Galois Theory

Richard Koch

September 12, 2021

Contents

Preface

The beginnings of algebra, and the discovery of the quadratic formula, are hidden in the mists of time. At first, algebra was written entirely with words: “the thing plus one equals two.” This “rhetorical algebra” was created in Babylonia and lasted until the early fifteenth century. Diophantus may have been the first person to use symbols for some of these words, in a book named *Arithmetica* written around 200 AD. Algebra was extensively developed by Arabic mathematicians starting around 700 AD, eventually giving the subject its name. The modern symbolic approach begins to appear in full dress in the works of Francois Viete toward the end of the 1500’s and Rene Descartes’ *La Geometrie* of 1637.

Signs of quadratic equations appear early in this long development, but it is difficult to pick a particular moment that the quadratic formula appears in the precise form we use today. Perhaps it is better to say that it has been a part of the subject for many centuries.

Cubic equations appear very early in this history, even in Babylonian times. By the 1500’s, with the quadratic formula in its modern form, mathematicians began searching for an analogous cubic formula. The first glimpse of the formula was seen by Scipione del Ferro around 1500, but he did not publish the result, which was found in his mathematical papers after his death. The formula was rediscovered by Niccol Tartaglia in 1530, and used by him in a mathematical contest against del Ferro’s son-in-law, who had discovered the formula in del Ferro’s papers. Later Tartaglia revealed the formula to Gerolamo Cardano, on condition that he not reveal the formula until Tartaglia published it. Cardano generally kept his word, but he did reveal the formula to an apprentice named Lodovico Ferrari, who used it to produce a formula for solving quartic equations. This put Cardano in a bind, since Tartaglia did not publish and Ferrari could not show his solution without revealing the cubic formula. Eventually Cardano recalled the ancient contest of Tartaglia and del Ferro’s son-in-law, and visited del Ferro’s widow, who turned out to have kept her husband room intact since his death. There Cardano found del Ferro’s solution, and published it while claiming not to break his oath to Tartaglia. All of this is covered in much greater detail in the wonderful book *Cardano, the Gambling Scholar* by Oystein Ore.

The quadratic formula can be obtained by introducing a new variable $y = x + \lambda$ and writing the quadratic in terms of y rather than x. Geometrically, this amounts to translating the graph of the quadratic by λ. If λ is chosen appropriately, the linear term in the equation for y cancels out and we obtain $y^2 + A = 0$, which is easily solved.

The same trick words for cubic equations, reducing an arbitrary cubic to the form

$$x^3 + Ax + B = 0$$

The cubic formula states that x is then equal to

$$x = \sqrt[3]{-\frac{B}{2} + \sqrt{\frac{B^2}{4} + \frac{A^3}{27}}} + \sqrt[3]{-\frac{B}{2} - \sqrt{\frac{B^2}{4} + \frac{A^3}{27}}}$$

This formula has some unexpected properties. If A and B are real, the cubic either has one real root or three real roots, except for trivial edge cases. If it has one real root, the expression under the square root is positive and then the formula gives the real root since any real number has a real cube root. But if there are three real roots, the expression under the square root is negative. This suggests that a separate formula may be required to handle the case of three real roots.

In 1572, Bombelli published *L'Algebra*. In this book he advanced algebraic notation and added many examples from Diophantus after becoming one of the first modern Europeans to rediscover this book. But Bombelli is mainly remembered today for a specific example in his book, the solution of the cubic equation $x^3 - 15x - 4 = 0$. He obtains $x = 4$, which can be easily checked. It is his method which is significant, for he obtains this root from the cubic formula. According to the formula,

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = \sqrt[3]{2 + 11i} + \sqrt[3]{2 - 11i}$$

Bombelli then notices that $(2 + i)^3 = 2 + 11i$ and $(2 - i)^3 = 2 - 11i$ and thus

$$x = (2 + i) + (2 - i) = 4$$

Before this book appeared, complex numbers were not taken seriously, and it is their use in the cubic formula which gave mathematicians the first glimpse of their importance. Later mathematicians computed complex cube roots using trigonometry and then the cubic formula could solve all cubic equations.

Incidentally, we will later prove that cubic equations with three real roots cannot be solved using only real radicals, so the cubic formula is the only game in town.

After Ferrari found a quartic formula, mathematicians tried to find formulas for higher degree equations. In the late 1700's, this work was taken on by Lagrange (who also created the metric system during the French Revolution). Lagrange created various linear combinations of the roots, called resolvents, and found that each satisfied a polynomial equation called its resolvent equation. Finding the resolvent by solving this equation often led to a complete solution of the original equation. In the quadratic, cubic, and quartic cases, resolvents exist whose resolvent equations have lower degree than the original equation, so there is a general inductive procedure to solve these equations. However, Lagrange was unable to find a general resolvent of lower degree for equations of order five. This led some mathematicians to conjecture that no formula exists for degrees five and higher. Gauss expresses this opinion in his book on number theory.

The first person to prove that fifth degree polynomials cannot be solved using radicals was Paulo Ruffini, in 1799. His paper was complicated and difficult to read, and few mathematicians were convinced. In more recent times his paper has been carefully studied; his proof had one significant gap but the essential ideas were there.

Early in his career, Niels Henrik Abel proposed a general method to solve fifth degree equations, and submitted a paper to that effect. The referee asked that he add a specific numerical example. While trying to construct that example, Abel found a mistake in his paper. In 1824, Abel published his first significant paper, "Memoir on algebraic equations, in which the impossibility of solving the general equation of the fifth degree is proved." Abel acknowledges Ruffini in this paper and then gives the first complete proof of impossibility. He published a more detailed proof in 1826 in *Crelle's Journal*. Nowadays, the theorem that fifth degree equations cannot be solved by radicals is usually attributed to Abel-Ruffini.

As Abel pointed out, the Abel-Ruffini argument only proves that there is no formula which solves all fifth degree polynomials. It might still be possible that the roots of any specific polynomial can be written as expressions that only involve radicals. Abel was working on this more specific problem when he died at age 29.

Ebvariste Galois found a different approach to these problems, which was capable of working for specific individual polynomials. In particular, his method shows that no solution of $x^5 - x - 1$ can be written in a form composed only of radicals.

Ebvariste Galois, born in 1811, became interested in mathematics when he was 14, and carefully read Lagrange's work on solving polynomial equations. In 1829, Galois submitted two papers on polynomial equations to the Academy of Sciences, where they were rejected by Augustin-Louis Cauchy. This rejection is very controversial, and many authors describe it as a case of the old guard refusing to accept new ideas from the young. But the Wikipedia article on Galois suggests that instead Cauchy recognized the importance of Galois' work and suggested combining the papers into one and submitting it for the Academy's Grand Prize.

At the time, France was in great political turmoil. Galois' father was the mayor of a small French town, and after a dispute with the village priest, he committed suicide on July 28, 1829.

Galois resubmitted his papers to the Academy in Febuary of 1830, and this time they were read by Fourier. But Fourier soon died and the papers were lost. Galois published three other papers in 1830, but in 1831 political events overwhelmed him and he was thrown in jail. In January of 1831, he again submitted his papers to the Academy, and this time they were read by Poisson. Poisson declared the papers "incomprehensible" but suggested that Galois publish all of his work on equations so the world could form a definitive opinion. Galois grew very angry when hearing of this report, but later was in the process of collecting his manuscripts while still in prison. He was released on April 29, 1832. On May 30, 1832, he was killed in a duel. The motivation behind this duel is controversial and unclear.

In 1846, Galois' paper "Memoir on conditions to solve an equation with radicals" was published by Liouville, who added some extra remarks praising the work, but completely missed the group theory at the core of Galois' arguments. Joseph Serret attended these Liouville talks, and included Galois' theory in an 1866 textbook. Credit for deciphering and modernizing the theory is generally given to Serret's pupil Camille Jordan, who described the theory in his 1870 book "Treatise on permutations and algebraic equations." Galois theory becomes a central topic in modern algebra only in the beginning of the twentieth century.

These remarks on Galois are taken from various Wikipedia articles. There are many books on Galois, and often his story is told in a completely different way, suggesting that he was betrayed by the mathematical elite in France. But it is astonishing how many great mathematicians interacted with Galois in his short life: Cauchy, Fourier, Poisson, Liouville, and others; none of them knew that he would soon die in a duel. If Galois had lived long enough to calm his political passions, it is easy to imagine that he would have joined this group as an equal.

It is time to turn from history to the actual theory created by these mathematicians. The modern approach to this theory as reformulated by Jordan and many others will be given in the following chapters, but here we'll give a glimpse of the central ideas in the language of the creators.

Although Galois theory works with great generality, let us assume that our polynomials have rational coefficients. If we like, we can multiply all coefficients by a common term and get integer coefficients, so we will deal with very explicit polynomials like

$$P(x) = x^5 - x - 1$$

$$P(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$$

In fact, these will be our two concrete examples, and we'll hint at reasons that the first cannot be solved by radicals but the second can be.

If $P(x)$ factors over Q as $P(x) = Q_1(x)Q_2(x)$, then it suffices to study the roots of Q_1 and Q_2 separately. So from now on, we will assume that all $P(x)$ are irreducible over Q.

In practice this creates difficulties because it can be hard to show that a polynomial is irreducible, but we leave that issue to the regular text.

While $P(x)$ is irreducible over Q, we can factor $P(x)$ over C and thus find its complex roots. So $P(x) = \prod(x-\theta_i)$ where the θ_i are complex and can be approximated numerically. These θ_i are not random complex numbers because several expressions involving them are rational numbers:

$$\theta_1 + \theta_2 + \theta_3 + \ldots + \theta_n$$

$$\theta_1\theta_2 + \theta_1\theta_3 + \ldots + \theta_{n-1}\theta_n$$

$$\ldots$$

$$\theta_1\theta_2 \ldots \theta_n$$

Indeed, when we multiply out, these are plus or minus the coefficients of our polynomial.

Next we ask if other expressions in the roots might be rational. For instance, what about

$$\theta_1^2 + \theta_2^2 + \ldots + \theta_n^2$$

$$\theta_1^3 + \theta_2^3 + \ldots + \theta_n^3$$

$$\ldots$$

The answer is given by "the fundamental theorem of symmetric polynomials," which we will later prove. A polynomial $Q(X_1, \ldots, X_n)$ is said to be *symmetric* if any permutation of the unknowns just reproduces the same polynomial. The fundamental theorem says that any symmetric polynomial Q with complex coefficients can be written uniquely as a polynomial R in the fundamental symmetric polynomials $(\sigma_1, \ldots, \sigma_n)$ where

$$\sigma_1(X_1, \ldots, X_n) = X_1 + \ldots + X_n$$

$$\sigma_2(X_1, \ldots, X_n) = X_1X_2 + X_1X_3 + \ldots + X_{n-1}X_n$$

$$\ldots$$

$$\sigma_n(X_1, \ldots, X_n) = X_1X_2 \ldots X_n$$

Moreover, if the original Q has rational or integer coefficients, so does R. It follows from this result that any symmetric polynomial evaluated at the roots of P is a rational number.

We now ask the key question. *Are there any other expressions in the roots which are rational?* This question is rather vague and we first make it more precise. Define

$$\widetilde{\mathcal{M}} = \{ \ T(X_1, \ldots, X_n) \mid T(\theta_1, \ldots, \theta_n) \in Q \ \}$$

where T is a polynomial with rational coefficients. This set contains all symmetric polynomials. Does it contain anything else?

The answer is rather awkward and shows that $\widetilde{\mathcal{M}}$ is not the correct set to study. Suppose S is a symmetric polynomial and suppose we multiply $S(X_1, \ldots, X_n)$ by X_1. The resulting polynomial is not symmetric. Does it belong to $\widetilde{\mathcal{M}}$? If $S(\theta_1, \ldots, \theta_n) \neq 0$, then clearly not because θ_1 isn't rational so its product by a non-zero rational cannot be rational. But if $S(\theta_1, \ldots, \theta_n) = 0$, then yes because θ_1 times zero is zero and thus rational.

Because of this problem, we define a slightly different set

$$\mathcal{M} = \{\ T(X_1, \ldots, X_n) \mid T(\theta_1, \ldots, \theta_n) = 0\ \}$$

where T is a polynomial with rational coefficients. If $S(X_1, \ldots, X_n)$ is any *symmetric* polynomial, then $S(\theta_1, \ldots, \theta_n)$ can be written as a polynomial with rational coefficients in the elementary symmetric functions and thus a polynomial in the coefficients of P and thus a specific rational number, and so

$$T(X_1, \ldots, X_n) = S(X_1, \ldots, X_n) - S(\theta_1, \ldots, \theta_n)$$

is in $\mathcal{M}$. We abbreviate this by saying that $\mathcal{M}$ "contains all symmetric polynomials." Moreover, $\mathcal{M}$ is now an ideal in the polynomial ring, so elements in $\mathcal{M}$ can be multiplied by any polynomial with rational coefficients. We then reformulate our question to ask if the symmetric polynomials generate this ideal. And if not, what other generators need be added.

It turns out that the answer to this question depends on the initial polynomial we are trying to solve by radicals. In degree five and higher and for a random $P(x)$, $\mathcal{M}$ is often generated by symmetric polynomials. But if a particular equation of high degree can be solved by radicals, then $\mathcal{M}$ will be larger and contain relations that are definitely not symmetric. We'll illustrate this with the polynomial $P(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$.

The roots of this polynomial can be expressed as radicals for the simple reason that

$$x^7 - 1 = (x-1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1)$$

and so the roots are all roots of unity. It isn't clear that

$$x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$$

is irreducible, but we will prove that it is. Indeed if p is prime, then

$$x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \ldots + x + 1)$$

and both factors are irreducible. This result isn't true in greater generality; for example

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

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

and so

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

Sticking with $X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$, let $\theta = e^{\frac{2\pi i}{7}}$. Then the roots are $\theta, \theta^2, \theta^3, \theta^4, \theta^5, \theta^6$. The final seventh root of unity is 1, which we have removed by factoring $X - 1$ out.

Notice that $X_2 - X_1^2 \in \mathcal{M}$ because the second root is the square of the first root. If we permute the X_i sending X_1 to X_2 and X_2 to X_3 and so forth, we convert $X_2 - X_1^2$ to $X_3 - X_2^2$, but this does not belong to $\mathcal{M}$ because the third root is θ^3, and it does not equal the square of the second root $(\theta^2)^2$. So this $\mathcal{M}$ contains elements, and surely generators, which are not symmetric. All symmetric polynomials are in the new $\mathcal{M}$, but now the set of rational relations is much larger. This is the rough sign that our equation is solvable by radicals.

Ruffini and Abel both studied objects like $\mathcal{M}$. Dealing with it is difficult because it contains an infinite number of polynomials. The great idea of Galois was to deal instead with the finite group $\mathcal{G}$ of permutations of the roots θ_i which preserve the set $\mathcal{M}$. This group is now called the *Galois group* of the polynomial $P(x)$. So to be precise, a permutation σ belongs to $\mathcal{G}$ if whenever $T(X_1, \ldots, T_n) \in \mathcal{M}$ then $T(X_{\sigma(1)}, \ldots, T_{\sigma(n)}) \in \mathcal{M}$.

If the symmetric polynomials generate all of $\mathcal{M}$, then $\mathcal{G}$ will be the full group of all permutations. Indeed, every element of $\mathcal{M}$ would then be a sum of terms of the form

$$T(X_1, \ldots, X_n)\Big[S(X_1, \ldots, X_n) - S(\theta_1, \ldots, \theta_n)\Big]$$

where T is an arbitrary polynomial with rational coefficients and S is a symmetric polynomial with rational coefficients. If we permute the X_i, the first T becomes a different polynomial. But it doesn't matter because S is symmetric and thus the second term is still zero. Thus our permutation belongs to $\mathcal{G}$.

But if $\mathcal{M}$ has additional generators which are not symmetric, then some permutations will not preserve these elements and the Galois group will be smaller. Note carefully the pattern here. If a polynomial can be solved by radicals, we expect its $\mathcal{M}$ to be larger than usual, and so its $\mathcal{G}$ to be smaller than usual.

We can confirm this in the special case of $X^6+X^5+X^4+X^3+X^2+X+1$. The equation $X^7-1=0$ has seven very symmetrical complex roots, shown on the left below. But when we remove the factor $X-1$, we remove one of these roots and obtain the unsymmetrical picture shown on the right. It looks like the resulting group might just be the identity. *However, the Galois group is not a group of ordinary symmetries in the complex plane. Instead, it is a group of "algebraic symmetries" that are not apparent in the picture of the roots.*

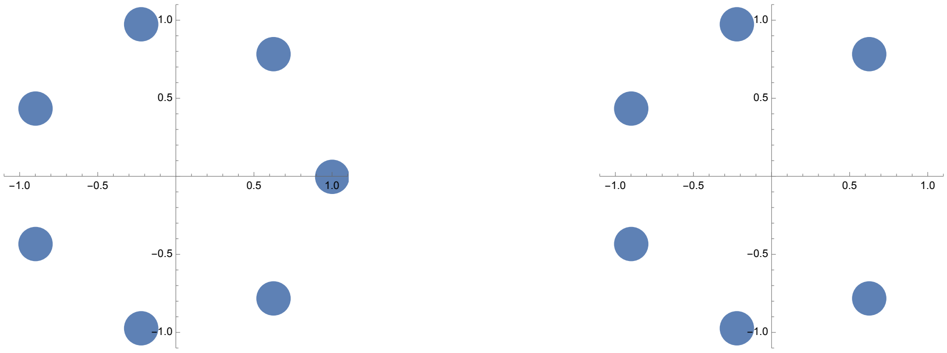


Figure 1: Seventh Roots of Unity

Roots of $X^6+X^5+X^4+X^3+X^2+X+1$

Our work will be motivated by a theorem we will prove later on: if $P(x)$ is irreducible, then the Galois group acts transitively on the roots. So the Galois permutation group in our case has permutations which carry θ to θ, to θ^2, to θ^3, to θ^4, to θ^5, and to θ^6. Once we know where θ goes, we can easily work out the complete permutation. For instance, if θ maps to θ^2, then θ^2 maps to θ^4, and θ^3 maps to θ^6 and θ^4 maps to $\theta^8=\theta$, and θ^5 maps to $\theta^{10}=\theta^3$ and θ^6 maps to $\theta^{12}=\theta^5$. A nicer way to write this is

$$\theta \to \theta^2 \to \theta^4 \to \theta$$

and

$$\theta^3 \to \theta^6 \to \theta^5 \to \theta^3$$

Even better, we can write this permutation in cycle notation:

$$(\theta, \theta^2, \theta^4)(\theta^3, \theta^6, \theta^5)$$

The permutation sending θ to θ^3 is even more interesting, for it has a single cycle:

$$(\theta, \theta^3, \theta^2, \theta^6, \theta^4.\theta^5)$$

So if we arrange the six roots around a wheel as below, then this permutation is one counterclockwise turn of the wheel. Since $\mathcal{G}$ is a group, all six turns of this wheel are in the group. And since these exhaust the possible images of θ, the complete Galois group must be the full rotational symmetries of this wheel, Z_6. Notice carefully that this placement of the roots is not at all the placement of the roots in the complex plane. The symmetry group Z_6 is a figment of our algebraic imagination, not a figment of our geometric imagination.

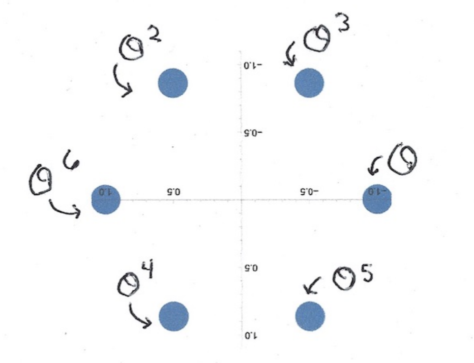


Figure 2: Galois Group of $X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$

Let us return to the general theory. The originators of our theory provided two tools, $\mathcal{M}$ and $\mathcal{G}$, which pick up subtle properties of the initial equation $F(x) = 0$. Amazingly, $\mathcal{G}$ completely determines whether our equation can be solved by radicals. We don't even have to understand how $\mathcal{G}$ acts as a group of permutations; it is enough to be given a list of elements of $\mathcal{G}$ and a multiplication table for these elements. This is the astonishing climax of Galois theory.

How does this determination work? If $\mathcal{H}$ is a subgroup of a group $\mathcal{G}$, we can form a set $\mathcal{G}/\mathcal{H}$, called *the quotient space.* There is a special kind of subgroup $\mathcal{H}$ called a *normal subgroup*, and when $\mathcal{H}$ is normal, the quotient space is also a group. So $\mathcal{G}$ is made from two smaller groups $\mathcal{H}$ and $\mathcal{G}/\mathcal{H}$. But some groups have no non-trivial normal subgroups. These groups are called *simple groups*; they are not made from smaller groups, and instead form the fundamental building blocks of group theory.

If $\mathcal{G}$ is an arbitrary finite group, we can find a chain of subgroups, each normal in the next element of the chain, leading from the identity to $\mathcal{G}$:

$$\{e\} = \mathcal{C}_0 \subset \mathcal{C}_1 \subset \ldots \subset \mathcal{C}_n = \mathcal{G}$$

To avoid trivialities, we require that each inclusion be strict, and that no inclusion can be expanding by adding an element in the middle. It then follows easily that the $\mathcal{C}_i/\mathcal{C}_{i-1}$ are simple groups; in a sense, they are the elementary pieces which create $\mathcal{G}$.

This "composition series" is not unique, but a fundamental theorem says that its length is determined by $\mathcal{G}$ and the simple quotients are determined by $\mathcal{G}$ up to order. So a crucial feature of any finite group is the list of simple groups which make it up.

What can be said about these simple groups? It is easy to determine all *abelian* simple groups; they are exactly Z_p for primes p. But the theory of non-abelian simple groups is much more difficult, The smallest such group is A_5, the group of even permutations on five letters, or equivalently the rotational symmetry group of the dodecahedron. This group has 60 elements.

Indeed, A_n is non-abelian and simple whenever $n \geq 5$. In the early twentieth century, a large number of matrix groups were shown to be simple. For example, $SL(n, F)$, the group of matrices with determinant one over a finite field F, modulo its center, is simple with a small list of exceptions. Eventually, toward the end of the century, a complete list of finite simple groups was obtained, with a proof requiring thousands of pages and dozens of authors; most are alternating groups or matrix groups, but there are 23 *sporadic groups* which appear out of thin air.

The Galois group of a polynomial $P(x)$ of degree n has order between n and $n!$ For a given n, this restricts the possibilities to a finite set of groups. In the quadratic case, $\mathcal{G} = Z_2$. In the cubic case, $\mathcal{G}$ is either Z_3 or D_3, where D_3 is the dihedral group of symmetries of an equilateral triangle. Notice that D_3 is not abelian.

It turns out that $F(x)$ can be solved by radicals whenever the Galois group is abelian. The D_3 example shows that this condition is not necessary, and the final beautiful result, says

Theorem 1 *An irreducible polynomial equation $F(x) = 0$ with rational coefficients can be solved by radicals if and only if all composition quotients of its Galois group are abelian, and thus of the form Z_p.*

A finite group with this property is called a *solvable group* for obvious reasons. The study of such groups is important in group theory even if the goal has nothing to do with Galois theory and solving equations.

In the evening of May 29, 1832, Galois wrote a letter to his friend Auguste Chevallier. A translation of this letter into English can be found at https://www.ias.ac.in/article/fulltext/reso/004/10/0093-0100. Near the beginning of this letter, Galois defines what it means for a subgroup of a group to be normal. Here is the beginning of the letter:

> My dear friend,
>
> I have studied several new ideas. In the theory of equations, I have determined which cases the equations are solvable by radicals, which has provided me with an opportunity to go into this theory in depth and describe all possible transformations on an equation, even when it is not solvable by radicals.
>
> All this can be put in three papers. The first one is written, and, in spite of what Poisson has said, I stand by it, with the corrections that I have indicated. The second contains rather interesting applications from the theory of equations. Here is a summary of the most important ones:
>
> 1. According to the propositions II and III of the first paper, one sees a great difference between adjoining, to an equation, one of the roots or all the roots of an auxiliary equation. In both the cases, the group of the equation can be partitioned by adjunction into groups such that one can pass from one to another by a self-transformation; but the condition that these groups have the same substitutions holds only in the second case. This is called proper decomposition.
>
> In other words, when a group G contains another, H, the group G can be partioned into groups each of which is obtained by operating on the permutations in H a self-transformation, in such a way that,
>
> $$G = H + HS + HS' + ..,$$
>
> And we can also partition into groups which have all similar substitutions, such that
>
> $$G = H + TH + T'H + ...$$
>
> These two types of decompositions generally do not coincide. When they do coincide, the decomposition is said to be proper. It is easy to see that, when the group of an equation is not susceptible to any proper decomposition, however well we might have transformed this equation, the groups of the transformed equations will always have the same number of permutations. On the contrary, when the group of an equation is susceptible to a proper decomposition in such a way that we can decompose it into M groups of N permutations, we can resolve the given equation by means of two equations: one will have a group of M permutations and the other, one of N permutations.
>
> Hence when we would have exhausted all possible proper decompositions on the group of an equation, we arrive at groups which can be transformed but for which the number of permutations will always be the same.

> If each of these groups has a prime number of permutations then the equation will be solvable by radicals; otherwise, not.
>
> The smallest number of permutations that an indecomposable group can have, when this number is not a prime number, is 5 4·3.

The letter continues with additional ideas, some about Galois theory and others about modular functions and elliptic function theory. It ends with

> You know, dear Auguste, that these subjects are not the only ones that I have explored. My principal meditations, for some time now, were directed to the application of the theory of ambiguity to transcendental analysis. But I don't have the time, and my ideas are not yet well developed in this area, which is immense.
>
> You will get this Letter printed in the *Revue encyclopedique.*
>
> I have often dared in my life to advance propositions about which I was not sure, but all that I have written down has been in my mind for over an year, and it would not be too much in my interest to make mistakes so that one suspects me of having announced theorems of which I would not have a complete proof.
>
> You make a public request to Jacobi and Gauss to give their opinion, not as to the truth but as to the importance of these theorems.
>
> After this, I hope there will be people who will profit by deciphering all this mess. I embrace you effusively.

The next day, Galois was killed in the duel. In the letter, he was clearly thinking of that duel. But he was also looking far into the future. The beginning remarks about groups and partitioning by $\mathcal{H}$ define normal subgroups for the first time. But even the definition of an abstract group is still far off, to be first given by Cayley in 1854.

We are now ready for the precise modern theory, but a word of warning. Although the ideas just sketched will play a crucial role for us, they will appear "in disguise" and may not at first be recognized.

We will start just as we did in the preface by fixing an irreducible $P(x)$ with rational coordinates. We will consider one of its complex roots θ and form the smallest field of complex numbers containing all rational numbers and this additional root. This field is called the *root field* $K(\theta)$. It is a vector space over Q and turns out to have dimension equal to the degree of P. We will obtain a very clear and precise understanding of this field.

Next we will form a second field by considering all of the complex roots $\theta_1, \dots, \theta_n$ and forming the smallest field of complex numbers containing all rational numbers and these additional θ_i. This field is called the *splitting field* $K(\theta_1, \dots, \theta_n)$. But annoyingly, we will obtain a much fuzzier view of this field. The field itself is precisely defined, but all sorts of obvious questions are hard to answer. For example, this field is also a vector space over Q, but if n is the degree of $F(x)$ then we will only be able to prove that $n \leq \dim K \leq n!$ So even if F only has degree 5, the dimension of this field is somewhere between 5 and 120. Quite a range of ignorance!

However, there is a reason for this difficulty, since studying the splitting field is just another way of studying $\mathcal{M}$. Indeed, recall that $\mathcal{M}$ is an ideal in the polynomial ring $Q[X_1, \ldots, X_n]$. It turns out that the splitting field is the quotient of the polynomial ring modulo $\mathcal{M}$. Thus $\mathcal{M}$ will appear in these notes in a very sneaky way.

Similarly, the Galois group will be defined to be the group of all automorphisms of the splitting field. An automorphism of K must fix all rational numbers and thus must permute the roots of $F(x) = 0$, and in this way we recover the description of the Galois group in this preface.

Confession When I completed a course on Galois theory in college, there were many loose ends. An extension $K \subset L$ was defined to be Galois if it was normal and separable; this sounded like a theorem, not a definition. I learned Abel's theorem that no general formula exists to solve fifth degree equations, but did not study specific polynomials like $x^5 - x - 1$. I could prove that angles cannot be trisected and circles cannot be squared, but with a gap because I could not prove that π is transcendental. Later, I found that more recent graduate students face the same problems. One of them, describing the general idea of the theory, told another that the Galois group contains *all* permutations of the roots — a pretty severe misunderstanding! Others imagined that the Galois group of $z^7 - 1$ is cyclic because — after all — the roots of unity form a cyclic picture in the plane.

So after I retired, I wrote these notes to straighten myself out. Whenever I got confused, I asked Google for help. It is amazing how many lecture courses are on the internet. I'd often try to puzzle out a result by reading one set of lectures, not understand a word, and then come across a second author's lectures with a clear presentation of the puzzling point.

I didn't write any exercises, and I didn't keep track of any references. Every beautiful formula here comes from someone else. My apologies!

Several years after writing the notes, I came across a book on Galois theory which had the same philosophy as these notes. It is Ian Stewart's *Galois Theory, Third Edition*, published by Chapman & Hall. If you read these notes, I recommend that you read Stewart's book afterward. Stewart has many, many historical details, particularly about Galois' life. He has a series of interesting exercises. He has more information on the work of Ruffini and Abel.

On the other hand, I'm happy I have these notes. They contain the answers to almost everything that puzzled me.

And so: Time to get to work.

Chapter 1

Arnold's Proof of the Abel-Ruffini Theorem

In 1963, Vladimir Arnold found a topological proof that there is no formula solving the general quintic equation by radicals. Arnold's proof avoids all of the abstract theory we develop later on, but it contains two key ideas from Galois theory in raw form, uncluttered by any polish we might later add. We sketch that proof now so the spirit of Arnold's analysis will hover over all of our later work.

Our sketch skips over some issues that will be mentioned at the end, so the argument in this section has some gaps.

There is a wonderful UTube video describing this proof. See https://www.youtube.com/watch?v=BSHv9Elk1MU.

Recall that complex radicals are multiple-valued. This follows from the multiple-valued nature of the logarithm;

$$\log(z) = \ln|z| + i\arg(z)$$

where $\arg(z)$ is only defined up to a multiple of 2π. All of this is a consequence of the polar form of complex numbers; any z can be written $z = re^{i\theta} = e^{\ln r}e^{i\theta} = e^{\ln|z|+i\theta}$ and so $\log(z) = \ln|z| + i\theta$. In turn,

$$\sqrt[n]{z} = e^{\frac{1}{n}\log(z)}$$

One way to deal with the multiple values of $\log z$ and $\sqrt[n]{z}$ is to remove the origin and negative x-axis and define the principal branch of the logarithm on the remaining set to be

$$\log z = \ln|z| + i\arg(z) \quad \text{where} \quad -\pi < \arg(z) < \pi$$

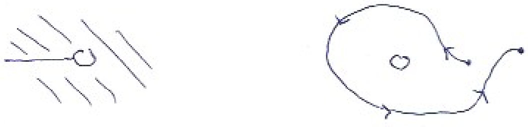


Figure 1.1: Restrict Domain OR Continuously Extend Along Curve

This trick also makes $\sqrt[n]{z}$ single-valued on the same set. See the picture on the left above.

In complex analysis, we often integrate expressions containing logarithms and radicals along paths in the complex plane. In this case, we deal with the multiple-valued nature of these functions in another way. Suppose we are integrating an expression containing $\log(z)$ and suppose our path $\gamma(t)$ never goes through the origin. We pick a value for the logarithm at the start of the path, and after that we require that the logarithm vary continuously with t. This completely determines the log along the remaining path. We can apply the same idea to radicals like $\sqrt[k]{z}$.

With this preliminary note out of the way, we are ready to begin studying polynomials and their roots. There are two natural pictures associated with a given polynomials: its roots in the complex plane, and its coefficients, also in this plane. See the picture below.

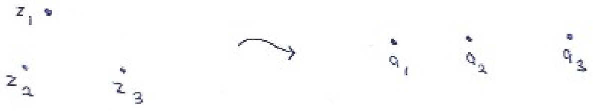


Figure 1.2: Roots of a Polynomial (left) and Coefficients of Polynomial (right)

There is an easy way to get from the roots to the coefficients, since we can just multiply out $(z - z_1)(z - z_2) \dots (z - z_k)$ to discover that $a_1 = -\sum z_i$ and so forth.

We are going to gradually move all of the roots, keeping them distinct. If we do that, the coefficients will also move. If these coefficients were originally real, they are likely to become complex. See the picture at the top of the next page.

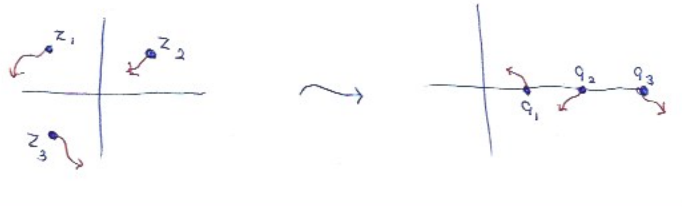


Figure 1.3: Moving the Roots Also Moves the Coefficients

Suppose that after such a motion, the roots return to the original root positions in some permuted order. Then the coefficients will return to their original values. See the picture below.

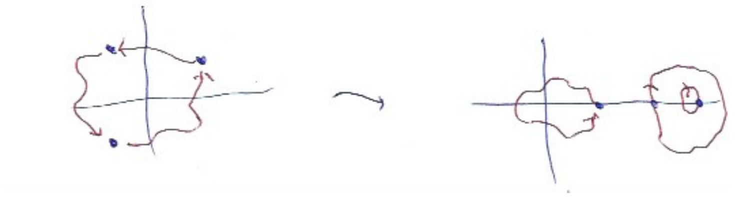


Figure 1.4: Permuting the Roots Returns the Coefficients to Original Values

Let us add to this picture a final map given by a conjectured general solution by radicals to $P(z) = 0$. The general solution will give multiple roots by choosing multiple values for the radicals, but let us pick an initial value for each radical so the general solution gives a particular root. Call this root z_1. Here is that picture. In the picture, the general solution is symbolized by a single radical sign, but of course it is a complicated expression with many sums, products, quotients, and radicals all mixed together.

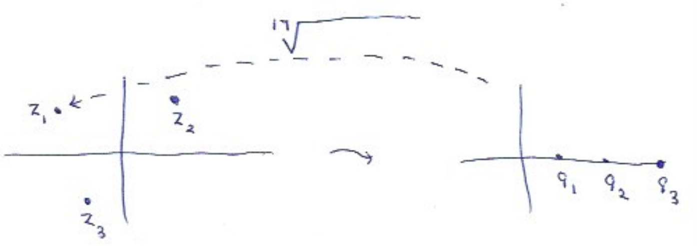


Figure 1.5: A Formula for Solving in Radicals Provides a Backward Map

Now imagine that we move the roots continuously. Then the coefficients move accordingly, and so the expressions inside the radical signs move continuously, and so the values of the radicals also move continuously and uniquely, always pointing to a root, and thus always pointing to the current position of z_1. So the previous picture is not static; we can turn on a motor and watch the roots, the coefficients, the radicals, and the solution move continuously, all in unison.

We have now arrived at the central point of Arnold's proof. Suppose the degree of our polynomial is at least three. We are first going to prove a special case: no general solution exists which contains just a single radical $\sqrt[k]{z}$. Inside this radical we allow an arbitrary expression formed from the coefficients by addition, subtraction, multiplication, and division. The radical does not have to appear alone; it can appear in an expression formed from the coefficients. It can even appear more than once but all appearances must be of the same radical.

Incidentally, there is a cubic formula, already mentioned in the Preface. But this formula has two different radicals, one a square root and one a cube root. Moreover, one radical is inside the other one. Both are not allowed in our preliminary lemma.

Form a specific motion of the roots as follows. First swap the roots in positions 1 and 2. Call this movement $\gamma(t)$. Then swap the roots in positions 1 and 3. Call this movement $\tau(t)$. Each of these paths on the left induces motions on the right. Pick the particular expression in the coefficients under the conjectured radical sign and follow it. Since the coefficients return to their original values, so does the path of this complicated expression. Since we have two paths on the left, γ and τ, we obtain two loops on the right, $\widehat{\gamma}$ and $\widehat{\tau}$.

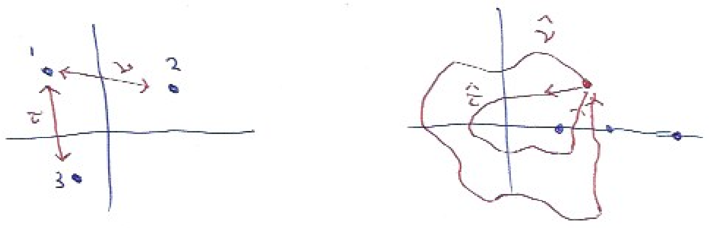


Figure 1.6: Transpositions on the Left Induce Loops on the Right

Consider the movement obtained by doing γ and then τ and then γ^{-1} and then τ^{-1}. So we swap the roots in positions 1 and 2, and then swap the roots in positions 1 and 3 and then unswap the roots in positions 1 and 2 and finally unswap the roots in positions 1 and 3. Please try this yourself. You may be surprised to discover that these actions cause the three roots to rotate by one click, as shown below.

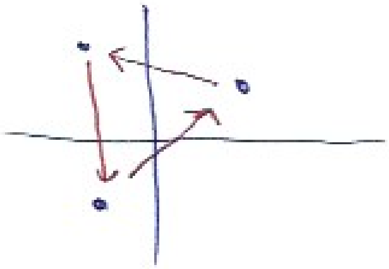


Figure 1.7: A Commutator of Transpositions Moves Everything

Over in coefficient space, the corresponding motion consists of four loops, $\widehat{\gamma}$ followed by $\widehat{\tau}$, followed by $\widehat{\gamma}$ in reverse, and then $\widehat{\tau}$ in reverse. But the resulting path doesn't wind around the origin at all, because each increase of the argument by a multiple of 2π is eventually undone. So on the right, the radical expression returns to its original value.

It follows that our conjectured formula giving z_1 does not move z_1, although we already know that z_1 moves to the original position of z_2. This contradiction proves the special case.

Notice that our argument proves much more. For cubics and higher, it rules out solutions containing many different radicals as long as the expressions inside these radicals can be obtained from the coefficients of the polynomial using only addition, subtraction, multiplication, and division.

However, expressions involving a radical inside another radical are not ruled out. Notice that the cubic formula, shown in the preface, does indeed have a square root inside a cube root. If we apply our argument to this situation, using a motion which is a commutator on the left, the inside radicals will return to their original value. So the outside radicals will be radicals of loops, but such radicals can end up at a different spot than they started, and thus can move z_1 to z_2.

Let us pause at this moment. Mathematicians who work with finite groups has many tools in their arsenal, but one of the most important is the notion of a *commutator*. If a and b are group elements, their commutator is $aba^{-1}b^{-1}$. If the group is abelian, this is the identity, but otherwise it can be very interesting. Readers who have solved Rubic's cube probably used a motion which rotated one face, and then a perpendicular face, and then unrotated the first face and then the second. That motion is a commutator, and it is useful because it only moves a few elements.

In Arnold's proof of the special case, the key idea is that the permutation group of the roots is not abelian and therefore we can produce an interesting commutator on the left side of our pictures. But the fundamental group of the space of loops around the origin is abelian, and thus commutators of loops on the right side of the diagram will be trivial.

Now let us return to the proof that equations of the fifth degree cannot be solved by radicals. Consider the special case where the formula contains a radical inside another radical. We must construct a permutation which moves z_1 on the left side of the diagram, but corresponds to loops on the right side of the diagram which preserve both the inner radical and the outer radical. We can do that using *commutators of commutators.* Suppose we find two paths on the left, both given by commutators. Then the corresponding loops on the right side will preserve the value of the inner radical. Now suppose we form the path given by the commutator of these commutators. This time, our argument shows that the outer radical is also preserved, and we have our contradiction.

Unfortunately, we also have to rule out radicals inside radicals inside radicals, and so forth. So we have to construct paths which actually move z_1 and are formed by a tower of "commutators of commutators of commutators of . . . of commutators".

To do this, we need at least five roots. Here is the beautiful and easy idea that finishes the argument. We know that three cycles are commutators of two transpositions. From now on, we build using only three cycles. The cycle (123) is a commutator. We will show that it is also a commutator of commutators. Here's the picture which shows that.

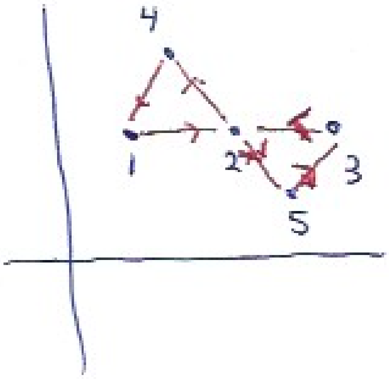


Figure 1.8: Every Three Cycle is a Commutator of Commutators

But then every three cycle is a commutator of commutators of commutators. And so forth. This finishes Arnold's argument.

Remark: The key moment of the proof comes with figure 1.5, where we claim that a general solution by radicals will produce a formula going backward. A careful reader might worry about two situations. First, is it possible that the algebraic expressions involving the roots will have singularities for special values. For instance, a term $\frac{1}{a_1^2+a_2+3}$ might vanish for exceptional values of the coefficients. And second, why are the expressions under the radicals always non-zero? If a path avoids the origin, we can find a continuous $\sqrt{z(t)}$, but there are uniqueness problems if $z(t)$ is sometimes zero.

We avoid some of these problems by only dealing with cases when the roots are all distinct. In the case of cubics, there will be a double root if and only if the discriminant

$$(z_1 - z_2)^2(z_1 - z_3)^2(z_2 - z_3)^2$$

is zero. This discriminant is symmetric in the roots and thus can be written as an expression in the coefficients by the fundamental theorem of symmetric polynomials. If

$$P(z) = z^3 + Az + B$$

the discriminant is $-4A^3 - 27B^2$.

The cubic formula involves $\sqrt{\frac{B^2}{4} + \frac{A^3}{27}}$, so this expression is zero if and only if the polynomial has a double root. But the formula also has cube roots of $-\frac{B}{2} \pm \sqrt{\frac{B^2}{4} + \frac{4A^3}{27}}$ and at least one of these terms will be zero exactly when

$$\left(\frac{B}{2}\right)^2 = \frac{B^2}{4} + \frac{4A^3}{27}$$

and thus when $A = 0$ and our equation is $z^3 + B = 0$. If $B \neq 0$, this has three distinct roots and thus might occur in our analysis.

Since the Abel-Ruffini theorem will be proved later in a completely different way, we will leave the Arnold proof as it is here, incomplete but extremely illuminating.

Chapter 2

Finite Group Theory

Before getting to the main subject, we prove some facts about group theory. These results will be used later in our study of generalizations of the quadratic formula.

In chemistry, every compound is built out of atoms. Knowing the atoms is not enough to determine the compound; for instance, graphite and diamond are both pure carbon. Similarly, we will show that every finite group is built out of ”simple groups”. But knowing the simple groups is not enough to determine the final group they build.

2.1 The Extension Problem; Simple Groups

Suppose we want to classify finite groups G.

It is natural to work by induction on the order of G. If H is a subgroup of G, then we already understand H. If H is normal, we also already understand G/H. Thus in the sequence

$$0 \to H \to G \to G/H \to 0$$

we understand the groups at the ends and need only fill in the middle group.

One choice for G is $H \times G/H$. There are usually others. The problem of constructing such G is called *the extension problem* in group theory; it is difficult. For example, suppose $H = Z_4$ and $G/H = Z_2$. Then G is a group of order 8. After some work, one can show that there are three G which fit in the sequence, $Z_4 \times Z_2, D_4$, and Q. Here D_4 is the dihedral group of order four, that is, the group of symmetries of a square. The group Q is the group of unit quaternions $\{\pm 1, \pm i, \pm j, \pm k\}$. Note that $\{\pm 1, \pm i\}$ is a normal subgroup isomorphic to Z_4. Note also that D_4 and Q are not isomorphic because D_4 has five elements of order two and Q has only one element of order two.

Luckily, we don't need to solve the extension problem for Galois theory. But suppose we completely understood this problem. What more would be needed to classify finite groups?

Some groups have no non-trivial normal subgroups. We call such groups *simple groups* and we would need to construct them another way. This has been one of the greatest research topics of the twentieth century, and there is now a complete list of all finite simple groups. A knowledge of this list, and a complete solution of the extension problem (which will never happen!) would produce a complete classification of all finite groups.

From the list we only need the *abelian* simple groups:

Theorem 2 *A finite abelian group is simple if and only if it equals Z_p for a prime p.*

Proof: All subgroups of an abelian group are normal, so it suffices to list all groups with no non-trivial subgroups. Certainly Z_p has no non-trivial subgroups, since every subgroup has order dividing p and thus equals $\{e\}$ or Z_p. Conversely if G is has no non-trivial subgroups and $g \neq e$ is in G, then the cyclic subgroup generated by g must be all of G, so G is cyclic of some order n. If n is not prime then it has non-trivial cyclic subgroups.

2.2 An Isomorphism Lemma

In the middle of the next section, we need a little lemma, so we prove it first.

Let G be a group with subgroups A and B. By definition AB is the set of all elements in G of the form $a_1b_1a_2b_2\ldots a_kb_k$ for varying k. This set is clearly a subgroup of G. If A and B are normal, so is AB because

$$g\left(a_1b_1a_2b_2\ldots a_kb_k\right)g^{-1} = \left(ga_1g^{-1}\right)\left(gb_1g^{-1}\right)\ldots\left(ga_kg^{-1}\right)\left(gb_kg^{-1}\right)$$

Lemma 1 *If A and B are normal,*

$$AB/B \cong A/(A \cap B)$$

Proof: We have a natural group homomorphism

$$A \to AB \to AB/B$$

and clearly this map sends $A \cap B$ to the identity. So it induces

$$A/(A \cap B) \to AB/B$$

The kernel of this map is e because if $a \in A$ maps to $e \in AB/B$, then $a \in A \cap B$. The map is onto because $a_ib_i \sim a_i$ in AB/B, so $a_1b_1\ldots a_kb_k \sim a_1\ldots a_k \in AB/B$.

2.3 Jordan-Holder

Continuing with the idealistic program of the previous sections, suppose G is an arbitrary finite group. Find a normal subgroup $H_1 \subset G$ unequal to G and as large as possible. It is easy to see that G/H_1 must be simple, since if $K \subset G/H_1$ is a non-trivial normal subgroup, the inverse image of K in G will be a normal subgroup $\tilde{K}$ with $H_1 \subset \tilde{K} \subset G$. If we understood the extension problem, we could construct G from H_1 and G/H_1.

Continue the process. Find a normal subgroup $H_2 \subset H_1$ unequal to H_1 and as large as possible. Then H_1/H_2 is simple. If we understood the extension problem, we could construct H_1 from H_2 and H_1/H_2.

Continuing in this vein, we eventually construct a *complete composition series*, that is, a chain of subgroups

$$\{e\} = H_n \subset H_{n-1} \subset \ldots \subset H_1 \subset G$$

with each H_i normal in H_{i-1} and as large as possible, and each H_{i-1}/H_i simple. The group G is constructed from the simple groups H_{i-1}/H_i by a series of group extensions.

Unfortunately, the H_i are not unique. For example, let $G = Z_6$. Let $Z_2 \subset Z_6$ be the subgroup $\{0, 3\}$ and let $Z_3 \subset Z_6$ be the subgroup $\{0, 2, 4\}$. Then we obtain two composition series

$$\{e\} \subset Z_2 \subset Z_6$$

$$\{e\} \subset Z_3 \subset Z_6$$

However, both series have "length 2"; the simple quotients in the first case are $Z_2, Z_3 = Z_6/Z_2$ and the simple quotients in the second case are $Z_3, Z_2 = Z_6/Z_3$ and thus the same up to order.

Theorem 3 (Jordan-Holder) *Any two complete composition series for a finite group have the same length, and their simple quotients are isomorphic up to order.*

Remark: This theorem is a generalization of the unique factorization theorem for integers. Indeed, if $n = p_1^{n_1} \ldots p_k^{n_k}$, it is easy to find a composition series for Z_n with simple quotients Z_{p_i}, each repeated n_i times.

Remark: Sometimes progress is made in mathematics by throwing information away until only the crucial information remains. Composition series allow us to throw away the intricate extension information until only the simple quotient information remains. It is this quotient information which is important in Galois theory.

In the previous section, we listed the three groups of order four obtained by extending Z_4 by Z_2. Notice that the simple quotients of all three groups are Z_2, Z_2, Z_2. So in this case, extension information is definitely thrown away.

Proof: We prove the theorem by induction on the order of G; the result is trivial for groups of order less than or equal to two.

In the induction step, suppose G has two composition series

$$\ldots \subset A_1 \subset A \subset G$$

$$\ldots \subset B_1 \subset B \subset G$$

If $A = B$, the theorem holds by induction, so suppose $A \neq B$. Then AB is a normal subgroup larger than A, and so $AB = G$. Notice that

$$A \cap B \subset A \subset G$$

$$A \cap B \subset B \subset G$$

By the isomorphism lemma, $G/A = AB/A = B/A \cap B$ and $G/B = AB/B = A/A \cap B$. It follows that the partial sequences above can be refined to composition sequences which are the same except at the beginning, and whose last two composition quotients are interchanged.

$$\ldots H_2 \subset A \cap B \subset A \subset G$$

$$\ldots H_2 \subset A \cap B \subset B \subset G$$

Compare these to the existing series

$$\ldots A_2 \subset A_1 \subset A \subset G$$

$$\ldots B_2 \subset B_1 \subset B \subset G$$

By induction, the theorem is true for A and B, so the lengths of the A_i and H_i series for A are equal, and the lengths of the B_i and H_i series for B are equal, and thus the lengths of the A_i and B_i series are equal. Moreover, the composition quotients for A are equal up to order, so $\{H_i/H_{i+1}\} = \{A_i/A_{i+1}\}$ up to order. Similarly, the composition quotients for B are equal up to order, so $\{H_i/H_{i+1}\} = \{B_i/B_{i+1}\}$ up to order. Adding G/A and A/A_1 to the first set, and G/B and B/B_1 to the second set and applying the isomorphism lemma as above finishes the proof.

Important Remark: In the material which follows, we often obtain series

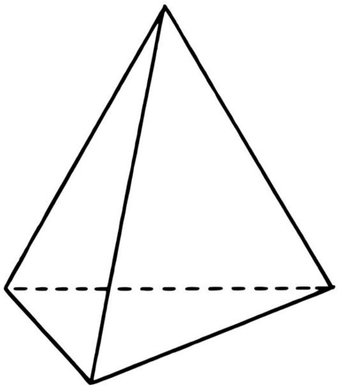
$$\ldots \subset G_2 \subset G_1 \subset G$$

in which the G_{i+1} are normal in G_i, but not necessarily maximal with this property. We still call these *composition series*. It is easy to see that every such series can be extended to a complete composition series by adding additional subgroups between the G_{i+1} and the G_i.

2.4 The Symmetric and Alternating Groups

Let us apply the previous theory to the symmetric group S_n, which has the alternating group A_n as a maximal normal subgroup. When $n = 2$, the alternating group is trivial and there is only one composition factor, Z_2. When $n = 3$, S_3 is isomorphic to the dihedral group D_3 and the group has two composition factors, Z_2 and Z_3.

When $n = 4$, S_4 is the group of all symmetries of a tetrahedron, and A_4 is the group of rotational symmetries of this tetrahedron. The group A_4 of order 12 has a normal subgroup $Z_2 \times Z_2$ of order 4, namely the group which rotates opposite pairs of lines by 180 degrees about their centers. Thus A_4 has composition factors Z_2, Z_2, Z_3.



We are going to prove that A_n is simple for $n \geq 5$. This will turn out to be the central reason that equations of degree five and higher cannot be solved by radicals.

It turns out that the group A_5 of order 60 is the smallest non-abelian simple group. This group is the group of rotational symmetries of a dodecahedron, so it is not surprising that the dodecahedron is part of the logo of the MAA.

A few simple observations will make our work easier. Every permutation can be written as a product of cycles: $\tau = (1,3)(2,5,7,8)(4,6)$. The sign of a cycle with k entries is $(-1)^{k-1}$ and the sign of a product of cycles is the product of the signs of the individual cycles. If σ is a permutation, an easily calculation shows that, for instance,

$$\sigma\tau\sigma^{-1} = \sigma \circ (1,3)(2,5,7,8)(4,6) \circ \sigma^{-1} = \Big(\sigma(1), \sigma(3)\Big) \Big(\sigma(2), \sigma(5), \sigma(7), \sigma(8)\Big) \Big(\sigma(4), \sigma(6)\Big)$$

Theorem 4 *If $n \geq 5$, A_n is simple. Thus the composition quotients of S_n are Z_2 and A_n.*

Proof of Theorem: It is well known that every element of S_n is a product of transpositions. Similarly, every element of the alternating group is a product of three cycles. Indeed, every element is a product of an even number of transpositions, so it suffices to show that the product of any two transpositions is a three cycle. If the transpositions contain a letter in common, then up to conjugation we have $(1,2)(2,3) = (1,3,2)$. Otherwise up to conjugation we have $(1,2)(3,4) = (3,2,4)(1,3,2)$.

Since $n \geq 5$, a normal subgroup of A_n with one three cycle contains all three cycles and consequently is all of A_n. For instance, if the normal subgroup contains $(1,2,3)$ and σ is a permutation taking i_1 to 1, i_2 to 2, and i_3 to 3, then after conjugation by σ it contains $(i_1, i_2, i_3) = \Big(\sigma(1), \sigma(2), \sigma(3)\Big)$. We need to make sure that the conjugating σ has even sign, but if not, choose two other indices i_4 and i_5 and multiply σ by the transposition which interchanges these extra indices.

We now prove that a nonzero normal subgroup of A_n is all of A_n. Suppose first that this subgroup contains an element g whose cycle notation has at least one cycle of length greater than three. For example, suppose the element is $g = (1,2,\dots,k)(...)$. Then the subgroup contains

$$g^{-1}\Big[(1,2,3)^{-1}g(1,2,3)\Big] = (2,3,k)$$

and so all three cycles, and so is all of A_n.

Suppose next that this subgroup contains an element g whose cycle notation has at least two cycles of length three. For instance, suppose $g = (1,2,3)(4,5,6)\dots$ is in the subgroup. Then

$$g^{-1}\Big[(1,2,4)^{-1}g(1,2,4)\Big] = (1,2,4,3,6)$$

and so by the previous step the subgroup is all of A_n.

Suppose next that the subgroup contains an element g whose cycle notation has only one cycle of length three and otherwise just transpositions. For instance, suppose $g = (1,2,3)\dots$ is in the subgroup. Then the subgroup contains $g^2 = (1,3,2)$ and so the subgroup is all of A_n.

Suppose finally that every element of the subgroup is a product of transpositions. Since there must be at least two of these transpositions, we can suppose $g = (1,2)(3,4)\dots$. Then

$$g\Big[(1,2,3)^{-1}g(1,2,3)\Big] = (1,4)(2,3)$$

is in the subgroup. So

$$(1,4)(2,3)\Big[(1,2,5)^{-1}(1,4)(2,3)(1,2,5)\Big] = (1,2,3,4,5)$$

is in the subgroup and consequently the subgroup is all of A_n. QED.

Remark: It is easy to connect the ideas in Arnold's proof to the ideas in this section. If G is a finite group, the set of all commutators in G forms a normal subgroup, and the quotient of G by this subgroup is abelian. We can continue by taking the commutator subgroup of the commutators, obtaining a composition series

$$C_n \subset C_{n-1} \subset \ldots \subset C_1 \subset G$$

If this sequence eventually gives $\{e\}$, then G has a composition series whose composition quotients are all abelian, and it easily follows that we can extend to a maximal composition series and thus all composition factors of G are Z_p. On the other hand, this does not happen if G has non-abelian composition factors, essentially because the commutator of any simple group is a non-trivial normal subgroup and thus the entire group, and thus every element of G can be written as a commutator of a commutator of a commutator of $\ldots$ for as long as necessary.

So the central idea of Arnold's proof is very close to the climactic theory of Galois theory which we prove much later on: every polynomial has an associated Galois group, and every root of the polynomial can be expressed by radicals if and only if all composition factors of the Galois group are abelian.

Chapter 3

The Quadratic, Cubic, and Quartic Formulas

3.1 The Quadratic Formula

Theorem 5 (Quadratic Formula) *The solutions of $ax^2 + bx + c = 0$ are*

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Proof: Introduce new coordinates $u = x - \lambda$ where λ is a constant to be chosen soon. Substituting in the equation gives

$$a(u + \lambda)^2 + b(u + \lambda) + c = au^2 + (2a\lambda + b)u + (a\lambda^2 + b\lambda + c) = 0$$

Now choose λ to make the coefficient of u vanish, so $\lambda = -\frac{b}{2a}$. The equation becomes

$$au^2 + \left(\frac{b^2}{4a} - \frac{b^2}{2a} + c\right) = 0$$

or

$$u^2 = \frac{b^2}{4a^2} - \frac{4ac}{4a^2}$$

This equation is easily solved:

$$u = \pm\frac{\sqrt{b^2 - 4ac}}{2a}$$

Hence

$$x = u + \lambda = \pm\frac{\sqrt{b^2 - 4ac}}{2a} - \frac{b}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

3.2 The Cubic Formula

We easily reduce an arbitrary cubic to the special form below, as will be seen shortly.

Theorem 6 (Cubic Formula) *The solutions of $x^3 + Ax + B = 0$ are*

$$x = \sqrt[3]{\frac{-B + \sqrt{B^2 + \frac{4A^3}{27}}}{2}} + \sqrt[3]{\frac{-B - \sqrt{B^2 + \frac{4A^3}{27}}}{2}}$$

Remark: Consider the equation $ax^3 + bx^2 + cx + d = 0$. Again introduce new coordinates $u = x - \lambda$ where λ is a constant. Substituting,

$$a(u + \lambda)^3 + b(u + \lambda)^2 + c(u + \lambda) + d = 0$$

or

$$au^3 + (3a\lambda + b)u^2 + (3a\lambda^2 + 2b\lambda + c)u + (a\lambda^3 + b\lambda^2 + c\lambda + d) = 0$$

Now choose λ to make the coefficient of u vanish, so $\lambda = -\frac{b}{3a}$ and the equation becomes

$$au^3 + \frac{3ac - b^2}{3a}u + \frac{2b^3 - 9abc + 27a^2d}{27a^2} = 0$$

Dividing by a gives

$$u^3 + \frac{3ac - b^2}{3a^2}u + \frac{2b^3 - 9abc + 27a^2d}{27a^3} = 0$$

Naming the coefficients in this equation A and B gives $u^3 + Au + B = 0$, and it suffices to solve this simpler equation.

All of this was known early in the Renaissance. In Italian, the word “cosa” means “thing.” So “cosa nostra” means ”our thing” in the gangster world, and the above equation was known in Renaissance Italy as the equation of the “cube and the cosa.” Many people tried to find the solution; even Omar Khayyam. The solution was discovered around 1500 by Scipione del Ferro. Independently, Niccolo Tartaglia found the solution, and later his solution was published by Gerolamo Cardano in his book *Ars Magna.*

Proof: To deduce the solution we introduce a trick. Write

$$x = u + \frac{\alpha}{u}$$

where α is a nonzero constant to be determined later. We can certainly write x in this form, since solving for u amounts to solving

$$u^2 - xu + \alpha = 0$$

and this equation has a nonzero solution. However, u might be complex even if x is real.

Substituting in the original equation gives

$$\left(u+\frac{\alpha}{u}\right)^3 + A\left(u+\frac{\alpha}{u}\right) + B = 0$$

or

$$u^3 + (3\alpha + A)u + (3\alpha^2 + A\alpha)\frac{1}{u} + B + \alpha^3\frac{1}{u^3} = 0$$

We now choose

$$\alpha = -\frac{A}{3}$$

and notice that the resulting equation is

$$u^3 + B - \frac{A^3}{27u^3} = 0$$

or

$$(u^3)^2 + Bu^3 - \frac{A^3}{27} = 0$$

but this is a quadratic equation in u^3. So

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

and

$$u = \sqrt[3]{\frac{-B \pm \sqrt{B^2 + \frac{4A^3}{27}}}{2}}$$

Therefore

$$x = \sqrt[3]{\frac{-B \pm \sqrt{B^2 + \frac{4A^3}{27}}}{2}} - \frac{A}{3\sqrt[3]{\frac{-B \pm \sqrt{B^2 + \frac{4A^3}{27}}}{2}}}$$

Our cubic has multiple roots, which can be obtained by taking different square roots and different cube roots; more detail will be given soon. But notice that the same cube root and the same square root must be used in both terms of the above formula. Let us replace the plus/minus sign with just a plus. The square root may be complex and thus can have two values with no easy way to distinguish one over the other.

Notice that

$$\left(\frac{-B+\sqrt{B^2+\frac{4A^3}{27}}}{2}\right)\left(\frac{-B-\sqrt{B^2+\frac{4A^3}{27}}}{2}\right)=\frac{B^2-(B^2-\frac{4A^3}{27})}{4}=\frac{A^3}{27}$$

and therefore

$$\sqrt[3]{\frac{-B+\sqrt{B^2+\frac{4A^3}{27}}}{2}}\;\sqrt[3]{\frac{-B-\sqrt{B^2+\frac{4A^3}{27}}}{2}}=\frac{A}{3}$$

provided we pick the value of the second cube root making this product $\frac{A}{3}$. This allows us to rewrite the formula in the form

$$x=\sqrt[3]{\frac{-B+\sqrt{B^2+\frac{4A^3}{27}}}{2}}+\sqrt[3]{\frac{-B-\sqrt{B^2+\frac{4A^3}{27}}}{2}}$$

Remark: In the final formula, we must choose one of the two values of the square root and use it in both places. Then we must select one of the three values of the first cube root; the second cube root will also have three values, but we must choose the value which makes the product of the two cube roots equals $\frac{A}{3}$.

Notice that if we change the sign of the square roots, then the two cubics are just interchanged, and their product is still $\frac{A}{3}$. So it suffices to fix one value for the square root, and then the three possible cube roots for the first term give the three possible roots.

Remark: By putting the equations of the previous paragraphs together, it is possible to write a single formula solving the general cubic and thus generalizing the quadratic formula. If you insist on doing so, go right ahead. In practice, we always simplify as above and solve the easier $x^3+Ax+B=0$.

Remark: In these equations, we are to choose one of the two values of the square root and use it consistently, and we are to choose one of the three values of the cube root and use it consistently. All told, then, there are six possible choices to produce the three solutions of the original cubic equation. So we expect that each root will be given twice by these formulas.

Remark: There are some unfortunate features of this solution. Consider what happens when the equation we want to solve is

$$x^3-2x-4=(x-2)(x^2+2x+2)=0$$

In this case $A=-2$ and $B=-4$ and

$$u^3=\frac{4\pm\sqrt{16-\frac{32}{27}}}{2}=2\pm 10\sqrt{\frac{1}{27}}$$

Taking the positive square root and the real cube root, we obtain

$$x = \sqrt[3]{2 + 10\sqrt{\frac{1}{27}}} + \sqrt[3]{2 - 10\sqrt{\frac{1}{27}}}$$

There is no hint here that the solution is an integer. I leave it to the reader to show that this expression equals 2.

A different problem is revealed by the example below, which shows that complex numbers can occur even if the final solution is real:

Example: Consider the equation $X^3 - 15X - 4 = 0$. The cubic formula asks us to compute

$$\frac{-B \pm \sqrt{B^2 + \frac{4A^3}{27}}}{2} = \frac{4 \pm \sqrt{16 - \frac{4 \cdot 15^3}{27}}}{2} = 2 \pm \sqrt{4 - 5^3} = 2 \pm \sqrt{-121} = 2 \pm 11i$$

According to the formula, the solution is then

$$X = \sqrt[3]{2 + 11i} + \sqrt[3]{2 - 11i}$$

According to the paper *On The Casus Irreducibilis of Solving the Cubic Equation* by Jay Villaneuva at Florida Memorial University, Bombelli noticed in 1550 that $(2+i)^3 = 2+11i$ and $(2-i)^3 = 2 - 11i$ and thus that

$$X = (2+i) + (2-i) = 4$$

which indeeds solves $X^3 - 15X - 4 = 0$.

Remark: If a cubic has a multiple root, it is easily solved. Otherwise a cubic with real roots either has one real root or three real roots. It turns out that when there is one real root, the expression under the square root sign is positive. The cubic formula then asks for the cube root of some real number, which always exists. So the cubic formula yields the real solution.

But if there are three real roots, then the expression under the square root sign is negative, and the cubic formula forces us to compute the cube root of a complex number. The final formula involves this cube root twice, and miraculously the complex part of the result cancels and we obtain real roots.

The case of the cubic formula when the cubic has three real roots is known as the *casus irreducibilis*. The real reason that mathematicians introduced complex numbers and began to take them seriously is that the cubic formula requires us to calculate cube roots of complex quantities in this case.

We will later prove that in the casus irreducibilis case, there is no expression involving only real roots giving the roots of the cubic. So the use of complex numbers in this case is essential.

Theorem 7 *If x^3+Ax+B has one real root and two imaginary roots, $B^2+\frac{4A^3}{27}$ is positive and u can be chosen to be real. If x^3+Ax+B has three distinct real roots, $B^2+\frac{4A^3}{27}$ is negative and u is necessarily complex.*

Proof: Call the roots x_1, x_2, and x_3. We are going to show that

$$[(x_1-x_2)(x_1-x_3)(x_2-x_3)]^2 = -27B^2-4A^3.$$

If the roots are real and distinct, it follows that $27B^2+4A^3$ is negative. On the other hand, if $x_1 = a, x_2 = b+ic, x_3 = b-ic$, then $(x_1-x_2)(x_1-x_3)(x_2-x_3) = [(a-b)^2+c^2](-2ic)$ and the square of this expression is $[(a-b)^2+c^2]^2(-4c^2)$ and so negative, so $27B^2+4A^3$ is positive.

Consider the expression $(x-x_1)(x-x_2)(x-x_3) = x^3+Ax+B$. Differentiate and then set x successively to x_1, x_2, x_3 to obtain

$$(x_1-x_2)(x_1-x_3) = 3x_1^2+A$$

$$(x_2-x_1)(x_2-x_3) = 3x_2^2+A$$

$$(x_3-x_1)(x_3-x_2) = 3x_3^2+A$$

Multiply left sides and right sides together to obtain

$$[(x_1-x_2)(x_1-x_3)(x_2-x_3)]^2 = -(3x_1^2+A)(3x_2^2+A)(3x_3^2+A)$$

The right side of this expression is

$$-27(x_1x_2x_3)^2-9A(x_1^2x_2^2+x_1^2x_3^2+x_2^2x_3^2)-3A^2(x_1^2+x_2^2+x_3^2)-A^3$$

and we want to show that it equals $-(27B^2+4A^3)$.

However $(x-x_1)(x-x_2)(x-x_3) = x^3+Ax+B$ and consequently

$$x_1+x_2+x_3 = 0$$

$$x_1x_2+x_1x_3+x_2x_3 = A$$

$$x_1x_2x_3 = -B$$

Squaring the first of these formulas gives $x_1^2+x_2^2+x_3^2+2(x_1x_2+x_1x_3+x_2x_3) = 0$ and consequently

$$x_1^2+x_2^2+x_3^2 = -2A$$

Squaring the second formula gives $x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2 + 2(x_1x_2x_3)(x_1 + x_2 + x_3) = A^2$ and consequently

$$x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2 = A^2$$

Thus the expression which interests us is

$$-27(x_1x_2x_3)^2 - 9A(x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2) - 3A^2(x_1^2 + x_2^2 + x_3^2) - A^3 =$$

$$-27B^2 - 9A(A^2) - 3A^2(-2A) - A^3 = -27B^2 - 4A^3$$

3.3 The Quartic Formula

The quartic formula was discovered by Lodovico Ferrari in 1540. Ferrari's formula reduces the quartic to a related cubic, which is then solved using the result of del Ferro and Tartaglia.

Ferrari was a student of Cardano, and he learned of Tartaglia's solution from Cardano. Unfortunately, Tartaglia revealed his solution to Cardano late at night over a glass of wine, and made Cardano swear an oath never to reveal his solution. Although Tartaglia claimed that he would soon publish the result, he never did. This put Cardano and Ferrari in a bind, because solving a quartic depended on Tartaglia's secret solution of the cubic.

This knot was untangled when Cardano realized that del Ferro, who by that time had died, had independently solved the cubic. So Cardano visited del Ferro's widow and discovered that she saved her husband's papers. Sacred oaths were taken seriously in the Renaissance, but they were also taken literally, so revealing del Ferro's formula was not the same thing as revealing Tartaglia's formula. In this way, Cardano published del Ferro's cubic formula, and Ferrari's quartic result became public.

Rather than using Ferrari's method, we discuss a simple method described in the Wikipedia article on the quartic. Begin with a general quartic equation

$$ax^4 + bx^3 + cx^2 + dx + e = 0$$

Divide the equation by a to obtain a corresponding equation with $a = 1$. Set $x = u - \frac{b}{4}$ and notice that the corresponding equation for u has the form

$$u^4 + Au^2 + Bu + C = 0$$

Now comes the main trick. We attempt to factor this as a product of two quadratics:

$$u^4 + Au^2 + Bu + C = (u^2 + pu + q)(u^2 + ru + s)$$

If we can find p, q, r, s, then we can solve the two quadratics by the quadratic formula and find all four roots of the quartic. Multiplying out, we find that we need to solve

$$0 = p + r$$

$$A = s + pr + q$$
$$B = ps + qr$$
$$C = qs$$

The first equation implies that

$$r = -p$$

so it suffices to solve

$$A = s - p^2 + q$$
$$B = ps - pq$$
$$C = qs$$

The first and second equations can be rewritten

$$A + p^2 = s + q$$
$$\frac{B}{p} = s - q$$
$$C = qs$$

We can solve these equations for s and q in terms of p:

$$s = \frac{1}{2}\left(A + p^2 + \frac{B}{p}\right)$$
$$q = \frac{1}{2}\left(A + p^2 - \frac{B}{p}\right)$$

Thus it suffices to find p satisfying the last equation

$$C = qs = \frac{1}{4}\left((A + p^2)^2 - \frac{B^2}{p^2}\right)$$

or equivalently

$$p^6 + 2Ap^4 + (A^2 - 4C)p^2 - B^2 = 0$$

This is a cubic equation in p^2. To solve the reduced quartic, we solve this equation for p^2, extract a root to find p, and use the result to compute r, s, q, and then solve the two resulting quadratics.

Remark: From here, it is possible to write down explicit formulas for the roots. I won't bother.

We should worry about edge cases. This happens when all solutions of the cubic are zero. That can only happen if $A = 0, C = 0$ and $B = 0$, but then the quartic is $u^4 = 0$, which has trivial solutions. We might worry that the cubic formula would produce a complicated expression which happens to simplify to zero. This would happen if the cubic has *one* zero solution, which happens if $B = 0$. But then the quartic is a quadratic equation in u^2, $u^4 + Au^2 + C = 0$, which can be solved by two applications of the quadratic formula.

Chapter 4

Field Extensions and Root Fields

4.1 Motivation for Field Theory

We now want to show that there cannot be equivalent formulas for equations of degree five and higher. This, of course, is much harder. To get started, we have to look at our earlier formulas from a more general point of view.

Consider the solution of a cubic. Starting with the coefficients A and B of the equation and the standard rational numbers, we use addition, subtraction, multiplication, and division to obtain the expression

$$B^2 + \frac{4A^3}{27}.$$

Then we form the square root of this expression:

$$\sqrt{B^2 + \frac{4A^3}{27}}$$

Using this new number, A, B, and the rationals, we use addition, subtraction, multiplication, and division to obtain a fancier expression:

$$\frac{-B + \sqrt{B^2 + \frac{4A^3}{27}}}{2}$$

Then we form the cube root of this expression:

$$\sqrt[3]{\frac{-B + \sqrt{B^2 + \frac{4A^3}{27}}}{2}}$$

Finally, starting from A, B, the rationals, and the two radicals, we apply additional, subtraction, multiplication, and division to obtain the actual root:

$$x = \sqrt[3]{2 + 10\sqrt{\frac{1}{27}}} + \frac{2}{3}\frac{1}{\sqrt[3]{2 + 10\sqrt{\frac{1}{27}}}}$$

It is convenient to think of this process in the following way. Imagine a large pot of numbers — indeed an infinite pot of numbers. When we begin, this pot contains the positive integers $1, 2, 3, \ldots$ and A and B. We want to add numbers to the pot until finally it contains the solutions to the cubic. We can add numbers in five ways: by adding two numbers already in the pot, by multiplying two numbers already in the pot, by subtracting two numbers already in the pot, by dividing two numbers already in the pot, and by adding a radical of a number already in the pot.

It turns out that adding a radical is by far the most important of these operations. So it is convenient to imagine the process as follows. First we add *all* numbers that can be obtained from the positive integers and A and B by adding, subtracting, multiplying, and dividing. This adds infinitely many numbers; call the resulting set K_0. Next compute a radical of *one* of the numbers in the pot and add that. This radical might be a square root, or a cube root, or whatever. After that, add all numbers which can be obtained from numbers now in the pot by adding, subtracting, multiplying, and dividing. Call the resulting set K_1, so $K_0 \subset K_1$.

Continue. Add a radical of a number currently in the pot, and then add all numbers that can be obtained from these by adding, subtracting, multiplying, and dividing. Call the new set K_2. In this way we obtain a finite chain of sets of numbers:

$$K_0 \subset K_1 \subset K_2 \subset \ldots \subset K_n$$

In the end, the roots of our equation should belong to K_n.

This way of thinking puts the emphasis on constructing the radical, since that's the process that gives the inclusion sign $\subset$. The process of adding, subtracting, multiplying, and dividing is hidden in the construction of K_i after the radical is added. In the following sections, we study these two processes in detail. We'll begin with the process of completing a set using addition, subtraction, multiplication, and division.

4.2 Fields

Definition 1 *A field is a set K with two binary operations $x, y \to x + y$ and $x, y \to x \cdot y$ satisfying the following axioms:*

1. *$x + y = y + x$ for all x, y*
2. *$(x + y) + z = x + (y + z)$ for all x, y, z*
3. *there is an element 0 such that $0 + x = x$ for all x*
4. *given x, there is an element $-x$ such that $x + (-x) = 0$*
5. *$x \cdot y = y \cdot x$ for all x, y*
6. *$(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all x, y, z*
7. *there is an element $1 \neq 0$ such that $1 \cdot x = x$ for all x*
8. *given $x \neq 0$, there is an element x^{-1} such that $x \cdot x^{-1} = 1$*
9. *$x \cdot (y + z) = (x \cdot y) + (x \cdot z)$ for all x, y, z*

Remark: It is easy to prove that 0 and 1 are unique, and that $(-x)$ and x^{-1} are uniquely determined by x.

Remark: Clearly the set of rational numbers Q, the set of real numbers R, and the set of complex numbers C are fields. Almost all of the fields in these notes are subsets $K \subset C$ with the standard addition and multiplication. Such a subset is a field if it satisfies five conditions:

1. $0 \in K$ and $1 \in K$
2. $x, y \in K$ implies $x + y \in K$
3. $x \in K$ implies $-x \in K$
4. $x, y \in K$ implies $x \cdot y \in K$
5. $x \in K$ and $x \neq 0$ implies $x^{-1} \in K$

Remark: The set Z_n of integers modulo n is a commutative ring, satisfying all field axioms except possibly axiom 8. If $n = ab$ has a non trivial factorization, then Z_n is not a field because a and b represent nonzero elements of Z_n whose product is zero. But if $n = p$ is prime, then Z_p is a field; indeed if $0 < a < p$, then a and p are relatively prime, so there exist integers A and B with $Aa + Bp = 1$ and consequently A represents a multiplicative inverse of a in Z_p.

Theorem 8 *Every field K has a smallest subfield, which is isomorphic to either Q or Z_p for some prime p.*

Proof: The intersection F of all subfields of K is a field, and thus the smallest subfield.

Define a map $\varphi : Z \to F$ by sending the generator $1 \in Z$ to $1 \in F$. This map is necessarily a ring homomorphism, and thus its kernel is an ideal, I. This ideal is prime, for if $ab \in I$, then $0 = \varphi(ab) = \varphi(a)\varphi(b)$, so $\varphi(a) = 0$ or $\varphi(b) = 0$ and thus $a \in I$ or $b \in I$. We conclude that $I = (0)$ or else $I = (p)$ for some prime p. In the first case $Z \subset F$ and we rapidly conclude that $Q \subset F$. In the second case $Z_p \subset F$. Since F is the smallest subfield, $Q = F$ or $Z_p = F$.

4.3 An Important Example

Example: In the previous section we started with $1 \in K$ and discovered that by adding all sums, additive inverses, products, and quotients we would end up with either Q or Z_p.

It is instructive to examine a similar example. Suppose we start with Q and $\sqrt{2}$ in a larger field K, and again add all sums, additive inverses, products, and quotients, to arrive at the smallest subfield of K containing Q and $\sqrt{2}$. In the argument, imagine that the larger K is either R or C.

If $a, b \in Q$, then surely we will obtain all $a+b\sqrt{2}$. Surprisingly, we get nothing more. Indeed the sum of two such elements is another, and the additive inverse of such an element is another. It is only slightly harder to see that the product of two such elements is another, essentially because $\sqrt{2} \cdot \sqrt{2} = 2 \in Q$. Finally the quotient of two such elements is another because

$$\frac{a+b\sqrt{2}}{c+d\sqrt{2}} = \frac{(a+b\sqrt{2})(c-d\sqrt{2})}{(c+d\sqrt{2})(c-d\sqrt{2})} = \frac{(ac-2bd)+(bc-ad)\sqrt{2}}{c^2-2d^2} = \left(\frac{ac-2bd}{c^2-2d^2}\right)+\left(\frac{bc-ad}{c^2-2d^2}\right)\sqrt{2}$$

Notice that if $c \neq 0$ or $d \neq 0$ then $c^2 - 2d^2 \neq 0$ because $\sqrt{2}$ is not rational.

4.4 Extension Fields

Imagine a very large field like the field of complex numbers, which we will pretend is the set of all possible numbers. Suppose we have a subfield K of this large field and an element a of the large field. Define $K(a)$ to be the set of all numbers which an be formed from K and a by adding, subtracting, multiplying, and dividing. We are going to obtain a very concrete description of such fields $K(a)$. Of course if $a \in K$ then $K(a)$ is just K, so assume that $a \notin K$.

There are two possibilities. If a satisfies a non-trivial polynomial $P(x)$ with coefficients in K, we say the extension $K(a)$ is *algebraic over* K. If there is no such polynomial, we say the extension is *transcendental over* K. For instance, if K is the field of rational numbers Q and $a = \sqrt{2}$, then a satisfies $x^2 - 2 = 0$. But if $K = Q$ and $a = \pi$, then there is no such polynomial.

We'll consider the easier transcendental case first. If $P(x)$ and $Q(x)$ are polynomials with coefficients in K and $Q(x)$ is not identically zero, then $P(a)$ and $Q(a)$ are certainly in the extension field $K(a)$. Moreover $Q(a) \neq 0$ because a is transcendental. So $\frac{P(a)}{Q(a)}$ is in this field. On the other hand, the set of all such elements is a field, and so

$$K(a) = \left\{ \frac{P(a)}{Q(a)} \mid P(x) \text{ and } Q(x) \in K[x] \right\}$$

This is not the end of the story, because it is conceivable that polynomials P_1, P_2, Q_1, and Q_2 exist with

$$\frac{P_1(a)}{Q_1(a)} = \frac{P_2(a)}{Q_2(a)}$$

But if so, then $P_1(a)Q_2(a) - P_2(a)Q_1(a) = 0$ and so the polynomial $P_1(x)Q_2(x) - P_2(x)Q_1(x)$ vanishes on $x = a$. Since a is transcendental, this polynomial must be trivial and $P_1(x)Q_2(x) = P_2(x)Q_1(x)$.

The quotient field of the integral domain $K[x]$ is by definition the set of all $\frac{P(x)}{Q(x)}$ with two such elements identified exactly if $P_1(x)Q_2(x) = P_2(x)Q_1(x)$. It is called *the field of rational functions over* K *in* x. We have therefore proved

Theorem 9 *If K is a subfield of a larger field which contains $a \notin K$, and if a is transcendental over K, then the field $K(a)$ is canonically isomorphic to the field of rational functions over K in x.*

4.5 Algebraic Extensions; Root Fields

We now study the more interesting case of $K(a)$ when a is algebraic over K. It is useful to keep in mind the special case when $K = Q$ and $a = \sqrt{2}$ studied at the end of section ten.

Since a is algebraic, it satisfies a polynomial $P(x)$ with coefficients in K. Factoring this polynomial if necessary, we can assume P is irreducible. Of all such irreducible polynomials vanishing on a, select P to have smallest degree and leading coefficient 1. We call this P the *minimal polynomial of* a.

The minimal polynomial divides any polynomial over K which vanishes on a. Indeed if $Q(x)$ is nontrivial and vanishes on a, we can divide $Q(x)$ by $P(x)$ to obtain

$$Q(x) = A(x)P(x) + R(x)$$

with $R(x)$ either zero or else of smaller degree than P. Since R vanishes on a and P is minimal, we conclude that R is identically zero and thus that P divides Q. If P and Q are both minimal, they both have leading coefficient one and must be equal.

Suppose the minimal polynomial has degree n and consider the set of expressions

$$K(a) = \left\{ \ r_0 + r_1 a + r_2 a^2 + \ldots + r_{n-1} a^{n-1} \mid r_i \in K \ \right\}$$

This set contains K and a and is clearly closed under addition. It is also closed under multiplication, for if we multiply two such expressions, we get another expression of the same form, but with degree possibly larger than $n-1$. However, a is a root of its minimal polynomial $P(x) = x^n + k_{n-1}x^{n-1} + \ldots + k_0$, so we can write

$$a^n = -k_0 - k_1 a - \ldots - k_{n-1} a^{n-1}$$

Thus $a^n \in K(a)$ and by the same technique all higher powers of a belong to this set.

Finally, the set $K(a)$ is also closed under division by the following lemma (justifying our notation):

Lemma 2 *Let K be a field, and suppose $K \subset L$ where L is an integral domain which is finite dimensional over K. Then L contains inverses of all nonzero elements and thus is itself a field.*

Proof: Let $b \in L$, $b \neq 0$. Consider the map $L \to L$ given by multiplication by b. This is a linear transformation over K. Since L is an integral domain, the map is one-to-one. By linear algebra, it is thus onto, so there exists $c \in L$ with $bc = 1$.

Remark: Notice that this simple lemma replaces the calculation at the end of section 3.3 showing that inverses exist in that special case.

4.6 Irreducible Polynomials over Q

In applications of Galois theory, we often assume the ground field is the field Q of rational numbers. We need an ample supply of irreducible polynomials over this field to give examples of the root fields just introduced. Such polynomials are usually found using two well-known results:

Theorem 10 (Gauss) *If a polynomial with integer coefficients factors over Q, then it factors over the integers.*

Theorem 11 (Eisenstein) *Let $P(X) = a_0 X^n + a_1 X^{n-1} + \ldots + a_n$ be a polynomial with integer coefficients. Fix a prime p and suppose*

1. *p does not divide a_0*
2. *p divides a_i for $i > 0$*
3. *p^2 does not divide a_n*

Then $P(X)$ is irreducible over the rationals.

Proof: Call a polynomial over the integers *primitive* if no prime divides all of its coefficients. If $P(X)$ is a nonzero polynomial over the integers, we can factor out integers that divide all coefficients, and eventually write $P(X) = c_P \tilde{P}$ where c_P is a positive integer and $\tilde{P}$ is primitive. This representation is clearly unique.

If $P(X)$ is a nonzero polynomial over Q, we can similarly factor out all denominators of individual terms and eventually write $P(X) = c_P \tilde{P}$ where c_P is a positive rational and $\tilde{P}$ is a primitive polynomial over the integers. This representation is again unique.

Now we come to the key point: the product of two primitive polynomials is again primitive. Indeed consider such a product

$$\left(a_0 X^m + a_1 X^{m-1} + \ldots + a_m\right)\left(b_0 X^n + b_1 X^{n-1} + \ldots + b_n\right)$$

If the result is not primitive, there is a prime p which divides all terms of the product. Since $a_0 X^m + \ldots + a_m$ is primitive, there is a j such that p does not divide a_j but divides $a_{j+1}, \ldots, a_m$. Note that j might be m. Similarly there is a k such that p does not divide b_k but does divide $b_{k+1}, \ldots, b_n$. Consider the term

$$\ldots + a_{j-1} b_{k+1} + a_j b_k + a_{j+1} b_{k-1} + \ldots$$

This coefficient in the product is not divisible by p because every subterm is divisble by p except $a_j b_k$, contradicting our assumption on p.

We can now prove Gauss' result. Suppose that $P(X)$ is a polynomial with integer coefficients which factors over the rationals as $Q(X)R(X)$. Then

$$c_P \tilde{P} = c_Q \tilde{Q} c_R \tilde{R} = (c_Q c_R) \tilde{Q} \tilde{R}.$$

Since $\tilde{Q}\tilde{R}$ is primitive and this representation is unique, we find that $\tilde{P} = \tilde{Q}\tilde{R}$. Since P has integer coefficients, c_P is an integer and $P = c_P \tilde{P} = c_P \tilde{Q} \tilde{R} = \left(c_P \tilde{Q}\right) \tilde{R}$.

We propose to prove Eisenstein's result using the notation of the proof just given. Suppose P is not irreducible over the rationals. Then it can be factored as a product of polynomials with integer coefficients:

$$\ldots + a_{j-1} b_{k+1} + a_j b_k + a_{j+1} b_{k-1} + \ldots$$

If the constant coefficient of the product is divisible by p but not p^2, then p must divide exactly one of $a_m b_n$. Say p divides a_m and p does not divide b_n. Find k such that p does not divide a_k but p divides $a_{k+1}, \ldots, a_m$. Such a k must exist, else p would divide all a_i and thus the highest coefficient of the product. One of the coefficients of the product is then

$$a_k b_n + a_{k+1} b_{n-1} + \ldots$$

and p does not divide this term because it divides all subterms except the first. So this must be the highest coefficient of the product, and $k + n = 0$. So the second term is constant and we really don't have a factorization.

Example: It follows from this result that there are root field extensions of the rational numbers with any degree. Indeed, $X^n + 2X^{n-1} + \ldots + 2$ is irreducible by Eisenstein's result and has degree n.

4.7 The Degree of a Field Extension

At the end of the previous section, we proved a result using linear algebra. This section is about the deeper structure behind that proof.

Suppose $K \subset L$ is a field extension. By forgetting some of the axioms of L, we see that L is a vector space over K. The dimension of this vector space is called the *degree of L over K* and denoted $[L : K]$.

Corollary 1 *If $K \subset K(a)$ is an extension, then $[K(a) : K]$ if infinite when a is transcendental over K, and finite if a is algebraic over K. In this second case, $[K(a) : K]$ is the degree of the minimal polynomial of a.*

Proof: This is an immediate consequence of previous results.

Theorem 12 *Suppose $K \subset L \subset M$. Then*

$$[M : K] = [M : L][L : K]$$

Proof: Suppose $[M : L]$ has degree p and let $m_1, \ldots, m_p$ be a basis for M over L. Suppose $[L : K]$ has degree q and let $l_1, \ldots, l_q$ be a basis of L over K. We will prove that $\{m_i l_j\}$ is a basis of M over K. The theorem immediately follows.

We first prove these elements linearly independent. Suppose $\sum_{i,j} a_{ij} l_i m_j = 0$ with $a_{ij} \in K$. Then $\sum_i \left(\sum_i a_{ij} l_i\right) m_j = 0$. Since the m_j are linearly independent over L, $\left(\sum_i a_{ij} l_i\right) = 0$ for each fixed j. Since the l_i are linearly independent over K, each $a_{ij} = 0$.

We now prove these elements generate M over K. Let $m \in L$. Since the m_j form a basis of M over L, we can write $m = \sum_j b_j m_j$ for coefficients $b_j \in L$. Since the l_i form a basis of L over K, each b_j can be written $b_j = \sum_i a_{ij} l_i$. So $m = \sum_{i,j} a_{ij} l_i m_j$.

4.8 Existence of Root Fields

This section contains an important new way to look at the results of section 3.5. In that section, we started with a "universal field" like the complex numbers and an element a in this field, produced a minimal polynomial $P(x)$, and constructed a larger extension field $K(a)$ containing K and the root a of P.

There is another way to think of this result. Suppose we do not have a universal field or a root a, but instead start with a field K and an irreducible polynomial P over K. We can then construct a new field L, whose elements are abstract symbols rather than complex numbers, with the properties that $K \subset L$, that $P(x)$ has a root in L, and that L is generated by this root and K. We call this the *abstract root field associated with* P.

The construction is very simple. Start with the set of all polynomials with coefficients in K. Denote this set by $K[X]$. It is a commutative integral domain. Let $\mathcal{J}$ be the ideal in this ring generated by the polynomial $P(X)$. Thus it consists of all polynomials of the form $P(X)R(X)$ for arbitrary polynomials R. It is easy to check that $\mathcal{J}$ is a prime ideal, and therefore the quotient ring $K[X]/\mathcal{J}$ is a field. This is our new field L. So

$$L = K[X]/\mathcal{J}$$

It is easy to make this construction look more concrete. If $A(X) \in K[X]$ is a polynomial, we can divide A by P to obtain $A(X) = P(X)Q(X) + R(X)$ where the degree of R is smaller than the degree of P. In the quotient field, A is equivalent to R because their difference is in the ideal $\mathcal{J}$. So every element in the quotient is equivalent to exactly one polynomial of degree smaller than n:

$$L = \left\{ \ k_0 + k_1 X + k_2 X^2 + \ldots + k_{n-1} X^{n-1} \mid k_i \in K \ \right\}$$

Notice that this expression is essentially the same as the expression for a general element of the root field $K(a)$:

$$L = \left\{ \ k_0 + k_1 a + k_2 a^2 + \ldots + k_{n-1} a^{n-1} \mid k_i \in K \ \right\}$$

Clearly addition in both fields is "the same". Remarkably, multiplication is also "the same". A special case will make this clear. Let us compare the operation of multiplying an arbitrary element of the abstract field by X and the operation of multiplying an arbitrary element of the root field by a. We then get two expressions

$$k_0X + k_1X^2 + \ldots + k_{n-1}X^n$$

and

$$k_0a + k_1a^2 + \ldots + k_{n-1}a^n$$

The trouble is that the first expression contains X^n and the second contains a^n. In the abstract case, $P(X) = X^n + p_1X^{n-1} + \ldots + p_n \in \mathcal{J}$ so we can write

$$X_n = -p_n - p_{n-1}X - \ldots - p_1X^{n-1}$$

In the second case, a has minimal polynomial $P(X)$, so $P(a) = a^n + p_1a^{n-1} + \ldots + p_n$ and we can write

$$a^n = -p_n - p_{n-1}a - \ldots - p_1a^{n-1}$$

These formulas allow us to write the product as a linear combination of $1, X, \ldots, X^{n-1}$ in the first case, and as a linear combination of $1, a, \ldots, a^{n-1}$ in the second case, and we get completely analogous results in the two cases.

It remains to show that the polynomial P has a root in the abstract case. Indeed $X \in L$, so we can insert this element into P to obtain $P(X)$. Since $P(X) \in \mathcal{J}$, the result is zero and X is a root of P.

So we have constructed an abstract field L, with $K \subset L$ such that P has a root in L.

Clearly this abstract field is isomorphic to the concrete root field $K(a)$ we constructed in a previous section. This isomorphism deserves a separate section.

Optional remark: Some people find the proof that X is a root of P in L unconvincing. These people have a point because our notation is ambiguous.

To clarify the argument, write the initial polynomial as $P(Y)$ and let the initial integral domain be $K[X]$. Here X and Y are two abstract symbols. Introduce the quotient field $K[X]/\mathcal{J}$. Distinguish elements of $K[X]$ and elements of the quotient field by writing

$$\overline{Q(X)}$$

to indicate the element of the quotient field represented by a polynomial $Q(X) \in K[X]$.

We want to form $P(\overline{X})$ by substituting $\overline{X}$ for Y in the formula for $P(Y)$. Clearly this can be done using representatives, so $P(X)$ represents the element $P(\overline{X}) \in L$.

But $P(X) \in \mathcal{J}$, so $P(X)$ represents zero, and thus

$$P(\overline{X}) = 0$$

(These overline symbols rapidly become distracting, so I urge the reader to adopt the more relaxed treatment in these notes.)

4.9 Isomorphism and Uniqueness

If K and L are fields, an *isomorphism* $\varphi : K \rightarrow L$ if a one-to-one and onto map which preserves addition and multiplication. That is

- $\varphi(x+y) = \varphi(x) + \varphi(y)$
- $\varphi(xy) = \varphi(x)\varphi(y)$

It is easy to prove that such a map preserves 0, 1, subtraction, and division.

Suppose K is a subfield of a universal field, which we suppose is the complex numbers. Suppose $P(x)$ is a polynomial irreducible over K. Since the complex numbers are algebraically closed, we can find a complex root a of $P(x)$, and form $K(a)$. We can also construct the abstract field L of the previous section. Comparing sections 12 and 14, it is clear that we have proved

Theorem 13 *There is a unique isomorphism $\varphi : L \rightarrow K(a)$ which is the identity on K and takes $X \in L$ to a.*

This theorem has a very important corollary:

Corollary 2 *Suppose K is a subfield of the complex numbers, $P(x)$ is a polynomial irreducible over K, and a and b are complex roots of P. Then there is a unique isomorphism $K(a) \rightarrow K(b)$ which is the identity on K and maps a to b.*

For example, consider $P(x) = x^2 - 2$, which has $\pm\sqrt{2}$ as roots. In this case, $K(\sqrt{2})$ and $K(-\sqrt{2})$ contain the same elements, but the isomorphism sends $k_1 + k_2\sqrt{2}$ to $k_1 - k_2\sqrt{2}$.

Consider $P(x) = x^3 - 2$. If $\theta = e^{\frac{2\pi i}{3}}$, the roots are $a = \sqrt[3]{2}, b = \sqrt[3]{2}\theta$, and $c = \sqrt[3]{2}\theta^2$, and the fields $K(a), K(b), K(c)$ do not contain the same elements. Geometrically as subsets of the plane, these fields look different; the first is a subset of the real numbers, while the other two are spread out across the entire plane. But as abstract fields, they are isomorphic. The reader might like to check this directly.

4.10 Putting It All Together

It is useful to reread sections ten through fifteen, because they illustrate two ways of thinking about fields which we will employ often. In the first few sections, we thought of fields contained in the complex numbers. Our polynomials often had rational coefficients, and we were particularly interested in fields $Q \subset K \subset C$ such that K could be obtained from Q by adding a finite number of algebraic elements.

But we might also be interested in extensions of Z_p. In this case, we haven't produced an analogue of the complex numbers. It is possible to prove that any field can be extended to an algebraically closed field, but the proof involves the axiom of choice and can be very abstract. We prefer to avoid this approach, and section fourteen allows us to do that.

Here is a simple example to make this concrete. Consider the field Z_2. It turns out that this field is contained in larger finite fields, and the number of elements in each such field is a power of two. Let us construct the simplest, a field with four elements.

To use the technique of section fourteen, we need to start with an irreducible polynomial $P(x)$. Notice that when this P is quadratic, the resulting $L = \{ \, r_0 + r_1 a \, \}$ where the $r_i \in Z_2$, and thus has four elements. So we must find an irreducible monic polynomial of degree two.

The complete list of such polynomaials is

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

The first three are reducible, but the last isn't. Let $P(x) = x^2 + x + 1$. Then we obtain a field with four elements, represented by the four polynomials 0, 1, x, and $1 + x$. It is trivial to write down an addition table. In the multiplication table, 0 times anything is zero and 1 times anything is that thing, and the only non-trivial products are

- $x \circ x = x^2 = x + 1$
- $x \circ 1 + x = x + x^2 = 1$
- $1 + x \circ 1 + x = x^2 + 1 = x$

For example, $x \circ x = x^2$, but since $x^2 + x + 1 = 0$ we have $x^2 = -x - 1$ and since $-1 = 1$ in Z_2, this equals $x + 1$.

Chapter 5

Splitting Fields

5.1 Factoring P

Recall that if a polynomial $P(x)$ has a root a, then it can be factored as $P(x) = (x-a)Q(x)$. Indeed, dividing by $x-a$ gives $P(x) = (x-a)Q(x)+R(x)$ where R has smaller degree than $x-a$ and thus is a constant. Setting $x = a$, we discover that the constant is zero.

In the previous sections, we started with an irreducible P over K, and constructed a field $K(a)$ containing a root a of P. Thus P factors over $K(a)$, but the nature of this factorization is unpredictable. In some cases, P suddenly factors completely, while in others the term $x-a$ factors out and the remaining polynomial $Q(x)$ is irreducible.

For example, consider $x^3 - 2$ over Q. One extension field is generated by $\sqrt[3]{(2)}$, but the two remaining roots are complex and thus not in this field. So $P(x)$ factors in $Q(\sqrt[3]{2})$ into

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

On the other hand, consider $x^5 - 1 = (x-1)(x^4 + x^3 + x^2 + x + 1)$. Ignoring the trivial factor $x-1$, we let $P(x) = x^4 + x^3 + x^2 + x + 1$. Later on we prove that this polynomial is irreducible. Notice that the complex roots of this polynomial are $\theta = e^{\frac{2\pi i}{5}}, \theta^2, \theta^3$, and θ^4. Thus P factors completely in $Q(\theta)$ as

$$(x-\theta)(x-\theta^2)(x-\theta^3)(x-\theta^4)$$

5.2 The Splitting Field

Definition 2 *Let K be a field and $P(x)$ a polynomial over K, not necessarily irreducible. An extension field L of K is called a* splitting field of $P(x)$ over K *if $P(x)$ factors completely* *into linear terms over L, so $P(x) = (x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n)$ and if L is the smallest subfield of L containing K and the α_i.*

Suppose as before that K is a subfield of the complex numbers and $P(x)$ is an irreducible polynomial over K. Let $a_1, \ldots, a_n$ be the complex roots of P, which exist by the fundamental theorem of algebra. Then $K(a_1, a_2, \ldots, a_n)$ is clearly a splitting field, and is contained in C.

Such a field exists because it can be defined as the intersection of all subfields containing these elements. Clearly this $K(a_1, a_2, \ldots, a_n)$ is unique.

Theorem 14 *Every element of $L = K(a_1, a_2, \ldots, a_n)$ is algebraic over K. In particular, if $P(x)$ is irreducible of degree n, then*

$$n \leq [L : K] \leq n!$$

Proof: We can construct $K(a_1, \ldots, a_n)$ as a chain of root fields, each of finite degree over K:

$$K \subset K(a_1) \subset K(a_1, a_2) \subset \ldots \subset K(a_1, a_2, \ldots, a_n)$$

so $K(a_1, \ldots, a_n)$ has finite degree over K. It immediately follows that each element is algebraic.

If $P(x)$ is irreducible, $[K(a_1) : K] = n$, so the first inequality is clear.

Factor $P(x)$ over $K(a_1)$ and write $P(x) = (x - a_1)Q(x)R(x)$ where a_2 is a root of the irreducible $Q(x)$ over $K(a_1)$. The degree of $Q(x)$ is at most $n-1$, so $[K(a_1, a_2) : K(a_1)]$ is at most $n-1$. So $[K(a_1, a_2) : K] = [K(a_1, a_2) : K(a_1)][K(a_1) : K]$ is at most $n(n-1)$.

Continue.

Remark: We have much less control over the splitting field than we had over the root field. For the root field, we know the exact degree over K and the exact structure. For the splitting field, we only know the degree within broad bounds, and we do not know the algebraic structure, or a basis over K.

We can also construct the splitting field in the situation when there is no large containing field like the complex numbers. Let $Q_1(x)$ be an irreducible factor of $P(x)$ over K, and construct the root field of Q_1 as in section fourteen. Call it K_1. It contains a root of $Q_1(X)$ and thus a root of $P(X)$. Earlier we called this root $\overline{X}$. But now the actual form of the elements of K_1 is irrelevant, so just call the root a_1.

Factor $P(x)$ over K_1. One term is $x - a_1$, and there may be other linear factors corresponding to other roots of $P(x)$ which are already in K_1. If all factors are linear, we are done. Otherwise let $Q_2(x)$ be an irreducible factor of $P(x)$ over K_1. Construct the root field of $Q_2(x)$ over K_1 and call it K_2. Then K_2 contains an additional root of $P(x)$ not in K_1. Call it a_2.

Continue in this manner until a field is constructed containing all roots of the original $P(x)$.

Theorem 15 *The splitting field is unique up to isomorphism.*

Proof: This is obvious in the situation when we have an enclosing field like C. But a proof is required if we use the abstract construction to obtain a splitting field, because the order in which we add roots isn't clear.

Our proof will gradually construct the required isomorphism. We start with two roots θ_1 and θ_2. By previous results, the root fields $K(\theta_1)$ and $K(\theta_2)$ are isomorphic. We extend this isomorphism to larger and larger subfields until we get an isomorphism between the complete splitting fields.

Our construction still gives something interesting when there is an enclosing field C, namely an isomorphism of the splitting field to itself mapping θ_1 to θ_2. An isomorphism from a field to itself is called an *automorphism.* Automorphisms are the central tool in Galois theory, and our proof of the above theorem will become the main tool used to construct them.

Because of the importance of this proof, we will give it in an unusual way. We give the general proof in the next section. In the following section, we go through the proof again, this time studying the special case $P(X) = X^3 - 2$, to make certain the central ideas are clear.

5.3 Proof that Splitting Fields Are Unique

Suppose we have two splitting fields L_1 and L_2 of a polynomial $P(x)$ over K. If $P(x)$ factors into linear terms over K, then $L_1 = L_2 = K$ and we are done.

Otherwise select an irreducible factor $Q(x)$ of $P(x)$ over K of degree at least two. Both L_1 and L_2 contain all roots of $P(x)$; choose one such root in each, calling them a_1 and b_1. Let $K(a_1) \subset L_1$ and $K(b_1) \subset L_2$ be the root fields, that is, the smallest subfields containing K and either a_1 or b_1. By previous results, these fields are isomorphic by an isomorphism which is the identity on K and takes a_1 to b_1, because both fields are isomorphic to the "abstract root field" constructed in section fourteen. Let $\varphi : K(a_1) \to K(b_1)$ be such an isomorphism.

Factor $P(x)$ over $K(a_1)$. The factors are polynomials with coefficients in $K(a_1)$, and consequently φ can be extended to this polynomial product and gives a corresponding factorization of $P(x)$ over $K(b_1)$. These factors are still irreducible, for otherwise we could find an additional factorization over $K(b_1)$ and then pull it back via φ^{-1} to a factorization over $K(a_1)$.

If all of the factors are linear, then all roots of $P(x)$ belong to $K(a_1)$ and $K(b_1)$, so $L_1 = K(a_1)$ and $L_2 = K(b_1)$ and they are isomorphic.

Otherwise let $Q_1(x)$ be an irreducible non-linear factor of $P(x)$ over $K(a_1)$ and let $Q_2(x)$ be the corresponding irreducible non-linear factor of $P(x)$ over $K(b_1)$. Since Q_1 is a factor of P and P factors completely in L_1, there is a root a_2 of Q_1 in L_1. Similarly there is a root b_2 of Q_2 in L_2. Form the root fields $K(a_1) \subset K(a_1, a_2) \subset L_1$ and $K(b_1) \subset K(b_1, b_2) \subset L_2$.

We now claim that we can extend φ to an isomorphism $\varphi : K(a_1, a_2) \rightarrow K(b_1, b_2)$ which is the identity on K and takes a_1 to b_1 and a_2 to b_2. If so, the rest of the proof will be obvious, for we can again factor $P(x)$ over $K(a_1, a_2)$ and proceed as before. Eventually, we will have isomorphic $K(a_1, a_2, \ldots, a_k)$ and $K(b_1, b_2, \ldots, b_k)$ over which P factors into linear terms. It will follow that the first is L_1 and the second is L_2 and they are isomorphic.

So we need a lemma. But this lemma will be applied again and again in later sections, so we promote it to a theorem. QED modulo:

Theorem 16 *Let $\varphi : K_1 \rightarrow K_2$ be an isomorphism, and let $P_1(X)$ be irreducible over K_1 and $P_2(X)$ be the image of this polynomial over K_2, obviously still irreducible.*

Suppose $K_1 \subset L_1$ and suppose $a \in L_1$ is a root of $P_1(X)$. Suppose $K_2 \subset L_2$ and $b \in L_2$ is a root of $P_2(X)$. Form $K_1 \subset K_1(a) \subset L_1$ and $K_2 \subset K_2(b) \subset L_2$.

Then φ extends to an isomorphism $K_1(a) \rightarrow K_2(b)$ mapping a to b.

Proof: A typical element of $K_1(a)$ is $k_0 + k_1 a + k_2 a^2 + \ldots + k_{n-1} a^{n-1}$ where the $k_i \in K_1$ and n is the degree of $P_1(X)$. A typical element of $K_2(b)$ is $k_0 + k_1 b + k_2 b^2 + \ldots + k_{n-1} b^{n-1}$ where the $k_i \in K_2$ and n is the degree of $P_2(X)$. Define our extended isomorphism to be

$$k_0 + k_1 a + \ldots + k_{n-1} a^{n-1} \rightarrow \varphi(k_0) + \varphi(k_1) b + \ldots + \varphi(k_{n-1}) b^{n-1}$$

This map is clearly one-to-one and onto, so it suffices to prove that it respects multiplication. But the multiplication rule in $K_1(a)$ is derived from the polynomial $P_1(X) = X^n + r_1 X^{n-1} + \ldots + r_n$ by repeated applications of the rule

$$a^n = -r_1 a^{n-1} - \ldots - r_n$$

and the multiplication rule in $K_2(X)$ is derived from the polynomial $P_2(X) = \varphi(P_1(X)) = X^n + \varphi(r_1) X^{n-1} + \ldots + \varphi(r_n)$ by repeated applications of the rule

$$b^n = -\varphi(r_1) b^{n-1} - \ldots - \varphi(r_n)$$

and our extended isomorphism maps the first of these to the second. QED.

Corollary 3 *Let L be the splitting field of an irreducible polynomial $P(X)$, and let θ_1 and θ_2 be roots of P in L. Then there is an automorphism φ of L mapping θ_1 to θ_2.*

Proof: This follows from the proof of theorem 14, since we began the proof by constructing an isomorphism of root fields $K(\theta_1) \to K(\theta_2)$ and ended by extending it to an automorphism of L.

5.4 Uniqueness of Splitting Fields, and $P(X) = X^3 - 2$

In this section we repeat the proof in the previous section, this time using the specific example $P(X) = X^3 - 2$ over the base field Q.

It is convenient to simplify the notation before starting the proof. Let $\alpha = \sqrt[3]{2}$ and $\omega = e^{\frac{2\pi i}{3}} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. Here α is real. The roots of our equation are $\theta_1, \theta_2, \theta_3 = \alpha, \omega\alpha$, and $\omega^2\alpha$.

Let's try to guess a basis for the splitting field. This field must contain $\alpha, \omega\alpha$, and $\omega^2\alpha$. It must contain quotients, so it also contains ω and ω^2. It must be closed under multiplication, so it contains powers of α. Since $\alpha^3 = 2$, we only need α and α^2. This leads to an initial guess that a basis is $\{1, \omega, \omega^2, \alpha, \omega\alpha, \omega^2\alpha, \alpha^2, \omega\alpha^2, \omega^2\alpha^2\}$.

However, $\omega^3 - 1 = 0 = (\omega - 1)(\omega^2 + \omega + 1)$ and so $\omega^2 = -1 - \omega$. Removing dependencies, a better guess is that a basis is $\{\ 1, \alpha, \alpha^2, \omega, \omega\alpha, \omega\alpha^2\ \}$. This indeed turns out to be a basis. Without giving details, we could prove it by showing that these elements are linearly independent, and that the set of linear combinations is closed under multiplication by each basis element. Division then takes care of itself by an earlier argument.

We won't use this guess because it follows from our more general considerations below.

Our P is already irreducible over Q. Let us arbitrarily select $\theta_1 = \alpha$ and $\theta_2 = \omega\alpha$. Form the root fields

$$Q(\theta_1) = \{\ q_0 + q_1\alpha + q_2\alpha^2\ \}$$

$$Q(\theta_2) = \{\ q_1 + q_1\omega\alpha + q_2\omega^2\alpha^2\ \}$$

They are isomorphic by a map φ fixing rational numbers and sending α to $\omega\alpha$.

Now factor $P(X)$ over $Q(\theta_1)$:

$$X^3 - 2 = (X - \alpha)(X^2 + \alpha X + \alpha^2)$$

The splitting field L_1 can be obtained from $Q(\theta_1)$ by adding a root of the quadratic term, which is irreducible over $Q(\theta_1)$ because its roots are complex and $Q(\theta_1)$ only contains real numbers. Once we add one such root, the polynomial $P(X)$ splits completely over L_1.

Since we know all roots of $X^3 - 2$, it is easy to guess the factorization of the quadratic term:

$$X^2 + \alpha X + \alpha^2 = (X - \omega\alpha)(X - \omega^2\alpha)$$

This is easily checked by direct multiplication.

Let us arbitrarily pick one of the roots of the quadratic term, say $\omega\alpha$. We can then write down the full splitting field starting with $Q(\theta_1)$ using our standard root construction:

$$Q(\alpha, \omega\alpha) = \{ \ (q_0 + q_1\alpha + q_2\alpha^2) + (q_3 + q_4\alpha + q_5\alpha^2)\omega\alpha \ \}$$

We now repeat this construction over $Q(\theta_2)$. Notice that our factorization of $X^3 - 2$ does not make sense over $Q(\theta_2)$; for instance example, $\alpha \notin Q(\theta_2)$. But φ maps $Q(\theta_1)$ to $Q(\theta_2)$, so it maps polynomials over $Q(\theta_1)$ to polynomials over $Q(\theta_2)$. Consequently, the corresponding factorization over $Q(\theta_2)$ is

$$X^3 - 2 = (X - \varphi(\alpha))(X^2 + \varphi(\alpha)X + \varphi(\alpha^2)) = (X - \omega\alpha)(X^2 + \omega\alpha + \omega^2\alpha^2)$$

Therefore, another way to get the full splitting field is to start with $Q(\theta_2)$ and add a root of $X^2 + \omega\alpha + \omega^2\alpha^2$.

It is easy to find the desired roots without resorting to the quadratic formula because we know all the roots of $X^3 - 2$; the roots of $X^2 + \alpha X + \alpha^2$ over $Q(\theta_2)$ are α and $\omega^2\alpha$, as is easily checked. Let us arbitrarily pick the root $\omega^2\alpha$. Then our standard root construction gives another version of the splitting field

$$Q(\omega\alpha, \omega^2\alpha) = \{ \ (q_0 + q_1\omega\alpha + q_2\omega^2\alpha^2) + (q_3 + q_4\omega\alpha + q_5\omega^2\alpha^2)\omega^2\alpha \ \}$$

The map φ was originally chosen to map $\alpha \to \omega\alpha$. We extend it to also map the root $\omega\alpha$ of $X^2 + \alpha X + \alpha^2$ to the root $\omega^2\alpha$ of $X^2 + \omega\alpha + \omega^2\alpha^2$. In the end, we obtain an isomorphism between our two representations of the splitting field of $X^3 - 2$:

$$(q_0 + q_1\alpha + q_2\alpha^2) + (q_3 + q_4\alpha + q_5\alpha^2)\omega\alpha \to (q_0 + q_1\omega\alpha + q_2\omega^2\alpha^2) + (q_3 + q_4\omega\alpha + q_5\omega^2\alpha^2)\omega^2\alpha$$

This isomorphism was our goal. But the previous formula gives more. It is easy to see directly that our two forms of the splitting field can be rewritten to have the basis we guessed originally: $\{ \ 1, \alpha, \alpha^2, \omega, \omega\alpha, \omega\alpha^2 \ \}$. After doing this rewritting, our isomorphism becomes an automorphism of this field mapping α to $\omega\alpha$.

Indeed, we can rewrite the isomorphism to get

$$q_0 + q_1\alpha + q_2\alpha^2 + g_3\omega\alpha + g_4\omega\alpha^2 + q_5 2\omega \to q_0 + q_1\omega\alpha + q_2(-\alpha^2 - \omega\alpha^2) + q_3(-\alpha - \omega\alpha) + q_4\alpha^2 + q_5 2\omega$$

and so our basis vectors map as follows

- $1 \to 1$
- $\alpha \to \omega\alpha$
- $\alpha^2 \to -\alpha^2 - \omega\alpha^2 = \omega^2\alpha$
- $\omega \to \omega$
- $\omega\alpha \to -\alpha - \omega\alpha = \omega^2\alpha$
- $\omega\alpha^2 \to \alpha^2$

Notice that this isomorphism permutes the roots of $X^3 - 2$; indeed

$$\alpha \to \omega\alpha \to \omega^2\alpha \to \alpha$$

Final Remark: Our proof that splitting fields are unique ends up finding a non-trivial isomorphism of the splitting field L for $X^3 - 2$. We made several random choices during the argument. Changing these choices would yield additional automorphisms of L. Indeed, a careful inventory of all the choices would give the complete automorphism group of L. We do this more generally in later sections.

Chapter 6

Finite Fields

6.1 Finite Fields

At this point, our field theory consists of three central ideas: the degree of an extension, the root field of an irreducible polynomial, and the splitting field of an arbitrary polynomial. In this intermission, we use this results to obtain a complete theory of finite fields.

If F is a finite field, then F contains a smallest subfield $Z_p \subset F$. Then F is a vector space over Z_p of some finite dimension n. It immediately follows that F contains p^n elements since every element has the form $a_1 f_1 + \ldots + a_n f_n$ for coefficients $a_i \in Z_p$ and a basis $f_1, \ldots f_n$. So

Theorem 17 *Every finite field has order p^n for some prime p and positive integer n.*

Remark: Thus there might be fields of orders $2, 3, 4, 5, 7, 8, 9, 11, 13, 16$, but certainly cannot be fields of orders $6, 10, 12, 14, 15$.

Theorem 18 *If p is a prime and n is a positive integer, there is a field of order p^n. This field is unique up to isomorphism.*

Proof: The nonzero elements of a field F of order p^n form a multiplicative subgroup of order $p^n - 1$, so every element except the identity satisfies $x^{p^n-1} - 1 = 0$. Multiplying by x, every element of the field would satisfy $P(x) = x^{p^n} - x = 0$.

Let F be a splitting field for this polynomial over Z_p. Factor the polynomial. If two terms are equal, then $P(x) = (x-a)^2 Q(z) = x^{p^n} - x$. Formally differentiating both sides,

$$2(x-a)Q(x) + (x-a)^2 Q'(x) = p^n x^{p^n-1} - 1 = -1$$

Substituting $x = a$ gives $0 = -1$, which is impossible. So all roots are distinct, and there are p^n roots.

On the other hand, the roots themselves form a field, and so the splitting field contains only roots and thus is a field with p^n elements. Indeed 0 and 1 are roots. If a and b are roots, then $(ab)^{p^n} = a^{p^n} b^{p^n} = ab$, so products of roots are roots. Also $(a+b)^p = a^p + b^p$ because all remaining binomial coefficients are divisible by p. So $(a+b)^{p^2} = ((a+b)^p)^p = (a^p + b^p)^p = a^{p^2} + b^{p^2}$, and so forth, so $(a+b)^{p^n} = a^{p^n} + b^{p^n} = a + b$. So sums of roots are roots. If a is a root, then $(-a)^{p^n} = (-1)^{p^n} a^{p^n} = (-1)^{p^n} a = -a$ is p is odd, and $-1 = 1$ if p is even, so additive inverses of roots are roots. So the set of all roots is a finite integral domain, and thus automatically closed under inverses and a field.

In short, the splitting field is a finite field of order p^n. Since the splitting field is unique up to isomorphism, a field of order p^n is unique up to isomorphism.

Theorem 19 *The multiplicative group of a finite field is cyclic and thus has a generator. More generally, any finite multiplicative subgroup of a field is cyclic.*

Example: Consider the field of order four described on page 20. The nonzero elements are $x, x^2 = x + 1$, and $x^3 = x^2 + x = 1$.

Proof: Many proofs are known. Here is one I like. Consider first the special case $Z_2 + Z_4$. We claim this cannot be a multiplicative subgroup of a field. For if it were, then every element would satisfy $x^4 - 1 = 0$, but this polynomial has at most 4 roots, not eight.

In general suppose G is a finite subgroup of a field. Any finite abelian group is a sum of groups of the form Z_{p^k}. Let $p_1, \ldots, p_s$ be the primes for G and let their highest order subgroups have orders $p_1^{k_1}, \ldots, p_s^{k_s}$. Let $m = p_1^{k^1} \cdot \ldots \cdot p_s^{k_x}$. Note that every element of G satisfies $x^m = 1$ and there are at most m roots of this polynomial. Consequently the $Z_{p_i^{s_i}}$ are the only groups involved, and there are no other summands with p_i of the same or lower order. But the sum of cyclic groups of prime power order, for *different* primes, is itself cyclic. QED.

Chapter 7

Beginning Galois Theory

7.1 Motivation for Galois Theory

Ruffini is the first mathematician to prove that the quadratic formula cannot be generalized to equations of degree five or higher, but his proof had a gap. This gap was filled by Abel, who gave the first complete proof. Galois gave a deeper proof by bringing ideas hidden in Abel's argument to the foreground. I don't know if Galois knew about Abel's or Ruffini's work.

Galois succeeded in finding a necessary and sufficient condition that a polynomial $P(x)$ be solvable by radicals. To do this, he had to find a deep property that is true of some polynomials but not others. In our previous work, with each irreducible $P(x)$ we associated two fields, the root field $K(\theta)$ and the splitting field $K(\theta_1, \ldots, \theta_n)$. All root fields had the same concrete structure, so it is unlikely that they reveal the required deep property of P. But splitting fields are not all the same. Sometimes adding one root gives all roots, while other times adding a root still leaves us far from the complete field. It turns out that the idea we need is hidden in the construction of the splitting field.

At the end of the previous chapter, we constructed the splitting field L of $X^3 - 2$ and in the process found an automorphism of this field mapping the root α to the root $\omega\alpha$. At several points in the argument, we made arbitrary choices. Changing these choices would have led to different automorphisms. In this particular case, there are six possible automorphisms.

If $K \subset L$, the automorphisms of L fixing each point of K form a group G, now called the *Galois group.* As it turns out, this group codifies the secrets of the splitting field. Galois proved that knowledge of the group completely determines whether or not the polynomial can be solved with radicals.

The mathematicians Ruffini, Abel, and Galois did not immediately introduce the Galois group. They were led to it by other considerations. It is useful to try to reconstruct their line of reasoning and we'll do that next.

Suppose we have an irreducible polynomial $P(X)$ with rational coefficients. By the fundamental theorem of algebra, this polynomial can be factored over the complex numbers. Let $\theta_1, \theta_2, \ldots, \theta_n$ be its roots. We can calculate these numerically. For instance, perhaps $\theta_1 = 0.43895 + 2.71834\ i$.

We only know approximate values for the roots. But the θ_i are far from arbitrary complex numbers because $\theta_1 + \ldots + \theta_n$ and $\theta_1\theta_2 \ldots \theta_n$ are rational. More generally

$$(X - \theta_1)(X - \theta_2) \ldots (X - \theta_n) = X^n - (\sum_i \theta_i)X^{n-1} + (\sum_{i<j} \theta_i\theta_j)X^{n-2} + \ldots \pm (\theta_1\theta_2 \cdot \theta_n)$$

and all of these coefficients are rational.

We can construct a large number of additional relations as follows. A polynomial $S(X_1, \ldots, X_n)$ is said to be *symmetric* if any permutation of the X_i gives the same polynomial. For instance, the following polynomials, known as the *elementary symmetric polynomials*, are symmetric:

- $\sigma_1 = X_1 + X_2 + \ldots + X_n$
- $\sigma_2 = X_1X_2 + X_1X_3 + \ldots + X_{n-1}X_n$
- $\sigma_i = \ldots$
- $\sigma_n = X_1X_2 \ldots X_n$

A key theorem about symmetric polynonials asserts that any symmetric polynomial can be written uniquely as a polynomial in the σ_i:

$$S(X_1, \ldots, X_n) = Q(\sigma_1, \ldots, \sigma_n)$$

If S has rational coefficients, so does Q. If S has integer coefficients, so does Q.

It follows that when S is *any* symmetric polynomial with rational coefficients, the expression $S(\theta_1, \ldots, \theta_n)$ is rational. This gives a gigantic number of relations satisfied by the roots of P. For example, $\theta_1^2 + \theta_2^2 + \ldots + \theta_n^2$ is rational.

For some polynomials, there are additional relations that arise for a completely different reason. Consider, for example, $X^7 - 1$. Factoring, we have

$$X^7 - 1 = (X - 1)(X^6 + X^5 + X^4 + X^3 + X^2 + X + 1)$$

We'll later show that the second factor is irreducible. The roots of the second factor are generated by $\theta = e^{\frac{2\pi i}{7}}$ and equal $\theta, \theta^2, \theta^3$, θ^4, θ^5 and θ^6.

These roots satisfy new relations, including

- $\theta_1^2 - \theta_2 = 0$
- $\theta_1^3 - \theta_3 = 0$
- $\theta_1^4 - \theta_4 = 0$
- $\theta_1^5 - \theta_5 = 0$
- $\theta_1^6 - \theta_6 = 0$

The new relations don't come from symmetric polynomials because they aren't invariant under all permutations of the roots. For instance, if we leave θ_1 fixed and interchange θ_2 and θ_3, the resulting equations are no longer all true.

On the other hand, some permutations leave these relations unchanged. Indeed, we can map θ_1 to any desired root, but if we want to retain the previous relations, then the remaining roots must map as in the rows of the following table:

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
θ_2	θ_4	θ_6	θ_1	θ_3	θ_5
θ_3	θ_6	θ_2	θ_5	θ_1	θ_4
θ_4	θ_1	θ_5	θ_2	θ_6	θ_3
θ_5	θ_3	θ_1	θ_6	θ_4	θ_2
θ_6	θ_5	θ_4	θ_3	θ_2	θ_1

For example, if we map $\theta_1 \to \theta_2$ and we want the equation $\theta_1^2 = \theta_2$ to be invariant, then the equation will become $\theta_2^2 = ?$ where ? is the image of θ_2, so we must map $\theta_2 \to \theta_4$.

These symmetries of the roots are *transitive*; any root can be taken to any other root by a symmetry. It is tempting to say "of course, since these are seventh roots of unity, and these occur at the vertices of a regular 7-gon, which can be rotated at will." *However!* When we factored, we removed 1 as a root, so geometrically there are no symmetries. The above symmetries are *algebraic*.

Each map from the top line to another line gives a permutation of the roots which preserves the equations listed earlier. Notice that the resulting six permutations form a group. The permutation $\theta_1 \to \theta_2$ represented by the second line of the matrix has order three in the group, for applying it three times will send $\theta_1 \to \theta_2 \to \theta_4 \to \theta_1$. On the other hand, the permutation sending $\theta_1 \to \theta_3$ has order six, since applying it six times sends $\theta_1 \to \theta_3 \to \theta_2 \to \theta_6 \to \theta_4 \to \theta_5 \to \theta_1$.

So surprisingly, the algebraic symmetries of the roots of $X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$ form the group Z_6, not at all like the full group S_6 of all permutations of the roots. Indeed Z_6 has order 6, while S_6 has order 720.

7.2 Motivation; Putting the Ideas Together

How, then, are the notions of splitting fields, relations among roots, and automorphisms related?

Suppose $P(X)$ has rational coefficients. The splitting field L consists of all sums and products of the roots θ_i. Therefore if $Q(X_1, X_2, \ldots, X_n)$ is a polynomial with rational coefficients, $Q(\theta_1, \theta_2, \ldots, \theta_n)$ is an element of L. Just as the root field is modeled by polynomials in $Q[X]$, the splitting field is modeled by polynomials in $Q[X_1, X_2, \ldots, X_n]$.

In the root field case, the root θ satisfies $P(\theta) = 0$, so the actual root field is

$$Q[X] / \Big(\text{multiples of } \ P(X)\Big)$$

Similarly let us define $\mathcal{J}$ to be the set of all polynomials $R(X_1, X_2, \ldots, X_n)$ with rational coefficients which satisfy $R(\theta_1, \theta_2, \ldots, \theta_n) = 0$. This is the set of all *relations* satisfied by the roots. It is easy to see that $\mathcal{J}$ is an ideal in the polynomial ring, and the splitting field is

$$L = \frac{Q[X_1, \ldots, X_n]}{\mathcal{J}} = \frac{Q[X_1, \ldots, X_n]}{(\text{ideal of relations } R)}$$

Among these relations are all symmetric polynomials. Galois' initial idea is that equations solvable by radicals will have many, many more relations than just these. Studying the relations is difficult because there are so many. Galois' second idea is that it is better to study the symmetry group G of the relations. By definition, a symmetry of the set of relations is a permutation σ of the X_i such that whenever $R(X_1, \ldots, X_n)$ is a relation, $R(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ is also a relation. As we get more relations, we get fewer symmetries. Often the group G is the full set S_n of all permutations, but for $X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$, it is much smaller, just Z_6. Studying G is much easier than studying the set of relations because G is a finite group.

A permutation of the roots which leaves the relations invariant obviously induces an automorphism of the full L. Conversely, each automorphism of L must send a root of P to another root and thus must permute the roots; to induce an automorphism, this permutation must preserves the relations.

In this way, we are led from the construction of the splitting field to the relations among the roots, and thus to the automorphisms of the splitting field.

7.3 The Galois Group

Definition 3 *Let $K \subset L$ be fields and suppose $[L : K]$ is finite. The Galois group G of this extension is the set of all automorphisms of L which leave all elements of K individually fixed.*

Remark: Unfortunately, this group may give no information. For example, consider $Q \subset Q(\sqrt[3]{2})$. An automorphism must map the polynomial $X^3 - 2$ back to itself, and thus map each root of this polynomial to another root. Since the field only contains the real root, the only automorphism is the identity map.

Theorem 20 *Let $P(X)$ be a polynomial over K, not necessarily irreducible. Let L be a splitting field of P.*

- *Every element of the Galois group of $K \subset L$ permutes the roots of P*
- *This permutation completely determines the element of G*
- *If $P(X)$ is irreducible over K and θ_i and θ_j are two roots, there is an element σ in the Galois group of the splitting field L such that $\sigma(\theta_i) = \theta_j$*

Proof: If $P(\theta) = 0$ and σ is a Galois automorphism, then σ fixes the coefficients of P and therefore $P(\sigma(\theta)) = 0$. Since the roots generate the splitting field, the second assertion is clear.

The third assertion was proved as theorem 13 in section 18 on splitting fields.

Remark: The last theorem of the previous section was the first of a string of theorems showing that non-trivial automorphisms exist. We now come to a somewhat difficult, but extremely important theorem which guarantees the existence of many more automorphisms:

Theorem 21 *Let $P(X)$ be a polynomial over K, not necessarily irreducible. Let L be a splitting field for $P(X)$. Suppose that $P(X)$ does not have multiple roots in L. Then*

$$|G| \geq [L : K]$$

and thus the Galois group has many elements.

Proof: Factor P and let $P_1(X)$ be one of the irreducible factors. This polynomial factors completely in L; call its roots $a_1, \ldots, a_{n_1}$. The root fields $K(a_1), \ldots, K(a_{n_1})$ inside L are isomorphic by a previous result. Let $\varphi_i : K(a_1) \to K(a_i)$ be such isomorphisms over K. In particular, φ_1 is the identity map from $K(a_1)$ to itself. We are going to extend these isomorphisms to automorphisms of L. We know that this will give $[K(a_1) : K]$ automorphisms because all the roots of P_1 are distinct.

Factor $P(X)$ over $K(a_1)$. If P factors completely into linear factors, then $K(a_1) = L$ and consequently $\varphi_i : L \to L$. This gives the same number of automorphisms as $[L : K] = [K(a_1) : K]$ and we are done.

Otherwise factor $P(X)$ over $K(a_1)$ as $Q_1 Q_2 \dots Q_s$ and let $P_2(X)$ be one of the nonlinear factors. We have isomorphisms $\varphi_i : K(a_1) \to K(a_i)$; in particular φ_1 is the identity. Applying these isomorphisms, we discover that $P(X)$ also factors as

$$\varphi_i(Q_1)\varphi_i(Q_2)\dots\varphi_i(Q_s)$$

over $K(a_i)$. In particular, $\varphi_i(P_2(X)$ is one of the factors over $K(a_i)$. For each i, including $i = 1$, let b_{ij} be all roots of $\varphi_i(P_2(X))$ in $K(a_i)$. The extension lemma on page 24 implies that we can uniquely extend the φ_i to isomorphisms $\varphi_{ij} : K(a_1, b_{11}) \to K(a_i, b_{ij})$. Notice that the map $K(a_1, b_{11})$ to itself is the identity.

Now factor $P(X)$ over $K(a_1, b_1)$. Suppose first that all factors are linear. Then all $K(a_i, b_{ij})$ are L and each $\varphi_{ij} : K(a_1, b_1) \to K(a_i, b_{ij})$ is an automorphism $L \to L$. These automorphisms are all different. Indeed, they take a_1 to a_i and the a_i are all different, so two equal automorphisms must be associated with the same i. But then the automorphisms map b_1 to elements b_{ij} for fixed i, and these are roots of $\varphi_{ij}(P_2(X)$, hence roots of $P(X)$, and so all different. The total number of automorphisms is the number of a_i multiplied by the number of b_{ij}, which equals the degree of $P_1(X)$ multiplied by the degree of $P_2(X)$, or $[K(a_1) : K][K(a_1, b_1) : K(a_1)] = [L, K]$.

If the factorization $P(X) = R_1(X)R_2(X)\dots R_p(X)$ of $P(X)$ over $K(a_1, b_1)$ has non-linear terms, fix one and call it $P_3(X)$. Applying the isomorphism $\varphi_{ij} : K(a_1, b_1) \to K(a_i, b_{ij})$, we discover that $P(X)$ also factors as

$$\varphi_{ij}(R_1)\varphi_{ij}(R_2)\dots\varphi_{ij}(R_p)$$

over $K(a_i, b_{ij})$. For each i and j, let c_{ijk} be all roots of $\varphi_{ij}(P_3(X))$ over $K(a_i, b_{ij})$.

Then for fixed i we have an isomorphism carrying a_1 to a_i, and once i is fixed we have a fixed extended isomorphism carrying b_2 to b_{ij} and for fixed i and j we have an isomorphism defined on $K(a_1, b_1, c_1)$ carrying c_1 to c_{ijk}. These isomorphisms extend to automorphisms of L. If two are equal, they have the same i since the a_i are distinct. They then have the same j since for fixed i the b_{ij} are distinct, and they have the same k since for fixed i and j the c_{ijk} are distinct. Counting as before, we have at least $\deg(P_1(X))\deg(P_2(X)\deg(P_3(X))$ automorphisms, and this last number is $K[a_1 : K][K(a_1, b_1) : K(a_1)][K(a_1, b_1, c_1] : K(a_1, b_1)]$. If we stop at this stage, this number is $[L : K]$ and we have enough automorphisms.

Otherwise continue step by step in the same way. QED.

Remark: In the next chapter we will prove that for any finite extension $K \subset L$, we have $|G| \le [L : K]$. Combining this with the previous theorem then gives the exact number of automorphisms for splitting fields of polynomials with no repeated roots.

Chapter 8

Galois Extensions and the Fundamental Theorem

8.1 Galois Extensions

Definition 4 *We say $K \subset L$ is a* Galois extension *is the only elements of L left fixed by every element of the Galois group are elements of K.*

Remark: The Galois theory developed in the next section works only for Galois extensions. So the remaining results in this section are crucial, showing that many important field extensions are Galois extensions.

The next developments come from the paper *On the Characterization of Galois Extensions* by Meinolf Geck, American Mathematical Monthly, August-September, 2014.

Lemma 3 *The Galois group of a finite extension $K \subset L$ is finite.*

Proof: Let $\alpha_1, \ldots, \alpha_n$ be a basis for L over K. An automorphism is completely determined by its action on the α_i, so it suffices to prove that each α_i can have only finitely many images. Let $P(X)$ be the minimal polynomial of α_i over K and notice that the image of α_i must also be a root of this minimal polynomial, which has finitely many roots.

Lemma 4 *If $K \subset L$ is a finite extension, then L is not the union of a finite number of strict subfields $K \subset M \subset L$.*

Proof: If K is infinite, an even stronger statement is true: a finite dimensional vector space over an infinite field cannot be a union of a finite number of proper subspaces. To prove this, pick a basis for L, and write the elements of L as n-tuples. Consider the set S of elements $(1, t, t^2, \ldots, t^{n-1}) \in L$. There are infinitely many such elements.

We can enlarge each subspace M to a subspace of codimension one, and thus to a subspace defined by an equation $a_0x_0 + a_1x_1 + \ldots + a_{n-1}x_{n-1} = 0$. The element $(1, t, \ldots, t^{n-1})$ belongs to this subspace only if $a_0 + a_1t + \ldots + a_{n-1}t^{n-1} = 0$, and this polynomial equation has only finitely many roots. So only finitely many elements of the set S belong to the finite union of subspaces.

If K is finite, we need to restrict to subfields $M \subset L$, but then everything in sight is a finite field. So the multiplicative group $L^\star$ of nonzero elements of L is cyclic. Let $g \in L$ be a generator. This g cannot belong to any strict subfield $M \subset L$. QED.

Corollary 4 *For an arbitrary finite extension $K \subset L$, there is an element $\theta \in L$ moved by every nontrivial element of the Galois group.*

Proof: For each nontrivial $g \in G$, the set M_g of fixed points of g is a proper subfield of L containing K. By the lemma, the union of the subfields is not the entire field.

Corollary 5 *We have $|G| \leq [L : K]$. If equality holds, there exists $\theta \in L$ such that $L = K(\theta)$, and L is a splitting field of the minimal polynomial of θ, which has no multiple roots.*

Proof: Choose θ as in the previous corollary and let $P(X)$ be its minimal polynomial over K. The Galois group must map θ to other roots of $P(X)$. Since no element of the Galois group except the identity leaves θ fixed, the number of distinct roots of $P(X)$ must be at least $|G|$, so the degree of P is at least $|G|$. Therefore

$$|G| \leq \deg(P) = [K(\theta) : K] \leq [L : K]$$

If equality holds, then $|G| = \deg(P)$ and $K(\theta) = L$. In particular, every root of P must equal $\sigma(\theta)$ for some $\sigma \in G$, so P splits and its roots are distinct.

Theorem 22 *Let $K \subset L$ be a field extension of finite degree. The following are equivalent:*

- $|G| = [L : K]$
- *$K \subset L$ is a Galois extension*
- *L is a splitting field over K for a polynomial $P(X)$ without multiple roots.*
- *$L = K(\theta)$ where θ is a root of a polynomial $P(X)$ which splits completely in L and does not have multiple roots.*

If the characteristic of the ground field is zero, we can add a fifth item

- *L is a splitting field over K for an arbitrary polynomial $P(X)$*

Proof of theorem: First a) implies b), for let M be the set of all elements of L fixed by every element of G. Clearly M is a subfield and $K \subset M \subset L$. Hence G is also the Galois group of L over M, and by corollary 5 we have $|G| \leq [L : M] \leq [L : K]$. The left and right sides of this equation are equal by assumption, so $M = K$.

Second, b) implies c) as follows. Find a basis $x_1, \ldots, x_n$ for L over K. For each i satisfying $1 \leq i \leq n$ and each $\sigma \in G$ we can form $\sigma(x_i) \in L$. The set of all these elements is a set which may have duplicates; let S be the same set with duplicates removed.

If we act on the $\sigma(x_i)$ on the left by a fixed $\tau \in G$, we just rearrange the elements. If a particular element occurs r times originally, the image of this element will occur r times. Consequently we can remove duplicates before or after multiplication by τ and get the same result. So the set S is invariant under the action of any fixed τ.

Form the polynomial

$$P(X) = \prod_S (X - \sigma(x_i))$$

The coefficients of this polynomial are left fixed by applying any τ. For example, the first coefficient is the sum of elements in S and the set S is invariant under multiplication by τ. Since $K \subset L$ is a Galois extension and the coefficients of P are invariant under the Galois group, all of these coefficients belong to K.

Since G contains the identity, the set S contains all x_i, so the field generated by the roots of P is L. Since P splits completely, it is a splitting field. By construction, P has no repeated roots.

Third, c) implies a). We already know that $|G| \leq [L : K]$ from Corollary 5. By theorem 20 in section 6.3, c) implies that $|G| \geq [L : K]$. Therefore $|G| = [L : K]$.

Condition d) implies condition c). This condition implies conditions a) and b), and condition a) implies condition d) by corollary 5.

Condition c) implies condition e). Conversely, if the ground field has characteristic zero, condition e) implies condition c) as follows.

If the characteristic of the ground field is zero, suppose $K \subset L$ is a splitting field of an arbitrary $P(X)$. Factor $P(X)$ over K into irreducible polynomials $P_1(X)P_2(X)\ldots P_j(X)$. If an irreducible factor occurs more than once, we can remove the redundant term and get a new polynomial $\tilde{P}$ with the same roots and the same splitting field. So we can assume there are no redundancies.

Fix i and consider $P_i(X)$. If θ is a root in a splitting field L, then P_i is the minimal polynomial of θ. If θ is a multiple root, we can write $P_i(X) = (X - \theta)^2 Q(X)$ over L. Taking formal derivatives, $P_i'(X) = 2(X - \theta)Q(X) + (X - \theta)^2 Q'(X)$ and consequently θ is a root of $P_i'(X)$. Writing $P_i(X) = X^k + a_1X^{k-1} + \ldots$, we have $P_i'(X) = kX^{k-1} + \ldots$. In a field of characteristic zero, $k \neq 0$ and so $P_i'(X)$ has smaller degree than P_i, a contradiction.

A root cannot be a root of two terms P_i and P_j, for otherwise they would both be minimal polynomials of the root, and thus equal up to a constant, so one of them would have been removed when we removed redundant terms. QED.

8.2 Fundamental Theorem of Galois Theory

Let $K \subset L$ be a fixed Galois extension with Galois group G. Suppose that K_1 is a field, $K \subset K_1 \subset L$. Associate this field with a subgroup of G, namely

$$\{g \in G \mid g \text{ fixes each element of } K_1\}$$

Conversely let H be a subgroup of G, $\{e\} \subset H \subset G$. Associate this subgroup with a subfield of L, namely

$$\{\ l \in L \mid h(l) = l \text{ for all } h \in H\}$$

Theorem 23 (The Fundamental Theorem of Galois Theory) *The maps just defined are inverse to each other and set up a one-to-one correspondence between the set of all subgroups of G and the set of all subfields of L containing K.*

Remark: This is astonishing, since fields are complicated and hard to pin down, while subgroups of a finite group can be enumerated mechanically.

Remark: We gather here a few extra features which are part of the theorem and will be proved as we prove the theorem.

- The correspondence is order reversing. The identity subgroup corresponds to L and the entire G corresponds to K.
- Suppose $K \subset M \subset L$ is a subfield of L corresponding to a subgroup H of G. Then the extension $M \subset L$ is a Galois extension, and its Galois group is H. The extension $K \subset M$ is a Galois extension if and only if H is normal, and in that case its Galois group is G/H.

Proof of Fundamental Theorem: Consider the map

$$\{\text{fields}\} \to \{\text{groups}\} \to \{\text{fields}\}$$

Suppose we start with a subfield M on the left. Since $K \subset L$ is a Galois extension, it is a splitting field for a polynomial $P(X)$ over K without multiple roots in L. This $P(X)$ certainly has coefficients in M, so L is the splitting field of P over M and there are no multiple roots. Hence $M \subset L$ is Galois.

The Galois group of $M \subset L$ is the set of all automorphisms of L which fix M. Such an automorphism certainly fixes K, so it belongs to the Galois group of $K \subset L$. Thus it is $\{ \sigma \in G \mid \sigma \text{ fixes M} \}$.

This is exactly the group H assigned to M by the Galois correspondence. The corresponding subfield attached to this group is the set of all elements of K fixed by the elements of H. Since $M \subset L$ is Galois, this set is M itself.

Consider the map

$$\{\text{groups}\} \to \{\text{fields}\} \to \{\text{groups}\}$$

Start with a subgroup H and consider the set M of all elements fixed by H. This $M \subset L$ is a Galois extension as above. It's Galois group is the set N of all automorphisms of L fixing M, and this is the group assigned to M on the right of the above sequence. So $|N| = [L : M]$. Notice that $H \subset N$.

By corollary 5 in section 7.1, there is a $\theta \in L$ such that $L = M(\theta)$. Define

$$P_1(X) = \prod_{\sigma \in H} \Big(X - \sigma(\theta) \Big)$$

By the standard argument, the coefficients of this polynomial are invariant under H and thus belong to M. Since θ is a root of $P_1(X)$, the minimal polynomial of θ over M divides $P_1(X)$ and thus has degree at most $|H|$. But $[L : M]$ equals the degree of this minimal polynomial, so $[L : M] \le |H|$. Therefore $|N| \le |H|$. Since $H \subset N$, we have $H = N$. QED.

Theorem 24 *If $K \subset L$ is a Galois extension and $K \subset M \subset L$ corresponds to a subgroup $H \subset G$, then $K \subset M$ is Galois if and only if H is normal in G, and in that case the Galois group of $K \subset M$ is G/H.*

Proof: Let $\sigma \in G$. If $K \subset M \subset L$ corresponds to $H \subset G$, then $\sigma(M)$ corresponds to $\sigma H \sigma^{-1}$, since an element of this group maps $\sigma(M)$ to M, and then leaves M fixed, and then maps it back to $\sigma(M)$. It follows that H is normal in G if an only if $\sigma(M) \subset M$ for all $\sigma \in G$.

Suppose H is normal, and thus $\sigma(M) \subset M$. It follows that every automorphism in G induces an automorphism of M. Two automorphisms σ_1 and σ_2 induce the same automorphism of M if and only if $\sigma_1^{-1}\sigma_2$ is the identity on M, or equivalently if and only if $\sigma_1^{-1}\sigma_2 \in H$, so the group G/H acts effectively on M. Notice that $|G| = |H||G/H|$ and $[L : K] = [L : M][M : K]$. Since $K \subset L$ and $M \subset L$ are Galois, $|G| = [L : K]$ and $|H| = [L : M]$. It follows that $|G/H| = [M : K]$, so M is Galois with Galois group G/H.

Conversely if $K \subset M$ is Galois, then there is a $\theta \in M$ such that $M = K(\theta)$ where θ is a root of a polynomial $P(X)$ which splits completely. If $\sigma \in G$, then σ maps θ to other roots of P, and thus maps M to itself, so H is normal. QED.

8.3 Important Note

The Galois theory developed above works for all fields, including fields of positive characteristic. The only cautionary note is that we cannot apply the theorem to a splitting field in characteristic $p > 0$ unless we know that it is the splitting field of a polynomial with no multiple roots.

8.4 An Example

Consider the polynomial $P(X) = X^3 - 2$ studied in section 4.4. It is irreducible over Q by Eisenstein's theorem. The splitting field of $P(X)$ is $Q(\alpha, \omega\alpha, \omega^2\alpha)$. Since it is a splitting field, $Q \subset L$ is Galois and the Galois group is a transitive permutation group on the roots and hence Z_3 or D_3. But complex conjugation is clearly an automorphism of degree 2, so $G = D_3$.

The subgroups of D_3 are $\{e\}$, three copies of Z_2 formed by the three reflections of the equilateral triangle, Z_3, and D_3. The Galois correspondence is order reversing, and the corresponding subfields of L are L, the three root fields $Q(\alpha)$, $Q(\omega\alpha)$, and $Q(\omega^2\alpha)$, a field corresponding to Z_3, and Q itself.

What subfield corresponds to Z_3? Notice that the cube roots of unity belong to L, and thus $M = Q(\theta)$ is a subfield with splitting polynomial $X^3 - 1 = (X - 1)(X^2 + X + 1)$. Hence $[Q(\theta) : Q] = 2$. This field is a Galois extension of Q, so it must correspond to a normal subgroup of D_3 and thus to Z_3. It's Galois group is $D_3/Z_3 = Z_2$.

8.5 Finite Fields Again

In section 16 we constructed a field with four elements. This construction works in general. Suppose we want a field L with p^n elements. Find an irreducible polynomial over Z_p of degree n. Then our field will be the root field of this polynomial.

We know that an irreducible polynomial of degree n exists for the following reason. We proved that there is a field L with p^n elements. The group $L^\star$ is cyclic with generator γ. So $L = Z_p(\gamma)$. The minimal polynomial $P(X)$ of γ has degree n.

Since γ satisfies $X^{p^n} - X = 0$, $P(X)$ is a factor of this polynomial.

Example: Suppose we want a field with 27 elements. We must find an irreducible polynomial over Z_3 with degree 3. It suffices to find a cubic with no roots in Z_3. Randomly trying examples, $X^3 + 2X + 2$ works. So a typical element of the field is $a_0 + a_1X + a_2X^2$ and $X^3 = -2X - 2 = X + 1$.

Definition 5 *The field with p^n elements is often denoted $GF(p^n)$.*

Definition 6 *Let $F = GF(p^n)$. The* Frobenius automorphism *is the map $\sigma : F \to F$ given by $a \to a^p$.*

Theorem 25 *The extension $Z_p \subset GF(p^n)$ is Galois with Galois group Z_n generated by the Frobenius automorphism.*

Proof: The extension is a splitting field for $X^{p^n} - X$. The roots of this polynomial are distinct, indeed exactly the elements of $GF(p^n)$. The Frobenius automorphism satisfies $\sigma(a + b) = (a + b)^p = a^p + b^p = \sigma(a) + \sigma(b)$ because all remaining coefficients in the expansion are divisible by p. It trivially preserves multiplication and takes nonzero elements to nonzero elements. It fixes each element of Z_p. Since it is a linear transformation over Z_p which is one-to-one, it is also onto.

We show that σ has order n in the full Galois group. Certainly its order divides this number, because every element $a \in GF(p^n)$ satisfies $a^{p^n} = a$. If σ has smaller order d, then every element of $GF(p^n)$ would satisfy $X^{p^d} - X = 0$, but this equation has at most p^d roots.

Finally, we show that every automorphism of $GF(p^n)$ is a power of σ. An automorphism of $GF(p^n)$ must fix 1, hence Z_p. Therefore the number of automorphisms is $[GF(p^n) : Z_p] = n$, and powers of σ give this many automorphisms. QED.

Remark: The fundamental theorem of Galois theory then reveals the lattice structure for the subfields of $GF(p^n)$. The subgroups of Z_n are all abelian and have the form $\{\text{all powers of } \sigma^d\}$ for $d|n$. The order of such a subgroup is $\frac{n}{d}$. The corresponding subfield is all elements satisfying $\sigma^d(a) = a^{p^d} = a$, but these elements exactly form $GF(p^d)$. It follows that the Galois group of $GF(p^n)$ over $GF(p^d)$ is $Z_{n/d}$, and generated by σ^d.

The finite subfields of $GF(p^n)$ are $GF(p^d)$ for d/n.

Chapter 9

Concrete Cases of the Theory

9.1 Cyclotomic Fields

We now want to prove Galois' theorem giving a necessary and sufficient condition that an equation be solvable in radicals. This proof requires partial calculation of a small number of specific Galois groups. In this chapter, we only carry calculations far enough to deduce what we will need in our discussion of solvability. A later chapter will completely compute the Galois groups discussed here.

Theorem 26 *Let n be a positive integer and consider $P(X) = X^n - 1$. Suppose the ground field K has characteristic zero or else that its characteristic does not divide n. Then the splitting field L of $P(X)$ is a Galois extension, and its Galois group is abelian.*

Proof: Some caution is required, because the polynomial P is reducible. For example, $X^4 - 1 = (x^2 - 1)(X^2 + 1) = (X - 1)(X + 1)(X^2 + 1)$.

Notice that P cannot have multiple roots by the standard argument, since its formal derivative nX^{n-1} is not zero. So $K \subset L$ is a Galois extension. Every root of P is an nth root of unity, and there are n distinct roots. So the roots are exactly the nth roots of unity. This is a finite subgroup of a field, hence cyclic.

Let ω be a generator of the group of nth units. If σ belongs to the Galois group, $\sigma(\omega) = \omega^i$ for some fixed i, $1 \leq i \leq n$. Then $\sigma(\omega^j) = \sigma(\omega)^j = \omega^{ij}$. It follows that i completely determines the automorphism. The map $\sigma \to i$ sends the Galois group of $K \subset L$ in a one-to-one manner into Z_n. If $\sigma \to i$ and $\tau \to j$, then $\tau\Big(\sigma\Big(\omega\Big)\Big) = \tau\Big(\omega^i\Big) = \tau(\omega)^i = \omega^{ij}$, so our map is a group homomorphism into the multiplicative group $Z_n^\star$ of invertible elements in Z_n, which is abelian. Since G is injected into an abelian group, it is abelian. The map in question need not be onto $Z_n^\star$, and certainly isn't when K already contains some roots of unity. QED.

9.2 Galois Group of a Radical Extension

Theorem 27 *Let K be a field containing all nth roots of unity. If the characteristic is a positive prime p, suppose that p does not divide n. Suppose $a \in K$, and let $L = K(\sqrt[n]{a})$. Then $K \subset L$ is a Galois extension, and its Galois group is cyclic.*

Proof: Since the roots of unity are solutions to $X^n - 1$ and p does not divide n, the standard argument via differentiation shows that the roots of this polynomial are distinct. Thus the nth roots of unity in K form a group of order n. Let ω be a primitive nth root of unity.

Then the roots of $X^n - a$ are $\omega^k \sqrt[n]{a}$, so L is a splitting field for F and thus a Galois extension. Notice that the $\omega^k \sqrt[n]{a}$ need not be linearly independent over K, and $X^n - a$ need not be irreducible.

Each element of the Galois group sends the root $\sqrt[n]{a}$ to some other root $\omega^k \sqrt[n]{a}$. Then $\omega^j \sqrt[n]{a}$ must map to $\omega^{j+k} \sqrt[n]{a}$, so the initial k determines the automorphism. Notice also that composition corresponds to addition of exponents, since $\sqrt[n]{a} \to \omega^k \sqrt[n]{a} \to \omega^{k+l} \sqrt[n]{a}$.

Hence the set of all k in $0 \le k < n$ corresponding to automorphisms of L is a subgroup of Z_n. All subgroups of Z_n are cyclic of some order d dividing n. It follows that the Galois group of $K \subset L$ is cyclic. QED.

9.3 Structure of Extensions with Cyclic Galois Group

Next we'd like to go backward and understand the structure of a Galois Extension if we only know that the Galois group is cyclic. The following theorem is a sort of converse of the previous result.

Theorem 28 *Let K be a field containing all nth roots of unity. If the characteristic is a positive prime p, suppose that p does not divide n. Let $K \subset L$ be a Galois extension with Galois group cyclic of order n. Then there exists a nonzero $a \in L$ such that L is a splitting field for $F(X) = X^n - a$ over K, and this polynomial has distinct roots.*

Proof: Let σ generate the Galois group.

Lemma 5 *The maps $1, \sigma, \sigma^2, \ldots, \sigma^{n-1} : L \to L$ are linearly independent over L, so it is impossible to choose a_i not all zero in L with $\sum a_i \sigma^i$ identically zero on L.*

Proof: Suppose we have n distinct automorphisms σ_i and ignore the assumption that these are the only automorphisms. We prove the maps independent by induction on n. To prove the induction step, suppose $\sigma_1, \ldots, \sigma_{n-1}$ independent. Let $\sum_{i=1}^{n} a_i \sigma_i = 0$. Since $\sigma_1 \neq \sigma_n$, we can find k_0 with $\sigma_1(k_0) \neq \sigma_n(k_0)$. Then $\sum a_i \sigma_i(kk_0) = 0$ so that

$$\sum a_i \sigma_i(k) \sigma_i(k_0) = 0$$

$$\sum a_i \sigma_i(k) \sigma_n(k_0) = 0$$

Subtracting,

$$\sum a_i \sigma_i(k) \Big(\sigma_i(k_0) - \sigma_n(k_0) \Big) = 0$$

The last term vanishes, so this is a dependence relation among the first $n-1$ terms. Hence all coefficients vanish. In particular, $a_1 \Big(\sigma_1(k_0) - \sigma_n(k_0) \Big) = 0$. Since $\sigma_1(k_0) \neq \sigma_n(k_0)$, we have $a_1 = 0$.

Repeat the argument using a different k_0 chosen so $\sigma_2(k_0) \neq \sigma_n(k_0)$. This time we conclude that $a_2 = 0$. Etc. Eventually all a_i vanish for $i < n$. We conclude that $a_n \sigma_n = 0$, and so $a_n = 0$. QED.

Proof of theorem: Let ω be a primitive nth root of unit and consider

$$1 + \omega\sigma + \omega^2\sigma^2 + \ldots + \omega^{n-1}\sigma^{n-1}$$

This expression is not identially zero, so we can find $\beta \in L$ making it nonzero:

$$\theta = \beta + \omega\sigma(\beta) + \omega^2\sigma^2(\beta) + \ldots + \omega^{n-1}\sigma^{n-1}(\beta) \neq 0$$

Applying $\omega\sigma$ to both sides gives

$$\omega\sigma(\theta) = \omega\sigma(\beta) + \omega^2\sigma^2(\beta) + \ldots + \omega^n\sigma^n(\beta)$$

The last term is β, so

$$\omega\sigma(\theta) = \beta + \omega\sigma(\beta) + \omega^2\sigma^2(\beta) + \ldots + \omega^{n-1}\sigma^{n-1}(\beta) = \theta$$

So $\sigma(\theta) = \omega^{-1}\theta$ and $\sigma(\theta^n) = \theta^n$. It follows that $\theta^n \in K$. Let $a = \theta^n$.

The element θ is then $\sqrt[n]{a}$. Consider the polynomial $F(X) = X^n - a$. The roots of this polynomial are $\omega^i \sqrt[n]{a}$, which are distinct. Thus $K(\theta)$ is a splitting field for $F(X)$ over K, and the roots of this polynomial in $K(\theta)$ are distinct. QED.

Chapter 10

Solving Polynomials by Radicals

10.1 Solving Polynomial Equations with Radicals

In this entire chapter, we assume that $P(X)$ is irreducible over a field of characteristic zero. Often this ground field is Q.

Definition 7 *Suppose $P(X)$ is an irreducible polynomial over a field K of characteristic zero. We say this polynomial can be solved by radicals if there are fields*

$$K \subset K_1 \subset K_2 \subset \ldots \subset K_m$$

such that each extension K_i is a root field over K_{i-1} of $X^n - a_i$ for some $a_i \in K_{i-1}$, and the splitting field of $P(X)$ over K is inside K_m.

Definition 8 *A finite group G is said to be* solvable *if each of its Jordan-Holder simple quotients is abelian, and hence cyclic of prime order p.*

Remark: Galois proved the beautiful theorem that an irreducible polynomial over a field of characteristic zero can be solved by radicals if and only if its Galois group is solvable. In particular, a polynomial with a Galois group equal to S_n cannot be solved by radicals if $n \geq 5$.

We will give a very easy proof of half of this theorem: if the polynomial can be solved by radicals, then the Galois group is solvable. Unfortunately, our proof is wrong, as we'll explain in the next section. It turns out that our error is exactly the gap in Ruffini's original proof that there is no formula solving all equations of degree five by radicals.

Theorem 29 (Proof below incorrect) *If $P(X)$ is an irreducible polynomial over a base field of characteristic zero, and P can be solved by radicals, then the Galois group of P is solvable.*

Proof: By taking the least common multiple of the degrees of the successive roots, we can find a common N such that each radical extension adds an Nth root. Instead of extending to a root field, let us extend to the full splitting field of $X^N - a$. Consequently we have a tower

$$K = K_1 \subset K_2 \subset \ldots \subset K_m$$

of Galois extensions, each a splitting field of $X^N - a_i$. Without loss of generality, we can suppose that the first extension $K_1 \subset K_2$ is a splitting field of $X^N - 1$, so that all subsequent fields contain all Nth roots of unity. By assumption the splitting field L of $P(X)$ is contained in K_n. Apply the fundamental theorem of Galois theory to obtain a corresponding reverse tower of Galois groups

$$G_1 \supset G_2 \supset \ldots \supset G_m = \{e\}$$

By the previous sections on the Galois group of a cyclotomic extension and the Galois group of a radical extension, we know that each intermediate composition factor is abelian. Refine to a complete composition series, which must still have abelian composition factors.

By assumption $K \subset L \subset K_m$, so $G_m \supset G \supset \{e\}$ where G is the Galois group of L. Refine this series to a complete composition series, and then apply the Jordan-Holder theorem. It follows that the composition quotients for G are all abelian. QED.

10.2 The Flaw

In the previous proof, we applied the fundamental theorem of Galois theory to the Galois extension $K_1 \subset K_m$ generated by a tower of splitting fields, and thus a tower of Galois extensions

$$K_1 \subset K_2 \subset \ldots \subset K_m$$

But why is $K_1 \subset K_m$ Galois?

It is tempting to guess that there is a simple lemma stating that if $K \subset L \subset M$ and each extension is Galois, then $K \subset M$ is Galois. But this lemma is false.

Consider, for example, the extensions $Q \subset Q(\sqrt{2}) \subset Q(\sqrt[4]{2})$. The first is Galois because it is a splitting field of $X^2 - 2$. The second is Galois because it is the splitting field of $X^2 - \sqrt{2}$. However the full extension is not Galois because $X^4 - 2$ is irreducible with roots $\pm\sqrt[4]{2}$ and $\pm i\sqrt[4]{2}$, so its splitting field contains complex numbers.

Remark: Looking back at the previous section, we notice that the proof only requires knowing that $K \subset K_m$ is Galois. We don't really need to know that $K_{i-1} \subset K_i$ is Galois.

So in the argument given in the previous section, we can assume that each $K_{i-1} \subset K_i$ is merely a root field of $X^N - a_i$, a more natural assumption.

10.3 The Fix

The following lemmas allow us to fix the argument.

Lemma 6 *Suppose $K \subset L$ is a Galois extension. If $F(X)$ is an irreducible polynomial over K and F has a root in L, then F factors completely in L.*

Proof: Let $\theta \in L$ be a root, and form $G(X) = \prod_{\sigma \in G} \Big(X - \sigma(\theta) \Big)$. By the standard arguments, $G(X)$ has coefficients in K. Since $F(X)$ is a minimal polynomial for θ and θ is a root of $G(X)$, $F(X)$ divides $G(X)$. Consequently, $F(X)$ splits completely. QED.

Lemma 7 *Suppose we can find a tower of extensions*

$$K \subset K_1 \subset K_2 \ldots \subset K_m$$

such that each K_i is obtained from K_{i-1} as a root field of a polynomial $X^n - a$, where $a \in K_{i-1}$. The exponents are allowed to vary from extension to extension. Then there is a splitting field M over K containing all of these fields, such that M can be obtained as a (possibly different) tower of radical extensions. If all of the original extensions satisfy $X^n - a = 0$ for a common n, we can assume this true for the new tower for M.

Remark: In this lemma, there is no hypothesis that the first extension adds roots of unity, and no requirement that each radical extension be Galois over its predecessor.

Proof: The tower of extensions has the form

$$K \subset K_1 = K(\alpha_1) \subset K_2 = K(\alpha_1, \alpha_2) \subset \ldots \subset K_m = K(\alpha_1, \alpha_2, \ldots, \alpha_m)$$

where $\alpha_i^{k_i} \in K_{i-1}$. For each i, let $L_i(X)$ be the minimal polynomial of α_i over K. Let M equal the splitting field of $L_1(X)L_2(X) \ldots L_m(X)$. Then M is Galois over K. Let $G = \{\sigma_1, \ldots, \sigma_t\}$ be the Galois group.

Each L_i is irreducible over K with a root $\alpha_i \in M$. It follows that each L_i factors completely over M. Let $\alpha_i = u_{i1}, u_{i2}, \ldots, u_{ik_i}$ be the roots of L_i. For each j, there is an isomorphism of root fields $\sigma : K(\alpha_i) \to K(u_{ij})$. This isomorphism can be extended to an automorphism of M. Hence the roots of L_i for a fixed i all have the form $\sigma(\alpha_i)$ for some σ in the Galois group. Conversely, automorphisms of L map roots of L_i to other roots of L_i. So the roots of L_i are precisely the $\sigma(\alpha_i)$.

The field M is generated by the roots of its splitting polynomial and thus by all

$$\{\sigma(\alpha_i) \mid \sigma \in G, i = 1, 2, \ldots, m\}$$

Let B_1 be the subfield of M generated by $\{\sigma(\alpha_1) \mid \sigma \in G\}$. Notice that

$$K \subset K(\sigma_1(\alpha_1)) \subset K(\sigma_1(\alpha_1), \sigma_2(\alpha_1)) \ldots \subset K(\sigma_1(\alpha_1), \sigma_2(\alpha_1), \ldots \sigma_t(\alpha_1)) = B_1$$

Each of these extensions is a radical extension because $\alpha_1^n = A \in K$ implies $\sigma_i(\alpha_1)^n = \sigma_i(\alpha_1^n) = \sigma_i(A) = A$. Some of our inclusions may be equalities, and there is no claim that n is the smallest exponent making $\sigma_i(\alpha_1)^n \in K$.

Now assume by induction that a radical extension B_i was constructed generated by all

$$\{\sigma(\alpha_j) \mid 1 \leq j \leq i \text{ and } \sigma \in G\}$$

Let

$$B_{i+1} = B_i(\{\sigma(\alpha_{i+1}) \mid \sigma \in G\})$$

To finish the proof, we need only prove that B_{i+1} is a radical extension of B_i. Notice that α_{i+1}^n is in the field generated by $\alpha_1, \ldots, \alpha_i$. Thus we can find a polynomial with coefficients in K such that $\alpha_{i+1}^n = P(\alpha_1, \ldots, \alpha_i)$. Applying $\sigma \in G$, we have $\sigma(\alpha_{i+1})^n = \sigma(\alpha_{i+1}^n) = \sigma P(\alpha_1, \ldots, \alpha_i) = P(\sigma(\alpha_1), \ldots, \sigma(\alpha_i))$. This element is in B_i since the $\sigma(\alpha_j)$ are in B_i and B_i is a field. So B_{i+1} is a radical extension of B_i. QED, and whew!

10.4 Galois' Theorem on Solving Via Radicals

Theorem 30 (Galois) *Let $P(X)$ be an irreducible polynomial over a field K of characteristic zero. If one root of $P(X)$ can be written in terms of radicals, then every root of $P(X)$ can be written that way.*

Let $K \subset L$ be the splitting field of P. Then $K \subset L$ is a Galois extension. The Galois group of this extension is solvable if and only if $P(X) = 0$ is solvable by radicals.

Proof: If one root of the equation is solvable by radicals, we can find a tower of extensions

$$K \subset K_1 \subset K_2 \ldots \subset K_m$$

such that each K_i is obtained from K_{i-1} as a root field of a polynomial $X^k - a$, where $a \in K_{i-1}$, such that $K \subset L \subset K_m$. By finding the least common multiple of the various k, we can find a common n and assume that all extensions are root fields of a polynomial of the form $X^n - a$. To simplify matters further, we can add to the start of the chain an extension $K \subset \tilde{K} \subset K_1$ where $\tilde{K}$ is a splitting field of $X^n - 1$. That is, we can assume that all nth roots of unity belong to the ground field.

We now apply the main theorem of the previous section. This theorem allows us to extend K_m to a field M such that $K \subset M$ is a Galois extension, and each intermediate field in the chain is a radical extension generated by a root of $X^n - a$ for our common n. Some care is required here because the first extension in our chain, $K \subset \tilde{K}$, need not itself be a radical extension. But it can clearly be written as a sequence of radical extensions. So in the construction of M, this sequence can be left alone, and only later extensions will require the addition of additional radical extensions to finally give M.

Let G be the Galois group of $K \subset M$. By the fundamental theorem of Galois theory, the sequence of radical extensions gives rise to a sequence of reverse inclusions

$$G \supset G_1 \supset G_2 \ldots \supset G_{n-1} \supset \{e\}$$

The sequence on the left, $G \supset G_1 \supset \{e\}$ corresponds to the field extensions $K \subset \tilde{K} \subset M$. Since $K \subset \tilde{K}$ is a splitting field and thus a Galois extension, G_1 is a normal subgroup of G and the quotient group G/G_1 is the Galois group of the cyclotomic extension $K \subset \tilde{K}$. By the main theorem in section 8.1, this group is abelian.

The remaining $G_i \supset G_{i+1} \supset M$ correspond to field extensions $K_i \subset K_{i+1} \subset M$. Each of these field extensions is a radical extension obtained by adding a root of $X^n - a = 0$. Since K_i contains all roots of unity, this extension completely splits the polynomial, and thus is a Galois extension. So G_{i+1} is a normal subgroup of G_i with quotient group the Galois group of $K_i \subset K_{i+1}$. By the main theorem in section 8.2, this Galois group is cyclic.

It follows that G has a composition series for which all composition quotients are abelian. Therefore a maximal composition series will have composition quotients Z_p for primes p. In particular, G is solvable.

Now we bring the polynomial $P(X)$ into play. If this polynomial has one root which can be expressed via radicals, then there is a radical extension as above which contains one root of $P(X)$. Since the final $K \subset M$ is a splitting field, $P(X)$ splits completely and thus all roots can be expressed by radicals.

The sequence $K \subset L \subset M$ corresponds to a sequence $G \supset H \supset \{e\}$ where H is the Galois group for $F(X)$. Extend this to a complete composition sequence for G. Since G is solvable, all composition quotients are abelian. In particular, all composition quotients for $H \supset \{e\}$ are abelian, so H is solvable.

Proof in the converse direction: Suppose the splitting field L of $F(X)$ has a Galois group H which is solvable. Then we can find a sequence

$$H \supset H_1 \supset H_2 \supset \ldots \supset \{e\}$$

where each $H_i \supset H_{i+1}$ is a normal subgroup and each quotient group is Z_p for some prime p. By the fundamental theorem of Galois theory, this corresponds to a sequence of fields

$$K \subset K_1 \subset K_2 \subset \ldots \subset L$$

Choose n a multiple of each $[K_{i+1}, K_i]$ and let $L \subset \tilde{L}$ be a splitting field for $X^n - 1$. The set of nth roots of unity in $\tilde{L}$ is cyclic of order n. Let ω be a generator. Note that ω could actually be in L or even in K. Extend each K_i by ω, obtaining

$$K \subset K(\omega) \subset K_1(\omega) \subset K_2(\omega) \subset \ldots \subset L(\omega)$$

We claim each of these extensions is a radical extension. In that case, L is inside such an extension, and thus F is solvable by radicals.

Certainly $K \subset K(\omega)$ comes from a sequence of radical extensions.

Since $H_i \supset H_{i+1} \supset \{e\}$ is normal, $K_i \subset K_{i+1} \subset L$ is Galois. So $K_i \subset K_{i+1}$ is a Galois extension with cyclic Galois group. We now claim that $K_{i-1}(\omega) \subset K_i(\omega)$ is Galois with Galois group a possibly different cyclic group. If this is true, then since $K_{i-1}(\omega)$ contains all nth roots of unity, the main theorem in section 8.3 asserts that the extension is radical, and we will be done.

Since $K_{i-1} \subset K_i$ is Galois, we can apply theorem 21 of section 7.1 to find a polynomial $G(X)$ over K_{i-1} such that $K_i = K_{i-1}(\theta)$ for a root θ of G, and such that the polynomial splits completely. Clearly $K_{i-1}(\omega) \subset K_i(\omega)$ is a splitting field of $G(X)$, and so it is a Galois extension. However, $G(X)$ may not be irreducible over $K_{i-1}(\omega)$. Factor it and let $G_1(X)$ be an irreducible factor with root θ.

Let $\sigma : K_i(\omega) \to K_i(\omega)$ be an automorphism over $K_{i-1}(\omega)$. Then σ is determined by $\sigma(\theta)$, and $\sigma(\theta)$ is another root of $G_1(X)$. This other root is a root of $G(X)$, so σ comes from an automorphism of K_i over K_{i-1}. The conclusion is that the automorphism group of $K_i(\omega)$ is a subgroup of the automorphism group of K_i. Since the automorphism group of K_i is cyclic, the automorphism group of $K_(\omega)$ is also cyclic, possibly of a different order. QED.

10.5 Solving Generic Equations by Radicals

Let $a_1, \ldots, a_n$ be arbitrary symbols. Form the field $K = Q(a_1, \ldots, a_n)$ consisting of all $\frac{P}{Q}$ where P and Q are polynomials in the symbols.

The equation $P(X) = X^n + a_1 X^{n-1} + \ldots + a_n = 0$ is called the *generic equation of degree n*. The quadratic formula gives a solution of the generic equation of degree 2:

$$\text{Solution of } X^2 + a_1 X + a_2\text{:} \quad \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2}$$

Similarly the cubic and quartic formulas give solutions of the generic equations of degree 3 and 4.

Theorem 31 *The generic equation of degree n cannot be solved by radicals if $n \geq 5$*

Proof: This follows from Galois' theorem and earlier results on the symmetric group, once we prove that the Galois group of the generic equation of degree n is the symmetric group of order n.

Let $\lambda_1, \ldots, \lambda_n$ be new arbitrary symbols, and form the field $Q(\lambda_1, \ldots, \lambda_n)$. Formally write

$$(X - \lambda_1)(X - \lambda_2) \ldots (X - \lambda_n) = X^n + a_1 X^{n-1} + \ldots + a_n$$

This holds if $a_i = (-1)^i \sigma_i(\lambda_1, \ldots, \lambda_n)$ where the σ_i are the elementary symmetric functions. Recall the definitions:

$$\begin{array}{rcl}
\sigma_1(X_1, \ldots, X_n) & = & X_1 + X_2 + \ldots + X_n \\
\sigma_2(X_1, \ldots, X_n) & = & X_1X_2 + X_1X_3 + \ldots + X_{n-1}X_n \\
\ldots & = & \ldots \\
\sigma_n(X_1, \ldots, X_n) & = & X_1X_2 \ldots X_n
\end{array}$$

Redefine the original field $Q(a_1, \ldots, a_n)$ to be

$$Q(-\sigma_1(\lambda_1, \ldots, \lambda_n), \ldots, \pm\sigma_n(\lambda_1, \ldots, \lambda_n))$$

so $Q(a_1, a_2, \ldots, a_n) \subset Q(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Clearly the field $Q(\lambda_1, \ldots, \lambda_n)$ is a splitting field of the polynomial $P(X) = (X - \lambda_1) \ldots (X - \lambda_n) = X^n + a_1 X^{n-1} + \ldots + a_n$. The elements of the Galois group permute these roots. Since the elementary symmetric functions are preserved by all permutations, all permutations work and the Galois group is the full symmetric group.

However, there is something else to prove. We have identified the transcendental extension i $Q(a_1, \ldots, a_n)$ with the subfield of $Q(\lambda_1, \ldots, \lambda_n)$ generated by the elementary symmetric polynomials. This is only legal if the σ_i are independent. So we must prove that whenever P is a polynomial such that $P(\sigma_1, \ldots, \sigma_n) = 0$, we have $P = 0$.

Lemma 8 *The elementary symmetric functions are algebraically independent; if P is a polynomial and $P(\sigma_1, \ldots, \sigma_n) = 0$, then P is identically zero.*

Proof: Each elementary symmetric function involves $X_1, X_2, \ldots, X_n$. We prove the theorem by induction on n. The result is clear when $n = 1$.

Assume the result for $n-1$ variables. To prove the induction step, suppose that $P(Y_1, \ldots, Y_n)$ is a polynomial of minimal total degree satisfied by the elementary symmetric functions $\lambda_1, \ldots, \lambda_n$, each a function of $X_1, \ldots, X_n$. Write $P = p_0(Y_1, \ldots, Y_{n-1}) + p_1(Y_1, \ldots, Y_{n-1})Y_n +$

$\ldots + p_k(Y_1, \ldots, Y_{n-1})Y_n^k$. If p_0 is zero, then $P = Y_n\tilde{Q}(Y_1, \ldots, Y_n)$ for a polynomial $\tilde{Q}$ of smaller degree and $\sigma_n Q(\sigma_1, \ldots, \sigma_n) = 0$. Since $\sigma_n = X_1 \ldots X_n$ in the integral domain $Q[X_1, \ldots, X_n]$, we have $\tilde{Q}(\sigma_1, \ldots, \sigma_n) = 0$, contradicting the assumption that P has smallest degree. So we can assume $p_0(\sigma_1, \ldots, \sigma_{n-1}) \neq 0$.

We have

$$p_0(\sigma_1, \ldots, \sigma_{n-1}) + p_1(\sigma_1, \ldots, \sigma_{n-1})\sigma_n + \ldots = 0$$

Each σ_i is a function of $X_1, \ldots, X_n$. Substitute 0 for X_n. Then $\sigma_n(X_1, \ldots, X_{n-1}, 0) = 0$, and we obtain $p_0(\sigma_1, \ldots, \sigma_{n-1}) = 0$. However $\sigma_i(X_1, X_2, \ldots, X_{n-1}, 0)$ are exactly the elementary symmetric functions in the first $n - 1$ variables, so we get the contradiction that $p_0 = 0$ by induction on n. QED.

10.6 A Polynomial Equation Which Cannot Be Solved by Radicals

We now know that there is no general formula solving an equation of degree $n \geq 5$. It is possible, however, that a solution can always be found involving radicals, but varying from equation to equation. To rule this out, we will find a quintic with integer coefficients whose solutions cannot be written in terms of radicals.

We will prove that $P(x) = x^5 - 4x + 2$ is not solvable by radicals. Note that $P(x)$ is irreducible over the rationals by Eisenstein's theorem, using $p = 2$.

Lemma 9 (Cauchy) *If a prime p divides the order of a group G, then G has an element of order p.*

Proof: Consider the set X of all p-tuples $(g_1, g_2, \ldots, g_p)$ whose product is the identity. The first $p - 1$ elements of such a tuple can be arbitrary, and these determine g_p. So X has $|G|^{p-1}$ elements. This number is divisible by p.

We can cyclically permute the elements in a particular p-tuple. Suppose we first return to the starting point after k steps. Clearly k divides p, so either $k = 1$ or $k = p$. If $k = 1$ then all terms are equal; otherwise all terms in the cyclic permutation are different. Since p divides the number of such terms and p divides the size of X, p must divide the number of $(g, g, \ldots, g)$ in X. Each represents an element of order p except $(e, e, \ldots, e)$. QED.

Calculation of the Galois Group: Below is a plot of $P(X)$:

This plot shows that the polynomial has three real roots between -2 and 2. It has no other real roots, for $x > 2$ implies that $P(x) = x(x^4 - 4) + 2 > 2$ and $x < -2$ implies that $P(x) = x(x^4 - 4) + 2 \leq 12x + 2 \leq -10$.

Consequently, P has two complex roots which are not real. They must be conjugates. The splitting field L is generated by the three real roots, and the two conjugate complex

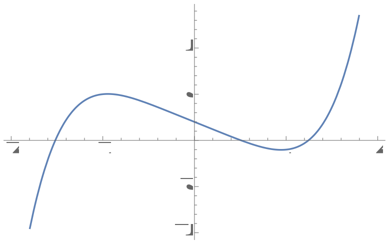


Figure 10.1: example caption

roots, so complex conjudation is an automorphism of $Q \subset L$. As a permutation, it is a two-cycle.

The extension $Q \subset L$ contains a root field $Q(\theta)$ for P. Since P is irreducible, the order of the root field is 5. It follows that $[L : K]$ is divisible by 5. Since $[L : K]$ is also the order of the Galois group, the order of the group is divisible by 5. Then by Cauchy's lemma, the Galois group contains an element σ of order five. Imagine this element written in cyclic notation. It must be a five cycle, else it would have order relatively prime to 5.

To complete the argument, we show that any subgroup G of S_5 containing a five cycle and a two cycle must be the entire group.

The 2-cycle and the 5-cycle have two elements in common. Number the roots so one common element is 1 and the 5-cycle is (1 2 3 4 5). The square of this element is (1 3 5 2 4), its cube is (1 4 2 5 3), etc. All of these elements except the identity are also 5-cycles whose powers give all these 5-cycles. The second element of one of these elements is the second element of the 2-cycle. By renumbering, we can suppose the 5-cycle is (1 2 3 4 5) and the 2-cycle is (1 2).

The set of all two cycles generates the full symmetric group S_n. To see this, write an arbitrary permutation in cycle notion and consider one of the cycles. By renumbering, it equals (1 2 . . . k) and

$$(12 \ldots k) = (1k) \ldots (13)(12)$$

So it suffices to show that our group contains all two cycles.

But $\sigma \tau \sigma^{-1}$ applies the permutation σ to the elements of τ. For instance $(12345)(12)(12345)^{-1} = (23)$.

It follows that our group contains (1 2), (2 3), (3 4), (4 5), and (5 1).

Also (1 3) = (2 3) (1 2) (2 3), (1 4) = (3 4) (1 3) (3 4), and (1 5) = (4 5) (1 4) (4 5). Also (2 4) = (3 4)(2 3) (3 4), (2 5) = (4, 5) (2 4) (4 5). Also (3 5) = (4 5) (3 4) (4 5).

So our group contains all 2-cycles and equals S_5. This group is not solvable, so $x^5 - 4x + 2$ cannot be solved with radicals.

Chapter 11

Straightedge and Compass Constructions

11.1 Constructions With Straightedge and Compass

Euclidean geometry is based on straightedge and compass constructions. The fundamental postulates immediately reveal this, since they proclaim the existence of lines through two points, and of circles with given center and radius. The very first theorem in Euclid asserts that equilateral triangles exist, proving this by a construction. Draw a base, draw circles with radius the base through the two endpoints of the base, and select the intersection of these circles as a third point.

It is well-known that many important constructions cannot be done with straightedge and compass; Galois theory provides exactly the tools needed to show this.

Our first step is to describe a straightedge and compass construction with sufficient rigor.

Definition 9 *A* construction problem *is a finite set of points in the plane, called the* **givens***, and another finite set of points in the plane, called the* **unknowns**. *Additional points may be added to the givens using one of the three methods listed below. The goal is to do this over and over until finally the unknowns are among the givens.*

- *If P_1 and Q_1 are distinct points already known, and if P_2 and Q_2 are distinct points already known, draw the line L_1 through P_1 and Q_1 and draw the line L_2 through P_2 and Q_2. If these lines are different and not parallel, their intersection point can be added to the givens.*
- *Suppose P_1 and P_2 are distinct points already known. Suppose P_3 is a known point.*

Suppose P_4 and P_5 are distinct points already known. Draw the circle C through P_3 with radius the distance from P_1 to P_2. Draw the line L through P_4 and P_5. This line and circle intersect in from 0 to 2 points. These intersection points can be added to the givens.

- *Suppose P_1 and P_2 are distinct points already known. Suppose P_3 is a known point. Suppose P_4 and P_5 are distinct points already known, and P_6 is a known point. Draw the circle through P_3 with radius the distance from P_1 to P_2. Draw the circle through P_4 with radius the distance from P_4 to P_5. If the two circles are not equal, they may intersect in from zero through two points. In this case, add any intersection points to the givens.*

Remark: At first it is not clear that these rules capture what is allowed during a construction. Consider the problem of bisecting an angle. We are actually given two intersecting lines, and we are asked to construct a third line.

However, we easily translate to the new language. We could describe this problem by giving the vertex P_1 and a second point P_2 on the first line. Once we know the second line, we could intersect it with the circle with center P_1 through P_2, so we can give the second line by giving the point P_3 where this circle intersects the second line. The angle bisector will then determine a final point on the circle, which becomes our unknown P_4.

Select coordinates so $P_1 = (0, 0)$ and $P_2 = (1, 0)$. Then $P_3 = (\cos\theta, \sin\theta)$. These three points are our givens. The unknown is $\left(\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\right)$.

Theorem 32

- *If only one point is given, it is impossible to construct other points.*
- *If at least two points P_1 and P_2 are given, we can choose orthonormal coordinates so $P_1 = (0, 0)$ and $P_2 = (1, 0)$. If $P_3 = (a, b)$ is a third point, we can construct $(a, 0)$ and $(b, 0)$. Conversely if these points are given, we can construct (a, b).*

Proof: The first item, and the first line of the second option, are obvious. So assume $P_1 = (0, 0)$ and $P_2 = (1, 0)$. Construct a perpendicular to the line through these points at P_1. If (a, b) is given, construct the perpendiculars from these points to the two coordinate lines, getting $(a, 0)$ and $(0, b)$. Using a circle with vertex the origin, construct $(b, 0)$ from $(0, b)$.

Conversely, given $(a, 0)$ and $(b, 0)$, draw the perpendicular at $(0, 0)$ as before. Using a circle through the origin, convert $(b, 0)$ to $(0, b)$. Construct perpendiculars to $(a, 0)$ and $(0, b)$. These perpendiculars intersect at (a, b). QED

Remark: Using this theorem, the general construction problem can be reformulated. Without loss of generality, we can take the givens to be finitely many real numbers, including in particular 0 and 1, and we can take the unknowns to be finitely many additional real numbers.

Theorem 33 *Given a known finite set of given real numbers, additional unknowns can be constructed using addition, subtraction, multiplication, division, and taking square roots of givens. Conversely, any constructible real can be formed by a finite combination of these operations.*

Proof: Adding and subtracting reals by compass is easy. Notice that multiplication only makes sense if we have a scale, for a^2 will be larger than a if $a > 1$ but smaller than a if $a < 1$. So the technique for multiplying must involve the segment of length 1. Once this is known, a construction is easy.

One is shown below. The idea of this construction is to form similar triangles and notice that $\frac{a}{1} = \frac{ab}{b}$. The reader can easily provide details.

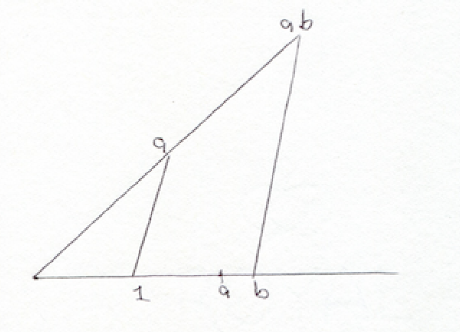


Figure 11.1: Multiply

The same diagram can be used to divide, using the similarity condition $\frac{a}{b} = \frac{\frac{a}{b}}{1}$.

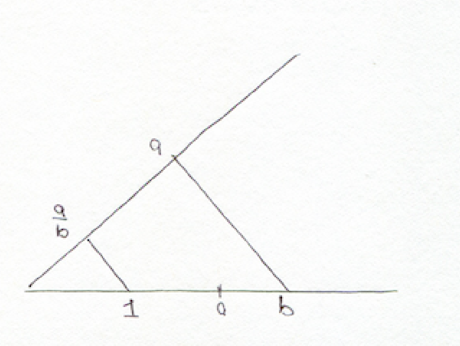


Figure 11.2: Divide

Calculating square roots also requires a unit. Below is one possible construction from http://math.stackexchange.com. Let the horizontal distance AC = 1 and draw the vertical line AD. Call its length L. By similar triangles, $\frac{a}{L} = \frac{L}{1}$ and so $L = \sqrt{a}$.

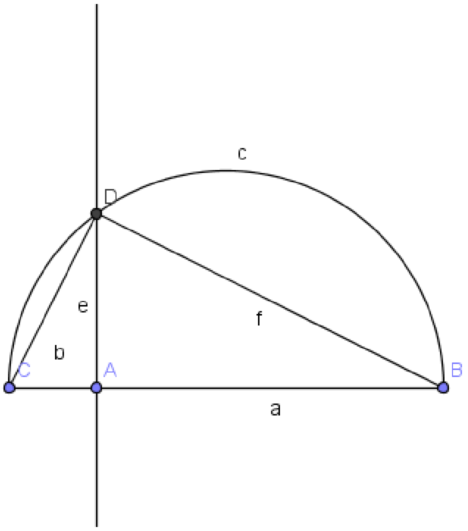


Figure 11.3: Square Root

Conversely, consider the operation of adding the intersection of two lines, when the coordinates of the given points and points constructed so far generate a field K. Each line is determined by two points with coordinates in K, so the equations of the lines have the form $ax + by + c = 0$ and $dx + ey + f = 0$ where $a, b, c, d, e, f \in K$. Ordinary high school algebra can be used to find the intersection point, and this algebra only requires addition, subtraction, multiplication, and division of the coordinates. So the coordinates of the intersection point are still in K.

Consider the operation of intersecting a line with a circle. Leaving the special case of a vertical line to the reader, we can assume that the line is given by $y = ax+b$ where $a, b \in K$. The equation of the circle is $(x - p)^2 + (y - q)^2 = r^2$ where r is the radius. This radius can be computed using the Pythagorian theorem applied to two points with coordinates in K, so $p, q, r^2 \in K$. To solve, we replace y by $ax + b$, obtaining $(x - p)^2 + (ax + b - q)^2 = r^2$. This is a quadratic equation in x, and solving for x only requires a square root. Once x is known, y can be found by solving the same formula, possibly requiring another square root. (There is no need to worry about cases where there is no solution, because if the line and circle do not intersect, we don't generate new points.)

Finally, the operation of intersecting two circles involves solving two equations $(x - p)^2 + (y - q)^2 = r^2$ and $(x - s)^2 + (y - t)^2 = u^2$ simultaneously. If we subtract one equation from the other, the terms x^2 and y^2 cancel out and we get a linear equation in x and y. Points on both circles lie on this line, so we are back to the previous case. QED

Remark: Summarizing, a straightedge and compass construction begins with *gives*; numbers which can be constructed from them using addition, subtraction, multiplication, and division. This gives a field $Q \subset K_1$. We can then construct the extension K_2 formed from K_1 by adding the square root of a positive number in K_1, and then extending to the field this root generates. Call this K_2. In general

Theorem 34

- *If a set of givens generates a field $Q \subset K_1$ and a set of unknowns generates a field $Q \subset L$, then we can construct the unknowns from the givens if and only if there is a sequence of extensions*
$$Q \subset K_1 \subset K_2 \subset \ldots \subset K_n$$
where each K_i is formed from K_{i-1} by adding a square root of a positive number in K_{i-1}, such that $L \subset K_n$.

- *In particular, the dimension of L over K_1 must be a power of* 2.

Proof: This is obvious from the discussion above.

11.2 Complex Extensions and Constructions

In the previous section, we reduced straightedge and compass constructions to the existence of a chain of real extensions $Q \subset K_1 \subset \ldots \subset K_n$. The results remain unchanged, however, if we work in the plane and use complex extensions.

We could then assume that the *givens* form a set of points in the plane. Label two of them as 0 and 1 and use this to identify the plane with the set of complex numbers. The *unknowns* form a larger subset of C. Using straightedge and compass constructions, we can add, subtract, multiply, and divide complex numbers. We can also find complex square roots, essentially because we can form real square roots and can bisect angles. Therefore

Theorem 35

- *If a set of givens generates a field $Q \subset K_1$ and a set of unknowns generates a field $Q \subset L$, then we can construct the unknowns from the givens if and only if there is a sequence of extensions*
$$Q \subset K_1 \subset K_2 \subset \ldots \subset K_n$$
where each K_i is formed from K_{i-1} by adding a square root of a complex number in K_{i-1}, such that $L \subset K_n$.

- *In particular, the dimension of L over K_1 must be a power of* 2.

11.3 Trisecting Angles; Doubling the Cube

Theorem 36 *It is not possible in general to trisect an angle using straightedge and compass. In particular, given two starting points* $(0,0)$ *and* $(1,0)$*, we can construct a* 60*-degree angle at the origin, but cannot construct a* 20*-degree angle at the origin.*

Proof: Construct an equilateral triangle at the origin using Euclid's first theorem. Intersect this triangle with the circle centered at the origin and with radius 1, to get the point $(\cos 60°, \sin 60°) = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. The unknown is the point $(\cos 20°, \sin 20°) = (\cos\theta, \sin\theta)$. By de Moivre,

$$(\cos\theta + i\sin\theta)^3 = (\cos 3\theta + i\sin 3\theta) = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

Comparing real parts, $\cos^3\theta - 3\cos\theta\sin^2\theta = \frac{1}{2}$. So $\cos^3\theta - 3\cos\theta(1-\cos^2\theta) = \frac{1}{2}$. If $X = \cos\theta$, then

$$4X^3 - 3X - \frac{1}{2} = 0$$

But $P(X) = 4x^3 - 3X - \frac{1}{2}$ is irreducible over Q since it does not have a rational root. Indeed if $\frac{a}{b}$ is written in lowest terms and solves the equation, then $8a^3 - 6ab^2 - b^3 = 0$. If p divides a, then p divides b, which is disallowed, so we can assume $a = 1$. Thus $8 - 6b^2 - b^3 = 0$. Since 2 divides $8 - 6b^2$, 2 divides b, so $b = 2B$. Then $8 - 24B^2 - 8B^3 = 0$ and so $1 - 3B^2 - B^3 = 0$. If p divides B, then p divides 1, so B is ± 1, but neither of these works.

So $Q \subset Q(\cos 20°)$ has degree 3 and that is not a power of 2. QED.

Warning: It does not follow that trisection is always impossible. For example, a 45° angle is easily trisected, because a 15° angle can be constructed by making an equilateral triangle and bisecting a vertex angle twice.

Remark: In an article on the web by J. J. O'Connor and E. F. Robertson, http://www-history.mcs.st-and.ac.uk/PrintHT/Doubling_the_cube.html, we read two accounts of the so-called Delian problem of ancient Greek mythology. Theon of Smyra quotes Eratosthenes, as translated by Heath:

> Eratosthenes, in his work entitled Platonicus relates that, when the god proclaimed to the Delians through the oracle that, in order to get rid of a plague, they should construct an altar double that of the existing one, their craftsmen fell into great perplexity in their efforts to discover how a solid could be made the double of a similar solid; they therefore went to ask Plato about it, and he replied that the oracle meant, not that the god wanted an altar of double the size, but that he wished, in setting them the task, to shame the Greeks for their neglect of mathematics and their contempt of geometry.

Eutocius writes

> Eratosthenes to King Ptolemy, greetings. The story goes that one of the ancient tragic poets represented Minos having a tomb built for Glaucus, and that when Minos found that the tomb measured a hundred feet on every side, he said "Too small is the tomb you have marked out as the royal resting place. Let it be twice as large. Without spoiling the form, quickly double each side of the tomb". This was clearly a mistake. For if the sides are doubled the surface is multiplied fourfold and the volume eightfold.

The Greek geometers took this ancient mythology to have the following meaning. Given the base of a cube, construct using straightedge and compass the base of a larger cube whose volume is twice the volume of the original cube.

Theorem 37 *It is impossible to double a cube using straightedge and compass.*

Proof: If we take the endpoints of the base of the original cube to be $(0, 0)$ and $(1, 0)$, then it is required that we construct $(a, 0)$ where $a^3 = 2$. Thus we are to form $Q(\sqrt[3]{2})$. But the minimal polynomial of this number is $X^3 - 2$, which is irreducible by Eisenstein. So $[Q(\sqrt[3]{2} : Q] = 3$, and the construction cannot be done. QED.

The most famous of all Greek construction problems is the problem of squaring the circle. Given the radius of a circle, the problem is to construct the side of a square with the same area. If the center of the circle is $(0, 0)$ and the circle goes through $(1, 0)$, we asked to construct a line of length $\sqrt{\pi}$, and so $(\sqrt{\pi}, 0)$.

Theorem 38 *It is impossible to square a circle using straightedge and compass.*

Proof: All constructed points are algebraic over Q. But π is not algebraic, and therefore $\sqrt{\pi}$ is not algebraic. This is proved in the next chapter.

Chapter 12

Irrationality and Transcendence of π and e

To complete the proof that it is impossible to square a circle, we must prove that π is transcendental. The proof uses $e^{i\pi} = -1$, so we study e and π simultaneously.

12.1 Irrationality e and π

Theorem 39 *The number e is irrational*

Proof: Write

$$e = 1 + 1 + \frac{1}{2!} + \ldots + \frac{1}{n!} + \frac{1}{(n+1)!}\left(1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \ldots\right)$$

Therefore

$$n!e = n!\left(1 + 1 + \frac{1}{2!} + \ldots + \frac{1}{n!}\right) + R$$

where $0 < R < \frac{1}{n+1}(e-1)$. If $e = \frac{a}{b}$ and n is larger then b, then $n!e$ is an integer. The number $n!\left(1 + 1 + \frac{1}{2!} + \ldots + \frac{1}{n!}\right)$ is an integer, so R is an integer. But as n increases, the estimate on R shows that it is positive and bounded by a series of numbers approaching 0. Such a series of R's cannot all be integers. QED.

Remark: The following beautiful proof is essentially due to Niven. It is not the original proof of the theorem, which was by Lambert using continued fractions.

Theorem 40 *π is irrational*

Proof: Consider

$$I_n = \int_0^\pi \frac{x^n(\pi - x)^n}{n!} \sin x \; dx$$

These integrals can be computed using integration by parts, and a short calculation shows that

- $I_0 = 2$
- $I_1 = 4$
- $I_2 = 24 - 2\pi^2$
- $I_3 = 240 - 24\pi^2$
- $I_4 = 3360 - 360\pi^2 + 2\pi^4$
- $I_5 = 60480 - 6720\pi^2 + 60\pi^4$

We claim that this pattern continues; I_n is a polynomial in π of degree at most n, with integer coefficients. An easy calculation gives the results in the table for $n = 0, 1$.

We now work by induction, and integrate by parts twice. Let $G(x) = \frac{x^n(\pi-x)^n}{n!}$. The boundary terms from integrating by parts will vanish because $F(x)$ and $F'(x)$ vanish at the limits of integration. So

$$I_n = \int G[x] \frac{d}{dx}(-\cos x) \; dx = \int G'(x) \cos x \; dx = \int G'(x) \frac{d}{dx} \sin x \; dx = -\int G''(x) \sin x \; dx$$

Moreover

$$G'(x) = \frac{x^{n-1}(\pi - x)^n}{(n-1)!} - \frac{x^n(\pi - x)^{n-1}}{(n-1)!}$$

and

$$G''(x) = \frac{x^{n-2}(\pi - x)^n}{(n-2)!} - 2n\frac{x^{n-1}(\pi - x)^{n-1}}{(n-1)!} + \frac{x^n(\pi - x)^{n-2}}{(n-2)!}$$

This expression equals

$$G''(x) = \Big((\pi - x)^2 + x^2\Big) \frac{x^{n-2}(\pi - x)^{n-2}}{(n-2)!} - 2n\frac{x^{n-1}(\pi - x)^{n-1}}{(n-1)!}$$

Notice that

$$(\pi - x)^2 + x^2 = \pi^2 - 2\pi x + x^2 + x^2 = \pi^2 - 2x(\pi - x)$$

It follows that

$$I_n = -\pi^2 I_{n-2} + 2(n-1)I_{n-1} + 2nI_{n-1} = (4n-2)I_{n-1} - \pi^2 I_{n-2}$$

and the lemma follows by induction.

Suppose $\pi = \frac{a}{b}$. Then $b^n I_n$ is an integer, because b^n clearers the denominators of all powers of $\frac{a}{b}$ of degree less than or equal to n. From the definition, $b^n I_n > 0$. However, by symmetry, the integrand is largest at $\frac{\pi}{2}$ when it is $\left(\frac{\pi}{2}\right)^{2n} \frac{1}{n!}$. Let $A = \frac{\pi^2}{4}$.

$$0 < b^n I_n \leq \pi \left(\frac{A}{1}\right)\left(\frac{A}{2}\right) \dots \left(\frac{A}{n}\right)$$

The right side of this expression goes to zero as n goes to infinity, so $b^n I_n$ cannot always be a positive integer. QED

12.2 Transcendence of e

The proofs of this section come from a 1970 course by Ian Steward.

Theorem 41 (Hermite) *e is transcendental*

Proof: Suppose

$$a_0 + a_1 e + a_2 e^2 + \dots + a_n e^n = 0$$

with the a_i integers.

The proof starts with a simple calculation from freshman calculus. For p an arbitrary positive integer, define

$$f(x) = \frac{x^{p-1}(x-1)^p(x-2)^p \dots (x-n)^p}{(p-1)!}$$

Define

$$F(x) = f(x) + f'(x) + f''(x) + \dots$$

where the terms eventually vanish because f is a polynomial. Then

$$\frac{d}{dx}\left(e^{-x}F(x)\right) = e^{-x}\left(F'(x) - F(x)\right) = -e^{-x}f(x)$$

and therefore

$$a_j \int_0^j e^{-x} f(x)\ dx = a_j \Big[- e^{-x}F(x) \Big]_0^j = a_j F[0] - a_j e^{-j} F[j]$$

Multiply by e^j and sum from $j = 0$ to $j = n$ to get

$$\sum_{j=0}^{n} a_j e^j \int_0^j e^{-x} f(x)\ dx = \sum_{j=0}^{n} a_j e^j F[0] - \sum_{j=0}^{n} a_j F[j]$$

$$= -\sum_{j=0}^{n} \sum_{i} a_j f^{(i)}(j)$$

In the end, we obtain the following result:

$$\sum_{j=0}^{n} a_j e^j \int_0^j e^{-x} f(x)\ dx = -\sum_{j=0}^{n} \sum_{i} a_j f^{(i)}(j)$$

This equation will lead to a contradiction as follows. The function $f(x)$ depends on an integer p. We will prove that as p goes to infinity, the left side of the equation goes to zero. We will then show that $f^{(i)}(j)$ is always an integer, and thus the right side is always an integer. Finally we restrict to the case that p is a prime larger than n. We'll show that$f^{(i)}(j)$ is divisible by p except when $j = 0$ and $i = p - 1$. We'll also show that the term with $j = 0$ and $i = p - 1$ is not divisible by p. It follows that the right side is a nonzero integer for all large primes, and we obtain a contradiction. The theorem follows.

Here are the details. First we estimate the left side of the equation. The term $\sum_{j=0}^{n} a_j e^j$ is a constant, so we need only estimate the integral. We can be fairly sloppy. Over the interval $[0, j] \subset [0, n]$ we have $|x - k| < n$ and consequently the integrand is at most n to the power $(p - 1) + np$. The term e^{-x} is bounded by 1. The length of the interval gives an extra n. so the total integral is smaller than $\frac{n^{1+(p-1)+np}}{(p-1)!} = \frac{n^{(n+1)p}}{(p-1)!}$. The term n^{n+1} is a constant C and this term is $\frac{C^p}{(p-1)!}$ which equals

$$C\frac{C}{1}\frac{C}{2}\dots\frac{C}{p}$$

As p increases, this term goes to zero.

Now the details for the right side. From now on, assume p prime. If $1 \leq j \leq n$, notice that the $f^{(i)}(j)$ will vanish unless the $(x - j)^p$ term was differentiated exactly p times, and in that case the $(p - 1)!$ in the denominator will be canceled and there will be an extra p in the numerator. When $(x - j)^p$ is differentiated p times, the other terms may also have been differentiated, but these will only contribute integer multiples of an integer already divisible by p.

The term $f^{(i)}(0)$ will vanish unless the term $x^{(p-1)}$ is differentiated exactly $p - 1$ times. In that case the $(p - 1)!$ in the denominator will be cancelled, but there will be no p in the numerator. Of course the remaining terms may also have been differentiated, but if a remaining term is differentiated even once, we'll get an extra p and the term will be divisible by p. The only exception is when x^{p-1} is differentiated $p - 1$ times and no other term is differentiated. In this case, the value at 0 is not divisible by p because $p > n$. QED.

12.3 Intermission: Fundamental Theorem of Symmetric Polynomials

We alluded to the following theorem earlier, but now we need it in the proof of the transcendence of π.

Theorem 42 *Let $P(X_1, \ldots, X_n)$ be a polynomial over a field K. Suppose P is symmetric in the sense that $P(X_{\tau(1)}, \ldots, X_{\tau(n)}) = P(X_1, \ldots, X_n)$ for all permutations $\tau \in S_n$. Then P can be written as a polynomial in the elementary symmetric functions:*

$$P(X_1, \ldots, X_n) = Q(\sigma_1(X_1, \ldots, X_n), \ldots, \sigma_n(X_1, \ldots, X_n))$$

for some polynomial Q. If P has integer coefficients, so does Q.

Preliminary Remark: When K is a finite field, a polynomial can vanish identically on $K \times \ldots \times K$ and yet not be the zero polynomial. This theorem is about polynomials, not about functions on $K \times \ldots \times K$. When we claim that two polynomials are equal, we are claiming that they have the same coefficients, not just that they take the same values.

Proof: We prove both parts at the same time by assuming that our polynomials have coefficients in a commutative ring with unit A. Define the *weight* of a polynomial to be the maximum of the weights of the monomials which occur in it, and define the weight of $X_1^{k_1} X_2^{k_2} \ldots X_n^{k_n}$ to be $k_1 + 2k_2 + \ldots + nk_n$.

We will prove a slightly stronger theorem, namely that Q exists and has weight at most the degree of P. This result will be proved by a double induction, first on the number of variables n, and then on the degree d of P.

The theorem is obvious if $n = 1$ because then there is only one variable and $Q = P$. So assume the theorem is true for all polynomials in $n - 1$ variables. We prove it true when there are n variables. The result is obvious if the degree of P is zero, so we can assume it true when the degree of a polynomial is smaller than d. Suppose our P has degree d.

Then $P(X_1, \ldots, X_{n-1}, 0)$ can be written as a polynomial G in the $n - 1$ elementary symmetric functions in $X_1, \ldots, X_{n-1}$. It is easy to see that these are exactly the first $n - 1$ elementary symmetric functions in $X_1, \ldots, X_n$, evaluated when $X_n = 0$. Form

$$P_1(X_1, \ldots, X_n) = P(X_1, \ldots, X_n) - G(\sigma_1(X_1, \ldots, X_n), \ldots, \sigma_{n-1}(X_1, \ldots, X_n))$$

Since the degree of $P(X_1, \ldots, X_{n-1}, 0)$ is at most d, the weight of G is at most d, and therefore the degree of $G(\sigma_1(X_1, \ldots, X_n), \ldots, \sigma_{n-1}(X_1, \ldots, X_n))$ is at most d.

It now suffices to prove the result for P_1, which vanishes when $X_n = 0$. Since we are dealing with polynomials rather than functions on the ground field, this statement actually means that every monomial in the polynomial contains X_n. By invariance under the symmetric group, every monomial contains all X_i. It follows that

$$P_1(X_1, \ldots, X_n) = \sigma_n(X_1, \ldots, X_n) P_2(X_1, \ldots, X_n)$$

The polynomial P_2 must be symmetric, and it must have degree $d - n$. So by induction, $P_2(X_1, \ldots, X_n) = G(\sigma_1, \ldots, \sigma_n)$ where the weight of G is at most $d - n$. It follows that $P_2 = \sigma_n G(\sigma_1, \ldots, \sigma_n)$ and the weight of the polynomial is at most d. QED.

12.4 Transcendence of π

Theorem 43 (Lindemann) *π is transcendental*

Proof: If π is algebraic, so is $i\pi$ and $e^{i\pi} + 1 = 0$. Let $\alpha_1, \ldots, \alpha_n$ be the roots of an irreducible polynomial over Q satisfied by $i\pi$. Then

$$(e^{\alpha_1} + 1)(e^{\alpha_2} + 1) \ldots (e^{\alpha_n} + 1) = 0$$

Rewrite this in the form

$$e^{\beta_1} + \ldots + e^{\beta_r} + k = 0$$

where k is a positive integer and the β_i are partial sums, like $\alpha_3 + \alpha_7 + \alpha_8$, of the $\alpha_1, \alpha_2, \ldots, \alpha_n$. The 1's in our formula correspond to those subsums that add to zero.

We claim there is a polynomial $P(X) = c_0 + c_1 X + \ldots + c_r X^r$ with integer coefficients such that the roots of P are exactly the nonzero β_i.

Indeed, the polynomial $Q(X)$ has integer coefficients and roots α_i. Call this $Q_1(X)$.

Define

$$Q_2(X) = \prod_{i<j} \Big(X - (\alpha_i + \alpha_j) \Big)$$

The coefficients of this polynomial are clearly symmetric polynomials in the α_i with integer coefficients. By the fundamental theorem of symmetric polynomials, each coefficient can be written as a polynomial in the elementary symmetric functions of the α_i with integer coefficients. These elementary symmetric functions of the α_i are exactly the coefficients of the original $Q(X)$, and so each is an integer. It follows that the coefficients of $Q_2(X)$ are integers.

Define

$$Q_3(X) = \prod_{i<j<k} \Big(X = (\alpha_i + \alpha_j + \alpha_k \Big)$$

By the previous argument, this polynomial has integer coefficients.

Continue. In the end, let $P(X) = Q_1(X)Q_2(X)\dots Q_n(X)$. This polynomial has integer coefficients and its roots are all sums of subsets of the α_i. Factor out as many X as possible, and redefine $P(X)$ to be the remaining polynomial, which now has no zero roots.

We are now in a position to apply the ideas used in the proof that e is transcendental. Let r be the degree of $P(X)$ and let c be the coefficient of the highest power of X in $P(X)$. If p is a positive integer, define

$$f(x) = c^{rp-1}x^{p-1}\frac{P(X)^p}{(p-1)!}$$

Define

$$F(x) = f(x) + f'(x) + f''(x) + \dots$$

where the terms eventually vanish because f is a polynomial. Then

$$\frac{d}{dx}\left(e^{-x}F(x)\right) = e^{-x}\left(F'(x) - F(x)\right) = -e^{-x}f(x)$$

and therefore

$$\int_0^y e^{-x}f(x)\ dx = \Big[-e^{-x}F(x)\Big]_0^y = F[0] - e^{-y}F[y]$$

We now diverge slightly from the algebra in the case of e. In the integral on the left, substitute $x = y\lambda$. Then we get $y\int_0^1 e^{-\lambda y}f(\lambda y)d\lambda = F[0] - e^{-y}F(y)$. Multiplying both sides by e^y gives

$$y\int_0^1 e^{y(1-\lambda)}f(\lambda y)\ d\lambda = e^yF(0) - F(y)$$

Substitute β_i for y and sum. Recall that $\sum e^{\beta_i} = -k$. So

$$\sum \beta_i \int_0^1 e^{\beta_i(1-\lambda)}f(\lambda\beta_i)\ d\lambda = -kF(0) - \sum F(\beta_i)$$

As before, we come to the moment when we can derive a contradiction. We claim that for all large primes p, the right hand side is a nonzero integer. But an easy calculation shows that the left side goes to zero as p goes to infinity.

Let us study the left side first. The only term in this expression depending on p is $f(\lambda\beta_i)$. Note that

$$f(x) = c^{r-1}P(X)\frac{(c^rxP(x))^{p-1}}{(p-1)!}$$

We need to evaluate this on a bounded set of the complex plane containing $\lambda\beta_i$ where $0 \le \lambda \le 1$. On this set, x and $P(X)$ are bounded. Clearly, then, the absolutely value of the expression is bounded by

$$B_1\frac{{B_2}^{p-1}}{(p-1)!} = B_1\frac{B_2}{1}\frac{B_2}{2}\cdots\frac{B_2}{p-1}$$

and this expression goes to zero as p goes to infinity.

Finally, consider the right hand side. We claim $-kF(0)$ is an integer not divisible by p for p large enough, and we claim that $\sum F(\beta_i)$ is an integer divisible by p. If these claims hold, then for all sufficiently large p, the right hand side is an integer not divisible by p, and thus a nonzero integer, and we have a contradiction.

Consider first $kF(0) = k\left(f(0) + f'(0) + f''(0) + \ldots\right)$. The only terms that matter at 0 are terms when x^{p-1} was differentiated exactly $p-1$ times. For exactly one of these terms, $P(X)^P$ was not differentiated. For the remaining terms, $P(X)^p$ was differentiated at least once. Differentiating x^{p-1} for $p-1$ times removed the $(p-1)!$ in the denominator; this term is not divisible by p for large p since $P(0) = c_0$ is not divisible by p. The remaining terms involve at least one $pP(X)^{p-1}$ and thus are divisible by p.

Therefore the full term $-kF(0)$ is an integer which is not divisible by p. To finish the argument, it suffices to show that $\sum F(\beta_i)$ is an integer which *is* divisible by p.

Since $P(\beta_i) = 0$, $f^{(i)}(\beta_i)$ can vanish only if the term $P(X)^p$ were differentiated at least p times. Each time a derivative is taken, many other algebraic terms form, but unless a term in the derivative has no $P(X)$ remaining, it will be zero at β_i. Consequently, the only terms in the derivatives which matter contain $p!$, which cancels out the $(p-1)!$ in the denominator and leaves an extra p. What is left after all of this differentiation is a polynomial $Q(X)$, possibly of very large degree, with integer coefficients. But $\sum Q(\beta_i)$ will then be an expression in terms of the form $\beta_1^k \beta_2^k \ldots \beta_b^k$. These are symmetric in the β_i and thus can be expressed as polynomials with integer coefficients in the elementary symmetric functions, and thus polynomials with integer coefficients in the coefficients of $P(X)$. So we get integers divisible by p.

Chapter 13

Special Cases

In this chapter, we obtain much more explicit results about the Galois groups of special extensions.

In particular, we obtain results on cyclotomic extensions. The results are required in the following chapter, when we return to straightedge and compass constructions and discuss regular polygons.

13.1 The Discriminant

Definition 10 *Let* $P(X) = a_0X^n + a_1X^{n-1} + \ldots + a_n$ *be a polynomial with roots* $\lambda_1, \lambda_2, \ldots, \lambda_n$*. The* discriminant *of the polynomial is the expression*

$$D = a_0^{2n-2} \prod_{i<j} (\lambda_j - \lambda_i)^2$$

Remark: This expression is invariant under all permutations of the roots, and thus expressible in terms of the elementary symmetric functions. These elementary functions give the coefficients of $P(X)$, so the discriminant can be written as a polynomial in the coefficients of $P(X)$. For example, the discriminant of $aX^2 + bX + C$ is

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

The discriminant of $aX^3 + bX^2 + cX + d$ is

$$b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$$

In particular the discriminant of $X^3 + aX + b$ is

$$-4a^3 - 27b^2$$

The discriminant of $ax^4 + bx^3 + cx^2 + dx + e$ is

$$256a^3e^3 - 192a^2bde^2 - 128a^2c^2e^2 + 144a^2cd^2e - 27a^2d^4 + 144ab^2ce^2 - 6ab^2d^2e$$

$$-80abc^2de + 18abcd^3 + 16ac^4e - 4ac^3d^2 - 27b^4e^2 + 18b^3cde - 4b^3d^3 - 4b^2c^3e + b^2c^2d^2$$

Theorem 44 *Let $K \subset L$ be a splitting field for an irreducible $P(X)$ without multiple roots over a field whose characteristic is not 2. Recall that the Galois group permutes the roots of P and thus is a subgroup of the symmetric group S_n. The Galois group is a subgroup of the alternating group A_n if and only if the discriminant of P has a square root in K.*

Proof: Since P does not have repeated roots, $D \neq 0$.

The expression

$$\sqrt{D} = a_0^{n-1} \prod_{i<j} (\lambda_j - \lambda_i)$$

belongs to the splitting field and has square D. It belongs to K if and only if it is invariant under every element of the Galois group G. If τ is a permutation of the roots,we clearly have $\tau(\sqrt{D}) = \mathrm{sgn}(\tau)\sqrt{D}$, so $\sqrt{D} \in K$ if and only if every element of the Galois group has sign equal one and thus is in A_n.

13.2 The Cubic Case

The Galois group of a cubic is a transitive subgroup of S_3. The full S_3 is just the dihedral group D_3, the symmetries of an equilateral triangle. This group has order 6, and contains three reflections and three rotations of 0, 120, and 240 degrees. A transitive subgroup must contain either 3 or 6 elements. If the group has three elements, these elements must be the rotations, because otherwise the group would contain two reflections, and any two reflections generate the full group. So the Galois group can only be Z_3 or D_3.

Theorem 45 *Let $P(X)$ be an irreducible cubic over a field of characteristic zero. Then $D \neq 0$. If D is a perfect square in K, the Galois group is Z_3. If D is not a perfect square, the Galois group is D_3*

Proof: This is a special case of the result in the previous section.

Theorem 46 *Let $P(X)$ be a polynomial with real coefficients. If $D < 0$, $P(X)$ has one real root and two conjugate complex roots. If $D = 0$, $P(X)$ has a repeated real root, and in particular is not irreducible. If $D > 0$, $P(X)$ has three real roots.*

Proof: It is clear that $D = 0$ if and only if $P(X)$ has a repeated root. In that case, P' divides P. If P has three real roots, each $(\lambda_j - \lambda_i)$ is real and nonzero, and their product squared is positive. If P has one real root and two complex roots, the complex roots must be conjugate, so the three terms $(\lambda_j - \lambda_i)$ contain one purely imaginary term, and two conjudate terms $e + fi$ and $e - fi$ with $f \neq 0$. The square of the purely imaginary term is negative, and the square of $(e + fi)(e - fi) = e^2 + f^2$ is positive, so $D < 0$.

Remark: Turn back to the solution of a cubic of the form $X^3 + aX + b$ on page 11. Notice that this solution contains $\sqrt{b^2 + \frac{4a^3}{27}} = \frac{1}{3\sqrt{3}}\sqrt{4a^3 + 27b^2} = \frac{1}{3\sqrt{3}}\sqrt{-D}$.

If $D < 0$, then our cubic has one real root and two complex roots. In this case the square root in the formula is real, and the formula involves a cube root of a real number, which always exists as a real number. So we can compute the real root without using complex numbers. In this case the Galois group is D_3.

But if $D > 0$, then there are three real roots, but the square root in the formula is a pure imaginary. We thus must use trigonometric functions to find the cube root. Miraculously, the complex parts of the two cube roots cancel out and the formula gives three real roots. The Galois group can be either Z_3 or D_3.

Remark: To prove the next result, we need

Lemma 10 *Let K be a field of characteristic zero, $a \in K$, and p a prime. Then $P(X) = X^p - a$ is irreducible over K if and only if a is not the pth power of an element in K.*

Proof: If $a = b^p$, then $X^p - a = X^p - b^p$ has a root b and therefore factors as $P(X) = (X - b)S(X)$.

Conversely suppose $P(X) = X^p - a = R(X)S(X)$ over K and the degree of $R(X)$ is $r < p$. Let L be the splitting field of $P(X)$. Then both P and R factor completely in L and roots of R are roots of P. So we can write $R(X) = (X - \lambda_1)\ldots(X - \lambda_r)$. Since the coefficients of this polynomial are in K, $\omega = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_r \in K$. We have $\omega^p = \lambda_1^p \cdot \ldots \cdot \lambda_r^p = a \cdot a \cdot \ldots \cdot a = a^r$.

Since r and p are relatively prime, we can find integers A and B with $1 = Ar + Bp$. Then $a = a^{Ar}a^{Bp} = (a^r)^A a^{Bp} = (\omega^p)^A a^{Bp} = \left(\omega^A a^B\right)^p$. It follows that $\sqrt[p]{a} = \omega^A a^B \in K$. QED.

Theorem 47 (Casus Irreducibilis) *If $F(X)$ is an irreducible cubic with real coefficients and $D > 0$, then the complex roots which occur in the cubic formula are unavoidable. It is impossible to write any root of $F(X)$ as a combination of real roots of various orders.*

Proof Let K be the base field; often $K = Q$. Consider the extension $K \subset K(\sqrt{D})$ where D is the discriminant of $P(X)$. Since $D > 0$, the left side still contains only real numbers.

Since $P(X)$ is irreducible of degree 3 over K, any root field of P has degree 3, and thus $K(\sqrt{D})$ does not contain any root of $P(X)$.

If the theorem is false, we can find a series of extensions

$$K(\sqrt{D}) \subset K_1 \subset \ldots \subset K_n$$

such that each extension is obtained by adding one real radical to the previous field, and such that one root of $P(X)$ belongs to K_n. We can suppose each radial extension adds a pth root for p a prime, since any root is a composition of such roots.

Clearly we can suppose that no root of $P(X)$ is in K_{n-1} by backing up the tower until we come to the transition point.

Let $K_{n-1} \subset L$ be the splitting field of $P(X)$ over K_{n-1}. This is a Galois extension. Since the discriminant D is a square in K_{n-1}, the Galois group is Z_3, so $[L, K_{n-1}] = 3$. But any root field $K_{n-1} \subset K_{n-1}(\lambda) \subset L$ also has degree 3. So any root extension over K_{n-1} contains all the roots of P. Since K_n contains a root, we must have $K_{n-1} \subset L \subset K_n$.

But $[K_n : K_{n-1}] = p$ by the previous lemma. Since $[K_n : K_{n-1}] = [K_n : L][L : K_{n-1}]$, we conclude that $p = 3$ and $L = K_n$. But then K_n is a Galois extension of K_{n-1}, and must contain all roots of $X^3 - a = 0$, which are $\sqrt[3]{a}$, $\sqrt[3]{a}\omega$ and $\sqrt[3]{a}\omega^2$ for $\omega = e^{\frac{2\pi i}{3}}$. Since ω is not real, this is a contradiction. QED.

13.3 Cyclotomic Fields

We begin with cyclotomic extensions, that is, extensions of the form $Q \subset Q(e^{\frac{2\pi i}{n}})$, and thus with the polynomial $P(X) = X^n - 1$. We restrict to a ground field Q and thus in particular to the characteristic zero case.

A common high school exercise factors this polynomial as $(X-1)(X^{n-1}+X^{n-2}+\ldots+1)$, but we can often continue factoring: $X^4-1 = (X^2-1)(X^2+1) = (X-1)(X+1)(X^2+1)$. In

To understand the general situation, notice that the set of all cyclotomic roots is a cyclic group Z_n generated by $\theta = e^{\frac{2\pi i}{n}}$. The elements of order n in this group are called the *primitive* nth roots of unity, and it turns out that they are the roots of an irreducible polynomial $\Phi_n(X)$ with integer coefficients. For instance, the primitive fourth roots of unity are $\pm i$, which are the roots of $X^2 + 1$. For each divisor d of n, the primitive dth roots of unity are exactly the elements of the group of order d . They satisfy $\Phi_d(X) = 0$. It follows that $X^n - 1 = \prod_{d|n} \Phi_x(X)$. For instance, the primitive square root of unity is -1, the root of $X + 1$, the primitive first root of unity is 1, the root of $X - 1$, and $X^4 - 1 = (X - 1)(X + 1)(X^2 + 1)$.

Theorem 48 *The polynomial $X^n - 1$ factors as*

$$\prod_{d|n} \Phi_d(X)$$

where the roots of Φ_d are exactly the primitive dth roots of unity. The polynomials Φ_d, called the cyclotomic polynomials*, are all irreducible over Q with integer coefficients.*

Proof: Define $\Phi_d(X) = \prod(X - \alpha_i)$ over all primitive dth roots of unity. This is a monic polynomial, but for a moment the coefficients are unclear. However, $X^n - 1 = \prod_{d|n} \Phi_d(X)$ for all n because every nth root of unity has some order d in the group of units.

These equations serve to recursively define the Φ_i. We have $\Phi_1 = X - 1$ and $X^2 - 1 = (X-1)(X+1) = \Phi_1(X)\Phi_2(X)$, so $\Phi_2(X) = X + 1$. In general

$$X^n - 1 = \Big(\prod_{d|n;d<l} \Phi_d(X) \Big) \Phi_n(X)$$

where the product is over known monic polynonials with integer coefficients, so by simple division we discover that Φ_n has integer coefficients.

Therefore we need only prove

Lemma 11 *Each $\Phi_n(X)$ is irreducible over Q.*

Proof: The proof isn't particularly difficult, but it is tricky.

Let ω be a primitive nth root of unity, with minimal polynomial $F(X)$ over Q. If we can prove that all primitive nth roots are roots of $F(X)$, then $F(X)$ must be divisible by Φ_n over C. Since both polynomials have rational coefficients, it must be divisible by Φ_n over Q. Since F is minimal, it must be equal to Φ_n.

Every nth root of unity equals ω^k for some k. The new root is primitive if and only if k is relatively prime to n. Indeed if j divides both k and n, then $\left(\omega^j\right)^{n/j} = 1$ and so ω^j is not a primitive root. Conversely if k and n are relatively prime, we can find integers a and b with $1 = ak + bn$ and then $\omega = \left(\omega^k\right)^a$ and so ω^k is primitive.

Therefore every primitive root can be obtained by starting with ω and a prime p not dividing n, forming ω^p, finding a prime q not dividing n, raising the new number to the qth power, etc. So it is enough to prove the following lemma:

Lemma 12 *Let $F(X)$ be the minimal polynomial for a primitive nth root of unity ω. Let p be a prime which does not divide n. Then $F(\omega^p) = 0$, ω^p is a primitive nth root, and $F(X)$ is the minimal polynomial for ω^p.*

Note that the lemma is clear provided we can prove that $F(\omega^p) = 0$.

Since ω is a root of $X^n - 1$, $F(X)$ divides $X^n - 1$ and $X^n - 1 = F(X)G(X)$. We claim F and G are monic with integer coefficients. This follows from Gauss' lemma on primitive polynomials.

If ω^p is not a root of F, then it must be a root of G. So ω is a root of $G(X^p)$. Since $F(X)$ is the minimal polynomial of ω, it follows that $G(X^p) = F(X)H(X)$. Repeating the Gauss technique, we see that since $G(X^p)$ has integer coefficients, so do $F(X)$ and $H(X)$.

Reduce modulo p to discover that in $Z_p[X]$ we have $X^n - 1 = F(X)G(X)$ and $G(X^p) = G(X)^p = F[X]H[X]$. Here we use the fact that $G(X^p) = G(X)^p$ over Z_p.

But then any irreducible factor of $F(X)$ is a factor of $G(X)^p$ and thus a factor of $G(X)$, so $X^n - 1$ has a repeated factor and thus over some extension field of Z_p, a multiple root. But this is impossible because $X^n - 1 = (X - \alpha)^2 K(X)$ implies by formal differentiation that $nX^{n-1} = 2(X - \alpha)K(X) + (X - \alpha)^2 K'(X)$ and so $n\alpha^{n-1} = 0$. This is impossible because $\alpha^n - 1 = 0$ implies $\alpha \neq 0$ and $n \neq 0$ because p does not divide n. QED.

Remark: Once we know that $\Phi_n(X)$ is irreducible over Q, we get a Galois extension with a known Galois group. Indeed

Theorem 49 *Let L be the splitting field of $X^n - 1$ over Q. Then L is a Galois extension, and its Galois group is the group of units in Z_n.*

Proof: Let ω be a primitive nth root of unity and consider the root field of $\Phi_n(X)$. This field contains all powers of ω and hence all roots of unity. In particular, the irreducible Φ_n splits completely, so it is a splitting field over a field of characteristic zero, and hence a Galois extension.

An automorphism of this field is completely determined by the image of ω. An image must be another root of $\Psi_n(X)$ and so another primitive nth root of unity. Since the Galois group is transitive, all such roots occur.

Thus an automorphism has the form $\omega \to \omega^k$ where k is only unique modulo n. The only k that can occur are those relatively prime to n. Said another way, k induces a unique element of Z_n and this element is a unit.

If j and k represent the automorphisms $\omega \to \omega^j$ and $\omega \to \omega^k$, then their product is given by

$$\omega \to \omega^k \to \left(\omega^j\right)^k = \omega^{jk}$$

so the group product is given by multiplication in the ring Z_n.

Remark: The determination of the group structure of the ring of units in Z_n is a famous topic of elementary number theory. If $n = p$ is prime, the group of units is $Z_p^\star$, which is isomorphic to Z_{p-1}. Other interesting cases are $Z_4^\star = Z_2$ and $Z_8^\star = Z_2 \times Z_2$. So the resulting group is not always cyclic.

The general result is given by the following theorem, which we will not prove:

Theorem 50

- *If $n = p_1^{k_1} p_2^{k_2} p_3^{k_3} \ldots$, then*

$$Z_n^\star = \left(Z_{p_1^{k_2}}\right)^\star \times \left(Z_{p_1^{k_2}}\right)^\star \times \left(Z_{p_1^{k_2}}\right)^\star \times \ldots$$

- *If p is odd,*

$$\left(Z_{p^k}\right)^\star = Z_{p^k(p^k-1)}$$

-

$$(Z_2)^\star = \{e\}$$

-

$$(Z_4)^\star = Z_2$$

- *For $k \geq 3$,*

$$(Z_{2^k})^\star = Z_2 \times Z_{2^{k-2}}$$

Remark: In particular, the Galois group of any cyclotomic extension is abelian. It follows that all subgroups are normal and abelian, and therefore any subgroup of the cyclotomic group produces a Galois extension $Q \subset K$ with abelian Galois group.

In 1853, Kronecker proved the converse of this theorem: Every Galois extension of Q with abelian Galois group is a subfield of $Q(\zeta_n)$ for some primitive root of unity ζ_m.

Kronecker's proof had gaps, by his own admission. The gaps were supposedly filled by Weber in 1886, but this proof also had an error. In 1896, Hilbert gave the first complete proof.

Every quadratic extension $Q \subset K$ has Galois group Z_2 and thus can be embedded in a cyclotomic extension. By comparing the factorization of prime ideals (p) in the quadratic extension and in the cyclotomic extension, Gauss' law of quadratic reciprocity turns out to be a consequence of Kronecker's theorem.

Weber's theorem led to the development of *class field theory*, which is essentially the number theory of Galois extensions $K \subset L$ with abelian Galois groups. This subject was developed in the first half of the twentieth century by Hilbert, Takagi, Artin, Hasse, and Chevalley. See the web document *History of Class Field Theory* by Keith Conrad. An important subfield of modern number theory flows from attempts to generalize class field theory to non-abelian extensions.

13.4 Galois Group of $X^p - a$ over Q

In our discussion of solving equations with radicals, we fixed a base field K containing all nth roots of unity. We then proved that the splitting field of $X^n - a$ over K has a cyclic Galois group.

We now generalize to the case when these roots of unity are missing. We'll consider the special case where $P(X) = X^p - a$, p is prime, and the base field is Q. Assume that a is not a pth power in Q. We'll determine the splitting field of P completely and compute its Galois group.

Since a is not a pth power, $P(X)$ is irreducible over Q by an earlier result, so the splitting field contains all of its roots and is generated by these roots. The roots of $P(X)$ are $\omega^k \sqrt[p]{a}$ where ω is a non-trivial pth root of unity. Dividing two roots, we find that the splitting field contains all roots of unity. So it is natural to construct the splitting field as an extension chain

$$Q \subset Q(\omega) \subset Q(\omega \sqrt[p]{a})$$

We know that $Q \subset Q(\omega)$ is a Galois extension with multiplicative Galois group $Z_p^\star$; this group is cyclic and isomorphic to Z_{p-1}. We'll continue to write it multiplicatively; in that case the nonzero $m \in Z_p^\star$ corresponds to the automorphism determined by $\omega^i \rightarrow \omega^{mi}$.

We know that $Q(\omega) \subset Q(\omega, \sqrt[p]{a}$ is also Galois with additive Galois group Z_p. The automorphism corresponding to $n \in Z_p$ maps $\omega^i \sqrt[p]{a} \rightarrow \omega^{i+n} \sqrt[p]{a}$.

Since there are p roots, the Galois group is a subgroup of the permutation group S_p. This group can have more than p elements; indeed $p \leq |G| \leq p!$. Moreover, $|G| = [Q(\omega, \sqrt[p]{a}) : Q]$. It is natural to suppose that the Galois group is some sort of product of the Galois groups of extensions in the above chain, and thus has order $p(p-1)$. Indeed $Q(\omega, \sqrt[p]{a})$ must contain higher powers of $\sqrt[p]{a}$. Writing $\alpha = \sqrt[p]{a}$, we rapidily discover that a basis is

$$\left\{ \omega_i \alpha^j \mid 0 \leq i \leq p-2 \text{ and } 0 \leq j < p \right\}$$

We require all of those powers of α because a root field of $X^p - a$ will have dimension p (recall that the root field adds only one root, not all roots). We only require powers of ω up to the $p-2$ power because ω satisfies $X^{p-1} + X^{p-2} + \ldots + X + 1 = 0$ and thus ω^{p-1} is a combination of lower powers of ω.

We now extend the operations of $m \in Z_p^\star$ and $n \in Z_p$ on roots of P in Q to the full field. Consider first $n \in Z_p$. We know this element fixes $Q(\omega)$ and maps α to $\omega^n \alpha$. It follows that it maps $\alpha^j \to \omega^{nj}\alpha^j$. So the full extension is

$$n \in Z_p : \quad \omega^i \alpha^j \to \omega^{i+nj}\alpha^j$$

Next consider $m \in Z_p^\star$. We know it maps $\omega^i \to \omega^{mi}$. Moreover, it maps $\alpha \to \omega^n \alpha$ for some n, since it preserves roots of $P(X)$. We can multiply the automorphism by any automorphism fixing $Q(\omega)$ and get another possible extension. One such automorphism maps $\alpha \to \omega^{-n}\alpha$. The composition of these maps fixes α. Let us take this extension. So

$$m \to Z_p^\star : \quad \omega^i \alpha^j \to \omega^{mi}\alpha^j$$

Form the set $Z_p^\star \times Z_p$. Typical elements of this product have the form $< m, n >$. Define an action of this element on the basis elements by letting m act first and following with the action of n:

$$\omega^i \alpha^j \to \omega^{mi}\alpha_j \to \omega^{mi+nj}\alpha^j$$

In particular, on roots $< m, n >$ acts via

$$< m, n >: \quad \omega^i \alpha \to \omega^{mi}\alpha \to \omega^{mi+n}\alpha$$

Now notice that

$$< m_2, n_2 >< m_1, n_1 >: \quad \omega^i \alpha \to \omega^{m_1 i + n_1}\alpha \to \omega^{m_2(m_1 i + n_1) + n_2}\alpha$$

We conclude that $Z_p^\star \times Z_p$ becomes a group with product

$$< m_2, n_2 >< m_1, n_1 >=< m_2 m_1, m_2 n_1 + n_2 >$$

This is a semi-direct product induced by $\varphi : Z_p^\star \to \mathrm{Aut}(Z_p)$ where $\varphi(m)(n) = mn$. Notice that $\varphi(m)$ preserves addition and is one-to-one, hence onto.

Note: In general, let G_1 and G_2 be groups and let $\varphi : G_1 \to \mathrm{Aut}(G_1)$ be a group homomorphism. We define the semi-direct product of the two groups with respect to φ to be the set $G_1 \times G_2$ with product

$$< \tilde{g}_1, \tilde{g}_2 >< g_1, g_2 >=< \tilde{g}_1 g_1, \varphi(\tilde{g}_1)(g_2) + \tilde{g}_2 >$$

It is easy to see that this semidirect product is a group. Both G_1 and G_2 are subgroups of this group. The group G_2 is a normal subgroup.

In our case $Z_p^\star \times Z_p$, the Z_p is a normal subgroup of the Galois group whose fixed field is $Q(\omega)$. Therefore $Q \subset Q(\omega)$ is Galois with Galois group $(Z_p^\star \times Z_p)/Z_p$. Clearly this quotient is $Z_p^\star$, as it ought to be.

Chapter 14

Constructing Regular Polygons

14.1 Constructing Regular Polygons

It is easy to inscribe a regular hexagon in a circle: take the compass used to draw the circle, and mark off six equal segments with the compass. Each straight segment joining two adjacent points has the same radius as the circle, so the interior angle is 60° and therefore six of these segments exactly fit around the circle. By joining every other point, a regular 3-gon can be inscribed.

Similarly, it is easy to inscribe a square. Draw the diameter, and the perpendicular to it at the center, and join the four intersection points of these lines with the circle. Bisecting each angle gives a regular 8-gon.

It is not so easy to inscribe a pentagon. Here's one way to proceed if you have completely forgotten what you learned in high school. We need to construct $\cos(\frac{2\pi}{5}) = \cos\theta$. Then $(\cos\theta + i\sin\theta)^5 = 1$. Taking imaginary parts, $5i\cos^4(\theta)\sin(\theta) - 10i\cos^2(\theta)\sin^3(\theta) + i\sin^5(\theta) = 0$. Factoring out $i\sin(\theta)$, we obtain

$$\left(5\cos^4(\theta) - 10\cos^2(\theta)(1-\cos^2(\theta)) + (1-\cos^2(\theta))^2\right) = 0$$

or

$$16\cos^4(\theta) - 12\cos^2(\theta) + 1 = 0$$

So $\cos\theta$ satisfies $16X^4 - 12X^2 + 1$. Consequently $X^2 = \frac{12\pm\sqrt{144-64}}{32} = \frac{12\pm\sqrt{80}}{32} = \frac{3\pm\sqrt{5}}{8}$. Since $\frac{3+\sqrt{5}}{8} = \frac{1}{4}\left(\frac{1+\sqrt{5}}{2}\right)^2$, we have $\cos\theta = \frac{1}{2}\left(\frac{1+\sqrt{5}}{2}\right)$. This number can be constructed by our methods, so a regular pentagon can be inscribed in a circle. Consult the internet for many beautiful constructions. Incidentally, $\frac{1+\sqrt{5}}{2}$ is the Golden Mean.

Theorem 51

- *If an n sided polygon can be inscribed in a circle using straightedge and compass, then a $2^k n$ sided polygon can also be inscribed*
- *If an m sided polygon and an n sided polygon can be inscribed and m and n are relatively prime, then an mn sided polygon can be inscribed.*

Proof: Bisect the angles of a polygon to double the number of sides. If m and n are relatively prime, integers a and b exist with $am + bn = 1$. These integers may be positive or negative or both. Then

$$a\left(\frac{2\pi}{n}\right) + b\left(\frac{2\pi}{m}\right) = (am + bn)\left(\frac{2\pi}{mn}\right) = \frac{2\pi}{mn}$$

Hence we can obtain a side of the mn-sided polygon by taking a multiple of the angle of one of the polygons and reversing by a different multiple of the angle of the other. QED.

Remark: It follows that up to 20, we can construct polygons with 3, 4, 5, 6, 8, 10, 12, 15, 16, and 20 sides, leaving 7, 9, 11, 13, 14, 17, 18, and 19 in doubt.

When he was very young, Gauss succeeded in constructing a 17 sided polygon. This was the first progress in the theory since the Greeks, and it caused Gauss to decide to become a mathematician. The theory of this polygon and its generalizations is covered in the last chapter of Gauss's book on number theory, written when he was 24 years old.

Theorem 52 *A polygon with n sides can be constructed using straightedge and compass if and only if*

$$n = 2^k p_1 p_2 \ldots p_k$$

where the p_k are distinct odd primes of the form $2^{2^n} + 1$.

Remark: This theorem rules out $n = 9, 18$ since the factorization of each contains a repeated off prime.

Primes of the form $2^{2^n} + 1$ are called *Fermat primes* because Fermat conjectured that all such numbers are prime. The first few such numbers are 3, 5, 17, 257, and 65537 and each is prime. The next number is 4294967297 which was too large to factor by hand in Fermat's day. Euler later proved that it is not prime by showing that it does not satisfy Fermat's little theorem for 3. No other prime has been found in the sequence of Fermat numbers, but it is unknown whether others exist. If one assumes that the $2^{2^n} + 1$ are random and takes into account the rough distribution of the primes, the likely number of primes in the sequence is only 4 or 5, but this is just a heuristic argument.

Proof: Starting with $(0, 0)$ and $(1, 0)$, we want to construct $\left(\cos\frac{2\pi}{n}, \sin\frac{2\pi}{n}\right)$. It suffices to construct the cosine, for then the sin can be found by taking one square root. Starting from Q, suppose a sequence of quadratic extensions can be found with the final one containing $\cos\frac{2\pi}{n}$. By taking one more quadratic extension (this one complex), we can get a tower whose last element contains $\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n} = \omega$ where ω is a primitive nth root of unity. The dimension of the final element of the tower over Q is a power of 2, so $[Q(\omega) : Q]$ must equal a power of 2.

In section 12.3 on cyclotomic fields, we discovered that $X^n - 1 = \prod_{d|n} \Phi_d(X)$ where each Φ_d is irreducible over Q. Moreover $[Q(\omega) : Q] = \text{degree}(\Phi_n(X))$. The numbers $\varphi(d) = \text{degree}(\Phi_d(X))$ define what is called the *Euler Phi Function.* Recall that $\varphi(d)$ is the number of generators of Z_d, or alternately the number of $a \in Z_d$ with a relatively prime to d.

Lemma 13

- *If p is prime, $\varphi(p^k) = p^k - p^{k-1} = p^k\left(1 - \frac{1}{p}\right)$*
- *if m and n are relatively prime, $\varphi(mn) = \varphi(m)\varphi(n)$*

Proof of theorem from lemma: Factor n into primes: $n = 2^k p_1^{k_1} \ldots p_m^{k_m}$ Then $\varphi(n)$ is a product of terms and each must be a power of 2 if an n-sided can be constructed. We have $\varphi(2^k) = 2^k - 2^{k-1} = 2^k(2-1)$, which is a power of 2. We have $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$. If $k > 1$, the first term is odd and the polygon is ruled out. If $k = 1$, the $p-1$ must be a power of 2, and thus $p = 2^k + 1$. But such a number cannot be prime unless k is itself a power of 2, for if b is odd prime then for

$$2^{2^k b} + 1 = (2^{2^k})^b + 1 = \left((2^{2^k} + 1\right)\left(\left(2^{2^k}\right)^{b-1} - \left(2^{2^k}\right)^{b-2} + \ldots - \left(2^{2^k}\right) + 1\right)$$

Consequently the condition of the theorem is necessary.

To prove that this condition is also sufficient, it suffices to study the case when $n = p = 2^{2^n}+1$ by the previous theorem. Consider $Q \subset Q(\omega)$. This is a Galois extension, with Galois group $Z_p^\star = Z_{p-1}$. Since $p = 2^{2^n}+1$, $p-1$ is a power of 2. The subgroups of Z_{p-1} correspond to divisors of this power of two, and therefore form a single chain $e \subset Z_2 \subset Z_4 \subset \ldots \subset Z^{p-1}$. Therefore the subfields of $Q(\omega)$ form a single chain $Q \subset K_1 \subset K_2 \subset \ldots \subset K_{n-1} \subset Q(\omega)$ and there are no other subfields to mess up this picture.

But we can easily find a candidate for K_{n-1}, namely $Q(\cos\frac{2\pi}{p})$. Indeed, the extension is generated by $X = i\sin\frac{2\pi}{p}$ and $X^2 = -\sin^2\frac{2\pi}{p} = \cos^2\frac{2\pi}{p} - 1$. So K_{n-1} is a field of real numbers, and therefore every other extension in the chain was obtained by adding a real square root. So $\cos\frac{2\pi}{p}$ can be constructed by straightedge and compass. QED.

Proof of lemma: Recall that $\varphi(p^n)$ is the number of $a \in Z_{p^n}$ relatively prime to p^n. There are p^n elements altogether; those not relatively prime to p^n are multiples of p and so those of the form pb for $1 \leq b < p^{n-1}$. Since there are p^{n-1} of the latter, there are $p^n - p^{n-1}$ relatively prime to p^n.

Suppose m and n are relatively prime. Then Z_{mn} is isomorphic to $Z_m \times Z_n$.

An element of Z_s is relatively prime to s if and only if it is a unit in the multiplicative group $Z_s^\star$. Indeed if a and s are relatively prime, we can find A and B with $Aa + Bs = 1$. Then in Z_s we have $Aa = 1$, so a is a unit. Conversely if $Aa = 1$ in Z_s, then $Aa - 1 = Bs$, so a and s are relatively prime.

The number of units in Z_{mn} is $\varphi(mn)$. Clearly $(a, b) \in Z_m \times Z_n$ is a multiplicative unit if and only if a is a unit in $Z_m^\star$ and b is a unit in $Z_n^s tar$ and the number of such units is $\varphi(m)\varphi(n)$.

Remark: We can now say more about trisection of angles:

Theorem 53 *There is a straightedge and compress construction to divide every angle into n equal pieces if and only if n is a power of 2.*

Proof: If there were such a construction and we applied it to the angle 2π, we'd obtain the vertices of an inscribed n-gon. Or analogously, we could apply the construction to a right angle and get a $4n$-gon, from which an n-gon is easily obtained. So $n = 2^k p_1 \dots p_k$ where the p_i are distinct odd primes. By connecting vertices with similar gaps, we could construct a p-gon. But then by dividing each interior angle into p pieces we'd have a p^p gone, and odd primes would be repeated, which is impossible. So no p_i can occur. QED.

Remark: In high school, we learn that a small number of angles have precise trigonometric values. For example, $\cos 0^\circ = 1$, $\cos 30\circ = \frac{\sqrt{3}}{2}$, $\cos 45^\circ = \frac{1}{\sqrt{2}}$, $\cos 60\circ = \frac{1}{2}$. An earlier result in these notes gave $\cos 72\circ = \frac{1}{2}\frac{1+\sqrt{5}}{2}$.

Gauss' results show that a small number of other values can be expressed using only square roots. In his number theory book, Gauss worked out the exact value of $\cos \frac{2\pi}{17}$. The value is

$$\cos \frac{2\pi}{17} = \frac{1}{16}\left[-1 + \sqrt{17} = \sqrt{34 - 2\sqrt{17}} + 2\sqrt{17 + 3\sqrt{17} - \sqrt{34 - 2\sqrt{17}} - 2\sqrt{34 + 2\sqrt{17}}}\right]$$

Chapter 15

Normal and Separable Extensions

15.1 Normal Extensions, Separable Extensions, and All That

In traditional treatments of Galois theory, an extension $K \subset L$ is called a Galois extension if L is *normal* and *separable* over K.

Definition 11 *A finite extension $K \subset L$ is said to be* normal *if whenever $Q(X)$ is an irreducible polynomial over K with a root in L, then $Q(X)$ splits completely over L.*

Theorem 54 *Let $K \subset L$ be a finite extension.*

- *If L is the splitting field of some polynomial $P(X)$, then L is normal.*
- *Conversely, if L is normal, then L is the splitting field of some polynomial $P(X)$ over K.*

Proof: Suppose L is a splitting field of $P(X)$, and suppose $Q(X)$ is irreducible over K and has a root in L. Let $K \subset M$ be the splitting field of $Q(X)P(X)$. Then M contains all roots of $P(X)$, and these generate L, so $K \subset L \subset M$. Suppose a is a root of $Q(X)$ in L. By assumption $Q(X)$ splits completely in M; let $b \in M$ be a second root.

By corollary 3 in section 4.3, there is an automorphism σ of M over K taking a to b. Since σ is an automorphism over K, it must map roots of $P(X)$ to other roots. But the roots of $P(X)$ generate L, so σ maps L to L. Hence $b \in L$. This holds for any root of $Q(X)$, so $Q(X)$ splits completely in L.

The other direction is essentially trivial. If $L = K$, we are done. Otherwise find $a_1 \in L$ and let $P_1(X)$ be its minimal polynomial over K. By assumption all roots of Q are in L. We have $K \subset K(a_1) \subset L$. If $K(a_1)$ is not L, find $a_2 \in L$ not in $K(a_1)$. Let $P_2(X)$ be its minimal polynomial. This polynomial splits completely in L. Then $K \subset K(a_1) \subset K(a_1, a_2) \subset L$. Continue. The sequence of fields is strictly increasing, so eventually we get L, and then $P(X) = P_1(X)P_2(X)\ldots$ is the splitting polynomial.

Definition 12 *Let $P(X)$ be a polynomial over a field K. We say $P(X)$ is* separable *if all of its roots are distinct in the splitting field of $P(X)$.*

Remark: So $P(X)$ is separable if and only if the roots are "separated" in L, i.e., there are no multiple roots.

Definition 13 *Let $K \subset L$ be a finite extension. We say the* extension is separable *if whenever $\theta \in L$ and $P(X)$ is its minimal polynomial, then $P(X)$ is separable.*

Theorem 55 *A finite extension $K \subset L$ is a Galois extension if and only if it is normal and separable.*

Proof: Suppose $K \subset L$ is normal and separable. Since the extension is normal, there is a polynomial $P(X)$ such that L is the splitting field of P over K. Factor P into irreducible polynomials $P_1(X)\ldots P_k(X)$ over K. If there are redundant factors, we can eliminate repetitions, so assume the factors are distinct.

Since the extension is separable, no irreducible factor has repeated roots. Suppose both P_i and P_j share a root. Then both are minimal polynomials of this root, and by uniqueness of the minimal polynomial they are equal. So $P(X)$ is a splitting polynomial with no repeated roots, and thus $K \subset L$ is Galois.

Conversely, suppose $K \subset L$ is Galois. Then there is a polynomial $P(X)$ making L a splitting field of P, so the extension is normal.

The most difficult step comes last; we want to prove that L is separable. Let $\alpha \in L$ and let $Q(X)$ be its minimal polynomial. We need to prove that all roots of Q are distinct. Since $K \subset L$ is normal, all roots of Q are in L.

Extend α to a set of generators of L: $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_n$. No assumption is being made about where these generators come from. We have a chain

$$K \subset K(\alpha_1) \subset K(\alpha_1, \alpha_2) \subset \ldots \subset K(\alpha_1, \ldots, \alpha_n) = L$$

For each i, let $Q_i(X)$ be the minimal polynomial of α_i over $K(\alpha_1, \ldots, \alpha_{i-1})$. In particular, $Q(X) = Q_1(X)$. Normality of $K \subset L$ again implies that all roots of $Q_i(X)$ are in L.

Let n_i be the number of distinct roots of Q_i in L and let d_i be the degree of Q_i. Clearly $n_i \leq d_i$.

We have $[L : K] = [K(\alpha_1, \ldots, \alpha_n) : K(\alpha_1, \ldots, \alpha_{n-1}] \ldots [K(\alpha_1) : K] = d_n d_{n-1} \ldots d_n$. Since $K \subset L$ is Galois, this is also the number of automorphisms of L. On the other hand, an automorphism σ must map α_1 to another root of Q_1. There are n_1 choices for this root, and thus n_1 homomorphisms $\varphi_i : K(\alpha_1) \to L$. We then extend these maps to $K(\alpha_1, \alpha_2)$ by choosing possible images of α_2. There are n_2 such choices, and so $n_1 n_2$ possible homomorphisms from $K(\alpha_1, \alpha_2)$ into L. Continue. We conclude that the total number of automorphisms of L is at most $n_1 n_2 \ldots n_n$. This can only equal $d_1 d_2 \ldots d_n$ is each $n_i = d_i$, and in particular $n_1 = d_1$. So $Q(X)$ cannot have repeated roots. QED.

Remark: Now the theorem implying that separability is a weak requirement, often automatically true.

Theorem 56

- *In characteristic zero, every irreducible polynomial is separable and every extension is separable.*
- *In characteristic p, an irreducible polynomial $P(X)$ is separable unless there is a $Q(X)$ with $P(X) = Q(X^p)$.*
- *If every element in K is a pth power, then every irreducible polynomial is separable and every extension is separable.*
- *If K is a finite field, every irreducible polynomial is separable and every extension is separable.*

Proof: Let α be a repeated root of an irreducible $P(X)$. Since $P(X)$ is irreducible, it is the minimal polynomial for α. But if α is a repeated root, then it is also a root of the formal derivative $P'(X)$, which cannot vanish in characteristic zero. Contradiction.

In characteristic p, the same argument works unless $P'(X)$ is identically zero. This can only happy if the only terms in P are $(X^p)^k$.

In a field of characteristic p, the map $\sigma : K \to K$ defined by $k \to k^p$ is a one-to-one homomorphism. Indeed, $(a+b)^p = a^p + b^p$ because all remaining binomial coefficients are divisible by p; clearly $(ab)^p = a^p b^p$. If $a^p = 0$, then $a = 0$.

Suppose each element of K is a pth power. Then if an irreducible polynomial $P(X)$ is not separable, it equals

$$Q(X^p) = a_0 + a_1 X^p + \ldots + a_k (X^p)^k = b_0^p + b_1^p X^p + \ldots + b_k^p (X^p)^k = \left(b_0 + b_1 X + \ldots + b_k X^k\right)^p$$

but this is not irreducible.

If $K = Z_p$, and $a \in K$, then $a^p = a$ by Fermat's Little Theorem. If L is a finite field of characteristic p, then $Z_p \subset K$ is finite dimensional and $\sigma(k) = k^p$ is a one-to-one map from $K \to K$, hence onto. QED.

Corollary 6 *If K has characteristic zero or K is a finite field, then every splitting field $K \subset L$ of an arbitrary polynomial $P(X)$ is a Galois extension.*

Proof: Obvious.

Definition 14 *Let $K \subset L$ be a finite extension. A* primitive element *is a $\theta \in L$ such that $L = K(\theta)$.*

Remark Several of our results guarantee the existence of a primitive element. They are all covered by the following:

Theorem 57 *If $K \subset L$ is finite and separable, it has a primitive element.*

Proof: The theorem is obvious if K and L are finite fields because $L^\star$ is cyclic. So we can assume that K is infinite.

Write $L = K(\alpha_1, \alpha_2, \ldots \alpha_n)$ where the α_i have no special properties. Notice that $K(\alpha_1, \ldots, \alpha_i)$ is separable for each i. We prove that each $K(\alpha_1, \ldots, \alpha_i)$ has a primitive element inductively. Clearly, it suffices to study the case $K(\alpha, \beta)$. In this case, we look for a primitive element of the form $\alpha + c\beta$.

Let $P_1(X)$ be the minimal polynomial of α over K and let $P_2(X)$ be the minimal polynomial of β over K and let $P(X)$ be the product of these polynomials. Then $P(X)$ is a polynomial over $K(\alpha, \beta)$. Let S be its splitting field, so

$$K \subset K(\alpha) \subset K(\alpha, \beta) \subset S$$

The key observation of the proof is that we can find exactly $[K(\alpha, \beta) : K]$ one-to-one homomorphisms $\sigma_i : K(\alpha, \beta) \to S$. Suppose for a moment that this has been done.

Next we find $c \in K$ such that $\sigma_i(\alpha + c\beta) \neq \sigma_j(\alpha + c\beta)$ for all $i \neq j$. This is easy. If the two expressions are equal, then

$$\sigma_i(\alpha) - \sigma_j(\alpha) = (-c)\,(\sigma_i(\beta) - \sigma_j(\beta))$$

This equation cannot hold if $\sigma_i(\beta) - \sigma_j(\beta)$ because it would imply that σ_i and σ_j agree on both α and β and thus everywhere, but $\sigma_i \neq \sigma_j$. Otherwise the equation rules out exactly one c for each i and j, and so only finitely many c's. Since K is infinite, we can find an appropriate c.

Consider

$$K \subset K(\alpha + c\beta) \subset K(\alpha, \beta) \subset S$$

Let $R(X)$ be the minimal polynomial of $\alpha + c\beta$ over K and let it have degree d. Since R has a root in the separable $K(\alpha, \beta)$, its roots in S are distinct. Any homomorphism $K(\alpha+c\beta) \to S$ must map $\alpha+c\beta$ to a root of R, so there are at most d such homomorphisms. But there are $[K(\alpha, \beta) : K]$ homomorphisms, so $[K(\alpha, \beta) : K] \leq d = [K(\alpha + c\beta) : K]$ and we conclude that $K(\alpha + c\beta) = K(\alpha, \beta)$.

Now we prove the existence of the σ_i.

Since $K \subset L$ is separable and $P_1(X)$ is irreducible over K, all of its roots are distinct in L. Call them $\alpha = \alpha_1, \ldots, \alpha_{d_1}$, where d_1 is the degree of $P_1(X)$. We can find exactly n one-to-one homomorphisms $\varphi_i : K(\alpha) \to S$, since such a map is determined by the image of α, which can be any root of $P_1(X)$.

Our goal is now to extend the φ_i to one-to-one homomorphisms $K(\alpha, \beta) \to S$, and count the number of ways this can be done. The polynomial $P_2(X)$ is irreducible over K, so all of its roots are distinct in S. It might be reducible in $K(\alpha_i)$. Factor $P_2(X)$ over $K(\alpha)$ and let $Q_2(X)$ be the irreducible factor with root β. Since P_2 has coefficients in K, $P = \varphi_i(P)$ factors in $K(\alpha_i)$, and $\varphi_i(Q_2(X))$ is a factor over $K(\alpha_i)$. Call the roots of this polynomial β_{ij} in S. For a fixed i, these roots are distinct and their number d_2 is the degree of $\varphi_i(Q_2(X))$, which equals the degree d_2 of $Q_2(X)$. We can extend φ to a map $K(\alpha, \beta) \to S$ sending α to α_i and sending β to some β_{ij}. The number of such maps is $d_1 d_2 = [K(\alpha : K][K(\alpha, \beta) : K(\alpha)]$. QED.

Corollary 7 *Every finite $K \subset L$ has a primitive element if the characteristic is zero, or if the fields are finite.*

Remark: Let $P(X)$ be a polynomial, and let $K \subset L$ be its splitting field. The Galois group of P is then a subgroup of the permutation group on the roots of P. We have found cases when this group is the full S_5. It is thus highly transitive, since any 5 elements can be taken to any other 5 elements.

This is a particular “permutation representation” of G, but it is not the only one. By the previous theorem, we can write the splitting field as $K \subset K(\gamma)$. Let $M(X)$ be the minimal polynomial of γ. Then each automorphism is completely determined by the value on γ, so the Galois group is a simply transitive subgroup on $5! = 120$ elements.

The Galois group is an abstract group; it makes sense to think of it as a permutation group only when we have a particular generating polynomial for the extension.

Chapter 16

Galois Theory and Reduction Modulo p

16.1 Preview

We often want to prove that a particular $P(X)$ over Q is irreducible and then find its Galois group. In these notes, we introduced a small number of tools to do these tasks. For instance, irreducibility was mainly proved using Eisenstein's theorem.

In this chapter, we discuss new methods for both calculations. The methods involve reducing modulo a prime p. We factor $P(X)$ over Z_p and compute the Galois groups of the resulting irreducible factors. Brute force methods and computer programs are available for these tasks. By varying the prime p, we gain additional information. Then this information can be assembled to provide information about the original problems over the rationals.

Consider the problem of factoring over Q. When a polynomial has rational coefficients, it does not make sense to reduce modulo p. But luckily Gauss associated to each polynomial with rational coefficients an associated primitive polynomial with integer coefficients, and proved that factoring the primitive polynomial over the integers is equivalent to factoring the original polynomial over the rational numbers. We can certainly reduce the primitive polynomial modulo a prime p. Suppose the resulting polynomial is irreducible. Then so is the primitive polynomial over Z, and also the original polynomial over Q.

But even if the reduced polynomial factors, we gain important information. For instance, suppose a primitive polynomial factors as a cubic times a quartic over Z_5, and factors as a quadratic times an irreducible polynomial of degree 6 over Z_7. Factorization over Z must be consistent with both of these results, and a little thought shows that our polynomial must be irreducible.

At the moment, it is unclear how we could start with a splitting field $Q \subset L$ and reduce modulo p. The essential idea is to introduce a notion of "integers" in both fields, so the field extension induces a ring inclusion $I(Q) = Z \subset I(L)$ and then reduce modulo p. Details we will provided in a later section. The ultimate theorem we obtain is due to Dedekind:

Theorem 58 (Dedekind) *Let $P(X)$ be a monic irreducible polynomial with integer coefficients. Suppose the reduction of P modulo a prime p factors as a product of irreducible polynomials over Z_p of degrees $d_1, d_2, \ldots, d_k$ over Z_p. Suppose this reduced polynomial has distinct roots in its splitting field over Z_p. If we regard the Galois group of the original P as a permutation group on its roots, then this group has an element τ that can be written as a product of k distinct cycles of degrees $d_1, d_2, \ldots, d_k$.*

16.2 $X^5 - X - 1$

We postpone the proof of Dedekind's theorem to the end of this chapter, and begin with a series of applications.

The most famous example of a polynomial which cannot be solved by radicals is

$$P(X) = X^5 - X - 1$$

This example was discovered by Artin. It has only one real root, so the techniques of section 9.6 do not suffice to find its Galois group. We will prove it irreducible with Galois group S_5.

We begin by proving this polynomial irreducible. The symbolic algebra program Mathematica has a built-in function to Factor polynomials modulo a prime. Let's factor over Z_5. Here is the command to run the function, and the resulting output:

```
Factor[X^5 - X - 1, Modulus -> 5]
X^5 + 4X + 4
```

It follows that $X^5 - X - 1$ is irreducible over Z_5, and thus irreducible over Q.

Just this once, imagine factoring over Z_5 by brute force. If $a \in Z_5$, $a^5 = a$. Therefore, for any a we have $a^5 - a - 1 = -1$. If follows that the reduced polynomial modulo 5 has no linear factors. If it factors at all, it must have one quadratic and one cubic factor.

W can ignore quadratic candidates which factor. So it suffices to divide by $X^2 + aX + b$ when $\sqrt{a^2 - 4b} \notin Z_5$, and thus when $a^2 + b = 2, 3$. If $a = 0$, this gives $X^2 + 2$ and $X^2 + 3$. If $a = 1$ it gives $X^2 + X + 1$ and $X^2 + X + 2$. If $a = 2$ it gives $X^2 + 2X + 3$ or $X^2 + 2X + 4$. If $a = 3$ it gives $X^2 + 3X + 3$ and $X^2 + 3X + 4$ and if $a = 4$ it gives $X^2 + 4X + 1$ and

X^2+4X+2. So it suffices to test the following list, and the reader can show that none of these is a factor of the reduced X^5-X-1.

$$X^2+2,\ X^2+3,\ X^2+X+1,\ X^2+X+2,\ X^2+2X+3,\ X^2+2X+4$$
$$X^2+3X+3,\ X^2+3X+4,\ X^2+4X+1,\ X^2+4X+2$$

We now compute the Galois group of X^5-X-1. The argument in section 9.6 showing that the group contains a 5-cycle is still valid here. But the argument that it has a 2-cycle fails. We will use Dedekind's theorem to find a 2-cycle. Then section 9.6 shows that the Galois group is the full S_5.

Let us ask Mathematica to factor X^5-X-1 over Z_7. Here is the session:

```
Factor[X^5 - X - 1, Modulus -> 7]
{X^2 + 6 X + 3} {X^3 + X^2 + 5X + 2}
```

We can apply Dedekind's theorem, since X^5-X-1 has non-zero formal derivative in Z_7 and thus cannot have multiple roots in an extension field. Dedekind's theorem allows us to find an element τ in the Galois group with disjoint 2-cycle and 3-cycle. An example would be $\tau=(12)(345)$. Notice that τ^3 then has a 2-cycle but no 3-cycle. Hence the Galois group has a 2-cycle and a 5-cycle, and so must be the full S_5.

Remark: Curiously, the similar polynomial X^5+X+1 is solvable by radicals. Indeed if ω is a primitive third root of unity, then $\omega^5=\omega^2$, so substituting in P gives $\omega^2+\omega+1=0$. It follows that X^2+X+1 divides P, and indeed

$$X^5+X+1=(X^2+X+1)(X^3-X^2+1)$$

Remark: If we replace X by $-X$ in X^5+X-1, we obtain $-(X^5+X+1)$ and it immediately follows that X^5+X-1 factors and is solvable by radicals. If we replace X by $-X$ in X^5-X+1, we obtain $-(X^5-X-1)$ and it immediately follows that X^5-X+1 has Galois group S_5 and is not solvable by radicals.

It turns out that all quintics can be reduced to X^5+aX+b by transformations which preserve the Galois group. If a or b is zero, the quintic can be solved by radicals. If $|a|=|b|=1$, we obtain one pair X^5+X+1 and X^5+X-1 of solvable quintics, and one pair X^5-X-1 and X^5-X+1 of unsolvable quintics.

16.3 A Tricky Point Concerning Straightedge and Compass Constructions

If $Q\subset L$ is an extension generated by constructible numbers, then the degree $[L:Q]$ is a power of 2. The converse of this theorem is false, and the purpose of this section is to explain.

Suppose first that $Q \subset L$ is a Galois extension of degree a power of 2. Then the Galois group of the extension has order a power of two, and thus is solvable by the Sylow theorems. So we can find a composition series with composition quotients Z_2. The fundamental theorem of Galois theory then gives a series of quadratic extensions $Q \subset K_1 \subset K_2 \subset \ldots \subset L$. So every element of L can be constructed by straightedge and compass.

However, suppose r is algebraic with minimal polynomial a power of 2. We can form the root field $Q(r)$, but this root extension need not be Galois, and we cannot employ Galois theory to obtain a tower of quadratic extensions. We now show by example that in this situation, r may not be constructable by straightedge and compass.

Consider the polynomial $X^4 + 8X + 12$ and let θ be a root. We claim that the polynomial is irreducible, so $[Q(\theta) : Q] = 4 = 2^2$. We also claim that the Galois group of the splitting field L is A_4. Assuming this is correct, suppose we can find a field $Q \subset K \subset Q(\theta) \subset L$. By the fundamental theorem of Galois theory, the field K would correspond to a subgroup of A_4 of order 6. But A_4 is the group of the 12 rotational symmetries of a tetrahedron, and this group contains no subgroup of order 6. QED.

To prove $X^4 + 8X + 12$ irreducible, we factor over Z_5 and Z_{17}.

```
Factor[X^4 + 8 X + 12, Modulus -> 5]
{X + 1} {X^3 + 4 X^2 + X + 2}

Factor[X^4 + 8 X + 12, Modulus -> 17]
{X^2 + 4 X + 7} {X^2 + 13 X + 9}
```

Over Z, $X^4 + 8X + 12$ cannot have a linear factor, since there is no linear factor over Z_{17}. It cannot have a quadratic factor, since there is no quadratic factor and no product of two linear factors, over Z_5. So $X^4 + 8X + 12$ is irreducible.

The discriminant of our polynomial is 576^2, so the Galois group is a subgroup of A_4. We want to prove that the Galois group is the full A_4, and for that it is enough to prove that the order of the group is divisible by both 3 and 4, since than it has at least 12 elements and thus must be all of A_4.

If L is the splitting field of P, we have $Q \subset Q(r) \subset L$. Then $|G| = [L : Q] = [L : Q(r)][Q(r) : Q]$. Since $[Q(r) : Q] = 4$, the order of G is divisible by 4.

Apply Dedekind's theorem. The derivative of our polynomial over Z_5 is not zero, so it has no multiple roots over an extension. By the above factorization, there is a τ in the Galois group with a one cycle and a three cycle. This element has order 3, so the order of the Galois group is divisible by 3.

16.4 Polynomials with Galois Group S_n

Theorem 59 *For each $n > 0$, there is a polynomial of degree n with integer coefficients whose Galois group is the full S_n.*

Proof: This result is trivial for $n = 1$ and $n = 2$.

Suppose p is prime and $n \geq 1$ is an integer. We first claim there is a monic polynomial over Z which is irreducible in $Z_p[X]$. Indeed, we proved that a field $GF(p^n)$ of order p^n exists. So $Z_p \subset GF(p^n)$ is a Galois extension. By an earlier result, there exists $\theta \in GF(p^n)$ generating this extension. The minimal polynomial of θ is monic in $Z_p[X]$ and irreducible there. Lift it to a monic polynomial over Z.

Find $P_1(X)$ a monic polynomial of degree n with integer coefficients such that P_1 is irreducible over Z_2. Find a monic polynomial P_2 of degree n which factors in $Z_3[X]$ into a linear polynomial and an irreducible polynomial of degree $n-1$. If n is odd, find a monic polynomial P_3 of degree n which factors over Z_5 as a product of an irreducible quadratic and an irreducible polynomial of degree $n-2$. If n is even and $n \neq 4$, find a similar polynomial which factors over Z_5 as an irreducible quadratic times a product of two irreducible polynomials of distinct odd degrees. If $n = 4$, let $P_3(X)$ be a product of an irreducible quadratic and two linear terms with distinct roots. Let $P(X) = -15P_1(X) + 10P_2(X) + 6P_3(X)$. Notice that this polynomial is monic of degree n. Since the polynomial reduced modulo 2 is irreducible, $P(X)$ is irreducible.

Notice that modulo 2, $P(X) = P_1(X)$. Also modulo 3, $P(X) = P_2(X)$. Finally modulo 5, $P(X) = P_3(X)$. Since $P_1(X)$ is irreducible over Z_2, all of its roots are distinct by theorem 55 in section 14.1. Since $P_2(X)$ factors over Z_3 into a linear factor and an irreducible factor, the linear factor cannot be one of the roots of the irreducible factor and the roots of the irreducible factor are all distinct by theorem 55. So all roots are distince. When n is odd, $P_3(X)$ factors over Z_5 as a product of an irreducible quadratic and an irreducible polynomial of degree $n-2$. Each irreducible factor has distinct roots by theorem 55, and the two polynomials cannot have a common root, else they would be minimal polynomials for this root, but one has even degree and one has odd degree.

If n is even and not four, $P_3(X)$ factors over Z_5 as an irreducible quadratic and two irreducible polynomials of distinct odd degrees. The same reasoning shows that all roots are distinct. If $n = 4$, $P_3(X)$ still factors into terms with distinct roots.

Hence we can apply Dedekinds theorem. It first implies that the Galois group has an n-cycle. Second, it implies that the group contains an $(n-1)$-cycle. When n is odd, it implies that the group has a 2-cycle and an (n - 2)-cycle. When n is even and not 4, it implies that the group has a 2-cycle and two odd cycles. When $n = 4$ it implies that the group has a 2-cycle.

By taking powers of elements, we conclude that the group always contains a 2-cycle, an n-cycle, and an (n - 1)-cycle. For instance, if it contains a 2-cycle σ multiplies by a disjoint cycle τ of odd order k, then $(\sigma\tau)^k = \sigma$.

We now prove that such a group must be all of S_n. Numbering the elements appropriately, we can assume that the $n-1$-cycle is $(1\ 2\ \ldots\ n-1)$. If the 2-cycle is $\sigma = (i\ j)$ with $1 \le i < j < n$, then find $\tau \in G$ mapping n to j; this is possible since G is transitive on roots. Then $\tau^{-1}\sigma\tau$ maps n to $\tau^{-1}(i)$ and $\tau^{-1}(i)$ to n and leaves everything else fixed. So this product is $(\tau^{-1}(i)\ n)$. In short, we may assume that G contains $(1\ 2\ 3\ \ldots\ n-1)$ and $(k\ n)$ for some $k < n$.

We now show that such a group must be all of S_n. Let $\tau = (1\ 2\ \ldots\ n-1)$ and let $\sigma = (k\ n)$. If $j < n$, we can find a power τ^s mapping j to k. Then $\tau^{-s}\sigma\tau^s$ maps j to n and maps n to j and leaves everything else fixed. So G contains $(j\ n)$ for all $j < n$.

Now notice that $(in)(jn)(in) = (ij)$. Hence all transpositions are in G, so $G = S_n$.

16.5 Every Finite G is a Galois Group

Theorem 60 *If G is a finite group, there is a Galois extension $Q \subset K \subset L$ such that the Galois group of $K \subset L$ is G*

Proof: Assume G has n elements. Define an action of G on itself by

$$g : g_1 \to gg_1$$

This is a group homomorphism from G to S_n, the permutation group on n symbols. It is one-to-one, because if $gg_1 = g_1$ for all g_1, then $g = e$. So we can identify G with its image in S_n. Notice that this subgroup has order n, while S_n has order $n!$.

Find a polynomial $P(X)$ whose splitting field $Q \subset L$ has Galois group S_n. By the fundamental theory of Galois theory, the subgroup G corresponds to a subfield $Q \subset M \subset L$. Indeed $M \subset L$ is a Galois extension with Galois group G. QED.

Remark: It is unknown whether every G is the Galois group of a Galois extension of Q. Hilbert proved that all S_n and all A_n are realizable. Shafarevich proved that all solvable groups are realizable. It is known that all sporadic groups except possibly the Mathieu group M_{23} are realizable.

16.6 Proof of Dedekind's Theorem

The proof we give is due to John Tate, from two web expositions of the proof.

Step 1: Let $K \subset L$ by the splitting field of $P(X)$ and denote the roots of P by $\{r_1, \ldots, r_n\}$. Define

$$D = \sum_{0 \le e_1, \ldots, e_n < n} Z r_1^{e_1} r_2^{e_2} \ldots r_n^{e_n}$$

Since the degree of P is n, powers of r_i of size n or larger can be written in terms of lower powers. Consequently D is a subring of L. It contains Z since $r_1^0 r_2^0 \ldots r_n^0 = 1$.

Notice that the $r_1^{e_1} \ldots r_n^{e_n}$ need not be linearly independent. But we easily deduce the general structure of D. It is a finitely generated abelian group. Moreover, it has no torsion elements since it is contained in a field of characteristic zero. So by the classification of finitely generated abelian groups, it equals $Z \times Z \times \ldots \times Z$. Said another way, it has a basis $d_1, \ldots, d_N$ of linearly independent elements and equals $Zd_1 \oplus Zd_2 \oplus \ldots \oplus Zd_N$.

Fix a prime p. Clearly pD is a proper ideal of D, since it contains linear combinations of the d_i whose coefficients are multiples of p. Let $\mathbf{m}$ be a maximal ideal containing pD. Then $E = \Big(D/\mathbf{m}\Big)$ is a finite field. This field has characteristic p, since p times any element of E is zero. So $\Big(Z/pZ\Big) \subset \Big(D/\mathbf{m}\Big)$.

Step 2: Let $\varphi : D \to D/\mathbf{m}$ be the natural map. We have $Z \subset D$; φ sends integers to their values modulo p, and thus maps $P(X)$ to $P_p(X)$. But P splits and all of its roots belong to D, so $\varphi(P) = P_p = \prod \Big(X - \varphi(r_i)\Big)$. We are assuming that P_p has distinct roots, so the $\varphi(r_i)$ are distinct.

Step 3: Let R be the set of roots of P in L, and let R_p be the set of roots of P_p in E. It follows from the above that φ maps R to R_p in a one-to-one and onto manner.

Step 3: Let G be the Galois group of $Q \subset L$. If $\sigma \in G$, , σ permutes the r_i. So it maps $D \to D$, inducing an automorphism of D. It follows that $\varphi \circ \sigma$ is a ring homomorphism from D to E. This map again maps R to R_p bijectively. Moreover, if $\sigma \neq \tau$, then $\varphi \circ \sigma$ and $\varphi \circ \tau$ are unequal because induce different maps from R to R_p. But we know that an element of the Galois group is completely determined by the resulting permutations of the roots.

Step 4: Let $G = \{\sigma_1, \ldots, \sigma_N\}$. It follows that $\{\varphi \circ \sigma_1, \varphi \circ \sigma_2, \ldots, \varphi \circ \sigma_n\}$ are distinct homomorphisms from D to E. Notice that one of these is the original φ since $e \in G$.

Step 5: We claim that the $\varphi \circ \sigma_i$ are the only homomorphisms from D to E sending the identity to the identity. This is the most difficult step in the proof, so we will postpone its proof until the bitter end.

Step 6: The Galois group of E is generated by the Frobenius automorphism $\pi(e) = e^p$. If $\varphi : D \to E$ is a homomorphism, then so is $\pi \circ \varphi$. Hence by the previous result there is an automorphism $\tau \in G$ such that $\pi \circ \varphi = \varphi \circ \tau$. But τ and π induce permutations of R and

R_p and φ sets up a bijection between these sets of roots. The equation $\pi \circ \varphi = \varphi \circ \tau$ then implies that τ and π have the same cycle structure as permutations.

Step 7: Notice that P_p is not irreducible; it factors as $P_1 P_2 \ldots \P_k$. Each of these has different roots, so the roots R_p separate into k pieces, each left invariant by π. On each of these pieces, π generates a transitive cycle. So the cycle structure of π corresponds to the factorization of P_p. This completes the argument.

Step 8: To finish, we need only prove step 5.

We have $D = Zd_1 \oplus Zd_2 \oplus \ldots \oplus Zd_N$. The d_i are linearly independent over Z, but then they are linearly independent over Q because a dependent relation $q_1 d_1 + \ldots + q_N d_N = 0$ could be written with common denominator $\frac{a_1}{b} d_1 + \ldots + \frac{a_N}{b} d_N = 0$, implying $a_1 d_1 + \ldots + a_N d_N = 0$.

Consider $Qd_1 + \ldots + Qd_N$. This is a ring containing Q and finite dimensional over Q, so it is a field. This field contains D, hence all powers of the roots of P. So it must be L. In particular, $[L : Q] = N$, so the Galois group G of $Q \subset L$ has N elements.

We want to prove that the N homomorphisms already introduced are all there are. Suppose not and suppose there is another, ψ_{N+1}.

Step 9:

Lemma 14 *The maps $\psi_1, \ldots, \psi_{N+1} : D \to E$ are linearly independent over E.*

Proof: We prove this by induction on the number of maps. It suffices to study the induction step. Suppose $\sum_{i=1}^{N+1} e_i \psi_i = 0$.

Since $\psi_1 \neq \psi_{N+1}$, they are not equal on some non-zero element k_0. Then $\sum e_i \psi_i(kk_0) = 0$ and so

$$\sum e_i \psi_i(k) \psi_i(k_0) = 0$$

$$\sum e_i \psi_i(k) \psi_{N+1}(k_0) = 0$$

Subtracting

$$\sum e_i \psi_i(k) \Big(\psi_i(k_0) - \psi_{N+1}(k_0) \Big) = 0$$

This is a dependence relation on the first N terms, so by induction all coefficients are zero. In particular, $e_1 \Big(\psi_1(k_0) - \psi_{N+1}(k_0) \Big) = 0$. By the choice of k_0 we conclude that $e_1 = 0$.

Repeat the argument, selecting k_0 such that $\psi_2(k_0) \neq \psi_{N+1}(k_0)$. We conclude that $e_2 = 0$. Continue. Eventually $e_i = 0$ for $i \leq N$. Consequently $e_{N+1} \psi_{N+1} = 0$. Since the ψ map the identity to itself, $e_{N+1} = 0$, proving the lemma.

Step 10: We have $N + 1$ linearly independent maps from D to E, where D has a basis $d_1, \ldots, d_N$ over Z, and these elements are actually linearly independent elements of L over Q. Consider the equations

$$\sum_{i=1}^{N+1} x_i \psi_i(d_j) = 0$$

These are N linear equations in unknowns $x_1, \ldots, x_{N+1}$ belonging to E. Consequently, they have a non-zero solution. This solution is a dependence relation among the ψ_i, which holds for each $d_i \in D$ and consequently is identically true in D. This contradiction proves step 5, and consequently the entire theorem. QED.