms.

In chapter 5, we will introduce the notion of a normal extension, in which the degree of the extension equals the number of automorphisms and show its equivalence to a Galois extension, which is defined as the splitting fields of a polynomial. Next we prove the main theorem of Galois theory, by which we will capture the symmetry of Galois extensions in a finite group. Because we have have a direct link between polynomials (and their roots), Galois extensions, and finite groups, we can now relate properties of this group to polynomials and

their field extensions (or solutions). Finally, we prove that the Galois group of the general quintic has no tower of normal subgroups that correspond to a solution by radicals. This proves it is impossible to solve the quintic by radicals. We also give an example of a specific polynomial over $\mathbb{Q}$ that is not solvable by radicals.

In chapter 6, we will describe various ways to transform the general quintic into "normal" forms. We will first eliminate some of the coefficients by Tschirnhaus transformations, reaching the principal quintic $x^5 + a_2x^2 + a_1x + a_0$ and the Bring-Jerrard quintic $x^5 + ax + a$ and finally the Brioschi quintic that expresses the coefficients in a parameter B that will play a great role in the next chapters.

In chapter 7, we will extensively treat the theory of the icosahedron. First we describe its rotation symmetry and reach the conclusion that the rotations form a group isomorphic to A_5. Then, we put the icosahedron inside a Riemann sphere and project it to a complex equatorial plane. The two rotations that generate A_5 now correspond to functions $\mathbb{C} \to \mathbb{C}$. Next we will give three complex polynomials $f(Z)$, $H(Z)$ and $T(Z)$[1] of degree 12, 20 and 30 that vanish at the "special points" (vertices, midpoints of the faces and midpoints of the edges). These polynomials are invariant under A_5 and satisfy the remarkable relationship $1728f^5 = H^3+T^2$. Afterwards we will describe the "form problem", which for given values for f, H and T, (satisfying the relationship) asks for a solution for Z. Finally, by linking the Brioschi parameter B to the invariants f, H and T, we will prove the equivalence of the form problem to the problem of solving the Brioschi quintic.

In chapter 8, we will develop the theory of elliptic curves. We start by drawing some cubic curves, embedded in $\mathbb{R}^2$ to give a geometrical intuition. Afterwards, we give an introduction to projective geometry, also shedding some more light on our use of homogeneous coordinates. We define an elliptic curve as a non-singular cubic curve in the projective plane and give a full description of the group structure. The points on the curve of order n form a subgroup $Z_n \times Z_n$, called the n-torsion. We will give a detailed description of the 2-, 3-, 4- and 5-torsion, spending attention to the algebra, while keeping in close contact to the geometry. Finally all of the elements in this work come together by linking the Brioschi parameter B (and thus f, H and T) to the j-invariant of an elliptic curve and using Galois theory to prove that the splitting field of the quintic is equal to the splitting field of $x(P) + x(2P)$ for P a 5-torsion point.

The use of 5-torsion points on an elliptic curve to solve the quintic has a nice analogy with the use radicals of to solve equations of degree lower than 5. Using radicals means that we allow a solution to $x^n = a$ to appear in our solution. By using the roots of unity, or in other words, the solutions to $x^n = 1$, we then find all solutions to the equation. The analogy with solving the quintic equation is that instead of the "torsion equation" $x^n = a$ on $\mathbb{C}^*$ for $n < 5$, we need the

[1] In fact we will use homogeneous (projective) coordinates, but this boils down to the same thing.

torsion equation $nP = Q$ for $n = 5$ on an elliptic curve $E : y^2 = x^3 + ax + b$. The parameter Q can be set to zero, but the parameters a, b for E depend on the equation.

The conclusion is that solving the quintic is equivalent to solving the form problem of the icosahedron and to finding the 5-torsion on an elliptic curve.

2 History

2.1 Mathematics in the ancient civilizations

Since the dawn of mathematics, mathematicians have occupied themselves with solving equations. This was done for their practical reasons such as economics and architecture, but also since the very beginning as a goal of its own. Motivated by pure curiosity, mathematicians pushed the limits of what equations could still be solved further and further.

The Babylonians (± 400 BC) had already developed some arithmetic for problems that in modern terms correspond to quadratic equations. They totally lacked however the notion of equation and negative numbers. The ancient Greek, or more specifically Euclid (± 300 BC), approached equations very much from a geometrical point of view. Plane geometry has the great advantage of being easy to imagine. However if the complexity rises, a higher level of abstraction, in particular equations with unknown variables and negative, are essential for further development.

Brahmagupta (598-665 AD), a Hindu mathematician, was the first to use unknown variables (most of the times the first letter of a color) and negative numbers.

The Arabs however did not know from these advances and still did not use negatives or zeros. However al-Khwarizmi (± 800) made the distinction between "roots" x, "squares of roots" x^2 and "numbers" to write down quadratic equations. His method for solving these equations rests on the geometrical method of completing the square.

In 1145 Savasorda published the first book in Europe that gives the complete solution of the quadratic equation.

2.2 A new Italian wave of development

Around 1500 a new wave of development started in Italy, where Luca Pacioli discussed quartic equations for the first time, although he was not able to solve all of them.

Soon afterwards, Scipione dal Ferro (1465-1526) solved all cubic equations of the form $x^3 + ax = b$. However without the use of the Hindu concept of negative numbers, he probably was not able to perform the substitution $y = x - c$ to bring all cubics into this form.

Instead of letting the world know about his breakthrough, he kept his discovery a complete secret until short before his death, when he told his secret to his student, Antonio Fior, who was not so good in keeping secrets and soon rumors

started to spread that the cubic was solved. Prompted by the rumors, Tartaglia managed to solve $x^3 + ax^2 = b$ and he made no secret of his discovery. Fior challenged Tartaglia to a public contest in which each gave the other 30 problems they had to solve within 40 or 50 days. Fior gave Tartaglia only problems of the form $x^3 + ax = b$, which he though Tartaglia was unable to solve. He solved them all in two hours and was declared the winner. Shortly before the problems were to be collected, Tartaglia discovered the general method for solving all cubics.

Cardano was just writing a book Practica Arithmeticae (1539) about the subject and was very interested in Tartaglia's discovery, so he invited him to a visit in Milan. After much persuading, Tartaglia revealed his secret to Cardano under the promise that he would not publish it, until Tartaglia published it himself. Cardano did not keep his promise and published it anyway. Even nowadays he often gets the credits for this discovery.

The quartic finally, was first solved by Lodovico Ferrari, a student of Cardano. This solution marked to some extend an endpoint in the process. The development of mathematics did not stop, it continued more actively than ever before, but despite great effort, solutions to higher degree equations did not come. It took centuries before it was realized why the quartic was the equation of highest degree that was soluble by radicals in its most general form.

2.3 A logical barrier

An important advancement was made by Baron Ehrenfield Walter von Tschirnhaus (1651-1708). He proposed to reduce some of the terms of the general equation. His ultimate goal was to give give a number of transformations that would eliminate all but the first and last term. The equation $x^n + c = 0$ could then be solved and as long as all of the transformations were of degree $< n$, the problem was solved by induction. He did not realize however that finding the coefficients of the system of all required transformations, was equivalent to the solution of an equation that rapidly increases in degree if n becomes large. Leibnitz did realize this and wrote in one of his letters that the total degree would become $(n-1)!$. His method could therefore never reach its intended goal of solving all polynomial equations in one variable. Cancelling the first few terms however was very useful and is still used today.

During the first half of the eighteenth century, more and more mathematicians realized that the old methods were unsuccessful and began working with the roots of the equation instead of with the coefficients. It seems very obvious to us now that if $x = a_1, a_2, \ldots$ are solutions to an equation, that equation must be equal to $c\prod(x - a_i)$, but back then, working formally with the unknown roots was a revolutionary new way of thinking that turned out to be the key element in many further developments.

Leibnitz used it to give the first algebraic proof of the method for solving the cubic by simply constructing the cubic from its roots with the formula above. All previous proofs were geometric.

Newton found formulas, now called Newton's identities, that express the coefficients in terms of the sums of the roots to an equation. This very strong connection between the roots and the coefficients actually leads to the conclusion that if the coefficients of a polynomial are general, or algebraically unrelated, the roots also satisfy as few relations as possible.

2.4 The tragedies of Ruffini, Abel and Galois

Ruffini (1756-1822) was the first to prove that the general fifth degree equation was unsolvable by radicals (square roots, cube roots, etc.). In 1799 he published a book with the title "General theory of equations in which it is shown that the algebraic solution of the general equation of degree greater than four is impossible" in which he first had to invent group theory, all by himself. He introduced the notions of the order of an element, conjugacy, the cycle decomposition of elements of permutation groups and the notions of primitive and imprimitive to finally give the proof that S_5 is not a solvable group and thus that the quintic cannot be solved by radicals[2].

Strangely though, his book was completely ignored by the whole mathematical community. In 1801 he sent a copy to Lagrange, who was the leading expert in the field, hoping for a review. Lagrange however did not respond, so because he was unsure if Lagrange received it, he sent him another copy of the book, asking if he might have erred himself in any of the proofs. Also if the things he wrote were already known, or if the book was useless for some other reason he prayed that it was pointed out to him. Lagrange did not respond.

Some mathematicians accepted Ruffini's proof, not because they understood it. Pietra Poali, professor in Pisa, for instance accepted it for dubious nationalistic reasons, mainly because Ruffini was Italian and developed a theory without the help of non-Italians. Others did not accept it and raised false objections. Ruffini figured that his book was too difficult and it has to be admitted that it was very revolutionary and not easy to understand.

In an attempt to make his ideas more transparent, he published further proofs in 1808 and 1813, which were in fact very elegant and far ahead of their time. Nobody responded.

The only mathematician who was influenced by Ruffini's book was Cauchy. Cauchy wrote a book on permutation groups between 1813-1815 in which he generalized many of Ruffini's ideas. Cauchy was probably the worst of all math-

[2]The proof contained a small gap that he would no doubt have been able to close, had somebody pointed it out to him.

ematicians in giving credit to others, but he did write to Ruffini, one year before Ruffini's death

... your memoir on the general resolution of equations is a work which has always seemed to me worthy of the attention of mathematicians and which, in my judgement, proves completely the impossibility of solving algebraically equations of higher than the fourth degree.

This is probably the only credit Ruffini ever got for his work.

Because Ruffini's ideas remained obscure to the mathematical community, the fact that the quintic is not solvable by radicals had to be discovered for a second time. In 1824 the Norwegian mathematician Niels Henrik Abel (1802-1829), who studied Cauchy's work on permutation groups, proved the impossibility of solving the general quintic in radicals. He thought this result would be impressive enough to gain respect from famous mathematicians in Germany and France, so he printed the proof on a pamphlet and took it with him on his travels through Europe. However, he had a lot of difficulties to get his ideas accepted. Gauss for instance was very unpleased with Abel's negative result on the quintic. It is uncertain why Gauss took this attitude towards Abel's work since he certainly never read it; the paper was found unopened after Gauss' death. Abel also made many other important contributions to algebra, especially to the theory of elliptic functions, but despite his great talent, the acknowledgement for his work came too late. Abel died in poverty at the age of 28.

Evariste Galois (1811-1832) also realized that the general quintic was unsolvable by radicals and he was the first of complete the process of capturing the "ambiguity" of the roots in a group, now called the Galois group. He wrote his ideas in a manuscript that he finished the night before he died tragically in a duel for his beloved Stéphanie.

Galois' brother and his friend Chevalier copied his mathematical papers and sent them to Gauss, Jacobi and others. It had been Galois' wish that Jacobi and Gauss should give their opinions on his work. No record exists of any comment of these men. However, by a strange coincidence and a great deal of luck, the papers reached Liouville, 11 years after Galois' death. Liouville announced that he had found in Galois' papers a concise solution of the problem: Given an irreducible equation of prime degree, decide whether or not it is soluble by radicals. Three years later he published Galois' work and the world could finally behold the beauty of Galois' theory.

2.5 The neverending story

Now it seems the story is over. The quintic has been proved to be unsolvable, so that's that. Still there is another chapter in the history of the theory of the quintic.

It starts with the independent discovery of the theory of elliptic functions by Abel and Jacobi in the 1820s. Instead of studying elliptic integrals, which arise in the determination of the arc length of an ellipse, they studied their inverse functions and found that they possess remarkable properties, like the fact that they are doubly periodic complex functions. As soon as they learned from each others results, they started to compete with each other, which resulted in many publications and a rapid development of a new branch of mathematics.

On 9 February 1828, Legendre, who was the leading expert on elliptic integrals until Jacobi and Abel surpassed him, wrote in a letter to Jacobi:

> *It gives me great satisfaction to see two young mathematicians such as you and [Abel] cultivate with such success a branch of analysis which for such a long time has been my favorite topic of study but which had not been received in my own country as well as it deserves. By your works you place yourselves in the ranks of the best analysts of our era.*

At first this theory was not understood by many others, but as the 19th century progressed, more and more mathematicians started to realize the importance of Abel's and Jacobi's work.

In 1858 this theory inspired Hermite to make a renewed attempt to solve the quintic. He first noted that the cubic can be solved by using period division by three on the periodic sine function. More precisely, the roots of

$$x^3 - 3x + 2a, \text{ where } a = \sin(\alpha)$$

can be described as

$$2\sin(\frac{\alpha}{3}),\ 2\sin(\frac{\alpha + 2\pi}{3}) \text{ and } 2\sin(\frac{\alpha + 4\pi}{3}).$$

He then designed a method to solve the Bring-Jerrard quintic

$$x^5 + x + a,$$

with period division by five on an elliptic function, which is doubly periodic. He also knew that every quintic equation can be reduced to one in Bring-Jerrard form by using only radicals, which means he could find a solution to all quintic equations, not with only radicals, but with elliptic functions. The quintic was finally solved. Still this is not the end of the story.

In 1878, Gordan described an alternative method to solve the quintic, based on invariant theory. He did not use the Bring-Jerrard quintic, like Hermite, but the Brioschi quintic

$$x^5 + 10Bx^3 + 45B^2x + B^2 = 0.$$

It has only one parameter and can be obtained from the general quintic by transformations using only radicals, just like the Bring-Jerrard quintic. Gordan related the Brioschi quintic to the symmetry functions of the icosahedron.

Shortly afterwards, also in 1878 Kiepert designed an algorithm for the solution of the quintic with elliptic functions. Later however Kiepert's work appeared to have been forgotten for more than a century until the Georgian chemist and mathematician R. Bruce King read the article of Kiepert and successfully implemented the algorithm on a modern PC in 1996. To quote King:

> *This work lay fallow for more than a century since the algorithm for roots of the general quintic equation appeared intractable before the era of computers. Many of the key ideas appear to have been forgotten by the subsequent generations of mathematicians during the past century so that some of the underlying mathematics has the status of a lost art.*

The work of Kiepert was probably overshadowed by that of Klein, who wrote the very successful book "Vorlesungen über das Ikosaeder", in which he presents the relations between the quintic and the icosahedron and a solution to the quintic in terms of elliptic and hypergeometric functions.

It also has to be mentioned that in 1926, Dickson worked out some of Klein's ideas in the book Modern Algebraic Theories and made the key ideas accessible to undergraduate students. After Dickson however, interest started to diminish. Serre wrote in 1978 that Dickson's book was still the most modern treatment of the theory. Recently however, with the increasing interest in elliptic curves, the interest in Klein's theory is also rising.

A very nice recent result, published by Edray Goins in 2003, is that some of the elliptic curves that are attached to quintic equations have remarkable properties. They are so called $\mathbb{Q}$-curves and they have an "absolutely irreducible mod 5" representation.

Results like this show that, even though a piece of mathematics has been studied for centuries, or even millennia, there is always the possibility of further research and deeper understanding.

3 Equations until fourth degree

3.1 Solving Equations

An important first step we have to make before we can start solving an equation is to begin with a field K. We need this field for the coefficients of our equation, but also for the multiplication and the addition. The field gives us everything we need to write down a polynomial equation in one variable x. For instance:

$$2x^2 + 3x^3 = -3x - 4 - x^2 \tag{1}$$

To solve this equation, we first reduce the equation to zero, sort the terms by powers of x and divide the equation by the coefficient of the largest power of x. In our example

$$x^3 + x^2 + 4/3 = 0 \tag{2}$$

The problem of solving an equation can thus be reduced to the problem of finding the zeroes (roots) of a polynomial with the first coefficient equal to 1.

Definition 3.1

$$x^n + a_1 x^{n-1} + ... + a_{n-1}x + a_n,$$

where $a_1, \ldots, a_n \in K$ is called a *monic polynomial of degree n over K*.

The roots of a first degree polynomial always lie in K, so solving first degree equations requires nothing else than the elements and operations already defined in K. With second degree polynomials however it can occur that the roots lie outside of K. To solve this problem, we extend K to a larger field L that, apart from K, also contains at least one of the roots of f. Of course we also need a lot of other elements to make L into field. The next example is meant as a preview of some of the terminology and concepts that will play a great role in the next chapters.

Example 3.2 Define f over $\mathbb{Q}$ as $f(x) = x^2 - 2$. Find solutions in $\mathbb{Q}$, or construct some field extension that contains them.

We do *not* suppose that $\sqrt{2}$ is already a known element of some large field like $\mathbb{R}$ or $\mathbb{C}$. We first prove that we cannot find a solution in $\mathbb{Q}$.

Proof: If there is an element $\alpha \in \mathbb{Q}$ such that $\alpha^2 = p^2/q^2 = 2$ for two relatively prime integers p and $q \neq 1$ then $p^2 = 2q^2$, and so p must have 2 as a divisor, but that means p^2 must have 2^2 as a divisor, so q must also have at least one 2 as a divisor, making p and q no longer relatively prime, which is a contradiction. The conclusion is that $\alpha \notin \mathbb{Q}$.

So, because no root of f has yet been defined in $\mathbb{Q}$, we can give one a name, let's say $\sqrt{2}$. We extend our field of coefficients $\mathbb{Q}$ to a larger field called $\mathbb{Q}(\sqrt{2})$ that apart from $\mathbb{Q}$ also contains $\sqrt{2}$. It turns out that $\{a + b\sqrt{2} | a, b \in \mathbb{Q}\}$ is a field that fulfills the criteria and what's more, it also contains the other root. The other root is namely the additive inverse of $\sqrt{2}$, $-\sqrt{2}$. At first sight it looks like we have a negative and a positive root, but actually this difference is just in the name. We are really free to interchange the names of the two roots.[3]

More precisely, $\mathbb{Q}(\sqrt{2}) = \{a+b\sqrt{2} | a, b \in \mathbb{Q}\}$ and $\mathbb{Q}(-\sqrt{2}) = \{a+b(-\sqrt{2}) | a, b \in \mathbb{Q}\}$ are isomorphic field extensions of $\mathbb{Q}$, under an isomorphism that is well defined by stating that it sends $\sqrt{2}$ to $-\sqrt{2}$ and leaves $\mathbb{Q}$ fixed. So the only difference in these two extensions is just the names of the elements, not the algebraic structure! These concepts are hints towards a bigger theory, explaining the close connection between roots of polynomials, field extensions and their isomorphisms. That theory will be studied later in much more detail and is the main subject of the first part of this paper.

Definition 3.3 A root of the polynomial $x^n - a$ over some field K containing a that lies in some field extension L is written as $\sqrt[n]{a}$. It is called an *n-th root of a* and $\sqrt[n]{}$ an *n-th root function*. We can also omit the part *function* or *of a* if it is clear or unimportant weather we mean the element or the function. In particular $\sqrt{} = \sqrt[2]{}$ is called a *square root* and $\sqrt[3]{}$ a *cube root*.

Note that there can be more than one n-th root in the field L (maximally n). In particular if L contains n different *n-th degree roots of unity*, which are solutions to the equation $x^n = 1$, called $\zeta_1, \ldots, \zeta_n$, the other roots of $x^n - a$ can be written as $\zeta_i \sqrt[n]{a}$ for $i = 1 \ldots n$.

As we will prove later, as long as $x^n - a$ is irreducible over K (which means that we cannot write it as a product of two polynomials with degree > 1 over K), the choice of an n-th degree root in some larger field L will not change (up to isomorphism) our field extension $K(\sqrt[n]{a})$ that we need to build to find a root of $x^n - a$. In fact we will see later that all fields $K(\alpha)$, with $\alpha \in L$ different roots of an irreducible polynomial, are isomorphic in much the same way as $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(-\sqrt{2})$ are isomorphic.

We saw that with the help of a square root function (and the fact that $\mathbb{Q}$ already contains the two second degree roots of unity, 1 and -1), we could find the roots of $x^2 - 2$ in the previous example. In fact it is a well known fact that square roots are sufficient to write down the solution to the general second degree equation over a general field K (provided $\mathrm{char}(K) \neq 2$, which means as much as $1 + 1 \neq 0$, so we can divide by 2 in K).

$$p(x) = x^2 + a_1 x + a_0 = 0 \tag{3}$$

[3] It is also possible to introduce $\sqrt{2} \in \mathbb{R}_{>0}$ in other ways, so that we can make the distinction between positive and negative.

has as a solution

$$x_{\pm} = \frac{-a_1}{2} \pm \frac{1}{2}\sqrt{a_1^2 - 4a_0} \tag{4}$$

$\Delta = a_1^2 - 4a_0$ is called the discriminant of p and it is easy to see that our field extension and the number of solutions in some extension L of K depend entirely on Δ. $\Delta = 0$ means there exist 1 solution in L, $0 \neq \sqrt{\Delta} \in L$ leads to 2 solutions in L and $\sqrt{\Delta} \notin L$ implies no solutions in L.

Similarly, the general third degree equation can be solved using a square root and a cube root. We shall look in more detail at this solution, first discovered by Tartaglia and, by breaking a promise of secrecy, first published by Cardano in detail.

3.2 The general cubic equation

We look for a solution of the equation

$$x^3 + a_1x^2 + a_2x + a_3 = 0$$

The coefficients a_1, a_2 and a_1 are taken from a general field K, char$(K) \neq 2, 3$. First we eliminate the x^2-term by a substitution

$$x = z - a_1/3$$

This leads to the equation

$$z^3 + b_2z + b_3 = 0$$

where

$$b_2 = \frac{3a_2 - a_1^2}{3}$$

$$b_3 = \frac{2a_1^3 - 9a_1a_2 + 27a_3}{27}$$

Now we make things seemingly more complicated by substituting

$$z = u + v$$

which leads to:

$$u^3 + v^3 + 3(uv + b_2/3)(u + v) + b_3$$

This equation will be satisfied if both of the following relationships hold

$$u^3 + v^3 = -b_3$$
$$uv = \frac{-b_2}{3}$$

Eliminating v from the above equations leads to the sixth degree equation

$$u^6 + b_3 u^3 - b_2^3/27 = 0$$

which is a *quadratic* equation in u^3. Using the formula for the second degree equation leads to

$$u^3 = \frac{-b_3}{2} \pm \sqrt{\frac{b_3^2}{4} + \frac{b_2^3}{27}}$$

Having to use the formula for the second degree equation, means we need an element that is the square root of the discriminant Δ in that formula. If that element is not yet present in our field K (or in other words: if Δ is not a square in K) we need to add this element to K, giving us an extension of K to $K(\sqrt{\Delta})$). Note that there are up to two possible choices for this square root, but again, just as in example 4.1, this choice doesn't influence $K(\sqrt{\Delta})$ up to isomorphism.

Now it is clear that the u we are searching for is a cube root of an element of $K(\sqrt{\Delta})$. We use our definition of the cube root and note that if $K(\sqrt{\Delta})$ doesn't contain any cube root of this element, we need to add one. We will later prove that if $K(\sqrt{\Delta})$ doesn't contain a cube root, there exist three isomorphic extensions that do contain such a cube root, for now we will assume it. Note further that adding one cube root doesn't necessarily mean our final extension contains all the roots! The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ for instance only contains one root of $x^3 - 2$. We can solve this problem by adding, apart from one cube root, also a third degree root of unity (unequal to 1) to our field. Its square will then give the other one, thus providing us with all third degree roots of unity and thus with all the u's we need.

What remains is to calculate back from u via v and z to x, but because all the equations are linear, we don't encounter any more problems there.

To make things more concrete we turn to $K = \mathbb{Q}$ and try to find the solutions in the already given field $\mathbb{C}$. There are up to two solutions for the square root and up to three for the cube root in $\mathbb{C}$. By convention we take the square and cube root with the smallest complex argument and find the other square root to be minus the first and the other two cube roots with the use of $\zeta_3 = e^{2\pi i/3}$, which is a primitive cube root of unity (primitive means that ζ_3^i for $i = 1, 2, 3$

give all the three different roots of unity). This will give all the solutions for u and thus for x.

Note finally that there are possibly six different solutions for u. Won't this give too many solutions? As we will see, the answer is no.

For convenience define p to be $-b_3/2$ and q to be $\sqrt{b_3^2/4 + b_2^3/27}$. We can now write down our six solutions for u:

$$u = \begin{array}{ccccc} \sqrt[3]{p+q} & , & \zeta_3\sqrt[3]{p+q} & , & \zeta_3^2\sqrt[3]{p+q} \\ \sqrt[3]{p-q} & , & \zeta_3\sqrt[3]{p-q} & , & \zeta_3^2\sqrt[3]{p-q} \end{array}$$

which have to satisfy

$$u^3 + v^3 = 2p$$
$$uv = \sqrt[3]{p^2 - q^2}$$

Now it is clear that there can only be three solutions for $z = u + v$, namely:

$$z_1 = \sqrt[3]{p+q} + \sqrt[3]{p-q}$$
$$z_2 = \zeta_3\sqrt[3]{p+q} + \zeta_3^2\sqrt[3]{p-q}$$
$$z_3 = \zeta_3^2\sqrt[3]{p+q} + \zeta_3\sqrt[3]{p-q}$$

3.3 The general quartic equation

To conclude this chapter, let's look at the general fourth degree equation, which was first solved by Ferrari, a student of Cardano.

$$x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 = 0 \tag{5}$$

As we will see, this does not give rise to new functions. Square roots and cube roots are sufficient to formulate the solution.

Determine the numbers a, b and k such that

$$x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 + (ax+b)^2 = (x^2 + \frac{a_1}{2}x + k)^2 \tag{6}$$

by comparing the coefficients of the equal powers of x. This leads to three equations that have to be solved simultaneously:

$$\begin{aligned} a^2 + a_2 &= 2k + a_1^2/4 \\ 2ab + a_3 &= ka_1 \\ b^2 + a_4 &= k^2 \end{aligned} \tag{7}$$

After eliminating a and b, the solution to this set of equations can be found to be

$$k^3 - \frac{1}{2}a_2k^2 + \frac{1}{4}(a_1a_3 - 4a_4)k + \frac{1}{8}(4a_2a_4 - a_3^2) = 0$$

This cubic equation is called the *Resolvent Cubic* and it can be solved with the solution of the general cubic. Substituting k in (7) gives a and b and all we need to do is combining (5) and (6) to get

$$\left(x^2 + \frac{a_1}{2}x + k\right)^2 = (ax + b)^2$$

Taking square roots gives the following quadratic equations:

$$x^2 + \frac{a_1}{2}x + k = ax + b$$
$$x^2 + \frac{a_1}{2}x + k = -ax - b$$

The four roots can now be found easily by using the formula for the general second degree equation.

So it turns out that we can solve the general quadratic equation with square roots and the general cubic and quartic equations with square and cube roots. A logical guess would be to state that all n-th degree equations can be solved with n-th degree roots. This guess turns out to be wrong. As we will see later, it fails for $n = 5$, but before we can understand why, we first need more field theory.

4 Field theory

4.1 Field extensions

In this chapter we will develop some theory around field extensions. We start with some definitions.

Definition 4.1 A *ring* is a commutative group under addition together with a multiplication that is associative and has a unity element, together with the distributive laws $x(y+z) = xy + xz$ and $(x+y)z = xz + yz$.

Example 4.2 The set $K[X]$ of all polynomials over a field K is a ring.

An ideal I is a non-empty subset of a ring R that is closed under $+$ and for which $ab \in I$ for all $a \in I$ and $b \in R$. If it can be generated by a finite set of elements $a_1, \ldots, a_n$, then it is written as $(a_1, \ldots, a_n)$.

Definition 4.3 A *field* is a non-zero ring with commutative multiplication for which every non-zero element has a multiplicative inverse.

The only two ideals in a field F are (0) and $(1) = F$.

Definition 4.4 If K and L are fields, L is called a *field extension* of K if $K \subset L$. We write "L/K" and read "L over K", or "L is an extension of K".

Definition 4.5 An element $\alpha \in L$ (with L an extension of K) is called *algebraic over* K if $\exists f \in K[X] : f(\alpha) = 0$. If every element of L is algebraic over K, L/K is called an *algebraic extension of* K.

Example 4.6 Some elements of $\mathbb{C}$ are algebraic, such as $\sqrt{2}$, $3 + \sqrt[3]{2}/8$, $3.5^{4.1}$ and $e^{\frac{2\pi i}{n}}$ (a root of $x^n - 1$), but not all of them. For instance π is non-algebraic (transcendental) over $\mathbb{Q}$. To prove that a number is transcendental can be very hard, but to prove that there are many transcendental numbers is easy, for there are only countably infinitely many numbers algebraic over $\mathbb{Q}$, but uncountably infinitely many complex numbers.

Definition 4.7 Let K be a field. L is called a *vector space over* K or a K-*vector space*, if L is an abelian group for $+$ together with a multiplication $K \times V : (x, v) \mapsto xv$ that satisfies the following conditions:

a) If $1 \in K$, then $1v = v$ for all $v \in V$
b) If $c \in K$, then $c(v + w) = cv + cw$ for all v, $w \in V$
c) If x, $y \in K$, then $(x + y)v = xv + yv$ for all $v \in V$
d) If x, $y \in K$, then $(xy)v = x(yv)$ for all $v \in V$

Definition 4.8 A set of elements in a vector space L is called a *basis* of L if they are linearly independent over K and if every element of L can be written as a linear combination of these elements over K. The number of elements in a basis is called the *dimension* of L.

Theorem 4.9 *If L/K is a field extension, then L is a K-vector space.*

Proof: L is a field, so certainly an abelian group. The 1 in K is also the 1 in L, because $K \subset L$, which proves a). Finally the multiplication of elements of L with those in K happens completely in L and L is a field, so it satisfies b), c) and d). □

Definition 4.10 If L/K is a field extension such that L is a finite-dimensional vector space over K, then L/K is called a *finite field extension*. We write the dimension $\dim_K L$ as $[L : K]$ and call it the *degree* of L over K.

Example 4.11 $\mathbb{C}/\mathbb{R}$ is finite with basis $\{1, i\}$, $\mathbb{C}/\mathbb{Q}$ is not finite and $\mathbb{Q}(\sqrt{2})$ is finite with basis $\{1, \sqrt{2}\}$.

Theorem 4.12 *For an extension L/K, $L = K \iff [L : K] = 1$*

Proof: If $L = K$, then $\{1\}$ $(1 \in L)$ is a basis for L as a vector space over K, so $[L : K] = 1$. In the other direction we have $[L : K] = 1$, so $\{\alpha\}$ with $\alpha \in L$ is a basis of L over K and because L is a field and thus contains a unity, the first condition for a vector space gives us $1_K \alpha = 1_L$, so $\alpha = 1$, so $L = 1K = K$. □

Theorem 4.13 *If L/K is finite, then L/K is algebraic.*

Proof: Suppose $[L : K] = n$ For all $\alpha \in L$ the elements $1, \alpha, \ldots, \alpha^n$ must be linearly dependent over K, otherwise $[L : K] \geq n + 1$. And this linear dependency gives us a relation $a_0 + a_1\alpha + \ldots + a_n\alpha^n = 0$, with $\alpha_i \in K$ and that means exactly that α is a root of the polynomial $f(X) = \sum_{i=0}^{n} a_i X^i \in K[X]$, so every $\alpha \in L$ is root of a polynomial over K and therefore algebraic. □

Theorem 4.14 *If $\alpha \in L$ is algebraic over K then the monic polynomial of smallest degree over K that has α as a root is unique.*

Proof: $I = \{f \in K[X] | f(\alpha) = 0\}$ is a non-zero ideal of $K[X]$. Suppose $0 \neq p \in I$ is a polynomial of smallest degree and $f \in I$, then by the Euclidian algorithm for polynomials [10, Lang, p.113], we can find polynomials $r, s \in K[X]$ such that $f = rp + s$, with $\deg s < \deg p$. Now $s = f - rp$ must also be in I because I is an ideal. Since p is of minimal degree, s must be zero and it follows that $f = rp$, so p is a generator of I.

If p_1 and p_2 are generators of I, then $p_1 = qp_2$, so $\deg p_1 \leq \deg p_2$ and by symmetry $\deg p_2 \leq \deg p_1$, so q must be constant and can be chosen in such a way

that the generator becomes monic. So there exists a unique monic polynomial of smallest degree that is a generator of the ideal of all polynomial for which α is a root. □

Definition 4.15 This polynomial is called the *minimal polynomial of α over K* and the degree of p is called the *degree of α over K*.

Theorem 4.16 *The minimal polynomial p of $\alpha \in L/K$ is irreducible over K.*

Proof: Suppose p reducible, then $p = fg$, with deg $f <$ deg $p >$ deg g. Furthermore $0 = p(\alpha) = f(\alpha)g(\alpha) \Rightarrow f(\alpha) = 0$ or $g(\alpha) = 0$, so (p) must contain an element with degree smaller than deg p, contradiction. □

Theorem 4.17 (Eisenstein's criterion) *Let*

$$f(X) = a_n X^n + \ldots + a_0$$

be a polynomial of degree ≥ 1 with integer coefficients. Let p be a prime and assume for all $i < n$

$$a_n \neq 0 \bmod p, \quad a_i = 0 \bmod p, \quad a_0 \neq 0 \bmod p^2.$$

Then f is irreducible over $\mathbb{Q}$.

Proof: For the proof see [10] Lang, p.139-140. □

Example 4.18 Let p be a prime number and ζ_p a p-th degree root of unity unequal to 1. ζ_p is a root of $X^p - 1$ and because it is not 1, also a root of

$$f(X) = \frac{X^p - 1}{X - 1} = X^{p-1} + \ldots + X + 1.$$

We want to show that this is the minimal polynomial of ζ_p over $\mathbb{Q}$, so we need to show it is irreducible over $\mathbb{Q}$. We substitute $X = Y + 1$ and find after some easy calculations that

$$g(Y) = \frac{(Y+1)^p - 1}{Y + 1 - 1} = \sum_{i=0}^{p-1} \binom{p}{i} Y^{p-i-1}.$$

Since $p | \binom{p}{i}$ for $i \neq 0, p$, Eisenstein's criterion implies that $g(Y)$ is irreducible over $\mathbb{Q}$, so therefore f is irreducible over $\mathbb{Q}$.

Remark 4.19 Suppose L/K a field extension and $\alpha \in L$. Just like we can make the polynomial ring $K[X]$, we can also make the ring

$$K[\alpha] := \{f(\alpha) | f \in K[X]\} \subset L \tag{8}$$

generated by α. If α is algebraic over K of degree n, this ring has $1, \alpha, \ldots \alpha^{n-1}$ as a basis.

Similarly, we define the ring generated by a finite number of elements $\{\alpha_1, \ldots \alpha_n\}$ of L by evaluating polynomials in n variables over K in $\alpha_1, \ldots \alpha_n$.

$$K[\alpha_1 \ldots \alpha_n] = \{f(\alpha_1, \ldots, \alpha_n) | f \in K[X_1, \ldots, X_n]\} \tag{9}$$

Theorem 4.20 *If $\alpha \in L/K$ is algebraic with minimal polynomial p, then*

$$K[X]/(p) \cong K[\alpha] \tag{10}$$

Proof: $I = (p) = \{f \in K[X] | f(\alpha) = 0\}$ is the kernel of the evaluation-in-α-homomorphism $\mathrm{ev}_\alpha : K[X] \to K$, $f(X) \mapsto f(\alpha)$, so $K[X]/(p) \cong \mathrm{im}\ \mathrm{ev}_\alpha = \{f(\alpha) | f \in K[X]\} = K[\alpha]$ under an isomorphism that evaluates the polynomials of $K[X]/(p)$ in α. [4] □

Example 4.21 Consider the field extension $\mathbb{C}/\mathbb{Q}$. Note that $i \in \mathbb{C}$ is algebraic over $\mathbb{Q}$ with minimal polynomial $x^2 + 1$. We can now construct the ring $\mathbb{Q}[X]/(X^2 + 1)$. From the previous theorem it now follows that this polynomial ring is isomorphic to

$$\mathbb{Q}[i] = \{f(i) | f \in \mathbb{Q}[X]\} = \{a + bi | a, b \in \mathbb{Q}\}$$

which is a field. So we have the field extensions $\mathbb{Q} \subset \mathbb{Q}[i] \subset \mathbb{C}$. The following theorem states that $K \subset K[X]/(p) \subset L$ is always a tower of field extensions.

Theorem 4.22 *Suppose $\alpha \in L/K$ algebraic over K, then*

$$K \subset K[\alpha] \subset L$$

is a tower of field extensions and $[K[\alpha] : K] = \deg_K \alpha$.

Proof: We have

$$K \subset K[\alpha] \subset L$$

and, as we said before in remark 4.19, $\{1, \alpha, \ldots, \alpha^n\}$ is a basis for $K[\alpha]$ over K, so what remains to be proven is that $K[\alpha]$ is a field. $K[\alpha]$ is a commutative ring that is not 0, so we only need multiplicative inverses.

Suppose $f(\alpha) \in K[\alpha]$. If $f(\alpha) \neq 0$, then the irreducible minimal polynomial p of α and f are relatively prime, so by the Euclidean algorithm there exist $a, b \in K[X]$ with

$$a(X)p(X) + b(X)f(X) = 1. \tag{11}$$

[4] This follows directly from the first isomorphism theorem, found in most elementary algebra textbooks, for instance [1].

If we evaluate this in $X = \alpha$, we get $b(\alpha)f(\alpha) = 1$, so $f(\alpha)$ has an inverse in $K[\alpha]$. Since every element of $K[\alpha]$ is of the form $f(\alpha)$ every element has an inverse, thus $K[\alpha]$ is a field. □

Remark 4.23 If α is algebraic, $K[\alpha]$ is a field and therefore equal to its own quotient field, which we denote by $K(\alpha)$. If α is transcendental $K[\alpha]$ is not a field and therefore not equal to $K(\alpha)$.

Remark 4.24 In an extension L/K, $K[\alpha_1, \ldots, \alpha_n]$ can be constructed by subsequently constructing the K-vector space $K[\alpha_1]$, the $K[\alpha_1]$-vector space $K[\alpha_1][\alpha_2]$, etc.

Example 4.25 Finite extensions are algebraic, but algebraic extensions need not to be finite. $L = \mathbb{Q}[\sqrt{2}, \sqrt[3]{2}, \sqrt[4]{2}, \ldots]$ is algebraic over $\mathbb{Q}$, but not finite, because $\mathbb{Q} \subset \mathbb{Q}[\sqrt[n]{2}] \subset L$ is a tower of field extensions which implies $[L : \mathbb{Q}] \geq [\mathbb{Q}[\sqrt[n]{2}] : \mathbb{Q}]$. Now $[\mathbb{Q}[\sqrt[n]{2} : \mathbb{Q}] = n$ because $x^n - 2$ is irreducible over $\mathbb{Q}$. So $[L : \mathbb{Q}] \geq n$ for all n and therefore infinite.

Theorem 4.26 *If $K \subset L_1 \subset L_2$ is a tower of field extensions, with L_2/K finite, then* $[L_2 : K] = [L_1 : K][L_2 : L_1]$

Proof: If $\{\alpha_i | i \in \mathbb{N}\}$ is a basis for L_1/K and $\{\beta_i | i \in \mathbb{N}\}$ a basis for L_2/L_1, then $\{\alpha_i\beta_j | i, j \in \mathbb{N}\}$ is a basis for L_2/K: this set generates L_2, because an element of L_2 can be written as linear combination of β_j with coefficients in L_1 and those coefficients can be written as linear combination of α_i with coefficients in K. It also is linearly independent: if $\sum a_{ij}\alpha_i\beta_j = 0$, then $\sum_j(\sum_i a_{ij}\alpha_i)\beta_j = 0$ is a combination for which the β_j are linearly independent for the coefficients $\sum_i a_{ij}\alpha_i$ in L_1, which means they must be zero for all j, and because the α_i are independent over K, we have $a_{ij} = 0$ for all i, j. □

Example 4.27 Consider $\alpha = \sqrt[3]{2}$ and $\beta = \alpha\zeta_3$, as elements of $\mathbb{C}$, where $\zeta_3 = e^{\frac{2\pi i}{3}} = \frac{-1+\sqrt{-3}}{2}$. They are both roots of the irreducible polynomial $X^3 - 2$ over $\mathbb{Q}$, so $[\mathbb{Q}(\alpha) : \mathbb{Q}] = [\mathbb{Q}(\beta) : \mathbb{Q}] = 3$. We also have $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\alpha, \zeta_3)$. $[\mathbb{Q}(\alpha, \zeta_3) : \mathbb{Q}(\alpha)]$ can be seen to be 2 because the degree of ζ_3 over $\mathbb{Q}(\alpha)$ is 2. This follows from the fact that it has minimal polynomial $X^2 + X + 1$ over $\mathbb{Q}(\alpha)$, which is irreducible, because it has no roots in $\mathbb{R}$, so in particular not in $\mathbb{Q}(\alpha) \subset \mathbb{R}$. We know now that $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$ and $[\mathbb{Q}(\alpha, \zeta_3) : \mathbb{Q}(\alpha)] = 2$, so because of theorem 4.26, $[\mathbb{Q}(\alpha, \beta) : \mathbb{Q}] = 3 \cdot 2 = 6$. In particular it is not equal to $[\mathbb{Q}(\alpha) : \mathbb{Q}][\mathbb{Q}(\beta) : \mathbb{Q}] = 9$.

In the next theorem it will be proved that, if we cannot find a root in our coefficient field, we do not need to have any previous knowledge of a larger field in which the polynomial does have a root. We can namely always *construct* a larger field from our coefficient field that contains a root.

Theorem 4.28 *Suppose K is a field and $f(X)$ is a non-constant polynomial over K. There exists a finite field extension L/K such that L contains a root of f.*

Proof: Suppose f is irreducible, then $K[X]/(f)$ is a ring for which the class X mod $f(X)$ is a root of f, because $f(X) = 0 \bmod X$. Just like in the proof of theorem 4.22 we use the Euclidian algorithm to show that $K[X]/(f)$ has multiplicative inverses, and hence must be a field and therefore a finite field extension of K that contains a root of f.

Suppose f is reducible, then $f = \prod {p_i}^{\nu_i}$ with p_i irreducible. The fields $K[X]/(p)$ are finite field extensions of K that contain a root of f. □

Remark 4.29 It is very important to see that, given an irreducible polynomial p over K, the field extension that we need to construct to find a root of p, is in fact the field $K[X]/(p)$ and that the root in this field is the class X mod p.

Now if $L = K[\alpha]$ is a larger field, generated by a root α of p, we know that p is the minimal polynomial of α over K, because p is irreducible. We also know by theorem 4.20, that $K[X]/(p)$ is isomorphic to $K[\alpha]$ under an isomorphism that evaluates the polynomials of $K[X]/(p)$ in α. Note that this means that although p has multiple roots in L it doesn't matter which one we take to construct $K[\alpha]$, because that field is always isomorphic to $K[X]/(p)$. The root X mod p is mapped to α by this isomorphism and in fact the basis $\{1, X, \ldots, X^{n-1}\}$ of $K[X]/(p)$ as a vector space over K, is mapped to $\{1, \alpha, \ldots, \alpha^{n-1}\}$, the basis of L as a vector space over K.

Definition 4.30 Suppose $f = \prod_{i=1}^{m} {p_i}^{\nu_i}$, with p_i irreducible. The m fields $K[X]/(p_i)$ are called the *root fields* of f.

Theorem 4.31 *All root fields of an irreducible polynomial p over K are isomorphic.*

Proof: As we saw in remark 4.29, the root fields of an irreducible polynomial are all isomorphic to $K[X]/(p)$, so we see that the root fields of a polynomial are uniquely defined by its irreducible factors. □

Example 4.32 We have another look at the polynomial $f = X^3 - 2$, which is irreducible over $\mathbb{Q}$. We can construct the root field $\mathbb{Q}[X]/(X^3 - 2)$ and note that, by theorem 4.28, it is a field extension that contains a root, namely X mod $X^3 - 2$. In the complex numbers f has three roots, namely $r_1 = \sqrt[3]{2}$, $r_2 = \sqrt[3]{2}\,\zeta_3$ and $r_3 = \sqrt[3]{2}\,\zeta_3^2$, where ζ_3 is a third degree root of unity. By theorem 4.31, the three field extensions $\mathbb{Q}[r_i]$, $i = 1, 2, 3$ are isomorphic to $\mathbb{Q}[X]/(X^3 - 2)$, under three isomorphisms that send X mod $X^3 - 2$ to r_i for $i = 1, 2, 3$ respectively.

Note that if we define $\sqrt[3]{2}$ as a positive real number, the three field extensions $\mathbb{Q}[r_i]$ are still isomorphic, but they are not the same subspaces of $\mathbb{C}$. For instance

$\mathbb{Q}[r_1]$ has $\{1, \sqrt[3]{2}, (\sqrt[3]{2})^2\}$ as a basis, so $\mathbb{Q}[r_1] \subset \mathbb{R}$ and $r_2 \notin \mathbb{R}$. Note furthermore that this means that $\mathbb{Q}[r_1]$ does not contain the roots r_2 or r_3.

The example above shows that although $\mathbb{Q}[X]/(X^3-2)$ is a field extension that contains a root, this does *not* mean $\mathbb{Q}[X]/(X^3-2)$ contains all the roots of f. If we want all the roots of f to be contained in our construction, we could for instance add a third degree root of unity. As we saw in example 4.27, such an element is a root of Y^2+Y+1, which is irreducible over $\mathbb{Q}[X]/(X^3-2)$. If we now follow our construction method rigourously, this leads us to construct the field extension

$$(\mathbb{Q}[X]/(X^3-2))[Y]/(Y^2+Y+1) \cong \mathbb{Q}[\sqrt[3]{2}, \zeta_3].$$

We could also have noticed that in $\mathbb{C}$ this third degree root of unity is equal to $\frac{-1+\sqrt{-3}}{2}$ and divide by the minimal polynomial of $\sqrt{-3}$ to get

$$(\mathbb{Q}[X]/(X^3-2))[Y]/(Y^2+3) \cong \mathbb{Q}[\sqrt[3]{2}, \sqrt{-3}].$$

Yet another possibility would have been to divide the root X in $\mathbb{Q}[X]/(X^3-2)$ out of the polynomial Y^3-2, by using the division algorithm for polynomials. This leads to $Y^3-2 = (Y-X)(Y^3+YX+X^2)$ and consequently to the construction:

$$(\mathbb{Q}[X]/(X^3-2))[Y]/(Y^2+YX+X^2) \cong \mathbb{Q}[\sqrt[3]{2}, \sqrt[3]{2}\,\zeta_3].$$

So we see that there are at least three different ways of constructing a field extension of degree 6 that contain the roots of X^3-2. Now the question is: are they the same? If we look at them as subspaces of $\mathbb{C}$ (the right hand sides), we notice that, because they contain all the roots of f, they must be subspaces of $\mathbb{Q}[r_1, r_2, r_3]$ and because $\mathbb{Q}[r_1, r_2, r_3]$ is also of degree 6 over $\mathbb{Q}$, we can apply theorem 4.12 and conclude that they are all the same subspaces of $\mathbb{C}$.

Definition 4.33 Let f be a polynomial over K and L a field extension of K that contains all the roots of f, so that f can be written as a product of linear factors over L. We say that f *splits completely* in L. Adding all the roots of f to K defines a subfield of L. This field is called the *splitting field* of f.

Theorem 4.34 *For every polynomial f over K there exists a splitting field.*

Proof: If f is a linear polynomial we are done. If not, decompose f into irreducible factors over K. Apply theorem 4.28 on one of its irreducible factors to find a field extension L that has a root of f. By doing this you always find at least one linear factor of f (we may find more than one, or find that other

irreducible factors become reducible, but that only speeds up the process). If there still exists an irreducible factor of degree > 1 over L, apply 4.28 again to construct a field extension of L that has a root of this irreducible factor. Do this until you have a field extension that has all the roots of f and therefore f spits completely into linear factors. The number of steps we have to take is less than the degree of f over K. □

Theorem 4.35 *Let f be a polynomial over $\mathbb{Q}$ and let $L \subset \mathbb{C}$ be its splitting field. L is a unique subset of $\mathbb{C}$.*

Proof: We know that all the roots of f are contained in $\mathbb{C}$, so our splitting field is the unique field attained by evaluating all polynomials in n variables over K in the roots $r_1, \ldots, r_n$ of f. □

The theorem also holds for the splitting field of a polynomial over a general field, but then "unique" must be seen as "unique up to isomorphism". The proof also becomes a little bit longer.

Theorem 4.36 *The splitting field L of a polynomial f over K is unique up to isomorphism.*

Proof: We prove this by induction on the degree of f. If f has degree 1, then $L = K$ and the theorem is trivial. Suppose that all splitting fields of polynomials of degree $< n$ are unique up to isomorphism, then we prove that splitting fields of n-th degree polynomials are isomorphic. Suppose f is of degree n. First decompose f into irreducible factors. If f has only linear factors over K, then $L = K$ and we are done, so let p be an irreducible factor of f. If a root α of p is added to K, we get $K[\alpha]$, which is a unique root field of p (theorem 4.31) and we note that L is also a splitting field of $\frac{p}{x-\alpha}$ over $K[\alpha]$, which, by the induction hypothesis, is unique up to isomorphism. □

To further expand our understanding of field extensions, we need to study their morphisms. This will be done in the next section.

4.2 Morphisms

When studying the algebraic structure of a mathematical object such as a group, a ring or a field extension, the first goal is to understand the *morphisms* between the objects, that is, the functions that map such an object to a similar object while preserving the algebraic structure. A homomorphism between two additive groups for instance preserves the addition and the zero and a ring homomorphism preserves the addition, the zero, the multiplication and the unit element. In the study of field extensions, we already used the concept of *isomorphism*, by which we actually meant a bijective morphism between field extensions or between vector spaces.

The next goal should be to understand how a mathematical object can be morphed *into itself*. How can such an object be changed, without changing it in an essential way? This can be compared to studying the symmetry of a geometrical figure, like a tetrahedron. There are twelve rotations which send the tetrahedron to itself. These rotations may move the individual points of the tetrahedron, but the object as a whole remains the same. These rotations form a *group*, which completely describes the symmetry of a tetrahedron. To quote the introduction from [1] Groups and symmetry: *Numbers measure size, groups measure symmetry.*

In a similar way the automorphisms on a splitting field form a group that describes the symmetry of a field extension. The elements of the field may be moved around, but the algebraic structure remains unchanged. Now because splitting fields are directly linked to polynomials, this group also describes the symmetry of a polynomial. By studying this group, called the *Galois group*, we will see that the general equation of degree 5 cannot be solved by radicals.

In this paragraph, we will start with the study of morphisms of field extensions (or vector spaces) and lay the basis for our final goal: a theory of automorphisms of splitting fields.

Definition 4.37 A *morphism* ϕ between two fields L and L' is a ring homomorphism between L and L', so it is a function $\phi : L \to L'$ such that

- $\phi(a+b) = \phi(a) + \phi(b)$
- $\phi(ab) = \phi(a)\phi(b)$
- $\phi(1) = 1$

Lemma 4.38 *A morphism $\sigma : L \to L'$ between two fields L and L' is injective.*

Proof: The kernel of σ is an ideal in L, so (0) or (1), but $\sigma(1) \neq 0$, so it is (0). □

Remark 4.39 Because a morphism $\sigma : L \to L'$ between fields is always injective, that means we can find a copy of L in L'. Therefore they are also called *embeddings*. Such a copy does not have to be unique. As we saw in example 4.32, there are three copies of $\mathbb{Q}[X]/(X^3 - 2)$ in $\mathbb{C}$. They are of course unique up to isomorphism.

Definition 4.40 If L/K and L'/K are field extensions of K, then a morphism $\sigma : L \to L'$ is called a *K-morphism* or a *morphism of field extensions over K* if σ is K-linear.

Note that K-linearity effectively makes a morphism between L and L' into a morphism between K-vector spaces, because if σ is K-linear it preserves all the conditions of a K-vector space.

Theorem 4.41 *$\sigma : L \to L'$ is a K-morphism if and only if $\sigma_{|K} = \mathrm{id}_K$.*

Proof: Let $k \in K$ and $l \in L$. If σ is K-linear, we have $\sigma(kl) = k\sigma(l)$ which implies $\sigma(k1) = k\sigma(1) = k$. On the other hand if $\sigma_{|K} = \mathrm{id}_K$, we have $\sigma(k) = k$, which implies $\sigma(kl) = \sigma(k)\sigma(l) = k\sigma(l)$. □

Remark 4.42 So we have proved that for a field extension L/K, the morphisms on the K-vector space L are injective morphisms that are the identity on K, in other words they leave K fixed. Now because every element in L can be written as a K-linear combination, a K-morphism is completely defined by stating what it does on the basis elements of L.

The following theorem shows a stronger result for the morphisms on an algebraic extension of the form $K[\alpha]$, called a *simple extension*, namely that a K-morphism on $K[\alpha]$ is completely defined by stating what it does on the roots of the minimal polynomial of α.

Theorem 4.43 *Let L/K be a field extension and $\alpha \in L$ an algebraic element over K with minimal polynomial p. The number of embeddings of $K[\alpha]$ into L is equal to the number of different roots of p in L and such an embedding can be defined by stating that it sends α to (another) root of p.*

Proof: If σ is an embedding of $K[\alpha]$ into L, then $p(\alpha) = 0$ implies $p(\sigma(\alpha)) = \sigma(p(\alpha)) = \sigma(0) = 0$, so $\sigma(\alpha)$ is also a root of p. So every embedding gives a root in L.

On the other hand, every root β of p in L, induces a morphism σ that sends the root field $K[\alpha]$ to $K[\beta]$ which is an isomorphism $K[\alpha] \to K[\beta]$ by theorem 31, so $\sigma : K[\alpha] \to L$ is an embedding. □

Remark 4.44 Note that if $L = K[\alpha]$, the theorem applies to embeddings of L into *itself*.

Definition 4.45 A morphism, resp. a K-morphism is called an *isomorphism* resp. a *K-isomorphism* if it is bijective.
An isomorphism, resp. a K-isomorphism is called an *automorphism* resp. a K-automorphism if it is a function from L to L.
The automorphisms, resp. the K-automorphisms on L form a group under composition that is written as $\mathrm{Aut}(L)$, resp. $\mathrm{Aut}_K(L)$.

Example 4.46

- The identity map $\mathrm{id} : L \to L$, $x \mapsto x$ is an automorphism. It is the identity element of $\mathrm{Aut}(L)$.

- Complex conjugation (sending $a+bi$ to $a-bi$) defines an $\mathbb{R}$-isomorphism σ on $\mathbb{C}$. Together with the identity, they are all $\mathbb{R}$-automorphisms on $\mathbb{C}$, because (theorem 4.43) the number of embeddings from $\mathbb{R}[i]$ into $\mathbb{C}$ is equal to the degree of the minimal polynomial of i, which is 2, so $\mathrm{Aut}_{\mathbb{R}}(\mathbb{C}) = \{\mathrm{id}, \sigma\} \cong \mathbb{Z}_2$, the cyclic group of order 2.

- The two embeddings of $\mathbb{Q}[\sqrt{2}\,]$ into itself send $\sqrt{2}$ either to itself or to $-\sqrt{2}$, the other root of x^2-2, so $\mathrm{Aut}_{\mathbb{Q}}\mathbb{Q}[\sqrt{2}\,] \cong \mathbb{Z}_2$.

- Similarly $\mathrm{Aut}_K(K[\sqrt{\alpha}\,]) \cong \mathbb{Z}_2$ if α is not a square in K.

- Let p be a prime and ζ_p a p-th degree root of unity. As we saw in example 4.18, the minimal polynomial of ζ_p is $x^{p-1} + \ldots + x + 1$, so the automorphisms on $\mathbb{Q}[\zeta_p]$ are induced by the $p-1$ embeddings that send ζ_p to another root of $x^{p-1}+\ldots+x+1$. For instance if $p=5$: the automorphism induced by $\zeta_5 \mapsto \zeta_5^2$, acts on the other roots as

$$\zeta_5^2 \mapsto \zeta_5^4 \qquad \zeta_5^3 \mapsto \zeta_5 \qquad \zeta_5^4 \mapsto \zeta_5^3.$$

  Note that composition of the automorphisms, boils down to multiplication modulo p, for instance the composition of the automorphism induced by $\zeta_p \mapsto \zeta_p^2$ by the automorphism induced by $\zeta_p \mapsto \zeta_p^3$ is induced by $\zeta_p \mapsto \zeta_p^6$. Therefore they form the multiplicative cyclic group $(\mathbb{Z}/p\mathbb{Z})^*$ of order $p-1$.

- There are no automorphisms on $\mathbb{Q}[\sqrt[3]{2}]$ except the identity, because there is only one root of x^3-2 contained in $\mathbb{Q}[\sqrt[3]{2}]$, so there is only one embedding of $\mathbb{Q}[\sqrt[3]{2}]$ into itself, which is the identity.

- Consider the extension $\mathbb{Q} \subset K \subset L$, where $K = \mathbb{Q}[\sqrt[3]{2}\,]$ and $L = \mathbb{Q}[\sqrt[3]{2}, \zeta_3]$, the splitting field of $f = x^3 - 2$ over $\mathbb{Q}$. K is a root field of f, so sending $\sqrt[3]{2}$ to another root ($\sqrt[3]{2}\,\zeta_3$ or $\sqrt[3]{2}\,\zeta_3^2$) of f induces three embeddings of K into L:

$$\sigma_1|_K = \mathrm{id} \qquad \sigma_2|_K : \sqrt[3]{2} \mapsto \sqrt[3]{2}\,\zeta_3 \qquad \sigma_3|_K : \sqrt[3]{2} \mapsto \sqrt[3]{2}\,\zeta_3^2$$

  These embeddings are not yet automorphisms on L because their domain is K, which is not the whole L, but they can be extended to automorphisms on L, by defining them on ζ_3, the primitive element of the second extension. Sending ζ_3 to itself gives three automorphisms, which act on the roots of f as follows:

$$\begin{array}{lcl}
\sigma_1 & = & \mathrm{id} \\
\sigma_2 & : & \sqrt[3]{2} \mapsto \sqrt[3]{2}\,\zeta_3 \quad , \quad \sqrt[3]{2}\,\zeta_3 \mapsto \sqrt[3]{2}\,\zeta_3^2 \quad , \quad \sqrt[3]{2}\,\zeta_3^2 \mapsto \sqrt[3]{2} \\
\sigma_3 & : & \sqrt[3]{2} \mapsto \sqrt[3]{2}\,\zeta_3^2 \quad , \quad \sqrt[3]{2}\,\zeta_3 \mapsto \sqrt[3]{2} \quad , \quad \sqrt[3]{2}\,\zeta_3^2 \mapsto \sqrt[3]{2}\,\zeta_3
\end{array}$$

  But we could also have extended the three embeddings on K by sending ζ_3 to ζ_3^2, the other root of the minimal polynomial Y^2+Y+1 of ζ_3 over

K. This gives three more automorphisms that act on the roots of f as follows.

$$\begin{array}{lcl} \tau_1 & : & \sqrt[3]{2} \mapsto \sqrt[3]{2} \quad , \quad \sqrt[3]{2}\,\zeta_3 \mapsto \sqrt[3]{2}\,\zeta_3^2 \quad , \quad \sqrt[3]{2}\,\zeta_3^2 \mapsto \sqrt[3]{2}\,\zeta_3 \\ \tau_2 & : & \sqrt[3]{2} \mapsto \sqrt[3]{2}\,\zeta_3^2 \quad , \quad \sqrt[3]{2}\,\zeta_3 \mapsto \sqrt[3]{2}\,\zeta_3 \quad , \quad \sqrt[3]{2}\,\zeta_3^2 \mapsto \sqrt[3]{2} \\ \tau_3 & : & \sqrt[3]{2} \mapsto \sqrt[3]{2}\,\zeta_3 \quad , \quad \sqrt[3]{2}\,\zeta_3 \mapsto \sqrt[3]{2}\,\zeta_3^2 \quad , \quad \sqrt[3]{2}\,\zeta_3^2 \mapsto \sqrt[3]{2}\,\zeta_3^2 \end{array}$$

This gives a group of 6 elements $\{\mathrm{id}, \sigma_2, \sigma_3, \tau_1, \tau_2, \tau_3\}$, generated by σ_2, which is of order 3 and τ_1, which is of order 2 (note that τ_1 is in fact the restriction of complex conjugation to L). Furthermore, σ and τ don't commute, since

$$\sigma_2\tau_1(\sqrt[3]{2}\zeta_3) = \sigma_2(\sqrt[3]{2}) = \sqrt[3]{2}\zeta_3$$

while

$$\tau_1\sigma_2(\sqrt[3]{2}\zeta_3) = \tau_1(\sqrt[3]{2}\zeta_3) = \sqrt[3]{2}\zeta_3^2$$

so this group of order 6 cannot be isomorphic to $\mathbb{Z}_6$, so the only possibility is that it is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_3 \cong D_3 \cong S_3$. So in fact *all* permutations of the three roots of f induce automorphisms on the splitting field of f.

We see in the examples above that if L is a splitting field of a polynomial, the degree of the extension (the dimension of L as a vector space over K) is equal to the number of elements in the group of automorphisms of L. If L is not a splitting field, this does not seem to be the case, as is seen in the case of $\mathbb{Q}[\sqrt[3]{2}]$. This holds in general.

Definition 4.47 Field extensions that are splitting fields of a polynomial are also called *normal extensions*.

Definition 4.48 A finite extension L/K is called a *Galois extension* if

$$\#\mathrm{Aut}_K(L) = [L : K].$$

In the case of a Galois extension $\#\mathrm{Aut}_K(L)$ is also written as $\mathrm{Gal}(L/K)$.

In the following section we will prove that for infinite fields a finite extension is Galois if and only if it is normal.

5 Galois theory

5.1 Normal extensions, Galois extensions

Convention: *In this chapter, we will assume all fields have characteristic 0.*

To prove the next theorem for fields of characteristic p, we need the additional condition of separability, which is not treated here. For more information, see for instance [10].

Theorem 5.1 *If L/K is a normal extension, then it is a Galois extension, so* $\#\mathrm{Gal}(L/K) = [L : K]$.

Proof: Let $\alpha_1, \ldots \alpha_n$ be the roots of the polynomial f of which L is the splitting field. Let $K = K_0$, $K_i = K[\alpha_1, \ldots, \alpha_i]$ $(i = 1, \ldots n)$ and let p_i be the minimal polynomial of α_i over K_{i-1}. We know that all the roots of p_i are contained in L, because otherwise K_i would contain elements not in L, but L is unique. So the number of K_{i-1}-embeddings $\sigma_i : K_i \to L$ is equal to the degree of the minimal polynomial p_i. Now each of the $\deg p_1$ K-embeddings of K_1 into L can be extended to a K-embedding of K_2 into L in $\deg p_2$ ways (namely by sending a root of p_2 to another root of p_2). By induction we see that the number of embeddings of L into L (K-automorphism on L) is equal to

$$\prod_{i=1}^{n} \deg p_i = \prod_{i=1}^{n} [K_i : K_{i-1}]$$

Which is equal to $[L : K]$ by theorem 5.26. □

So we proved that normal extension implies Galois extension. To prove the converse, we need the following lemma.

Lemma 5.2 *If L and L' are field extensions of K, then the number of different K-morphisms $\sigma : L \to L'$ is smaller than or equal to $[L : K]$. In particular, we have* $\mathrm{Aut}_K(L/K) \leq [L : K]$.

Proof: For a simple extension $L = K(\alpha)$ this follows from theorem 5.39, because in that case, every morphism must send α to another root of its minimal polynomial p over K and there exist at most $[L : K]$ different roots of p.

For a general finite extension the proof is more difficult and requires e.g. Dedekind's lemma, which will not be treated here (see [3]). We will later prove independently that every extension can be written as a simple extension, so then the proof above becomes sufficient. □

Now we hope that by demanding the inequality above to be an equality (which would make L/K Galois), we end up getting a splitting field. In that case the converse of theorem 5.1 would be proven.

Theorem 5.3 *If a finite extension L/K is a Galois extension, it is a normal extension.*

Proof: Let L/K be Galois, $L = K(\alpha_1, \dots \alpha_n)$ and p_i be the minimal polynomial of α_i. We can construct a field extension M in which $f = \prod_{i=1}^{n} p_i$ splits completely into linear factors. We want to prove that all roots of f are contained in L, in which case L would be the splitting field of f and therefore a normal extension.

Suppose $\beta \in M$ is a root of f that is not contained in L. Without loss of generality, we can assume β to be the root of p_1 (we can rearrange the α's if necessary). Now we know by theorem 5.2 that the number of K-morphisms from $L \to M$ is smaller than or equal to $[L : K]$. Because every K-morphism of L gives such a morphism and since $\#\mathrm{Gal}(L/K) = [L : K]$, it follows that all morphisms from $L \to M$ are elements of $\#\mathrm{Gal}(L/K)$. Now by theorem 5.43, we know that there exists a morphism $\sigma : K[\alpha_2 \dots \alpha_n][\alpha_1] \to M$ that maps α_1 to β. Since $\sigma \in \#\mathrm{Gal}(L/K)$, it follows that $\sigma(\alpha_1) = \beta \in L$. So we proved that every root β of f is contained in L, so L is the splitting field of f and L is a normal extension of K. □

Theorem 5.4 *A finite extension is normal if and only if it is Galois.*

Proof: Combine theorem 5.1 and theorem 5.3. □

There is yet another requirement that an extension can fulfill that is equivalent to it being Galois or normal.

Theorem 5.5 *If L/K is Galois, then L contains a splitting field of all irreducible polynomials over K that have a root in L. In other words: if an irreducible polynomial over K has one root in L, it has all its roots in L.*

Proof: This follows directly from the proof of theorem 5.3: the other roots are found by applying a morphism σ in the Galois group that sends a root of an irreducible polynomial to another root of that polynomial. □

5.2 Main theorem

In our attempt to understand the symmetry of field extensions and polynomials, we now have three equivalent requirements for a Galois extension. The next goal is to break up that symmetry into smaller parts. We want to do this because

the steps that need to be taken, or in other words the radicals that are needed, to solve a polynomial equation is related to breaking up the splitting field in smaller parts, which in turn is related to breaking up the Galois group. If the Galois group of L over K is G, then the Galois group of L over $M \subset L$ must be something smaller, so let's look at fields that lie between K and L and their Galois groups. This finally leads to the proof of the main theorem of Galois theory: the correspondence between intermediate fields and subgroups of the Galois group.

Theorem 5.6 *If $K \subset L' \subset L$ is a tower of field extensions and L/K is Galois, then L/L' is Galois and its Galois group* $\mathrm{Gal}(L/L')$ *is a subgroup of* $\mathrm{Gal}(L/K)$.

For the first half of the theorem note that if L is a splitting field of a polynomial over K, it is also a splitting field of a polynomial over L', so L/L' is also Galois.

For the second half, note that the automorphisms that leave L' fixed form a subgroup of the automorphisms that leave K fixed.

Remark 5.7 Although L/L' is always Galois in the theorem above, this doesn't have to be the case for L'/K. The well known extension $\mathbb{Q} \subset \mathbb{Q}[\sqrt[3]{2}] \subset \mathbb{Q}[\sqrt[3]{2}, \sqrt[3]{2}\,\zeta_3]$ is an example of this.

Remark 5.8 Although not every extension is Galois, every extension is contained in a Galois extension, namely in the splitting field of f from the proof of theorem 5.3. This field is called a *Galois closure*.

Definition 5.9 Let L/K be a Galois extension and $H \subset \mathrm{Gal}(L/K)$. The field $L^H = \{z \in L \mid \sigma(z) = z \text{ for all } \sigma \in H\}$ is an *intermediate field* (a subfield of L that contains K), which is also called the *fixed field* of H.

The next lemma states that the number of elements in a subgroup of the Galois group is equal to the degree of the fixed field, defined by that subgroup. We need this lemma for the proof of the main theorem, which comes directly after it.

Lemma 5.10 *Let L/K be a Galois extension and $H \subset \mathrm{Gal}(L/K)$, then*

$$\#\mathrm{Gal}(L/L^H) = [L : L^H] = \#H$$

and $\mathrm{Gal}(L/L^H) = H$.

Proof: The proof is not given here but can be found in [3]. □

We now have everything we need to prove the main theorem of Galois theory: the link between subgroups of the Galois group and intermediate fields.

Theorem 5.11 (Main theorem of Galois theory) *Let L/K be Galois, then*

$$\mathrm{Gal}(L/K) \supset H \mapsto M = L^H$$

defines a bijection between subgroups of $\mathrm{Gal}(L/K)$ *and fixed fields (extensions of K that are subfields of L).*

Proof: It is clear that for every subgroup H of $\mathrm{Gal}(L/K)$, we can define the field that is fixed under all automorphisms of L and for every intermediate field M, we can define the subgroup of all automorphism that leave M fixed, so this function is well defined in both directions. We will first prove surjectivity.

Let M be a field such that $K \subset M \subset L$. From theorem 5.6, we know that L/M is Galois. We also know that $M \subset M^{\mathrm{Gal}(L/M)}$, because the elements of $\mathrm{Gal}(L/M)$ leave at least M fixed. Furthermore we know that $\#\mathrm{Gal}(L/M) = [L : M]$ because L/M is Galois and from theorem 5.10 it follows that $[L : L^{\mathrm{Gal}(L/M)}] = [L : M]$, so (apply theorem 3.12) we have that $L^{\mathrm{Gal}(L/M)} = M$.

Now we prove injectivity. Let H be a subgroup of $\mathrm{Gal}(L/K)$ and $H' \subset \mathrm{Gal}(L/K)$. Suppose $L^H = L^{H'} = M$. By theorem 5.10 we know that $\#H = \#H' = [L : M]$. Note that L/M is a Galois extension (theorem 5.6), so $\#\mathrm{Gal}(L/M) = [L : M]$. Note furthermore that we chose H and H' to be subgroups of this Galois group, so it follows from $\#H = \#H' = \#\mathrm{Gal}(L/M)$ that $H = \mathrm{Gal}(L/M) = H'$. □

Remark 5.12 This theorem is so important because it translates problems about intermediate fields to problems about finite groups. The theory of finite groups is very well known and has some very strong results (e.g. the Sylow theorems). Galois theory lead to a great interest in the study of finite groups, which eventually lead to the classification of all finite simple groups.[5]

For us this theorem is also important, because for every polynomial f, the splitting field is Galois, so to understand something about the fields that lie between the field of coefficients and the splitting field (and thus can tell us something about a solution of $f(x) = 0$), we need to understand the subgroups of the Galois group.

We already know that if L/K is Galois and $K \subset M \subset L$, that L/M is Galois, but M/K need not be Galois. The next theorem gives a condition on $\mathrm{Gal}(L/K)$ so that it *is* Galois. We suppose that the reader is familiar with some elementary group theory.

Theorem 5.13 *Let L/K be a Galois extension and $K \subset M \subset L$. Then M/K is Galois if and only if* $\mathrm{Gal}(L/M)$ *is a normal subgroup of $G = \mathrm{Gal}(L/K)$. In*

[5]For a long time some doubts remained on whether there exists a complete and correct proof, due to the sheer length and complexity of the published work and the fact that parts of the supposed proof remain unpublished. The latest news is that the work has finally been finished this year (2005).

that case we have

$$\mathrm{Gal}(M/K) \cong \mathrm{Gal}(L/K) \,/\, \mathrm{Gal}(L/M).$$

Proof: We use the condition for a Galois extension from theorem 5.5. If M is a normal extension of K then if an irreducible polynomial over K has one root in M, it has all its roots in M. Furthermore we know that all elements in G can be defined by stating that they send roots of an irreducible polynomial over K to another root of that polynomial. So M/K is Galois if and only if $\sigma(M) = M$ for all $\sigma \in G$.

Now let's look at the Galois groups of M and of $\sigma(M)$. Suppose $H = \mathrm{Gal}(L/M)$, then $M = \{x \,|\, h(x) = x \; \forall h \in H\}$. Now if h leaves x fixed, then we know that $\sigma h \sigma^{-1}$ leaves σx fixed, so $\sigma(M) = \{\sigma x \,|\, (\sigma h \sigma^{-1})(\sigma x) = \sigma x \; \forall h \in H\}$. So we see that M is the fixed field of H and $\sigma(M)$ is the fixed field of $\sigma H \sigma^{-1}$, so $\sigma(M) = M$ if and only if $\sigma H \sigma^{-1} = H$ for all $\sigma \in G$, which is exactly the definition of H being a normal subgroup of G.

Finally there is a group homomorphism $\rho : G \to \mathrm{Gal}(M/K)$, namely restriction : $\sigma \mapsto \sigma|_M$. This is surjective because every $\sigma|_M$ can be extended to L (see [3]) and its kernel is H, so $\mathrm{Gal}(M/K) = G/H$. □

There is another very remarkable consequence of the main theorem, which is the theorem of primitive elements. Galois in fact proved this theorem before the main theorem and many of his proofs relied on it. We will give two proofs. The first proof uses the main theorem, which has the benefit that the proof becomes shorter and easier, and the second proof is independent of the main theorem, which has the benefit that we do not need Dedekind's lemma in the proof of theorem 5.2.

Theorem 5.14 (Primitive elements) *Let L/K be a finite extension.*

1. *There are only a finite number of intermediate fields M, such that $K \subset M \subset L$.*
2. *L is a simple extension. There exists an element $\gamma \in L$ such that $L = K[\gamma]$.*

Proof: 1. This follows directly from the main theorem, because $\mathrm{Gal}(L/K)$ has only a finite number of elements.
2. It is not difficult to see that it is sufficient to prove this for finite extensions of the form $K[\alpha, \beta]$.

Let $0 \neq c \in K$ and consider the elements $\alpha + c\beta \in K[\alpha, \beta]$. Because $K[\alpha + c\beta] \subset K[\alpha, \beta]$ for all c and because there are infinitely many c and only a finite number of intermediate fields (point 1), there must be $c, c' \in K$, with $c \neq c'$ such that

$$K[\alpha + c\beta] = K[\alpha + c'\beta].$$

But then $(c-c')\beta \in K[\alpha+c\beta]$, which implies $\beta \in K[\alpha+c\beta]$, because $c-c' \neq 0$. We also have $\alpha = (\alpha+c\beta)-c\beta \in K[\alpha+c\beta]$, so it follows that $K[\alpha+c\beta] = K[\alpha,\beta]$ which we had to prove. □

We now give a proof that does not rely on the main theorem.

Proof: Suppose $L = K[\alpha,\beta]$ is of degree n and σ_i, $i = 1,\ldots,n$ are the n extensions of id_K to L. Consider the polynomial

$$f(X) = \prod_{i=1}^{n}\prod_{j\neq i}(\sigma_j\alpha - \sigma_i\alpha + X(\sigma_j\beta - \sigma_i\beta)). \tag{12}$$

This is not the zero polynomial, so it has only a finite number of roots. Therefore there exists a $c \in K$ such that $f(c) \neq 0$ and that means that $\sigma_i\alpha + c\sigma_i\beta$ are different for all $i = 1,\ldots,n$. Suppose $\gamma = \alpha + c\beta$, then $K[\gamma]$ has n different embeddings in L that are the identity on K, namely the restrictions of the σ_i's. Consequently $[K[\gamma] : K] \geq n$ and because $K[\gamma] \subset K[\alpha,\beta]$, we get $K[\gamma] = K[\alpha,\beta]$. □

Note that this is a constructive proof. As soon as we found our c, we have found $\gamma = \alpha + c\beta$.

Example 5.15 Consider the extension $\mathbb{Q} \subset \mathbb{Q}[\sqrt{2},\sqrt{3}] = L$. Following the steps in the proof given above, we are now able to find a primitive element. The four extensions of the identity on $\mathbb{Q}$ to L are

$$\begin{array}{cc} \sigma_1 = \mathrm{id}_L & \sigma_2 : \sqrt{2} \mapsto -\sqrt{2},\ \sqrt{3} \mapsto \sqrt{3} \\ \sigma_3 : \sqrt{2} \mapsto \sqrt{2},\ \sqrt{3} \mapsto -\sqrt{3} & \sigma_4 : \sqrt{2} \mapsto -\sqrt{2},\ \sqrt{3} \mapsto -\sqrt{3}. \end{array}$$

The polynomial from equation 12 now becomes (for some $d \in L$, $p \in \mathbb{N}$):

$$f(X) = dX^p(\sqrt{2}+\sqrt{3}X)(-\sqrt{2}+\sqrt{3}X)(\sqrt{2}-\sqrt{3}X)(-\sqrt{2}-\sqrt{3}X)$$

We see that $f(c) \neq 0$ for $c = 1$, so $\mathbb{Q} \subset \mathbb{Q}[\sqrt{2},\sqrt{3}] = \mathbb{Q}[\sqrt{2}+\sqrt{3}]$.

5.3 The Galois group of polynomials

The following definition should come as no surprise.

Definition 5.16 The *Galois group of a polynomial* f over a field K is defined as the Galois group of a splitting field L of f over K : $\mathrm{Gal}(f/K) := \mathrm{Gal}(L/K)$.

The following theorem gives a modernized version of the definition Galois gave of the Galois group. The problem is that it uses infinitely many "polynomial

relations between the roots". The modern concept of a splitting field brings it back to a finite dimensional vector space and the main theorem reduces it to the study of a finite group. The advantage of Galois' definition is that it stays closer to the polynomials, which are the main object of study in this paper.

Theorem 5.17 *Suppose* $X = \{\alpha_1 \ldots, \alpha_n\}$ *are the roots of a polynomial* $f \in K[X]$ *in a splitting field* L *of* f*. Then* $\mathrm{Gal}(f/K)$ *is isomorphic to the group*

$$H = \{\sigma \in S_X | \forall g \in K[x_1, \ldots, x_n] : g(\alpha_1 \ldots, \alpha_n) = 0 \Rightarrow g(\sigma(\alpha_1), \ldots, \sigma(\alpha_n)) = 0\}$$

that is, the group of permutations of the roots of f *that preserve all the polynomial relations between those roots.*

Proof: Indeed we can see an element σ of $G := \mathrm{Gal}(f/K)$ as a permutation of the roots of f, because if $f(\alpha) = 0$, then $\sigma(f(\alpha)) = f(\sigma(\alpha)) = 0$, because f has coefficients in K. Furthermore the polynomial relations are preserved, because g has coefficients in K. So we have a group homomorphism of G into H, which is clearly injective. Finally if $\sigma \in H$ is a relation preserving permutation of the roots, then we define a field extension of L as follows: any element of L is a polynomial $g(\alpha_1, \ldots, \alpha_n)$ that we send to $g(\sigma(\alpha_1), \ldots, \sigma(\alpha_n))$. This is clearly a non-zero ring homomorphism, which is well defined, for if $g(\alpha_1, \ldots, \alpha_n) = 0$, than also its image under σ. □

Consequence 5.18 $\#\mathrm{Gal}(f/K) \leq \deg(f)!$

Proof: The number of permutations of the roots of f is at most $\#S_X = \deg(f)!$. □

Now that we have enough understanding of the Galois group of a polynomial, we can start relating properties of the Galois group to properties of the polynomials. We already saw that subgroups of the Galois group relate to intermediate fields and that normal subgroups relate to intermediate fields which are normal extensions. Another property I would like to show is that transitivity relates to irreducibility.

We saw that the elements of a Galois group G are permutations on the set of roots X of a polynomial, so every element of G can be seen as a function $g : X \to X$. We say that G *acts on* X.

Definition 5.19 A permutation group G on a set of elements $X = \{X_1, \ldots, X_n\}$ is said to be *transitive* if it has only one *orbit*, where an orbit of an element $x \in X$ is defined as $G(x) = \{g(x) \in X : g \in G\}$.

Note that the orbits define a partition of X.

Example 5.20 The group $\{e, (123)(132), (45), (123)(45), (132)(45)\}$ has two orbits, $\{1, 2, 3\}$ and $\{4, 5\}$, whereas S_n is a transitive group because it has only one orbit.

Theorem 5.21 *A polynomial has a transitive Galois group if and only if it is irreducible.*

Proof: Let p over K be an irreducible polynomial, r a root of p and L the splitting field of p. Then by theorem 3.43, there exist embeddings of $K[r]$ into L that send a root of p to any other root of p. These embeddings extend to automorphisms on the splitting field of p, so the Galois group contains all the elements that send a root of p to any other root of p and is therefore a transitive group.

On the other hand, if a polynomial p has a transitive Galois group, then the automorphisms on L send a root of p to any of the other roots of the same irreducible factor, so p must be irreducible. □

Example 5.22 Consider the reducible polynomial

$$f = x^5 + x^3 - 2x^2 - 2 = (x^3 - 2)(x^2 + 1)$$

over $\mathbb{Q}$. Its roots in $\mathbb{C}$ are $\{\sqrt[3]{2}, \sqrt[3]{2}\zeta_3, \sqrt[3]{2}\zeta_3^2, i, -i\}$. As we saw in example 3.46, the Galois group of $x^3 - 2$ is $S_3 = \{e, (123), (132), (12), (13), (23)\}$. The other irreducible factor, gives us another morphism, namely sending i to $-i$. We can extend this to an automorphism (complex conjugation) by also sending ζ_3 to ζ_3^2 and leaving $\sqrt[3]{2}$ fixed, so $(23)(45)$ must also be in the Galois group of f. Composition with the other elements gives us at least the elements

$$G = \left\{ \begin{array}{cccccc} e, & (123), & (132), & (12), & (13), & (23), \\ (45), & (123)(45), & (132)(45), & (12)(45), & (13)(45), & (23)(45) \end{array} \right\}$$

Now the degree of the splitting field M of $x^3 - 2$ over $\mathbb{Q}$ is 6 and the degree of M over the splitting field of f is 2, because $X^2 - 1$ is irreducible over M, so the degree of L over $\mathbb{Q}$ must be 12 so the Galois group contains 12 elements, so G is the Galois group of f. Note that this group has two orbits: $\{1, 2, 3\}$ and $\{4, 5\}$, which correspond to the two irreducible factors of f.

5.4 Unsolvability of the quintic by radicals

This paragraph will deal with the final goal of this chapter: the proof that the general equation of degree 5 or more is not *solvable by radicals*. We will see how this relates to a property of the Galois group of that polynomial, namely that the Galois group has to be a *solvable group*. It turns out that the Galois group of the general fifth degree polynomial is not a solvable group and that therefore

the general fifth degree equation cannot be solved by radicals. There even exist specific polynomials of degree 5 that are not solvable by radicals. We will give an example at the end of this paragraph.

Remark 5.23 Intuitively a polynomial f over K is solvable by radicals if all the roots of f can be obtained from elements of K by applying the operations

$$+\ ,\ -\ ,\ \times\ ,\ \sqrt[n]{\cdot}$$

a finite number of times.

Example 5.24 All equations of degree less than 5, as well as $X^n - a$ are solvable by radicals.

It is not difficult to see that the condition that the roots of f over K can be expressed as radical functions of elements of K, is equivalent to the condition that there exists a tower of field extensions that starts with K, is then extended with n-th roots, and finally ends in a splitting field of L. Therefore the following definition is equivalent to our intuitive "definition" 5.23.

Definition 5.25 A polynomial f over K with splitting field L is *solvable (by radicals)* if and only if there exists a tower of field extensions

$$K = K_0 \subset K_1 \subset \cdots \subset K_m \supset L$$

with $K_i = K_{i-1}(\alpha_i)$ and $\exists n_i$ with $\alpha_i^{n_i} \in K_{i-1}$.

Definition 5.26 A group G is *solvable* if there exists a tower of subgroups

$$e = H_0 \subset \cdots \subset H_n = G$$

such that

- the subsequent groups are normal subgroups of the next one: $H_i \lhd H_{i+1}$ for all $i = 0, \ldots, n-1$
- the subsequent quotients H_{i+1}/H_i are commutative.

Theorem 5.27 *A solvable polynomial f over $K \subset \mathbb{C}$ has a solvable Galois group* $\mathrm{Gal}(f/K)$.

Proof: Let L be the splitting field of f. Because f is solvable we have a tower of field extensions

$$K = K_0 \subset K_1 \subset \cdots \subset K_m \supset L$$

with $K_i = K_{i-1}(\alpha_i)$. By bringing the roots of unity to the front of the tower, incorporating them in one extension and possibly enlarging the tower, we see that there is also a tower

$$K = K_0 \subset K_1 \subset \cdots \subset K_m \supset L$$

with K_1 splitting field of $X^n - 1$ over K_0 for $n = n_1 \cdot \cdots \cdot n_m$ and for $i > 1$, $K_i = K_{i-1}[\alpha_i]$ with $\alpha_i^{n_i} = \beta_{i-1} \in K_{i-1}$.

K_1 is a splitting field, so a Galois extension and in example 3.46 we saw that $\mathrm{Gal}(K_1/K_0) \cong (\mathbb{Z}/n\mathbb{Z})^*$ and thus in particular a commutative group.

Now look at K_i/K_{i-1} for $i > 1$. K_i is a root field of $X^{n_i} - \beta_{i-1}$ over K_{i-1} and K_{i-1} contains K_1, so in particular it contains $\zeta_{n_i} = e^{\frac{2\pi i}{n_i}}$. Choose a root $\alpha = \sqrt[n_i]{\beta_{i-1}}$. Now all the roots of $X^{n_i} - \beta_{i-1}$ are given by $\alpha\zeta_{n_i}^a$ for $a = 0, \ldots, n_i - 1$. They are all contained in K_i, so K_i/K_{i-1} is a Galois extension. Now for all $\sigma, \tau \in \mathrm{Gal}(K_i/K_{i-1})$, we have $\sigma(\alpha) = \alpha\zeta_{n_i}^a$, $\tau(\alpha) = \alpha\zeta_{n_i}^b$, for some a and b, so it follows that $\sigma\tau = \tau\sigma$, so $\mathrm{Gal}(K_i/K_{i-1})$ is also commutative.

So we have proved that all extensions $K_{i+1} \subset K_i$ in the tower are Galois extensions with commutative Galois groups.

The extensions $K \subset L \subset K_m$ are all Galois, so suppose $H = \mathrm{Gal}(K_m/L)$ and $G = \mathrm{Gal}(K_m/K)$. We know by the main theorem that $\mathrm{Gal}(L/K) \cong G/H$.

We first prove that G is a solvable group. We have the following tower of subgroups of G:

$$\{e\} = \mathrm{Gal}(K_m/K_m) \subset \cdots \subset \mathrm{Gal}(K_m/K_0) = G.$$

From theorem 5.13 it follows that the subgroups are all normal subgroups and that the subsequent quotients are

$$\mathrm{Gal}(K_m/K_i)/\mathrm{Gal}(K_m/K_{i+1}) \cong \mathrm{Gal}(K_{i+1}/K_i)$$

which are all commutative as we already proved, so G is a solvable group.

To prove that $\mathrm{Gal}(L/K) = G/H$ is solvable we apply a theorem from group theory that states that the quotient of a solvable group G by one of its normal subgroups H is again a solvable group. For the proof, suppose that

$$\{e\} = G_0 \subset \cdots \subset G_{m-1} \subset G_m \subset G$$

where G_i/G_{i-1} is commutative. We take this chain modulo H and find that:

$$e = G_0/(G_0 \cap H) \subset \cdots \subset G_{m-1}/(G_{m-1} \cap H) \subset G_m/(G_m \cap H) \subset G.$$

The subsequent quotients are

$$(G_i/(G_i \cap H))/(G_{i-1}/(G_{i-1} \cap H)) \cong G_iH/G_{i-1}H$$

where the last group is the image under the homomorphism "modulo H" and consequently also commutative. □

Remark 5.28 The converse is also true: a polynomial with a solvable Galois group is solvable. The proof is not included here. This raises the question on how to solve solvable equations. This topic will be not be treated here. For more information, I refer to the article [5].

Definition 5.29 Suppose that $t_1, \ldots, t_n$ are all algebraically independent over K, which means that $K[t_1, \ldots, t_n]$ is isomorphic to the polynomial ring in n variables over K. We define the *general n-th degree polynomial over K* as*:*

$$f_{\text{gen}}(X) = \prod_{i=1}^{n}(X - t_i) = \sum_{i=0}^{n} s_i X^i.$$

A solution in radicals is exactly an expression of all t_i in terms of radical functions of s_i over the field $K(s_0, \ldots, s_{n-1})$ (we take $s_n = 1$). We saw in chapter 2 that the general second degree, the third degree and the fourth degree polynomial are all solvable by radicals.

We now calculate the Galois group of the general n-th degree equation and we will see that for $n > 4$ this is not a solvable group.

Theorem 5.30 $\text{Gal}(f_{\text{gen}}/K(s_0, \ldots, s_{n-1})) \cong S_n$.

Proof: $L = K(t_1, \ldots, t_n)$ is the splitting field of f_{gen}. Furthermore we know that $\text{Gal}(f_{\text{gen}}/K(s_0, \ldots, s_{n-1})) \subset S_n$, because an element of the Galois group gives a permutation of the roots (see theorem 5.17). On the other hand for every $\sigma \in S_n$ a permutation on $\{1, \ldots, n\}$ there is an isomorphism

$$\phi : K[t_1, \ldots, t_n] \to K[t_{\sigma(1)}, \ldots, t_{\sigma(n)}], t_i \mapsto t_{\sigma(i)}$$

because the $\{t_i\}$ are algebraically independent. So there is also an isomorphism between their quotient fields $K(t_1, \ldots, t_n)$ and $K(t_{\sigma(i)}, \ldots, t_{\sigma(n)})$ (both equal to L) for all $\sigma \in S_n$. Now because $\{s_i\}$ remain unchanged if the $\{t_i\}$ are permuted (see the definition of s_i), all ϕ's are automorphisms of L over $K(s_0, \ldots, s_n)$. So all permutations $\sigma \in S_n$ are in the Galois group, which implies that $\text{Gal}(f_{\text{gen}}/K(s_0, \ldots, s_{n-1})) \cong S_n$. □

Theorem 5.31 *S_n is not a solvable group for $n \geq 5$, so the general n-th degree equation is not solvable by radicals for $n \geq 5$.*

Proof: Suppose H is a subgroup containing all 3-cycles and N is a normal subgroup of H such that H/N is commutative. We want to prove that N also contains all 3-cycles.

Suppose $\sigma = (ijk)$ and $\tau = (krs)$ are two 3-cycles for all possible choices of 5 different i, j, k, r and s. Note that this is possible because $n \geq 5$. For all such choices $\sigma, \tau \in H$, because H contains all 3-cycles. It is easily verified that $\sigma\tau\sigma^{-1}\tau^{-1} = (rki)$.

Consider the group homomorphism $\phi : H \to H/N$ with kernel N. Because H/N is commutative, we have $\sigma\tau\sigma^{-1}\tau^{-1} = 1$, so $\sigma\tau\sigma^{-1}\tau^{-1} \in \ker(\phi)$. Consequently N contains all 3-cycles (rki) for all possible choices of three different r, k and i.

Now if S_n is a solvable group, there is a tower

$$S_n = H_0 \supset H_1 \supset \cdots \supset H_r = \{e\}$$

with subsequent quotients commutative. Because S_n contains all 3-cycles, we now know by induction that H_r contains all 3-cycles, which is a contradiction. □

Theorem 5.32 For $n < 5$, S_n is a solvable group.

Proof: S_1 and S_2 are commutative. We denote with $A \rhd B$ that B is a normal subgroup of A.

$$S_3 = D_3 \rhd \mathbb{Z}_3 \rhd \{e\},$$

with $D_3/\mathbb{Z}_3 = \mathbb{Z}_2$ and $\mathbb{Z}_3/\{e\} = \mathbb{Z}_3$. And finally

$$S_4 \rhd A_4 \rhd \mathbb{Z}_2 \times \mathbb{Z}_2 \rhd \{e\}$$

with $S_4/A_4 = \mathbb{Z}_3$, $A_4/(\mathbb{Z}_2 \times \mathbb{Z}_2) = \mathbb{Z}_3$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ is commutative. □

The last thing we want to prove in this chapter is that there exist specific fifth degree polynomials that are not solvable by radicals. For that we first prove the following lemmas.

Lemma 5.33 *For a prime p, S_p is generated by any 2-cycle τ and a p-cycle σ.*

Proof: It is not difficult to see that S_p is generated by (1 2) and σ.[6] The difficulty lies in proving that we can make (1 2) with τ and σ. We renumber so that $\tau = (1\ n)$ and $\sigma = (1\ 2 \ldots p)$. We first note that

$$\sigma(1\ n)\sigma^{-1} = (2\ n+1).$$

Now replace σ by σ^{n-1} to make the first number equal to n.

$$m_1 = \sigma^{n-1}\tau\sigma^{-(n-1)} = (n\ 2n-1)$$

so if we repeat this procedure k times we get

$$m_k = (1 + k(n-1) \quad n + k(n-1)).$$

Let $m_0 = (1\ n)$ and calculate

$$m_0 m_1 \ldots m_k = (1 \quad n + k(n-1)).$$

Because p is prime, $n+k(n-1)$ runs through all integers mod p, so in particular, there is a k such that we get (1 2). □

[6] It is entertaining to convince yourself of this fact by writing the numbers 1 to 7 on small pieces of paper and see that you can put them in any order by moving the first two and moving the last to the front.

Lemma 5.34 *If p is prime and f is an irreducible polynomial over $\mathbb{Q}$ of degree p with exactly $p-2$ real roots, then $G = \mathrm{Gal}(f/\mathbb{Q}) \cong S_p$.*

Proof: Suppose L is a splitting field of f and $\alpha \in L$ a root of f. We know now that $[\mathbb{Q}[\alpha] : \mathbb{Q}] = \deg f = p$, so p is a divisor of $[L : \mathbb{Q}] = |G|$. Cauchy's theorem (a well know result from group theory) states that there is an element of order p, so G contains a p-cycle.

There also is a transposition (a 2-cycle), namely the one switching the two complex roots and leaving all real roots fixed, or in other words the restriction of complex conjugation to L.

We apply lemma 5.32 to finish the proof. □

We can now prove the existence of a specific fifth degree polynomial unsolvable by radicals.

Theorem 5.35 *The polynomial $f(x) = x^5 - 6x + 3$ over $\mathbb{Q}$ is not solvable by radicals.*

Proof: This polynomial is irreducible because of Eisenstein's criterium (theorem 3.17) for $p = 3$. From the graph of f we conclude that f has at least three simple real roots. If there are four or more, then f' has three or more, so f'' has two or more, but $f'' = 20x^3$. It follows that f exactly three real roots and we can apply lemma 5.34, so the Galois group of f is S_5 which is not a solvable group. □

6 Reduction of the quintic to normal forms

6.1 Tschirnhaus transformations

We have proven that the general fifth degree equation is not solvable by radicals, but perhaps there are other functions that do allow us to write down a solution of the general fifth degree equation. It turns out that we need elliptic curves, but before we start developing the theory of elliptic curves, we will look if we can transform the general equation to a form that is easier to handle. We can do this using so called Tschirnhaus transformations.

Definition 6.1 The transformation of an algebraic equation

$$x^n + a_1 x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \tag{13}$$

by introducing a new variable

$$y = x^m + b_1 x^{m-1} + \ldots + b_{m-1}x + b_m \tag{14}$$

to an equation of the form

$$y^n + c_1 y^{n-1} + \ldots + c_{n-1}y + c_n = 0 \tag{15}$$

is called a *Tschirnhaus transformation.*

Tschirnhaus aimed with this method to cancel out some of the coefficients of (13) by a suitable choice of $b_1, \ldots, b_m$.

Example 6.2 The first coefficient of equation (13) can be cancelled by a transformation $y = x + a_1/n$ as can easily be verified by substituting $x = y - a_1/n$ in (13).

So if we take (14) to be a first degree polynomial $y = x + b_1$, we can cancel out one term of (13) with a suitable choice of b_1. Now, Tschirnhaus reasoned, if we take (14) to be a polynomial of degree $n-1$, the choices of the coefficients $b_1, \ldots, b_{n-1}$ give us $n-1$ degrees of freedom that can be used to fulfill the $n-1$ conditions $c_1 = c_2 = \ldots = c_{n-1} = 0$. This would transform (13) into $y^n + c_0$, which is solvable by radicals. Plugging in the solution $y = \sqrt[n]{c_0}$ into (14), we then obtain a solution of (13) by solving an equation of degree $n-1$.

By induction it follows that the general n-th degree equation is solvable by radicals, which is not the case, so there must be something wrong. The problem is that the conditions which ensure that all the coefficients $c_1, \ldots, c_{n-1}$ vanish, yield a system of equations of various degrees in the parameters b_i and this system is very difficult to solve. In fact, as is explained in [15], solving this system actually amounts to solving a single equation of degree $(n-1)!$, so it

appears that this method does not work for $n > 3$, unless the resulting equation of degree $(n-1)!$ has some particular features which makes it reducible to equations of degree less than n. This turns out to be the case for $n = 4$: the resulting sextic can be seen to factorize into a product of factors of degree 2 whose coefficients are solutions of cubic equations, but for $n \geq 5$ no such simplification is apparent.

We can still use Tschirnhaus transformations to cancel out some more coefficients and make the resulting equation easier to handle. We will see that a transformation $y = x^2 + ax + b$ with a suitable choice of a and b can be used to cancel out the first two coefficients. To calculate the right a and b and the resulting coefficients we need Newton's identities.

Theorem 6.3 (Newton's identities) *Let $x_1, \ldots, x_n$ be the roots of the equation*

$$x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n = 0$$

and let $s_k = \sum_{i=1}^{n} x_i^k$. Then the following relations hold:

$$s_1 = -a_1$$

$$s_2 = -a_1 s_1 - 2a_2$$

$$s_3 = -a_1 s_2 - a_2 s_1 - 3a_3$$

and in general for $k \leq n$

$$s_k = -\sum_{i=1}^{k-1} a_i s_{k-i} - k s_k$$

and for $k > n$

$$s_k = -\sum_{i=1}^{k-1} a_i s_{k-i}$$

Proof: A proof can be found in [15] p.54-55. □

Using these relations, we can find a Tschirnhaus transformation that cancels out two terms in the general n-th degree equation.

6.2 Principal equation

Theorem 6.4 *By a linear transformation or by a quadratic Tschirnhaus transformation, whose coefficients can be found by solving a quadratic equation, the general n-th degree equation* (13) *can be transformed to a* principle equation, *that is an equation of the form*

$$x^n + a_3 x^{n-3} + a_4 x^{n-4} + \ldots + a_n = 0 \tag{16}$$

Proof: We saw in example 5.2 that with the help of a linear transformation we could transform (13) into an equation in which $a_1 = 0$. Suppose that $a_2 \neq 0$. We know from Newton's identities that $s_1 = 0$ and $s_2 = -2a_2$. Use the transformation $y = x^2 + ax + b$. Write y_k for the value of y when $x = x_k$. Then $\sum y_k = s_2 + nb$ will be zero if we take $b = -s_2/n$. Next,

$$y^2 = x^4 + 2ax^3 + (a^2 + 2b)x^2 + 2abx + b^2,$$

which implies

$$\sum y_k^2 = s_4 + 2as_3 + (a^2 + 2b)s_2 + nb^2 = 0$$

is a quadratic equation for a in which the coefficient s_2 of a^2 is not zero, and hence has two roots from which we can pick one. Note that $\sum y_k^2 = 0$ means that also the term for x^{n-2} is cancelled out in the resulting equation (15). We may compute similarly $\sum y_k^3$, $\sum y_k^4$, ... and then compute by Newton's identities the coefficients of our resulting equation, which is of the form (16). □

For $n = 5$ we call (16) the *principal quintic.*

6.3 Bring-Jerrard normal form

We shall now transform the principal equation (16) to an equation of the form

$$x^n + a_4x^{n-4} + \ldots + a_n = 0 \tag{17}$$

in which the coefficients of x^{n-1}, x^{n-2} and x^{n-3} are all zero.

Theorem 6.5 *By means of a Tschirnhaus transformation whose coefficients can be found by solving a cubic equation and three quadratic equations, the principal equation (16) can be transformed to (17).*

Proof: Consider (16). By hypothesis $a_1 = a_2 = 0$. If also $s_3 = 0$, then by Newton's identities, $a_1 = a_2 = a_3 = 0$ and no transformation is necessary. Next, let $s_3 \neq 0$. The functions

$$g(x) = x^3 + ax^2 - s_3/n, \qquad h(x) = x^4 + bx^2 - s_4/n \tag{18}$$

evidently have the property $\sum g = 0$ and $\sum h = 0$, where $\sum g$ denotes $\sum g(x_k)$. The conditions for $\sum xg = 0$, $\sum xh = 0$ are $s_4 + as_3 = 0$, $s_5 + bs_3$ and may be satisfied by choice of a and b. The function

$$\psi(x) = zg + wh \tag{19}$$

will have the property $\sum \psi^2 = 0$ if

$$z^2 \sum g^2 + 2zw \sum gh + w^2 \sum h^2 = 0 \tag{20}$$

This homogeneous quadratic equation[7] in z and w can evidently be satisfied by values of z and w not both zero. Then x and ψ have the properties

$$\sum x = 0, \quad \sum x^2 = 0, \quad \sum \psi = 0, \quad \sum \psi^2 = 0, \quad \sum x\psi = 0. \tag{21}$$

Write $y = ux + v\psi$. Then $\sum y = 0$, $\sum y^2 = 0$ for every u and v. The condition for $\sum y^3 = 0$ is a cubic equation in u/v and hence can be satisfied by values of u and v not both zero. In the resulting equation in y, the coefficients of y^{n-1}, y^{n-2}, y^{n-3} are therefore all zero. □

This theorem is usually ascribed to Jerrard, but for $n = 5$ it was obtained much earlier by E.S. Bring. We therefore call

$$y^5 + dy + e = 0 \tag{22}$$

the *Bring-Jerrard normal form of the quintic equation.*

If we now apply a transformation $y = (e/d)x$, and then set $a = d^5/e^4$ we get a form that involves only a single parameter:

$$x^5 + ax + a = 0. \tag{23}$$

6.4 Brioschi normal form

We saw that any quintic equation can be reduced to a normal form involving a single parameter by means of transformations involving only square roots and a cube root. By means of a transformation involving only square roots, we shall now reduce any sufficiently general quintic equation to a remarkable form which also involves a single parameter and which plays a central role in the theory of quintic equations.

By sufficiently general we mean that certain polynomials in the coefficients are non-zero, so the transformation does not involve division by 0. For convenience we will assume therefore that the general quintic has algebraically independent coefficients. This also implies that the roots are algebraically independent, because we have Newton's identities, so the general quintic has Galois group S_5.

We will see in the next chapter that the Brioschi quintic has Galois group A_5, so the square root in the transformation is really essential.

Theorem 6.6 *The principal quintic*

$$x^5 + a_3x^2 + a_4x + a_5, \tag{24}$$

[7] A homogeneous polynomial of degree d is a linear combination of monomials $x_1^{n_1}x_2^{n_2}\ldots$ with $\sum n_i = d$. See paragraph 8.1 for more details.

with algebraically independent coefficients and $a_5 \neq 0$ *can be reduced to the* Brioschi normal form

$$x^5 + 10Bx^3 + 45B^2x + B^2 = 0 \tag{25}$$

by means of a transformation involving at most one square root.

Proof: We start with a principal quintic $f(x)$, having therefore $s_1 = 0$, $s_2 = 0$. We assume that $s_3 \neq 0$, and use the polynomial (19) having the properties (20). We shall first prove the existence of constants p, q, r, a, b, t (p and q not both zero) such that

$$p\,\psi^2 + 2qx\psi + rx^2 - a\psi - bx + t \equiv 0 \bmod f(x), \tag{26}$$

which means that the right hand side is equal to $k(x)f(x)$ for some polynomial $k(x)$ in x. It also means that in the roots of f the right hand side equals 0.

First suppose the solution (w, z) to (19) has $w = 0$. We may then assume $z = 1$, so ψ is the cubic g. From ψ^2 we eliminate x^6 and x^5 by subtracting $k(x)f(x)$ for some $k(x)$, then eliminate x^4 by adding $2qx\psi$ for some q and finally x^3 by adding $-a\psi$ for some a. We get (26) with $p = 1$ for some r, b and t.

Second, suppose $w \neq 0$. If we take $w = 1$, we are still able to choose z such that (20) and (21) hold. We write

$$\psi = x^4 + zx^3 + dx^2 + e.$$

For some second degree polynomial Q in x, we have

$$C := x\psi - f(x) - z\psi = kx^3 + Q, \qquad k = d - z^2.$$

If $k = 0$, this gives (26) with $p = 0$ and $2q = 1$. If $k \neq 0$, we eliminate x^5 and higher powers of x from ψ^2 by means of $f(x) = 0$, then x^4 by means of ψ and finally x^3 by means of C and obtain (26) with $p = 1$. This finishes the proof of (26).

Inserting the five roots of $f(x) = 0$ into (26), summing and applying (21), we see that $t = 0$.

If χ and ϕ are linear functions of x and ψ, the relations (21) imply

$$\sum \chi = 0, \quad \sum \chi^2 = 0, \quad \sum \phi = 0, \quad \sum \phi^2 = 0, \quad \sum \chi\phi = 0, \tag{27}$$

while (26) implies

$$p\,\chi^2 + 2q\phi\chi + r\phi^2 - a\chi - b\phi \equiv 0 \bmod f(x). \tag{28}$$

It can be verified that we have the following identity

$$\begin{aligned} mF &= (d\chi + e\phi)^2 - c(a\chi + b\phi)^2 \\ F &= p\chi^2 + 2q\phi\chi + \phi^2 \end{aligned} \tag{29}$$

where

$$m = pb^2 - 2qab + ra^2, \quad c = q^2 - pr, \quad d = bp - aq, \quad e = bq - ar. \tag{30}$$

The possibility that a, b, c, d, e or m equals 0 can be excluded because this would lead to a polynomial relation between the coefficients of f, contradicting the assumption that they are algebraically independent. Therefore we can safely apply the following transformation to $f(x) = 0$:

$$y = \frac{d\chi + e\phi}{a\chi + b\phi}. \tag{31}$$

Note that all parameters d, e, a and b can be determined constructively from f, but χ and ϕ are still some unspecified linear combinations of x and ψ. The only restriction we put on the choice for χ and ϕ is that $a\chi + b\psi$ does not vanish for any root of f. Denote the resulting equation by

$$y^5 + a_1y^4 + a_2y^3 + a_3y^2 + a_4y + a_5 = 0. \tag{32}$$

We determine its coefficients as follows. By (31),

$$y + \sqrt{c} = \frac{d\chi + e\phi + \sqrt{c}(a\chi + b\phi)}{a\chi + b\phi}.$$

By combining (28) and (29) we get

$$m(a\chi + b\phi) = (d\chi + e\phi)^2 - c(a\chi + b\phi)^2 \equiv 0 \bmod f(x).$$

Remember that the possibilities that m or c equals 0 were excluded. Hence, by division,

$$\frac{m}{y + \sqrt{c}} \equiv d\chi + e\phi - \sqrt{c}(a\chi + b\phi) \bmod f(x).$$

By (27), this implies

$$\sum z = 0, \quad \sum z^2 = 0 \quad \text{if} \quad z = \frac{1}{y + \sqrt{c}}.$$

Hence if we replace y by $(1 - z\sqrt{c})/z$ in (32) we obtain

$$\begin{array}{lllllll} (1 - z\sqrt{c})^5 & + & a_1z\,(1 - z\sqrt{c})^4 & + & a_2z^2(1 - z\sqrt{c})^3 & + \\ a_3z^3(1 - z\sqrt{c})^2 & + & a_4z^4(1 - z\sqrt{c}) & + & a_5z^5 = 0. & \end{array}$$

Since these hold for both values of $\sqrt{c}$, we get

$$\begin{array}{cc} 5c^2 - 4a_1c + a_4 = 0, & 4a_1c + 2a_3 = 0 \\ 10c + 3a_2 = 0, & 6a_1c + a_3 = 0 \end{array}$$

which implies $a_1 = a_3 = 0$, $a_2 = -10/3c$, $a_4 = 5c^2$. The last coefficient, a_5 depends on the choices of χ and ψ and can be calculated by substituting (31) into (32).

Write $c = -3C$. Then (32) becomes

$$y^5 + 10Cy^3 + 45C^2y + a_5 = 0 \tag{33}$$

Remember that we excluded the possibility $c = 0$ and $a_5 = 0$, so we can introduce a new variable

$$x = y\,C^2/a_5.$$

We replace y by xa_5/C^2, multiply by C^{10}/a_5^5 and introduce the constant

$$B = C^5/a_5^2.$$

This transforms (33) to the desired equation (25). □

Remark 6.7 The assumption of algebraically independent coefficients is actually too strong. The whole procedure can be carried out as long as a, b are not both zero, d, e are not both zero and $m \neq 0$. If $c = 0$ the procedure even leads to a solvable equation of the form $x^5 + a_5$. Also if $a_5 = 0$ it is clear we get a solvable equation.

7 Solving the quintic by using the icosahedron

The Brioschi normal form of the quintic equation has very close connections to the symmetry functions of the icosahedron. For that reason we will set up some theory about the icosahedron. We will first discuss its rotation group, which will turn out to be a group that consists of quotients of linear transformations. Next we will discuss invariants of this group and the relations between them. Finally we will discuss so called *form problems*. In particular, we will see that solving the Brioschi quintic is equivalent to solving the form problem of the icosahedron.

7.1 The symmetry group of the icosahedron

Consider a regular icosahedron I in $\mathbb{R}^3$ with center O at $(0,0,0)$, top vertex V at $(0,0,1)$.

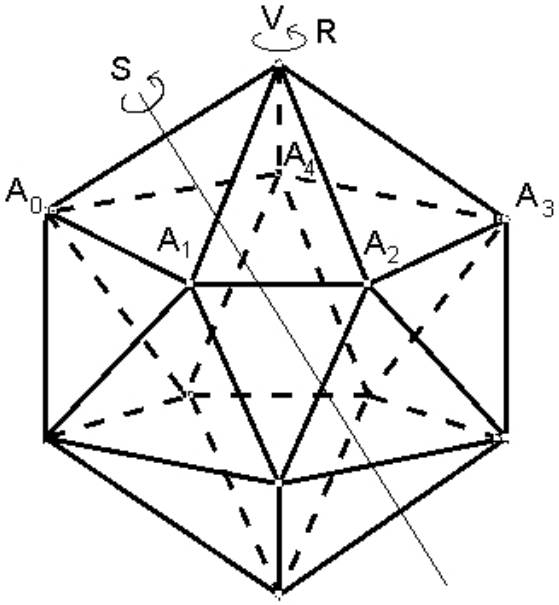

Fig. 7.1

We want to describe the group of all rotations that leave I unchanged. Note first that a rotation R of $2\pi/5$ around the axis through O and V leaves I unchanged. We define positive rotation as counterclockwise rotation when viewed from V. Note further that the rotation S of π around the axis through the midpoint of A_0V and O also leaves I unchanged.

Theorem 7.1 *The group R_{60} of all rotations that leave I unchanged is a group of order 60, generated by R and S.*

Proof: We first prove that with R and S, we can make the five rotations of π around the axis through the midpoint of A_iV for $i = 0, \ldots, 4$. Such a rotation will be denoted by $[A_iV]$.

Since R carries the midpoint of A_0V to that of A_1V, $R^{-1}SR$ leaves the latter point fixed. Since it is not the identity (V does not remain fixed), it must be $[A_1V]$. Similarly, $R^{-j}SR^j = [A_jV]$ for $j = 0, \ldots, 4$.

We can make five rotations that end up with V on the top. By applying $[A_iV]$ and then R^j for some j, we see that we can also make five rotations that end up with A_i on top for all A_i. We will now continue to prove that we can get all vertices on top. By applying $[A_iV]$ for a proper i when A_j is on top for some j, we see that we can also get all the vertices connected to A_i on top, so in particular we can get the vertices connected to V' on top. Finally by applying $[A_iV]$ again when one of the vertices connected to V' is on top, we can get V' on top.

So we can get all 12 vertices on top and we can make all 5 rotations end up with that vertex on top. These are all rotations because every rotation ends up with some vertex on top and S is the only rotation that leaves the top fixed. Therefore all 5×12 rotations of R_{60} can be made with R and S. □

Remark 7.2 Since S interchanges V and A_0, as well as A_1 and A_4, the product $RS = (V\ A_0A_4)$ is a rotation of order 3 whose axis joins O to the mid point of the face VA_0A_4. It now follows that R_{60} consists of

- The identity
- 15 rotations of order 2 about diameters passing through the mid points of the 15 pairs of opposite edges
- 10×2 rotations of order 3 about diameters passing through the mid points of the 10 pairs of opposite faces
- 6×4 rotations of order 5 about diameters passing through the 6 pairs of opposite vertices.

It follows that R_{60} is isomorphic to A_5 and that R corresponds to a permutation of the form (1 2 3 4 5), whereas S corresponds to a permutation of the form (2 3)(4 5).

7.2 Projection of the Riemann sphere to the complex plane

A regular polyhedron, like the icosahedron can be represented as points on the surface of a unit sphere, called the *Riemann sphere*.

We can project the sphere

$$S_2 = \{(x, y, z) \,|\, x^2 + y^2 + z^2 = 1\} \tag{34}$$

to its equatorial (complex) plane, by drawing the line from $N = (0, 0, 1)$ through a point $P = (x, y, z)$ on the sphere and take the intersection of this line with the equatorial plane. Every point P on the sphere thus corresponds to a point $Z = a + bi$ in the complex plane, provided we agree to identify all points at infinity, so that $Z = \infty$ corresponds to N.

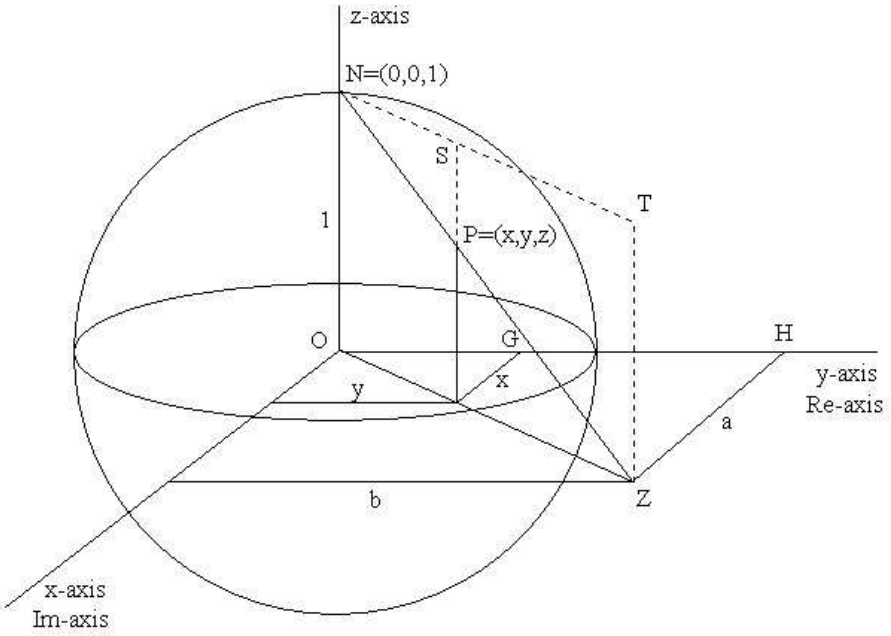

Fig. 7.2

We see that the triangles NSP and NTZ are similar, as are the triangles OFG and OZH. Furthermore they have the same proportion, because $NS = OF$ and $NT = OZ$. Therefore we get

$$SP : TZ = (1 - z) : 1 = NS : NT = OF : OZ = x : a = y : b.$$

Which leads to

$$a = \frac{x}{1 - z} \quad , \quad b = \frac{y}{1 - z} \quad , \quad Z = a + bi. \tag{35}$$

Conversely, by using (34) we find that

$$1 + a^2 + b^2 = \frac{(1 - z)^2 + x^2 + y^2}{(1 - z)^2} = \frac{1 - 2z + 1}{(1 - z)^2} = \frac{2}{1 - z}, \tag{36}$$

which leads to the following expression of x, y and z in terms of a and b

$$x = 2a/R \ , \ y = 2b/R \ , \ z = 1 - 2/R \ , \ R = 1 + a^2 + b^2, \tag{37}$$

or in terms of Z and $\overline{Z} = a - bi$

$$x = \frac{Z + \overline{Z}}{1 + Z\overline{Z}} \ , \ y = \frac{i(\overline{Z} - Z)}{1 + Z\overline{Z}} \ , \ z = \frac{Z\overline{Z} - 1}{1 + Z\overline{Z}}. \tag{38}$$

These formulas establish a one-to-one correspondence between the points of the unit sphere and the points of the complex plane.

It is a fact (see [4] p.223) that all rotations of the Riemann sphere correspond to fractional linear transformations on the complex plane. In particular, the rotations R and S of the previous paragraph that generate the symmetry group of the icosahedron correspond to such transformations. We will now calculate which ones.

We take the coordinate axis as in the last paragraph and we embed the icosahedron I into R^3 as in figure 7.1. Furthermore we let the vertex A_0 correspond to a real negative number n.

Theorem 7.3 *Let $\zeta = e^{i\pi/5}$, $m = \zeta + \zeta^4$, and $n = \zeta^2 + \zeta^3$. The rotations R and S correspond to transformations*

$$R(Z) = \zeta Z, \tag{39}$$

$$S(Z) = \frac{nZ + 1}{Z - n}. \tag{40}$$

Furthermore all vertices of I are located at

$$Z = 0, \ \infty, \ \zeta^k n, \ \zeta^k m, \quad \textit{where} \quad k = 0, \ldots, 4 \tag{41}$$

Proof: The fact that $R(Z) = \zeta Z$ is obvious, because ζ is exactly a rotation of $2\pi i/5$ counterclockwise around OV when seen from V.

Applying powers of R to n, we see that A_i corresponds to $\zeta^k n$ for $k = 0, \ldots, 4$. From (38) with $Z = \overline{Z}$ it follows that $-A_0$, the opposite point of A_0, corresponds to $m = -1/n$ and therefore the vertices $-A_k$ correspond to $\zeta^k m$. Together with the fact that V corresponds to infinity and V' to 0, we have proved the theorem except for the values of m and n.

Since S interchanges A_0 and V, which correspond to n and ∞ and interchanges the opposite points $m = -1/n$ and 0, we see that $S(n) = \infty$ and $S(m) = 0$, thus

$$S(Z) = \frac{nZ + 1}{Z - n}.$$

Since S interchanges A_1 and A_4, $Z = \zeta n$ implies $S(Z) = \zeta^4 n$. Therefore $1 = n^2(1 - \zeta - \zeta^4) = n^2(\zeta + \zeta^4)$ and because $\zeta + \zeta^4$ is positive, but n negative, we see that $n(\zeta + \zeta^4) = -1$, so $n = \zeta^2 + \zeta^3$ and $m = \zeta + \zeta^3$. □

7.3 Invariants of the symmetry group of the icosahedron

Before calculating the invariants under the rotations, we switch to homogeneous variables by setting $Z = u/v$. Note that $Z = \infty$ now corresponds to $v = 0$. We see that the polynomial that vanishes at all vertices (41) is

$$f = uv(u^5 - m^5v^5)(u^5 - n^5v^5).$$

f is invariant under all rotations of R_{60}.

We want to calculate the coefficients of f. First we write out

$$f = uv(u^{10} - (m^5 + n^5)u^5v^5 + m^5n^5v^{10}).$$

To calculate $m^5 + n^5$, we note that m and n are the two roots of $x^2 + x - 1$. Write s_r for the sum of the r-th powers of the roots of this polynomial, multiply by x^{r-2} and sum. We get

$$s_r + s_{r-1} - s_{r-2} = 0.$$

Using this recursion formula for $r = 5$, we get

$$s_5 = -s_4 + s_3 = 2s_3 - s_2 = -3s_2 + 2s_1 = 5s_1 - 6 = -11.$$

Calculating n^5m^5 is easy because $nm = -1$, so we get

$$f = uv(u^{10} + 11u^5v^5 - v^{10}). \tag{42}$$

The Hessian (named after Otto Hesse) of $f(u, v)$ is defined as the determinant

$$\begin{vmatrix} \frac{\partial^2 f}{\partial u^2} & \frac{\partial^2 f}{\partial u \partial v} \\ \frac{\partial^2 f}{\partial u \partial v} & \frac{\partial^2 f}{\partial v^2} \end{vmatrix}$$

which, applied to (42), gives $121H$ where

$$H = -u^{20} - v^{20} + 228(u^{15}v^5 - u^5v^{15}) - 494u^{10}v^{10}. \tag{43}$$

After writing H in inhomogeneous form (divide H by v^{20} and set $Z = u/v$), we see that all powers of Z are multiples of 5, so H is invariant under $R(Z) = \zeta_5 Z$. H is also invariant under the transformation $F(u, v) = (-v, u)$, or in inhomogeneous form $F(Z) = -1/Z = -\overline{Z}$. It is not difficult to see that this transformation is rotation of 180° around the x-axis and therefore belongs to R_{60}. We will prove in the next paragraph that it is of the form (1 2)(3 4). Because R is of the form (1 2 3 4 5), R and F generate A_5, so H is invariant under $A_5 \cong R_{60}$.

The Jacobian of f and H is defined as

$$\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial H}{\partial u} & \frac{\partial H}{\partial v} \end{vmatrix}$$

which yields $20T$ where

$$T = u^{30} + v^{30} + 522(u^{25}v^5 - u^5v^{25}) - 10005(u^{20}v^{10} + u^{10}v^{20}). \qquad (44)$$

We calculate $R(T)$ and $F(T)$ as we did for H and find that T is invariant under R_{60}.

Remark 7.4 In fact f, H and T are invariant under the linear homogeneous icosahedral group H_{120}, because by switching to homogeneous variables, we got the extra transformation $(u, v) \mapsto (-u, -v)$, which also leaves f, H and T invariant. In inhomogeneous variables, this would be the identity, because $z = \frac{u}{v} = \frac{-u}{-v}$, but in homogeneous variables it is an extra transformation of order 2. If we divide H_{120} by its normal subgroup of order 2, we get $R_{60} \cong A_5$.

We will now give a geometric meaning to the invariants f, H and T.

Remark 7.5 A *special point* of the Riemann sphere is defined as a point which takes fewer than 60 positions under the group R_{60}. Therefore it must be a point on the axis of one the rotations other than the identity. It follows from remark 7.2 that every special point belongs to one of the following sets: the 12 vertices (at which f vanishes), the central projections on the sphere of the mid points of the 20 faces and those of the mid points of the 30 edges. f, H and T all have degrees < 60, so they must vanish at special points. Since H is of degree 20 and T is of degree 30, H must vanishes at the central projections on the sphere of the mid points of the 20 faces and T at those of the mid points of the 30 edges.

Theorem 7.6 *The invariants satisfy the identity*

$$T^2 = 1728f^5 - H^3 \qquad (45)$$

Proof: Consider any homogeneous polynomial in u and v which is invariant under H_{120}. If it vanishes for a special point, it has one of the factors f, H or T. After removing such factors, we obtain an invariant quotient Q, which vanishes for no special point. Let f and g denote any two of the absolute invariants f^5, H^3 or T^2, each of degree 60. Then also $f - cg$ is an invariant, if c is any constant. The condition that it shall vanish at U, V uniquely determines c, since z' is not a special point and therefore $g(U, V) \neq 0$. Hence Q has the factor $f - cg$. The quotient is either a constant or has another such factor. So every polynomial in u and v which is invariant under H_{120} is a product of factors f, H, T, $f - cg$, where the c's are constants $\neq 0$. Hence an invariant of degree

60 which vanishes at no special point can be expressed in each of the forms $a(f^5 - cH^3)$, $b(f^5 - dT^2)$, where a, b, c, d, are constants $\neq 0$. Thus f^5, H^3 and T^2 satisfy a linear identity

$$kf^5 + lH^3 + mT^2 = 0.$$

By comparing the coefficients in inhomogeneous form for Z^{60}, we find that $l = m = 1$ and by comparing the coefficients for Z^5, we find that

$$-k - 3 \times -228 + 2 \times -522 = 0$$

and thus $k = -1728$. It follows that $T^2 = 1728f^5 - H^3$. □

To conclude this paragraph, we give f, H and T in their inhomogeneous forms.

$$f = Z(Z^{10} + 11Z^5 - 1) \tag{42'}$$

$$H = -(Z^{20} + 1) + 228(Z^{15} - Z^5) - 494Z^{10} \tag{43'}$$

$$T = (Z^{30} + 1) + 522(Z^{25} - Z^5) - 10005(Z^{20} - Z^{10}) \tag{44'}$$

7.4 The form problem of the icosahedron

For every regular polyhedron, there exist three invariants under the group of rotations that leave that polyhedron fixed, namely the one vanishing at the vertices, the one vanishing at the central projections of the midpoints of the edges and the one vanishing at those of the faces. The form problem for a certain regular polyhedron is the problem of finding the pairs (u, v), when the invariants are given. For the icosahedron, this amounts to the problem of finding the 120 pairs u, v, when f, H and T are given in accordance to equation (45).

Assume we that the values for f, H and T are known and satisfy (45). We first try to find a value for Z such that the invariant equations (42'), (43') and (44') hold. We care about finding (u, v) later.

We could for instance try to find a solution to $f = Z(Z^5 - m)(Z^5 - n)$, but we can also write out something like $fH^2 + T^5$ in terms of Z and try to find Z so that it takes the given value. For reasons that are perhaps a little obscure now, but that will become clear later, we write out

$$B = -f^5/T^2. \tag{46}$$

in terms of Z and consider the problem of finding a Z for the given B. The equation we chose to solve has now become $f^5 + BT^2 = 0$, which in homogeneous form is equal to

$$Z^5(Z^{10} + 11Z^5 - 1)^5 + B[Z^{30} + 1 + 522(Z^{25} - Z^5) - 10005(Z^{20} + Z^{10})]^2 = 0. \tag{47}$$

This equation of degree 60, having single parameter B is called the *icosahedral equation*.

When we have found Z, we find the corresponding (u, v) by defining

$$X(u, v) = fH/T.$$

Since X is of degree 2 in u, v, we have

$$v^2 = X(u, v)/X(Z, 1), \qquad Z = u/v,$$

Which can be solved by taking the square root. So if we know Z, we find

$$v = \pm\sqrt{\frac{fH}{TX(Z, 1)}}, \text{ and } u = vZ.$$

Hence the form problem for the icosahedron has been reformulated to the problem of finding Z, when f, H and T are given, or in other words, finding a solution to (47).

Our next step will be to prove that that the icosahedral equation can be reduced to the Brioschi normal form of the quintic, but before we can do that, we first need to prove the following theorem.

Theorem 7.7 *The rotation group of the icosahedron R_{60} is isomorphic to the group of all even permutations that permute, the five octahedra $t_0, \ldots, t_4$ with vertices on the midpoints of the edges of the icosahedron.*

Proof: Consider again f from equation (42'). We note that f is invariant under the transformations

$$F(u, v) = (-v, u), \text{ so } F(Z) = -1/Z. \tag{48}$$

Therefore the corresponding rotation F leaves I unchanged, so that (48) belongs to H_{120}. We note that it sends $A_0 = n$ to m and leaves i and $-i$ fixed. Therefore it is rotation of π around the x-axis. Remember that S is rotation of π around OP, where P is the mid point of A_0V. The product SF must therefore be rotation of π around the axis which is a common perpendicular to the x-axis and OP. Note that these three rotation axis are perpendicular. The points where they intersect the sphere are therefore the vertices of a regular octahedron.

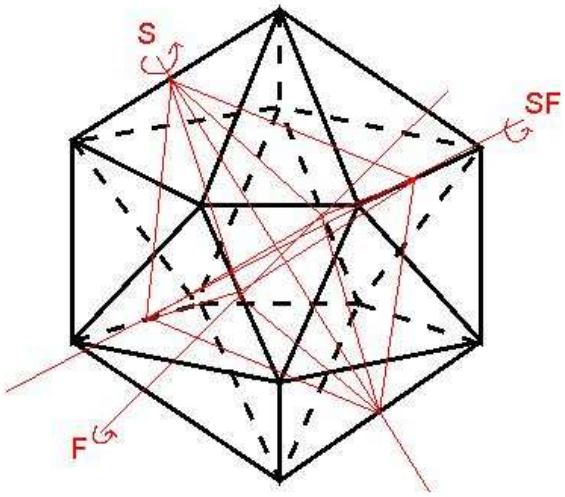

Fig. 7.3

The fixed points of F can be found by setting $F(Z) = Z$, which gives $Z^2+1 = 0$. Similarly we can find the fixed points of S from setting $S(Z) = Z$ in (40). Finally for SF we calculate $S(-1/Z) = Z$. After setting $Z = u/v$ we find that the fixed points of F, S and SF are respectively those for which

$$A_0 = u^2 + v^2, \quad B_0 = u^2 - 2nuv - v^2, \quad C_0 = u^2 - 2muv - v^2 \tag{49}$$

vanish. Hence the points for which

$$t_0 = A_0B_0C_0 = u^6 + 2u^5v - 5u^4v^2 - 5u^2v^4 - 2uv^5 + v^6 \tag{50}$$

vanish correspond to the vertices of the octahedron. If we now apply R to this octahedron, we find another octahedron with one vertex at the mid point of A_1V. We see that by applying R^k for $k = 0, \ldots 4$ we find 5 octahedra.

To find the homogeneous polynomials that vanish at the vertices of these octahedra, we first write R^k in one of its homogeneous forms

$$R^k : \quad U = \zeta^{3k}u, \quad V = \zeta^{2k}v$$

and apply this to (49) to find

$$A_k = \zeta^k u^2 + \zeta^{4k} k v^2, \quad B_k = \zeta^k u^2 - 2nuv - \zeta^{4k} v^2, \quad C_k = \zeta^k u^2 - 2muv - \zeta^{4k} v^2.$$

We write $t_k = A_kB_kC_k$.

Since the rotations of R_{60} merely changes the places of the 5 octahedra, R and S permute the 5 octahedra. With some geometrical insight and help of the following picture

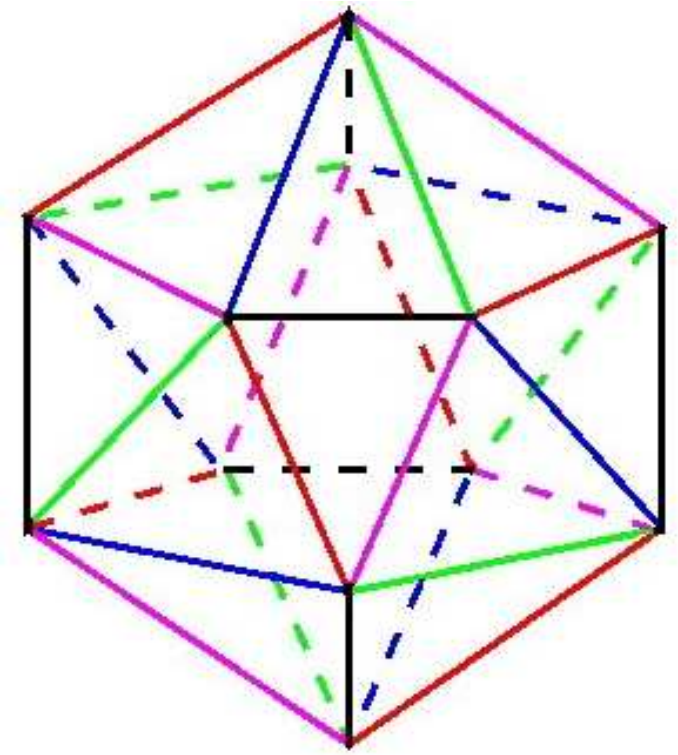

Fig. 7.4

one can see that S interchanges the octahedra t_1 and t_2 as well as the octahedra t_3 and t_4. Therefore we get

$$R = (t_0 t_1 t_2 t_3 t_4), \qquad S = (t_0)(t_1 t_2)(t_3 t_4).$$

Together these permutations generate A_5. □

Theorem 7.8 *The octahedral functions $t_0, \ldots, t_4$ are the roots of a quintic equation in the Brioschi normal form (25).*

Proof: Writing out the expression

$$(t - t_0)(t - t_1)(t - t_2)(t - t_3)(t - t_4) = 0$$

gives a fifth degree polynomial

$$t^5 + c_1 t^4 + c_2 t^3 + c^3 t^2 + c^4 t + c^5 = 0$$

whose coefficients are symmetric polynomials of the roots $t_0, \ldots, t_1$. Because t_i is of degree 6 in u, v, we see that c_k is of degree $6k$ in u, v. The largest $6k$ is 30, which is < 60, so from remark 7.4 it follows that c_k is a product of factors f, H

and T, whose degrees are 12, 20 and 30. Since no such product is of degree 6 or 18, we have $c_1 = c_3 = 0$. Also, $c_2 = af$, $c_4 = bf^2$, $c_5 = cT$, where a, b, c are constants.

Let s_k be the sum of the k-th powers of the roots. Since $c_1 = c_3 = 0$ two of Newton's identities reduce to $s_2 + 2c_2 = 0$ and $s_4 + c_2 s_2 + 4c_4 = 0$. Thus

$$s_2 + 2af = 0, \quad s_4 + (4b - 2a^2)f^2 = 0, \quad cT + t_0 t_1 t_2 t_3 t_4 = 0. \tag{51}$$

We want to calculate a, b and c by comparing terms from s_2, s_4 and $t_0 t_1 t_2 t_3 t_4$. Writing out a few terms of $t_k = A_k B_k C_k$ gives

$$t_k = \zeta^{3k} u^6 + 2\zeta^{2k} u^5 v - 5\zeta^k u^4 v^2 + \dots$$

and thus

$$t_k^2 = \zeta^k u^{12} + 4u^{11} v - 6\zeta^{4k} u^{10} v^2 + \dots$$
$$t_k^4 = \zeta^{2k} u^{24} + 8\zeta^k u^{23} v + 4u^2 2v^2 + \dots.$$

Comparing the coefficients of $u^{11}v$, $u^{22}v^2$ and u^{30} in the respective functions (51) gives

$$20 + 2a = 0, \qquad 20 + 4b - 2a^2 = 0, \qquad c + 1 = 0$$

which implies $a = -10$, $b = 45$, $c = -1$. This means that we have

$$t^5 - 10ft^3 + 45f^2 t - T = 0. \tag{52}$$

Introduce in place of t the new variable $x = -tf^2/T$, which is of degree zero in u and v and hence a function of Z. Replace t by $-xT/f^2$, multiply all terms by $-f^{10}/T^5$ and write B for $-f^5/T^2$, in accordance to (46). We get the Brioschi quintic

$$x^5 + 10Bx^3 + 45B^2 x + B^2 = 0.$$

□

Note that because the B we have here is precisely the B we have used to formulate the icosahedral equation, we have reduced the icosahedral equation (and thus the form problem) to the Brioschi quintic equation.

We now have enough to conclude with a theorem that sums up all results in this and the previous chapter.

Theorem 7.9 *Solving the general quintic equation is equivalent (up to radicals) to solving the form problem of the icosahedron.*

Proof: We first prove we can solve the general quintic if we know a solution to the form problem. From the coefficients of the general quintic, we determine the coefficients of a corresponding principal quintic, by following the procedure in the proof of theorem 6.4. From the coefficients of this principal quintic we determine the parameter B of the Brioschi quintic by following the proof of theorem 6.6.

Now if we are somehow able to solve the form problem of the icosahedron, we can find (u, v) for given values of f and T. We choose B, such that $B = f^5/T^2$ and consider u and v known. From (u, v), we calculate a solution t_0 to (52). Therefore $x_0 = t_0 f^2/T$ is a solution to the Brioschi quintic as desired.

On the other hand, if we are able to solve the general quintic equation, then in particular, we can solve (52) for any given invariants f, H and T such that $1728f^5 = T^2 + H^3$. The $t_0, \ldots, t_4$, can be used in $t_k = A_k B_k C_k$ to obtain 5 equations of degree 6. We can use the first few to lower the degree of the last one and obtain a solvable equation in (u, v). This means we can solve the form problem. □

8 Solving the quintic by using elliptic curves

In this section we will develop some theory on elliptic curves, which are cubic curves of the form $y^2 = x^3 + ax + b$ in the projective plane. We will see that the points on such a curve form a group under a remarkable addition that is defined by a geometric construction on the graph.

Next we will look at group elements of finite order, called torsion points. We will focus our attention both on the algebra as on the geometry of these points.

Following this, we will see how we can associate to each Brioschi quintic a certain elliptic curve in such a way that finding the 5-torsion points on this elliptic curve is equivalent to solving the quintic. We will also see that the associated elliptic curves have a lot of connections to the icosahedron, in particular to the functions f, H and T.

Finally we will look at the field extensions that arise from the various quintics, the icosahedral equation and the 5-torsion polynomial and we will use Galois theory to bring all of them together.

8.1 Introduction to elliptic curves in the projective plane

We will study cubic curves of the form

$$C : y^2 = x^3 + ax + b. \tag{53}$$

If we consider solutions in $\mathbb{R}^2$, we can draw the graph. We can distinguish four different curves:

1. $f(x)$ has three real roots, for instance $y^2 = x^3 - x$, with $\Delta = 64$:

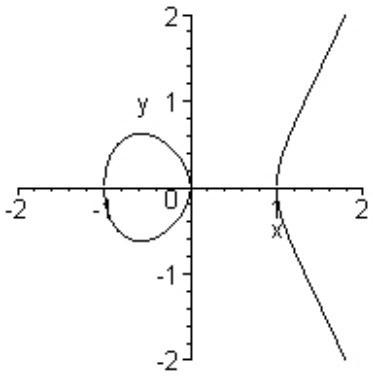

Fig. 8.1

2. $f(x)$ has one real and two complex roots, for instance $y^2 = x^3 - 3x + 3$, with $\Delta = -2160$ or $y^2 = x^3 + x$, with $\Delta = -64$:

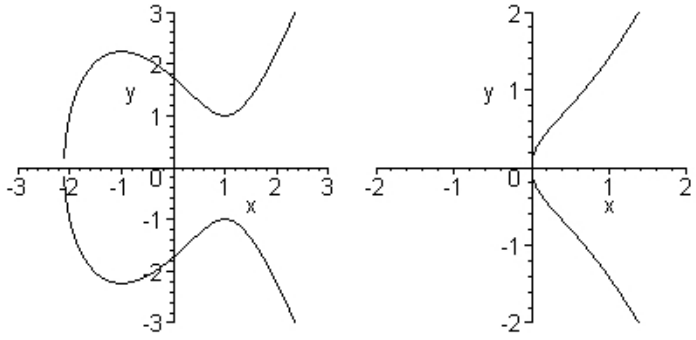

Fig. 8.2

3. $f(x)$ has a double root, for instance $y^2 = x^3 + x^2$, with $\Delta = 0$:

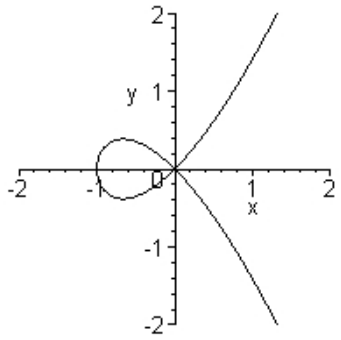

Fig. 8.3

4. $f(x)$ has a triple root, for instance $y^2 = x^3$, with $\Delta = 0$:

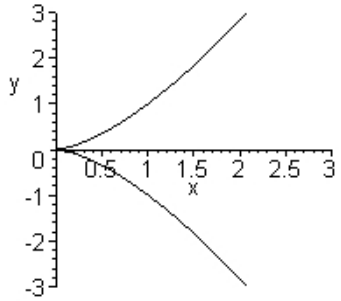

Fig. 8.4

A curve $C(x, y) = 0$ is called a *singular curve* if both partial derivatives vanish at a certain point. If we write (53) as $C(x, y) = y^2 - f(x) = 0$ and calculate the partial derivatives

$$\frac{\partial F}{\partial x} = -f'(x), \qquad \frac{\partial F}{\partial y} = 2y,$$

we see that this happens only on the last two curves. On the other hand, we see that on the first two curves, every point has a well-defined tangent line.

In section 2.2 we saw that the discriminant Δ of f is equal to

$$\Delta = \frac{a^3}{27} + \frac{b^2}{4}.$$

This means that C is singular if and only if $\Delta = 0$.

We are only interested in non-singular curves. In fact, we want to study non-singular curves over the so called *projective plane* $\mathbb{P}^2$. We give two definitions of $\mathbb{P}^2$, one algebraic and one geometric. First the algebraic one:

Definition 8.1 (Projective Plane 1)

$$\mathbb{P}^2 = \frac{\{(X, Y, Z) : \ X, Y, Z \text{ not all zero}\}}{\sim},$$

where $\sim$ is an equivalence relation defined by the rule that $(x, y, z) \sim (X, Y, Z)$ if there is a non-zero t such that

$$X = tx, \quad Y = ty, \quad Z = tz. \tag{54}$$

The numbers X, Y, Z are called *homogeneous coordinates.*

Remark 8.2 The reason for calling X, Y, Z homogeneous coordinates is that, if we have an algebraic curve of degree d in the affine plane defined by $f(x, y) = 0$ and we apply the substitutions $x = X/Z,\ y = Y/Z$, we get a homogeneous curve in X, Y, Z (after multiplication by Z^d). On the other hand, if we have a homogeneous curve defined by $F(X, Y, Z) = 0$, we can find its affine part by dividing it by Z^d and substituting $X = Zx,\ Y = Zy$.

If we apply these substitutions to (53), we can formulate our first definition of an elliptic curve.

Definition 8.3 (Elliptic Curve 1) An *elliptic curve* $E(K) \subset \mathbb{P}^2$ over a field K is a set of points that satisfy a non-singular homogeneous equation in the projective plane of the form

$$E : Y^2Z = X^3 + aX^2Z + bXZ^2 + cZ^3, \quad a, b, c \in K. \tag{55}$$

Note that if $Z \neq 0$, we can divide by Z^3 to get back equation (53). This holds in general, for if we have a homogeneous equation in the projective plane and make the substitution (54) with $t = 1/Z$, we see that $(X, Y, Z) \sim (X/Z, Y/Z, 1)$, so the part of $\mathbb{P}^2$ with $Z \neq 0$ is just the two dimensional *Affine plane*. We will call (53) the *affine part* of (55).

If $Z = 0$ however, equation (55) has a solution that (53) seems to lack. In fact this solution lies "at infinity". This term comes from the fact that, if we take the limit for $Z \to 0$ in the substitution $(x, y) = (X/Z, Y/Z)$, the coordinates x and y go to infinity. In general, the points with $Z = 0$ form a line $(X, Y, 0) \sim (1, Y/X, 0)$, together with a point $(0, 1, 0)$ that lies "at the end of the line". Together these point form a *projective line at infinity*, which we will denote by $\mathbb{P}$. This leads to our second definition of the projective plane.

Definition 8.4 (Projective Plane 2)

$$\mathbb{P}^2 = \mathbb{A}^2 \cup \mathbb{P}.$$

The geometric intuition we should follow here is that we have the usual affine plane, together with a set of (non-oriented) *directions* in $\mathbb{A}^2$. The direction of a line $ax+by = 0$ is best characterized by its derivative dy/dx. Note that, because a direction is an element of $\mathbb{P}$, we don't exclude vertical lines with $dy/dx = \infty$. So we can define a direction as a set of equivalence classes of all lines that have the same derivative as a given line. This means that two "parallel" lines in $\mathbb{P}^2$, have a point in common, namely their direction point, which is a point *at infinity*. In fact $\mathbb{P}^2$ has no parallel lines at all!

Because the points at infinity also form a projective line, every line in $\mathbb{P}^2$ intersects the line at infinity in a single point corresponding to its direction. Furthermore it is not difficult to see that there is a unique line going through every two distinct points. This means that every two distinct lines intersect each other at exactly one point. Bezout's theorem states a more general result:

Theorem 8.5 (Bezout's theorem) *Two projective curves C_1, C_2 with no common components intersect at* $(\deg C_1)(\deg C_2)$ *points in $\mathbb{C}$, counting multiplicity.*

Proof: A detailed proof can be found in Silverman [1] page 242-251. □

Now where does the elliptic curve (54) intersect the line at infinity? The right intuition we should follow here is that the curve is somehow "missing" a point that lies at the end of the curve. If we move along the curve the derivative dy/dx goes to $\pm\infty$, so it must be the infinite direction $(0, 1, 0)$. Because the line at infinity intersects the curve only at this point, Bezout's theorem theorem implies this is an intersection of multiplicity three.

We can now formulate a simpler definition of an elliptic curve that is equivalent to our first definition.

Definition 8.6 (Elliptic Curve 2) An *elliptic curve* E over a field K is a set of points (x, y) in K^2 that satisfy a non-singular cubic equation of the form

$$E : y^2 = x^3 + ax + b, \quad a, b \in K$$

together with the infinite direction point $\mathcal{O} \in \mathbb{P}^2$.

8.2 The group structure

In this section we will see that points on an elliptic curve form a group under an addition law that is best described by a geometrical construction on the graph of the curve. We define this addition as follows:

Definition 8.7 (Group law) To add two points P and Q, we first draw the line through P and Q and find a third intersection point $P * Q$. Then we draw the line through $P * Q$ and $\mathcal{O}$, which is just the vertical line through $P * Q$ and intersect it with the curve to find $P + Q$. An elliptic curve is symmetric about the x axis, so to find $P + Q$, we only have to reflect $P * Q$ in the x-axis. We illustrate this procedure with the following picture:

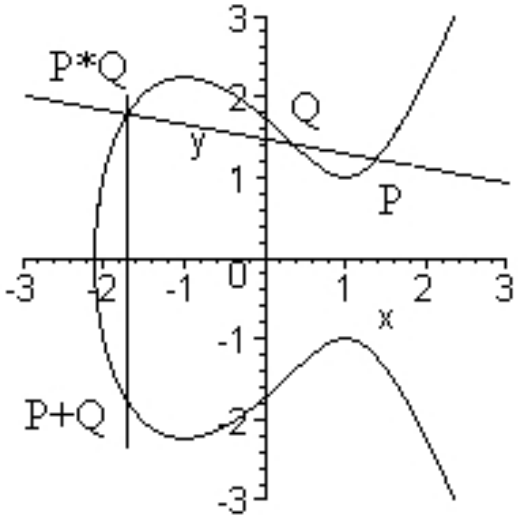

Fig. 8.5

Remark 8.8 Note that P and Q need not be distinct. If $P = Q$, the curve intersects the line through P and Q twice at P, which means that this line is the tangent line to the curve. Note furthermore that, because we are working in $\mathbb{P}^2$ and take multiplicity into account, we always find a third intersection point $P * Q$, so our addition is well-defined. Moreover it is immediately clear that it is commutative.

The zero element of this addition is the element $\mathcal{O}$. A look at figure ?? (and considering some special cases) shows that $P + \mathcal{O} = P$ for all P on E. Finding the additive inverse is also not difficult. If, for a point $P = (x, y)$, we define $-P = (x, -y)$, we find that $P + (-P) = \mathcal{O}$, so all we need to check to complete our definition of the group is associativity. This turns out to be more difficult. We first give explicit formulas for the addition.

We start with an elliptic curve E of the form (58) with two distinct points P and Q on E and set

$$P = (x_1, y_1), \quad Q = (x_2, y_2), \quad P * Q = (x_3, y_3), \quad P + Q = (x_3, -y_3).$$

We first assume that $P \neq Q$, so we can look at the equation of the line joining P and Q. This line has the equation

$$y = \lambda x + \nu, \quad \text{where} \quad \lambda = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{and} \quad \nu = y_1 - \lambda x_1 = y_2 - \lambda x_2.$$

To find the x-coordinate of the third intersection point on this line, we substitute

$$y^2 = (\lambda x + \nu)^2 = x^3 + ax + b.$$

Working out the brackets and putting everything on one side yields

$$x^3 - \lambda x^2 + (a - 2\lambda\nu)x + b - \nu^2 = 0.$$

This is a cubic equation in x and its three roots x_1, x_2, x_3 give us the x coordinates of the three intersections. Thus

$$x^3 - \lambda x^2 + (a - 2\lambda\nu)x + b - \nu^2 = (x - x_1)(x - x_2)(x - x_3) = 0.$$

Comparing the coefficients of the x^2 term on both sides, we find that $\lambda^2 = x_1 + x_2 + x_3$, so we can express x_3 in terms of x_1 and x_2 and then calculate y_3 with the equation of the line:

$$x_3 = \lambda^2 - x_1 - x_2, \qquad y_3 = \lambda x_3 + \nu. \tag{56}$$

These formulas give a way to compute the sum of two points. Note that the result of this computation does not depend on the coefficients a and b of the elliptic curve. This means that, if we have two distinct points on an elliptic curve of the form (58) and if we assume that the addition does not involve intersections with multiplicity > 1, we can compute the sum without having to know the coefficients of the curve!

If $P = Q$, we can calculate $P + Q$ by calculating an equation for the intersection of the tangent line to P with E. By implicit differentiation we find that

$$\lambda = \frac{dy}{dx} = \frac{f'(x)}{2y} = \frac{3x^2 + a}{2y}.$$

If we substitute this into the formulas we found earlier, put everything over a common denominator, and replace y^2 by $f(x)$, we find that

$$x \text{ coordinate of } 2(x,y) = \frac{x^4 - 2ax^2 - 8bx + a^2}{4x^3 + 4ax + 4b}.$$

This formula is often called the *duplication formula*.

It is now possible to check that the addition is associative by direct calculation with the formulas above. This is a lot of tedious work and there are a lot of special cases to consider, like when one point is the negative of the other or when two points coincide. On top of that there is very little learning in writing or reading such a proof. To my knowledge no such a proof has ever been published.

Fortunately there exist more elegant ways of proving associativity. We need the following lemma in which we use the "language" of Bezout's theorem, so we count points by their multiplicities and allow complex and infinite points.

Lemma 8.9 *Let C, C_1 and C_2 be three cubic curves. Suppose C goes through eight of the nine intersection points of C_1 and C_2, then C goes through the ninth intersection point.*

Sketch of proof:[8] The trick is to consider the problem of constructing a cubic curve which goes through a certain set of points. To define a cubic curve

$$ax^3 + bx^2 + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0$$

we have to give ten coefficients. If we multiply all the coefficients by a constant, then we get the same curve, so really the set of all possible cubics is, so to speak, 9-dimensional. Now if we want the cubic to go through a point, that imposes one linear condition on the coefficients so the set of all possible cubics that go through one given point is 8-dimensional. Each time you impose a condition that the cubic should go through a given point, that imposes an extra linear condition on the coefficients. Thus the family of all cubics which go through the eight points of intersection of the two given cubics C_1 and C_2 forms a 1-dimensional family.

Let $F_1(x,y) = 0$ and $F_2(x,y) = 0$ be the cubic equations giving C_1 and C_2. We can then find cubics going through the eight points by taking linear combinations $\lambda_1 F_1(x,y) + \lambda_2 F_2(x,y)$. Because the cubics going through the eight points form a 1-dimensional family and because the set of cubics $\lambda_1 F_1(x,y) + \lambda_2 F_2(x,y)$ is a 1-dimensional family, we see that the cubic C has an equation $\lambda_1 F_1(x,y) + \lambda_2 F_2(x,y) = 0$ for a suitable choice of λ_1, λ_2.

Now how about the ninth point? Since that ninth point is on both C_1 and C_2, we know that $F_1(x,y)$ and $F_2(x,y)$ vanish at that point. It follows that

[8] Silverman [1] page 17

$\lambda_1 F_1(x,y) + \lambda_2 F_2(x,y)$ also vanishes there and this means C also contains that point. □

Now we are ready to prove associativity.

Theorem 8.10 *The group law on an elliptic curve E is associative. In other words, for all points A, B, C on E, we have*

$$(A+B)+C = A+(B+C).$$

Proof: It is enough to show that

$$-((A+B)+C) = -(A+(B+C)).$$

Consider the following lines:

L_1	is the line through	A,	B,	$-(A+B)$
L_2	is the line through	$A+B$,	C,	$-((A+B)+C)$
L_3	is the line through	$B+C$,	$\mathcal{O}$,	$-(B+C)$
N_1	is the line through	$A+B$,	B,	$-(A+B)$
N_2	is the line through	B,	C,	$-(B+C)$
N_3	is the line through	A,	$B+C$,	$-(A+(B+C)$

We can draw a picture to represent all the above information. We also label a point D where L_2 intersects N_3.

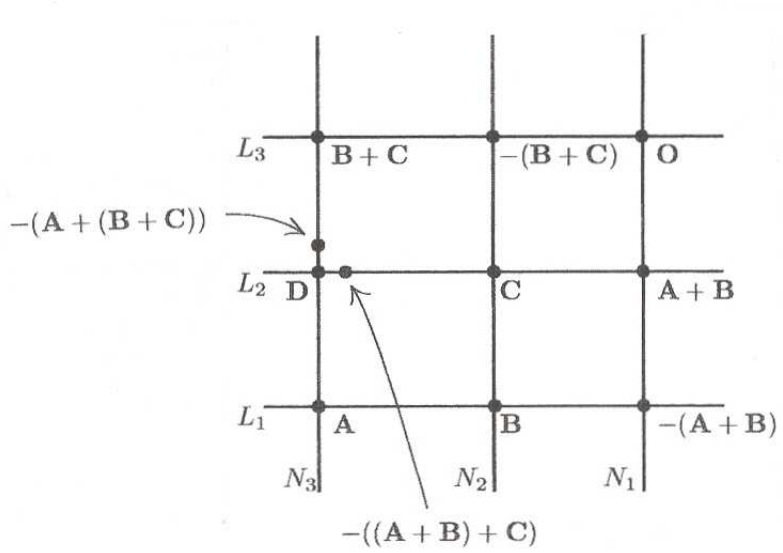

Fig. 8.6

Remember that our elliptic curve is in the background, also passing through these points:

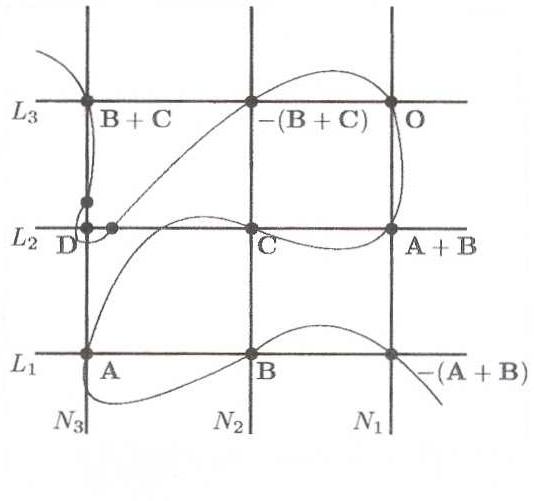

Fig. 8.7

Note that these pictures should not be taken as accurate drawings, but as a reminder of which lines pass through which points. Also note that we don't know E passes through D. This is what we want to show.

By the definitions of L_2 and N_3, we know that $-((A+B)+C)$ lies on L_2 and $-(A+(B+C))$ lies on N_3. But we'd like these points to be equal. In fact we will prove they are both equal to D, which is $L_2 \cap N_3$.

Now we have the following two cubic curves

$$L_1L_2L_3 = 0 \quad \text{and} \quad N_1N_2N_3 = 0.$$

We know by construction that these both pass through the eight points

$$\begin{array}{c} \mathcal{O}, \quad A, \quad B, \quad C, \\ A+B, \quad B+C, \quad -(A+B), \quad -(B+C). \end{array}$$

By Bezout's theorem we know that two cubics intersect in 9 points and we call the ninth point D, so since E goes through these 8 points, lemma 8.9 implies it also passes through D. Now on $N_1N_2N_3 \cap E$ we have the points

$$\begin{array}{c} \mathcal{O}, \quad A, \quad B, \quad C, \\ A+B, \quad B+C, \quad -(A+B), \quad -(B+C), \\ -(A+(B+C)), \quad D. \end{array}$$

But since there are only 9 points on the intersection of two cubics, two of these must be equal. By definition the first 8 are different and D is not equal to any of these, so we have $D = -(A + (B + C))$.

Similarly, by considering the 10 labelled points on $L_1L_2L_3 \cap E$, we find $D = -((A + B) + C)$.

So we have

$$-(A + (B + C)) = D = -((A + B) + C).$$

This completes the proof of associativity. □

Since we showed in remark 8.8 that the addition is well-defined, commutative and that we have an identity element as well as additive inverses, our completion of the proof of associativity gives us the following theorem.

Theorem 8.11 *The group law from definition 8.7 defines a commutative group on an elliptic curve.*

8.3 Torsion

In this section we will study a special part of the group on the elliptic curve, namely the part consisting of elements of finite order.

Definition 8.12 *Torsion points* are points of finite order on an elliptic curve. We say that a point has order m if

$$mP = \underbrace{P + P + \ldots + P}_{m \text{ times}} = \mathcal{O}.$$

Let E be an elliptic curve over a field of characteristic 0 of the form

$$E : y^2 = f(x) = x^3 + ax + b, \tag{57}$$

together with a point $\mathcal{O}$ at infinity.

2-Torsion: The points of order two are easily found, because if $P + P = \mathcal{O}$, then $P = -P$, so because $-(x, y) = (x, -y)$, these are the points with $y = 0$ and x-coordinate a root of f. Geometrically this means that the tangent line to points on the x-axis intersect E at $\mathcal{O}$.

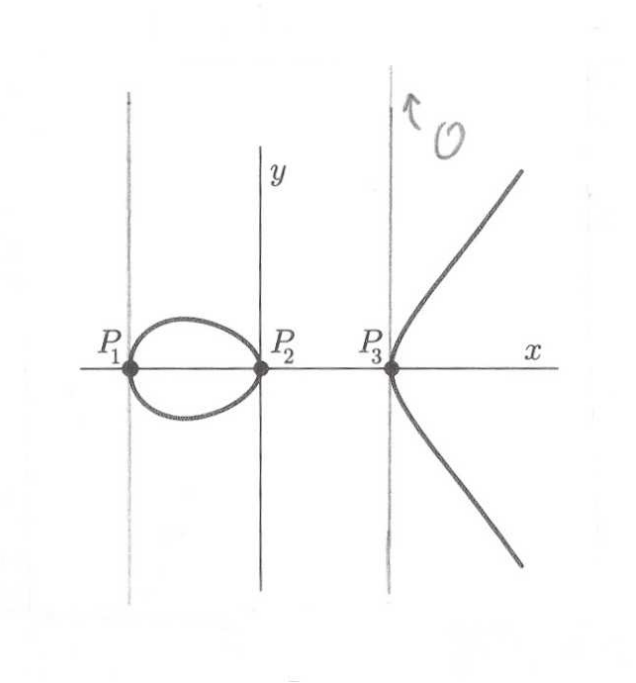

Fig. 8.8

If we allow complex coordinates, f has exactly three roots, since E is by definition non-singular. These three points of order two, together with $\mathcal{O}$ form a group of four elements. Since there exists only one non-cyclic group of order four, it must be isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. We call this group the *2-torsion* and denote it by $E[2]$.

3-Torsion: Next we try to find the points $P = (x, y)$ of order three, which satisfy $3P = \mathcal{O}$ or equivalently $2P = -P$. If we write $x(P)$ for the x-coordinate of P, we see that we have to solve $x(2P) = x(-P) = x(P) = x$. We can express $x(2P)$ in terms of x by using the duplication formula from the previous paragraph:

$$x(2P) = \frac{x^4 - 2ax^2 - 8bx + a^2}{4x^3 + 4ax + 4b} = \frac{x^4 - 2ax^2 - 8bx + a^2}{4y^2}. \tag{58}$$

If we set this equal to x and cross multiply (this is allowed since $y \neq 0$ for 3-torsion points), we get a polynomial $\psi_3(x)$ of which the roots correspond to the x-coordinates of the 3-torsion points:

$$\psi_3(x) = 3x^4 + 6ax^2 + 12bx - a^2.$$

We want to prove that all roots of ψ_3 are different. For that reason, we will now deduce an alternative expression for ψ_3. It follows directly from formula (56)

that

$$x(2P) = \frac{f'(x)^2}{4y^2} - 2x$$

By setting this equal to x, multiplying by $-4y^2 = -4f(x)$ (which is allowed since $y \neq 0$) and substituting $6x = f''(x)$, we see that

$$\psi_3(x) = 2f(x)f''(x) - f'(x)^2.$$

It now follows that a common root to $\psi_3(x)$ and $\psi_3'(x) = 12f(x)$ would be a common root to f and f', which is impossible, because E is non-singular.

We saw that ψ_3 has four different roots. Because none of these lie on the x-axis, every solution for x gives two points on E, namely $P = (x, y)$ and $-P = (x, -y)$. Together with $\mathcal{O}$, we see that the 3-torsion contains 9 points. Since there is only one (abelian) group with 9 elements such that every element has order dividing three, it must be $\mathbb{Z}_3 \times \mathbb{Z}_3$.

There is also a nice geometric way to describe the points of order three: they are inflection points, points where the tangent line to the cubic has a triple order contact. This follows from $2P = -P$. Geometrically this means that when we draw the tangent at point P, then take the third intersection and connect it with $\mathcal{O}$, we get $-P$. This is only the case if the third intersection point is the same point P.

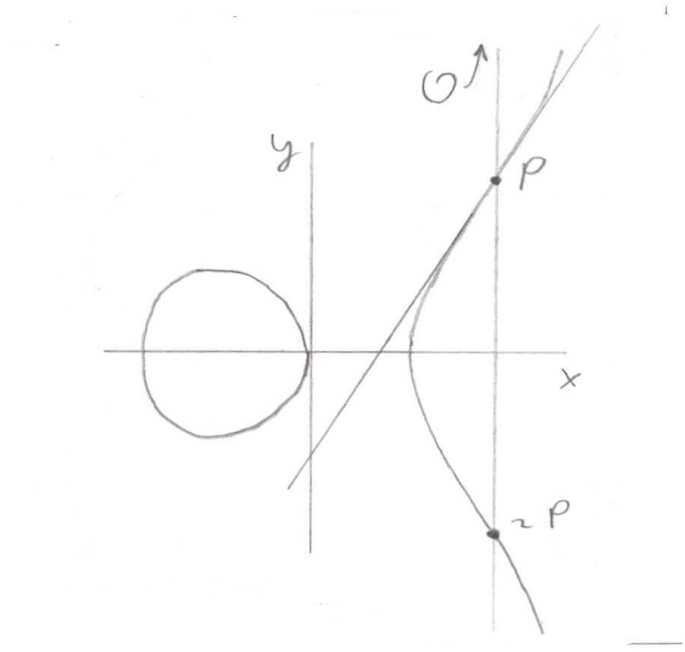

Fig. 8.9

4-Torsion: Before we start doing the algebra of 4-torsion points, we will look at their geometry. If P is a 4-torsion point, then $2P$ must be a 2-torsion point. This means that the tangent line to P must intersect E at a point on the x-axis.

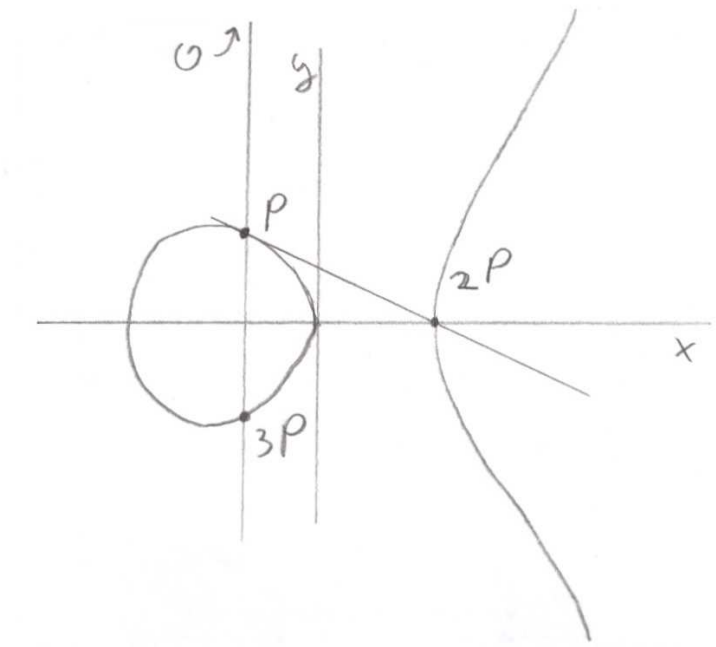

Fig. 8.10

Now lets look at the algebra. The x-coordinate of $2P$ is given by the duplication formula, so if we plug this formula into $f(x)$, it must equal zero. After multiplication by $(4y^2)^3$, we find that the x-coordinates of 4-torsion points, that are no 2-torsion points and neither the point at infinity, satisfy a 12-th degree polynomial in x. Using a symbolic calculator, we find that it is the square of the 6-th degree polynomial

$$\psi_4(x) = x^6 + 5ax^4 + 20bx^3 - 5ax^2 - 4abx - 8b^2 - a^3 = 0.$$

The 6 different roots of this polynomial yield 12 points on E, so together with the three 2 torsion points and $\mathcal{O}$, this gives an abelian group with 16 elements (12 of order 4, 3 of order 2 and the identity) so it follows that the 4-torsion is isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_4$.

5-Torsion: Now we come to our final goal: understanding the 5-torsion points on elliptic curves. Again we start with the geometry. The 5-torsion points must be points for which $3P = -2P$ and $4P = -P$ so $x(3P) = x(2P)$ and $x(4P) = x(P)$. Furthermore the tangent line to P must hit the curve at $-2P = 3P$ and the tangent line to $3P$ must hit the curve at $-6P = 4P = -P$.

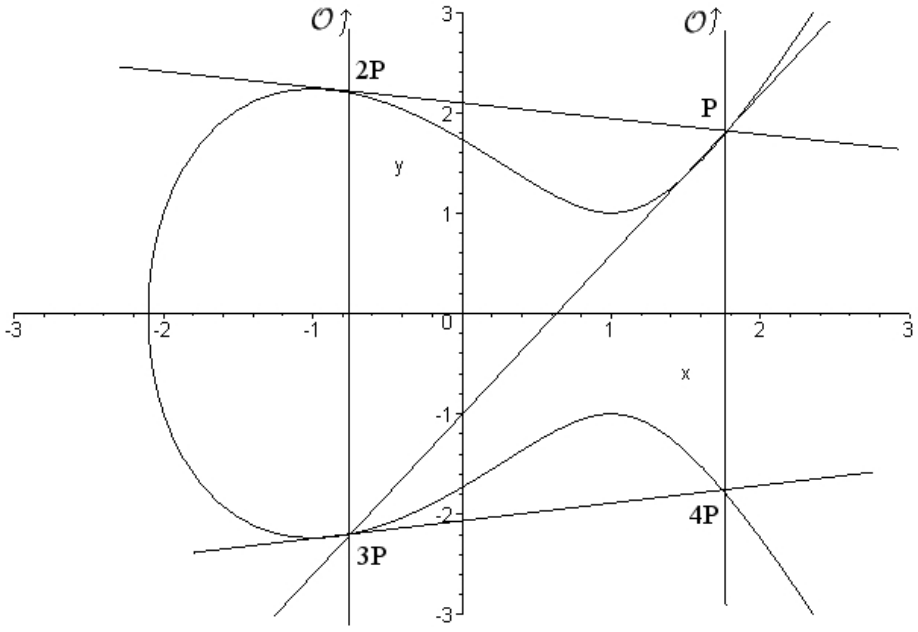

Fig. 8.11

The duplication formula gives the x-coordinate of $2P$ and $3P$, and if we duplicate these points again, the x-coordinate has to be equal to the x-coordinate of P and $-P$, which is x. So we have to solve $x(2(2P)) = x$. So to find the x-coordinates of the 5-torsion points, we can plug the duplication formula into the duplication formula and set it equal to x. This leads to a 16-th degree polynomial. Later in this paragraph, we will prove it has 12 distinct roots which satisfy

$$\begin{aligned}\psi_5(x) = 32(x^3 + ax + b)^2(x^6 + 5ax^4 + 20bx^3 - 5ax^2 - 4abx - 8b^2a^3) \\ -(3x^4 + 6ax^2 + 12bx - a^2)^3 = 0.\end{aligned} \tag{59}$$

Because we have 12 distinct x-coordinates for 5-torsion points, the 5-torsion contains 24 points of order 5 plus the point $\mathcal{O}$ at infinity, so it is an abelian group of order 25 with 24 points of order 5. From the classification of finite groups it follows that it must be isomorphic to $\mathbb{Z}_5 \times \mathbb{Z}_5$.

Remark 8.13 This result could also be obtained by using the fact that $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$, where $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ is a certain lattice. It turns out that the addition on the elliptic curve corresponds to ordinary addition in the complex plane modulo this lattice. This way it is easy to see that the n-torsion forms a group isomorphic to $\mathbb{Z}_n \times \mathbb{Z}_n$ (see figure 8.12). For more information see Silverman [1].

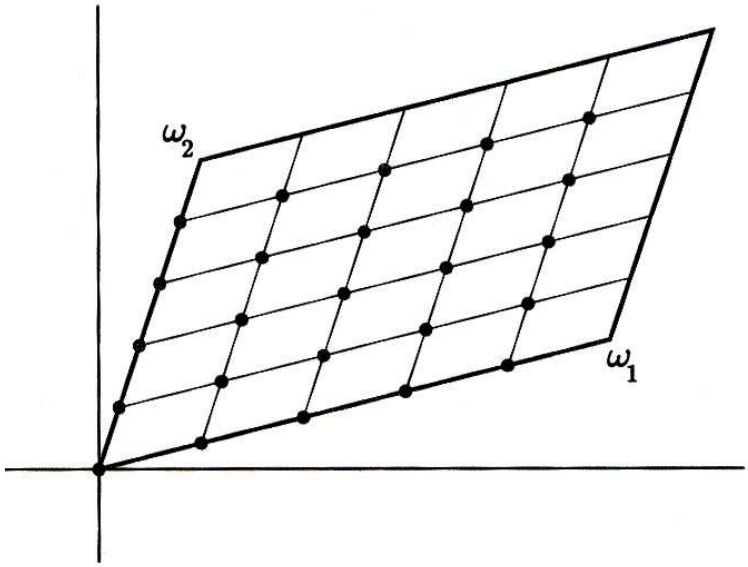

Fig. 8.12

n-Torsion: The process above can be continued and it is possible to derive formulas for the x-coordinates of n-torsion points from the duplication formula by induction. However there is a method that leads to a much faster algorithm and easier formulas. Moreover this method has the great advantage that the polynomials it generates have no multiple roots. The proof involves some more advanced methods from algebraic geometry and complex function theory, but it is understandable without to much extra theory.

Theorem 8.14 *Let* $mP = m(x, y) = (x_m, y_m)$, *then*

$$x_m = \frac{x\psi_m^2 - \psi_{m+1}\psi_{m-1}}{\psi_m^2}, \qquad y_m = \frac{\psi_{2m}}{2\psi_m^4},$$

where

$$\psi_0 = 0, \quad \psi_1 = 1, \quad \psi_2 = 2y$$
$$\psi_3 = 3x^4 + 6ax^2 + 12bx - a^2$$
$$\psi_4 = 4y(x^6 + 5ax^4 + 20bx^3 - 5ax^2 - 4abx - 8b^2 - a^3)$$

$$\psi_{2n+1} = \psi_n^3\psi_{n+2} - \psi_{n+1}^3\psi_{n-1}$$
$$y\psi_{2n} = \psi_n(\psi_{n-1}^2\psi_{n+2} - \psi_{n+1}^2\psi_{n-2})$$

Proof: We use the fact that a function is defined, up to a constant factor, by its zeros and poles. We determine the constant by looking at the behavior at

$x = \mathcal{O}$ using the *local uniformiser*

$$t = x/y.$$

The function ψ_m is defined by

- it has a simple zero at all m-torsion points $P \neq \mathcal{O}$
- it behaves like mt^{-m^2+1} at $\mathcal{O}$.

More precisely

- if m is odd, there are $1/2(m^2-1)$ pairs $(a_j, \pm b_j)$ of m-division pairs and
$$\psi_m = m\prod(x - a_j)$$
- if m is even, the three 2-torsion points are m-torsion points and there are $1/2(m^2-4)$ pairs $(a_j, \pm b_j)$, $b_j \neq 0$. Then
$$\psi_m = my\prod(x - a_j).$$

Now for all m, even or odd, we have

$$x_m \sim m^{-2}t^{-2}, \quad y_m \sim m^{-3}t^{-3}$$

at $\mathcal{O}$, and

$$\psi_m^2 x_m$$

has no poles except at $\mathcal{O}$.

Further, $x_m - x$ vanishes at P only if $(m+1)P = \mathcal{O}$ or if $(m-1)P = \mathcal{O}$. Therefore

$$x_m - x = \frac{\psi_{m+1}\psi_{m-1}}{\psi_m^2}, \tag{60}$$

where the constant is right, since both sides behave like $(m^2-1)/m^2t^2$ at $\mathcal{O}$. This gives the formula for x_m. That for y_m follows immediately from the specification of the poles and zeros.

It remains to give the recurrence relation. For integers l, m we have $x_l = x_m$ precisely when either $(l+m)P = \mathcal{O}$ or $(l-m)P = \mathcal{O}$. Therefore

$$x_l - x_m = \frac{\psi_{m+1}\psi_{m-1}}{\psi_l^2\psi_m^2},$$

in which we determined the constant by the behavior at $\mathcal{O}$. Furthermore we have

$$x_l - x_m = (x - x_m) - (x - x_l).$$

Therefore, by (60), we have

$$\psi_l^2\psi_{m+1}\psi_{m-1} - \psi_m^2\psi_{l+1}\psi_{l-1} = \psi_{m+l}\psi_{m-l}.$$

If we put $l = n$, $m = n+1$, we get $\psi_{m-l} = 1$ and

$$\psi_{2n+1} = \psi_n^3\psi_{n+2} - \psi_{n+1}^3\psi_{n-1}.$$

And if we put $l = n-1$, $m = n+1$, we get $\psi_{m-l} = \psi_2 = y$ and

$$y\psi_{2n} = \psi_n(\psi_{n-1}^2\psi_{n+2} - \psi_{n+1}^2\psi_{n-2})$$

□

Remark 8.15 It is now an easy task to derive ψ_5 and prove (59).

8.4 Isomorphism and the j-invariant

To what extent is an elliptic curve unique? To answer this question, we have to look at isomorphisms between elliptic curves.

Definition 8.16 Two elliptic curves E and E' in Weierstrass form are said to be isomorphic over K, if there exists a change of variables

$$x = f(x'), \qquad y = g(y')$$

with f, g rational functions, that transforms E into E'.

Theorem 8.17 *The change of variables in the definition above can be written as*

$$x = u^2x', \quad y = u^3y' \qquad \text{for some } u \in K - 0 \tag{61}$$

Proof: The proof can be found in [13] p. 49-50.

We will see in this paragraph that, up to isomorphism, elliptic curves can be characterized by a single parameter called the j-invariant, which is a function of the coefficients of the curve and is invariant under the change of variables (61). More precisely, we will prove that isomorphic elliptic curves have the same j-invariant and that elliptic curves with the same j-invariant are isomorphic[9]. Furthermore, we will prove there that for every $j_0 \in K$, there exists an elliptic curve with j-invariant j_0.

We again consider an elliptic curve E over K ($\text{char}(K) = 0$) of the form

$$E : y^2 = x^3 + ax + b \tag{62}$$

[9]the isomorphism involves taking fourth and sixth roots of the coefficients, so in fact they are isomorphic over $\bar{K}$, the algebraic closure of K.

together with a point $\mathcal{O}$ at infinity.

We associate to this curve the quantities

$$\Delta = \frac{a^3}{27} + \frac{b^2}{4}, \qquad j = \frac{(4a)^3}{\Delta}$$

and prove the following theorems.

Theorem 8.18 *Two elliptic curves are isomorphic over $\bar{K}$ if and only if they have the same j-invariant.*

Proof: Suppose we have two isomorphic elliptic curves E and E' of the form (62), then by definition there exists a change of variables

$$x = u^2 x', \quad y = u^3 y' \qquad \text{for some } u \neq 0$$

and then

$$u^4 a' = a, \qquad u^6 b' = b, \qquad u^{12}\Delta = \Delta'.$$

So we have

$$j = \frac{(4a)^3}{\Delta} = \frac{(4u^4 a')^3}{u^{12}\Delta'} = \frac{(4a')^3}{\Delta'} = j'.$$

On the other hand suppose we have two elliptic curves E and E' with the same j-invariant, say with equations

$$E : y^2 = x^3 + ax + b, \qquad E' : y'^2 = x'^3 + a'x' + b'.$$

Then

$$\frac{(4a)^3}{a^3/27 + b^2/4} = \frac{(4a')^3}{a'^3/27 + b'^2/4},$$

which yields

$$a^3 b'^2 = a'^3 b^2.$$

We look for an isomorphism of the form $(x, y) = (u^2 x', u^3 y')$, and consider three cases.

Case 1. $a = 0$ ($j = 0$). Then $b \neq 0$ (since $\Delta \neq 0$), so $a' = 0$, and we obtain an isomorphism using $u = (b/b')^{1/6}$.

Case 2. $b = 0$ ($j = 1728$). Then $a \neq 0$, so $b' = 0$, and we take $u = (a/a')^{1/4}$.

Case 3. $ab \neq 0$ ($j \neq 0, 1728$). Then $a'b' \neq 0$ (since if one of them is zero, then they both are, contradicting $\Delta' \neq 0$). Hence taking $u = (a/a')^{1/4} = (b/b')^{1/6}$ gives the desired isomorphism. □

Theorem 8.19 *If $j_0 \in \bar{K}$, then there exists an elliptic curve (defined over $K(j_0)$) with j-invariant equal to j_0.*

Proof: For $j_0 = 0$ we have the curve

$$E : y^2 = x^3 + 1, \qquad j = 0.$$

For $j_0 = 1728$ we have the curve

$$E : y^2 = x^3 + x, \qquad j = 1728.$$

Now assume that $j_0 \neq 0, 1728$ and consider the curve

$$E : y^2 = x^3 + \frac{3j_0}{1728 - j_0}x + \frac{2j_0}{1728 - j_0}. \tag{63}$$

An easy computation yields that $j = j_0$. □

8.5 Solution of the unsolvable quintic

There exists a nice analogy between the methods for solving the equations of degree up to four and the method for solving the quintic.

The use of 5-torsion points on an elliptic curve to solve the quintic has a nice analogy with the use radicals of to solve equations of degree lower than 5. Using radicals means that we allow a solution to $x^n = a$ to appear in our solution. By using the roots of unity, or in other words, the solutions to $x^n = 1$, we then find all solutions to the equation. The analogy with solving the quintic is that instead of the "torsion equation" $x^n = a$ on $\mathbb{C}^*$ for $n < 5$, we need the torsion equation $nP = Q$ for $n = 5$ on an elliptic curve E. The parameter Q can be set to zero, but the parameters a, b for E depend on the equation.

In this paragraph, we will see that the splitting field of the general quintic equation, defined over a field $F \subseteq \mathbb{C}$ is contained in a field that can be obtained by adjoining to F some radicals and the x-coordinates of the 5-torsion of some appropriate elliptic curve. First of all, we will explain how the icosahedral equation (47) can be linked to a certain elliptic curve.

Remember from the previous chapter that, up to a square root, solving the Brioschi quintic with parameter B is equivalent to solving the icosahedral equation

$$\mathcal{I}(Z) = f^5 + BT^2 = 0.$$

Now if we define the invariant $j = H^3/f^5$, we get a similar equation, also of degree 60 and with the same roots

$$\mathcal{J}(Z) = -H^3 + jf^5 = 0, \tag{64}$$

with j as only parameter. The relation between B and j can be deduced from

the relation between the invariants of the icosahedron.

$$\begin{aligned} 1728f^5 &= H^3 + T^2 \\ 1728f^5 &= jf^5 - f^5/B \\ 1728 &= j - 1/B \qquad (\text{if } f \neq 0) \\ j &= 1728 + 1/B \end{aligned} \tag{65}$$

Now we let j be the j-invariant of an elliptic curve E_j. We know from theorem 8.18 that (62) is such a curve. Note that if $j = 0$ we have $H = 0$, if $j = 1728$, we have $T = 0$ and if $j = \infty$, we have $f = 0$. So in each case, we find that a special point of the icosahedron yields a solution to the form problem and thus the quintic is solvable by radicals. Suppose this is not the case. We can then write this curve as

$$E_j : y^2 = x^3 + ax + b, \quad a = \frac{3j}{1728 - j},\ b = \frac{2j}{1728 - j}. \tag{66}$$

We want to prove that finding the 5-torsion of this curve is equivalent to solving the general quintic (up to some harmless radicals). We first look at the geometry of the 5-torsion to obtain polynomial ψ_5^* of degree six.

Theorem 8.20 *Let E_j be the elliptic curve (66) mentioned above and let P be a 5-torsion point on this curve. The element $y := x(P) + x(2P)$ is a root of the 6-th degree polynomial*

$$\psi_5^*(x) = x^6 + 20ax^4 + 160bx^3 - 80a^2x^2 - 128abx - 80b^2. \tag{67}$$

Proof: If we write out $y = x(P) + x(2P)$ with the duplication formula, we get

$$y = x + \frac{x^4 - 2ax^2 - 8bx + a^2}{4x^3 + 4ax + 4b}$$

Furthermore we have

$$\begin{aligned} \psi_5(x) = 32(x^3 + ax + b)^2(x^6 + 5ax^4 + 20bx^3 - 5ax^2 - 4abx - 8b^2a^3) \\ -(3x^4 + 6ax^2 + 12bx - a^2)^3 = 0. \end{aligned}$$

Now if we eliminate x from these equations with Mathematica, and factorize the result we get

$$83886080(4a^3 + 27b^2)^6(80b^2 + 128aby + 80a^2y^2 - 160by^3 - 20ay^4 - y^6)^2.$$

□

Now let

$$q = x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$$

be a general quintic over F with algebraically independent coefficients, so that $\mathrm{Gal}(q/F) \cong S_5$. If we replace F by $F' = F(\sqrt{\mathrm{disc}(q)})$, then $\mathrm{Gal}(q/F') \cong A_5$,

which is a *simple* group, hence not solvable. This is depicted in the following diagram:

$$\begin{array}{cc} & F(q) \\ A_5 & \vert \\ & F' = F(\sqrt{\mathrm{disc}(q)}) \\ \mathbb{Z}_2 & \vert \\ & F \end{array}$$

Let q_p, q_B and $\mathcal{J}$ be resp. the principal quintic, Brioschi quintic and icosahedral equation corresponding to q.

In previous paragraphs we proved the following things:

- After adjoining a square root $\sqrt{k}$, solving $q = 0$ is equivalent to solving $q_p = 0$.
- After adjoining another square root $\sqrt{l}$, solving $q_p = 0$ is equivalent to solving $q_B = 0$.
- Splitting q_B is equivalent to the form problem.
- After adjoining another square root $\sqrt{m}$, the form problem is equivalent to solving $\mathcal{J} = 0$.

So lets work over the field $N := F(\sqrt{\mathrm{d}isc(q)}, \sqrt{k}, \sqrt{l}, \sqrt{m})$. We will clarify this with the following diagram:

$$\begin{array}{ccccc} & & N(q) & & A_5 \\ & / & \vert & \backslash & \\ & F'(q)- & L & -N & \\ & \backslash & \vert & / & \\ A_5 & & F' & & \end{array}$$

We see that $N(q) = N(\mathcal{J}) = N(q_B)$ all have Galois group A_5 over N because $F'(q) \cap N = F'$. This follows from the fact that if $F'(q) \cap N = L$ for some intermediate field L, then L/F' would be Galois (multi-quadratic) and $\mathrm{Gal}(F'(q)/L) \lhd A_5$, but since A_5 is simple it has no normal subgroups other than $\{e\}$ and A_5, so L must be equal to F'.

Now for a given B, we look at E_j (with $j = 1728 + 1/B$) and ψ_5^* for this E_j. Let

$$\alpha = x(P) + x(2P)$$

for a certain $P \in E_j[5]$ be a root of ψ_5^*. If we can prove that $\alpha \in N(\mathcal{J})$ and $\alpha \notin N$, we will have proved that $N(q) = N(\psi_5^*)$ and thereby that solving the 5-torsion equation is equivalent to solving the quintic (up to radicals).

Lemma 8.21 *We have an explicit expression of α in terms of Z, so*

$$\alpha \in N(\mathcal{J})$$

Proof: We define the function

$$r(Z) = \frac{-125Z^6}{f} = \frac{-125Z^5}{Z^{10} + 11Z^5 - 1} \tag{68}$$

where the variable Z, as usual, is a solution to the icosahedral equation. It is constructed in such a way that it is invariant under the icosahedral rotations $R(Z) = \zeta_5 Z$ and $F(Z) = -1/Z$, so it is invariant under the group D_5, which is generated by R and F. Therefore, it satisfies a polynomial of degree 6. In fact, by calculating its coefficients with the same method we used for ψ_5^*, we find that

$$(r^2 + 10r + 5)^3 = jr. \tag{69}$$

Furthermore, we can verify computationally that

$$\alpha = -2\frac{r^2 + 10r + 5}{r^2 + 4r - 1}$$

by checking that it is indeed a root of ψ_5^*. This gives the desired expression of α in terms of Z. □

Lemma 8.22

$$\alpha \notin N$$

Proof: $\mathrm{Gal}(N(q)/N) = A_5$, but the action of this Galois group on α is non-trivial, since $S(\alpha) \neq \alpha$. Therefore $\alpha \notin N$. □

Theorem 8.23 *Suppose q is a general quintic over F and suppose we have associated an elliptic curve to it by calculating the j-invariant from the coefficients of q. There exists a radical extension N of F such that $N(q) = N(\psi_5^*)$.*

Proof: Set $N = F(\sqrt{disc(q)}, \sqrt{k}, \sqrt{l}, \sqrt{m})$ as before. Since the Galois group of $N(q)$ over N is A_5 and thus a simple group, we know that the normal closure of $N(\alpha)$ cannot lie properly between N and $N(q)$ (see the diagram above). Now since the two lemmas imply $N \neq N(\alpha) \subset N(q)$, we must have $N(q) = N(\psi_5^*)$. □

We have now proved our final goal, namely that solving the general quintic is equivalent (up to some harmless radicals) to finding the 5-torsion of an elliptic curve.

After this final result, I feel compelled to make some remarks about the beauty of this theory as a whole. The various forms of the quintic, the icosahedron and the 5-torsion on an elliptic curve, are mathematical objects that seem unrelated

at first sight, but on closer inspection can be seen to have very deep and beautiful connections. It is almost like a symphony in which all the instruments play a complex melody of their own and simultaneously interact with all of the other instruments in perfect harmony.

References

[1] Armstrong, M. A., *Groups and Symmetry*, Springer-Verlag, New York, 1988.

[2] Cassels, J.W.S. *Lectures on elliptic curves*, London Mathematical Society Student Texts 24, 1991.

[3] Cornelissen, Gunther *Galois theorie*, Universiteit Utrecht, 2004 (collegedictaat)

[4] Dickson, Leonard, *Algebraic Theories*, Dover Publications, 1959 (unaltered reprint of *Modern Algebraic Theories*, 1926).

[5] Dummit, D.S., Solving Solvable equations, Mathematics of Computation, Volume 57, number 195, pages 387-401, July 1991.

[6] Goins, E. H., *Elliptic Curves and Icosahedral Galois Representations*, Ph.D. thesis, Stanford University, 1999.

[7] Goins, E. H., *Icosahedral* $\mathbb{Q}$*-curve extensions*, Math. Research Letters pp. 205-217, March-May 2003.

[8] King, R. Bruce, *Beyond the Quartic Equation*, Birkhäuser, Boston, 1996.

[9] Klute, Annette, *Icosahedral Galois extensions and elliptic curves*, Manuscripta Math., 93(3):301-324, 1997.

[10] Lang, Serge *Undergraduate Algebra, Second Edition*, Springer-Verlag, New York, 1987-1990.

[11] O'Connor, J.J. & Robertson, E.F., http://www-history.mcs.st-andrews.ac.uk, University of St Andrews, Scotland 1996-2005.

[12] Serre, J.-P. *Extensions icosaédriques. In: Oeuvres III, pages 550-554.* Springer-Verlag, 1986 (the letter was written in 1979).

[13] Silverman, Joseph H.*The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986.

[14] Silverman, Joseph H. & Tate, John, *Rational Points on Elliptic Curves*, Springer-Verlag, New York, 1992.

[15] Tignol, Jean-Pierre, *Galois' theory of algebraic equations*, Institut de Mathématique Pure et Appliqueé, UCL, Louvain-la-Neuve, Belgium 1988 (translation by the author of the French version from 1980).