Graduate Algebra

S. Paul Smith
Department of Mathematics
University of Washington
Seattle, WA 98195, USA
smith@math.washington.edu

December 7, 2013

Contents

Chapter 1

Origins of Modern Algebra

Modern algebra was developed to solve equations.

In this chapter we discuss some of the questions that gave rise to modern algebra. I assume you are already familiar with some of the language of modern algebra: groups, rings, ideals, homomorphisms, fields, vector spaces, and so on. I have minimized the number of definitions in this chapter. You will find most of the basic definitions and properties of rings, modules, and homomorphisms in chapter **??**.

The phrase "modern algebra" is vague, but is commonly used to describe the material in van der Waerden's 1930 book *Moderne Algebra*. Van der Waerden (1903-1996) studied at Amsterdam as an undergraduate, and then spent the academic year 1923-24 at Göttingen where he had the good fortune of attending Noether's algebra course. He then spent the fall semester of 2004 in Hamburg where he attended Artin's lectures. He received his Ph.D. from Amsterdam in 1926.

Among the primary developers of the material in Van der Waerden's book were Noether, Dedekind, Weber, Hilbert, Lasker, Macaulay, Steinitz, Artin, Krull, and Wedderburn, (on rings, ideals, and modules), Schur, Frobenius, Burnside, Schreier, and Galois (on groups and their representations).

Van der Waerden's book is a marvel, as fresh today as when it was written. Although hundreds of books covering similar ground have been written since, none cast the original into shadow.

1.1 From $\mathbb{N}$ to $\mathbb{Z}$ to $\mathbb{Q}$ to $\overline{\mathbb{Q}}$, $\mathbb{R}$ and $\mathbb{C}$

I disagree with the following quotation:

> Die ganze Zahl schuf der liebe Gott, alles Übrige ist Menschenwerk.
>
> God created the integers, all else is the work of man.
>
> *Kronecker*

Even the integers are the work of man. No doubt the first mathematical achievement of man was to recognize when two non-empty sets had the same

cardinality. Then came the abstraction, picking a single label, one, two, three, et cetera, to name/describe sets having the appropriate cardinality. Thus arose the natural numbers $1, 2, 3, \ldots$.

Several primitive cultures have had no numbers beyond one, two, and three. Those cultures with more extended numbering systems have not always had a notion of zero.

The creation of the natural numbers was motivated by man's desire to understand and manipulate the world. Mathematics is a practical art.

Many equations can be solved within the integers. One can postulate simple arithmetic problems arising from everyday life that can be solved within the integers. A typical example might be *find an integer x such that $x + 27 = 30$*. At a slightly more sophisticated level, one can imagine simple division problems, such as *find x such that $3x = 60$*, that can also be solved within the positive integers. A mild modification, such as $3x = 67$, leads to the idea of division with remainder, and suggests how mankind was led to the rational numbers.

One can imagine the forces that prompted the notion of negative integers.

1.1.1 Linear equations and rational numbers

The construction of the rationals $\mathbb{Q}$ from the integers $\mathbb{Z}$ can be formalized in such a way that a similar process applied to any domain[1] produces its field of fractions (see section 1.9). The next result summarizes the utility of the rational numbers in terms of solving certain kinds of equations. Notice that the result holds true if any field is substituted for the rationals.

Theorem 1.1. *If a, b, c are rational numbers with $a \neq 0$, then there is a unique rational number x such that $ax + b = c$.*

1.1.2 Quadratic equations

After linear equations come quadratics.

One of the great historical events concerning quadratics is Euclid's famous proof that $\sqrt{2}$ is not rational.

Theorem 1.2. *There is no rational number whose square is two.*

Proof. Suppose to the contrary that x is a rational number such that $x^2 = 2$. Write $x = a/b$ where a and b are integers. By cancelling common factors, we may assume that a and b have no common factor. Now, $2b^2 = a^2$, so 2 divides a^2. Hence 2 divides a, and we may write $a = 2c$. Hence $2b^2 = 4c^2$, and $b^2 = 2c^2$. It follows that b^2, and hence b, is even. Thus a and b are both even. This contradicts the hypothesis that they have no common factor, so we conclude that 2 cannot be a square in $\mathbb{Q}$. □

Undoubtedly, Euclid was motivated by the problem of computing the length of the hypotenuese of the isoceles right triangle with sides of length one.

[1] By domain I mean a commutative ring with the property that $ab \neq 0$ if $a \neq 0$ and $b \neq 0$.

The key point in Euclid's proof is that every non-zero element of $\mathbb{Q}$ can be written as a/b with a and b having no common factor. That fact is a consequence of a still more elementary fact, which we summarize in the next theorem.

Theorem 1.3. *Every non-zero integer can be written in an essentially unique way as a product of primes,*

$$p_1^{i_1} \cdots p_n^{i_n}$$

where $p_1, \ldots, p_n$ are primes.

By a prime we mean an integer p whose only divisors are ± 1 and $\pm p$. Thus, the primes are $\{\pm 2, \pm 3, \pm 5, \cdots\}$. When we say "essentially unique" we mean that factorizations $6 = 2.3 = 3.2 = (-3).(-1).2 = 1.(-2).3.(-1)$ are to be viewed as the same; they differ only by order and the inclusion of the terms ± 1.

Two integers are relatively prime if the only numbers that divide both of them are ± 1.

This theme, the unique factorization of integers and their relatives, reappeared often in the early development of modern algebra, and it remains a staple of introductory algebra courses.

That the Greek's view of numbers and algebra was intimately connected to geometry is well documented. They had no problem accepting the existence of numbers of the form $\sqrt{d}$ with d rational because, by Pythagoras's theorem, if the lengths of two sides of a right-angle triangle are rational numbers the length of the third side is of the form $\sqrt{d}$ for some rational number d. Accepting such numbers on an (almost) equal footing with the rationals allowed the solution of a range of quadratic equations with rational coefficients.

In modern parlance, the Greeks were happy computing in fields such as $\mathbb{Q}(\sqrt{d})$ when d is a positive rational number.

1.1.3 The field $\mathbb{Q}(\sqrt{d})$

Let d be a rational number that is not the square of a rational number. We define

$$\mathbb{Q}(\sqrt{d}) := \{\alpha + \beta\sqrt{d} \mid \alpha, \beta \in \mathbb{Q}\}.$$

It is easy to see that this subset of $\mathbb{C}$ is closed under multiplication and addition, meaning that the product and sum of two numbers in $\mathbb{Q}(\sqrt{d})$ belong to $\mathbb{Q}(\sqrt{d})$. For that reason we call $\mathbb{Q}(\sqrt{d})$ a subring of $\mathbb{C}$. A more subtle point is that the inverse (in $\mathbb{C}$) of a non-zero element of $\mathbb{Q}(\sqrt{d})$ belongs to $\mathbb{Q}(\sqrt{d})$. This follows from the calculation

$$\begin{aligned}
\frac{1}{a+b\sqrt{d}} &= \frac{1}{a+b\sqrt{d}} \cdot \frac{a-b\sqrt{d}}{a-b\sqrt{d}} \\
&= \frac{a-b\sqrt{d}}{a^2-bd^2} \\
&= \left(\frac{a}{a^2-bd^2}\right) - \left(\frac{b}{a^2-bd^2}\right)\sqrt{d}.
\end{aligned}$$

Notice that the denominator is non-zero if a or b is non-zero Thus $\mathbb{Q}(\sqrt{d})$ is a *field*.[2]

1.1.4 Quadratic equations and $\sqrt{-1}$

The reason why the equation $x^2 = -1$ has no solution in $\mathbb{Q}$ is quite different than the reason why $x^2 = 2$ has no solution. One can imagine that the ancients were unconcerned by the fact that $x^2 = -1$ has no rational solution. It probably seemed a foolish waste of time to even consider that a problem.

It is less apparent that an equation such as $x^2 + 2x + 2 = 0$ has no rational solution, and the discovery of this fact must surely have been intimately related to the discovery of the general solution to a quadratic equation. Several ancient cultures independently discovered the result that

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{1-1}$$

gives the two solutions to the quadratic equation $ax^2 + bx + c = 0$. It follows from the formula that the equation has no real-number solution if $b^2 - 4ac < 0$.

This, after many centuries, led to the invention/discovery of $\sqrt{-1}$ and eventually to the notion of complex numbers. This in turn leads to the following question: *if $f(x)$ a polynomial with coefficients in a field k, is there a field K containing k in which f has a zero?* We take up this question in section 1.4.

Having discovered the formula (1-1) for the roots of a quadratic polynomial attention turned to the question of whether there are analogous formulas for the solutions to higher degree polynomials. This was question was the primary force behind the development of algebra from 1300-1800. Eventually, Galois (1811-1832) gave a comprehensive solution to this problem, though no one understood his work at the time.[3] The first comprehensive account of his work was published in the OctoberNovember 1846 issue of the *Journal des mathmatiques pures et appliques*. Apparently, Galois was the first person to use the word *groupe* for a collection of permutations closed under composition.

Once the ancients had realized that one could pass beyond the rationals $\mathbb{Q}$ to include roots of rational numbers and more complicated expressions built from such roots, it was natural to ask if this gave "all" numbers. This question is crystallized by asking whether π is the zero of a polynomial with rational coefficients. More generally, this leads the distinction between algebraic and transcendental elements over an arbitrary field.

---

[2] A field is a commutative ring in which every non-zero element has an inverse.

[3] The night before the duel that lead to his death, Galois stayed up all night composing what would become his mathematical testament, the famous letter to Auguste Chevalier outlining his ideas. Hermann Weyl, one of the greatest mathematicians of the 20th century, said of this testament, "This letter, if judged by the novelty and profundity of ideas it contains, is perhaps the most substantial piece of writing in the whole literature of mankind."

1.2 Divisibility and Factorization

After learning to count and add, children learn to multiply and divide. Questions about division and factorization are of primary importance in all rings. A great part of the impetus for the development of modern abstract algebra arose from problems of division and factorization, especially in the rings most closely related to the integers, the rings of integers in number fields.

The simplest number field is $\mathbb{Q}$ itself, and *"number field"* is just another name for a ring that contains $\mathbb{Q}$ and is a finite dimensional vector space over it. For example, the fields $\mathbb{Q}(\sqrt{d})$ are number fields.

Let $d \in \mathbb{Z}$ and suppose that d is not a square. The ring of integers in $\mathbb{Q}(\sqrt{d})$ consists of those elements ξ in $\mathbb{Q}(\sqrt{d})$ such that ξ is a zero of a polynomial of the form

$$x^2 + bx + c$$

where $b, c \in \mathbb{Q}$. Every ordinary integer $n \in \mathbb{Z}$ is an integer in $\mathbb{Q}(\sqrt{d})$ because it is a zero of the polynomial $x^2 - n^2$. It is also clear that $\sqrt{d}$ is an integer in $\mathbb{Q}(\sqrt{d})$ because it is a zero of $x^2 - d$. In fact, if $m, n \in \mathbb{Z}$, then $m + n\sqrt{d}$ is an integer in $\mathbb{Q}(\sqrt{d})$ because

$$\big(m + n\sqrt{d}\big)^2 - 2m\big(m + n\sqrt{d}\big) + m^2 - n^2 = 0.$$

Hence $\mathbb{Z}[\sqrt{d}]$ consists of integers $\mathbb{Q}(\sqrt{d})$. If $d \equiv 2, 3 (\text{mod } 4)$, then $\mathbb{Z}[\sqrt{d}]$ is the entire ring of integers $\mathbb{Q}(\sqrt{d})$. However, if $d \equiv 1 (\text{mod } 4)$ the ring of integers $\mathbb{Q}(\sqrt{d})$ is $\mathbb{Z}\big[\frac{1+\sqrt{d}}{2}\big]$. We leave the reader to verify these claims.

Quite possibly, the expectation was that the prime factorization theorem in the integers, Theorem 1.3, would extend to the rings of integers in number fields. That hope proved over-optimistic,[4] but its failure led Kummer (1810-1893) to the invention of *ideals* in 1843. Ideals are certain *subsets* of a ring. The word ideal is short for *idealized number*. There is a notion of product of ideals and the vindication for this more subtle notion is the breathtakingly beautiful result that every non-zero ideal in a ring of integers in a number field is a product of prime ideals in a unique way. For the usual ring of integers this result reduces to Theorem 1.3.

1.2.1 Quadratic extensions of $\mathbb{Z}$

Let d be an integer that is not a square. We define

$$\mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}.$$

This is a subset of $\mathbb{C}$ and is closed under multiplication, addition, and subtraction, meaning that the product, sum, and difference, of two elements in $\mathbb{Z}[\sqrt{d}]$ belongs to $\mathbb{Z}[\sqrt{d}]$. Hence $\mathbb{Z}[\sqrt{d}]$ is a ring.

[4] Gabriel Lamé (1795-1870) thought he had proved Fermat's Last Theorem but he had overlooked the fact that unique factorization failed in some of the rings he was working with. This was perhaps understandable because he was primarily an applied mathematician. Nevertheless, he did verify Fermat's Last Theorem for the case $n = 7$.

The notion of division makes sense in any ring: if $b = ac$ we say that b is a multiple of a, and a divides b, and write $a|b$. Strictly speaking we should be explicit about the ring because b can be a multiple of a in one ring but not in another. Hence if a and b are elements of a ring R, we say that a divides b in R if $b = ar$ for some $r \in R$.

Every element divides zero. Zero divides no elements other than itself.

At the other end of the spectrum, 1 divides every element. But 1 is not the only element with this property.

An element u in a ring R is a unit in R if there is an element $v \in R$ such that $uv = vu = 1$. We call v the inverse of u and denote it by u^{-1}. For example, $2 + \sqrt{3}$ is a unit in $\mathbb{Z}[\sqrt{3}]$ because $(2 + \sqrt{3})(2 - \sqrt{3}) = 1$. The inverse of an element is unique because if v and b are inverses, then $b = b(uv) = (bu)v = v$.

Exercise.

1. Show that if $uv = wu = 1$, then $v = w$.

2. Let V be an infinite dimensional vector space. Give an example of a linear map $u : V \to V$ such that there is an element $v : V \to V$ such that $uv = 1$, but $vu \neq 1$. Here 1 denotes the identity map.

3. Show that u divides every element of R if and only if it is a unit.

Let d be a non-square integer. Let $x = a + b\sqrt{d}$ be an element of $\mathbb{Z}[\sqrt{d}]$. The norm of x is

$$N(x) = a^2 - b^2 d.$$

Since d is not a square, $N(x) = 0 \Leftrightarrow x = 0$. The other important property of the norm is that $N(xy) = N(x)N(y)$.

Because the norm is an integer, a factorization $a = xy$ in $\mathbb{Z}[\sqrt{d}]$ implies the the factorization $N(a) = N(x)N(y)$ in $\mathbb{Z}$. This gives us a tool for studying factorization questions in $\mathbb{Z}[\sqrt{d}]$.

If d is a negative integer the norm of an element x in $\mathbb{Z}[\sqrt{-d}]$ is equal to $x\bar{x} = |x|^2$, where $\bar{x}$ is its complex conjugate.

Lemma 2.1. *Let d be a negative integer.*

1. *The element $x = a + b\sqrt{d}$ is a unit in $\mathbb{Z}[\sqrt{d}]$ if and only if $N(x) = 1$.*

2. *The units in $\mathbb{Z}[i]$ are $\{\pm 1, \pm i\}$.*

3. *If $d \neq -1$, the units in $\mathbb{Z}[\sqrt{d}]$ are $\{\pm 1\}$.*

Proof. Since $d < 0$, $N(x) \geq 0$. Certainly, if x is a unit, then $1 = N(1) = N(xx^{-1}) = N(x)N(x^{-1})$, so we conclude that $N(x) = 1$. Conversely, suppose that $N(x) = 1$. Then $x \neq 0$, and it has an inverse in $\mathbb{C}$, namely

$$x^{-1} = \frac{1}{a + b\sqrt{d}} \cdot \frac{a - b\sqrt{d}}{a - b\sqrt{d}} = \frac{a - b\sqrt{d}}{a^2 - b^2 d} = a - b\sqrt{d}.$$

This belongs to $\mathbb{Z}[\sqrt{d}]$ so x is a unit in $\mathbb{Z}[\sqrt{d}]$.

The only way $a^2 - b^2 d$ can equal 1 is if $a^2 = 1$ and $b = 0$, leading to the units ± 1, or if $a = 0$, $d = -1$ and $b^2 = 1$, leading to the units $\pm i$ in $\mathbb{Z}[i]$. □

1.2.2 Irreducible and prime elements

Definition 2.2. *Let R be a commutative ring. A non-zero non-unit $a \in R$ is* irreducible *if in every factorization $a = bc$ either b or c is a unit. A non-zero non-unit $p \in R$ is called a* prime *if in every division $p|bc$ either $p|b$ or $p|c$.* ◊

A prime is irreducible: if $p = bc$ then, perhaps after relabelling the factors, $p|b$, so $b = pu$ and $p = puc$, so $1 = uc$, whence c is a unit.

The converse is not always true though: an irreducible need not be prime.

In particular, as we will now show, $\mathbb{Z}[\sqrt{-5}]$ contains irreducible elements that are not prime. We will show that

1. 2, 3, $1 + \sqrt{-5}$, and $1 + \sqrt{-5}$, are irreducible;
2. that none of them is a unit multiple of another;
3. that none of them is prime.

Since

$$(1 + \sqrt{-5})(1 - \sqrt{-5}) = 6 = 2.3$$

the number 6 is not a product of primes, and is not a product of irreducible elements in a unique way.

The key tool is the norm, $N : \mathbb{Z}[\sqrt{-5}] \to \mathbb{Z}$ defined by

$$N(a + b\sqrt{-5}) = a^2 + 5b^2.$$

Thus, if $\bar{z}$ is the complex conjugate of a number $z \in \mathbb{Z}[\sqrt{-5}]$, then $N(z) = z\bar{z}$. Thus $N(xy) = N(x)N(y)$.

If u is a unit in $\mathbb{Z}[\sqrt{-5}]$, then $N(u)N(u^{-1}) = N(1) = 1$ so $N(u)$ is ± 1. However, if a and b are integers such that $a^2 + 5b^2 = 1$, then $b = 0$ and $a = \pm 1$. Therefore ± 1 are the only units in $\mathbb{Z}[\sqrt{-5}]$. This proves the claim (2).

If if a and b are integers, then $a^2 + 5b^2$ cannot equal 2 or 3. Suppose that $x, y \in \mathbb{Z}[\sqrt{-5}]$ are such that $xy = 1 + \sqrt{-5}$. Then $N(x)N(y) = N(xy) = N(1 + \sqrt{-5}) = 6$, so either $N(x)$ or $N(y)$ is 1. Thus either x or y is a unit. Therefore $1 + \sqrt{-5}$ is an irreducible element of $\mathbb{Z}[\sqrt{-5}]$. Similar arguments show that 2, 3, and $1 + \sqrt{-5}$, are irreducible.

This leads to the question of identifying those domains in which every irreducible element is prime. The answer appears in Lemma 10.2.

Notice that 2 is not a prime in $\mathbb{Z}[i]$ because $2 = (1 + i)(1 - i)$. However, $i + i$ and $1 - i$ are both irreducible because, for example, if $1 + i = xy$, then $N(x)N(y) = N(1+i) = 2$ so the norm of either x or y is equal to ± 1, and hence eiher x or y is a unit.

Exercise. Is 2 prime in $\mathbb{Z}[i]$? Describe exactly which prime integers remain prime in $\mathbb{Z}[i]$.

1.2.3 Greatest common divisors

Let R be a domain. A greatest common divisor of two elements $a, b \in R$ is an element $d \in R$ such that

1. $d|a$ and $d|b$, and
2. if $e|a$ and $e|b$, then $e|d$.

We write $d = \gcd(a, b)$, or just $d = (a, b)$. We say that greatest common divisors exist in R if every pair of elements in R has a greatest common divisor in R.

The greatest common divisor is not unique. For example, in the ring of integers, both 2 and -2 are greatest common divisors of 6 and 10. Similarly, in $\mathbb{Z}[i]$ both 2 and $2i$ are greatest common divisors of 4 and 6. In general, if d and d' are two greatest common divisors of a and b, then each is a unit multiple of the other: because each divides the other, we have $d' = du$ and $d = d'v$, so $d(uv - 1) = 0$, whence $uv = 1$.

To obtain uniqueness of a greatest common divisor we need some additional structure on R. For example, in $\mathbb{Z}$ if we also insist that the greatest common divisor be positive, then it becomes unique.

Actually, we haven't even shown that greatest common divisors exist in $\mathbb{Z}$ or $\mathbb{Z}[\sqrt{d}]$. There is something to do here.

We can define the greatest common divisor of any collection of elements by saying that d is a greatest common divisor of $a_1, \ldots, a_n$ if it divides each a_i, and if e is any element of R dividing all of them, then e necessarily divides d.

1.3 Fields

I assume you are familiar with fields such as the real numbers, $\mathbb{R}$, the rational numbers, $\mathbb{Q}$, and the complex numbers $\mathbb{C}$. A field k is a non-empty set of elements that can be added and multiplied with the usual rules holding. The crucial feature of a field is that every non-zero element of it has an inverse; that is, if α is a non-zero element of k, there is an element α^{-1} in k such that $\alpha^{-1}\alpha = \alpha\alpha^{-1} = 1$.

Exercise. Look up the definition of a field in a textbook. Ponder these points:

1. the examples came before the definition;
2. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, have some common properties;
3. abstracting from these properties leads to a definition which captures the salient features of those examples.

It would be foolish to develop a theory of fields if these were the only examples. Fields abound. Finite fields, the simplest examples of which appear in the next exercise, play a central role in number theory, and in applications

of algebra to communications, coding theory, and other computer-related areas. Number fields, the simplest examples of which are the quadratic extensions $\mathbb{Q}(\sqrt{d})$ of the rationals, occupy a central place in number theory and arithmetic questions. Function fields, being fields consisting of ratios of functions defined on geometric objects, are central to algebraic geometry, complex analysis, and other areas.

Exercise. Let p be a positive prime number. There is a unique[5] field with exactly p elements. The elements of $\mathbb{F}_p$ are subsets of the integers; they are denoted $[0], [1], \ldots, [p-1]$ and defined as follows:

$$[i] := \{a \in \mathbb{Z} \mid p \text{ divides } a - i\}. \tag{3-2}$$

Actually, we use (3-2) to define $[i]$ for every $i \in \mathbb{Z}$, but because $[i] = [i + np]$, we only obtain p different $[i]$'s. Define

$$[i] + [j] := [i + j] \text{ and} \tag{3-3}$$
$$[i] \cdot [j] := [ij]. \tag{3-4}$$

Show the definitions of $+$ and $\cdot$ in $\mathbb{F}_p$ are unambiguous: i.e., if $[i] = [i']$ and $[j] = [j']$ show that $[i] + [j] = [i'] + [j']$ and $[i][j] = [i'][j']$.

To show that $\mathbb{F}_p$ is a field one must show that every non-zero element in $\mathbb{F}_p$ has an inverse: if the integer i is not divisible by p, show there is an integer j such that $ij - 1$ is divisible by p, and hence that $[i][j] = 1 = [1]$. We write $[j] = [i]^{-1}$.

Exercise. Let n be a positive integer and $\zeta = e^{2\pi i/n}$. Show that $\mathbb{Q}(\zeta) := \mathbb{Q} \oplus \mathbb{Q}\zeta \oplus \cdots \oplus \mathbb{Q}\zeta^{n-1}$ is a subfield of $\mathbb{C}$.

Exercise. Think of six interesting questions about the fields $\mathbb{F}_p$, $\mathbb{Q}(\sqrt{d})$, and $\mathbb{Q}(\zeta)$.

Exercise. The field of rational functions in one variable, denoted $k(x)$, consists of all ratios p/q where p and q are polynomials in x having coefficients in k, and $q \neq 0$. We add and multiply these in the obvious way. The inverse of a non-zero element p/q is q/p. This is the field of rational functions on the affine line over k. Likewise, the field $k(x, y)$ of rational functions on the affine plane over k consists of all ratios p/q where p and q are polynomials in the variables x and y, and $q \neq 0$. Are the fields $k(x)$ and $k(x, y)$ isomorphic? What does the word "isomorphic" mean in this context?

Later, we will examine fields in some detail, but for now we treat them as a necessary preliminary for our discussion of polynomials. Fields provide the coefficients for polynomials.

The letter k is often used to denote a field because German mathematicians, who were the first to examine fields in some detail, called a field *ein körper* (körper=body, cf. "corpse"). Despite this nomenclature, the study of fields remains a lively topic.

[5] When we say unique we really mean that all fields with p elements are isomorphic.

Notes on notation. The same symbol is often used for different things in mathematics. If the author is doing a good job, the context will provide enough information to interpret the symbol unambiguously. For example, in (3-3), the + on the left-hand of the = sign is different from the + on the right-hand side. The + on the right-hand side is the usual addition in $\mathbb{Z}$, but + on the left-hand is the new addition in $\mathbb{F}_p$. We are using the old addition in $\mathbb{Z}$ to define the new addition in $\mathbb{F}_p$.

We will use the symbol 0 to denote the zero element in all the rings we meet, so you need to be alert as to which zero is being meant. Likewise, the symbol 1 is used to denote the unit element in a ring. So, rather than writing [1] or [0] for the unit and zero in $\mathbb{F}_p$, we simply write 1 or 0.

If i is an integer, we might write $\bar{i}$, for the element $[i]$ of $\mathbb{F}_p$. If we could all agree to be careful, we could even write i for $[i]$. Think of the time and effort we would save by doing this; the price is eternal vigilance...with apologies to Thomas Jefferson "The price of liberty is eternal vigilance".

Exercise. Suppose that R is a ring containing a field k as a subring. Show that the addition and multiplication on R give it the structure of a vector space over k.

Exercise. Sometimes we write (a, b) for the greatest common divisor of two integers a and b. This notation is also used to denote the ideal generated by a and b. Show there is an equality of ideals, $(a, b) = (d)$, if d is a greatest common divisor of a and b.

1.4 The polynomial ring in one variable

Throughout this section k denotes a field.

In this section we show that the ring of polynomials in one variable with coefficients in k behaves rather like the ring of integers. Our initial focus is on questions of division and factorization. There is a well-behaved notion of division with remainder, and even a version of the Euclidean algorithm. There are polynomials that behave like prime numbers—the so-called *irreducible* polynomials—and a version of Theorem 1.3 saying that every polynomial is a product of irreducible polynomials in an essentially unique way.

Let R be a commutative ring. To begin with you might think of R being one of the rings in the previous sections, perhaps the integers, or the rationals, or the reals, or one of the more exotic examples like $\mathbb{F}_p$ or $\mathbb{Z}[\sqrt{d}]$ or $\mathbb{Q}(\sqrt{d})$.

Polynomials in one variable, say x, with coefficients in R can be added and multiplied in the obvious way to produce another polynomial with coefficients in R.

We write $R[x]$ for the set of all polynomials in x with coefficients in R. An element of $R[x]$ is an expression

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where the coefficients a_i belong to R. Two polynomials are considered to be the same only if all their coefficients are the same.

Addition and multiplication are defined in the obvious way. In this way $R[x]$ becomes a ring, with zero element the zero polynomial 0, and identity element the constant polynomial 1.

Definition 4.1. *Let R be a ring. The* polynomial ring with coefficients in R, *which we denote by $R[x]$, consists of all formal expressions*

$$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_n x^n$$

where $\alpha_0, \ldots, \alpha_n \in R$, and this is made into a ring by defining the sum and product of two polynomials by

$$\sum \alpha_i x^i + \sum \beta_i x^i := \sum (\alpha_i + \beta_i) x^i$$

and

$$\left(\sum \alpha_i x^i\right)\left(\sum \beta_i x^i\right) := \sum_n \left(\sum_{j=0}^{n} (\alpha_j \beta_{n-j}\right) x^n.$$

We call $\alpha_0, \ldots, \alpha_n$ the coefficients *of $\sum_{i=0}^{n} \alpha_i x^i$. We say that two polynomials are* equal *if and only if they have the same coefficients.*

We call x an indeterminate. ◊

We leave it to the reader to check that $R[x]$ is a ring.

We are particularly interested in the case when R is a field.

1.4.1 Division with remainder

Recall that if a and b are integers with b non-zero, then there are integers q and r such that $a = bq + r$ and $0 \leq r < |b|$. We usually call r the remainder. This result plays a key role in arithmetic. To show that there is an analogous result for $k[x]$ we need a notion of "size" to replace absolute value.

The degree of a non-zero element $f = a_n x^n + \cdots + a_1 x + a_0$ in $R[x]$ is n provided that $a_n \neq 0$. In that case we call a_n the leading coefficient of f. If $f = 0$ it is convenient to define its degree to be $-\infty$. It is a trivial observation that the units in $k[x]$ are precisely the polynomials of degree zero.

Lemma 4.2. *Let R be a domain and let $f, g \in R[x]$. Then*

1. $\deg(f + g) \leq \max\{\deg f, \deg g\}$;
2. $\deg(fg) = \deg f + \deg g$;
3. *$R[x]$ is a domain.*

Proposition 4.3. *If f and g are non-zero elements of $k[x]$ such that f is non-zero, then there are unique polynomials q and r such that*

$$g = fq + r \qquad \text{and} \qquad \deg r < \deg f.$$

Proof. Existence. We argue by induction on $\deg g$. If $g = 0$, we can take $q = r = 0$. If $\deg g < \deg f$, we can take $q = 0$ and $r = g$. If $m = \deg g \geq \deg f = n$, we can write

$$g = \alpha x^m + \cdots \text{lower degree terms}$$
$$f = \beta x^n + \cdots \text{lower degree terms}.$$

Since

$$\deg\big(g - \alpha\beta^{-1}x^{m-n}f\big) < \deg g,$$

we may apply the induction hypothesis to $g - \alpha\beta^{-1}x^{m-n}f$.

Uniqueness. If $g = fq+r = fq'+r'$, then $f(q-q') = r'-r$. But $\deg(r'-r) < \deg f$, so this implies that $r' - r = 0$. Hence $q' = q$ also. □

Proposition 4.4. *Every pair of non-zero elements in $k[x]$ has a greatest common divisor.*

Proof. To prove this, we need to introduce the Euclidean algorithm. The Euclidean algorithm is a constructive method that produces the greatest common divisor of two polynomials, as we now show. □

1.4.2 The Euclidean algorithm

Let f and g be elements of $k[x]$ with f non-zero. By repeatedly using Proposition 4.3 we may write

$$\begin{aligned} g &= fq_1 + r_1 && \text{with} && \deg r_1 < \deg f, \\ f &= r_1q_2 + r_2 && \text{with} && \deg r_2 < \deg r_1, \\ r_1 &= r_2q_3 + r_3 && \text{with} && \deg r_3 < \deg r_2, \\ \cdots &\quad \cdots . \end{aligned}$$

Since the degrees of the remainders r_i are strictly decreasing, this process must stop. Stopping means that the remainder must eventually be zero. If $r_{t+2} = 0$, and we set $r_{-1} = g$ and $r_0 = f$, then the general equation becomes

$$r_i = r_{i+1}q_{i+2} + r_{i+2} \qquad \text{with} \qquad \deg r_{i+2} < \deg r_{i+1}, \tag{4-5}$$

and the last equation becomes

$$r_t = r_{t+1}q_{t+2}.$$

Claim: $r_{t+1} = \gcd(f, g)$. Proof: Since r_{t+1} divides r_t, it follows from (4-5) that r_{t+1} also divides r_{t-1}. By descending induction, (4-5) implies that r_{t+1} divides all r_i, $i \geq -1$. In particular, r_{t+1} divides f and g. On the other hand, if e divides both f and g, then it divides r_1. If e divides r_i and r_{i+1}, then it follows from (4-5) that it also divides r_{i+2}. By induction, e divides r_{t+1}. Hence r_{t+1} is a greatest common divisor of f and g. ◊

This proceedure for finding the greatest common divisor of f and g is called the Euclidean algorithm. It completes the proof of Proposition 4.4.

If K is a field containing k, then $K[x]$ contains $k[x]$. Hence, if f and g belong to $k[x]$, we can ask for their greatest common divisor in $k[x]$, and for their greatest common divisor in $K[x]$. These are the same. This is because the uniqueness of q and r in Proposition 4.3 ensures that carrying out the Euclidean algorithm in $k[x]$ for a pair $f, g \in k[x]$ produces exactly the same result as carrying out the Euclidean algorithm in $K[x]$ for that pair.

Proposition 4.5. *Let d be a greatest common divisor in $k[x]$ of non-zero elements f and g. Then $d = af + bg$ for some a and b.*

Proof. Since a greatest common divisor is unique up to a scalar multiple, we can assume that $d = r_{t+1}$, the last remainder produced by Euclidean algorithm. Working backwards, we have

$$r_{t+1} = r_{t-1} - r_t q_{t+1} = r_{t-1} - (r_{t-2} - r_{t-1}q_t)q_{t+1} = \cdots ,$$

and so on. Eventually we obtain an expression in which every term is a multiple of either $r_0 = f$ or $r_{-1} = g$. Hence the result. □

Let $f \in k[x]$. We write (f) for the set of all multiples of f. That is,

$$(f) = \{fg \mid g \in k[x]\}.$$

It is clear that (f) contains zero. The sum and difference of two multiples of f are multiples of f. Any multiple of a multiple of f is a multiple of f. Hence (f) is an ideal of $k[x]$. We call it the principal ideal generated by f.

Theorem 4.6. *Every ideal in $k[x]$ is principal.*

Proof. The zero ideal consists of all multiples of zero, so is principal. If I is a non-zero ideal, choose a non-zero element f in it of minimal degree. Clearly $(f) \subset I$. If g is an element of I, we may write $g = fq + r$ with $\deg r < \deg f$. However, r equals $g - fq$, so belongs to I; because the degree of f was minimal, we conclude that $r = 0$. Hence $g \in (f)$. Thus $I = (f)$. □

Notice that (f) is generated by λf if λ is a non-zero element of k. Conversely, if $(f) = (g)$, then g and f must be multiples of each other, so $g = \lambda f$ for some non-zero λ in k. Hence, if I is a non-zero ideal in $k[x]$, there is a *unique* monic polynomial f such that $I = (f)$.

The next result is one way to recognize some irreducible polynomials.

Proposition 4.7 (Eisenstein's criterion)**.** *Let $f = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$. Suppose there is a prime p such that*

1. *p does not divide a_n,*
2. *p divides all the other coefficients,*

3. p^2 does not divide a_0.

Then f is irreducible in $\mathbb{Q}[x]$.

Proof. Suppose to the contrary that f is not irreducible in $\mathbb{Q}[x]$. By Lemma 10.5(1), $f = gh$ for some polynomials $g, h \in \mathbb{Z}[x]$ of *positive* degree.

Passing to $\mathbb{Z}_p = \mathbb{Z}/(p)$ and $\mathbb{Z}_p[x]$, this implies that $\bar{f} = \bar{g}\bar{h}$ in $\mathbb{Z}_p[x]$, where $\bar{f}$ denotes the image of f in $\mathbb{Z}_p[x]$ (i.e., the polynomial obtained by reducing all the coefficients of f modulo p). Thus $\bar{g}\bar{h} = \bar{a}_n x^n \neq 0$. Hence all the coefficients of g and h, except their leading ones, are divisible by p. In particular, their constant terms, say b_0 and c_0 are divisible by p. Hence p^2 divides $b_0 c_0 = a_0$, a contradiction. □

Remark. The following generalization is proved in roughly the same way: if $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial with coefficients in a commutative domain R and there is a prime ideal $\mathfrak{p}$ containing all the a_i but $a_0 \notin \mathfrak{p}^2$, then f is irreducible $R[x]$ (and in $k[x]$ if R is a UFD, where $k = \operatorname{Fract} R$).

One of the most important applications of Eisenstein's criterion is to prove the irreducibility of the cyclotomic polynomials

$$x^{p-1} + \cdots + x + 1,$$

where p is a prime. Notice that the zeroes of this are the p^{th} roots of unity $e^{2n\pi i/p}$, $1 \leq n \leq p-1$.

Corollary 4.8. *Let p be a prime. The polynomial $x^{p-1}+\cdots+x+1$ is irreducible.*

Proof. Write $f(x) = x^{p-1} + \cdots + x + 1$. If we substitute $y = x - 1$ into the equality $(x-1)f(x) = x^p - 1$, we get

$$yf(y+1) = (y+1)^p - 1 = y^p + \binom{p}{p-1} y^{p-1} + \cdots + \binom{p}{1} y.$$

If $1 \leq i \leq p-1$, then p divides $\binom{p}{i}$, so factoring out y shows that $f(y+1)$ satisfies Eisenstein's criterion, and is therefore irreducible. Hence $f(x)$ is irreducible. □

1.4.3 Quotient rings of $k[x]$

Quotient rings of $k[x]$ present a psychological obstruction for the beginner because the elements of these quotient rings are *subsets* of $k[x]$. It seems a leap is required to this of a set, especially an infinite set, as a single element.

Let's warm up to this by looking again at the finite fields $\mathbb{F}_p$ defined earlier. These are, in fact, quotient rings of $\mathbb{Z}$, although they weren't presented in that way. The first example is the field with two elements, $\mathbb{F}_2$. The ring $\mathbb{Z}$ is the disjoint union of two subsets, the even integers and the odd integers. As you know, a product of an even and an odd number is even, the sum of two odd numbers is even, and so on. We can display this in an addition and multiplication table:

+	even	odd
even	even	odd
odd	odd	even

×	even	odd
even	even	even
odd	even	odd

This is exactly the same as the addition and multiplication tables for $\mathbb{F}_2$. Only the labelling of the elements is different:

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

There is nothing special about the prime number two, except for the fact that the English language has the words *even* and *odd*. For example, consider the prime 3. We can write $\mathbb{Z}$ as the disjoint union of three sets, one consisting of the multiples of three, one consisting of the numbers that leave a remainder of 1 when divided by three, and the last consisting of the numbers that leave a remainder of 2 when divided by three. If we write $3\mathbb{Z}$ for the the set of multiples of 3, the other two sets can be written as $1 + 3\mathbb{Z}$ and $2 + 3\mathbb{Z}$ where

$$a + 3\mathbb{Z} = \{a + b \mid b \in 3\mathbb{Z}\} = \{a + 3n \mid n \in \mathbb{Z}\}.$$

We call these *cosets* of $3\mathbb{Z}$ in $\mathbb{Z}$. One now has analogues of the fact that even+odd = odd, odd×odd=odd, and so on. These can be gathered into the multiplication table for the field $\mathbb{F}_3$ as follows (we write $[a]$ for $a + 3\mathbb{Z}$):

+	[0]	[1]	[2]
[0]	[0]	[1]	[2]
[1]	[1]	[2]	[0]
[1]	[2]	[0]	[1]

×	[0]	[1]	[2]
[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]
[2]	[0]	[2]	[1]

The ring $\mathbb{F}_p$ is sometimes denoted by $\mathbb{Z}/p\mathbb{Z}$.

The basic idea can be extended to any ring. In particular, if I is an ideal in $k[x]$ there is a ring $k[x]/I$ whose elements are the cosets

$$a + I = \{a + f \mid f \in I\}$$

and the addition and multiplication are defined by

$$(a + I) + (b + I) = (a + b) + I, \qquad (a + I) \times (b + I) = ab + I.$$

Lemma 4.9. *If f is a polynomial of degree $n \geq 0$, then $\dim_k k[x]/(f) = n$, and the images of $1, x, \ldots, x^{n-1}$ are a basis for $k[x]/(f)$.*

Proof. The natural homomorphism $\pi : k[x] \to k[x]/(f)$ sends k to an isomorphic copy of itself in $k[x]/(f)$, so we think of k as a subring of $k[x]/(f)$. Multiplication in $k[x]/(f)$ therefore gives $k[x]/(f)$ the structure of a k-vector space. Since the powers of x are a basis for $k[x]$, their images span $k[x]/(f)$.

If g is any element of $k[x]$, then $g = af + r$ for some $a \in k[x]$ and some r of degree $< n$. Since $\pi(g) = \pi(r)$ and since r is a linear combination of $1, x, \ldots, x^{n-1}$, $\{\pi(x^i) \mid 0 \leq i \leq n-1\}$ spans $k[x]/(f)$. These elements are linearly independent too because the only linear combination of $1, x, \ldots, x^{n-1}$ that belongs to (f) is $0.1 + 0.x + \cdots + 0.x^{n-1}$. □

An ideal in a ring R that is not equal to R and is contained in no ideals other than itself and R is called a maximal ideal.

One sees easily that $(x - \lambda)$ is a maximal ideal of $k[x]$, but if k is not algebraically closed, there will be other maximal ideals. For example, $(x^2 + 1)$ is a maximal ideal in $\mathbb{R}[x]$.

Lemma 4.10. *An ideal I in a ring R is maximal if and only if R/I is a field.*

Proof. Suppose that I is maximal. A non-zero element of R/I can be written as $[a + I]$ for some $a \notin I$. Since I is maximal $aR + I = R$. Hence there are elements $b \in R$ and $c \in I$ such that $1 = ab + c$. In R/I,

$$[a + I][b + I] = [ab + I] = [1 - c + I] = [1 + I] = 1_{R/I}.$$

Hence $[b + I]$ is the inverse in R/I of $[a + I]$. This shows that R/I is a field.

Conversely, suppose that R/I is a field. Let J be an ideal of R that is strictly larger than I. There is an element $a \in J \backslash I$. Since $[a + I]$ is a non-zero element of R/I, it has an inverse, say $[b + I]$. Since

$$1_{R/I} = [1 + I] = [a + I][b + I] = [ab + I],$$

$1 - ab \in I$, and $1 \in aR + I \subset J$. Hence $J = R$, showing that I is maximal. □

Algebraic and transcendental elements. Let K be a field and k a subfield of K. An element $a \in K$ is said to be algebraic over k if it is a zero of a non-zero polynomial with coefficients in k. That is, if

$$\lambda_n a^n + \lambda_{n-1} a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0$$

for some $\lambda_0, \ldots, \lambda_n \in k$, not all zero. An equivalent way of saying this is that the homomorphism $\varepsilon : k[x] \to K$ given by $\varepsilon(f) = f(a)$ is not injective.

If a is not algebraic over k we say it is transcendental over k.

We say that k is algebraically closed if the only elements algebraic over k (whatever K may be) are the elements of k itself.

Proposition 4.11. *Let k be a field. The following are equivalent:*

1. *k is algebraically closed;*
2. *the only irreducible polynomials in $k[x]$ are the degree one polynomials;*
3. *every polynomial in $k[x]$ of positive degree has a zero in k.*

1.5 Fields of fractions

The formal construction of the ring $\mathbb{Q}$ as "the field of fractions" of $\mathbb{Z}$ may be copied for any commutative domain R.

So, suppose that R is a commutative ring in which evry product of non-zero elements is non-zero. let $R^* = R - \{0\}$. Define a relation on $R \times R^*$ by

$$(a, b) \sim (c, d) \qquad \text{if } ad = bc. \tag{5-6}$$

This is an equivalence relation. we will write $[a, b]$ for the equivalence class containing (a, b), and write $R \times R^*/\sim$ for the set of equivalence classes.

Lemma 5.1. *$R \times R^*/\sim$ becomes a commutative ring under the definitions*

$$[a/b] + [c/d] := [ad + bc/bd],$$
$$[a/b].[c/d] := [ac/bd].$$

The zero element is $[0/1]$ and the identity is $[1/1]$.

Proof. First one must check that these binary operations are defined unambiguously. Then one must check that all the ring axioms hold. I have done that, and you should do this at least once in your life too. □

Proposition 5.2. *The map $r \mapsto [r/1]$ is an injective ring homomorphism $\iota : R \to R \times R^*/\sim$. Identifying R with its image, every non-zero element of R is a unit in $R \times R^*/\sim$, namely $r^{-1} = [1/r]$ if $r \neq 0$. Every element in $R \times R^*/\sim$ is of the form rs^{-1} for some $r \in R$ and $s \in R^*$. If $\varphi : R \to F$ is any homomorphism from R to a field F there is a unique homomorphism $a : R \times R^*/\sim \to F$ such that $\varphi = \alpha \circ \iota$.*

We call the ring $R \times R^*/\sim$ the field of fractions of R and denote it by Fract R.

1.6 Zeroes of polynomials

One of the great motivating problems for the development of algebra was the question of finding the zeroes, or roots, of a polynomial in one variable.

The question of whether an element $\alpha \in k$ is a zero of a polynomial $f \in k[x]$ can be expressed formally as follows: *is f in the kernel of the ring homomorphism $\varepsilon_\alpha : k[x] \to k$ defined by*

$$\varepsilon_\alpha(f) = f(\alpha)?$$

You should check that ε_α is a ring homomorphism; indeed, the ring structure on $k[x]$ is defined just so this is a homomorphism. The kernel of ε_α is an ideal that contains $x - \alpha$ and therefore the ideal $(x - \alpha)$. However, $(x - \alpha)$ is a maximal ideal. We therefore have the following result.

Lemma 6.1. *If $f \in k[x]$, then $x - \alpha$ divides f if and only if $f(\alpha) = 0$.*

Definition 6.2. *Let $\alpha \in k$ and $0 \neq f \in k[x]$. We say that α is a* zero of f of multiplicity *n if $(x-\alpha)^n$ divides f but $(x-\alpha)^{n+1}$ does not.* $\Diamond$

Proposition 6.3. *Let f be a monic polynomial in $k[x]$. If $\alpha_1, \ldots, \alpha_r$ are the distinct zeroes of f, and α_i is a zero of multiplicity n_i, then*

$$f = (x-\alpha_1)^{n_1} \cdots (x-\alpha_r)^{n_r} g$$

where g is a polynomial having no zeroes in k.

Proof. We argue by induction on the number of zeroes and multiplicity, cancelling a factor of the form $x-\alpha$ at each step. $\square$

The next result is very devious. It shows that if f is a non-constant polynomial with coefficients in a field k, then there is a larger field K in which f has a zero. Of course, the first example that comes to mind is the polynomial x^2+1 in which case $\mathbb{C}$, the field of complex numbers, contains a zero of the polynomial. However, notice that the proof is essentially a tautology.

Proposition 6.4. *Let f be a non-constant polynomial in $k[x]$. Then f has a zero in some extension field $K \supset k$.*

Proof. Since every polynomial is a product of irreducible polynomials it suffices to prove this when f is irreducible. When f is irreducible (f) is a maximal ideal, so $K := k[x]/(f)$ is a field.

Let $\pi : k[x] \to K$ denote the natural map, and write $\bar{x} = \pi(x)$. If $f = \sum_{i=0}^{n} \lambda_i x^i$, then

$$f(\bar{x}) = \sum_{i=0}^{n} \lambda_i \bar{x}^i = \pi\Big(\sum_{i=0}^{n} \lambda_i x^i\Big) = \pi(f) = 0.$$

Hence $\bar{x}$ is a zero of f. $\square$

This result suggests that we undertake a systematic examination of fields that contain a given field k.

A field K is called an extension of a field k if k is a subfield of K. We give a more formal definition of an extension field on page **??**.

If K is an extension of k, the action of k on K by multiplication makes K into a k-vector space. We may therefore define the degree of K over k to be

$$[K : k] = \dim_k K.$$

We say that K is a finite extension if $[K : k] < \infty$.

The trivial observation that K is a vector space over k already has important consequences. For example, if p is a prime and K is a finite extension of $\mathbb{F}_p$, the field of p elements, then $[K : \mathbb{F}_p] = n$ implies that $|K| = |\mathbb{F}_p^n| = |\mathbb{F}_p|^n = p^n$. On the other hand, if K is a finite field, then the map $\mathbb{Z} \to K$ sending 1 to 1 has a non-zero kernel which must be of the form (p) for some prime p, so K is an extension of $\mathbb{F}_p$. Hence a finite field must have cardinality p^n for some prime p and integer $n \geq 1$.

It is natural to ask if there is a field of cardinality p^n. As we shall see later, there is a unique field of cardinality p^n up to isomorphism. We write $\mathbb{F}_{p^n}$ for the field with p^n elements.

If f is an irreducible polynomial in $\mathbb{F}_p[x]$ of degree n, then $\mathbb{F}_p[x]/(f) \cong \mathbb{F}_{p^n}$. You should try some examples with small n and p and see what you can find.

Using the fact that an extension field is a vector space over any of its subfields, if $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$, then $m|n$ because if $[\mathbb{F}_{p^n} : \mathbb{F}_{p^m}] = r$, then $\mathbb{F}_{p^n}$ is isomorphic to the r-dimensional vector space over $\mathbb{F}_{p^m}$, and therefore

$$p^n = |\mathbb{F}_{p^n}| = |\mathbb{F}_{p^m}|^r = p^{mr}.$$

If K is an extension of k, and $\alpha_1, \ldots, \alpha_n \in K$ we write

$$k(\alpha_1, \ldots, \alpha_n)$$

for the smallest subfield of K that contains k and $\alpha_1, \ldots, \alpha_n$.

Example 6.5. *Write $\omega = \sqrt{2} + \sqrt{3}$, and $K = \mathbb{Q}(\omega)$. What is $[K : \mathbb{Q}]$? Since $\omega^2 = 5 + 2\sqrt{6}$ and $(\omega^2 - 5)^2 = 24$, the minimal polynomial of ω divides $(x^2 - 5)^2 - 24$. Hence $[\mathbb{Q}(\omega) : \mathbb{Q}] \leq 4$. The computation of ω^2 shows that $\sqrt{6} \in \mathbb{Q}(\omega)$; taking a $\mathbb{Q}$-linear combination of $\sqrt{6}\omega$ and ω shows that $\sqrt{2}$ and $\sqrt{3}$ are in $\mathbb{Q}(\omega)$. Hence $\mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\omega)$. But $[\mathbb{Q}(\sqrt{2} : \mathbb{Q}] = 2$, so $[\mathbb{Q}(\omega) : \mathbb{Q}]$ is even. An elementary computation shows that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$, whence $\mathbb{Q}(\omega)$ is strictly larger than $\mathbb{Q}(\sqrt{2})$. It follows that $[\mathbb{Q}(\omega) : \mathbb{Q}] \geq 4$. Hence $[\mathbb{Q}(\omega) : \mathbb{Q}] = 4$ and the minimal polynomial of ω is $(x^2 - 5)^2 - 24$.* ◊

Example 6.6. *Let's construct $\mathbb{F}_{25}$. Since 2 is not a square in $\mathbb{F}_5$, $x^2 - 2$ is an irreducible polynomial in $\mathbb{F}_5[x]$, whence $\mathbb{F}_{25} \cong \mathbb{F}_5[x]/(x^2 - 2)$. Write λ for the image of x in $\mathbb{F}_{25}$. Viewing $\mathbb{F}_{25}$ as a two-dimensional vector space over $\mathbb{F}_5$, we have $\mathbb{F}_{25} = \mathbb{F}_5 \oplus \mathbb{F}_5\lambda$, so every element of $\mathbb{F}_{25}$ can be written uniquely as*

$$a + b\lambda, \qquad a, b \in \mathbb{F}_5$$

and the multiplication is given by

$$(a + b\lambda)(c + d\lambda) = ac + bd\lambda^2 + (ad + bc)\lambda = ac + 2bd + (ad + bc)\lambda.$$

Notice that 3 is not a square in $\mathbb{F}_5$. We can ask whether it has a square root in $\mathbb{F}_{25}$. Now $(a + b\lambda)^2 = 3$ if and only if

$$a^2 + 2b^2 = 3 \qquad \text{and} \qquad 2ab = 0.$$

We see that $a = 0$ and $b = 2$ is a solution. Thus $(2\lambda)^2 = 3$. ◊

Proposition 6.7. *Let $k \subset K \subset L$ be fields. Then $[L : k] = [L : K][K : k]$ if any two of these degrees are finite (and then the third is also finite).*

Proof. If $[L : k]$ is finite, then the other two degrees are finite, so suppose that $[L : K]$ and $[K : k]$ are finite. If $\{\alpha_1, \ldots, \alpha_n\}$ is a k-basis for K and $\{\beta_1, \ldots, \beta_m\}$ is a K-basis for L, then

$$\begin{aligned} L &= K\beta_1 \oplus \cdots \oplus K\beta_m \\ &= (k\alpha_1 \oplus \cdots \oplus k\alpha_n)\beta_1 \oplus \cdots \oplus (k\alpha_1 \oplus \cdots \oplus k\alpha_n)\beta_m \\ &= \bigoplus_{i=1}^{m} \bigoplus_{j=1}^{n} k\alpha_j\beta_i, \end{aligned}$$

so $\{\alpha_j\beta_i\}$ is a k-basis for L. Hence $[L : k] = mn = [L : K][K : k]$. □

If R is any ring containing k as a subring and $\alpha \in R$, there is a unique ring homomorphism $\psi : k[x] \to R$ which is the identity on k and sends x to α. If ψ is injective we say that α is transcendental over k, otherwise we say that α is algebraic over k, and we call the unique monic generator of the ideal $\ker \psi$ the minimal polynomial of α.

The element α is transcendental if and only if $\{1, \alpha, \alpha^2, \ldots\}$ is linearly independent over k. If α is algebraic, then there is a linear dependence relation between $1, \alpha, \ldots, \alpha^n$ where n is the degree of the minimal polynomial of α.

If K is an extension field of k, and $\alpha \in K$ is algebraic over k, then the image of $k[x]$ is a domain and has finite dimension over k (equal to the degree of the minimal polynomial of α). Hence that image is a field, and we deduce that the minimal polynomial of α is irreducible. Furthermore, the image of $k[x]$ is equal to $k(\alpha)$, the subfield of K generated by k and α.

Proposition 6.8. *Let K be an extension of k and $\alpha \in K$ an element that is algebraic over k. Then*

$$[k(\alpha) : k] = \deg p$$

where p is the minimal polynomial of α over k.

Example 6.9. *For each positive integer write $\zeta_n = e^{2\pi i/n}$. Since $x^n = 1$, the minimal polynomial of ζ_n divides $x^n - 1$. Since $x^n - 1 = (x-1)(x^{n-1} + \cdots + x + 1)$, it follows that the minimal polynomial of ζ_n divides $x^{n-1} + \cdots + x + 1$. If $n = p$ is prime, the polynomial $x^{p-1} + \cdots + x + 1$ is irreducible by 4.8, so is the minimal polynomial of ζ_p.* ◊

We say that K is an algebraic extension of k if every element of K is algebraic over k.

Lemma 6.10. *If K is a finite extension of k, then it is an algebraic extension.*

Proof. If $\alpha \in k$, then the map $k[x] \to K$, $x \mapsto \alpha$, cannot be injective for dimension reasons, so α is algebraic. □

Remember that $\overline{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$, is an algebraic extension of $\mathbb{Q}$ that is not a finite extension.

Lemma 6.11. *Let K be an extension of k. Then $[K : k] < \infty$ if and only if K is an algebraic extension of k and $K = k(\alpha_1, \ldots, \alpha_n)$ for some $\alpha_1, \ldots, \alpha_n$.*

Proof. ($\Rightarrow$) By Lemma 6.10, K is algebraic over k. Also, if $\alpha_1, \ldots, \alpha_n$ is a k-basis for K, then $K = k(\alpha_1, \ldots, \alpha_n)$.

($\Leftarrow$) We argue by induction on n, the case $n = 0$ being trivial. Set $L = k(\alpha_1, \ldots, \alpha_{n-1})$. The induction hypothesis implies that $[L : k] < \infty$. Now $K = L(\alpha_n) \cong L[x]/(p)$ where p is the minimal polynomial of α_n over L. Hence $[K : L] < \infty$. Thus $[K : k] = [K : L][L : k] < \infty$. □

Three impossible constructions. The ancients asked whether it was possible, using only a straightedge and compass, to double a cube, trisect an angle, and square the circle.

Using a straightedge and compass one can

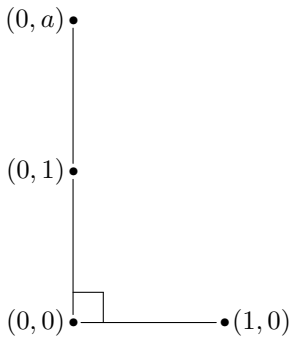
1. draw a straight line through two given points, and
2. draw a circle with given center and radius.

Starting with a line segment of length one, we can construct the lattice $\mathbb{Z}^2$ of points (a, b) with integer coordinates in the plane $\mathbb{R}^2$. For example, it is easy to construct the integer points on the x-axis, and the construct a perpendicular, and continue doing the obvious thing to obtain $\mathbb{Z}^2$.

A point in the plane is constructible if it can be obtained by repeating the constructions (1) and (2) in some order a suitable number of times. More precisely, suppose that at some stage we have constructed the points $\mathcal{P}$ (initially $\mathcal{P}$ consists of the points in the lattice $\mathbb{Z}^2$); using (1) we can draw straight lines between distinct points of $\mathcal{P}$ and the intersection points of such lines are costructible; using (2) we can draw a circle with center $p \in \mathcal{P}$ and radius equal to a segment joining two points of $\mathcal{P}$; the points of intersection of such circles are constructible; the points of intersection of such circles with the lines between points of $\mathcal{P}$ are also constructible.

We say that $a \in \mathbb{R}$ is constructible if there is $b \in \mathbb{R}$ such that either (a, b) or (b, a) is constructible. Thus a is constructible if and only if we can construct a line segment of length a.

The constructible numbers form a field. Obviously the sum and difference of two constructible numbers is constructible. Products can be constructed by constructing similar triangles. Inverses can be constructed similarly. To see this, suppose that $a > 1$ has been constructed, and consider the problem of constructing a^{-1}. First construct line segments as below:



Now construct two similar triangles by drawing the line through $(0, a)$ and $(1, 0)$ to give the big hypotenuese, and then construct a line parallel to that through $(0, 1)$ giving the smaller triangle; that line will meet the base at $(a^{-1}, 0)$, showing that a^{-1} is constructible.

Since the integers are constructible, $\mathbb{Q}$ is constructible.

Now suppose that we have constructed all elements of a field k lying between $\mathbb{Q}$ and $\mathbb{R}$, $\mathbb{Q} \subset k \subset \mathbb{R}$. That is the points constructed so far contain all $(a, b) \in k^2 \subset \mathbb{R}^2$. If we make a single new construction to obtain a point (a, b) then a belongs to $k(\sqrt{d})$ for some $d \in k$; similarly for b. For example, Pythagoras's theorem shows that the length of the line segment joining two points of k^2 is of length $\sqrt{d}$ for some $d \in k$.

Corollary 6.12. *If $a \in \mathbb{R}$ is constructible, then there is a sequence of fields $\mathbb{Q} = k_0 \subset k_1 \subset \cdots \subset k_n$ such that $[k_i : k_{i-1}] = 2$ for all i. In particular, the degree of the minimal polynomial of a is of the form 2^m.*

Proof. Since $\mathbb{Q} \subset \mathbb{Q}(a) \subset k_n$, $2^n = [k_n : \mathbb{Q}] = [k : \mathbb{Q}(a)][\mathbb{Q}(a) : \mathbb{Q}]$. □

Cannot double a cube. Suppose our original cube has sides of length one. The cube of twice the volume has sides of length $2^{1/3}$. The minimal polynomial of $2^{1/3}$ is $x^3 - 2$, so $[\mathbb{Q}(2^{1/3}) : \mathbb{Q}] = 3 \neq 2^m$.

$\pi/3$ cannot be trisected. If it were possible to construct $\pi/9$, then $a = 2\cos(\frac{\pi}{9})$ would be constructible. But substituting $\theta = \frac{\pi}{9}$ into the identity $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ gives $8\alpha^3 - 6\alpha - 1 = 0$. Since $8x^3 - 6x - 1$ is irreducible over $\mathbb{Q}$ (you can check it has no zero in $\mathbb{Z}_5$) it is the minimal polynomial of α. Hence $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3 \neq 2^m$.

Cannot square the circle. The square with area equal to the circle of radius one has sides of length $\sqrt{\pi}$. If this were constructible, then π would be constructible, and hence algebraic over $\mathbb{Q}$. But F. Lindemann proved that π is transcendental in 1882.

1.7 Pythagoras and integers

This section is preparation for the discussion of Fermat's last theorem that appears in the next section.

As already suggested in section 1.1, the fact that integers may be uniquely factored as a product of primes is a powerful tool for the analysis and solution of integer equations. For example, $2x = 3$ has no solutions in the integers because 3 is not divisible by 2. Slightly more thought shows that there are no integers x and y such that $3x^2 + y^2 = 54$—if (x, y) were a solution to this equation, then y would be divisible by 3 so, writing $y = 3a$, we would have $x^2 + 3a^2 = 18$, whence 3 divides x, and writing $x = 3b$, we get $3b^2 + a^2 = 6$, so a is divisible by 3 and, writing $a = 3c$, we obtain $b^2 + 3c^2 = 2$, an equation that obviously has no integer solutions.

The proof of the following result is elementary, albeit tedious, but notice how it uses the unique factorization property of the integers in an essential way.

Proposition 7.1. *Let v and w be relatively prime integers. If vw is a square, then both v and w are squares.*

Proof. It is helpful in this proof to assume that all the numbers appearing in it are positive. Let's do that. It is not essential, but it makes things a little cleaner.

Before proving the result, we show that the uniqueness of factorization implies that if a prime p divides a product ab, then it must divide either a or b. This is because if $ab = px$, and we write $a = p_1^{i_1} \cdots p_m^{i_m}$, $b = q_1^{j_1} \cdots q_n^{j_n}$, and $x = r_1^{k_1} \cdots r_t^{k_t}$, as products of positive powers of distinct primes, then

$$pr_1^{k_1} \cdots r_t^{k_t} = p_1^{i_1} \cdots p_m^{i_m} q_1^{j_1} \cdots q_n^{j_n}$$

so, by the uniqueness of factorization into primes, $p \in \{p_1, \ldots, p_m, q_1, \ldots, q_n\}$. Thus p divides either a or b.

If the result fails we can pick a smallest pair v and w for which the result fails. Write $vw = z^2$.

Write $v = p_1^{i_1} \cdots p_m^{i_m}$ and $w = q_1^{j_1} \cdots q_n^{j_n}$ as products of positive powers of primes. By hypothesis, p_1 does not divide w so is not equal to any of the q_ks. But p_1 divides z^2, so by the previous observation, p_1 divides z. Hence p_1^2 divides $z^2 = vw$. So we can write $vw = p_1^2 r_1^{k_1} \cdots r_t^{k_t}$ for some primes r_i. But uniqueness of factorization says this is the same as the factorization $vw = p_1^{i_1} \cdots p_m^{i_m} q_1^{j_1} \cdots q_n^{j_n}$, so we conclude that $i_1 \geq 2$, whence p_1^2 divides v. This yields an equation $(v/p_1^2)w = (z/p_1)^2$ in integers in which v/p_1^2 is smaller than v, contradicting our original choice of v and w. We conclude that no such v and w can exist, so this proves the result. □

A high point of the application of unique factorization is the classification of the integer solutions to the equation

$$x^2 + y^2 = z^2. \tag{7-7}$$

This equation, motivated by Pythagoras's Theorem, was studied in antiquity, and complete solutions to it were independently found by several ancient cultures.

We will restrict our attention to positive integer solutions because all others can be obtained from these in an obvious way. First, observe that if each of x, y, and z, is divisible by a number d, then xd^{-1}, yd^{-1}, zd^{-1} is also a solution to the equation. Thus every solution is obtained from a primitive solution, that is one in which x, y, and z have no common factor. It therefore suffices to classify the primitive solutions. However, if d divides two of x, y, and z, it must divide the third, so if x, y, z is a primitive solution, the greatest common divisor of any two of x, y, and z, is one. Hence, at most one of x, y, and z, is even, and at least two are odd. But if two of them are odd, the other must be even because a sum or difference of two odd numbers is even. However, if x and y are odd, then both x^2 and y^2 leave a remainder of one when divided by four, and z^2 must therefore leave a remainder of two when divided by four. But this is impossible,

so we conclude that either x or y is even, and z is odd. We can assume without loss of generality that x is even and y is odd.

Now rewrite the equation as $x^2 = (z+y)(z-y)$. Because x, $y+z$, and $y-z$ are all even, there are integers u, v, and w, such that $x = 2u$, $y+z = 2v$, and $y-z = 2w$. Hence $u^2 = vw$.

I claim that v and w are relatively prime, because if p divided them both it would divide $v+w = y$ and $v-w = z$, which are relatively prime. Now, unique factorization implies that **if a product vw of relatively prime numbers v and w is a square, then v and w must be squares**. Hence there are integers a and b such that $v = a^2$ and $w = b^2$, and a and b must be relatively prime because v and w are. It follows that $y = a^2 + b^2$ and $z = a^2 - b^2$. Now,

$$x^2 = (y+z)(y-z) = 4vw = 4a^2b^2,$$

and $x = 2ab$.

Since we required x, y, and z, to be positive, a and b are positive and $a > b$. We have therefore proved the following result.

Theorem 7.2. *A complete list of the positive primitive solutions to the equation $x^2 + y^2 = z^2$ is given by*

$$x = 2ab, \quad y = a^2 + b^2, \quad z = a^2 - b^2,$$

where a and b are arbitrary positive integers with $a > b$.

Exercises.

1. Suppose that $a_1, \ldots, a_r$ are pairwise relatively prime integers, i.e., $\gcd(a_i, a_j) = 1$ for all $i \neq j$. If b is an integer such that $a_1 a_2 \cdots a_r = b^n$, show that each a_i is an n^{th}-power of an integer. This is an easy exercise, but notice that your proof depends on the fact that every integer can be written as a product of primes in a *unique* way.

2. Show that $\mathbb{Z}[\sqrt{-3}] := \{a + b\sqrt{-3} \mid a, b \in \mathbb{Z}\}$ is a ring; i.e., that sums and products of elements of this form are again of this form.

3. We call a non-zero element $u \in \mathbb{Z}[\sqrt{-3}]$ a unit if u^{-1} belongs to $\mathbb{Z}[\sqrt{-3}]$. Find all the units in $\mathbb{Z}[\sqrt{-3}]$.

4. Show that if $a + b\sqrt{-3}$ divides both $1 + \sqrt{-3}$ and $1 - \sqrt{-3}$ in $\mathbb{Z}[\sqrt{-3}]$, then $a + b\sqrt{-3}$ is a unit.

5. By the previous exercise, $v = 1 + \sqrt{-3}$ and $w = 1 - \sqrt{-3}$ are relatively prime in $\mathbb{Z}[\sqrt{-3}]$. Show that vw is a square in $\mathbb{Z}[\sqrt{-3}]$ despite the fact that neither v nor w is a square in $\mathbb{Z}[\sqrt{-3}]$.

6. Show that the smallest subring of $\mathbb{C}$ containing $\mathbb{Z}$ and $\frac{1}{2}(1 + \sqrt{-3})$ is

   $$R = \{a + \tfrac{b}{2}(1 + \sqrt{-3}) \mid a, b \in \mathbb{Z}\}.$$

   Is this the same as

   $$\{\tfrac{1}{2}(a + b\sqrt{-3}) \mid a, b \in \mathbb{Z}\}?$$

7. Let $\zeta_3 = e^{2\pi i/3}$. Show that the smallest subring of $\mathbb{C}$ containing $\mathbb{Z}$ and ζ_3 is the ring R in the previous exercise. We write $\mathbb{Z}[\zeta_3]$ for this ring.

8. Show that every element in $\mathbb{Z}[\zeta_3]$ satisfies a monic polynomial with coefficients in $\mathbb{Z}$.

9. Working in the ring

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$$

show that

(a) neither $4 + 4\sqrt{-5}$ nor $9 - 9\sqrt{-5}$ is a cube;

(b) their product is a cube;

(c) if $u = a + b\sqrt{-5}$ divides both $4 + 4\sqrt{-5}$ and $9 - 9\sqrt{-5}$, then $\mathbb{Z}[\sqrt{-5}]$ contains u^{-1}.

1.8 Fermat's Last Theorem

The initial impetus for the development of abstract algebra came from number theory, especially attempts to prove Fermat's conjecture that if p is an integer ≥ 3 then there no non-zero integers x, y, and z, such that

$$x^p + y^p = z^p. \tag{8-8}$$

The tale has been told many times. I won't repeat it. The book *Fermat's Last Theorem* by H.M. Edwards is an excellent historical account. The account below is taken from Edwards's book.

In his papers Fermat left a proof that there are indeed no solutions when $p = 4$. It therefore suffices to establish his conjecture in the case when p is a prime number, so we shall assume that p is prime in our discussion from now on.

One observes easily that if there is a solution to (8-8) in which x, y, and z have a common factor, one may cancel that factor to obtain another solution with smaller x, y, and z. Thus, if one wants to argue by contradiction, one can assume that one has a solution in which no two of x, y, and z has a common factor. Furthermore, one can assume that if there are solutions, then there is a "smallest" solution in an appropriate sense.

The proceedings of the meeting of the Paris Academy on March 1, 1847, serve to illustrate the forces driving the early development of abstract algebra. Lamé announced that he had a proof of Fermat's conjecture. The starting point of his approach was the factorization

$$x^p + y^p = (x + y)(x + \zeta_p y) \cdots (x + \zeta_p^{p-1} y) \tag{8-9}$$

where $\zeta_p = e^{2\pi i/p}$. He planned to split the argument into two cases: if the factors $(x + \zeta_p^i y)$ are pairwise relatively prime, then the fact that their product

is a p^{th}-power implies that each $(x + \zeta_p^i y)$ is a p^{th}-power; on the other hand, if the factors are not pairwise relatively prime, Lamé planned to show that they shared a common factor, and then dividig through by that common factor obtain a smaller solution to (8-8).

Liouville objected to Lamé's claim that the only way a product of relatively prime "numbers" could be a p^{th}-power was if each number was itself a p^{th}-power. In modern language, Lamé needed to prove that every number in the ring $\mathbb{Z}[\zeta_p]$, that is every number of the form

$$a_0 + a_1\zeta_p + \cdots + a_{p-1}\zeta_p^{p-1} \qquad (a_0, \ldots, a_{p-1} \in \mathbb{Z}),$$

could be written as a product of primes in a unique way.

Liouville's objection can be appreciated by a consideration of Euler's "proof" for $p = 3$ that appeared in his 1770 book on algebra. By straightforward and solid arguments (see pages 40-41 of Edwards's book), Euler shows that if there is a solution to $x^3 + y^3 = z^3$, then there exist relatively prime integers u and v, one odd and one even, such that

$$2u(u^2 + 3v^2) = \text{ a cube.} \tag{8-10}$$

Euler's proof then breaks into two cases depending on whether or not 3 divides u. Let's consider the case where 3 does not divide u. Because u and v are relatively prime and $u^2 + 3v^2$ is odd, it follows easily that $2u$ and $u^2 + 3v^2$ are relatively prime. Hence $2u$ and $u^2 + 3v^2$ are cubes. As Edwards explains on page 41 of his book, "one way to find cubes of the form $u^2 + 3v^2$ is to choose a, b at random and to set

$$u = a^3 - 9ab^2 \qquad v = 3a^2b - 3b^3$$

so that $u^2 + 3v^2 = (a^2 + 3b^2)^3$. The major gap [...] in Euler's proof is [his claim] that this is the *only* way that $u^2 + 3v^2$ can be a cube". The remainder of Euler's proof is solid.

Euler tried to justify his claim about cubes of the form $u^2 + 3v^2$ by arguments involving numbers of the form

$$a + b\sqrt{-3} \qquad (a, b \in \mathbb{Z}).$$

An exercise in the previous section showed that the set of these numbers is a ring, denoted $\mathbb{Z}[\sqrt{-3}]$. Euler's argument was based on the factorization

$$u^2 + 3v^2 = (u + v\sqrt{-3})(u - v\sqrt{-3})$$

in $\mathbb{Z}[\sqrt{-3}]$. Euler noted that if one of these factors is a cube, say $u + v\sqrt{-3} = (a + b\sqrt{-3})^3$, then $u = a^3 - 9ab^2$ and $v = 3a^2b - 3b^3$. He also observed that $u - v\sqrt{-3} = (a - b\sqrt{-3})^3$ and so

$$\begin{aligned} u^2 + 3v^2 &= (u + v\sqrt{-3})(u - v\sqrt{-3}) \\ &= (a + b\sqrt{-3})^3(a - b\sqrt{-3})^3 \\ &= (a^2 + 3b^2)^3. \end{aligned}$$

That is, if $u+v\sqrt{-3} = (a+b\sqrt{-3})^3$, then $u^2+3v^2 = (a^2+3b^2)^3$ is a cube. Euler's error was to take this *sufficent* condition for $u^2 + 3v^2$ to be a cube and treat it as if it were a *necessary* condition. One finds in Euler's book the statement that "[if x and y are relatively prime integers and] $x^2 + cy^2$ is [...] a cube, one can certainly conclude that [...] $x + y\sqrt{-c}$ and $x - y\sqrt{-c}$ must be cubes, because they are relatively prime in that x and y have no common factor." In this generality, Euler's statement is false—

Edwards's speculates that when Euler wrote to Goldbach in 1753 that he had proves the $p = 3$ case of Fermat's last theorem, he had in mind an argument that did not involve the factorization in $\mathbb{Z}[\sqrt{-3}]$.

One can imagine that the proof Fermat had in mind when he made his historical marginal note was based on the factorization (8-9). In any case, this factorization was used by Lagrange, by Euler in proving Fermat's theorem for $n = 3$, by Gauss for $n = 5$, and by Dirichlet for $n = 14$. In his proof for $n = 3$, Euler assumed that $\mathbb{Z}[\zeta_3]$ is a UFD; since $\mathbb{Z}[\zeta_3]$ is a UFD, Euler's proof is correct. It can be shown that Fermat's Theorem holds if $\mathbb{Z}[\zeta_n]$ is a UFD; unfortunately (or fortunately, depending on your point of view) it is not a UFD for all values of n. Kummer was the first to realize this, and he developed the theory of ideals (= ideal numbers) to recover this lack of unique factorization: every ideal in $\mathbb{Z}[\zeta_n]$ can be written as a product of prime ideals in a unique way.

Exercise. Is $\mathbb{Z}[\zeta_n]$ isomorphic to $\mathbb{Z}[x]/(x^n - 1)$? If not, what is the relation between these rings?

Exercise. The failure of unique factorization in $\mathbb{Z}[\sqrt{-5}]$ can be repaired in some sense. The *ideal* generated by 6 can be written in a unique way as a product of prime ideals:

$$(6) = (2, 1 + \sqrt{-5})^2.(3, 1 + \sqrt{-5}).(3, 1 - \sqrt{-5}).$$

To see that $(2, 1+\sqrt{-5})$, $(3, 1+\sqrt{-5})$, and $(3, 1-\sqrt{-5})$ are prime ideals compute the quotient ring and check that it is a domain.

1.9 Domains and fields

A ring R is a domain, or an integral domain, if every pair of non-zero elements in R has a non-zero product. Thus, in a domain, a product xy can only be zero if either x or y is zero. We can therefore cancel in a domain: if $ax = ay$ and $a \neq 0$, then $x = y$ because $a(x - y) = 0$.

The ring of integers is a domain: a product of non-zero integers is non-zero. A field is a domain because if x is a non-zero element and $xy = 0$, then $0 = x^{-1}.0 = x^{-1}.xy = 1.y = y$. Every subring of a field is a domain. We will soon show that every domain is a subring of a field.

It is easy to show that $\mathbb{Z}/(a)$ is a domain if and only if a is a prime number or zero. For example, $\mathbb{Z}/(6)$ is not a domain because $[2+(6)].[3+(6)] = [6+(6)] = 0$.

A simple geometric example of a commutative ring that is not a domain is provided by the ring of k-valued functions on a space X that has more than one

element: if x and y are different points of X, then the product of the non-zero functions f and g, defined by $f(x) = g(y) = 1$ and $f(X\backslash\{x\}) = g(X\backslash\{y\}) = 1$, is zero. Another geometric example occurs for the ring of continuous $\mathbb{R}$-valued functions on the topological space $X \subset \mathbb{R}^2$ that is the union of the usual x- and y-axes; the functions f and g that take, respectively, the x- and y-coordinates of a point $p \in X$ are both non-zero, but their product is zero. This last example is a baby example of the general fact that the coordinate ring of an affine algebraic variety is a domain if and only if the variety is irreducible.

The next two exercises show that under appropriate finiteness conditions a domain must in fact be a field. These exercises can be considered as warm ups for Proposition 12.6 which gives another finiteness condition that ensures a domain is a field.

Exercise. Show that a finite commutative domain is a field.

Exercise. Let R be a commutative domain containing a field k. Show that R is a field if $\dim_k R < \infty$.

As a consequence of the previous exercise, the rings $\mathbb{Q}[\sqrt{d}]$ are actually fields. More generally, if R is any subring of $\mathbb{C}$ containing $\mathbb{Q}$, R is a field if $\dim_{\mathbb{Q}} R < \infty$.

Fields of fractions.

Just as the field of rational numbers can be constructed from the ring of integers, so too does every commutative domain R have a field of fractions, denoted $\operatorname{Fract} R$, the elements of which can be written as fractions a/b with $a, b \in R$ and $b \neq 0$. You probably don't need much persuasion to believe this, but the formal construction of $\operatorname{Fract} R$ is as follows.

Let R be a domain. Define an equivalence relation on the cartesian product $R \times R\backslash\{0\}$ as follows:

$$(a, b) \sim (c, d) \quad \text{if } ad = cb.$$

Check this is an equivalence relation. We denote the equivalence class of (a, b) by a/b. We now impose a ring structure on the set of equivalence classes by defining addition by

$$(a/b) + (c/d) = (ad + bc)/bd,$$

and multiplication by

$$(a/b).(c/d) = ac/bd.$$

Check that these two binary operations are well-defined. Check that under $+$, the equivalence classes form an abelian group with zero element $0/1$. Check that $1/1$ is an identity element for the multiplication. Check that the equivalence classes form a ring with identity. We denote this ring by $\operatorname{Fract} R$ and call it the field of fractions of R.

Check that there is an injective homomorphism $\rho : R \to \operatorname{Fract} R$ defined by $r \mapsto r/1$. We usually identify R with its image in $\operatorname{Fract} R$ under this map, and think of R as a subring of $\operatorname{Fract} R$. Each non-zero element $b \in R$ has an inverse in $\operatorname{Fract} R$, namely $1/b$. We often write b^{-1} for $1/b$.

Exercise. Let R be a commutative domain, and $\varphi : R \to S$ a ring homomorphism such that $\varphi(b)$ is a unit in S for every non-zero element $b \in R$. Show there is a unique ring homomorphism $\psi : \operatorname{Fract} R \to S$ such that $\varphi = \psi\rho$, where $\rho : R \to \operatorname{Fract} R$ is the map in the previous paragraph.

It is often useful when dealing with a domain R to conisder rings between R and $\operatorname{Fract} R$ that are obtained by inverting only some of the non-zero elements in R. We now consider this matter.

If $\mathcal{S}$ is a subset of R that does not contain zero and R' is a ring lying between R and $\operatorname{Fract} R$ in which every element of $\mathcal{S}$ is a unit, then every product of elements from $\mathcal{S}$ is also a unit in R' because $(st)^{-1} = s^{-1}t^{-1}$. It therefore makes sense to assume that $\mathcal{S}$ is closed under multiplication; such a subset of R is said to be multiplicatively closed.

Proposition 9.1. *Let R be a commutative domain and $\mathcal{S}$ a mutiplicatively closed subset of R that does not contain 0. Then*

$$R_{\mathcal{S}} := R[\mathcal{S}^{-1}] := \{as^{-1} \mid a \in R,\ s \in \mathcal{S}\}$$

is a subring of $\operatorname{Fract} R$ *containing R. Moreover,*

1. *$R[\mathcal{S}^{-1}]$ is the smallest subring of* $\operatorname{Fract} R$ *containing R in which every element of $\mathcal{S}$ is a unit;*

2. *every ideal of $R_{\mathcal{S}}$ is of the form $IR_{\mathcal{S}}$ for some ideal I in R;*

3. *if J is an ideal of $R_{\mathcal{S}}$, then $J = (J \cap R)R_{\mathcal{S}}$;*

4. *if every ideal of R is finitely generated, so is every ideal of $R_{\mathcal{S}}$.*

Proof. To see that $R_{\mathcal{S}}$ is a ring, we need to check that it is closed under products and sums. If $x, y \in R_{\mathcal{S}}$, we can write $x = as^{-1}$ and $y = bt^{-1}$ for some $a, b \in R$ and $s, t \in \mathcal{S}$. It follows that $xy = ab(st)^{-1}$ and $x + y = (at + bs)(st)^{-1}$. Since $st \in \mathcal{S}$, xy and $x + y$ belong to $R_{\mathcal{S}}$.

(1) If $s \in \mathcal{S}$, then $s^{-1} = 1.s^{-1}$ belongs to $R_{\mathcal{S}}$, so every element of $\mathcal{S}$ is a unit in $R_{\mathcal{S}}$. On the other hand, if R' is a ring lying between R and $\operatorname{Fract} R$ and contains s^{-1} for each $s \in \mathcal{S}$, then R' contains as^{-1} for every $a \in R$, so contains $R_{\mathcal{S}}$.

(2) and (3). Let J be an ideal of $R_{\mathcal{S}}$. Then $I := R \cap J$ is an ideal of R, so (2) follows from (3). Since J contains I, it contains $IR_{\mathcal{S}}$. To prove the converse, suppose that $x \in J$. Then $x = as^{-1}$ for some $a \in R$ and $s \in \mathcal{S}$. Since J is an ideal it contains $xs = a$. Thus $a \in J \cap R = I$, and $as^{-1} \in IR_{\mathcal{S}}$.

(4) If J is an ideal of $R_{\mathcal{S}}$, then $J \cap R$ is a finitely generated ideal of R by hypothesis, so $J = (J \cap R)R_{\mathcal{S}}$ is generated as an ideal of $R_{\mathcal{S}}$ by that same finite set of generators. □

For example, the field of fractions of $\mathbb{Z}[\sqrt{d}]$ is $\mathbb{Q}(\sqrt{d})$.

1.10 Unique factorization domains

Definition 10.1. *A commutative domain R is a* unique factorization domain*, or* UFD*, if every element of R can be written uniquely as a product of irreducible elements, and the irreducibles that occur in the factorization are unique up to order and multiplication by units.* ◊

To see what "uniqueness" means in this definition, consider the factorizations

$$6 = 2.3 = (-3).(-2) = (-3).(-1).(2).(-1).(-1)$$

in $\mathbb{Z}$. The uniqueness means this: if we have two factorizations of an element as a product of irreducibles, and x is an irreducible appearing in one of those factorizations, then some unit multiple of x must appear in the other factorization.

Lemma 10.2. *In a UFD, primes and irreducibles are the same.*

Proof. We observed on page 9 that a prime is irreducible.

Suppose that x is an irreducible and that $x|bc$. Then $bc = xy$ for some y. We can write each of b, c, and y, as a product of irreducibles. Doing so gives two factorizations of bc as a product of irreducibles. Buy the uniqueness of such a factorization, at least one of the irreducibles in the factorizations of b and c must be a unit mutiple of x. But that implies that x divides either b or c, thus showing that x is prime. □

This puts into perspective the non-unique factorization

$$6 = 2.3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

in the ring $\mathbb{Z}[\sqrt{-5}]$.

The proof that $k[x_1, \cdots, x_n]$ is a UFD proceeds by induction on the number of variables. We will show that $R[x]$ is a UFD if R is.

Definition 10.3. *Let R be a UFD, and $R[x]$ the polynomial ring over it. The* content *of $f = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n \in R[x]$ is*

$$c(f) := \gcd(\alpha_0, \ldots, \alpha_n).$$

This is only well-defined up to a unit multiple, so when we write $c(f) = c(g)$ we will mean that $c(f)$ is a unit multiple of $c(g)$.

We call f a primitive *polynomial if $c(f)$ is a unit.* ◊

Remark. If $f \in R[x]$, then $c(f)^{-1}f$ is primitive so every polynomial is a scalar multiple of a primitive polynomial. Thus, every irreducible polynomial in $R[x]$ is primitive.

Lemma 10.4 (Gauss)**.** *Let R be a UFD. If $f, g \in R[x]$, then*

1. *$c(fg) = uc(f)c(g)$ for some unit u, and*

2. a product of primitive polynomials is primitive.

Proof. By the above remark, we can write $f = af_1$ and $g = bg_1$ where $a, b \in R$ and $f_1, g_1 \in R[x]$ are primitive. It suffices to prove that $f_1 g_1$ is primitive. Hence (1) follows from (2).

To prove (2), let $f = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ and $g = \beta_0 + \beta_1 x + \cdots + \beta_m x^m$ be primitive. Suppose, contrary to what we wish to show, that fg is not primitive. Let $p \in R$ be an irreducible that divides all the coefficients of fg. Since f and g are primitive we can choose s and t minimal such that α_s and β_t are not divisible by p. Since p is prime, it does not divide $\alpha_s \beta_t$. The coefficient of x^{s+t} in fg is

$$\cdots + \alpha_{s-1}\beta_{t+1} + \alpha_s \beta_t + \alpha_{s+1}\beta_{t-1} + \cdots .$$

Since p divides all the terms in this sum except $\alpha_s \beta_t$, it does not divide the coefficient of x^{s+t}. This is a contradiction. □

Lemma 10.5. *Let R be a UFD, write $k = \operatorname{Fract} R$, and let $f \in R[X]$.*

1. *Suppose $f = gh$ with $g, h \in k[x]$. Then there exist $G, H \in R[x]$ such that $f = GH$, $\deg G = \deg g$, and $\deg H = \deg h$.*

2. *Suppose f is primitive. Then f is irreducible in $R[x]$ if and only if it is irreducible in $k[x]$.*

Proof. (1) ($\Leftarrow$) This is obvious.

($\Rightarrow$) We may, without loss of generality, assume f is primitive. Write $g = \sum_{i=0}^{n} \alpha_i \beta_i^{-1} x^i$ where all α_i and β_i are in R. Let β be the product of all the β_is, and set $\gamma_i = \beta \alpha_i \beta_i^{-1}$. Thus $g = \beta^{-1} \sum_{i=0}^{n} \gamma_i x^i$ where each $\gamma_i \in R$. Since $\beta g \in R[x]$, we can write it as $\beta g = c(\beta g) G$ where $G \in R[x]$ is primitive. Hence $g = \beta^{-1} c(\beta g) G$. In a similar way, we can write $h = \delta^{-1} c(\delta h) H$ where $\delta \in R$, $\delta h \in R[x]$, and $H \in R[x]$ is primitive.

Because $\beta\delta f = \beta\delta gh = c(\beta g)c(\delta h)GH$ belongs to $R[x]$, we can take the content of both sides and apply Gauss's Lemma to conclude that $\beta\delta = c(\beta g)c(\delta h)$, whence $f = GH$. This is a factorization of f in $R[x]$ and $\deg G = \deg g$.

(2) ($\Leftarrow$) If $f \in R[x]$ factors as $f = GH$ in $R[x]$, then that factorization is also a factorization in $k[x]$, so either G or h is a unit in $k[x]$. If $G \in k$, then $G \in k \cap R[x] = R$, so G divides all the coefficients of $f = GH$. Since f is primitive in $R[x]$ it follows that G is a unit. Hence f is irreducible in $R[x]$.

($\Rightarrow$) Assume f is irreducible in $R[x]$. Suppose $f = gh$ where $g, h \in k[x]$. We must show that the degree of either g or h is zero. By (1) we can write $f = GH$, where $G, H \in R[x]$, $\deg G = \deg g$, and $\deg H = \deg h$. But f is irreducible in $R[x]$ so either G or H is a unit in $R[x]$. Suppose G is a unit in $R[x]$. Then $G \in R \subset k$, and hence $g \in k$ too, showing that f is irreducible in $k[x]$. □

Exercise. Let R be a domain in which every element is a product of irreducibles. Show that R is a UFD if and only if every irreducible is prime.

Theorem 10.6. *If R is a UFD, so is $R[x]$.*

Proof. Let $k = \operatorname{Fract} R$.

Let $f \in R[x]$. We first show that f is a product of irreducibles. Write $f = \alpha f_1$ with $\alpha \in R$ and $f_1 \in R[x]$ primitive. By hypothesis, α is a product of irreducibles in R; these irreducibles remain irreducible in $R[x]$, so it suffices to show that f_1 is a product of irreducibles. Replacing f by f_1, we can therefore assume that f is primitive.

Then $f = g_1 \cdots g_n$ where each g_i is an irreducible in $k[x]$. By the proof of Lemma 10.5, we can write $g_i = \beta_i^{-1}\gamma_i h_i$, where $\beta_i, \gamma_i \in R$ and $h_i \in R[x]$ is primitive. Since g_i is irreducible in $k[x]$, so is h_i. By Lemma 10.5, h_i is therefore irreducible in $R[x]$. Taking the content of $\beta_1 \cdots \beta_n f = \gamma_1 \cdots \gamma_n h_1 \cdots h_n$ gives $\beta_1 \cdots \beta_n = \gamma_1 \cdots \gamma_n$, whence $f = h_1 \cdots h_n$. This expresses f as a product of irreducibles in $R[x]$.

To show $R[x]$ is a UFD it suffices to show that every irreducible is prime. Suppose $p \in R[x]$ is irreducible and divides ab. Since p is irreducible in $R[x]$ it is primitive, and hence irreducible in $k[x]$. Since $k[x]$ is a UFD, p divides either a or b in $k[x]$. We may assume that p divides b. Hence $b = pd$ with $c \in k[x]$. It now suffices to show that $d \in R[x]$. Write $b = c(b)b_1$ with $b_1 \in R[x]$ primitive. By the proof of Lemma 10.5, $d = \alpha\beta^{-1}d_1$ with $\alpha, \beta \in R$ and $d_1 \in R[x]$ is primitive. Taking the content of $\beta c(b)b_1 = \alpha p d_1$ gives $\beta c(b) = \alpha$, so $d = c(b)d_1$ belongs to $R[x]$. □

Example 10.7. *The ring $R = k[t, t^{1/2}, t^{1/4}, \cdots]$ is a domain in which prime and irreducible elements are the same but it is not a UFD. It fails to be a UFD because some elements, t for example, cannot be written as a product of irreducibles. To see that every irreducible is prime, suppose that x is irreducible and that $x|yz$. There is a suitably large n such that x, y, and z, all belong to $= k[t, t^{1/2}, \cdots, t^{1/2^n}] = k[t^{1/2^n}]$ which is a polynomial ring in one variable (so a UFD); since x is still irreducible as an element of $k[t^{1/2^n}]$ it is prime in $k[t^{1/2^n}]$, so must divide either y or z; hence x is prime in R. Notice that R is not noetherian: the chain $(t) \subset (t^{1/2}) \subset (t^{1/4}) \subset \ldots$ does not stabilize.* ◊

1.11 Principal ideal domains

Recall that every ideal in $\mathbb{Z}$ is of the form (d) for some d. Similarly, every ideal in $k[x]$ is of the form (f) (Theorem 4.6).

An ideal of the form (r) in a ring R is said to be principal.

Definition 11.1. *A* principal ideal domain *is a domain in which every ideal is principal, i.e., every ideal consists of multiples of a single element.* ◊

Using the Euclidean algorithm is the standard method to show that a ring is a principal ideal domain. The argument in Theorem 4.6 is typical.

Principal ideal domains. Principal ideal domains abound. They are of great importance in both number theory and algebraic geometry. Later we will study them in detail.

Number theorists are interested in finite extension fields of $\mathbb{Q}$. By definition, such a field k is a subfield of $\mathbb{C}$ that is obtained by adjoining to $\mathbb{Q}$ the zeroes of a polynomial in $\mathbb{Q}[x]$. One then has the notion of the subring of integers in k; by that one means the elements of k that are a zero of a *monic* polynomial in $\mathbb{Z}[x]$. It is remarkable that such elements form a ring. If you don't believe this try to see why $\alpha+\beta$ and $\alpha\beta$ are zeroes of monic polynomials in $\mathbb{Z}[x]$ given that α and β are. It is an important question to decide when such a ring of integers is a principal ideal domain. Here is an easy example. Adjoining to $\mathbb{Q}$ the zeroes of x^2+1 gives the field $\mathbb{Q}(i)=\{a+bi \mid a,b\in\mathbb{Q}\}$. The ring of integers in this is $\mathbb{Z}[i]=\{a+bi \mid a,b\in\mathbb{Z}\}$. To show that $\mathbb{Z}[i]$ is a principal ideal domain, one uses a version of the Euclidean algorithm. They key point is to introduce a notion of "size" that allows us to prove an analogue of Proposition 4.3 saying that when we divide we can obtain a remainder that is smaller than the number we are dividing by.

Determining the ring of integers is more subtle than the example of $\mathbb{Z}[i]$ suggests. For example, the ring of integers in $\mathbb{Q}(\sqrt{5})$ is $\mathbb{Z}[\frac{1}{2}(1+\sqrt{5})]$. Is this a principal ideal domain?

The polynomial ring in two variables is not a principal ideal domain. The ideal $(x^n, x^{n-1}y, \ldots, xy^{n-1}, y^n)$ can not be generated by less than $n+1$ elements. Try to prove this. If it proves difficult, start with the case (x,y). You should also pay attention to the fact that the $n+1$ generators I listed are all homogeneous, so $I=\oplus_{d=n}^{\infty} I\cap k[x,y]_d$.

Proposition 11.2. *Let R be a principal ideal domain. Then*

1. *greatest common divisors exist in R;*
2. *if $d=\gcd(a,b)$, then $d=ax+by$ for some $x,y\in R$;*
3. *every irreducible in R is prime.*

Proof. (1) The ideal $aR+bR$ is principal, so is equal to dR for some $d\in R$. Clearly, $d=ax+by$ for some $x,y\in R$, so it remains to show that d is a greatest common divisor of a and b. First, since a and b belong to dR, they are both divisible by d. Second, if e divides both a and b, then $aR+bR$ is contained in eR, so d is a multiple of e. Hence d is a greatest common divisor of a and b.

(2) Let a be irreducible, and suppose that $a|bc$. To show that a is prime, we must show it divides either b or c. Suppose a does not divide b. Let $d=ax+by=\gcd(a,b)$. Since d divides a, either d is a unit or $a=du$ with u a unit. But $d|b$, so the second alternative implies that $a|b$. Hence d must be a unit. Since a divides bc, it therefore divides $acxd^{-1}+bcyd^{-1}=c(ax+by)d^{-1}=c$. Hence a is prime. □

Theorem 11.3. *Every principal ideal domain is a UFD.*

Proof. Let R be a PID and a a non-zero non-unit in R. We must show that a is a product of irreducibles in a unique way.

Uniqueness. Suppose that $a=a_1\cdots a_m=b_1\cdots b_n$ and that each a_i and b_j is irreducible. Without loss of generality we can assume that $m\le n$. If $m=1$,

then we would be done. By Proposition 11.2, a_1 divides some b_j; relabel the b_js so that $a_1|b_1$. Since a_1 and b_1 are irreducible, $b_1 = a_1 u$ for some unit u. Thus $a_2 \cdots a_m = (ub_2) \cdots b_n$. If $m = 1$, we would have $1 = (ub_2) \cdots b_n$ so n would have to be one also, and we would be finished. However, if $m > 1$ and by an induction argument we can reduce to the case $m = 1$.

Existence. Suppose to the contrary that a is not a product of irreducibles. Then a is not irreducible, so $a = a_1 b_1$ with a_1 and b_1 non-units. Since a is not a product of irreducibles, at least one of a_1 and b_1 is not a product of irreducibles. Relabelling if necessary, we can assume that a_1 is not a product of irreducibles. Thus a_1 is not irreducible, and we may write $a_1 = a_2 b_2$ with a_2 and b_2 non-units.

Continuing in this way, we obtain a sequence $a_1, a_2, \ldots$ of irreducible elements, and factroizations $a_i = a_{i+1} b_{i+1}$ into a product of non-units. This yields a chain

$$Ra \subset Ra_1 \subset Ra_2 \subset \cdots$$

of ideals. The union of an ascending chain of ideals is an ideal of R, and it is a principal ideal, say Rz, by hypothesis. Now z must belong to some Ra_i, but then $Rz \subset Ra_i \subset Ra_{i+1} \subset Rz$, so these ideals are equal. In particular, $a_{i+1} \in Ra_i$, so $a_{i+1} = a_i u$. It follows that $a_i = a_{i+1} b_{i+1} = a_i u b_{i+1}$, whence b_{i+1} is a unit. This is a contradiction.

We conclude that a must be a product of irreducibles. □

Proposition 11.4. *Let f be an element in a principal ideal domain R. The following are equivalent:*

1. *f is irreducible;*
2. *(f) is a maximal ideal;*
3. *$R/(f)$ is a field;*
4. *f is a prime.*

Proof. Lemma 4.10 shows that conditions (2) and (3) are equivalent. Theorem 11.3 and Lemma 10.2 shows that conditions (2) and (4) are equivalent.

(1) ⇒ (2). Suppose J is an ideal of R that contains (f). By hypothesis, J is principal, say $J = (g)$. Thus $f = gh$ for some $h \in R$. Since f is irreducible either g is a unit, in which case $J = R$, or h is a unit, in which case $g = fh^{-1}$ and $(g) = (f)$.

(2) ⇒ (1). Suppose that $f = gh$. Then $(f) \subset (g)$ so either $(g) = R$, in which case g is a unit, or $(g) = (f)$, in which case $g = fv$ for some $v \in R$ and $hv = 1$ so h is a unit. Thus f is irreducible. □

1.12 Integrality

Let $R \subset S$ be commutative rings. We say that $a \in S$ is integral over R if it satisfies a monic polynomial with coefficients in R; that is, if there are elements

$\lambda_0, \ldots, \lambda_{n-1}$ in R such that

$$a^n + \lambda_{n-1}a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0.$$

It is clear that every element of R is integral over R, because $a \in R$ is a zero of the monic polynomial $x - a \in R[x]$.

If $d \in \mathbb{Z}$, then all n^{th} roots of d are integral over $\mathbb{Z}$ because they are zeroes of the monic polynomials $x^n - d$.

If the only elements of S that are integral over R are the elements of R, we say that R is integrally closed in S. A domain that is integrally closed in its field of fractions is sometimes said to be integrally closed.

The integral closure of R in S is the set of all elements of S that are integral over R. We sometimes write $\bar{R}$ for this integral closure even though the notation does not indicate its dependence on S. Proposition 12.5 shows that $\bar{R}$ is a ring. To see that this is really not obvious, try to show that $\sqrt{2}+\sqrt{3}+\sqrt{5}$ is integral over $\mathbb{Z}$.

For example, $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$. To see this, suppose that $q \in \mathbb{Q}$ satisfies

$$q^n + \lambda_{n-1}q^{n-1} + \cdots + \lambda_1 q + \lambda_0 = 0$$

with all λ_i in $\mathbb{Z}$. Write $q = a/b$ with $a, b \in \mathbb{Z}$. We can assume that a and b have no common factor. Since

$$a^n + \lambda_{n-1}a^{n-1}b + \cdots + \lambda_1 ab^{n-1} + \lambda_0 b^n = 0,$$

every prime dividing b must also divide a^n and hence a. Since a and b have no common prime factor, we conclude that $b = \pm 1$, whence $q \in \mathbb{Z}$.

This argument depends only on the fact that $\mathbb{Z}$ is a UFD. Hence we have the next result.

Proposition 12.1. *A UFD is integrally closed.*

In particular, the polynomial ring $k[t]$ is integrally closed. In contrast, its subring $k[t^2, t^3]$ is not: for example, t is a zero of the monic polynomial $x^2 - t^2$ with coefficients in $k[t^2, t^3]$. The next result shows that *every* element of $k[t]$ is integral over $k[t^2, t^3]$. Since Fract $k[t^2, t^3]$ contains $t^{-1} = t^2(t^3)^{-1}$, it follows that Fract $k[t^2, t^3] = k(t)$. Hence $k[t]$ is the integral closure of $k[t^2, t^3]$.

Proposition 12.2. *Let $R \subset S$ be commutative rings, and $a \in S$. The following are equivalent:*

1. *a is integral over R;*
2. *$R[a]$ is a finitely generated R-module;*
3. *there is a subring S' of S such that $R[a] \subset S' \subset S$ and S' is a finitely generated R-module.*

Proof. (1) $\Rightarrow$ (2) If $a^n + \lambda_{n-1}a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0$, then $R[a] = R + Ra + \cdots + Ra^{n-1}$.

(2) $\Rightarrow$ (3) Take $S' = R[a]$.

(3) $\Rightarrow$ (1) Write $S' = Rs_1 + \cdots + Rs_n$ where $s_1 = 1$. For each i, $as_i \in S'$, so $as_i = \sum_j \lambda_{ij} s_j$ for some λ_{ij}s in R. Rewrite this equation as $\sum_{j=1}^{n}(a\delta_{ij} - \lambda_{ij})s_j = 0$. Let M be the $n \times n$ matrix with ij^{th} entry $a\delta_{ij} - \lambda_{ij}$, set $\Delta = \det M$ and write $\underline{s}$ for the column vector $(s_1, \ldots, s_n)^{\mathsf{T}}$. Thus $M\underline{s} = 0$, and

$$0 = (M^{\text{adj}})M\underline{s} = \Delta\underline{s}$$

where M^{adj} is the adjoint matrix. Hence $\Delta s_i = 0$ for all i; in particular, $0 = \Delta s_1 = \Delta = \det(a\delta_{ij} - \lambda_{ij})$. Writing out this determinant explicitly gives a monic polynomial of degree n in a with coefficients in R. Hence a is integral over R. □

Corollary 12.3. *Let $R \subset S$ be commutative rings and $a_1, \ldots, a_n$ elements of S. If all the a_is are integral over R, then $R[a_1, \ldots, a_n]$ is a finitely generated R-module.*

Proof. We argue by induction on n, the case $n = 1$ being given by Proposition 12.2. The induction hypothesis is that $R[a_1, \ldots, a_{n-1}]$ is a finitely generated R-module, say equal to $Rs_1 + \cdots + Rs_m$. Since a_n is integral over R, it is integral over $R[a_1, \ldots, a_{n-1}]$, so Proposition 12.2 shows that $R[a_1, \ldots, a_n]$ is a finitely generated $R[a_1, \ldots, a_{n-1}]$-module, say

$$R[a_1, \ldots, a_n] = R[a_1, \ldots, a_{n-1}]t_1 + \cdots + R[a_1, \ldots, a_{n-1}]t_k.$$

It follows that $R[a_1, \ldots, a_n] = \sum_{i=1}^{m}\sum_{j=1}^{k} Rs_i t_j$. □

Let $R \subset S$ be commutative rings. We say that S is integral over R if every element of S is integral over R.

Corollary 12.4. *Let $R \subset S$ be commutative rings. If S is a finitely generated R-algebra, then S is integral over R if and only if it is a finitely generated R-module.*

Proposition 12.5. *Let $R \subset S$ be commutative rings. The integral closure of R in S is a subring of S.*

Proof. Write $\bar{R}$ for the integral closure of R in S. Thus,

$$\bar{R} = \{a \in S \mid a \text{ is integral over } R\}.$$

Obviously, $R \subset \bar{R}$. We must show that if $a, b \in \bar{R}$, then $a \pm b$ and ab are in $\bar{R}$. By Proposition 12.2, $R[a]$ and $R[b]$ are finitely generated R-modules, say $R[a] = Ra_1 + \cdots + Ra_m$ and $R[b] = Rb_1 + \cdots + Rb_n$. We can assume that $a_1 = b_1 = 1$. This ensures that the finitely generated R-module

$$S' := \sum_{i=1}^{m}\sum_{j=1}^{n} Ra_i b_j$$

is in fact a ring. Both $R[a]$ and $R[b]$ are subrings of S', so $a \pm b$ and ab belong to S'. By Proposition 12.2, $a \pm b$ and ab are integral over R. □

The next result should be compared with the two exercises in section 1.9 which showed that suitable finiteness conditions imply that a domain must be a field.

Proposition 12.6. *Let T be a domain and R a subring of T such that T is integral over R. Then R is a field if and only if T is.*

Proof. ($\Rightarrow$) Let a be a non-zero element of T. Let n be minimal such that

$$a^n + \lambda_{n-1}a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0$$

for some elements $\lambda_{n-1}, \ldots, \lambda_0 \in R$. We must have $\lambda_0 \neq 0$ because if it were not, we could cancel of a common factor of a and so reduce the minimal n. But now λ_0 is a unit in R, so

$$a(-\lambda_{n-1}a^{n-1} + \cdots - \lambda_1)\lambda_0^{-1} = 1,$$

thus showing that a is a unit in T.

($\Leftarrow$) Let b be a non-zero element of R. Then $b^{-1} \in T$, so satisfies a monic polynomial

$$b^{-n} + \lambda_{n-1}b^{-n+1} + \cdots + \lambda_1 b^{-1} + \lambda_0 = 0$$

with coefficients in R. Multiplying through by b^{n-1} gives

$$b^{-1} = -\lambda_{n-1} - \lambda_{n-2}b - \cdots - \lambda_1 b^{n-2} - \lambda_0 b^{n-1},$$

thus showing that $b^{-1} \in R$, and hence that R is a field. □ waerden

Example 12.7. *The polynomial ring $k[t]$ is a PID, hence a UFD, and is therefore integrally closed in its field of fractions $k(t)$. The field $k(t)$ is also the field of fracions of the subring $R = k[t^2, t^3] = k + kt^2 + kt^3 + kt^4 + \cdots$ of $k[t]$. Now R is not integrally closed in $k(t)$ because t is integral over R but not an element of R. It is a zero of the polynomial $x^2 - t^2 \in R[x]$. Because $k[t] = R[t]$ it therefore follows from Proposition 12.2 that every element of $k[t] = R[t]$ is integral over R. Hence the integral closure of R in $k(t)$ is $k[t]$.* ◊

Example 12.8. *Since $\mathbb{Z}$ is a UFD it is integrally closed in $\mathbb{Q}$. However, it is not integrally closed in the larger field $\mathbb{Q}(\sqrt{2})$ because $\sqrt{2}$ is integral over $\mathbb{Z}$. It is a zero of $x^2 - 2 \in \mathbb{Z}[x]$. One can show that the integral closure of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{2}$ is equal to $\mathbb{Z}[\sqrt{2}] = \mathbb{Z} \oplus \mathbb{Z}\sqrt{2}$.*

On the basis of the example of $\mathbb{Q}(\sqrt{2})$ one might guess that the integral closure of $\mathbb{Z}in\mathbb{Q}(\sqrt{-3})$ is $\mathbb{Z}[\sqrt{-3}]$. That would be wrong because although $\mathbb{Z}[\sqrt{-3}]$ is intgral over $\mathbb{Z}$ it is not the integral closure. For example, $\alpha = \frac{1}{2}(1 + \sqrt{-3})$ is integral over $\mathbb{Z}$. It is a zero of $x^2 - x + 1 \in \mathbb{Z}$. ◊

1.13 Integers in number fields

Definition 13.1. *A subfield F of $\mathbb{C}$ is called a* number field *if* $\dim_{\mathbb{Q}} F < \infty$. *The integral closure of $\mathbb{Z}$ in F is called the* ring of integers in F. $\Diamond$

1.14 Transcendental extensions

Define transcendence degree and show well defined.

Chapter 2

Field Extensions

Throughout this chapter k denotes a field and K an extension field of k.

Let L/k be an extension. To explain what we mean when we say that a polynomial $g \in k[x]$ has a zero in L we must first be a little more precise about what we mean when we say that L is an extension of k. An extension of k is a pair (L, ϕ) consisting of a field L and a homomorphism $\phi : k \to L$. The polynomial $\phi(g) \in L[x]$ so if $\alpha \in L$ we can form the evaluation $\phi(g)(\alpha)$; if $\phi(g)(\alpha) = 0$ we say that α is a zero of g but we should really say that α *is a zero of* $\phi(g)$.

Proposition 0.1. *If $g \in k[x]$ is irreducible, then there is an extension L/k in which g has a zero.*

Proof. Define $L = k[x]/(g)$. We will show that $\alpha := x + (g) \in L$ is a zero of g. There are homomorphisms

□

If X is any set and K a field, then $K^X := \{f : X \to K\}$ is a ring under pointwise addition and multiplication. We make K^X a K-vector space by defining $(\lambda \cdot f)(x) := \lambda f(x)$ for all $f \in K^X$ and $\lambda \in K$.

Theorem 0.2. *Let $\psi_1, \ldots, \psi_n : L \to K$ be distinct ring homomorphisms between the fields L and K. Then $\psi_1, \ldots, \psi_n$ are linearly independent over K.*

Proof. We argue by induction on n. The case $n = 1$ is trivial. Suppose the theorem is true for all sets of $\leq n - 1$ distinct homomorphisms. Suppose $\{\psi_1, \ldots, \psi_n\}$ is not linearly independent over K. Then $\psi_1 = \sum_{i=2}^{n} a_i \psi_i$ for some $a_2, \ldots, a_n \in K$ *all of which are non-zero.* If b, c are in L, then

$$\psi_1(b)\psi_1(c) = \psi_1(b) \sum_{i=2}^{n} a_i \psi_i(c)$$

and

$$\psi_1(b)\psi_1(c) = \psi_1(bc) = \sum_{i=2}^{n} a_i \psi_i(bc) = \sum_{i=2}^{n} a_i \psi_i(b)\psi_i(c).$$

Subtracting, we get

$$0 = \sum_{i=2}^{n} a_i[\psi_1(b) - \psi_i(b)]\psi_i(c).$$

This holds for all $c \in L$ so $\sum_{i=2}^{n} a_i[\psi_1(b) - \psi_i(b)]\psi_i = 0$. But $\{\psi_2, \ldots, \psi_n\}$ is linearly independent over K so $a_i[\psi_1(b) - \psi_i(b)] = 0$ for all $i = 2, \ldots, n$ and all $b \in L$. Since $a_i \neq 0$, $\psi_1(b) = \psi_i(b)$ for all $b \in L$, whence $\psi_1 = \psi_i$, contradicting the fact that the ψ_i's are distinct. □

2.1 Splitting Fields

Definition 1.1. *A polynomial* splits over k *if it is a product of linear polynomials in* $k[x]$. ◊

Let $\psi : k \to K$ be a homomorphism between two fields. There is a unique extension of ψ to a ring homomorphism $k[x] \to K[x]$ that we also denote by ψ; explicitly,

$$\psi\left(\sum_{i=0}^{n} \lambda_i x^i\right) = \sum_{i=0}^{n} \psi(\lambda_i) x^i.$$

Hence it makes sense to ask if a polynomial in $k[x]$ has a zero in K. Similarly, it makes sense to ask if a polynomial in $k[x]$ splits in $K[x]$.

Definition 1.2. *Let $f \in k[x]$ be a polynomial of degree ≥ 1. An extension K/k is called a* splitting field *for f over k if f splits over K and if L is an intermediate field, say $k \subset L \subset K$, and f splits in $L[x]$, then $L = K$.* ◊

The second condition in the definition could be replaced by the requirement that $K = k(\alpha_1, \ldots, \alpha_n)$ where $\alpha_1, \ldots, \alpha_n$ are the zeroes of f in K.

The main result in this section is the existence and uniqueness up to isomorphism of splitting fields.

Remarks. 1. If $k \subset L \subset K$, and K is a splitting field for $f \in k[x]$, then K is also a splitting field for f over L. The converse is false as one sees by taking $f = x^2 + 1$ and $k = \mathbb{Q} \subset L = \mathbb{R} \subset K = \mathbb{C}$.

2. Let K be a splitting field for f over k. If F is an extension of K and α is a zero of f in F, then $\alpha \in K$. To see this, write $f = \beta(x - \alpha_1) \ldots (x - \alpha_n)$ with $\beta \in k$ and $\alpha_1, \ldots, \alpha_n \in K$, and observe that $0 = f(\alpha) = \beta(\alpha - \alpha_1) \ldots (\alpha - \alpha_n)$, so $\alpha = \alpha_i$ for some i.

3. Let K be a splitting field for f over k, and let α be a zero of f in K. Then $f = (x - \alpha)g$ for some $g \in K[x]$. Because f splits in K, so does g. Hence K is a splitting field for g over $k(\alpha)$.

Theorem 1.3. *Let k be a field and $f \in k[x]$. Then f has a splitting field, say K/k, and $[K : k] \leq (\deg f)!$.*

Proof. Induction on $n = \deg f$. If $\deg f = 1$, then $f = \alpha x + \beta$ with $\alpha, \beta \in k$, so f already splits in k, so we can take $K = k$.

Suppose that $n > 1$. If f is already split we may take $K = k$, so we may assume that f has an ireducible factor, say g, of degree ≥ 2. By Proposition **??**.6.4, g has a zero in the extension field $k(\alpha) = k[x]/(g)$; the degree of this extension is $\deg g \leq n$. Now write $f = (x - \alpha)h$ where $h \in k(\alpha)[x]$. Since $\deg h = n - 1$, the induction hypothesis says there is an extension $L/k(\alpha)$ over which h splits, and $[L : k(\alpha)] \leq (n-1)!$. Certainly f also splits over L, and $[L : k] = [L : k(\alpha)][k(\alpha) : k] \leq n!$. If $\alpha_1, \ldots, \alpha_n$ are the zeroes of f in L, then $k(\alpha_1, \ldots, \alpha_n)$ is a splitting field for f over k. □

The proof of Theorem 1.3 involves a choice of an irreducible factor of f. It is conceivable that choosing a different factor might produce a different splitting field. Before showing that is not the case, and hence that a splitting field is unique up to isomorphism, we need the following lemma.

Lemma 1.4. *Let $\varphi : k \to k'$ be an isomorphism of fields. Let $f \in k[x]$ be irreducible. If α is zero of f in some extension of k and β is an extension of $\varphi(f)$ in some extension of k', then there is an isomorphism $\psi : k(\alpha) \to k'(\beta)$ such that $\psi|_k = \mathrm{id}_k$ and $\psi(\alpha) = \beta$.*

Proof. The following picture describes the situation we are considering:

$$\begin{array}{ccc} k(\alpha) & \overset{\Phi}{- - \to} & k(\beta) \\ \big| & & \big| \\ k & \xrightarrow[\varphi]{} & k' \end{array}$$

The map φ extends to an isomorphism $k[x] \to k'[x]$ and sends (f) to $(\varphi(f))$, so induces an isomorphism between the quotient rings by these ideals. The composition of the obvious isomorphisms

$$k(\alpha) \to k[x]/(f) \to k'[x]/(\varphi(f)) \to k'(\beta)$$

is the desired isomorphism. □

It helps to draw a picture when considering results like that in Lemma 1.4. The basic picture looks like

$$\begin{array}{ccc} K & \overset{\Phi}{- - \to} & K' \\ \big| & & \big| \\ k & \xrightarrow[\varphi]{} & k' \end{array}$$

where K/k and K'/k' are extensions, $\varphi : k \to k'$ is a given homomorphism, and Φ is a possible extension of φ. In Lemma 1.4, $K = k(\alpha)$ and $K' = k'(\beta)$.

Theorem 1.5. *Let k be a field and $f \in k[x]$. Let $\varphi : k \to k'$ be an isomorphism of fields. Let K/k be a splitting field for f, and let K'/k' be an extension such that $\varphi(f)$ splits in K'. Then*

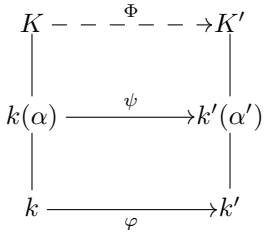
1. *there is a homomorphism* $\Phi : K \to K'$ *such that* $\Phi|_k = \varphi$*;*

2. *if* K' *is a splitting field for* $\varphi(f)$ *over* k'*, then* $K' \cong K$*.*

Proof. We argue by induction on $[K : k]$. If $[K : k] = 1$, then $K = k$, and we can take $\Phi = \varphi$. Our induction hypothesis is that the theorem holds for all field extensions of degree $\leq n - 1$.

(1) Suppose that $[K : k] > 1$. Write $f = p_1 \dots p_n$ as a product of irreducibles $p_i \in k[x]$. Then $\varphi(f) = \varphi(p_1) \dots \varphi(p_n)$, and each $\varphi(p_i)$ is irreducible in $k'[x]$.

Some p_i, say p_1, is not linear. Let $\alpha \in K$ be a zero of p_1, and let $\beta \in K'$ be a zero of $\varphi(p_1)$. By Lemma 1.4, there is an isomorphism $\psi : k(\alpha) \to k'(\beta)$ such that $\psi|_k = \varphi$. Now K is a splitting field for f over $k(\alpha)$ and $\varphi(f)$ splits over $k'(\beta)$ so we can consider the following diagram:



Since $[K : k(\alpha)] < [K : k]$ the induction hypothesis applies to the top half of the diagram giving a homomorphism $\Phi : K \to K'$ such that $\Phi|_{k(\alpha)} = \psi$. In particular, $\Phi|_k = \psi|_k = \varphi$.

(2) Certainly Φ is injective, so it remains to show it is surjective. However, if $f = (x - \alpha_1) \dots (x - \alpha_n)$ then $\varphi(f) = (x - \varphi(\alpha_1)) \dots (x - \varphi(\alpha_n))$; since K and K' are splitting fields $K = k(\alpha_1, \dots, \alpha_n)$ and $K' = k'(\varphi(\alpha_1), \dots, \varphi(\alpha_n))$, Φ is also surjective, and hence an isomorphism. □

Theorem 1.6. *A polynomial of positive degree has a unique splitting field up to isomorphism.*

2.2 Normal extensions

Definition 2.1. *A finite extension* K/k *is* normal *if every irreducible polynomial in* $k[x]$ *that has a zero in* K *actually splits over* K*.* ◊

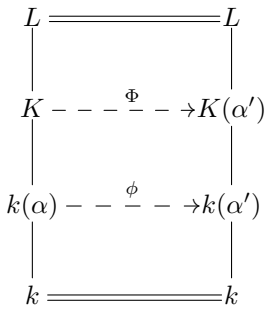
Theorem 2.2. *An extension* K/k *is normal if and only if* K *is a splitting field for some polynomial in* $k[x]$*.*

Proof. (⇒) Let K/k be a finite extension. Write $K = k(\alpha_1, \dots, \alpha_n)$. The minimal polynomial p_i of α_i has a zero in K so splits in K. Hence $f = p_1 \cdots p_n$ splits in K. But K is generated by the zeroes of f, so K is the splitting field of f over k.

(⇐) Suppose $K = k(\alpha_1, \dots, \alpha_n)$ is the splitting field of a degree n polynomial $g \in k$ where $\alpha_1, \dots, \alpha_n$ are the zeroes of g. Let $f \in k[x]$ be irreducible and suppose that $\alpha \in K$ is a zero of f. We must show that f splits in K.

Think of $fg \in K[x]$, and let L be the splitting field for fg over K. Suppose $\alpha' \in L$ is a zero of f. We want to show that α' is in K.

The following picture will help:



Since α and α' are zeroes of f, there is an isomorphism $\psi : k(\alpha) \to k(\beta')$ such that $\psi(\alpha) = \alpha'$ and $\psi|_k = \mathrm{id}_k$ (Lemma 1.4).

We will apply Theorem 1.5. Since $g \in k(\alpha)[x]$, $\phi(g) \in k(\alpha')[x]$. Since K is a splitting field for g over $k(\alpha)$ and $\phi(g)$ splits in $K(\alpha')$ there is a map $\Phi : K \to K(\alpha')$ such that $\Phi|_{k(\alpha)} = \phi$. Hence $\Phi|_k = \phi|_k = \mathrm{id}_k$. Therefore $\Phi(g) = g$ which implies that $0 = \Phi(g(\alpha_i)) = g(\Phi(\alpha_i))$; but K is a splitting field for g over k so $\Phi(\alpha_i) \in K$. Hence $\Phi(K) \subset K$. In particular, $\alpha' = \Phi(\alpha) \in K$; thus, *all* the zeroes of f belong to K. □

The next result shows that a finite extension K/k can be embedded in a unique smallest normal extension L/k. The extension L/k is called the normal closure of K/k.

Theorem 2.3. *Let K/k be a finite extension. Then there is a finite extension L/K such that*

1. *L is normal over k, and*
2. *if $K \subset F \subset L$ and F is normal over k, then $F = L$, and*
3. *if L'/K is a finite extension that satisfies (1) and (2), then there is a K-isomorphism $\Phi : L \to L'$.*

Proof. (1) Write $K = k(\alpha_1, \ldots, \alpha_n)$, let $p_i \in k[x]$ be the minimal polynomial of α_i, set $p = p_1 p_2 \cdots p_n$, and let L be the splitting field of p over K. Then $k \subset K \subset L$, L is the splitting field for p over k, and $[L : k] < \infty$. By Theorem 2.2, L is normal over k.

(2) If $K \subset F \subset L$ and F is normal over k, then each p_i splits in F because it is irreducible and has a zero in F, whence p splits in F, so $F = L$.

(3) If L'/K is a finite extension that satisfies (1) and (2), then each p_i splits in L', so by Theorem 1.5 there is a map $\Phi : L \to L'$ such that $\Phi|_K = \mathrm{id}_K$. By (2) applied to the extensions $K \subset \Phi(L) \subset L'$, it follows that $\Phi(L) = L'$, whence Φ is an isomorphism as claimed. □

Example 2.4. *Let p be an odd prime. What is the splitting field of $x^p - 2$ over $\mathbb{Q}$ and what is its degree?*

By Eisenstein's criterion $f = x^p - 2$ is irreducible. Let α be the real p^{th} root of 2. Then f is the minimal polynomial of α over $\mathbb{Q}$, so $[\mathbb{Q}(\alpha) : \mathbb{Q}] = p$.

If $\beta \in \mathbb{C}$ is a zero of f, then $(\beta\alpha^{-1})^p = 1$, so $\beta\alpha^{-1}$ is a zero of $x^p - 1 = (x-1)(x^{p-1} + \cdots + x + 1)$. The polynomial $x^{p-1} + \cdots + x + 1$ is irreducible. Let $\zeta = e^{2\pi i/p}$. The zeroes of f are $\alpha, \zeta\alpha, \zeta^2\alpha, \ldots, \zeta^{p-1}\alpha$, so the splitting field of f over $\mathbb{Q}$ is $L := \mathbb{Q}(\alpha, \zeta)$.

Since L contains both $\mathbb{Q}(\alpha)$ and $\mathbb{Q}(\zeta)$ its degree over $\mathbb{Q}$ is divisible by both p and $p-1$. Hence $p(p-1)$ divides $[L : \mathbb{Q}]$. Let g be the minimal polynomial of α over $\mathbb{Q}(\zeta)$. Then g divides $x^p - 2$ so has degree $\leq p$. Therefore

$$[L : \mathbb{Q}] = [\mathbb{Q}(\zeta)(\alpha) : \mathbb{Q}(\zeta)][\mathbb{Q}(\zeta) : \mathbb{Q}] = (p-1)\deg g \leq (p-1)p,$$

whence $[L : \mathbb{Q}] = p(p-1)$. ◊

2.3 Finite fields

We already saw parts of the next result in §1.6.

Theorem 3.1. *Let K be a finite field. Then*

1. *$|K| = p^n$ for some positive integer n, where $p = \operatorname{char} K$;*
2. *K is the splitting field for $x^{p^n} - x$ over $\mathbb{F}_p$;*
3. *any field of order p^n is isomorphic to K.*

Proof. Since $\operatorname{char} K = p$, $\mathbb{F}_p$ is a subfield of K. Hence K is a finite dimensional vector space over $\mathbb{F}_p$, and so has p^n elements where $n = \dim_{\mathbb{F}_p} K$.

It follows that $K^* := K\backslash\{0\}$ is an abelian group of order $p^n - 1$. Hence $\lambda^{p^n-1} = 1$ for every non-zero $\lambda \in K$. It follows that every element of K is a zero of $x^{p^n} - x$. In other words, $x^{p^n} - x$ has p^n distinct zeroes in K. Hence K is the splitting field for $x^{p^n} - x$ over $\mathbb{F}_p$. It now follows from Theorem 1.5 that any field of order p^n must be isomorphic to K. □

Proposition 3.2. *The multiplicative group of non-zero elements in a finite field is cyclic.*

Proof. Suppose that $|K| = p^n$. Write $e = p^n - 1 = q_1^{n_1} \cdots q_t^{n_t}$ as a product of powers of distinct primes. We will show there is an element in $K\backslash\{0\}$ of order e. Define

$$e_i = eq_i^{-1} \qquad \text{and} \qquad d_i = eq_i^{-n_i}.$$

Since $e_i < e$, there is some $\alpha_i \in K$ that is not a zero of $x^{e_i} - 1$. Define $\beta_i = (\alpha_i)^{d_i}$ and $\beta = \beta_1 \cdots \beta_t$. The order of β_i divides $q_i^{n_i}$, but if it were smaller then α_i would be a zero of $x^{e_i} - 1$; hence the order of β_i is $q_i^{n_i}$. It follows that the order of β is e. □

Theorem 3.1 does *not* show that a field of order p^n exists. It just shows what it has to be if it exists.

We will write $\mathbb{F}_{p^n}$ for the field of p^n elements (if it exists!). To prove its existence we will show that $x^{p^n} - x$ has p^n distinct zeroes, and that the set of these zeroes is equal to the splitting field of $x^{p^n} - x$ over $\mathbb{F}_p$.

To see whether a polynomial has repeated zeroes we look at its derivative.

Definition 3.3. *The* formal derivative *of a polynomial* $f = a_0 + a_1x + \cdots + a_nx^n$ *in* $k[x]$ *is*

$$f' = D(f) := a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$

$\Diamond$

Notice that $D(fg) = f'g + fg'$ and $D(\lambda f) = \lambda f'$ if $\lambda \in k$.

Lemma 3.4. *Let* f *be a non-zero polynomial in* $k[x]$*. Then* f *has a multiple zero in some extension field of* k *if and only if* $\gcd\{f, f'\} \neq 1$*.*

Proof. ($\Rightarrow$) Let K be an extension of k, and suppose that $\alpha \in K$ is a multiple zero of f. Then $f = (x - \alpha)^2 g$ for some $g \in K[x]$. Hence $x - \alpha$ divides both f and f' in $K[x]$. As remarked on page 15, $\gcd\{f, f'\}$ is the same in $K[x]$ as in $k[x]$, so $\gcd\{f, f'\} \neq 1$.

($\Leftarrow$) Let K/k be a splitting field for ff'. Suppose $\gcd\{f, f'\} \neq 1$. Then f and f' have a common factor of the form $x - \alpha$ in $K[x]$. Write $f = (x - \alpha)g$. Then $f' = (x - \alpha)g' + g$ so $x - \alpha$ divides g. Hence $(x - \alpha)^2$ divides f. □

Proposition 3.5. *The polynomial* $x^{p^n} - x \in \mathbb{F}_p[x]$ *has* p^n *distinct zeroes in its splitting field.*

Proof. Since the derivative of $x^{p^n} - x$ is 1, the result follows from the previous lemma. □

Lemma 3.6. *Let* p *be a prime and* R *a commutative ring in which* $p = 0$*. Then the map* $\phi : R \to R$ *defined by* $\phi(a) = a^p$ *is a ring homomorphism.*

Proof. Certainly $\phi(1) = 1$. It is clear that $\phi(ab) = \phi(a)\phi(b)$, and

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

The integers $\binom{p}{i}$ are divisible by p whenever $1 \leq i \leq p - 1$, so are zero in R. Hence $\phi(a + b) = \phi(a) + \phi(b)$. □

We call the map ϕ in Lemma 3.6 the Frobenius map.

If K is a field of characteristic p, then ϕ is injective.

Corollary 3.7. *If* K *is a finite field of characteristic* p*, then every element of* K *is a* p^{th} *power.*

Proof. The Frobenius map is injective, hence surjective. □

Theorem 3.8. *For each prime p and positive integer n, there is a unique field with p^n elements, namely the splitting field of $x^{p^n} - x$.*

Proof. Let K be the splitting field of $x^{p^n} - x$ over $\mathbb{F}_p$. Let $\phi : K \to K$ be the Frobenius map. Notice that $\alpha \in K$ is a zero of $x^{p^n} - x$ if and only if $\phi^n(\alpha) = \alpha$. Hence, if α and β are zeroes of $x^{p^n} - x$, so are $\alpha \pm \beta$, $\alpha\beta$ and α^{-1}. Hence the zeroes of $x^{p^n} - x$ are a subfield of K. Since K is generated over $\mathbb{F}_p$ by the zeroes of $x^{p^n} - x$, we conclude that K is exactly the set of zeroes of $x^{p^n} - x$. □

We write $\mathbb{F}_{p^n}$ for "the" field with p^n elements.

Proposition 3.9. *$\mathbb{F}_{p^n}$ has a subfield isomorphic to $\mathbb{F}_{p^m}$ if and only if m divides n.*

Proof. ($\Rightarrow$) If $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$, then $\mathbb{F}_{p^n}$ is a vector space over $\mathbb{F}_{p^m}$ so isomorphic to $\left(\mathbb{F}_{p^m}\right)^d$ for some integer d. But the number of elements in K^d is $|K|^d$ so $p^n = |\mathbb{F}_{p^n}| = |\mathbb{F}_{p^m}|^d = p^{md}$. Thus $n = md$.

($\Leftarrow$) Suppose that $n = md$. There is an element $\alpha \in \mathbb{F}_{p^n}^\times$ whose order is $p^n - 1$. Since $p^n - 1 = (p^m)^d - 1$, $p^m - 1$ divides $p^n - 1$. Let $s \in \mathbb{N}$ be such that $p^n - 1 = (p^m - 1)s$. Now $(\alpha^s)^{p^m - 1} = \alpha^{s(p^m - 1)} = \alpha^{p^n - 1} = 1$ so $(\alpha^s)^{p^m} = \alpha^{sp^m} = \alpha^s$. Hence α^s is a root of the polynomial $x^{p^m} - x$.

MORE TO SAY □

2.4 Separability

Let's begin with a warning: an irreducible polynomial can have multiple zeroes in its splitting field. For example, let $k = \mathbb{F}_p(t)$ be the rational function field over $\mathbb{F}_p$, and let $f = x^p - t \in k[x]$. By Eisenstein's criterion applied to $f \in k[t][x]$, f is irreducible but over the extension field $K = k(t^{1/p})$ we have $f = (x - t^{1/p})^p$. This behavior causes problems.

Definition 4.1. *A polynomial $f \in k[x]$ is* separable *if none of its irreducible factors has a multiple zero in its splitting field.*

If every $f \in k[x]$ is separable, we say that k is a perfect *field.*

Let K be an extension of k An element $\alpha \in K$ is separable *over k if its minimal polynomial is separable. We say that K is a* separable extension *of k if every element in it is separable over k.* ◊

We have just seen that $\mathbb{F}_p(t)$ is not perfect: $t^{1/p}$ is not separable over $\mathbb{F}_p(t)$, and $\mathbb{F}_p(t^{1/p})$ is not a separable extension of $\mathbb{F}_p(t)$.

Lemma 4.2. *If $k \subset F \subset K$ are fields and K/k is separable, so are K/F and F/k.*

Proof. It follows at once form the definition that F/k is separable if K/k is. On the other hand, if $\alpha \in K$ its minimal polynomial over F divides its minimal polynomial over k, so has distinct zeroes. Hence α is separable over F. □

It is quite a bit harder to prove the converse of this lemma.

Proposition 4.3. *An irreducible polynomial f is separable if and only if $f' \neq 0$.*

Proof. Since f is irreducible, $\gcd\{f, f'\}$ is either 1 or f. By Lemma 3.4, f is separable if and only if $\gcd\{f, f'\} = 1$. Thus f is separable if and only if $\gcd\{f, f'\} \neq f$. However, $\gcd\{f, f'\} = f$ if and only if $f' = 0$. □

Proposition 4.4. *Fields of characteristic zero are perfect.*

Proof. Let f be an irreducible polynomial with coefficients in a field of characteristic zero. Since the characteristic is zero, the derivative f' is not zero. Since f is irreducible and $\deg f' < \deg f$, it follows that $\gcd(f, f') = 1$, whence f has no multiple zeroes by Lemma 3.4. □

Theorem 4.5. *A field of characteristic $p > 0$ is perfect if and only if every element of it is a p^{th} power (if and only if the Frobenius map is surjective).*

Proof. Let k be the field in question and write $k^p = \{\alpha^p \mid \alpha \in k\}$.

($\Rightarrow$) Let $\alpha \in k$ and set $f = x^p - \alpha$. Since $f' = 0$, $\gcd(f, f') = f$ so Lemma 3.4 implies that f has a multiple zero. But k is perfect so irreducible polynomials in $k[x]$ do not have multiple zeroes. Hence f is reducible. Write $f = gh$ with $1 \leq \deg g \leq p - 1$. Let K be the splitting field for f over k. If $\lambda \in K$ is a zero of f, then $\lambda^p = \alpha$ so

$$f = x^p - \alpha = x^p - \lambda^p = (x - \lambda)^p = gh.$$

Hence $g = (x - \lambda)^d$ with $1 \leq d \leq p - 1$. But the coefficients of g belong to k, so $\lambda^d \in k$. Also $\lambda^p = \alpha \in k$. Since $(p, d) = 1$, there are integers u and v such that $du + pv = 1$. Hence

$$\lambda = (\lambda^d)^u (\lambda^p)^v \in k.$$

In particular, $\alpha = \lambda^p$ is the p^{th}-power of an element in k.

($\Leftarrow$) Suppose to the contrary that k is not perfect. Let $f = \sum \alpha_j x^j$ be an irreducible polynomial in $k[x]$ having a repeated zero in some extension field. Then $\deg(\gcd(f, f')) \geq 1$ but f is irreducible so $\gcd(f, f') = f$. Hence $f' = 0$. It follows that $\alpha_j = 0$ if p does not divide j. Hence $f = \beta_0 + \beta_1 x^p + \beta_2 x^{2p} + \cdots$. By hypothesis there are elements $\gamma_i \in k$ such that $\gamma_i^p = \beta_i$ so $f = (\gamma_0 + \gamma_1 x + \gamma_2 x^2 + \cdots)^p$. This contradicts the irreducibility of f. We conclude that k must be perfect. □

Corollary 4.6. *A finite field is perfect.*

Every algebraic extension of a finite field, and every extension of a characteristic zero field is a separable extension.

Definition 4.7. *If $K = k(\alpha)$ we say that K is a* simple extension *of k and that α is a primitive element of K over k.* ◊

For example, if $(n, d) = 1$ then $\alpha = e^{2\pi id/n}$ is a primitive element for the extension of $\mathbb{Q}$ obtained by adjoining all n^{th} roots of one.

Theorem 4.8 (The primitive element theorem)**.** *A finite separable extension is simple. In particular, if k is finite or of characteristic zero every finite extension of k is of the form $k(\alpha)$.*

Proof. Let K be a finite separable extension of k.

If k is finite so is K whence $K - \{0\}$ is a cyclic group; . if α is a generator of this group, then $K = k(\alpha)$.

For the remainder of the proof we assume k is infinite.

By induction it suffices to show that a finite separable extension of the form $k(\alpha, \beta)$ is equal to $k(\gamma)$ for a suitable γ. If one of $k(\alpha)$ and $k(\beta)$ is contained in the other the result is obvious so we assume that is not the case. This assumption implies that neither α nor β is in k.

Let f and g be the minimal polynomials of α and β over k. Write $m = \deg f$ and $n = \deg g$, and let $\{\alpha = \alpha_1, \ldots, \alpha_m\}$ and $\{\beta = \beta_1, \ldots, \beta_n\}$ be the zeroes of f and g respectively. Since the extension is separable, the α_is are distinct and the β_js are distinct. Since neither α nor β is in k both m and n are ≥ 2.

Consider the $mn - 1$ equations

$$(\alpha - \alpha_i) + (\beta - \beta_j)X = 0, \qquad 1 \leq i \leq m, 1 \leq j \leq n,\ (i, j) \neq (1, 1).$$

Since k is infinite there is $\lambda \in k$ that is not a solution to any of these equations.

Define $\gamma := \alpha + \beta\lambda \in k(\alpha, \beta)$.

The polynomials $g(x)$ and $f(\gamma - \lambda x)$ belong to $k(\gamma)[x]$ and β is a zero of each of them. If $g(x)$ and $f(\gamma - \lambda x)$ have a common zero $\xi \neq \beta$ in some extension of K, then $\xi = \beta_j$ for some $j > 1$ and $\gamma - \lambda\xi = \alpha_i$ for some i, whence $\alpha + \beta\lambda - \lambda\beta_j = \alpha_i$ which contradicts the choice of λ. Since $g(\beta) = 0$ and $(x - \beta)^2$ does not divide $g(x)$, $\gcd(g(x), f(\gamma - \lambda x)) = x - \beta$. But the gcd of two polynomials in $k(\gamma)[x]$ belongs to $k(\gamma)[x]$ (see the remark on page 15) so $\beta \in k(\gamma)$. Hence $\alpha = \gamma - \lambda\beta$ is also in $k(\gamma)$, and we conclude that $k(\alpha, \beta) = k(\gamma)$. □

Corollary 4.9. *If K/k is a finite, normal, separable extension, then K is the splitting field of an irreducible separable polynomial over k.*

Proof. By the Primitive Element Theorem, $K = k(\alpha)$. Let f be the minimal polynomial of α. Then f is separable and irreducible of degree $[K : k]$. Since f has one zero in K and K/k is normal, f splits in K. Hence K is the splitting field for f. □

2.5 Automorphisms of separable extensions

The notion of separabilty was defined in terms of individual elements. However, it eventually proves more useful to be able to characterize whether or not an

extension K/k is separable in terms of automorphisms of K. We shall see that non-separable extensions are more rigid than separable ones in the sense that they possess far fewer automorphisms.

The next example illustrates the matter.

Example 5.1. *The extension $K/k = \mathbb{F}_p(t^{1/p})/\mathbb{F}_p(t)$ is not separable. If ϕ is an automorphism of this extension then $\phi(t^{1/p})$ has the same minimal polynomial as $t^{1/p}$, namely $x^p - t = (x - t^{1/p})^p$, so $\phi(t^{1/p}) = t^{1/p}$, whence $\phi = \mathrm{id}_K$. Hence* $\mathrm{Aut}(K/k) = \{1\}$. ◊

Compare this to a separable extension like $\mathbb{Q}(\sqrt{2})$ where there is a $\mathbb{Q}$-automorphism of $\mathbb{Q}(\sqrt{2})$ sending $\sqrt{2} \to -\sqrt{2}$. Of course, $\mathbb{Q}(\sqrt[3]{\ })$ does not have any automorphisms but for a rather different reason: although the minimal polynomial of $\sqrt[3]{2}$ has three distinct zeroes, only one of them is in $\mathbb{Q}(\sqrt[3]{2})$. Hence to really understand the difference between separable and non-separable extensions from the perspective of automorphisms we should focus on separable extensions that are normal.

Proposition 5.2. *Let $k(\alpha)$ be a degree d extension of k and let $f \in k[x]$ be the minimal polynomial of α. Let $\phi : k \to F$ be a homomorphism, and suppose that $\phi(f)$ splits over F.*

1. *If α is separable over k there are exactly d distinct extensions of ϕ to maps $\phi_i : k(\alpha) \to F$ such that $\phi_i|_k = \phi$.*

2. *If α is not separable there are $< d$ such extensions.*

Proof. Let $\beta_1, \ldots, \beta_n$ be the distinct zeroes of f in F. Then $n \leq d = \deg f$ and $n = d$ if and only if α is separable over k.

Any homomorphism $\psi : k(\alpha) \to F$ that extends ϕ is completely determined by $\psi(\alpha)$. But $\psi(\alpha) = \beta_i$ for some i, so there are at most n different ψs. By Lemma 1.4, there *is* such an extension for all i. Hence there are exactly n extensions of ϕ, and the result follows. □

Theorem 5.3. *Let K/k be a degree d extension and let $\phi : k \to F$ be a homomorphism. Suppose the minimal polynomials of all the elements in K split in F.*

1. *If K/k is separable there are exactly d distinct extensions of ϕ to maps $\phi_i : K \to F$ such that $\phi_i|_k = \phi$.*

2. *If K/k is not separable there are $< d$ such extensions.*

Proof. (1) By the Primitive Element theorem $K = k(\alpha)$, so the result follows from part (1) of Proposition 5.2.

(2) We argue by induction on d. Since K/k is not separable, $d > 1$. Let $\alpha \in K$ be non-separable over k. Set $s = [k(\alpha) : k]$ and $t = [K : k(\alpha)]$, so $st = d$ and $t < d$.

Every extension of ϕ to K can be obtained in two steps: first extend ϕ to $k(\alpha)$, then extend the extension from $k(\alpha)$ to K. By part (2) of Proposition 5.2, there are $< s$ extensions of ϕ to $k(\alpha)$, and by the induction hypothesis and (1), each of those extensions has $\leq t$ extensions to K, giving a total of $< d$ extensions of ϕ to K. □

Theorem 5.4. *The following conditions on a finite extension K/k are equivalent:*

1. *K/k is a normal and separable;*
2. *$|\operatorname{Aut}(K/k)| = [K : k]$;*
3. *$K^{\operatorname{Aut}(K/k)} = k$.*

If K/k is normal but not separable, then $|\operatorname{Aut}(K/k)| < [K : k]$.

Proof. The final sentence in the statement of this theorem follows from Theorem 5.3(2).

(1) ⇒ (2) If we set $F = K$, then the hypotheses of Theorem 5.3(1) hold and the conclusion of that result gives (2).

(2) ⇔ (3) Artin's Theorem says that

$$[K : K^{\operatorname{Aut}(K/k)}] = |\operatorname{Aut}(K/k)|,$$

so (2) and (3) are equivalent.

(2) ⇒ (1) Write $d = [K : k]$ and let $\phi_1, \ldots, \phi_d$ be the distinct elements in $\operatorname{Aut}(K/k)$. If F denotes the normal closure of K, then the hypotheses of Theorem 5.3 hold, and each ψ_i may be considered as a map $\phi_i : K \to F$ that extends the inclusion $\phi : k \to F$. The conclusion of Theorem 5.3 then shows that K/k is separable.

By the Primitive Element Theorem, $K = k(\alpha)$ for some $\alpha \in K$. Let $p \in k[x]$ be the minimal polynomial of α. Since the ϕ_is are distinct, the elements $\phi_1(\alpha), \ldots, \phi_d(\alpha)$ are distinct. Hence p has $d = \deg p$ distinct zeroes in K, whence K is a splitting field for p. Hence K/k is normal. □

Theorem 5.5. *An extension generated by separable elements is separable.*

Proof. Let $K = k(\alpha_1, \ldots, \alpha_n)$ and suppose that each α_i is separable over k. We argue by induction on n. If $n = 0$, there is nothing to do, so suppose that $k' = k(\alpha_1, \ldots, \alpha_{n-1})$ is separable over k, set $\alpha = \alpha_n$ so that $K = k'(\alpha)$ and consider the extensions k'/k and K/k'.

Let F be the normal closure of k' over k (see page 45). Let $\phi : k \to F$ be the inclusion; the hypotheses of Theorem 5.3 now hold, so there are exactly $[k' : k]$ distinct homomorphisms $\phi_i : k' \to F$ such that $\phi_i|_k = \mathrm{id}_k$.

By hypothesis α is separable over k and hence over k'. By Proposition 5.2 applied to $k'(\alpha)$, there are exactly $[k'(\alpha) : k']$ extensions of each ϕ_i to homomorphisms $K = k'(\alpha) \to F$. This gives a total of $[k' : k].[k'(\alpha) : k'] = [K : k]$ homomorphisms $\psi : K \to F$ such that $\psi|_k = \mathrm{id}_k$. By part (2) of Theorem 5.3, K/k is separable. □

Corollary 5.6. *The splitting field of a separable polynomial is a separable extension.*

Proof. Let $f \in k[x]$ be a separable polynomial. Its splitting field is generated by the zeroes of the irreducible factors of f, and by hypothesis these zeroes are separable. Theorem 5.5 implies that this splitting field is separable over k. $\square$

Chapter 3

Galois Theory

Let K be a finite separable normal extension of k, i.e., a Galois extension. The fundamental idea of Galois theory is the interplay between subfields of K/k and subgroups of $\mathrm{Gal}(K/k)$.

The fundamental theorem gives a containment-reversing bijection between subgroups of $\mathrm{Gal}(K/k)$ and subfields of K/k that contain k. Under the bijection normal subgroups correspond to normal extensions of k.

The bijection can be stated in a more precise way if we use the language of lattices.

Let X be a set. A finite lattice of subsets of X is a finite collection $\mathcal{L}$ of subsets of X that is closed under intersections and unions. A union of two subgroups is not a group so this notion isn't quite right for Galois theory. However, if H and H' are subgroups of a group G there is a unique smallest subgroup of G that contains $H \cup H'$. With this slight modification and a similar one for intermediate subfields we can make the precise statement: the lattice of subgroups of $\mathrm{Gal}(K/k)$ is anti-isomorphic to the lattice of intermediate extensions of K/k.

3.1 The Galois correspondence

If K is an extension of k we write

$$\mathrm{Gal}(K/k) = \mathrm{Aut}(K/k),$$

and call this the Galois group of the extension.

Definition 1.1. *A finite, normal, separable extension K/k is called a* Galois extension.

Lemma 1.2. *Let K/k be a Galois extension and F an intermediate field. Then K/F is a Galois extension and* $\mathrm{Gal}(K/F)$ *is the subgroup* $\{\psi \in \mathrm{Gal}(K/k) \mid \psi|_F = \mathrm{id}_F\}$ *of* $\mathrm{Gal}(K/k)$.

Proof. Because K/k is normal it is a splitting field for some $f \in k[x] \subset F[x]$. Thus K/F is a splitting field for f over F, and hence normal. Because K/k is separable so is K/F (Lemma 2.4.2).

Every F-linear automorphism of K is k-linear it is obvious that $\mathrm{Gal}(K/F)$ is the subset of $\mathrm{Gal}(K/k)$ described in the statement of the lemma. □

Theorem 1.3 (The Galois correspondence)**.** *Suppose K/k is a Galois extension. There is an order-reversing bijection*

$$\{\textit{intermediate fields } F,\ k \subset F \subset K\} \longleftrightarrow \{\textit{subgroups of } \mathrm{Gal}(K/k)\}$$

implemented by

$$F \mapsto \mathrm{Gal}(K/F)$$

and

$$K^H \leftarrow H.$$

Furthermore,

1. *F/k is normal if and only if $\mathrm{Gal}(K/F)$ is a normal subgroup of $\mathrm{Gal}(K/k)$, and*

2. *if F/k is normal, then $\mathrm{Gal}(F/k) \cong \mathrm{Gal}(K/k)/\mathrm{Gal}(K/F)$.*

Proof. We must show that the compositions $F \mapsto \mathrm{Gal}(K/F) \mapsto K^{\mathrm{Gal}(K/F)}$ and $H \mapsto K^H \mapsto \mathrm{Gal}(K/K^H)$ are the identities.

Since K/k is a separable normal extension so is K/F. Hence by Theorem 2.5.4, $K^{\mathrm{Gal}(K/F)} = F$.

Let H be a subgroup of $\mathrm{Gal}(K/k)$. By Artin's Theorem, $[K : K^H] = |H|$. But K/K^H is a Galois extension so $|\mathrm{Gal}(K/K^H)| = [K : K^H]$, by Theorem 2.5.4. Hence $|H| = |\mathrm{Gal}(K/K^H)|$. But $H \subset \mathrm{Gal}(K/K^H)$ so $H = \mathrm{Gal}(K/K^H)$.

"Order-reversing" means that if $H \subset H'$, then $K^H \supset K^{H'}$. Actually, the lattices involved in the bijection are anti-isomorphic. This is obvious.

(2) Suppose F/k is normal.

We will show that there is a well-defined surjective group homomorphism $\Psi : \mathrm{Gal}(K/k) \to \mathrm{Gal}(F/k)$, $\Psi(\sigma) := \sigma|_F$, whose kernel is $\mathrm{Gal}(K/F)$. Once that is done we will have an isomorphism

$$\frac{\mathrm{Gal}(K/k)}{\mathrm{Gal}(K/F)} \cong \mathrm{Gal}(F/k).$$

This will prove (2) and also prove that $\mathrm{Gal}(K/F)$ is a normal subgroup of $\mathrm{Gal}(K/k)$.

To prove Ψ is well-defined we must show that $\sigma(\alpha) \in F$ for all $\sigma \in \mathrm{Gal}(K/k)$ and $\alpha \in F$. Let $f \in k[x]$ be the minimal polynomial of α. Then f splits in F; but $\sigma(\alpha)$ is a zero of f so $\sigma(\alpha) \in F$. Thus $\sigma(F) \subset F$.

The kernel of Ψ is $\{\psi \in \mathrm{Gal}(K/k) \mid \psi|_F = \mathrm{id}_F\}$ which is equal to $\mathrm{Gal}(K/F)$.

Finally, Ψ is surjective because if $\theta \in \mathrm{Gal}(F/k)$, then $\theta : F \to K$ extends to a homomorphism $\widetilde{\theta} : K \to K$ and, of course, $\Psi(\widetilde{\theta}) = \theta$.

(1) ($\Leftarrow$) Let H be a normal subgroup of $\mathrm{Gal}(K/k)$. To show that K^H/k is normal, suppose that f is an irreducible polynomial in $k[x]$ and that $\alpha \in K^H$ is a zero of f. We will show that f splits in K^H.

Certainly f splits in K. Let $\beta \in K$ be a zero of f. There is $\sigma \in \mathrm{Gal}(K/k)$ such that $\sigma(\alpha) = \beta$. If $\varphi \in H$, then $\sigma^{-1}\varphi\sigma \in H$, so $\alpha = \sigma^{-1}\varphi\sigma(\alpha) = \sigma^{-1}\eta(\beta)$, whence $\varphi(\beta) = \sigma(\alpha) = \beta$. Therefore $\beta \in K^H$.

($\Rightarrow$) This was proved when we proved (2). □

Remarks. Consider a Galois extension K/k and an intermediate extension $k \subset F \subset K$.

1. We will always think of $\mathrm{Gal}(K/F)$ as the subgroup of $\mathrm{Gal}(K/k)$ consisting of automorphisms of K/k that are the identity on F.

2. Notice that $[F : k] = |\mathrm{Gal}(K/k)|/|\mathrm{Gal}(K/F)|$.

3. Since $\mathrm{Gal}(K/k)$ is finite it has only a finite number of subgroups. Hence there are only a finite number of intermediate extensions F. This is not apriori obvious because one has intermediate extensions $k(\alpha)$ for all $\alpha \in K$ and there are infinitely many choices for α if k is infinite. For example, it is not at all obvious why only finitely many different extensions of $\mathbb{Q}$ appear as $\mathbb{Q}(a\sqrt[5]{2} + b\sqrt[7]{-11} + c\sqrt[3]{17})$ as $a, b, c \in \mathbb{Z}$ vary!

4. The surjectivity of Ψ in the proof of part (2) of Theorem 1.3 says, as we already know, that every $\theta \in \mathrm{Gal}(F/k)$ can be extended to an automorphism of K/k in $[K : F]$ different ways.

3.2 Elementary examples

Proposition 2.1. *The Galois group of $\mathbb{F}_{p^n}/\mathbb{F}_p$ is the cyclic group of order n generated by the Frobenius automorphism, $\sigma(a) = a^p$.*

Proof. Since $\mathbb{F}_{p^n}$ is the splitting field for $x^{p^n} - x$ over $\mathbb{F}_p$, it is a normal extension. It is a separable extension also because $\mathbb{F}_p$ is finite. Hence $\mathbb{F}_{p^n}/\mathbb{F}_p$ is a Galois extension of degree n.

The Frobenius automorphism is an automorphism of any field of characteristic p, and it fixes the elements of $\mathbb{F}_p$ because $\mathbb{F}_p - \{0\}$ is a group of order $p-1$ (so $a^{p-1} = a$ for all $0 \neq a \in \mathbb{F}_p$). Hence $\sigma \in \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$.

Now, $\sigma^r(a) = a^{p^r}$, so $\sigma^n = 1$. However, if $\sigma^r = 1$, then $a^{p^r} = a$ for all $a \in \mathbb{F}_{p^n}$, so every element in $\mathbb{F}_{p^n}$ is a zero of $x^{p^r} - x$; the degree of a polynomial is at least as big as its number of zeroes, so $p^r \geq |\mathbb{F}_{p^n}| = p^n$, whence $r \geq n$. It follows that the order of σ is n, so the subgroup of $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$ has order $\geq n$. However, $|\mathrm{Gal}(\mathbb{F}^{p^n}/\mathbb{F}_p)| = [\mathbb{F}_{p^n} : \mathbb{F}_p] = n$, so $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p) = \langle\sigma\rangle$. □

Example 2.2. *If $n = mr$, the Galois group of $\mathbb{F}_{p^n}/\mathbb{F}_{p^m}$ is the cyclic group of order r generated by the m^{th} power of the Frobenius automorphism. In other words,*

$$\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_{p^m}) = \langle\sigma^m\rangle$$

where $\sigma^m(a) = a^{p^m}$. ◊

Example 2.3. *The splitting field of $f = (x^2 - 2)(x^2 - 3)$ over $\mathbb{Q}$ is $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. This is a Galois extension and $|\mathrm{Gal}(K/\mathbb{Q})| = [K : \mathbb{Q}] = 4$. There are only two groups with four elements, $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$.*

There are elements σ, τ in $\mathrm{Gal}(K/\mathbb{Q})$ defined by

$$\begin{aligned} \sigma(\sqrt{2}) &= -\sqrt{2} \qquad \sigma(\sqrt{3}) = \sqrt{3} \\ \tau(\sqrt{2}) &= \sqrt{2} \qquad \tau(\sqrt{3}) = -\sqrt{3}, \end{aligned}$$

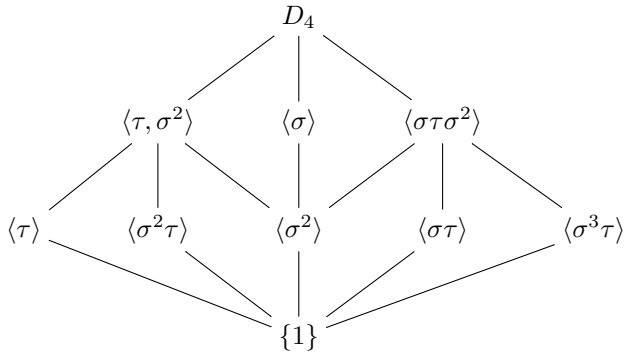
so $\mathrm{Gal}(K/\mathbb{Q}) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. Its subgroup lattice is This is easy to compute. For example, if $H = \{1, \sigma\tau\}$, then $K^H = \mathbb{Q}(\sqrt{6})$. ◊

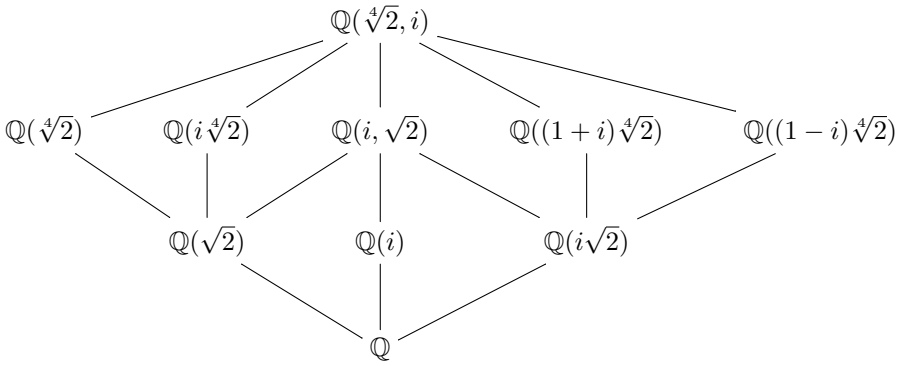
Example 2.4. *Let K be a splitting field for $f = x^4 - 2 \in \mathbb{Q}[x]$.*

Set $\alpha = 2^{1/4}$. The zeroes of f are $\pm\alpha$ and $\pm i\alpha$ so $K = \mathbb{Q}(\alpha, i) = \mathbb{Q}(\alpha)(i)$, whence $[K : \mathbb{Q}] = 8$. Hence $G = \mathrm{Gal}(K/\mathbb{Q})$ has eight elements. Elements of G must permute the zeroes of f, so G is a subgroup of the symmetric group S_4. Since α and $i\alpha$ have the same minimal polynomial over $\mathbb{Q}(i)$, namely $x^4 - 2$, there is a $\sigma \in G$ such that $\sigma(i) = i$ and $\sigma(\alpha) = i\alpha$. Similarly, there is a $\tau \in G$ such that $\tau(\alpha) = \alpha$ and $\tau(i) = -i$.

We will now show that G is isomorphic to the dihedral group D_4, the symmetry group of the square. A straighforward calculation shows that $\sigma^4 = \tau^2 = 1$ and $\tau\sigma\tau = \sigma^{-1}$, so there is a homomorphism $D_4 \to G$. Notice that $\sigma^2 \neq 1$; since a non-trivial normal subgroup of D_4 contains σ^2, the map $D_4 \to G$ is injective and hence surjective. Thus $G = \langle \sigma, \tau \rangle \cong D_4$.

Before determining the intermediate fields we make some calculations. Since $\sigma^2(\alpha) = -\alpha$, $\sigma^2(\alpha^2) = \alpha^2$. Also $\sigma\tau(i) = -i$, $\sigma\tau(\alpha) = i\alpha$, and $\sigma\tau(i\alpha) = \alpha$, so $\sigma\tau(\alpha + i\alpha) = \alpha + i\alpha$.





As we said above, every non-trivial normal subgroup of D_4 contains σ^2. Since $D_4/\langle\sigma^2\rangle \cong \mathbb{Z}_4$, the non-trivial normal subgroups of G are $\langle\sigma^2\rangle$ and those of index two, namely

$$\{1, \sigma^2\},\ \{1, \sigma, \sigma^2, \sigma^3\},\ \{1, \sigma^2, \tau, \tau\sigma^2\},\ \{1, \sigma^2, \tau\sigma, \tau\sigma^3\}.$$

The corresponding intermediate fields are the normal extensions of $\mathbb{Q}$, namely

$$\begin{aligned} K^{\langle\sigma^2\rangle} &= \mathbb{Q}(i, \alpha^2) = \mathbb{Q}(i, \sqrt{2}), \\ K^{\langle\sigma\rangle} &= \mathbb{Q}(i), \\ K^{\langle\tau, \sigma^2\rangle} &= K^{\langle\sigma^2\rangle} \cap K^{\langle\tau\rangle} = \mathbb{Q}(\sqrt{2}), \\ K^{\langle\sigma\tau, \sigma^2\rangle} &= \mathbb{Q}(i\sqrt{2}). \end{aligned}$$

The other intermediate fields are

$$\begin{aligned} K^{\langle\tau\rangle} &= \mathbb{Q}(\alpha), \\ K^{\langle\sigma\tau\rangle} &= \mathbb{Q}((1+i)\alpha), \\ K^{\langle\sigma^2\tau\rangle} &= \mathbb{Q}(i\alpha), \\ K^{\langle\sigma^3\tau\rangle} &= \mathbb{Q}((1-i)\alpha), \end{aligned}$$

all of which are degree four extensions of $\mathbb{Q}$. ◊

Exercises.
Let β be an element of $\mathbb{F}_4$ that is not in $\mathbb{F}_2$.

1. Find the minimal polynomial of β over $\mathbb{F}_2$.
2. Show that $x^2 + (\beta + 1)x + 1$ is irreducible in $\mathbb{F}_4[x]$.
3. Is the cubic $x^3 + x^2 + \beta \in \mathbb{F}_4[x]$ irreducible? If not, find its factors.

4. Show that $\mathbb{F}_{16}$ contains an element α that is a primitive fifth root of one over $\mathbb{F}_2$, and that $\mathbb{F}_{16} = \mathbb{F}_2(\alpha)$. Find the minimal polynomial of α over $\mathbb{F}_4$, and show that α^4 is the other zero of this polynomial.

5. Show that α is a zero of $x^3 + x^2 + \beta \in \mathbb{F}_4[x]$.

6. Show that the Galois group of $x^5 - 1$ over $\mathbb{F}_4$ is $\mathbb{Z}_2$.

7. Factor $x^4 + x^3 + x^2 + x + 1$ over $\mathbb{F}_4$.

8. You have shown above that $\mathbb{F}_{16} = \mathbb{F}_4(\alpha)$ where α is a primitive fifth root of 1. Does there exist an element $\alpha \in \mathrm{Gal}(x^5 - 1/\mathbb{F}_4)$ such that $\sigma(\alpha) = \alpha^2$?

3.3 Polynomials of degree ≤ 4

In this section we determine the Galois groups of quadratics, cubics, and some quartics.

Notation. The Galois group of a separable polynomial $f \in k[x]$ is $\mathrm{Gal}(f/k) := \mathrm{Gal}(K/k)$ where K is a splitting field of f.

Definition 3.1. *A subgroup H of the symmetric group S_n is* transitive *if, given any $i, j \in \{1, \ldots, n\}$, there is an $\eta \in H$ such that $\eta(i) = j$.* ♢

Lemma 3.2. *Let $f \in k[x]$ be a separable, irreducible, polyomial of degree n. Then $\mathrm{Gal}(f/k)$ is isomorphic to a transitive subgroup of the symmetric group S_n, and n divides $|\mathrm{Gal}(f/k)|$.*

Proof. Let K be the splitting field of f and write $K = k(\alpha_1, \ldots, \alpha_n)$ where the α_i are the distinct zeroes of f. Because f is irreducible it is the minimal polynomial of each α_i, so $[k(\alpha) : k] = \deg f = n$. Hence n divides $[K : k] = |\mathrm{Gal}(K/k)|$.

If $\sigma \in \mathrm{Gal}(K/k)$, then the minimal polynomial of $\sigma(\alpha_i)$ is the same as the minimal polynomial of α_i so is f. Hence $\sigma(\alpha_i) = \alpha_j$ for some j. Hence $\mathrm{Gal}(K/k)$ permutes the zeroes of f and this gives a homomorphism $\mathrm{Gal}(K/k) \to S_n$. It is injective because if σ is the identity on the α_is it is the identity on $k(\alpha_1, \ldots, a_n)$. Because α_i and α_j have the same minimal polynomial there is an automorphism of K sending α_i to α_j (Theorem 1.5). Hence $\mathrm{Gal}(K/k)$ is a transitive subgroup of the symmetric group. □

The discriminant. Let $f \in k[x]$ be a monic, irreducible, separable polynomial. The element

$$\delta(f) := \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)$$

depends (up to a sign) on the order in which we label the zeroes of f, so is not an invariant of f. However, the discriminant of f, which is defined to be

$$D(f) := \delta(f)^2$$

is independent of the labelling of the zeroes of f.

If $\tau \in S_n$ is a transposition, then $\tau(\delta) = -\delta$ and $\tau(D) = D$.

It follows that D is invariant under $\mathrm{Gal}(f/k)$, so belongs to $K^{\mathrm{Gal}(K/k)} = k$.

Lemma 3.3. *Let $f \in k[x]$ be monic, irreducible, and separable. Then $\mathrm{Gal}(f/k)$ is contained in the alternating group if and only if the discriminant $D(f)$ is a square in k.*

Proof. Clearly $D(f)$ is a square in k if and only if $\delta(f)$ belongs to k; if and only if $\delta(f)$ is fixed by every element of $\mathrm{Gal}(f/k)$. But $\delta(f)$ is fixed by even permutations and sent to $-\delta(f)$ by odd permutations, so $\delta(f) \in k$ if and only if $\mathrm{Gal}(f/k)$ consists of even permutations. □

Quadratic Polynomials. Let $f = (x-\alpha)(x-\beta) = x^2 + bx + c \in k[x]$. Then $b = -(\alpha+\beta)$ and $c = \alpha\beta$. The discriminant is

$$(\alpha-\beta)^2 = \alpha^2 - 2\alpha\beta + \beta^2 = (\alpha+\beta)^2 - 4\alpha\beta = b^2 - 4c.$$

Since $S_2 \cong \mathbb{Z}_2$, $\mathbb{A}_2 = \{1\}$. The lemma says that $\mathrm{Gal}(f/k)$ is trivial if and only if $D(f)$ is a square, i.e., if and only if $b^2 - 4c$ is a square in k. In other words f splits in k if and only if $b^2 - 4c$ is a square, a result known to the ancients.

Cubic Polynomials. If $f = x^3 + ax^2 + bx + c$, a tedious computation gives

$$D(f) = a^2b^2 - 4b^3 - 4a^3c + 18abc - 27c^2.$$

You should do this tedious computation at least once in your life: write $f = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$, express each of a, b, and c in terms of $\alpha_1, \alpha_2, \alpha_3$, then multiply out $D(f) = (\alpha_1-\alpha_2)^2(\alpha_1-\alpha_3)^2(\alpha_2-\alpha_3)^2$ and rewrite it in terms of a, b, c.

We can always make a linear change of variable to bring a cubic polynomial into the form $f = x^3 + px + q$. Doing so, the discriminant takes the simpler form

$$-4p^3 - 27q^2.$$

The only transitive subgroups of S_3 are S_3 itself and A_3, so the Galois group of an irreducible, separable cubic is either S_3 or $A_3 \cong \mathbb{Z}_3$, with the two possibilities being determined by whether $D(f)$ is or is not a square in k.

Quartic Polynomials. The transitive subgroups of S_4 are:

S_4

the six conjugates of $\langle(1234)\rangle \cong \mathbb{Z}_4$

the three conjugates of $H = \langle(1234), (12)(34)\rangle \cong D_4$

A_4

$V = \{1, (12)(34), (13)(24), (14)(23)\} \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.

The last two are the only ones contained in A_4.

Proposition 3.4. *Let*

$$f = x^4 + ax^2 + b \in \mathbb{Q}[x]$$

be irreducible and suppose that $k/\mathbb{Q}$ is an extension that does not contain a zero of f. Because f is a quadratic in x^2 its zeroes are $\{\pm\alpha, \pm\beta\}$ in its splitting field. Then

$$\mathrm{Gal}(f/k) \cong \begin{cases} \mathbb{Z}_4 & \text{if } \alpha\beta^{-1} - \beta\alpha^{-1} \in k \\ \mathbb{Z}_2 \times \mathbb{Z}_2 & \text{if } \alpha\beta \in k \\ D_4 & \text{otherwise.} \end{cases}$$

Proof. Let's write $G = \mathrm{Gal}(f/k)$.

Notice that G can not contain a 3-cycle because such an element would fix one of the zeroes, say α, but would not fix $-\alpha$. That is absurd. Hence G can not equal to S_4 or A_4.

To compute the discriminant notice that

$$x^4 + ax^2 + b = (x^2 - \alpha^2)(x^2 - \beta^2) = x^4 - (\alpha^2 + \beta^2)x^2 + \alpha^2\beta^2$$

so $-a = \alpha^2 + \beta^2$ and $b = \alpha^2\beta^2$. Therefore

$$\begin{aligned} \delta &= (\alpha + \alpha)(\alpha - \beta)(\alpha + \beta)(-\alpha - \beta)(-\alpha + \beta)(-\beta - \beta) \\ &= -4\alpha\beta(\alpha^2 - \beta^2)^2 \\ &= -4\alpha\beta\big(b^2 - 4a\big). \end{aligned}$$

Hence $\alpha\beta \in k \iff \delta \in k \iff G \subset A_4 \iff G \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.

Suppose k contains $\phi := \alpha\beta^{-1} - \alpha^{-1}\beta$ and that $V \subset G$. Then G contains an automorphism σ such that $\sigma(\alpha) = \beta$. But $\sigma(\phi) = -\phi$, so $\phi \notin k$. This is a contradiction. Since H and its conjugates contain V we conclude that $G \cong \mathbb{Z}_4$ if $\phi \in k$.

The 4-cycles

$$\begin{aligned} &\alpha \mapsto -\alpha \mapsto -\beta \mapsto \beta, \\ &\alpha \mapsto \beta \mapsto -\beta \mapsto -\alpha, \\ &\alpha \mapsto -\beta \mapsto \beta \mapsto -\alpha, \end{aligned}$$

can not belong to G because they do not extend to automorphisms of the splitting field: for example, if σ is the first of these 4-cycles, then $\sigma(\alpha)\sigma(\beta) = -\alpha^2$ whereas $\sigma(-\alpha)\sigma(-\beta) = \beta^2$ so there is no sensible way to define $\sigma(\alpha\beta)$.

The other three 4-cycles all fix ϕ so, if $G \cong \mathbb{Z}_4$, then $\phi \in k$. Thus, $\phi \in k \iff G \cong \mathbb{Z}_4$. □

3.4 Generic Polynomials

The symmetric group S_n acts on the polynomial ring $k[t_1, \ldots, t_n]$ in the obvious way:

$$\sigma(t_i) := t_{\sigma(i)}.$$

We extend the action by requiring σ to act as the identity on k and to be an automorphism of the polynomial ring. Thus, if $\sigma = (123)$, then the action of σ on a polynomial f is given by replacing each t_1 by t_2, each t_2 by t_3, and each t_3 by t_1, and leaving the other t_js untouched.

The action extends to the function field $k(t_1, \ldots, t_n)$.

A polynomial f which is invariant under S_n, that is, $\sigma(f) = f$ for all $\sigma \in S_n$, is called a symmetric polynomial. The invariants $k[t_1, \ldots, t_n]^{S_n}$ form a subring of $k[t_1, \ldots, t_n]$. Among the invariants are the elementary symmetric polynomials defined as follows:

$$\begin{aligned}
\varepsilon_0 &= 1 \\
\varepsilon_1 &= t_1 + \cdots + t_n \\
\varepsilon_2 &= \sum_{p<q} t_p t_q = t_1t_2 + t_1t_3 + \cdots + t_2t_3 + \cdots + t_{n-1}t_n \\
\varepsilon_3 &= \sum_{p<q<r} t_p t_q t_r \\
&\vdots \quad \vdots \\
\varepsilon_n &= t_1 t_2 \cdots t_n.
\end{aligned}$$

Theorem 4.1. *1.* $k[t_1, \cdots, t_n]^{S_n} = k[\varepsilon_1, \cdots, \varepsilon_n]$;

2. $k[\varepsilon_1, \cdots, \varepsilon_n]$ *is a polynomial ring in* n *variables;*

3. $k(t_1, \cdots, t_n)^{S_n} = k(\varepsilon_1, \cdots, \varepsilon_n)$;

4. $k(t_1, \cdots, t_n)$ *is a Galois extension of* $k(\varepsilon_1, \cdots, \varepsilon_n)$ *with Galois group* S_n.

Proof. (3) Let $f, g \in k[t_1, \ldots, t_n]$ and suppose f/g is fixed by all $\sigma \in S_n$. Define

$$H := \prod_{\sigma \in S_n \sigma \neq 1} \sigma(g).$$

It is clear that $gH \in k[t_1, \ldots, t_n]^{S_n}$. Since $f/g = fH/gH$ is fixed by S_n it follows that fH is fixed by S_n. Hence $f/g = fH/gH \in k(\varepsilon_1, \ldots, \varepsilon_n)$.

(4) It is clear that

$$\begin{aligned}
f :=& (x - t_1) \cdots (x - t_n) \qquad \text{and belongs to } k(t_1, \ldots, t_n)[x] \\
=& x^n - \varepsilon_1 x^{n-1} + \varepsilon_2 x^{n-2} - \cdots + (-1)^{n-1}\varepsilon_{n-1}x + (-1)^n \varepsilon_n \in k(\varepsilon_1, \ldots, \varepsilon_n)[x]
\end{aligned}$$

and its splitting field over $k(\varepsilon_1, \ldots, \varepsilon_n)$ is $k(t_1, \ldots, t_n)$. Hence $k(t_1, \ldots, t_n)$ is a normal extension of $k(\varepsilon_1, \ldots, \varepsilon_n)$. Since the roots of f are distinct, f is separable, and therefore $k(t_1, \ldots, t_n)$ is a separable extension of $k(\varepsilon_1, \ldots, \varepsilon_n)$.

The Galois group for the extension is a subgroup of S_n. But S_n itself acts as automorphisms of the extension $k(t_1, \ldots, t_n)/k(\varepsilon_1, \ldots, \varepsilon_n)$, so the Galois group must equal S_n. □

The last part of the theorem tells us that the degree of the extension $k(t_1,\dots,t_n)/k(\varepsilon_1,\dots,\varepsilon_n)$ is $n!$. One can prove that fact directly, i.e., without appealing to Galois theory. You might like to think about finding $n!$ *nice looking* elements in $k[t_1,\dots,t_n]$ that provide a basis for the extension $k(t_1,\dots,t_n)/k(\varepsilon_1,\dots,\varepsilon_n)$.

Part (1) of Theorem 4.1 says that every polynomial that is invariant under the action of the symmetric group can be written as a polynomial in the elementary symmetric polynomials $\varepsilon_1,\cdots,\varepsilon_n$. For example, the polynomials $t_1^d+\cdots+t_n^d$ are obviously invariant. We have

$$t_1^2+\cdots+t_n^2=\varepsilon_1^2-2\varepsilon_2.$$

Try finding analogous expressions for $d=3,4,\dots$.

Lemma 4.2. *Let* $f=(x-t_1)\cdots(x-t_n)\in k(t_1,\dots,t_n)[x]$. *Then*

$$f=x^n-\varepsilon_1x^{n-1}+\varepsilon_2x^{n-2}-\cdots+(-1)^{n-1}\varepsilon_{n-1}x+(-1)^n\varepsilon_n.$$

Proof. Straightforward. □

We can view these previous two results in another way. Let $s_1,\dots,s_n$ be indeterminates over k, and consider the field $k(s_1,\dots,s_n)$. The general polynomial of degree n is

$$f=x^n-s_1x^{n-1}+s_2x^{n-2}-\cdots+(-1)^{n-1}s_{n-1}x+(-1)^ns_n.$$

It belongs to $k(s_1,\dots,s_n)[x]$.

Theorem 4.3. $\mathrm{Gal}(f/k(s_1,\dots,s_n))\cong S_n$.

Proof. Let $t_1,\dots,t_n$ be indeterminates and $\varepsilon_1,\cdots,\varepsilon_n$ the elementary symmetric polynomials in $t_1,\dots,t_n$. By Theorem 4.1, $k(s_1,\dots,s_n)$ is isomorphic to the subfield $k(\varepsilon_1,\cdots,\varepsilon_n)$ of $K=k(t_1,\cdots,t_n)$. We identify $k(s_1,\dots,s_n)$ with $k(\varepsilon_1,\cdots,\varepsilon_n)$. By Lemma 4.2, f splits over K. In fact $f=(x-t_1)\cdots(x-t_n)$, so $K/k(s_1,\dots,s_n)$ is the splitting field of f. By Theorem 4.1(3), the Galois group of f is isomorphic to S_n. □

Corollary 4.4 (Abel, Galois)**.** *Let* $n\geq 5$. *The general polynomial of degree* n *is not solvable by radicals.*

Proof. We will prove this in the next chapter—a polynomial is solvable by radicals if and only if its Galois group is solvable. However, the symmetric group S_n is not solvable if $n\geq 5$. □

Chapter 4

Solvability by radicals

A polynomial $f \in k[x]$ is solvable by radicals if the zeroes of f are given by a "formula" that involves only $+$, $-$, $\times$, $\div$, n^{th} roots, and elements of k and the coefficients of f. The paradigmatic example is that the zeroes of $ax^2 + bx + c$ are given by

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

provided that $a \neq 0$ and $\operatorname{char} k \neq 2$. (What happens when $\operatorname{char} k = 2$?)

The formula for the zeroes of a quadratic polynomial was known to several ancient civilizations. A formula for the cubics was not found until the 16th century, and is generally credited to Scipio Ferro and Niccolo Tartaglia, although the first published solution appeared in a book by Cardano. The zeroes of $x^3 + px + q$ are

$$A - B,\ \omega A - \omega^2 B,\ \omega^2 A - \omega B,$$

where ω is a primitive cube root of unity, say $\frac{1}{2}(-1 + \sqrt{-3})$, and

$$A = \sqrt[3]{-\frac{q}{2}} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^2}$$

$$B = \sqrt[3]{+\frac{q}{2}} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^2}$$

and the cube roots are chosen so that $AB = -3p$. That is quite some formula! Notice though that if p and q belong to some field k, the three zeroes belong to the field $k(\omega, \alpha, A, B)$ where $\alpha = \sqrt{-3D}$ and $A^3, B^3 \in k(\omega, \alpha)$. In other words the zeroes belong to a field K that is obtained by successively adjoining roots of elements.

One very interesting aspect of this formula, an aspect that was a great puzzle at the time, is that the three roots could all be real numbers even if $\sqrt{-3D}$ is *not* a real number.

This was soon followed by a general solution to the quartic, and attention soon shifted to the quintic. Despite intense efforts during the 17th century no general solution was found and the suspicion then arose that there might not be

a formula for the solution to the general quintic. This was confirmed by Ruffini and Abel.

4.1 Roots of unity

The polynomials $x^n - 1$ and their Galois groups are of fundamental importance. Understanding them is a necessary preparation for Galois's theorem that (with a proviso on the characteristic of the field) a polynomial is solvable by radicals if and only if its Galois group is solvable.

Definition 1.1. *An element ζ in a field is called a* primitive n^{th} root of unity *if $\zeta^n = 1$ but $\zeta^i \neq 1$ for every $1 \leq i \leq n-1$.* ♢

Proposition 1.2. *Let k be a field of characteristic $p > 0$. If p divides n, then k does not contain a primitive n^{th} root of unity.*

Proof. Write $n = pd$. Then $x^n - 1 = (x^d - 1)^p$, so every n^{th} root of unity is a d^{th} root of unity. □

In particular, if $\operatorname{char} k$ divides n, no extension of k can contain a primitive n^{th} root of unity. The next proposition shows that if $\operatorname{char} k$ does not divide n then there is an extension of k containing a primitive n^{th} root of unity.

Lemma 1.3. *The n^{th} roots of unity form a cyclic subgroup of the multiplicative group of a field.*

Proof. The set Γ of n^{th} roots of unity in k is a subgroup of the multiplicative group $k^\times = k - \{0\}$. Let $m > 0$ be minimal such that $a^m = 1$ for all $a \in \Gamma$. Then $|\Gamma| \leq m$ because $x^m - 1$ has at most m zeroes in k. However, Γ is a finite abelian group so isomorphic to $\mathbb{Z}_{a_1} \oplus \cdots \oplus \mathbb{Z}_{a_t}$ for some integers a_i, so m is the least common multiple of the a_is, whence $m \leq a_1 \cdots a_t = |\Gamma|$. Hence $|\Gamma| = m$, so the a_is are relatively prime and $\Gamma \cong \mathbb{Z}_m$. □

The group of units in $\mathbb{Z}/8$ is equal to $\{1, 3, 5, 7\}$. These are the square roots of unity, also the 4^{th} roots of unity, etc. The group of units is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. In particular, it is not a cyclic group.

Proposition 1.4. *If the characteristic of k does not divide n, then*

1. *$x^n - 1$ is separable over k;*
2. *the splitting field of $x^n - 1$ contains a primitive n^{th} root of unity;*
3. *$\operatorname{Gal}(x^n - 1/k)$ is abelian;*
4. *$\operatorname{Gal}(x^n - 1/k)$ is isomorphic to a subgroup of the group of units in $\mathbb{Z}/(n)$.*[1]

[1] It need not be the full group of units because, for example, k might already contain ω in which case the Galois group is trivial.

Proof. (1) and (2). Since $\operatorname{char} k$ does not divide n, the derivative of $x^n - 1$ is non-zero, hence relatively prime to $x^n - 1$. Thus $x^n - 1$ has n distinct zeroes and these form a cyclic group $\{1, \omega, \ldots, \omega^{n-1}\}$ by Lemma 1.3. In particular, ω is a primitive n^{th} root of unity.

(3) and (4) Let K denote the splitting field of $x^n - 1$. If $\sigma \in \operatorname{Gal}(K/k)$, then $\sigma(\omega)$ is also a zero of $x^n - 1$, so $\sigma(\omega) = \omega^{i(\sigma)}$ for a unique $i(\sigma) \in \mathbb{Z}_n$. A simple calculation shows that the map $i : \operatorname{Gal}(K/k) \to \mathbb{Z}_n$, $\sigma \mapsto i(\sigma)$, satisfies $i(\sigma\tau) = i(\sigma)i(\tau)$ and $i(\sigma^{-1}) = i(\sigma)^{-1}$, so is a homomorphism to the group of units in $\mathbb{Z}_n$. □

Corollary 1.5. *Let p be a prime and k a field of characteristic not p. Then $\operatorname{Gal}(x^p - 1/k)$ is isomorphic to a subgroup of $\mathbb{Z}_{p-1}$, hence cyclic.*

Proof. The group of units in the ring $\mathbb{Z}_p$ is isomorphic to $\mathbb{Z}_{p-1}$ by Proposition 2.3.2. A subgroup of a cyclic group is cyclic. □

As the next example illustrates, $\operatorname{Gal}(x^p - 1/k)$ can be strictly smaller than $\mathbb{Z}/p - 1$.

Proposition 1.6. $\operatorname{Gal}(x^5 - 1/\mathbb{F}_4) \cong \mathbb{Z}_2$.

Proof. It is easy to see that $f = x^4 + x^3 + x^2 + x + 1$ is irreducible over $\mathbb{F}_2$. Hence if α is a zero of f, $[\mathbb{F}_2(\alpha) : \mathbb{F}_2] = 4$. The four distinct zeroes of f are $\alpha, \alpha^2, \alpha^3, \alpha^4$, so $\mathbb{F}_2(\alpha)$ is the splitting field for f over $\mathbb{F}_2$ and $\mathbb{F}_2(\alpha)$ is a Galois extension of $\mathbb{F}_2$. Since $\mathbb{F}_2(\alpha) \cong \mathbb{F}_{16}$, it has a subfield isomorphic to $\mathbb{F}_4$, and $[\mathbb{F}_2(\alpha) : \mathbb{F}_4] = 2$. Hence $\mathbb{F}_2(\alpha)$ is the splitting field of the separable polynomial $x^5 - 1$ over $\mathbb{F}_4$, and $\operatorname{Gal}(x^5 - 1/\mathbb{F}_4) = \operatorname{Gal}(\mathbb{F}_2(\alpha)/\mathbb{F}_4) \cong \mathbb{Z}_2$. □

The reason the Galois group is $\mathbb{Z}_2$ is because f is not irreducible over $\mathbb{F}_4$: we can write $\mathbb{F}_4 = \{0, 1, \beta, \beta + 1\}$ with $\beta^2 = \beta + 1$; then $f = (x^2 + \beta x + 1)(x^2 + (\beta + 1)x + 1)$.

4.2 Solvability by radicals

First the formality.

Definition 2.1. *A polynomial $f \in k[x]$ is* solvable by radicals *if there is a chain of field extensions*

$$k = K_0 \subset K_1 \subset \cdots \subset K_r \tag{2-1}$$

of the form

$$K_{i+1} = K_i\big(\sqrt[n_i]{a_i}\big)$$

for various positive integers n_i and elements $a_i \in K_i$, $0 \le i \le r - 1$, such that f splits in K_r.

In this situation we call each K_{i+1}/K_i a simple radical extension*, K_r/k a* radical extension*, and (2-1) a* radical sequence*.* ◊

Definition 2.2. *We call K/k a* cyclic extension *if it is Galois with cyclic Galois group.* ◊

Theorem 2.3. *Let k be a field containing a primitive n^{th} root of unity. Let K/k be an extension of degree n. Then K/k is a cyclic extension if and only if $K = k(\sqrt[n]{a})$ for some $a \in k$.*

Proof. Let ζ be a primitive n^{th} root of 1 in k.

($\Leftarrow$) Suppose $K = k(\alpha)$ where $\alpha = \sqrt[n]{a}$. Then K/k is the splitting field of the polynomial $x^n - a \in k[x]$ because it has n distinct zeroes $\alpha, \zeta\alpha, \ldots, \zeta^{n-1}\alpha$ in K. Hence K/k is a Galois extension.

If $\sigma \in \mathrm{Gal}(K/k)$, then $\sigma(\alpha) = \zeta^{i(\sigma)}\alpha$ for some $i(\sigma) \in \mathbb{Z}_n$. The map $\sigma \mapsto i(\sigma)$ is a homomorphism $\mathrm{Gal}(K/k) \to \mathbb{Z}_n$. It is injective, and hence an isomorphism because $|\mathrm{Gal}(K/k)| = [K : k] = n = |\mathbb{Z}_n|$.

($\Rightarrow$) Suppose $\mathrm{Gal}(K/k) = \langle\sigma\rangle \cong \mathbb{Z}_n$. By Dedekind's Proposition **??**, the distinct maps $\mathrm{id}_K, \sigma, \sigma^2, \ldots, \sigma^{n-1} : K \to K$ are linearly independent over K. Hence $\mathrm{id} + \zeta^{-1}\sigma + \cdots + \zeta^{1-n}\sigma^{n-1} \neq 0$, i.e.,

$$\alpha := \beta + \zeta^{-1}\sigma(\beta) + \cdots + \zeta^{1-n}\sigma^{n-1}(\beta)$$

is non-zero for some $\beta \in K$. Since $\sigma(\alpha) = \zeta\alpha$, α^n is invariant under σ, hence invariant under $\mathrm{Gal}(K/k)$. Thus $\alpha^n = a \in k$. Because $x^n - a \in k[x]$ has n distinct zeroes $\alpha, \zeta\alpha, \ldots, \zeta^{n-1}\alpha$ in K, it is separable over k and K is its splitting field.

Every element of $\mathrm{Gal}(K/k)$ sends $k(\alpha)$ to $k(\alpha)$, so restriction gives a homomorphism $\mathrm{Gal}(K/k) \to \mathrm{Gal}(k(\alpha)/k)$, $\phi \mapsto \phi|_{k(\alpha)}$, which is clearly injective. Hence $[k(\alpha) : k] \geq n$, and it follows that $K = k(\alpha) = k(\sqrt[n]{a})$. □

Lemma 2.4. *Let F/k be an extension. If $f \in k[x]$ is separable, then* $\mathrm{Gal}(f/F)$ *is isomorphic to a subgroup of* $\mathrm{Gal}(f/k)$.

Proof. First observe that f remains separable over F because its irreducible factors over F divide its irreducible factors over k, and therefore have no repeated zeroes.

Let $L = F(\alpha_1, \ldots, \alpha_n)$ be a splitting field for f over F, where $\alpha_1, \ldots, \alpha_n$ are the zeroes of f. Then $K := k(\alpha_1, \ldots, \alpha_n)$ is a splitting field for f over k.

If $\sigma \in \mathrm{Gal}(L/F) = \mathrm{Gal}(f/F)$, then each $\sigma(\alpha_i)$ is a zero of f so equal to some α_j. Hence $\sigma(K) \subset K$. The map $\sigma \mapsto \sigma|_K$ is a group homomorphism $\mathrm{Gal}(f/F) \to \mathrm{Gal}(K/k) = \mathrm{Gal}(f/k)$. This map is injective because if $\sigma|_K = \mathrm{id}_K$, then $\sigma(\alpha_i) = \alpha_i$ for all i whence $\sigma = \mathrm{id}_L$. □

Theorem 2.5. *Let $f \in k[x]$ and suppose that* $\mathrm{char}\, k$ *does not divide* $\deg f$. *If* $\mathrm{Gal}(f/k)$ *is solvable, then f is solvable by radicals.*

Proof. Set $n = (\deg f)!$. By Proposition 1.4, there is an extension F/k generated by a primitive n^{th} root of unity. By Lemma 2.4, $\mathrm{Gal}(f/F)$ is isomorphic to a subgroup of $\mathrm{Gal}(f/k)$ so is also solvable. Since F/k is a radical extension, it suffices to show that f is solvable by radicals over F. Hence we can, and will, assume that k contains a primitive n^{th} root of unity.

Let K/k be a splitting field of f, and choose a solvable series

$$\mathrm{Gal}(K/k) = G_0 \supset G_1 \supset \ldots \supset G_r = \{1\}$$

such that each G_i/G_{i+1} is cyclic. Define $K_i := K^{G_i}$ to obtain a sequence of fields

$$k = K_0 \subset K_1 \subset \ldots \subset K_r = K$$

with each K_{i+1}/K_i a cyclic extension whose degree, d say, divides n. Since K_i contains a primitive n^{th} root of unity, $K_{i+1} = K_i\big(\sqrt[d]{a}\big)$ for some $a \in K_i$ by Theorem 2.3. Hence K is a radical extension of k. □

Lemma 2.6. *Let K/k be a radical extension and L/k its normal closure. Then L/k is a radical extension.*[2]

Proof. Let $K = k(\alpha_1, \ldots, \alpha_r)$ be a radical extension of k. Let $K_i = k(\alpha_1, \ldots, \alpha_{i-1})$ and suppose $\alpha_i^{n_i} = a_i \in K_i$. Then α_i is a root of $x^{n_i} - a_i \in K_i[x]$.

Let f_i be the minimal polynomial of α_i over k and let $f := f_1 f_2 \ldots f_r$. Let L/k be a splitting field for f that contains K. Thus, L/k is generated by $\alpha_1, \ldots, \alpha_r$ and the other roots of f. Write $G := \mathrm{Gal}(L/k) = \{1, \sigma, \tau, \ldots\}$.

Define $L_1 := k(\alpha_1, \sigma(\alpha_1), \tau(\alpha_1), \ldots)$. Then L_1/k is a radical extension because $\sigma(\alpha_1)^{n_1} = \sigma(\alpha_1^{n_1}) = \sigma(a_1) = a_1$. Define $L_2 := L_1(\alpha_2, \sigma(\alpha_2), \tau(\alpha_2), \ldots)$. Now L_2/L_1 is a radical extension because

$$\sigma(\alpha_2)^{n_2} = \sigma(\alpha_2^{n_2}) = \sigma(a_2) \in \sigma(k(\alpha_1)) \subset \sigma(L_1) = L_1.$$

We continue in this way. Finally, $L_r = L_{r-1}(\alpha_r, \sigma(\alpha_r), \tau(\alpha_r), \ldots)$ which is a radical extension of L_{r-1} and hence a radical extension of k.

Now we will show that $L_r = L$. Let β be a root of f. Suppose β is a root of f_i. There is an isomorphism $\theta : k(\alpha_i) \to k(\beta)$ such that $\theta(\alpha_i) = \beta$. Since $L/k(\alpha_i)$ and $L/k(\beta)$ are normal extensions, θ extends to an automorphism, σ say, of L. Since $\sigma \in G$, $\sigma(\alpha_i) \in L_i$; i.e., $\beta \in L_i \subset L_r$. Since L/k is generated by all such β, $L = L_r$. □

Remark. We need a more precise version of Lemma 2.6. With the notation in its proof, suppose that $\alpha_i^{n_i} \in k(\alpha_1, \ldots, \alpha_{i-1})$ for all i. Then L is built from k by adjoining only n_i^{th} roots too.

Theorem 2.7. *Suppose that $f \in k[x]$ is solvable by radicals, say by taking various n_i^{th} roots. If $\mathrm{char}\, k$ does not divide any of the n_is, then $\mathrm{Gal}(f/k)$ is solvable.*

Proof. Let n be the least common multiple of the various n_is. Then $\mathrm{char}\, k$ does not divide n. By Lemma 2.6 and the hypothesis, there is a radical normal extension L/k in which f splits. By the remark after Lemma 2.6, L can be constructed by successively adjoining n^{th} roots.

Claim: It is enough to prove the theorem when k contains a primitive n^{th} root of unity. Proof: The splitting field for $x^n - 1$ over L is equal to $K = L(\xi)$ where

[2]The slogan is that a normal closure of a radical extension is radical.

ξ is a primitive n^{th} root of unity. Since L/k and $L(\xi)/L$ are radical extensions, $L(\xi)/k$ is a radical extension. Hence $L(\xi)/k(\xi)$ is a radical extension.

If $\sigma \in \mathrm{Gal}(L(\xi)/k)$, then $\sigma(\xi)$ is a primitive n^{th} root of unity and therefore a power of ξ. In particular, $\sigma(\xi) \in k(\xi)$. Therefore σ sends $k(\xi)$ to itself. Hence there is a restriction homomorphism

$$\mathrm{Gal}(L(\xi)/k) \to \mathrm{Gal}(k(\xi)/k).$$

The kernel of this is $\mathrm{Gal}(L(\xi)/k(\xi))$ and the image is abelian, hence solvable. It therefore suffices to show that $\mathrm{Gal}(L(\xi)/k(\xi))$ is solvable. $\diamond$

From now on we assume that k contains a primitive n^{th} root of unity, ξ say, and L/k is a radical normal extension

Let F/k be a splitting field for f that is a subfield of L/k. Since F/k is a normal extension, every automorphism of L sends F to itself. This gives a homomorphism

$$\mathrm{Gal}(L/k) \to \mathrm{Gal}(F/k) = \mathrm{Gal}(f/k)$$

with kernel $\mathrm{Gal}(L/F)$. To show that $\mathrm{Gal}(f/k)$ is solvable it it suffices to show that $\mathrm{Gal}(L/k)$ is solvable.

There is a tower of fields

$$k = L_0 \subset L_1 \subset \cdots \subset L_r = L$$

in which each $L_{i+1} = L_i\big(\sqrt[n]{a_i}\big)$ for some $a_i \in L_i$.

Let $G_i := \{\sigma \in \mathrm{Gal}(L/k) \mid \sigma\big|_{L_i} = \mathrm{id}_{L_i}\} \subset \mathrm{Gal}(L/L_i)$. There is a chain

$$\mathrm{Gal}(L/k) = G_0 \supset G_1 \supset \cdots \supset G_r = \{1\}.$$

Suppose $\sigma \in G_i$. Because $\sigma(\sqrt[n]{a_i})$ is another n^{th} root of a_i, $\sigma(\sqrt[n]{a_i}) = \xi^j \sqrt[n]{a_i}$ which is in L_{i+1} also. Hence $\sigma(L_{i+1}) \subset L_{i+1}$. Therefore restriction gives a homomorphism $G_i \to \mathrm{Gal}(L_{i+1}/L_i)$. The kernel of this homomorphism is G_i. Hence G_{i+1} is a normal subgroup of G_i and the quotient G_i/G_{i+1} is isomorphic to a subgroup of $\mathrm{Gal}(L_{i+1}/L_i)$. By Theorem 2.3, L_{i+1}/L_i is a cyclic extension so $\mathrm{Gal}(L_{i+1}/L_i)$ is a cyclic group. Hence G_i/G_{i+1} is abelian. It follows that $\mathrm{Gal}(L/k)$ is solvable. $\square$

Example 2.8. *Let $k = \mathbb{F}_p(t)$ be the rational function field over the field of p elements. If $f = x^p - x - t \in k[x]$, then $\mathrm{Gal}(f/k) \cong \mathbb{Z}_p$, so is solvable. However, f is not solvable by radicals.*

Let α be a zero of f. Then

$$(\alpha+1)^p - (\alpha+1) - t = \alpha^p + 1 - \alpha - 1 - t = 0$$

so $\alpha, \alpha+1, \alpha+2, \ldots, \alpha+p-1$ are the distinct zeros of f. Hence f is separable. The splitting field of f is $k(\alpha)$ and the minimal polynomial of α is f, so $[k(\alpha) : k] = p$. Hence the Galois group is $\mathbb{Z}_p$.

However, f is not solvable by radicals.

Because $\mathbb{Z}_p$ is a simple group, if f is solvable by radicals, then $k(\alpha) = k(\beta)$ where β is a root of $x^p - a$ for some $a \in k$. But $x^p - a = (x - \beta)^p$ so $x^p - a$ is not separable and therefore $k(\beta)$ is not a separable extension of k. This would contradict the fact that $k(\alpha)$ is a separable extension of k.

$\Diamond$

The proof of the next result uses Cauchy's Theorem, which will be proved later, and a technical result on the symmetric group that will also be proved later. In particular, we use the fact that the symmetric group S_n is not solvable if $n \geq 5$.

Theorem 2.9. *Let $f \in \mathbb{Q}[x]$ be an irreducible polynomial of prime degree, say p. If f has exactly two non-real zeroes, then $\mathrm{Gal}(f/\mathbb{Q}) \cong S_p$. In particular, if $p \geq 5$, then f is not solvable by radicals.*

Proof. By the Fundamental Theorem of Algebra f splits in $\mathbb{C}$. Let K be the splitting field for f over $\mathbb{Q}$, and set $G = \mathrm{Gal}(K/\mathbb{Q}) = \mathrm{Gal}(f/\mathbb{Q})$. By Lemma 2.4, the action of G on the zeroes of f gives an injective group homomorphism $G \to S_p$.

By Lemma 3.2, p divides $[K : \mathbb{Q}] = |G|$ so, by Cauchy's Theorem, G has an element σ of order p. This must be a p-cycle, say $\sigma = (12 \dots p)$. Complex conjugation is a $\mathbb{Q}$-automorphism of $\mathbb{C}$, so restricts to a $\mathbb{Q}$-automorphism of K. But conjugation fixes the $p - 2$ real zeroes of f, and permutes the two non-real zeroes so is a 2-cycle. Let's assume this transposition is $(1n)$.

Notice that σ^n is again a p-cycle and $\sigma^n = (1\, n\, 2n{-}1 \cdots)$, so after relabelling we can assume that G contains $(12 \dots p)$ and (12). But these two elements generate all of S_p so $G = S_p$. $\square$

Exercise. In S_4, (13) and (1234) do *not* generate S_4, so check the "without loss of generality" claim in the last sentence of the proof—you need to use the fact that p is prime.

Example 2.10. *The polynomial $f = x^5 - 6x + 3 \in \mathbb{Q}[x]$ is not solvable by radicals. By Eisenstein's criterion f is irreducible. Since $f'(x) = 5x^4 - 6$ has only two real zeroes, say $\pm\alpha$, f has at most three real zeroes by Rolle's Theorem. Since $f(-\alpha) > 0 > f(\alpha)$ and $f(x) to \pm\infty$ as $x \to \pm\infty$, f has three real zeroes. Hence $\mathrm{Gal}(f/\mathbb{Q}) \cong S_5$.* $\Diamond$

4.3 Cyclotomic polynomials

Definition 3.1. *For each positive integer n define*

$$\zeta_n := e^{2\pi i/n}.$$

We call $\mathbb{Q}(\zeta_n)$ the n^{th} cyclotomic extension *of $\mathbb{Q}$.*

The minimal polynomial of ζ_n over $\mathbb{Q}$ is called the n^{th} cyclotomic polynomial *and is denoted by $\Phi_n(x)$.* $\Diamond$

Claim: $\mathbb{Q}(\xi_m, \xi_n) = \mathbb{Q}(\xi_\ell)$ where $\ell = \text{lcm}\{m, n\}$. To see this, first write $\ell = ma = nb$ where $(a, b) = 1$. Then ξ_ℓ^a is a primitive n^{th} root of unity, so $\xi_n \in \mathbb{Q}(\xi_\ell)$. Similarly, $\xi_m \in \mathbb{Q}(\xi_\ell)$. So it remains to prove that $\xi_\ell \in \mathbb{Q}(\xi_m, \xi_n)$. It suffices to show that $\mathbb{Q}(\xi_m, \xi_n)$ contains a primitive ℓ^{th} root of unity. Notice that $\xi_\ell^a = \xi_n$ and $\xi_\ell^b = \xi_m$. There are integers c and d such that $1 = ac + bd$. Hence $\xi_\ell = \xi_n^c \xi_m^d$. $\Diamond$

We write

$$\mu_n := \{\text{the group of } n^{\text{th}} \text{ roots of unity}\} \subset \mathbb{C}.$$

There is an isomorphism of groups $\mathbb{Z}_n \to \mu_n$ defined by $a \mapsto \xi_n^a$. If $1 \leq a < n$, ξ_n^a is a primitive d^{th} root of unity where $d = \gcd(a, n)$.

If we view $\mathbb{Z}_n$ as a ring and write U_n for its group of units, then ξ_n^a is a primitive n^{th} root of unity if and only if $a \in U_n$. Hence the set of primitive n^{th} roots of unity is the image in μ_n of U_n.

Lemma 3.2. *The n^{th} cyclotomic polynomial is*

$$\Phi_n(x) = \prod_{\substack{\xi \in \mu_n \\ \xi \text{ primitive}}} (x - \xi) = \prod_{\substack{1 \leq a < n \\ \gcd(a,n)=1}} (x - \xi_n^a)$$

Proof. It is clear that these two products are equal.

□

Definition 3.3. *The* Euler ϕ-function $\phi : \mathbb{N} \to \mathbb{N}$ *is defined by*

$$\phi(n) := \textit{the number of integers } 1 \leq m \leq n \textit{ such that } (m, n) = 1.$$

$\Diamond$

Corollary 3.4. *Let n be a positive integer. Then*

1. $\deg \Phi_n(x) = \phi(n)$;
2. $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \phi(n)$;
3. $\text{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ *is isomorphic to the group of units in* $\mathbb{Z}_n$.

Proof. The set $U_n := \{m \mid 1 \leq m \leq n \text{ and } (m, n) = 1\}$ has cardinality $\phi(n)$. The set U_n has two other descriptions: it is the set of m such that ζ_n^m is a primitive n^{th} root of unity; its image in $\mathbb{Z}_n$ is the group of units. The map $\text{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \to \mathbb{Z}_n$, $\sigma \mapsto i(\sigma)$ defined by $\sigma(\zeta_n) = \zeta_n^{i(\sigma)}$, is an isomorphism to U_n. □

Our goal is to show that if k is an algebraically closed field and R and S are k-algebras that are integral domains, then $R \otimes_k S$ is an integral domain. In geometric terms, the product of two irreducible varieties over an algebraically closed field is an irreducible variety.

We need a preliminary result that uses Gauss's Lemma so let's recall that. If R is a UFD with field of fractions F and $f \in R[t]$, then f is irreducible in $F[t]$ if and only if it is irreducible and primitive in $R[t]$. A polynomial in $R[t]$ is primitive if the gcd of its coefficients is a unit (or 1, if you prefer). We will use Gauss's Lemma in the following form: if f is an irreducible polynomial in $F[t]$ there is an element $c \in R - \{0\}$ such that cf is in $R[t]$ and is irreducible and primitive in $R[t]$.

In the next lemma R will be $k[x_1, \ldots, x_n]$ and $F = k(x_1, \ldots, x_n)$.

Lemma 3.5. *Let K/k be an extension such that k is algebraically closed in K. Let $x_1, \ldots, x_n$ be indeterminates and let $f \in k(x_1, \ldots, x_n)[t]$. If f is irreducible it remains irreducible as an element of $K(x_1, \ldots, x_n)[t]$.*

Proof. By the remarks prior to the lemma we can replace f by cf for some non-zero $c \in k[x_1, \ldots, x_n]$ and it suffices to prove that cf is irreducible in $K(x_1, \ldots, x_n)[t]$ so will assume that has been done, i.e., $f \in k[x_1, \ldots, x_n][t]$ is irreducible and primitive.

The gcd of a set of elements in $k[x_1, \ldots, x_n]$ can be computed in $K[x_1, \ldots, x_n]$ but that gcd is the same as their gcd computed in $k[x_1, \ldots, x_n]$. This is a consequence of the fact that gcd can be computed by the Euclidean algorithm and if one applies it to two elements in $k[x_1, \ldots, x_n]$ all the remainders and quotients belong to $k[x_1, \ldots, x_n]$. Hence f is primitive as an element in $K[x_1, \ldots, x_n][t]$. Thus, to prove the lemma it suffices to show that f is irreducible in $K[x_1, \ldots, x_n][t]$.

Suppose $g, h \in K[x_1, \ldots, x_n][t]$ and $f = gh$.

We will now work in the polynomial ring $K[x_1, \ldots, x_n, t]$ with the usual notion of total degree.

The leading coefficient of an element in $K[x_1, \ldots, x_n, t]$ is taken with respect to total degree and the ordering $t < x_1 < \cdots < x_n$. The leading coefficient of f belongs to k.

Suppose $u \in K$ is the leading coefficient of g. Replace g by $u^{-1}g$ and h by uh. The leading coefficient of g is now 1, and since the leading coefficient of f is in k it follows that the leading coefficient of h is also in k.

Let d be an integer such that $f \in k[x_1, \ldots, x_n, t]_{<d}$, the set of polynomials having total degree $< d$. The K-algebra homomorphism

$$\theta : K[x_1, \ldots, x_n, t] \to K[t], \qquad \theta(t) = t,\ \theta(x_i) = t^{d^i},\ 1 \le i \le n,$$

is injective on $K[x_1, \ldots, x_n, t]_{<d}$ and $\theta(f)$ is a polynomial in t with coefficients in k. Both g and h belong to $K[x_1, \ldots, x_n, t]_{<d}$ since f does. Let $\bar{K}$ be an algebraic closure of K and factor

$$\theta(f) = \alpha \prod_{i \in I} (t - \alpha_i)$$

where $\alpha \in k$ and $\alpha_i \in \bar{K}$. Since each α_i is a zero of a polynomial each α_i is algebraic over k. The coefficients of $\theta(g)$ and $\theta(h)$ are polynomials in the α_is so they too are algebraic over k. But $\theta(g), \theta(h) \in K[t]$ and k is algebraically closed in K so $\theta(g), \theta(h) \in k[t]$. Hence $g, h \in k[x_1, \ldots, x_n][t]$. But f is irreducible in $k[x_1, \ldots, x_n][t]$ so either g or h is a unit. □

Proposition 3.6. *Let k be an algebraically closed field and R and S integral domains containing a copy of k. Then $R \otimes_k S$ is an integral domain.*

Proof. Let F be the fields of fractions of R.

The natural map $R \otimes_k S \to F \otimes_k S$ is injective so it suffices to prove that $F \otimes_k S$ is an integral domain. Suppose it is not.

Then there is a finitely generated subalgebra $k[y_1, \ldots, y_m] \subset S$ such that $F \otimes_k k[y_1, \ldots, y_m]$ is not a domain. Now Fract $k[y_1, \ldots, y_m] = k(x_1, \ldots, x_n)(\alpha)$ where $x_1, \ldots, x_n$ are algebraically independent over k and α is algebraic over $k(x_1, \ldots, x_n)$; only a single α is needed because k and hence $k(x_1, \ldots, x_n)$ is separably closed.

Let f be the minimal polynomial of α over $k(x_1, \ldots, x_n)$. Then

$$F \otimes_k \frac{k(x_1, \ldots)[t]}{(f)}$$

is not a domain. It follows that f is not irreducible as an element of $F(x_1, \ldots, x_n)[t]$ even though it is irreducible as an element of $k(x_1, \ldots)[t]$. This contradicts the lemma so we conclude that $R \otimes_k S$ must be an integral domain. □

Corollary 3.7. *Let k be an algebraically closed field, R an integral domain containing k, and S any commutative k-algebra. If $\mathfrak{p}$ is a prime ideal in S, then $R \otimes_k (S/\mathfrak{p})$ is an integral domain.*

Chapter 5

Group theory

I assume you already know some group theory.

5.1 Some reminders

Assumed knowledge: The definitions of a group, group homomorphism, subgroup, left and right coset, normal subgroup, quotient group, kernel of a homomorphism, center, cyclic group, order of an element, symmetric group, cycle decomposition, transposition, even/odd permutation, alternating group, Lagrange's Theorem, Isomorphism theorems relating to a homomorphism $f : G \to H$, etc.

If N is a normal subgroup of a group G there is a bijection between the subgroups of G containing N and the subgroups of G/N. If H is a subgroup of G containing N the corresponding subgroup of G/N is H/N; furthermore, H is normal in G if and only if H/N is normal in G/N; and in that case, $G/H \cong (G/N)/(H/N)$.

For the most part I will write groups multiplicatively—thus, the product of two elements will be denote by juxtaposition, gh. Sometimes if the group is abelian it makes sense to write the group operation additively as we usually do with the integers—the sum is denoted by $g + h$. Sometimes it is not clear which notation is best. For example, the binary operation in the cyclic group of order n, $\mathbb{Z}_n$, is best written as $+$ when we are thinking of $\mathbb{Z}_n$ as a quotient of the integers or when we think of $\mathbb{Z}_n$ as a ring.

The n^{th} roots of unity in $\mathbb{C}$ form a group under multiplication and we denote this group by μ_n. Although μ_n is abelian we write its group operation multiplicatively. If we choose a primitive n^{th} root of unity, say ε, there is a group isomorphism $\mathbb{Z}_n \to \mu_n$ given by $a \mapsto \varepsilon^a$. Sometimes we will simply write $\{\pm 1\}$ for the group $\mathbb{Z}_2$.

Most of the time when we deal with a group we do not know whether or assume that it is abelian so it makes sense to write it multiplicatively.

The direct product of two groups G and H is denoted by $G \times H$ and is defined to the cartesian product with group operation

$$(g, h).(g', h') := (gg', hh').$$

It is easy to check that this is a group.

If G and H are abelian we often call their direct product the direct sum and denote it by $G \oplus H$. The reason for this is that G and H are $\mathbb{Z}$-modules with the action given by $n.g = g^n$.

Example 1.1. *There is an isomorphism* $(\mathbb{R}^+, \cdot) \times \mathbb{R}/\mathbb{Z} \to (\mathbb{C}^*, \cdot)$, $(r, [\theta + \mathbb{Z}]) \mapsto re^{2\pi i\theta}$. *If you prefer, there is an isomorphism* $(\mathbb{R}^+, \cdot) \times \mathbb{R}/2\pi\mathbb{Z} \to (\mathbb{C}^*, \cdot)$, $(r, [\theta + 2\pi\mathbb{Z}]) \mapsto re^{i\theta}$. ♢

The automorphism group of a group G consists of all group isomorphisms $\psi : G \to G$ and is denoted $\operatorname{Aut} G$. It is obviously a group under composition. Sometimes if $\sigma \in \operatorname{Aut} G$ it is common to write g^σ for $\sigma(g)$. The danger in doing this is that $g^{\sigma\tau} = (g^\tau)^\sigma$! The notation g^σ should not seem too odd because it is what we do already in some situations: consider the circle group

$$U(1) = \{z \in \mathbb{C} \mid |z| = 1\}$$

of complex numbers of absolute value one under multiplication and the automorphism sending each z to its inverse z^{-1}. Thus $\operatorname{Aut} U(1)$ has a subgroup isomorphic to $\mathbb{Z}_2$ that we usually denote by $\{\pm 1\}$ and the action of the automorphism -1 is denoted by $z \mapsto z^{-1}$ rather than $(-1)(z)$!

The circle group is often denoted $U(1)$ because it is the first in the family of unitary groups which are denoted $U(n)$, $n \geq 1$.

But do be wary that with these conventions $g^{\sigma\tau} = (g^\tau)^\sigma$!

Proposition 1.2. *If p is a prime, then* $\operatorname{Aut}_{\mathsf{gp}}(\mathbb{Z}_p) \cong \mathbb{Z}_{p-1}$.

Proof. Let's write $\mathbb{Z}_p$ multiplicatively by identifying it with μ_p, the set of p^{th} roots of unity in $\mathbb{C}^\times$.

If $1 \leq i \leq p-1$, the map $\theta_i : \mathbb{Z}_p \to \mathbb{Z}_p$ defined by $\theta_i(\varepsilon^j) = \varepsilon^{ij}$ is a group homomorphism. Because the only subgroups of $\mathbb{Z}_p$ are itself and the trivial subgroup, θ_i is both injective and surjective. Thus each θ_i belongs to $\operatorname{Aut} \mathbb{Z}_p$.

Now define $\phi : \mathbb{F}_p^\times \to \operatorname{Aut} \mathbb{Z}_p$ from the multiplicative group of non-zero elements of $\mathbb{F}_p$ by

$$\phi(\bar{i}) := \theta_i.$$

Here i as an integer and $\bar{i}$ is its image in $\mathbb{Z}/(p) = \mathbb{F}_p$. Since

$$\Big(\phi(\bar{i}) \circ \phi(\bar{j})\Big)(\varepsilon) = (\varepsilon^i)^j = \varepsilon^{ij} = \phi(\overline{ij})(\varepsilon),$$

ϕ is a group homomorphism. We already know that $\mathbb{F}_p^\times \cong \mathbb{Z}_{p-1}$, so it remains to show that ϕ is an isomorphism.

If $\theta \in \operatorname{Aut} \mathbb{Z}_p$ then θ is completely determined by its action on a generator, say $\zeta \in \mu_p$. If $\theta(\zeta) = \zeta^k$, then

$$\theta(\zeta^r) = \theta(\zeta)^r = (\zeta^k)^r = (\zeta^r)^k = \theta_k(\zeta^r)$$

so $\theta = \theta_k$. Hence ϕ is surjective.

If $\theta_i = \mathrm{id}$, then $\varepsilon^i = \varepsilon$ for all $\varepsilon \in \mu_p$, so $\varepsilon^{i-1} = 1$ and p must divide $i - 1$. Hence, $i = 1$ and we conclude that ϕ is injective. □

5.2 Semi-direct products

Suppose that $\varphi : H \to \operatorname{Aut} N$ is a group homomorphism. We define the semi-direct product

$$N \rtimes_\varphi H$$

to be the Cartesian product $N \times H$ with multiplication

$$(x, a).(y, b) = (xy^{\varphi(a)}, ab)$$

for $x, y \in N$ and $a, b \in H$. It is a little burdensome to carry the φ notation everywhere so we often supress it and write

$$N \rtimes H$$

and

$$(x, a).(y, b) = (xy^a, ab). \tag{2-1}$$

One should check that this product is associative:

$$((x, a).(y, b))(z, c) = (xy^a, ab)(z, c) = (xy^a z^{ab}, abc)$$

and

$$(x, a).((y, b).(z, c)) = (x, a)(yz^b, bc) = (x(yz^b)^a, abc) = (xy^a(z^b)^a, abc)$$

and this equal to the other product because $(z^b)^a = z^{ab}$. Because $N \rtimes H$ contains copies of both N and H as subgroups, namely $\{(x, 1) \mid x \in N\}$ and $\{(1, a) \mid a \in H\}$, it is common to identify N and H with those subgroups. Thus we say that N and H are subgroups of $N \rtimes H$. It then makes sense to simply write the elements of $N \rtimes H$ as xa with $x \in N$ and $a \in H$. Notice that $ax = x^a a$. The menmonic I use to remember the multiplication rule (**??**) is that elements of $N \rtimes H$ like to be written as xa with the H-piece a on the right, but if I find an element ax with $a \in H$ and $x \in N$, when I move the a to the right it twists the x as it moves past it—$ax = x^a a$.

Notice that if $a \in H \equiv (1, H) \subset N \rtimes H$, then N is stable under conjugation by $a \equiv (1, a)$, and $ana^{-1} = \varphi(a)(n)$ for all $n \in N$.

Example 2.1 (The dihedral groups)**.** *Let N be a cyclic group generated by τ. Let $s \in \operatorname{Aut} N$ be the automorphism of order two defined by $s(\tau) = \tau^{-1}$. Let $H = \{1, \sigma\} \cong \mathbb{Z}_2$ and let $\phi : H \to \operatorname{Aut} G$ be given by $\phi(\sigma) = s$. Then the semi-direct product $N \rtimes H$ is isomorphic to the dihedral group*

$$D = \langle \tau, \sigma \mid \sigma^2 = 1,\ \sigma\tau\sigma = \tau^{-1},\ \tau^n = 1 \rangle.$$

This includes the infinite dihedral group which occurs when $n = \infty$, i.e., $N \cong \mathbb{Z}$. ◊

If H is a subgroup of $\operatorname{Aut} N$, we may form $N \rtimes H$.

It is possible for the groups $N \rtimes_\phi H$ and $N \rtimes_\psi H$ to be isomorphic even if ϕ and ψ are different homomorphisms.

Proposition 2.2. *Let $\phi : H \to \operatorname{Aut} N$ be a group homomorphism and $\beta \in \operatorname{Aut} H$. Then $N \rtimes_\phi H \cong N \rtimes_{\phi\beta} H$.*

Proof. Define

$$\Phi : N \rtimes_{\phi\beta} H \cong N \rtimes_\phi H$$

by

$$\Phi(x, a) := (x, \beta(a)).$$

This is obviously a bijective map from $N \times H$ to itself. Also

$$\begin{aligned}
\Phi\Big((x,a)(y,b)\Big) &= \Phi\Big(\big(x\phi\beta(a)(y), ab\big)\Big) \\
&= \Big(x\phi\beta(a)(y), \beta(ab)\Big) \\
&= (x, \beta(a))(y, \beta(b)) \\
&= \Phi\Big((x,a)\Big)\Phi\Big((y,b)\Big)
\end{aligned}$$

so Φ is a group homomorphism. □

Example 2.3. *View elements of $V = \mathbb{Z}_n \oplus \mathbb{Z}_n$ as column vectors forming a group under addition. Then $\mathrm{GL}(2, \mathbb{Z}_n)$ the group of invertible 2×2 matrices with entries in the ring $\mathbb{Z}_n$ acts on V by left multiplication. Let define $\phi : \mathbb{Z}_n \to \mathrm{GL}(2, \mathbb{Z}_n)$ be the map*

$$\phi(c) = \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix}.$$

Thus, with our notation above, $\mathbb{Z}_n$ acts on V by $(a, b)^a := (a, b - ac)$. The associated semidirect product $G = V \rtimes \mathbb{Z}_n$ is isomorphic to the group of unipotent upper triangular matrices via the map

$$(a, b, c) \mapsto \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}.$$

This is an example of a discrete Heisenberg group. ◊

Example 2.4. *Let H and Z be abelian groups in which the group operations are written additively and multiplicatively respectively. Suppose that $\psi : H \times H \to Z$ is a function. Then the following two conditions are equivalent:*

1. *$\psi(0,0) = 1$ and $\psi(a,b)\psi(a+b,c) = \psi(a,b+c)\psi(b,c)$ for all $a,b,c \in H$;*
2. *$G = Z \times H$ is a group under the operation*

$$(w,a).(z,b) = (wz\psi(a,b), a+b).$$

Suppose these conditions are satisfied. Then there is an "exact" sequence $1 \to Z \to G \to H \to 1$ given by $z \mapsto (z,0)$ and $(z,a) \mapsto a$.

Let U_n denote the group of units in $\mathbb{Z}_n$, and set $Z = \mathbb{Z}_n$. Consider the ring of 2×2 matrices over $\mathbb{Z}_n$ and take the multiplicative Let G be the subgroup of the additive group of 2×2 matrices over $\mathbb{Z}_n$ consisting of the elements

$$G := \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a, c \in U_n,\ b \in \mathbb{Z}_n \right\}.$$

Then G is isomorphic to the group constructed above with $H = U_n \times U_n$ and $Z = \mathbb{Z}_n$. ◊

Groups of order 8. The abelian ones are $\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. The non-abelian ones are D_4, the dihedral group that is the symmetry group of the square, and $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ the quaternion group sitting inside Hamilton's ring of quaternions $\mathbb{H}$. Recall that $\mathbb{H}$ is the 4-dimensional $\mathbb{R}$-vectorspace with basis $1, i, j, k$ made into a ring via the multiplication rules

$$i^2 = j^2 = k^2 = -1 \qquad \text{and} \qquad ij = k,\ jk = i,\ ki = j.$$

Let's write

$$D_4 = \langle \sigma, \tau \mid \sigma^4 = \tau^2 = 1,\ \tau\sigma\tau = \sigma^{-1} \}.$$

Thus $D_4 = \{1, \tau, \sigma^i, \sigma^i\tau \mid 1 \le i \le 3\}$. The groups look similar: the center of D_4 is $\{1, \sigma^2\}$ and the center of Q is $\{\pm 1\}$, and the two groups have conjugacy classes of the same sizes:

$$\begin{aligned} D_4 :&\quad \{1\},\ \{\sigma^2\},\ \{\sigma, \sigma^{-1}\},\ \{\tau, \sigma^2\tau\},\ \{\sigma\tau, \sigma^3\tau\} \\ Q :&\quad \{1\},\ \{-1\},\ \{\pm i\},\ \{\pm j\},\ \{\pm k\} \end{aligned}$$

We saw above that D_4 is a semidirect product $\mathbb{Z}_4 \rtimes \mathbb{Z}_2$. Although $N = \langle i \rangle$ is a normal cyclic subgroup of Q of order 4, Q is *not* a semidirect product $N \rtimes H$ because the elements not in N, namely $\pm j, \pm k$, all have order 4. This (more or less) shows that Q is not isomorphic to D_4 because it cannot be written as a semi-direct product $\mathbb{Z}_4 \rtimes \mathbb{Z}_2$.

Proposition 2.5. *Let G be a group containing a normal subgroup N and a subgroup H such that $N \cap H = \{1\}$ and $G = NH$. Then $G \cong N \rtimes_\phi H$ where $\phi : H \to \operatorname{Aut} N$ is given by $\phi(h)(n) := hnh^{-1}$.*

Proof. To see that ϕ is a group homomorphism: $\phi(a)\phi(b)(n) = a(bnb^{-1})a^{-1} = \phi(ab)(n)$. □

5.3 The symmetric group

Definition 3.1. *The n^{th}* symmetric group*, denoted S_n, is the group of all permutations of $\{1, 2, \ldots, n\}$.* ◊

We always think of elements of S_n as acting on $\{1, 2, \ldots, n\}$.

Let $\sigma \in S_n$. The orbit of $i \in \{1, 2, \ldots, n\}$ under the action of $\sigma \in S_n$ is $\{i, \sigma(i), \sigma^2(i), \ldots\}$. Obviously $\{1, 2, \ldots, n\}$ is the disjoint union of its orbits under σ.

Notation. If $\sigma, \tau \in S_n$, the product $\sigma\tau$ means first do τ, then do σ. Thus, we think of permutations as acting on $\{1, 2, \ldots, n\}$ from the left. Not all books adopt this convention (e.g., P.M. Cohn's book uses the opposite convention). The permutation $\sigma \in S_9$ defined by

$$\sigma(1) = 1,\ \sigma(2) = 5, \sigma(3) = 7,\ \sigma(4) = 8,\ \sigma(5) = 2,$$
$$\sigma(6) = 4,\ \sigma(7) = 6, \sigma(8) = 9,\ \sigma(9) = 3,$$

is denoted by

$$\sigma = \begin{pmatrix} 1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 1\,5\,7\,8\,2\,4\,6\,9\,3 \end{pmatrix}.$$

We adopt the notation

$$(a\,b\,c\,\ldots\,z) := \begin{pmatrix} a\,b\,c\,\ldots\,z \\ b\,c\,\ldots\,z\,a \end{pmatrix}.$$

A permutation of this form is called a cycle. With our convention that $\sigma\tau$ means first do τ then σ, we have $(12)(23) = (123)$. For example,

$$\begin{pmatrix} 1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 1\,5\,7\,8\,2\,4\,6\,9\,3 \end{pmatrix} = (2\,5)(3\,7\,6\,4\,8\,9\,3).$$

Two cycles σ and τ are disjoint if they can be written as $\sigma = (abc\ldots z)$ and $\tau = (a'b'c'\ldots z')$ with $\{a, b, c, \ldots, z\} \cap \{a', b', c', \ldots, z'\} = \phi$. Disjoint cycles commute with each other. The length of the cycle $(abc\ldots z)$ is the cardinality of $\{a, b, c, \ldots, z\}$. A cycle of length k is called a k-cycle. A 2-cycle is called a transposition.

Lemma 3.2. *Every element of S_n can be written as a product of disjoint cycles in a unique way up to order.*

Proof. Let $\sigma \in S_n$. Write $\{1, 2, \ldots, n\}$ as the disjoint union of its σ-orbits, say $O_1 \cup \cdots \cup O_m$. Let τ_i be the cycle that is the identity on all O_j other than O_i and acts on O_i as does σ: thus, if $a \in O_i$, then $\tau_i = (a\,\sigma(a),\, \sigma^2(a)\,\ldots)$. Then $\sigma = \tau_1 \ldots \tau_m$. □

Lemma 3.3. *Every permutation can be written as a product of transpositions.*

Proof. Every cycle is a product of transpositions because, for example, $(1\,2\,\ldots\,m{-}1\,m) = (1\,m)(1\,m{-}1)\cdots(1\,3)(1\,2)$. But every permutation is a product of cycles, so the result follows. □

The Lemma can be read as saying that S_n is generated by transpositions. However, one can be efficient and generate it with just $n-1$ transpositions. Show that $S_n = \langle (1\,2), (2\,3), \ldots, (n-1\,n)\rangle$.

Partitions. A partition of a positive integer n is a collection of positive integers $n_1, \ldots, n_k$ such that $n_1 + \cdots + n_k = n$. The order of the integers is not important. It is often convenient to denote a partition by writing, for example, $(1^3 2 3^2 5)$ to denote the partition $1, 1, 1, 2, 3, 3, 5$ of 16. Each element of S_n determines a partition of n by taking the size of its orbits.

Lemma 3.4. *Two elements of S_n are conjugate if and only if they determine the same partition of n; that is, if and only if the have orbits of the same size.*

Proof. Suppose that σ and τ yield the same partition of n. Then, we can write $\{1, \ldots, n\}$ as a disjoint union in two ways, say

$$\{1, \ldots, n\} = A_1 \sqcup \ldots \sqcup A_r = B_1 \sqcup \ldots \sqcup B_r,$$

where $|A_i| = |B_i|$ for all i, and the elements of each A_i (resp., each B_i) consist of a single σ-orbit (resp., τ-orbit). Fix elements $a_i \in A_i$ and $b_i \in B_i$ for all i. It is obvious that there is an element $\eta \in S_n$ such that $\eta(A_i) = B_i$ for all i, and even more precisely $\eta(\sigma^j(a_i)) = \tau^j(b_i)$ for all i and j. In particular, $\eta(a_i) = b_i$, so $\tau = \eta\sigma\eta^{-1}$.

The converse is obvious. □

The bijection between conjugacy classes and partitions is fundamental to the analysis of the symmetric group.

Example 3.5. *The conjugacy classes in S_5 are as follows:*

Partition	*Element in the conjugacy class*	*Size of conjugacy class*
5	(12345)	24
1, 4	(1234)	30
2, 3	(12)(345)	20
1, 1, 3	(123)	20
1, 2, 2	(12)(34)	15
1, 1, 1, 2	(12)	10
1, 1, 1, 1, 1	1	1.

We will use this list to find the normal subgroups of S_5 in Proposition 3.8. ◊

The nest result is useful for recognizing when a subgroup of the symmetric group is actually the whole group. We will use it when show that certain polynomials do not have a solution in radicals (see ???).

Proposition 3.6. *S_n is generated by (12) and $(12\cdots n)$.*

Proof. Let H be the subgroup generated by $a = (12)$ and $b = (12\cdots n)$. Then H contains $bab^{-1} = (23)$ and hence, by induction, $(i\, i+1)$ for all i. Thus H contains $(12)(23)(12) = (13)$ and $(13)(34)(13) = (14)$, and so on. That is, $(1i) \in H$ for all i. Hence, if $i \neq j$, H contains $(1i)(1j)(1i) = (ij)$. Since every element of S_n is a product of transpositions we conclude that $H = S_n$, as claimed. □

Lemma 3.7. *If σ is written as a product of transpositions in two different ways, say $\sigma = \alpha_1 \cdots \alpha_m = \beta_1 \cdots \beta_r$, then $m \equiv r \pmod 2$.*

A permutation is even if it is a product of an even number of transpositions, and is odd if it is a product of an odd number of transpositions. The previous lemma ensures that this definition is unambiguous. The set of even permutations form a subgroup of S_n called the alternating group and denoted by A_n.

There is a well-defined group homomorphism

$$sgn : S_n \to \{\pm 1\}$$

defined by

$$sgn(\sigma) = \begin{cases} +1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}.$$

The kernel of this homomorphism is obvious A_n. Hence A_n is a normal subgroup of S_n of index 2.

We now use the list of conjugacy classes in Example 3.5 to find the normal subgroups of S_5.

Proposition 3.8. *The only normal subgroups of S_5 are A_5, S_5, and $\{1\}$.*

Proof. Let H be a normal subgroup that is neither S_5 nor $\{1\}$. Since $gHg^{-1} \subset H$ for every $g \in S_5$, H is a union of conjugacy classes. The conjugacy classes have sizes 1, 10, 15, 20, 24, and 30. The class of size one must belong to H because it consists of the identity element.

The order of H is a divisor of $|S_5| = 120$. Since 120 is not divisible by $1+10$, or $1+15$, or $1+20$, or $1+24$, or $1+10+15$, or $1+30$, the only possibilities for $|H|$ are $40 = 1+15+24$ and $60 = 1+15+24+20$.

If $|H| = 40 = 24+15+1$, then H contains every 5-cycle, so contains (12345), and contains every element corresponding to the partition $1, 2, 2$, so contains $(12)(34)$. Hence H contains their product (135). Since (135) has order 3, $|H| \neq 40$.

Thus $|H| = 60$ and H contains the conjugacy classes of (12345), $(12)(34)$, and (135). But the union of these conjugacy classes is A_5. Hence $H = A_5$. □

Lemma 3.9. *A_n is generated by $(123), (124), \ldots, (12n)$.*

Proof. By its very definition A_n is generated by the elements $(ab)(cd) = (cad)(abc)$ and $(ac)(ab) = (abc)$. So it suffices to show that each (abc) belongs to the subgroup generated by $\{(12m) \mid 3 \leq m \leq n\}$. If $\{b, c\} \cap \{1, 2\} = \varnothing$, then

$(1bc) = (12c)^{-1}(12b)(12c)$ and $(2bc) = (12b)(12c)(12b)^{-1}$. If $\{a, b, c\} \cap \{1, 2\} = \varnothing$, then $(abc) = (12a)(2bc)(12a)^{-1}$. The result now follows. □

Proposition 3.10. *If $n \geq 5$, then A_n is a simple group.*

Proof. Let H be a non-trivial normal subgroup of A_5.

Suppose that H contains a 3-cycle, say (123). If $i \notin \{1, 2, 3\}$, then H contains $(12)(3i)(123)(3i)(12) = (1i2)$ and its square $(12i)$. It now follows from Lemma 3.9 that $H = A_n$.

Choose $1 \neq \alpha \in H$ fixing as many elements of $\{1, \ldots, n\}$ as possible. Since α is even it is not a transposition. Write α as a product of disjoint cycles.

Suppose that only 2-cycles occur in the cycle decomposition of α. Suppose first that $\alpha = (12)(34)\cdots$. If $\alpha = (12)(34)$, then H contains $(543)\alpha(543)^{-1}\alpha^{-1} = (345)$. If $\alpha \neq (12)(34)$, then $\alpha = (12)(34)(56)(78)\cdots$. Hence H contains $(543)\alpha(543)^{-1}\alpha^{-1} = (36)(45)$ and applying the previous argument to $(36)(45)$ in place of $(12)(34)$ we see that H contains a 3-cycle.

If α moves just 3 elements then it must be a 3-cycle.

We may now assume that the cycle decomposition for α contains a d-cycle with $d \geq 3$, say $\alpha = (123\cdots)\cdots$. If α moves exactly four elements, then it must be $(123i)$; but this is not even, so α moves at least 5 elements, say $1, 2, 3, 4, 5$. Since H is normal it contains $\beta = (543)\alpha(543)^{-1}$, and hence $\beta\alpha^{-1}$. Notice that $(543)\alpha$ sends 2 to 5 and $\alpha(543)$ sends 2 to 3, so $\beta\alpha^{-1} \neq 1$. Now $\beta\alpha^{-1}$ fixes every element that α fixes, and also fixes 2; so $\beta\alpha^{-1}$ fixes more elements that α. This contradicts our choice of α. □

Exercise. Show that $S_3 \cong \mathrm{GL}_2(\mathbb{F}_2)$.

5.4 Actions

Often a group acts as permutations of a set.

This is typically how groups arise: as certain kinds of permutations of a set with structure, and the group consists of permutations that preserve the structure. This is roughly what we mean when we speak of a symmetry group.

We will exploit this perspective in this section.

A permutation of a set X is a bijective map $X \to X$. The set $S(X)$ of all permutations of X is a group under composition of maps. If X has n elements we call the set of all permutations of X the symmetric group on n letters, and denote it by S_n.

An action of a group G on a set X is a group homomorphism $G \to S(X)$. If $g \in G$ and $x \in X$ we write $g.x$ for the image of x under the action of g. Thus $1.x = x$, and $g.(h.x) = (gh).x$ for all $x \in X$, and all $g, h \in G$.

Example 4.1. 1. *The* dihedral group *D_n of order $2n$, ($n \geq 3$). Let V be the set of vertices of a regular n-gon P. Consider all rigid motions of P that send V to V. Think of P as a piece of wood that you may pick up, rotate, turn over, and put down again so that it is in its original position, i.e., the vertices are placed*

on top of the original vertices. The set of all such motions is called the dihedral group D_n. *If we think of D_n acting on V we see that it is a subgroup of S_n.*

Label the vertices $1, 2, \ldots, n$ in a clockwise sequence. The position of the P after a rigid motion is determined by the new position of 1 after the motion and whether the vertices are now labelled in the clockwise or counter-clockwise order. Hence D_n has $2n$ elements.

If clockwise rotation by $2\pi/n$ radians is denoted by τ, and the flip about some fixed axis is denoted by σ, then D_n is generated by σ and τ. Now $\tau^n = \sigma^2 = 1$, and $\sigma\tau\sigma^{-1} = \tau^{-1}$. It is not hard to convince oneself that D_n consists of the $2n$ distinct elements $\{\tau^i, \sigma\tau^i \mid 0 \le i \le n-1\}$.

For example, D_3 acts on the vertices of an equilateral triangle, and $D_3 \cong S_3$.

2. *The* general linear group *$GL_n(k)$ or $GL(n,k)$ is the group of all invertible k-linear maps $k^n \to k^n$.*

3. *If K/k is a Galois extension, then $\mathrm{Gal}(K/k)$ acts on K. It also acts on the intermediate fields lying between k and K. If $K = k(\alpha)$, then $\mathrm{Gal}(K/k)$ permutes the zeroes of the minimal polynomial of α.*

This is the historical origin of groups: the ancients considered the permutations of the zeroes of a polynomial.

4. *A group G acts on itself by left multiplication, $g.x = gx$. If $|G| = n$, this action gives a homomorphism $G \to S_n$.*

Notice that the action of G on itself by right multiplication is not always an action according to our definition because then $g.(h.x) = (hg).x$. (It is an action if G is commutative.) However, if we define $g.x = xg^{-1}$, this is an action of G on itself.

5. *A group G acts on itself by conjugation: $g.x = gxg^{-1}$. This action defines a group homomorphism $G \to \mathrm{Aut}\, G$, the automorphism group of G. The kernel of this is*

$$\{g \in G \mid gx = xg \text{ for all } x \in G\} = Z(G),$$

the center of G. The image of this homomophism is called the group of inner automorphisms of G, and is denoted by $\mathrm{Inn}(G)$.

6. *A group G acts on the set of its subgroups by congugation, $g.H = gHg^{-1} = \{ghg^{-1} \mid h \in H\}$. Subgroups H and H' are said to be* conjugate *if $H' = gHg^{-1}$ for som $g \in G$.*

7. *If H is a subgroup of G, then G acts on the set of left cosets of H by $g.xH = (gx)H$.*

8. *Suppose that G acts on a set X, and R is the ring of k-valued functions on X. Then there is an action of G on R by*

$$(g.f)(x) = f(g^{-1}.x) \qquad \text{for } g \in G,\ f \in R, \text{ and } x \in X.$$

It is essential to use g^{-1} here so one gets a left action of G on R. Notice that each g acts as an automorphism of the ring R. In fact, the automorphisms of a ring form a group, denoted by $\mathrm{Aut}\, R$, and this action of G on R can be interpreted as a group homomorphism $G \to \mathrm{Aut}\, R$. It is usual for X to have some additional structure and for the G action to preserve that structure and R to consist of functions that are related to that structure. For example, X

may be a topological space and R might be the ring of all continuous $\mathbb{R}$-valued (or $\mathbb{C}$-valued) functions on X; when X is a topological space we usually only consider continuous G-actions on X; i.e., the map $x \mapsto g.x$ is required to be continuous for all $g \in G$; in this case, $g.f$ is a continuous function whenever f is continuous so $g.f$ belongs to R if f does.

9. *Consider the previous example, but now suppose that X is an irreducible algebraic variety over a field k. Suppose further that the G action on X is such that the map $x \mapsto g.x$ is a morphism of varieties for each $g \in G$. Each morphism corresponds to a homomorphism of rings $\mathcal{O}(X) \to \mathcal{O}(X)$; explicitly it is $f \mapsto g^{-1}.f$. Hence we obtain a G-action on $\mathcal{O}(X)$. We extend this to an action of G on $k(X)$ in the natural way $g.(a/b) = (g.a)/(g.b)$. It makes sense to consider the invariants $k(X)^G$ and $\mathcal{O}(X)^G$.*

More ..... ◊

Definition 4.2. *Let G be a group acting on a set X. The* orbit *of $x \in X$ is $G.x = \{g.x \mid g \in G\}$. The* stabilizer *of x is* $\mathrm{Stab}_G(x) = \{g \in G \mid g.x = x\}$.

Notice that the stabilizer of x is a subgroup of G.

The orbits partition X; X is the disjoint union of its orbits. This provides an equivalence relation on X,

$$x \sim y \Leftrightarrow y \in G.x \Leftrightarrow G.x = G.y.$$

Example 4.3. *We may view the general linear group $GL_n(k)$ as the set of invertible $n \times n$ matrices. It therefore acts on the space of $n \times n$ matrices $M_n(k)$ by $g.A = gAg^{-1}$. This is an important example and it motivates a lot of mathematics. The finer aspects of it are a topic for current research.*

What are the orbits? In each orbit find a "nice" element. Jordan normal form, which we discuss next quarter, gives such a nice element, and is useful in answering other questions about this action.

Proposition 4.4. *Let G be a finite group acting on a set X. If $x \in X$, then*

1. $|G| = |G.x| \times |Stab_G x|$;
2. $|G.x| = |G : \mathrm{Stab}_G x|$;
3. $|G.x|$ *divides* $|G|$.

Proof. Let $g, h \in G$, and set $S = \mathrm{Stab}_G x$. Then

$$g.x = h.x \Leftrightarrow g^{-1}h.x = x \Leftrightarrow g^{-1}h \in S \Leftrightarrow gS = hS.$$

Therefore the map

$$\phi : \{\text{left cosets of } S \text{ in } G\} \longrightarrow Gx, \qquad \phi(gG_x) := gx$$

is a well-defined bijection. Thus $|G.x|$ is equal to the number of left cosets of S in G; that number is $|G : S| = |G|/|S|$. □

The following is a trivial consequence, but its triviality belies its significance.

Lemma 4.5 (The Orbit Formula)**.** *Let G be a finite group acting on a finite set X. Let $X_1, \ldots, X_n$ be the distinct G-orbits in X, and for each i choose $x_i \in X_i$. Then*

$$|X| = \sum_{i=1}^{n} |X_i| = \sum_{i=1}^{n} |G : \mathrm{Stab}_G(x_i)|.$$

Notice that if G acts on X and x and y belong to the same orbit, then their stabilizers are conjugate: if $y = g.x$ and $S = \mathrm{Stab}_G(x)$, then $\mathrm{Stab}_G(y) = gSg^{-1}$. Conversely, if two subgroups of G are conjugate to one another, then one is a stabilizer if and only if the other is.

Definition 4.6. *The* conjugacy class *of $x \in G$ is*

$$C_G(x) = \{gxg^{-1} \mid g \in G\},$$

and the centralizer *of $x \in G$ is*

$$Z_G(x) = \{g \in G \mid gx = xg\}.$$

These are, respectively, the orbit and the stabilizer of x under the action of G on itself by conjugation.

The center of a group is denoted by $Z(G)$; by definition it consists of those elements z such that $zg = gz$ for all $g \in G$. Notice that the center of a group consists of exactly those elements whose conjugacy classes have size one.

The next result follows at once from the definition and Proposition 4.4.

Lemma 4.7. *Let x be an element of a finite group G. Then*

$$|G| = |Z_G(x)| \times |C_G(x)|.$$

In particular, the number of conjugates of x equals $|G : Z_G(x)|$, which divides $|G|$.

Proposition 4.8 (The Class Formula)**.** *Let G be a finite group and let $C_1, \ldots, C_n$ be the distinct conjugacy classes in G. Then*

1. *$|G| = \sum_{i=1}^{n} |C_i|$;*
2. *If $Z(G)$ denotes the center of G, then*

$$|G| = |Z(G)| + \sum_{|C_i|>1} |C_i|. \tag{4-2}$$

Proof. Since G is the disjoint union of its orbits, $|G| = \sum_{i=1}^{n} |C_i|$ and this can be written as

$$|G| = \sum_{|C_i|=1} |C_i| + \sum_{|C_i|>1} |C_i|.$$

However, $Z(G)$ is the disjoint union of those C_i having cardinality one. □

Theorem 4.9 (Cauchy's Theorem)**.** *Let p be a prime. If p divides the order of G, then G has an element of order p.*

Proof. Let $\mathbb{Z}/p$ act on

$$X := \{(x_1, \ldots, x_p) \mid x_i \in G,\ x_1 \cdots x_p = 1\} \subseteq G^p$$

by cyclic permutations, i.e., fix a generator ξ for $\mathbb{Z}/p$ and define

$$\xi \cdot (x_1, \ldots, x_p) = (x_p, x_1, \ldots, x_{p-1}).$$

The stabilizer of a point in X is a subgroup of $\mathbb{Z}/p$ so is either $\mathbb{Z}/p$ or the trivial subgroup. The number of elements in an orbit is therefore either 1 or p. An element in X is completely determined by its first $p-1$ terms, which can be anything, so $|X| = |G^{p-1}|$. In particular, $|X|$ is divisible by p. Since $|X|$ is the sum of sizes of the distinct orbits, and non-trivial orbits have p elements, the number of orbits of size 1 is divisible by p. There is at least one orbits of size 1, namely $(1, \ldots, 1)$, so there are at least p orbits of size 1. Every orbit of size 1 is of the form $(a, \ldots, a)$ for some $a \in G$. Hence there is $a \in G - \{1\}$ such that $a^p = 1$. □

Another proof of Cauchy's Theorem. (This proof uses the classification of finite abelian groups.)

We argue by induction on the order of G. If G has a proper subgroup whose order is divisible by p, we may apply the induction hypothesis to that subgroup to obtain the result. So, we may assume G has no such subgroup.

If $x \in G - Z(G)$, then p does not divide the order of the proper subgroup $Z_G(x)$, so divides $|C_G(x)|$. It follows from (4-2) that p divides $|Z(G)|$. Hence $Z(G) = G$; thus G is abelian.

Suppose that G and $\{1\}$ are the only subgroups of G. Let $x \in G - \{1\}$. Then $G = \langle x \rangle$, the subgroup generated by x; hence the order of x is $|G|$, which equals pn, so x^n has order p.

Now suppose that G and $\{1\}$ are not the only subgroups of G. We can choose a proper subgroup H of largest possible order. Then H must be a maximal subgroup and, if $|H| = m$, then $(p, m) = 1$. By the induction hypothesis applied to G/H, there is an element $x \in G - H$ such that $x^p \in H$. Since H is maximal, $G = \langle H, x \rangle$, and $|G| = pm$. If $x^m = 1$, choose $a, b \in \mathbb{Z}$ such that $ap + bm = 1$. Then

$$x = x^{ap+bm} = (x^m)^a (x^p)^b \in H,$$

contradicting our choice of x. Hence $x^m \neq 1$, and

$$(x^m)^p = x^{mp} = x^{|G|} = 1,$$

so x^m has order p.

5.4.1 Small groups

A fundamental problem that has driven the development of finite group theory since its infancy is that of classification: classify all finite groups.

Here "small" does not mean that $|G|$ is small, though that is a reasonable place to begin, but that $|G|$ is a product of a small number of primes. For example, if p is prime and $|G| = p$, then $G \cong \mathbb{Z}_p$. The next result is the first step towards classifying groups of order p^n. I believe the classification of groups of order p^n is still not solved; I don't even know if it is reasonable to hope for some sort of classification.

Proposition 4.10. *If the order of a group is a power of a prime, then its center is non-trivial; i.e., it contains a non-identity element.*

Proof. Suppose that $|G| = p^r$. If $C_1, \ldots, C_m$ are the conjugacy classes with more than one element, then $|C_i| = p^{r_i}$ for some $r_i > 1$. It therefore follows from (4-2) that p divides $|Z(G)|$. □

Proposition 4.11. *If p is prime and $|G| = p^2$, then G is isomorphic to either $\mathbb{Z}_p \times \mathbb{Z}_p$ or $\mathbb{Z}/p^2$.*

Proof. Let G be a group with p^2 elements. If G is abelian, then G is isomorphic to either $\mathbb{Z}_p \times \mathbb{Z}_p$ or $\mathbb{Z}/p^2$. We will show that G *must* be abelian. By Proposition 4.10, $Z(G) \neq \{1\}$. If $Z(G) = G$, then G is abelian. The remaining alternative is that $|Z(G)| = p$. However, in that case $G/Z(G)$ has size p so is isomorphic to $\mathbb{Z}_p$ and is there generated by a single element. It follows that G is generated by two elements, say z which generates $Z(G)$ and $y \in G - Z(G)$. But x commutes with y so the group $\langle x, y\rangle$ is abelian, i.e., G is abelian. □

A group of order p^3 need not be abelian. The quaternion group,

$$Q := \{\pm 1, \pm i, \pm j, \pm k\} \subset \mathbb{H}^\times,$$

where $\mathbb{H}^\times$ is Hamilton's group of non-zero quaternions, has size 8 and is not abelian.

There is another non-abelian group with 8 elements? What is it? Show it is not isomorphic to the quaternion group.

You might like to think of the next two cases, groups of size p^3 for $p = 3$ and $p = 5$. Can you classify them.

The next case is $|G| = pq$ where p and q are distinct primes.

Proposition 4.12. *Let p and q be primes with $p > q$. If G is a group with pq elements, then G has a unique subgroup of order p and that subgroup is normal.*

Proof. By Cauchy's Theorem, G has an element of order p. Let N be the subgroup it generates.

Suppose N is not normal. Then G has another subgroup of order p, $M = gNg^{-1}$ for some $g \in G$. Both G and M act on $X := \{gNg^{-1} \mid g \in G\}$ by conjugation, $g \cdot N = gNg^{-1}$ We have $|G| = |\mathrm{Orb}_G(N)| \times |\mathrm{Stab}_G(N)|$. But $N \subseteq$

$\mathrm{Stab}_G(N) = N_G(N) \subsetneqq G$ so $\mathrm{Stab}_G(N) = N$. Therefore $|\mathrm{Orb}_G(N)| = q$. Hence $|\mathrm{Orb}_M(N)| \leq q < p = |M|$. But $|\mathrm{Orb}_M(N)|$ divides $|M|$ so $|\mathrm{Orb}_M(N)| = 1$. Hence $M \subseteq \mathrm{Stab}_G(N) = N$. It follows that $M = N$. This contradiction implies that N is a normal subgroup of G.

Suppose M is another subgroup of order p. Let $x \in M - \{1\}$. Then x has order p. If x were in N it would generate both N and M which is absurd. Hence $x \in N$. The image of x in G/N, which is isomorphic to $\mathbb{Z}/q$, is not the identity element so its order must be q. But the order of the image of x in G/N divides the order of x; i.e., q divides p. That is absurd so we conclude that N is the only subgroup of G having p elements. □

Theorem 4.13. *Let p and q be primes such that q divides $p-1$. Then there is a unique non-abelian group of order pq, namely the semi-direct product*

$$\langle \sigma, \tau \mid \sigma^p = \tau^q = 1,\ \tau\sigma\tau^{-1} = \sigma^s \text{ where } s^q \equiv 1 (\mathit{mod}\, p) \rangle. \tag{4-3}$$

Proof. By Proposition 1.2, $\mathrm{Aut}\,\mathbb{Z}_p \cong \mathbb{Z}_{p-1}$. By Cauchy's Theorem, $\mathbb{Z}_{p-1}$ has an element of order q. Hence there is a non-trivial homomorphism $\phi : \mathbb{Z}_q \to \mathrm{Aut}(\mathbb{Z}_p)$ and therefore a non-abelian group $\mathbb{Z}_p \rtimes_\phi \mathbb{Z}_q$ having pq elements.

Let G be a non-abelian group of order pq. Let N be the normal subgroup of G having p elements. Then $G \cong N \rtimes (G/N) \cong \mathbb{Z}_p \rtimes_\psi \mathbb{Z}_q$ for some non-trivial homomorphism $\psi : \mathbb{Z}_q \to \mathrm{Aut}(\mathbb{Z}_p)$.

Since $\mathbb{Z}_{p-1}$ is cyclic it has a *unique* subgroup with q elements. Therefore $\psi = \phi\beta$ for some $\beta \in \mathrm{Aut}(\mathbb{Z}_q)$. By Proposition 2.2,

$$\mathbb{Z}_p \rtimes_\phi \mathbb{Z}_q \cong \mathbb{Z}_p \rtimes_{\phi\beta} \mathbb{Z}_q = \mathbb{Z}_p \rtimes_\psi \mathbb{Z}_q.$$

Define the elements $\sigma := (1,0)$ and $\tau := (0,1)$ in $\mathbb{Z}_p \rtimes_\phi \mathbb{Z}_q$. We will use multiplicative notation for the product in $\mathbb{Z}_p \rtimes_\phi \mathbb{Z}_q$; this might be a little confusing because we use additive notation for $\mathbb{Z}_p$ and $\mathbb{Z}_q$; thus, the product $(a,b)(c,d)$ of two elements in $\mathbb{Z}_p \rtimes_\phi \mathbb{Z}_q$ is $(a + \phi(c)(b), b + d)$. Now $\sigma^p = \tau^q = 1$ and

$$\tau\sigma\tau^{-1} = (0,1)(1,0)(0,-1) = (\phi(1)(1), 0) = (s,0) = \sigma^s$$

for some integer s. However, $\tau^q = 1$ so $\sigma = \tau^q\sigma\tau^{-q} = \sigma^{s^q}$. Therefore s^q is equal to 1 modulo p. □

Proposition 4.14. *Let p and q be distinct primes and G a group with pq elements. Without loss of generality, assume $q < p$. If q does not divide $p-1$, then $G \cong \mathbb{Z}_p \times \mathbb{Z}_q \cong \mathbb{Z}_{pq}$.*

Proof. By Proposition 4.12, G has a normal subgroup, N say, with p elements. Hence $G \cong \mathbb{Z}_p \rtimes_\phi \mathbb{Z}_q$ for some $\phi \in \mathrm{Aut}(\mathbb{Z}_p)$ with the property that $\phi^q = \mathrm{id}_{\mathbb{Z}_p}$. Since $\mathrm{Aut}\,\mathbb{Z}_p \cong \mathbb{Z}_{p-1}$ and q does not divide $p-1$, $\phi = \mathrm{id}_{\mathbb{Z}_p}$. Hence $G \cong \mathbb{Z}_{pq}$. □

Theorem 4.13 is quite useful for classifying groups of small order. For example, making a table of small primes p, and primes q dividing $p-1$, we obtain

the following

p	q	pq
3	2	6
5	2	10
7	2, 3	14, 21
11	2, 5	22, 55
13	2, 3	26, 39
17	2,	34
19	2, 3	38, 57
23	2, 11	46, 253
29	2, 7	58, 206

Hence we have described the unique non-abelian group of the orders appearing in the right-hand column. We now know the structure of all groups of order $n \leq 30$ except for $n \in \{8, 12, 16, 18, 20, 24, 27, 28, 30\}$.

Classifying those groups is a nice problem.

5.5 The Sylow Theorems

Let p be a prime. A p-group is a group in which the order of every element is a power of p. p-groups are of great importance, and it is an interesting problem to describe their structure. By Cauchy's Theorem, the order of a finite p-group is a power of p so has a non-trivial center by Proposition 4.10.

Finite p-groups are the building blocks for finite groups.

Definition 5.1. *Let G be a finite group. If the order of G is $p^n t$ where $(p, t) = 1$, a p-*Sylow subgroup *of G is a subgroup of order p^n.*

Let H be a subgroup of G. Its normalizer in G, denoted $N_G(H)$, is largest subgroup of G that contains H as a normal subgroup.

If we let G act on its subgroups by conjugation, then $N_G(H) = \mathrm{Stab}_G H$. Hence the number of conjugates of H in G is $|G : N_G(H)| = |G|/|N_G(H)|$.

Theorem 5.2 (Sylow's First Theorem, 1872)**.** *Let G be a finite group of order $p^n t$, where $(p, t) = 1$. Let H be a subgroup of G of order p^i with $i < n$. Then there exists a subgroup K of G such that*

1. *H is normal in K, and*
2. *$|K| = p^{i+1}$.*

In particular, p-Sylow subgroups exist, and every p-subgroup of G is contained in a p-Sylow subgroup.

Proof. We will show that p divides the order of the group $N_G(H)/H$; Cauchy's Theorem will then provide an element $x \in N_G(H) - H$ such that $x^p \in H$, whence $\langle H, x \rangle$ will be the desired K.

Let H act on $X = \{aH \mid a \in G\}$ by left multiplication. Since $|X| = |G|/|H|$, p divides $|X|$. But

$$|X| = \sum_{\text{distinct orbits}} |H.x| = \sum_{\text{distinct orbits}} |H.(aH)|$$

and $|H.aH| = |H : \mathrm{Stab}_H(aH)| = p^j$ for some $j \leq i$. Therefore the number of orbits of size one is divisible by p.

Notice that $|H.aH| = 1$ if and only if $h.aH = aH$ for all $h \in H$, if and only if $a^{-1}ha \in H$ for all $h \in H$, if and only if $a \in N_G(H)$, if and only if $aH \in N_G(H)/H$. Therefore the number of orbits of size one is $|N_G(H)/H|$; so this is divisible by p as claimed. □

Lemma 5.3. *Let P be a p-Sylow subgroup of G, and suppose that $x \in G$ has order p^i. If $xPx^{-1} = P$, then $x \in P$. In particular, if P is normal in G, then it is the only p-Sylow subgroup.*

Proof. Suppose to the contrary that $x \notin P$. Then P is normal in the strictly larger group $\langle P, x\rangle$. But $\langle P, x\rangle/P$ is a cyclic group generated by $\bar{x}$, the image of x, so its order is equal to the order of $\bar{x}$, which must divide the order of x, and so is equal to p^m for some $m \geq 1$. But this implies that p^m divides $|G|/|P|$ which is false since P is a p-Sylow subgroup. Hence x belongs to P, as claimed.

Now suppose that P is normal in G. If Q is another p-Sylow subgroup, pick $x \in Q$. The order of x is a power of p, so the first part of the lemma applies, showing that $x \in P$. Hence $Q \subset P$. But $|Q| = |P|$, so $Q = P$. □

Theorem 5.4 (Sylow's Second and Third Theorems)**.** *Let G be a finite group.*

1. *Any two p-Sylow subgroups of G are conjugate.*
2. *The number of p-Sylow subgroups is of the form $np + 1$.*
3. *The number of p-Sylow subgroups divides $|G|$.*

Proof. Let P be a p-Sylow subgroup of G, and let $X = \{P = P_0, P_1, \ldots, P_t\}$ be the distinct conjugates of P in G.

Let P act on X by conjugation; that is, $x.P_i = xP_ix^{-1}$ for $x \in P$. One orbit is $P = P_0$ and, by Lemma 5.3, this is the only orbit consisting of one element. The orbit of P_i has size $|P : \mathrm{Stab}_P(P_i)|$, so is divisible by p if $i \neq 0$. But $|X| = \sum |\text{distinct orbits}|$, so $|X| = np + 1$ for some $n \in \mathbb{N}$.

Suppose Q is a p-Sylow subgroup of G that does not belong to X. Let Q act on X by conjugation. There is no orbit of size one because $xP_ix^{-1} = P_i$ for all $x \in Q$, then $Q \subset P_i$ by Lemma 5.3. Hence, by the same argument as in the previous paragraph, p divides $|X|$; this contradicts the previous paragraph, so we conclude that X consists of all P-Sylow subgroups. This proves (1) and (2).

Let G act on X by conjugation. There is only one orbit, so $|X| = |G : \mathrm{Stab}_G(P)|$ which divides $|G|$, so (3) holds. □

5.6 Using Sylow's Theorems

Sylow's theorems are the rock on which all deeper analysis of finite groups is built. This section contains some illustrative examples.

Theorem 6.1. *Let p and q be primes such that $q|p-1$. Then there is a unique non-abelian group of order pq, namely*

$$\langle \sigma, \tau \mid \sigma^p = \tau^q = 1,\ \tau\sigma\tau^{-1} = \sigma^s \text{ where } s^q \equiv 1 (mod\, p) \rangle. \tag{6-4}$$

Proof. Suppose that G is a non-abelian group of order pq. The number of p-Sylow subgroups is congruent to $1(\mathrm{mod})p$ and divides pq, so there is a unique p-Sylow subgroup, say N, and it is therefore normal in G. Let H be any q-Sylow subgroup of G. Then $H \cap N = \{1\}$ and $NH = G$. Hence $G \cong N \rtimes H$. The result now follows from Proposition 2.5 and Corollary **??**. □

Example 6.2. *If $|G| = 28$, then G is not simple. Since $28 = 2^2.7$, and the number of 7-Sylow subgroups is $\equiv 1 (mod 7)$ and divides 28, there is exactly one 7-Sylow subgroup. Since all 7-Sylow subgroups are conjugate, that 7-Sylow subgroup must be normal.* ◊

Example 6.3. *If $|G| = 56 = 2^3.7$, then G is not simple. The number of 7-Sylow subgroups divides 56 and is congruent to $1 (mod 7)$, so is either 1 or 8. If there were only one 7-Sylow subgroup it would be a normal subgroup, so G would not be simple.*

Supose there were eight 7-Sylow subgroups. Since a 7-Sylow subgroup is isomorphic to $\mathbb{Z}_7$, the intersection of two distinct 7-Sylow subgroups would equal $\{1\}$. Hence the union, say V, of the eight distinct 7-Sylow subgroups contains $1 + 8 \times 6 = 49$ elements. Now let P be a 2-Sylow subgroup. None of the eight elements in P has order 7, so $P \cap V = \{1\}$, whence $V \cup P = G$. It follows that P is the unique 2-Sylow subgroup, and is therefore normal in G. ◊

Lemma 6.4. *If $|G| = p^n q$ where p and q are distinct primes with $q < p$, then G is not simple.*

Proof. The number of p-Sylow subgroups is $\equiv 1 (\mathrm{mod}\, p)$ and divides q, so must be one. That p-Sylow subgroup is therefore normal. □

A possible project in a first course on group theory is to find all simple groups of order ≤ 100. The previous lemma allows one to exclude quite a lot of the possibilities

q	p	there is no simple group of order
2	3	$6, 18, 54$
2	5	$10, 50$
2	7	$14, 98$
2	≥ 11	$22, 26, 34, 38, 46, 58, 62, 74, 86, 94$
3	5	$15, 75$
3	≥ 7	$21, 33, 39, 51, 57, 69, 87, 93$
5	≥ 7	$35, 55, 65, 85, 95$

Lemma 6.5. *If H is a subgroup of G of index two, then H is normal in G.*

Proof. Because G is the disjoint union of the cosets of H we have

$$G = H \sqcup Hx = H \sqcup xH,$$

where $x \notin H$. Thus $xH = xH$ and $xHx^{-1} = H$. □

We now use Sylow's theorems to prove that $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$. Our proof will use two facts:

1. if the degree of $f \in \mathbb{C}[x]$ is two, then f splits in $\mathbb{C}$;
2. if the degree of $f \in \mathbb{R}[x]$ is odd, then f has a zero in $\mathbb{R}$.

Theorem 6.6 (The Fundamental Theorem of Algebra)**.** *Every polynomial in $\mathbb{C}[x]$ is a product of linear polynomials. That is, $\mathbb{C}$ is algebraically closed.*

Proof. It is enough to show that if $\mathbb{R} \subset \mathbb{C} \subset K$ is a finite normal extension of $\mathbb{R}$, then $K = \mathbb{C}$.

Write $G = \mathrm{Gal}(K/\mathbb{R})$, and $|G| = [K : \mathbb{R}] = 2^n s$ with s odd.

Let P be a 2-Sylow subgroup of G, and let K^P be the fixed field of P. Then $[K^P : \mathbb{R}] = |G|/|P| = s$. Since s is odd, $[\mathbb{R}(\alpha) : \mathbb{R}]$ is odd for all $\alpha \in K^P$. Hence the minimal polynomial of α over $\mathbb{R}$ has odd degree, and therefore has a zero in $\mathbb{R}$. But the minimal polynomial is also irreducible, so it has degree one. Hence $\alpha \in \mathbb{R}$, and therefore $K^P = \mathbb{R}$, and $s = 1$.

It follows that $[K : \mathbb{C}] = 2^{n-1}$. Because K is normal over $\mathbb{R}$ it is normal, and hence Galois, over $\mathbb{C}$. Write $G' = \mathrm{Gal}(K/\mathbb{C})$. Thus $|G'| = 2^{n-1}$. If $K \neq \mathbb{C}$, then Sylow's First Theorem provides a subgroup H of G' of index two. Hence $[K^H : \mathbb{C}] = |G'|/|H| = 2$. But this contradicts the fact (1). □

Theorem 6.7. *Let $(A, +)$ be a finite abelian group of order $n = p_1^{r_1} \cdots p_n^{r_n}$ where the p_is are distinct primes and each r_i is ≥ 1. For each prime p, A has a unique p-Sylow subgroup, namely*

$$A_p := \{a \in A \mid \textit{the order of } a \textit{ is } p^j \textit{ for some } j\},$$

and $A = A_{p_1} \oplus \cdots \oplus A_{p_n}$.

Proof. Since A is a abelian all its subgroups are normal, so for each prime p dividing n there is a unique p-Sylow subgroup. Each element of that p-Sylow subgroup has order a power of p and, conversely, every element of A having order a power of p belongs to a p-Sylow (Sylow's First Theorem). Hence the p-Sylow subgroups are the sugroups A_p.

To see that the sum of the A_ps is direct suppose that $0 = a_1 + \cdots + a_n$ with each $a_i \in A_{p_i}$. If some $a_i \neq 0$, we can assume after relabelling that $a_1 \neq 0$. Write $m = n/p_1^{r_1}$. Then $0 = m.0 = ma_1$; but this implies that m divides $p_1^{r_1}$ which is absurd.

It remains to show that the sum of all the A_ps is A. Let $0 \neq a \in A$. The order of a is of the form $m = p_1^{s_1} \cdots p_n^{s_n}$ for suitable integers s_i. Set

$d_i = m/p^{s_i}$. Then $\gcd(d_1, \ldots, d_n) = 1$. Hence there exist integers $t_1, \ldots, t_n$ such that $t_1 q_1 + \cdots + t_n d_n = 1$. Thus $a = (t_1 q_1)a + \cdots + (t_n d_n)a$. The order of $(t_i d_i)a = (t_i m/p^{s_i})a$ is a power of p_i, so $(t_i d_i)a \in A_{p_i}$. Hence $a \in A_{p_1} + \cdots + A_{p_n}$. □

5.7 Simple Groups

One of the outstanding algebraic achievements of the 20^{th} century is the classification of finite simple groups. Recall that a group is simple if its only normal subgroups are itself and $\{1\}$. If G is any finite group there is a chain of subgroups

$$G = G_0 \supset G_1 \supset \cdots \supset G_n = \{1\}$$

such that each G_{i+1} is normal in G_i and G_i/G_{i+1} is simple. To construct such a chain start by choosing G_1 to be a maximal normal subgroup of G, and take G_2 to be a maximal normal subgroup of G_1, and so on. Thus one sees that simple groups are the building blocks of all finite groups and a classification of all finite groups might proceed by finding all the simple ones and then understanding how a simple group can be glued on top of another group.

The precise problem that encapsulates the last step is as follows: Fix two finite groups N and H and classify the groups G that contain a copy of N as a normal subgroup such that $G/N \cong H$. This is a hard problem. It contains, at the very least, the problem of classifying semi-direct products $N \rtimes H$.

Let's think look for small simple groups. First there are the cyclic groups $\mathbb{Z}_p$, p prime. Then we can run through the integers starting at 4 and try to use the results in the previous section to eliminate various numbers n, i.e., find n such that there is not a simple group of size n.

Proposition 7.1. *If G is a simple group with 60 elements, then $G \cong A_5$.*

Proof. Let G be a simple group of size 60. Notice first that $60 = 2^2.3.5$.

Claim: If G has a subgroup of index 5, then $G \cong A_5$. Proof: If H is a subgroup of index 5, then G acts on the set of left cosets of H and they form a single orbit; the action of G on this set of 5 cosets provides a non-trivial group homomorphism $G \to S_5$ which is injective because G is simple. So we can think of G as a subgroup of S_5, and since $|S_5| = 120$, G is of index two and hence normal in S_5. By Proposition 3.8, $G = A_5$. ◊

Now we will prove G has a subgroup of index 5.

Consider first the 5-sylow subgroups. There must be six of them and the intersection of any two of them is $\{1\}$. Their union has $6 \times (5-1) = 24$ elements of order 5.

Now consider the 2-sylow subgoups. The possibilities for their number is 1, 3, 5, and 15. There can't be just one because it would then be a normal subgroup. Neither can there be just three of them because then the action of G by conjugation on the set of 2-Sylow subgroups would provide a non-trivial homomorphism $G \to S_3$ and, because $|G| > |S_3|$, the kernel of this map would

be a proper normal subgroup of G. Hence there are either 5 or 15 2-sylow subgroups.

If there are five 2-sylow subgroups they form a single orbit under G acting by conjugation and hence we obtain a homomorphism $G \to S_5$. Now argue as before that the image of this map is A_5.

Now suppose there are 15 2-sylow subgroups. If the intersection of any two of these is $\{1\}$, then the union of the 2-sylow subgroups contains $15 \times (2^2 - 1) = 45$ elements of order either two or four. But $45+24 > 60$, so this cannot be the case. Hence there are two 2-sylow subgroups, say H and K such that $H \cap K \neq \{1\}$. Now $H \cap K$ is central in H and K and hence in the subgroup they generate, $\langle H, K\rangle$. The center of a group is always a normal subgroup, so the simplicity of G implies that $\langle H, K\rangle \neq G$. Since H is a proper subgroup of $\langle H, K\rangle$, and $4 = |H|$ divides $|\langle H, K\rangle|$, we conclude that $|\langle H, K\rangle| \geq 8$; of course it can't be 8 because 8 does not divide 60, so $|\langle H, K\rangle| \geq 12$. If the order of this were > 12 its index would be $d < 5$ and the action of G on the cosets of $\langle H, K\rangle$ would provide a homomorphism $G \to S_d$ which would have a kernel. That can't happen, so we conclude that $|\langle H, K\rangle| = 12$, and hence G has a subgroup of index 5. □

The next smallest simple group (apart from the cyclic ones) has order 168. It is $\mathrm{PSL}(2, 7)$ the projective special linear group of 2×2 matrices over $\mathbb{F}_7$.

Definition 7.2. *For any field k and integer $n \geq 1$, the* special linear group $\mathrm{SL}(n, k)$ *consists of the $n \times n$ matrices over k of determinant one. The* projective special linear group *is*

$$\mathrm{PSL}(n, k) := \mathrm{SL}(n, k)/\mathit{center}$$

◊

If $n > 1$, $\mathrm{SL}(n, k)$ is never simple because it has non-trivial center: its center consists of those matrices ξI where $\xi \in k$ is an n^{th} root of unity.

The center of $\mathrm{SL}(2, 7)$ is $\{\pm 1\}$ so $|\mathrm{PSL}(2, 7)| = \frac{1}{2}|\mathrm{SL}(2, 7)|$. Let's count: there are $7^2 - 1 = 48$ choices for the first column of $g \in \mathrm{GL}(2, 7)$; then there are $7^2 - 7$ choices for the second column. Hence $|\mathrm{GL}(2, 7)| = (7^2 - 1)(7^2 - 7)$. Now, multiplying the first column of $g \in \mathrm{GL}(2, 7)$ by a non-zero $\xi \in \mathbb{F}_7$ produces a new element of $\mathrm{GL}(2, 7)$ whose determinant is ξ times $\det g$. Hence, there is a unique ξ such that the new matrix has determinant one. Hence

$$|\mathrm{SL}(2, 7)| = \frac{|\mathrm{GL}(2, 7)|}{|\mathbb{F}_7 - \{0\}|} = (7^2 - 1) \times 7.$$

It follows at once that $|\mathrm{PSL}(2, 7)| = 168$.

Theorem 7.3. *If $p > 4$, then $\mathrm{PSL}(2, p)$ is simple.*

Generally speaking, the more factors $|G|$ has the greater the range of possibilities for G.

Proposition 7.4. *There is not a simple group with 144 elements.*

Proof. Let G be a simple group with 144 elements. Since $144 = 2^4 \times 3^2$, the number of 2-Sylow subgroups is either 1, 3, or 9, and the number of 3-Sylow subgroups is either 1, 4, or 16. Let X be the set of 2-Sylow subgroups and Y the set of 3-Sylow subgroups. The actions of G on X and Y by conjugation give homomorphisms from G to the symmetric groups on $|X|$ and $|Y|$ elements. Since 3! and 4! are smaller than 144, G must have 9 2-Sylow subgroups and 16 3-Sylow subgroups.

If $P \cap Q = \{1\}$ for all pairs of different 3-Sylow subgroups P and Q, then the union of the 3-Sylow subgroups would have $16 \times (9-1) + 1 = 129$ elements; by Sylow's first theorem every element of order 3 belongs to a 3-Sylow subgroup, so there are $144 - 129 = 15$ elements of order a power of 2. But the 2-Sylow subgroups have 16 elements so this would force there to be a unique, and therefore normal, 2-Sylow subgroup. This does not happen so there are 3-Sylow subgroups P and Q such that $P \cap Q \neq \{1\}$. But P and Q have 3^2 elements so are abelian. Hence the elements in $P \cap Q$ commute with all elements in $\langle P, Q\rangle$, the subgroup generated by P and Q. In particular, $\langle P, Q\rangle \subset N_G(P \cap Q)$. The order of $N_G(P \cap Q)$ is a multiple of 9 and divides 144, so the index of $N_G(P \cap Q)$ in G is either 1, 2, 4, or 8. It can't be 1, 2, or 4 because then the action of G by left multiplication on the cosets of $N_G(P \cap Q)$ would give a homomorphism from G to a symmetric group with 1, 2, or 24, elements and that homomorphism would have a kernel. That doesn't happen so $[G : N_G(P \cap Q)] = 8$.

Hence $|N_G(P \cap Q)| = 2 \times 3^2$. However, a group with 2×3^2 has a unique 3-Sylow subgroup. But $N_G(P \cap Q)$ contains both P and Q.

This contradiction implies that G can not be simple. □

5.8 Solvable groups

Definition 8.1. *A group G is* solvable *if there exists a finite chain of subgroups*

$$G = G_0 \supset G_1 \supset \cdots \supset G_n = \{1\}$$

such that each G_{i+1} is a normal subgroup of G_i and G_i/G_{i+1} is abelian. We call such a chain of subgroups a solvable chain. ◊

If H is a finitely generated abelian group, then there is a finite chain of subgroups $H = H_0 \supset H_1 \supset \cdots \supset H_m = \{0\}$ such that each H_i/H_{i+1} is cyclic: if $H = \mathbb{Z}h_1 + \cdots + \mathbb{Z}h_m$, the chain of submodules $0 \subset \mathbb{Z}h_1 \subset \mathbb{Z}h_1 + \mathbb{Z}h_2 \subset \cdots$ has cyclic slices.

Hence, in the definition of a solvable group we can insist that the quotients G_i/G_{i+1} are cyclic.

Proposition 8.2. *Let N be a normal subgroup of a group G. Then G is solvable if and only if N and G/N are solvable.*

Proof. ($\Rightarrow$) Let $G = G_0 \supset G_1 \supset \cdots \supset G_n = \{1\}$ be a solvable chain.

Claim: the chain $N = N \cap G_0 \supset N \cap G_1 \supset \cdots \supset N \cap G_n = \{1\}$ is a solvable chain. Let $\pi_i : G_i \to G_i/G_{i+1}$ be the natural map. The restriction of π_i to

$N \cap G_i$ has kernel $N \cap G_{i+1}$, so $N \cap G_{i+1}$ is a normal subgroup of $N \cap G_i$, and the quotient $N \cap G_i / N \cap G_{i+1}$ is isomorphic to a subgroup of G_i/G_{i+1}, so is abelian.

Let $\psi : G \to G/N$ be the natural map. To show that G/N is solvable, we show that the chain $G/N = \psi(G_0) \supset \psi(G_1) \supset \cdots \supset \psi(G_n) = \{1\}$ is a solvable chain. Because G_{i+1} is a normal subgroup of G_i, $\psi(G_{i+1})$ is a normal subgroup of $\psi(G_i)$. The map $G_i \to \psi(G_i) \to \psi(G_i)/\psi(G_{i+1})$ is surjective and sends G_{i+1} to zero, so induces a surjective map $G_i/G_{i+1} \to \psi(G_i)/\psi(G_{i+1})$. Therefore $\psi(G_i)/\psi(G_{i+1})$ is abelian.

($\Leftarrow$) The map $H \mapsto H/N$ from subgroups of G containing N to subgroups of G/N is a bijection. Furthermore, H is normal in G if and only if H/N is normal in G/N, and if $H \supset K \supset N$, then $H/K \cong (H/N)/(K/N)$.

Thus, a solvable chain in G/N corresponds to a chain $G = G_0 \supset G_1 \supset \cdots \supset G_n = N$ of subgroups in G.

If $N = N_0 \supset N_1 \supset \cdots \supset N_m = \{1\}$ is a solvable chain in N, then

$$G = G_0 \supset G_1 \supset \cdots \supset G_n = N \supset N_1 \supset \cdots \supset N_m = \{1\}$$

is a solvable chain in G. □

The proof that G solvable implies N solvable did not use the fact that N is normal. In fact, every subgroup of a solvable group is solvable.

5.9 Some important groups

GL(n), $SL(n)$, $O(n)$, $SO(n)$, $SP(n)$, Spin groups, $PGL(n)$, $PGL(2, \mathbb{F}_7)$, $U(n)$ etc...

Tori, ...

Example 9.1. *Let p be a prime and $q = p^r$. We want to consider the finite general linear group*

$$G = \mathrm{GL}_n(\mathbb{F}_q).$$

First let's compute its order.

By its very construction $\mathrm{GL}_n(\mathbb{F}_q)$ acts on the n-dimensional vector space $V = \mathbb{F}_q^n$. We view elements of V as column vectors so that $\mathrm{GL}_n(\mathbb{F}_q)$ acts from the left by multiplication.

Fix an ordered basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ for V. Each $g \in \mathrm{GL}_n(\mathbb{F}_q)$ sends $\mathcal{B}$ to a new ordered basis $g.\mathcal{B} = \{g.v_1, \ldots, g.v_n\}$. Every ordered basis is of the form $g.\mathcal{B}$ for some g, and $g.\mathcal{B} = g'.\mathcal{B}$ if and only if $g = g'$. Thus the ordered bases form a single orbit under the action of $\mathrm{GL}_n(\mathbb{F}_q)$ and the stabilizer of each ordered basis is trivial. Hence

$$|\mathrm{GL}_n(\mathbb{F}_q)| = \textit{the number of ordered bases.}$$

Let's count the ordered bases. Choose $0 \neq v_1 \in V$. Since $|V| = q^n$ there are $q^n - 1$ possible choices for v_1. Having chosen v_1, choose v_2 such that $\{v_1, v_2\}$ is

linearly independent. Since v_2 may be any element of $V - \mathbb{F}_q v_1$, there are $q^n - q$ choices for v_2. Now choose v_3 such that $\{v_1, v_2, v_3\}$ is linearly independent; since v_3 may be any element in $V - \mathbb{F}_q v_1 + \mathbb{F}_q v_2$, there are $q^n - q^2$ choices for v_3. Continuing in this way we see that the number of possible ordered bases is

$$(q^n-1)(q^n-q)\cdots(q^n-q^{n-1}) = q^{\frac{1}{2}n(n-1)}(q^n-1)(q^{n-1}-1)\cdots(q-1) = |\operatorname{GL}_n(\mathbb{F}_q)|.$$

It follows from this that the order of a p-Sylow subgroup of $\operatorname{GL}_n(\mathbb{F}_q)$ is $q^{\frac{1}{2}n(n-1)}$. The order of the upper triangular subgroup

$$\begin{pmatrix} 1 & * & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ 0 & 0 & 1 & \cdots & * \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

is obviously $q^{n-1}q^{n-2}\cdots q^2.q = q^{\frac{1}{2}n(n-1)}$, so this is a p-Sylow subgroup. All other p-Sylow subgroups are conjugate to this one, so if N is another p-Sylow subgroup, there is a choice of basis for V in which N is this upper triangular subgroup. ♢

5.10 Fun with $\mathbb{F}_1$

Coming soon to a theater near you!

Fix a group G and a field k.

A k-linear representation of G is a k-vector space V together with a left action of G on V by k-linear maps. We may denote the action of $g \in G$ on $v \in V$ either by $g.v$ or, more formally, as $\rho(g)(v)$ where $\rho : G \to GL(V)$ is the group homomorphism determined by the action. Thus, we sometimes denote a representation by a pair (V, ρ).

Usually, k is understood so we simply speak of representations of G or, more briefly, of G-modules.

If U and V are G-modules, a linear map $\phi : U \to V$ is said to be a G-module homomorphism of G-equivariant if

$$\phi(g.u) = g.\phi(u)$$

for all $u \in U$ and $g \in G$.

There are some other aspects of G-actions that should be discussed briefly although they are not purely algebraic matters.

The basic point is that group actions abound, and one often wishes to describe the orbit spaces. Moreover, if the space X being acted on has some structure beyond being a set and the group G acts on X so as to preserve that structure (for example, X is a manifold and elements of G act as diffeomorphisms) one wants to put such a structure on the orbit space X/G too. That is not always possible, but one would still like to do the best possible.

For example, consider the group $\mathbb{Z}_2 = \{\pm 1\}$ acting on $\mathbb{R}^2$ with -1 acting as multiplication by -1. Is there a way to give the orbit space $\mathbb{R}^2/\mathbb{Z}_2$ the structure of a manifold so that the map $\mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}_2$ sending each point to its orbit is smooth? The simplest realization of the orbit space is as the cone $Q := \{(u, v, w) \in \mathbb{R}^3 \mid uw = v^2\}$, and the map $\mathbb{R}^2 \to Q$ defined by $(x, y) \mapsto (x^2, xy, z^2)$ is the quotient map. Of course the cone is not a submanifold of $\mathbb{R}^3$ because of the singularity at the cone.

Projective spaces. Perhaps the most important orbit spaces in geometry are the projective spaces, $\mathbb{RP}^n$ and $\mathbb{CP}^n$. We outline a construction of these as orbit spaces.

Consider a field k, and the action of the multiplicative group $(k^\times, \cdot)$ on the punctured vector space $k^{n+1} - \{0\}$ according to the rule

$$\lambda.(x_0, \ldots, x_n) = (\lambda x_0, \ldots, \lambda x_n).$$

The orbits are the punctured lines through the origin; they are in bijection with the 1-dimensional subspaces of k^{n+1}. The space of orbits is called the n-dimensional projective space over k, and is denoted by $\mathbb{P}^n_k$. Let's write $[x_0, \ldots, x_n]$ for the point in $\mathbb{P}^n_k$ that represents the orbit that is the line through $(x_0, \ldots, x_n)$.

Here is one motivation for introducing $\mathbb{P}^n$. You already know that the set of simultaneous solutions to some system of polynomial equations $f_1 = \cdots = f_r = 0$, where each $f_i \in k[x_0, \ldots, x_n]$, is an affine algebraic subvariety of $\mathbb{A}^{n+1} = k^{n+1}$. Let's write X for this subvariety. If each of the f_is is homogeneous in $x_0, \ldots, x_n$, then it is clear that $f_i(\lambda x_0, \ldots, \lambda x_n) = \lambda^{\deg f_i} f_i(x_0, \ldots, x_n)$, so that

X is a union of lines through the origin. Hence, we might as well understand the image of X in $\mathbb{P}^n_k$. The images of such Xs are called projective algebraic varieties and their study forms the subject matter of algebraic geometry.

There is a reason for preferring $\mathbb{P}^n$ to $\mathbb{A}^n$: it is compact. Compactness is a finiteness property and one always has better results for compact than for non-compact spaces. A paradigmatic example of this is Poincaré duality, a result for compact connected manifolds that has no suitable analogue without the compactness hypothesis.

If you want a little practice consider the action of $U(1) = \{z \in \mathbb{C} \mid |z| = 1$ on the unit 3-sphere S^3, realized as $\{(z_1, z_2) \in \mathbb{C}^2 \mid |z_1|^2 + |z_2|^2 = 1\}$ in $\mathbb{C}^2$, defined by $\lambda.(z_1, z_2) = (\lambda z_1, \lambda z_2)$. Show that the map $\pi : S^3 \to \mathbb{CP}^1$, the complex projective line, defined by $\pi(z_1, z_2) = [z_1, z_2]$ has fibers the $U(1)$-orbits, and hence that $\mathbb{CP}^1 \cong SS^3/U(1)$. You can view π as being a fibration over $\mathbb{CP}^1$ with fibers isomorphic to $U(1) \equiv S^1$. Show that $\mathbb{CP}^2 \cong S^2$, the 2-sphere. This is the first example of a Hopf fibration; there are similar ones $S^7 \to S^4$ and $S^{15} \to S^8$ arising from Hamilton's quaternions and Cayley's octonions.

Conjugacy classes of matrices. A standard result in a graduate algebra course is to classify $n \times n$ matrices up to conjugation over an algebraically closed field k. This is equivalent to classifying finite dimensional modules over $k[x]$. You can think of this as classifying the orbits when $\mathrm{GL}(n)$ acts by conjugation on $M_n(k)$ the set of $n \times n$ matrices. The result is encapsulated in the famous Jordan normal form.

A similar sounding problem is to classify pairs of commuting $n \times n$ matrices $\{(A, B) \mid AB = BA\}$ up to simultaneous conjugation $g.(A, B) = (gAg^{-1}, gBg^{-1})$. This is an unsolved problem. Get the message that classifying orbits is an important but generally hard problem.

Exercises.

1. Let β be an element of $\mathbb{F}_4$ that is not in $\mathbb{F}_2$.
   (a) Find the minimal polynomial of β over $\mathbb{F}_2$.
   (b) Show that $x^2 + (\beta + 1)x + 1$ is irreducible in $\mathbb{F}_4[x]$.
   (c) Is the cubic $x^3 + x^2 + \beta \in \mathbb{F}_4[x]$ irreducible? If not, find its factors.
   (d) Show that $\mathbb{F}_{16}$ contains an element α that is a primitive fifth root of one over $\mathbb{F}_2$, and that $\mathbb{F}_{16} = \mathbb{F}_2(\alpha)$. Find the minimal polynomial of α over $\mathbb{F}_4$, and show that α^4 is the other zero of this polynomial.
   (e) Show that α is a zero of $x^3 + x^2 + \beta \in \mathbb{F}_4[x]$.
   (f) Show that the Galois group of $x^5 - 1$ over $\mathbb{F}_4$ is $\mathbb{Z}_2$.
   (g) Factor $x^4 + x^3 + x^2 + x + 1$ over $\mathbb{F}_4$.
   (h) You have shown above that $\mathbb{F}_{16} = \mathbb{F}_4(\alpha)$ where α is a primitive fifth root of 1. Does there exist an element $\alpha \in \mathrm{Gal}(x^5 - 1/\mathbb{F}_4)$ such that $\sigma(\alpha) = \alpha^2$?