Galois Theory

Tom Leinster, University of Edinburgh

Version of 28 March 2022

Note to the reader

These are the 2021–22 course notes for Galois Theory (MATH10080).

Structure Each chapter corresponds to one week of the semester. You are expected to read Chapter *n* before the lectures in Week *n*, except for Chapter 1. I may make small changes to these notes as we go along (e.g. to correct errors), so I recommend that you download a fresh copy before you start each week's reading.

Exercises looking like this are sprinkled through the notes. The idea is that you try them immediately, before you continue reading.

Most of them are meant to be quick and easy, much easier than assignment or workshop questions. If you can do them, you can take it as a sign that you're following successfully. For those that defeat you, talk with others in the class, ask on Piazza, or ask me.

I promise you that if you make a habit of trying every exercise, you'll enjoy the course more and understand it better than if you don't.

Digressions like this are optional and not examinable, but might interest you. They're usually on points that *I* find interesting, and often describe connections between Galois theory and other parts of mathematics.

Here you'll see titles of relevant videos. These are entirely optional but may help your understanding.

References to theorem numbers, page numbers, etc., are clickable links.

What to prioritize You know by now that the most important things in almost any course are the *definitions* and the results called *Theorem*. But I also want to emphasize the *proofs*. This course presents a wonderful body of theory, and the idea is that you learn it all, including the proofs that are its beating heart.

A closed-book exam would test that by asking you to reproduce some proofs. Your exam will be open book, so it can't ask you to *reproduce* proofs, but it will test something arguably harder: that you *understand* them. So the proofs will need your attention and energy.

Compulsory prerequisites   To take this course, you must have already taken these two courses:

- Honours Algebra: We'll need some abstract linear algebra, corresponding to Chapter 1 of that course. We'll also need everything from Honours Algebra about rings and polynomials (Chapter 3 there), including ideals, quotient rings (factor rings), the universal property of quotient rings, and the first isomorphism theorem for rings.

- Group Theory: From that course, we'll need fundamentals such as normal subgroups, quotient groups, the universal property of quotient groups, and the first isomorphism theorem for groups. You should know lots about the symmetric groups S_n, alternating groups A_n, and cyclic groups C_n, as well as a little about the dihedral groups D_n, and I hope you can list all of the groups of order < 8 without having to think too hard.

  Chapter 8 of Group Theory, on solvable groups, will be crucial for us. For example, you'll need to understand what it means that S_4 is solvable but A_5 is not.

  We won't need anything on free groups, the Sylow theorems, or the Jordan–Hölder theorem.

If you're a visiting or MSc student and didn't take those courses, please contact me so that we can decide whether your background is suitable.

Mistakes   I'll be grateful to hear of mistakes in these notes (Tom.Leinster@ ed.ac.uk), even if it's something very small and even if you're not sure.

Chapter 1

Overview of Galois theory

This chapter stands apart from all the others,

Modern treatments of Galois theory take advantage of several well-developed branches of algebra: the theories of groups, rings, fields, and vector spaces. This is as it should be! However, assembling all the algebraic apparatus will take us several weeks, during which it's easy to lose sight of what it's all for.

Galois theory came from two basic insights:

- every polynomial has a symmetry group;

- this group determines whether the polynomial can be solved by radicals (in a sense I'll define).

In this chapter, I'll explain these two ideas in as short and low-tech a way as I can manage. In Chapter 2 we'll start again, beginning the modern approach that will take up the rest of the course. But I hope that all through that long build-up, you'll keep in mind the fundamental ideas you learn in this chapter.

1.1 The view of $\mathbb{C}$ from $\mathbb{Q}$

Imagine you lived several centuries ago, before the discovery of complex numbers. Your whole mathematical world is the real numbers, and there is no square root of -1. This situation frustrates you, and you decide to do something about it.

So, you invent a new symbol i (for 'imaginary') and decree that $i^2 = -1$. You still want to be able to do all the usual arithmetic operations ($+$, $\times$, etc.), and you want to keep all the rules that govern them (associativity, commutativity, etc.). So you're also forced to introduce new numbers such as $2 + 3 \times i$, and you end up with what today we call the complex numbers.

So far, so good. But then you notice something strange. When you invented the complex numbers, you only intended to introduce one square root of -1. But

4

accidentally, you introduced a second one at the same time: $-i$. (You wait centuries for a square root of -1, then two come along at once.) Maybe that's not so strange in itself; after all, positive reals have two square roots too. But then you realize something genuinely weird:

> *There's nothing you can do to distinguish i from −i.*

Try as you might, you can't find any reasonable statement that's true for i but not $-i$. For example, you notice that i is a solution of

$$z^3 - 3z^2 - 16z - 3 = \frac{17}{z},$$

but then you realize that $-i$ is too.

Of course, there are *unreasonable* statements that are true for i but not $-i$, such as '$z = i$'. We should restrict to statements that only refer to the known world of real numbers. More precisely, let's consider statements of the form

$$\frac{p_1(z)}{p_2(z)} = \frac{p_3(z)}{p_4(z)},$$

where p_1, p_2, p_3, p_4 are polynomials with *real* coefficients. Any such equation can be rearranged to give

$$p(z) = 0,$$

where again p is a polynomial with real coefficients, so we might as well just consider statements of that form. The point is that if $p(i) = 0$ then $p(-i) = 0$.

Let's make this formal. We could say that two complex numbers are 'indistinguishable when seen from $\mathbb{R}$' if they satisfy the same polynomials over $\mathbb{R}$. But the official term is 'conjugate':

Definition 1.1.1 Two complex numbers z and z' are **conjugate over** $\mathbb{R}$ if for all polynomials p with coefficients in $\mathbb{R}$,

$$p(z) = 0 \iff p(z') = 0.$$

For example, i and $-i$ are conjugate over $\mathbb{R}$. This follows from a more general result, stating that conjugacy in this new sense is closely related to complex conjugacy:

Lemma 1.1.2 *Let $z, z' \in \mathbb{C}$. Then z and z' are conjugate over $\mathbb{R}$ if and only if $z' = z$ or $z' = \overline{z}$.*

Proof 'Only if': suppose that z and z' are conjugate over $\mathbb{R}$. Write $z = x + iy$ with $x, y \in \mathbb{R}$. Then $(z - x)^2 + y^2 = 0$. Since x and y are real, conjugacy implies that $(z' - x)^2 + y^2 = 0$, so $z' - x = \pm iy$, so $z' = x \pm iy$.

'If': obviously z is conjugate to itself, so it's enough to prove that z is conjugate to $\bar{z}$. I'll give two proofs. Each one teaches us a lesson that will be valuable later.

First proof: recall that complex conjugation satisfies

$$\overline{w_1 + w_2} = \overline{w_1} + \overline{w_2}, \qquad \overline{w_1 \cdot w_2} = \overline{w_1} \cdot \overline{w_2}$$

for all $w_1, w_2 \in \mathbb{C}$. Also, $\bar{a} = a$ for all $a \in \mathbb{R}$. It follows by induction that for any polynomial p over $\mathbb{R}$,

$$\overline{p(w)} = p(\overline{w})$$

for all $w \in \mathbb{C}$. So

$$p(z) = 0 \iff \overline{p(z)} = \bar{0} \iff p(\bar{z}) = 0.$$

Second proof: write $z = x + iy$ with $x, y \in \mathbb{R}$. Let p be a polynomial over $\mathbb{R}$ such that $p(z) = 0$. We will prove that $p(\bar{z}) = 0$. This is trivial if $y = 0$, so suppose that $y \neq 0$.

Consider the real polynomial $m(t) = (t - x)^2 + y^2$. Then $m(z) = 0$. You know from Honours Algebra that

$$p(t) = m(t)q(t) + r(t) \tag{1.1}$$

for some real polynomials q and r with $\deg(r) < \deg(m) = 2$ (so r is either a constant or of degree 1). Putting $t = z$ in (1.1) gives $r(z) = 0$. It's easy to see that this is impossible unless r is the zero polynomial (using the assumption that $y \neq 0$). So $p(t) = m(t)q(t)$. But $m(\bar{z}) = 0$, so $p(\bar{z}) = 0$, as required.

We have just shown that for all polynomials p over $\mathbb{R}$, if $p(z) = 0$ then $p(\bar{z}) = 0$. Exchanging the roles of z and $\bar{z}$ proves the converse. Hence z and $\bar{z}$ are conjugate over $\mathbb{R}$. $\qquad\square$

Exercise 1.1.3 Both proofs of 'if' contain little gaps: 'It follows by induction' in the first proof, and 'it's easy to see' in the second. Fill them.

Digression 1.1.4 With complex analysis in mind, we could imagine a stricter definition of conjugacy in which polynomials are replaced by arbitrary convergent power series (still with coefficients in $\mathbb{R}$). This would allow functions such as exp, cos and sin, and equations such as $\exp(i\pi) = -1$.

But this apparently different definition of conjugacy is, in fact, equivalent. A complex number is still conjugate to exactly itself and its complex conjugate. (For example, $\exp(-i\pi) = -1$ too.) Do you see why?

Lemma 1.1.2 tells us that conjugacy over $\mathbb{R}$ is rather simple. But the same idea becomes much more interesting if we replace $\mathbb{R}$ by $\mathbb{Q}$. And in this course, we will mainly focus on polynomials over $\mathbb{Q}$.

Define **conjugacy over** $\mathbb{Q}$ by replacing $\mathbb{R}$ by $\mathbb{Q}$ in Definition 1.1.1. Again, when you see the words 'conjugate over $\mathbb{Q}$', you can think to yourself 'indistinguishable when seen from $\mathbb{Q}$'. From now on, I will usually just say 'conjugate', dropping the 'over $\mathbb{Q}$'.

Example 1.1.5 I claim that $\sqrt{2}$ and $-\sqrt{2}$ are conjugate. And I'll give you two different proofs, closely analogous to the two proofs of the 'if' part of Lemma 1.1.2.

First proof: write

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

Example 1.1.5

For $w \in \mathbb{Q}(\sqrt{2})$, there are *unique* $a, b \in \mathbb{Q}$ such that $w = a + b\sqrt{2}$, because $\sqrt{2}$ is irrational. So it is logically valid to define

$$\widetilde{w} = a - b\sqrt{2} \in \mathbb{Q}(\sqrt{2}).$$

(Question: what did the uniqueness of a and b have to do with the logical validity of that definition?) Now, $\mathbb{Q}(\sqrt{2})$ is closed under addition and multiplication, and it is straightforward to check that

$$\widetilde{w_1 + w_2} = \widetilde{w_1} + \widetilde{w_2}, \qquad \widetilde{w_1 \cdot w_2} = \widetilde{w_1} \cdot \widetilde{w_2}$$

for all $w_1, w_2 \in \mathbb{Q}(\sqrt{2})$. Also, $\widetilde{a} = a$ for all $a \in \mathbb{Q}$. So just as in the proof of Lemma 1.1.2, it follows that w and $\widetilde{w}$ are conjugate for every $w \in \mathbb{Q}(\sqrt{2})$. In particular, $\sqrt{2}$ is conjugate to ('indistinguishable from') $-\sqrt{2}$.

Second proof: let $p = p(t)$ be a polynomial with coefficients in $\mathbb{Q}$ such that $p(\sqrt{2}) = 0$. You know from Honours Algebra that

$$p(t) = (t^2 - 2)q(t) + r(t)$$

for some polynomials $q(t)$ and $r(t)$ over $\mathbb{Q}$ with $\deg r < 2$. Putting $t = \sqrt{2}$ gives $r(\sqrt{2}) = 0$. But $\sqrt{2}$ is irrational and $r(t)$ is of the form $at + b$ with $a, b \in \mathbb{Q}$, so r must be the zero polynomial. Hence $p(t) = (t^2 - 2)q(t)$, giving $p(-\sqrt{2}) = 0$.

We have just shown that for all polynomials p over $\mathbb{Q}$, if $p(\sqrt{2}) = 0$ then $p(-\sqrt{2}) = 0$. The same argument with the roles of $\sqrt{2}$ and $-\sqrt{2}$ reversed proves the converse. Hence $\pm\sqrt{2}$ are conjugate.

> **Exercise 1.1.6** Let $z \in \mathbb{Q}$. Show that z is not conjugate to z' for any complex number $z' \neq z$.

One thing that makes conjugacy more subtle over $\mathbb{Q}$ than over $\mathbb{R}$ is that over $\mathbb{Q}$, more than two numbers can be conjugate:
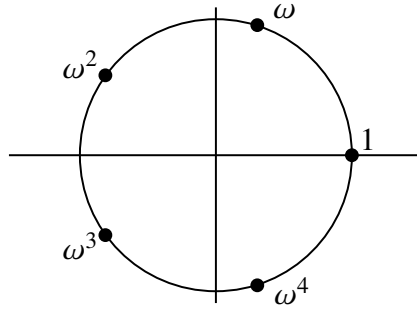
Figure 1.1: The 5th roots of unity.

Example 1.1.7 The 5th roots of unity are

$$1, \omega, \omega^2, \omega^3, \omega^4,$$

where $\omega = e^{2\pi i/5}$ (Figure 1.1). Now 1 is not conjugate to any of the rest, since it is a root of the polynomial $t - 1$ and the others are not. (See also Exercise 1.1.6.) But it turns out that $\omega, \omega^2, \omega^3, \omega^4$ are all conjugate to each other.

Complex conjugate numbers are conjugate over $\mathbb{R}$, so they're certainly conjugate over $\mathbb{Q}$. (If you've got a pair of complex numbers that you can't tell apart using only the reals, you certainly can't tell them apart using only the rationals.) Since $\omega^4 = 1/\omega = \overline{\omega}$, it follows that ω and ω^4 are conjugate over $\mathbb{Q}$. By the same argument, ω^2 and ω^3 are conjugate. What's not so obvious is that ω and ω^2 are conjugate. I know two proofs, which are like the two proofs of Lemma 1.1.2 and Example 1.1.5. But we're not equipped to do either yet.

Example 1.1.8 More generally, let p be any prime and put $\omega = e^{2\pi i/p}$. Then $\omega, \omega^2, \ldots, \omega^{p-1}$ are all conjugate to one another.

So far, we have asked when *one* complex number can be distinguished from another, using only polynomials over $\mathbb{Q}$. But what about more than one?

Definition 1.1.9 Let $k \geq 0$ and let $(z_1, \ldots, z_k)$ and $(z'_1, \ldots, z'_k)$ be k-tuples of complex numbers. Then $(z_1, \ldots, z_k)$ and $(z'_1, \ldots, z'_k)$ are **conjugate over** $\mathbb{Q}$ if for all polynomials $p(t_1, \ldots, t_k)$ over $\mathbb{Q}$ in k variables,

$$p(z_1, \ldots, z_k) = 0 \iff p(z'_1, \ldots, z'_k) = 0.$$

When $k = 1$, this is just the earlier definition of conjugacy.

Example 1.1.11 For any $z_1, \ldots, z_k \in \mathbb{C}$, the k-tuples $(z_1, \ldots, z_k)$ and $(\overline{z_1}, \ldots, \overline{z_k})$ are conjugate. For let $p(t_1, \ldots, t_k)$ be a polynomial over $\mathbb{Q}$. Then

$$\overline{p(z_1, \ldots, z_k)} = p(\overline{z_1}, \ldots, \overline{z_k})$$

since the coefficients of p are real, by a similar argument to the one in the first proof of Lemma 1.1.2. Hence

$$p(z_1, \ldots, z_k) = 0 \iff p(\overline{z_1}, \ldots, \overline{z_k}) = 0,$$

which is what we had to prove.

Example 1.1.12 Let $\omega = e^{2\pi i/5}$, as in Example 1.1.7. Then

$$(\omega, \omega^2, \omega^3, \omega^4) \quad \text{and} \quad (\omega^4, \omega^3, \omega^2, \omega)$$

are conjugate, by Example 1.1.11. It can also be shown that

$$(\omega, \omega^2, \omega^3, \omega^4) \quad \text{and} \quad (\omega^2, \omega^4, \omega, \omega^3)$$

are conjugate, although the proof is beyond us for now. But

$$(\omega, \omega^2, \omega^3, \omega^4) \quad \text{and} \quad (\omega^2, \omega, \omega^3, \omega^4) \tag{1.2}$$

are *not* conjugate, since if we put $p(t_1, t_2, t_3, t_4) = t_2 - t_1^2$ then

$$p(\omega, \omega^2, \omega^3, \omega^4) = 0 \neq p(\omega^2, \omega, \omega^3, \omega^4).$$

Warning 1.1.13 The converse of Exercise 1.1.10 is false: just because z_i and z'_i are conjugate for all i, it doesn't follow that $(z_1, \ldots, z_k)$ and $(z'_1, \ldots, z'_k)$ are conjugate. For we saw in Example 1.1.7 that ω, ω^2, ω^3 and ω^4 are all conjugate to each other, but we just saw that the 4-tuples (1.2) are not conjugate.

1.2 Every polynomial has a symmetry group. . .

We are now ready to describe the first main idea of Galois theory: every polynomial has a symmetry group.

Definition 1.2.1 Let f be a polynomial with coefficients in $\mathbb{Q}$. Write $\alpha_1, \ldots, \alpha_k$ for its distinct roots in $\mathbb{C}$. The **Galois group** of f is

$$\mathrm{Gal}(f) = \{\sigma \in S_k : (\alpha_1, \ldots, \alpha_k) \text{ and } (\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)}) \text{ are conjugate}\}.$$

'Distinct roots' means that we ignore any repetition of roots: e.g. if $f(t) = t^5(t-1)^9$ then $k = 2$ and $\{\alpha_1, \alpha_2\} = \{0, 1\}$.

Exercise 1.2.2 Show that $\mathrm{Gal}(f)$ is a subgroup of S_k. (Hint: if you permute the variables of a polynomial, you get another polynomial.)

Digression 1.2.3 I brushed something under the carpet. The definition of $\mathrm{Gal}(f)$ depends on the order in which the roots are listed. Different orderings gives different subgroups of S_k. However, these subgroups are all *conjugate* to each other (conjugacy in the sense of group theory!), and therefore isomorphic as abstract groups. So $\mathrm{Gal}(f)$ is well-defined as an abstract group, independently of the choice of ordering.

Example 1.2.4 Let f be a polynomial over $\mathbb{Q}$ whose complex roots $\alpha_1, \ldots, \alpha_k$ are all rational. If $\sigma \in \mathrm{Gal}(f)$ then $\alpha_{\sigma(i)}$ and α_i are conjugate for each i, by Exercise 1.1.10. But since they are rational, that forces $\alpha_{\sigma(i)} = \alpha_i$ (by Exercise 1.1.6), and since $\alpha_1, \ldots, \alpha_k$ are distinct, $\sigma(i) = i$. Hence $\sigma = \mathrm{id}$. So the Galois group of f is trivial.

Example 1.2.5 Let f be a quadratic over $\mathbb{Q}$. If f has rational roots then as we have just seen, $\mathrm{Gal}(f)$ is trivial. If f has two non-real roots then they are complex conjugate, so $\mathrm{Gal}(f) = S_2$ by Example 1.1.11. The remaining case is where f has two distinct roots that are real but not rational, and it can be shown that in that case too, $\mathrm{Gal}(f) = S_2$.

Warning 1.2.6 On terminology: note that just now I said 'non-real'. Sometimes people casually say 'complex' to mean 'not real'. But try not to do this yourself. It makes as little sense as saying 'real' to mean 'irrational', or 'rational' to mean 'not an integer'.

Example 1.2.7 Let $f(t) = t^4 + t^3 + t^2 + t + 1$. Then $(t-1)f(t) = t^5 - 1$, so f has roots $\omega, \omega^2, \omega^3, \omega^4$ where $\omega = e^{2\pi i/5}$. We saw in Example 1.1.12 that

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \in \mathrm{Gal}(f), \qquad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix} \notin \mathrm{Gal}(f).$$

In fact, it can be shown that

$$\mathrm{Gal}(f) = \left\langle \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \right\rangle \cong C_4.$$

Example 1.2.8 Let $f(t) = t^3 + bt^2 + ct + d$ be a cubic over $\mathbb{Q}$ with no rational roots. Then

$$\text{Gal}(f) \cong \begin{cases} A_3 & \text{if } \sqrt{-27d^2 + 18bcd - 4c^3 - 4b^3d + b^2c^2} \in \mathbb{Q}, \\ S_3 & \text{otherwise.} \end{cases}$$

This appears as Proposition 22.4 in Stewart, but is way beyond us for now. Calculating Galois groups is hard.

Galois groups, intuitively

1.3   . . . which determines whether it can be solved

Here we meet the second main idea of Galois theory: the Galois group of a polynomial determines whether it can be solved. More exactly, it determines whether the polynomial can be 'solved by radicals'.

To explain what this means, let's begin with the quadratic formula. The roots of a quadratic $at^2 + bt + c$ are

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

After much struggling, it was discovered that there is a similar formula for cubics $at^3 + bt^2 + ct + d$: the roots are given by

$$\frac{\sqrt[3]{-27a^2d+9abc-2b^3+3a\sqrt{3(27a^2d^2-18abcd+4ac^3+4b^3d-b^2c^2)}} + \sqrt[3]{-27a^2d+9abc-2b^3-3a\sqrt{3(27a^2d^2-18abcd+4ac^3+4b^3d-b^2c^2)}}}{3\sqrt[3]{2}a}.$$

(No, you don't need to memorize that!) This is a complicated formula, and there's also something strange about it. Any nonzero complex number has three cube roots, and there are two $\sqrt[3]{}$ signs in the formula (ignoring the $\sqrt[3]{2}$ in the denominator), so it looks as if the formula gives *nine* roots for the cubic. But a cubic can only have three roots. What's going on?

It turns out that some of the nine aren't roots of the cubic at all. You have to choose your cube roots carefully. Section 1.4 of Stewart's book has much more on this point, as well as an explanation of how the cubic formula was obtained. We won't be going into this ourselves.

As Stewart also explains, there is a similar but even more complicated formula for quartics (polynomials of degree 4).

Digression 1.3.1 Stewart doesn't actually write out the explicit formula for the cubic, let alone the much worse one for the quartic. He just describes algorithms by which they can be solved. But if you unwind the algorithm for the cubic, you get the formula above. I have done this exercise once and do not recommend it.

11

Once mathematicians discovered how to solve quartics, they naturally looked for a formula for quintics (polynomials of degree 5). But it was eventually proved by Abel and Ruffini, in the early 19th century, that there is *no* formula like the quadratic, cubic or quartic formula for polynomials of degree ≥ 5. A bit more precisely, there is no formula for the roots in terms of the coefficients that uses only the usual arithmetic operations $(+, -, \times, \div)$ and kth roots (for integers k).

Spectacular as this result was, Galois went further—and so will we.

Informally, let us say that a complex number is **radical** if it can be obtained from the rationals using only the usual arithmetic operations and kth roots. For example,

$$\frac{\frac{1}{2} + \sqrt[3]{\sqrt[3]{2} - \sqrt[2]{7}}}{\sqrt[4]{6 + \sqrt[5]{\frac{2}{3}}}}$$

is radical, whichever square root, cube root, etc., we choose. A polynomial over $\mathbb{Q}$ is **solvable (or soluble) by radicals** if all of its complex roots are radical.

Example 1.3.2 Every quadratic over $\mathbb{Q}$ is solvable by radicals. This follows from the quadratic formula: $(-b \pm \sqrt{b^2 - 4ac})/2a$ is visibly a radical number.

Example 1.3.3 Similarly, the cubic formula shows that every cubic over $\mathbb{Q}$ is solvable by radicals. The same goes for quartics.

Example 1.3.4 *Some* quintics are solvable by radicals. For instance,

$$(t - 1)(t - 2)(t - 3)(t - 4)(t - 5)$$

is solvable by radicals, since all its roots are rational and, therefore, radical. A bit less trivially, $(t - 123)^5 + 456$ is solvable by radicals, since its roots are the five complex numbers $123 + \sqrt[5]{-456}$, which are all radical.

What determines whether a polynomial is solvable by radicals? Galois's amazing achievement was to answer this question completely:

Theorem 1.3.5 (Galois) *Let f be a polynomial over $\mathbb{Q}$. Then*

$$f \text{ is solvable by radicals} \iff \text{Gal}(f) \text{ is a solvable group}.$$

Example 1.3.6 Definition 1.2.1 implies that if f has degree n then $\text{Gal}(f)$ is isomorphic to a subgroup of S_n. You saw in Group Theory that S_4 is solvable, and that every subgroup of a solvable group is solvable. Hence the Galois group of any polynomial of degree ≤ 4 is solvable. It follows from Theorem 1.3.5 that every polynomial of degree ≤ 4 is solvable by radicals.

Example 1.3.7 Put $f(t) = t^5 - 6t + 3$. Later we'll show that $\mathrm{Gal}(f) = S_5$. You saw in Group Theory that S_5 is *not* solvable. Hence f is not solvable by radicals.

If there was a quintic formula then *all* quintics would be solvable by radicals, for the same reason as in Examples 1.3.2 and 1.3.3. But since this is not the case, there is no quintic formula.

Galois's result is much sharper than Abel and Ruffini's. They proved that there is no formula providing a solution by radicals of *every* quintic, whereas Galois found a way of determining *which* quintics (and higher) can be solved by radicals and which cannot.

> **Digression 1.3.8** From the point of view of modern numerical computation, this is all a bit odd. Computationally speaking, there is probably not much difference between solving $t^5 + 3 = 0$ to 100 decimal places (that is, finding $\sqrt[5]{-3}$) and solving $t^5 - 6t + 3 = 0$ to 100 decimal places (that is, solving a polynomial that isn't solvable by radicals). Numerical computation and abstract algebra have different ideas about what is easy and what is hard!

<div align="center">*   *   *</div>

This completes our overview of Galois theory. What's next?

Mathematics increasingly emphasizes *abstraction* over *calculation*. Individual mathematicians' tastes vary, but the historical trend is clear. In the case of Galois theory, this means dealing with *abstract algebraic structures*, principally fields, instead of manipulating *explicit algebraic expressions* such as polynomials. The cubic formula already gave you a taste of how hairy that can get.

Developing Galois theory using abstract algebraic structures helps us to see its connections to other parts of mathematics, and also has some fringe benefits. For example, we'll solve some notorious geometry problems that perplexed the ancient Greeks and remained unsolved for millennia. For that and many other things, we'll need some of the theory of groups, rings and fields—and that's what's next.

Chapter 2

Group actions, rings and fields

We now start again. This chapter is a mixture of revision and material that is likely to be new to you. The revision is from Fundamentals of Pure Mathematics, Honours Algebra, and Introduction to Number Theory (if you took it, which I won't assume). Because much of it is revision, it's a longer chapter than usual.

2.1 Group actions

Let's begin with a definition from Fundamentals of Pure Mathematics (Figure 2.1).

Definition 2.1.1 Let G be a group and X a set. An **action** of G on X is a function $G \times X \to X$, written as $(g, x) \mapsto gx$, such that

$$(gh)x = g(hx)$$

for all $g, h \in G$ and $x \in X$, and

$$1x = x$$

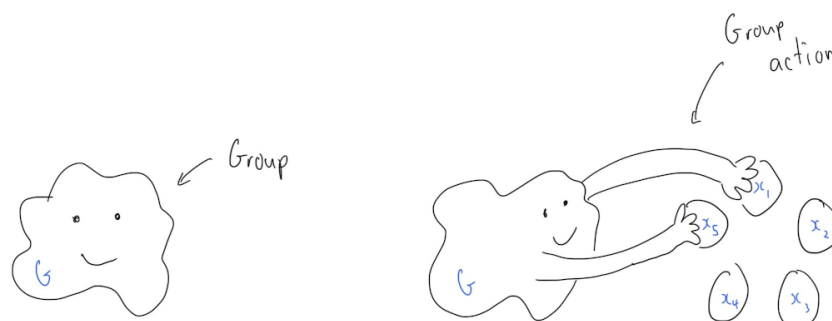for all $x \in X$. Here 1 denotes the identity element of G.

Figure 2.1: Action of a group G on a set X. (Image adapted from @rowvector.)

14

Examples 2.1.2   i. Let X be a set. There is a group $\mathrm{Sym}(X)$ whose elements are the bijections $X \to X$, with composition as the group operation and the identity function $\mathrm{id}_X \colon X \to X$ as the identity of the group. When $X = \{1, \ldots, n\}$, this group is nothing but S_n.

There is an action of $\mathrm{Sym}(X)$ on X defined by

$$
\begin{array}{rcl}
\mathrm{Sym}(X) \times X & \to & X \\
(g, x) & \mapsto & g(x).
\end{array}
$$

Acting on X is what $\mathrm{Sym}(X)$ was born to do!

ii. Similar examples can be given for many kinds of mathematical object, not just sets. Generally, an **automorphism** of an object X is an isomorphism $X \to X$ (preserving whatever structure X has), and the automorphisms of X form a group $\mathrm{Aut}(X)$ under composition. It acts on X just as in (i): $gx = g(x)$, for $g \in \mathrm{Aut}(X)$ and $x \in X$.

For instance, when X is a real vector space, the linear automorphisms form a group $\mathrm{Aut}(X)$ which acts on the vector space X. When X is finite-dimensional, we can describe this action in more concrete terms. Writing $n = \dim X$, the vector space X is isomorphic to $\mathbb{R}^n$, whose elements we will view as column vectors. The group $\mathrm{Aut}(X)$ is isomorphic to the group of $n \times n$ real invertible matrices under multiplication, usually called $\mathrm{GL}_n(\mathbb{R})$ ('general linear' group). Under these isomorphisms, the action of $\mathrm{Aut}(X)$ on X becomes

$$
\begin{array}{rcl}
\mathrm{GL}_n(\mathbb{R}) \times \mathbb{R}^n & \to & \mathbb{R}^n \\
(M, \mathbf{v}) & \mapsto & M\mathbf{v},
\end{array}
$$

where $M\mathbf{v}$ is the usual matrix product.

iii. Let G be the 48-element group of isometries (rotations and reflections) of a cube. Then G acts on the 6-element set of faces of the cube: any isometry maps faces to faces. It also acts in a similar way on the 12-element set of edges, the 8-element set of vertices, and a little less obviously, the 4-element set of long diagonals. (The **long diagonals** are the lines between a vertex and its opposite, furthest-away, vertex.)

iv. For any group G and set X, the **trivial action** of G on X is given by $gx = x$ for all g and x. Nothing moves anything!

Take an action of a group G on a set X. Every group element g gives rise to a function

$$
\bar{g} \colon X \to X
$$

defined by
$$\bar{g}(x) = gx.$$

In fact, $\bar{g}$ is a bijection, because $\overline{g^{-1}}$ is the inverse function of $\bar{g}$. So $\bar{g} \in \mathrm{Sym}(X)$ for each $g \in G$. For instance, consider the usual action of the isometry group G of the cube on the set X of faces (Example 2.1.2(iii)). If g is a particular isometry, then $\bar{g}$ is whatever permutation of the set of faces the isometry induces.

We have just seen that whenever G acts on X, every element g of the group G gives rise to an element $\bar{g}$ of the group $\mathrm{Sym}(X)$. So, we have defined a function

$$\Sigma: \quad G \quad \to \quad \mathrm{Sym}(X)$$
$$g \quad \mapsto \quad \bar{g}.$$

You can check that Σ is a group homomorphism.

Exercise 2.1.3 Check that $\bar{g}$ is a bijection for each $g \in G$. Also check that Σ is a homomorphism.

In summary: any action of a group G on X gives rise to a homomorphism $G \to \mathrm{Sym}(X)$, in a natural way.

Examples 2.1.4   i. Let X be a set, and consider the action of $\mathrm{Sym}(X)$ on X described in Example 2.1.2(i). For each $g \in \mathrm{Sym}(X)$, the function $\bar{g}: X \to X$ is just g itself. Hence the homomorphism $\Sigma: \mathrm{Sym}(X) \to \mathrm{Sym}(X)$ is the identity.

ii. Similarly, take a real vector space X and consider the action of $\mathrm{Aut}(X)$ on X described in Example 2.1.2(ii). The resulting homomorphism $\Sigma: \mathrm{Aut}(X) \to \mathrm{Sym}(X)$ is the inclusion; that is, $\Sigma(g) = g$ for all $g \in \mathrm{Aut}(X)$. (The domain of Σ is the group of *linear* bijections $X \to X$, whereas the codomain is the group of *all* bijections $X \to X$.)

iii. Consider the usual action of the isometry group G of the cube on the set X of edges (Example 2.1.2(iii)). Since X has 12 elements, $\mathrm{Sym}(X) \cong S_{12}$, and Σ amounts to a homomorphism $G \to S_{12}$.

iv. The trivial action of a group G on a set X (Example 2.1.2(iv)) corresponds to the trivial homomorphism $G \to \mathrm{Sym}(X)$.

Remark 2.1.5 When X is finite, we often choose an ordering of its elements, writing $X = \{x_1, \dots, x_k\}$. Then $\mathrm{Sym}(X) \cong S_k$ (assuming the x_is are all distinct). For each $g \in G$ and $i \in \{1, \dots, k\}$, the element gx_i of X must be equal to x_j for some j. Write that j as $\sigma_g(i)$, so that

$$gx_i = x_{\sigma_g(i)}.$$

Then $\sigma_g \in S_k$, and the composite homomorphism

$$G \xrightarrow{\ \Sigma\ } \mathrm{Sym}(X) \cong S_k$$

is $g \mapsto \sigma_g$.

Digression 2.1.6 In fact, an action of G on X is *the same thing as* a homomorphism $G \to \mathrm{Sym}(X)$. What I mean is that there is a natural one-to-one correspondence between actions of G on X and homomorphisms $G \to \mathrm{Sym}(X)$. Some books even *define* an action of G on X to be a homomorphism $G \to \mathrm{Sym}(X)$.

In detail: we've just seen how an action of G on X gives rise to a homomorphism $\Sigma \colon G \to \mathrm{Sym}(X)$. In the other direction, take any homomorphism $\Sigma \colon G \to \mathrm{Sym}(X)$. Define a function $G \times X \to X$ by

$$(g, x) \mapsto \big(\Sigma(g)\big)(x).$$

(To make sense of the right-hand side: $\Sigma(g)$ is an element of the group $\mathrm{Sym}(X)$, which is the set of bijections $X \to X$, so we can apply the function $\Sigma(g)$ to the element x to obtain another element $(\Sigma(g))(x)$ of X.) You can check that this function $G \times X \to X$ is an action of G on X. So, we've now seen how to convert an action into a homomorphism and vice versa. These two processes are mutually inverse. Hence actions of G on X correspond one-to-one with homomorphisms $G \to \mathrm{Sym}(X)$.

At the purely set-theoretic level (ignoring the group structures), the key is that for any sets A, B and C, there's a natural bijection

$$C^{A \times B} \cong (C^B)^A.$$

Here C^B means the set of functions $B \to C$. The general proof is very similar to what we've just done (where $A = G$ and $B = C = X$). In words, a function $A \times B \to C$ can be seen as a way of assigning to each element of A a function $B \to C$. In a picture:
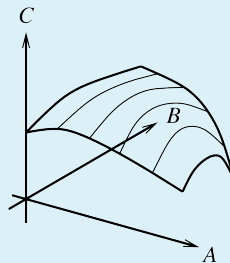
Here $A = B = C = \mathbb{R}$. By slicing up the surface as shown, a function $\mathbb{R}^2 \to \mathbb{R}$ can be seen as a function from $\mathbb{R}$ to {functions $\mathbb{R} \to \mathbb{R}$}.

Definition 2.1.7 An action of a group G on a set X is **faithful** if for $g, h \in G$,

$$gx = hx \text{ for all } x \in X \Longrightarrow g = h.$$

Faithfulness means that if two elements of the group *do* the same, they *are* the same. Here are some other ways to express it.

Lemma 2.1.8 *For an action of a group G on a set X, the following are equivalent:*

  i. the action is faithful;

  ii. for $g \in G$, if $gx = x$ for all $x \in X$ then $g = 1$;

  iii. the homomorphism $\Sigma \colon G \to \mathrm{Sym}(X)$ is injective;

  iv. $\ker \Sigma$ is trivial.

Proof Faithfulness states that whenever $g, h \in G$ with $\bar{g} = \bar{h}$, then $g = h$. But $\Sigma(g) = \bar{g}$, so (i) $\Longleftrightarrow$ (iii). Similarly, (ii) $\Longleftrightarrow$ (iv). Finally, it is a standard fact that a homomorphism is injective if and only if its kernel is trivial, so (iii) $\Longleftrightarrow$ (iv). $\square$

Many common actions are faithful:

Examples 2.1.9   i. The natural action of $\mathrm{Sym}(X)$ on a set X (Examples 2.1.2(i) and 2.1.4(i)) is faithful, since the corresponding homomorphism id$\colon \mathrm{Sym}(X) \to \mathrm{Sym}(X)$ is injective.

  ii. Similarly, the natural action of $\mathrm{Aut}(X)$ on a vector space (Examples 2.1.2(ii) and 2.1.4(ii)) is faithful, since the corresponding homomorphism $\mathrm{Aut}(X) \to \mathrm{Sym}(X)$ is injective.

  iii. The action of the isometry group G of the cube on the set of faces (Examples 2.1.2(iii) and 2.1.4(iii)) is faithful, since an isometry is determined by its effect on faces. The same is true for edges and vertices.

  But the action of G on the 4-element set X of long diagonals is not faithful: for G has 48 elements, whereas $\mathrm{Sym}(X)$ has only $4! = 24$ elements, so the homomorphism $\Sigma \colon G \to \mathrm{Sym}(X)$ cannot be injective.

  iv. The trivial action of a group G on a set X is never faithful unless G itself is trivial, since $gx = x$ for all $g \in G$ and $x \in X$.

> **Exercise 2.1.10** Example 2.1.9(iii) shows that the action of the isometry cube G of the cube on the set X of long diagonals is not faithful. By Lemma 2.1.8, there must be some non-identity isometry of the cube that fixes all four long diagonals. In fact, there is exactly one. What is it?

When a group G acts faithfully on a set X, there is a copy of G sitting inside $\mathrm{Sym}(X)$ as a subgroup (a 'faithful representation' of G):

Lemma 2.1.11 *Let G be a group acting faithfully on a set X. Then G is isomorphic to the subgroup*

$$\mathrm{im}\,\Sigma = \{\bar{g} : g \in G\}$$

of $\mathrm{Sym}(X)$, where $\Sigma\colon G \to \mathrm{Sym}(X)$ and $\bar{g}$ are defined as above.

Proof By Lemma 2.1.8, Σ is injective, and it is a general group-theoretic fact that any injective homomorphism $\varphi\colon G \to H$ induces an isomorphism between G and $\mathrm{im}\,\varphi$. $\qquad\square$

Example 2.1.12 Consider the usual action of the isometry group G of the cube on the 8-element set X of vertices. As we have seen, this action is faithful. Hence the associated homomorphism

$$\begin{array}{cccc} \Sigma: & G & \to & \mathrm{Sym}(X) \\ & g & \mapsto & \bar{g} \end{array}$$

induces an isomorphism between G and the subgroup $\{\bar{g} : g \in G\}$ of $\mathrm{Sym}(X)$. The subgroup consists of all permutations of the set of vertices that come from some isometry. For instance, there is no isometry that exchanges two vertices but leaves the rest fixed, so this subgroup contains no 2-cycles.

Remark 2.1.13 How does Lemma 2.1.11 look when X is a finite set with elements $x_1, \ldots, x_k$? Then $\mathrm{Sym}(X) \cong S_k$, and as in Remark 2.1.5, we can write $gx_i = x_{\sigma_g(i)}$. It follows from that lemma and remark that G is isomorphic to the subgroup $\{\sigma_g : g \in G\}$ of S_k (which *is* a subgroup). The isomorphism is given by $g \mapsto \sigma_g$.

Faithfulness is about which elements of the group fix everything in the set. We can also ask which elements of the set are fixed by everything in the group—or more generally, by some prescribed set S of group elements.

Definition 2.1.14 Let G be a group acting on a set X. Let $S \subseteq G$. The **fixed set** of S is

$$\mathrm{Fix}(S) = \{x \in X : sx = x \text{ for all } s \in S\}.$$

Later, we'll need the following lemma.

Lemma 2.1.15 *Let G be a group acting on a set X, let $S \subseteq G$, and let $g \in G$. Then $\mathrm{Fix}(gSg^{-1}) = g\,\mathrm{Fix}(S)$.*

Here $gSg^{-1} = \{gsg^{-1} : s \in S\}$ and $g\,\mathrm{Fix}(S) = \{gx : x \in \mathrm{Fix}(S)\}$.

Proof For $x \in X$, we have

$$
\begin{aligned}
x \in \mathrm{Fix}(gSg^{-1}) &\iff gsg^{-1}x = x \text{ for all } s \in S \\
&\iff sg^{-1}x = g^{-1}x \text{ for all } s \in S \\
&\iff g^{-1}x \in \mathrm{Fix}(S) \\
&\iff x \in g\,\mathrm{Fix}(S). \qquad \qquad \square
\end{aligned}
$$

2.2 Rings

We'll begin this part with some stuff you know—but with a twist.

In this course, the word **ring** means commutative ring with 1 (multiplicative identity). Noncommutative rings and rings without 1 are important in some parts of mathematics, but since we'll be focusing on commutative rings with 1, it will be easier to just call them 'rings'.

Example 2.2.1 There are many ways of building new rings from old. One of the most fundamental is that from any ring R, we can build the ring $R[t]$ of polynomials over R. We will define $R[t]$ formally and study it in detail in Chapter 3.

Given rings R and S, a **homomorphism** from R to S is a function $\varphi \colon R \to S$ satisfying the equations

$$
\varphi(r + r') = \varphi(r) + \varphi(r'), \qquad \varphi(0) = 0, \qquad \qquad \varphi(-r) = -\varphi(r),
$$
$$
\varphi(rr') = \varphi(r)\varphi(r'), \qquad \varphi(1) = 1 \text{ (note this!)}
$$

for all $r, r' \in R$. For example, complex conjugation is a homomorphism $\mathbb{C} \to \mathbb{C}$. It is a very useful lemma that if

$$
\varphi(r + r') = \varphi(r) + \varphi(r'), \qquad \varphi(rr') = \varphi(r)\varphi(r'), \qquad \varphi(1) = 1
$$

for all $r, r' \in R$ then φ is a homomorphism. In other words, to show that φ is a homomorphism, you only need to check it preserves $+$, $\cdot$ and 1; preservation of 0 and negatives then comes for free. But you *do* need to check it preserves 1. That doesn't follow from the other conditions.

A **subring** of a ring R is a subset $S \subseteq R$ that contains 0 and 1 and is closed under addition, multiplication and negatives. Whenever S is a subring of R, the inclusion $\iota \colon S \to R$ (defined by $\iota(s) = s$) is a homomorphism.

Lemma 2.2.3 *Let R be a ring and let $\mathscr{S}$ be any set (perhaps infinite) of subrings of R. Then their intersection $\bigcap_{S \in \mathscr{S}} S$ is also a subring of R.*

In contrast, in the Honours Algebra setup, even the intersection of *two* subrings need not be a subring.

Proof Write $T = \bigcap_{S \in \mathscr{S}} S$.

For each $S \in \mathscr{S}$, we have $0 \in S$ since S is a subring. Hence $0 \in T$ by definition of intersection.

Let $r, s \in T$. For each $S \in \mathscr{S}$, we have $r, s \in S$ by definition of intersection, so $r + s \in S$ since S is a subring. Hence $r + s \in T$ by definition of intersection.

Similar arguments show that $r \in T \Rightarrow -r \in T$, that $1 \in T$, and that $r, s \in T \Rightarrow rs \in T$. $\qquad\square$

Example 2.2.4 For any ring R, there is exactly one homomorphism $\mathbb{Z} \to R$. Here is a sketch of the proof.

To show there is *at least* one homomorphism $\chi \colon \mathbb{Z} \to R$, we construct one. Define χ inductively on integers $n \geq 0$ by $\chi(0) = 0$ and $\chi(n + 1) = \chi(n) + 1_R$. Thus,

$$\chi(n) = 1_R + \cdots + 1_R.$$

Define χ on negative integers n by $\chi(n) = -\chi(-n)$. A series of tedious checks shows that χ is indeed a ring homomorphism.

To show there is *only* one homomorphism $\mathbb{Z} \to R$, let φ be any homomorphism $\mathbb{Z} \to R$; we have to prove that $\varphi = \chi$. Certainly $\varphi(0) = 0 = \chi(0)$. Next prove by induction on n that $\varphi(n) = \chi(n)$ for nonnegative integers n. I leave the details to you, but the crucial point is that *because homomorphisms preserve* 1, we must have

$$\varphi(n + 1) = \varphi(n) + \varphi(1) = \varphi(n) + 1_R$$

for all $n \geq 0$. Once we have shown that φ and χ agree on the nonnegative integers, it follows that for negative n,

$$\varphi(n) = -\varphi(-n) = -\chi(-n) = \chi(n).$$

Hence $\varphi(n) = \chi(n)$ for all $n \in \mathbb{Z}$; that is, $\varphi = \chi$.

The meaning of '$n \cdot 1$', and Exercise 2.3.13

Usually we write $\chi(n)$ as $n \cdot 1_R$, or simply as n if it is clear from the context that n is to be interpreted as an element of R. So for $n \geq 0$,

$$n \cdot 1_R = \underbrace{1_R + \cdots + 1_R}_{n \text{ times}}.$$

The dot in the expression '$n \cdot 1_R$' is not multiplication in any ring, since $n \in \mathbb{Z}$ but $1_R \in R$. It's just notation.

Every ring homomorphism $\varphi \colon R \to S$ has an image $\operatorname{im} \varphi$, which is a subring of S, and a kernel $\ker \varphi$, which is an ideal of R.

> **Warning 2.2.5** Subrings are analogous to subgroups, and ideals are analogous to normal subgroups. But whereas normal subgroups are a special kind of subgroup, ideals are *not* a special kind of subring! Subrings must contain 1, but most ideals don't.

> **Exercise 2.2.6** Prove that the only subring of a ring R that is also an ideal is R itself.

Quotient rings

Given an ideal $I \trianglelefteq R$, we obtain the quotient ring or factor ring R/I and the canonical homomorphism $\pi_I \colon R \to R/I$, which is surjective and has kernel I.

As explained in Honours Algebra, the quotient ring together with the canonical homomorphism has a 'universal property': given any ring S and any homomorphism $\varphi \colon R \to S$ satisfying $\ker \varphi \supseteq I$, there is exactly one homomorphism $\bar{\varphi} \colon R/I \to S$ such that this diagram commutes:

$$
\begin{array}{ccc}
 & R & \\
\pi_I \downarrow & & \searrow \varphi \\
R/I & \dashrightarrow_{\bar{\varphi}} & S.
\end{array}
$$

(For a diagram to **commute** means that whenever there are two different paths from one object to another, the composites along the two paths are equal. Here, it means that $\varphi = \bar{\varphi} \circ \pi_I$.) The first isomorphism theorem says that if φ is surjective and has kernel *equal* to I then $\bar{\varphi}$ is an isomorphism. So $\pi_I \colon R \to R/I$ is essentially the only surjective homomorphism out of R with kernel I.

An **integral domain** is a ring R such that $0_R \neq 1_R$ and for $r, r' \in R$,

$$rr' = 0 \Rightarrow r = 0 \text{ or } r' = 0.$$

Equivalently, an integral domain is a nontrivial ring in which cancellation is valid: $rs = r's$ implies $r = r'$ or $s = 0$.

23

Let Y be a subset of a ring R. The **ideal $\langle Y \rangle$ generated by Y** is defined as the intersection of all the ideals of R containing Y. You can show that any intersection of ideals is an ideal (much as for subrings in Lemma 2.2.3). So $\langle Y \rangle$ *is* an ideal. We can also characterize $\langle Y \rangle$ as the smallest ideal of R containing Y. That is, $\langle Y \rangle$ is an ideal containing Y, and if I is another ideal containing Y then $\langle Y \rangle \subseteq I$.

This definition of the ideal generated by Y is top-down: we obtain $\langle Y \rangle$ as the intersection of bigger ideals. But there is also a useful bottom-up description of $\langle Y \rangle$. Here it is when Y is finite.

Lemma 2.2.11 *Let R be a ring and let $Y = \{r_1, \ldots, r_n\}$ be a finite subset. Then*

$$\langle Y \rangle = \{a_1 r_1 + \cdots + a_n r_n : a_1, \ldots, a_n \in R\}.$$

Proof Write I for the right-hand side. It is straightforward to check that I is an ideal of R, and it contains Y because, for instance, $r_1 = 1 r_1 + 0 r_2 + \cdots + 0 r_n$.

Now let J be any ideal of R containing Y. Let $a_1, \ldots, a_n \in R$. For each i, we have $r_i \in J$ since J contains Y, and so $a_i r_i \in J$ since J is an ideal. Hence $\sum a_i r_i \in J$, again since J is an ideal. So $I \subseteq J$.

Hence I is the smallest ideal of R containing Y, that is, $I = \langle Y \rangle$. $\qquad \square$

Digression 2.2.12 A similar interplay between top-down and bottom-up appears in other parts of mathematics.

For example, in topology, the closure of a subset of a metric or topological space is the intersection of all closed subsets containing it. In linear algebra,

the span of a subset of a vector space is the intersection of all linear subspaces containing it. In group theory, the subgroup generated by a subset of a group is the intersection of all subgroups containing it.

These are all top-down definitions, but there are equivalent bottom-up definitions, describing explicitly which elements belong to the subset. Sometimes we're lucky and those descriptions are simple. For instance, closures can easily be described in terms of limit points, and spans are just sets of linear combinations. But sometimes it gets more complicated. For example, the subgroup of a group G generated by a subset Y can be described *informally* as the set of elements of G that can be obtained from Y by taking products and inverses, but expressing that precisely is a little bit fiddly.

It's worth getting comfortable with the top-down style of definition, as it works well in cases where the bottom-up approach is prohibitively complicated, and we'll need it later.

When $Y = \{r_1, \ldots, r_n\}$, we write $\langle Y \rangle$ as $\langle r_1, \ldots, r_n \rangle$ rather than $\langle \{r_1, \ldots, r_n\} \rangle$. In particular, when $n = 1$, Lemma 2.2.11 implies that

$$\langle r \rangle = \{ar : a \in R\}.$$

Ideals of the form $\langle r \rangle$ are called **principal ideals**. A **principal ideal domain** is an integral domain in which every ideal is principal.

Example 2.2.13 $\mathbb{Z}$ is a principal ideal domain. Indeed, if $I \trianglelefteq \mathbb{Z}$ then either $I = \{0\}$, in which case $I = \langle 0 \rangle$, or I contains some positive integer, in which case we can define n to be the least positive integer in I and use the division algorithm to show that $I = \langle n \rangle$.

Exercise 2.2.14 Fill in the details of Example 2.2.13.

Let r and s be elements of a ring R. We say that r **divides** s, and write $r \mid s$, if there exists $a \in R$ such that $s = ar$. This condition is equivalent to $s \in \langle r \rangle$, and to $\langle s \rangle \subseteq \langle r \rangle$.

An element $u \in R$ is a **unit** if it has a multiplicative inverse, or equivalently if $\langle u \rangle = R$. The units form a group $R^\times$ under multiplication. For instance, $\mathbb{Z}^\times = \{1, -1\}$.

Exercise 2.2.15 Let r and s be elements of an integral domain. Show that $r \mid s \mid r \iff \langle r \rangle = \langle s \rangle \iff s = ur$ for some unit u.

Elements r and s of a ring are **coprime** if for $a \in R$,

$$a \mid r \text{ and } a \mid s \Rightarrow a \text{ is a unit.}$$

Proposition 2.2.16 *Let R be a principal ideal domain and $r, s \in R$. Then*

$$r \text{ and } s \text{ are coprime} \iff ar + bs = 1 \text{ for some } a, b \in R.$$

Proof $\Rightarrow$: suppose that r and s are coprime. Since R is a principal ideal domain, $\langle r, s \rangle = \langle u \rangle$ for some $u \in R$. Since $r \in \langle r, s \rangle = \langle u \rangle$, we must have $u \mid r$, and similarly $u \mid s$. But r and s are coprime, so u is a unit. Hence $1 \in \langle u \rangle = \langle r, s \rangle$. But by Lemma 2.2.11,

$$\langle r, s \rangle = \{ ar + bs : a, b \in R \},$$

and the result follows.

$\Leftarrow$: suppose that $ar + bs = 1$ for some $a, b \in R$. If $u \in R$ with $u \mid r$ and $u \mid s$ then $u \mid (ar + bs) = 1$, so u is a unit. Hence r and s are coprime. $\square$

2.3 Fields

A **field** is a ring K in which $0 \neq 1$ and every nonzero element is a unit. Equivalently, it is a ring such that $K^\times = K \setminus \{0\}$. Every field is an integral domain.

> **Exercise 2.3.1** Write down all the examples of fields that you know.

As we go on, we'll see several ways of making new fields out of old. Here's the simplest.

Example 2.3.2 Let K be a field. A **rational expression** over K is a ratio of two polynomials

$$\frac{f(t)}{g(t)},$$

where $f(t), g(t) \in K[t]$ with $g \neq 0$. Two such expressions, f_1/g_1 and f_2/g_2, are regarded as equal if $f_1 g_2 = f_2 g_1$ in $K[t]$. So formally, a rational expression is an equivalence class of pairs (f, g) under the equivalence relation in the last sentence. The set of rational expressions over K is denoted by $K(t)$.

Rational expressions are added, subtracted and multiplied in the ways you'd expect, making $K(t)$ into a field. We will look at it more carefully in Chapter 3.

A field K has exactly two ideals: $\{0\}$ and K. For if $\{0\} \neq I \lhd K$ then $u \in I$ for some $u \neq 0$; but then u is a unit, so $\langle u \rangle = K$, so $I = K$.

Lemma 2.3.3 *Every homomorphism between fields is injective.*

A 'homomorphism between fields' means a *ring* homomorphism.

Proof Let $\varphi\colon K \to L$ be a homomorphism between fields. Then $\ker\varphi \lhd K$, so $\ker\varphi$ is either $\{0\}$ or K. If $\ker\varphi = K$ then $\varphi(1) = 0$; but $\varphi(1) = 1$ by definition of homomorphism, so $0 = 1$ in L, contradicting the assumption that L is a field. Hence $\ker\varphi = \{0\}$, that is, φ is injective. $\qquad\square$

Warning 2.3.4 With the Honours Algebra definition of homomorphism, Lemma 2.3.3 would be false, since the map with constant value 0 would be a homomorphism.

Exercise 2.3.5 Let $\varphi\colon K \to L$ be a homomorphism of fields and let $0 \neq a \in K$. Prove that $\varphi(a^{-1}) = \varphi(a)^{-1}$. Why is $\varphi(a)^{-1}$ defined?

A **subfield** of a field K is a subring that is a field.

Lemma 2.3.6 *Let $\varphi\colon K \to L$ be a homomorphism between fields.*

   i. For any subfield K' of K, the image $\varphi K'$ is a subfield of L.

   ii. For any subfield L' of L, the preimage $\varphi^{-1} L'$ is a subfield of K.

Proof For (i), you know from Proposition 3.4.29 of Honours Algebra that $\varphi K'$ is a subring of L, and you can use Exercise 2.3.5 above to show that if $0 \neq b \in \varphi K'$ then $b^{-1} \in \varphi K'$. The proof of (ii) is similar. $\qquad\square$

Whenever we have a collection of homomorphisms between the same pair of fields, we get a subfield in the following way.

Definition 2.3.7 Let X and Y be sets, and let $S \subseteq \{\text{functions } X \to Y\}$. The **equalizer** of S is

$$\mathrm{Eq}(S) = \{x \in X : f(x) = g(x) \text{ for all } f, g \in S\}.$$

In other words, it is the part of X where all the functions in S are *equal*.

Lemma 2.3.8 *Let K and L be fields, and let $S \subseteq \{\text{homomorphisms } K \to L\}$. Then $\mathrm{Eq}(S)$ is a subfield of K.*

Proof We must show that $0, 1 \in \mathrm{Eq}(S)$, that if $a \in \mathrm{Eq}(S)$ then $-a \in \mathrm{Eq}(S)$ and $1/a \in \mathrm{Eq}(S)$ (for $a \neq 0$), and that if $a, b \in \mathrm{Eq}(S)$ then $a + b, ab \in \mathrm{Eq}(S)$. I will show just the last of these, leaving the rest to you.

Suppose that $a, b \in \mathrm{Eq}(S)$. For all $\varphi, \theta \in S$, we have

$$\varphi(ab) = \varphi(a)\varphi(b) = \theta(a)\theta(b) = \theta(ab),$$

so $ab \in \mathrm{Eq}(S)$. $\qquad\square$

Example 2.3.9 Let $K = L = \mathbb{C}$. Let $S = \{\mathrm{id}_{\mathbb{C}}, \kappa\}$, where $\kappa \colon \mathbb{C} \to \mathbb{C}$ is complex conjugation. Then

$$\mathrm{Eq}(S) = \{z \in \mathbb{C} : z = \bar{z}\} = \mathbb{R},$$

and $\mathbb{R}$ is indeed a subfield of $\mathbb{C}$.

Next we ask: when is $1 + \cdots + 1$ equal to 0?

Let R be any ring. By Example 2.2.4, there is a unique homomorphism $\chi \colon \mathbb{Z} \to R$. Its kernel is an ideal of the principal ideal domain $\mathbb{Z}$. Hence $\ker \chi = \langle n \rangle$ for a unique integer $n \geq 0$. This n is called the **characteristic** of R, and written as char R. So for $m \in \mathbb{Z}$, we have $m \cdot 1_R = 0$ if and only if m is a multiple of char R. Or equivalently,

$$\mathrm{char}\, R = \begin{cases} \text{the least } n > 0 \text{ such that } n \cdot 1_R = 0_R, & \text{if such an } n \text{ exists;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

The concept of characteristic is mostly used in the case of fields.

Examples 2.3.10   i. $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ all have characteristic 0.

ii. For a prime number p, we write $\mathbb{F}_p$ for the field $\mathbb{Z}/\langle p \rangle$ of integers modulo p. Then char $\mathbb{F}_p = p$.

iii. For any field K, the field $K(t)$ of rational expressions has the same characteristic as K.

Lemma 2.3.11 *The characteristic of an integral domain is* 0 *or a prime number.*

Proof Let R be an integral domain and write $n = \mathrm{char}\, R$. Suppose that $n > 0$; we must prove that n is prime.

Since $1 \neq 0$ in an integral domain, $n \neq 1$. (Remember that 1 is not a prime! So that step was necessary.) Now let $k, m > 0$ with $km = n$. Writing χ for the unique homomorphism $\mathbb{Z} \to R$, we have

$$\chi(k)\chi(m) = \chi(km) = \chi(n) = 0,$$

and R is an integral domain, so $\chi(k) = 0$ or $\chi(m) = 0$. WLOG, $\chi(k) = 0$. But $\ker \chi = \langle n \rangle$, so $n \mid k$, so $k = n$. Hence n is prime. $\qquad \square$

In particular, the characteristic of a field is always 0 or a prime. But there is no way of mapping between fields of different characteristics:

Lemma 2.3.12 *Let* $\varphi \colon K \to L$ *be a homomorphism of fields. Then* char $K =$ char L.

Proof Write χ_K and χ_L for the unique homomorphisms from $\mathbb{Z}$ to K and L, respectively. Since χ_L is the *unique* homomorphism $\mathbb{Z} \to L$, the triangle

$$
\begin{array}{ccc}
 & \mathbb{Z} & \\
\chi_K \swarrow & & \searrow \chi_L \\
K \xrightarrow{\quad \varphi \quad} & & L
\end{array}
$$

commutes. (Concretely, this says that $\varphi(n \cdot 1_K) = n \cdot 1_L$ for all $n \in \mathbb{Z}$.) Hence $\ker(\varphi \circ \chi_K) = \ker \chi_L$. But φ is injective by Lemma 2.3.3, so $\ker(\varphi \circ \chi_K) = \ker \chi_K$. Hence $\ker \chi_K = \ker \chi_L$, or equivalently, char K = char L. $\qquad\square$

For example, the inclusion $\mathbb{Q} \to \mathbb{R}$ is a homomorphism of fields, and both have characteristic 0.

> **Exercise 2.3.13** This proof of Lemma 2.3.12 is quite abstract. Find a more concrete proof, taking equation (2.2) as your definition of characteristic. (You will still need the fact that φ is injective.)

The meaning of '$n \cdot 1$', and Exercise 2.3.13

The **prime subfield** of K is the intersection of all the subfields of K. It is straightforward to show that any intersection of subfields is a subfield (much as in Lemma 2.2.3). Hence the prime subfield *is* a subfield. It is the smallest subfield of K, in the sense that any other subfield of K contains it.

Concretely ('bottom-up'), the prime subfield of K is

$$
\left\{ \frac{m \cdot 1_K}{n \cdot 1_K} : m, n \in \mathbb{Z} \text{ with } n \cdot 1_K \neq 0 \right\}.
$$

To see this, first note that this set is a subfield of K. It is the smallest subfield of K: for if L is a subfield of K then $1_K \in L$ by definition of subfield, so $m \cdot 1_K \in L$ for all integers m, so $(m \cdot 1_K)/(n \cdot 1_K) \in L$ for all integers m and n such that $n \cdot 1_K \neq 0$.

Examples 2.3.14   i. The field $\mathbb{Q}$ has no proper subfields, so the prime subfield of $\mathbb{Q}$ is $\mathbb{Q}$ itself.

ii. Let p be a prime. The field $\mathbb{F}_p$ has no proper subfields, so the prime subfield of $\mathbb{F}_p$ is $\mathbb{F}_p$ itself.

> **Exercise 2.3.15** What is the prime subfield of $\mathbb{R}$? Of $\mathbb{C}$?

The prime subfields appearing in Examples 2.3.14 were $\mathbb{Q}$ and $\mathbb{F}_p$. In fact, these are the *only* prime subfields of anything:

Lemma 2.3.16 *Let K be a field.*

    i. If $\operatorname{char} K = 0$ *then the prime subfield of K is $\mathbb{Q}$.*

    ii. If $\operatorname{char} K = p > 0$ *then the prime subfield of K is $\mathbb{F}_p$.*

In the statement of this lemma, as so often in mathematics, the word 'is' means 'is isomorphic to'. I hope you're comfortable with that by now.

Proof For (i), suppose that $\operatorname{char} K = 0$. By definition of characteristic, $n \cdot 1_K \neq 0$ for all integers $n \geq 0$. One can check that there is a well-defined homomorphism $\varphi \colon \mathbb{Q} \to K$ defined by $m/n \mapsto (m \cdot 1_K)/(n \cdot 1_K)$. (The check uses the fact that $\chi \colon n \mapsto n \cdot 1_K$ is a homomorphism.) Now φ is injective (being a homomorphism of fields), so $\operatorname{im} \varphi \cong \mathbb{Q}$. But $\operatorname{im} \varphi$ is a subfield of K, and since $\mathbb{Q}$ has no proper subfields, it is the prime subfield.

For (ii), suppose that $\operatorname{char} K = p > 0$. By Lemma 2.3.11, p is prime. The unique homomorphism $\chi \colon \mathbb{Z} \to K$ has kernel $\langle p \rangle$, by definition. By the first isomorphism theorem, $\operatorname{im} \chi \cong \mathbb{Z}/\langle p \rangle = \mathbb{F}_p$. But $\operatorname{im} \chi$ is a subfield of K, and since $\mathbb{F}_p$ has no proper subfields, it is the prime subfield. $\qquad\square$

Lemma 2.3.17 *Every finite field has positive characteristic.*

Proof By Lemma 2.3.16, a field of characteristic 0 contains a copy of $\mathbb{Q}$ and is therefore infinite. $\qquad\square$

> **Warning 2.3.18** There are also *infinite* fields of positive characteristic. An example is the field $\mathbb{F}_p(t)$ of rational expressions over $\mathbb{F}_p$.

Square roots usually come in pairs: how many times in your life have you written a $\pm$ sign before a $\sqrt{}$? But in characteristic 2, plus and minus are the same, so the two square roots become one. We'll see that this pattern persists: pth roots behave strangely in characteristic p. First, an important little lemma:

Lemma 2.3.19 *Let p be a prime and $0 < i < p$. Then $p \mid \binom{p}{i}$.*

For example, the 7th row of Pascal's triangle is $1, 7, 21, 35, 35, 21, 7, 1$, and the lemma predicts that 7 divides all of these numbers apart from the first and last.

Proof We have $i!(p - i)!\binom{p}{i} = p!$. Now p divides $p!$ but not $i!$ or $(p - i)!$ (since p is prime and $0 < i < p$), so p must divide $\binom{p}{i}$. $\qquad\square$

Proposition 2.3.20 *Let p be a prime number and R a ring of characteristic p.*

i. The function

$$\theta \colon \begin{array}{ccc} R & \to & R \\ r & \mapsto & r^p \end{array}$$

is a homomorphism.

ii. If R is a field then θ is injective.

iii. If R is a finite field then θ is an automorphism of R.

Proof For (i), certainly θ preserves multiplication and 1. To show that θ preserves addition, let $r, s \in R$: then by Lemma 2.3.19 and the hypothesis that char $R = p$,

$$\theta(r + s) = (r + s)^p = \sum_{i=0}^{p} \binom{p}{i} r^i s^{p-i} = r^p + s^p = \theta(r) + \theta(s).$$

Now (ii) follows since every homomorphism between fields is injective, and (iii) since every injection from a finite set to itself is bijective. $\qquad\square$

The homomorphism $\theta \colon r \mapsto r^p$ is called the **Frobenius map**, or, in the case of finite fields, the **Frobenius automorphism**.

That θ is a homomorphism is a shocker. Writing $(x + y)^n = x^n + y^n$ is a classic algebra mistake. But here, it's true!

Example 2.3.21 The Frobenius automorphism of $\mathbb{F}_p = \mathbb{Z}/\langle p \rangle$ is not very interesting. When G is a finite group of order n, Lagrange's theorem implies that $g^n = 1$ for all $g \in G$. Applying this to the multiplicative group $\mathbb{F}_p^\times = \mathbb{F}_p \setminus \{0\}$ gives $a^{p-1} = 1$ whenever $0 \neq a \in \mathbb{F}_p$. It follows that $a^p = a$ for all $a \in \mathbb{F}_p$. That is, θ is the identity. Everything is its own pth root!

Unfortunately, we can't give any interesting examples of the Frobenius map just now, because we have so few examples of fields. That will change later.

Corollary 2.3.22 *Let p be a prime number.*

i. In a field of characteristic p, every element has at most one pth root.

ii. In a finite field of characteristic p, every element has exactly one pth root.

Proof Part (i) says that the Frobenius map is injective, and part (ii) says that it is bijective, as Proposition 2.3.20 states. $\qquad\square$

Examples 2.3.23 i. In a field of characteristic 2, every element has at most one square root.

ii. In $\mathbb{C}$, there are p different pth roots of unity. But in a field of characteristic p, there is only one: 1 itself.

iii. Let K be a field of characteristic p and $a \in K$. Corollary 2.3.22(i) says that a has *at most* one pth root. It may have none. For instance, you'll show in Exercise 3.1.13 that the element t of $\mathbb{F}_p(t)$ has no pth root.

Here's a construction that will let us manufacture many more examples of fields.

Building blocks

An element r of a ring R is **irreducible** if r is not 0 or a unit, and if for $a, b \in R$,

$$r = ab \implies a \text{ or } b \text{ is a unit.}$$

For example, the irreducibles in $\mathbb{Z}$ are $\pm 2, \pm 3, \pm 5, \ldots$. An element of a ring is **reducible** if it is not 0, a unit, or irreducible.

Warning 2.3.24 The 0 and units of a ring count as neither reducible nor irreducible, in much the same way that the integers 0 and 1 are neither prime nor composite.

Exercise 2.3.25 What are the irreducible elements of a field?

Proposition 2.3.26 *Let R be a principal ideal domain and $0 \neq r \in R$. Then*

$$r \text{ is irreducible} \iff R/\langle r \rangle \text{ is a field.}$$

Proof Write π for the canonical homomorphism $R \to R/\langle r \rangle$.

$\Rightarrow$: suppose that r is irreducible. To show that $1_{R/\langle r \rangle} \neq 0_{R/\langle r \rangle}$, note that since r is not a unit, $1_R \notin \langle r \rangle = \ker \pi$, so

$$1_{R/\langle r \rangle} = \pi(1_R) \neq 0_{R/\langle r \rangle}.$$

Next we have to show that every nonzero element of $R/\langle r \rangle$ is a unit, or equivalently that $\pi(s)$ is a unit whenever $s \in R$ with $s \notin \langle r \rangle$. We have $r \nmid s$, and r is irreducible, so r and s are coprime. Hence by Proposition 2.2.16 and the assumption that R is a principal ideal domain, we can choose $a, b \in R$ such that

$$ar + bs = 1_R.$$

Applying π to each side gives

$$\pi(a)\pi(r) + \pi(b)\pi(s) = 1_{R/\langle r \rangle}.$$

But $\pi(r) = 0$, so $\pi(b)\pi(s) = 1$, so $\pi(s)$ is a unit.

$\Leftarrow$: suppose that $R/\langle r \rangle$ is a field. Then $1_{R/\langle r \rangle} \neq 0_{R/\langle r \rangle}$, that is, $\pi(1_R) \notin \ker \pi = \langle r \rangle$, that is, $r \nmid 1_R$. Hence r is not a unit.

Next we have to show that if $a, b \in R$ with $r = ab$ then a or b is a unit. We have

$$0 = \pi(r) = \pi(a)\pi(b)$$

and $R/\langle r \rangle$ is an integral domain, so WLOG $\pi(a) = 0$. Then $a \in \ker \pi = \langle r \rangle$, so $a = rb'$ for some $b' \in R$. This gives

$$r = ab = rb'b.$$

But $r \neq 0$ by hypothesis, and R is an integral domain, so $b'b = 1$. Hence b is a unit. $\qquad\square$

Example 2.3.27 When n is an integer, $\mathbb{Z}/\langle n \rangle$ is a field if and only if n is irreducible (that is, $\pm$ a prime number).

Proposition 2.3.26 enables us to construct fields from irreducible elements. . . but irreducible elements *of a principal ideal domain*. Right now that's not much help, because we don't have many examples of principal ideal domains. But we will do soon.

Chapter 3

Polynomials

This chapter revisits and develops some themes you met in Honours Algebra. Before you begin, it may help you to reread Section 3.3 (Polynomials) of the Honours Algebra notes.

3.1 The ring of polynomials

You already know the definition of polynomial, but I want to make a point by phrasing it in an unfamiliar way.

Definition 3.1.1 Let R be a ring. A **polynomial over** R is an infinite sequence $(a_0, a_1, a_2, \ldots)$ of elements of R such that $\{i : a_i \neq 0\}$ is finite.

The set of polynomials over R forms a ring as follows:

$$(a_0, a_1, \ldots) + (b_0, b_1, \ldots) = (a_0 + b_0, a_1 + b_1, \ldots), \tag{3.1}$$

$$(a_0, a_1, \ldots) \cdot (b_0, b_1, \ldots) = (c_0, c_1, \ldots) \tag{3.2}$$

$$\text{where } c_k = \sum_{i,j \,:\, i+j=k} a_i b_j, \tag{3.3}$$

the zero of the ring is $(0, 0, \ldots)$, and the multiplicative identity is $(1, 0, 0, \ldots)$.

Of course, we almost always write $(a_0, a_1, a_2, \ldots)$ as $a_0 + a_1 t + a_2 t^2 + \cdots$, or the same with some other symbol in place of t. In that notation, formulas (3.1) and (3.2) look like the usual formulas for addition and multiplication of polynomials. Nevertheless:

> **Warning 3.1.2** A polynomial is not a function!
>
> A polynomial *gives rise to* a function, as we'll recall in a moment. But a polynomial itself is a purely formal object. To emphasize this, we sometimes call the symbol t an **indeterminate** rather than a 'variable'.

The set of polynomials over R is written as $R[t]$ (or $R[u]$, $R[x]$, etc.). Since $R[t]$ is itself a ring S, we can consider the ring $S[u] = (R[t])[u]$, usually written as $R[t, u]$. And then there's $R[t, u, v] = (R[t, u])[v]$, and so on.

Polynomials are typically written as either f or $f(t)$, interchangeably. A polynomial $f = (a_0, a_1, \ldots)$ over R gives rise to a function

$$
\begin{aligned}
R &\to R \\
r &\mapsto a_0 + a_1 r + a_2 r^2 + \cdots.
\end{aligned}
$$

(The sum on the right-hand side makes sense because only finitely many a_is are nonzero.) This function is usually denoted by f too. But calling it that is slightly dangerous, because:

> **Warning 3.1.3** Different polynomials can give rise to the same function. For example, consider $t, t^2 \in \mathbb{F}_2[t]$. They are different polynomials: going back to Definition 3.1.1, they're alternative notation for the sequences
>
> $$(0, 1, 0, 0, \ldots) \quad \text{and} \quad (0, 0, 1, 0, \ldots),$$
>
> which are plainly not the same. On the other hand, they induce the same *function* $\mathbb{F}_2 \to \mathbb{F}_2$, because $a = a^2$ for all (both) $a \in \mathbb{F}_2$.

> **Exercise 3.1.4** Show that whenever R is a finite nontrivial ring, it is possible to find distinct polynomials over R that induce the same function $R \to R$. (Hint: are there finitely or infinitely many polynomials over R? Functions $R \to R$?)

Remark 3.1.5 In Example 2.3.2, we met the field $K(t)$ of rational expressions over a field K. People sometimes say 'rational function' to mean 'rational expression'. But just as for polynomials, I want to emphasize that *rational expressions are not functions*. For instance, $1/(t - 1)$ is a totally respectable element of $K(t)$. You don't have to worry about what happens 'when $t = 1$', because t is just a formal symbol (a mark on a piece of paper), not a variable. And $1/(t - 1)$ is just a formal expression, not a function.

The ring of polynomials has a universal property: a homomorphism from $R[t]$ to some other ring B is determined by its effect on scalars and on t itself, in the following sense.

Proposition 3.1.6 (Universal property of the polynomial ring) *Let R and B be rings. For every homomorphism $\varphi\colon R \to B$ and every $b \in B$, there is exactly one homomorphism $\theta\colon R[t] \to B$ such that*

$$\theta(a) = \varphi(a) \quad \text{for all } a \in R, \tag{3.4}$$
$$\theta(t) = b. \tag{3.5}$$

On the left-hand side of (3.4), the 'a' means the polynomial $a + 0t + 0t^2 + \cdots$.

Proof To show there is *at most one* such θ, take any homomorphism $\theta\colon R[t] \to B$ satisfying (3.4) and (3.5). Then for every polynomial $\sum_i a_i t^i$ over R,

$$\theta\left(\sum_i a_i t^i\right) = \sum_i \theta(a_i)\theta(t)^i \qquad \text{since } \theta \text{ is a homomorphism}$$
$$= \sum_i \varphi(a_i)b^i \qquad \text{by (3.4) and (3.5).}$$

So θ is uniquely determined.

To show there is *at least one* such θ, define a function $\theta\colon R[t] \to B$ by

$$\theta\left(\sum_i a_i t^i\right) = \sum_i \varphi(a_i)b^i$$

($\sum_i a_i t^i \in R[t]$). Then θ clearly satisfies conditions (3.4) and (3.5). It remains to check that θ is a homomorphism. I will do the worst part of this, which is to check that θ preserves multiplication, and leave the rest to you.

So, take polynomials $f(t) = \sum_i a_i t^i$ and $g(t) = \sum_j b_j t^j$. Then $f(t)g(t) = \sum_k c_k t^k$, where c_k is as defined in equation (3.3). We have

$$\theta(fg) = \theta\left(\sum_k c_k t^k\right)$$
$$= \sum_k \varphi(c_k)b^k \qquad\qquad\qquad \text{by definition of } \theta$$
$$= \sum_k \varphi\left(\sum_{i,j\,:\,i+j=k} a_i b_j\right)b^k \qquad\qquad \text{by definition of } c_k$$
$$= \sum_k \sum_{i,j\,:\,i+j=k} \varphi(a_i)\varphi(b_j)b^k \qquad \text{since } \varphi \text{ is a homomorphism}$$
$$= \sum_{i,j} \varphi(a_i)\varphi(b_j)b^{i+j}$$
$$= \left(\sum_i \varphi(a_i)b^i\right)\left(\sum_j \varphi(b_j)b^j\right)$$
$$= \theta(f)\theta(g) \qquad\qquad\qquad\qquad \text{by definition of } \theta. \qquad \square$$

Here are three uses for the universal property of the ring of polynomials. First:

Definition 3.1.7 Let $\varphi \colon R \to S$ be a ring homomorphism. The **induced homomorphism**

$$\varphi_* \colon R[t] \to S[t]$$

is the unique homomorphism $R[t] \to S[t]$ such that $\varphi_*(a) = \varphi(a)$ for all $a \in R$ and $\varphi_*(t) = t$.

The universal property guarantees that there is one and only one homomorphism φ_* with these properties. Concretely,

$$\varphi_*\Big(\sum_i a_i t^i\Big) = \sum_i \varphi(a_i) t^i$$

for all $\sum_i a_i t^i \in R[t]$.

Second, let R be a ring and $r \in R$. By the universal property, there is a unique homomorphism $\mathrm{ev}_r \colon R[t] \to R$ such that $\mathrm{ev}_r(a) = a$ for all $a \in R$ and $\mathrm{ev}_r(t) = r$. Concretely,

$$\mathrm{ev}_r\Big(\sum_i a_i t^i\Big) = \sum_i a_i r^i$$

for all $\sum_i a_i t^i \in R[t]$. This map ev_r is called **evaluation at r**.

(The notation $\sum a_i t^i$ for what is officially $(a_0, a_1, \ldots)$ makes it look *obvious* that we can evaluate a polynomial at an element, and that this gives a homomorphism: *of course* $(f \cdot g)(r) = f(r)g(r)$, for instance! But that's only because of the notation: there was actually something to prove here.)

Third, let R be a ring and $c \in R$. For any $f(t) \in R[t]$, we can 'substitute $t = u + c$' to get a polynomial in u. What exactly does this mean? Formally, there is a unique homomorphism $\theta \colon R[t] \to R[u]$ such that $\theta(a) = a$ for all $a \in R$ and $\theta(t) = u + c$. Concretely,

$$\theta\Big(\sum_i a_i t^i\Big) = \sum_i a_i (u + c)^i.$$

This particular substitution is invertible. Informally, the inverse is 'substitute $u = t - c$'. Formally, there is a unique homomorphism $\theta' \colon R[u] \to R[t]$ such that $\theta'(a) = a$ for all $a \in R$ and $\theta'(u) = t - c$. These maps θ and θ' carrying out the substitutions are inverse to each other, as you can deduce from either the universal property or the concrete descriptions. So, the substitution maps

$$R[t] \underset{\theta'}{\overset{\theta}{\rightleftarrows}} R[u] \tag{3.6}$$

define an isomorphism between $R[t]$ and $R[u]$. For example, since isomorphism preserve irreducibility (and everything else that matters!), $f(t)$ is irreducible if and only if $f(t - c)$ is irreducible.

> **Exercise 3.1.8** What happens to everything in the previous paragraph if we substitute $t = u^2 + c$ instead?

The rest of this section is about degree.

Definition 3.1.9 The **degree**, $\deg(f)$, of a nonzero polynomial $f(t) = \sum a_i t^i$ is the largest $n \geq 0$ such that $a_n \neq 0$. By convention, $\deg(0) = -\infty$, where $-\infty$ is a formal symbol which we give the properties

$$-\infty < n, \qquad (-\infty) + n = -\infty, \qquad (-\infty) + (-\infty) = -\infty$$

for all integers n.

> **Digression 3.1.10** Defining $\deg(0)$ like this is helpful because it allows us to make statements about *all* polynomials without annoying exceptions for the zero polynomial (e.g. Lemma 3.1.11(i)).
>
> But putting $\deg(0) = -\infty$ also makes intuitive sense. At least for polynomials over $\mathbb{R}$, the degree of a nonzero polynomial tells us how fast it grows: when x is large, $f(x)$ behaves roughly like $x^{\deg(f)}$. What about the zero polynomial? Well, whether or not x is large, $0(x) = 0$. And $x^{-\infty}$ can sensibly be interpreted as $\lim_{r \to -\infty} x^r = 0$, so it makes sense to put $\deg(0) = -\infty$.

Lemma 3.1.11 *Let R be an integral domain. Then:*

 i. $\deg(fg) = \deg(f) + \deg(g)$ *for all* $f, g \in R[t]$;

 ii. $R[t]$ *is an integral domain.*

Proof This was proved in Honours Algebra (Section 3.3). □

Example 3.1.12 For any integral domain R, the ring $R[t_1, \ldots, t_n]$ of polynomials over R in n variables is also an integral domain, by Lemma 3.1.11(ii) and induction. In particular, this is true when R is a field.

> **Exercise 3.1.13** Let p be a prime and consider the field $\mathbb{F}_p(t)$ of rational expressions over $\mathbb{F}_p$. Show that t has no pth root in $\mathbb{F}_p(t)$. (Hint: consider degrees of polynomials.)

The one and only polynomial of degree $-\infty$ is the zero polynomial. The polynomials of degree 0 are the nonzero constants. The polynomials of degree > 0 are, therefore, the nonconstant polynomials.

Lemma 3.1.14 *Let K be a field. Then:*

  i. the units in $K[t]$ are the nonzero constants;

  ii. $f \in K[t]$ is irreducible if and only if f is nonconstant and cannot be expressed as a product of two nonconstant polynomials.

Proof Part (i) was also in Honours Algebra (Section 3.3), and part (ii) follows from the general definition of irreducible element of a ring. □

3.2 Factorizing polynomials

Every nonzero integer can be expressed as a product of primes in an essentially unique way. But the analogous statement is not true in all rings, or even all integral domains. Some rings have elements that can't be expressed as a product of irreducibles at all. In other rings, factorizations into irreducibles exist but are not unique. (By 'not unique' I mean more than just changing the order of the factors or multiplying them by units.)

The big theorem of this section is that, happily, every polynomial over a field *can* be factorized into irreducibles, essentially uniquely.

We begin with a result on division of polynomials from Section 3.3 of Honours Algebra.

Proposition 3.2.1 *Let K be a field and $f, g \in K[t]$. Then there is exactly one pair of polynomials $q, r \in K[t]$ such that $f = qg + r$ and $\deg(r) < \deg(g)$.* □

We use this to prove an extremely useful fact:

Proposition 3.2.2 *Let K be a field. Then $K[t]$ is a principal ideal domain.*

Proof First, $K[t]$ is an integral domain, by Lemma 3.1.11(ii).

Now let $I \trianglelefteq K[t]$. If $I = \{0\}$ then $I = \langle 0 \rangle$. Otherwise, put $d = \min\{\deg(f) : 0 \neq f \in I\}$ and choose $g \in I$ such that $\deg(g) = d$.

I claim that $I = \langle g \rangle$. To prove this, let $f \in I$; we must show that $g \mid f$. By Proposition 3.2.1, $f = qg + r$ for some $q, r \in K[t]$ with $\deg(r) < d$. Now $r = f - qg \in I$ since $f, g \in I$, so the minimality of d implies that $r = 0$. Hence $f = qg$, as required. □

If you struggled with Exercise 2.2.14, that proof should give you a clue.

Warning 3.2.3 We just saw that $K[t]$ is a principal ideal domain, and we saw in Example 3.1.12 that $K[t_1, \ldots, t_n]$ is an *integral* domain. But it is not a *principal ideal* domain if $n > 1$. For example, the ideal

$$\langle t_1, t_2 \rangle = \{ f(t_1, t_2) \in \mathbb{Q}[t_1, t_2] : f \text{ has constant term } 0 \}$$

of $\mathbb{Q}[t_1, t_2]$ is not principal.

Also, Proposition 3.2.2 really needed the hypothesis that K is a *field*; it's not enough for it to be a principal ideal domain. For example, $\mathbb{Z}$ is a principal ideal domain, but in $\mathbb{Z}[t]$, the ideal

$$\langle 2, t \rangle = \{ f(t) \in \mathbb{Z}[t] : \text{the constant term of } f \text{ is even} \}$$

is not principal.

Exercise 3.2.4 Prove that the ideals in Warning 3.2.3 are indeed not principal.

Exercise 3.2.4: a non-principal ideal

At the end of Chapter 2, I promised I'd give you a way of manufacturing lots of new fields. Here it is!

Corollary 3.2.5 *Let K be a field and let $0 \neq f \in K[t]$. Then*

$$f \text{ is irreducible } \iff K[t]/\langle f \rangle \text{ is a field}.$$

Proof This follows from Propositions 2.3.26 and 3.2.2. □

To make new fields using Corollary 3.2.5, we'll need a way of knowing which polynomials are irreducible. That's the topic of Section 3.3. But for now, let's stick to our mission: proving that every polynomial factorizes into irreducibles in an essentially unique way.

To achieve our mission, we'll need two more lemmas.

Lemma 3.2.6 *Let K be a field and let $f(t) \in K[t]$ be a nonconstant polynomial. Then $f(t)$ is divisible by some irreducible in $K[t]$.*

Proof Let g be a nonconstant polynomial of smallest possible degree such that $g \mid f$. (For this to make sense, there must be at least one nonconstant polynomial dividing f, and there is: f.) I claim that g is irreducible. Proof: if $g = g_1 g_2$ then each g_i divides f, so by minimality of $\deg(g)$, each g_i has degree 0 or $\deg(g)$. They cannot both have degree $\deg(g)$, since $\deg(g_1) + \deg(g_2) = \deg(g) > 0$. So at least one has degree 0, which by Lemma 3.1.14(i) means that it is a unit. □

Lemma 3.2.7 *Let K be a field and $f, g, h \in K[t]$. Suppose that f is irreducible and $f \mid gh$. Then $f \mid g$ or $f \mid h$.*

This behaviour is familiar in the integers: if a prime p divides some product ab, then $p \mid a$ or $p \mid b$. In fact, our proof works in any principal ideal domain.

Proof Suppose that $f \nmid g$. Since f is irreducible, f and g are coprime. Since $K[t]$ is a principal ideal domain, Proposition 2.2.16 implies that there are $p, q \in K[t]$ satisfying

$$pf + qg = 1.$$

Multiplying both sides by h gives

$$pfh + qgh = h.$$

But $f \mid pfh$ and $f \mid gh$, so $f \mid h$. $\qquad\square$

Theorem 3.2.8 *Let K be a field and $0 \neq f \in K[t]$. Then*

$$f = a f_1 f_2 \cdots f_n$$

for some $n \geq 0$, $a \in K$ and monic irreducibles $f_1, \ldots, f_n \in K[t]$. Moreover, n and a are uniquely determined by f, and $f_1, \ldots, f_n$ are uniquely determined up to reordering.

In the case $n = 0$, the product $f_1 \cdots f_n$ should be interpreted as 1 (as in Digression 2.2.10). **Monic** means that the leading coefficient is 1.

Proof First we prove that such a factorization exists, by induction on $\deg(f)$. If $\deg(f) = 0$ then f is a constant a and we take $n = 0$. Now suppose that $\deg(f) > 0$ and assume the result for polynomials of smaller degree. By Lemma 3.2.6, there is an irreducible g dividing f, and we can assume that g is monic by dividing by a constant if necessary. Then f/g is a nonzero polynomial of smaller degree than f, so by inductive hypothesis,

$$f/g = a h_1 \cdots h_m$$

for some $a \in K$ and monic irreducibles $h_1, \ldots, h_m$. Rearranging gives

$$f = a h_1 \cdots h_m g,$$

completing the induction.

Now we prove uniqueness, again by induction on $\deg(f)$. If $\deg(f) = 0$ then f is a constant a and the only possible factorization is the one with $n = 0$. Now suppose that $\deg(f) > 0$, and take two factorizations

$$af_1 \cdots f_n = f = bg_1 \cdots g_m \tag{3.7}$$

where $a, b \in K$ and f_i, g_j are monic irreducible. Since $\deg(f) > 0$, we have $n, m \geq 1$. Now $f_n \mid bg_1 \cdots g_m$, so by Lemma 3.2.7, $f_n \mid g_j$ for some j. By rearranging, we can assume that $j = m$. But g_m is also irreducible, so $f_n = cg_m$ for some nonzero $c \in K$, and both f_n and g_m are monic, so $c = 1$. Hence $f_n = g_m$. Cancelling in (3.7) (which we can do as $K[t]$ is an integral domain) gives

$$af_1 \cdots f_{n-1} = bg_1 \cdots g_{m-1}.$$

By inductive hypothesis, $n - 1 = m - 1$, $a = b$, and the lists $f_1, \ldots, f_{n-1}$ and $g_1, \ldots, g_{m-1}$ are the same up to reordering. This completes the induction. $\qquad \square$

One way to find an irreducible factor of a polynomial $f(t) \in K[t]$ is to find a **root** (an element $a \in K$ such that $f(a) = 0$):

Lemma 3.2.9 *Let K be a field, $f(t) \in K[t]$ and $a \in K$. Then*

$$f(a) = 0 \iff (t - a) \mid f(t).$$

Proof $\Rightarrow$: suppose that $f(a) = 0$. By Proposition 3.2.1,

$$f(t) = (t - a)q(t) + r(t) \tag{3.8}$$

for some $q, r \in K[t]$ with $\deg(r) < 1$. Then r is a constant, so putting $t = a$ in (3.8) gives $r = 0$.

$\Leftarrow$: if $f(t) = (t - a)q(t)$ for some polynomial q then $f(a) = 0$. $\qquad \square$

A field is **algebraically closed** if every nonconstant polynomial has at least one root. For example, $\mathbb{C}$ is algebraically closed (the fundamental theorem of algebra). A straightforward induction shows:

Lemma 3.2.10 *Let K be an algebraically closed field and $0 \neq f \in K[t]$. Then*

$$f(t) = c(t - a_1)^{m_1} \cdots (t - a_k)^{m_k},$$

where c is the leading coefficient of f, and $a_1, \ldots, a_k$ are the distinct roots of f in K, and $m_1, \ldots, m_k \geq 1$. $\qquad \square$

3.3 Irreducible polynomials

Determining whether an integer is prime is generally hard, so it's no surprise that determining whether a polynomial is irreducible is hard too. This section presents a few techniques for doing so.

Let's begin with the simplest cases. Recall Lemma 3.1.14(ii): a polynomial over a field is irreducible if and only if it is nonconstant (has degree > 0) and cannot be expressed as a product of two nonconstant polynomials.

Lemma 3.3.1 *Let K be a field and $f \in K[t]$.*

  i. *If f is constant then f is not irreducible.*

  ii. *If $\deg(f) = 1$ then f is irreducible.*

  iii. *If $\deg(f) \geq 2$ and f has a root then f is reducible.*

  iv. *If $\deg(f) \in \{2, 3\}$ and f has no root then f is irreducible.*

Proof Parts (i) and (ii) follow from what we just recalled, and (iii) follows from Lemma 3.2.9. For (iv), suppose for a contradiction that $f = gh$ with $\deg(g), \deg(h) \geq 1$. We have $\deg(g) + \deg(h) \in \{2, 3\}$, so without loss of generality, $\deg(g) = 1$. Also without loss of generality, g is monic, say $g(t) = t + a$; but then $f(-a) = 0$, a contradiction. $\qquad\square$

> **Warning 3.3.2** To show a polynomial is irreducible, it's generally *not* enough to show it has no root. The converse of (iii) is false! For instance, $(t^2 + 1)^2 \in \mathbb{Q}[t]$ has no root but is reducible.

> **Warning 3.3.3** Make sure you've digested Warning 3.3.2!
>
> This is an extremely common mistake.

Examples 3.3.4 i. Let p be a prime. Then $f(t) = 1 + t + \cdots + t^{p-1} \in \mathbb{F}_p[t]$ is reducible, since $f(1) = 0$.

  ii. Let $f(t) = t^3 - 10 \in \mathbb{Q}[t]$. Then $\deg(f) = 3$ and f has no root in $\mathbb{Q}$, so f is irreducible by part (iv) of the lemma.

  iii. Over $\mathbb{C}$ or any other algebraically closed field, the irreducibles are exactly the polynomials of degree 1.

From now on we focus on $K = \mathbb{Q}$. Any polynomial over $\mathbb{Q}$ can be multiplied by a nonzero integer to get a polynomial over $\mathbb{Z}$, and that's often a helpful move, so we'll look at $\mathbb{Z}[t]$ too.

Definition 3.3.6 A polynomial over $\mathbb{Z}$ is **primitive** if its coefficients have no common divisor except for ± 1.

For example, $15 + 6t + 10t^2$ is primitive but $15 + 6t + 30t^2$ is not.

Lemma 3.3.7 *Let $f(t) \in \mathbb{Q}[t]$. Then there exist a primitive polynomial $F(t) \in \mathbb{Z}[t]$ and $\alpha \in \mathbb{Q}$ such that $f = \alpha F$.*

Proof Write $f(t) = \sum_i (a_i/b_i)t^i$, where $a_i \in \mathbb{Z}$ and $0 \neq b_i \in \mathbb{Z}$. Take any common multiple b of the b_is; then writing $c_i = a_i b/b_i \in \mathbb{Z}$, we have $f(t) = (1/b) \sum c_i t^i$. Now let c be the greatest common divisor of the c_is, put $d_i = c_i/c \in \mathbb{Z}$, and put $F(t) = \sum d_i t^i$. Then $F(t)$ is primitive and $f(t) = (c/b)F(t)$. $\qquad\square$

If the coefficients of a polynomial $f(t) \in \mathbb{Q}[t]$ happen to all be integers, the word 'irreducible' could mean two things: irreducibility in the ring $\mathbb{Q}[t]$ or in the ring $\mathbb{Z}[t]$. We say that f is irreducible **over** $\mathbb{Q}$ or $\mathbb{Z}$ to distinguish between the two.

Suppose we have a polynomial over $\mathbb{Z}$ that's irreducible over $\mathbb{Z}$. In principle it could still be reducible over $\mathbb{Q}$: although there's no nontrivial way of factorizing it over $\mathbb{Z}$, perhaps it can be factorized when you give yourself the freedom of non-integer coefficients. But the next result tells us that you can't.

Lemma 3.3.8 (Gauss)   *i. The product of two primitive polynomials over $\mathbb{Z}$ is primitive.*

ii. If a polynomial over $\mathbb{Z}$ is irreducible over $\mathbb{Z}$, it is irreducible over $\mathbb{Q}$.

Proof For (i), let f and g be primitive polynomials over $\mathbb{Z}$. Let p be a prime number. (We're going to show that p doesn't divide all the coefficients of fg.) Write $\pi \colon \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ for the canonical homomorphism, which induces a homomorphism $\pi_* \colon \mathbb{Z}[t] \to \mathbb{F}_p[t]$ as in Definition 3.1.7.

Since f is primitive, p does not divide all the coefficients of f. Equivalently, $\pi_*(f) \neq 0$. Similarly, $\pi_*(g) \neq 0$. But $\mathbb{F}_p[t]$ is an integral domain, so

$$\pi_*(fg) = \pi_*(f)\pi_*(g) \neq 0,$$

so p does not divide all the coefficients of fg. This holds for all primes p, so fg is primitive.

For (ii), let $f \in \mathbb{Z}[t]$ be a polynomial irreducible over $\mathbb{Z}$. Let $g, h \in \mathbb{Q}[t]$ with $f = gh$. By Lemma 3.3.7, $g = \alpha G$ and $h = \beta H$ for some $\alpha, \beta \in \mathbb{Q}$ and primitive $G, H \in \mathbb{Z}[t]$. Then $\alpha\beta = m/n$ for some coprime integers m and n, giving

$$nf = mGH.$$

(All three of these polynomials are over $\mathbb{Z}$.) Now n divides every coefficient of nf, hence every coefficient of mGH. Since m and n are coprime, n divides every coefficient of GH. But GH is primitive by (i), so $n = \pm 1$, so $f = \pm mGH$. Since f is irreducible over $\mathbb{Z}$, either G or H is constant, so g or h is constant, as required. □

Gauss's lemma quickly leads to a test for irreducibility. It involves taking a polynomial over $\mathbb{Z}$ and reducing it mod p, for some prime p. This means applying the map $\pi_*\colon \mathbb{Z}[t] \to \mathbb{F}_p[t]$ from the last proof. As we saw after Definition 3.1.7, if $f(t) = \sum a_i t^i$ then $\pi_*(f)(t) = \sum \pi(a_i)t^i$, where $\pi(a_i)$ is the congruence class of a_i mod p. I'll write $\pi(a)$ as $\bar{a}$ and $\pi_*(f)$ as $\bar{f}$. That is, $\bar{f}$ is 'f mod p'.

Proposition 3.3.9 (Mod p method) *Let* $f(t) = a_0 + a_1 t + \cdots + a_n t^n \in \mathbb{Z}[t]$. *If there is some prime p such that $p \nmid a_n$ and $\bar{f} \in \mathbb{F}_p[t]$ is irreducible, then f is irreducible over $\mathbb{Q}$.*

I'll give some examples first, then the proof.

Examples 3.3.10   i. Let's use the mod p method to show that $f(t) = 9 + 14t - 8t^3$ is irreducible over $\mathbb{Q}$. Take $p = 7$: then $\bar{f}(t) = 2 - t^3 \in \mathbb{F}_7[t]$, so it's enough to show that $2 - t^3$ is irreducible over $\mathbb{F}_7$. Since this has degree 3, it's enough to show that $t^3 = 2$ has no solution in $\mathbb{F}_7$ (by Lemma 3.3.1(iv)). And you can easily check this by computing 0^3, $(\pm 1)^3$, $(\pm 2)^3$ and $(\pm 3)^3$ mod 7.

ii. The condition in Proposition 3.3.9 that $p \nmid a_n$ can't be dropped. For instance, consider $f(t) = 6t^2 + t$ and $p = 2$.

> **Warning 3.3.11** Take $f(t)$ as in Example 3.3.10(i), but this time take $p = 3$. Then $\bar{f}(t) = -t + t^3 \in \mathbb{F}_3[t]$, which is reducible. But that doesn't mean f is reducible! The mod p method only ever lets you show that a polynomial is *irreducible* over $\mathbb{Q}$, not reducible.

Proof of Proposition 3.3.9 Take a prime p satisfying the stated conditions.

First suppose that f is primitive. By Gauss's lemma, it is enough to prove that f is irreducible over $\mathbb{Z}$.

Since $\bar{f}$ is irreducible, $\deg(\bar{f}) > 0$, so $\deg(f) > 0$.

Let $f = gh$ in $\mathbb{Z}[t]$. We have $\bar{f} = \bar{g}\bar{h}$ and $\bar{f}$ is irreducible, so without loss of generality, $\bar{g}$ is constant. The leading coefficient of f is the product of the leading

coefficients of g and h, and is not divisible by p, so the leading coefficient of g is not divisible by p. Hence $\deg(g) = \deg(\overline{g})$. But $\deg(\overline{g}) = 0$, so $\deg(g) = 0$, so $g \in \mathbb{Z}[t]$ is a constant $b \in \mathbb{Z}$. Finally, $f = gh = bh$ and f is primitive, so $b = \pm 1$, which is a unit in $\mathbb{Z}[t]$. It follows that f is irreducible over $\mathbb{Z}$.

Now take an arbitrary f satisfying the hypotheses. We have $f = cF$ where $c \in \mathbb{Z}$ is the greatest common divisor of the coefficients and $F \in \mathbb{Z}[t]$ is primitive. Then $\overline{f} = \overline{c}\overline{F}$, and $\overline{c}$ is a unit in $\mathbb{F}_p$ because $p \nmid c$. Since $\overline{f}$ is irreducible, this implies that $\overline{F}$ is irreducible, and so by what we've just proved, F is irreducible over $\mathbb{Q}$. But $c \neq 0$, so c is a unit in $\mathbb{Q}$, so $f = cF$ is also irreducible over $\mathbb{Q}$. $\qquad\square$

We finish with an irreducibility test that turns out to be surprisingly powerful.

Proposition 3.3.12 (Eisenstein's criterion) *Let* $f(t) = a_0 + \cdots + a_n t^n \in \mathbb{Z}[t]$,

with $n \geq 1$. Suppose there exists a prime p such that:

- $p \nmid a_n$;

- $p \mid a_i$ *for all* $i \in \{0, \ldots, n-1\}$;

- $p^2 \nmid a_0$.

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

To prove this, we will use the concept of the **codegree** $\mathrm{codeg}(f)$ of a polynomial $f(t) = \sum_i a_i t^i$, which is defined to be the least i such that $a_i \neq 0$ (if $f \neq 0$), or as the formal symbol ∞ if $f = 0$. For polynomials f and g over an integral domain,

$$\mathrm{codeg}(fg) = \mathrm{codeg}(f) + \mathrm{codeg}(g).$$

Clearly $\mathrm{codeg}(f) \leq \deg(f)$ unless $f = 0$.

Proof By Gauss's lemma, it is enough to show f is irreducible over $\mathbb{Z}$. Let $g, h \in \mathbb{Z}[t]$ with $f = gh$. Continue to write $\overline{f}(t) \in \mathbb{F}_p[t]$ for f reduced mod p; then $\overline{f} = \overline{g}\overline{h}$. Since

$$p^2 \nmid a_0 = f(0) = g(0)h(0),$$

we may assume without loss of generality that $p \nmid g(0)$. Hence $\mathrm{codeg}(\overline{g}) = 0$. Also, $\mathrm{codeg}(\overline{f}) = n$, since p divides each of $a_0, \ldots, a_{n-1}$ but not a_n. So

$$n = \mathrm{codeg}(\overline{f}) = \mathrm{codeg}(\overline{g}) + \mathrm{codeg}(\overline{h}) = \mathrm{codeg}(\overline{h}) \leq \deg(\overline{h}) \leq \deg(h), \quad (3.9)$$

giving $n \leq \deg(h)$. But $f = gh$ with $\deg(f) = n$, so $\deg(h) = n$ and $\deg(g) = 0.\square$

Example 3.3.14 Let

$$g(t) = \frac{2}{9}t^5 - \frac{5}{3}t^4 + t^3 + \frac{1}{3} \in \mathbb{Q}[t].$$

Then g is irreducible over $\mathbb{Q}$ if and only if

$$9g(t) = 2t^5 - 15t^4 + 9t^3 + 3$$

is irreducible over $\mathbb{Q}$, which it is by Eisenstein's criterion with $p = 3$.

Testing for irreducibility

Exercise 3.3.15 Use Eisenstein's criterion to show that for every $n \geq 1$, there is an irreducible polynomial over $\mathbb{Q}$ of degree n.

I'll give you one more example, and it's an important one.

Example 3.3.16 Let p be a prime. The **pth cyclotomic polynomial** is

$$\Phi_p(t) = 1 + t + \cdots + t^{p-1} = \frac{t^p - 1}{t - 1}. \tag{3.10}$$

I claim that Φ_p is irreducible. We can't apply Eisenstein to Φ_p as it stands, because whichever prime we choose (whether it's p or another one) doesn't divide any of the coefficients. However, we saw on p. 38 that $\Phi_p(t)$ is irreducible if and only if $\Phi_p(t - c)$ is irreducible, for any $c \in \mathbb{Q}$. We'll take $c = -1$. We have

$$\begin{aligned}
\Phi_p(t + 1) &= \frac{(t + 1)^p - 1}{(t + 1) - 1} \\
&= \frac{1}{t} \sum_{i=1}^{p} \binom{p}{i} t^i \\
&= p + \binom{p}{2}t + \cdots + \binom{p}{p - 1}t^{p-2} + t^{p-1}.
\end{aligned}$$

So $\Phi_p(t + 1)$ is irreducible by Eisenstein's criterion and Lemma 2.3.19, hence $\Phi_p(t)$ is irreducible too.

Digression 3.3.17 I defined the pth cyclotomic polynomial Φ_p only when p is prime. The definition of Φ_n for general $n \geq 1$ is not the obvious generalization of (3.10). It's this:

$$\Phi_n(t) = \prod_\zeta (t - \zeta),$$

where the product runs over all primitive nth roots of unity ζ. (In this context, 'primitive' means that n is the smallest number satisfying $\zeta^n = 1$; it's a different usage from 'primitive polynomial'.)

Many surprising things are true. It's not obvious that the coefficients of Φ_n are real, but they are. Even given that they're real, it's not obvious that they're rational, but they are. Even given that they're rational, it's not obvious that they're integers, but they are. The degree of Φ_n is $\varphi(n)$, the number of integers between 1 and n that are coprime with n (Euler's function). It's also true that the polynomial Φ_n is irreducible for *all* n, not just primes.

Some of these things are quite hard to prove, and results from Galois theory help. We won't get into all of this, but you can read more here.

Chapter 4

Field extensions

Roughly speaking, an 'extension' of a field K is a field M that contains K as a subfield. It's not much of an exaggeration to say that field extensions are the central objects of Galois theory, in much the same way that vector spaces are the central objects of linear algebra.

It will be a while before it becomes truly clear why field extensions are so important, but here are a couple of indications:

- For any polynomial f over $\mathbb{Q}$, we can take the smallest subfield M of $\mathbb{C}$ that contains all the complex roots of f, and that's an extension of $\mathbb{Q}$.

- For any irreducible polynomial f over a field K, the quotient ring $M = K[t]/\langle f \rangle$ is a field. The constant polynomials form a subfield of M isomorphic to K, so M is an extension of K.

It's important to distinguish between these two types of example. The first extends $\mathbb{Q}$ by *all* the roots of f, whereas the second extends K by just *one* root of f—as we'll see.

4.1   Definition and examples

Before we do anything else, we need to think about some set theory. What follows might seem trivial, but it's worth taking the time to get it straight.

Given a set A and a subset $B \subseteq A$, there is an **inclusion** function $\iota \colon B \to A$ defined by $\iota(b) = b$ for all $b \in B$. (Remember that any function has a specified domain and codomain, so this is not the same as the identity on B.) This inclusion function is injective.

On the other hand, given any injective function between sets, say $\varphi \colon X \to A$, the image $\operatorname{im} \varphi$ is a subset of A, and there is a bijection $\varphi' \colon X \to \operatorname{im} \varphi$ given by

$\varphi'(x) = \varphi(x)$ ($x \in X$). Hence the set X is isomorphic to (in bijection with) the subset $\operatorname{im} \varphi$ of A.

So given any subset of A, we get an injection into A, and vice versa. These two back-and-forth processes are mutually inverse (up to isomorphism), so subsets and injections are more or less the same thing.

Now here's an example to show you that the concept of subset is not as clear-cut as it might seem—at least when you look at what mathematicians actually *do*, rather than what we claim we do.

- It's common to define the set $\mathbb{C}$ as $\mathbb{R}^2$.

- Everyone treats $\mathbb{R}$ as a subset of $\mathbb{C}$.

- But almost no one would say that $\mathbb{R}$ is a subset of $\mathbb{R}^2$. (If you think it is, and you agree that $\mathbb{R}$ has an element called 6, then you must think that $\mathbb{R}^2$ has an element called 6—which you probably don't.)

So in truth, while we almost always 'know what we mean', the common conventions are inconsistent.

Probably you're thinking that this all seems rather distant from 'real mathematics'. Nothing important should depend on whether $\mathbb{R}$ is *literally* a subset of $\mathbb{C}$. I agree! But the challenge is to set up the formal definitions so that we never have to worry about irrelevant-seeming questions like this again. And the solution is to work with injections rather than subsets.

So: we intuitively want to define an 'extension' of a field K as a field M that contains K as a subfield. But if we defined it that way, we'd run into the annoying question of whether $\mathbb{C}$ really is an extension of $\mathbb{R}$. So instead, we define an extension of K to be a field M together with an injective homomorphism $K \to M$. Lemma 2.3.3 tells us that *every* homomorphism between fields is injective, so our actual definition is as follows.

Definition 4.1.1 Let K be a field. An **extension** of K is a field M together with a homomorphism $\iota \colon K \to M$.

Often we blur the distinction between injections and subsets, speaking as if K is literally a subfield of M and ι is the inclusion. We then write $M : K$ (read 'M over K') to mean that M is an extension of K, not bothering to mention ι.

Examples 4.1.2   i. The field $\mathbb{C}$, together with the inclusion $\iota \colon \mathbb{Q} \to \mathbb{C}$, is an extension of $\mathbb{Q}$. Similarly, there are field extensions $\mathbb{C} : \mathbb{R}$ and $\mathbb{R} : \mathbb{Q}$.

ii. Let
$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

Then $\mathbb{Q}(\sqrt{2})$ is a subring of $\mathbb{C}$ (easily), and in fact it's a subfield: for if $(a, b) \neq (0, 0)$ then

$$\frac{1}{a + b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2}$$

(noting that the denominators are not 0 because $\sqrt{2}$ is irrational). So we have an extension $\mathbb{C} : \mathbb{Q}(\sqrt{2})$. Also, because $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2})$, we have another extension $\mathbb{Q}(\sqrt{2}) : \mathbb{Q}$.

iii. Write

$$\mathbb{Q}(\sqrt{2}, i) = \{a + b\sqrt{2} + ci + d\sqrt{2}i : a, b, c, d \in \mathbb{Q}.\}$$

By direct calculation or later theory (which will make it much easier), $\mathbb{Q}(\sqrt{2}, i)$ is also a subfield of $\mathbb{C}$, so we have extensions $\mathbb{C} : \mathbb{Q}(\sqrt{2}, i)$ and $\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}$.

iv. Let K be a field, and consider the field $K(t)$ of rational expressions over K (Example 2.3.2). There is a homomorphism $\iota \colon K \to K(t)$ given by $\iota(a) = a/1$ ($a \in K$). In other words, $K(t)$ contains a copy of K as the constant rational expressions. So, we have a field extension $K(t) : K$.

v. There is a homomorphism $\kappa \colon \mathbb{C} \to \mathbb{C}$ defined by $\kappa(z) = \overline{z}$. So $\mathbb{C}$ together with κ is an extension of $\mathbb{C}$! You might feel that this example obeys the letter but not the spirit of Definition 4.1.1, but it *is* an example.

Exercise 4.1.3 Find two examples of fields K such that $\mathbb{Q} \subsetneqq K \subsetneqq \mathbb{Q}(\sqrt{2}, i)$. (The symbol $\subsetneqq$ means proper subset.)

Sometimes we fix a field K and think about fields that contain it—extensions of K. Other times, we fix a field K and think about fields it contains—subfields of K. It may be that we are given a mere sub*set* X of K and want to generate a subfield from it. Recalling the top-down/bottom-up distinction of Digression 2.2.12, we define this as follows.

Definition 4.1.4 Let K be a field and X a subset of K. The subfield of K **generated by** X is the intersection of all the subfields of K containing X.

Let F be the subfield of K generated by X. Since any intersection of subfields is a subfield, F really is a subfield of K. It contains X. By definition of intersection, F is the *smallest* subfield of K containing X, in the sense that any subfield of K containing X contains F.

Examples 4.1.6   i. The subfield of K generated by $\varnothing$ is the prime subfield of K.

ii. Let L be the subfield of $\mathbb{C}$ generated by $\{i\}$. I claim that

$$L = \{a + bi : a, b \in \mathbb{Q}\}.$$

To prove this, we have to show that L is the smallest subfield of $\mathbb{C}$ containing i. First, it *is* a subfield of $\mathbb{C}$ (by an argument similar to Example 4.1.2(ii)) and it contains $0 + 1i = i$. Now let L' be any subfield of $\mathbb{C}$ containing i. Then L' contains the prime subfield of $\mathbb{C}$ (by definition of prime subfield), which is $\mathbb{Q}$. So whenever $a, b \in \mathbb{Q}$, we have $a, b, i \in L'$ and so $a + bi \in L'$. Hence $L \subseteq L'$, as required.

iii. A very similar argument shows that the subfield of $\mathbb{C}$ generated by $\sqrt{2}$ is what we have been calling $\mathbb{Q}(\sqrt{2})$.

Exercise 4.1.7 What is the subfield of $\mathbb{C}$ generated by $\{7/8\}$? By $\{2 + 3i\}$? By $\mathbb{R} \cup \{i\}$?

We will be *very* interested in chains of fields

$$K \subseteq L \subseteq M$$

in which K and M are regarded as fixed and L as variable. You can think of K as the floor, M as the ceiling, and L as varying in between.

Definition 4.1.8 Let $M : K$ be a field extension and $Y \subseteq M$. We write $K(Y)$ for the subfield of M generated by $K \cup Y$, and call it K with Y **adjoined**.

So, $K(Y)$ is the smallest subfield of M containing both K and Y.

When Y is a finite set $\{\alpha_1, \ldots, \alpha_n\}$, we write $K(\{\alpha_1, \ldots, \alpha_n\})$ as $K(\alpha_1, \ldots, \alpha_n)$.

Examples 4.1.9   i. Take $M : K$ to be $\mathbb{C} : \mathbb{Q}$ and $Y = \{\sqrt{2}\}$. By definition, $K(Y)$ is the smallest subfield of $\mathbb{C}$ containing $\mathbb{Q} \cup \{\sqrt{2}\}$. But *every* subfield of $\mathbb{C}$ contains $\mathbb{Q}$: that's what it means for $\mathbb{Q}$ to be the prime subfield of $\mathbb{C}$. So, $K(Y)$ is the smallest subfield of $\mathbb{C}$ containing $\sqrt{2}$. By Example 4.1.6(iii), that's exactly what we've been calling $\mathbb{Q}(\sqrt{2})$ all along. We refer to $\mathbb{Q}(\sqrt{2})$ as '$\mathbb{Q}$ with $\sqrt{2}$ adjoined'.

ii. Similarly, $\mathbb{Q}$ with i adjoined is

$$\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$$

(Example 4.1.6(ii)), and $\mathbb{Q}$ with $\{\sqrt{2}, i\}$ adjoined is the subfield denoted by $\mathbb{Q}(\sqrt{2}, i)$ in Example 4.1.2(iii).

iii. Let M be a field and $X \subseteq M$. Write K for the prime subfield of M. Then $K(X)$ is the smallest subfield of M containing K and X. But *every* subfield of M contains K, by definition of prime subfield. So $K(X)$ is the smallest subfield of M containing X; that is, it's the subfield of M generated by X.

We already saw this argument in (i), in the case $M = \mathbb{C}$ and $X = \{\sqrt{2}\}$.

iv. Let K be any field and let M be the field $K(t)$ of rational expressions over K, which is an extension of K. You might worry that there's some ambiguity in the notation: $K(t)$ could *either* mean the field of rational expressions over K (as defined in Example 4.1.2(iv)) *or* the subfield of $K(t)$ obtained by adjoining the element t of $K(t)$ to K (as in Definition 4.1.8).

In fact, they're the same. In other words, the smallest subfield of $K(t)$ containing K and t is $K(t)$ itself. Or equivalently, the *only* subfield of $K(t)$ containing K and t is $K(t)$ itself. To see this, let L be any such subfield. For any polynomial $f(t) = \sum a_i t^i$ over K, we have $f(t) \in L$, since $a_i, t \in L$ and L is closed under multiplication and addition. Hence for any polynomials $f(t), g(t)$ over K with $g(t) \neq 0$, we have $f(t), g(t) \in L$, so $f(t)/g(t) \in L$ as L is closed under division by nonzero elements. So $L = K(t)$.

> **Warning 4.1.10** It is *not* true in general that
>
> $$K(\alpha) = \{a + b\alpha : a, b \in K\} \qquad \text{(false!)} \qquad (4.1)$$
>
> Examples like $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(i)$ do satisfy this, but that's only because $\sqrt{2}$ and i satisfy *quadratic* equations. Certainly the right-hand side is a *subset* of $K(\alpha)$, but in general it's much smaller, and isn't a subfield.
>
> You've just seen an example: the field $K(t)$ of rational expressions is much bigger than the set $\{a + bt : a, b \in K\}$ of polynomials of degree ≤ 1. And that set of polynomials isn't closed under multiplication.
>
> Another example: let ξ be the real cube root of 2. You can show that ξ^2 cannot be expressed as $a + b\xi$ for any $a, b \in \mathbb{Q}$ (a fact we'll come back to in Example 4.2.11(ii)). But $\xi \in \mathbb{Q}(\xi)$, so $\xi^2 \in \mathbb{Q}(\xi)$, so (4.1) fails in this case. In fact,
>
> $$\mathbb{Q}(\xi) = \{a + b\xi + c\xi^2 : a, b, c \in \mathbb{Q}\}.$$
>
> We'll see why next week.

Exercise 4.1.11 Let $M : K$ be a field extension. Show that $K(Y \cup Z) = (K(Y))(Z)$ whenever $Y, Z \subseteq M$. (For example, $K(\alpha, \beta) = (K(\alpha))(\beta)$ whenever $\alpha, \beta \in M$.)

Remark 4.1.12 For a field extension $M : K$, I'll generally use small Greek letters $\alpha, \beta, \ldots$ for elements of M and small English letters $a, b, \ldots$ for elements of K.

4.2 Algebraic and transcendental elements

A complex number α is said to be 'algebraic' if

$$a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$$

for some rational numbers a_i, not all zero. (You may have seen this definition with 'integer' instead of 'rational number'. It makes no difference, as you can always clear the denominators.) This concept generalizes to arbitrary field extensions:

Definition 4.2.1 Let $M : K$ be a field extension and $\alpha \in M$. Then α is **algebraic** over K if there exists $f \in K[t]$ such that $f(\alpha) = 0$ but $f \neq 0$, and **transcendental** otherwise.

Exercise 4.2.2 Show that every element of K is algebraic over K.

Examples 4.2.3   i. Let $n \geq 1$. Then $e^{2\pi i/n} \in \mathbb{C}$ is algebraic over $\mathbb{Q}$, since $f(t) = t^n - 1$ is a nonzero polynomial such that $f(e^{2\pi i/n}) = 0$.

   ii. The numbers π and e are both transcendental over $\mathbb{Q}$. Both statements are hard to prove (and we won't prove them). By Exercise 4.2.2, any complex number transcendental over $\mathbb{Q}$ is irrational. Proving the irrationality of π and e is already a challenge; proving they're transcendental is even harder.

   iii. Although π is transcendental over $\mathbb{Q}$, it is algebraic over $\mathbb{R}$, since it's an *element* of $\mathbb{R}$. (Again, we're using Exercise 4.2.2.) Moral: you shouldn't say an element of a field is just 'algebraic' or 'transcendental'; you should say it's 'algebraic/transcendental *over K*', specifying your K. Or at least, you should do this when there's any danger of confusion.

   iv. Take the field $K(t)$ of rational expressions over a field K. Then $t \in K(t)$ is transcendental over K, since $f(t) = 0 \iff f = 0$.

The set of complex numbers algebraic over $\mathbb{Q}$ is written as $\overline{\mathbb{Q}}$. It's a fact that $\overline{\mathbb{Q}}$ is a subfield of $\mathbb{C}$, but this is extremely hard to prove by elementary means. Next week I'll show you that with a surprisingly small amount of abstract algebra, you can transform this from a very hard problem into an easy one (Proposition 5.2.7).

So that you appreciate the miracle later, I give you this unusual exercise now.

Exercise 4.2.4 Attempt to prove any part of the statement that $\overline{\mathbb{Q}}$ is a subfield of $\mathbb{C}$. For example, try to show that $\overline{\mathbb{Q}}$ is closed under addition, or multiplication, or reciprocals. I have no idea how to do any of these using only our current tools, but it's definitely worth a few minutes of doomed effort to get a sense of the difficulties.

Digression 4.2.5 The field $\overline{\mathbb{Q}}$ is, in fact, algebraically closed, as you'll see in Workshop 3. So you might ask whether it's possible for *every* field K to build an algebraically closed field containing K. It turns out that it is. Better still, there is a unique 'smallest' algebraically closed field containing K, called its **algebraic closure** $\overline{K}$. For example, the algebraic closure of $\mathbb{Q}$ is $\overline{\mathbb{Q}}$. We won't have time to do algebraic closure properly, but you can read about it in most Galois theory texts.

Let $M : K$ be a field extension and $\alpha \in M$. An **annihilating polynomial** of α is a polynomial $f \in K[t]$ such that $f(\alpha) = 0$. So, α is algebraic if and only if it has some nonzero annihilating polynomial.

It is natural to ask not only *whether* α is annihilated by some nonzero polynomial, but *which* polynomials annihilate it. The situation is pleasantly simple:

Lemma 4.2.6 *Let $M : K$ be a field extension and $\alpha \in M$. Then there is a polynomial $m(t) \in K[t]$ such that*

$$\langle m \rangle = \{\text{annihilating polynomials of } \alpha \text{ over } K\}. \tag{4.2}$$

If α is transcendental over K then $m = 0$. If α is algebraic over K then there is a unique monic polynomial m satisfying (4.2).

Proof By the universal property of polynomial rings (Proposition 3.1.6), there is a unique homomorphism

$$\theta \colon K[t] \to M$$

such that $\theta(a) = a$ for all $a \in K$ and $\theta(t) = \alpha$. (Here we're taking the 'φ' of Proposition 3.1.6 to be the inclusion $K \to M$.) Then

$$\theta\left(\sum a_i t^i\right) = \sum a_i \alpha^i$$

for all $\sum a_i t^i \in K[t]$, so

$$\ker \theta = \{\text{annihilating polynomials of } \alpha \text{ over } K\}.$$

But $\ker \theta$ is an ideal of the principal ideal domain $K[t]$ (using Proposition 3.2.2), so $\ker \theta = \langle m \rangle$ for some $m \in K[t]$.

If α is transcendental then $\ker \theta = \{0\}$, so $m = 0$.

If α is algebraic then $m \neq 0$. Multiplying a polynomial by a nonzero constant does not change the ideal it generates (by Exercise 2.2.15 and Lemma 3.1.14(i)), so we can assume that m is monic. It remains to prove that m is the *only* monic polynomial such that $\langle m \rangle = \ker \theta$. If $\widetilde{m}$ is another monic polynomial such that $\langle \widetilde{m} \rangle = \ker \theta$ then $\widetilde{m} = cm$ for some nonzero constant c (again by Exercise 2.2.15 and Lemma 3.1.14(i)), and both are monic, so $c = 1$ and $\widetilde{m} = m$. □

Definition 4.2.7 Let $M : K$ be a field extension and let $\alpha \in M$ be algebraic over K. The **minimal polynomial** of α is the unique monic polynomial m satisfying (4.2).

Warning 4.2.8 We do not define the minimal polynomial of a transcendental element. So for an arbitrary field extension $M : K$, some elements of M may have no minimal polynomial.

Exercise 4.2.9 What is the minimal polynomial of an element of K?

This is an important definition, so we give some equivalent conditions.

Lemma 4.2.10 *Let $M : K$ be a field extension, let $\alpha \in M$ be algebraic over K, and let $m \in K[t]$ be a monic polynomial. The following are equivalent:*

   i. m is the minimal polynomial of α over K;

  ii. $m(\alpha) = 0$, and $m \mid f$ for all annihilating polynomials f of α over K;

 iii. $m(\alpha) = 0$, and $\deg(m) \leq \deg(f)$ for all nonzero annihilating polynomials f of α over K;

 iv. $m(\alpha) = 0$ and m is irreducible over K.

Part (iii) says the minimal polynomial is a monic annihilating polynomial of least degree.

Proof (i)$\Rightarrow$(ii) follows from the definition of minimal polynomial.

(ii)$\Rightarrow$(iii) because if $m \mid f \neq 0$ then $\deg(m) \leq \deg(f)$.

(iii)$\Rightarrow$(iv): assume (iii). First, m is not constant: for if m is constant then $m = 1$ (since m is monic); but $m(\alpha) = 0$, so $1 = 0$ in K, a contradiction. Next, suppose that $m = fg$ for some $f, g \in K[t]$. Then $0 = m(\alpha) = f(\alpha)g(\alpha)$, so without loss of generality, $f(\alpha) = 0$. By (iii), $\deg(f) \geq \deg(m)$, so $\deg(f) = \deg(m)$ and $\deg(g) = 0$. This proves (iv).

(iv)$\Rightarrow$(i): assume (iv), and write m_α for the minimal polynomial of α. We have $m_\alpha \mid m$ by definition of m_α and since $m(\alpha) = 0$. But m is irreducible and m_α is not constant, so m is a nonzero constant multiple of m_α. Since both are monic, $m = m_\alpha$, proving (i). $\qquad\qquad\square$

Examples 4.2.11   i. The minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $t^2 - 2$. There are several ways to see this.

One argument: $t^2 - 2$ is a monic annihilating polynomial of $\sqrt{2}$, and no nonzero polynomial of degree ≤ 1 over $\mathbb{Q}$ annihilates $\sqrt{2}$ since it is irrational. Then use Lemma 4.2.10(iii).

Another: $t^2 - 2$ is an irreducible monic annihilating polynomial. It is irreducible because $t^2 - 2$ has degree 2 and has no rational roots (using Lemma 3.3.1(iv)). Then use Lemma 4.2.10(iv).

ii. The minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $t^3 - 2$. This will follow from Lemma 4.2.10(iv) as long as $t^3 - 2$ is irreducible, which you can show using either Lemma 3.3.1(iv) or Eisenstein.

But unlike in (i), it's *not* so easy to show directly that $t^3 - 2$ is the annihilating polynomial of least degree. Try proving with your bare hands that $\sqrt[3]{2}$ satisfies no quadratic equation over $\mathbb{Q}$, i.e. that the equation

$$\sqrt[3]{2}^2 = a\sqrt[3]{2} + b$$

has no solution for $a, b \in \mathbb{Q}$. It's not impossible, but it's a mess. (You naturally begin by cubing both sides, but look what happens next...) So the theory really gets us something here.

iii. Let p be a prime number, and put $\omega = e^{2\pi i/p} \in \mathbb{C}$. Then ω is a root of $t^p - 1$, but that is not the minimal polynomial of ω, since it is reducible:

$$t^p - 1 = (t - 1)m(t)$$

where

$$m(t) = t^{p-1} + \cdots + t + 1.$$

Since $\omega^p - 1 = 0$ but $\omega - 1 \neq 0$, we must have $m(\omega) = 0$. By Example 3.3.16, m is irreducible over $\mathbb{Q}$. Hence m is the minimal polynomial of ω over $\mathbb{Q}$.

4.3 Simple extensions

Suppose I give you a field K and a nonconstant polynomial f over K. Can you find an extension of K containing a root of f?

If $K = \mathbb{Q}$, it's easy. The fundamental theorem of algebra guarantees that f has a root α in $\mathbb{C}$, so you can take your extension to be $\mathbb{C}$. Or, if you're feeling economical, you can take $\mathbb{Q}(\alpha)$ as your extension, that being the *smallest* subfield of $\mathbb{C}$ containing your root α.

But what if K is not $\mathbb{Q}$?

It's a bit like this. Say you want to go rock-climbing. If you live next to Arthur's Seat, no problem: just walk out of your door and get started. There's a ready-made solution. But if you live in the middle of the fields in the Netherlands, you're going to have to build your own climbing wall.

When $K = \mathbb{Q}$, we have a ready-made algebraically closed field $\mathbb{C}$ containing K, so it's easy to find an extension of K containing a root of f. For a general K, it's not so easy. We're going to have to build an extension of our own. But it's not so hard either!

Rather than taking a general polynomial f, we will just consider irreducibles. (Theorem 3.2.8 guarantees that f has some irreducible factor m, and any root of m is automatically a root of f.) We will also restrict to *monic* irreducibles, which makes no real difference to anything.

So, we have a field K and a monic irreducible polynomial $m \in K[t]$. We are trying to construct an extension M of K and an element $\alpha \in M$ such that $m(\alpha) = 0$. By Lemma 4.2.10, m will then be the minimal polynomial of α.

This construction can be done as follows. By Corollary 3.2.5, the quotient $K[t]/\langle m \rangle$ is a field. We have ring homomorphisms

$$K \to K[t] \xrightarrow{\pi} K[t]/\langle m \rangle, \tag{4.3}$$

where the first homomorphism sends $a \in K$ to the constant polynomial $a \in K[t]$ and π is the canonical homomorphism. Their composite is a homomorphism of fields $K \to K[t]/\langle m \rangle$. So, we have a field extension $\big(K[t]/\langle m \rangle\big) : K$. And one of the elements of $K[t]/\langle m \rangle$ is $\pi(t)$, which I will call α.

For a polynomial $\sum a_i t^i \in K[t]$,

$$\pi\left(\sum_i a_i t^i\right) = \sum_i a_i \alpha^i. \tag{4.4}$$

Since π is surjective, every element of $K[t]/\langle m \rangle$ is of the form $\sum a_i \alpha^i$. You can think of $K[t]/\langle m \rangle$ as the ring of polynomials over K, but with two polynomials seen as equal if they differ by a multiple of m. (Compare how you think of $\mathbb{Z}/\langle p \rangle$.)

Part (i) of the following lemma says that α is a root of m, and that if we're looking for an extension of K containing a root of m, then $K[t]/\langle m \rangle$ is an economical choice: it's no bigger than it needs to be.

Part (ii) answers an analogous but easier question: what if we start with a field K and want to extend it by an element that satisfies *no* nonzero polynomial over K?

Lemma 4.3.1 *Let K be a field.*

i. *Let $m \in K[t]$ be monic and irreducible. Write $\alpha \in K[t]/\langle m \rangle$ for the image of t under the canonical homomorphism $K[t] \to K[t]/\langle m \rangle$. Then α has minimal polynomial m over K, and $K[t]/\langle m \rangle$ is generated by α over K.*

ii. *The element t of the field $K(t)$ of rational expressions over K is transcendental over K, and $K(t)$ is generated by t over K.*

In part (i), we are viewing $K[t]/\langle m \rangle$ as an extension of K, as in (4.3).

Proof For (i), write $M = K[t]/\langle m \rangle$. Equation (4.4) implies that the set of annihilating polynomials of α over K is $\ker \pi$, which is $\langle m \rangle$. So m is by definition the minimal polynomial of α over K.

Any subfield L of M containing K and α contains every polynomial in α over K, so $L = M$. Hence M is generated by α over K.

For (ii), we have already seen that t is transcendental over K (Example 4.2.3(iv)).

Let L be a subfield of $K(t)$ containing K and t. Then any polynomials $f, g \in K[t]$ are in L, so if $g \neq 0$ then $f/g \in L$. Hence $L = M$, and M is generated by t over K. $\qquad\square$

So far, we've seen that given a monic irreducible polynomial m over a field K, we can build an extension of K containing a root of m. In fact, there are many such extensions. For instance:

Example 4.3.2 If $K = \mathbb{Q}$ and $m(t) = t^2 - 2$ then all three of the extensions $\mathbb{Q}(\sqrt{2})$, $\mathbb{R}$ and $\mathbb{C}$ contain a root of m.

But $K[t]/\langle m \rangle$ is the *canonical* or *minimal* choice. In fact, $K[t]/\langle m \rangle$ has a universal property. To express it, we need a definition.

Definition 4.3.3 Let K be a field, and let $\iota\colon K \to M$ and $\iota'\colon K \to M'$ be extensions of K. A homomorphism $\varphi\colon M \to M'$ is said to be a **homomorphism over K** if

$$
\begin{array}{ccc}
M & \xrightarrow{\ \varphi\ } & M' \\
& \nwarrow_{\iota} \quad \nearrow_{\iota'} & \\
& K &
\end{array}
$$

commutes.

For the triangle to commute means that $\varphi(\iota(a)) = \iota'(a)$ for all $a \in K$. Very often, the homomorphisms ι and ι' are thought of as inclusions, and we write both $\iota(a)$ and $\iota'(a)$ as just a. Then for φ to be a homomorphism over K means that $\varphi(a) = a$ for all $a \in K$.

Example 4.3.4 Define $\kappa \colon \mathbb{C} \to \mathbb{C}$ by $\kappa(z) = \overline{z}$. Then κ is a homomorphism, and it is a homomorphism over $\mathbb{R}$ since $\overline{a} = a$ for all $a \in \mathbb{R}$.

Exercise 4.3.5 Let $M : K$ and $L : K$ be field extensions, and let $\varphi \colon M \to L$ be a homomorphism over K. Show that if $\alpha \in M$ has minimal polynomial m over K then $\varphi(\alpha) \in L$ also has minimal polynomial m over K.

Here's an extremely useful lemma about homomorphisms over a field.

Lemma 4.3.6 *Let M and M' be extensions of a field K, and let $\varphi, \psi \colon M \to M'$ be homomorphisms over K. Let Y be a subset of M such that $M = K(Y)$. If $\varphi(\alpha) = \psi(\alpha)$ for all $\alpha \in Y$ then $\varphi = \psi$.*

Proof We have $\varphi(a) = a = \psi(a)$ for all $a \in K$, since φ and ψ are homomorphisms over K. But we are assuming that $\varphi(\alpha) = \psi(\alpha)$ for all $\alpha \in Y$, so $K \cup Y$ is a subset of the equalizer $\mathrm{Eq}\{\varphi, \psi\}$ (Definition 2.3.7). Hence by Lemma 2.3.8, $\mathrm{Eq}\{\varphi, \psi\}$ is a subfield of M containing $K \cup Y$. But $K(Y)$ is the smallest subfield of M containing $K \cup Y$, so $\mathrm{Eq}\{\varphi, \psi\} = K(Y) = M$. Hence $\varphi = \psi$. $\qquad\square$

Now we can formulate the universal property of $K[t]/\langle m \rangle$, and similarly that of $K(t)$.

Proposition 4.3.7 (Universal properties of $K[t]/\langle m \rangle$ and $K(t)$) *Let K be a field.*

i. *Let $m \in K[t]$ be monic and irreducible, let $L : K$ be an extension of K, and let $\beta \in L$ with minimal polynomial m. Write α for the image of t under the canonical homomorphism $K[t] \to K[t]/\langle m \rangle$. Then there is exactly one homomorphism $\varphi \colon K[t]/\langle m \rangle \to L$ over K such that $\varphi(\alpha) = \beta$.*

ii. *Let $L : K$ be an extension of K, and let $\beta \in L$ be transcendental. Then there is exactly one homomorphism $\varphi \colon K(t) \to L$ over K such that $\varphi(t) = \beta$.*

Diagram for (i):

$$
\begin{array}{ccc}
 & & \beta \\
\alpha & & \\
 & \varphi & L \\
K[t]/\langle m \rangle & & \\
 & & \\
 & K & \\
\end{array}
$$
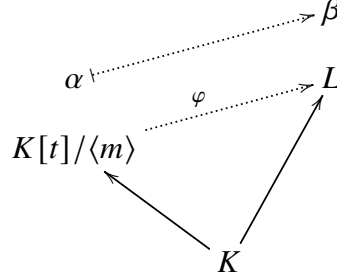
I've drawn L higher than $K[t]/\langle m \rangle$ to convey the idea that L may be bigger. Before I give the proof, here's an example.

Example 4.3.8 Let $K = \mathbb{Q}$ and $m(t) = t^2 - 2$. Let $L = \mathbb{C}$ and let $\beta = -\sqrt{2} \in \mathbb{C}$. Proposition 4.3.7(i) tells us that there is a unique field homomorphism

$$\varphi \colon \mathbb{Q}[t]/\langle t^2 - 2 \rangle \to \mathbb{C}$$

over $\mathbb{Q}$ mapping the equivalence class of t to $-\sqrt{2}$.

Proof of Proposition 4.3.7 For (i), first we show there is *at least* one homomorphism $\varphi \colon K[t]/\langle m \rangle \to L$ over K such that $\varphi(\alpha) = \beta$. By the universal property of polynomial rings (Proposition 3.1.6), there is exactly one homomorphism $\theta \colon K[t] \to L$ such that $\theta(a) = a$ for all $a \in K$ and $\theta(t) = \beta$. Then $\theta(m(t)) = m(\beta) = 0$, so $\langle m \rangle \subseteq \ker \theta$. Hence by the universal property of quotients (p. 22), there is exactly one homomorphism $\varphi \colon K[t]/\langle m \rangle \to L$ such that

$$
\begin{array}{ccc}
K[t] & & \\
\pi \downarrow & \searrow^{\theta} & \\
K[t]/\langle m \rangle & \xrightarrow{\varphi} & L \\
\end{array}
$$

commutes. Then φ is a homomorphism over K, since for all $a \in K$ we have

$$\varphi(a) = \varphi(\pi(a)) = \theta(a) = a.$$

Moreover,

$$\varphi(\alpha) = \varphi(\pi(t)) = \theta(t) = \beta,$$

so $\varphi(\alpha) = \beta$.

Now we show there is *at most* one homomorphism $K[t]/\langle m \rangle \to L$ over K such that $\alpha \mapsto \beta$. Let φ and φ' be two such. Then $\varphi(\alpha) = \varphi'(\alpha)$, and α generates $K[t]/\langle m \rangle$ over K (Lemma 4.3.1(i)), so $\varphi = \varphi'$ by Lemma 4.3.6.

For (ii), first we show there is *at least* one homomorphism $\varphi \colon K(t) \to L$ over K such that $\varphi(t) = \beta$. Every element of $K(t)$ can be represented as f/g where

$f, g \in K[t]$ with $g \neq 0$. Since β is transcendental over K, we have $g(\beta) \neq 0$, and so $f(\beta)/g(\beta)$ is a well-defined element of L. One can check that this gives a well-defined homomorphism

$$\begin{aligned} \varphi \colon \quad K(t) &\rightarrow \quad L \\ \frac{f(t)}{g(t)} &\mapsto \quad \frac{f(\beta)}{g(\beta)}. \end{aligned}$$

Evidently φ is a homomorphism over K (that is, $\varphi(a) = a$ for all $a \in K$), and evidently $\varphi(t) = \beta$.

The proof that there is *at most* one homomorphism $K(t) \rightarrow L$ over K such that $t \mapsto \beta$ is similar to the uniqueness proof in part (i). □

Exercise 4.3.9 Fill in the details of the last paragraph of that proof.

We know now that for a monic irreducible polynomial m over K, the extension $K[t]/\langle m \rangle$ contains a root of m and is generated by that root. As we're about to see, Proposition 4.3.7 implies that $K[t]/\langle m \rangle$ is the *only* extension of K with this property. But the word 'only' has to be interpreted in an up-to-isomorphism sense (as you're used to from statements like 'there is only one group of order 5'). The appropriate notion of isomorphism is as follows.

Let M and M' be extensions of a field K. A homomorphism $\varphi \colon M \rightarrow M'$ is an **isomorphism over K** if it is a homomorphism over K and an isomorphism of fields. (You can check that φ^{-1} is then also a homomorphism over K.) If such a φ exists, we say that M and M' are **isomorphic over K**.

Warning 4.3.10 Let M and M' be extensions of a field K. It can happen that M and M' are isomorphic, but not isomorphic over K. In other words, just because it's possible to find an isomorphism $\varphi \colon M \rightarrow M'$, it doesn't mean you can find one making the triangle in Definition 4.3.3 commute. A question on Workshop 3 leads you through a counterexample.

Corollary 4.3.11 *Let K be a field.*

i. *Let $m \in K[t]$ be monic and irreducible, let $L : K$ be an extension of K, and let $\beta \in L$ with minimal polynomial m and with $L = K(\beta)$. Write α for the image of t under the canonical homomorphism $K[t] \rightarrow K[t]/\langle m \rangle$. Then there is exactly one isomorphism $\varphi \colon K[t]/\langle m \rangle \rightarrow L$ over K such that $\varphi(\alpha) = \beta$.*

ii. Let $L : K$ be an extension of K, and let $\beta \in L$ be transcendental with $L = K(\beta)$. Then there is exactly one isomorphism $\varphi \colon K(t) \to L$ over K such that $\varphi(t) = \beta$.

(Spot the differences between this corollary and Proposition 4.3.7...)

Proof For (i), Proposition 4.3.7(i) implies that there is a unique homomorphism $\varphi \colon K[t]/\langle m \rangle \to L$ over K such that $\varphi(\alpha) = \beta$. So we only have to show that φ is an isomorphism. Since homomorphisms of fields are injective, we need only show that φ is surjective. Now by Lemma 2.3.6(i), $\operatorname{im} \varphi$ is a subfield of L, and it contains both K (since φ is a homomorphism over K) and β (since $\varphi(\alpha) = \beta$). But $L = K(\beta)$, so $\operatorname{im} \varphi = L$.

The proof of (ii) is similar. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Examples 4.3.12   i. Let m be a monic irreducible polynomial over $\mathbb{Q}$. Choose a complex root β of m. Then the subfield $\mathbb{Q}(\beta)$ of $\mathbb{C}$ is an extension of $\mathbb{Q}$ generated by β. So by Corollary 4.3.11(i), $\mathbb{Q}[t]/\langle m \rangle \cong \mathbb{Q}(\beta)$.

ii. Let β be a transcendental complex number. Then by Corollary 4.3.11(ii), the field $\mathbb{Q}(t)$ of rational expressions is isomorphic to $\mathbb{Q}(\beta) \subseteq \mathbb{C}$.

Field extensions generated by a single element have a special name.

Definition 4.3.13 A field extension $M : K$ is **simple** if there exists $\alpha \in M$ such that $M = K(\alpha)$.

Examples 4.3.14   i. Surprisingly many extensions are simple. For instance, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}$ is a simple extension (despite appearances), because in fact $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$.

ii. $K(t) : K$ is simple, where $K(t)$ is the field of rational expressions over K.

> **Exercise 4.3.15** Prove that $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$. Hint: begin by finding $(\sqrt{2} + \sqrt{3})^3$.

How to understand simple algebraic extensions

We've now shown that simple extensions can be classified completely:

Theorem 4.3.16 (Classification of simple extensions) *Let K be a field.*

i. Let $m \in K[t]$ be a monic irreducible polynomial. Then there exist an extension $M : K$ and an algebraic element $\alpha \in M$ such that $M = K(\alpha)$ and α has minimal polynomial m over K.

Moreover, if (M, α) and (M', α') are two such pairs, there is exactly one isomorphism $\varphi \colon M \to M'$ over K such that $\varphi(\alpha) = \alpha'$.

Proof For (i), we can take $M = K[t]/\langle m \rangle$ and α to be the image of t under the canonical homomorphism $K[t] \to M$. Lemma 4.3.1(i) implies that α has minimal polynomial m over K and that $M = K(\alpha)$, and Corollary 4.3.11(i) gives 'Moreover'.

Part (ii) follows from Lemma 4.3.1(ii) and Corollary 4.3.11(ii) in the same way. □

Conclusion: given any field K (not necessarily $\mathbb{Q}$!) and any monic irreducible $m(t) \in K[t]$, we can say the words '**adjoin to K a root α of m**', and this unambiguously defines an extension $K(\alpha) : K$. (At least, unambiguously up to isomorphism over K—but who could want more?) Similarly, we can unambiguously adjoin to K a transcendental element.

Examples 4.3.17   i. Let K be any field not containing a square root of 2. Then $t^2 - 2$ is irreducible over K. So we can adjoin to K a root of $t^2 - 2$, giving an extension $K(\sqrt{2}) : K$.

We have already seen this example many times when $K = \mathbb{Q}$, in which case $K(\sqrt{2})$ can be seen as a subfield of $\mathbb{C}$. But the construction works for *any* K. For instance, 2 has no square root in $\mathbb{F}_3$, so there is an extension $\mathbb{F}_3(\sqrt{2})$ of $\mathbb{F}_3$. It can be constructed as $\mathbb{F}_3[t]/\langle t^2 - 2 \rangle$.

ii. The polynomial $m(t) = 1 + t + t^2$ is irreducible over $\mathbb{F}_2$, so we may adjoin to $\mathbb{F}_2$ a root α of m. Then $\mathbb{F}_2(\alpha) = \mathbb{F}_2[t]/\langle 1 + t + t^2 \rangle$.

> **Exercise 4.3.18** How many elements does the field $\mathbb{F}_3(\sqrt{2})$ have? What about $\mathbb{F}_2(\alpha)$, where α is a root of $1 + t + t^2$?

> **Warning 4.3.19** Take $K = \mathbb{Q}$ and $m(t) = t^3 - 2$, which is irreducible. Write $\alpha_1, \alpha_2, \alpha_3$ for the roots of m in $\mathbb{C}$. Then $\mathbb{Q}(\alpha_1)$, $\mathbb{Q}(\alpha_2)$ and $\mathbb{Q}(\alpha_3)$ are all different *as subsets of* $\mathbb{C}$. For example, one of the α_i is the real cube root of 2 (say α_1), which implies that $\mathbb{Q}(\alpha_1) \subseteq \mathbb{R}$, whereas the other two are not real, so $\mathbb{Q}(\alpha_i) \not\subseteq \mathbb{R}$ for $i \neq 1$. However, $\mathbb{Q}(\alpha_1) : \mathbb{Q}$, $\mathbb{Q}(\alpha_2) : \mathbb{Q}$ and $\mathbb{Q}(\alpha_3) : \mathbb{Q}$ are all isomorphic *as abstract field extensions* of $\mathbb{Q}$. This follows from Theorem 4.3.16, since all the α_i have the same minimal polynomial, m.

You're already very familiar with this kind of situation in other branches of algebra. For instance, in linear algebra, take three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in $\mathbb{R}^2$, none a scalar multiple of any other. Then span($\mathbf{v}_1$), span($\mathbf{v}_2$) and span($\mathbf{v}_3$) are all different *as subsets of* $\mathbb{R}^2$, but they are all isomorphic *as abstract vector spaces* (since they're all 1-dimensional). A similar example could be given with a group containing several subgroups that are all isomorphic.

You've seen that Galois theory involves aspects of group theory and ring theory. In the next chapter, you'll see how linear algebra enters the picture too.

Chapter 5

Degree

We've seen that if you adjoin to $\mathbb{Q}$ a *square* root of 2, then each element of the resulting field can be specified using *two* rational numbers, a and b:

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

I also mentioned that if you adjoin to $\mathbb{Q}$ a *cube* root of 2, then it takes *three* rational numbers to specify each element of the resulting field:

$$\mathbb{Q}(\sqrt[3]{2}) = \left\{a + b\sqrt[3]{2} + c\sqrt[3]{2}^2 : a, b, c \in \mathbb{Q}\right\}$$

(Warning 4.1.10). This might lead us to suspect that $\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}$ is in some sense a 'bigger' extension than $\mathbb{Q}(\sqrt{2}) : \mathbb{Q}$.

The first thing we'll do in this chapter is to make this intuition rigorous. We'll define the 'degree' of an extension and see that $\mathbb{Q}(\sqrt{2}) : \mathbb{Q}$ and $\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}$ have degrees 2 and 3, respectively.

The concept of degree is incredibly useful, and not only in Galois theory. In fact, I'll show you how it can be used to solve three problems that remained unsolved for literally millennia, since the time of the ancient Greeks.

5.1 The degree of an extension

Let $M : K$ be a field extension. Then M is a vector space over K in a natural way. Addition and subtraction in the vector space M are the same as in the field M. Scalar multiplication in the vector space is just multiplication of elements of M by elements of K, which makes sense because K is embedded as a subfield of M.

This little observation is amazingly useful. It is an excellent illustration of a powerful mathematical technique: forgetting. When we view M as a vector space over K rather than a field extension of K, we are forgetting how to multiply together elements of M that aren't in K.

Definition 5.1.1 The **degree** $[M : K]$ of a field extension $M : K$ is the dimension of M as a vector space over K.

If M is a *finite-dimensional* vector space over K, it's clear what this means. If M is infinite-dimensional over K, we write $[M : K] = \infty$, where ∞ is a formal symbol which we give the properties

$$n < \infty, \qquad n \cdot \infty = \infty \quad (n \geq 1), \qquad \infty \cdot \infty = \infty$$

for integers n. An extension $M : K$ is **finite** if $[M : K] < \infty$.

> **Digression 5.1.2** You know that whenever V is a finite-dimensional vector space, (i) there exists a basis of V, and (ii) there is a bijection between any two bases. This makes it possible to define the dimension of a vector space as the number of elements in a basis. In fact, both (i) and (ii) are true for *every* vector space, not just the finite-dimensional ones. So we can define the dimension of an arbitrary vector space as the 'number' of elements in a basis, where now 'number' means cardinal, i.e. isomorphism class of sets.
>
> We could interpret Definition 5.1.1 using this general definition of dimension. For instance, suppose we had one field extension $M : K$ such that M had a countably infinite basis over K, and another, $M' : K$, such that M' had an uncountably infinite basis over K. Then $[M : K]$ and $[M' : K]$ would be different.
>
> However, we'll lump all the infinite-dimensional extensions together and say that their degrees are all ∞. We'll mostly be dealing with finite extensions anyway, and won't need to distinguish between sizes of ∞. It's a bit like the difference between a house that costs a million pounds and a house that costs ten million: although the difference is huge, most of us would lump them together in a single category called 'unaffordable'.

Examples 5.1.3   i. Every field M contains at least one nonzero element, namely, 1. So $[M : K] \geq 1$ for every field extension $M : K$.

If $M = K$ then $\{1\}$ is a basis, so $[M : K] = 1$. On the other hand, if $[M : K] = 1$ then the one-element linearly independent set $\{1\}$ must be a basis, which implies that every element of M is equal to $a \cdot 1 = a$ for some $a \in K$, and so $M = K$. Hence

$$[M : K] = 1 \iff M = K.$$

ii. Every element of $\mathbb{C}$ is equal to $x + yi$ for a unique pair (x, y) of elements of $\mathbb{R}$. That is, $\{1, i\}$ is a basis of $\mathbb{C}$ over $\mathbb{R}$. Hence $[\mathbb{C} : \mathbb{R}] = 2$.

iii. Let K be a field and $K(t)$ the field of rational expressions over K. Then $1, t, t^2, \ldots$ are linearly independent over K, so $[K(t) : K] = \infty$.

Warning 5.1.4 The degree $[K : K]$ of K over itself is 1, not 0. Degrees of extensions are never 0. See Example 5.1.3(i).

Theorem 5.1.5 *Let $K(\alpha) : K$ be a simple extension, with α algebraic over K. Write $m \in K[t]$ for the minimal polynomial of α and $n = \deg(m)$. Then*

$$1, \alpha, \ldots, \alpha^{n-1}$$

is a basis of $K(\alpha)$ over K. In particular, $[K(\alpha) : K] = \deg(m)$.

Proof To show that $1, \alpha, \ldots, \alpha^{n-1}$ is a basis of $K(\alpha)$ over K, we will show that every element of $K(\alpha)$ can be expressed as a K-linear combination of $1, \alpha, \ldots, \alpha^{n-1}$ in a unique way.

By Lemma 4.3.1(i) and Theorem 4.3.16(i), we might as well take $K(\alpha) = K[t]/\langle m \rangle$ and $\alpha = \pi(t)$, where $\pi \colon K[t] \to K[t]/\langle m \rangle$ is the canonical homomorphism.

Since π is surjective, every element of $K(\alpha)$ is equal to $\pi(f)$ for some $f \in K[t]$. By Proposition 3.2.1, there are unique $q, r \in K[t]$ such that $f = qm + r$ and $\deg(r) < n$. In particular, there is a unique polynomial $r \in K[t]$ such that $f - r \in \langle m \rangle$ and $\deg(r) < n$. Equivalently, there are unique $a_0, \ldots, a_{n-1} \in K$ such that

$$f(t) - \left(a_0 + a_1 t + \cdots + a_{n-1} t^{n-1}\right) \in \langle m \rangle.$$

Equivalently, there are unique $a_0, \ldots, a_{n-1} \in K$ such that

$$\pi(f) = \pi\left(a_0 + a_1 t + \cdots + a_{n-1} t^{n-1}\right).$$

Equivalently (since $\pi(t) = \alpha$), there are unique $a_0, \ldots, a_{n-1} \in K$ such that

$$\pi(f) = a_0 + a_1 \alpha + \cdots a_{n-1} \alpha^{n-1},$$

as required. $\qquad\qquad\square$

Examples 5.1.6 i. Let $\alpha \in \mathbb{C}$ be an algebraic number over $\mathbb{Q}$ whose minimal polynomial is quadratic. Then by Theorem 5.1.5,

$$\mathbb{Q}(\alpha) = \{a + b\alpha : a, b \in \mathbb{Q}\}.$$

We've already seen this in many examples, such as $\alpha = \sqrt{2}$ and $\alpha = i$.

ii. Let p be a prime. We saw in Example 4.2.11(iii) that $e^{2\pi i/p}$ has minimal polynomial $1+t+\cdots+t^{p-1}$. This has degree $p-1$, so $[\mathbb{Q}(e^{2\pi i/p}) : \mathbb{Q}] = p-1$.

> **Warning 5.1.7** For a prime p, the degree of $e^{2\pi i/p}$ over $\mathbb{Q}$ is $p - 1$, not p!

Example 5.1.8 Apart from finite fields of the form $\mathbb{F}_p$, the simplest finite field is $\mathbb{F}_2(\alpha)$, where α is a root of the irreducible polynomial $1 + t + t^2$ over $\mathbb{F}_2$ (Example 4.3.17(ii)). By Theorem 5.1.5,

$$\mathbb{F}_2(\alpha) = \{a + b\alpha : a, b \in \mathbb{F}_2\} = \{0, 1, \alpha, 1 + \alpha\}.$$

Since $1 + \alpha + \alpha^2 = 0$ and $\mathbb{F}_2(\alpha)$ has characteristic 2,

$$\alpha^2 = 1 + \alpha, \qquad (1 + \alpha)^2 = \alpha.$$

So the Frobenius automorphism of $\mathbb{F}_2(\alpha)$ interchanges α and $1 + \alpha$. Like all automorphisms, it fixes 0 and 1.

> **Exercise 5.1.9** Write out the addition and multiplication tables of $\mathbb{F}_2(\alpha)$.

Theorem 5.1.5 implies that when $\alpha \in M$ is algebraic over K, with minimal polynomial of degree n, the subset $\{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_i \in K\}$ is a subfield of M. This isn't particularly obvious: for instance, why is it closed under taking reciprocals? But it's true.

Corollary 5.1.10 *Let $M : K$ be a field extension and $\alpha \in M$. Then*

$$\deg_K(\alpha) < \infty \iff \alpha \text{ is algebraic over } K.$$

Proof For $\Rightarrow$, we prove the contrapositive. By Theorem 4.3.16(ii), if α is transcendental over K then $K(\alpha)$ is isomorphic over K to the field $K(t)$ of rational expressions, and $[K(t) : K] = \infty$ by Example 5.1.3(iii).

Theorem 5.1.5 gives $\Leftarrow$. $\qquad\square$

For a field extension $M : K$ and $\alpha \in M$, the **degree** of α over K is $[K(\alpha) : K]$. We write it as $\deg_K(\alpha)$. So, $\deg_K(\alpha) < \infty$ if and only if α is algebraic over K, and in that case, $\deg_K(\alpha)$ is the degree of the minimal polynomial of α over K.
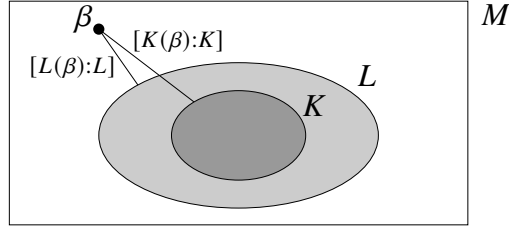
Figure 5.1: Visualization of Corollary 5.1.12 (not to be taken too seriously).

Example 5.1.11 Let ξ be the real cube root of 2. By Example 4.2.11(ii), the minimal polynomial of ξ over $\mathbb{Q}$ is $t^3 - 2$, so $\deg_{\mathbb{Q}}(\xi) = 3$. It follows that $\mathbb{Q}(\xi) \neq \{a + b\xi : a, b \in \mathbb{Q}\}$, since otherwise the two-element set $\{1, \xi\}$ would span the three-dimensional vector space $\mathbb{Q}(\xi)$. So we have another proof that $2^{2/3}$ cannot be written as a $\mathbb{Q}$-linear combination of 1 and $2^{1/3}$. As observed in Example 4.2.11(ii), this is messy to prove directly.

Theorem 5.1.5 is powerful. Here are two more of its corollaries.

Corollary 5.1.12 *Let $M : L : K$ be field extensions and $\beta \in M$. Then $[L(\beta) : L] \leq [K(\beta) : K]$.*

Informally, I think of Corollary 5.1.12 as in Figure 5.1. The degree of β over K measures how far β is from being in K. Since L contains K, it might be that β is closer to L than to K (i.e. $[L(\beta) : L] < [K(\beta) : K]$), and it's certainly no further away.

Proof If $[K(\beta) : K] = \infty$ then the inequality is clear. Otherwise, β is algebraic over K (by Corollary 5.1.10), with minimal polynomial $m \in K[t]$, say. Then m is an annihilating polynomial for β over L, so the minimal polynomial of β over L has degree $\leq \deg(m)$. The result follows from Theorem 5.1.5. $\square$

Exercise 5.1.13 Give an example to show that the inequality in Corollary 5.1.12 can be strict. Your example can be as trivial as you like.

Corollary 5.1.14 *Let $M : K$ be a field extension. Let $\alpha_1, \ldots, \alpha_n \in M$, with α_i algebraic over K of degree d_i. Then every element $\alpha \in K(\alpha_1, \ldots, \alpha_n)$ can be expressed as a polynomial in $\alpha_1, \ldots, \alpha_n$ over K. More exactly,*

$$\alpha = \sum_{r_1, \ldots, r_n} c_{r_1, \ldots, r_n} \alpha_1^{r_1} \cdots \alpha_n^{r_n}$$

for some $c_{r_1, \ldots, r_n} \in K$, where r_i ranges over $0, \ldots, d_i - 1$.

70

For example, here's what this says in the case $n = 2$. Let $M : K$ be a field extension, and take algebraic elements α_1, α_2 of M. Write d_1 and d_2 for their degrees over K. Then every element of $K(\alpha_1, \alpha_2)$ is equal to

$$\sum_{r=0}^{d_1-1} \sum_{s=0}^{d_2-1} c_{rs} \alpha_1^r \alpha_2^s$$

for some coefficients $c_{rs} \in K$. A fundamental point in the proof is that a polynomial in two variables can be seen as a polynomial in one variable whose coefficients are themselves polynomials in one variable, and similarly for more than two variables.

Proof When $n = 0$, this is trivial. Now let $n \geq 1$ and suppose inductively that the result holds for $n - 1$. Let

$$\alpha \in K(\alpha_1, \dots, \alpha_n) = \big(K(\alpha_1, \dots, \alpha_{n-1})\big)(\alpha_n).$$

By Theorem 5.1.5 applied to the extension $(K(\alpha_1, \dots, \alpha_{n-1}))(\alpha_n)$: $K(\alpha_1, \dots, \alpha_{n-1})$, noting that $\deg_{K(\alpha_1,\dots,\alpha_{n-1})}(\alpha_n) \leq \deg_K(\alpha_n) = d_n$, we have

$$\alpha = \sum_{r=0}^{d_n-1} c_r \alpha_n^r \tag{5.1}$$

for some $c_0, \dots, c_{d_n-1} \in K(\alpha_1, \dots, \alpha_{n-1})$. By inductive hypothesis, for each r we have

$$c_r = \sum_{r_1,\dots,r_{n-1}} c_{r_1,\dots,r_{n-1},r} \alpha_1^{r_1} \cdots \alpha_{n-1}^{r_{n-1}} \tag{5.2}$$

for some $c_{r_1,\dots,r_{n-1},r} \in K$, where r_i ranges over $0, \dots, d_i - 1$. Substituting (5.2) into (5.1) completes the induction. $\qquad\square$

Example 5.1.15 Back in Example 4.1.9(ii), I claimed that

$$\mathbb{Q}(\sqrt{2}, i) = \{a + b\sqrt{2} + ci + d\sqrt{2}i : a, b, c, d \in \mathbb{Q}\}.$$

Corollary 5.1.14 applied to $\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}$ proves this, since $\deg_{\mathbb{Q}}(\sqrt{2}) = \deg_{\mathbb{Q}}(i) = 2$.

> ✏️ **Exercise 5.1.16** Let $M : K$ be a field extension and α a transcendental element of M. Can every element of $K(\alpha)$ be represented as a polynomial in α over K?

For extensions obtained by adjoining several elements, the following result is invaluable.

> **Theorem 5.1.17 (Tower law)** *Let $M : L : K$ be field extensions.*
>
> i. *If $(\alpha_i)_{i \in I}$ is a basis of L over K and $(\beta_j)_{j \in J}$ is a basis of M over L, then $(\alpha_i \beta_j)_{(i,j) \in I \times J}$ is a basis of M over K.*
>
> ii. *$M : K$ is finite $\iff$ $M : L$ and $L : K$ are finite.*
>
> iii. *$[M : K] = [M : L][L : K]$.*

The sets I and J here could be infinite. I'll say that a family $(a_i)_{i \in I}$ of elements of a field is **finitely supported** if the set $\{i \in I : a_i \neq 0\}$ is finite.

Proof To prove (i), we show that $(\alpha_i \beta_j)_{(i,j) \in I \times J}$ is a linearly independent spanning set of M over K.

For linear independence, let $(c_{ij})_{(i,j) \in I \times J}$ be a finitely supported family of elements of K such that $\sum_{i,j} c_{ij} \alpha_i \beta_j = 0$. Then $\sum_j (\sum_i c_{ij} \alpha_i) \beta_j = 0$, with $\sum_i c_{ij} \alpha_i \in L$ for each $j \in J$. Since $(\beta_j)_{j \in J}$ is linearly independent over L, we have $\sum_i c_{ij} \alpha_i = 0$ for each $j \in J$. But $(\alpha_i)_{i \in I}$ is linearly independent over K, so $c_{ij} = 0$ for each $i \in I$ and $j \in J$.

To show $(\alpha_i \beta_j)_{(i,j) \in I \times J}$ spans M over K, let $e \in M$. Since $(\beta_j)_{j \in J}$ spans M over L, we have $e = \sum_j d_j \beta_j$ for some finitely supported family $(d_j)_{j \in J}$ of elements of L. Since $(\alpha_i)_{i \in I}$ spans L over K, for each $j \in J$ we have $d_j = \sum_i c_{ij} \alpha_i$ for some finitely supported family $(c_{ij})_{i \in I}$ of K. Hence $e = \sum_{i,j} c_{ij} \alpha_i \beta_j$, as required.

Parts (ii) and (iii) follow. $\qquad\square$

Example 5.1.18 What is $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}]$? The tower law gives

$$
\begin{aligned}
\left[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}\right] &= \left[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})\right]\left[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}\right] \\
&= 2\left[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})\right].
\end{aligned}
$$

Now on the one hand,

$$
\left[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})\right] \leq \left[\mathbb{Q}(\sqrt{3}) : \mathbb{Q}\right] = 2
$$

by Corollary 5.1.12. On the other, $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$, so $\mathbb{Q}(\sqrt{2}, \sqrt{3}) \neq \mathbb{Q}(\sqrt{2})$, so $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})] > 1$ by Example 5.1.3(i). So $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})] = 2$, giving the answer: $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$.

By the same argument as in Example 5.1.15, $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ spans $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$. But we have just shown that $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ has dimension 4 over $\mathbb{Q}$. Hence this spanning set is a basis. That is, for every element $\alpha \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$, there is one *and only one* 4-tuple (a, b, c, d) of rational numbers such that

$$
\alpha = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}.
$$

Corollary 5.1.19 *Let $M : L' : L : K$ be field extensions. If $M : K$ is finite then $[L' : L]$ divides $[M : K]$.*

Proof By the tower law twice, $[M : K] = [M : L'][L' : L][L : K]$. □

That result might remind you of Lagrange's theorem on group orders. The resemblance is no coincidence, as we'll see.

> ✏️ **Exercise 5.1.20** Show that a field extension whose degree is a prime number must be simple.

That result might remind you of the fact that a group of prime order must be cyclic, and that's no coincidence either!

A second corollary of the tower law:

Corollary 5.1.21 *Let $M : K$ be a field extension and $\alpha_1, \ldots, \alpha_n \in M$. Then*

$$[K(\alpha_1, \ldots, \alpha_n) : K] \leq [K(\alpha_1) : K] \cdots [K(\alpha_n) : K].$$

Proof By the tower law and then Corollary 5.1.12,

$$
\begin{aligned}
&[K(\alpha_1, \ldots, \alpha_n) : K] \\
&\quad = [K(\alpha_1, \ldots, \alpha_n) : K(\alpha_1, \ldots, \alpha_{n-1})] \cdots [K(\alpha_1, \alpha_2) : K(\alpha_1)][K(\alpha_1) : K] \\
&\quad \leq [K(\alpha_n) : K] \cdots [K(\alpha_2) : K][K(\alpha_1) : K].
\end{aligned}
$$
□

Example 5.1.22 What is $[\mathbb{Q}(12^{1/4}, 6^{1/15}) : \mathbb{Q}]$? You can check (hint, hint) that $\deg_{\mathbb{Q}}(12^{1/4}) = 4$ and $\deg_{\mathbb{Q}}(6^{1/15}) = 15$. So by Corollary 5.1.19, $[\mathbb{Q}(12^{1/4}, 6^{1/15}) : \mathbb{Q}]$ is divisible by 4 and 15. But also, Corollary 5.1.21 implies that $[\mathbb{Q}(12^{1/4}, 6^{1/15}) : \mathbb{Q}] \leq 4 \times 15 = 60$. Since 4 and 15 are coprime, the answer is 60.

> ✏️ **Exercise 5.1.23** Generalize Example 5.1.22. In other words, what general result does the argument of Example 5.1.22 prove, not involving the particular numbers chosen there?

5.2 Algebraic extensions

We defined a field extension $M : K$ to be finite if $[M : K] < \infty$, that is, M is finite-dimensional as a vector space over K. Here are two related conditions.
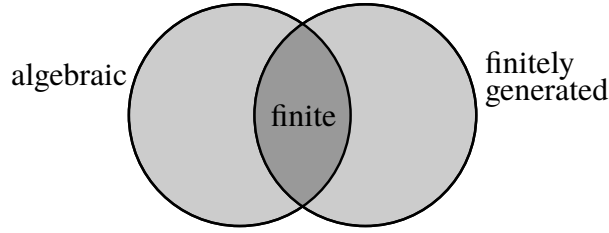
Figure 5.2: Finiteness conditions on a field extension

Definition 5.2.1 A field extension $M : K$ is **finitely generated** if $M = K(Y)$ for some finite subset $Y \subseteq M$.

Definition 5.2.2 A field extension $M : K$ is **algebraic** if every element of M is algebraic over K.

Recall from Corollary 5.1.10 that α is algebraic over K if and only if $K(\alpha) : K$ is finite. So for a field extension to be algebraic is also a kind of finiteness condition.

Examples 5.2.3   i. For any field K, the extension $K(t) : K$ is finitely generated (take the 'Y' above to be $\{t\}$) but not finite, by Corollary 5.1.10.

ii. In Section 4.2 you met the set $\overline{\mathbb{Q}}$ of complex numbers algebraic over $\mathbb{Q}$. We'll very soon prove that it's a subfield of $\mathbb{C}$. It is algebraic over $\mathbb{Q}$, by definition. But you'll show in Workshop 3 that it is not finite over $\mathbb{Q}$.

Our three finiteness conditions are related as follows (Figure 5.2).

Proposition 5.2.4 *The following conditions on a field extension $M : K$ are equivalent:*

i. $M : K$ is finite;

ii. $M : K$ is finitely generated and algebraic;

iii. $M = K(\alpha_1, \ldots, \alpha_n)$ for some finite set $\{\alpha_1, \ldots, \alpha_n\}$ of elements of M algebraic over K.

Proof (i)$\Rightarrow$(ii): suppose that $M : K$ is finite.

To show that $M : K$ is finitely generated, take a basis $\alpha_1, \ldots, \alpha_n$ of M over K. Every subfield L of M containing K is a K-linear subspace of M, so if $\alpha_1, \ldots, \alpha_n \in L$ then $L = M$. This proves that the only subfield of M containing $K \cup \{\alpha_1, \ldots, \alpha_n\}$ is M itself; that is, $M = K(\alpha_1, \ldots, \alpha_n)$. So $M : K$ is finitely generated.

To show that $M : K$ is algebraic, let $\alpha \in M$. Then by part (ii) of the tower law (Theorem 5.1.17), $K(\alpha) : K$ is finite, so by Corollary 5.1.10, α is algebraic over K.

(ii)$\Rightarrow$(iii) is immediate from the definitions.

(iii)$\Rightarrow$(i): suppose that $M = K(\alpha_1, \ldots, \alpha_n)$ for some $\alpha_i \in M$ algebraic over K. Then
$$[M : K] \leq [K(\alpha_1) : K] \cdots [K(\alpha_n) : K]$$
by Corollary 5.1.21. For each i, we have $[K(\alpha_i) : K] < \infty$ since α_i is algebraic over K (using Corollary 5.1.10 again). So $[M : K] < \infty$. $\qquad\square$

We already saw that when $M = K(\alpha_1, \ldots, \alpha_n)$ with each α_i algebraic, every element of M is a polynomial in $\alpha_1, \ldots, \alpha_n$ (Corollary 5.1.14). So for any finite extension $M : K$, there is some finite set of elements such that everything in M can be expressed as a polynomial over K in these elements.

> **Exercise 5.2.5** Let $M : K$ be a field extension and $K \subseteq L \subseteq M$. In the proof of Proposition 5.2.4, I said that *if L is a subfield of M then L is a K-linear subspace of M*. Why is that true? And is the converse also true? Give a proof or a counterexample.

Corollary 5.2.6 *Let $K(\alpha) : K$ be a simple extension. The following are equivalent:*

  i. $K(\alpha) : K$ *is finite;*

  ii. $K(\alpha) : K$ *is algebraic;*

  iii. α *is algebraic over K.*

Proof (i)$\Rightarrow$(ii) follows from (i)$\Rightarrow$(ii) of Proposition 5.2.4.

(ii)$\Rightarrow$(iii) is immediate from the definitions.

(iii)$\Rightarrow$(i) follows from (iii)$\Rightarrow$(i) of Proposition 5.2.4. $\qquad\square$

Here's a spectacular application of Corollary 5.2.6.

Proposition 5.2.7 $\overline{\mathbb{Q}}$ *is a subfield of $\mathbb{C}$.*

Proof By Corollary 5.2.6,
$$\overline{\mathbb{Q}} = \{\alpha \in \mathbb{C} : [\mathbb{Q}(\alpha) : \mathbb{Q}] < \infty\}.$$
For all $\alpha, \beta \in \overline{\mathbb{Q}}$,
$$[\mathbb{Q}(\alpha, \beta) : \mathbb{Q}] \leq [\mathbb{Q}(\alpha) : \mathbb{Q}][\mathbb{Q}(\beta) : \mathbb{Q}] < \infty$$

by Corollary 5.1.21. Now $\alpha + \beta \in \mathbb{Q}(\alpha, \beta)$, so $\mathbb{Q}(\alpha + \beta) \subseteq \mathbb{Q}(\alpha, \beta)$, so

$$[\mathbb{Q}(\alpha + \beta) : \mathbb{Q}] \leq [\mathbb{Q}(\alpha, \beta) : \mathbb{Q}] < \infty,$$

giving $\alpha + \beta \in \overline{\mathbb{Q}}$. Similarly, $\alpha \cdot \beta \in \overline{\mathbb{Q}}$. For all $\alpha \in \overline{\mathbb{Q}}$,

$$[\mathbb{Q}(-\alpha) : \mathbb{Q}] = [\mathbb{Q}(\alpha) : \mathbb{Q}] < \infty,$$

giving $-\alpha \in \overline{\mathbb{Q}}$. Similarly, $1/\alpha \in \overline{\mathbb{Q}}$ (if $\alpha \neq 0$). And clearly $0, 1 \in \overline{\mathbb{Q}}$. $\qquad \square$

If you did Exercise 4.2.4, you'll appreciate how hard that result is to prove from first principles, and how amazing it is that the proof above is so clean and simple.

> **Exercise 5.2.8** Let $M : K$ be a field extension, and write L for the set of elements of M algebraic over K. By imitating the proof of Proposition 5.2.7, prove that L is a subfield of M.

5.3 Ruler and compass constructions

This section is a truly wonderful application of the algebra we've developed so far. Using it, we will solve problems that lay unsolved for thousands of years.

The arguments here have a lot in common with those we'll use in Chapter 9 for the problem of solving polynomials by radicals. It's well worth getting used to these arguments now, since the polynomial problem involves some extra subtleties that will need your full attention then. In other words, treat this as a warm-up.

The ancient Greeks developed planar geometry to an extraordinary degree, discovering how to perform a very wide range of constructions using only ruler and compasses. But there were three particular constructions that they couldn't figure out how to do using only these instruments:

- **Trisect the angle:** given an angle θ, construct the angle $\theta/3$.

- **Duplicate the cube:** given a length, construct a new length whose cube is twice the cube of the original. That is, given two points distance L apart, construct two points distance $\sqrt[3]{2}L$ apart.

- **Square the circle:** given a circle, construct a square with the same area. That is, given two points distance L apart, construct two points distance $\sqrt{\pi}L$ apart.

The challenge of finding constructions lay unanswered for millennia. And it wasn't for lack of attention. My Galois theory lecture notes from when I was an undergraduate contain the following words:

Thomas Hobbes claimed to have solved these. John Wallis disagreed. A 17th century pamphlet war ensued.

Twitter users may conclude that human nature has not changed.

It turns out that the reason why no one could find a way to do these constructions is that they're impossible. We'll prove it using field theory.

In order to prove that you *can't* do these things using ruler and compasses, it's necessary to know that you *can* do certain other things using ruler and compasses. I'll take some simple constructions for granted (but there's a video if you want the details).

Ruler and compass constructions

> **Digression 5.3.1** The standard phrase is 'ruler and compass constructions', but it's slightly misleading. A ruler has distance markings on it, whereas for the problems of ancient Greece, you're supposed to use only a 'straight edge': a ruler without markings (and no, you're not allowed to mark it). As Stewart explains (Section 7.1), with a marked or markable straight edge, you *can* solve all three problems. Also, for what it's worth, an instrument for drawing circles is strictly speaking a *pair* of compasses. But like everyone else, we'll say 'ruler and compass'—
>
> —when we really mean 'straight edge and compasses'—
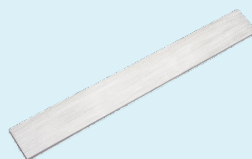
The problems as stated above are maybe not quite precise; let's formalize them.

Starting from a subset Σ of the plane, our instruments allow the following constructions:

- given two distinct points A, B of Σ, draw the (infinite) line through A and B;

- given two distinct points A, B of Σ, draw the circle with centre A passing through B.

A point in the plane is **immediately constructible** from Σ if it is a point of intersection between two distinct lines, or two distinct circles, or a line and a

circle, of the form above. A point C in the plane is **constructible** from Σ if there is a finite sequence $C_1, \ldots, C_n = C$ of points such that C_i is immediately constructible from $\Sigma \cup \{C_1, \ldots, C_{i-1}\}$ for each i. Broadly, the question is: which points are constructible from which?

The key idea of the solution is that when you're doing Greek-style geometry, then in terms of coordinates, you're repeatedly solving linear or quadratic equations. (The Greeks didn't use coordinates, but we will.) This is because equations of lines and circles are linear or quadratic.

For instance, suppose we start with the points $(0, 0)$ and $(1, 0)$. Draw the circle with centre $(0, 0)$ passing through $(1, 0)$, and vice versa. The intersection points of the two circles are $(1/2, \pm\sqrt{3}/2)$, where the square root came from solving a quadratic. If we do further ruler and compass constructions, we might end up with coordinates like $\sqrt{\sqrt{2} + \sqrt{3}}$. But we can never get coordinates like $\sqrt[3]{2}$, because where would a cube root come from? We're only solving quadratics here.

To translate this idea into field terms, we make the following definition. Take a subfield $K \subseteq \mathbb{R}$. We say that the extension $K : \mathbb{Q}$ is **iterated quadratic** if there is some finite sequence of subfields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$$

such that $[K_i : K_{i-1}] = 2$ for all $i \in \{1, \ldots, n\}$.

Example 5.3.2 $\mathbb{Q}\left(\sqrt{\sqrt{2} + \sqrt{3}}\right)$ is an iterated quadratic extension of $\mathbb{Q}$, because we have a chain of subfields

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3}) \subseteq \mathbb{Q}\left(\sqrt{\sqrt{2} + \sqrt{3}}\right)$$

where each has degree 2 over the last. (For the equality, see Exercise 4.3.15.)

There is an iterated quadratic extension of $\mathbb{Q}$ containing $\sqrt{\sqrt{2} + \sqrt{3}}$, and by the same argument, there is one containing $\sqrt{\sqrt{5} + \sqrt{7}}$. Is there one containing both? We will prove a general result guaranteeing that there is. The following terminology will be useful.

Definition 5.3.3 Let L and L' be subfields of a field M. The **compositum** LL' of L and L' is the subfield of M generated by $L \cup L'$.

That is, LL' is the smallest subfield of M containing both L and L'. In the notation of Definition 4.1.8, we could also write LL' as either $L(L')$ or $L'(L)$.

Example 5.3.4 The compositum of the subfields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ of $\mathbb{R}$ is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

To show that any two iterated quadratic extensions of $\mathbb{Q}$ can be merged into one, we first consider extensions of degree 2.

Lemma 5.3.6 *Let $M : K$ be a field extension and let L, L' be subfields of M containing K. If $[L : K] = 2$ then $[LL' : L'] \in \{1, 2\}$.*
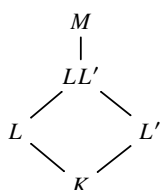
Actually, it is true more generally that $[LL' : L'] \leq [L : K]$ (Workshop 3), but we will not need this fact.

Proof Choose some $\beta \in L \setminus K$. By applying the tower law to $L : K(\beta) : K$ and using the hypothesis that $[L : K] = 2$, we see that $K(\beta) = L$.

Next we show that $LL' = L'(\beta)$. Certainly $L'(\beta) \subseteq LL'$, since $L' \subseteq LL'$ and $\beta \in L \subseteq LL'$. Conversely, $L'(\beta)$ is a subfield of M that contains both $K(\beta) = L$ and L', so it contains LL'. Hence $LL' = L'(\beta)$, as claimed.

It follows that $[LL' : L'] = [L'(\beta) : L'] \leq [K(\beta) : K] = 2$, where the inequality comes from Corollary 5.1.12. $\qquad\square$

Lemma 5.3.8 *Let K and L be subfields of $\mathbb{R}$ such that the extensions $K : \mathbb{Q}$ and $L : \mathbb{Q}$ are iterated quadratic. Then there is some subfield M of $\mathbb{R}$ such that the extension $M : \mathbb{Q}$ is iterated quadratic and $K, L \subseteq M$.*

Proof Take subfields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K \subseteq \mathbb{R}, \qquad \mathbb{Q} = L_0 \subseteq L_1 \subseteq \cdots \subseteq L_m = L \subseteq \mathbb{R}$$

with $[K_i : K_{i-1}] = 2 = [L_j : L_{j-1}]$ for all i, j. Consider the chain of subfields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K = KL_0 \subseteq KL_1 \subseteq \cdots \subseteq KL_m = KL \qquad (5.3)$$

of $\mathbb{R}$. It is enough to show that KL is an iterated quadratic extension of K.

In the chain (5.3), $[K_i : K_{i-1}] = 2$ for all i. Moreover, for each j we have $[L_j : L_{j-1}] = 2$, so Lemma 5.3.6 implies that $[KL_j : KL_{j-1}] \in \{1, 2\}$ (taking the 'K' of that lemma to be L_{j-1}). Hence in (5.3), all the successive degrees are 1 or 2. An extension of degree 1 is an equality, so by ignoring repeats, we see that $KL : \mathbb{Q}$ is an iterated quadratic extension. $\qquad\square$

79

The general theory of ruler and compass constructibility starts with any set $\Sigma \subseteq \mathbb{R}^2$ of given points. But for simplicity, we will stick to the case where Σ consists of two points, and we'll choose our coordinate axes so that they have coordinates $(0,0)$ and $(1,0)$. This will still enable us to solve the notorious problems of ancient Greece.

Proposition 5.3.9 *Let $(x,y) \in \mathbb{R}^2$. If (x,y) is constructible from $\{(0,0),(1,0)\}$ then there is an iterated quadratic extension of $\mathbb{Q}$ containing x and y.*

Proof Suppose that (x,y) is constructible from $\{(0,0),(1,0)\}$ in n steps. If $n = 0$ then (x,y) is $(0,0)$ or $(1,0)$, so $x, y \in \mathbb{Q}$, and $\mathbb{Q}$ is trivially an iterated quadratic extension of $\mathbb{Q}$.

Now let $n \geq 1$. Suppose inductively that each coordinate of each point constructible from $\{(0,0),(1,0)\}$ in $< n$ steps lies in some iterated quadratic extension of $\mathbb{Q}$. By definition, (x,y) is an intersection point of two distinct lines/circles through points constructible in $< n$ steps. By inductive hypothesis, each coordinate of each of those points lies in some iterated quadratic extension of $\mathbb{Q}$, so by Lemma 5.3.8, there is an iterated quadratic extension L of $\mathbb{Q}$ containing all the points' coordinates. The coefficients in the equations of the lines/circles then also lie in L.

We now show that $\deg_L(x) \in \{1,2\}$.

If (x,y) is the intersection point of two distinct lines, then x and y satisfy two linearly independent equations

$$ax + by + c = 0,$$
$$a'x + b'y + c' = 0$$

with $a, b, c, a', b', c' \in L$. Solving gives $x \in L$. (In more detail, x is a rational function of $a, b, \ldots$—write it down if you want!—and so $x \in L$.)

If (x,y) is an intersection point of a line and a circle, then

$$ax + by + c = 0,$$
$$x^2 + y^2 + dx + ey + f = 0$$

with $a, b, c, d, e, f \in L$. If $b = 0$ then $a \neq 0$ and $x = -c/a \in L$. Otherwise, we can eliminate y to give a quadratic over L satisfied by x, so that $\deg_L(x) \in \{1,2\}$.

If (x,y) is an intersection point of two circles, then

$$x^2 + y^2 + dx + ey + f = 0,$$
$$x^2 + y^2 + d'x + e'y + f' = 0$$

with $d, e, f, d', e', f' \in L$. Subtracting, we reduce to the case of a line and a circle, again giving $\deg_L(x) \in \{1,2\}$.
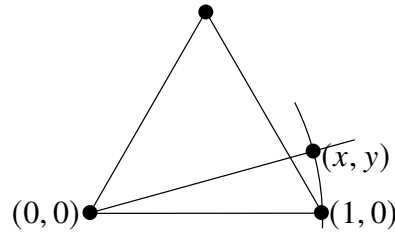
Figure 5.3: The impossibility of trisecting $60°$.

Hence $\deg_L(x) \in \{1, 2\}$. If $\deg_L(x) = 1$ then $x \in L$, and L is an iterated quadratic extension of $\mathbb{Q}$. If $\deg_L(x) = 2$, i.e. $[L(x) : L] = 2$, then $L(x)$ is an iterated quadratic extension of $\mathbb{Q}$. In either case, x lies in some iterated quadratic extension of $\mathbb{Q}$. The same is true of y. Hence by Lemma 5.3.8, there is an iterated quadratic extension of $\mathbb{Q}$ containing x and y. This completes the induction. $\qquad\square$

Theorem 5.3.10 *Let $(x, y) \in \mathbb{R}^2$. If (x, y) is constructible from $\{(0, 0), (1, 0)\}$ then x and y are algebraic over $\mathbb{Q}$, and their degrees over $\mathbb{Q}$ are powers of 2.*

Proof By Proposition 5.3.9, there is an iterated quadratic extension M of $\mathbb{Q}$ with $x \in M$. Then $[M : \mathbb{Q}] = 2^n$ for some $n \geq 0$, by the tower law. But then $\deg_{\mathbb{Q}}(x) = [\mathbb{Q}(x) : \mathbb{Q}]$ divides 2^n by Corollary 5.1.19, and is therefore a power of 2. And similarly for y. $\qquad\square$

Now we solve the problems of ancient Greece.

Proposition 5.3.11 *The angle cannot be trisected by ruler and compass.*

Proof Suppose it can be. Construct an equilateral triangle with $(0, 0)$ and $(1, 0)$ as two of its vertices (which can be done by ruler and compass; Figure 5.3). Trisect the angle of the triangle at $(0, 0)$. Plot the point (x, y) where the trisector meets the circle with centre $(0, 0)$ through $(1, 0)$. Then $x = \cos(\pi/9)$, so by Theorem 5.3.10, $\deg_{\mathbb{Q}}(\cos(\pi/9))$ is a power of 2. But you showed in Workshop 2 that $\deg_{\mathbb{Q}}(\cos(\pi/9)) = 3$, a contradiction. $\qquad\square$

Proposition 5.3.12 *The cube cannot be duplicated by ruler and compass.*

Proof Suppose it can be. Since $(0, 0)$ and $(1, 0)$ are distance 1 apart, we can construct from them two points A and B distance $\sqrt[3]{2}$ apart. From A and B we can construct, using ruler and compass, the point $(\sqrt[3]{2}, 0)$. So $\deg_{\mathbb{Q}}(\sqrt[3]{2})$ is a power of 2, by Theorem 5.3.10. But $\deg_{\mathbb{Q}}(\sqrt[3]{2}) = 3$ by Example 5.1.11, a contradiction. $\qquad\square$

Proposition 5.3.13 *The circle cannot be squared by ruler and compass.*

This one is the most outrageously impossible, yet the hardest to prove.

Proof Suppose it can be. Since the circle with centre $(0,0)$ through $(1,0)$ has area π, we can construct by ruler and compass a square with side-length $\sqrt{\pi}$, and from that, we can construct by ruler and compass the point $(\sqrt{\pi}, 0)$. So by Theorem 5.3.10, $\sqrt{\pi}$ is algebraic over $\mathbb{Q}$ with degree a power of 2. Since $\overline{\mathbb{Q}}$ is a subfield of $\mathbb{C}$, it follows that π is algebraic over $\mathbb{Q}$. But it is a (hard) theorem that π is transcendental over $\mathbb{Q}$. $\square$

> **Digression 5.3.14** Stewart has a nice alternative approach to all this, in his Chapter 7. He treats the plane as the *complex* plane, and he shows that the set of all points in $\mathbb{C}$ constructible from 0 and 1 is a subfield. In fact, it is the smallest subfield of $\mathbb{C}$ closed under taking square roots. He calls it $\mathbb{Q}^{\mathrm{py}}$, the 'Pythagorean closure' of $\mathbb{Q}$. It can also be described as the set of complex numbers contained in some iterated quadratic extension of $\mathbb{Q}$.

There is one more famous ruler and compass problem: for which integers n is the regular n-sided polygon constructible, starting from just a pair of points in the plane?

The answer has to do with **Fermat primes**, which are prime numbers of the form $2^u + 1$ for some $u \geq 1$. A little exercise in number theory shows that if $2^u + 1$ is prime then u must itself be a power of 2. The only known Fermat primes are

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

Whether there are any others is a longstanding open question. In any case, it can be shown that the regular n-sided polygon is constructible if and only if

$$n = 2^r p_1 \cdots p_k$$

for some $r, k \geq 0$ and distinct Fermat primes $p_1, \ldots, p_k$.

We will not do the proof, but it involves cyclotomic polynomials. A glimpse of the connection: let p be a prime such that the regular p-sided polygon is constructible. Consider the regular p-sided polygon inscribed in the unit circle in $\mathbb{C}$, with one of its vertices at 1. Then another vertex is at $e^{2\pi i/p}$, and from constructibility, one can deduce that $\deg_{\mathbb{Q}}(e^{2\pi i/p})$ is a power of 2. But we saw in Example 5.1.6(ii) that $\deg_{\mathbb{Q}}(e^{2\pi i/p}) = p - 1$. So $p - 1$ is a power of 2, that is, p is a Fermat prime. Field theory, number theory and Euclidean geometry come together!

Chapter 6

Splitting fields

Introduction to Week 6

In Chapter 1, we met a definition of the symmetry group of a polynomial over $\mathbb{Q}$. It was phrased in terms of conjugate tuples, it was possibly a little mysterious, and it was definitely difficult to work with (e.g. we couldn't compute the symmetry group of $1 + t + t^2 + t^3 + t^4$).

In this chapter, we're going to give a different but equivalent definition of the symmetry group of a polynomial. It's a two-step process:

1. We show how every polynomial f over K gives rise to an extension of K, called the 'splitting field' of f.

2. We show how every field extension has a symmetry group.

The symmetry group, or 'Galois group', of a polynomial is then defined to be the symmetry group of its splitting field extension.

How does these two steps work?

1. When $K = \mathbb{Q}$, the splitting field of f is the smallest subfield of $\mathbb{C}$ containing all the complex roots of f. For a general field K, it's constructed by adding the roots of f one at a time, using simple extensions, until we obtain an extension of K in which f splits into linear factors.

2. The symmetry group of a field extension $M : K$ is defined as the group of automorphisms of M over K. This is the same idea you've seen many times before, for symmetry groups of other mathematical objects.

Why bother? Why not define the symmetry group of f directly, as in Chapter 1?

- Because this strategy works over every field K, not just $\mathbb{Q}$.

- Because there are field extensions that do not arise from a polynomial, and their symmetry groups are sometimes important. For example, an important

structure in number theory, somewhat mysterious to this day, is the symmetry group of the algebraic numbers $\overline{\mathbb{Q}}$ over $\mathbb{Q}$.

- Because using abstract algebra means you can cut down on explicit calculations with polynomials. (By way of analogy, you've seen how abstract linear algebra with vector spaces and linear maps allows you to cut down on calculations with matrices.) It also reveals connections with other parts of mathematics.

6.1 Extending homomorphisms

In your degree so far, you'll have picked up the general principle that for many kinds of mathematical *object* (such as groups, rings, fields, vector spaces, modules, metric spaces, topological spaces, measure spaces, ...), it's important to consider the appropriate notion of *mapping* between them (such as homomorphisms, linear maps, continuous maps, ...). And since Chapter 4, you've known that the basic objects of Galois theory are field extensions.

So it's no surprise that sooner or later, we have to think about mappings from one field extension to another. That moment is now. We'll need what's in this section in order to establish the basic facts about splitting fields.

When we think about a field extension $M : K$, we generally regard the field K as our starting point and M as a field that extends it. Similarly, we might start with a *homomorphism* $\psi : K \to K'$ between fields, together with extensions M of K and M' of K', and look for a *homomorphism* $M \to M'$ that extends ψ. The language is as follows.

Definition 6.1.1 Let $\iota : K \to M$ and $\iota' : K' \to M'$ be field extensions. Let $\psi : K \to K'$ be a homomorphism of fields. A homomorphism $\varphi : M \to M'$ **extends** ψ if the square

Extension problems

$$
\begin{array}{ccc}
M & \xrightarrow{\ \varphi\ } & M' \\
{\scriptstyle \iota}\Big\uparrow & & \Big\uparrow{\scriptstyle \iota'} \\
K & \xrightarrow[\ \psi\ ]{} & K'
\end{array}
$$

commutes ($\varphi \circ \iota = \iota' \circ \psi$).

Here I've used the strict definition of a field extension as a homomorphism ι of fields (Definition 4.1.1). Most of the time we view K as a subset of M and K' as a subset of M', with ι and ι' being the inclusions. In that case, for φ to extend ψ just means that

$$\varphi(a) = \psi(a) \text{ for all } a \in K.$$

Example 6.1.2 Let M and M' be two extensions of a field K. For a homomorphism $\varphi\colon M \to M'$ to extend id_K means that φ is a homomorphism over K.

The basic questions about extending homomorphisms are: given the two field extensions and the homomorphism ψ, is there some φ that extends ψ? If so, how many extensions φ are there?

We'll get to these questions later. In this section, we simply prove two general results about extensions of field homomorphisms.

Recall that any ring homomorphism $\psi\colon R \to S$ induces a homomorphism $\psi_*\colon R[t] \to S[t]$ (Definition 3.1.7). To reduce clutter, I'll write $\psi_*(f)$ as $\psi_* f$.

$$\alpha \quad \begin{array}{c} M \xrightarrow{\varphi} M' \\ \uparrow \qquad \uparrow \\ f \quad K \xrightarrow[\psi]{} K' \end{array} \quad \varphi(\alpha) \\ \qquad\qquad\qquad \psi_* f$$

Lemma 6.1.3 *Let* $M : K$ *and* $M' : K'$ *be field extensions, let* $\psi\colon K \to K'$ *be a homomorphism, and let* $\varphi\colon M \to M'$ *be a homomorphism extending* ψ. *Let* $\alpha \in M$ *and* $f(t) \in K[t]$. *Then*

$$f(\alpha) = 0 \iff (\psi_* f)(\varphi(\alpha)) = 0.$$

Proof Write $f(t) = \sum_i a_i t^i$, where $a_i \in K$. Then $(\psi_* f)(t) = \sum_i \psi(a_i) t^i \in K'[t]$, so

$$(\psi_* f)(\varphi(\alpha)) = \sum_i \psi(a_i)\varphi(\alpha)^i = \sum_i \varphi(a_i)\varphi(\alpha)^i = \varphi(f(\alpha)),$$

where the second equality holds because φ extends ψ. Since φ is injective (Lemma 2.3.3), the result follows. $\qquad\square$

Example 6.1.4 Let M and M' be extensions of a field K, and let $\varphi\colon M \to M'$ be a homomorphism over K. Then the annihilating polynomials of an element $\alpha \in M$ are the same as those of $\varphi(\alpha)$. This is the case $\psi = \mathrm{id}_K$ of Lemma 6.1.3.

Exercise 6.1.5 Show that if a ring homomorphism ψ is injective then so is ψ_*, and if ψ is an isomorphism then so is ψ_*.

An isomorphism between fields, rings, groups, vector spaces, etc., can be understood as simply a renaming of the elements. For example, if I tell you that the ring R is left Noetherian but not right Artinian, and that S is isomorphic to R, then you can deduce that S is left Noetherian but not right Artinian *without having the slightest idea what those words mean*. Just as long as they don't depend on the names of the elements of the ring concerned (which such definitions never do), you're fine.

Proposition 6.1.6 *Let* $\psi\colon K \to K'$ *be an isomorphism of fields. Let* $K(\alpha) : K$ *be a simple extension where* α *has minimal polynomial* m *over* K, *and let* $K'(\alpha') : K'$ *be a simple extension where* α' *has minimal polynomial* $\psi_* m$ *over* K'. *Then there is exactly one isomorphism* $\varphi\colon K(\alpha) \to K'(\alpha')$ *that extends* ψ *and satisfies* $\varphi(\alpha) = \alpha'$.

Diagram:

$$
\begin{array}{ccc}
K(\alpha) & \overset{\varphi}{\underset{\cong}{\dashrightarrow}} & K'(\alpha') \\
\uparrow & & \uparrow \\
K & \underset{\psi}{\overset{\cong}{\longrightarrow}} & K'
\end{array}
$$

We often use a dotted arrow to denote a map whose existence is part of the conclusion of a theorem.

Proof View $K'(\alpha')$ as an extension of K via the composite homomorphism $K \xrightarrow{\psi} K' \to K'(\alpha')$. Then the minimal polynomial of α' over K is m. (If this isn't intuitively clear to you, think of the isomorphism ψ as renaming.) Hence by the classification of simple extensions, Theorem 4.3.16, there is exactly one isomorphism $\varphi \colon K(\alpha) \to K'(\alpha')$ over K such that $\varphi(\alpha) = \alpha'$. The result follows.□

6.2 Existence and uniqueness of splitting fields

Let f be a polynomial over a field K. Informally, a splitting field for f is an extension of K where f has all its roots, and which is no bigger than it needs to be.

Warning 6.2.1 If f is irreducible, we know how to create an extension of K where f has at least *one* root: take the simple extension $K[t]/\langle f \rangle$, in which the equivalence class of t is a root of f (Lemma 4.3.1(i)).

But $K[t]/\langle f \rangle$ is not usually a splitting field for f. For example, take $K = \mathbb{Q}$ and $f(t) = t^3 - 2$, as in Warning 4.3.19. Write ξ for the real cube root of 2. (Half the counterexamples in Galois theory involve the real cube root of 2.) Then $\mathbb{Q}[t]/\langle f \rangle$ is isomorphic to the subfield $\mathbb{Q}(\xi)$ of $\mathbb{R}$, which only contains *one* root of f: the other two are non-real, hence not in $\mathbb{Q}(\xi)$.

Definition 6.2.2 Let f be a polynomial over a field M. Then f **splits** in M if

$$
f(t) = \beta(t - \alpha_1) \cdots (t - \alpha_n)
$$

for some $n \geq 0$ and $\beta, \alpha_1, \ldots, \alpha_n \in M$.

Equivalently, f splits in M if all its irreducible factors in $M[t]$ are linear.

Examples 6.2.3   i. A field M is algebraically closed if and only if every polynomial over M splits in M.

ii. Let $f(t) = t^4 - 4t^2 - 5$. Then f splits in $\mathbb{Q}(i, \sqrt{5})$, since

$$
\begin{aligned}
f(t) &= (t^2 + 1)(t^2 - 5) \\
&= (t - i)(t + i)(t - \sqrt{5})(t + \sqrt{5}).
\end{aligned}
$$

But f does not split in $\mathbb{Q}(i)$, as its factorization into irreducibles in $\mathbb{Q}(i)[t]$ is

$$
f(t) = (t - i)(t + i)(t^2 - 5),
$$

which contains a nonlinear factor.

> **Warning 6.2.4** As Example 6.2.3(ii) shows, a polynomial over M may have one root or even several roots in M, but still not split in M.

Example 6.2.5 Let $M = \mathbb{F}_2(\alpha)$, where α is a root of $f(t) = 1 + t + t^2$, as in Example 4.3.17(ii). We have

$$
f(1 + \alpha) = 1 + (1 + \alpha) + (1 + 2\alpha + \alpha^2) = 1 + \alpha + \alpha^2 = 0,
$$

so f has two distinct roots in M, giving

$$
f(t) = (t - \alpha)(t - (1 + \alpha))
$$

in $M[t]$. Hence f splits in M.

In *this* example, adjoining one root of f gave us a second root for free. But this doesn't typically happen (Warning 6.2.1).

Definition 6.2.6 Let f be a nonzero polynomial over a field K. A **splitting field** of f over K is an extension M of K such that:

i. f splits in M;

ii. $M = K(\alpha_1, \ldots, \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are the roots of f in M.

> **Exercise 6.2.7** Show that (ii) can equivalently be replaced by: 'if L is a subfield of M containing K, and f splits in L, then $L = M$'.

Examples 6.2.8 i. Let $0 \neq f \in \mathbb{Q}[t]$. Write $\alpha_1, \ldots, \alpha_n$ for the complex roots of f. Then $\mathbb{Q}(\alpha_1, \ldots, \alpha_n)$, the smallest subfield of $\mathbb{C}$ containing $\alpha_1, \ldots, \alpha_n$, is a splitting field of f over $\mathbb{Q}$.

Splitting fields over $\mathbb{Q}$ are easy because we have a ready-made algebraically closed field containing $\mathbb{Q}$, namely, $\mathbb{C}$.

ii. If a polynomial $f \in K[t]$ splits in K then K itself is a splitting field of f over K. For instance, since $\mathbb{C}$ is algebraically closed, it is a splitting field of every nonzero polynomial over $\mathbb{C}$.

iii. Let $f(t) = t^3 - 2 \in \mathbb{Q}[t]$. Its complex roots are ξ, $\omega\xi$ and $\omega^2\xi$, where ξ is the real cube root of 2 and $\omega = e^{2\pi i/3}$. Hence a splitting field of f over $\mathbb{Q}$ is

$$\mathbb{Q}(\xi, \omega\xi, \omega^2\xi) = \mathbb{Q}(\xi, \omega).$$

Now $\deg_{\mathbb{Q}}(\xi) = 3$ as f is irreducible, and $\deg_{\mathbb{Q}}(\omega) = 2$ as ω has minimal polynomial $1 + t + t^2$. By an argument like that in Example 5.1.22, it follows that $[\mathbb{Q}(\xi, \omega) : \mathbb{Q}] = 6$. On the other hand, $[\mathbb{Q}(\xi) : \mathbb{Q}] = 3$. So again, the extension we get by adjoining *all* the roots of f is bigger than the one we get by adjoining just *one* root of f.

iv. Take $f(t) = 1 + t + t^2 \in \mathbb{F}_2[t]$, as in Example 6.2.5. By Theorem 5.1.5, $\{1, \alpha\}$ is a basis of $\mathbb{F}_2(\alpha)$ over $\mathbb{F}_2$, so

$$\begin{aligned}\mathbb{F}_2(\alpha) &= \{0, 1, \alpha, 1 + \alpha\} \\ &= \mathbb{F}_2 \cup \{\text{the roots of } f \text{ in } \mathbb{F}_2(\alpha)\}.\end{aligned}$$

Hence $\mathbb{F}_2(\alpha)$ is a splitting field of f over $\mathbb{F}_2$.

Exercise 6.2.9 In Example 6.2.8(iii), I said that $\mathbb{Q}(\xi, \omega\xi, \omega^2\xi) = \mathbb{Q}(\xi, \omega)$. Why is that true?

Our mission for the rest of this section is to show that every nonzero polynomial f has exactly one splitting field. So that's actually two tasks: first, show that f has *at least* one splitting field, then, show that f has *only* one splitting field. The first task is easy, and in fact we prove a little bit more:

Lemma 6.2.10 *Let $f \neq 0$ be a polynomial over a field K. Then there exists a splitting field M of f over K such that $[M : K] \leq \deg(f)!$.*

Proof We prove this by induction on $\deg(f)$, for all fields K simultaneously.

If $\deg(f) = 0$ then K is a splitting field of f over K, and the result holds trivially.

Now suppose that $\deg(f) \geq 1$. We may choose an irreducible factor m of f. By Theorem 4.3.16, there is an extension $K(\alpha)$ of K with $m(\alpha) = 0$. Then $(t - \alpha) \mid f(t)$ in $K(\alpha)[t]$, giving a polynomial $g(t) = f(t)/(t - \alpha)$ over $K(\alpha)$.

We have $\deg(g) = \deg(f) - 1$, so by inductive hypothesis, there is a splitting field M of g over $K(\alpha)$ with $[M : K(\alpha)] \leq \deg(g)!$. Then M is a splitting field of f over K. (Check that you understand why.) Also, by the tower law,

$$[M : K] = [M : K(\alpha)][K(\alpha) : K] \leq (\deg(f) - 1)! \cdot \deg(m) \leq \deg(f)!,$$

completing the induction. □

Proving that every polynomial has *only* one splitting field is harder. As ever, 'only one' has to be understood up to isomorphism: after all, if you're given a splitting field, you can always rename its elements to get an isomorphic copy that's not literally identical to the original one. But isomorphism is all that matters.

Our proof of the uniqueness of splitting fields depends on the following result, which will also be useful for other purposes as we head towards the fundamental theorem of Galois theory.

Proposition 6.2.11 *Let $\psi : K \to K'$ be an isomorphism of fields, let $0 \neq f \in K[t]$, let M be a splitting field of f over K, and let M' be a splitting field of $\psi_* f$ over K'. Then:*

$$\begin{array}{ccc} M & \overset{\varphi}{\dashrightarrow} & M' \\ \uparrow & & \uparrow \\ f \quad K & \underset{\psi}{\longrightarrow} & K' \quad \psi_* f \end{array}$$
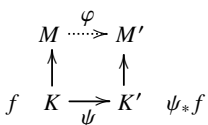
i. there exists an isomorphism $\varphi : M \to M'$ extending ψ;

ii. there are at most $[M : K]$ such extensions φ.

We'll often use this result in the case where $K' = K$ and $\psi = \mathrm{id}_K$. (What does it say then?)

Proof We prove both statements by induction on $\deg(f)$. If $\deg(f) = 0$ then both field extensions are trivial, so there is exactly one isomorphism φ extending ψ.

Now suppose that $\deg(f) \geq 1$. We can choose a monic irreducible factor m of f. Then m splits in M since f does and $m \mid f$; choose a root $\alpha \in M$ of m. We have $f(\alpha) = 0$, so $(t - \alpha) \mid f(t)$ in $K(\alpha)[t]$, giving a polynomial $g(t) = f(t)/(t - \alpha)$ over $K(\alpha)$. Then M is a splitting field of g over $K(\alpha)$, and $\deg(g) = \deg(f) - 1$.

Also, $\psi_* m$ splits in M' since $\psi_* f$ does and $\psi_* m \mid \psi_* f$. Write $\alpha'_1, \ldots, \alpha'_s$ for the distinct roots of $\psi_* m$ in M'. Note that

$$1 \leq s \leq \deg(\psi_* m) = \deg(m). \tag{6.1}$$

Since ψ_* is an isomorphism, $\psi_* m$ is monic and irreducible, and is therefore the minimal polynomial of α'_j for each $j \in \{1, \ldots, s\}$. Hence by Proposition 6.1.6, for each j, there is a unique isomorphism $\theta_j : K(\alpha) \to K'(\alpha'_j)$ that extends ψ and satisfies $\theta_j(\alpha) = \alpha'_j$. (See diagram below.)

For each $j \in \{1, \ldots, s\}$, we have a polynomial

$$\theta_{j*}(g) = \frac{\theta_{j*}(f)}{\theta_{j*}(t - \alpha)} = \frac{\psi_* f}{t - \alpha'_j}$$

over $K'(\alpha'_j)$, and M' is a splitting field of $\psi_* f$ over K', so M' is also a splitting field of $\theta_{j*}(g)$ over $K'(\alpha'_j)$.

To prove that there is at least one isomorphism φ extending ψ, choose any $j \in \{1, \ldots, s\}$ (as we may since $s \geq 1$). By applying the inductive hypothesis to g and θ_j, there is an isomorphism φ extending θ_j:

$$
\begin{array}{ccc}
M & \overset{\varphi}{\dashrightarrow} & M' \\
\uparrow & & \uparrow \\
K(\alpha) & \overset{\theta_j}{\longrightarrow} & K'(\alpha'_j) \\
\uparrow & & \uparrow \\
K & \overset{}{\longrightarrow} & K' \\
& \psi &
\end{array}
$$

But then φ also extends ψ, as required.

To prove there are at most $[M : K]$ isomorphisms $\varphi \colon M \to M'$ extending ψ, first note that any such φ satisfies $(\psi_* f)(\varphi(\alpha)) = 0$ (by Lemma 6.1.3), so $\varphi(\alpha) = \alpha'_j$ for some $j \in \{1, \ldots, s\}$. Hence

(number of isos φ extending ψ) =

$$\sum_{j=1}^{s} (\text{number of isos } \varphi \text{ extending } \psi \text{ such that } \varphi(\alpha) = \alpha'_j).$$

If φ extends ψ then $\varphi K = \psi K = K'$, and if also $\varphi(\alpha) = \alpha'_j$ then $\varphi(K(\alpha)) = K'(\alpha'_j)$. Since homomorphisms of fields are injective, φ then restricts to an isomorphism $K(\alpha) \to K'(\alpha'_j)$ satisfying $\alpha \mapsto \alpha'_j$. By the uniqueness part of Proposition 6.1.6, this restricted isomorphism must be θ_j. Thus, φ extends θ_j. Hence

$$(\text{number of isos } \varphi \text{ extending } \psi) = \sum_{j=1}^{s} (\text{number of isos } \varphi \text{ extending } \theta_j).$$

For each j, the number of isomorphisms φ extending θ_j is $\leq [M : K(\alpha)]$, by inductive hypothesis. So, using the tower law and (6.1),

$$(\text{number of isos } \varphi \text{ extending } \psi) \leq s \cdot [M : K(\alpha)] = s \cdot \frac{[M : K]}{\deg(m)} \leq [M : K],$$

completing the induction. $\qquad\square$

This brings us to the foundational result on splitting fields. Recall that an **automorphism** of an object X is an isomorphism $X \to X$.

Theorem 6.2.13 *Let f be a nonzero polynomial over a field K. Then:*

    i. *there exists a splitting field of f over K;*

    ii. *any two splitting fields of f are isomorphic over K;*

    iii. *when M is a splitting field of f over K,*

$$(\text{number of automorphisms of } M \text{ over } K) \leq [M : K] \leq \deg(f)!.$$

Proof Part (i) is immediate from Lemma 6.2.10, and part (ii) follows from Proposition 6.2.11 by taking $K' = K$ and $\psi = \mathrm{id}_K$. The first inequality in (iii) follows from Proposition 6.2.11 by taking $K' = K$, $M' = M$ and $\psi = \mathrm{id}_K$, and the second follows from Lemma 6.2.10. $\qquad\square$

Up to now we have been saying 'a' splitting field. Theorem 6.2.13(ii) gives us the right to speak of *the* splitting field of a given polynomial f over a given field K. We write it as $\mathrm{SF}_K(f)$.

We finish with a left over lemma that will be useful later.

Lemma 6.2.14   i. *Let $M : S : K$ be field extensions, $0 \neq f \in K[t]$, and $Y \subseteq M$. Suppose that S is the splitting field of f over K. Then $S(Y)$ is the splitting field of f over $K(Y)$.*

    ii. *Let $f \neq 0$ be a polynomial over a field K, and let L be a subfield of $\mathrm{SF}_K(f)$ containing K (so that $\mathrm{SF}_K(f) : L : K$). Then $\mathrm{SF}_K(f)$ is the splitting field of f over L.*

Proof For (i), f splits in S, hence in $S(Y)$. Writing X for the set of roots of f in S, we have $S = K(X)$ and so $S(Y) = K(X)(Y) = K(X \cup Y) = K(Y)(X)$; that is, $S(Y)$ is generated over $K(Y)$ by X. This proves (i), and (ii) follows by taking $M = \mathrm{SF}_K(f)$ and $Y = L$. $\qquad\square$

6.3 The Galois group

Before you get stuck into this section, you may want to review Section 2.1, especially the parts about homomorphisms $G \to \mathrm{Sym}(X)$. We'll need all of it.

What gives Galois theory its special flavour is the use of groups to study fields and polynomials. Here is the central definition.

Definition 6.3.1 The **Galois group** $\mathrm{Gal}(M : K)$ of a field extension $M : K$ is the group of automorphisms of M over K, with composition as the group operation.

> **Exercise 6.3.2** Check that this really does define a group.

In other words, an element of $\mathrm{Gal}(M : K)$ is an isomorphism $\theta \colon M \to M$ such that $\theta(a) = a$ for all $a \in K$.

Examples 6.3.3   i. What is $\mathrm{Gal}(\mathbb{C} : \mathbb{R})$? Certainly the identity is an automorphism of $\mathbb{C}$ over $\mathbb{R}$. So is complex conjugation κ, as implicitly shown in the first proof of Lemma 1.1.2. So $\{\mathrm{id}, \kappa\} \subseteq \mathrm{Gal}(\mathbb{C} : \mathbb{R})$. I claim that $\mathrm{Gal}(\mathbb{C} : \mathbb{R})$ has no other elements. For let $\theta \in \mathrm{Gal}(\mathbb{C} : \mathbb{R})$. Then

$$(\theta(i))^2 = \theta(i^2) = \theta(-1) = -\theta(1) = -1$$

as θ is a homomorphism, so $\theta(i) = \pm i$. If $\theta(i) = i$ then $\theta = \mathrm{id}$, by Lemma 4.3.6 and the fact that $\mathbb{C} = \mathbb{R}(i)$. Similarly, if $\theta(i) = -i$ then $\theta = \kappa$. So $\mathrm{Gal}(\mathbb{C} : \mathbb{R}) = \{\mathrm{id}, \kappa\} \cong C_2$.

ii. Let ξ be the real cube root of 2. For each $\theta \in \mathrm{Gal}(\mathbb{Q}(\xi) : \mathbb{Q})$, we have

$$(\theta(\xi))^3 = \theta(\xi^3) = \theta(2) = 2$$

and $\theta(\xi) \in \mathbb{Q}(\xi) \subseteq \mathbb{R}$, so $\theta(\xi) = \xi$. It follows from Lemma 4.3.6 that $\theta = \mathrm{id}$. Hence $\mathrm{Gal}(\mathbb{Q}(\xi) : \mathbb{Q})$ is trivial.

> **Exercise 6.3.4** Prove that $\mathrm{Gal}(\mathbb{Q}(e^{2\pi i/3}) : \mathbb{Q}) = \{\mathrm{id}, \kappa\}$, where $\kappa(z) = \bar{z}$. (Hint: imitate Example 6.3.3(i).)

The Galois group of a polynomial is defined to be the Galois group of its splitting field extension:

Definition 6.3.5 Let f be a nonzero polynomial over a field K. The **Galois group** $\mathrm{Gal}_K(f)$ of f over K is $\mathrm{Gal}(\mathrm{SF}_K(f) : K)$.

So the definitions fit together like this:

$$\text{polynomial} \longmapsto \text{field extension} \longmapsto \text{group.}$$

We will soon prove that Definition 6.3.5 is equivalent to the definition of Galois group in Chapter 1, where we went straight from polynomials to groups.

Theorem 6.2.13(iii) says that

$$|\operatorname{Gal}_K(f)| \leq [\operatorname{SF}_K(f) : K] \leq \deg(f)!. \tag{6.2}$$

In particular, $\operatorname{Gal}_K(f)$ is always a *finite* group.

Examples 6.3.6   i. If $f \in K[t]$ splits in K then $\operatorname{SF}_K(f) = K$ (Example 6.2.8(ii)), so $\operatorname{Gal}_K(f)$ is trivial. In particular, the Galois group of any polynomial over an algebraically closed field is trivial.

ii. $\operatorname{Gal}_{\mathbb{Q}}(t^2 + 1) = \operatorname{Gal}(\mathbb{Q}(i) : \mathbb{Q}) = \{\text{id}, \kappa\} \cong C_2$, where κ is complex conjugation on $\mathbb{Q}(i)$. The second equality is proved by the same argument as in Example 6.3.3(i), replacing $\mathbb{C} : \mathbb{R}$ by $\mathbb{Q}(i) : \mathbb{Q}$.

Calculating the Galois group with bare hands, part 1

Calculating the Galois group with bare hands, part 2

iii. Let $f(t) = (t^2 + 1)(t^2 - 2)$. Then $\operatorname{Gal}_{\mathbb{Q}}(f)$ is the group of automorphisms of $\mathbb{Q}(i, \sqrt{2})$ over $\mathbb{Q}$. Similar arguments to those in Examples 6.3.3 show that every $\theta \in \operatorname{Gal}_{\mathbb{Q}}(f)$ must satisfy $\theta(i) = \pm i$ and $\theta(\sqrt{2}) = \pm\sqrt{2}$, and that the two choices of sign determine θ completely. And one can show that all four choices are possible, so that $|\operatorname{Gal}_{\mathbb{Q}}(f)| = 4$. There are two groups of order four, C_4 and $C_2 \times C_2$. But each element of $\operatorname{Gal}_{\mathbb{Q}}(f)$ has order 1 or 2, so $\operatorname{Gal}_{\mathbb{Q}}(f)$ is not C_4, so $\operatorname{Gal}_{\mathbb{Q}}(f) \cong C_2 \times C_2$.

I've been sketchy with the details here, because it's not really sensible to try to calculate Galois groups until we have a few more tools at our disposal. We start to assemble them now.

By definition, $\operatorname{Gal}_K(f)$ acts on $\operatorname{SF}_K(f)$ (Example 2.1.2(ii)). The action is

$$(\theta, \alpha) \mapsto \theta(\alpha)$$

($\theta \in \operatorname{Gal}_K(f)$, $\alpha \in \operatorname{SF}_K(f)$). In the examples so far, we've seen that if α is a root of f then so is $\theta(\alpha)$ for every $\theta \in \operatorname{Gal}_K(f)$. This is true in general: the action of $\operatorname{Gal}_K(f)$ on $\operatorname{SF}_K(f)$ restricts to an action on the set of roots. In a slogan: *the Galois group permutes the roots.*

Lemma 6.3.7 *Let f be a nonzero polynomial over a field K. Then the action of $\operatorname{Gal}_K(f)$ on $\operatorname{SF}_K(f)$ restricts to an action on the set of roots of f in $\operatorname{SF}_K(f)$.*

Proof We have to show that if $\theta \in \mathrm{Gal}_K(f)$ and α is a root of f in $\mathrm{SF}_K(f)$ then $\theta(\alpha)$ is also a root. This follows from Example 6.1.4. □

Better still, the Galois group acts *faithfully* on the roots:

Lemma 6.3.8 *Let f be a nonzero polynomial over a field K. Then the action of $\mathrm{Gal}_K(f)$ on the roots of f is faithful.*

Proof Write X for the set of roots of f in $\mathrm{SF}_K(f)$. Then $\mathrm{SF}_K(f) = K(X)$. Hence by Lemma 4.3.6, if $\theta \in \mathrm{Gal}_K(f)$ with $\theta(x) = x$ for all $x \in X$, then $\theta = \mathrm{id}$. □

In other words, an element of the Galois group of f is completely determined by how it permutes the roots of f. So you can view elements of the Galois group as *being* permutations of the roots.

However, not every permutation of the roots belongs to the Galois group. To understand the situation, recall Remark 2.1.13, which tells us the following. Suppose that $f \in K[t]$ has distinct roots $\alpha_1, \ldots, \alpha_k$ in its splitting field. For each $\theta \in \mathrm{Gal}_K(f)$, there is a permutation $\sigma_\theta \in S_k$ defined by

$$\theta(\alpha_i) = \alpha_{\sigma_\theta(i)}$$

($i \in \{1, \ldots, k\}$). Then $\mathrm{Gal}_K(f)$ is isomorphic to the subgroup $\{\sigma_\theta : \theta \in \mathrm{Gal}_K(f)\}$ of S_k (and this is indeed a subgroup). The isomorphism is given by $\theta \mapsto \sigma_\theta$.

All this talk of the Galois group as a subgroup of S_k may have set your antennae tingling. Back in Chapter 1, we provisionally *defined* the Galois group to be a certain subgroup of S_k (Definition 1.2.1). We can now show that the two definitions are equivalent.

That definition was in terms of conjugacy. Let's now make the concept of conjugacy official, also generalizing from $\mathbb{Q}$ to an arbitrary field.

Definition 6.3.9 Let $M : K$ be a field extension, let $k \geq 0$, and let $(\alpha_1, \ldots, \alpha_k)$ and $(\alpha'_1, \ldots, \alpha'_k)$ be k-tuples of elements of M. Then $(\alpha_1, \ldots, \alpha_k)$ and $(\alpha'_1, \ldots, \alpha'_k)$ are **conjugate** over K if for all $p \in K[t_1, \ldots, t_k]$,

$$p(\alpha_1, \ldots, \alpha_k) = 0 \iff p(\alpha'_1, \ldots, \alpha'_k) = 0.$$

In the case $k = 1$, we omit the brackets and say that α and α' are conjugate to mean that (α) and (α') are.

We now show that the two definitions of the Galois group of f are equivalent.

Proposition 6.3.10 *Let f be a nonzero polynomial over a field K, with distinct roots $\alpha_1, \ldots, \alpha_k$ in $\mathrm{SF}_K(f)$. Then*

$$\{\sigma \in S_k : (\alpha_1, \ldots, \alpha_k) \text{ and } (\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)}) \text{ are conjugate over } K\} \quad (6.3)$$

is a subgroup of S_k isomorphic to $\mathrm{Gal}_K(f)$.

94

Proof As above, each $\theta \in \text{Gal}_K(f)$ gives rise to a permutation $\sigma_\theta \in S_k$, defined by $\theta(\alpha_i) = \alpha_{\sigma_\theta(i)}$. For the purposes of this proof, let us say that a permutation $\sigma \in S_k$ is 'good' if it belongs to the set (6.3). By Remark 2.1.13, it suffices to show that a permutation σ is good if and only if $\sigma = \sigma_\theta$ for some $\theta \in \text{Gal}_K(f)$.

First suppose that $\sigma = \sigma_\theta$ for some $\theta \in \text{Gal}_K(f)$. For every $p \in K[t_1, \ldots, t_k]$,

$$p(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)}) = p(\theta(\alpha_1), \ldots, \theta(\alpha_k)) = \theta(p(\alpha_1, \ldots, \alpha_k)),$$

where the first equality is by definition of σ_θ and the second is because θ is a homomorphism over K. But θ is an isomorphism, so it follows that

$$p(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)}) = 0 \iff p(\alpha_1, \ldots, \alpha_k) = 0.$$

Hence σ is good.

Conversely, suppose that σ is good. By Corollary 5.1.14, every element of $\text{SF}_K(f)$ can be expressed as $p(\alpha_1, \ldots, \alpha_k)$ for some $p \in K[t_1, \ldots, t_k]$. Now for $p, q \in K[t_1, \ldots, t_k]$, we have

$$p(\alpha_1, \ldots, \alpha_k) = q(\alpha_1, \ldots, \alpha_k) \iff p(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)}) = q(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)})$$

(by applying Definition 6.3.9 of conjugacy with $p - q$ as the 'p'). So there is a well-defined, injective function $\theta \colon \text{SF}_K(f) \to \text{SF}_K(f)$ satisfying

$$\theta(p(\alpha_1, \ldots, \alpha_k)) = p(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)}) \tag{6.4}$$

for all $p \in K[t_1, \ldots, t_k]$. Moreover, θ is surjective because σ is a permutation, and $\theta(a) = a$ for all $a \in K$ (by taking $p = a$ in (6.4)), and $\theta(\alpha_i) = \alpha_{\sigma(i)}$ for all i (by taking $p = t_i$ in (6.4)). You can check that θ is a homomorphism of fields. Hence $\theta \in \text{Gal}_K(f)$ with $\sigma_\theta = \sigma$, as required. $\square$

Exercise 6.3.11 I skipped two small bits in that proof: 'θ is surjective because σ is a permutation' (why?), and 'You can check that θ is a homomorphism of fields'. Fill in the gaps.

It's important in Galois theory to be able to move between fields. For example, you might start with a polynomial whose coefficients belong to one field K, but later decide to interpret the coefficients as belonging to some larger field L. Here's what happens to the Galois group when you do that.

Corollary 6.3.12 *Let $L : K$ be a field extension and $0 \neq f \in K[t]$. Then $\text{Gal}_L(f)$ is isomorphic to a subgroup of $\text{Gal}_K(f)$.*

Proof This follows from Proposition 6.3.10 together with the observation that if two k-tuples are conjugate over L, they are conjugate over K. $\square$

Example 6.3.13 Let's find the Galois group of $f(t) = (t^2 + 1)(t^2 - 2)$ over $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ in turn.

In Example 6.3.6(iii), we saw that $\mathrm{Gal}_{\mathbb{Q}}(f) \cong C_2 \times C_2$.

Since both roots of $t^2 - 2$ are real, $\mathrm{SF}_{\mathbb{R}}(f) = \mathrm{SF}_{\mathbb{R}}(t^2 + 1) = \mathbb{C}$. So $\mathrm{Gal}_{\mathbb{R}}(f) = \mathrm{Gal}(\mathbb{C} : \mathbb{R}) \cong C_2$, where the last step is by Example 6.3.3(i).

Finally, $\mathrm{Gal}_{\mathbb{C}}(f)$ is trivial since $\mathbb{C}$ is algebraically closed (Example 6.3.6(i)).

So as Corollary 6.3.12 predicts, $\mathrm{Gal}_{\mathbb{C}}(f)$ is isomorphic to a subgroup of $\mathrm{Gal}_{\mathbb{R}}(f)$, which is isomorphic to a subgroup of $\mathrm{Gal}_{\mathbb{Q}}(f)$.

Corollary 6.3.14 *Let f be a nonzero polynomial over a field K, with k distinct roots in $\mathrm{SF}_K(f)$. Then $|\mathrm{Gal}_K(f)|$ divides $k!$.*

Proof By Proposition 6.3.10, $\mathrm{Gal}_K(f)$ is isomorphic to a subgroup of S_k, which has $k!$ elements. The result follows from Lagrange's theorem. □

The inequalities (6.2) already gave us $|\mathrm{Gal}_K(f)| \leq \deg(f)!$. Corollary 6.3.14 improves on this in two respects. First, it implies that $|\mathrm{Gal}_K(f)| \leq k!$. It's always the case that $k \leq \deg(f)$ in all cases, and $k < \deg(f)$ if f has repeated roots in its splitting field. A trivial example: if $f(t) = t^2$ then $k = 1$ and $\deg(f) = 2$. Second, it tells us that $|\mathrm{Gal}_K(f)|$ is not only less than or equal to $k!$, but a factor of it.

Galois theory is about the interplay between field extensions and groups. In the next chapter, we'll see that just as every field extension gives rise to a group of automorphisms (its Galois group), every group of automorphisms gives rise to a field extension. We'll also go deeper into the different types of field extension: normal extensions (the mirror image of normal subgroups) and separable extensions (which have to do with repeated roots). All of that will lead us towards the fundamental theorem of Galois theory.

Chapter 7

Preparation for the fundamental theorem

Very roughly, the fundamental theorem of Galois theory says that you can tell a lot about a field extension by looking at its Galois group. A bit more specifically, it says that the subgroups and quotients of $\mathrm{Gal}(M : K)$, and their orders, give us information about the subfields of M containing K, and their degrees. For example, one part of the fundamental theorem is that

$$[M : K] = |\mathrm{Gal}(M : K)|.$$

The theorem doesn't hold for all extensions, just those that are 'nice enough'. Crucially, this includes splitting field extensions $\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q}$ of polynomials f over $\mathbb{Q}$—the starting point of classical Galois theory.

Let's dip our toes into the water by thinking about why it might be true that $[M : K] = |\mathrm{Gal}(M : K)|$, at least for extensions that are nice enough.

The easiest nontrivial extensions are the simple algebraic extensions, $M = K(\alpha)$. Write m for the minimal polynomial of α over K and $\alpha_1, \alpha_2, \ldots, \alpha_s$ for the distinct roots of m in M. For every element φ of $\mathrm{Gal}(M : K)$, we have $m(\varphi(\alpha)) = 0$ by Example 6.1.4, and so $\varphi(\alpha) = \alpha_j$ for some $j \in \{1, \ldots, s\}$. On the other hand, for each $j \in \{1, \ldots, s\}$, there is exactly one $\varphi \in \mathrm{Gal}(M : K)$ such that $\varphi(\alpha) = \alpha_j$, by Proposition 6.1.6. So $|\mathrm{Gal}(M : K)| = s$.

On the other hand, $[M : K] = \deg(m)$. So $[M : K] = |\mathrm{Gal}(M : K)|$ if and only if $\deg(m)$ is equal to s, the number of distinct roots of m in M. Certainly $s \leq \deg(m)$. But are s and $\deg(m)$ equal?

There are two reasons why they might not be. First, m might not split in M. For instance, if $K = \mathbb{Q}$ and $\alpha = \sqrt[3]{2}$ then $m(t) = t^3 - 2$, which has only one root in $\mathbb{Q}(\sqrt[3]{2})$, so $|\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q})| = 1 < 3 = \deg(m)$. An algebraic extension is called 'normal' if this problem doesn't occur, that is, if the minimal polynomial of every element does split. That's what Section 7.1 is about.

Second, we might have $s < \deg(m)$ because some of the roots of m in M are repeated. If they are, the number s of *distinct* roots will be less then $\deg(m)$. An algebraic extension is called 'separable' if this problem doesn't occur, that is, if the minimal polynomial of every element has no repeated roots in its splitting field. That's what Section 7.2 is about.

If we take any finite extension $M : K$ (not necessarily simple) that is both normal and separable, then it is indeed true that $|\operatorname{Gal}(M : K)| = [M : K]$. And in fact, these conditions are enough to make the whole fundamental theorem work, as we'll see next week.

I hesitated before putting normality and separability into the same chapter, because you should think of them in quite different ways:

- Normality has a clear conceptual meaning, and its importance was recognized by Galois himself. Despite the name, most field extensions aren't normal. Normality isn't something to be taken for granted.

- In contrast, Galois never considered separability, because it holds automatically over $\mathbb{Q}$ (his focus), and in fact over any field of characteristic 0, as well as any finite field. It takes some work to find an extension that *isn't* separable. You can view separability as more of a technicality.

There's one more concept in this chapter: the 'fixed field' of a group of automorphisms (Section 7.3). Every Galois theory text I've seen contains at least one proof that makes you ask 'how did anyone think of that?' I would argue that the proof of Theorem 7.3.3 is the one and only truly ingenious argument in this course: maybe not the hardest, but the most ingenious. This is not a compliment.

7.1 Normality

Definition 7.1.1 An algebraic field extension $M : K$ is **normal** if for all $\alpha \in M$, the minimal polynomial of α splits in M.

We also say *M **is normal over** K* to mean that $M : K$ is normal.

Lemma 7.1.2 *Let $M : K$ be an algebraic extension. Then $M : K$ is normal if and only if every irreducible polynomial over K either has no roots in M or splits in M.*

Put another way, normality means that any irreducible polynomial over K with at least *one* root in M has *all* its roots in M.

Proof Suppose that $M : K$ is normal, and let f be an irreducible polynomial over K. If f has a root α in M then the minimal polynomial of α is f/c, where $c \in K$

is the leading coefficient of f. Since $M : K$ is normal, f/c splits in M, so f does too.

Conversely, suppose that every irreducible polynomial over K either has no roots in M or splits in M. Let $\alpha \in M$. Then the minimal polynomial of α has at least one root in M (namely, α), so it splits in M. $\qquad\square$

Examples 7.1.3   i. Let $\xi = \sqrt[3]{2} \in \mathbb{R}$, and consider $\mathbb{Q}(\xi) : \mathbb{Q}$. The minimal polynomial of ξ over $\mathbb{Q}$ is $t^3 - 2$, whose roots in $\mathbb{C}$ are $\xi \in \mathbb{R}$ and $\omega\xi, \omega^2\xi \in \mathbb{C} \setminus \mathbb{R}$, where $\omega = e^{2\pi i/3}$. Since $\mathbb{Q}(\xi) \subseteq \mathbb{R}$, the minimal polynomial $t^3 - 2$ does not split in $\mathbb{Q}(\xi)$. Hence $\mathbb{Q}(\xi)$ is not normal over $\mathbb{Q}$.

Alternatively, using the equivalent condition in Lemma 7.1.2, $\mathbb{Q}(\xi) : \mathbb{Q}$ is not normal because $t^3 - 2$ is an irreducible polynomial over $\mathbb{Q}$ that has a root in $\mathbb{Q}(\xi)$ but does not split there.

One way to think about the non-normality of $\mathbb{Q}(\xi) : \mathbb{Q}$ is as follows. The three roots of $t^3 - 2$ are conjugate ('indistinguishable') over $\mathbb{Q}$, since they have the same minimal polynomial. But if they're indistinguishable, it would be strange for an extension to contain some but not all of them—that would be making a distinction between elements that are supposed to be indistinguishable. In this sense, $\mathbb{Q}(\xi) : \mathbb{Q}$ is 'abnormal'.

ii. Let f be a nonzero polynomial over a field K. Then $\mathrm{SF}_K(f) : K$ is always normal, as we shall see (Theorem 7.1.5).

iii. Every extension of degree 2 is normal (just as, in group theory, every subgroup of index 2 is normal). You'll be asked to show this in Workshop 4, but you also know enough to do it now.

What does it mean to be normal?

> **Exercise 7.1.4** What happens if you drop the word 'irreducible' from Lemma 7.1.2? Is it still true?

Normality of field extensions is intimately related to normality of subgroups, and conjugacy in field extensions is also related to conjugacy in groups. (The video 'What does it mean to be normal?' explains both kinds of normality and conjugacy in intuitive terms.)

Here's the first of our two theorems about normal extensions. It describes which extensions arise as splitting field extensions.

> **Theorem 7.1.5** *Let $M : K$ be a field extension. Then*
>
> $$M = \mathrm{SF}_K(f) \text{ for some nonzero } f \in K[t] \iff M : K \text{ is finite and normal}.$$

$$SF_M(m)$$

$$M = K(\alpha_1, \ldots, \alpha_n) \xdashrightarrow[\cong]{\varphi} K(\alpha_1, \ldots, \alpha_n, \varepsilon)$$

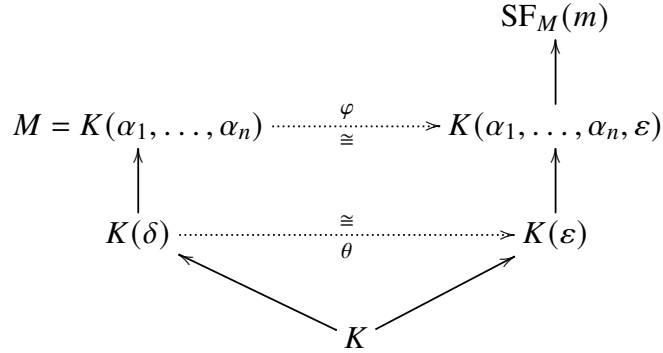$$K(\delta) \xdashrightarrow[\theta]{\cong} K(\varepsilon)$$

$$K$$

Figure 7.1: Maps used in the proof that splitting field extensions are normal.

Proof For $\Leftarrow$, suppose that $M : K$ is finite and normal. By finiteness, there is a basis $\alpha_1, \ldots, \alpha_n$ of M over K, and each α_i is algebraic over K (by Proposition 5.2.4). For each i, let m_i be the minimal polynomial of α_i over K; then by normality, m_i splits in M. Hence $f = m_1 m_2 \cdots m_n \in K[t]$ splits in M. The set of roots of f in M contains $\{\alpha_1, \ldots, \alpha_n\}$, and $M = K(\alpha_1, \ldots, \alpha_n)$, so M is generated over K by the set of roots of f in M. Thus, M is a splitting field of f over K.

For $\Rightarrow$, take a nonzero $f \in K[t]$ such that $M = SF_K(f)$. Write $\alpha_1, \ldots, \alpha_n$ for the roots of f in M. Then $M = K(\alpha_1, \ldots, \alpha_n)$. Each α_i is algebraic over K (since $f \neq 0$), so by Proposition 5.2.4, $M : K$ is finite.

We now show that $M : K$ is normal, which is the most substantial part of the proof (Figure 7.1). Let $\delta \in M$, with minimal polynomial $m \in K[t]$. Certainly m splits in $SF_M(m)$, so to show that m splits in M, it is enough to show that every root ε of m in $SF_M(m)$ lies in M.

Splitting field extensions are normal

Since m is a monic irreducible annihilating polynomial of ε over K, it is the minimal polynomial of ε over K. Hence by Theorem 4.3.16, there is an isomorphism $\theta \colon K(\delta) \to K(\varepsilon)$ over K such that $\theta(\delta) = \varepsilon$. Now observe that:

- $M = SF_{K(\delta)}(f)$, by Lemma 6.2.14(ii);

- $K(\alpha_1, \ldots, \alpha_n, \varepsilon) = SF_{K(\varepsilon)}(f)$, by Lemma 6.2.14(i);

- $\theta_* f = f$, since $f \in K[t]$ and θ is a homomorphism over K.

So we can apply Proposition 6.2.11, which implies that there is an isomorphism $\varphi \colon M \to K(\alpha_1, \ldots, \alpha_n, \varepsilon)$ extending θ. It is an isomorphism over K, since θ is.

Since $\delta \in K(\alpha_1, \ldots, \alpha_n)$ and φ is a homomorphism over K, we have $\varphi(\delta) \in K(\varphi(\alpha_1), \ldots, \varphi(\alpha_n))$. Now $\varphi(\delta) = \theta(\delta) = \varepsilon$, so $\varepsilon \in K(\varphi(\alpha_1), \ldots, \varphi(\alpha_n))$. Moreover, for each i we have $f(\varphi(\alpha_i)) = 0$ (by Example 6.1.4) and so $\varphi(\alpha_i) \in \{\alpha_1, \ldots, \alpha_n\}$. Hence $\varepsilon \in K(\alpha_1, \ldots, \alpha_n) = M$, as required. $\square$

Corollary 7.1.6 *Let $M : L : K$ be field extensions. If $M : K$ is finite and normal then so is $M : L$.*

Proof Follows from Theorem 7.1.5 and Lemma 6.2.14(ii). □

> **Warning 7.1.7** It does *not* follow that $L : K$ is normal. For instance, consider $\mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3}) : \mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}$. The first field is the splitting field of $t^3 - 2$ over $\mathbb{Q}$, and therefore normal over $\mathbb{Q}$, but $\mathbb{Q}(\sqrt[3]{2})$ is not (Example 7.1.3(i)).

Theorem 7.1.5 is the first of two theorems about normality. The second is to do with the action of the Galois group of an extension.

> **Warning 7.1.8** By definition, the Galois group $\mathrm{Gal}(M : K)$ of an extension $M : K$ acts on M. But if M is the splitting field of some polynomial f over K then the action of $\mathrm{Gal}(M : K)$ on M restricts to an action on the roots of f (a finite set), as we saw in Section 6.3. So there are *two* actions of the Galois group in play, one the restriction of the other. Both are important.

When a group acts on a set, a basic question is: what are the orbits? For $\mathrm{Gal}(M : K)$ acting on M, the answer is: the conjugacy classes of M over K. Or at least, that's the case when $M : K$ is finite and normal:

Proposition 7.1.9 *Let $M : K$ be a finite normal extension and $\alpha, \alpha' \in M$. Then*

$$\alpha \text{ and } \alpha' \text{ are conjugate over } K \iff \alpha' = \varphi(\alpha) \text{ for some } \varphi \in \mathrm{Gal}(M : K).$$

Proof For $\Leftarrow$, let $\varphi \in \mathrm{Gal}(M : K)$ with $\alpha' = \varphi(\alpha)$. Then α and α' are conjugate over K, by Example 6.1.4.

For $\Rightarrow$, suppose that α and α' are conjugate over K. Since $M : K$ is finite, both are algebraic over K, and since they are conjugate over K, they have the same minimal polynomial $m \in K[t]$. By Theorem 4.3.16, there is an isomorphism $\theta \colon K(\alpha) \to K(\alpha')$ over K such that $\theta(\alpha) = \alpha'$ (see diagram below).

By Theorem 7.1.5, M is the splitting field of some polynomial f over K. Hence M is also the splitting field of f over both $K(\alpha)$ and $K(\alpha')$, by Lemma 6.2.14(ii). Moreover, $\theta_* f = f$ since θ is a homomorphism over K and f is a polynomial over K. So by Proposition 6.2.11(i), there is an automorphism φ of M extending θ:

$$
\begin{array}{ccc}
M & \overset{\varphi}{\underset{\cong}{\dashrightarrow}} & M \\
\uparrow & & \uparrow \\
K(\alpha) & \overset{\cong}{\underset{\theta}{\dashrightarrow}} & K(\alpha') \\
& \nwarrow \quad \nearrow & \\
& K &
\end{array}
$$
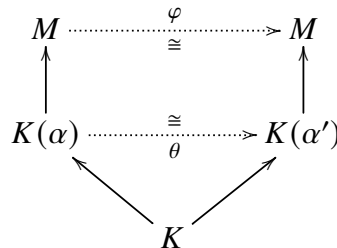
Then $\varphi \in \mathrm{Gal}(M : K)$ with $\varphi(\alpha) = \theta(\alpha) = \alpha'$, as required. $\qquad\square$

Example 7.1.10 Consider the finite normal extension $\mathbb{C} : \mathbb{R}$. Let $\alpha, \alpha' \in \mathbb{C}$. Lemma 1.1.2 states that α and α' are conjugate over $\mathbb{R}$ if and only if α' is either α or $\bar{\alpha}$. Example 6.3.3(i) states that $\mathrm{Gal}(\mathbb{C} : \mathbb{R}) = \{\mathrm{id}, \kappa\}$, where κ is complex conjugation. This confirms Proposition 7.1.9 in the case $\mathbb{C} : \mathbb{R}$.

Proposition 7.1.9 is about the action of $\mathrm{Gal}(M : K)$ on the whole field M, but it has a powerful corollary involving the action of the Galois group on just the roots of an irreducible polynomial f, in the case $M = \mathrm{SF}_K(f)$:

Corollary 7.1.11 *Let f be an irreducible polynomial over a field K. Then the action of $\mathrm{Gal}_K(f)$ on the roots of f in $\mathrm{SF}_K(f)$ is transitive.*

Recall what **transitive** means, for an action of a group G on a set X: for all $x, x' \in X$, there exists $g \in G$ such that $gx = x'$.

Proof Since f is irreducible, the roots of f in $\mathrm{SF}_K(f)$ all have the same minimal polynomial, namely, f divided by its leading coefficient. So they are all conjugate over K. Since $\mathrm{SF}_K(f) : K$ is finite and normal (by Theorem 7.1.5), the result follows from Proposition 7.1.9. $\qquad\square$

> ✏️ **Exercise 7.1.12** Show by example that Corollary 7.1.11 becomes false if you drop the word 'irreducible'.

Example 7.1.13 Let $f(t) = 1 + t + \cdots + t^{p-1} \in \mathbb{Q}[t]$, where p is prime. Since $(1 - t)f(t) = 1 - t^p$, the roots of f in $\mathbb{C}$ are $\omega, \omega^2, \ldots, \omega^{p-1}$, where $\omega = e^{2\pi i/p}$. By Example 3.3.16, f is irreducible over $\mathbb{Q}$. Hence by Corollary 7.1.11, for each $i \in \{1, \ldots, p - 1\}$, there is some $\varphi \in \mathrm{Gal}_\mathbb{Q}(f)$ such that $\varphi(\omega) = \omega^i$.

This is spectacular! Until now, we've been unable to prove such things without a huge amount of explicit checking, which, moreover, only works on a case-by-case basis. For example, if you watched the video 'Calculating Galois groups with bare hands, part 2', you'll have seen how much tedious calculation went into the single case $p = 5, i = 2$:

102
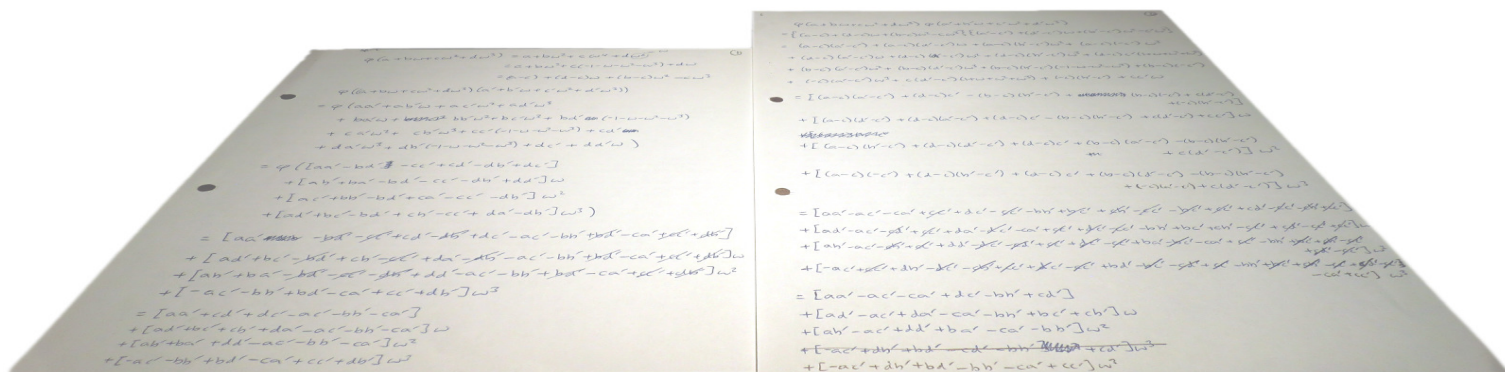
But the theorems we've proved make all this unnecessary.

In fact, for each $i \in \{1, \ldots, p-1\}$, there's exactly *one* element φ_i of $\mathrm{Gal}_{\mathbb{Q}}(f)$ such that $\varphi_i(\omega) = \omega^i$. For since $\mathrm{SF}_{\mathbb{Q}}(f) = \mathbb{Q}(\omega)$, two elements of $\mathrm{Gal}_{\mathbb{Q}}(f)$ that take the same value on ω must be equal. Hence

$$\mathrm{Gal}_{\mathbb{Q}}(f) = \{\varphi_1, \ldots, \varphi_{p-1}\}.$$

You'll see in Workshop 5 that $\mathrm{Gal}_{\mathbb{Q}}(f) \cong C_{p-1}$.

Example 7.1.14 Let's calculate $G = \mathrm{Gal}_{\mathbb{Q}}(t^3 - 2)$. Since $t^3 - 2$ has 3 distinct roots in $\mathbb{C}$, it has 3 distinct roots in its splitting field. By Proposition 6.3.10, G is isomorphic to a subgroup of S_3. Now G acts transitively on the 3 roots, so it has at least 3 elements, so it is isomorphic to either A_3 or S_3. Since two of the roots are non-real complex conjugates, one of the elements of G is complex conjugation, which has order 2. Hence 2 divides $|G|$, forcing $G \cong S_3$.

We now show how a normal field extension gives rise to a normal subgroup. Whenever in life you meet a normal subgroup, you should immediately want to form the quotient, so we do that too.

> **Theorem 7.1.15** *Let $M : L : K$ be field extensions with $M : K$ finite and normal.*
>
> i. *$L : K$ is a normal extension $\iff \varphi L = L$ for all $\varphi \in \mathrm{Gal}(M : K)$.*
>
> ii. *If $L : K$ is a normal extension then $\mathrm{Gal}(M : L)$ is a normal subgroup of $\mathrm{Gal}(M : K)$ and*
> $$\frac{\mathrm{Gal}(M : K)}{\mathrm{Gal}(M : L)} \cong \mathrm{Gal}(L : K).$$

Before the proof, here's some context and explanation.

Part (i) answers the question implicit in Warning 7.1.7: we know from Corollary 7.1.6 that $M : L$ is normal, but when is $L : K$ normal? The notation φL means $\{\varphi(\alpha) : \alpha \in L\}$. For φL to be equal to L means that φ fixes L *as a set* (in other words, permutes it within itself), not that φ fixes each element of L.

In part (ii), it's true for all $M : L : K$ that $\mathrm{Gal}(M : L)$ is a *subset* of $\mathrm{Gal}(M : K)$, since

$$\mathrm{Gal}(M : L) = \{\text{automorphisms } \varphi \text{ of } M \text{ such that } \varphi(\alpha) = \alpha \text{ for all } \alpha \in L\}$$
$$\subseteq \{\text{automorphisms } \varphi \text{ of } M \text{ such that } \varphi(\alpha) = \alpha \text{ for all } \alpha \in K\}$$
$$= \mathrm{Gal}(M : K).$$

And it's always true that $\mathrm{Gal}(M : L)$ is a *subgroup* of $\mathrm{Gal}(M : K)$, as you can easily check. But part (ii) tells us something much more substantial: it's a *normal* subgroup when $L : K$ is a normal extension.

Proof of Theorem 7.1.15 For (i), first suppose that L is normal over K, and let $\varphi \in \mathrm{Gal}(M : K)$. For all $\alpha \in L$, Proposition 7.1.9 implies that α and $\varphi(\alpha)$ are conjugate over K, so they have the same minimal polynomial, so $\varphi(\alpha) \in L$ by normality. Hence $\varphi L \subseteq L$. The same argument with φ^{-1} in place of φ gives $\varphi^{-1} L \subseteq L$, and applying φ to each side then gives $L \subseteq \varphi L$. So $\varphi L = L$.

Conversely, suppose that $\varphi L = L$ for all $\varphi \in \mathrm{Gal}(M : K)$. Let $\alpha \in L$ with minimal polynomial m. Since $M : K$ is normal, m splits in M. Each root α' of m in M is conjugate to α over K, so by Proposition 7.1.9, $\alpha' = \varphi(\alpha)$ for some $\varphi \in \mathrm{Gal}(M : K)$, giving $\alpha' \in \varphi L = L$. Hence m splits in L and $L : K$ is normal.

For (ii), suppose that $L : K$ is normal. To prove that $\mathrm{Gal}(M : L)$ is a normal subgroup of $\mathrm{Gal}(M : K)$, let $\varphi \in \mathrm{Gal}(M : K)$ and $\theta \in \mathrm{Gal}(M : L)$. We show that $\varphi^{-1} \theta \varphi \in \mathrm{Gal}(M : L)$, or equivalently,

$$\varphi^{-1} \theta \varphi(\alpha) = \alpha \text{ for all } \alpha \in L,$$

or equivalently,

$$\theta \varphi(\alpha) = \varphi(\alpha) \text{ for all } \alpha \in L.$$

But by (i), $\varphi(\alpha) \in L$ for all $\alpha \in L$, so $\theta(\varphi(\alpha)) = \varphi(\alpha)$ since $\theta \in \mathrm{Gal}(M : L)$. This completes the proof that $\mathrm{Gal}(M : L) \trianglelefteq \mathrm{Gal}(M : K)$.

Finally, we prove the statement on quotients (still supposing that $L : K$ is a normal extension). Every automorphism φ of M over K satisfies $\varphi L = L$ (by (i)), and therefore restricts to an automorphism $\hat{\varphi}$ of L. The function

$$\nu: \quad \mathrm{Gal}(M : K) \quad \to \quad \mathrm{Gal}(L : K)$$
$$\varphi \quad \mapsto \quad \hat{\varphi}$$

is a group homomorphism, since it preserves composition. Its kernel is $\mathrm{Gal}(M : L)$, by definition. If we can prove that ν is surjective then the last part of the theorem will follow from the first isomorphism theorem.

To prove that ν is surjective, we must show that each automorphism ψ of L over K extends to an automorphism φ of M:
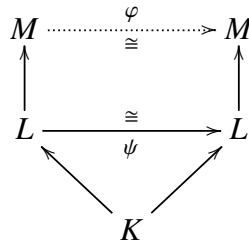
The argument is similar to the second half of the proof of Proposition 7.1.9. By Theorem 7.1.5, M is the splitting field of some $f \in K[t]$. Then M is also the splittting field of f over L. Also, $\psi_* f = f$ since ψ is a homomorphism over K and

104

f is a polynomial over K. So by Proposition 6.2.11(i), there is an automorphism φ of M extending ψ, as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Example 7.1.16 Take $M : L : K$ to be

$$\mathbb{Q}(\xi, \omega) : \mathbb{Q}(\omega) : \mathbb{Q},$$

where $\xi = \sqrt[3]{2}$ and $\omega = e^{2\pi i/3}$. As you will recognize by now, $\mathbb{Q}(\xi, \omega)$ is the splitting field of $t^3 - 2$ over $\mathbb{Q}$, so it is a finite normal extension of $\mathbb{Q}$ by Theorem 7.1.5.

Also, $\mathbb{Q}(\omega)$ is the splitting field of $t^2 + t + 1$ over $\mathbb{Q}$, so it too is a normal extension of $\mathbb{Q}$. Part (i) of Theorem 7.1.15 implies that every element of $\mathrm{Gal}_{\mathbb{Q}}(t^3 - 2)$ restricts to an automorphism of $\mathbb{Q}(\omega)$.

Part (ii) implies that

$$\mathrm{Gal}\big(\mathbb{Q}(\xi, \omega) : \mathbb{Q}(\omega)\big) \trianglelefteq \mathrm{Gal}\big(\mathbb{Q}(\xi, \omega) : \mathbb{Q}\big)$$

and that

$$\frac{\mathrm{Gal}\big(\mathbb{Q}(\xi, \omega) : \mathbb{Q}\big)}{\mathrm{Gal}\big(\mathbb{Q}(\xi, \omega) : \mathbb{Q}(\omega)\big)} \cong \mathrm{Gal}(\mathbb{Q}(\omega) : \mathbb{Q}). \qquad\qquad (7.1)$$

What does this say explicitly? We showed in Example 7.1.14 that $\mathrm{Gal}(\mathbb{Q}(\xi, \omega) : \mathbb{Q}) \cong S_3$. That is, each element of the Galois group permutes the three roots

$$\xi, \ \omega\xi, \ \omega^2\xi$$

of $t^3 - 2$, and all six permutations are realized by some element of the Galois group. An element of $\mathrm{Gal}(\mathbb{Q}(\xi, \omega) : \mathbb{Q})$ that fixes ω is determined by which of the three roots ξ is mapped to, so $\mathrm{Gal}(\mathbb{Q}(\xi, \omega) : \mathbb{Q}(\omega)) \cong A_3$. Finally, $\mathrm{Gal}(\mathbb{Q}(\omega) : \mathbb{Q}) \cong C_2$ by Example 7.1.13. So in this case, the isomorphism (7.1) states that

$$\frac{S_3}{A_3} \cong C_2.$$

Exercise 7.1.17 Draw a diagram showing the three roots of $t^3 - 2$ and the elements of $H = \mathrm{Gal}(\mathbb{Q}(\xi, \omega) : \mathbb{Q}(\omega))$ acting on them. There is a simple geometric description of the elements of $\mathrm{Gal}(\mathbb{Q}(\xi, \omega) : \mathbb{Q})$ that belong to the subgroup H. What is it?

7.2 Separability

Theorem 6.2.13 implies that $|\mathrm{Gal}(M : K)| \leq [M : K]$ whenever $M : K$ is a splitting field extension. Why is this an inequality, not an equality? The answer

can be traced back to the proof of Proposition 6.2.11 on extension of isomorphisms. There, we had an irreducible polynomial called $\psi_* m$, and we wrote s for the number of distinct roots of $\psi_* m$ in its splitting field. Ultimately, the source of the inequality was the fact that $s \leq \deg(\psi_* m)$.

But is this last inequality actually an equality? That is, does an irreducible polynomial of degree d always have d distinct roots in its splitting field? Certainly it has d roots when counted *with multiplicity*. But there will be fewer than d *distinct* roots if any of the roots are repeated (have multiplicity ≥ 2). The question is whether this can ever happen.

Formally, for a polynomial $f(t) \in K[t]$ and a root α of f in some extension M of K, we say that α is a **repeated** root if $(t - \alpha)^2 \mid f(t)$ in $M[t]$.

Exercise 7.2.1 Try to find an example of an irreducible polynomial of degree d with fewer than d distinct roots in its splitting field. Or if you can't, see if you can prove that this is impossible over $\mathbb{Q}$: that is, an irreducible over $\mathbb{Q}$ has no repeated roots in $\mathbb{C}$. Both are quite hard, but ten minutes spent trying may help you to appreciate what's to come.

Definition 7.2.2 An irreducible polynomial over a field is **separable** if it has no repeated roots in its splitting field.

Equivalently, an irreducible polynomial $f \in K[t]$ is separable if it splits into *distinct* linear factors in $\mathrm{SF}_K(f)$:

$$f(t) = a(t - \alpha_1) \cdots (t - \alpha_n)$$

for some $a \in K$ and *distinct* $\alpha_1, \ldots, \alpha_n \in \mathrm{SF}_K(f)$. Put another way, an irreducible f is separable if and only if it has $\deg(f)$ distinct roots in its splitting field.

Example 7.2.3 $t^3 - 2 \in \mathbb{Q}[t]$ is separable, since it has 3 distinct roots in $\mathbb{C}$, hence in its splitting field.

Example 7.2.4 This is an example of an irreducible polynomial that's *in*separable. It's a little bit complicated, but it's the simplest example there is.

Let p be a prime number. We will consider the field $K = \mathbb{F}_p(u)$ of rational expressions over $\mathbb{F}_p$ in an indeterminate (variable symbol) u. Put $f(t) = t^p - u \in K[t]$. We will show that f is an inseparable irreducible polynomial.

By definition, f has at least one root α in its splitting field. But the roots of f are the pth roots of u, and in fields of characteristic p, each element has at most one pth root (Corollary 2.3.22(i)). So α is the *only* root of f in $\mathrm{SF}_K(f)$, despite f having degree $p > 1$. Alternatively, we can argue like this:

$$f(t) = t^p - u = t^p - \alpha^p = (t - \alpha)^p,$$

106

where the last step comes from the Frobenius map of $\mathrm{SF}_K(f)$ being a homomorphism (Proposition 2.3.20(i)).

We now show that f is irreducible over K. Suppose it is reducible. The unique factorization of f into irreducible polynomials over $\mathrm{SF}_K(f)$ is $f(t) = (t - \alpha)^p$, so any nontrivial factorization of f in $K[t]$ is of the form

$$f(t) = (t - \alpha)^i (t - \alpha)^{p-i}$$

where $0 < i < p$ and both factors belong to $K[t]$. The coefficient of t^{i-1} in $(t - \alpha)^i$ is $-i\alpha$, so $-i\alpha \in K$. But i is invertible in K, so $\alpha \in K$. Hence u has a pth root in $K = \mathbb{F}_p(u)$, contradicting Exercise 3.1.13.

> **Warning 7.2.5** Definition 7.2.2 is only a definition of separability for *irreducible* polynomials. There *is* a definition of separability for arbitrary polynomials, but it's not simply Definition 7.2.2 with the word 'irreducible' deleted. We won't need it, but here it is: an arbitrary polynomial is called separable if each of its irreducible factors is separable. So t^2 is separable, even though it has a repeated root.

In real analysis, we can test whether a root is repeated by asking whether the derivative is 0 there:
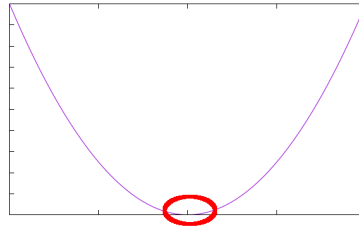
Over an arbitrary field, there's no general definition of the derivative of a function, as there's no meaningful notion of limit. But even without limits, we can differentiate polynomials in the following sense.

Definition 7.2.6 Let K be a field and let $f(t) = \sum_{i=0}^n a_i t^i \in K[t]$. The **formal derivative** of f is

$$(Df)(t) = \sum_{i=1}^n i a_i t^{i-1} \in K[t].$$

We use Df rather than f' to remind ourselves not to take the familiar properties of differentiation for granted. Nevertheless, the usual basic laws hold:

Lemma 7.2.7 *Let K be a field. Then*

$$D(f + g) = Df + Dg, \qquad D(fg) = f \cdot Dg + Df \cdot g, \qquad Da = 0$$

for all $f, g \in K[t]$ and $a \in K$. $\qquad\square$

The real analysis test for repetition of roots has an algebraic analogue:

Lemma 7.2.9 *Let f be a nonzero polynomial over a field K. The following are equivalent:*

   i. f has a repeated root in $\mathrm{SF}_K(f)$;

   ii. f and Df have a common root in $\mathrm{SF}_K(f)$;

   iii. f and Df have a nonconstant common factor in $K[t]$.

Proof (i)$\Longrightarrow$(ii): suppose that f has a repeated root α in $\mathrm{SF}_K(f)$. Then $f(t) = (t - \alpha)^2 g(t)$ for some $g(t) \in (\mathrm{SF}_K(f))[t]$. Hence

$$(Df)(t) = (t - \alpha)\{2g(t) + (t - \alpha) \cdot (Dg)(t)\},$$

so α is a common root of f and Df in $\mathrm{SF}_K(f)$.

   (ii)$\Longrightarrow$(iii): suppose that f and Df have a common root α in $\mathrm{SF}_K(f)$. Then α is algebraic over K (since $f \neq 0$), and the minimal polynomial of α over K is then a nonconstant common factor of f and Df in $K[t]$.

   (iii)$\Longrightarrow$(ii): if f and Df have a nonconstant common factor g then g splits in $\mathrm{SF}_K(f)$, and any root of g in $\mathrm{SF}_K(f)$ is a common root of f and Df.

   (ii)$\Longrightarrow$(i): suppose that f and Df have a common root $\alpha \in \mathrm{SF}_K(f)$. Then $f(t) = (t - \alpha)g(t)$ for some $g \in (\mathrm{SF}_K(f))[t]$, giving

$$(Df)(t) = g(t) + (t - \alpha) \cdot (Dg)(t).$$

But $(Df)(\alpha) = 0$, so $g(\alpha) = 0$, so $g(t) = (t - \alpha)h(t)$ for some $h \in (\mathrm{SF}_K(f))[t]$. Hence $f(t) = (t - \alpha)^2 h(t)$, and α is a repeated root of f in its splitting field. $\square$

The point of Lemma 7.2.9 is that condition (iii) allows us to test for repetition of roots in $\mathrm{SF}_K(f)$ without ever leaving $K[t]$, or even knowing what $\mathrm{SF}_K(f)$ is.

Proposition 7.2.10 *Let f be an irreducible polynomial over a field. Then f is inseparable if and only if $Df = 0$.*

Proof This follows from (i) $\Longleftrightarrow$ (iii) in Lemma 7.2.9. Since f is irreducible, f and Df have a nonconstant common factor if and only if f divides Df; but $\deg(Df) < \deg(f)$, so $f \mid Df$ if and only if $Df = 0$. $\square$

Corollary 7.2.11 *Let K be a field.*

i. If char $K = 0$ *then every irreducible polynomial over K is separable.*

ii. If char $K = p > 0$ *then an irreducible polynomial $f \in K[t]$ is inseparable if and only if*

$$f(t) = b_0 + b_1 t^p + \cdots + b_r t^{rp}$$

for some $b_0, \ldots, b_r \in K$.

In other words, the only irreducible polynomials that are inseparable are the polynomials in t^p in characteristic p. Inevitably, Example 7.2.4 is of this form.

Proof Let $f(t) = \sum a_i t^i$ be an irreducible polynomial. Then f is inseparable if and only if $Df = 0$, if and only if $ia_i = 0$ for all $i \geq 1$. If char $K = 0$, this implies that $a_i = 0$ for all $i \geq 1$, so f is constant, which contradicts f being irreducible. If char $K = p$, then $ia_i = 0$ for all $i \geq 1$ is equivalent to $a_i = 0$ whenever $p \nmid i$. □

Remark 7.2.12 In the final chapter we will show that every irreducible polynomial over a finite field is separable. So, it is only over infinite fields of characteristic p that you have to worry about inseparability.

We now build up to showing that $|\operatorname{Gal}(M : K)| = [M : K]$ whenever $M : K$ is a finite normal extension in which the minimal polynomial of every element of M is separable. First, some terminology:

Definition 7.2.13 Let $M : K$ be an algebraic extension. An element of M is **separable** over K if its minimal polynomial over K is separable. The extension $M : K$ is **separable** if every element of M is separable over K.

Examples 7.2.14   i. Every algebraic extension of fields of characteristic 0 is separable, by Corollary 7.2.11.

ii. Every algebraic extension of a finite field is separable, by Remark 7.2.12.

iii. The splitting field of $t^p - u$ over $\mathbb{F}_p(u)$ is inseparable. Indeed, the element denoted by α in Example 7.2.4 is inseparable over $\mathbb{F}_p(u)$, since its minimal polynomial is the inseparable polynomial $t^p - u$.

> **Exercise 7.2.15** Let $M : L : K$ be field extensions. Show that if $M : K$ is algebraic then so are $M : L$ and $L : K$.

Lemma 7.2.16 *Let $M : L : K$ be field extensions, with $M : K$ algebraic. If $M : K$ is separable then so are $M : L$ and $L : K$.*

Proof Both $M : L$ and $L : K$ are algebraic by Exercise 7.2.15, so it does make sense to ask whether they are separable. (We only defined what it means for an *algebraic* extension to be separable.) That $L : K$ is separable is immediate from the definition. To show that $M : L$ is separable, let $\alpha \in M$. Write m_L and m_K for the minimal polynomials of α over L and K, respectively. Then m_K is an annihilating polynomial of α over L, so $m_L \mid m_K$ in $L[t]$. Since $M : K$ is separable, m_K splits into distinct linear factors in $\mathrm{SF}_K(m_K)$. Since $m_L \mid m_K$, so does m_L. Hence $m_L \in L[t]$ is separable, so α is separable over L. $\qquad\square$

As hinted in the introduction to this section, we will prove that $|\mathrm{Gal}(M : K)| = [M : K]$ by refining Proposition 6.2.11.

$$M \overset{\varphi}{\dashrightarrow} M'$$
$$\uparrow \qquad \uparrow$$
$$f \quad K \underset{\psi}{\rightarrow} K' \quad \psi_* f$$

Proposition 7.2.17 *Let $\psi : K \to K'$ be an isomorphism of fields, let $0 \neq f \in K[t]$, let M be a splitting field of f over K, and let M' be a splitting field of $\psi_* f$ over K'. Suppose that the extension $M' : K'$ is separable. Then there are exactly $[M : K]$ isomorphisms $\varphi : M \to M'$ extending ψ.*

Proof This is almost the same as the proof of Proposition 6.2.11, but with the inequality $s \leq \deg(\psi_* m)$ replaced by an equality, which holds by separability. For the inductive hypothesis to go through, we need the extension $M' : K'(\alpha_j')$ to be separable, and this follows from the separability of $M' : K'$ by Lemma 7.2.16. $\quad\square$

> **Theorem 7.2.18** $|\mathrm{Gal}(M : K)| = [M : K]$ *for every finite normal separable extension $M : K$.*

Proof By Theorem 7.1.5, $M = \mathrm{SF}_K(f)$ for some $f \in K[t]$. The result follows from Proposition 7.2.17, taking $M' = M$, $K' = K$, and $\psi = \mathrm{id}_K$. $\qquad\square$

Examples 7.2.19   i. $|\mathrm{Gal}_K(f)| = [\mathrm{SF}_K(f) : K]$ for any nonzero polynomial f over a field K of characteristic 0.

For instance, if $f(t) = t^3 - 2$ then $[\mathrm{SF}_\mathbb{Q}(f) : \mathbb{Q}] = 6$ by a similar argument to Example 5.1.22, using that $\mathrm{SF}_\mathbb{Q}(f)$ contains elements of degree 2 and 3 over $\mathbb{Q}$. Hence $|\mathrm{Gal}_\mathbb{Q}(f)| = 6$. But $\mathrm{Gal}_\mathbb{Q}(f)$ embeds into S_3 by Proposition 6.3.10, so $\mathrm{Gal}_\mathbb{Q}(f) \cong S_3$. We already proved this in Example 7.1.16, by a different argument.

ii. Consider $K = \mathbb{F}_p(u)$ and $M = \mathrm{SF}_K(t^p - u)$. With notation as in Example 7.2.4, we have $M = K(\alpha)$, so $[M : K] = \deg_K(\alpha) = p$. On the other hand, $|\mathrm{Gal}(M : K)| = 1$ by Corollary 6.3.14. So Theorem 7.2.18 fails if we drop the separability hypothesis.

Digression 7.2.20 With some effort, one can show that in any algebraic extension $M : K$, the separable elements form a subfield of M. (See Stewart, Theorem 17.22.) It follows that a finite extension $K(\alpha_1, \ldots, \alpha_n) : K$ is separable if and only if each α_i is. Hence a splitting field extension $\mathrm{SF}_K(f) : K$ is separable if and only if every root of f is separable in $\mathrm{SF}_K(f)$, which itself is equivalent to f being separable in the sense of Warning 7.2.5.

So: $\mathrm{SF}_K(f)$ is separable over K if and only if f is separable over K. Thus, the different meanings of 'separable' interact nicely.

Digression 7.2.21 It's a stunning fact that every finite separable extension is simple. This is called the theorem of the primitive element. For instance, whenever $\alpha_1, \ldots, \alpha_n$ are complex numbers algebraic over $\mathbb{Q}$, there is some $\alpha \in \mathbb{C}$ (a 'primitive element') such that $\mathbb{Q}(\alpha_1, \ldots, \alpha_n) = \mathbb{Q}(\alpha)$. We saw one case of this in Example 4.3.14(i): $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$.

The theorem of the primitive element was at the heart of most early accounts of Galois theory, and is used in many modern treatments too, but not this one.

7.3 Fixed fields

Write $\mathrm{Aut}(M)$ for the group of automorphisms of a field M. Then $\mathrm{Aut}(M)$ acts naturally on M (Example 2.1.2(ii)). Given a subset S of $\mathrm{Aut}(M)$, we can consider the set $\mathrm{Fix}(S)$ of elements of M fixed by S (Definition 2.1.14).

Lemma 7.3.1 $\mathrm{Fix}(S)$ *is a subfield of M, for any $S \subseteq \mathrm{Aut}(M)$.*

Proof $\mathrm{Fix}(S)$ is the equalizer $\mathrm{Eq}(S \cup \{\mathrm{id}_M\})$, which is a subfield of M by Lemma 2.3.8. $\qquad\qquad\square$

For this reason, we call $\mathrm{Fix}(S)$ the **fixed field** of S.

Exercise 7.3.2 Using Lemma 7.3.1, show that every automorphism of a field is an automorphism over its prime subfield. In other words, $\mathrm{Aut}(M) = \mathrm{Gal}(M : K)$ whenever M is a field with prime subfield K.

Here's the big, ingenious, result about fixed fields. It will play a crucial part in the proof of the fundamental theorem of Galois theory.

> **Theorem 7.3.3** *Let M be a field and H a finite subgroup of* $\mathrm{Aut}(M)$. *Then* $[M : \mathrm{Fix}(H)] \leq |H|$.

So the smaller $|H|$ is, the smaller $[M : \mathrm{Fix}(H)]$ must be, which means that $\mathrm{Fix}(H)$ must be bigger. In other words, the smaller $|H|$ is, the more of M must be fixed by H.

Proof Write $n = |H|$. It is enough to prove that any $n + 1$ elements $\alpha_0, \ldots, \alpha_n$ of M are linearly dependent over $\mathrm{Fix}(H)$.

Write

$$W = \left\{ (x_0, \ldots, x_n) \in M^{n+1} : x_0 \theta(\alpha_0) + \cdots + x_n \theta(\alpha_n) = 0 \text{ for all } \theta \in H \right\}.$$

The size of fixed fields

Then W is defined by n homogeneous linear equations in $n + 1$ variables, so it is a nontrivial M-linear subspace of M^{n+1}.

Claim: let $(x_0, \ldots, x_n) \in W$ and $\varphi \in H$. Then $(\varphi(x_0), \ldots, \varphi(x_n)) \in W$.

Proof: For all $\theta \in H$, we have

$$x_0 (\varphi^{-1} \circ \theta)(\alpha_0) + \cdots + x_n (\varphi^{-1} \circ \theta)(\alpha_n) = 0$$

since $\varphi^{-1} \circ \theta \in H$. Applying φ to both sides gives that for all $\theta \in H$,

$$\varphi(x_0)\theta(\alpha_0) + \cdots + \varphi(x_n)\theta(\alpha_n) = 0,$$

proving the claim.

Define the *length* of a nonzero vector $\mathbf{x} = (x_0, \ldots, x_n)$ to be the unique number $\ell \in \{0, \ldots, n\}$ such that $x_\ell \neq 0$ but $x_{\ell+1} = \cdots = x_n = 0$. Since W is nontrivial, we can choose an element $\mathbf{x}$ of W of minimal length, ℓ. Since W is closed under scalar multiplication by M, we may assume that $x_\ell = 1$. Thus, $\mathbf{x}$ is of the form $(x_0, \ldots, x_{\ell-1}, 1, 0, \ldots, 0)$. Since $\mathbf{x}$ is a nonzero element of W of *minimal* length, the only element of W of the form $(y_0, \ldots, y_{\ell-1}, 0, 0, \ldots, 0)$ is $\mathbf{0}$.

We now show that $x_i \in \mathrm{Fix}(H)$ for all i. Let $\varphi \in H$. By the claim, $(\varphi(x_0), \ldots, \varphi(x_n)) \in W$. Put

$$\mathbf{y} = (\varphi(x_0) - x_0, \ldots, \varphi(x_n) - x_n).$$

Since W is a linear subspace, $\mathbf{y} \in W$. Now $\varphi(x_i) - x_i = 0 - 0 = 0$ for all $i > \ell$ and $\varphi(x_\ell) - x_\ell = \varphi(1) - 1 = 0$, so by the last sentence of the previous paragraph, $\mathbf{y} = \mathbf{0}$. In other words, $\varphi(x_i) = x_i$ for all i. This holds for all $\varphi \in H$, so $x_i \in \mathrm{Fix}(H)$ for all i.

We have shown that W contains a nonzero vector $\mathbf{x} \in \mathrm{Fix}(H)^{n+1}$. But taking $\theta = \mathrm{id}$ in the definition of W gives $\sum x_i \alpha_i = 0$. Hence $\alpha_0, \ldots, \alpha_n$ are linearly dependent over $\mathrm{Fix}(H)$. $\qquad\square$

Example 7.3.4 Write $\kappa \colon \mathbb{C} \to \mathbb{C}$ for complex conjugation. Then $H = \{\mathrm{id}, \kappa\}$ is a subgroup of $\mathrm{Aut}(\mathbb{C})$, and Theorem 7.3.3 predicts that $[\mathbb{C} : \mathrm{Fix}(H)] \leq 2$. Since $\mathrm{Fix}(H) = \mathbb{R}$, this is true.

Exercise 7.3.5 Find another example of Theorem 7.3.3.

Digression 7.3.6 In fact, Theorem 7.3.3 is an equality: $[M : \mathrm{Fix}(H)] = |H|$. This is proved directly in many Galois theory books (e.g. Stewart, Theorem 10.5). In our approach, it will be a *consequence* of the fundamental theorem of Galois theory rather than a step on the way to proving it.

The reverse inequality, $[M : \mathrm{Fix}(H)] \geq |H|$, is closely related to the result called 'linear independence of characters'. (A good reference is Lang, *Algebra*, 3rd edition, Theorem 4.1.) Another instance of linear independence of characters is that the functions $x \mapsto e^{2\pi i n x}$ ($n \in \mathbb{Z}$) on $\mathbb{R}$ are linearly independent, a fundamental fact in the theory of Fourier series.

We finish by adding a further connecting strand between the concepts of normal extension and normal subgroup, complementary to the strands in Theorem 7.1.15.

Proposition 7.3.7 *Let $M : K$ be a finite normal extension and H a normal subgroup of* $\mathrm{Gal}(M : K)$. *Then* $\mathrm{Fix}(H)$ *is a normal extension of K.*

Proof Since every element of H is an automorphism over K, the subfield $\mathrm{Fix}(H)$ of M contains K. For each $\varphi \in \mathrm{Gal}(M : K)$, we have

$$\varphi \, \mathrm{Fix}(H) = \mathrm{Fix}(\varphi H \varphi^{-1}) = \mathrm{Fix}(H),$$

where the first equality holds by Lemma 2.1.15 and the second because H is normal in $\mathrm{Gal}(M : K)$. Hence by Theorem 7.1.15(i), $\mathrm{Fix}(H) : K$ is a normal extension. $\square$

The stage is now set for the central result of the course: the fundamental theorem of Galois theory.

Chapter 8

The fundamental theorem of Galois theory

We've been building up to this moment all semester. Let's do it!

8.1 Introducing the Galois correspondence

Let $M : K$ be a field extension, with K viewed as a subfield of M, as usual.

An **intermediate field** of $M : K$ is a subfield of M containing K. Write

$$\mathscr{F} = \{\text{intermediate fields of } M : K\}.$$

For $L \in \mathscr{F}$, we draw diagrams like this:

$$
\begin{array}{c}
M \\
| \\
L \\
| \\
K,
\end{array}
$$

with the bigger fields higher up.

Also write

$$\mathscr{G} = \{\text{subgroups of } \mathrm{Gal}(M : K)\}.$$

For $H \in \mathscr{G}$, we draw diagrams like this:

$$
\begin{array}{c}
1 \\
| \\
H \\
| \\
\mathrm{Gal}(M : K).
\end{array}
$$

114

Here 1 denotes the trivial subgroup and the bigger groups are *lower down*. It will become clear soon why we're using opposite conventions.

For $L \in \mathscr{F}$, the group $\mathrm{Gal}(M : L)$ consists of all automorphisms φ of M that fix each element of L. Since $K \subseteq L$, any such φ certainly fixes each element of K. Hence $\mathrm{Gal}(M : L)$ is a subgroup of $\mathrm{Gal}(M : K)$. This process defines a function

$$\mathrm{Gal}(M : -) \colon \quad \begin{array}{ccc} \mathscr{F} & \to & \mathscr{G} \\ L & \mapsto & \mathrm{Gal}(M : L). \end{array}$$

In the expression $\mathrm{Gal}(M : -)$, the symbol $-$ should be seen as a blank space into which arguments can be inserted.

Warning 8.1.1 The group we're associating with L is $\mathrm{Gal}(M : L)$, not $\mathrm{Gal}(L : K)$! Both groups matter, but only one is a subgroup of $\mathrm{Gal}(M : K)$, which is what we're interested in here.

We showed just now that $\mathrm{Gal}(M : L)$ is a subgroup of $\mathrm{Gal}(M : K)$. If you wanted to show that $\mathrm{Gal}(L : K)$ is (isomorphic to) a subgroup of $\mathrm{Gal}(M : K)$—which it isn't—then you'd probably do it by trying to prove that every automorphism of L over K extends uniquely to M. And that's false. For instance, when $L = K$, the identity on L typically has *many* extensions to M: they're the elements of $\mathrm{Gal}(M : K)$.

Although $\mathrm{Gal}(L : K)$ isn't a subgroup of $\mathrm{Gal}(M : K)$, it is a *quotient* of it, at least when both extensions are finite and normal. We saw this in Theorem 7.1.15, and we'll come back to it in Section 8.2.

In the other direction, for $H \in \mathscr{G}$, the subfield $\mathrm{Fix}(H)$ of M contains K. Indeed, $H \subseteq \mathrm{Gal}(M : K)$, and by definition, every element of $\mathrm{Gal}(M : K)$ fixes every element of K, so $\mathrm{Fix}(H) \supseteq K$. Hence $\mathrm{Fix}(H)$ is an intermediate field of $M : K$. This process defines a function

$$\mathrm{Fix} \colon \quad \begin{array}{ccc} \mathscr{G} & \to & \mathscr{F} \\ H & \mapsto & \mathrm{Fix}(H). \end{array}$$

We have now defined functions

$$\mathscr{F} \underset{\mathrm{Fix}}{\overset{\mathrm{Gal}(M:-)}{\rightleftarrows}} \mathscr{G}.$$

The fundamental theorem of Galois theory tells us how these functions behave: how the concepts of Galois group and fixed field interact. It will bring together most of the big results we've proved so far, and will assume that the extension is finite, normal and separable. But first, let's say the simple things that are true for all extensions:

115

$$
\begin{array}{ccc}
M & \quad & 1 \\
| & & | \\
L_2 & & \mathrm{Gal}(M : L_2) \\
| & & | \\
L_1 & & \mathrm{Gal}(M : L_1) \\
| & & | \\
K & & \mathrm{Gal}(M : K)
\end{array}
$$
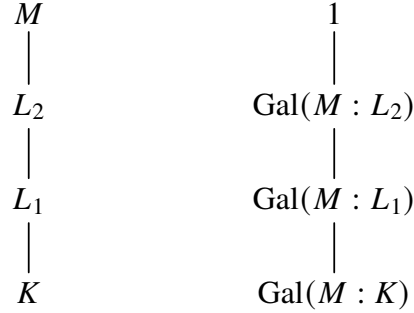
Figure 8.1: The function $L \mapsto \mathrm{Gal}(M : L)$ is order-reversing (Lemma 8.1.2(i)).

Lemma 8.1.2 *Let $M : K$ be a field extension, and define $\mathscr{F}$ and $\mathscr{G}$ as above.*

i. For $L_1, L_2 \in \mathscr{F}$,

$$L_1 \subseteq L_2 \Longrightarrow \mathrm{Gal}(M : L_1) \supseteq \mathrm{Gal}(M : L_2)$$

(Figure 8.1). For $H_1, H_2 \in \mathscr{G}$,

$$H_1 \subseteq H_2 \Longrightarrow \mathrm{Fix}(H_1) \supseteq \mathrm{Fix}(H_2).$$

ii. For $L \in \mathscr{F}$ and $H \in \mathscr{G}$,

$$L \subseteq \mathrm{Fix}(H) \iff H \subseteq \mathrm{Gal}(M : L).$$

iii. For all $L \in \mathscr{F}$,
$$L \subseteq \mathrm{Fix}(\mathrm{Gal}(M : L)).$$

For all $H \in \mathscr{G}$,
$$H \subseteq \mathrm{Gal}(M : \mathrm{Fix}(H)).$$

Warning 8.1.3 In part (i), the functions $\mathrm{Gal}(M : -)$ and Fix *reverse* inclusions. The bigger you make L, the smaller you make $\mathrm{Gal}(M : L)$, because it gets harder for an automorphism to fix everything in L. And the bigger you make H, the smaller you make $\mathrm{Fix}(H)$, because it gets harder for an element of M to be fixed by everything in H. That's why the field and group diagrams are opposite ways up.

Proof (i): I leave the first half as an exercise. For the second, suppose that $H_1 \subseteq H_2$, and let $\alpha \in \text{Fix}(H_2)$. Then $\theta(\alpha) = \alpha$ for all $\theta \in H_2$, so $\theta(\alpha) = \alpha$ for all $\theta \in H_1$, so $\alpha \in \text{Fix}(H_1)$.

(ii): both sides are equivalent to the statement that $\theta(\alpha) = \alpha$ for all $\theta \in H$ and $\alpha \in L$.

(iii): the first statement follows from the $\Leftarrow$ direction of (ii) by taking $H = \text{Gal}(M : L)$, and the second follows from the $\Rightarrow$ direction of (ii) by taking $L = \text{Fix}(H)$. (Or, they can be proved directly.) $\qquad\square$

> **Exercise 8.1.4** Prove the first half of Lemma 8.1.2(i).

> **Exercise 8.1.5** Draw a diagram like Figure 8.1 for the second half of Lemma 8.1.2(i).

> **Digression 8.1.6** If you've done some algebraic geometry, the formal structure of Lemma 8.1.2 might seem familiar. Given a field K and a natural number n, we can form the set $\mathscr{F}$ of subsets of K^n and the set $\mathscr{G}$ of ideals of $K[t_1, \ldots, t_n]$, and there are functions $\mathscr{F} \rightleftarrows \mathscr{G}$ defined by taking the annihilating ideal of a subset of K^n and the zero-set of an ideal of $K[t_1, \ldots, t_n]$. The analogue of Lemma 8.1.2 holds.
>
> In general, a pair of ordered sets $\mathscr{F}$ and $\mathscr{G}$ equipped with functions $\mathscr{F} \rightleftarrows \mathscr{G}$ satisfying the properties in Lemma 8.1.2 is called a **Galois connection**. This in turn is a special case of the category-theoretic notion of adjoint functors.

The functions

$$\mathscr{F} \xrightleftarrows[\text{Fix}]{\text{Gal}(M:-)} \mathscr{G}.$$

are called the **Galois correspondence** for $M : K$. This terminology is mostly used in the case where the functions are **mutually inverse**, meaning that

$$L = \text{Fix}(\text{Gal}(M : L)), \qquad H = \text{Gal}(M : \text{Fix}(H))$$

for all $L \in \mathscr{F}$ and $H \in \mathscr{G}$. We saw in Lemma 8.1.2(iii) that in both cases, the left-hand side is a subset of the right-hand side. But they are not always equal:

Example 8.1.7 Let $M : K$ be $\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}$. Since $[M : K]$ is 3, which is a prime number, the tower law implies that there are no nontrivial intermediate fields: $\mathscr{F} = \{M, K\}$. We saw in Example 6.3.3(ii) that $\text{Gal}(M : K)$ is trivial, so $\mathscr{G} = \{\text{Gal}(M : K)\}$. Hence $\mathscr{F}$ has two elements and $\mathscr{G}$ has only one. This

makes it impossible for there to be mutually inverse functions between $\mathscr{F}$ and $\mathscr{G}$. Specifically, what goes wrong is that

$$\mathrm{Fix}\big(\mathrm{Gal}\big(\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}\big)\big) = \mathrm{Fix}\big(\big\{\mathrm{id}_{\mathbb{Q}(\sqrt[3]{2})}\big\}\big) = \mathbb{Q}(\sqrt[3]{2}) \neq \mathbb{Q}.$$

Exercise 8.1.8 Let p be a prime number, let $K = \mathbb{F}_p(u)$, and let M be the splitting field of $t^p - u$ over K, as in Examples 7.2.4 and 7.2.19(ii). Prove that $\mathrm{Gal}(M : -)$ and Fix are not mutually inverse.

If $\mathrm{Gal}(M : -)$ and Fix *are* mutually inverse then they set up a one-to-one correspondence between the set $\mathscr{F}$ of intermediate fields of $M : K$ and the set $\mathscr{G}$ of subgroups of $\mathrm{Gal}(M : K)$. The fundamental theorem of Galois theory tells us that this dream comes true when $M : K$ is finite, normal and separable. And it tells us more besides.

8.2 The theorem

The moment has come.

Theorem 8.2.1 (Fundamental theorem of Galois theory) *Let $M : K$ be a finite normal separable extension. Write*

$$\mathscr{F} = \{intermediate\ fields\ of\ M : K\},$$
$$\mathscr{G} = \{subgroups\ of\ \mathrm{Gal}(M : K)\}.$$

i. The functions $\mathscr{F} \underset{\mathrm{Fix}}{\overset{\mathrm{Gal}(M:-)}{\rightleftarrows}} \mathscr{G}$ are mutually inverse.

ii. $|\mathrm{Gal}(M : L)| = [M : L]$ for all $L \in \mathscr{F}$, and $[M : \mathrm{Fix}(H)] = |H|$ for all $H \in \mathscr{G}$.

iii. Let $L \in \mathscr{F}$. Then

$$L\ is\ a\ normal\ extension\ of\ K$$
$$\Longleftrightarrow \mathrm{Gal}(M : L)\ is\ a\ normal\ subgroup\ of\ \mathrm{Gal}(M : K),$$

and in that case,

$$\frac{\mathrm{Gal}(M : K)}{\mathrm{Gal}(M : L)} \cong \mathrm{Gal}(L : K).$$

Proof First note that for each $L \in \mathscr{F}$, the extension $M : L$ is finite and normal (by Corollary 7.1.6) and separable (by Lemma 7.2.16). Also, the group $\mathrm{Gal}(M : K)$ is finite (by Theorem 7.2.18), so every subgroup is finite too.

We prove (i) and (ii) together. First let $H \in \mathscr{G}$. We have

$$|H| \leq |\mathrm{Gal}(M : \mathrm{Fix}(H))| = [M : \mathrm{Fix}(H)] \leq |H|, \tag{8.1}$$

where the first inequality holds because $H \subseteq \mathrm{Gal}(M : \mathrm{Fix}(H))$ (Lemma 8.1.2(iii)), the equality follows from Theorem 7.2.18 (since $M : \mathrm{Fix}(H)$ is finite, normal and separable), and the second inequality follows from Theorem 7.3.3 (since H is finite). So equality holds throughout (8.1), giving

$$H = \mathrm{Gal}(M : \mathrm{Fix}(H)), \qquad [M : \mathrm{Fix}(H)] = |H|.$$

Now let $L \in \mathscr{F}$. We have

$$[M : \mathrm{Fix}(\mathrm{Gal}(M : L))] = |\mathrm{Gal}(M : L)| = [M : L],$$

where the first equality follows from the previous paragraph by taking $H = \mathrm{Gal}(M : L)$, and the second follows from Theorem 7.2.18. But $L \subseteq \mathrm{Fix}(\mathrm{Gal}(M : L))$ by Lemma 8.1.2(iii), so $L = \mathrm{Fix}(\mathrm{Gal}(M : L))$ by Workshop 3, q. 1. This completes the proof of (i) and (ii).

We have already proved most of (iii) as Theorem 7.1.15(ii). It only remains to show that whenever L is an intermediate field such that $\mathrm{Gal}(M : L)$ is a normal subgroup of $\mathrm{Gal}(M : K)$, then L is a normal extension of K. By Proposition 7.3.7, $\mathrm{Fix}(\mathrm{Gal}(M : L)) : K$ is normal. But by (i), $\mathrm{Fix}(\mathrm{Gal}(M : L)) = L$, so $L : K$ is normal, as required. $\qquad\square$

The fundamental theorem of Galois theory is about field extensions that are finite, normal and separable. Let's take a moment to think about what those conditions mean.

An extension $M : K$ is finite and normal if and only if M is the splitting field of some polynomial over K (Theorem 7.1.5). So, the theorem can be understood as a result about splitting fields of polynomials.

Not every splitting field extension is separable (Example 7.2.14(iii)). However, we know of two settings where separability is guaranteed. The first is fields of characteristic zero (Example 7.2.14(i)). The most important of these is $\mathbb{Q}$, which is our focus in this chapter: we'll consider examples in which $M : K$ is the splitting field extension of a polynomial over $K = \mathbb{Q}$. The second is where the fields are finite (Example 7.2.14(ii)). We'll come to finite fields in the final chapter.

Digression 8.2.2 Normality and separability are core requirements of Galois theory, but there are extensions of the fundamental theorem (well beyond this course) in which the finiteness condition on $M : K$ is relaxed.

The first level of relaxation replaces 'finite' by 'algebraic'. Then $\mathrm{Gal}(M : K)$ is no longer a finite group, but it does acquire an interesting topology. One example is where M is the algebraic closure $\overline{K}$ of K, and $\mathrm{Gal}(\overline{K} : K)$ is called the **absolute Galois group** of K. It contains *all* splitting fields of polynomials over K, so to study it is to study all polynomials over K at once.

Going further, we can even drop the condition that the extension is algebraic. In this realm, we need the notion of 'transcendence degree', which counts how many algebraically independent elements can be found in the extension. You may have met this if you're taking Algebraic Geometry.

You'll want to see some examples! Section 8.3 is devoted to a single example of the fundamental theorem, showing every aspect of the theorem in all its glory. I'll give a couple of simpler examples in a moment, but before that, it's helpful to review some of what we did earlier:

Remark 8.2.3 When working out the details of the Galois correspondence for a polynomial $f \in K[t]$, it's not only the fundamental theorem that's useful. Some of our earlier results also come in handy, such as the following.

   i. Lemmas 6.3.7 and 6.3.8 state that $\mathrm{Gal}_K(f)$ acts faithfully on the set of roots of f in $\mathrm{SF}_K(f)$. That is, an element of the Galois group can be understood as a permutation of the roots.

   ii. Corollary 6.3.14 states that $|\mathrm{Gal}_K(f)|$ divides $k!$, where k is the number of distinct roots of f in its splitting field.

   iii. Let α and β be roots of f in $\mathrm{SF}_K(f)$. Then there is an element of the Galois group mapping α to β if and only if α and β are conjugate over K (have the same minimal polynomial). This follows from Proposition 7.1.9.

   iv. In particular, when f is irreducible, the action of the Galois group on the roots is transitive (Corollary 7.1.11). See Example 7.1.13 for an illustration of the power of this principle.

Example 8.2.4 Let $M : K$ be a normal separable extension of prime degree p. By the fundamental theorem, $|\mathrm{Gal}(M : K)| = [M : K] = p$. Every group of prime order is cyclic, so $\mathrm{Gal}(M : K) \cong C_p$. By the tower law, $M : K$ has no nontrivial intermediate fields, and by Lagrange's theorem, $\mathrm{Gal}(M : K)$ has no nontrivial

subgroups. So $\mathscr{F} = \{M, K\}$ and $\mathscr{G} = \{1, \mathrm{Gal}(M : K)\}$:

$$
\begin{array}{ccc}
M & \qquad & 1 \\
| & & | \\
K & & \mathrm{Gal}(M : K)
\end{array}
$$

Both M and K are normal extensions of K, and both 1 and $\mathrm{Gal}(M : K)$ are normal subgroups of $\mathrm{Gal}(M : K)$.

Example 8.2.5 Let $f(t) = (t^2 + 1)(t^2 - 2) \in \mathbb{Q}[t]$. Put $M = \mathrm{SF}_\mathbb{Q}(f) = \mathbb{Q}(\sqrt{2}, i)$ and $G = \mathrm{Gal}(M : K) = \mathrm{Gal}_\mathbb{Q}(f)$. Then $M : K$ is a finite normal separable extension, so the fundamental theorem applies. We already calculated G in a sketchy way in Example 6.3.6(iii). Let's do it again in full, using what we now know.

First,

$$[M : K] = \big[\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}(\sqrt{2})\big]\big[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}\big] = 2 \times 2 = 4$$

(much as in Example 5.1.18).

Now consider how G acts on the set $\{\pm\sqrt{2}, \pm i\}$ of roots of f. The conjugacy class of $\sqrt{2}$ is $\{\sqrt{2}, -\sqrt{2}\}$, so for each $\varphi \in G$ we have $\varphi(\sqrt{2}) = \pm\sqrt{2}$. Similarly, $\varphi(i) = \pm i$ for each $\varphi \in G$. The two choices of sign determine φ entirely, so $|G| \leq 4$. But by the fundamental theorem, $|G| = [M : K] = 4$, so each of the four possibilities does in fact occur. So $G = \{\mathrm{id}, \varphi_{+-}, \varphi_{-+}, \varphi_{--}\}$, where

$$
\begin{array}{ccc}
\varphi_{+-}(\sqrt{2}) = \sqrt{2}, & \varphi_{-+}(\sqrt{2}) = -\sqrt{2}, & \varphi_{--}(\sqrt{2}) = -\sqrt{2}, \\
\varphi_{+-}(i) = -i, & \varphi_{-+}(i) = i, & \varphi_{--}(i) = -i.
\end{array}
$$

The only two groups of order 4 are C_4 and $C_2 \times C_2$, and each element of G has order 1 or 2, so $G \cong C_2 \times C_2$.

The subgroups of G are

$$
\begin{array}{ccc}
& 1 & \\
\langle\varphi_{+-}\rangle \quad & \langle\varphi_{-+}\rangle & \quad \langle\varphi_{--}\rangle \\
& G &
\end{array}
\tag{8.2}
$$

where lines indicate inclusions. Here $\langle\varphi_{+-}\rangle$ is the subgroup generated by φ_{+-}, which is $\{\mathrm{id}, \varphi_{+-}\}$, and similarly for φ_{-+} and φ_{--}.

What are the fixed fields of these subgroups? The fundamental theorem implies that $\text{Fix}(G) = \mathbb{Q}$. Also, $\varphi_{+-}(\sqrt{2}) = \sqrt{2}$, so $\mathbb{Q}(\sqrt{2}) \subseteq \text{Fix}(\langle \varphi_{+-} \rangle)$. But

$$\left[\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}(\sqrt{2}) \right] = 2 = |\langle \varphi_{+-} \rangle| = \left[\mathbb{Q}(\sqrt{2}, i) : \text{Fix}(\langle \varphi_{+-} \rangle) \right]$$

(where the last step is by the fundamental theorem), so $\mathbb{Q}(\sqrt{2}) = \text{Fix}(\langle \varphi_{+-} \rangle)$.

Let's reflect for a moment on the argument in the last paragraph, because it's one you'll need to master. We have a subgroup H of $\text{Gal}(M : K)$ (here, $H = \langle \varphi_{+-} \rangle$) and we want to find its fixed field. We do this in three steps:

- First, we spot some elements $\alpha_1, \ldots, \alpha_r$ fixed by H. (Here, $r = 1$ and $\alpha_1 = \sqrt{2}$.) It follows that $K(\alpha_1, \ldots, \alpha_r) \subseteq \text{Fix}(H)$.

- Next, we check that $[M : K(\alpha_1, \ldots, \alpha_r)] = |H|$. If they're not equal, that means we didn't spot enough elements fixed by H.

- Finally, we apply a standard argument:

$$[M : K(\alpha_1, \ldots, \alpha_r)] = |H| = [M : \text{Fix}(H)]$$

(using the fundamental theorem), so by the tower law,

$$[\text{Fix}(H) : K(\alpha_1, \ldots, \alpha_r)] = \frac{[M : K(\alpha_1, \ldots, \alpha_r)]}{[M : \text{Fix}(H)]} = 1,$$

giving $\text{Fix}(H) = K(\alpha_1, \ldots, \alpha_r)$.

The strategy is similar to one you've met in linear algebra: to prove that two subspaces of a vector space are equal, show that one is a subspace of the other and that they have the same dimension.

Similar arguments apply to φ_{-+} and φ_{--}, so the fixed fields of the groups in diagram (8.2) are

$$(8.3)$$

Equivalently, the groups in (8.2) are the Galois groups of $\mathbb{Q}(\sqrt{2}, i)$ over the fields in (8.3). For instance, $\text{Gal}(\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}(i)) = \langle \varphi_{-+} \rangle$.

Since the overall Galois group $G \cong C_2 \times C_2$ is abelian, every subgroup is normal. Hence all the extensions in diagram (8.3) are normal too.

> **Exercise 8.2.6** In this particular example, one can also see more directly that all the extensions in (8.3) are normal. How?

Like any big theorem, the fundamental theorem of Galois theory has some important corollaries. Here's one.

Corollary 8.2.7 *Let $M : K$ be a finite normal separable extension. Then for every $\alpha \in M \setminus K$, there is some automorphism φ of M over K such that $\varphi(\alpha) \neq \alpha$.*

Proof Theorem 8.2.1(i) implies that $\mathrm{Fix}(\mathrm{Gal}(M : K)) = K$. Now $\alpha \notin K$, so $\alpha \notin \mathrm{Fix}(\mathrm{Gal}(M : K))$, which is what had to be proved. $\square$

Example 8.2.8 For any $f \in \mathbb{Q}[t]$ and irrational $\alpha \in \mathrm{SF}_{\mathbb{Q}}(f)$, there is some $\varphi \in \mathrm{Gal}_{\mathbb{Q}}(f)$ that does not fix α. This is clear if $\alpha \notin \mathbb{R}$, as we can take φ to be complex conjugation restricted to $\mathrm{SF}_{\mathbb{Q}}(f)$. But it is not so obvious otherwise.

8.3 A specific example

Chapter 13 of Stewart's book opens with these words:

> The extension that we discuss is a favourite with writers on Galois theory, because of its archetypal quality. A simpler example would be too small to illustrate the theory adequately, and anything more complicated would be unwieldy. The example is the Galois group of the splitting field of $t^4 - 2$ over $\mathbb{Q}$.

We go through the same example here. My presentation of it is different from Stewart's, so you can consult his book if anything that follows is unclear.

Write $f(t) = t^4 - 2 \in \mathbb{Q}[t]$, which is irreducible by Eisenstein's criterion. Write $G = \mathrm{Gal}_{\mathbb{Q}}(f)$.

Splitting field Write ξ for the unique real positive root of f. Then the roots of f are $\pm\xi$ and $\pm\xi i$ (Figure 8.2). So $\mathrm{SF}_{\mathbb{Q}}(f) = \mathbb{Q}(\xi, \xi i) = \mathbb{Q}(\xi, i)$. We have

$$[\mathbb{Q}(\xi, i) : \mathbb{Q}] = [\mathbb{Q}(\xi, i) : \mathbb{Q}(\xi)][\mathbb{Q}(\xi) : \mathbb{Q}] = 2 \times 4 = 8,$$

where the first factor is 2 because $\mathbb{Q}(\xi) \subseteq \mathbb{R}$ and the second factor is 4 because f is the minimal polynomial of ξ over $\mathbb{Q}$ (being irreducible). By the fundamental theorem, $|G| = 8$.
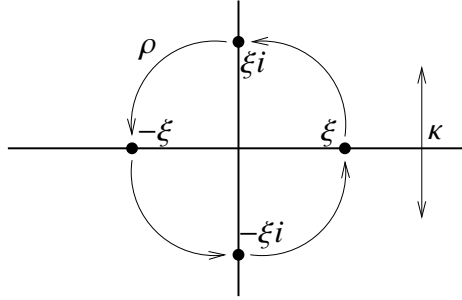
Figure 8.2: The roots of f, and the effects on them of $\rho, \kappa \in \mathrm{Gal}_{\mathbb{Q}}(f)$.

Galois group   We now look for the 8 elements of the Galois group. We'll use the principle that if $\varphi, \theta \in G$ with $\varphi(\xi) = \theta(\xi)$ and $\varphi(i) = \theta(i)$ then $\varphi = \theta$ (by Lemma 4.3.6).

Complex conjugation on $\mathbb{C}$ restricts to an automorphism κ of $\mathbb{Q}(\xi, i)$ over $\mathbb{Q}$, giving an element $\kappa \in G$ of order 2.

I now claim that G has an element ρ satisfying $\rho(\xi) = \xi i$ and $\rho(i) = i$. In that case, ρ will act on the roots of f as follows:

$$\xi \mapsto \xi i \mapsto -\xi \mapsto -\xi i \mapsto \xi$$

(Figure 8.2). This element ρ will have order 4.

Proof of claim: since f is irreducible, G acts transitively on the roots of f in $\mathrm{SF}_{\mathbb{Q}}(f)$, so there is some $\varphi \in G$ such that $\varphi(\xi) = \xi i$. The conjugacy class of i over $\mathbb{Q}$ is $\{\pm i\}$, so $\varphi(i) = \pm i$. If $\varphi(i) = i$ then we can take $\rho = \varphi$. If $\varphi(i) = -i$ then

$$(\varphi \circ \kappa)(\xi) = \varphi(\xi) = \xi i, \qquad (\varphi \circ \kappa)(i) = \varphi(-i) = -\varphi(i) = i,$$

so we can take $\rho = \varphi \circ \kappa$.

(From now on, I will usually omit the $\circ$ sign and write things like $\varphi\kappa$ instead. Of course, juxtaposition is also used to mean multiplication, as in ξi. But confusion shouldn't arise: automorphisms are composed and numbers are multiplied.)

Figure 8.2 might make us wonder if G is the dihedral group D_4, the symmetry group of the square. We will see that it is.

> ⚠ **Warning 8.3.1** The symmetry group of a regular n-sided polygon has $2n$ elements: n rotations and n reflections. Some authors call it D_n and others call it D_{2n}. I will call it D_n, as in the Group Theory course.

If $G \cong D_4$, we should have $\kappa\rho = \rho^{-1}\kappa$. (This is one of the defining equations of the dihedral group; you saw it in Example 3.2.12 of Group Theory.) Let's check

$\varphi \in G$	$\varphi(\xi)$	$\varphi(i)$	$\varphi(\xi i)$	order	geometric description (see Warning 8.3.2)
id	ξ	i	ξi	1	identity
ρ	ξi	i	$-\xi$	4	rotation by $\pi/2$
ρ^2	$-\xi$	i	$-\xi i$	2	rotation by π
$\rho^3 = \rho^{-1}$	$-\xi i$	i	ξ	4	rotation by $-\pi/2$
κ	ξ	$-i$	$-\xi i$	2	reflection in real axis
$\kappa\rho = \rho^{-1}\kappa$	$-\xi i$	$-i$	$-\xi$	2	reflection in axis through $1 - i$
$\kappa\rho^2 = \rho^2\kappa$	$-\xi$	$-i$	ξi	2	reflection in imaginary axis
$\kappa\rho^{-1} = \rho\kappa$	ξi	$-i$	ξ	2	reflection in axis through $1 + i$

Figure 8.3: The Galois group of $t^4 - 2$ over $\mathbb{Q}$.

this. We have

$$\kappa\rho(\xi) = \overline{\overline{\xi i}} = -\xi i, \qquad \rho^{-1}\kappa(\xi) = \rho^{-1}(\xi) = -\xi i,$$
$$\kappa\rho(i) = \kappa(i) = -i, \qquad \rho^{-1}\kappa(i) = \rho^{-1}(-i) = -i,$$

so $\kappa\rho$ and $\rho^{-1}\kappa$ are equal on ξ and i, so $\kappa\rho = \rho^{-1}\kappa$. It follows that $\kappa\rho^r = \rho^{-r}\kappa$ for all $r \in \mathbb{Z}$.

Figure 8.3 shows the effect of 8 elements of G on ξ, i and ξi. Since no two of them have the same effect on both ξ and i, they are all *distinct* elements of G. Since $|G| = 8$, they are the only elements of G. So $G \cong D_4$.

> **Warning 8.3.2** The 'geometric description' in Figure 8.3 applies only to the roots, not the whole of the splitting field $\mathbb{Q}(\xi, i)$. For example, ρ^2 is rotation by π *on the set of roots*, but it is not rotation by π on the rest of $\mathbb{Q}(\xi, i)$: it fixes each element of $\mathbb{Q}$, for instance.

Subgroups of the Galois group   Since $|G| = 8$, any nontrivial proper subgroup of G has order 2 or 4. Let's look in turn at subgroups of order 2 and 4, also determining which ones are normal. This is pure group theory, with no mention of fields.

- The subgroups of order 2 are of the form $\langle\varphi\rangle = \{\mathrm{id}, \varphi\}$ where $\varphi \in G$ has order 2. So, they are

$$\langle\rho^2\rangle, \ \langle\kappa\rangle, \ \langle\kappa\rho\rangle, \ \langle\kappa\rho^2\rangle, \ \langle\kappa\rho^{-1}\rangle.$$

  If you watched the video 'What does it mean to be normal?', you may be able to guess which of these subgroups are normal in G, the symmetry group

of the square. It should be those that can be specified without referring to particular vertices or edges of the square. So, just the first should be normal. Let's check.

We know that $\kappa\rho^2 = \rho^2\kappa$, so ρ^2 commutes with both κ and ρ, which generate G. Hence ρ^2 is in the centre of G (commutes with everything in G). It follows that $\langle\rho^2\rangle$ is a normal subgroup of G. On the other hand, for each $r \in \mathbb{Z}$, the subgroup $\langle\kappa\rho^r\rangle$ is not normal, since

$$\rho(\kappa\rho^r)\rho^{-1} = (\rho\kappa)\rho^{r-1} = (\kappa\rho^{-1})\rho^{r-1} = \kappa\rho^{r-2} \notin \langle\kappa\rho^r\rangle.$$

- The subgroups of G of order 4 are isomorphic to either C_4 or $C_2 \times C_2$, since these are the only groups of order 4.

  The only elements of G of order 4 are $\rho^{\pm 1}$, so the only subgroup of G isomorphic to C_4 is $\langle\rho\rangle = \{\mathrm{id}, \rho, \rho^2, \rho^3 = \rho^{-1}\}$.

  Now consider subgroups H of G isomorphic to $C_2 \times C_2$.

> **Exercise 8.3.3** Show that every such H must contain ρ^2. (Hint: think geometrically.)

We have $\rho^2 \in H$, and both other nonidentity elements of H have order 2, so they are of the form $\kappa\rho^r$ for some $r \in \mathbb{Z}$. The two such subgroups H are

$$\langle\kappa, \rho^2\rangle = \{\mathrm{id}, \kappa, \rho^2, \kappa\rho^2\},$$
$$\langle\kappa\rho, \rho^2\rangle = \{\mathrm{id}, \kappa\rho, \rho^2, \kappa\rho^{-1}\}.$$

Finally, any subgroup of index 2 of any group is normal, so all the subgroups of G of order 4 are normal.

Hence the subgroup structure of $G \cong D_4$ is as follows, where a box around a subgroup means that it is normal in G.

	order
1	order 1
$\langle\kappa\rangle \quad \langle\kappa\rho^2\rangle \quad \langle\rho^2\rangle \quad \langle\kappa\rho\rangle \quad \langle\kappa\rho^{-1}\rangle$	order 2
$\langle\kappa, \rho^2\rangle \cong C_2 \times C_2 \quad \langle\rho\rangle \cong C_4 \quad \langle\kappa\rho, \rho^2\rangle \cong C_2 \times C_2$	order 4
$G = \langle\kappa, \rho\rangle \cong D_4$	order 8
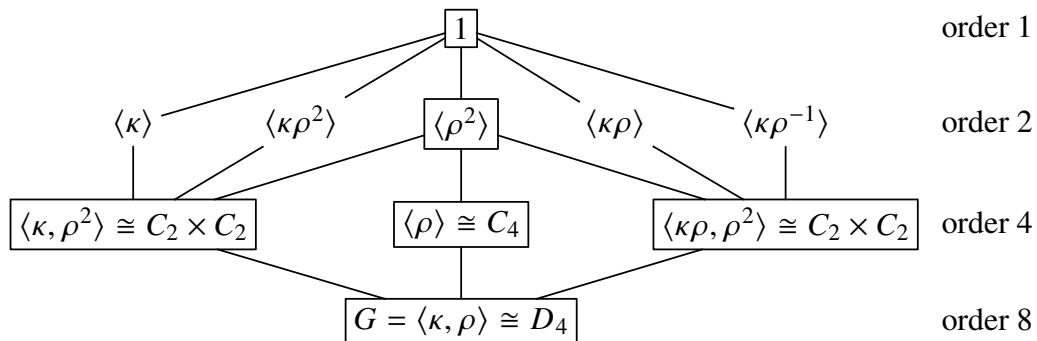
Fixed fields We now find $\mathrm{Fix}(H)$ for each $H \in \mathcal{G}$, again considering the subgroups of orders 2 and 4 in turn. We'll use the same three-step strategy as in Example 8.2.5.
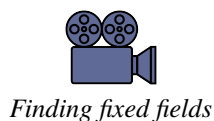
- Order 2: take $\mathrm{Fix}\langle\kappa\rangle$ (officially $\mathrm{Fix}(\langle\kappa\rangle)$, but let's drop the brackets). We have $\kappa(\xi) = \xi$, so $\xi \in \mathrm{Fix}\langle\kappa\rangle$, so $\mathbb{Q}(\xi) \subseteq \mathrm{Fix}\langle\kappa\rangle$. But $[\mathbb{Q}(\xi, i) : \mathbb{Q}(\xi)] = 2$, and by the fundamental theorem, $[\mathbb{Q}(\xi, i) : \mathrm{Fix}\langle\kappa\rangle] = |\langle\kappa\rangle| = 2$, so $\mathrm{Fix}\langle\kappa\rangle = \mathbb{Q}(\xi)$.

  Finding fixed fields

  The same argument shows that for any $\varphi \in G$ of order 2, if we can spot some $\alpha \in \mathbb{Q}(\xi, i)$ such that $\varphi(\alpha) = \alpha$ and $[\mathbb{Q}(\xi, i) : \mathbb{Q}(\alpha)] \leq 2$, then $\mathrm{Fix}\langle\varphi\rangle = \mathbb{Q}(\alpha)$. For $\varphi = \kappa\rho^2$, we can take $\alpha = \xi i$ (by Figure 8.3). We have $\deg_{\mathbb{Q}}(\xi i) = 4$ since ξi is a root of f, so $[\mathbb{Q}(\xi i) : \mathbb{Q}] = 4$, or equivalently, $[\mathbb{Q}(\xi, i) : \mathbb{Q}(\xi i)] = 2$. Hence $\mathrm{Fix}\langle\kappa\rho^2\rangle = \mathbb{Q}(\xi i)$.

  > **Exercise 8.3.4** I took a small liberty in the sentence beginning 'The same argument', because it included an inequality but the previous argument didn't. Prove the statement made in that sentence.

  It is maybe not so easy to spot an α for $\kappa\rho$, but the geometric description in Figure 8.3 suggests taking $\alpha = \xi(1 - i)$. And indeed, one can check that $\kappa\rho$ fixes $\xi(1 - i)$. One can also check that $\xi(1 - i)$ is not the root of any nonzero quadratic over $\mathbb{Q}$, so $\deg_{\mathbb{Q}}(\xi(1 - i))$ is ≥ 4 (since it divides 8), so $[\mathbb{Q}(\xi, i) : \mathbb{Q}(\xi(1-i))] \leq 8/4 = 2$. Hence $\mathrm{Fix}\langle\kappa\rho\rangle = \mathbb{Q}(\xi(1-i))$. Similarly, $\mathrm{Fix}\langle\kappa\rho^{-1}\rangle = \mathbb{Q}(\xi(1 + i))$.

  Finally,
  $$\rho^2(\xi^2) = (\rho^2(\xi))^2 = (-\xi)^2 = \xi^2, \quad \rho^2(i) = i,$$
  so $\mathbb{Q}(\xi^2, i) \subseteq \mathrm{Fix}\langle\rho^2\rangle$. But $[\mathbb{Q}(\xi, i) : \mathbb{Q}(\xi^2, i)] = 2$, so $\mathrm{Fix}\langle\rho^2\rangle = \mathbb{Q}(\xi^2, i)$.

- Order 4: for $H = \langle\kappa, \rho^2\rangle$, note that ξ^2 is fixed by both κ and ρ^2, so $\xi^2 \in \mathrm{Fix}(H)$, so $\mathbb{Q}(\xi^2) \subseteq \mathrm{Fix}(H)$. But $\xi^2 \notin \mathbb{Q}$, so $[\mathbb{Q}(\xi^2) : \mathbb{Q}] \geq 2$, so $[\mathbb{Q}(\xi, i) : \mathbb{Q}(\xi^2)] \leq 4$. The fundamental theorem guarantees that
  $$[\mathbb{Q}(\xi, i) : \mathrm{Fix}(H)] = |H| = 4,$$
  so $\mathrm{Fix}(H) = \mathbb{Q}(\xi^2)$.

  The same argument applies to the other two subgroups H of order 4: if we can spot an element $\alpha \in \mathbb{Q}(\xi, i) \setminus \mathbb{Q}$ fixed by the generators of H, then $\mathrm{Fix}(H) = \mathbb{Q}(\alpha)$. This gives $\mathrm{Fix}\langle\rho\rangle = \mathbb{Q}(i)$ and $\mathrm{Fix}\langle\kappa\rho, \rho^2\rangle = \mathbb{Q}(\xi^2 i)$.

In summary, the fixed fields of the subgroups of G are as follows.

$$
\begin{array}{ccccc}
& & \boxed{\mathbb{Q}(\xi,i)} & & \text{degree 1} \\
\mathbb{Q}(\xi) & \mathbb{Q}(\xi i) & \boxed{\mathbb{Q}(\xi^2,i)} & \mathbb{Q}(\xi(1-i)) & \mathbb{Q}(\xi(1+i)) \quad \text{degree 2} \\
\boxed{\mathbb{Q}(\xi^2)} & & \boxed{\mathbb{Q}(i)} & & \boxed{\mathbb{Q}(\xi^2 i)} \quad \text{degree 4} \\
& & \boxed{\mathbb{Q}} & & \text{degree 8}
\end{array}
$$
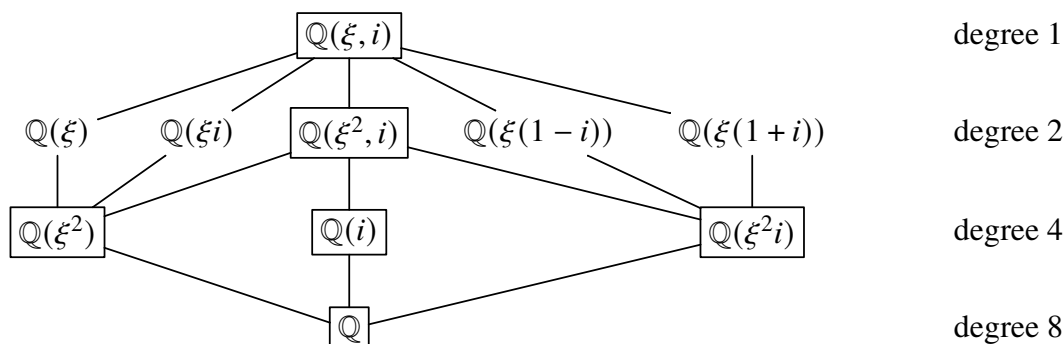
On the right, 'degree' means the degree of $\mathbb{Q}(\xi,i)$ over the subfield concerned, *not* the degree over $\mathbb{Q}$. The fundamental theorem implies that the Galois group of $\mathbb{Q}(\xi,i)$ over each intermediate field is the subgroup of G in the same position in the earlier diagram. For example, $\mathrm{Gal}(\mathbb{Q}(\xi,i) : \mathbb{Q}(\xi^2,i)) = \langle\rho^2\rangle$. It also implies that the intermediate fields that are normal over $\mathbb{Q}$ are the boxed ones.

Quotients   Finally, the fundamental theorem tells us that

$$
\frac{\mathrm{Gal}(\mathbb{Q}(\xi,i) : \mathbb{Q})}{\mathrm{Gal}(\mathbb{Q}(\xi,i) : L)} \cong \mathrm{Gal}(L : \mathbb{Q})
$$

whenever L is an intermediate field normal over $\mathbb{Q}$.

For $L = \mathbb{Q}(\xi^2, i)$, this gives

$$
G/\langle\rho^2\rangle \cong \mathrm{Gal}(\mathbb{Q}(\xi^2,i) : \mathbb{Q}). \tag{8.4}
$$

The left-hand side is the quotient of D_4 by a subgroup isomorphic to C_2. It has order 4, but it has no element of order 4: for the only elements of G of order 4 are $\rho^{\pm 1}$, whose images in $G/\langle\rho^2\rangle$ have order 2. Hence $G/\langle\rho^2\rangle \cong C_2 \times C_2$. On the other hand, $\mathbb{Q}(\xi^2, i)$ is the splitting field over $\mathbb{Q}$ of $(t^2 - 2)(t^2 + 1)$, which by Example 8.2.5 has Galois group $C_2 \times C_2$. This confirms the isomorphism (8.4).

The other three intermediate fields normal over $\mathbb{Q}$, I leave to you:

Normal subgroups and normal extensions

> **Exercise 8.3.5** Choose one of $\mathbb{Q}(\xi^2)$, $\mathbb{Q}(i)$ or $\mathbb{Q}(\xi^2 i)$, and do the same for it as I just did for $\mathbb{Q}(\xi^2, i)$.

As you've now seen, it can take quite some time to work through a particular example of the Galois correspondence. You'll get practice at doing this in workshops.

Beyond examples, there are at least two other uses of the fundamental theorem. The first is to resolve the old question on solvability of polynomials by radicals, which we met back in Chapter 1. The second is to work out the structure of finite fields. We will carry out these two missions in the remaining two weeks.

Chapter 9

Solvability by radicals

We began this course with a notorious old problem: can every polynomial be solved by radicals? Theorem 1.3.5 gave the answer and more: not only is it impossible to find a *general formula* that does it, but we can tell which *specific* polynomials can be solved by radicals.

Theorem 1.3.5 states that a polynomial over $\mathbb{Q}$ is solvable by radicals if and only if it has the right kind of Galois group—a solvable one. In degree 5 and higher, there are polynomials that have the wrong kind of group. These polynomials are not, therefore, solvable by radicals.

We'll prove one half of this 'if and only if' statement: if f is solvable by radicals then $\mathrm{Gal}_{\mathbb{Q}}(f)$ is solvable. This is the half that's needed to show that some polynomials are *not* solvable by radicals. The proof of the other direction is in Chapter 18 of Stewart's book, but we won't do it.

If you're taking Algebraic Topology, you'll already be familiar with the idea that groups can be used to solve problems that seem to have nothing to do with groups. You have a problem about some objects (such as topological spaces or field extensions), you associate groups with those objects (maybe their fundamental groups or their Galois groups), you translate your original problem into a problem about groups, and you solve that instead. For example, the question of whether $\mathbb{R}^2$ and $\mathbb{R}^3$ are homeomorphic is quite difficult using only general topology; but using algebraic topology, we can answer 'no' by noticing that the fundamental group of $\mathbb{R}^2$ with a point removed is not isomorphic to the fundamental group of $\mathbb{R}^3$ with a point removed. In much the same way, we'll answer a difficult question about field extensions by converting it into a question about groups.

For this chapter, you'll need what you know about solvable groups. At a minimum, you'll need the definition, the fact that any quotient of a solvable group is solvable, and the fact that S_5 is not solvable.

9.1 Radicals

We speak of square roots, cube roots, and so on, but we also speak about roots of polynomials. To distinguish between these two related usages, we will use the word **radical** for square roots etc. (*Radical* comes from the Latin for root. A radish is a root, and a change or policy is radical if it gets to the root of the matter.)

Back in Chapter 1, I said that a complex number is called radical if 'it can be obtained from the rationals using only the usual arithmetic operations [addition, subtraction, multiplication and division] and kth roots [for $k \geq 1$]'. As an example, I said that

$$\frac{\frac{1}{2} + \sqrt[3]{\sqrt[3]{2} - \sqrt[2]{7}}}{\sqrt[4]{6 + \sqrt[5]{\frac{2}{3}}}} \tag{9.1}$$

is radical, whichever square root, cube root, etc., we choose (p. 12). Let's now make this definition precise.

The first point is that the notation $\sqrt[n]{z}$ or $z^{1/n}$ is highly dangerous:

> **Warning 9.1.1** Let z be a complex number and $n \geq 2$. Then there is **no single number called** $\sqrt[n]{z}$ **or** $z^{1/n}$. There are n elements α of $\mathbb{C}$ such that $\alpha^n = z$. So, the notation $\sqrt[n]{z}$ or $z^{1/n}$ makes no sense if it is intended to denote a single complex number. It is simply invalid.
>
> When z belongs to the set $\mathbb{R}^+$ of nonnegative reals, the convention is that $\sqrt[n]{z}$ or $z^{1/n}$ denotes the unique $\alpha \in \mathbb{R}^+$ such that $\alpha^n = z$. There is also a widespread convention that when z is a negative real and n is odd, $\sqrt[n]{z}$ or $z^{1/n}$ denotes the unique real α such that $\alpha^n = z$. In these cases, there is a sensible and systematic way of choosing one of the nth roots of z. But for a general z and n, there is not.
>
> Complex analysis has a lot to say about different choices of nth roots. But we don't need to go into that. We simply treat all the nth roots of z on an equal footing, not attempting to pick out any of them as special.

With this warning in mind, we define the radical numbers without using notation like $\sqrt[n]{z}$ or $z^{1/n}$. It is a 'top-down' definition, in the sense of Section 2.2. Loosely, it says that the radical numbers form the smallest subfield of $\mathbb{C}$ closed under taking square roots, cube roots, etc.

The definition of radical number

Definition 9.1.2 Let $\mathbb{Q}^{\mathrm{rad}}$ be the smallest subfield of $\mathbb{C}$ such that for $\alpha \in \mathbb{C}$,

$$\alpha^n \in \mathbb{Q}^{\mathrm{rad}} \text{ for some } n \geq 1 \;\Rightarrow\; \alpha \in \mathbb{Q}^{\mathrm{rad}}. \tag{9.2}$$

A complex number is **radical** if it belongs to $\mathbb{Q}^{\mathrm{rad}}$.

So any rational number is radical; any nth root of a radical number is radical; the sum, product, difference or quotient of radical numbers is radical; and there are no more radical numbers than can be obtained by those rules.

For the definition of $\mathbb{Q}^{\text{rad}}$ to make sense, we need there to *be* a smallest subfield of $\mathbb{C}$ with the property (9.2). This will be true as long as the intersection of any family of subfields of $\mathbb{C}$ satisfying (9.2) is again a subfield of $\mathbb{C}$ satisfying (9.2): for then $\mathbb{Q}^{\text{rad}}$ is the intersection of *all* subfields of $\mathbb{C}$ satisfying (9.2).

> **Exercise 9.1.3** Check that the intersection of any family of subfields of $\mathbb{C}$ satisfying (9.2) is again a subfield of $\mathbb{C}$ satisfying (9.2). (That any intersection of subfields is a subfield is a fact we met back on p. 29; the new aspect is (9.2).)

Example 9.1.4 Consider again the expression (9.1). It's not quite as random as it looks. I chose it so that the various radicals are covered by one of the two conventions mentioned in Warning 9.1.1: they're all nth roots of positive reals except for $\sqrt[3]{\sqrt[3]{2} - \sqrt[2]{7}}$, which is an odd root of a negative real. Let z be the number (9.1), choosing the radicals according to those conventions.

I claim that z is radical, or equivalently that z belongs to every subfield K of $\mathbb{C}$ satisfying (9.2).

First, $\mathbb{Q} \subseteq K$ since $\mathbb{Q}$ is the prime subfield of $\mathbb{C}$. So $2/3 \in K$, and so $\sqrt[5]{2/3} \in K$ by (9.2). Also, $6 \in K$ and K is a field, so $6 + \sqrt[5]{2/3} \in K$. But then by (9.2) again, the denominator of (9.1) is in K. A similar argument shows that the numerator is in K. Hence $z \in K$.

Definition 9.1.5 A nonzero polynomial over $\mathbb{Q}$ is **solvable by radicals** if all of its complex roots are radical.

The simplest nontrivial example of a polynomial solvable by radicals is something of the form $t^n - a$, where $a \in \mathbb{Q}$. The theorem we're heading for is that *any* polynomial solvable by radicals has solvable Galois group, and if that's true then the group $\text{Gal}_{\mathbb{Q}}(t^n - a)$ must be solvable. Let's consider that group now. The results we prove about it will form part of the proof of the big theorem.

We begin with the case $a = 1$.

Lemma 9.1.6 *For all $n \geq 1$, the group $\text{Gal}_{\mathbb{Q}}(t^n - 1)$ is abelian.*

Proof Write $\omega = e^{2\pi i/n}$. The complex roots of $t^n - 1$ are $1, \omega, \ldots, \omega^{n-1}$, so $\text{SF}_{\mathbb{Q}}(t^n - 1) = \mathbb{Q}(\omega)$.

Let $\varphi, \theta \in \text{Gal}_{\mathbb{Q}}(t^n - 1)$. Since φ permutes the roots of $t^n - 1$, we have $\varphi(\omega) = \omega^i$ for some $i \in \mathbb{Z}$. Similarly, $\theta(\omega) = \omega^j$ for some $j \in \mathbb{Z}$. Hence

$$(\varphi \circ \theta)(\omega) = \varphi(\omega^j) = \varphi(\omega)^j = \omega^{ij},$$

131

and similarly $(\theta \circ \varphi)(\omega) = \omega^{ij}$. So $(\varphi \circ \theta)(\omega) = (\theta \circ \varphi)(\omega)$. Since $\mathrm{SF}_{\mathbb{Q}}(t^n - 1) = \mathbb{Q}(\omega)$, it follows that $\theta \circ \varphi = \varphi \circ \theta$. $\qquad\square$

> **Exercise 9.1.7** In the last sentence of that proof, how exactly does it 'follow'?

Much more can be said about the Galois group of $t^n - 1$, and you'll see a bit more at Workshop 5. But this is all we need for our purposes.

Now that we've considered $t^n - 1$, let's do $t^n - a$ for an arbitrary a.

Lemma 9.1.8 *Let K be a field and $n \geq 1$. Suppose that $t^n - 1$ splits in K. Then $\mathrm{Gal}_K(t^n - a)$ is abelian for all $a \in K$.*

The hypothesis that $t^n - 1$ splits in K might seem so restrictive as to make this lemma useless. For instance, it doesn't hold in $\mathbb{Q}$ or even $\mathbb{R}$ (for $n > 2$). Nevertheless, this turns out to be the key lemma in the whole story of solvability by radicals.

Proof If $a = 0$ then $\mathrm{Gal}_K(t^n - a)$ is trivial; suppose otherwise.

Choose a root ξ of $t^n - a$ in $\mathrm{SF}_K(t^n - a)$. For any other root ν, we have $(\nu/\xi)^n = a/a = 1$ (valid since $a \neq 0$), and $t^n - 1$ splits in K, so $\nu/\xi \in K$.

It follows that $\mathrm{SF}_K(t^n - a) = K(\xi)$. Moreover, given $\varphi, \theta \in \mathrm{Gal}_K(t^n - a)$, we have $\varphi(\xi)/\xi \in K$ (since $\varphi(\xi)$ is a root of $t^n - a$), so

$$(\theta \circ \varphi)(\xi) = \theta\left(\frac{\varphi(\xi)}{\xi} \cdot \xi\right) = \frac{\varphi(\xi)}{\xi} \cdot \theta(\xi) = \frac{\varphi(\xi)\theta(\xi)}{\xi}.$$

Similarly, $(\varphi \circ \theta)(\xi) = \varphi(\xi)\theta(\xi)/\xi$, so $(\theta \circ \varphi)(\xi) = (\varphi \circ \theta)(\xi)$. Since $\mathrm{SF}_K(t^n - a) = K(\xi)$, it follows that $\varphi \circ \theta = \theta \circ \varphi$. $\qquad\square$

> **Warning 9.1.9** For $a \in \mathbb{Q}$, the Galois group of $t^n - a$ over $\mathbb{Q}$ is *not* usually abelian. For instance, we saw in Example 7.1.14 that $\mathrm{Gal}_{\mathbb{Q}}(t^3 - 2)$ is the nonabelian group S_3.

> **Exercise 9.1.10** What does the proof of Lemma 9.1.8 tell you about the eigenvectors and eigenvalues of the elements of $\mathrm{Gal}_K(t^n - a)$?

> **Exercise 9.1.11** Use Lemmas 9.1.6 and 9.1.8 to show that $\mathrm{Gal}_{\mathbb{Q}}(t^n - a)$ is solvable for all $a \in \mathbb{Q}$.
>
> This is harder than most of these exercises, but I recommend it as a way of getting into the right frame of mind for the theory that's coming in Section 9.2.

9.2   Solvable polynomials have solvable groups

Here we'll prove that every polynomial over $\mathbb{Q}$ that is solvable by radicals has solvable Galois group.

You know by now that in Galois theory, we tend not to jump straight from polynomials to groups. We go via the intermediate stage of field extensions, as in the diagram

$$\text{polynomial} \longmapsto \text{field extension} \longmapsto \text{group}$$

that I first drew after the definition of $\text{Gal}_K(f)$ (p. 93). That is, we understand polynomials through their splitting field extensions.

So it shouldn't be a surprise that we do the same here, defining a notion of 'solvable extension' and showing (roughly speaking) that

$$\text{solvable polynomial} \longmapsto \text{solvable extension} \longmapsto \text{solvable group}.$$

Solvable polynomials have solvable groups: a map

In other words, we'll define 'solvable extension' in such a way that (i) if $f \in \mathbb{Q}[t]$ is a polynomial solvable by radicals then $\text{SF}_\mathbb{Q}(f) : \mathbb{Q}$ is a solvable extension, and (ii) if $M : K$ is a solvable extension then $\text{Gal}(M : K)$ is a solvable group. Hence if f is solvable by radicals then $\text{Gal}_\mathbb{Q}(f)$ is solvable—the result we're aiming for.

Definition 9.2.1 Let $M : K$ be a finite normal separable extension. Then $M : K$ is **solvable** (or M is **solvable over** K) if there exist $r \geq 0$ and intermediate fields

$$K = L_0 \subseteq L_1 \subseteq \cdots \subseteq L_r = M$$

such that $L_i : L_{i-1}$ is normal and $\text{Gal}(L_i : L_{i-1})$ is abelian for each $i \in \{1, \ldots, r\}$.

Exercise 9.2.2 Let $N : M : K$ be extensions, with $N : M$, $M : K$ and $N : K$ all finite, normal and separable. Show that if $N : M$ and $M : K$ are solvable then so is $N : K$.

We will focus on subfields of $\mathbb{C}$, where separability is automatic (Example 7.2.14(i)).

Example 9.2.3 Let $a \in \mathbb{Q}$ and $n \geq 1$. Then $\mathrm{SF}_\mathbb{Q}(t^n - a) : \mathbb{Q}$ is a finite normal separable extension, being a splitting field extension over $\mathbb{Q}$. I claim that it is solvable.

Proof: if $a = 0$ then $\mathrm{SF}_\mathbb{Q}(t^n - a) = \mathbb{Q}$, and $\mathbb{Q} : \mathbb{Q}$ is solvable (taking $r = 0$ and $L_0 = \mathbb{Q}$ in Definition 9.2.1). Now assume that $a \neq 0$. Choose a complex root ξ of $t^n - a$ and write $\omega = e^{2\pi i/n}$. Then the complex roots of $t^n - a$ are

$$\xi, \omega\xi, \ldots, \omega^{n-1}\xi.$$

So $\mathrm{SF}_\mathbb{Q}(t^n - a)$ contains $(\omega^i\xi)/\xi = \omega^i$ for all i, and so $t^n - 1$ splits in $\mathrm{SF}_\mathbb{Q}(t^n - a)$. Hence

$$\mathbb{Q} \subseteq \mathrm{SF}_\mathbb{Q}(t^n - 1) \subseteq \mathrm{SF}_\mathbb{Q}(t^n - a).$$

Now $\mathrm{SF}_\mathbb{Q}(t^n - 1) : \mathbb{Q}$ is normal (being a splitting field extension) and has abelian Galois group by Lemma 9.1.6. Also $\mathrm{SF}_\mathbb{Q}(t^n - a) : \mathrm{SF}_\mathbb{Q}(t^n - 1)$ is normal (being the splitting field extension of $t^n - a$ over $\mathrm{SF}_\mathbb{Q}(t^n - 1)$, by Lemma 6.2.14(ii)), and has abelian Galois group by Lemma 9.1.8. So $\mathrm{SF}_\mathbb{Q}(t^n - a) : \mathbb{Q}$ is a solvable extension, as claimed.

The definition of solvable extension bears a striking resemblance to the definition of solvable group. Indeed:

Lemma 9.2.4 *Let $M : K$ be a finite normal separable extension. Then*

$$M : K \text{ is solvable} \iff \mathrm{Gal}(M : K) \text{ is solvable.}$$

Proof We will only need the $\Rightarrow$ direction, and that is all I prove here. For the converse, see Workshop 5.

Suppose that $M : K$ is solvable. Then there are intermediate fields

$$K = L_0 \subseteq L_1 \subseteq \cdots \subseteq L_r = M$$

such that each extension $L_i : L_{i-1}$ is normal with abelian Galois group. For each $i \in \{1, \ldots, r\}$, the extension $M : L_{i-1}$ is finite, normal and separable (by Corollary 7.1.6 and Lemma 7.2.16), so we can apply the fundamental theorem of Galois theory to it. Since $L_i : L_{i-1}$ is a normal extension, $\mathrm{Gal}(M : L_i)$ is a normal subgroup of $\mathrm{Gal}(M : L_{i-1})$ and

$$\frac{\mathrm{Gal}(M : L_{i-1})}{\mathrm{Gal}(M : L_i)} \cong \mathrm{Gal}(L_i : L_{i-1}).$$

By hypothesis, the right-hand side is abelian, so the left-hand side is too. So the sequence of subgroups

$$\mathrm{Gal}(M : K) = \mathrm{Gal}(M : L_0) \supseteq \mathrm{Gal}(M : L_1) \supseteq \cdots \supseteq \mathrm{Gal}(M : L_r) = 1$$

exhibits $\mathrm{Gal}(M : K)$ as a solvable group. $\square$

According to the story I'm telling, solvability by radicals of a polynomial should correspond to solvability of its splitting field extension. Thus, the subfields of $\mathbb{C}$ that are solvable over $\mathbb{Q}$ should be exactly the splitting fields $\mathrm{SF}_{\mathbb{Q}}(f)$ of polynomials f that are solvable by radicals. (This is indeed true, though we won't entirely prove it.) Now if $f, g \in \mathbb{Q}[t]$ are both solvable by radicals then so is fg, and $\mathrm{SF}_{\mathbb{Q}}(fg)$ is a solvable extension of $\mathbb{Q}$ containing both $\mathrm{SF}_{\mathbb{Q}}(f)$ and $\mathrm{SF}_{\mathbb{Q}}(g)$. So it should be the case that for any two subfields of $\mathbb{C}$ solvable over $\mathbb{Q}$, there is some larger subfield, also solvable over $\mathbb{Q}$, containing both.

The following pair of lemmas proves this. They use the notion of compositum (Definition 5.3.3).

Lemma 9.2.6 *Let $M : K$ be a field extension and let L and L' be intermediate fields.*
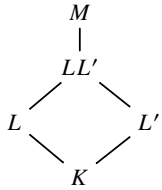
i. If $L : K$ and $L' : K$ are finite and normal, then so is $LL' : K$.

ii. If $L : K$ is finite and normal, then so is $LL' : L'$.

iii. If $L : K$ is finite and normal with abelian Galois group, then so is $LL' : L'$.

Proof For (i), we have $L = \mathrm{SF}_K(f)$ and $L' = \mathrm{SF}_K(f')$ for some $f, f' \in K[t]$. Now LL' is the subfield of M generated by $L \cup L'$, or equivalently by the roots of f and f'. So $LL' = \mathrm{SF}_K(ff')$, which is finite and normal over K.

For (ii), we have $L = \mathrm{SF}_K(f)$ for some $f \in K[t]$. Then $LL' = \mathrm{SF}_{L'}(f)$ by Lemma 6.2.14(i) (with $S = L$ and $Y = L'$), so LL' is finite and normal over L'. Now $\mathrm{Gal}(LL' : L') = \mathrm{Gal}_{L'}(f)$, which by Corollary 6.3.12 is isomorphic to a subgroup of $\mathrm{Gal}_K(f) = \mathrm{Gal}(L : K)$. So if $\mathrm{Gal}(L : K)$ is abelian then so is $\mathrm{Gal}(LL' : L')$, proving (iii). $\qquad\square$

Lemma 9.2.7 *Let L and M be subfields of $\mathbb{C}$ such that the extensions $L : \mathbb{Q}$ and $M : \mathbb{Q}$ are finite, normal and solvable. Then there is some subfield N of $\mathbb{C}$ such that $N : \mathbb{Q}$ is finite, normal and solvable and $L, M \subseteq N$.*

Both the statement and the proof of this lemma should remind you of Lemma 5.3.8 on ruler and compass constructions.

Proof Take subfields

$$\mathbb{Q} = L_0 \subseteq \cdots \subseteq L_r = L, \qquad \mathbb{Q} = M_0 \subseteq \cdots \subseteq M_s = M$$

such that $L_i : L_{i-1}$ is normal with abelian Galois group for each i, and similarly for $M_j : M_{j-1}$. There is a chain of subfields

$$\mathbb{Q} = L_0 \subseteq \cdots \subseteq L_r = L = LM_0 \subseteq \cdots \subseteq LM_s = LM \qquad (9.3)$$

of $\mathbb{C}$. Put $N = LM$. Certainly $L, M \subseteq N$. We show that $N : \mathbb{Q}$ is finite, normal and solvable.

That $N : \mathbb{Q}$ is finite and normal follows from Lemma 9.2.6(i).

To see that $N : \mathbb{Q}$ is solvable, we show that each successive extension in (9.3) is normal with abelian Galois group. For those to the left of L, this is immediate. For those to the right, let $j \in \{1, \ldots, s\}$. Since $M_j : M_{j-1}$ is finite and normal with abelian Galois group, so is $LM_j : LM_{j-1}$ by Lemma 9.2.6(iii). $\qquad \square$

The set of radical numbers is a subfield of $\mathbb{C}$ closed under taking nth roots. So if the story I'm telling is right, the set of complex numbers that can be reached from $\mathbb{Q}$ by solvable extensions should also be a subfield of $\mathbb{C}$ closed under taking nth roots. That's an informal description of our next two results. Write

$$\mathbb{Q}^{\mathrm{sol}} = \{\alpha \in \mathbb{C} : \alpha \in L \text{ for some subfield } L \subseteq \mathbb{C} \text{ that is finite, normal and solvable}$$
$$\text{over } \mathbb{Q}\}.$$

Lemma 9.2.8 $\mathbb{Q}^{\mathrm{sol}}$ *is a subfield of* $\mathbb{C}$.

Proof This is similar to the proof that the algebraic numbers form a subfield of $\mathbb{C}$ (Proposition 5.2.7). Let $\alpha, \beta \in \mathbb{Q}^{\mathrm{sol}}$. Then $\alpha \in L$ and $\beta \in M$ for some L, M that are finite, normal and solvable over $\mathbb{Q}$. By Lemma 9.2.7, $\alpha, \beta \in N$ for some N that is finite, normal and solvable over $\mathbb{Q}$. Then $\alpha + \beta \in N$, so $\alpha + \beta \in \mathbb{Q}^{\mathrm{sol}}$, and similarly $\alpha \cdot \beta \in \mathbb{Q}^{\mathrm{sol}}$. This shows that $\mathbb{Q}^{\mathrm{sol}}$ is closed under addition and multiplication. The other parts of the proof (negatives, reciprocals, 0 and 1) are straightforward. $\qquad \square$

Lemma 9.2.9 *Let* $\alpha \in \mathbb{C}$ *and* $n \geq 1$. *If* $\alpha^n \in \mathbb{Q}^{\mathrm{sol}}$ *then* $\alpha \in \mathbb{Q}^{\mathrm{sol}}$.

The proof (below) is slightly subtle. Here's why.

Let L be a subfield of $\mathbb{C}$ that's finite, normal and solvable over $\mathbb{Q}$, and take $\alpha \in \mathbb{C}$ and $n \geq 1$ such that $\alpha^n \in L$. To find some larger M that contains α itself and is also solvable over $\mathbb{Q}$, we could try putting $M = \mathrm{SF}_L(t^n - \alpha^n)$. But the problem is that $M : \mathbb{Q}$ is not in general normal. And normality is part of the definition of $\mathbb{Q}^{\mathrm{sol}}$, ultimately because it's essential if we want to use the fundamental theorem of Galois theory. The basic problem is this:

Warning 9.2.10 A normal extension of a normal extension is not in general normal, just as a normal subgroup of a normal subgroup is not in general normal.

An example should clarify.

Example 9.2.11 Put $\alpha = \sqrt[4]{2}$ and $n = 2$. We have $\alpha^2 = \sqrt{2} \in \mathbb{Q}(\sqrt{2})$, and $\mathbb{Q}(\sqrt{2}) : \mathbb{Q}$ is finite, normal and solvable (since its Galois group is the abelian group C_2), so $\alpha^2 \in \mathbb{Q}^{\text{sol}}$. Hence, according to Lemma 9.2.9, $\alpha = \sqrt[4]{2}$ should be contained in some finite normal solvable extension M of $\mathbb{Q}$.

How can we find such an M? We *can't* take $M = \text{SF}_{\mathbb{Q}(\sqrt{2})}(t^2 - \sqrt{2})$, since this is $\mathbb{Q}(\sqrt[4]{2})$, which is not normal over $\mathbb{Q}$ (for the same reason that $\mathbb{Q}(\sqrt[3]{2})$ isn't).

To find a bigger M, still finite and solvable over $\mathbb{Q}$ but also normal, we have to adjoin a square root not just of $\sqrt{2}$ *but also of its conjugate*, $-\sqrt{2}$. This is the crucial point: the whole idea of normality is that conjugates are treated equally. (Normal behaviour means that anything you do for one element, you do for all its conjugates.) The result is $\mathbb{Q}(\sqrt[4]{2}, i) = \text{SF}_{\mathbb{Q}}(t^4 - 2)$, which is indeed a finite, solvable and normal extension of $\mathbb{Q}$ containing $\sqrt[4]{2}$.

Proof of Lemma 9.2.9 Write $a = \alpha^n \in \mathbb{Q}^{\text{sol}}$. Choose a subfield K of $\mathbb{C}$ such that $a \in K$ and $K : \mathbb{Q}$ is finite, normal and solvable.

Step 1: enlarge K to a field in which $t^n - 1$ splits. Put $L = \text{SF}_K(t^n - 1) \subseteq \mathbb{C}$.

Since $K : \mathbb{Q}$ is finite and normal, $K = \text{SF}_{\mathbb{Q}}(f)$ for some nonzero $f \in \mathbb{Q}[t]$, and then $L = \text{SF}_{\mathbb{Q}}\big((t^n - 1)f(t)\big)$. Hence $L : \mathbb{Q}$ is finite and normal. The Galois group of $L : K$ is $\text{Gal}_K(t^n - 1)$, which is isomorphic to a subgroup of $\text{Gal}_{\mathbb{Q}}(t^n - 1)$ (by Corollary 6.3.12), which is abelian (by Lemma 9.1.6). Hence $L : K$ is a normal extension with abelian Galois group. Also, $K : \mathbb{Q}$ is solvable. It follows from the definition of solvable extension that $L : \mathbb{Q}$ is solvable.

In summary, L is a subfield of $\mathbb{C}$ such that $a \in L$ and $L : \mathbb{Q}$ is finite, normal and solvable, *and*, moreover, $t^n - 1$ splits in L. We now forget about K.

Step 2: adjoin the nth roots of the conjugates of a. Write $m \in \mathbb{Q}[t]$ for the minimal polynomial of a over $\mathbb{Q}$, and put $M = \text{SF}_L(m(t^n)) \subseteq \mathbb{C}$. Then $\alpha \in M$, as $m(\alpha^n) = m(a) = 0$. We show that $M : \mathbb{Q}$ is finite, normal and solvable.

Since $L : \mathbb{Q}$ is finite and normal, $L = \text{SF}_{\mathbb{Q}}(g)$ for some nonzero $g \in \mathbb{Q}[t]$. Then $M = \text{SF}_{\mathbb{Q}}(g(t)m(t^n))$, so $M : \mathbb{Q}$ is finite and normal. Moreover, $M : L$ is finite and normal, being a splitting field extension.

Now to show that $M : \mathbb{Q}$ is solvable, it is enough to show that $M : L$ is solvable, by Exercise 9.2.2. Since $L : \mathbb{Q}$ is normal and $m \in \mathbb{Q}[t]$ is the minimal polynomial

of $a \in L$, it follows by definition of normality that m splits in L, say

$$m(t) = \prod_{i=1}^{r} (t - a_i)$$

($a_i \in L$). Define subfields $L_0 \subseteq \cdots \subseteq L_r$ of $\mathbb{C}$ by

$$
\begin{aligned}
L_0 &= L \\
L_1 &= \mathrm{SF}_{L_0}(t^n - a_1) \\
L_2 &= \mathrm{SF}_{L_1}(t^n - a_2) \\
&\ \ \vdots \\
L_r &= \mathrm{SF}_{L_{r-1}}(t^n - a_r).
\end{aligned}
$$

Then

$$L_i = L\big(\{\beta \in M : \beta^n \in \{a_1, \ldots, a_i\}\}\big).$$

In particular, $L_r = M$. For each $i \in \{1, \ldots, s\}$, the extension $L_i : L_{i-1}$ is finite and normal (being a splitting field extension), and its Galois group is abelian (by Lemma 9.1.8 and the fact that $t^n - 1$ splits in $L \subseteq L_{i-1}$). So $M : L$ is solvable. $\square$

Now we can relate the set $\mathbb{Q}^{\mathrm{rad}}$ of radical numbers, defined in terms of basic arithmetic operations, to the set $\mathbb{Q}^{\mathrm{sol}}$, defined in terms of field extensions.

Proposition 9.2.12 $\mathbb{Q}^{\mathrm{rad}} \subseteq \mathbb{Q}^{\mathrm{sol}}$. *That is, every radical number is contained in some subfield of $\mathbb{C}$ that is a finite, normal, solvable extension of $\mathbb{Q}$.*

In fact, $\mathbb{Q}^{\mathrm{rad}}$ and $\mathbb{Q}^{\mathrm{sol}}$ are equal, but we won't prove this.

Proof By Lemmas 9.2.8 and 9.2.9, $\mathbb{Q}^{\mathrm{sol}}$ is a subfield of $\mathbb{C}$ such that $\alpha^n \in \mathbb{Q}^{\mathrm{sol}} \implies \alpha \in \mathbb{Q}^{\mathrm{sol}}$. The result follows from the definition of $\mathbb{Q}^{\mathrm{rad}}$. $\square$

This brings us to the main result of this chapter. Notice that it doesn't mention field extensions: it goes straight from polynomials to groups.

Theorem 9.2.13 *Let $0 \neq f \in \mathbb{Q}[t]$. If the polynomial f is solvable by radicals then the group $\mathrm{Gal}_{\mathbb{Q}}(f)$ is solvable.*

Proof Suppose that f is solvable by radicals. Write $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ for its roots. For each i, we have $\alpha_i \in \mathbb{Q}^{\mathrm{rad}}$ (by definition of solvability by radicals), hence $\alpha_i \in \mathbb{Q}^{\mathrm{sol}}$ (by Proposition 9.2.12). So each of $\alpha_1, \ldots, \alpha_n$ is contained in some subfield of $\mathbb{C}$ that is finite, normal and solvable over $\mathbb{Q}$. By Lemma 9.2.7, there is some subfield M of $\mathbb{C}$ that is finite, normal and solvable over $\mathbb{Q}$ and contains all of $\alpha_1, \ldots, \alpha_n$. Then $\mathbb{Q}(\alpha_1, \ldots, \alpha_n) \subseteq M$; that is, $\mathrm{SF}_{\mathbb{Q}}(f) \subseteq M$.

138

By Lemma 9.2.4, $\mathrm{Gal}(M : \mathbb{Q})$ is solvable. Now $\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q}$ is normal, so by the fundamental theorem of Galois theory, $\mathrm{Gal}(\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q})$ is a quotient of $\mathrm{Gal}(M : \mathbb{Q})$. But $\mathrm{Gal}(\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q}) = \mathrm{Gal}_{\mathbb{Q}}(f)$, and a quotient of a solvable group is solvable, so $\mathrm{Gal}_{\mathbb{Q}}(f)$ is solvable. $\qquad\square$

Examples 9.2.14 i. For $a \in \mathbb{Q}$ and $n \geq 1$, the polynomial $t^n - a$ is solvable by radicals, so the group $\mathrm{Gal}_{\mathbb{Q}}(t^n - a)$ is solvable. You may already have proved this in Exercise 9.1.11. It also follows from Example 9.2.3 and Lemma 9.2.4.

ii. Let $a_1, \ldots, a_k \in \mathbb{Q}$ and $n_1, \ldots, n_k \geq 1$. Each of the polynomials $t^{n_i} - a_i$ is solvable by radicals, so their product is too. Hence $\mathrm{Gal}_{\mathbb{Q}}\big(\prod_i (t^{n_i} - a_i)\big)$ is a solvable group.

Theorem 9.2.13 is most sensational in its contrapositive form: if $\mathrm{Gal}_{\mathbb{Q}}(f)$ is *not* solvable then f is *not* solvable by radicals. That's the subject of the next section.

> **Digression 9.2.15** The converse of Theorem 9.2.13 is also true: if $\mathrm{Gal}_{\mathbb{Q}}(f)$ is solvable then f is solvable by radicals. You can even unwind the proof to obtain an explicit formula for the solving the quartic by radicals (Stewart, Chapter 18).
>
> For this, we have to deduce properties of a field extension from assumptions about its Galois group. A solvable group is built up from abelian groups, and every finite abelian group is a direct sum of cyclic groups. The key step in proving the converse of Theorem 9.2.13 has come to be known as 'Hilbert's Theorem 90' (Stewart's Theorem 18.18), which gives information about field extensions whose Galois groups are cyclic.

> **Digression 9.2.16** The proof of Theorem 9.2.13 might not have ended quite how you expected. Given my explanations earlier in the chapter, you might justifiably have imagined we were going to show that when the polynomial f is solvable by radicals, the extension $\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q}$ is solvable. That's not what we did. We showed that $\mathrm{SF}_{\mathbb{Q}}(f)$ is contained in some larger subfield M such that $M : \mathbb{Q}$ is solvable, then used that to prove the solvability of the group $\mathrm{Gal}_{\mathbb{Q}}(f)$.
>
> But all is right with the world: $\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q}$ *is* a solvable extension. Indeed, its Galois group $\mathrm{Gal}_{\mathbb{Q}}(f)$ is solvable, so Lemma 9.2.4 implies that $\mathrm{SF}_{\mathbb{Q}}(f) : \mathbb{Q}$ is solvable too.

9.3   An unsolvable polynomial

Here we give a specific example of a polynomial over $\mathbb{Q}$ that is not solvable by radicals. By Theorem 9.2.13, our task is to construct a polynomial whose Galois

group is not solvable. The smallest non-solvable group is A_5 (of order 60). Our polynomial has Galois group S_5 (of order 120), which is also non-solvable.

Finding Galois groups is hard, and we will use a whole box of tools and tricks, from Cauchy's theorem on groups to Rolle's theorem on differentiable functions.

First we prove a useful general fact on the order of Galois groups.

Lemma 9.3.1 *Let f be an irreducible polynomial over a field K, with $\mathrm{SF}_K(f) : K$ separable. Then $\deg(f)$ divides $|\mathrm{Gal}_K(f)|$.*

Proof Let α be a root of f in $\mathrm{SF}_K(f)$. By irreducibility, $\deg(f) = [K(\alpha) : K]$, which divides $[\mathrm{SF}_K(f) : K]$ by the tower law, which is equal to $|\mathrm{Gal}_K(f)|$ by Theorem 7.2.18 (using separability). $\qquad\square$

Next, we need some results about the symmetric group S_n. I assume you know that S_n is generated by the 'adjacent transpositions' $(12), (23), \ldots, (n-1\, n)$. This may have been proved in Fundamentals of Pure Mathematics, and as the Group Theory notes say (p. 58):

> This is intuitively clear: suppose you have n people lined up and you want them to switch into a different order. To put them in the order you want them, it's clearly enough to have people move up and down the line; and each time a person moves one place, they switch places with the person next to them.

Here's a different way of generating S_n.

Lemma 9.3.2 *For $n \geq 2$, the symmetric group S_n is generated by (12) and $(12 \ldots n)$.*

Proof We have
$$(12 \ldots n)(12)(12 \ldots n)^{-1} = (23),$$
either by direct calculation or the general fact that $\sigma(a_1 \ldots a_k)\sigma^{-1} = (\sigma(a_1) \ldots \sigma(a_k))$ for any $\sigma \in S_n$ and cycle $(a_1 \ldots a_k)$. So any subgroup H of S_n containing (12) and $(12 \ldots n)$ also contains (23). By the same argument, H also contains $(34), \ldots, (n-1\, n)$. But the adjacent transpositions generate S_n, so $H = S_n$. $\qquad\square$

Lemma 9.3.3 *Let p be a prime number, and let $f \in \mathbb{Q}[t]$ be an irreducible polynomial of degree p with exactly $p - 2$ real roots. Then $\mathrm{Gal}_{\mathbb{Q}}(f) \cong S_p$.*

Proof Since $\operatorname{char}\mathbb{Q} = 0$ and f is irreducible, f is separable and therefore has p distinct roots in $\mathbb{C}$. By Proposition 6.3.10, the action of $\mathrm{Gal}_{\mathbb{Q}}(f)$ on the roots of f in $\mathbb{C}$ defines an isomorphism between $\mathrm{Gal}_{\mathbb{Q}}(f)$ and a subgroup H of S_p. Since
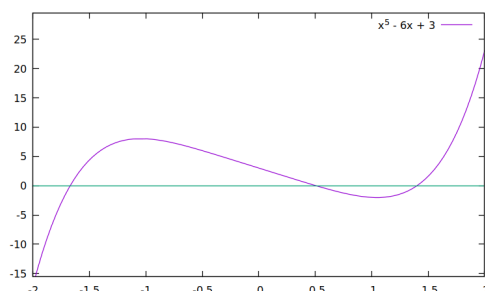
Figure 9.1: The function $x \mapsto x^5 - 6x + 3$.

f is irreducible, p divides $|\operatorname{Gal}_{\mathbb{Q}}(f)| = |H|$ (by Lemma 9.3.1). So by Cauchy's theorem, H has an element σ of order p. Then σ is a p-cycle, since these are the only elements of S_p of order p.

The complex conjugate of any root of f is also a root of f, so complex conjugation restricts to an automorphism of $\operatorname{SF}_{\mathbb{Q}}(f)$ over $\mathbb{Q}$. Exactly two of the roots of f are non-real; complex conjugation transposes them and fixes the rest. So H contains a transposition τ.

Without loss of generality, $\tau = (12)$. Since σ is a p-cycle, $\sigma^r(1) = 2$ for some $r \in \{1, \ldots, p-1\}$. Since p is prime, σ^r also has order p, so it is a p-cycle. Now without loss of generality, $\sigma^r = (123 \ldots p)$. So $(12), (12 \ldots p) \in H$, forcing $H = S_p$ by Lemma 9.3.2. Hence $\operatorname{Gal}_{\mathbb{Q}}(f) \cong S_p$. $\qquad\square$

Exercise 9.3.4 Explain why, in the last paragraph, σ^r has order p.

Theorem 9.3.5 *Not every polynomial over $\mathbb{Q}$ of degree 5 is solvable by radicals.*

Proof We show that $f(t) = t^5 - 6t + 3$ satisfies the conditions of Lemma 9.3.3. Then $\operatorname{Gal}_{\mathbb{Q}}(f)$ is S_5, which is not solvable, so by Theorem 9.2.13, f is not solvable by radicals.

Evidently $\deg(f)$ is the prime number 5, and f is irreducible by Eisenstein's criterion with prime 3. It remains to prove that f has exactly 3 real roots. This is where we use some analysis, considering f as a function $\mathbb{R} \to \mathbb{R}$ (Figure 9.1).

We have

$$\lim_{x \to -\infty} f(x) = -\infty, \qquad f(0) > 0, \qquad f(1) < 0, \qquad \lim_{x \to \infty} f(x) = \infty,$$

and f is continuous on $\mathbb{R}$, so by the intermediate value theorem, f has at least 3 real roots. On the other hand, $f'(x) = 5x^4 - 6$ has only 2 real roots ($\pm\sqrt[4]{6/5}$), so

141

by Rolle's theorem, f has at most 3 real roots. Hence f has exactly 3 real roots, as required. □

Exercise 9.3.6 Prove that for every $n \geq 5$, there is some polynomial of degree n that is not solvable by radicals.

Example 9.3.7 There are also quintics with Galois group A_5. These are not solvable by radicals, since A_5 is not a solvable group. One example, although we won't prove it, is $t^5 + 20t + 16$.

Digression 9.3.8 We now know that some polynomials f over $\mathbb{Q}$ are not solvable by radicals, which means that not *all* their complex roots are radical.

Could it be that some of the roots are radical and others are not? Yes: simply take a polynomial g that is not solvable by radicals and put $f(t) = tg(t)$. Then the roots of f are 0 (which is radical) together with the roots of g (which are not all radical).

But what if f is irreducible? In that case, either all the roots of f are radical or none of them are. This follows from the fact that the extension $\mathbb{Q}^{\mathrm{rad}} : \mathbb{Q}$ is normal, which we will not prove.

Digression 9.3.9 There are many similarities between the theory of constructibility of points by ruler and compass and the theory of solvability of polynomials by radicals. In both cases, the challenge is to construct some things (points in the plane or roots of polynomials) using only certain tools (ruler and compass or a machine for taking nth roots). In both cases, there were difficult questions of constructibility that remained open for a very long time, and in both cases, they were solved by field theory.

The solutions have something in common too. For the geometry problem, we used iterated quadratic extensions, and for the polynomial problem, we used solvable extensions, which could reasonably be called iterated abelian extensions. For the geometry problem, we showed that the coordinates of any point constructible by ruler and compass satisfy a certain condition on their degree over $\mathbb{Q}$ (Theorem 5.3.10); for the polynomial problem, we showed that any polynomial solvable by radicals satisfies a certain condition on its Galois group over $\mathbb{Q}$. There are other similarities: compare Lemmas 5.3.8 and 9.2.7, for example, and maybe you can find more similarities still.

We have now used the fundamental theorem of Galois theory to solve a major problem about $\mathbb{Q}$. What else can we do with it?

The fundamental theorem is about *separable* extensions. Our two main sources of separable extensions are:

- fields of characteristic 0 such as $\mathbb{Q}$ (Example 7.2.14(i)), which we've explored extensively already;

- finite fields (Example 7.2.14(ii)), which we've barely touched.

In the next and final chapter, we'll use the fundamental theorem and other results we've proved to explore the world of finite fields. In contrast to the intricately complicated world of finite groups, finite fields are almost shockingly simple.

Chapter 10

Finite fields

This chapter is dessert. Through this semester, we've developed a lot of sophisticated theory for general fields. All of it works for finite fields, but becomes much simpler there. It's a miniature world in which life is sweet. For example:

- If we want to apply the fundamental theorem of Galois theory to a field extension $M : K$, we first have to ask whether it is finite, and whether it is normal, and whether it is separable. When M and K are finite fields, all three conditions are automatic.

- There are many fields of different kinds, and to classify them all would be a near-impossible task. But for finite fields, the classification is very simple. We know exactly what finite fields there are.

- The Galois correspondence for arbitrary field extensions can also be complicated. But again, it's simple when the fields are finite. Their Galois groups are very easy (they're all cyclic), we know what their subgroups are, and it's easy to describe all the subfields of any given finite field.

So although the world of finite fields is not trivial, there's a lot about it that's surprisingly straightforward.

We've already encountered two aspects of finite fields that may seem counter-intuitive. First, they always have positive characteristic, which means they satisfy some equation like $1 + \cdots + 1 = 0$ (Lemma 2.3.17). Second, any element of a finite field of characteristic p has precisely one pth root (Corollary 2.3.22(ii)), making finite fields quite unlike $\mathbb{C}$, $\mathbb{R}$ or $\mathbb{Q}$. But the behaviour of pth roots and pth powers is fundamental to all of finite fields' nice properties.

10.1 Classification of finite fields

If you try to write down a formula for the number of groups or rings with a given number of elements, you'll find that it's hard and the results are quite strange. For instance, more than 99% of the first 50 billion groups have order 1024.

But fields turn out to be much, much easier. We'll obtain a complete classification of finite fields in the next two pages.

The **order** of a finite field M is its cardinality, or number of elements, $|M|$.

Warning 10.1.1 *Order* and *degree* mean different things. For instance, if the order of a field is 9, then its degree over its prime subfield $\mathbb{F}_3$ is 2.

Lemma 10.1.2 *Let M be a finite field. Then* char M *is a prime number p, and* $|M| = p^n$ *where $n = [M : \mathbb{F}_p] \geq 1$.*

In particular, the order of a finite field is a prime power.

Proof By Lemmas 2.3.11 and 2.3.17, char M is a prime number p. By Lemma 2.3.16, M has prime subfield $\mathbb{F}_p$. Since M is finite, $1 \leq [M : \mathbb{F}_p] < \infty$; write $n = [M : \mathbb{F}_p]$. As a vector space over $\mathbb{F}_p$, then, M is n-dimensional and so isomorphic to $\mathbb{F}_p^n$. But $|\mathbb{F}_p^n| = |\mathbb{F}_p|^n = p^n$, so $|M| = p^n$. $\qquad\square$

Example 10.1.3 There is no field of order 6, since 6 is not a prime power.

Lemma 10.1.2 prompts two questions: given a prime power p^n, *is* there some field of order p^n? And if so, how many are there?

To answer them, we need to recall the Frobenius automorphism θ of a finite field (Proposition 2.3.20).

Exercise 10.1.4 Work out the values of the Frobenius automorphism on the field $\mathbb{F}_3(\sqrt{2})$, which you first met in Exercise 4.3.18.

The answer to the first of these questions is yes:

Lemma 10.1.5 *Let p be a prime number and $n \geq 1$. Then the splitting field of* $t^{p^n} - t$ *over $\mathbb{F}_p$ has order p^n.*

Proof Put $f(t) = t^{p^n} - t \in \mathbb{F}_p[t]$ and $M = \mathrm{SF}_{\mathbb{F}_p}(f)$. Then $Df = -1$ (since $n \geq 1$), so by (i)$\Rightarrow$(ii) of Lemma 7.2.9, f has no repeated roots in M. Hence M has at least p^n elements.

Write θ for the Frobenius map of M and $\theta^n = \theta \circ \cdots \circ \theta$. Then $\theta^n(\alpha) = \alpha^{p^n}$ for all α, so the set L of roots of f in M is $\mathrm{Fix}\{\theta^n\}$. Since θ is a homomorphism,

Lemma 7.3.1 implies that L is a subfield of M. Hence by definition of splitting field, $L = M$; that is, every element of M is a root of f. And since $\deg(f) = p^n$, it follows that M has at most p^n elements. $\qquad\square$

As for the second question, there is exactly *one* field of each prime power order. To show this, we need a lemma.

Lemma 10.1.6 *Let M be a finite field of order q. Then $\alpha^q = \alpha$ for all $\alpha \in M$.*

The proof uses the same argument as in Example 2.3.21.

Proof The multiplicative group $M^\times = M \setminus \{0\}$ has order $q - 1$, so Lagrange's theorem implies that $\alpha^{q-1} = 1$ for all $\alpha \in M^\times$. Hence $\alpha^q = \alpha$ whenever $0 \neq \alpha \in M$, and clearly the equation holds for $\alpha = 0$ too. $\qquad\square$

> **Exercise 10.1.7** Verify directly that $\beta^4 = \beta$ for all β in the 4-element field $\mathbb{F}_2(\alpha)$ of Example 5.1.8.

Lemma 10.1.8 *Every finite field of order q is a splitting field of $t^q - t$ over $\mathbb{F}_p$.*

Proof Let M be a field of order q. By Lemma 10.1.2, $q = p^n$ for some prime p and $n \geq 1$, and char $M = p$. Hence M has prime subfield $\mathbb{F}_p$. By Lemma 10.1.6, every element of M is a root of $f(t) = t^{p^n} - t$. So f has $|M|$ distinct roots in M; but $|M| = p^n = \deg(f)$, so f splits in M. The set of roots of f in M generates M, since it is *equal* to M. Hence M is a splitting field of f. $\qquad\square$

Together, these results completely classify the finite fields.

> **Theorem 10.1.9 (Classification of finite fields)**   *i. Every finite field has order p^n for some prime p and integer $n \geq 1$.*
>
> *ii. For each prime p and integer $n \geq 1$, there is exactly one field of order p^n, up to isomorphism. It has characteristic p and is a splitting field for $t^{p^n} - t$ over $\mathbb{F}_p$.*

Proof This is immediate from the results above together with the uniqueness of splitting fields (Theorem 6.2.13(ii)). $\qquad\square$

When $q > 1$ is a prime power, we write $\mathbb{F}_q$ for the one and only field of order q.

We now know exactly how many finite fields there are of each order. But generally in algebra, it's important to think not just about the *objects* (such as vector spaces, groups, modules, rings, fields, ...), but also the *maps* (homomorphisms) between objects. So now that we've counted the finite fields, it's natural to try to count the homomorphisms between finite fields. Field homomorphisms are injective, so this boils down to counting subfields and automorphisms. Galois theory is very well equipped to do that! We'll come to this in the final section. But first, we look at another way in which finite fields are very simple.

10.2 Multiplicative structure

The multiplicative group $K^\times$ of a finite field K is as easy as can be:

Proposition 10.2.1 *For an arbitrary field K, every finite subgroup of $K^\times$ is cyclic. In particular, if K is finite then $K^\times$ is cyclic.*

The multiplicative group of a finite field is cyclic

Proof This was Theorem 5.1.13 and Corollary 5.1.14 of Group Theory. □

Example 10.2.2 In examples earlier in the course, we frequently used the nth root of unity $\omega = e^{2\pi i/n} \in \mathbb{C}$, which has the property that every other nth root of unity is a power of ω.

Can we find an analogue of ω in an arbitrary field K? It's not obvious how to generalize the formula $e^{2\pi i/n}$, since the exponential is a concept from complex analysis. But Proposition 10.2.1 solves our problem. For $n \geq 1$, put

$$U_n(K) = \{\alpha \in K : \alpha^n = 1\}.$$

Then $U_n(K)$ is a subgroup of $K^\times$, and is finite since its elements are roots of $t^n - 1$. So by Proposition 10.2.1, $U_n(K)$ is cyclic. Let ω be a generator of $U_n(K)$. Then every nth root of unity in K is a power of ω, which is what we were aiming for.

Note, however, that $U_n(K)$ may have fewer than n elements, or equivalently, the order of ω may be less than n. For instance, if char $K = p$ then $U_p(K)$ is trivial and $\omega = 1$, by Example 2.3.23(ii).

Exercise 10.2.3 Let K be a field and let H be a finite subgroup of $K^\times$ of order n. Prove that $H \subseteq U_n(K)$.

147

Example 10.2.4 The group $\mathbb{F}_p^{\times}$ is cyclic, for any prime p. This means that there is some $\omega \in \{1, \ldots, p - 1\}$ such that $\omega, \omega^2, \ldots$ runs through all elements of $\{1, \ldots, p - 1\}$ when taken mod p. In number theory, such an ω is called a **primitive root** mod p (another usage of the word 'primitive'). For instance, you can check that 3 is a primitive root mod 7, but 2 is not, since $2^3 \equiv 1 \pmod{7}$.

Finding the primitive roots mod p is one aspect of finite fields that is *not* trivial.

Corollary 10.2.5 *Every extension of one finite field over another is simple.*

Proof Let $M : K$ be an extension with M finite. By Proposition 10.2.1, the group $M^{\times}$ is generated by some element $\alpha \in M^{\times}$. Then $M = K(\alpha)$. $\square$

This is yet another pleasant aspect of finite fields.

> **Exercise 10.2.6** In the proof of Corollary 10.2.5, once we know that the group $M^{\times}$ is generated by α, how does it follow that $M = K(\alpha)$?

> **Digression 10.2.7** In Digression 7.2.21, I mentioned the theorem of the primitive element: every finite separable extension $M : K$ is simple. One of the standard proofs involves splitting into two cases, according to whether M is finite or infinite. We've just done the finite case.

Corollary 10.2.8 *For every prime number p and integer $n \geq 1$, there exists an irreducible polynomial over $\mathbb{F}_p$ of degree n.*

Proof The field $\mathbb{F}_{p^n}$ has prime subfield $\mathbb{F}_p$. By Corollary 10.2.5, the extension $\mathbb{F}_{p^n} : \mathbb{F}_p$ is simple, say $\mathbb{F}_{p^n} = \mathbb{F}_p(\alpha)$. The minimal polynomial of α over $\mathbb{F}_p$ is irreducible of degree $[\mathbb{F}_p(\alpha) : \mathbb{F}_p] = [\mathbb{F}_{p^n} : \mathbb{F}_p] = n$. $\square$

This is not obvious. For example, can you find an irreducible polynomial of degree 100 over $\mathbb{F}_{31}$?

10.3 Galois groups for finite fields

Here we work out the Galois correspondence for $\mathbb{F}_{p^n} : \mathbb{F}_p$.

> **Warning 10.3.1** The term 'finite field extension' means an extension $M : K$ that's finite in the sense defined on p. 67: M is finite-*dimensional* as a vector space over K. It doesn't mean that M and K are finite fields. But the safest policy is to avoid this term entirely.

The three hypotheses of the fundamental theorem of Galois theory are always satisfied when both fields in the extension are finite:

Lemma 10.3.2 *Let $M : K$ be a field extension.*

   i. If K is finite then $M : K$ is separable.

   ii. If M is also finite then $M : K$ is finite and normal.

Proof For (i), we show that every irreducible polynomial f over K is separable. Write $p = \operatorname{char} K > 0$, and suppose for a contradiction that f is inseparable. By Corollary 7.2.11,

$$f(t) = b_0 + b_1 t^p + \cdots + b_r t^{rp}$$

for some $b_0, \ldots, b_r \in K$. For each i, there is a (unique) pth root c_i of b_i in K, by Corollary 2.3.22(ii). Then

$$f(t) = c_0^p + c_1^p t^p + \cdots + c_r^p t^{rp}.$$

But by Proposition 2.3.20(i), the function $g \mapsto g^p$ is a homomorphism $K[t] \to K[t]$, so

$$f(t) = (c_0 + c_1 t + \cdots + c_r t^r)^p.$$

This contradicts f being irreducible.

    For (ii), suppose that M is finite. Write $p = \operatorname{char} M > 0$. By Theorem 10.1.9, M is a splitting field over $\mathbb{F}_p$, so by Lemma 6.2.14(ii), it is also a splitting field over K. Hence $M : K$ is finite and normal, by Theorem 7.1.5. $\qquad\square$

    Part (i) fulfils the promise made in Remark 7.2.12 and Example 7.2.14(ii), and the lemma as a whole lets us use the fundamental theorem freely in the world of finite fields. We now work out the Galois correspondence for the extension $\mathbb{F}_{p^n} : \mathbb{F}_p$ of an arbitrary finite field over its prime subfield.

Proposition 10.3.3 *Let p be a prime and $n \geq 1$. Then $\operatorname{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p)$ is cyclic of order n, generated by the Frobenius automorphism of $\mathbb{F}_{p^n}$.*

    By Workshop 4, q. 4, $\operatorname{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p)$ is the group of *all* automorphisms of $\mathbb{F}_{p^n}$.

Proof Write θ for the Frobenius automorphism of $\mathbb{F}_{p^n}$; then $\theta \in \operatorname{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p)$. First we calculate the order of θ. By Lemma 10.1.6, $\alpha^{p^n} = \alpha$ for all $\alpha \in \mathbb{F}_{p^n}$, or equivalently, $\theta^n = \operatorname{id}$. If m is a positive integer such that $\theta^m = \operatorname{id}$ then $\alpha^{p^m} = \alpha$ for all $\alpha \in \mathbb{F}_{p^n}$, so the polynomial $t^{p^m} - t$ has p^n roots in $\mathbb{F}_{p^n}$, so $p^n \leq p^m$, so $n \leq m$. Hence θ has order n.

    On the other hand, $[\mathbb{F}_{p^n} : \mathbb{F}_p] = n$, so by the fundamental theorem of Galois theory, $|\operatorname{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p)| = n$. The result follows. $\qquad\square$

In Fundamentals of Pure Mathematics or Group Theory, you presumably saw that the cyclic group of order n has exactly one subgroup of order k for each divisor k of n. (And by Lagrange's theorem, there are no subgroups of other orders.)

In the case at hand, $\mathrm{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p) = \langle \theta \rangle \cong C_n$, and when $k \mid n$, the unique subgroup of order k is $\langle \theta^{n/k} \rangle$.

Proposition 10.3.6 *Let p be a prime and $n \geq 1$. Then $\mathbb{F}_{p^n}$ has exactly one subfield of order p^m for each divisor m of n, and no others. It is*

$$\left\{ \alpha \in \mathbb{F}_{p^n} : \alpha^{p^m} = \alpha \right\}.$$

Proof The subfields of $\mathbb{F}_{p^n}$ are the intermediate fields of $\mathbb{F}_{p^n} : \mathbb{F}_p$, which by the fundamental theorem of Galois theory are precisely the fixed fields $\mathrm{Fix}(H)$ of subgroups H of $\mathrm{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p)$. Any such H is of the form $\langle \theta^{n/k} \rangle$ with $k \mid n$, and

$$\mathrm{Fix}\langle \theta^{n/k} \rangle = \left\{ \alpha \in \mathbb{F}_{p^n} : \alpha^{p^{n/k}} = \alpha \right\}.$$

The tower law and the fundamental theorem give

$$[\mathrm{Fix}\langle \theta^{n/k} \rangle : \mathbb{F}_p] = \frac{[\mathbb{F}_{p^n} : \mathbb{F}_p]}{[\mathbb{F}_{p^n} : \mathrm{Fix}\langle \theta^{n/k} \rangle]} = \frac{n}{|\langle \theta^{n/k} \rangle|} = \frac{n}{k},$$

so $|\mathrm{Fix}\langle \theta^{n/k} \rangle| = p^{n/k}$. As k runs through the divisors of n, the quotient n/k also runs through the divisors of n, so putting $m = n/k$ gives the result. $\square$

Warning 10.3.7 The subfields of $\mathbb{F}_{p^n}$ are of the form $\mathbb{F}_{p^m}$ where m *divides* n, not $m \leq n$. For instance, $\mathbb{F}_8$ has no subfield isomorphic to $\mathbb{F}_4$ (that is, no 4-element subfield), since $8 = 2^3$, $4 = 2^2$, and $2 \nmid 3$.

Let m be a divisor of n. By Proposition 10.3.6, $\mathbb{F}_{p^n}$ has exactly one subfield isomorphic to $\mathbb{F}_{p^m}$. We can therefore speak of the extension $\mathbb{F}_{p^n} : \mathbb{F}_{p^m}$ without ambiguity. Since $\mathbb{F}_{p^m} = \mathrm{Fix}\langle \theta^m \rangle$ (by Proposition 10.3.6) and $\langle \theta^m \rangle \cong C_{n/m}$, it follows from the fundamental theorem that

$$\mathrm{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_{p^m}) \cong C_{n/m}. \tag{10.1}$$

So in working out the Galois correspondence for $\mathbb{F}_{p^n} : \mathbb{F}_p$, we have accidentally derived the Galois group of a completely arbitrary extension of finite fields. Another way to phrase (10.1) is:

Proposition 10.3.8 *Let $M : K$ be a field extension with M finite. Then* $\mathrm{Gal}(M : K)$ *is cyclic of order* $[M : K]$. □

In the Galois correspondence for $\mathbb{F}_{p^n} : \mathbb{F}_p$, all the extensions and subgroups involved are normal, either by Lemma 10.3.2 or because cyclic groups are abelian. For $m \mid n$, the isomorphism
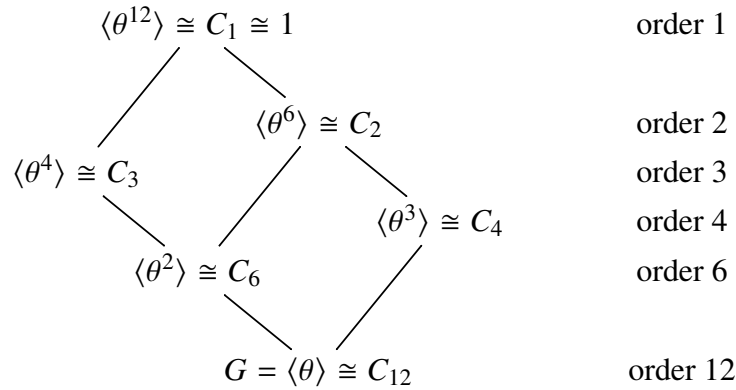
$$\frac{\mathrm{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_p)}{\mathrm{Gal}(\mathbb{F}_{p^n} : \mathbb{F}_{p^m})} \cong \mathrm{Gal}(\mathbb{F}_{p^m} : \mathbb{F}_p)$$

supplied by the fundamental theorem amounts to
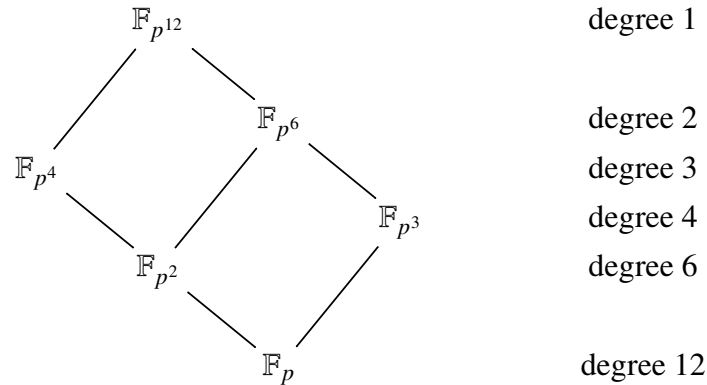
$$\frac{C_n}{C_{n/m}} \cong C_m.$$

Alternatively, substituting $k = n/m$, this is $C_n/C_k \cong C_{n/k}$.

Example 10.3.9 Consider the Galois correspondence for $\mathbb{F}_{p^{12}} : \mathbb{F}_p$, where p is any prime. Writing θ for the Frobenius automorphism of $\mathbb{F}_{p^{12}}$, the subgroups of $G = \mathrm{Gal}(\mathbb{F}_{p^{12}} : \mathbb{F}_p)$ are

$$\langle \theta^{12} \rangle \cong C_1 \cong 1 \qquad \qquad \text{order } 1$$

$$\langle \theta^6 \rangle \cong C_2 \qquad \qquad \text{order } 2$$

$$\langle \theta^4 \rangle \cong C_3 \qquad \qquad \text{order } 3$$

$$\langle \theta^3 \rangle \cong C_4 \qquad \qquad \text{order } 4$$

$$\langle \theta^2 \rangle \cong C_6 \qquad \qquad \text{order } 6$$

$$G = \langle \theta \rangle \cong C_{12} \qquad \qquad \text{order } 12$$

Their fixed fields are

$$\mathbb{F}_{p^{12}} \qquad \qquad \text{degree } 1$$

$$\mathbb{F}_{p^6} \qquad \qquad \text{degree } 2$$

$$\mathbb{F}_{p^4} \qquad \qquad \text{degree } 3$$

$$\mathbb{F}_{p^3} \qquad \qquad \text{degree } 4$$

$$\mathbb{F}_{p^2} \qquad \qquad \text{degree } 6$$

$$\mathbb{F}_p \qquad \qquad \text{degree } 12$$

Here, 'degree' means the degree of $\mathbb{F}_{p^{12}}$ over the subfield, and (for instance) the subfield of $\mathbb{F}_{p^{12}}$ called $\mathbb{F}_{p^4}$ is

$$\left\{ \alpha \in \mathbb{F}_{p^{12}} : \alpha^{p^4} = \alpha \right\}.$$

The Galois group $\mathrm{Gal}(\mathbb{F}_{p^{12}} : \mathbb{F}_{p^4})$ is $\langle \theta^4 \rangle \cong C_3$, and similarly for the other subfields.

Exercise 10.3.10 What do the diagrams of Example 10.3.9 look like for p^8 in place of p^{12}? What about p^{432}? (Be systematic!)

Ordered sets

In Workshop 5, you'll be asked to work through the Galois correspondence for an arbitrary extension $\mathbb{F}_{p^n} : \mathbb{F}_{p^m}$ of finite fields, but there's not much more to do: almost all the work is contained in the case $m = 1$ that we have just done.

*   *   *