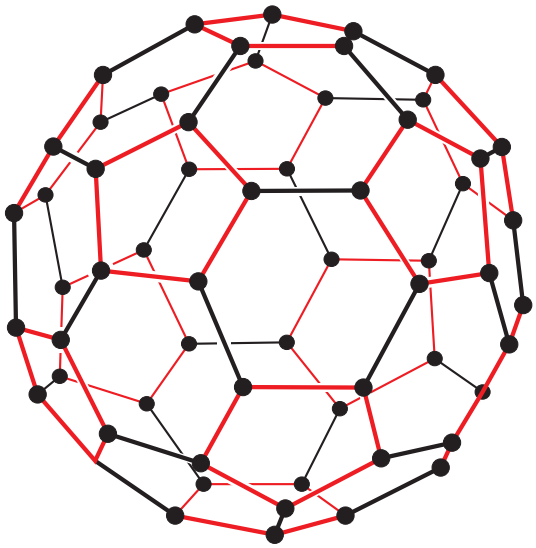
Symmetries of Equations: An Introduction to Galois Theory

Brent Everitt



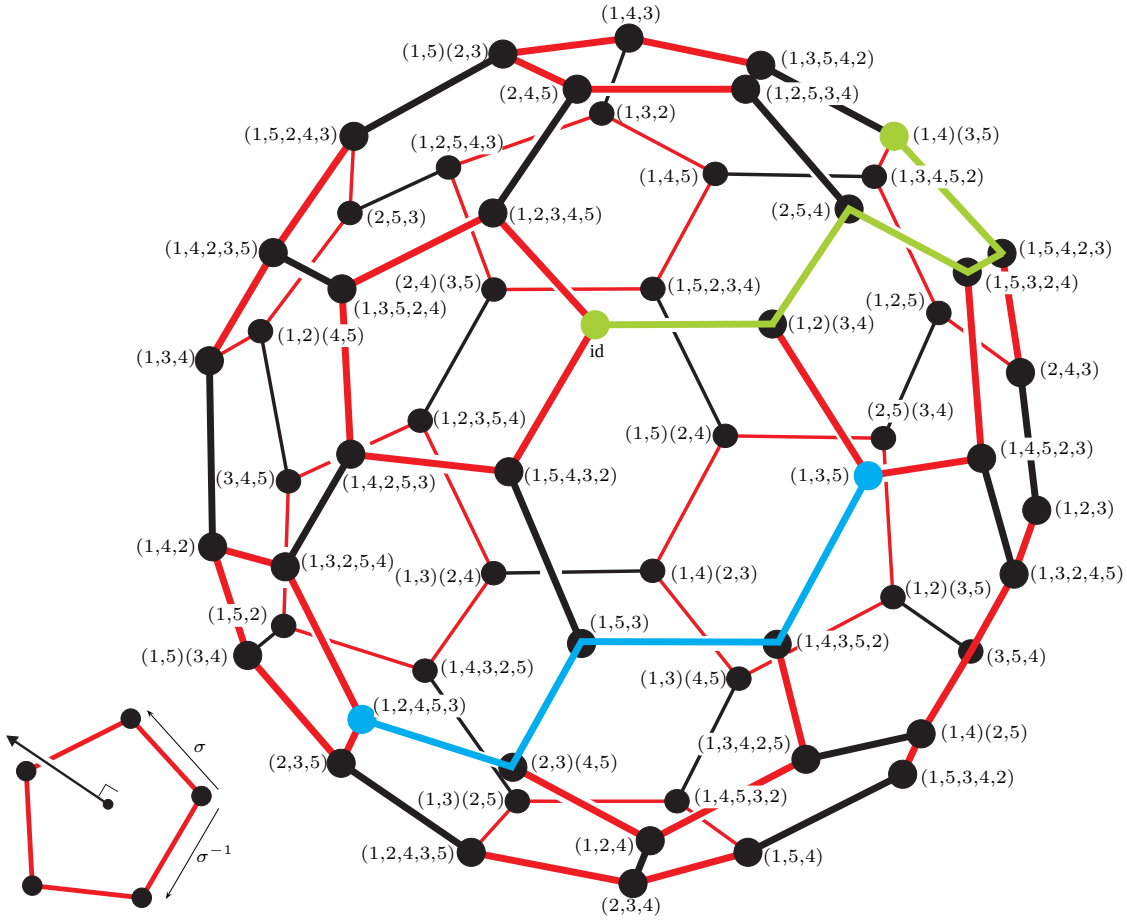
Contents

Department of Mathematics, University of York, York YO10 5DD, England.

THE COVER shows the Cayley graph for the smallest non-Abelian simple group, the alternating group A_5 (see §**11.**). We will see in §**16.** that the simplicity of this group means there is no algebraic expression for any of the roots of the polynomial $x^5 - 4x + 2$ using the algebraic ingredients,

$$\frac{a}{b} \in \mathbb{Q}, +, -, \times, \div, \sqrt[2]{}, \sqrt[3]{}, \sqrt[4]{}, \sqrt[5]{}, \ldots,$$

so therefore there can be no formula for the solutions of $ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0$ that works for all possible $a, b, c, d, e, f \in \mathbb{C}$.



The Cayley graph is a visual depiction of the multiplication in the group A_5. The vertices correspond to the elements of the group as marked, the red edges to the particular element $\sigma = (1, 2, 3, 4, 5)$ and the black edges to $\tau = (1, 2)(3, 4)$. The red pentagonal faces are oriented anticlockwise with respect to the outward pointing normal vector (use the right-hand rule), so that crossing a red edge in an anticlockwise direction corresponds to σ and crossing in a clockwise direction corresponds to σ^{-1} (as in the diagram).

The edges depict multiplication on the right: crossing a red edge anticlockwise (repectively clockwise) multiplies the label of the start vertex by σ (resp. σ^{-1}) to give the label of the finish vertex; crossing a black edge in either direction multiplies the label of the start vertex by τ to give the label of the finish vertex[1].

Thus, the green sequence of edges gives the decomposition $(1, 4)(3, 5) = \tau\sigma^{-2}\tau\sigma^{-1}$ and the blue sequence shows that $(1, 2, 4, 5, 3)\tau\sigma^{-2}\tau = (1, 3, 5)$.

It is a curious coincidence that the Cayley graph of the simplest non-Abelian simple group is the Buckminsterfullerene molecule: the simplest known pure form of Carbon.

[1]The reason for the lack of orientation on the black edges is because the permutation $\tau = \tau^{-1}$.

§1. What is Galois Theory?

A quadratic equation $ax^2 + bx + c = 0$ has exactly two (possibly repeated) solutions in the complex numbers. We can even write an algebraic expression for them, thanks to a formula that first appears in the ninth century book *Hisab al-jabr w'al-muqabala* by Abu Abd-Allah ibn Musa al'Khwarizmi, and written in modern notation as,

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

Less familiar maybe, $ax^3 + bx^2 + cx + d = 0$ has three $\mathbb{C}$-solutions, and they too can be expressed algebraically using Cardano's formula. For instance, one solution turns out to be,

$$-\frac{b}{3a} + \sqrt[3]{-\frac{1}{2}\left(\frac{2b^3}{27a^3} - \frac{bc}{a^2} + \frac{d}{a}\right) + \sqrt{\frac{1}{4}\left(\frac{2b^3}{27a^3} - \frac{bc}{a^2} + \frac{d}{a}\right)^2 + \frac{1}{27}\left(\frac{c}{a} - \frac{b^2}{3a^2}\right)^3}} + \sqrt[3]{-\frac{1}{2}\left(\frac{2b^3}{27a^3} - \frac{bc}{a^2} + \frac{d}{a}\right) - \sqrt{\frac{1}{4}\left(\frac{2b^3}{27a^3} - \frac{bc}{a^2} + \frac{d}{a}\right)^2 + \frac{1}{27}\left(\frac{c}{a} - \frac{b^2}{3a^2}\right)^3}},$$

and the other two have similarly horrendous expressions. There is an even more complicated formula, attributed to Descartes, for the roots of a quartic polynomial equation.

What is mildly miraculous is not that the solutions exist, but they can always be expressed algebraically in terms of the coefficients and the basic algebraic operations,

$$+, -, \times, \div, \sqrt{\ }, \sqrt[3]{\ }, \sqrt[4]{\ }, \sqrt[5]{\ }, \ldots$$

By the turn of the 19th century, no equivalent formula for the solutions to a quintic (degree five) polynomial equation had materialised, and it was Abels who had the crucial realisation: *no such formula exists!*

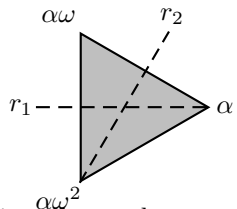
Such a statement can be interpreted in a number of ways. Does it mean that there are always algebraic expressions for the roots of quintic polynomials, but their form is too complex for one *single* formula to describe all the possibilities? It would therefore be necessary to have a number, maybe even infinitely many, formulas. The reality turns out to be far worse: there are specific polynomials, such as $x^5 - 4x + 2$, whose solutions *cannot be expressed algebraically in any way whatsoever*. There is no formula for the roots of just this single polynomial, never mind all the others.

A few decades after Abel's bombshell, Evaristé Galois started thinking about the deeper problem: *why* don't these formulae exist? Thus Galois theory was originally motivated by the desire to understand, in a much more precise way than they hitherto had been, the solutions to polynomial equations.

Galois' idea was this: *study the solutions by studying their "symmetries"*. Nowadays, when we hear the word symmetry, we normally think of group theory rather than number theory. Actually, to reach his conclusions, Galois kind of invented group theory along the way. In studying the symmetries of the solutions to a polynomial, Galois theory establishes a link between these two areas of mathematics. We illustrate the idea, in a somewhat loose manner, with an example.

The symmetries of the solutions to $x^3 - 2 = 0$.

(1.1) We work in $\mathbb{C}$. Let α be the real cube root of 2, ie: $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and, $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. Note that ω is a cube root of 1, and so $\omega^3 = 1$.



The three solutions to $x^3 - 2 = 0$ (or *roots* of $x^3 - 2$) are the complex numbers $\alpha, \alpha\omega$ and $\alpha\omega^2$, forming the vertices of the equilateral triangle shown. The triangle has what we might call "geometric symmetries": three reflections, a counter-clockwise rotation through $\frac{1}{3}$ of a turn, a counter-clockwise rotation through $\frac{2}{3}$ of a turn and a counter-clockwise rotation through $\frac{3}{3}$ of a turn $=$ the identity symmetry. Notice for now that if r_1 and r_2 are the reflections in the lines shown, the geometrical symmetries are r_1, r_2, $r_2r_1r_2$, r_2r_1, $(r_2r_1)^2$ and $(r_2r_1)^3 = 1$ (read these expressions from right to left).

The symmetries referred to in the preamble are not so much geometric as "number theoretic". It will take a little explaining before we see what this means.

(1.2) A *field* is a set F with two operations, called, *purely for convenience*, $+$ and $\times$, such that for any $a, b, c \in F$,

1. $a + b$ and $a \times b$ ($= ab$ from now on) are uniquely defined elements of F,
2. $a + (b + c) = (a + b) + c$,
3. $a + b = b + a$,
4. there is an element $0 \in F$ such that $0 + a = a$,
5. for any $a \in F$ there is an element $-a \in F$ with $(-a) + a = 0$,
6. $a(bc) = (ab)c$,
7. $ab = ba$,
8. there is an element $1 \in F \setminus \{0\}$ with $1 \times a = a$,
9. for any $a \neq 0 \in F$ there is an $a^{-1} \in F$ with $aa^{-1} = 1$,
10. $a(b + c) = ab + ac$.

A field is just a set of things that you can add, subtract, multiply and divide so that the "usual" rules of algebra are satisfied. Familiar examples of fields are $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$; familiar examples of non-fields are $\mathbb{Z}$, polynomials and matrices (you can't in general divide integers, polynomials and matrices to get integers, polynomials or matrices).

(1.3) A *subfield* of a field F is a subset that also forms a field under the same $+$ and $\times$. Thus, $\mathbb{Q}$ is a subfield of $\mathbb{R}$ which is in turn a subfield of $\mathbb{C}$, and so on. On the other hand, $\mathbb{Q} \cup \{\sqrt{2}\}$ is not a subfield of $\mathbb{R}$: it is certainly a subset but axiom 1 fails, as both 1 and $\sqrt{2}$ are elements but $1 + \sqrt{2}$ is not.

Definition. If F is a subfield of the complex numbers $\mathbb{C}$ and $\beta \in \mathbb{C}$, then $F(\beta)$, is the "smallest" subfield of $\mathbb{C}$ that contains both F and the number β.

What do we mean by smallest? That there is no other field F' having the same properties as $F(\beta)$ which is smaller, ie: no F' with $F \subset F'$ and $\beta \in F'$ too, but F' properly $\subset F(\beta)$. It is usually more useful to say it the other way around:

$$\text{If } F' \text{ is a subfield that also contains } F \text{ and } \beta, \text{ then } F' \text{ contains } F(\beta) \text{ too.} \qquad (*)$$

Loosely speaking, $F(\beta)$ is all the complex numbers we get by adding, subtracting, multiplying and dividing the elements of F and β together in all possible ways.

(1.4) To illustrate with some trivial examples, $\mathbb{R}(i)$ can be shown to be all of $\mathbb{C}$: it must contain all expressions of the form bi for $b \in \mathbb{R}$, and hence all expressions of the form $a + bi$ with $a, b \in \mathbb{R}$, and this accounts for all the complex numbers; $\mathbb{Q}(2)$ is equally clearly just $\mathbb{Q}$ back again.

Slightly less trivially, $\mathbb{Q}(\sqrt{2})$, the smallest subfield of $\mathbb{C}$ containing all the rational numbers and $\sqrt{2}$ is a field that is strictly bigger than $\mathbb{Q}$ (eg: it contains $\sqrt{2}$) but is much, much smaller than all of $\mathbb{R}$.

Exercise 1 Show that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$.

(1.5) Returning to the symmetries of the solutions to $x^3 - 2 = 0$, we look at the field $\mathbb{Q}(\alpha, \omega)$, where $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, as before. Since $\mathbb{Q}(\alpha, \omega)$ is by definition a field, and fields are closed under $+$ and $\times$, we have

$$\alpha \in \mathbb{Q}(\alpha, \omega) \text{ and } \omega \in \mathbb{Q}(\alpha, \omega) \Rightarrow \alpha \times \omega = \alpha\omega, \alpha \times \omega \times \omega = \alpha\omega^2 \in \mathbb{Q}(\alpha, \omega) \text{ too.}$$

So, $\mathbb{Q}(\alpha, \omega)$ contains all the solutions to the equation $x^3 - 2 = 0$. On the other hand,

Exercise 2 Show that $\mathbb{Q}(\alpha, \omega)$ has "just enough" numbers in it to solve the equation $x^3 - 2 = 0$. More precisely, $\mathbb{Q}(\alpha, \omega)$ is the *smallest* (in the sense (*))subfield of $\mathbb{C}$ that contains all the solutions to this equation. (*hint*: you may find it useful to do Exercise 5 first).

(1.6) A very loose definition of a symmetry of the solutions of $x^3 - 2 = 0$ is that it is a "rearrangement" of $\mathbb{Q}(\alpha, \omega)$ that does not disturb (or is compatible with) the $+$ and $\times$.

To see an example, consider the two fields $\mathbb{Q}(\alpha, \omega)$ and $\mathbb{Q}(\alpha, \omega^2)$. Despite first appearances they are actually the same: certainly

$$\alpha, \omega \in \mathbb{Q}(\alpha, \omega) \Rightarrow \alpha, \omega^2 \in \mathbb{Q}(\alpha, \omega).$$

But $\mathbb{Q}(\alpha, \omega^2)$ is the smallest field containing $\mathbb{Q}, \alpha$ and ω^2, so by (*),

$$\mathbb{Q}(\alpha, \omega^2) \subseteq \mathbb{Q}(\alpha, \omega).$$

Conversely,

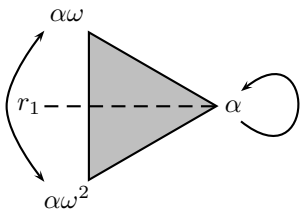
$$\alpha, \omega^2 \times \omega^2 = \omega^4 = \omega \in \mathbb{Q}(\alpha, \omega^2) \Rightarrow \mathbb{Q}(\alpha, \omega) \subseteq \mathbb{Q}(\alpha, \omega^2).$$

Remember that $\omega^3 = 1$ so $\omega^4 = \omega$. Thus $\mathbb{Q}(\alpha, \omega)$ and $\mathbb{Q}(\alpha, \omega^2)$ are indeed the same. In fact, we should think of $\mathbb{Q}(\alpha, \omega)$ and $\mathbb{Q}(\alpha, \omega^2)$ as two different ways of looking at the same field, or more suggestively, the same field viewed from two different angles.

Whenever we hear the phrase, "the same field viewed from two different angles", we should immediately think that a symmetry is lurking–a symmetry that moves the field from the one point of view to the other. In the case above, there should be a symmetry of the field $\mathbb{Q}(\alpha, \omega)$ that puts it into the form $\mathbb{Q}(\alpha, \omega^2)$. Surely this symmetry should send

$$\alpha \mapsto \alpha, \text{ and } \omega \mapsto \omega^2.$$

We haven't yet defined what we mean by, "is compatible with the $+$ and $\times$". It will turn out to mean that if α and ω are sent to α and ω^2 respectively, then $\alpha \times \omega$ should go to $\alpha \times \omega^2$; similarly $\alpha \times \omega \times \omega$ should go to $\alpha \times \omega^2 \times \omega^2 = \alpha\omega^4 = \alpha\omega$. The symmetry thus moves the vertices of the equilateral triangle determined by the roots in the same way that the reflection r_1 of the triangle does[2]:



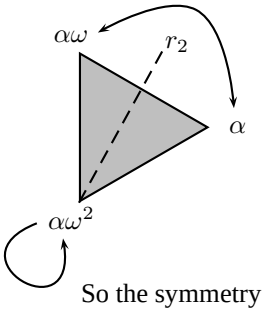
(1.7) In exactly the same way, we can consider the fields $\mathbb{Q}(\alpha\omega, \omega^2)$ and $\mathbb{Q}(\alpha, \omega)$. We have

$$\alpha, \omega \in \mathbb{Q}(\alpha, \omega) \Rightarrow \omega^2, \alpha\omega \in \mathbb{Q}(\alpha, \omega) \Rightarrow \mathbb{Q}(\alpha\omega, \omega^2) \subseteq \mathbb{Q}(\alpha, \omega);$$

and conversely, $\alpha\omega, \omega^2 \in \mathbb{Q}(\alpha\omega, \omega^2) \Rightarrow \alpha\omega\omega^2 = \alpha\omega^3 = \alpha \in \mathbb{Q}(\alpha\omega, \omega^2)$, and hence also

$$\alpha^{-1}\alpha\omega = \omega \in \mathbb{Q}(\alpha\omega, \omega^2) \Rightarrow \mathbb{Q}(\alpha, \omega) \subseteq \mathbb{Q}(\alpha\omega, \omega^2).$$

[2]This compatability also means that it would have made no sense to have the symmetry send $\alpha \mapsto \omega^2$ and $\omega \mapsto \alpha$. A symmetry should not fundamentally change the algebra of the field, so that if an element like ω cubes to give 1, then its image under the symmetry should too: but α *doesn't* cube to give 1.



Thus, $\mathbb{Q}(\alpha, \omega)$ and $\mathbb{Q}(\alpha\omega, \omega^2)$ are the same field, and we can define another symmetry that sends

$$\alpha \mapsto \alpha\omega, \text{ and } \omega \mapsto \omega^2.$$

To be compatible with the $+$ and $\times$,

$$\alpha \times \omega \mapsto \alpha\omega \times \omega^2 = \alpha\omega^3 = \alpha, \text{ and } \alpha \times \omega \times \omega \mapsto \alpha\omega \times \omega^2 \times \omega^2 = \alpha\omega^5 = \alpha\omega^2.$$

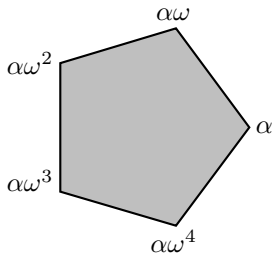
So the symmetry is like the reflection r_2 of the triangle:

Finally, if we have two symmetries of the solutions to some equation, we would like their composition to be a symmetry too. So if the symmetries r_1 and r_2 of the original triangle are to be considered, so should $r_2 r_1 r_2, r_1 r_2, (r_1 r_2)^2$ and $(r_1 r_2)^3 = 1$.

(1.8) The symmetries of the solutions to $x^3 - 2 = 0$ include all the geometrical symmetries of the equilateral triangle. We will see much later that any symmetry of the solutions is uniquely determined as a *permutation* of the solutions. Since there are $3! = 6$ of these, we have accounted for all of them. It would appear then that the solutions to $x^3 - 2 = 0$ have symmetry *precisely* the geometrical symmetries of the equilateral triangle.

(1.9) If this was always the case, things would be very simple: Galois theory would just be the study of the "shapes" formed by the roots of polynomials, and the symmetries of those shapes. It would be a branch of planar geometry.

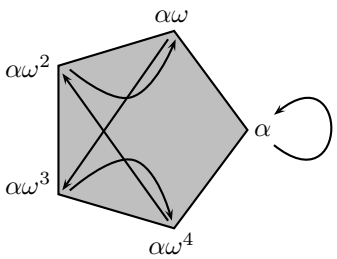
But things are not so simple. If we look at the solutions to $x^5 - 2 = 0$, something quite different happens:



$$\alpha = \sqrt[5]{2}$$

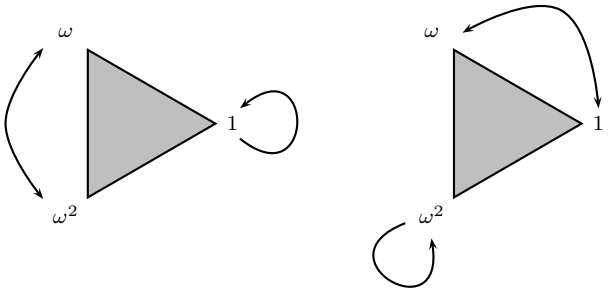
$$\omega = \frac{\sqrt{5}-1}{4} + \frac{\sqrt{2}\sqrt{5+\sqrt{5}}}{4} i$$

We will see later on how to obtain these expressions for the roots. A pentagon has 10 geometric symmetries, and you can check that all arise as symmetries of the roots of $x^5 - 2$ using the same reasoning as in the previous example. But this reasoning also gives a symmetry that moves the vertices of the pentagon according to:



This is not a geometrical symmetry! Later we will see that for $p > 2$ a prime number, the solutions to $x^p - 2 = 0$ have $p(p-1)$ symmetries. While agreeing with the six obtained for $x^3 - 2 = 0$, it gives *twenty* for $x^5 - 2 = 0$. In fact, it was a bit of a fluke that all the number theoretic symmetries were also geometric ones for $x^3 - 2 = 0$. A p-gon has $2p$ geometrical symmetries and $2p \leq p(p-1)$ with equality *only* when $p = 3$.

Exercise 3 Show that the figure on the left depicts a symmetry of the solutions to $x^3 - 1 = 0$, but the one on the right does not.



Further Exercises for §**1.**

Exercise 4 You already know that the 3-rd roots of 1 are 1 and $-\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$. What about the p-th roots for higher primes?

1. If $\omega \neq 1$ is a 5-th root it satisfies $\omega^4 + \omega^3 + \omega^2 + \omega + 1 = 0$. Let $u = \omega + \omega^{-1}$. Find a quadratic polynomial satisfied by u, and solve it to obtain u.
2. Find another quadratic satisfied this time by ω, with *coefficients involving* u, and solve it to find explicit expressions for the four primitive 5-th roots of 1.
3. Repeat the process with the 7-th roots of 1.

factoid: the n-th roots of 1 can be expressed in terms of field operations and extraction of pure roots of rationals for any n. The details (which are a little complicated!) were finally completed by the work of Gauss and Galois.

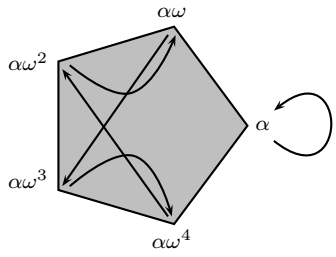
Exercise 5

1. Let F be a field such that the element
$$\underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} \neq 0,$$
for any $n > 0$. Argueing intuitively, show that F contains a copy of the rational numbers $\mathbb{Q}$ (see also §**4.**).
2. Give an example of a field where
$$\underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = 0,$$
for some n.

Exercise 6 Let $\alpha = \sqrt[6]{5} \in \mathbb{R}$ and $\omega = \frac{1}{2} + \frac{\sqrt{3}}{2}i$. Show that $\mathbb{Q}(\alpha, \omega)$, $\mathbb{Q}(\alpha\omega^2, \omega^5)$ and $\mathbb{Q}(\alpha\omega^4, \omega^5)$ are all the same field.

Exercise 7

1. Show that there is a symmetry of the solutions to $x^5 - 2 = 0$ that moves the vertices of the pentagon according to:



where $\alpha = \sqrt[5]{2}$, and $\omega^5 = 1, \omega \in \mathbb{C}$.

2. Show that the solutions in $\mathbb{C}$ to the equation $x^6 - 5 = 0$ have 12 symmetries.

§2. Polynomials, Rings and Polynomial Rings

(2.1) There are a number of basic facts about polynomials that we will need. Suppose F is a field ($\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$ will do for now). A polynomial over F is an expression of the form

$$f = a_0 + a_1 x + \cdots a_n x^n,$$

where the $a_i \in F$ and x is a "formal symbol" (sometimes called an indeterminate). We don't tend to think of x as a variable–it is purely an object on which to perform algebraic manipulations. Denote the set of all polynomials over F by $F[x]$. If $a_n \neq 0$, then n is called the *degree*[3] of f, written $\deg(f)$. If the leading coefficient $a_n = 1$, then f is *monic*.

(2.2) We can add and multiply elements of $F[x]$ in the usual way:

$$\text{if } f = \sum_{i=0}^{n} a_i x^i \text{ and } g = \sum_{i=0}^{m} b_i x^i,$$

then,

$$f + g = \sum_{i=0}^{\max(m,n)} (a_i + b_i) x^i \text{ and } fg = \sum_{k=0}^{m+n} c_k x^k \text{ where } c_k = \sum_{i+j=k} a_i b_j. \tag{1}$$

that is, $c_k = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0$. The arithmetic of the coefficients (ie: how to work out $a_i + b_i, a_i b_j$ and so on) is just that of the field F.

Exercise 8 Convince yourself that this multiplication is really just the "expanding brackets" multiplication of polynomials that you know so well!

(2.3) The polynomials $F[x]$ together with this addition form an example of a,

Definition. A *group* is a set G endowed with an operation $\oplus$ such that for all $a, b \in G$,

1. $a \oplus b$ is a uniquely defined element of G (closure);
2. $a \oplus (b \oplus c) = (a \oplus b) \oplus c$ (associativity);
3. there is an $e \in G$ such that $e \oplus a = a = a \oplus e$ (identity),;
4. for any $a \in G$ there is an $a^{-1} \in G$ with $a \oplus a^{-1} = e = a^{-1} \oplus a$ (inverses).

A group that also satisfies $a \oplus b = b \oplus a$ for all $a, b \in G$ (commutativity) is said to be *Abelian*.

With polynomials, the operation $\oplus$ is just the regular addition of polynomials. When the group operation is "familiar" addition it is customary to use the symbols: $+$ for $\oplus$; 0 for e and $-$ for inverses. Thus the identity of $F[x]$ as a group is the zero polynomial and inverses are given by

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

Its also easy to see that $F[x]$ forms an abelian group: for $f + g = g + f$ exactly when $a_i + b_i = b_i + a_i$ for all i. But the coefficients of our polynomials come from the field F, and addition is always commutative in a field.

[3] In one of those triumphs of notation over intuition for which Mathematics is justifiably famous, define $\deg(0) = -\infty$, whereas $\deg(\lambda) = 0$ if $\lambda \in F$ is not zero. The arithmetic of degrees is then just the arithmetic of non-negative integers, except we also need to decree that $-\infty + n = -\infty$.

(2.4) If we want to think about multiplication as well, we need the formal concept of,

Definition. A *ring* is a set R endowed with two operations $\oplus$ and $\otimes$ such that for all $a, b \in R$,

1. R is an Abelian group under $\oplus$;
2. for any $a, b \in R$, $a \otimes b$ is a uniquely determined element of R (closure of $\otimes$);
3. $a \otimes (b \otimes c) = (a \otimes b) \otimes c$ (associativity of $\otimes$);
4. there is an $f \in R$ such that $f \otimes a = a = a \otimes f$ (identity of $\otimes$);
5. $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$ (the distributive law).

Loosely, a ring is a set on which you can *add* ($\oplus$), *subtract* (the inverse of $\oplus$ in the Abelian group) and *multiply* ($\otimes$), but *not* necessarily divide (there is no inverse axiom for $\otimes$).

Here are some well known examples of rings:

$$\mathbb{Z}, F[x] \text{ for } F \text{ a field}, \mathbb{Z}_n \text{ and } M_n(F),$$

where $\mathbb{Z}_n$ is addition and multiplication of integers modulo n and $M_n(F)$ are the $n \times n$ matrices, with entries from F, together with the usual addition and multiplication of matrices.

A ring is *commutative* if the second operation $\otimes$ is commutative: $a \otimes b = b \otimes a$ for all a, b.

Exercise 9

1. Show that $fg = gf$ for polynomials $f, g \in F[x]$, hence $F[x]$ is a commutative ring.
2. Show that $\mathbb{Z}$ and $\mathbb{Z}_n$ are commutative rings, but $M_n(F)$ is not for *any* field F if $n > 2$.

(2.5) The observation that $\mathbb{Z}$ and $F[x]$ are both commutative rings is not just some vacuous formalism. A concrete way of putting it it this: *at a very fundamental level, integers and polynomials share the same algebraic properties*.

When we work with polynomials, we need to be able to add and multiply the coefficients of the polynomials in a way that doesn't produce any nasty surprises–in other words, the coefficients have to satisfy the basic rules of algebra that we all know and love. But these basic rules of algebra can be found among the axioms of a ring. *Thus, to work with polynomials successfully, all we need is that the coefficients come from a ring*.

This observation means that for a ring R, we can form the set of all polynomials with coefficients from R and add and multiply them together as we did above. In fact, we are just repeating what we did above, but are replacing the field F with a ring R. In practice, rather than allowing our coefficients to some from an arbitrary ring, we take R to be commutative. Since we are so used to our coefficients commuting with each other, this is probably a prudent precaution. This all leads to,

Definition. Denote by $R[x]$ the set of all polynomials with coefficients from some commutative ring R, together with the $+$ and $\times$ defined at (1).

Exercise 10

1. Show that $R[x]$ forms a ring.
2. Since $R[x]$ forms a ring, we can consider polynomials with coefficients from $R[x]$: take a new variable, say y, and consider $R[x][y]$. Show that this is just the set of polynomials in two variables x and y together with the 'obvious' $+$ and $\times$.

(2.6) A commutative ring R is called an *integral domain* iff for any $a, b \in R$ with $a \otimes b = e$, we have $a = e$, or $b = e$ or both. Clearly $\mathbb{Z}$ is an integral domain.

Exercise 11

1. Show that any field F is an integral domain.
2. For what values of n is $\mathbb{Z}_n$ an integral domain?

Lemma 1 *Let $f, g \in R[x]$, R an integral domain. Then*

1. $\deg(fg) = \deg(f) + \deg(g)$.
2. *$R[x]$ is an integral domain.*

The second part means that given polynomials f and g (with coefficients from an integral domain), we have $fg = 0 \Rightarrow f = 0$ or $g = 0$. You have been implicitly using this fact for some time now when you solve polynomial equations by factorising them.

Proof: We have

$$fg = \sum_{k=0}^{m+n} c_k x^k \text{ where } c_k = \sum_{i+j=k} a_i b_j,$$

so in particular $c_{m+n} = a_n b_m \neq 0$ as R is an integral domain. Thus $\deg(fg) \geq m + n$ and since the reverse inequality is obvious, we have part (1) of the Lemma. Part (2) now follows immediately since $fg = 0 \Rightarrow \deg(fg) = -\infty \Rightarrow \deg f + \deg g = -\infty$, which can only happen if at least one of f or g has degree $= -\infty$ (see the footnote at the bottom of the first page). □

All your life you have been happily adding the degrees of polynomials when you multiply them. But as the result above shows, *this is only possible when the coefficients of the polynomial come from an integral domain*. For example, $\mathbb{Z}_6$, the integers under addition and multiplication modulo 6, is a ring that is not an integral domain (as $2 \times 3 = 0$ for example), and sure enough,

$$(3x + 1)(2x + 1) = 5x + 1,$$

where all of this is happening in $\mathbb{Z}_6[x]$.

(2.7) Although we cannot necessarily divide two polynomials and get another polynomial, we *can* divide upto a possible "error term", or, as it is more commonly called, a remainder.

Theorem A (The division algorithm). *Suppose f and g are elements of $R[x]$ where the leading coefficient of g has a multiplicative inverse in the ring R. Then there exist q and r in $R[x]$ (quotient and remainder) such that*

$$f = qg + r,$$

where either $r = 0$ or the degree of r is $<$ the degree of g.

When R is a field (where you may be more used to doing long division) all the non-zero coefficients of a polynomial have multiplicative inverses (as they lie in a field) so the condition on g becomes $g \neq 0$.

Actually the name of the theorem is not very apt: it merely guarantees the existence of a quotient and remainder. It doesn't give us any idea how to find them (in other words, an algorithm). Compare the theorem with what you know about $\mathbb{Z}$. There, we can also divide to get a remainder: when you divide 17 by 3, it goes 5 times with remainder 2; in other words, $17 = 5 \times 3 + 2$. With integers, we are used to the remainder being smaller than the integer we are dividing by; in $R[x]$ this condition is replaced by the degree of the remainder being strictly smaller than the degree of the divisor.

Proof: For all $q \in R[x]$, consider those polynomials of the form $f - gq$ and choose one, say r, of smallest degree. Let $d = \deg r$ and $m = \deg g$. We claim that $d < m$. This will give the result, as the r chosen has he form $r = f - gq$ for some q, giving $f = gq + r$. Suppose that $d \geq m$ and consider

$$\bar{r} = (r_d)(g_m^{-1})x^{(d-m)}g,$$

a polynomial since $d - m \geq 0$. Notice also that we have used the fact that the leading coefficient of g has a multiplicative inverse. The leading term of $\bar{r}$ is $r_d x^d$, which is also the leading term of r. Thus, $r - \bar{r}$ has degree $< d$. But $r - \bar{r} = f - gq - r_d g_m^{-1} x^{d-m} g$ by definition, which equals $f - g(q - r_d g_m^{-1} x^{d-m}) = f - g\bar{q}$, say. Thus $r - \bar{r}$ has the form $f - g\bar{q}$ too, but with smaller degree than r, which was of minimal degree amongst all polynomials of this form–this is our desired contradiction. □

Exercise 12

1. If R is an integral domain, show that the quotient and remainder are unique.
2. Show that the quotient and remainder are not unique when you divide polynomials in $\mathbb{Z}_6[x]$.

(2.8) Other familiar concepts from $\mathbb{Z}$ are those of divisors, common divisors and greatest common divisors. Since we need no more algebra to define these notions than is enshrined in the axioms for a ring, it should come as no surprise that these concepts carry pretty much straight over to polynomial rings. We will state these in the setting of polynomials from $F[x]$ for F a field.

Definition. For $f, g \in F[x]$, we say that f divides g iff $g = fh$ for some $h \in F[x]$. Write $f \mid g$.

Definition. Let $f, g \in F[x]$. Suppose that d is a polynomial satisfying

1. d is a common divisor of f and g, ie: $d \mid f$ and $d \mid g$;
2. d is the greatest common divisor in the sense that any other common divisor must divide d (and so in particular be smaller!), ie: if $c \mid f$ and $c \mid g$ then $c \mid d$;
3. d is monic.

As with the division algorithm, we have tweaked the definition from $\mathbb{Z}$ to make it work in $F[x]$. The reason is that we want *the* gcd to be unique. In $\mathbb{Z}$ you ensure this by insisting that all gcd's are positive, otherwise, -3 would make a perfectly good gcd for 6 and 27; in $F[x]$ we go for the monic condition (otherwise if d was a gcd of f and g, then $17 \times d$ would be too).

(2.9) $x^2 - 1$ and $2x^3 - 2x^2 - 4x \in \mathbb{Q}[x]$ have greatest common divisor $x + 1$: it is certainly a common divisor as $x^2 - 1 = (x+1)(x-1)$ and $2x^3 - 2x^2 - 4x = 2x(x+1)(x-2)$. From the two factorisations, any other common divisor must have the form $\lambda(x + 1)$ for some $\lambda \in \mathbb{Q}$, and so divides $x + 1$.

(2.10)

Theorem 1 *Any two $f, g \in F[x]$ have a greatest common divisor d. Moreover, there are $a_0, b_0 \in F[x]$ such that*

$$d = a_0 f + b_0 g.$$

Compare this with $\mathbb{Z}$! In fact, one may replace $F[x]$ by $\mathbb{Z}$ in the following proof to obtain the corresponding fact for the integers.

Proof: Consider the set $I = \{af + bg \mid a, b \in F[x]\}$. Let $d \in I$ be a monic polynomial with minimal degree. Then $d \in I$ gives that $d = a_0 f + b_0 g$ for some $a_0, b_0 \in F[x]$. We claim that d is the gcd of f and g. The following two basic facts are easy to verify:

1. The set I is a subgroup of the Abelian group $F[x]$–exercise.
2. If $u \in I$ and $w \in F[x]$ then $uw \in I$, since $wu = w(af + bg) = (wa)f + (wb)g \in I$.

Consider now the set $P = \{hd \mid h \in F[x]\}$. Since $d \in I$ and by the second observation above, $hd \in I$, and we have $P \subseteq I$. Conversely, if $u \in I$ then by the division algorithm, $u = qd + r$ where $r = 0$ or $\deg(r) < \deg(d)$. Now, $r = u - qd$ and $d \in I$, so $qd \in I$ by (2). But $u \in I$ and $qd \in I$ so $u - dq = r \in I$ by (1) above. Thus, if $\deg(r) < \deg(d)$ we would have a contradiction to the degree of d being minimal, and so we must have $r = 0$, giving $u = qd$. This means that any element of I is a multiple of d, so $I \subseteq P$.

Now that we know that I is just the set of all multiples of d, and since letting $a = 1, b = 0$ or $a = 0, b = 1$ gives that $f, g \in I$, we have that d is a common divisor of f and g. Finally, if d' is another common divisor, then $f = u_1 d'$ and $g = u_2 d'$, and since $d = a_0 f + b_0 g$, we have $d = a_0 u_1 d' + b_0 u_2 d' = d'(a_0 u_1 + b_0 u_2)$ giving $d' \mid d$. Thus d is indeed the greatest common divisor. □

(2.11) We have one more thing to say about polynomial rings. First, we need to recall a fundamental notion:

Definition. Let R and S be rings. A mapping $\varphi : R \to S$ is called a *ring homomorphism* if and only if for all $a, b \in R$,

1. $\varphi(a+b) = \varphi(a) + \varphi(b)$;
2. $\varphi(ab) = \varphi(a)\varphi(b)$;
3. $\varphi(f) = f$ (where f is the multiplicative identity in R).

In any ring of interest to us, the last item translates as $\varphi(1) = 1$. Why do we need this but not $\varphi(0) = 0$? Actually it's quite simple: we have $\varphi(0) = \varphi(0+0) = \varphi(0) + \varphi(0)$ and since S is an Abelian group under addition, we can cancel (we are using the existence of inverses under addition!) to get $\varphi(0) = 0$. We can't do this to get $\varphi(1) = 1$ as we don't have inverses under multiplication, so we need to enshrine the desired property in the definition.

You should think of a homomorphism as being like an "algebraic analogy", or a way of transferring algebraic properties; the algebra in the image of φ is analogous to the algebra of R.

(2.12) We will have much more to say about general homomorphisms later on. For now, let's look at one in particular. Let $R[x]$ be a ring of polynomials over a commutative ring R, and let $\lambda \in R$. Define a mapping $\varepsilon_\lambda : R[x] \to R$ by

$$\varepsilon_\lambda(f) = f(\lambda) \stackrel{\text{def}}{=} a_0 + a_1\lambda + \cdots + a_n\lambda^n.$$

ie: substitute λ into f. This is a ring homomorphism from $R[x]$ to R, called the *evaluation at λ homomorphism*: to see this, certainly $\varepsilon_\lambda(1) = 1$, and I'll leave $\varepsilon_\lambda(f+g) = \varepsilon_\lambda(f) + \varepsilon_\lambda(g)$ to you as its not hard. Now,

$$\varepsilon_\lambda(fg) = \varepsilon_\lambda\left(\sum_{k=0}^{m+n} c_k x^k\right) = \sum_{k=0}^{m+n} c_k\lambda^k \text{ where } c_k = \sum_{i+j=k} a_i b_j.$$

But $\sum_{k=0}^{m+n} c_k\lambda^k = \left(\sum_{i=0}^{n} a_i\lambda^i\right)\left(\sum_{j=0}^{m} b_j\lambda^j\right) = \varepsilon_\lambda(f)\varepsilon_\lambda(g)$ and we are done.

One consequence of ε_λ being a homomorphism is that given a factorisation of a polynomial, say $f = gh$, we have $\varepsilon_\lambda(f) = \varepsilon_\lambda(g)\varepsilon_\lambda(h)$, ie: if we substitute λ into f *we get the same answer as when we substitute into g and h and multiply the answers.* This is another fact that appears to be trivial at first sight–you would have instinctively done this anyway no doubt.

Further Exercises for §**2.**

Exercise 13 Let f, g be polynomials over the field F and $f = gh$. Show that h is also a polynomial over F.

Exercise 14 Let $\sigma : R \longrightarrow S$ be a homomorphism of rings. Define $\sigma^* : R[x] \longrightarrow S[x]$ by

$$\sigma^* : \sum_i a_i x^i \mapsto \sum_i \sigma(a_i) x^i.$$

Show that σ^* is a homomorphism.

Exercise 15 let R be a ring and define $\partial : R[x] \longrightarrow R[x]$ by

$$\partial : \sum_{k=0}^{n} a_k x^k \mapsto \sum_{k=1}^{n} (ka_k) x^{k-1} \text{ and } \partial(\lambda) = 0,$$

for any constant λ. (Ring a bell?) Show that $\partial(f+g) = \partial(f) + \partial(g)$ and $\partial(fg) = \partial(f)g + f\partial(g)$. The map ∂ is called the *formal derivative.*

§3. Roots and Irreducibility

(3.1) Much of the material in this section is familiar in the setting of polynomials with $\mathbb{R}$ coefficients. The point is that these results are still true for polynomials with coefficients coming from an arbitrary field F, and quite often, for polynomials with coefficients from a ring R.

Let

$$f = a_0 + a_1x + \cdots + a_nx^n$$

be a polynomial in $R[x]$ for R a ring. We say that $\lambda \in R$ is a *root* of f if

$$f(\lambda) = a_0 + a_1\lambda + \cdots + a_n\lambda^n = 0 \text{ in } R.$$

As a trivial example, the polynomial $x^2 + 1$ is in all three rings $\mathbb{Q}[x], \mathbb{R}[x]$ and $\mathbb{C}[x]$. It has no roots in either $\mathbb{Q}$ or $\mathbb{R}$, but two in $\mathbb{C}$.

Aside. "I thought that we weren't thinking of x as a variable!", I hear you say. In fact we don't need to, as long as we are prepared to think a little more abstractly about something we have been happily doing intuitively for a while now. Here is how: we say that λ is a root of f if and only if there is a homomorphism $\varphi : R[x] \to R$ such that φ restricts to the identity on R, ie: $\varphi(\alpha) = \alpha$ for all $\alpha \in R$, and also that $\varphi(x) = \lambda$ and $\varphi(f) = 0$. In fact you see that the homomorphism needed is the evaluation homomorphism ε_λ.

(3.2)

The Factor Theorem. *An element $\lambda \in R$ is a root of f if and only if $f = (x - \lambda)g$ for some $g \in R[x]$.*

In English, λ is a root exactly when $x - \lambda$ is a factor.

Proof: This is an illustration of the power of the division algorithm, Theorem A. Suppose that f has the form $(x - \lambda)g$ for some $g \in R[x]$. Then

$$f(\lambda) = (\lambda - \lambda)g(\lambda) = 0.g(\lambda) = 0,$$

so that λ is indeed a root (notice we used that ε_λ is a homomorphism, ie: that $\varepsilon_\lambda(f) = \varepsilon_\lambda(x - \lambda)\varepsilon_\lambda(g)$). On the other hand, by the division algorithm, we can divide f by the polynomial $x - \lambda$ to get,

$$f = (x - \lambda)g + \mu,$$

where $\mu \in R$ (we can use the division algorithm, as the leading coefficient of $x - \lambda$, being 1, has an inverse in R). Since $f(\lambda) = 0$, we must also have $(\lambda - \lambda)g + \mu = 0$, hence $\mu = 0$. Thus $f = (x - \lambda)g$ as required. □

(3.3) Here is another result that you probably already know to be true for polynomials over the reals, complexes, etc. Reassuringly, it is true for polynomials with coefficients from (almost) *any* ring.

Theorem 2 *Let $f \in R[x]$ be a non-zero polynomial with coefficients from the integral domain R. Then f has at most $\deg(f)$ roots in R.*

Proof: We use induction on the degree which is ≥ 0 since f is non-zero. If $\deg(f) = 0$ then $f = \mu$ a nonzero constant in R, which clearly has no roots, so the result holds. Assume $\deg(f) \geq 1$ and that the result is true for any polynomial of degree $< \deg(f)$. If f has no roots in R then we are done. Otherwise, f has a root $\lambda \in R$ and

$$f = (x - \lambda)g,$$

for some $g \in R[x]$ by the factor theorem. Moreover, as R is an integral domain, $f(\mu) = 0$ iff either $\mu - \lambda = 0$ or $g(\mu) = 0$, so the roots of f are λ, together with the roots of g. Since the degree of g must be $\deg(f) - 1$ (by Lemma 1, again using the fact that R is an integral domain), it has at most $\deg(f) - 1$ roots by the inductive hypothesis, and these combined with λ give at most $\deg(f)$ roots for f. □

(3.4) As the theorem indicates, a cherished fact such as this might not be true if the coefficients of our polynomial do not come from an integral domain. For instance, if $R = \mathbb{Z}_6$, then the quadratic polynomial $(x-1)(x-2) = x^2 + 3x + 2$ has roots $1, 2, 4$ and 5 in $\mathbb{Z}_6$.

(3.5) Notice that when we say that $f \in R[x]$, all we are claiming is that the ring R is big enough to contain the coefficients of f. So $x^2 + 1$ is equally at home in $\mathbb{Q}[x], \mathbb{R}[x]$ and $\mathbb{C}[x]$ (not to mention $\mathbb{Q}(i)[x] \ldots$).

This observation and the theorem mean that a polynomial has at most its degree number of roots *in any ring that contains its coefficients*. Put another way, we may become comfortable with the idea of creating "new" numbers to solve equations (for example, the creation of $\mathbb{C}$ to solve $x^2 + 1 = 0$), but there will always be a limit to our inventiveness–you will never find more than two solutions to $x^2 + 1 = 0$, now matter how many "new numbers" you make up.

Exercise 16 A polynomial like $x^2 + 2x + 1 = (x+1)^2$ has 1 as a *repeated root*. It's derivative, in the sense of elementary calculus, is $2(x+1)$, which also has 1 as a root. In general, and in light of the Factor Theorem, call $\lambda \in F$ a repeated root of f iff $f = (x-\lambda)^k g$ for some $k > 1$.

1. Using the formal derivative ∂ (see Exercise 15), show that λ is a repeated root of f if and only if λ is a root of $\partial(f)$.
2. Show that f has no repeated roots, ie: the roots of f are distinct, if and only if $\gcd(f, \partial(f)) = 1$.

(3.6) For reasons that will become clearer later, a very important role is played by polynomials that cannot be "factorised".

Definition. Let F be a field and $f \in F[x]$ a non-constant polynomial. A *non-trivial factorisation* of f is an expression of the form $f = gh$, where $g, h \in F[x]$ and $\deg g, \deg h \geq 1$. Say f is *reducible over* F iff it has a non-trivial factorisation, and *irreducible over* F otherwise.

Thus, a polynomial over a field F is irreducible precisely when it cannot be written as a product of non-constant polynomials. Another way of putting it is to say that $f \in F[x]$ is irreducible precisely when it is divisible only by a constant μ, or $\mu \times f$.

Aside. You can also talk about polynomials being irreducible over a ring (eg: over $\mathbb{Z}$). The definition is slightly more complicated however: let $f \in R[x]$ a non-constant polynomial with coefficients from the ring R. A *non-trivial factorisation* of f is an expression of the form $f = gh$, where $g, h \in R[x]$ and either,

1. $\deg g, \deg h \geq 1$, or
2. if say $g = \lambda \in R$ is a constant, then λ has *no* multiplicative inverse in R.

Say f is *reducible* over R iff it has a non-trivial factorisation, and *irreducible* over R otherwise. Notice that if $R = F$ a field, then the second possibility never arises, as every non-zero element of F has a multiplicative inverse.

The reason for the extra complication in the definition is that if $\lambda \in R$ is a constant which does have a multiplicative inverse in R, then you can always write

$$f = \lambda(\lambda^{-1} f).$$

So pulling out such constants is too easy! As an example, $3x + 3 = 3(x+1)$ is a non-trivial factorisation in $\mathbb{Z}[x]$ but a trivial one in $\mathbb{Q}[x]$.

The "over F" that follows reducible or irreducible is *crucial*; polynomials are never absolutely reducible or irreducible in any sense. An obvious example is $x^2 + 1$, which is irreducible over $\mathbb{R}$ but reducible over $\mathbb{C}$.

(3.7) There is an exception to this, and it is that a linear polynomial $f = ax + b \in F[x]$ is irreducible over any field F: if $f = gh$ then since $\deg f = 1$, we cannot have both $\deg(g), \deg(h) \geq 1$, for then $\deg(gh) = \deg(g) + \deg(h) \geq 1 + 1 = 2$, a contradiction. Thus, one of g or h must be a constant with f thus irreducible over F. So maybe we can qualify the statement above: linear polynomials are absolutely irreducible (we don't need to mention the field), but that's it!

Exercise 17

1. Let F be a field and $\lambda \in F$. Show that f is an irreducible polynomial over F if and only if λf is irreducible over F for any $\lambda \neq 0$.
2. Show that if $f(x + \lambda)$ is irreducible over F then $f(x)$ is too.

(3.8) There is the famous,

Fundamental Theorem of Algebra. *Any non-constant $f \in \mathbb{C}[x]$ has a root in $\mathbb{C}$.*

So if $f \in \mathbb{C}[x]$ has $\deg f \geq 2$, then f has a root in $\mathbb{C}$, hence a linear factor over $\mathbb{C}$, hence is reducible over $\mathbb{C}$. Thus, the only irreducible polynomials over $\mathbb{C}$ are the linear ones.

Aside. Actually, the fundamental theorem of algebra has been described as neither fundamental nor about algebra! Later we will be able to prove it from something known as the Galois correspondence, which also happens to be called the Fundamental Theorem of Galois Theory. Now, if you take the view that Galois theory is a subset of algebra, then it does seem rather odd that a theorem supposedly fundamental to all of algebra can be proved from a theorem that is merely fundamental to a *part* of it.

Exercise 18 Show that if f is irreducible over $\mathbb{R}$ then f is either linear or quadratic.

(3.9) A very common mistake is to think that having no roots in F is the same thing as being irreducible over F. In fact, the two are not even remotely the same thing.

Just because a polynomial is irreducible over F does not mean that it has no roots in the field: we saw above that a linear polynomial $ax + b$ is always irreducible, and yet has a root in F, namely $-b/a$. It is true though that if a polynomial f has degree ≥ 2 and had a root in F, then by the factor theorem it would have a linear factor so would be reducible. Thus, if $\deg(f) \geq 2$ and f is irreducible over F, then f has no roots in F.

A polynomial that has no roots in F is not necessarily irreducible over the field: the polynomial $x^4 + 2x^2 + 1 = (x^2 + 1)^2$ is reducible over $\mathbb{Q}$, but with roots $\pm i \notin \mathbb{Q}$.

(3.10) There is no general method for deciding if a polynomial over an arbitrary field F is irreducible: the situation is not dissimilar to that of integration in calculus. There is no list of rules that collectively apply to all situations. The best we can hope for is an ever expanding list of techniques, of which this is the first:

Proposition 1 *Let F be a field and $f \in F[x]$ be a polynomial of degree ≤ 3. If f has no roots in F then it is irreducible over F.*

Proof: Arguing by contradiction, if f is reducible then $f = gh$ with $\deg g, \deg h \geq 1$. Since $\deg g + \deg h = \deg f \leq 3$, we must have for g say, that $\deg g = 1$. Thus $f = (ax + b)h$ and f has the root $(-b \times a^{-1})$. □

Exercise 19 We need a new field to play with. Let p be a prime and $\mathbb{F}_p$ the set $\{0, 1 \ldots, p-1\}$. Define addition and multiplication on this set to be addition and multiplication of integers modulo p.

1. Verify that $\mathbb{F}_p$ is a field by checking the axioms. The only tricky one is the existence of inverses under multiplication: use the gcd theorem from §**2.** (but for $\mathbb{Z}$ rather than polynomials).
2. Show that a field F is an integral domain. Hence, show that if n is *not* prime, then the addition and multiplication of integers modulo n is *not* a field.

(3.11) Consider polynomials with coefficients from, say, $\mathbb{F}_2$, ie: the ring $\mathbb{F}_2[x]$, and in particular, the polynomial

$$f = x^4 + x + 1 \in \mathbb{F}_2[x].$$

Now, $0^4 + 0 + 1 \neq 0 \neq 1^4 + 1 + 1$, so f has no roots in $\mathbb{F}_2$. Although this is a good start, we are in no position to finish, as the Proposition above does not apply to quartics. But we can certainly say that any factorisation of f over $\mathbb{F}_2$, if there is one, must be as a product of two quadratics. Moreover, these quadratics must themselves be irreducible over $\mathbb{F}_2$, for if not, they would factor into linear factors and the factor theorem would give roots of f.

There are only four quadratics over $\mathbb{F}_2$:

$$x^2, x^2 + 1, x^2 + x \text{ and } x^2 + x + 1.$$

The first two are reducible as they have roots 0 and 1 respectively; the third is also reducible with both 0 and 1 as roots. By the Proposition above, the last is irreducible. Thus, any factorisation of f into irreducible quadratics must in fact be of the form,

$$(x^2 + x + 1)(x^2 + x + 1).$$

Unfortunately, f *doesn't* factorise this way (just expand the brackets). Thus f is irreducible over $\mathbb{F}_2$.

(3.12) As we delve deeper into Galois theory, it will transpire that $\mathbb{Q}$ is where much of the action happens. Consequently, determining the irreducibility of polynomials over $\mathbb{Q}$ will be of great importance. The first useful test for irreducibility over $\mathbb{Q}$ has the following main ingredient: *to see if a polynomial can be factorised over $\mathbb{Q}$ it suffices to see whether it can be factorised over $\mathbb{Z}$.*

First we recall Exercise 14, which is used a number of times in these notes so is worth placing in a,

Lemma 2 *Let $\sigma : R \to S$ be a homomorphism of rings. Define $\sigma^* : R[x] \to S[x]$ by*

$$\sigma^* : \sum_i a_i x^i \mapsto \sum_i \sigma(a_i) x^i.$$

Then σ^* is a homomorphism.

Lemma 3 (Gauss) *Let f be a polynomial with integer coefficients. Then f can be factorised non-trivially as a product of polynomials with integer coefficients if and only if it can be factorised non-trivially as a product of polynomials with rational coefficients.*

Proof: If the polynomial can be written as a product of $\mathbb{Z}$-polynomials then it clearly can as a product of $\mathbb{Q}$-polynomials as integers are rational! Suppose on the otherhand that $f = gh$ in $\mathbb{Q}[x]$ is a non-trivial factorisation. By multiplying through by a multiple of the denominators of the coefficients of g we get a polynomial $g_1 = mg$ with $\mathbb{Z}$-coefficients. Similarly we have $h_1 = nh \in \mathbb{Z}[x]$ and so

$$mnf = g_1 h_1 \in \mathbb{Z}[x]. \tag{2}$$

Now let p be a prime dividing mn, and consider the homomorphism $\sigma : \mathbb{Z} \to \mathbb{F}_p$ given by $\sigma(k) = k \mod p$. Then by the lemma above, the map $\sigma^* : \mathbb{Z}[x] \to \mathbb{F}_p[x]$ given by

$$\sigma^* : \sum_i a_i x^i \mapsto \sum_i \sigma(a_i) x^i,$$

is a homomorphism. Applying the homomorphism to (2) gives $0 = \sigma^*(g_1)\sigma^*(h_1)$ in $\mathbb{F}_p[x]$, as $mn \equiv 0 \mod p$. As the ring $\mathbb{F}_p[x]$ is an integral domain the only way that this can happen is if one of the polynomials is equal to the zero polynomial in $\mathbb{F}_p[x]$, ie: one of the original polynomials, say g_1, has all of its coefficients divisible by p. Thus we have $g_1 = pg_2$ with $g_2 \in \mathbb{Z}[x]$, and (2) becomes

$$\frac{mn}{p} f = g_2 h_1.$$

Working our way through all the prime factors of mn in this way, we can remove the factor of mn from (2) and obtain a factorisation of f into polynomials with $\mathbb{Z}$-coefficients. □

So to determine whether a polynomial with $\mathbb{Z}$-coefficients is *irreducible* over $\mathbb{Q}$, one need only check that it has no non-trivial factorisations with all the coefficients integers.

Eisenstein Irreducibility Theorem. *Let*

$$f = c_n x^n + \cdots + c_1 x + c_0,$$

be a non-linear polynomial with integer coefficients. If there is a prime p that divides all the c_i for $i < n$, does not divide c_n and such that p^2 does not divide c_0, then f is irreducible over $\mathbb{Q}$.

Proof: By virtue of the fact above, we need only show that under the conditions stated, there is no factorisation of f using integer coefficients. Suppose otherwise, ie: $f = gh$ with

$$g = a_r x^r + \cdots + a_0 \text{ and } h = b_s x^s + \cdots + b_0,$$

and the $a_i, b_i \in \mathbb{Z}$. Expanding gh and equating coefficients,

$$\begin{aligned} c_0 &= a_0 b_0 \\ c_1 &= a_0 b_1 + a_1 b_0 \\ &\vdots \\ c_i &= a_0 b_i + a_1 b_{i-1} + \cdots + a_i b_0 \\ &\vdots \\ c_n &= a_r b_s. \end{aligned}$$

By hypothesis, $p \,|\, c_0$. Write both a_0 and b_0 as a product of primes, so if $p \,|\, c_0$, ie: $p \,|\, a_0 b_0$, then p must be one of the primes in this factorisation, hence divides one of a_0 or b_0. Thus, either $p \,|\, a_0$ or $p \,|\, b_0$, *but not both* (for then p^2 would divide c_0). Assume that it is $p \,|\, a_0$ that we have. Next, $p \,|\, c_1$, and this coupled with $p \,|\, a_0$ gives $p \,|\, c_1 - a_0 b_1 = a_1 b_0$ (If we had assumed $p \,|\, b_0$, we would still reach this conclusion). Again, p must divide one of the these last two factors, and since we've already decided that it doesn't divide b_0, it must be a_1 that it divides. Continuing in this manner, we get that p divides all the coefficients of g, and in particular, a_r. But then p divides $a_r b_s = c_n$, the contradiction we were after. □

As a meta-mathematical comment, the proof of Eisenstein irreducibility is a nice example of the manner in which mathematics is created. You start with as few assumptions as possible (in this case that p divides some of the coefficients of f) and proceed towards some sort of conclusion, imposing extra conditions as and when you need them. In this way the correct statement of the theorem writes itself in an organic fashion.

(3.13) To show the power of the result, we get immediately that

$$x^4 - 5x^3 + 10x^2 + 25x - 35,$$

is irreducible over $\mathbb{Q}$, a fact not easily shown another way. Even more useful, we have

$$x^n - p,$$

is irreducible over $\mathbb{Q}$ for any prime p. Thus, we can find polynomials over $\mathbb{Q}$ of arbitrary large degree that are irreducible, which is to be contrasted strongly with the situation for polynomials over $\mathbb{R}$ or $\mathbb{C}$.

(3.14) It turns out that there is a fundamental connection between the multitude of irreducible polynomials over $\mathbb{Q}$ (and the relative paucity of them over $\mathbb{R}$ and $\mathbb{C}$) and the empirical observation that there are lots of fields a "little bigger" than $\mathbb{Q}$ (for example, $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\alpha, \omega)$ from §**1.**), but very few fields a "little bigger" than $\mathbb{R}$ or $\mathbb{C}$.

(3.15) Another useful tool arises when you have polynomials with coefficients from some ring R and a homomorphism from R to some field F. If the homomorphism is applied to all the coefficients of the polynomial (turning it from a polynomial with R-coefficients into a polynomial with F-coefficients), then a *reducible polynomial cannot turn into an irreducible one.* The precise statement goes by the name of:

The Reduction Test. *Let R be an integral domain, F a field and $\sigma : R \to F$ a ring homomorphism. Let $\sigma^* : R[x] \to F[x]$ be given by*

$$\sigma^* : \sum_i a_i x^i \mapsto \sum_i \sigma(a_i) x^i.$$

be the homomorphism of Lemma 2. Moreover, let $p \in R[x]$ be such that

1. $\deg \sigma^*(p) = \deg(p)$, *and*

2. $\sigma^*(p)$ *is irreducible over* F.

Then p *cannot be written as a product* gh *with* $g, h \in R[x]$ *and* $\deg g, \deg h < \deg p$.

Although it is stated in some generality, the reduction test is very useful for determining the irreducibility of polynomials over $\mathbb{Q}$. As an example, take $R = \mathbb{Z}$; $F = \mathbb{F}_5$ and $p = 8x^3 - 6x - 1 \in \mathbb{Z}[x]$. For σ, take reduction modulo 5, ie: $\sigma(n) = n \bmod 5$. It is not hard to show that σ is a homomorphism. Since $\sigma(8) = 3 \bmod 5$, and so on, we get

$$\sigma^*(p) = 3x^3 + 4x + 4 \in \mathbb{F}_5[x].$$

Clearly, the degree has not changed, and by substituting the five elements of $\mathbb{F}_5$ into $\sigma^*(p)$, one can see that it has no roots in $\mathbb{F}_5$. Since the polynomial is a cubic, it must therefore be irreducible over $\mathbb{F}_5$. Thus, by the reduction test, $8x^3 - 6x - 1$ cannot be written as a product of smaller degree polynomials with $\mathbb{Z}$-coefficients. But by Gauss' lemma, this gives that this polynomial is irreducible over $\mathbb{Q}$.

$\mathbb{F}_5$ was chosen because with $\mathbb{F}_2$ instead, condition (i) fails; with $\mathbb{F}_3$, condition (ii) fails.

Proof: Suppose on the contrary that $p = gh$ with $\deg g, \deg h < \deg p$. Then $\sigma^*(p) = \sigma^*(gh) = \sigma^*(g)\sigma^*(h)$, the last part because σ^* is a homomorphism. Now $\sigma^*(p)$ is irreducible, so the only way it can factorise like this is if one of the factors, $\sigma^*(g)$ say, is a constant, hence $\deg \sigma^*(g) = 0$. Then

$$\deg p = \deg \sigma^*(p) = \deg \sigma^*(g)\sigma^*(h) = \deg \sigma^*(g) + \deg \sigma^*(h) = \deg \sigma^*(h) \leq \deg h < \deg p,$$

a contradiction. That $\deg \sigma^*(h) \leq \deg h$ rather than equality necessarily, is because the homomorphism σ may send some of the coefficients of h (including quite possibly the leading one) to $0 \in F$. □

(3.16) Our final tool requires a little more set-up. We've already observed the similarity between polynomials and integers. The idea of irreducibility in $\mathbb{Z}$ is just that of a prime number, and perhaps this goes some way to indicating its importance for polynomials as well. One thing we know about integers is that they can be written uniquely as a product of primes. We would hope that something similar is true for polynomials, and it is in certain situations. For the next few results, we deal only with polynomials $f \in F[x]$ for F a field (they are actually true in more generality, but this is beyond the scope of these notes). In what follows, it is worth comparing the situation with what you know about $\mathbb{Z}$.

Lemma 4 1. *If* $\gcd(f, g) = 1$ *and* $f \mid gh$ *then* $f \mid h$.

2. *If* f *is irreducible and monic, then for any* g *monic with* $g \mid f$ *we have either* $g = 1$ *or* $g = f$.

3. *If* g *is irreducible and monic and* g *does not divide* f*, then* $\gcd(g, f) = 1$.

4. *If* g *is irreducible and monic and* $g \mid f_1 f_2 \dots f_n$ *then* $g | f_i$ *for some* i.

Proof: 1. Since $\gcd(f, g) = 1$ there are $a, b \in F[x]$ such that $1 = af + bg$, hence $h = afh + bgh$. We have that $f \mid bgh$ by assumption, and it clearly divides afh, hence it divides $afh + bgh = h$ also.

2. If g divides f and f is irreducible, then by definition g must be either a constant or a constant multiple of f. But f is monic, so $g = 1$ or $g = f$ are the only possibilities.

3. The $\gcd$ of f and g is certainly a divisor of g, and hence by irreducibility must be either a constant, or a constant times g. As g is also monic, the gcd must in fact be either 1 or g itself, and since g does not divide f it cannot be g, so must be 1.

4. Proceed by induction, with the first step for $n = 1$ being immediate. Since $g \mid f_1 f_2 \dots f_n = (f_1 f_2 \dots f_{n-1}) f_n$, we either have $g \mid f_n$, in which case we are finished, or not, in which case $\gcd(g, f_n) = 1$ by part (3). But then part (1) gives that $g \mid f_1 f_2 \dots f_{n-1}$, and the inductive hypothesis kicks in.

□

Perhaps the best way of summarising the lemma is this: monic irreducible polynomials are like the "prime numbers" of $F[x]$.

(3.17) And just as any integer can be decomposed uniquely as a product of primes, so too can any polynomial as a product of irreducible polynomials:

Unique factorisation in $F[x]$. *Every polynomial in $F[x]$ can be written in the form*

$$\lambda p_1 p_2 \dots p_r,$$

where λ is a constant and the p_i are monic and irreducible $\in F[x]$. Moreover, if $\mu q_1 q_2 \dots q_s$ is another factorisation with the q_j monic and irreducible, then $r = s$, $\lambda = \mu$ and the q_j are just a rearrangement of the p_i.

The last part says that the factorisation is unique, except for trivial matters like the order you write down the factors. Like many such results in mathematics, the first impression is that the existence of the factorisation is the useful part, but in fact it is the uniqueness that really is.

Proof: To get the factorisation in the first place is easy enough: just keep factorising reducible polynomials until they become irreducible. At the end, pull out the coefficient of the leading term in each factor, and place them all at the front.

For uniqueness, suppose that

$$\lambda p_1 p_2 \dots p_r = \mu q_1 q_2 \dots q_s.$$

Then p_r divides $\mu q_1 q_2 \dots q_s$ which by Lemma 4 part (4) means that $p_r \mid q_i$ for some i. Reorder the q's so that it is $p_r \mid q_s$ that in fact we have. Since both p_r and q_s are monic, irreducible, and hence non-constant, $p_r = q_s$, which leaves us with

$$\lambda p_1 p_2 \dots p_{r-1} = \mu q_1 q_2 \dots q_{s-1}.$$

This gives $r = s$ straight away: if say $s > r$, then repetition of the above leads to $\lambda = \mu q_1 q_2 \dots q_{s-r}$, which is absurd, as consideration of degrees gives different answers for each side. Similarly if $r > s$. But then we also have that the p's are just a rearrangement of the q's, and canceling down to $\lambda p_1 = \mu q_1$, that $\lambda = \mu$. □

(3.18) It is worth repeating that everything depends on the ambient field F, even the uniqueness of the decomposition. For example, $x^4 - 4$ decomposes as,

$$\begin{array}{l}(x^2+2)(x^2-2) \text{ in } \mathbb{Q}[x],\\ (x^2+2)(x-\sqrt{2})(x+\sqrt{2}) \text{ in } \mathbb{R}[x] \text{ and}\\ (x-\sqrt{2}i)(x+\sqrt{2}i)(x-\sqrt{2})(x+\sqrt{2}) \text{ in } \mathbb{C}[x].\end{array}$$

To illustrate how unique factorisation can be used to determine irreducibility, we have in $\mathbb{C}[x]$ that,

$$x^2 + 2 = (x - \sqrt{2}i)(x + \sqrt{2}i).$$

Since the factors on the right are not in $\mathbb{R}[x]$ we have an inkling that this polynomial is irreducible over $\mathbb{R}$. To make this more precise, any factorisation in $\mathbb{R}[x]$ would be of the form

$$x^2 + 2 = (x - \lambda_1)(x - \lambda_2)$$

with the $\lambda_i \in \mathbb{R}$. But this would be a factorisation in $\mathbb{C}[x]$ too, and there is only one such by unique factorisation. This forces the λ_i to be $\sqrt{2}i$ and $-\sqrt{2}i$, contradicting $\lambda_i \in \mathbb{R}$. Hence $x^2 + 2$ is indeed irreducible over $\mathbb{R}$. Similarly, $x^2 - 2$ is irreducible over $\mathbb{Q}$.

Exercise 20 Formulate the example above into a general Theorem.

Further Exercises for §**3.**

Exercise 21 Prove that if a polynomial equation has all its coefficients in $\mathbb{C}$ then it must have all its roots in $\mathbb{C}$.

Exercise 22

1. Let $f = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial in $\mathbb{R}[x]$, that is, all the $a_i \in \mathbb{R}$. Show that complex roots of f occur in conjugate pairs, ie: $\zeta \in \mathbb{C}$ is a root of f if and only if $\bar{\zeta}$ is.
2. Find an example of a polynomial in $\mathbb{C}[x]$ for which part (a) is not true.

Exercise 23

1. Let m, n and k be integers with m and n relatively prime (ie: $\gcd(m, n) = 1$). Show that if m divides nk then m must divide k (*hint*: there are two methods here. One is to use Lemma 4 but in $\mathbb{Z}$. The other is to use the fact that any integer can be written uniquely as a product of primes. Do this for m and n, and ask yourself what it means for this factorisation that m and n are relatively prime).
2. Show that if m/n is a root of $a_0 + a_1 x + ... + a_r x^r$, $a_i \in \mathbb{Z}$, where m and n are relatively prime integers, then $m|a_0$ and $n|a_r$ (*hint*: use the first part!).
3. Deduce that if $a_r = 1$ then m/n is in fact an integer.

moral: If a monic polynomial with integer coefficients has a rational root m/n, then this rational number is in fact an integer.

Exercise 24 If $m \in \mathbb{Z}$ is not a perfect square, show that $x^2 - m$ is irreducible over $\mathbb{Q}$ (note: it is *not* enough to merely assume that under the conditions stated $\sqrt{m}$ is not a rational number).

Exercise 25 Find the greatest common divisor of $f(x) = x^3 - 6x^2 + x + 4$ and $g(x) = x^5 - 6x + 1$ (*hint*: look at linear factors of $f(x)$).

Exercise 26 Determine which of the following polynomials are irreducible over the stated field:

1. $1 + x^8$ over $\mathbb{R}$;
2. $1 + x^2 + x^4 + x^6 + x^8 + x^{10}$ over $\mathbb{Q}$ (**hint**: Let $y = x^2$ and factorise $y^n - 1$);
3. $x^4 + 15x^3 + 7$ over $\mathbb{R}$ (**hint**: use the intermediate value theorem from analysis);
4. $x^{n+1} + (n+2)!\, x^n + \cdots + (i+2)!\, x^i + \cdots + 3!\, x + 2!$ over $\mathbb{Q}$.
5. $x^2 + 1$ over $\mathbb{F}_7$.
6. Let $\mathbb{F}$ be the field of order 8 from §**4.**, and let $\mathbb{F}[X]$ be polynomials with coefficients from $\mathbb{F}$ and indeterminate X. Is $X^3 + (\alpha^2 + \alpha)X + (\alpha^2 + \alpha + 1)$ irreducible over $\mathbb{F}$?
7. $a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$ over $\mathbb{Q}$ where the $a_i \in \mathbb{Z}$; a_3, a_2 are even and a_4, a_1, a_0 are odd.

Exercise 27 If p is a prime integer, prove that p is a divisor of $\begin{pmatrix} p \\ i \end{pmatrix}$, for $0 < i < p$.

Exercise 28 Show that

$$x^{p-1} + px^{p-2} + \cdots + \begin{pmatrix} p \\ i \end{pmatrix} x^{p-i-1} + \cdots + p,$$

is irreducible over $\mathbb{Q}$.

Exercise 29 A complex number ω is an n-th *root of unity* if $\omega^n = 1$. It is a *primitive* n-th root of unity if $\omega^n = 1$, but $\omega^r \neq 1$ for any $0 < r < n$. So for example, $\pm 1, \pm i$ are the 4-th roots of 1, but only $\pm i$ are primitive 4-th roots.

Convince yourself that for any n,

$$\omega = \cos\frac{2\pi}{n} + i \sin\frac{2\pi}{n}$$

is an n-th root of 1. In fact, the other n-th roots are $\omega^2, \ldots, \omega^n = 1$.

1. Show that if ω is a *primitive* n-th root of 1 then ω is a root of the polynomial
$$x^{n-1} + x^{n-2} + \cdots + x + 1. \tag{3}$$
2. Show that for (3) to be irreducible over $\mathbb{Q}$, n cannot be even.
3. Show that a polynomial $f(x)$ is irreducible over a field F if $f(x+1)$ is irreducible over F.
4. Finally, if
$$\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$$
for p a prime number, show that $\Phi_p(x+1)$ is irreducible over $\mathbb{Q}$, and hence $\Phi_p(x)$ is too (*hint*: consider $x^p - 1$ and use the binomial theorem, Exercise 27 and Eisenstein).

The polynomial $\Phi_p(x)$ is called the *p-th cyclotomic polynomial*.

§4. Fields I: Basics, extensions and concrete examples

(4.1) This course is primarily the study of solutions to polynomial equations. Broadly speaking, questions in this direction can be restated as questions about fields. It is to these that we now turn.

(4.2) We remembered the definition of a field in Lecture §**1.**. Since then we have become more familar with rings, so we can restate the definition as:

Definition. A *field* is a set F with two operations, $\oplus$ and $\otimes$, such that for any $a, b, c \in F$,

1. F is an Abelian group under $\oplus$ (with $\oplus$ normally called just $+$, e called 0, and a^{-1} called $-a$),
2. $F \setminus \{0\}$ is an Abelian group under $\otimes$ (with $\otimes$ normally called just $\times$, and f called 1),
3. the two operations are linked by the distributive law.

The two groups are called the *additive* and *multiplicative* groups of the field. In particular, we will write F^* to denote the multiplicative group (ie: F^* is the group with elements $F \setminus \{0\}$ and operation the multiplication from the field). Even more succinctly,

Definition. A *field* is a set F with two operations, $\oplus$ and $\otimes$, such that for any $a, b, c \in F$,

1. F is a commutative ring under $\oplus$ and $\otimes$ (with $\oplus$ normally called just $+$, e called 0, inverses under $\oplus$ called $-a$, $\otimes$ just $\times$, and f called 1),
2. for any $a \in F \setminus \{e\}$ there is an $a^{-1} \in F$ with $a \otimes a^{-1} = f = a^{-1} \otimes a$,

In particular, *a field is a very special kind of ring*.

(4.3) More concepts from the first lecture that can now be properly defined are:

Definition. Let F and E be fields with F a subfield of E. We call E an *extension* of F. The standard notation for an extension is to write E/F, but in these notes we will use the more concrete $F \subseteq E$, being mindful at all times that this means F is a subfield of E, and not just a subset.

If $\beta \in E$, we write, as in §**1.**, $F(\beta)$ for the smallest subfield of E containing both F and β (so in particular $F(\beta)$ is an extension of F). In general, if $\beta_1, \ldots, \beta_k \in E$, define $F(\beta_1, \ldots, \beta_k) = F(\beta_1, \ldots, \beta_{k-1})(\beta_k)$.

We say that β has is *adjoined* to F to obtain $F(\beta)$. The last bit of the definition just says that to adjoin several elements to a field, you just adjoin them one at a time[4]. Finally, if we have an extension $F \subset E$ and there is a $\beta \in E$ such that $E = F(\beta)$, then we call E a *simple extension* of F.

(4.4) Trivially, $\mathbb{R}$ is an extension of $\mathbb{Q}$; $\mathbb{C}$ is an extension of $\mathbb{R}$, and so on. Any field is equally trivially an extension of itself!

(4.5) Let $\mathbb{F}_2$ be the field of integers modulo 2 arithmetic. Let α be an "abstract symbol" that can be multiplied so that it has the following property: $\alpha \times \alpha \times \alpha = \alpha^3 = \alpha + 1$ (a bit like decreeing that the imaginary i squares to give -1). Let

$$\mathbb{F} = \{a + b\alpha + c\alpha^2 \,|\, a, b, c \in \mathbb{F}_2\},$$

Define addition on $\mathbb{F}$ by: $(a_1 + b_1\alpha + c_1\alpha^2) + (a_2 + b_2\alpha + c_2\alpha^2) = (a_1 + a_2) + (b_1 + b_2)\alpha + (c_1 + c_2)\alpha^2$, where the addition of coefficients happens in $\mathbb{F}_2$. For multiplication, "expand" the expression $(a_1 + b_1\alpha + c_1\alpha^2)(a_2 + b_2\alpha + c_2\alpha^2)$ like you would a polynomial with α the indeterminate, so that $\alpha\alpha\alpha = \alpha^3$, the coefficients are dealt with using the arithmetic from $\mathbb{F}_2$, and so on. Replace any α^3 that result using the rule $\alpha^3 = \alpha + 1$.

[4] Although the definition has you adjoining them in a particular order, the order doesn't matter.

For example,

$$(1+\alpha+\alpha^2)+(\alpha+\alpha^2)=1 \text{ and } (1+\alpha+\alpha^2)(\alpha+\alpha^2)=\alpha+\alpha^4=\alpha+\alpha(\alpha+1)=\alpha^2.$$

It turns out that $\mathbb{F}$ forms a field with this addition and multiplication, see Exercise 40. For now we content ourselves with the following observation: taking those elements of $\mathbb{F}$ with $b=c=0$, we obtain (an isomorphic) copy of $\mathbb{F}_2$ inside of $\mathbb{F}$.

Thus, we have an extension of $\mathbb{F}_2$ that contains 8 elements.

(4.6) Certainly, $\mathbb{Q}(\sqrt{2})$ is a simple extension of $\mathbb{Q}$. On the other hand, $\mathbb{Q}(\sqrt{2},\sqrt{3})$ would appear not to be; but looking at the definition closely you see that a simple extension is one *that can be obtained* by adjoining one element.

Consider now $\mathbb{Q}(\sqrt{2}+\sqrt{3})$: certainly $\sqrt{2}+\sqrt{3}\in\mathbb{Q}(\sqrt{2},\sqrt{3})$, and so $\mathbb{Q}(\sqrt{2}+\sqrt{3})\subset\mathbb{Q}(\sqrt{2},\sqrt{3})$. On the other hand,

$$(\sqrt{2}+\sqrt{3})^3=11\sqrt{2}+9\sqrt{3},$$

as is readily checked using the Binomial Theorem. Since $(\sqrt{2}+\sqrt{3})^3\in\mathbb{Q}(\sqrt{2}+\sqrt{3})$, we get

$$(11\sqrt{2}+9\sqrt{3})-9(\sqrt{2}+\sqrt{3})\in\mathbb{Q}(\sqrt{2}+\sqrt{3})\Rightarrow 2\sqrt{2}\in\mathbb{Q}(\sqrt{2}+\sqrt{3}).$$

And so $\sqrt{2}\in\mathbb{Q}(\sqrt{2}+\sqrt{3})$ as $\dfrac{1}{2}$ is there too. Similarly it can be shown that $\sqrt{3}\in\mathbb{Q}(\sqrt{2}+\sqrt{3})$. The upshot is that $\mathbb{Q}(\sqrt{2},\sqrt{3})\subset\mathbb{Q}(\sqrt{2}+\sqrt{3})$. So $\mathbb{Q}(\sqrt{2},\sqrt{3})$ *is* a simple extension! It didn't appear to be as we hadn't written it the right way. We will see more precisely at the end of §**9.** when extensions are simple.

(4.7) What do the elements of the field $\mathbb{Q}(\sqrt{2})$ actually look like? Later we will be answer this question in a general and completely satisfactory manner, but for now we can feel our way towards an ad-hoc answer.

Certainly $\sqrt{2}$ and any $b\in\mathbb{Q}$ are in $\mathbb{Q}(\sqrt{2})$ by definition. Since fields are closed under $\times$, any number of the form $b\sqrt{2}\in\mathbb{Q}(\sqrt{2})$. Similarly, fields are closed under $+$, so any $a+b\sqrt{2}\in\mathbb{Q}(\sqrt{2})$ for $a\in\mathbb{Q}$. Thus, the set

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

But $\mathbb{F}$ is a field in its own right using the usual addition and multiplication of complex numbers. This is easily checked from the axioms; for instance, the inverse of $a+b\sqrt{2}$ can be calculated:

$$\frac{1}{a+b\sqrt{2}}\times\frac{a-b\sqrt{2}}{a-b\sqrt{2}}=\frac{a-b\sqrt{2}}{a^2-2b^2}=\frac{a}{a^2-2b^2}-\frac{b}{a^2-2b^2}\sqrt{2}\in\mathbb{F},$$

and you can check the other axioms for yourself. We also have $\mathbb{Q}\subset\mathbb{F}$ (letting $b=0$) and $\sqrt{2}\in\mathbb{F}$ (letting $a=0,b=1$). Since $\mathbb{Q}(\sqrt{2})$ is the smallest field having these two properties, we have $\mathbb{Q}(\sqrt{2})\subseteq\mathbb{F}$. Thus,

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

Exercise 30 Let α be a complex number such that $\alpha^3=1$ and consider the set

$$\mathbb{F}=\{a_0+a_1\alpha+a_2\alpha^2\,|\,a_i\in\mathbb{Q}\}$$

1. By row reducing the matrix,
$$\begin{pmatrix} a_0 & 2a_2 & 2a_1 & 1\\ a_1 & a_0 & 2a_2 & 0\\ a_2 & a_1 & a_0 & 0\end{pmatrix}$$
find an element of $\mathbb{F}$ that is the inverse under multiplication of $a_0+a_1\alpha+a_2\alpha^2$.
2. Show that $\mathbb{F}$ is a field, hence $\mathbb{Q}(\alpha)=\mathbb{F}$.

(4.8) The previous exercise shows that the following two fields have the form,

$$\mathbb{Q}(\sqrt[3]{2}) = \{a + b\sqrt[3]{2} + c\sqrt[3]{2}^2 \mid a, b, c \in \mathbb{Q}\} \text{ and } \mathbb{Q}(\beta) = \{a + b\beta + c\beta^2 \mid a, b, c \in \mathbb{Q}\},$$

where

$$\beta = \frac{-\sqrt[3]{2} + \sqrt[3]{2}\sqrt{3}i}{2} \in \mathbb{C}.$$

Observe for now that these two fields are different. The first is clearly completely contained in $\mathbb{R}$, but the second contains β, which is obviously complex but not real.

(4.9) A bijective homomorphism of rings $\varphi : R \to S$ is called an *isomorphism*.

A silly but instructive example is given by the Roman ring, whose elements are

$$\{\ldots, -V, -IV, -III, -II, -I, 0, I, II, III, IV, V, \cdots\},$$

and with addition and mutiplication giving such things as $IX + IV = XIII$ and $IX \times VI = LIV, \ldots$ Obviously the ring is isomorphic to $\mathbb{Z}$, and it is this idea of a trivial relabelling that is captured by the idea of an isomorphism–two rings are isomorphic if they are really the same, just written in different languages! The translation is carried out by the mapping φ.

It seems a sensible enough idea, but we place a huge emphasis on the way things are labelled, often without even realising that we are doing it. The two fields above are a good example, for,

$$\mathbb{Q}(\sqrt[3]{2}) \text{ and } \mathbb{Q}\left(\frac{-\sqrt[3]{2} + \sqrt[3]{2}\sqrt{3}i}{2}\right) \text{ are isomorphic!}$$

(we'll see why in §**6.**). To illustrate how we might now come unstuck, suppose we were to formulate the following,

"Definition". *A subfield of* $\mathbb{C}$ *is called* real *if and only if it is contained in* $\mathbb{R}$.

So $\mathbb{Q}(\sqrt[3]{2})$ is a real field, but $\mathbb{Q}\left(\frac{-\sqrt[3]{2} + \sqrt[3]{2}\sqrt{3}i}{2}\right)$ is not. But they are the same field! A definition should not depend on the way the elements are labelled. The problem is that we have become too bogged down in the minutiae of real and complex numbers and we need to think about fields in a more abstract way.

(4.10) The previous example has motivated the direction of the next few sections. In the remainder of this section we introduce a few more concepts associated with fields.

It is well known that $\sqrt{2}$ and π are both irrational real numbers. Nevertheless, from an algebraic point of view, $\sqrt{2}$ is slightly more tractable than π, as it is a root of a very simple equation $x^2 - 2$, whereas there is no polynomial with integer coefficients having π as a root (this is not obvious).

Let $F \subseteq E$ be an extension of fields and $\alpha \in E$. Call α *algebraic over* F if and only if

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n = 0,$$

for some $a_0, a_1, \ldots, a_n \in F$. In otherwords, α is a root of the polynomial $f = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ in $F[x]$. If α is not algebraic, ie: not the root of any polynomial with F-coefficients, then we say that it is *transcendental* over F.

As the story of Galois theory develops, we will see that it is the algebraic elements over F that are the most easily understood. It is tempting to think of them as having expressions in terms of elements of F, the four field operations $+, -, \times, \div$ and roots $\sqrt{}, \sqrt[3]{}, \ldots, \sqrt[n]{}, \ldots$, but as we shall see in §**16.**, the situation is much more subtle than that. Indeed there are algebraic numbers that cannot be expressed *algebraically*. For now it is best just to stick to the definition and not read too much into it.

(4.11) Some simple examples:

$$\sqrt{2},\ \frac{1+\sqrt{5}}{2} \text{ and } \sqrt[5]{\sqrt{2}+5\sqrt[3]{3}},$$

are algebraic over $\mathbb{Q}$, whereas π and e are transcendental over $\mathbb{Q}$; π is however algebraic over $\mathbb{Q}(\pi)$.

(4.12) A field can obviously contain many subfields: if we look at $\mathbb{C}$, it contains $\mathbb{Q}(\sqrt{2}), \mathbb{R}, \ldots$. It also contains $\mathbb{Q}$, *but no subfields that are smaller than this*, in the usual sense that they are properly contained in $\mathbb{Q}$. Indeed, *any* subfield of $\mathbb{C}$ contains $\mathbb{Q}$. So, $\mathbb{Q}$ is the "smallest" subfield of the complex numbers.

For any field F, the *prime subfield* $\mathbb{F}$ of F is the smallest subfield of F in the sense that if F' is any subfield with $F' \subseteq \mathbb{F}$ then $F' = \mathbb{F}$.

Exercise 31 Show that the prime subfield can also be defined as the intersection of all the subfields of F. Thus in particular, the prime subfield is *contained in every subfield of* F.

Exercise 32 Consider the field of rational numbers $\mathbb{Q}$ or the finite field $\mathbb{F}_p$ having p elements. Show that neither of these fields contain a proper subfield (hint: for $\mathbb{F}_p$, consider the additive group and use Lagrange's Theorem from §**11.**. For $\mathbb{Q}$, any subfield must contain 1, and show that it must then be all of $\mathbb{Q}$).

Whatever the prime subfield is, it must contain 1, hence any expression of the form $1+1+\cdots+1$ for any number of summands. If no such expression equals the 0 in the field, then we have infinitely many distinct such elements, and their inverses under addition, so what we have is basically a copy of $\mathbb{Z}$ in F. Otherwise, if n is the smallest number of summands for which such an expression is equal to 0, then the elements

$$1, 1+1, 1+1+1, \ldots, \underbrace{1+1+\cdots+1}_{n \text{ times}} = 0,$$

forms a copy of $\mathbb{Z}_n$ inside of F.

These comments can be made precise as in the following exercise. It looks ahead a little, requiring the first isomorphism theorem for rings in §**5.**

Exercise 33 Let F be a field and define a map $\mathbb{Z} \to F$ by

$$n \mapsto \begin{cases} 0, \text{ if } n = 0, \\ 1+\cdots+1, (n \text{ times}), \text{ if } n > 0 \\ -1-\cdots-1, (n \text{ times}), \text{ if } n < 0. \end{cases}$$

Show that the map is a homomorphism. If the kernel consists of just $\{0\}$, then show that F contains $\mathbb{Z}$ as a subring. Otherwise, let n be the smallest positive integer contained in the kernel, and show that F contains $\mathbb{Z}_n$ as a subring. As F is a field, hence an integral domain, show that we must have $n = p$ a prime in this situation.

Thus *any field contains a subring isomorphic to* $\mathbb{Z}$ *or to* $\mathbb{Z}_p$ *for some prime* p. But the ring $\mathbb{Z}_p$ is the field $\mathbb{F}_p$, and we saw in Exercise 32 that $\mathbb{F}_p$ contains no subfields. The conclusion is that in the second case above, the prime subfield is this copy of $\mathbb{F}_p$. In the first case, $\mathbb{Z}$ is obviously not a field, but each m in this copy of $\mathbb{Z}$ has an inverse $1/m$ in F, and the product of this with any other n gives an element $m/n \in F$. The set of all such elements obtained is a copy of $\mathbb{Q}$ inside F.

Exercise 34 Make these loose statements precise: let F be a field and R a subring of F with $\varphi : \mathbb{Z} \to R$ an isomorphism of rings (this is what we mean when we say that F contains a copy of $\mathbb{Z}$). Show that this can be extended to an isomorphism $\widehat{\varphi} : \mathbb{Q} \to F' \subseteq F$ with $\widehat{\varphi}|_{\mathbb{Z}} = \varphi$.

(4.13) Putting it all together we get: *the prime subfield of a field is isomorphic either to the rationals* $\mathbb{Q}$ *or to the finite field* $\mathbb{F}_p$ *for some prime* p. Define the *characteristic* of a field to be 0 if the prime subfield is $\mathbb{Q}$ or p if the prime subfield is $\mathbb{F}_p$. Thus fields like $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ have characteristic zero, and indeed, any field of characteristic zero must be infinte, to contain $\mathbb{Q}$. Fields like $\mathbb{F}_2, \mathbb{F}_3 \ldots$ and the field $\mathbb{F}$ of order 8 given above have characteristic $2, 3$ and 2 respectively.

Exercise 35 Show that a field F has characteristic $p > 0$ if and only if p is the smallest number of summands such that the expression $1+1+\cdots+1$ is equal to 0. Show that F has characteristic 0 if and only if no such expression is equal to 0.

Thus, all fields of characteristic 0 are infinite, and the only examples we know of fields of characteristic $p > 0$ are finite. It is *not true though that a field of characteristic $p > 0$ must be finite*. There are some examples of infinite fields of characteristic $p > 0$ below.

Exercise 36 Suppose that f is an irreducible polynomial over a field F of characteristic 0. Recalling Exercise 16, show that the roots of f in any extension E of F are *distinct*.

(4.14) A natural question is to ask what fields contain the integers $\mathbb{Z}$?[5] Obviously the rationals $\mathbb{Q}$ do, and indeed by Exercise 34, as soon as a field contains a copy of $\mathbb{Z}$ it must also contain a copy of $\mathbb{Q}$.

It turns out that we can also construct $\mathbb{Q}$ abstractly from $\mathbb{Z}$ without having to first position it inside another field: consider the set

$$\mathbb{F} = \{(a, b) \,|\, a, b \in \mathbb{Z}, b \neq 0, \text{ and } (a, b) = (c, d) \text{ iff } ad = bc\}.$$

In otherwords, we take all ordered pairs of elements from $\mathbb{Z}$, but think of two ordered pairs (a, b) and (c, d) as being the same if $ad = bc$, eg: think of $(0, 1)$ and $(0, 2)$ as being the same element of $\mathbb{F}$, and similarly $(1, 1)$ and $(3, 3)$

Aside. One makes these loose statements more preicse by defining an equivalence relation on the set of ordered pairs $\mathbb{Z} \times \mathbb{Z}$ as $(a, b) \sim (c, d)$ if and only if $ad = bc$. The elements of $\mathbb{F}$ are then the equivalence classes under this relation. We will nevertheless stick with the looser formulation.

Define addition and multiplication on $\mathbb{F}$ in the following way:

$$(a, b) + (c, d) = (ad + bc, bd) \text{ and } (a, b)(c, d) = (ac, bd).$$

Exercise 37

1. Show that these definitions are "well-defined", ie: if $(a, b) = (a', b')$ and $(c, d) = (c', d')$, then $(a, b) + (c, d) = (a', b') + (c', d')$ and $(a, b)(c, d) = (a', b')(c', d')$–in otherwords, if two pairs are thought of as being the same, it doesn't matter which one we use in the arithmetic as we get the same answer.
2. Show that $\mathbb{F}$ is a field.
3. Now define a map $\varphi : \mathbb{F} \to \mathbb{Q}$ by $\varphi(a, b) = a/b$. Show that the map is well defined, ie: if $(a, b) = (a', b')$ then $\varphi(a, b) = \varphi(a', b')$. Show that φ is an isomorphism from $\mathbb{F}$ to $\mathbb{Q}$.

This construction can be generalised as the following Exercise shows:

Exercise 38 Repeat the construction above with $\mathbb{Z}$ replaced by an arbitrary integral domain R. The resulting field is called the *field of fractions of R*.

The field of fractions construction provides some very interesting examples of fields, possibly new in the reader's experience. Let $F[x]$ be the ring of polynomials with F-coefficients where F is any field. The field of fractions of this integral domain has elements of the form $f(x)/g(x)$ for f and g polynomials, in otherwords, rational functions with F-coefficients. The field is denoted $F(x)$ and is called the *field of rational functions over F*.

An infinite field of characteristic p: If $\mathbb{F}_p$ is the finite field of order p, then the field of rational functions $\mathbb{F}_p(x)$ is obviously infinite (it contains for example all the polynomials over $\mathbb{F}_p$, of which there are an infinite number). Moreover, the rational function 1 adds to itself p times to give 0.

A field properly containing the complex numbers: any field F is properly contained in $F(x)$, even $F = \mathbb{C}$.

Further Exercises for §**4.**

Exercise 39 Let $\mathbb{F}$ be the set of all matrices of the form $\begin{bmatrix} a & b \\ 2b & a \end{bmatrix}$ where a, b are in the field $\mathbb{F}_5$. Define addition and multiplication to be the usual addition and multiplication of matrices (and also the addition and multiplication in $\mathbb{F}_5$). Show that $\mathbb{F}$ is a field. How many elements does it have?

[5] or more precisely, which fields contain an *isomorphic copy* of the integers.

Exercise 40 Let $\mathbb{F}_2$ be the field of integers modulo 2, and α be an "abstract symbol" that can be multiplied so that it has the following property: $\alpha \times \alpha \times \alpha = \alpha^3 = \alpha + 1$ (a bit like decreeing that the imaginary i squares to give -1). Let

$$\mathbb{F} = \{a + b\alpha + c\alpha^2 \mid a, b, c \in \mathbb{F}_2\},$$

Define addition on $\mathbb{F}$ by: $(a_1 + b_1\alpha + c_1\alpha^2) + (a_2 + b_2\alpha + c_2\alpha^2) = (a_1 + a_2) + (b_1 + b_2)\alpha + (c_1 + c_2)\alpha^2$, where the addition of coefficients happens in $\mathbb{F}_2$. For multiplication, "expand" the expression $(a_1 + b_1\alpha + c_1\alpha^2)(a_2 + b_2\alpha + c_2\alpha^2)$ like you would a polynomial with α the indeterminate, the coefficients are dealt with using the arithmetic from $\mathbb{F}_2$, and so on. Replace any α^3 that result using the rule above.

1. Write down all the elements of $\mathbb{F}$.
2. Write out the addition and multiplication tables for $\mathbb{F}$ (ie: the tables with rows and columns indexed by the elements of $\mathbb{F}$, with the entry in the i-th row and j-th column the sum/product of the i-th and j-th elements of the field). Hence show that $\mathbb{F}$ is a field (you can assume that the addition and multiplication are associative as well as the distributive law, as these are a bit tedious to verify!) Using your tables, find the inverses (under multiplication) of the elements $1 + \alpha$ and $1 + \alpha + \alpha^2$, ie: find
$$\frac{1}{1+\alpha} \text{ and } \frac{1}{1+\alpha+\alpha^2} \text{ in } \mathbb{F}.$$
3. Is the extension $\mathbb{F}_2 \subset \mathbb{F}$ a simple one?

Exercise 41 Take the set $\mathbb{F}$ of the previous exercise, and define addition/multiplication in the same way except that the rule for simplification is now $\alpha^3 = \alpha^2 + \alpha + 1$. Show that in this case you *don't* get a field.

Exercise 42 Verify the claim in lectures that the set $\mathbb{F} = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$.

Exercise 43 Verify the claim in lectures that $\mathbb{Q}(\sqrt[3]{2}) = \{a + b(\sqrt[3]{2}) + c(\sqrt[3]{2})^2 \mid a, b, c \in \mathbb{Q}\}$.

Exercise 44 Find a complex number α such that $\mathbb{Q}(\sqrt{2}, i) = \mathbb{Q}(\alpha)$.

Exercise 45 Is $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{7})$ a simple extension of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, $\mathbb{Q}(\sqrt{2})$ or even of $\mathbb{Q}$?

Exercise 46 Let ∇ be an "abstract symbol" that has the following property: $\nabla^2 = -\nabla - 1$ (a bit like i squaring to give -1). Let

$$\mathbb{F} = \{a + b\nabla \mid a, b \in \mathbb{R}\},$$

and define an addition on $\mathbb{F}$ by: $(a_1 + b_1\nabla) + (a_2 + b_2\nabla) = (a_1 + a_2) + (b_1 + b_2)\nabla$. For multiplication, expand the expression $(a_1 + b_1\nabla)(a_2 + b_2\nabla)$ normally (treating ∇ like an indeterminate, so that $\nabla\nabla = \nabla^2$, and so on), and replace the resulting ∇^2 using the rule above. Show that $\mathbb{F}$ is a field, and is just the complex numbers $\mathbb{C}$. Do exactly the same thing, but with symbol $\triangle$ satisfying $\triangle^2 = \sqrt{2}\triangle - \sqrt[3]{5}$. Show that you *still* get the complex numbers.

§5. Rings II: Quotients

(5.1) Let R be a commutative ring. A subset $\mathfrak{a} \subseteq R$ is called a *principal ideal* iff there is some $r \in R$ such that

$$\mathfrak{a} = \{pr \,|\, p \in R\}.$$

In other words, $\mathfrak{a}$ is precisely the set of all multiples of some fixed element r. Denote such an ideal by $\langle r \rangle$.

The name is "principal" ideal as there are more general kinds of ideal. Nevertheless, in all rings we will be concerned with, every ideal in this more general sense turns out to be a principal ideal, so we will drop the principal from now on and just say "ideal".

Aside. Here is the more general notion: $\mathfrak{a} \subseteq R$ is an ideal iff $\mathfrak{a}$ is a subgroup of the abelian group $(R, \oplus)$ and for any $s \in R$ we have $s\mathfrak{a} = \{sp \,|\, p \in \mathfrak{a}\} \subseteq \mathfrak{a}$ and similarly $\mathfrak{a}s \subseteq \mathfrak{a}$. For example, if $R = \mathbb{Z}[x]$, then the set $\mathfrak{a}$ of polynomials with even coefficients and no constant term form an ideal. But there is no single polynomial f in $\mathbb{Z}[x]$ such that every polynomial in $\mathfrak{a}$ is a multiple of f.

Notice that if $s \in R$ and $\langle r \rangle$ is an ideal, then $s\langle r \rangle = \{s(pr) \,|\, p \in R\} = \{(sp)r \,|\, p \in R\}$ which is $\subseteq \langle r \rangle$. Similarly, $\langle r \rangle s \subseteq \langle r \rangle$ (as R is commutative). In other words, multiplying the elements of an ideal by an arbitrary element of the ring gives elements of the ideal.

(5.2) In any ring there are the trivial ideals $\langle 0 \rangle = \{0\}$ and $\langle 1 \rangle = R$, the second one as any element of R is a multiple of 1.

Exercise 47

1. Show that the only ideals in a field F are the two trivial ones (hint: use the property of ideals mentioned at the end of the last paragraph).
2. If R is a commutative ring whose only ideals are $\{0\}$ and R, then show that R is a field.
3. Show that in the non-commutative ring $M_n(F)$ of $n \times n$ matrices with entries from the field F there are only the two trivial ideals, but that $M_n(F)$ is not a field.

(5.3) For another example, consider the ring $\mathbb{Q}[x]$, the number $\sqrt{2} \in \mathbb{R}$, and the evaluation homomorphism $\varepsilon_{\sqrt{2}} : \mathbb{Q}[x] \longrightarrow \mathbb{R}$ given by

$$\varepsilon_{\sqrt{2}}(a_n x^n + \cdots + a_0) = a_n(\sqrt{2})^n + \cdots + a_0.$$

(see §**2.**). Let $\mathfrak{a}$ be the set of all polynomials in $\mathbb{Q}[x]$ that are sent to $0 \in \mathbb{R}$ by this map. Certainly $x^2 - 2 \in \mathfrak{a}$ (as $\sqrt{2}^2 - 2 = 0$). If $f = (x^2 - 2)g \in \mathbb{Q}[x]$, then

$$\varepsilon_{\sqrt{2}}(f) = \varepsilon_{\sqrt{2}}(x^2 - 2)\varepsilon_{\sqrt{2}}(g) = 0 \times \varepsilon_{\sqrt{2}}(g) = 0,$$

using the fact that $\varepsilon_{\sqrt{2}}$ is a homomorphism. Thus, $f \in \mathfrak{a}$, and so the ideal $\langle x^2 - 2 \rangle$ is $\subseteq \mathfrak{a}$.

Conversely, if h is sent to 0 by $\varepsilon_{\sqrt{2}}$, ie: $h \in \mathfrak{a}$, we can divide it by $x^2 - 2$ using the division algorithm,

$$h = (x^2 - 2)q + r,$$

where $\deg r < 2$, so that $r = ax + b$ for some $a, b \in \mathbb{Q}$. But since $\varepsilon_{\sqrt{2}}(h) = 0$ we have

$$(\sqrt{2}^2 - 2)q(\sqrt{2}) + r(\sqrt{2}) = 0 \Rightarrow r(\sqrt{2}) = 0 \Rightarrow a\sqrt{2} + b = 0.$$

If $a \neq 0$, then $\sqrt{2} \in \mathbb{Q}$ as $a, b \in \mathbb{Q}$, which is plainly nonsense. Thus $a = 0$, hence $b = 0$ too, so that $r = 0$, and hence $h = (x^2 - 2)q \in \langle x^2 - 2 \rangle$, and we get that $\mathfrak{a} \subseteq \langle x^2 - 2 \rangle$.

The conclusion is that the set of polynomials in $\mathbb{Q}[x]$ sent to zero by the evaluation homomorphism $\varepsilon_{\sqrt{2}}$ is an ideal.

(5.4) This in fact always happens. Remember that if R, S are rings and $\varphi : R \longrightarrow S$ a ring homomorphism, then the *kernel* of φ, denoted $\ker\varphi$, is the set of all elements of R sent to $0 \in S$ by φ, ie:

$$\ker\varphi = \{r \in R \,|\, \varphi(r) = 0 \in S\}.$$

Proposition 2 *If F is a field and S a ring then the kernel of a homomorphism $\varphi : F[x] \longrightarrow S$ is an ideal.*

Proof: Choose $g \in \ker\varphi$ non-zero of smallest degree (which we can do since the degrees of polynomials come from the set $\mathbb{Z}^+ \cup \{-\infty\}$). We claim that $\ker\varphi = \langle g\rangle$, for which we need to show that these two sets are mutually contained within each other. On the one hand, if $pg \in \langle g\rangle$ then

$$\varphi(pg) = \varphi(p)\varphi(g) = \varphi(p) \times 0 = 0,$$

since $g \in \ker\varphi$. Thus, $\langle g\rangle \subseteq \ker\varphi$.

On the other hand, let $f \in \ker\varphi$ and use the division algorithm to divide it by g,

$$f = qg + r,$$

where $\deg r < \deg g$. Now, $r = f - qg \Rightarrow \varphi(r) = \varphi(f - qg) = \varphi(f) - \varphi(q)\varphi(g) = 0 - \varphi(q).0 = 0$, since both $f, g \in \ker\varphi$. Thus, r is also in the kernel of φ. If r was a non-zero polynomial, then we would have a contradiction because $\deg r < \deg g$, but g was chosen from $\ker\varphi$ to have smallest degree. Thus we must have that $r = 0$, hence $f = qg \in \langle g\rangle$, ie: $\ker\varphi \subseteq \langle g\rangle$. □

(5.5) Let $\langle f\rangle \subset F[x]$ be an ideal and $g \in F[x]$ any polynomial. The set

$$g + \langle f\rangle = \{g + h \,|\, h \in \langle f\rangle\},$$

is called the *coset of $\langle f\rangle$ with representative g* (or the coset of $\langle f\rangle$ *determined* by g).

(5.6) As an example, consider the ideal $\mathfrak{a} = \langle x\rangle$ in $\mathbb{F}_2[x]$. Thus, $\mathfrak{a}$ is the set of all multiples of x, which is just the same thing as the collection of polynomials in $\mathbb{F}_2[x]$ *that have no constant term*. What are the cosets of $\mathfrak{a}$? Let g be any polynomial and consider the coset $g + \langle x\rangle$. The only possibilities are that g has no constant term, or it does, in which case this term is 1 (we are in $\mathbb{F}_2[x]$).

If g has no constant term, then

$$g + \langle x\rangle = \langle x\rangle.$$

For, $g+$ a polynomial with no constant term is another polynomial with no constant term, ie: $g + \langle x\rangle \subseteq \langle x\rangle$. On the other hand, if $p \in \langle x\rangle$ is any polynomial with no constant term, then $p - g \in \langle x\rangle$ so $p = g + (p - g) \in g + \langle x\rangle$, ie: $\langle x\rangle \subseteq g + \langle x\rangle$.

If g does have a constant term, you can convince yourself in exactly the same way that,

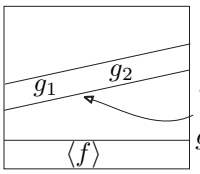
$$g + \langle x\rangle = 1 + \langle x\rangle.$$

Thus, there are only two cosets of $\langle x\rangle$ in $\mathbb{F}_2[x]$, namely the ideal $\mathfrak{a} = \langle x\rangle$ itself and $1 + \mathfrak{a}$; in English, the first coset consists of those polynomials *without* constant term, and the second those *with* a constant term.

Notice that these two cosets are completely disjoint, but every polynomial is in one or the other of them of them.

(5.7) Here are some basic properties of cosets:

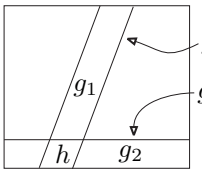
1. Every polynomial g is in some coset of $\langle f\rangle$, for $g = g + 0 \times f \in g + \langle f\rangle$.
2. For any q, we have $qf + \langle f\rangle = \langle f\rangle$, so multiples of f get "absorbed" into the ideal $\langle f\rangle$.

3. 

$g_1 + \langle f \rangle$ = $g_2 + \langle f \rangle$

The following three things are equivalent: (i). g_1 and g_2 lie in the same coset of $\langle f \rangle$; (ii). $g_1 + \langle f \rangle = g_2 + \langle f \rangle$; (iii). g_1 and g_2 differ by a multiple of f. To see this: (iii) $\Rightarrow$ (ii) If $g_1 - g_2 = pf$ then $g_1 = g_2 + pf$ so that $g_1 + \langle f \rangle = g_2 + pf + \langle f \rangle = g_2 + \langle f \rangle$; (ii) $\Rightarrow$ (i) Since $g_1 \in g_1 + \langle f \rangle$ and $g_2 \in g_2 + \langle f \rangle$, and these cosets are equal we have that g_1, g_2 lie in the same coset; (i) $\Rightarrow$ (iii) If g_1 and g_2 lie in the same coset, ie: $g_1, g_2 \in h + \langle f \rangle$, then each $g_i = h + p_i f \Rightarrow g_1 - g_2 = (p_1 - p_2)f$.

It can perhaps best be summarised by saying that g_1 and g_2 lie in the same coset if and only if this coset has the two different names, $g_1 + \langle f \rangle$ and $g_2 + \langle f \rangle$, as in the picture.

4. 

The situation in the picture opposite *never* happens. If the two cosets pictured are called respectively $g_1 + \langle f \rangle$ and $g_2 + \langle f \rangle$, then h is in both and so differs from g_1 and g_2 by multiples of f, ie: $g_1 - h = p_1 f$ and $h - g_2 = p_2 f$, so that $g_1 - g_2 = (p_1 + p_2)f$. Since g_1 and g_2 differ by a multiple of f, we have $g_1 + \langle f \rangle = g_2 + \langle f \rangle$.

Thus, the *cosets of an ideal partition the ring*.

(5.8) As an example of all these ideas, let $x^2 - 2 \in \mathbb{Q}[x]$ and consider the ideal

$$\langle x^2 - 2 \rangle = \{p(x^2 - 2) \mid p \in \mathbb{Q}[x]\}.$$

Certainly, $(x^3 - 2x + 15) + \langle x^2 - 2 \rangle$ is a coset, but is it written in the nicest possible form? If we divide by $x^2 - 2$:

$$x^3 - 2x + 15 = x(x^2 - 2) + 15,$$

we have that $x^3 - 2x + 15$ and 15 differ by a multiple of $x^2 - 2$. That gives

$$(x^3 - 2x + 15) + \langle x^2 - 2 \rangle = 15 + \langle x^2 - 2 \rangle.$$

(5.9) If we look again at the example of the coset $\langle x \rangle$ in $\mathbb{F}_2[x]$, there were only two cosets,

$$\langle x \rangle = 0 + \langle x \rangle \text{ and } 1 + \langle x \rangle,$$

that corresponded to the polynomials with constant term 0 and the polynomials with constant term 1 (these are the only possibilities for the coefficients in $\mathbb{F}_2[x]$!) We could try "adding" and "multiplying" these two cosets together according to,

$$(0 + \langle x \rangle) + (0 + \langle x \rangle) = 0 + \langle x \rangle, (1 + \langle x \rangle) + (0 + \langle x \rangle) = 1 + \langle x \rangle, (1 + \langle x \rangle) + (1 + \langle x \rangle) = 0 + \langle x \rangle,$$

and so on, where all we have done is to add the representatives of the cosets together using the addition from $\mathbb{F}_2$. Similarly for multiplying the cosets. This looks awfully like $\mathbb{F}_2$, but with $0 + \langle x \rangle$ and $1 + \langle x \rangle$ replacing 0 and 1.

(5.10) In fact this always happens. Let $\langle f \rangle$ be an ideal in $F[x]$, and define an addition and multiplication of cosets of $\langle f \rangle$ by,

$$(g_1 + \langle f \rangle) + (g_2 + \langle f \rangle) = (g_1 + g_2) + \langle f \rangle \text{ and } (g_1 + \langle f \rangle)(g_2 + \langle f \rangle) = (g_1 g_2) + \langle f \rangle,$$

where the addition and multiplication of the g_i's is happening in $F[x]$.

Theorem 3 *The set of cosets $F[x]/\langle f \rangle$ together with the $+$ and $\times$ above is a ring.*

Call this the *quotient ring* of $F[x]$ by the ideal $\langle f \rangle$. All our rings have a "zero", a "one", and so on, and for the quotient ring these are,

element of a ring	corresponding element in $F[x]/\langle f \rangle$
a	$g + \langle f \rangle$
$-a$	$(-g) + \langle f \rangle$
0	$0 + \langle f \rangle = \langle f \rangle$
1	$1 + \langle f \rangle$

Exercise 48 To prove this theorem,

1. Show that the addition of cosets is *well defined*, ie: if $g_i' + \langle f \rangle = g_i + \langle f \rangle$, then
$$(g_1' + g_2') + \langle f \rangle = (g_1 + g_2) + \langle f \rangle.$$
2. Similarly, show that the multiplication is well defined. Actually, it is because of this and the previous part that we can only take the quotients of polynomials by ideals, and not just any old subring.
3. Now verify the axioms for a ring.

Notice that the quotient is a ring, but not necessarily a field, So the motivating example above, where the quotient turned out to be the field $\mathbb{F}_2$ was a little special.

(5.11) Let $x^2 + 1 \in \mathbb{R}[x]$, and look at the ideal $\langle x^2 + 1 \rangle$. We want to see what the quotient $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ looks like. First, any coset can be put into a nice form: for example,

$$x^4 + x^2 + x + 1 + \langle x^2 + 1 \rangle = x^2(x^2 + 1) + (x + 1) + \langle x^2 + 1 \rangle,$$

where we have divided $x^4 + x^2 + x + 1$ by $x^2 + 1$ using the division algorithm. But

$$x^2(x^2 + 1) + (x + 1) + \langle x^2 + 1 \rangle = x + 1 + \langle x^2 + 1 \rangle,$$

as the multiple of x^2+1 gets absorbed into the ideal. In fact, for any $g \in \mathbb{R}[x]$ we can make this argument,

$$g + \langle x^2 + 1 \rangle = q(x^2 + 1) + (ax + b) + \langle x^2 + 1 \rangle = ax + b + \langle x^2 + 1 \rangle,$$

for some $a, b \in \mathbb{R}$, so the set of cosets can be written as

$$\mathbb{R}[x]/\langle x^2 + 1 \rangle = \{ax + b + \langle x^2 + 1 \rangle \mid a, b \in \mathbb{R}\}.$$

Now take two elements of the quotient, say $(x+1) + \langle x^2 + 1 \rangle$ and $(2x-3) + \langle x^2 + 1 \rangle$, and add/multiply them together:

$$\Big\{(x + 1) + \langle x^2 + 1 \rangle\Big\} + \Big\{(2x - 3) + \langle x^2 + 1 \rangle\Big\} = 3x - 2 + \langle x^2 + 1 \rangle,$$

and

$$\begin{aligned}\Big\{(x + 1) + \langle x^2 + 1 \rangle\Big\} \times \Big\{(2x - 3) + \langle x^2 + 1 \rangle\Big\} &= (2x^2 - x - 3) + \langle x^2 + 1 \rangle \\ &= 2(x^2 + 1) + (-x - 5) + \langle x^2 + 1 \rangle \\ &= -x - 5 + \langle x^2 + 1 \rangle.\end{aligned}$$

Now "squint" your eyes, so that the "$+\langle x^2 + 1 \rangle$" part in the above disappears, and $ax + b + \langle x^2 + 1 \rangle$ becomes the complex number $ai + b \in \mathbb{C}$. Then

$$(i + 1) + (2i - 3) = 3i - 2 \text{ and } (i + 1)(2i - 3) = -i - 5.$$

The addition and multiplication of cosets in $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ looks exactly like the addition and multiplication of complex numbers!

(5.12) In order to see what quotient rings *really* look like, you need to use the,

First Isomorphism Theorem. *Let R, S be rings and $\varphi : R \longrightarrow S$ a ring homomorphism. Then*

$$R/\mathit{ker}\,\varphi \cong \mathit{Im}\varphi \subset S,$$

where the isomorphism $\widehat{\varphi} : R/\mathit{ker}\,\varphi \to \mathit{Im}\varphi$ is given by $\widehat{\varphi}(r + \mathit{ker}\,\varphi) = \varphi(r)$.

Aside. For any ring R and homomorphism φ, $\ker\varphi$ is an ideal of R in the more general sense mentioned at the beginning of the section. Thus it makes sense to take the quotient $R/\ker\varphi$.

(5.13) Getting back to the example above, let $R = \mathbb{R}[x]$ and $S = \mathbb{C}$. Let the homomorphism φ be the evaluation at i homomorphism,

$$\varepsilon_i : \left(\sum a_k x^k\right) \mapsto \sum a_k (i)^k.$$

In exactly the same way as an earlier example, one can show that

$$\ker\varepsilon_i = \langle x^2 + 1\rangle.$$

On the other hand, if $ai + b \in \mathbb{C}$, then $ai + b = \varepsilon_i(ax + b)$, so the image of the homomorphism ε_i is all of $\mathbb{C}$. Feeding all this into the first homomorphism theorem gives,

$$\mathbb{R}[x]/\langle x^2 + 1\rangle \cong \mathbb{C}.$$

Exercise 49 Going back to the general case of an ideal $\mathfrak{a}$ in a ring R, consider the map $\eta : R \to R/\mathfrak{a}$ given by,

$$\eta(r) = r + \mathfrak{a},$$

sending an element of R to the coset of $\mathfrak{a}$ determined by it.

1. Show that η is a homomorphism.
2. Show that if $\mathfrak{b}$ is an ideal in R containing $\mathfrak{a}$ then $\eta(\mathfrak{b})$ is an ideal of $R/\mathfrak{a}$.
3. Show that if $\mathfrak{b}'$ is an ideal of $R/\mathfrak{a}$ then there is an ideal $\mathfrak{b}$ of R, containing $\mathfrak{a}$, such that $\eta(\mathfrak{b}) = \mathfrak{b}'$.
4. Show that in this way, η is a bijection between the ideals of R containing $\mathfrak{a}$ and the ideals of $R/\mathfrak{a}$.

Further Exercises for §**5.**

Exercise 50 Let $\theta : R \to S$ be a ring homomorphism. Show that,

1. $\theta(0) = 0$ (**hint**: consider $\theta(0 + 0)$),
2. θ is injective (ie: 1 to 1) if and only if $\ker\theta = \{0\}$.

Exercise 51 Determine which of these maps are ring homomorphisms.

1. The map $\theta : \mathbb{Z} \to \mathbb{Z}$ given by $\theta(n) = 2n$.
2. The map $\theta : \mathbb{Z} \to \mathbb{Z}$ given by $\theta(n) = -n$.
3. The map $\theta : \mathbb{R} \to \mathbb{R}$ given by $\theta(x) = |x|$.
4. The map $\theta : \mathbb{C} \to \mathbb{C}$ given by $\theta(z) = \overline{z}$ (*i.e.* complex conjugation).
5. The map $\theta : \mathbb{C} \to \mathrm{Mat}_2(\mathbb{R})$ defined by $\theta(x + iy) = \begin{bmatrix} x & y \\ -y & x \end{bmatrix}$.

Exercise 52 Determine whether the following maps are ring homomorphisms.

1. $\theta : \mathbb{Z}[x] \to \mathbb{Z}$ given by $\theta(f(x)) = f(0)$.
2. $\theta : \mathbb{Z}[x] \to \mathbb{Z}$ given by $\theta(f(x)) = f(1)$.
3. $\theta : \mathbb{Z}[x] \to \mathbb{Z}[x]$ given by $\theta(f(x)) = f(-x)$.
4. $\theta : \mathbb{Z}[x] \to \mathbb{Z}$ given by $\theta(f(x)) = f(2)^2$.

Exercise 53 Let $\phi = (1 + \sqrt{5})/2$ (in fact the *Golden Number*).

1. Show that the kernel of the evaluation map $\epsilon_\phi : \mathbb{Q}[x] \to \mathbb{C}$ (given by $\epsilon_\phi(f) = f(\phi)$) is the ideal $\langle x^2 - x - 1\rangle$.
2. Show that $\mathbb{Q}(\phi) = \{a + b\phi \mid a, b \in \mathbb{Q}\}$.
3. Show that $\mathbb{Q}(\phi)$ is the image in $\mathbb{C}$ of the map ϵ_ϕ.

§6. Fields II: Constructions and more examples

(6.1) An ideal $\langle f\rangle$ is *maximal* if and only if $\langle f\rangle \subset F[x]$ and the only ideals of $F[x]$ containing it are itself and the whole ring $F[x]$, ie:

$$\langle f\rangle \subseteq \mathfrak{a} \subseteq F[x],$$

with $\mathfrak{a}$ an ideal implies that $\mathfrak{a} = \langle f\rangle$ or $\mathfrak{a} = F[x]$.

(6.2) The principle result of this section is,

Theorem B (Constructing Fields). *The quotient ring $F[x]/\langle f\rangle$ is a field if and only if $\langle f\rangle$ is a maximal ideal.*

Proof: By Exercise 47, a commutative ring R is a field if and only if the only ideals of R are the trivial one $\{0\}$ and the whole ring R. Thus the quotient $F[x]/\langle f\rangle$ is a field if and only if its only ideals are the trivial one $\langle f\rangle$ and the whole ring $F[x]/\langle f\rangle$. By Exercise 49, there is a one to one correspondence between the ideals of the quotient $F[x]/\langle f\rangle$ and the ideals of $F[x]$ that contain $\langle f\rangle$. Thus $F[x]/\langle f\rangle$ has only the two trivial ideals precisely when there are only two ideals of $F[x]$ containing $\langle f\rangle$, which is the same as saying that $\langle f\rangle$ is maximal. □

(6.3) Suppose now that f is an irreducible polynomial over F, and let $\langle f\rangle \subseteq I \subseteq F[x]$ with I an ideal. Then $I = \langle h\rangle$ giving $\langle f\rangle \subseteq \langle h\rangle$, and so h divides f. Since f is irreducible this means that h must be either a constant $\lambda \in F$ or λf, so that the ideal I is either $\langle\lambda\rangle$ or $\langle\lambda f\rangle$. But $\langle\lambda f\rangle$ is just the same as the ideal $\langle f\rangle$. On the otherhand, any polynomial g can be written as a multiple of λ, just by setting $g = \lambda(\lambda^{-1}g)$, and so $\langle\lambda\rangle = F[x]$.

Thus, if f is an irreducible polynomial, then the ideal $\langle f\rangle$ is a maximal one. Conversely, if $\langle f\rangle$ is maximal and h divides f, then $\langle f\rangle \subseteq \langle h\rangle$, so that by maximality $\langle h\rangle = \langle f\rangle$ or $\langle h\rangle = F[x]$.

Exercise 54 Show that $\langle f\rangle = \langle h\rangle$ if and only if $h = \lambda f$ for some constant $\lambda \in F$. Similarly, $\langle h\rangle = F[x]$ if and only if $h = \lambda$ some constant.

Thus, the ideal $\langle f\rangle$ is maximal precisely when f is irreducible, giving,

Corollary. *$F[x]/\langle f\rangle$ is a field if and only if f is an irreducible polynomial over F.*

(6.4) The polynomial x^2+1 is irreducible over the reals $\mathbb{R}$, so the quotient ring $\mathbb{R}[x]/\langle x^2+1\rangle$ is a field.

(6.5) The polynomial x^2-2x+2 has roots $1\pm i$, hence is irreducible over $\mathbb{R}$, giving the field,

$$\mathbb{R}[x]/\langle x^2-2x+2\rangle.$$

Consider the evaluation map $\varepsilon_{1+i}\mathbb{R}[x] \to \mathbb{C}$ given as usual by $\varepsilon_{1+i}(f) = f(1+i)$. In exactly the same way as the example for $\varepsilon_{\sqrt{2}}$ in §**5.**, one can show that $\ker\varepsilon_{1+i} = \langle x^2-2x+2\rangle$. Moreover, $a+bi = \varepsilon_{1+i}(a-b+bx)$ so that the evaluation map is onto $\mathbb{C}$. Thus, by the first isomorphism theorem we get that,

$$\mathbb{R}[x]/\langle x^2-2x+2\rangle \cong \mathbb{C}.$$

What this means is that one can construct the complex numbers in the following (slightly non-standard) way: start with the reals $\mathbb{R}$, and define a new symbol, ∇ say, which is defined by the property,

$$\nabla^2 = 2\nabla - 2.$$

Now consider all expressions of the form $c+d\nabla$ for $c,d\in\mathbb{R}$. Add and multiply two such expressions together as follows:

$$\begin{aligned}(c_1+d_1\nabla)+(c_2+d_2\nabla) &= (c_1+c_2)+(d_1+d_2)\nabla\\ (c_1+d_1\nabla)(c_2+d_2\nabla) &= c_1c_2+(c_1d_2+d_1c_2)\nabla+d_1d_2\nabla^2\\ &= c_1c_2+(c_1d_2+d_1c_2)\nabla+d_1d_2(2\nabla-2)\\ &= (c_1c_2-2d_1d_2)+(c_1d_2+d_1c_2+2d_1d_2)\nabla.\end{aligned}$$

Exercise 55 By solving the equations $cx - 2dy = 1$ and $cy + dx + 2dy = 0$ for x and y in terms of c and d, find the inverse of the element $c + d\nabla$.

Exercise 56 According to Exercise 18, if f is irreducible over $\mathbb{R}$ then f must be either quadratic or linear. Suppose that $f = ax^2 + bx + c$ is an irreducible quadratic over $\mathbb{R}$. Show that the field $\mathbb{R}[x]/\langle zx^2 + bx + c\rangle \cong \mathbb{C}$.

(6.6) We saw in §**3.** that the polynomial $x^4 + x + 1$ was irreducible over the field $\mathbb{F}_2$. Thus the quotient

$$\mathbb{F}_2[x]/\langle x^4 + x + 1\rangle,$$

is a field. Each of its elements is a coset of the form $g + \langle x^4 + x + 1\rangle$. Use the division algorithm, dividing g by $x^4 + x + 1$ to get

$$g + \langle x^4 + x + 1\rangle = q(x^4 + x + 1) + r + \langle x^4 + x + 1\rangle = r + \langle x^4 + x + 1\rangle,$$

where the remainder r must be of the form $ax^3 + bx^2 + cx + d$, for $a, b, c, d \in \mathbb{F}_2$. Thus every element of the field has the form $ax^3 + bx^2 + cx + d + \langle x^4 + x + 1\rangle$, of which there are at most 16 possibilities (2 choices for a, 2 choices for b, ...).

Indeed these 16 are all distinct, for if

$$a_1x^3 + b_1x^2 + c_1x + d_1 + \langle x^4 + x + 1\rangle = a_2x^3 + b_2x^2 + c_2x + d_2 + \langle x^4 + x + 1\rangle$$

then,

$$\begin{aligned}(a_1 - a_2)x^3 + (b_1 - b_2)x^2 + (c_1 - c_2)x + (d_1 - d_2) + \langle x^4 + x + 1\rangle \\ = \langle x^4 + x + 1\rangle \Leftrightarrow (a_1 - a_2)x^3 + (b_1 - b_2)x^2 + (c_1 - c_2)x + (d_1 - d_2) \in \langle x^4 + x + 1\rangle.\end{aligned}$$

Since the non-zero elements of the ideal are multiples of a degree four polynomial, they have degrees that are at least four. Thus the only way the cubic can be an element is if it is the zero polynomial. In particular, $a_1 - a_2 = \cdots = d_1 - d_2 = 0$ so the two cosets are the same.

The upshot is that the quotient ring is a field with 16 elements.

(6.7) Returning to the general situation of a quotient $F[x]/\langle f\rangle$ by an irreducible polynomial f, the resulting field contains a copy of the original field F, obtained by taking the cosets $\lambda + \langle f\rangle$.

Exercise 57 Show that the map $\lambda \mapsto \lambda + \langle f\rangle$ is an injective homomorphism $F \to F[x]/\langle f\rangle$, and so F is isomorphic to its image in $F[x]/\langle f\rangle$.

Blurring the distinction between the original F and this copy inside $F[x]/\langle f\rangle$, we get that $F \subseteq F[x]/\langle f\rangle$ is an extension of fields.

(6.8) Generalising the example of the field of order 16 above, if $\mathbb{F}_p$ is the finite field with p elements and $f \in \mathbb{F}_p[x]$ is an irreducible polynomial of degree d, then the quotient $\mathbb{F}_p[x]/\langle f\rangle$ is a field containing p^d elements. They have the form,

$$a_{d-1}x^{d-1} + \cdots a_0 + \langle f\rangle,$$

where $f = b_dx^d + \cdots + b_1x + b_0$ and the $a_i \in \mathbb{F}_p$. Any two such are distinct by exactly the same argument as above. Letting $\alpha = x + \langle f\rangle$ and replacing $\mathbb{F}_p$ by its copy in $\mathbb{F}_p[x]/\langle f\rangle$ (ie: identifying $\lambda \in \mathbb{F}_p$ with $\lambda + \langle f\rangle \in \mathbb{F}_p[x]/\langle f\rangle$), we have,

$$\mathbb{F}_p[x]/\langle f\rangle = \{a_{d-1}\alpha^{d+1} + \cdots a_0 \mid a_i \in \mathbb{F}_p\},$$

where two such expressions are added and multiplied like "polynomials" in α. The only proviso is that since $f + \langle f\rangle = \langle f\rangle$, we have the "rule" $b_d\alpha^d + \cdots + b_1\alpha + b_0 = 0$, which allows us to remove any powers of α bigger than d that occur in such expressions.

Call α a *generator* for the finite field.

(6.9) The polynomial $x^3 + x + 1$ is irreducible over the field $\mathbb{F}_2$ (it is a cubic and has no roots) so that

$$\mathbb{F}_2[x]/\langle x^3 + x + 1\rangle,$$

is a field with $2^3 = 8$ elements of the form $\mathbb{F} = \{a + b\alpha + c\alpha^2 \,|\, a, b, c \in \mathbb{F}_2\}$ subject to the rule $\alpha^3 + \alpha + 1 = 0$, ie: $\alpha^3 = \alpha + 1$. This is the field $\mathbb{F}$ of order 8 in §**4.**.

Exercise 58 Construct fields with exactly:

1. 125 elements;
2. 49 elements;
3. 81 elements;
4. 243 elements.

(6.10) Theorem B and its Corollary clears up a little mystery that has lingered since the end of §**4.** Remember from there that the fields

$$\mathbb{Q}(\sqrt[3]{2}) \text{ and } \mathbb{Q}\left(\frac{-\sqrt[3]{2} + \sqrt[3]{2}\sqrt{3}i}{2}\right),$$

were different (that is, their elements were different), but isomorphic? The polynomial $x^3 - 2$ is irreducible over $\mathbb{Q}$, either by Eisenstein, or by observing that its roots do not lie in $\mathbb{Q}$. Thus

$$\mathbb{Q}[x]/\langle x^3 - 2\rangle,$$

is an extension field of $\mathbb{Q}$. Consider the two evaluation homomorphisms $\varepsilon_{\sqrt[3]{2}} : \mathbb{Q}[x] \to \mathbb{C}$ and $\varepsilon_\beta : \mathbb{Q}[x] \to \mathbb{C}$ where β is the complex number adjoined to $\mathbb{Q}$ in the second extension above. Since, and this is the key bit,

$$\sqrt[3]{2} \text{ and } \frac{-\sqrt[3]{2} + \sqrt[3]{2}\sqrt{3}i}{2}$$

are both roots of the polynomial $x^3 - 2$, we can show in a similar manner to examples at the end of §**5.** that $\ker \varepsilon_{\sqrt[3]{2}} \cong \langle x^3 - 2\rangle \cong \ker \varepsilon_\beta$. Thus,

$$\begin{array}{ccccc}
\mathbb{Q}[x]/\ker \varepsilon_{\sqrt[3]{2}} & \xleftarrow{\ =\ } & \mathbb{Q}[x]/\langle x^3 - 2\rangle & \xrightarrow{\ =\ } & \mathbb{Q}[x]/\ker \varepsilon_\beta \\
\downarrow \cong & \longleftarrow & \text{1st Isomorphism Theorem} & \longrightarrow & \cong \downarrow \\
\mathrm{Im}\varepsilon_{\sqrt[3]{2}} & & & & \mathrm{Im}\varepsilon_\beta
\end{array}$$

We can see what the image of $\varepsilon_{\sqrt[3]{2}}$ must be by considering the diagram,

$$\begin{array}{ccc}
a_n x^n + \cdots + a_1 x + a_0 & \xrightarrow[\text{algorithm}]{\text{division}} & q(x^3 - 2) + (a + bx + cx^2) \\
\downarrow \varepsilon_{\sqrt[3]{2}} & & \downarrow \varepsilon_{\sqrt[3]{2}} \\
a_n(\sqrt[3]{2})^n + \cdots + a_1\sqrt[3]{2} + a_0 & \overset{=}{\dashrightarrow} & (a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2)
\end{array}$$

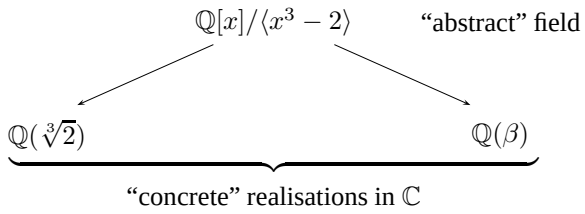
The point is that $\varepsilon_{\sqrt[3]{2}}$ is a ring homomorphism, so that

$$\begin{aligned}
\varepsilon_{\sqrt[3]{2}}(a_n x^n + \cdots + a_1 x + a_0) &= \varepsilon_{\sqrt[3]{2}}(q(x^3 - 2) + (a + bx + cx^2)) \\
&= \varepsilon_{\sqrt[3]{2}}(q)\varepsilon_{\sqrt[3]{2}}(x^3 - 2) + \varepsilon_{\sqrt[3]{2}}(a + bx + cx^2) \\
&= \varepsilon_{\sqrt[3]{2}}(q).0 + \varepsilon_{\sqrt[3]{2}}(a + bx + cx^2) = a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2.
\end{aligned}$$

Pictures like the one above, where you travel two routes but end up in the same place are called *commutative diagrams*. The result of the argument provided by the diagram is that

$$\mathrm{Im}\varepsilon_{\sqrt[3]{2}} \subseteq \{a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2 \in \mathbb{C} \mid a, b, c \in \mathbb{Q}\} = \mathbb{Q}(\sqrt[3]{2}).$$

On the other hand, any complex number of the form $a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2$ is the image of $a + bx + cx^2$. Thus $\mathrm{Im}\varepsilon_{\sqrt[3]{2}} = \mathbb{Q}(\sqrt[3]{2})$. Similarly one can show that $\mathrm{Im}\varepsilon_\beta = \mathbb{Q}(\beta)$. Filling this information into the first of the two diagrams above gives the claimed isomorphism between $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\beta)$:



(6.11) A special place is reserved in number theory for those fields of the form $\mathbb{Q}[x]/\langle f\rangle$, for f an irreducible polynomial over the rationals $\mathbb{Q}$. Such a field is called a *number field*, and their detailed study is the subject of algebraic number theory.

Suppose that β is a root of the polynomial f and consider the subfield of $\mathbb{C}$ given by $\mathbb{Q}(\beta)$. The reasoning in the example above can be extended to show that the two fields $\mathbb{Q}[x]/\langle f\rangle$ and $\mathbb{Q}(\beta)$ are isomorphic (see also Theorem D). Indeed, if $\{\beta_1, \ldots, \beta_n\}$ are the roots of f, then we have n mutually isomorphic fields inside $\mathbb{C}$, namely $\mathbb{Q}(\beta_1), \ldots, \mathbb{Q}(\beta_n)$. The isomorphisms from $\mathbb{Q}[x]/\langle f\rangle$ to each of these are called the *Galois monomorphisms* of the number field $\mathbb{Q}[x]/\langle f\rangle$.

(6.12) Returning to some generality, the observation that the field $F[x]/\langle f\rangle$ is an extension of F has far-reaching consequences that goes by the name of,

Kronecker's Theorem. *Let f be a polynomial in $F[x]$. Then there is an extension of F containing a root of f.*

Proof: If f is not irreducible over F, then factorise as $f = gh$ with g irreducible over F and proceed as below but with g instead of f. The result will be an extension containing a root of g, and hence of f. Thus we may suppose that f is irreducible over F and $f = a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0$ with the $a_i \in F$. Replace F by its isomorphic copy in the quotient $F[x]/\langle f\rangle$, so that instead of a_i, we write $a_i + \langle f\rangle$, ie,

$$f = (a_n + \langle f\rangle)x^n + (a_{n-1} + \langle f\rangle)x^{n-1} + \cdots + (a_1 + \langle f\rangle)x + (a_0 + \langle f\rangle).$$

Consider the field $E = F[x]/\langle f\rangle$ which is an extension of F and the element $x+\langle f\rangle \in E$. If we substitute $x + \langle f\rangle$ into the polynomial then we perform all our arithmetic in E, ie: we perform the arithmetic of cosets, and bear in mind that the zero of this field is the coset $\langle f\rangle$. Thus,

$$\begin{aligned} f(x + \langle f\rangle) &= (a_n + \langle f\rangle)(x + \langle f\rangle)^n + (a_{n-1} + \langle f\rangle)(x + \langle f\rangle)^{n-1} + \cdots + (a_1 + \langle f\rangle)(x + \langle f\rangle) + (a_0 + \langle f\rangle) \\ &= (a_n x^n + \langle f\rangle) + (a_{n-1}x^{n-1} + \langle f\rangle) + \cdots + (a_1 x + \langle f\rangle) + (a_0 + \langle f\rangle) \\ &= (a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0) + \langle f\rangle = f + \langle f\rangle = \langle f\rangle, \end{aligned}$$

which in the field E translates as $f(\mu) = 0$ for $\mu = x + \langle f\rangle$. □

Corollary. *Let f be a polynomial in $F[x]$. Then there is an extension of F that contains all the roots of f.*

Proof: Repeat the process described in the proof of Kronecker's Theorem at most $\deg f$ number of times, until the desired field is obtained.

Further Exercises for §**6.**

Exercise 59 Show that $x^4+x^3+x^2+x+1$ is irreducible over $\mathbb{F}_3$. How many elements does the resulting extension of $\mathbb{F}_3$ have?

Exercise 60 As linear polynomials are always irreducible, show that the field $F[x]/\langle ax+b\rangle$ is isomorphic to F.

Exercise 61

1. Show that $1+2x+x^3 \in \mathbb{F}_3[x]$ is irreducible and hence that $\mathbb{F} = \mathbb{F}_3[x]/\langle 1+2x+x^3\rangle$ is a field.
2. Show that every coset can be written uniquely in the form $(a+bx+cx^2)+\langle 1+2x+x^3\rangle$ with $a,b,c \in \mathbb{F}_3$.
3. Deduce that the field $\mathbb{F}$ has exactly 27 elements.

Exercise 62 Find an irreducible polynomial $f(x)$ in $\mathbb{F}_5[x]$ of degree 2. Show that $\mathbb{F}_5[x]/\langle f(x)\rangle$ is a field with 25 elements.

Exercise 63

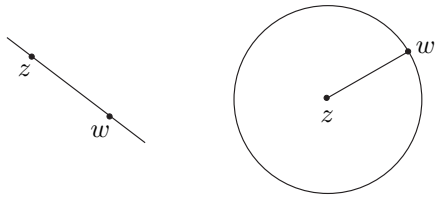
1. Show that the polynomial x^3-3x+6 is irreducible over $\mathbb{Q}$.
2. Hence, or otherwise, if
$$\alpha = \sqrt[3]{2\sqrt{2}-3}, \beta = -\sqrt[3]{2\sqrt{2}+3} \text{ and } \omega = -\frac{1}{2}+\frac{\sqrt{3}}{2}i,$$
prove that
   (a) the fields $\mathbb{Q}(\alpha+\beta)$ and $\mathbb{Q}(\omega\alpha+\overline{\omega}\beta)$ are *distinct* (that is, their elements are different), but,
   (b) $\mathbb{Q}(\alpha+\beta)$ and $\mathbb{Q}(\omega\alpha+\overline{\omega}\beta)$ are *isomorphic*.

You may assume that $\omega\alpha+\overline{\omega}\beta$ is not a real number.

§7. Ruler and Compass constructions I

If we allow ourselves a slightly fanciful historical interlude, we can imagine that the earliest civilizations to embrace agriculture came up against the problem of subdividing arable land into portions to be worked. Thus the Babylonians for instance would have needed the basics of surveying at their disposal. The most basic surveying instruments are wooden pegs and rope, with which you can do two very basic things: two pegs can be set a distance apart and the rope stretched between them; alternatively, one of the pegs can be kept stationary and we can take the path traced by the other as you walk around keeping the rope stretched taut. In otherwords, we can draw a line through two points and draw a circle centered at one point and passing through the other.

(7.1) Instead of the Euphrates river valley, we work in the complex plane $\mathbb{C}$. We are thus able, given two numbers $z, w \in \mathbb{C}$, to draw a line through them using a straight edge, or to place one end of a compass at z, and draw the circle passing through w:

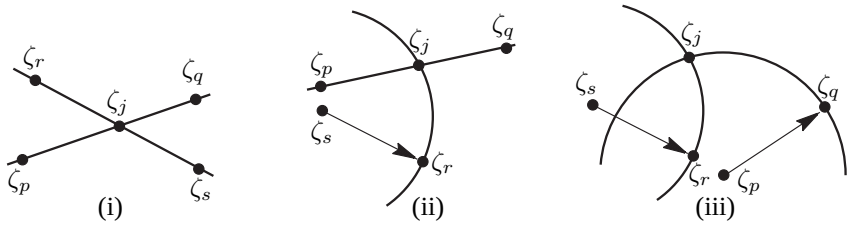


Notice that neither of these operations involves any "measuring".

(7.2) With these two constructions in hand, we call a complex number ζ *constructible* iff there is a sequence of numbers

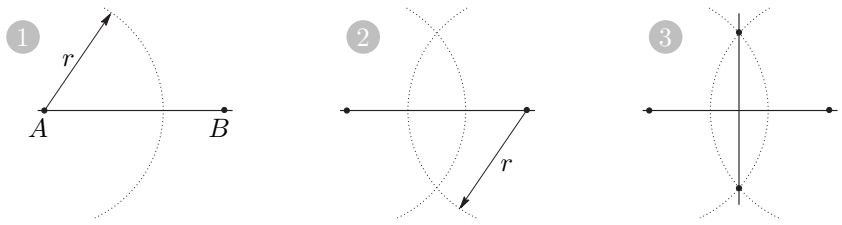
$$0, 1, i = \zeta_1, \zeta_2, \ldots, \zeta_n = \zeta,$$

with ζ_j obtained from earlier numbers in the sequence in one of the three following ways:



In these pictures, p, q, r and s are all $< j$. Notice that we are given the three numbers $0, 1, i$ "for free", so that they are indisputably constructible. The reasoning is this: if you stand in a plane (not $\mathbb{R}^2$ or $\mathbb{C}$, but a plane without coordinates), then your position can be taken as 0; decree a direction to be the real axis and a distance along it to be length 1; construct the perpendicular bisector of the segment from -1 to 1 (as in the next paragraph) and measure a unit distance along this new axis (in either direction) to get i.

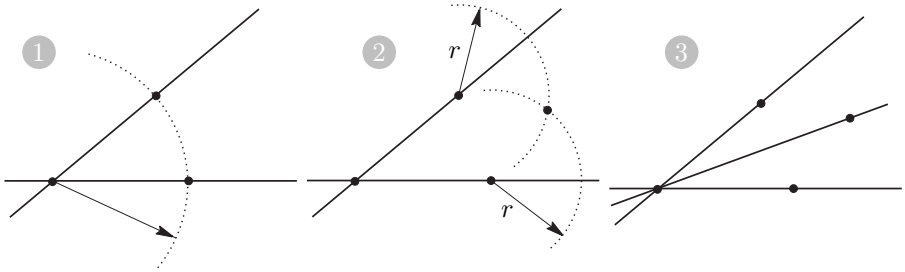
(7.3) The basic two moves are a little restrictive for the purposes of determining which numbers are constructible. There are a number of other constructions though, that follow immediately from them. For instance, we can construct the perpendicular bisector of a segment AB by the following three steps:



The pictures are supposed to be self-explanatory, modulo the following conventions. A ray, centered at some point and tracing out a dotted circle is obviously meant to describe usage of the compass. If the ray is marked r, as in the first two pictures above, this does not mean that the radius of the circle has been set to some length r, as we can not do this. It merely means that in passing from the first picture to the second, the setting on the compass is kept the same.

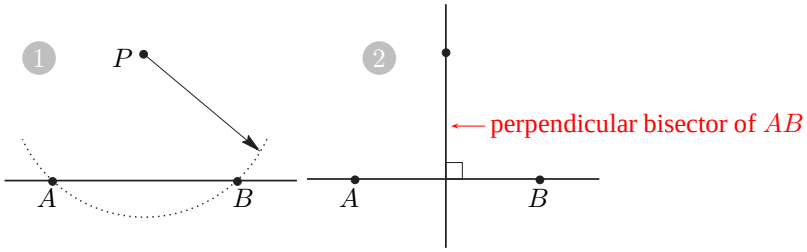
We can convince ourselves that the construction works as follows: think of the set S of points in $\mathbb{C}$ that are an equal distance from both A and B. After a moments thought, you see that this must be the perpendicular bisector of the line segment $\overline{AB}$ that we are constructing. Lines are determined by any two of their points, so if we can find two points equidistant from A and B, and we draw a line through them, this must be the set S that we want (and hence the perpendicular bisector). But the intersections of the two circular arcs are clearly equidistant from A and B, so we are done.

(7.4) As well as bisecting segments, we can bisect angles, ie: if two lines meet in some angle we can construct a third line meeting these in angles that are each half the original one:

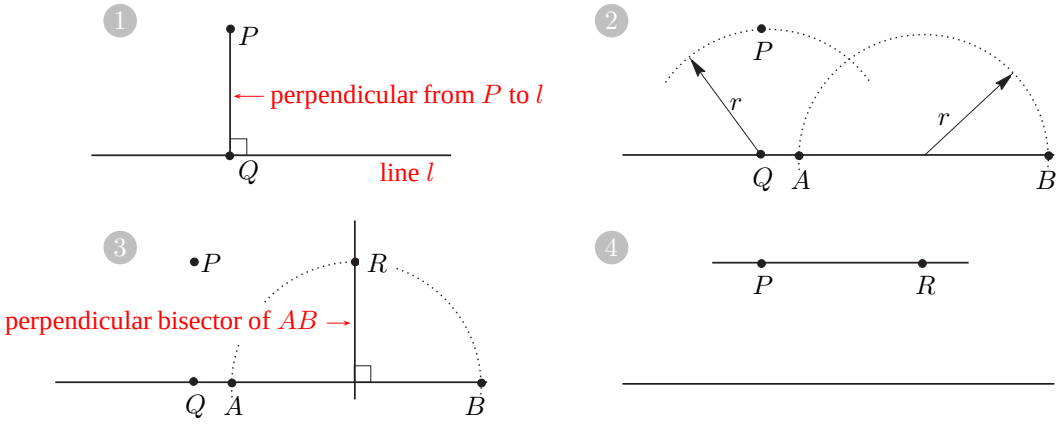


It is worth repeating that none of the angles in this picture can be measured. Nevertheless, the two new ones must each be half the old one.

(7.5) Given a line and a point P not on it, we can construct a new line passing through P and perpendicular to the line. We describe this as "dropping a perpendicular from a point to a line":



(7.6) Given a line l and a point P not on it we can construct a new line through P parallel to l:

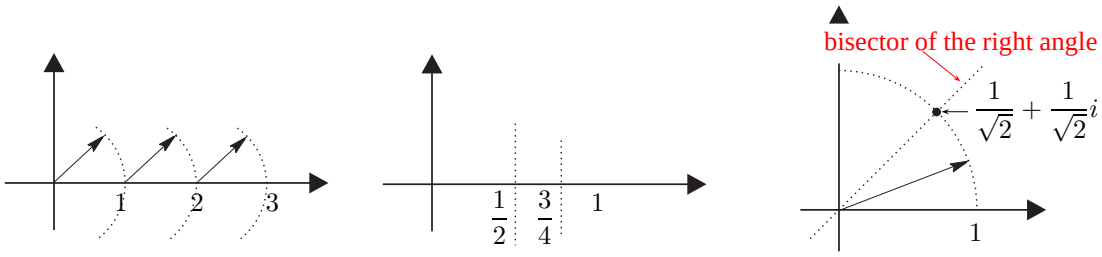


Perhaps some explanation wouldn't go amiss: the first step is to drop a perpendicular from P to the line l, meeting it at the new point Q. Next, set your compass to the distance from P to Q, and transfer this circular distance along the line to some point, drawing a semicircle that meets l at the points A and B. Construct the perpendicular bisector of the segment from A to B, which meets the semicircle at the new point R. Finally, draw a line through the points P and R. It should be obvious that P are R are equidistant from the line l, hence the line through them is parallel to l.

(7.7) Here are some basic examples that show that the numbers

$$3, \frac{3}{4} \text{ and } \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i$$

are constructible:



In the second example, we have bisected the segment from 0 to 1 and then the segment from $\frac{1}{2}$ to 1.

(7.8) Looking at the construction of $\frac{3}{4}$ above, it is less clear how one might construct the number $\frac{27}{129}$, or the golden ratio,

$$\phi = \frac{1+\sqrt{5}}{2}.$$

Nevertheless, these numbers *are* constructible, and the reason is the first non-trivial fact about constructible numbers: they can be added, subtracted, multiplied and divided[6]. Defining K to be the set of constructible numbers in $\mathbb{C}$, we have,

Theorem C. *K is a subfield of $\mathbb{C}$.*

Proof: The proof proceeds in two steps: as it is easier to deal with real numbers rather than complex, we show that the theorem can be reduced to the real case, and then show that the real constructible numbers form a subfield of $\mathbb{R}$.

First observe that $\zeta \in K$ precisely when $\mathrm{Re}\zeta$ and $\mathrm{Im}\zeta$ are in K. For, if $\zeta \in K$ then dropping perpendiculars to the real and imaginary axes give the numbers $\mathrm{Re}\zeta$ and $\mathrm{Im}\zeta i$, the second of which can be transferred to the real axis by drawing the circle centered at 0 passing through $\mathrm{Im}\zeta i$. On the otherhand, if we have $\mathrm{Re}\zeta$ and $\mathrm{Im}\zeta$ on the real axis, then we have $\mathrm{Im}\zeta i$ too, and constructing a line through $\mathrm{Re}\zeta$ parallel to the imaginary axis and a line through $\mathrm{Im}\zeta i$ parallel to the real axis gives ζ.

We now reduce the Theorem to the real case by showing that K is a subfield of $\mathbb{C}$ if and only if $K \cap \mathbb{R}$ is a subfield of $\mathbb{R}$. As the intersection of two subfields of $\mathbb{C}$ is a subfield of $\mathbb{C}$, the "only if" case is immediate.

Suppose then that the real constructible numbers form a subfield of the reals. We show that K is then a subfield of $\mathbb{C}$, for which we need to show that if z, w are constructible complex numbers then so are

[6] In principle you can now throw away your calculator, and perform arithmetic operations with ruler and compass! This is not as far-fetched as it sounds, even if it is a little impractical. To compute $\cos x$ of a constructable number x, construct as many terms of the Taylor series,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

as you need (your calculator only ever gives you approximations anyway).

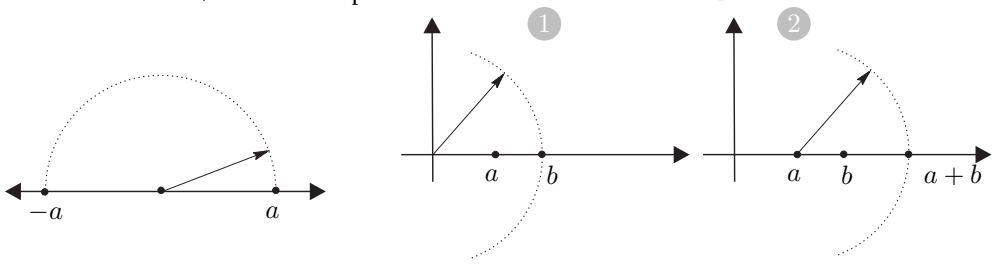
$z+w, -z, zw$ and $1/z$. By the observation above we have that the real and imaginary parts of z and w are real constructible numbers. Then

$$\begin{aligned} -z &= -\mathrm{Re}z - \mathrm{Im}zi \\ z+w &= (\mathrm{Re}z + \mathrm{Re}w) + (\mathrm{Im}z + \mathrm{Im}w)i \\ zw &= (\mathrm{Re}z\mathrm{Re}w - \mathrm{Im}z\mathrm{Im}w) + (\mathrm{Re}z\mathrm{Im}w + \mathrm{Im}w\mathrm{Re}z)i \\ \frac{1}{z} &= \frac{\mathrm{Re}z}{\mathrm{Re}z^2 + \mathrm{Im}z^2} - \frac{\mathrm{Im}z}{\mathrm{Re}z^2 + \mathrm{Im}z^2}i, \end{aligned}$$

As the constructible numbers form a subfield of $\mathbb{R}$, hence are closed under the four basic field operations, the real and imaginary parts of z and w are constuctible. Thus, the complex numbers are constructible as their real and imaginary parts are.

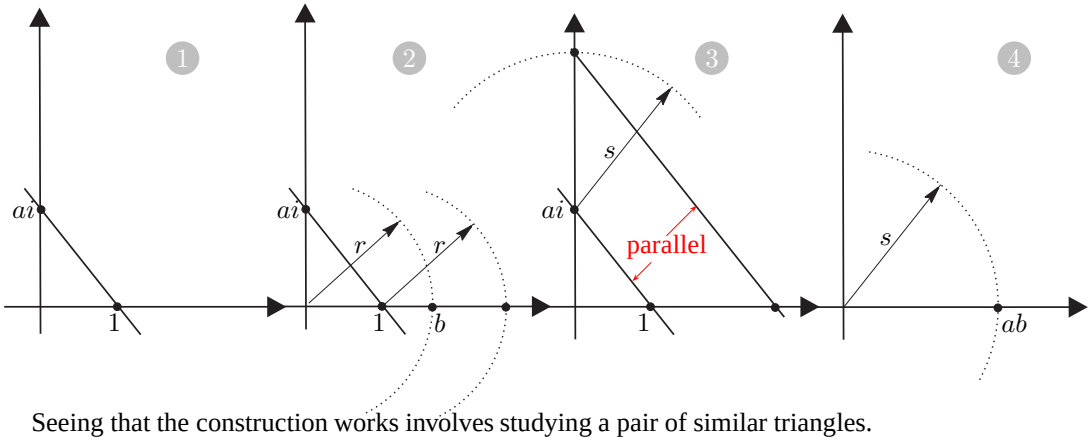
In light of this, it suffices to show that the real constructible numbers are a subfield of the reals, for which we need to show that if $a, b \in K \cap \mathbb{R}$ then so are $-a, a+b, ab$ and $1/a$.

1. $K \cap \mathbb{R}$ *is closed under* $+$ *and* $-$: The picture below left shows that if $a \in \mathbb{K} \cap \mathbb{R}$ then so is $-a$.



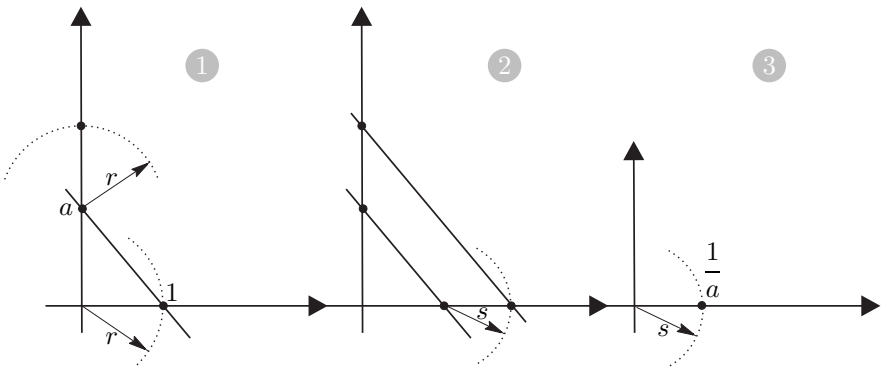
   Similarly, the two on the right give $a, b \in K \cap \mathbb{R} \Rightarrow a+b \in K \cap \mathbb{R}$.

2. $K \cap \mathbb{R}$ *is closed under* $\times$, as can be seen by following through the steps:



   Seeing that the construction works involves studying a pair of similar triangles.

3. $K \cap \mathbb{R}$ *is closed under* $\div$, which is of course just the previous construction backwards:



□

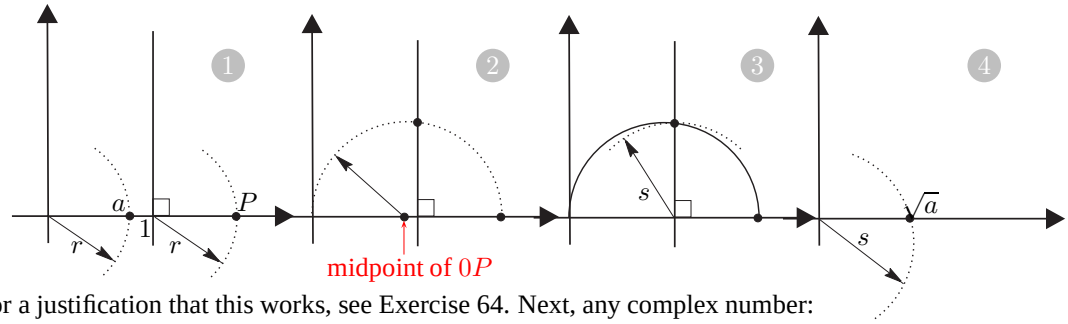
(7.9) As any subfield of $\mathbb{C}$ contains $\mathbb{Q}$ we thus have the,

Corollary. *Any rational number is constructible.*

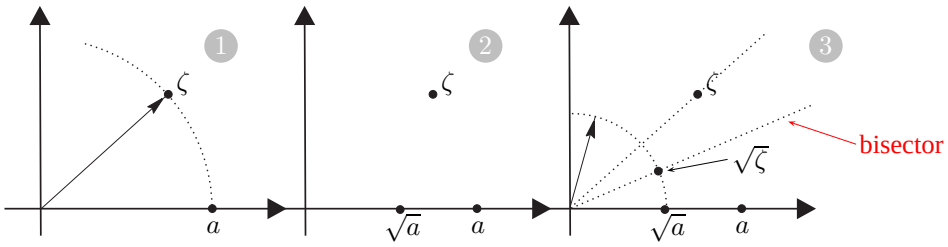
(7.10) Not only can we perform the four basic arithmetic operations with constructible numbers, but we can extract square roots too:

Theorem 4 *If $\zeta \in K$ then $\sqrt{\zeta} \in K$.*

Proof: First of all, we can construct the square root of any real number:

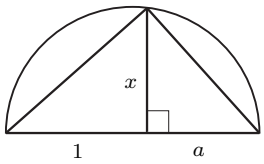


For a justification that this works, see Exercise 64. Next, any complex number:



where we have used the construction of real square roots in the second step. □

Exercise 64 Show that in the following picture,



the length $x = \sqrt{a}$.

Constructing angles and polygons

(7.11) We say that an angle can be constructed when we can construct two lines intersecting in that angle.

Exercise 65

1. Show that we can always assume that one of the lines giving an angle is the positive real axis.
2. Show that an *angle* θ can be constructed if and only if the *number* $\cos\theta$ can be constructed. Do the same for $\sin\theta$ and $\tan\theta$.

Exercise 66 Show that if φ, θ are constructible angles then so are $\varphi + \theta$ and $\varphi - \theta$.

(7.12) A *regular n-sided polygon* or *regular n-gon* is a polygon in $\mathbb{C}$ with n sides of equal length and n interior angles of equal size.

Exercise 67 Show that a regular n-gon can be constructed centered at $0 \in \mathbb{C}$ if and only if the angle $\frac{2\pi}{n}$ can be constructed. Show that a regular n-gon can be constructed centered at $0 \in \mathbb{C}$ if and only if the complex number

$$\zeta = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n},$$

can be constructed.

Exercise 68 Show that if an n-gon and an m-gon can be constructed for n and m relatively prime, then so can a mn-gon (hint: use the $\mathbb{Z}$-version of Theorem 1).

(7.13) For what n can one construct a regular n-gon? It makes sense to consider first the p-gons for p a prime. The complete answer even to this question will not be revealed until §**15.**. It turns out that the p-gons that can be constructed are *extremely rare*. Nevertheless, the first two (odd) primes do work.

Exercise 69 Show that a regular 3-gon, ie: an equilateral triangle, can be constructed with any side length. Using Exercises 4 and 67, show that a regular 5-gon can also be constructed.

(7.14) Here is a "proof" that a regular 17-gon is constructible. Gauss proved the following remarkable identity, which is still found in trigonometric tables: $\cos\frac{\pi}{17} =$

$$\frac{1}{8}\sqrt{2\left(2\sqrt{\sqrt{\frac{17(17-\sqrt{17})}{2}}-\sqrt{\frac{17-\sqrt{17}}{2}}-4\sqrt{34+2\sqrt{17}}+3\sqrt{17}+17}+\sqrt{34+2\sqrt{17}}+\sqrt{17}+15\right)}$$

Thus the number $\cos\pi/17$ can be constructed as this expression involves only integers, the four field operations and square roots, all of which are operations we can perform with a ruler and compass. Hence, by Exercise 65(2) the angle $\pi/17$ can be constructed and so adding it to itself gives the angle $2\pi/17$. Exercise 67 then gives that the 17-gon is constructible.

Further Exercises for §**7.**

Exercise 70 Using the fact that the constructible numbers include $\mathbb{Q}$, show that any given line segment can be trisected in length.

Exercise 71 Show that if you can construct a regular n-sided polygon, then you can also construct a regular $2^k n$-sided polygon for any $k \geq 1$.

Exercise 72 Show that $\cos\theta$ is constructible if and only if $\sin\theta$ is.

Exercise 73 If a, b and c are constructible numbers (ie: in K), show that the roots of the quadratic equation $ax^2 + bx + c$ are also constructible.

§8. Linear Algebra I: Dimensions

We have met rings and and fields so far in our study of Galois Theory. Time for our third algebraic object: vector spaces.

(8.1) A *vector space over a field* F is a set V of *vectors* together with two operations: addition $u, v \mapsto u + v$ of vectors and scalar multiplication $\lambda, v \mapsto \lambda v$ of a vector by an element λ of the field F, such that,

1. $(u + v) + w = u + (v + w)$, for all $u, v, w \in V$;
2. There exists a zero vector: $0 \in V$ s.t. $v + 0 = v = 0 + v$ for all $v \in V$,
3. Every $v \in V$ has a negative $-v$ s.t. $v + (-v) = 0 = -v + v$, for all $v \in V$.
4. $u + v = v + u$, for all $u, v \in V$.
5. $\lambda(u + v) = \lambda u + \lambda v$, for all u, v and $\lambda \in F$;
6. $(\lambda + \mu)v = \lambda v + \mu v$,
7. $\lambda(\mu v) = (\lambda\mu)v$,
8. $1v = v$ for $1 \in F$.

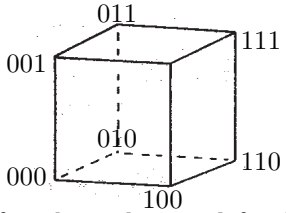
Aside. Alternatively we can say that the set V of vectors forms an Abelian group under $+$ (these are the first four axioms) together with the scalar multiplication which satisfies the last four axioms.

(8.2) The set $\mathbb{R}^2$ of 2×1 column vectors is a well known real vector space under the normal addition and scalar multiplication of vectors. Alternatively, the complex numbers $\mathbb{C}$ form a vector space over $\mathbb{R}$, and of course these two spaces are the same space after making the identification,

$$\begin{bmatrix} a \\ b \end{bmatrix} \leftrightarrow a + bi.$$

The complex numbers also form a vector space *over themselves*: addition of complex numbers gives an Abelian group and now we can scalar multiply a complex number by another one, just using the usual multiplication of complex numbers. It may seem a little confusing (especially the idea of thinking of a complex number as being *both* a vector and a scalar) but from a purely formal point of view, it satisfies the axioms and so is an admissible example. As we shall see below, the idea of thinking of the same *set* of objects as a vector space over two different fields is an important one for Galois Theory.

(8.3) We can consider vector spaces over finite fields too:



Consider the set of all 3-tuples where the coordinates come from the field $\mathbb{F}_2$, so are either 0 or 1, and add two such coordinate-wise, using the addition from $\mathbb{F}_2$. Scalar multiply a tuple coordinate-wise using the multiplication from $\mathbb{F}_2$. As there are only two possibilities for each coordinate and three coordinates in total, we get a total of $2^3 = 8$ elements in this space. Indeed, the elements can be arranged around the vertices of a cube as shown at left, where we have abbreviated so that abc is the vector with the three coordinates $a, b, c \in \mathbb{F}_2$.

(8.4) We saw in §**4.** that the field $\mathbb{Q}(\sqrt{2})$ consisted precisely of those elements of $\mathbb{C}$ of the form $a + b\sqrt{2}$ for $a, b \in \mathbb{Q}$. By making the identification,

$$a + b\sqrt{2} \quad \leftrightarrow \quad \begin{bmatrix} a \\ b \end{bmatrix} \begin{matrix} \longleftarrow \text{ coordinate in ``1 direction''} \\ \longleftarrow \text{ coordinate in ``}\sqrt{2}\text{ direction''} \end{matrix}$$

we realise $\mathbb{Q}(\sqrt{2})$ as a vector space over $\mathbb{Q}$. It is easy to check that the vector space operations match up with $(a+b\sqrt{2})+(c+d\sqrt{2})=(a+c)+(b+d)\sqrt{2}$ corresponding to,

$$\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a+c \\ b+d \end{bmatrix},$$

and $c(a+b\sqrt{2})=ac+bc\sqrt{2}$ corresponding to,

$$c\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ac \\ bc \end{bmatrix}.$$

(8.5) The polynomial x^3-2 is irreducible over $\mathbb{Q}$ so that $\mathbb{Q}[x]/\langle x^3-2\rangle$ is a field with elements having the form $(a+bx+cx^2)+\langle x^3-2\rangle$. It becomes a $\mathbb{Q}$-vector space under the identification,

$$(a+bx+cx^2)+\langle x^3-2\rangle \leftrightarrow \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{matrix} \longleftarrow \text{ coordinate in } ``1+\langle x^3-2\rangle \text{ direction}" \\ \longleftarrow \text{ coordinate in } ``x+\langle x^3-2\rangle \text{ direction}" \\ \longleftarrow \text{ coordinate in } ``x^2+\langle x^3-2\rangle \text{ direction}" \end{matrix}$$

(Check for yourself that the addition and scalar multiplications match up).

(8.6) The previous two are special cases of the following situation: if $F \subseteq E$ is an extension of fields then E *can be turned into a vector space over*F in the following way: the "vectors" are the elements of E and the scalars are obviously the elements of F. Addition of vectors is just the addition of elements in E, and to scalar multiply a $v \in E$ by a $\lambda \in F$, just multiply λv using the multiplication of the field E. That the first four axioms for a vector space hold follows from the addition of the field E, and the second four from the multiplication of the field E.

(8.7) Some more fundamental notions to do with vector spaces: for $v_1,\ldots,v_n \in V$ vectors, any vector of the form

$$\alpha_1 v_1 + \ldots + \alpha_n v_n,$$

for $\alpha_1,\ldots,\alpha_n \in F$, is a *linear combination* of the $v_1,\ldots,v_n$. The *linear span* of $v_1,\ldots,v_n \in V$ is the set of all linear combinations of these vectors,

$$\text{span}\{v_1,\ldots,v_n\} = \left\{ \sum_{j=1}^{n} \alpha_j v_j : \alpha_j \in F \right\}.$$

Say $v_1,\ldots,v_n$ *span* V when $V = \text{span}\{v_1,\ldots,v_n\}$.

A set of vectors $v_1,\ldots,v_n \in V$ is *linearly dependent* if and only if there exist scalars $\alpha_1,\ldots,\alpha_n$, *not all zero*, s.t.

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0.$$

The vectors $v_1,\ldots,v_n$ are *linearly independent* otherwise, ie: whenever $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$ implies that the α_i are all 0.

(8.8) In the examples above, the complex numbers $\mathbb{C}$ are spanned, as a vector space over $\mathbb{R}$, by the two elements $\{1,i\}$, and indeed by any two complex numbers that are not scalar multiples of each other. As a vector space over $\mathbb{C}$, the complex numbers are spanned by *just one element*, for example, any element $\zeta \in \mathbb{C}$ can be written as $\zeta \times 1$, so that every element is a complex scalar multiple of 1. Indeed, $\mathbb{C}$ is spanned as a complex vector space by any single one of its elements, except for 0. The moral is that in changing the field of scalars, you need to keep your wits about you.

(8.9) A *basis* for V is a linear independent set $\{v_j : j \in J\}$ (here J is a not necessarily finite index set), that spans V. We say V is *finite dimensional* if it has a finite basis.

It can be proved that there is a 1-1 correspondence between the elements of any two bases for a vector space V. Correspondingly, whenever V is finite dimensional we define *the dimension of* V to be $\dim(V) =$ number of elements in any basis.

(8.10) Thus $\mathbb{C}$ is 2-dimensional as a vector space over $\mathbb{R}$ but 1-dimensional as a vector space over $\mathbb{C}$. We will see later in this section that $\mathbb{C}$ is *infinite* dimensional as a vector space over $\mathbb{Q}$.

With the other examples above, $\mathbb{Q}(\sqrt{2})$ is 2-dimensional over $\mathbb{Q}$ with basis $\{1, \sqrt{2}\}$ and $\mathbb{Q}[x]/\langle x^3 - 2\rangle$ is 3-dimensional over $\mathbb{Q}$ with basis the cosets

$$1 + \langle x^3 - 2\rangle, x + \langle x^3 - 2\rangle \text{ and } x^2 + \langle x^3 - 2\rangle.$$

In Exercise 128 in §**14.**, we will see that if $\alpha = \sqrt[4]{2}$, then $\mathbb{Q}(\alpha, i)$ is a 2-dimensional space over $\mathbb{Q}(\alpha)$ or $\mathbb{Q}(\alpha i)$ or even $\mathbb{Q}((1+i)\alpha)$; a 4-dimensional space over $\mathbb{Q}(i)$ or $\mathbb{Q}(i\alpha^2)$, and an 8-dimensional space over $\mathbb{Q}$ (and these are almost, but not quite, all the possibilities; see the exercise for the full story).

(8.11) Vector spaces, like groups, rings and fields, are algebraic objects, and so like these other examples, there is a notion of a *homomorphism* of vector spaces. This is a map $\varphi : V_1 \to V_2$ that preserves any operations we may have, which in the case of vector spaces is the addition and scalar multiplications:

$$\varphi(u + v) = \varphi(u) + \varphi(v), \varphi(\lambda v) = \lambda\varphi(v) \text{ for } u, v \in V \text{ and } \lambda \in F.$$

For historical reasons, homomorphisms of vector are more commonly called *linear maps*.

Aside. Although we don't need these concepts here, there is an algebraic theory of vector spaces akin to that for groups, rings and fields. For example, linear maps have *kernels*, there are *quotients* of vector spaces, and so on. There is a first isomorphism theorem for vector spaces, which as usual reads as $V/\ker\varphi \cong \text{image}\,\varphi$. We get things like $\dim(V_1/V_2) = \dim(V_1) - \dim(V_2)$, so in particular for a linear map,

$$\dim(V) = \dim(\ker\varphi) + \dim(\text{image}\,\varphi).$$

In the *linear* theory of vector spaces (rather than the algebraic theory), $\dim(\text{image}\,\varphi)$ is called the *rank*, $\dim(\ker\varphi)$ the *nullility* and $\dim(V)$ the number of columns of a matrix. So the first isomorphism theorem for vector spaces translates into the mantra, "rank + nullility = the number of columns".

(8.12) Let $F \subseteq E$ be an extension of fields. Consider E as a vector space over F, and define the *degree of the extension* to be the dimension of this vector space, denoted $[E : F]$. Call $F \subseteq E$ a *finite extension* if the degree is finite.

(8.13) The extensions $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2})$ and $\mathbb{Q} \subseteq \mathbb{Q}[x]/\langle x^3 - 2\rangle$ have degrees 2 and 3.

(8.14) It is no coincidence that the degree of extensions of the form $F \subseteq F[x]/\langle f\rangle$ turn out to be the same as the degree of the polynomial f as the next result shows.

Theorem 5 *Let f be an irreducible polynomial in $F[x]$ of degree d. Then the extension,*

$$F \subseteq F[x]/\langle f\rangle,$$

has degree d.

Hence the name degree!

Proof: Replace, as usual, the field F by its copy in $F[x]/\langle f\rangle$, so that $\lambda \in F$ becomes $\lambda + \langle f\rangle$. Consider the set of cosets,

$$B = \{1 + \langle f\rangle, x + \langle f\rangle, x^2 + \langle f\rangle, \ldots, x^{d-1} + \langle f\rangle\}.$$

Then B is a basis for $F[x]/\langle f\rangle$ over F, for which we have to show that it spans the field/vector space and is linearly independent. To see that it spans, consider a typical element, which has the form,

$$g + \langle f\rangle = (qf + r)\langle f\rangle = r + \langle f\rangle = (a_0 + a_1x + \cdots + a_{d-1}x^{d-1}) + \langle f\rangle.$$

using the division algorithm and basic properties of cosets. This is turn gives,

$$(a_0+a_1x+\cdots+a_{d-1}x^{d-1})+\langle f\rangle = (a_0+\langle f\rangle)(1+\langle f\rangle)+(a_1+\langle f\rangle)(x+\langle f\rangle)+\cdots+(a_{d-1}+\langle f\rangle)(x^{d-1}+\langle f\rangle),$$

where the last is an F-linear combination of the elements of B. Thus this sets spans the space.

For linear independence, suppose we have an F-linear combination of the elements of B giving zero, ie:

$$(b_0 + \langle f\rangle)(1 + \langle f\rangle) + (b_1 + \langle f\rangle)(x + \langle f\rangle) + \cdots + (b_{d-1} + \langle f\rangle)(x^{d-1} + \langle f\rangle) = \langle f\rangle,$$

remembering that the zero of the field $F[x]/\langle f\rangle$ is the coset $0 + \langle f\rangle = \langle f\rangle$. Multiplying and adding all the cosets on the left hand side gives,

$$(b_0 + b_1x + \cdots + b_{d-1}x^{d-1}) + \langle f\rangle = \langle f\rangle,$$

so that $b_0 + b_1x + \cdots + b_{d-1}x^{d-1} \in \langle f\rangle$ (using another basic property of cosets). The elements of $\langle f\rangle$, being multiples of f, must have degree at least d, except for the zero polynomial. On the other hand $b_0 + b_1x + \cdots + b_{d-1}x^{d-1}$ has degree $\leq d - 1$. Thus it must be the zero polynomial, giving that all the b_i are zero, and that the set B is linearly independent over F as claimed. □

(8.15) What is the degree of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\pi)$? If it was finite, say $[\mathbb{Q}(\pi) : \mathbb{Q}] = d$, then any collection of more than d elements would be linearly dependent. In particular, the $d + 1$ elements,

$$1, \pi, \pi^2, \ldots, \pi^d,$$

are dependent over $\mathbb{Q}$, so that $a_0 + a_1\pi + a_2\pi^2 + \ldots + a_d\pi^d = 0$ for some $a_0, a_1, \ldots, a_d \in \mathbb{Q}$, not all zero, and so π is a root of the polynomial $a_0 + a_1x + a_2x^2 + \ldots + a_dx^d$, which contradicts π being transcendental over $\mathbb{Q}$. Thus, the degree of the extension is infinite, and so for $\mathbb{Q} \subset \mathbb{Q}(\alpha)$ to be finite, we clearly cannot have that α is transcendental over $\mathbb{Q}$.

(8.16) In fact this is always true:

Proposition 3 *Let $F \subseteq E$ and $\alpha \in E$. If the extension $F \subseteq F(\alpha)$ is finite, then α is algebraic over F.*

Proof: The proof is very similar to the example above. Suppose that the extension $F \subseteq F(\alpha)$ has degree n, so that any collection of $n + 1$ elements of $F(\alpha)$ must be linearly dependent. In particular the $n + 1$ elements

$$1, \alpha, \alpha^2, \ldots, \alpha^n$$

are dependent over F, so that there are $a_0, a_1, \ldots, a_n$ in F with

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

giving that α is algebraic over F as claimed. □

Thus, any field E that contains transcendentals over F will be infinite dimensional over F. In particular, $\mathbb{R}$ and $\mathbb{C}$ are infinite dimensional over $\mathbb{Q}$.

(8.17) The converse to Proposition 3 is partly true, as we summarise now in an important result

Theorem D (Complete Description of Simple Extensions). *Let $F \subseteq E$ and $\alpha \in E$ be algebraic over F. Then,*

1. *There is a unique polynomial $f \in F[x]$ that is monic, irreducible over F, and has α as a root;*
2. *The simple extension $F(\alpha)$ is isomorphic to the quotient $F[x]/\langle f\rangle$;*
3. *if $\deg f = d$, then the extension $F \subseteq F(\alpha)$ has degree d with basis $\{1, \alpha, \alpha^2, \ldots, \alpha^{d-1}\}$, and so,*

$$F(\alpha) = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{d-1}\alpha^{d-1} \mid a_0, \ldots, a_{d-1} \in F\}.$$

Proof: Hopefully most of the proof will be recognisable from the specific examples we have discussed already. As α is algebraic over F there is at least one F-polynomial having α as a root. Choose f' to be a non-zero one having smallest degree. This polynomial must then be irreducible over F, for if not, we have $f' = gh$ with $\deg(g), \deg(h) < \deg(f')$, and α must be a root of one of g or h, contradicting the original choice of f'. Divide through by the leading coefficient of f', to get f, a monic, irreducible (by Exercise 17) F-polynomial, having α as a root. If f_1, f_2 are polynomials with these properties then $f_1 - f_2$ has degree strictly less than either f_1 or f_2 and still has α as a root, so the only possibility is that $f_1 - f_2$ is zero, hence f is unique.

Consider the evaluation homomorphism $\varepsilon_\alpha : F[x] \to E$ defined as usual by $\varepsilon_\alpha(f) = f(\alpha)$. To show that the kernel of this homomorphism is the ideal $\langle f \rangle$ is completely analogous to the example at the beginning of Section §**5.**: clearly $\langle f \rangle$ is contained in the kernel, as any multiple of f must evaluate to zero when α is substituted into it. On the otherhand, if h is in the kernel of ε_α, then by division algorithm,

$$h = qf + r,$$

with $\deg(r) < \deg(f)$. Finding the ε_α image of both sides gives $0 = \varepsilon_\alpha(h) = \varepsilon_\alpha(qf) + \varepsilon_\alpha(r) = \varepsilon_\alpha(r)$, so that r has α as a root. As f is minimal with this property, we must have that $r = 0$, so that $h = qf$, ie: h is in the ideal $\langle f \rangle$, and so the kernel is contained in this ideal. Thus, $\ker \varepsilon_\alpha = \langle f \rangle$.

In particular we have an isomorphism $\widehat{\varepsilon_\alpha} : F[x]/\langle f \rangle \to \mathrm{Im}\varepsilon_\alpha$, given by,

$$\widehat{\varepsilon_\alpha}(g + \langle f \rangle) = \varepsilon_\alpha(g) = g(\alpha),$$

with the left hand side a field as f is irreducible over F. Thus, $\mathrm{Im}\varepsilon_\alpha$ is a subfield of E. Clearly, both the element α ($\varepsilon_\alpha(x) = \alpha$) and the field F ($\varepsilon_\alpha(\lambda) = \lambda$) are contained in $\mathrm{Im}\varepsilon_\alpha$, hence $F(\alpha)$ is too as $\mathrm{Im}\varepsilon_\alpha$ is subfield of E, and $F(\alpha)$ is the smallest one enjoying these two properties. Conversely, if $g = \sum a_i x^i \in F[x]$ then $\varepsilon_\alpha(g) = \sum a_i \alpha^i$, which is an element of $F(\alpha)$ as fields are closed under sums and products. Hence $\mathrm{Im}\varepsilon_\alpha \subseteq F(\alpha)$ and so these two are the same. Thus $\widehat{\varepsilon_\alpha}$ is an isomorphism between $F[x]/\langle f \rangle$ and $F(\alpha)$.

This final part follows immediately from Theorem 5, where we showed that the set of cosets

$$\{1 + \langle f \rangle, x + \langle f \rangle, x^2 + \langle f \rangle, \ldots, x^{d-1} + \langle f \rangle\},$$

formed a basis for $F[x]/\langle f \rangle$ over F. Their images under $\widehat{\varepsilon_\alpha}$, namely $\{1, \alpha, \alpha^2, \ldots, \alpha^{d-1}\}$, must then form a basis for $F(\alpha)$ over F. □

Notice from the proof of the first part of Theorem D that the polynomial f has the smallest degree of any polynomial having α as a root. For this reason it is called the *minimum polynomial* of α over F.

(8.18) An important property of the minimum polynomial is that it divides *any* other F-polynomial that has α as a root. Suppose that g is such an F-polynomial. By unique factorisation in $F[x]$, we can decompose g as

$$g = \lambda f_1 f_2 \ldots f_k,$$

where the f_i are monic and irreducible over F. Being a root of g, the element α must be a root of one of the f_i. By uniqueness, this f_i must be the minimum polynomial of α over F.

(8.19) It is labouring the point, but to find the degree of a simple extension $F \subseteq F(\alpha)$, you want to find the degree of the minimum polynomial over F of α.

How do you find this polynomial? Its simple: guess! A sensible first guess is a polynomial with F-coefficients that has α as root. If your guess is also monic and irreducible, then you have guessed right–Theorem D says there is only once such polynomial! If your guess is not monic, then replace it by a suitable scalar multiple.

Thus, the only way you can go wrong is if you inadvertently guess a polynomial that is not irreducible. In this case, *your next guess should be a factor of your first guess.* In this way, the search for minimum polynomials is no harder than determining irreducibility.

(8.20) As an example of this process, consider the minimum polynomial over $\mathbb{Q}$ of the p-th root of 1,

$$\cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p},$$

for p a prime. Your first guess is $x^p - 1$ which satisfies all the criteria bar irreducibility as $x - 1$ is a factor. Factorising,

$$x^p - 1 = (x-1)\Phi_p(x),$$

for Φ_p the p-th cyclotomic polynomial shown to be irreducible over $\mathbb{Q}$ in Exercise 29.

(8.21) How does one find the degree of extensions $F \subseteq F(\alpha_1, \ldots, \alpha_k)$ that are not simple, but the result of adjoining several elements? Such extensions are just a sequence of simple extensions. one after the other. If we can find the degrees of each of these simple extensions, all we need is a way to patch the answers together. The result that does this is called the,

Tower Law. *Let $F \subseteq E \subseteq L$ be a sequence or "tower" of extensions. If both of the intermediate extensions $F \subseteq E$ and $E \subseteq L$ are finite, then $F \subseteq L$ is too, and indeed*

$$[L : F] = [L : E][E : F].$$

(8.22) Before tackling the proof of the tower law, consider the example of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}, i)$, which is nothing other than a sequence of two simple extensions,

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{Q}(\sqrt[3]{2}, i).$$

We can use Theorem D to find the degrees of each of these individual simple extensions. Firstly, the minimum polynomial over $\mathbb{Q}$ of $\sqrt[3]{2}$ must $x^3 - 2$, for this polynomial is monic in $\mathbb{Q}[x]$ with $\sqrt[3]{2}$ as a root and irreducible over $\mathbb{Q}$ by Eisenstein (using $p = 2$). Thus the first of the two extensions above has degree $\deg(x^3 - 2) = 3$ and $\{1, \sqrt[3]{2}, (\sqrt[3]{2})^2\}$ is a basis for $\mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$.

Now let $\mathbb{F} = \mathbb{Q}(\sqrt[3]{2})$ so that the second extension is $\mathbb{F} \subseteq \mathbb{F}(i)$ and where the minimum polynomial of i over $\mathbb{F}$ must be $x^2 + 1$: it is monic in $\mathbb{F}[x]$ with i as a root, and irreducible over $\mathbb{F}$ as its two roots $\pm i$ are not in $\mathbb{F}$ (as $\mathbb{F} \subset \mathbb{R}$). Thus Theorem D again gives that $\mathbb{F} \subseteq \mathbb{F}(i)$ is a degree $\deg(x^2 + 1) = 2$ extension with $\{1, i\}$ a basis for $\mathbb{F}(i)$ over $\mathbb{F}$.

Now consider the elements,

$$\{1, \sqrt[3]{2}, (\sqrt[3]{2})^2, i, \sqrt[3]{2}i, (\sqrt[3]{2})^2 i\},$$

obtained by multiplying the two bases together. The claim is that they form a basis for $\mathbb{Q}(\sqrt[3]{2}, i) = \mathbb{F}(i)$ over $\mathbb{Q}$, so we need to show that the $\mathbb{Q}$-span of these six is all of this field and that they are linearly independent over $\mathbb{Q}$. For the first, let x be an arbitrary element of $\mathbb{Q}(\sqrt[3]{2}, i) = \mathbb{F}(i)$. As $\{1, i\}$ is a basis for $\mathbb{F}(i)$ over $\mathbb{F}$, x can be expressed as an $\mathbb{F}$-linear combination,

$$x = a + bi, a, b \in \mathbb{F}.$$

As $\{1, \sqrt[3]{2}, (\sqrt[3]{2})^2\}$ is a basis for $\mathbb{F}$ over $\mathbb{Q}$, both a and b can be expressed as $\mathbb{Q}$-linear combinations,

$$a = a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2, b = b_0 + b_1\sqrt[3]{2} + b_2(\sqrt[3]{2})^2,$$

with the $a_i, b_i \in \mathbb{Q}$. This gives,

$$x = a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 + b_0 i + b_1\sqrt[3]{2}i + b_2(\sqrt[3]{2})^2 i,$$

a $\mathbb{Q}$-linear combination for x, and so these six elements do indeed span the $\mathbb{Q}$-vector space $\mathbb{Q}(\sqrt[3]{2}, i)$.

Suppose we have,

$$a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 + b_0 i + b_1 a_3\sqrt[3]{2}i + b_2(\sqrt[3]{2})^2 i = 0,$$

with the $a_i, b_i \in \mathbb{Q}$. Gathering together real and imaginary parts,

$$(a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2) + (b_0 + b_1\sqrt[3]{2} + b_2(\sqrt[3]{2})^2)i = a + bi = 0,$$

for a and b now elements of $\mathbb{F}$. As $\{1, i\}$ independent over $\mathbb{F}$ we must have that the coefficients in this last expression are zero, ie: that $a = b = 0$. This now gives,

$$a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 = 0 = b_0 + b_1\sqrt[3]{2} + b_2(\sqrt[3]{2})^2,$$

and as $\{1, \sqrt[3]{2}, (\sqrt[3]{2})^2\}$ are independent over $\mathbb{Q}$ we deduce that all the coefficients in these two expressions are zero, ie: that $a_0 = a_1 = a_2 = b_0 = b_1 = b_2 = 0$, so that the six elements are independent and form a basis as claimed.

This certainly agrees with the answer given to us by the tower law in that,

$$6 = [\mathbb{Q}(\sqrt[3]{2}, i) : \mathbb{Q}] = 3 \times 2 = [\mathbb{Q}(\sqrt[3]{2}, i) : \mathbb{Q}(\sqrt[3]{2})][\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}].$$

(8.23) The example above is more than a specific verification of the tower law. It also shows us exactly how to prove it:

Proof: Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis for E as an F-vector space and $\{\beta_1, \beta_2, \ldots, \beta_m\}$ a basis for L as an E-vector space, both containing a finite number of elements as these extensions are finite by assumption. We then show that the mn elements

$$\{\alpha_i\beta_j\}, 1 \leq i \leq n, 1 \leq j \leq m,$$

form a basis for the F-vector space L. Working "backwards" as in the example above, if x is any element of L we can express it as an E-linear combination of the $\{\beta_1, \ldots, \beta_m\}$,

$$x = \sum_{i=1}^{m} a_i\beta_i,$$

where, as they are elements of E, each of the a_i can be expressed as F-linear combinations of the $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$,

$$a_i = \sum_{j=1}^{n} b_{ij}\alpha_j \Rightarrow x = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}\alpha_j\beta_i.$$

Thus the elements $\{\alpha_i\beta_j\}$ span the field L. If we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}\alpha_j\beta_i = 0,$$

with the $b_{ij} \in F$, we can collect together all the β_1 terms, all the β_2 terms, and so on (much as we took real and imaginary parts in the example), to obtain an E-linear combination,

$$\left(\sum_{j=1}^{n} b_{1j}\alpha_j\right)\beta_1 + \left(\sum_{j=1}^{n} b_{2j}\alpha_j\right)\beta_2 + \cdots \left(\sum_{j=1}^{n} b_{mj}\alpha_j\right)\beta_m = 0.$$

The independence of the β_i over E forces all the coefficients to be zero so that

$$\left(\sum_{j=1}^{n} b_{1j}\alpha_j\right) = \cdots = \left(\sum_{j=1}^{n} b_{mj}\alpha_j\right) = 0,$$

and the independence of the α_j over F forces all the coefficients in each of these to be zero too, ie: $b_{ij} = 0$ for all i, j. □

Further Exercises for §**8.**

Exercise 74 Show that the following two fields are isomorphic:

$$\mathbb{Q}\left(\cos\frac{2\pi}{p}+\sin\frac{2\pi}{p}i\right) \text{ and } \mathbb{Q}\left(\cos\frac{4\pi}{p}+\sin\frac{4\pi}{p}i\right)$$

where p is an (odd) prime number.

Exercise 75

1. Show that if $F \subseteq L$ are fields with $[L : F] = 1$ then $L = F$.
2. Let $F \subseteq L \subseteq E$ be fields with $[E : F] = [L : F]$. Show that $E = L$.

Exercise 76 Let $\mathbb{F} = \mathbb{Q}(a)$, where $a^3 = 2$. Express $(1 + a)^{-1}$ and $(a^4 + 1)(a^2 + 1)^{-1}$ in the form $ba^2 + ca + d$, where b, d, c are in $\mathbb{Q}$.

Exercise 77 Let $\alpha = \sqrt[3]{5}$. Express the following elements of $\mathbb{Q}(\alpha)$ as polynomials of degree at most 2 in α (with coefficients in $\mathbb{Q}$):

1. $1/\alpha$.
2. $\alpha^5 - \alpha^6$.
3. $\alpha/(\alpha^2 + 1)$.

Exercise 78 Find the minimum polynomial over $\mathbb{Q}$ of $\alpha = \sqrt{2} + \sqrt{-2}$. Show that the following are elements of the field $\mathbb{Q}(\alpha)$ and express them as polynomials in α (with coefficients in $\mathbb{Q}$) of degree at most 3:

1. $\sqrt{2}$.
2. $\sqrt{-2}$.
3. i.
4. $\alpha^5 + 4\alpha + 3$.
5. $1/\alpha$.
6. $(2\alpha + 3)/(\alpha^2 + 2\alpha + 2)$.

Exercise 79 Find the minimum polynomials over $\mathbb{Q}$ of the following numbers:

1. $1 + i$.
2. $\sqrt[3]{7}$.
3. $\sqrt[4]{5}$.
4. $\sqrt{2} + i$.
5. $\sqrt{2} + \sqrt[3]{3}$.

Exercise 80 Find the minimum polynomial over $\mathbb{Q}$ of the following:

1. $\sqrt{7}$.
2. $(\sqrt{11} + 3)/2$.
3. $(i\sqrt{3} - 1)/2$.

Exercise 81 For each of the following fields L and F, find $[L : F]$ and compute a basis for L over F.

1. $L = \mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$, $F = \mathbb{Q}$;
2. $L = \mathbb{Q}(\sqrt[4]{2}, i)$, $F = \mathbb{Q}(i)$;
3. $L = \mathbb{Q}(\xi)$, $F = \mathbb{Q}$, where ξ is a primitive complex 7th root of unity;
4. $L = \mathbb{Q}(i, \sqrt{3}, \omega)$, $F = \mathbb{Q}$, where ω is a primitive complex cube root of unity.

Exercise 82 Let $a = e^{\pi i/4}$. Find $[F(a) : F]$ when $F = \mathbb{R}$ and when $F = \mathbb{Q}$.

§9. Fields III: A Menagerie

This section collects together a number of miscellaneous but important concepts and examples of fields.

Splitting Fields

(9.1) In the first lecture we were interested in fields containing just enough numbers to solve some polynomial equation.

Suppose f is a polynomial with F coefficients. We say that f *splits* in an extension $F \subseteq E$ iff we can factorise

$$f = \prod_{i=1}^{\deg f} (x - \alpha_i),$$

in the polynomial ring $E[x]$. Thus f splits in E precisely when E contains all the roots $\{\alpha_1, \alpha_2, \ldots, \alpha_{\deg f}\}$ of f.

There will in general be many such extension fields: we are after the smallest one. Call E a *splitting field* for f over F, if f splits in E and $E = F(\alpha_1, \alpha_2, \ldots, \alpha_{\deg f})$, where $\{\alpha_1, \alpha_2, \ldots, \alpha_{\deg f}\}$ are the roots of f.

Exercise 83 Show that E is a splitting field of the polynomial f over F if and only if f splits in E but not in any subfield of E containing F (so in this sense, E is the *smallest* field containing F and all the roots).

(9.2) Our example from the first lecture again: the polynomial was $x^3 - 2$ with roots $\alpha, \alpha\omega, \alpha\omega^2$ where $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and

$$\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i.$$

Thus a splitting field for f over $\mathbb{Q}$ is given by $\mathbb{Q}(\alpha, \alpha\omega, \alpha\omega^2)$, which is the same thing as $\mathbb{Q}(\alpha, \omega)$.

(9.3) The example above shows that we can always find a splitting field for a polynomial over F: by Kronecker's Theorem we can find an extension E of F that contains all the roots $\{\alpha_1, \alpha_2, \ldots, \alpha_{\deg f}\}$ of f, and so adjoining them to F gives a splitting field $F(\alpha_1, \alpha_2, \ldots, \alpha_{\deg f}) \subseteq E$.

Aside. In §**12.** we will prove (Theorem **??**) that an isomorphism of a field to itself $\sigma : F \to F$ can always be extended to an isomorphism $\widehat{\sigma} : E_1 \to E_2$ where E_1 is a splitting field of some polynomial f over F and E_2 is another splitting field of this polynomial. Thus, *any two splitting fields of a polynomial over F are isomorphic.*

Finite Fields

We have already met a number of examples of finite fields: the $\mathbb{F}_p$ of course, and a few others such as the field of order 8 in §**4.**.

(9.4) To get more examples, we saw in §**6.** that by taking irreducible polynomials over finite fields we could, in principle, construct fields with a prime power number of elements. The idea was to find a polynomial of degree n, irreducible over the field $\mathbb{F}_p$, giving a field of order p^n. Here is a very concrete example of that idea.

Consider the polynomial $f = x^2 + x + 2 \in \mathbb{F}_3[x]$. Substituting the three elements of $\mathbb{F}_3$ into f gives

$$0^2 + 0 + 2 = 2, 1^2 + 1 + 2 = 1 \text{ and } 2^2 + 2 + 2 = 2,$$

so that f has no roots in $\mathbb{F}_3$. Being a quadratic, this gives that f is irreducible over the field $\mathbb{F}_3$, and so $\mathbb{F}_3[x]/\langle x^2 + x + 2\rangle$ is a field of order 3^2 called, say $\mathbb{F}_9$.

Let $\alpha = x + \langle x^2 + x + 2\rangle$ in $\mathbb{F}_9$ be a generator for this field as in §**6.**, so that the elements of $\mathbb{F}_9$ have the form $a + b\alpha$ with $a, b \in \mathbb{F}_3$ and multiplication satisfying the rule $\alpha^2 + \alpha + 2 = 0$, or equivalently[7],

[7] Note that $-1 = 2$ and $-2 = 1$ in $\mathbb{F}_3$.

$\alpha^2 = 2\alpha + 1$. Now let X be a new indeterminate, and consider the polynomials $\mathbb{F}_9[X]$ over $\mathbb{F}_9$ in this new variable. In particular,

$$g = X^3 + (2\alpha + 1)X + 1.$$

As g is a cubic, it will be irreducible over $\mathbb{F}_9$ precisely when it has no roots in this field, which can be verified as usual by a straight, albeit tedious, substitution:

$$\begin{aligned}
g(0) &= 1,\\
g(1) &= 1^3 + 2\alpha + 1 + 1 = 2\alpha,\\
g(2) &= 2^3 + 2(2\alpha + 1) + 1 = \alpha + 2,\\
g(\alpha) &= \alpha^3 + \alpha(2\alpha + 1) + 1 = \alpha(2\alpha + 1) + \alpha(2\alpha + 1) + 1 = 4\alpha^2 + 2\alpha + 1 = 2\alpha + 1 + 2\alpha + 1\\
&= \alpha + 2,\\
g(\alpha + 1) &= (\alpha + 1)^3 + (\alpha + 1)(2\alpha + 1) + 1 = \alpha^3 + 1 + 2\alpha^2 + 1 + 1 = \alpha(2\alpha + 1) + 2(2\alpha + 1)\\
&= (\alpha + 2)(2\alpha + 1) = 2(\alpha^2 + \alpha + 1) = 1,\\
g(\alpha + 2) &= (\alpha + 2)^3 + (\alpha + 2)(2\alpha + 1) + 1 = \alpha^3 + 2 + (\alpha + 2)(2\alpha + 1) + 1 = (2\alpha + 2)(2\alpha + 1)\\
&= \alpha^2 + 2 = (2\alpha + 1) + 2 = 2\alpha,\\
g(2\alpha) &= (2\alpha)^3 + 2\alpha(2\alpha + 1) + 1 = 2\alpha(2\alpha + 1) + 2\alpha(2\alpha + 1) + 1 = \alpha(2\alpha + 1) + 1\\
&= 2\alpha^2 + \alpha + 1 = \alpha + 2 + \alpha + 1 = 2\alpha,\\
g(2\alpha + 1) &= (2\alpha + 1)^3 + (2\alpha + 1)(2\alpha + 1) + 1 = 2\alpha^3 + 1 + \alpha^2 + \alpha + 1 + 1 = 2\alpha(2\alpha + 1) + \alpha^2 + \alpha\\
&= \alpha^2 + 2\alpha + \alpha^2 + \alpha = \alpha + 2,\\
g(2\alpha + 2) &= (2\alpha + 2)^3 + (2\alpha + 2)(2\alpha + 1) + 1 = 2\alpha^3 + \alpha + \alpha^2 + 2 + 1 = 2\alpha(2\alpha + 1) + \alpha + \alpha^2\\
&= 2\alpha^2 = \alpha + 2.
\end{aligned}$$

We have a used an energy saving device in these computations as summarised in the following exercise.

Exercise 84 If F is a field of characteristic $p > 0$, then $(a + b)^p = a^p + b^p$ (hint: refer to Exercise 27).

Thus g is irreducible over $\mathbb{F}_9$, giving a field

$$\mathbb{F}_9[X]/\langle X^3 + (2\alpha + 1)X + 1\rangle$$

of order $9^3 = 3^6 = 729$, called say $\mathbb{F}_{729}$. As we have a sequence of extensions $\mathbb{F}_3 \subseteq \mathbb{F}_9 \subseteq \mathbb{F}_{729}$, we can view $\mathbb{F}_{729}$ in two ways. Using the extension $\mathbb{F}_9 \subseteq \mathbb{F}_{729}$, the elements have the form,

$$A_0 + A_1\beta + A_2\beta^2,$$

where the $A_i \in \mathbb{F}_9$ and $\beta = X + \langle g\rangle$. Multiplication uses the rule $\beta^3 = (\alpha + 2)\beta + 2$. Alternatively, the extension $\mathbb{F}_3 \subseteq \mathbb{F}_{729}$ has, by the tower law, elements of the form,

$$a_0 + a_1\beta + a_2\beta^2 + a_3\alpha + a_4\alpha\beta + a_5\alpha\beta^2,$$

with the $a_i \in \mathbb{F}_3$.

Exercise 85

1. Construct a field $\mathbb{F}_8$ with 8 elements by showing that $x^3 + x + 1$ is irreducible over $\mathbb{F}_2$.
2. Find a cubic polynomial that is irreducible in $\mathbb{F}_8[x]$ (hint: refer to Exercise 26).
3. Hence, or otherwise, construct a field with $2^9 = 512$ elements.

(9.5) Recall that the prime subfield of a field is the smallest subfield, and is isomorphic to $\mathbb{F}_p$ for some p or to $\mathbb{Q}$. In particular, the prime subfield of a finite field F must be isomorphic to $\mathbb{F}_p$.

Using the ideas from §**8.**, we have an extension of fields $\mathbb{F}_p \subseteq F$ and hence the finite field F forms a vector space over the field $\mathbb{F}_p$. This space must be finite dimensional (for F to be finite), so each element of F can be written uniquely as a linear combination,

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n,$$

of some basis vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ with the $a_i \in \mathbb{F}_p$. In particular there are p choices for each a_i, and the choices are independent, giving p^n elements of F in total.

Thus, *a finite field must have p^n elements for some prime p.*

(9.6) Here is an extended example that shows the converse, ie: gives a standard construction of a field with $q = p^n$ elements for any prime p and positive integer n. Consider first the polynomial $x^q - x$ over the field $\mathbb{F}_p$ of p elements.

Let L be an extension of the field $\mathbb{F}_p$ containing all the roots of the polynomial, as guaranteed us by the Corollary to Kronecker's Theorem. In Exercise 16 we used the formal derivative whether a polynomial has distinct roots. We have that $\partial(x^q - x) = qx^{q-1} - 1 = p^n x^{p^n - 1} - 1 = -1$ as $p^n = 0$ in $\mathbb{F}_p$. Clearly the constant polynomial -1 has no roots in L, and so the original polynomial $x^q - x$ has no repeated roots in L by Exercise 16.

In fact, the p^n distinct roots of $x^q - x$ in L form a subfield, and this is the field of order p^n that we are after. To show this, we need that if λ, μ are roots, then so are $-\lambda$, $\lambda + \mu$, $\lambda\mu$ and λ^{-1}.

Firstly, $(-\lambda)^q - (-\lambda) = (-1)^q\lambda^q + \lambda$. If p is a prime, then it is either $p = 2$ or odd, in which case we have two cases to consider. If $p = 2$, then $-1 = 1$ in $\mathbb{F}_2$, so that $(-1)^q\lambda^q + \lambda = \lambda^q + \lambda = \lambda + \lambda$ (as λ is a root of $x^q - x$ so that $\lambda^q = \lambda$) $= 2\lambda = 0$. If p is odd then $(-1)^q = -1$ and $(-1)^q\lambda^q + \lambda = -\lambda^q + \lambda = -\lambda + \lambda = 0$. In either case, $-\lambda$ is also a root of the polynomial $x^q - x$.

Next,

$$(\lambda + \mu)^q = \sum_{i=0}^{q} \left(\begin{array}{c} q \\ i \end{array} \right) \lambda^i \mu^{q-i} = \lambda^q + \mu^q + q(\text{other terms}),$$

where in $\mathbb{F}_p$ we have that $q = p^n = 0$. Thus $(\lambda + \mu)^q = \lambda^q + \mu^q$. Substituting $\lambda + \mu$ into our polynomial then gives

$$(\lambda + \mu)^q - (\lambda + \mu) = \lambda^q + \mu^q - \lambda - \mu = 0,$$

as both λ and μ are roots so that $\lambda^q - \lambda = 0 = \mu^q - \mu$. Thus $\lambda + \mu$ is also a root of the polynomial.

Now, $(\lambda\mu)^q - \lambda\mu = \lambda^q\mu^q - \lambda\mu = \lambda\mu - \lambda\mu = 0$. Finally, $(\lambda^{-1})^q - (\lambda^{-1}) = (\lambda^q)^{-1} - (\lambda^{-1}) = \lambda^{-1} - \lambda^{-1} = 0$. In both cases we have used $\lambda^q = \lambda$.

Thus the $q = p^n$ roots of the polynomial in L form a subfield as claimed, and we have constructed a field with this many elements.

(9.7) Looking back at this example, we let L be an extension of $\mathbb{F}_p$ containing all the roots of the polynomial $x^q - x$. In particular, if these roots are $\{\alpha_1, \ldots, \alpha_q\}$, then $\mathbb{F}_p(\alpha_1, \ldots, \alpha_q)$ is a, hence *the*, splitting field over $\mathbb{F}_p$ of the polynomial. In the example we constructed the subfield $\mathbb{F}$ of L consisting of the roots of $x^q - x$. As any subfield contains $\mathbb{F}_p$, we have $\mathbb{F}_p(\alpha_1, \ldots, \alpha_q) \subseteq \mathbb{F}$, whereas $\mathbb{F} = \{\alpha_1, \ldots, \alpha_q\}$ so that $\mathbb{F} \subseteq \mathbb{F}_p(\alpha_1, \ldots, \alpha_q)$. Hence the field we constructed in the example *was* the splitting field over $\mathbb{F}_p$ of the polynomial $x^q - q$.

If F is now an arbitrary field with q elements, then it has prime subfield $\mathbb{F}_p$. Moreover, as the multiplicative group of F has order $q - 1$, by Lagrange's Theorem (see §**11.**), every element of F satisfies $x^{q-1} = 1$, hence is a root of the $\mathbb{F}_p$-polynomial $x^q = x \Leftrightarrow x^q - x = 0$. Thus, a finite field of order q is the splitting field over $\mathbb{F}_p$ of the polynomial $x^q - x$, and by the uniqueness of such things, *any two fields of order q are isomorphic*.

(9.8) We finish with a fact about finite fields that will prove useful later on. Remember that a field is, among other things, two groups spliced together in a compatible way: the elements form a group

under addition (the *additive group*) and the non-zero elements form a group under multiplication (the *multiplicative group*) .

Looking at the complex numbers as an example, we can find a number of *finite* subgroups of the multiplicative group $\mathbb{C}^*$ of $\mathbb{C}$ by considering roots of 1. For any n, the powers of the n-th root of 1,

$$\omega = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n},$$

form a subgroup of $\mathbb{C}^*$ of order n. Indeed, by definition, this subgroup is cyclic.

Proposition 1 *Let F be any field and G a finite subgroup of the multiplicative group F^* of F. Then G is a cyclic group.*

In particular, if F is now a finite field, then the whole multiplicative group F^* of F is finite. Hence *the multiplicative group of a finite field is cyclic*.

Proof: By Exercise 97 the order in an Abelian group of the element gh is the lowest common multiple of the orders of g and h. As G is finite we can write a list $1 = m_1, m_2, \ldots, m_k$ of all the possible orders of elements and find $1 = g_1, g_2, \ldots, g_k$ such that g_i has order m_i. Thus $g_1 g_2 \ldots g_k$ has order the lowest common multiple of all the possible orders in the group. Thus, if we call this order m, there is an element g of the group of order m, and any other element h satisfies $h^m = 1$. Hence every element of the group is a root of $x^m - 1$, and since this polynomial has at most m roots in F, the order of G must be $\leq m$. As $g \in G$ has order m its powers must exhaust the whole group, hence G is cyclic. □

Algebraically closed fields

(9.9) In the first part of this section we dealt with fields in which a particular polynomial of interest split into linear factors. On the otherhand, there are fields like the complex numbers in which *any* polynomial splits.

A field F is said to be *algebraically closed* if and only if every (non-constant) polynomial over F splits in F.

(9.10) If F is algebraically closed and α is algebraic over F then there is a polynomial with F-coefficients having α as a root. As F is algebraically closed, this polynomial splits in F, so that in particular α is *in* F. This explains the terminology: an algebraically closed field is *closed* with respect to the taking of algebraic elements. Contrast this with fields like $\mathbb{Q}$, over which there are algebraic elements like $\sqrt{2}$ that are not contained in $\mathbb{Q}$.

Exercise 86 Show that the following are equivalent:

1. F is algebraically closed;
2. every non-constant polynomial over F has a root in F;
3. the irreducible polynomials over F are precisely the linear ones;
4. if $F \subseteq E$ is a finite extension then $E = F$.

Theorem 6 *Every field is contained in an algebraically closed one.*

Proof (sketch): The full proof is beyond the scope of these notes, although the technical difficulties are not algebraic or number theoretical, but set theoretical. If the field is countable, the proof sort of runs as follows: there are countably many polynomials over a countable field, so take the union of all the splitting fields of these polynomials. Note that for a finite field, this is an infinite union, so an algebraically closed field containing even a finite field is very large. □

Simple extensions

(9.11) We saw in §**4.** that an extension like $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3})$ is, despite appearances, simple. It is certainly a finite extension, and this turns out to give simplicity as we now show:

Theorem 7 *Let $F \subseteq E$ be a finite extension so that the roots of any irreducible polynomial $f \in E[x]$ are distinct. Then E is simple, ie: $E = F(\alpha)$ for some $\alpha \in E$.*

The following proof is for the case that F is *infinite*.

Proof: Let $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a basis for E over F and consider the field $F_1 = F(\alpha_3, \ldots, \alpha_k)$, so that $E = F_1(\alpha_1, \alpha_2)$. We will show that $F_1(\alpha_1, \alpha_2)$ is a simple extension of F_1, ie: that $F_1(\alpha_1, \alpha_2) = F_1(\theta)$ for some $\theta \in E$. Thus $E = F(\alpha_1, \alpha_2, \ldots, \alpha_k) = F(\theta, \alpha_3 \ldots, \alpha_k)$, and so by repeatedly applying this procedure, E is a simple extension.

Let f_1, f_2 be the minimum polynomials over F_1 of α_1 and α_2, and let L be an algebraically closed field containing of the field F. As the α_i are algebraic over F, we have that the fields F_1 and E are contained in L too. In particular the polynomials f_1 and f_2 split in L,

$$f_1 = \prod_{i=1}^{\deg f_1} (x - \beta_i), f_2 = \prod_{i=1}^{\deg f_2} (x - \delta_i),$$

with $\beta_1 = \alpha_1$ and $\delta_1 = \alpha_2$. As the roots of these polynomials are distinct we have that $\beta_i \neq \beta_j$ and $\delta_i \neq \delta_j$ for all $i \neq j$. For any i and any $j \neq 1$, the equation. $\beta_i + x\delta_j = \beta_1 + x\delta_1$ has precisely *one* solution in F_1, namely

$$x = \frac{\beta_i - \beta_1}{\delta_1 - \delta_j}.$$

(notice that if we had $\delta_j = \delta_1$ then there would be infinitely many solutions to the equation $\beta_1 + x\delta_j = \beta_1 + x\delta_1$). As there only finitely many such equations and infinitely many elements of F_1, there must be an $c \in F_1$ which is a solution to *none* of them, ie: such that,

$$\beta_i + c\delta_j \neq \beta_1 + c\delta_1$$

for any i and any $j \neq 1$. Let $\theta = \beta_1 + c\delta_1 = \alpha_1 + c\alpha_2$, and we show that $F_1(\alpha_1, \alpha_2) = F_1(\theta) = F_1(\alpha_1 + c\alpha_2)$.

Clearly $\alpha_1 + c\alpha_2 \in F_1(\alpha_1, \alpha_2)$ so that $F_1(\alpha_1 + c\alpha_2) \subseteq F_1(\alpha_1, \alpha_2)$. We will show that $\alpha_2 \in F_1(\alpha_1 + c\alpha_2) = F_1(\theta)$, for then if so, $\alpha_1 + c\alpha_2 - c\alpha_2 = \alpha_1 \in F_1(\alpha_1 + c\alpha_2)$, and so $F_1(\alpha_1, \alpha_2) \subseteq F_1(\alpha_1 + c\alpha_2)$.

We have $0 = f_1(\alpha_1) = f_1(\theta - c\alpha_2)$, so if we let $r(t) \in F_1(\theta)[t]$ be given by $r(t) = f_1(\theta - ct)$, then we have that α_2 is a root of both $r(t)$ and $f_2(x)$. If γ is another common root of r and f_2, then γ is one of the δ_j, and $\theta - c\gamma$ (being a root of f_1) is one of the β_i, so that,

$$\gamma = \delta_j \text{ and } \theta - c\gamma = \beta_i \Rightarrow \beta_i + c\delta_j = \beta_1 + c\delta_1,$$

a contradiction. Thus r and f_2 have just the single common root α_2. Let h be the minimum polynomial of α_2 over $F_1(\theta)$, so that h divides both r and f_2 (recall that the minimum polynomial divides any other polynomial having α_2 as a root). This means that h must have degree one, for a higher degree would give more than one common root for r and f_2, ie: $h = t + b$ for some $b \in F_1(\theta)$. As $h(\alpha_2) = 0$ we thus get that $\alpha_2 = -b$ and so $\alpha_2 \in F_1(\theta)$ as required. □

The theorem is true for finite extensions of *finite* fields (even without the condition on the roots of the polynomials), but we omit the proof here. We saw in Exercise 36 that irreducible polynomials over fields of characteristic 0 have distinct roots. Thus, *any finite extension of a field of characteristic zero is simple*. For example, if $\alpha_1, \ldots, \alpha_k$ are algebraic over $\mathbb{Q}$, then $\mathbb{Q}(\alpha_1, \ldots, \alpha_k) = \mathbb{Q}(\theta)$ for some θ. This is a fundamental fact in algebraic number theory, the proof of which we have merely adapted.

§10. Ruler and Compass constructions II

(10.1) The degree of an extension is the only concept we need to completely answer the question of which complex numbers are constructible:

Theorem E. *The number $\zeta \in \mathbb{C}$ is constructible if and only if there exists a sequence of field extensions,*

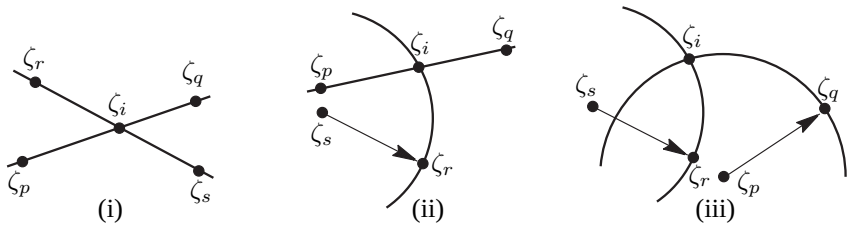
$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n,$$

such that $\mathbb{Q}(\zeta)$ is a subfield of K_n, and each K_i is a degree two extension of K_{i-1}.

Proof: ($\Rightarrow$) We prove the "only if" part first. Recall that ζ is constructible if and only if there is a sequence of numbers

$$0, 1, i = \zeta_1, \zeta_2, \ldots, \zeta_n = \zeta,$$

with ζ_i obtained from earlier numbers in the sequence in one of the three forms,



with $p, q, r, s \in \{1, 2, \ldots, n-1\}$. Let K_j be the field $\mathbb{Q}(\zeta_1, \ldots, \zeta_j)$, giving a "tower" of extensions,

$$\mathbb{Q} \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n.$$

We will show the following two things: (a). each of the fields K_j is closed under conjugation, ie: if $z \in K_j$ then $\bar{z} \in K_j$, and (b). the degree of each extension $K_{j-1} \subseteq K_j$ is at most two. Part (a) is just a technical convenience, the main point of which is illustrated by the exercise following the proof. We will prove it by induction: $K_1 = \mathbb{Q}(i)$ is clearly closed under conjugation, so we assume that K_i is closed under conjugation if $i < j$.

Suppose that ζ_j is obtained as in case (i), ie: as the intersection of two straight lines. The Cartesian equation for one of the straight lines is $y = m_1 x + c_1$, and suppose this line passes through the points ζ_p, ζ_q, with $\zeta_p, \zeta_q \in K_{j-1}$. As this field is closed under conjugation, Exercise 87 gives that the real and imaginary parts of ζ_p and ζ_q are in K_{j-1} too. As ζ_p, ζ_q lie on the line with this equation we have,

$$\left.\begin{aligned} \text{Im}\zeta_q &= m_1 \text{Re}\zeta_q + c_1 \\ \text{Im}\zeta_p &= m_1 \text{Re}\zeta_p + c_1 \end{aligned}\right\} \Rightarrow m_1 = \frac{\text{Im}\zeta_p - \text{Im}\zeta_q}{\text{Re}\zeta_p - \text{Re}\zeta_q} \in K_{j-1} \text{ and } c_1 = \text{Im}\zeta_p - m_1 \text{Re}\zeta_p \in K_{j-1}$$

(unless the line is vertical with equation $x = c_1$, in which case $c_1 = \text{Re}\zeta_p \in K_{j-1}$). Similarly if the equation of the other line is $y = m_2 x + c_2$, we have $m_2, c_2 \in K_{j-1}$. As ζ_j lies on both these lines we have

$$\left.\begin{aligned} \text{Im}\zeta_j &= m_2 \text{Re}\zeta_j + c_2 \\ \text{Im}\zeta_j &= m_1 \text{Re}\zeta_j + c_1 \end{aligned}\right\} \; m_i, c_i \in K_{j-1} \Rightarrow \text{Re}\zeta_j = \frac{c_2 - c_1}{m_1 - m_2} \text{ and } \text{Im}\zeta_j = \frac{m_1(c_2 - c_1)}{m_1 - m_2} + c_1,$$

and so $\text{Re}\zeta_j$ and $\text{Im}\zeta_j$ are in K_{j-1} as well. As this field is closed under conjugation we have that $\zeta_j \in K_{j-1}$ too, so that in fact $K_j = K_{j-1}(\zeta_j) = K_{j-1}$. Thus the degree of the extension $K_{j-1} \subseteq K_j$, being one, is certainly ≤ 2. Moreover, K_j is closed under conjugation as K_{j-1} is.

For case (ii), suppose that the line has equation $y = mx + c$ and the circle equation $(x - \text{Re}\zeta_s)^2 + (y - \text{Im}\zeta_s)^2 = r^2$, where $r^2 = (\text{Re}\zeta_r - \text{Re}\zeta_s)^2 + (\text{Im}\zeta_r - \text{Im}\zeta_s)^2$, so that in particular $r^2 \in K_{j-1}$

as $\zeta_p, \zeta_q, \zeta_r, \zeta_s \in K_{j-1}$ hence their real and imaginary parts are too. As ζ_j lies on the line we have $\text{Im}\zeta_j = m\text{Re}\zeta_j + c$, and it lies on the circle too, so that,

$$(\text{Re}\zeta_j - \text{Re}\zeta_s)^2 + (m\text{Re}\zeta_j + c - \text{Im}\zeta_s)^2 = r^2.$$

Thus the polynomial $(x - \text{Re}\zeta_s)^2 + (mx + c - \text{Im}\zeta_s)^2 = r^2$ is a quadratic with K_{j-1} coefficients having $\text{Re}\zeta_j$ as a root. As the minimum polynomial over K_{j-1} of $\text{Re}\zeta_j$ divides any other K_{j-1}-polynomial having $\text{Re}\zeta_j$ as a root, we get that this minimum polynomial has degree ≤ 2. Theorem D then gives,

$$[K_{j-1}(\text{Re}\zeta_j) : K_{j-1}] \leq 2.$$

In fact, $\text{Im}\zeta_j \in K_{j-1}(\text{Re}\zeta_j)$ as $\text{Im}\zeta_j = m\text{Re}\zeta_j + c$, thus ζ_j itself is in $K_{j-1}(\text{Re}\zeta_j)$, as i also is. Hence we have the sequence,

$$K_{j-1} \subseteq K_j = K_{j-1}(\zeta_j) \subseteq K_{j-1}(\text{Re}\zeta_j),$$

giving that the degree of the extension $K_{j-1} \subseteq K_j$ is also ≤ 2 by the tower law. Finally, we show that the field K_j is closed under conjugation, for which we can assume that the degree is two (it is trivially the case if the degree is one). Now, $K_j = K_{j-1}(\zeta_j) = K_{j-1}(\text{Re}\zeta_j)$, so that in particular ζ_j and its real part are in K_j, hence its imaginary part

$$\text{Im}\zeta_j = \frac{\zeta_j - \text{Re}\zeta_j}{i},$$

is too. The upshot is that $\text{Re}\zeta_j - i\text{Im}\zeta_j = \bar{\zeta_j}$ is in K_j, and as the elements of this field have the form $a + b\zeta_j$ with $a, b \in K_{j-1}$, we get that it is indeed closed under conjugation.

Finally, case (iii). As ζ_j lies on both circles we have,

$$(\text{Re}\zeta_j - \text{Re}\zeta_s)^2 + (\text{Im}\zeta_j - \text{Im}\zeta_s)^2 = r^2 \text{ and } (\text{Re}\zeta_j - \text{Re}\zeta_p)^2 + (\text{Im}\zeta_j - \text{Im}\zeta_p)^2 = s^2,$$

with both r^2 and s^2 in K_{j-1} for the same reason as in case (ii). Expanding both expressions, they contain terms of the form $\text{Re}\zeta_j^2 + \text{Im}\zeta_j^2$, and equating leads to,

$$\text{Im}\zeta_j = \frac{\beta_1}{\alpha}\text{Re}\zeta_j + \frac{\beta_2}{\alpha}, \text{ where } \alpha = 2(\text{Im}\zeta_s - \text{Im}\zeta_p), \beta_1 = 2(\text{Re}\zeta_p - \text{Re}\zeta_s)$$
$$\text{and } \beta_2 = \text{Re}\zeta_s^2 + \text{Im}\zeta_s^2 - (\text{Re}\zeta_p^2 + \text{Im}\zeta_p^2) + s^2 - r^2.$$

Combining this K_{j-1}-expression for $\text{Im}\zeta_j$ with the first of the two circle equations above puts us into a similar situation as part (ii), from which the result follows in the same way.

($\Leftarrow$) Now for the "if" part. Suppose that we have a tower of fields,

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n,$$

with $\mathbb{Q}(\zeta)$ in K_n. Each K_j is a simple extension $K_j = K_{j-1}(\zeta_j)$, so $K_j = \mathbb{Q}(\zeta_1, \ldots, \zeta_j)$, and in particular, $K_n = \mathbb{Q}(\zeta_1, \ldots, \zeta_n)$. We may as well assume that $\mathbb{Q}(\zeta)$ is not contained in K_{n-1}, so that $\zeta \notin K_{n-1}$. As $\mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}, \zeta) \subseteq \mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}, \zeta_n)$, we have that $\mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}, \zeta)$ is a degree two extension of $\mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1})$, so by Exercise 75, part 2,

$$\mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}, \zeta) = \mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}, \zeta_n).$$

Thus, the tower of extensions has the form,

$$\mathbb{Q} \subseteq \mathbb{Q}(\zeta_1) \subseteq \cdots \subseteq \mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}) \subseteq \mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1}, \zeta).$$

It suffices to prove therefore, that whenever we have an extension $K \subseteq K(\theta)$ of degree two, then there are finitely many elements of K from which θ can be constructed in a finite number of steps. For if so, then ζ can be constructed from finitely many elements of $\mathbb{Q}(\zeta_1, \ldots, \zeta_{n-1})$, each of which in turn can be constructed from finitely many elements of $\mathbb{Q}(\zeta_1, \ldots, \zeta_{n-2})$, and so on.

Given $K \subseteq K(\theta)$ as above then, the minimum polynomial of θ over K has the form $x^2 + bx + c$, with $b, c \in K$, so that θ is one of,

$$\frac{-1 \pm \sqrt{b^2 - 4c}}{2},$$

which can be constructed from $1, 2, 4, b, c \in K$, using the arithmetical and square root constructions of §**7.** □

Exercise 87 Let K be a field such that

$$\mathbb{Q}(i) \subseteq K \subseteq \mathbb{C},$$

as well as being closed under conjugation, ie: if $z \in K$ then $\bar{z} \in K$. Show that $z \in K$ if and only if the real and imaginary parts of z are in K.

(10.2) It is much easier to use the "only if" part of the Theorem, which shows when numbers *cannot* be constructed, so we restate this part as a separate,

Corollary. *If $\zeta \in \mathbb{C}$ is constructible then the degree of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\zeta)$ must be a power of two.*

To use the "if" part, in otherwords, to show that numbers *can* be constructed by finding a tower of fields as in Theorem E, is a little harder. We will need to know a great deal more about the fields stacked in between $\mathbb{Q}$ and $\mathbb{Q}(\zeta)$ before we can do this. The Galois Correspondence in §**14.** will give us the control to do this, so we postpone any attempts at using the "if" part of Theorem E until then.

Proof: If ζ is constructible then we have the tower of degree two extensions as given in Theorem E, with $\zeta \in K_n$. Thus we have the sequence of extensions $\mathbb{Q} \subseteq \mathbb{Q}(\zeta) \subseteq K_n$, which by the tower law gives,

$$[K_n : \mathbb{Q}] = [K_n : \mathbb{Q}(\zeta)][\mathbb{Q}(\zeta) : \mathbb{Q}].$$

Thus $[\mathbb{Q}(\zeta) : \mathbb{Q}]$ divides $[K_n : \mathbb{Q}]$, which is a power of two, so $[\mathbb{Q}(\zeta) : \mathbb{Q}]$ must also be a power of two. □

(10.3) Notice that the Corollary is only stated in one direction. Indeed, the converse, that if the extension has degree a power of two, then the number is constructible, *is not true*.

(10.4) A regular p-gon, for p a prime, can be constructed, by Exercise **??**, precisely when the complex number $\zeta = \cos(2\pi/p) + i\sin(2\pi/p)$ can be constructed, so we need to find the degree of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\zeta)$. By Exercise 29, the minimum polynomial of ζ over $\mathbb{Q}$ is the p-th cyclotomic polynomial,

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1.$$

Thus the degree of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\zeta)$ is $p - 1$, hence by the Corollary to Theorem E we require, for the p-gon to be constructible, that $p - 1$ is a power of two. In otherwords, the prime p is of the form

$$p = 2^n + 1.$$

Actually, even more can be said. If m is odd, the polynomial $x^m + 1$ has -1 as a root, thus can be factorised as $x^m + 1 = (x + 1)(x^{m-1} - x^{m-2} + x^{m-3} - \cdots - x + 1)$. Thus if $n = mk$ for m odd, we have

$$2^n + 1 = (2^k)^m + 1 = (2^k + 1)((2^k)^{m-1} - (2^k)^{m-2} + (2^k)^{m-3} - \cdots - (2^k) + 1),$$

so that $2^n + 1$ cannot be prime unless n has no odd divisors, which means that n itself must be a power of two.

Thus for a p-gon to be constructible, we must have that p is a prime number of the form

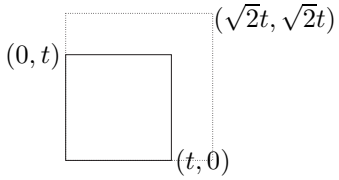
$$2^{2^t} + 1,$$

a so-called *Fermat prime*. Such primes are extremely rare: the only ones $< 10^{900}$ are

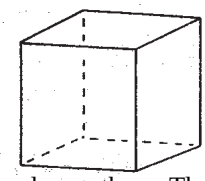
$$3, 5, 17, 257 \text{ and } 65537.$$

We will see in §**15.** that the converse is true: if p is a Fermat prime, then a regular p-gon can be constructed!

(10.5) A square plot of land can always be doubled in area using a ruler and compass:



Whatever the side length t of the original square is, it is constructible: just set the compass to the side length. As $\sqrt{2}$ is also a constructible number, we can construct the point with coordinates $(\sqrt{2}t, \sqrt{2}t)$, hence doubling the area.



What about a regular cube: is there a similar procedure? Suppose the original cube has side length 1, so that the task is to produce a new cube of *volume* 2. If this could be accomplished via a ruler and compass construction, then by setting the compass to the side length of the new cube, we would have constructed $\sqrt[3]{2}$. But the minimum polynomial over $\mathbb{Q}$ of $\sqrt[3]{2}$ is clearly $x^3 - 2$, with the extension $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2})$ thus having degree three. Thus, such a construction cannot be possible.

(10.6) The subset $\square^n$ of $\mathbb{R}^n$ given by

$$\square^n = \{x \in \mathbb{R}^n \mid |x_i| \leq \frac{t}{2} \text{ for all } i\}$$

is an n-dimensional cube of side length t having volume t^n. In particular, in 4-dimensions we have the so-called *hypercube*,

8-cell or hypercube, the vertices of which can be placed on the 3-sphere S^3 in $\mathbb{R}^4$. Stereographically projecting S^3 to $\mathbb{R}^3$ gives a picture as at right. It is the shadow cast by a hypercube on a 3-dimensional table top sitting in the 4-dimensional sun.

which can always be doubled in volume because the point with coordinates $(\sqrt[4]{2}t, \sqrt[4]{2}t, \sqrt[4]{2}t, \sqrt[4]{2}t)$ can be constructed!

(10.7) One of our basic ruler and compass constructions was to bisect an angle. It is therefore natural to ask if there is a construction that *trisects* an arbitrary angle. Certainly there are particular angles that can be trisected, for instance, if the angle ϕ is constructible then the angle 3ϕ can be trisected!

However, the angle $\pi/3$ cannot be trisected, as we show by demonstrating that the angle $\pi/9$ cannot be constructed.

Exercise 88 Evaluate the complex number $(\cos\phi + i\sin\phi)^3$ in two different ways: using the binomial theorem and De Moivre's theorem. By equating real parts, deduce that

$$\cos 3\phi = 4\cos^3\phi - 3\cos\phi.$$

Derive a similar expression for $\cos 5\phi$ and $\cos 7\phi$. What about the general case?

We have from Exercise **??** that the angle $\pi/9$ is constructible precisely when the complex number $\cos\pi/9$ can be constructed, for which it is necessary in turn that the degree of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\cos\pi/9)$ be a power of two. Using Exercise 88 with $\phi = \pi/9$ we get that,

$$\cos\frac{\pi}{3} = 4\cos^3\frac{\pi}{9} - 3\cos\frac{\pi}{9} \Leftrightarrow 1 = 8\cos^3\frac{\pi}{9} - 6\cos\frac{\pi}{9}.$$

Thus, if $u = 2\cos(\pi/9)$, then $u^3 - 3u - 1 = 0$. This polynomial is irreducible over $\mathbb{Q}$ by the reduction test applied with $p = 2$, so it is the minimum polynomial over $\mathbb{Q}$ of $2\cos(\pi/9)$. Thus, the extension $\mathbb{Q} \subseteq \mathbb{Q}(2\cos(\pi/9)) = \mathbb{Q}(\cos(\pi/9))$ has degree three, and the angle $\pi/9$ cannot be constructed.

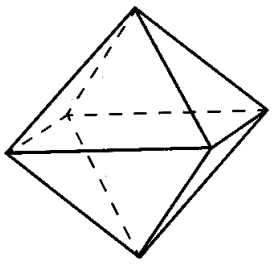
We will be able to say more about which angles of the form π/n can be constructed in §**15.**.

Exercise 89

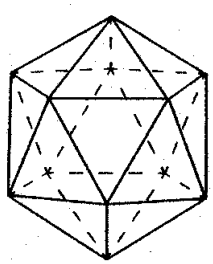
1. Can you construct an angle of 40^0?
2. Assuming 72^0 is constructible, what about 24^0 and 8^0?

Further Exercises for §**10.**

Exercise 90 The *octahedron* and *icosahedron* are two of the five Platonic solids.



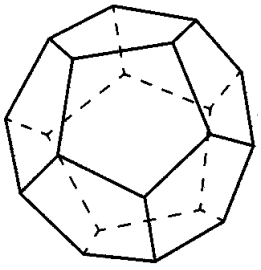
$$V_O = \frac{x^3\sqrt{2}}{3}$$



$$V_I = \frac{5x^3(3+\sqrt{5})}{12}$$

The volume of each is given by the formula, where x is the length of any edge. Show that in each case, there is no general method, using a ruler and compass, to construct a new solid from a given one, and having *twice* the volume.

Exercise 91 Consider a regular dodecahedron with volume as given.



$$V_D = \frac{x^3(15+7\sqrt{5})}{4}$$

Show that there is no general method, using a ruler and compass, to construct a new dodecahedron from a given one, and having *five times* the volume.

Exercise 92 Let S_O, S_D and S_I be the surface areas of the three Platonic solids of Exercise 90. If,

$$S_O = 2x^2\sqrt{3}, S_D = 3x^2\sqrt{5(5+2\sqrt{5})} \text{ and } S_I = 5x^2\sqrt{3},$$

determine whether or not a solid can be constructed from a given one with twice the surface area.

Exercise 93 Using the identity

$$\cos 5\theta = 16\cos^5\theta - 20\cos^3\theta + 5\cos\theta.$$

Show that is is impossible, using a ruler and compass, to *quinsect* (that is, divide into 5 equal parts) any angle ψ that satisfies,

$$\cos\psi = \frac{5}{6}$$

Exercise 94 Using the identity,

$$\cos 7\theta = 64\cos^7\theta - 112\cos^5\theta + 56\cos^3\theta - 7\cos\theta$$

show that it is impossible, using ruler and compass, to *septsect* (that is, divide into *seven* equal parts) any angle φ such that

$$\cos\varphi = \frac{7}{8}$$

Exercise 95 Show that if a general angle can be n-sected (that is, divided into n equal parts) then a regular n-gon can be constructed. Use this to re-deduce the result of the last exercise and to obtain conditions on a prime p, such that a general angle can be divided into p equal parts.

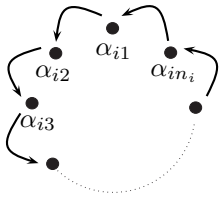
§11. Groups I: A Miscellany

As the title suggests, this section collects together some (carefully chosen) random facts about groups.

(11.1) A *permutation* of a set X is a bijection $X \to X$. Mostly we are interested in the case where X is finite, say $X = \{1, 2, \ldots, n\}$, so that a permutation is just a rearrangement of these numbers. Permutations are most compactly written using the cycle notation

$$(\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1n_1})(\alpha_{21}, \alpha_{22}, \ldots, \alpha_{2n_2}) \ldots (\alpha_{k1}, \alpha_{k2}, \ldots, \alpha_{kn_k})$$

where the α_{ij} are elements of $\{1, 2, \ldots, n\}$. Each $(\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in_i})$ indicates that the points are permuted in a cycle as



The cumulative effect of the cycles is obtained by dealing with them from right to left, eg: $(1, 2)(1, 2, 4, 3)$ $(1, 3)(2, 4) = (1, 2, 3)$. *A permutation can always be rewritten so that the points in the cycles are all distinct.*

The set of all permutations of the set X forms a group under composition of bijections called the *symmetric group* S_X, or S_n if $X = \{1, 2, \ldots, n\}$.

(11.2) Any permutation can be written as a composition of permutations where just two things are swapped, and everything else is left fixed. In other words, any permutation can be written as a composition of *transpositions* of the form (a, b):

$$(\alpha_1, \alpha_2, \ldots, \alpha_i) = (\alpha_1, \alpha_i)(\alpha_i, \alpha_{i-1}) \ldots (\alpha_1, \alpha_3)(\alpha_1, \alpha_2).$$

Indeed, much more is true: there may be several ways that a permutation can be decomposed into transpositions like this, and different ways may not involve the same number of transpositions, but any two such decompositions will either both involve an even number of transpositions or both an odd number.

We can thus, without any ambiguity, call a permutation *even* if it can be decomposed into an even number of transpositions, and *odd* otherwise. The *Alternating group* A_n consists of all those elements of S_n that are even.

Exercise 96 Show that A_n is indeed a group comprising exactly half of the elements of S_n. Show that the odd elements in S_n *do not* form a group.

Exercise 97 Recall that the *order* of an element g of a group G is the least n such that $g^n = 1$. Show that if G is Abelian then $(gh)^n = g^n h^n$ for every $g, h \in G$. Deduce that the order of gh is then the lowest common multiple of the orders of g and h.

(11.3) If G is a group and $\{g_1, g_2, \ldots, g_n\}$ are elements of G, then we say that the g_i *generate* G when every element $g \in G$ can be obtained as a product

$$g = g_{i_1}^{\pm 1} g_{i_2}^{\pm 1} \ldots g_{i_k}^{\pm 1},$$

of the g_i and their inverses. Write $G = \langle g_1, g_2, \ldots, g_n \rangle$.

(11.4) We find generators for the symmetric and alternating groups. Firstly, we have already seen that the transpositions (a, b) generate S_n, for any permutation can be written as a product

$$(\alpha_1, \alpha_2, \ldots, \alpha_i) = (\alpha_1, \alpha_i)(\alpha_i, \alpha_{i-1}) \ldots (\alpha_1, \alpha_3)(\alpha_1, \alpha_2).$$

Symbol	Name
$\mathbb{Z}_p$	cyclic
A_n	alternating

notes: p is a prime;
$n \neq 1, 2, 4$

Table 1: The simplest two families of simple groups

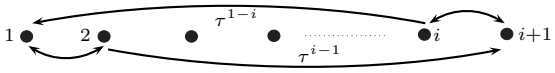
of transpositions. The transpositions themselves can be expressed in terms of just some of them: letting (i, j) be our transposition now with $i < j$, we have

$$(i, j) = (i, i+1)(i+1, i+2)\dots(j-1, j-2)(j, j-1)\dots(i+1, i+2)(i, i+1)$$

where the easiest way to see that this works is to consider the picture,

and perform the swaps indicated in the picture in the following order: do the swaps across the top first, from left to right, and then the swaps along the bottom from right to left. Any number strictly in between i and j moves one place to the right and then one place to the left, with net effect that it remains stationary. The point i is moved progressively along to j by the top swaps, but then stays there. Similarly j stays put for a while but is then moved progressively rightwards by the bottom swaps.

Substituting this new expression for each transposition gives any permutation in S_n as a product of transpositions of the form $(i, i+1)$. But in fact even these transpositions can be further reduced, by transferring i and $i+1$ to the two points 1 and 2, performing the swap between these two and transferring the answer back to i and $i+1$. Indeed if $\tau = (1, 2, \dots, n)$ then the picture,



gives $(i, i+1) = \tau^{i-1}(1, 2)\tau^{1-i}$ as τ^{i-1} sends 1 to i, 2 to $i+1$ and so on, while τ^{1-i} is its inverse.

The conclusion is that S_n is generated by *just two* permutations, namely $(1, 2)$ and $(1, 2, \dots, n)$.

Exercise 98 Show that the Alternating group is generated by the permutations of the form (a, b, c). Show that in fact just the 3-cycles of the form $(1, 2, i)$ will suffice.

(11.5) Lagrange's theorem tells us that if G is a finite group and H a subgroup of G, then the order $|H|$ of H divides the order $|G|$ of G. The converse, that if a *subset* of a group has size dividing the order of the group then it is a subgroup *is false*.

Exercise 99 By considering the Alternating group A_4, justify this statement.

Exercise 100 Show that if G is a cyclic group, then the converse to Lagrange's theorem *is* true, ie: if G has order n and k divides n then G has a subgroup of order k.

Exercise 101 Use Lagrange's Theorem to show that if a group G has order a prime number p, then G is isomorphic to a cyclic group. Thus, *any two groups of order p are isomorphic*.

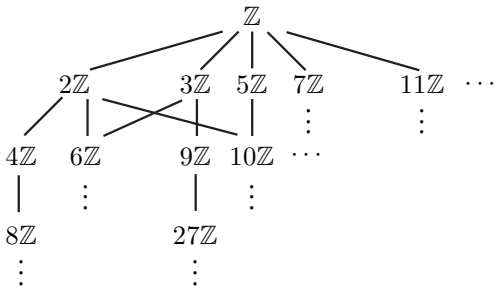
There is however a *partial* converse to Lagrange's Theorem, due to the Norwegian Peter Sylow[8]

Sylow's First Theorem. *Suppose G is a finite group of order $p^k m$, where p does not divide m (ie: k is the largest power of p dividing the order of G). Then G has a subgroup of order p^i for any $1 \leq i \leq k$.*

[8]pronounced Soo-lov, not Si-low.

(11.6) It is often useful to consider *all* the subgroups of a group at once, rather than just one at a time. The information is summarised in the *subgroup lattice*, which is a diagram depicting all the subgroups and the relations between them. Specifically, if H_1, H_2 are subgroups of G with $H_1 \subseteq H_2$, place H_2 higher in the diagram than H_1 with a line connecting them like so,

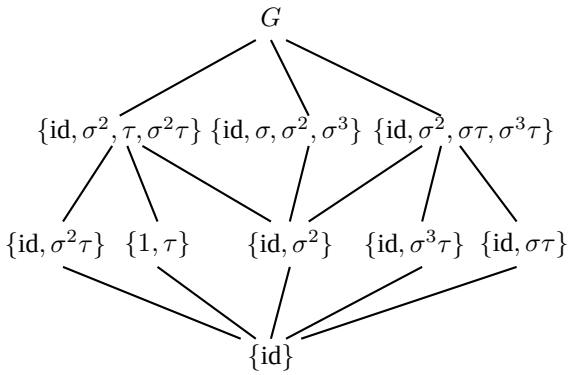
At the very base of the diagram is the trivial subgroup $\{\mathrm{id}\}$ and at the apex is the other trivial subgroup, namely G itself. For example, the subgroups of the integers $\mathbb{Z}$ all have the form $n\mathbb{Z}$ for some n (ie: the multiples of n) and arrange themselves into the lattice:



As another example, the group of symmetries of a square consists of the eight elements,

$$\{\mathrm{id}, \sigma, \sigma^2, \sigma^3, \tau, \sigma\tau, \sigma^2\tau, \sigma^3\tau\},$$

where σ is a rotation anticlockwise through $\frac{1}{4}$ of a turn and τ is a reflection in the horizontal axis. The subgroup lattice looks like,



(11.7) Suppose we have a finite group G and a sequence of subgroups $H_0 = \{1\}, H_1, \ldots, H_{n-1}, H_n = G$ arranged as follows:

$$\{1\} = H_0 \lhd H_1 \lhd \cdots \lhd H_{n-1} \lhd H_n = G,$$

ie: H_0 is a normal subgroup of H_1, which is in turn a normal subgroup of H_2, and so on. In fact, we can always ensure this if the group is finite: find a normal subgroup of G, then a normal subgroup of that normal subgroup, and so on. Eventually the process must stop with the identity subgroup.

Whenever we have normal subgroups we get new groups by taking the quotient. Given a sequence like the above then, we get a sequence of quotient groups,

$$H_1/H_0, H_2/H_1, \ldots, H_n/H_{n-1}.$$

In principle these quotient groups could be anything. In the quite special situation that they all turn out to be Abelian, call the group G *soluble*.

Symbol	Name	Discovered
$\mathrm{PSL}_n\mathbb{F}_q$	projective	1870
$\mathrm{PSP}_{2n}\mathbb{F}_q$	simplectic	1870
$\mathrm{P}\Omega^+_{2n}$	orthogonal	1870
$\mathrm{P}\Omega_{2n+1}$	orthogonal	1870
$E_6(q)$	Chevalley	1955
$E_7(q)$	Chevalley	1955
$E_8(q)$	Chevalley	1955
$F_4(q)$	Chevalley	1955
$G_2(q)$	Chevalley	1955
${}^2A_n(q^2) = \mathrm{PSU}_n\mathbb{F}_{q^2}$	unitary or twisted Chevalley	1870
${}^2D_n(q^2) = \mathrm{P}\Omega^-_{2n}$	orthogonal or twisted Chevalley	1870
${}^2E_6(q^2)$	twisted Chevalley	c. 1960
${}^3D_4(q^3)$	twisted Chevalley	c. 1960
${}^2B_2(2^{2e+1})$	Suzuki	1960
${}^2G_2(2^{2e+1})$	Ree	1961
${}^2F_4(2^{2e+1})$	Ree	1961
notes: n and e are $\in \mathbb{Z}$	There are some restrictions on n	
q is a prime power;	and q, left off here for clarity.	

Table 2: The simple groups of Lie type

(11.8) If G is an Abelian group, then consider the sequence of subgroups,

$$\{1\} \lhd G,$$

(note that the trivial subgroup is *always* a normal subgroup). There is only one quotient group to consider here, namely $G/\{1\} \cong G$, an Abelian group. Thus Abelian groups themselves are soluble, and indeed, one can think of solubility as a generalisation of Abelian.

(11.9) For another example, take the symmetries, both rotations and reflections, of a regular n-gon in the plane, and the sequence,

$$\{1\} \lhd \{\text{rotations}\} \lhd \{\text{rotations and reflections}\}$$

To convince ourselves first of all that this is indeed a proper sequence, we need that the rotations form a normal subgroup of the full group of symmetries. That they form a subgroup is not hard to see, and the normality follows from the fact that the rotations comprise half of all the symmetries and Exercise 110.

Moreover, the rotations are isomorphic as a group to the cyclic group $\mathbb{Z}_n$, and so the quotients of this sequence are

$$\{\text{rotations}\}/\{1\} \cong \{\text{rotations}\} \cong \mathbb{Z}_n \text{ and } \{\text{rotations and reflections}\}/\{\text{rotations}\} \cong \mathbb{Z}_2,$$

both Abelian groups. Thus the dihedral groups are soluble[9].

(11.10) It turns out, although for quite technical reasons (see the next couple of exercises) that a subgroup of a soluble group is also soluble.

Exercise 102 Let H be a subgroup and N a normal subgroup of some group G and,

$$NH = \{nh \,|\, n \in N, h \in H\}.$$

[9]Groups like this, where you have a 2-step sequence $\{1\} \lhd H \lhd G$, with Abelian quotients are sometimes called *meta-Abelian.*

1. Define a map $\varphi : H \to NH/N$ by $\varphi(h) = Nh$. Show that φ is an onto homomorphism with kernel $N \cap H$.
2. Use the first isomorphism theorem for groups to deduce that $H/H \cap N$ is isomorphic to NH/H.

(This is called the *second isomorphism* or *diamond isomorphism* theorem. Why diamond? Draw a picture of all the subgroups–the theorem says that two "sides" of a diamond are isomorphic).

Exercise 103 Let G be a soluble group with series,

$$\{1\} = H_0 \lhd H_1 \lhd \cdots \lhd H_{n-1} \lhd H_n = G,$$

and K a subgroup of G. Intersect K with all the H_i and use the second isomorphism theorem to show that

$$\{1\} = H_0 \cap K \lhd H_1 \cap K \lhd \cdots \lhd H_{n-1} \cap K \lhd H_n \cap K = K,$$

is a series with Abelian quotients for K, hence K is soluble too.

(11.11) In some sense the antithesis of the soluble groups are the *simple* groups: groups G whose only normal subgroups are the trivial subgroup $\{1\}$ and the whole group G. These two are always normal subgroups, so one could say that a group is simple when it has no *non-trivial* normal subgroups.

Whenever we have normal subgroups we can take quotients, so another way of putting it to say that a group is simple whenever its only quotients are itself $G/\{1\} \cong G$ and the trivial group $G/G \cong \{1\}$. In this way simple groups are analogous to prime numbers, which are integers whose only quotients[10] are themselves $p/1 = p$ and the trivial integer $p/p = 1$.

The reason that simple groups are at the opposite end of a spectrum to soluble ones is this: if G is non-Abelian and simple, then G *cannot* be soluble. For, the only sequence of normal subgroups that G can have is

$$\{1\} \lhd G,$$

and as G is non-Abelian the quotients of this sequence are non-Abelian. Thus, non-Abelian simple groups provide a ready source of non-soluble groups.

(11.12) So what are these groups then? Amazingly, there is a complete list, compiled over approximately 150 years, through the efforts of over a 100 mathematicians, and running to roughly 15000 pages of research articles. It is quite possibly the greatest taxonomic (if not necessarily conceptual) achievement of 20th Century Mathematics. The list is contained in Tables 1-3.

Exercise 104 Show that if p is a prime number then the cyclic group $\mathbb{Z}_p$ has *no non-trivial subgroups whatsoever*, and so is certainly a simple group.

(11.13) Looking at Table 1 we see that the Alternating groups A_n are simple for $n \neq 1, 2$ or 4. Thus these Alternating groups are not soluble, and as any subgroup of a soluble group is soluble, any group *containing* the Alternating group will also not be soluble. Thus, *the symmetric groups S_n are not soluble if $n \neq 1, 2$ or 4.*.

Exercise 105 Show that the previous statement is *not quite* correct in that the symmetric group S_3 is soluble.

(11.14) Tables 2 and 3 list the really interesting simple groups. The groups of Lie type are basically groups of matrices whose entries come from finite fields. We have already seen that if $q = p^n$ a prime power, then there is a field $\mathbb{F}_q$ with $q = p^n$ elements. The group $\mathrm{SL}_n\mathbb{F}_q$ consists of the $n \times n$ matrices with entries from this field and the usual matrix multiplication. Unfortunately this group is not simple as the subset

$$N = \{\lambda I_n \mid \lambda \in \mathbb{F}_q\},$$

consisting of all scalar multiples of the identity matrix forms a normal subgroup. But it turns out that this is the biggest normal subgroup you can find in the sense that the quotient group,

$$\mathrm{SL}_n\mathbb{F}_q/N,$$

[10] Obviously the way it is normally put is to say that the only divisors are itself and one, but as the notion of divisor does not carry over quite so easily to group theory, we use quotients instead.

Symbol	Name	Discovered	Order
1. *First generation of the Happy Family*			
M_{11}	Mathieu	1861	$2^4\,3^2\,5\,11$
M_{12}	Mathieu	1861	$2^4\,3^3\,5\,11$
M_{22}	Mathieu	1873	$2^7\,3^2\,5\,7\,11$
M_{23}	Mathieu	1873	$2^7\,3^2\,5\,7\,11\,23$
M_{24}	Mathieu	1873	$2^{10}\,3^3\,5\,7\,11\,23$
2. *Second generation of the Happy Family*			
HJ	Hall-Janko	1968	$2^7\,3^3\,5^2\,7$
HiS	Higman-Sims	1968	$2^9\,3^2\,5^3\,7\,11$
McL	McLaughlin	1969	$2^7\,3^6\,5^3\,7\,11$
Suz	Suzuki	1969	$2^{13}3^7\,5^2\,7\,11\,13$
Co_1	Conway	1969	$2^{21}\,3^9\,5^4\,7^2\,11\,13\,23$
Co_2	Conway	1969?	$2^{18}\,3^6\,5^3\,7\,11\,23$
Co_3	Conway	1969?	$2^{10}\,3^7\,5^3\,7\,11\,23$
3. *Third generation of the Happy Family*			
He	Held	1968	$2^{10}\,3^2\,5^2\,7^3\,17$
Fi_{22}	Fischer	1968	$2^{17}\,3^9\,5^2\,7\,11\,13$
Fi_{23}	Fischer	1968	$2^{18}\,3^{13}\,5^2\,7\,11\,13\,17\,23$
Fi_{24}	Fischer	1968	$2^{21}\,3^{16}\,5^2\,7^3\,11\,13\,17\,23\,29$
F_5	Harada-Norton	1973	$2^{14}\,3^6\,5^6\,7\,11\,19$
F_3	Thompson	1973	$2^{15}\,3^{10}\,5^3\,7^2\,13\,19\,31$
F_2	Fischer or “Baby Monster”	1973	$2^{41}\,3^{13}\,5^6\,7^2\,11\,13\,17\,19\,23\,47$
$\mathbb{M}$	Fischer-Griess or “Friendly Giant” or “Monster”	1973	$\approx 10^{55}$
4. *The Pariahs.*			
J_1	Janko	1965	$2^3\,5\,7\,11\,19$
J_3	Janko	1968	$2^7\,3^5\,5\,17\,19$
J_4	Janko	1975	$2^{21}\,3^3\,5\,7\,11^3\,23\,29\,31\,37\,43$
Ly	Lyons	1969	$2^8\,3^7\,5^6\,7\,11\,31\,37\,67$
Ru	Rudvalis	1972	$2^{14}\,3^3\,5^3\,7\,13\,29$
O’N	O’Nan	1973	$2^9\,3^4\,5\,7^3\,11\,19\,31$

Table 3: The sporadic simple groups

has no non-trivial normal subgroups, ie: is a simple group. It is denoted $\mathrm{PSL}_n\mathbb{F}_q$, and called the n-dimensional projective[11] special linear group over $\mathbb{F}_q$. The remaining groups in Table 2 come from more complicated constructions.

Table 3 lists groups that don't seem to fall into any of the categories described so far–for this reason they are called the "sporadic" simple groups. They arise from various (often quite complicated) constructions that are well beyond the remit of these notes. The most interesting of them is the largest one–the Monster simple group (which actually contains quite a few of the others as subgroups). The Monster has a number of fascinating connections with a diverse range of mathematical areas, including number theory (where it plays a central role in something called "Monstrous Moonshine") and even Mathematical Physics.

All of this notwithstanding, the simple groups in Tables 2 and 3 are all non-Abelian, hence provide ready examples of non-soluble groups.

Further Exercises for §**11.**

Exercise 106 Show that any subgroup of an abelian group is normal.

Exercise 107 Let G be a group. Show that $G/G \cong \{\mathrm{id}\}$ and $G/\{\mathrm{id}\} \cong G$.

Exercise 108 Let n be a positive integer that is *not* prime (sometimes called a composite integer). Show that the cyclic group $\mathbb{Z}_n$ is *not* simple.

Exercise 109 Show that A_2 and A_4 are not simple groups, but A_3 is.

Exercise 110 Let G be a group and H a subgroup such that H has exactly *two* cosets in G. Let C_2 be the group of order two with elements $\{-1, 1\}$ and operation just usual multiplication. Define a map $f : G \longrightarrow C_2$ by

$$f(g) = \begin{cases} 1 & g \in H \\ -1 & g \notin H \end{cases}$$

1. Show that f is a homomorphism.
2. Deduce that H is a normal subgroup.

Exercise 111 Consider the group of symmetries (rotations and reflections) of a regular n-sided polygon for $n \geq 3$. Show that this is not a simple group.

Exercise 112 Show that S_2 is simple but S_n isn't for $n \geq 3$. Show that A_n has no subgroups of index 2 for $n \geq 5$.

Exercise 113 Show that if G is abelian and simple then it is cyclic. Deduce that if G is simple and not $\mathbb{Z}_p$ then G is non-Abelian.

Exercise 114 For each of the following groups G, draw the subgroup lattice $\mathscr{L}_G$:

1. $G =$ the group of symmetries of a square, pentagon or hexagon.
2. $G =$ the cyclic group $\{1, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$ where $\sigma^i\sigma^j = \sigma^{i+j \bmod n}$ and $\sigma^n = 1$.

[11] the name "projective" comes fom the fact that the group is the group of symmetries of projective geometry over the field $\mathbb{F}_q$.

§12. Groups II: Symmetries of Fields

We are finally in a position to introduce the idea of symmetry into the solutions of polynomial equations.

(12.1) An *automorphism* of a field F is an isomorphism $\sigma : F \to F$, ie: a bijective map from F to F such that $\sigma(a+b) = \sigma(a) + \sigma(b)$ and $\sigma(ab) = \sigma(a)\sigma(b)$ for all $a, b \in F$.

We commented in §**4.** that an isomorphism of fields (indeed of any algebraic object) was just a relabelling of the elements using different symbols. The "algebra" is identical though. An automorphism is then a relabelling that is achieved merely by moving the elements of F around amongst themselves. So it is a way of picking the field up and placing it down upon itself so it looks like the same field: it is thus a *symmetry* of the field.

Exercise 115 Show that if σ is an automorphism of the field F then $\sigma(0) = 0$ and $\sigma(1) = 1$.

(12.2) A familiar example of a field symmetry/automorphism is complex conjugation: the map $z \mapsto \overline{z}$ is an automorphism of $\mathbb{C}$, for, from elementary complex analysis we have,

$$\overline{z+w} = \overline{z} + \overline{w} \text{ and } \overline{zw} = \overline{z}\,\overline{w},$$

with conjugation a bijection $\mathbb{C} \to \mathbb{C}$. This symmetry captures the idea that from an algebraic point of view, we could have just as easily adjoined $-i$ to $\mathbb{R}$, rather than i, to obtain the complex numbers (they look the same upside down as right side up!).

We will see at the end of this section that if a non-trivial automorphism of $\mathbb{C}$ fixes pointwise the real numbers, then it must be complex conjugation. If we drop the requirement that $\mathbb{R}$ be fixed then there may be more possibilities: if we only insist that $\mathbb{Q}$ is fixed pointwise, there are infinitely many.

(12.3) Every field F has a prime subfield that is either $\mathbb{F}_p$ or $\mathbb{Q}$. Every element of the prime subfield has the form,

$$\frac{\overbrace{1+1+\cdots+1}^{m \text{ times}}}{\underbrace{1+1+\cdots+1}_{n \text{ times}}}.$$

If σ is now an automorphism of F we have

$$\sigma\left(\frac{\overbrace{1+1+\cdots+1}^{m \text{ times}}}{\underbrace{1+1+\cdots+1}_{n \text{ times}}}\right) = \sigma(\overbrace{1+1+\cdots+1}^{m})\sigma\left(\frac{1}{\underbrace{1+1+\cdots+1}_{n}}\right)$$

$$= (\overbrace{\sigma(1)+\sigma(1)+\cdots+\sigma(1)}^{m})\left(\frac{1}{\underbrace{\sigma(1)+\sigma(1)+\cdots+\sigma(1)}_{n}}\right) = \frac{\overbrace{1+1+\cdots+1}^{m \text{ times}}}{\underbrace{1+1+\cdots+1}_{n \text{ times}}}.$$

as $\sigma(1) = 1$. Thus the elements of the prime subfield are *fixed pointwise* by the automorphism.

Exercise 116 We saw above that the map $a + bi \mapsto a - bi$ is an automorphism of $\mathbb{C}$. Show that $a + bi \mapsto -a + bi$ is not an automorphism of $\mathbb{C}$.

(12.4) Symmetries of things normally arrange themselves into a group, and field symmetries are no exception. We could talk just of the symmetry group of a field, but it turns out to be more instructive to make a slightly more elaborate definition that takes into consideration not just fields, but their extensions:

Definition. Let $F \subseteq E$ be an extension of fields. The automorphisms of the field E that fix pointwise the elements of F form a group under composition called the *Galois group of E over F* and denoted $\mathrm{Gal}\,(E/F)$.

Thus an element σ of Gal (E/F) has the property that $\sigma(\lambda) = \lambda$ for all $\lambda \in F$.

Exercise 117 For $F \subset E$ fields, show that the set of automorphisms Gal (E/F) of E that fix F pointwise do indeed form a group under composition.

(12.5) Consider as an example the field $\mathbb{Q}(\sqrt{2}, i)$. From the proof of the tower law, a basis for this field over $\mathbb{Q}$ is given by $\{1, \sqrt{2}, i, \sqrt{2}i\}$, so that the elements of the field are, by Theorem D,

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

Suppose we have a symmetry $\sigma \in$ Gal $(\mathbb{Q}(\sqrt{2}, i)/\mathbb{Q})$ and consider its effect on a typical element,

$$\sigma(a+b\sqrt{2}+ci+d\sqrt{2}i) = \sigma(a)+\sigma(b)\sigma(\sqrt{2})+\sigma(c)\sigma(i)+\sigma(d)\sigma(\sqrt{2}i) = a+b\sigma(\sqrt{2})+c\sigma(i)+d\sigma(\sqrt{2}i)$$

using the properties of automorphisms and the fact that σ fixes rational numbers. Thus, the symmetry σ *is completely determined by its effect on the basis elements* $\{1, \sqrt{2}, i, \sqrt{2}i\}$, in that once their images are decided, then σ is uniquely known.

Aside. We can see that this is really no surprise. When we have fields $F \subseteq E$, then among other things, E is a vector space over F. Given a symmetry $\sigma \in$ Gal (E/F), then this is, among other things, a linear map of vector spaces $E \to E$, and we know from linear algebra that such things are completely determined by their effect on a basis.

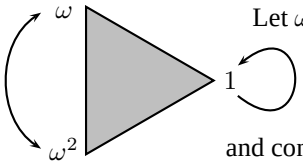
Actually we can say even more. Clearly $\sigma(1) = 1$ is always true, and $\sigma(\sqrt{2}i) = \sigma(\sqrt{2})\sigma(i)$. Thus the symmetry σ *is completely determined by its effect on* $\sqrt{2}$ *and* i, the elements adjoined to $\mathbb{Q}$.

(12.6) And indeed this is a general fact. If $F \subseteq F(\alpha_1, \alpha_2, \ldots, \alpha_k) = E$ and $\sigma \in$ Gal (E/F), then σ is completely determined by its effect on the $\alpha_1, \ldots, \alpha_k$. For, suppose that $\{\beta_1, \ldots, \beta_n\}$ is a basis for E over F, so that σ is completely determined as above by its effect on the β_i. From the proof of the tower law, each β_i is a product of the form,

$$\beta_i = \alpha_1^{i_1}\alpha_2^{i_2}\ldots\alpha_k^{i_k},$$

and so $\sigma(\beta_i) = \sigma(\alpha_1)^{i_1}\sigma(\alpha_2)^{i_2}\ldots\sigma(\alpha_k)^{i_k}$. Thus $\sigma(\beta)$ in turn is determined by the $\sigma(\alpha_i)$.

(12.7) The structure of Galois groups can sometimes be determined via ad-hoc arguments, at least in very simple cases.



Let ω be the primitive cube root of 1,

$$\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i,$$

and consider the extension $\mathbb{Q} \subseteq \mathbb{Q}(\omega)$. Although the most obvious polynomial that ω is a root of is $x^3 - 1$, this is reducible, so the minimum polynomial of ω over $\mathbb{Q}$ is in fact $x^2 + x + 1$ (see Exercise 29 where we showed $1 + x + x^2 + \cdots + x^{p-1}$ to be irreducible over $\mathbb{Q}$ for p a prime). Thus by Theorem D, $\mathbb{Q}(\omega) = \{a + b\omega \mid a, b \in \mathbb{Q}\}$, giving that $\mathbb{Q}(\omega)$ is 2-dimensional over $\mathbb{Q}$ with basis $\{1, \omega\}$. Suppose that $\sigma \in$ Gal $(\mathbb{Q}(\omega)/\mathbb{Q})$, whose effect we now know is completely determined by where it sends ω. Suppose $\sigma(\omega) = a + b\omega$ for some $a, b \in \mathbb{Q}$ to be determined. On the one hand we have $\sigma(\omega^3) = \sigma(1) = 1$, while on the other,

$$\sigma(\omega^3) = \sigma(\omega)^3 = (a + b\omega)^3 = (a^3 + b^3 - 3ab^2) + (3a^2b - 3ab^2)\omega.$$

The last bit uses the fact that $\omega^2 = -\omega - 1$.

One of the consequences of $\{1, \omega\}$ being a basis for $\mathbb{Q}(\omega)$ over $\mathbb{Q}$ is that elements have *unique* expressions as linear combinations of these two basis elements (this is a consequence of linear independence). This means that given two expressions for an element as linear combinations of 1 and ω, we can "equate the 1 and ω parts[12]". Thus,

$$1 = \sigma(\omega^3) = (a^3 + b^3 - 3ab^2) + (3a^2b - 3ab^2)\omega, \text{ so that } a^3 + b^3 - 3ab^2 = 1 \text{ and } 3a^2b - 3ab^2 = 0.$$

[12] Just as we equate real and imaginary parts of complex numbers, and for the same reason: $\{1, i\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$. On the other hand, we could *not* do this for two expressions in terms of $1, \omega$ and ω^2

Solving these last two equations (in $\mathbb{Q}$!) gives three solutions $a = 0, b = 1$, $a = 1, b = 0$ and $a = -1, b = -1$, coresponding to $\sigma(\omega) = \omega$ and $\sigma(\omega) = -1 - \omega = \omega^2$ (the middle solution gives $\sigma(\omega) = 1$ which is impossible as σ is a bijection and we already have that $\sigma(1) = 1$). Thus the Galois group $\mathrm{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q}) = \{\mathrm{id}, \sigma\}$ has order two.

(12.8) Our first major tool for unpicking the structure of Galois groups is,

Theorem F (The Extension Theorem). *Let F_1, F_2 be fields and $\tau : F_1 \to F_2$ an isomorphism between them with $\tau^* : F_1[x] \to F_2[x]$ the resulting homomorphism of rings given by $\tau^*(\sum a_i x^i) = \sum \tau(a_i) x^i$. If α is algebraic over F_1, then the isomorphism τ extends to an isomorphism $\sigma : F_1(\alpha) \to F_2(\beta)$ with $\sigma(\alpha) = \beta$ if and only if β is a root of $\tau^*(f)$, where f is the minimum polynomial of α over F_1.*

The elements α and β are assumed to lie in some extensions $F_i \subseteq E_i$ of the two fields, and when we say that τ extends to σ we mean that $\sigma|_F = \tau$.

$$\begin{array}{ccc} F(\alpha) & \xrightarrow{\sigma} & F(\alpha) \\ \uparrow & & \uparrow \\ F & \xrightarrow{\mathrm{id}} & F \end{array}$$

The theorem is quite technical, but nevertheless has an intuitive meaning. Suppose we have the special case where $F_1 = F_2 = F$ and σ is the identity isomorphism (hence σ^* is also the identity map). Then we have an extension $\widehat{\sigma} : F(\alpha) \to F(\beta)$ *precisely when β is a root of the minimum polynomial f of α over F*. Indeed we can say even more: if β is an element of $F(\alpha)$, then $F(\beta) \subseteq F(\alpha)$, is an F-vector subspace, but since f must also be the minimum polynomial of β, $F(\beta)$ is $(\deg f)$-dimensional over F, just like $F(\alpha)$, and so $F(\beta) = F(\alpha)$. Thus $\widehat{\sigma}$ is an automorphism of $F(\alpha)$ fixing F pointwise, and so an element of the Galois group $\mathrm{Gal}\,(F(\alpha)/F)$. Summarising everything we know about Galois groups so far,

Corollary. *Let α be algebraic over F with minimum polynomial f over F. A map $\sigma : F(\alpha) \to F(\alpha)$ is an element of the Galois group $\mathrm{Gal}\,(F(\alpha)/F)$ if and only if for any $\sum_{k=0}^{\deg f - 1} a_k \alpha^k \in F(\alpha)$ we have,*

$$\sigma\left(\sum_{k=0}^{\deg f - 1} a_k \alpha^k\right) = \sum_{k=0}^{\deg f - 1} a_k \beta^k,$$

where β is also a root of f contained in $F(\alpha)$.

Thus the elements of the Galois group permute the roots of the minimum polynomial that are contained in $F(\alpha)$ amongst themselves.

Proof of the Extension Theorem: We give a "grungy" proof that nevertheless makes the situation nice and concrete. For the only if part, we have that if $f = \sum a_i x^i$ with $f(\alpha) = 0$, then $\sum a_i \alpha^i = 0$ in E_1 so that

$$\sigma\left(\sum a_i \alpha^i\right) = 0 \Rightarrow \sum \sigma(a_i)\sigma(\alpha)^i = 0 \Rightarrow \sum \tau(a_i)\beta^i = 0 \text{ in } E_2,$$

giving that β is a root of $\tau^*(f)$ as claimed.

For the if part, we need to define an isomorphism with the desired properties. The elements of $F(\alpha)$ all have the form $\sum_{i=0}^{d-1} a_i \alpha^i$, where $d = \deg f$. Define σ by

$$\sigma\left(\sum_{i=0}^{m} a_i \alpha^i\right) = \sum_{i=0}^{m} \tau(a_i)\beta^i, \tag{4}$$

for any m. From this definition we see that $\sigma(a) = \tau(a)$ for any $a \in F_1$ and also that $\sigma(\alpha) = \beta$.

The proof then proceeds in three parts. *σ is well-defined and 1-1*: Suppose we have two expressions

$$\sum a_i \alpha^i = \sum b_i \alpha^i,$$

representing the same element of $F_1(\alpha)$. Thus $\sum (a_i - b_i)\alpha^i = 0$ giving that α is a root of the polynomial $g = \sum (a_i - b_i) x^i \in F_1[x]$. As f is the minimum polynomial of α over F_1 we must have that f is a factor of g, and so,

$$g = fh \Leftrightarrow \tau^*(g) = \tau^*(fh) = \tau^*(f)\tau^*(h).$$

Hence $\tau^*(f)$ is a factor of $\tau^*(g)$. As β is a root of $\tau^*(f)$ it must then also be a root of $\tau^*(g)$, ie;

$$\tau^*(g)(\beta) = 0 \Leftrightarrow \sum \tau(a_i - b_i)\beta^i = 0 \Leftrightarrow \sum \tau(a_i)\beta^i = \sum \tau(b_i)\beta^i \Leftrightarrow \sigma\left(\sum a_i\alpha^i\right) = \sigma\left(\sum b_i\alpha^i\right).$$

Thus the two expressions for the same element are sent to the same element of $F_2(\beta)$, giving that σ is well-defined. This is the "only if" part of all the equivalences above, with the "if" part giving that σ is 1-1.

σ is a homomorphism: we need to show that σ respects the addition and multiplication in the field $F_1(\alpha)$. Let

$$\lambda = \sum_{i=0}^{n} a_i\alpha^i \text{ and } \mu = \sum_{i=0}^{m} b_i\alpha^i,$$

be two elements. Then

$$\sigma(\lambda+\mu) = \sigma\left(\sum_{i=0}^{\max\{m,n\}} (a_i+b_i)\alpha^i\right) = \sum_{i=0}^{\max\{m,n\}} \tau(a_i+b_i)\beta^i$$
$$= \sum_{i=0}^{n} \tau(a_i)\beta^i + \sum_{i=0}^{m} \tau(b_i)\beta^i = \sigma(\lambda) + \sigma(\mu).$$

Similarly,

$$\sigma(\lambda\mu) = \sigma\left(\sum_{k=0}^{n+m}\left(\sum_{i+j=k} a_ib_j\right)\alpha^k\right) = \sum_{k=0}^{n+m}\tau\left(\sum_{i+j=k} a_ib_j\right)\beta^k = \sum_{k=0}^{n+m}\left(\sum_{i+j=k}\tau(a_i)\tau(b_j)\right)\beta^k$$
$$= \sum_{i=0}^{n}\tau(a_i)\beta^i \sum_{j=0}^{m}\tau(b_j)\beta^j = \sigma(\lambda)\sigma(\mu).$$

One comment: in both cases we had σ of an expression, and we replaced this by the definition given at (4). Certainly in the case of multiplication, the expression was quite possibly not of the form a polynomial in α of degree $< d$. If we had defined σ for just these expressions we wouldn't have been able to use (4) as it stands. Thus we defined σ for *any* expression, but this then requires we show the definition to be well-defined, for an element has many different expressions as polynomials in α, if we relax the condition that these expressions have degree $< d$.

σ is onto: Certainly we have $\sigma(F_1(\alpha))$ is contained in $F_2(\beta)$ by the definition at (4)–the right hand side is contained in $F_2(\beta)$. On the otherhand, any $\mu \in F_2$ arises as the image $\tau(\lambda)$ for some $\lambda \in F_1$, as τ is onto. Also $\beta = \sigma(\alpha)$ by definition, so $F_2, \beta \in \sigma(F_1(\alpha))$, hence $F_2(\beta) \subseteq \sigma(F_1(\alpha))$, giving that the image $\sigma(F_1(\alpha))$ is $F_2(\beta)$. □

(12.9) If we compute instead the Galois group of the extension $\mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2})$, we have the freedom to send $\sqrt[3]{2}$ to those roots of its minimum polynomial over $\mathbb{Q}$ that are also contained in the field $\mathbb{Q}(\sqrt[3]{2})$. This minimum polynomial is $x^3 - 2$ which has the roots $\alpha, \alpha\omega$ and $\alpha\omega^2$ for $\alpha = \sqrt[3]{2}$ and

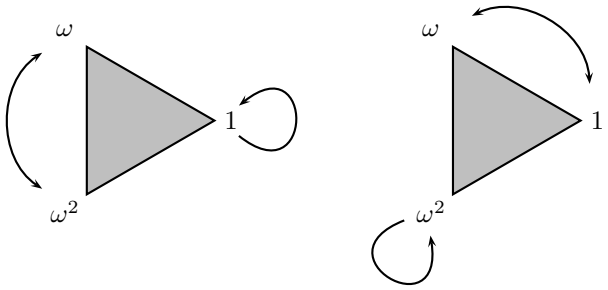
$$\omega = -\frac{1}{2} + \frac{\sqrt[3]{2}}{2}i.$$

But now the roots $\alpha\omega$ and $\alpha\omega^2$ are not contained in $\mathbb{Q}(\sqrt[3]{2})$ as this field contains only real numbers while these roots are clearly non-real. Thus the only possible image for α is α itself, giving that Gal $(\mathbb{Q}(\alpha)/\mathbb{Q})$ is just the trivial group.

(12.10) Returning to the example we calculated in an ad-hoc fashion immediately before the extension theorem, any automorphism of $\mathbb{Q}(\omega)$ that fixes $\mathbb{Q}$ pointwise is determined by where it sends ω, and this must be to a root of the minimum polynomial over $\mathbb{Q}$ of ω. As this polynomial is $1 + x + x^2$ with roots ω and ω^2, we get that the possible automorphisms send ω to itself or to ω^2, ie:

$$\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q}) = \{\text{id}, \sigma\},$$

where $\sigma(a+b\omega) = a+b\omega^2$ is in fact just complex conjugation. This answers Exercise 3, in showing that the left hand figure depicts an automorphism, but the right hand figure does not:



(12.11) The only if part of the proof of the Extension Theorem is usefully stated as a separate,

Corollary. *If g is a polynomial with F-coefficients and with a root $\alpha \in E$, then for any $\sigma \in Gal\,(E/F)$, the image $\sigma(\alpha)$ is also a root of g.*

An immediate and important consequence is,

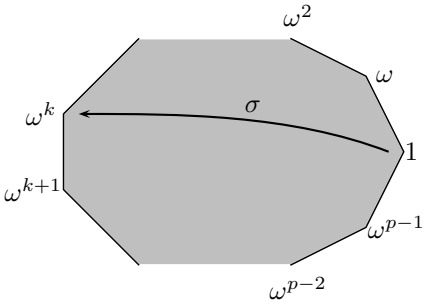
Corollary. *If $F \subseteq E$ is a finite extension then the Galois group Gal (E/F) is finite.*

Proof: If $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ is a basis for E over F, then we have $E = F(\alpha_1, \alpha_2, \ldots, \alpha_k)$, and by Proposition 3, each of the α_i is algebraic over F with some minimum polynomial $f_i \in F[x]$. If σ is an element of the Galois group, then σ is completely determined by where it sends each α_i, for which there are only finitely many possiblities: to the roots of f_i. □

(12.12) Let p be a prime and let

$$\omega = \cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p},$$

be a root of 1.



By the Extension Theorem we have an element of the Galois group Gal $(\mathbb{Q}(\omega)/\mathbb{Q})$ precisely when it sends ω to some root contained in $\mathbb{Q}(\omega)$ of its minimum polynomial over $\mathbb{Q}$. The minimum polynomial is the p-th cyclotomic polynomial,

$$\Phi_p = 1 + x + x^2 + \cdots + x^{p-1},$$

as we saw in Exercise 29, with roots the other roots of 1 (except for 1 itself) namely $\omega, \omega^2, \ldots, \omega^{p-1}$. Clearly all these roots are contained in $\mathbb{Q}(\omega)$, and so we are free to send ω to any one of them. Thus, the Galois group has order $p-1$, with an element corresponding to each of the possible images of ω. If $\sigma(\omega) = \omega^k$ then $\sigma^i(\omega) = \omega^{k^i}$, where $\omega^p = 1$.

We saw in §**9.** that the multiplicative group of the finite field $\mathbb{F}_p$ is cyclic. In otherwords, there is a k with $1 < k < p$, such that the powers k^i of k exhaust all of the non-zero elements of $\mathbb{F}_p$, ie: the powers k^i run through $\{1, 2, \ldots, p-1\}$ modulo p, or k generates $\mathbb{F}_p^*$.

Putting the previous two paragraphs together, if we take a σ in the Galois group with $\sigma(\omega) = \omega^k$ for k a generator of $\mathbb{F}_p^*$, then the elements,

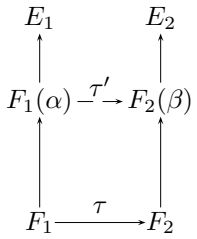
$$\sigma(\omega), \sigma^2(\omega), \ldots, \sigma^{p-1}(\omega),$$

run through the roots $\{\omega, \omega^2, \ldots, \omega^{p-1}\}$. Thus the elements $\sigma, \sigma^2, \ldots, \sigma^{p-1}$ exhaust the Galois group, and so *the Galois group of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\omega)$ is cyclic*. This (I imagine) is the reason behind the term cyclotomic.

(12.13) The extension theorem gives the *existence* of extended automorphisms, but also indicates the number of such extensions: there is one for each root of $\sigma^*(f)$ contained in E_2. Making this more precise:

Theorem G. *Let $\tau : F_1 \to F_2$ be an isomorphism and $F_1 \subseteq E_1$ and $F_2 \subseteq E_2$ be extensions with E_1 a splitting field of some polynomial f over F_1 and E_2 a splitting field of $\tau^*(f)$ over F_2. Assume that the roots of $\tau^*(f)$ in E_2 are distinct. Then the number of extensions of τ to an isomorphism $\sigma : E_1 \to E_2$ is equal to the degree of the extension $F_2 \subseteq E_2$.*

Proof: The proof proceeds by induction and the Extension Theorem.



We have that $E_1 = F_1(\alpha_1, \ldots, \alpha_{\deg f})$ as it is a splitting field for f over F_1. Let $\alpha = \alpha_1$ and consider the extension $F_1 \subseteq F_1(\alpha)$ and the picture at left where β is some element of E_2. We have by the Extension Theorem that τ extends to an isomorphism $\tau' : F_1(\alpha) \to F_2(\beta)$ if and only if β is a root in E_2 of $\tau^*(p)$, where p is the minimum polynomial of α over F_1. In particular, $F_2(\beta)$ is isomorphic as a vector space over F_2 to the vector space $F_1(\alpha)$ over F_1 ($\cong F_2$), and so they must have the same dimension. Thus any polynomial that has β as a root, has, by Theorem D, degree at least that of p, so that $\tau^*(p)$ does. On the otherhand we always have $\deg \tau^*(p) < \deg p$. Thus $\tau^*(p)$ must be the minimum polynomial of β over F_2.

As α is a root of f we have that p divides f, ie: $f = ph$ in $F_1[x]$ so that $\tau^*(f) = \tau^*(p)\tau^*(h)$ in $F_2[x]$ giving that $\tau^*(p)$ divides $\tau^*(f)$. As the roots of $\tau^*(f)$ are distinct, those of $\tau^*(p)$ must be too.

Thus the number of possible extensions τ', which is equal to the number of distinct roots of $\tau^*(p)$, must in fact be equal to the degree of $\tau^*(p)$, which is in turn the degree of the extension $[F_2(\beta) : F_2] > 1$. By the tower law,

$$[E_2 : F_2] = [E_2 : F_2(\beta)][F_2(\beta) : F_2],$$

and by induction, any isomorphism $\tau' : F_1(\alpha) \to F_2(\beta)$ will have

$$[E_2 : F_2(\beta)] = \frac{[E_2 : F_2]}{[F_2(\beta) : F_2]},$$

extensions to an isomorphism $\sigma : E_1 \to E_2$. Finally then, starting from the very bottom, τ extends to $[F_2(\beta) : F_2]$ possible τ''s, and extending each in turn gives,

$$[F_2(\beta) : F_2]\frac{[E_2 : F_2]}{[F_2(\beta) : F_2]} = [E_2 : F_2],$$

extensions in total. □

The reason that the roots of $\tau^*(f)$ need to be distinct is that we can then relate the number of automorphisms to degrees of extensions by passing through the midway house of the roots of polyomials. If the polynomial has repeated roots then the number of automorphisms would be less that the degree of the extension and so the set-up is less conveniently described.

Thus the requirement in the Theorem, and later in the notes, that the roots be distinct is not essential to the theory *per se*, but allows the theorems to be stated in a nice way.

(12.14) To summarise where we are at, Theorem D gives a connection between the degrees of field extensions and minimum polynomials, while the Extension Theorem and Theorem connect minimum polynomials with the number of automorphisms of a field. Perhaps the following theorem is not then so surprising:

Corollary. *Let f be a polynomial over F and $E = F(\alpha_1, \ldots, \alpha_m)$ its splitting field over F with the roots α_i of f distinct. Then*

$$|Gal\,(E/F)| = [E : F].$$

The polynomial f is contained in the *ring* $F[x]$ over the *field* F, with E a *vector space* over F and Gal (E/F) the *group* of automorphisms. In this concise statement appears all the major algebraic objects of undergraduate mathematics.

Proof: The result follows immediately from Theorem by letting $F_1 = F_2 = F$, $E_1 = E_2 = E$ and τ be the identity automorphism. This gives that there are $[E : F]$ extensions of the identity automorphism $F \to F$ to automorphisms of E. On the otherhand, any automorphism of E fixing F pointwise is an extension of the identity automorphism on F, and so we obtain all the elements of the Galois group this way. □

(12.15) The criterion that E be a splitting field is important in using Theorem and it's Corollary properly. If you consider the extension $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2})$, then σ is an element of the Galois group Gal $(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ precisely when it sends $\sqrt[3]{2}$ to a root contained in $\mathbb{Q}(\sqrt[3]{2})$ of its minimum polynomial over $\mathbb{Q}$. As these roots are $\sqrt[3]{2}$ itself and the other two are complex and the field $\mathbb{Q}(\sqrt[3]{2})$ is completely contained in $\mathbb{R}$, the only automorphism we can have is the one that sends $\sqrt[3]{2}$ to itself, ie: the identity automorphism.

Thus the Galois group has order 1, whereas the degree of the extension is 3.

(12.16)

Theorem 8 *Let f be a polynomial over F and $E = F(\alpha_1, \ldots, \alpha_m)$ its splitting field over F with the roots α_i of f distinct. Moreover, suppose that*

$$[E : F] = \prod_i [F(\alpha_i) : F].$$

Then there is a $\sigma \in$ Gal (E/F) with $\sigma(\alpha_i) = \beta_i$ if and only if β_i is a root of the minimum polynomial over F of α_i.

Proof: That is is necessary for β_i to be a root of f_i has already been established. On the other hand, the condition on the degree of the extension means that the order of the Galois group Gal (E/F) is equal to the product of the degrees of the f_i, and so for the correct number of automorphisms to be realised, it must be possible to send α_i to any root of its minimum polynomial f_i. □

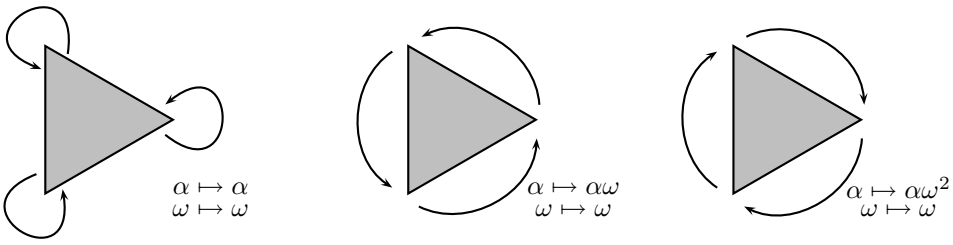
(12.17) In the first lecture we looked at the automorphisms of $\mathbb{Q}(\alpha, \omega)$ for

$$\alpha = \sqrt[3]{2} \in \mathbb{R} \text{ and } \omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i,$$

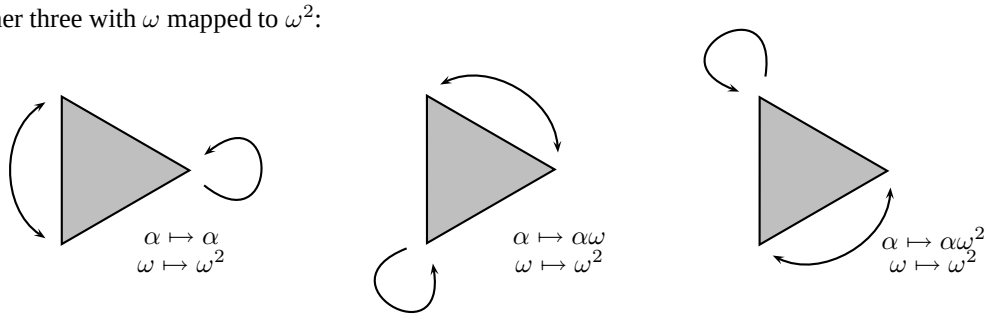
which in our new language translates as finding the elements of the Galois group Gal $(\mathbb{Q}(\alpha, \omega)/\mathbb{Q})$. The minimum polynomial of α over $\mathbb{Q}$ is $x^3 - 2$ with roots $\alpha, \alpha\omega, \alpha\omega^2$ and the minimum polynomial of ω over $\mathbb{Q}$ is $1 + x + x^2$ with roots ω, ω^2. The Tower law then gives that

$$[\mathbb{Q}(\alpha, \omega) : \mathbb{Q}] = [\mathbb{Q}(\alpha, \omega) : \mathbb{Q}(\alpha)][\mathbb{Q}(\alpha) : \mathbb{Q}] = [\mathbb{Q}(\omega) : \mathbb{Q}][\mathbb{Q}(\alpha) : \mathbb{Q}].$$

Thus by the Theorem above, we may send α to anyone of $\alpha, \alpha\omega, \alpha\omega^2$ and ω to any one of ω, ω^2 and get an automorphism. This gives six possible automorphisms, agreeing with the six we found in Lecture 1, one for each symmetry of the equilateral triangle formed by the roots in $\mathbb{C}$. Following this through with the vertices of the triangle, we have three automorphisms with ω mapped to itself:



and another three with ω mapped to ω^2:

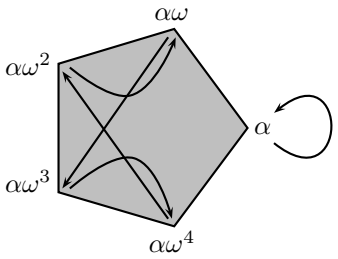


Exercise 118 Let $\alpha = \sqrt[5]{2}$ and $\omega = \cos(2\pi/5) + i\sin(2\pi/5)$ (so that $\alpha^5 = 2$ and $\omega^5 = 1$). Letting $\beta = \alpha + \omega$, eliminate radicals by considering the expression $(\beta - \omega)^5 = 2$ and find a polynomial of degree 20 having β as a root. Show that this polynomial is irreducible over $\mathbb{Q}$ and hence that

$$[\mathbb{Q}(\alpha + \omega) : \mathbb{Q}] = [\mathbb{Q}(\alpha) : \mathbb{Q}][\mathbb{Q}(\omega) : \mathbb{Q}].$$

Finally, show that $\mathbb{Q}(\alpha + \omega) = \mathbb{Q}(\alpha, \omega)$.

(12.18) Returning to some of the other examples from the first lecture, the extension $\mathbb{Q} \subset \mathbb{Q}(\alpha, \omega)$ satisfies the criterion of the Theorem above, where $\alpha = \sqrt[5]{2}$ and ω is a primitive 5-th root of 1. Thus an automorphism is free to send α to any root of the polynomial $x^5 - 2$ and ω to any root of the 5-th cyclotomic polynomial $1 + x + x^2 + x^3 + x^4$. Thus there are twenty elements of the Galois group in total.



$$\alpha = \sqrt[5]{2}$$

$$\omega = \frac{\sqrt{5}-1}{4} + \frac{\sqrt{2}\sqrt{5+\sqrt{5}}}{4}i$$

In particular we have the automorphism that sends α to itself and ω to ω^3 as depicted in the picture.

(12.19) So far we have only considered the Galois groups of fields, but if we are to get closer to the spirit of the first lecture, then we should be more interested in the Galois groups of *polynomials*. In the first lecture we achieved this using the smallest field containing the roots of the polynomial, and indeed we define: the *Galois group of the polynomial* $f \in F[x]$ is the Galois group $\mathrm{Gal}\,(E/F)$ of the splitting field E of f. Denote the group by $\mathrm{Gal}\,(f)$.

Proposition 4 *The Galois group of a polynomial of degree d is isomorphic to a subgroup of the symmetric group S_d.*

Proof: If $\{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ are the roots of f, then $\mathrm{Gal}\,(f) = \mathrm{Gal}\,(E/F)$ where the splitting field E is given by $E = F(\alpha_1, \alpha_2, \ldots, \alpha_d)$. Any element of $\mathrm{Gal}\,(f)$ is determined by where it sends each α_i, which must be to some root of its minimum polynomial over F. For any i, this minimum polynomial divides f (recall that the minimum polynomial of α divides any polynomial having α as a root) so its roots are contained amongst the roots of f, ie: amongst the $\{\alpha_1, \alpha_2, \ldots, \alpha_d\}$. Thus, any element of the Galois group can be identified with a permutation of these d roots. Different automorphisms correspond to different permutations, as the effect of the automorphism on the roots determines the whole automorphism. Thus $\mathrm{Gal}\,(f)$ may be identified with a subgroup of the group of permutations of the d roots, which is clearly isomorphic to S_d. □

Aside. There is a slick algebraic (some would say *proper*) way to put this, although it loses a little of the intuitive nature of what is going on. As any element of Galois group defines a permutation of the roots, define a map $\mathrm{Gal}\,(E/F) \to \mathrm{Sym}\{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ by sending a $\sigma \in \mathrm{Gal}\,(E/F)$ to this permutation. As the group operation is composition of maps in both these groups, we get that this is a homomorphism. If $\sigma \in \mathrm{Gal}\,(E/F)$ is sent to the identity permutation, then as an automorphism it fixes all the roots, so must be the identity automorphism, ie: the kernel of the homomorphism is trivial. The first isomorphism theorem for groups then gives that $\mathrm{Gal}\,(E/F)/\{\mathrm{id}\} \cong \mathrm{Gal}\,(E/F) \cong H$, a subgroup of $\mathrm{Sym}\{\alpha_1, \alpha_2, \ldots, \alpha_d\} \cong S_d$.

Further Exercises for §**12.**

Exercise 119 Show that the following Galois groups have the given orders:

1. $|\text{Gal}\,(\mathbb{Q}(\sqrt{2})/\mathbb{Q})| = 2$.
2. $|\text{Gal}\,(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})| = 1$.
3. $|\text{Gal}\,(\mathbb{Q}(-\frac{1}{2} + \frac{\sqrt{3}}{2}i)/\mathbb{Q})| = 2$.
4. $|\text{Gal}\,(\mathbb{Q}(\sqrt[3]{2}, -\frac{1}{2} + \frac{\sqrt{3}}{2}i)/\mathbb{Q})| = 6$.

Exercise 120 Find the orders of the following Galois groups:

1. $\text{Gal}\,(L/\mathbb{Q})$, where L is the splitting field of the polynomial $x - 2$.
2. $\text{Gal}\,(L/\mathbb{Q})$, where L is the splitting field of the polynomial $x^2 - 2$.
3. $\text{Gal}\,(L/\mathbb{Q})$, where L is the splitting field of the polynomial $x^5 - 2$.
4. $\text{Gal}\,(L/\mathbb{Q})$, where L is the splitting field of the polynomial $1 + x + x^2 + x^3 + x^4$.
5. $\text{Gal}\,(L/\mathbb{Q})$, where L is the splitting field of the polynomial $1 + x^2 + x^4$ (**hint**: $(x^2 - 1)(1 + x^2 + x^4) = x^6 - 1$).

Exercise 121 Let $p > 2$ be a prime number. Show that

1. $|\text{Gal}\,(\mathbb{Q}(\cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p})/\mathbb{Q})| = p - 1$.
2. $|\text{Gal}\,(L/\mathbb{Q})| = p(p-1)$, where L is the splitting field of the polynomial $x^p - 2$. Compare the answer when $p = 3$ and 5 to Lecture §**1.**.

Exercise 122 Let $p_1, \ldots, p_m$ be distinct primes. Show that,

$$\text{Gal}\,(\mathbb{Q}(\sqrt{p_1}, \ldots, \sqrt{p_m})/\mathbb{Q}) \cong \underbrace{\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2}_{m \text{ times}}$$

§13. Linear Algebra II: Solving equations

This section exists purely to provide some of the technical results we need for the big theorem of the next. It can be skipped over on a first reading.

(13.1) Let V be a n-dimensional vector space over the field F, with fixed basis $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. A *homogenous linear equation* over F is an equation of the form,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0,$$

with the a_i in F. The vector $v \in V$ is a solution if

$$v = \sum_{i=1}^{n} t_i\alpha_i \Rightarrow a_1t_1 + a_2t_2 + \cdots + a_nt_n = 0.$$

Call a system of linear equations,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0, \\ &\vdots \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n &= 0, \end{aligned}$$

independent over F iff the collection of vectors,

$$v_1 = \sum a_{1j}\alpha_j, v_2 = \sum a_{2j}\alpha_j, \ldots, v_k = \sum a_{kj}\alpha_j,$$

in V are independent.

Exercise 123

1. Let S be an independent system of equations in n unknowns. Show that S has the unique solution $v = 0$ in V.
2. Let S be a system of independent equations in V and let S' be a *proper* subset of the equations. Show that the set of solutions in V to S is a *proper* subspace of the set of solutions in V to S'.

Exercise 124 Let $F \subseteq E$ be an extension of fields and B a finite set. Let V_F be the F-vector space with basis B, ie: the elements of V_F are formal sums

$$\sum \lambda_i b_i,$$

with the $\lambda_i \in F$ and the $b_i \in B$. Formal sums are added together and multiplied by $\lambda \in F$ in the obvious way. Similarly let V_E be the E-vector space with basis B, and identify V_F with a subset (it is not a subspace) of V_E in the obvious way. Now let S, S' be systems of equations in V_E as in the previous exercise. Show that the conclusion reached there is still true when looking at the solution sets in V_F.

Exercise 125 Let F be a field and $\alpha_1, \ldots, \alpha_{n+1}$ distinct elements of it. Show that

$$\det \begin{pmatrix} \alpha_1^n & \cdots & \alpha_1 & 1 \\ \vdots & & \vdots & \vdots \\ \alpha_{n+1}^n & \cdots & \alpha_{n+1} & 1 \end{pmatrix} \neq 0.$$

(hint: suppose not, and find a polynomial of degree n having $n+1$ distinct roots in F, thus contradicting Theorem 2). This is called the *Vandermonde determinant*.

Lemma 5 *Let F be a field and $f, g \in F[x]$ two polynomials of degree n over F. Suppose that there exist $n+1$ distinct values $\alpha_i \in F$, such that $f(\alpha_i) = g(\alpha_i)$ for all i. Then $f = g$.*

Proof: Let

$$f(x) = \sum a_i x^i \text{ and } g(x) = \sum b_i x^i.$$

We get $n+1$ expressions of the form

$$\sum a_i \alpha_j^i = \sum b_i \alpha_j^i \Leftrightarrow \sum a_j^i y_i = 0,$$

where $y_i = a_i - b_i$. Think of these last as $n+1$ equations in the $n+1$ unknowns y_i. The matrix of coefficients is

$$\begin{pmatrix} \alpha_1^n & \cdots & \alpha_1 & 1 \\ \vdots & & \vdots & \vdots \\ \alpha_{n+1}^n & \cdots & \alpha_{n+1} & 1 \end{pmatrix}$$

which is the matrix given in Exercise 125. Its determinant is non-zero, and thus the system of equations has the unique solution $y_i = 0$ for all i, so that $a_i = b_i$ for all i and hence $f = g$. □

(13.2) Here is the result we will require in the next section.

Theorem 9 *Let $F \subseteq E = F(\alpha)$ be a simple extension of fields with the minimum polynomial of α over F having distinct roots. Let $\{\sigma_1, \sigma_2 \ldots, \sigma_k\}$ be distinct non-identity elements of the Galois group Gal (E/F). Then*

$$\sigma_1(x) = \sigma_2(x) = \cdots = \sigma_k(x) = x,$$

is a system of independent linear equations over E.

Proof: By Theorem D we have a basis $\{1, \alpha, \alpha^2, \ldots, \alpha^d\}$ for E over F where the minimum polynomial f of α has degree $d+1$. Thus any $x \in E$ has the form

$$x = x_0 + x_1\alpha + x_2\alpha^2 + \cdots + x_d\alpha^d,$$

for some $x_i \in F$. By the Extension Theorem, the elements of the Galois group send α to roots of f. Suppose these roots are $\{\alpha = \alpha_0, \alpha_1, \ldots, \alpha_d\}$ where $\sigma_i(\alpha) = \alpha_i$ (as none of the σ_i are the identity, we have that no σ_i sends α to itself). Then x satisfies $\sigma_i(x) = x$ if and only if,

$$(\alpha_0 - \alpha_i)x_1 + (\alpha_0^2 - \alpha_i^2)x_2 + \cdots + (\alpha_0^d - \alpha_i^d)x_d = 0.$$

Thus we have a system of equations $Ax = 0$ where the matrix of coefficients A is made up of rows from the larger $d \times d$ matrix $\widehat{A}$ given by,

$$\widehat{A} = \begin{pmatrix} \alpha_0 - \alpha_1 & \alpha_0^2 - \alpha_1^2 & \cdots & \alpha_0^d - \alpha_1^d \\ \alpha_0 - \alpha_2 & \alpha_0^2 - \alpha_2^2 & \cdots & \alpha_0^d - \alpha_2^d \\ \vdots & \vdots & & \vdots \\ \alpha_0 - \alpha_d & \alpha_0^2 - \alpha_d^2 & \cdots & \alpha_0^d - \alpha_d^d \end{pmatrix}.$$

Suppose we have $\widehat{A}b = 0$ for some vector $b \in E^n$, so that

$$b_0\alpha_0 + b_1\alpha_0^2 + \cdots + b_d\alpha_0^d = b_0\alpha_i + b_1\alpha_i^2 + \cdots + b_d\alpha_i^d,$$

for each $1 \leq i \leq d$. Thus if $f = b_0x + b_1x^2 + \cdots + b_dx^d$, then we have $f(\alpha_0) = f(\alpha_1) = f(\alpha_2) = \cdots = f(\alpha_d) = a$, say. Thus the degree d polynomial $g = f - a$ agrees with the zero polynomial at $d+1$ distinct values, hence by the lemma, must be the zero polynomial, and so all the a_i are zero. Thus, the system of equations $\widehat{A}x = 0$, hence the system $Ax = 0$, is independent. □

§14. The Fundamental Theorem of Galois Theory

(14.1) In §**10.** we saw that a complex number ζ was constructible precisely when there was a tower of fields,

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n,$$

with each K_i a degree two extension of K_{i-1} and $\mathbb{Q}(\zeta)$ contained in the top field. All the examples we have given so far that use this result have showed that no such tower exists. In otherwords, they have been negative results, and for good reason: to use the theorem positively, to show that a number can actually *be* constructed, requires a knowledge of the fields sandwiched in between $\mathbb{Q}$ and $\mathbb{Q}(\zeta)$. In this section we prove the theorem that gives us that knowledge.

(14.2) First we need a picture of all the fields we are interested in, analogous to the picture of all the subgroups of a group that we drew in §**11.**.

Let $F \subseteq E$ be an extension of fields. Call K an *intermediate field* precisely when K is an extension of F and E is an extension of K: ie: $F \subseteq K \subseteq E$. The *lattice of intermediate fields* is a diagram such that if K_1 and K_2 are two such fields and $K_1 \subseteq K_2$, then K_2 is placed higher in the diagram than K_1, with a line connecting them as shown at left. Denote this lattice by $\mathscr{L}(E/F)$.

(14.3) From now on we will work in the following situation: we have an extension $F \subseteq E$ such that *every irreducible polynomial over F has distinct roots in E*. For example we saw in Exercise 36 that this is the case if F has characteristic 0. It is also true if F is a finite field, although we omit the proof here.

Thus, the following theorem includes in its remit the fields we have spent most of the time considering: subfields of $\mathbb{C}$ and finite fields. It is only examples like $\mathbb{F}_p(t)$, the rational function field over $\mathbb{F}_p$ (being infinite of characteristic $p > 0$, see Exercise 38) that are left out in the cold.

The Galois Correspondence (part 1). *Let $F \subseteq E$ be a finite extension as above with E the splitting field over F of some polynomial $f \in F[x]$, and $G = Gal\,(E/F)$ its Galois group. Let $\mathscr{L}(G)$ and $\mathscr{L}(E/F)$ be the subgroup and intermediate field lattices.*

1. *For any subgroup H of G, let*

$$E^H = \{\lambda \in E \,|\, \sigma(\lambda) = \lambda \text{ for all } \sigma \in H\}.$$

   Then E^H is an intermediate field, called the fixed field of H.

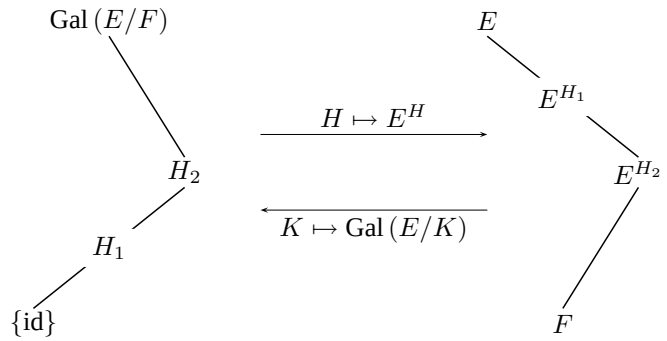
2. *For any intermediate field, $Gal\,(E/K)$ is a subgroup of G.*

3. *The maps $H \mapsto E^H$ and $K \mapsto Gal\,(E/K)$ are mutual inverses, hence bijections $\mathscr{L}(G) \leftrightarrow \mathscr{L}(E/F)$.*

4. *Both maps reverse order: if $H_1 \subseteq H_2 \subseteq G$ then $F \subseteq E^{H_2} \subseteq E^{H_1} \subseteq E$, and if $F \subseteq K_1 \subseteq K_2 \subseteq E$ then $Gal\,(E/K_2) \subseteq Gal\,(E/K_1) \subseteq G$.*

5. *The degree of the extension $E^H \subseteq E$ is equal to the order $|H|$ of the subgroup H.*

In otherwords, once you know the lattice of subgroups of the Galois group, you can find the lattice of intermediate fields just by turning it upside down (and vice-versa)! Schematically,



There are a few other things worth noticing. The whole Galois group Gal (E/F) fixes F pointwise, so its fixed field is F, while the trivial subgroup, consisting of just the identity automorphism, fixes everything, hence its fixed field is all of E. Thus, the largest subgroup corresponds to the smallest intermediate field and the smallest subgroup to the largest intermediate field.

The Theorem also says that the two maps $X \mapsto E^X$ and $Y \mapsto \text{Gal}\,(E/Y)$ are bijections, hence in particular are 1-1: if $E^{H_1} = E^{H_2}$ then $H_1 = H_2$ and if Gal $(E/K_1) =$ Gal (E/K_2) then $K_1 = K_2$.

If the upside down nature of the correspondence seems puzzling, it is simply linear algebra (and indeed this is how we will prove it). If H is a subgroup, think of the fixed field E^H as the set of solutions in E to the equations,

$$\sigma(x) = x, \sigma \in H.$$

The more equations you have, the greater the number of conditions being imposed on x, hence the smaller the number of solutions. So larger subgroups should correspond to smaller intermediate fields. That the correspondence is exact, so that as soon as we add *just one more* equation the number of solutions strictly decreases, will follow from §**13.** as these equations are linear and *independent*.

Proof: In the situation described, where E is a finite extension of F, the extension must be simple by Theorem 7, ie: of the form $F \subseteq F(\alpha)$ for some α algebraic over F.

For the first part, we have $E^H \subseteq E$ by definition, and $F \subseteq E^H$, as every element of G, so in particular, every element of H fixes F. If $\lambda, \mu \in E^H$ then $\sigma(\lambda + \mu) = \sigma(\lambda) + \sigma(\mu) = \lambda + \mu$ and similarly for $\sigma(\lambda\mu)$ and $\sigma(1/\lambda)$. Thus E^H is an intermediate field.

If an automorphism of E fixes the intermediate field K pointwise, then it certainly fixes the field F pointwise. Thus Gal $(E/K) \subseteq$ Gal (E/F) and we indeed have a map $\mathscr{L}(E/F) \to \mathscr{L}(G)$ given by $K \mapsto$ Gal (E/K). If λ is fixed by every automorphism in H_2, then it is fixed by every automorphism in H_1 and so $E^{H_2} \subseteq E^{H_1}$. If σ fixes every element of K_2 pointwise then it fixes every element of K_1 pointwise too, so that Gal $(E/K_2) \subseteq$ Gal (E/K_1).

To show that the two maps are inverses of each other, we take a subgroup H and show that their composition,

$$H \to E^H \to \text{Gal}\,(E/E^H),$$

gets us back to where we started, ie: that Gal $(E/E^H) = H$. This will then give the desired bijection.

By definition, every element of H fixes E^H pointwise, and since Gal (E/E^H) consists of *all* the automorphisms of E that fix E^H pointwise, we have that $H \subseteq$ Gal (E/E^H). In fact, both of the subgroups H and Gal (E/E^H) have the same fixed field, ie: $E^{\text{Gal}\,(E/E^H)} = E^H$. To see this, certainly any $\sigma \in$ Gal (E/E^H) fixes E^H pointwise by definition, so $E^H \subseteq E^{\text{Gal}\,(E/E^H)}$. On the otherhand, as $H \subseteq$ Gal (E/E^H) and the maps reverse order, we have, $E^{\text{Gal}\,(E/E^H)} \subseteq E^H$.

By the results of §**13.**, the elements of the fixed field $E^{\text{Gal}\,(E/E^H)}$ are obtained by solving the system of linear equations $\sigma(x) = x$ for all $\sigma \in$ Gal (E/E^H), and these equations are independent. In particular, a *proper* subset of these equations has a strictly larger solution set. We already have that $H \subseteq$ Gal (E/E^H), so suppose that H is a proper subgroup of Gal (E/E^H). The fixed field E^H would then properly contain the fixed field $E^{\text{Gal}\,(E/E^H)}$. As this contradicts $E^H = E^{\text{Gal}\,(E/E^H)}$, we must have that $H =$ Gal (E/E^H). Thus the map $H \mapsto E^H$ is a bijection as desired.

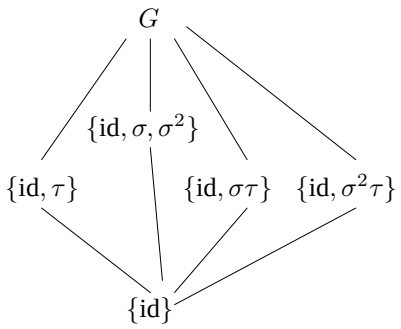
As E is a splitting field we can apply Theorem G to get $|\mathrm{Gal}\,(E/E^H)| = [E : E^H]$, where we now have that $\mathrm{Gal}\,(E/E^H) = H$, so that $|H| = [E : E^H]$. □

(14.4) We are certainly long overdue an example. In §**12.** we reverified the example of the first lecture to show that the splitting field $\mathbb{Q}(\alpha, \omega)$ of the polynomial $x^3 - 2$ had Galois group,

$$G = \mathrm{Gal}\,(\mathbb{Q}(\alpha, \omega)/\mathbb{Q}) = \{\mathrm{id}, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\},$$

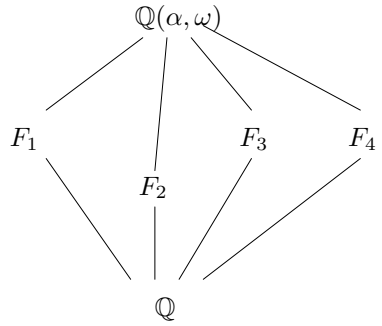
where $\sigma(\alpha) = \alpha\omega, \sigma(\omega) = \omega$ and $\tau(\alpha) = \alpha, \tau(\omega) = \omega^2$.

We claim that the subgroup lattice $\mathscr{L}(G)$ is,



Firstly, the subsets given are easily seen to be subgroups, so we just need to check that the picture is complete. Let H be an arbitrary subgroup of G and suppose that H contains the element σ. Then it must contain all the powers of σ, hence must contain the subgroup $\{\mathrm{id}, \sigma, \sigma^2\}$. Thus the order of H is constrained by $3 \leq |H| \leq 6$, and by Lagrange's Theorem $|H|$ divides 6, so we must have $|H| = 3$ or 6. Thus H must equal $\{\mathrm{id}, \sigma, \sigma^2\}$ or be all of G. This completely describes all the subgroups that contain the element σ. The same argument (and conclusion) applies to the subgroups containing σ^2. Thus we are left to describe the subgroups containing any one of the three "reflections" $\tau, \sigma\tau, \sigma^2\tau$ but not σ or σ^2. Let H be a subgroup containing τ. As H contains $\{\mathrm{id}, \tau\}$, and by Lagrange, it has order $2, 3$ or 6. The only one of these three possibilities not already in the lattice is the order 3 case, so we show that this is not possible. To have order 3, H must also contain one of $\sigma\tau$ or $\sigma^2\tau$. If the former, then it also contains $\sigma\tau\tau = \sigma$, a contradiction, and similarly for the other case. Thus the lattice $\mathscr{L}(G)$ is as depicted[13].

The Galois Correspondence now gives the lattice $\mathscr{L}(E/F)$ of intermediate fields to be,



with F_2 the fixed field of the subgroup $\{\mathrm{id}, \sigma, \sigma^2\}$ and the others the fixed fields (in no particular order) of the three order two subgroups. By the fourth part of the Galois correspondence, each of the extensions $F_i \subseteq \mathbb{Q}(\alpha, \omega)$ has degree the order of the appropriate subgroup, so that $\mathbb{Q}(\alpha, \omega)$ is a degree three extension of F_2, and a degree two extension of the other intermediate fields.

Suppose that F_1 is the fixed field of the subgroup $\{\mathrm{id}, \tau\}$. We find an explicit description of its elements. From the Tower law, a basis for $\mathbb{Q}(\alpha, \omega)$ over $\mathbb{Q}$ is given by

$$\{1, \alpha, \alpha^2, \omega, \alpha\omega, \alpha^2\omega\},$$

[13] In general such arguments become more complicated as the order of the Galois group increases.

so that an arbitrary element x of $\mathbb{Q}(\alpha,\omega)$ has the form,

$$x = a_0 + a_1\alpha + a_2\alpha^2 + a_3\omega + a_4\alpha\omega + a_5\alpha^2\omega,$$

with the $a_i \in \mathbb{Q}$. The element x is in F_1 if and only if $\tau(x) = x$ where,

$$\begin{aligned}\tau(x) &= a_0 + a_1\alpha + a_2\alpha^2 + a_3\omega^2 + a_4\alpha\omega^2 + a_5\alpha^2\omega^2\\ &= a_0 + a_1\alpha + a_2\alpha^2 + a_3(-1-\omega) + a_4\alpha(-1-\omega) + a_5\alpha^2(-1-\omega)\\ &= (a_0 - a_3) + (a_1 - a_4)\alpha + (a_2 - a_5)\alpha^2 - a_3\omega - a_4\alpha\omega^2 - a_5\alpha^2\omega.\end{aligned}$$

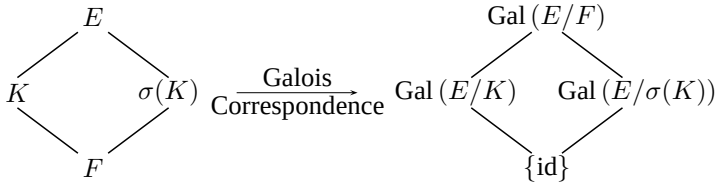
Because we are using a basis, we can equate coefficients to get,

$$a_0 - a_3 = a_0, a_1 - a_4 = a_1, a_2 - a_5 = a_2, -a_3 = a_3, -a_4 = a_4 \text{ and } -a_5 = a_5.$$

Thus, $a_3 = a_4 = a_5 = 0$ and a_0, a_1, a_2 are completely arbitrary. Hence x has the form $x = a_0 + a_1\alpha + a_2\alpha^2$ so is an element of $\mathbb{Q}(\alpha)$. This gives that $F_1 \subseteq \mathbb{Q}(\alpha)$. On the otherhand, τ fixes $\mathbb{Q}$ pointwise and fixes α by definition, hence fixes every element of $\mathbb{Q}(\alpha)$. This gives that $\mathbb{Q}(\alpha) \subseteq F_1$ and so $F_1 = \mathbb{Q}(\alpha)$.

(14.5) The Galois correspondence allows us to "model" the subgroups of the Galois group by the intermediate fields (and vice-versa). But there are subgroups and there are subgroups: what about the normal subgroups? As they are slightly special, they should correspond to slightly special intermediate fields. Is the Galois correspondence sensitive enough to spot the difference?

Let $F \subseteq E$ be an extension of fields with Galois group Gal (E/F), and let K be an intermediate field and $\sigma \in$ Gal (E/F). The image of K by the automorphism σ is another intermediate field, and so we get the picture below left. By the Galois correspondence, there are subgroups Gal (E/K) and Gal $(E/\sigma(K))$ corresponding to the two intermediate fields as shown below right:



The two intermediate fields are then related by,

Proposition 5 *The subgroups Gal (E/K) and Gal $(E/\sigma(K))$ are conjugate, indeed,*

$$\textit{Gal}\,(E/\sigma(K)) = \sigma^{-1}\textit{Gal}\,(E/K)\sigma.$$

(We are reading expressions in a group from left to right).

Proof: If $x \in \sigma(K)$, then $x = \sigma(y)$ for some $y \in K$. Thus if $\overline{\sigma} \in$ Gal (E/K), then $\sigma^{-1}\overline{\sigma}\sigma$ (read from left to right) fixes x, and so is contained in Gal $(E/\sigma(K))$. Thus σ^{-1}Gal $(E/K)\sigma \subseteq$ Gal $(E/\sigma(K))$. The proof of the opposite inclusion is the same. □

(14.6) Remembering that a subgroup N of G is normal when $g^{-1}Ng = N$ for all $g \in G$ (see §**11.**). We have σ^{-1}Gal $(E/K)\sigma =$ Gal $(E/\sigma(K))$, and this in turn will clearly equal Gal (E/K) when $\sigma(K) = K$ for all σ. So this is the kind of intermediate field that picks out normal subgroups: one that is sent to itself by any automorphism[14].

If every automorphism sends K to itself then any automorphism of E restricts to an automorphism of K as well. This is all summarised in the second part of the Galois correspondence:

The Galois Correspondence (part 2). *Under the assumptions of the first part of the Galois correspondence, let K be an intermediate field. Then, $\sigma(K) = K$ for all $\sigma \in$ Gal (E/F) if and only if Gal (E/K) is a normal subgroup of Gal (E/F), and in this case,*

$$\textit{Gal}\,(E/F)/\textit{Gal}\,(E/K) \cong \textit{Gal}\,(K/F).$$

[14]Note that this is different from saying that the field is fixed pointwise, which is a far stronger property

Proof: If $\sigma(K) = K$ for all σ then by Proposition 5, $\sigma^{-1}\text{Gal}\,(E/K)\sigma = \text{Gal}\,(E/\sigma(K)) = \text{Gal}\,(E/K)$ for all σ and so Gal (E/K) is normal. On the otherhand, if Gal (E/K) is normal then Proposition 5 gives that Gal $(E/\sigma(K)) = \text{Gal}\,(E/K)$ for all σ, where $X \mapsto \text{Gal}\,(E/X)$ is a 1-1 map by the first part of the Galois correspondence, hence we have $\sigma(K) = K$ for all σ.

Now define a map Gal $(E/F) \to$ Gal (K/F) by taking an automorphism σ of E fixing F pointwise and restricting it to $K \subseteq E$. We get an automorphism of K as $\sigma(K) = K$ for any σ. The map is a homomorphism between these two groups rather trivially, as the same operation, namely composition of automorphisms, is being used in both. An element σ is in the kernel of the homomorphism if and only if it restricts to the identity map on K (ie: fixes K pointwise when restricted) which happens if and only if σ is in Gal (E/K). If σ is an automorphism of K fixing F pointwise then by Theorem **??**, it can be extended to an automorphism of E fixing F pointwise, ie: any element of the Galois group Gal (K/F) can be obtained by restricting an element of Gal (E/F). Thus the map is onto and the isomorphism follows by the first isomorphism theorem. □

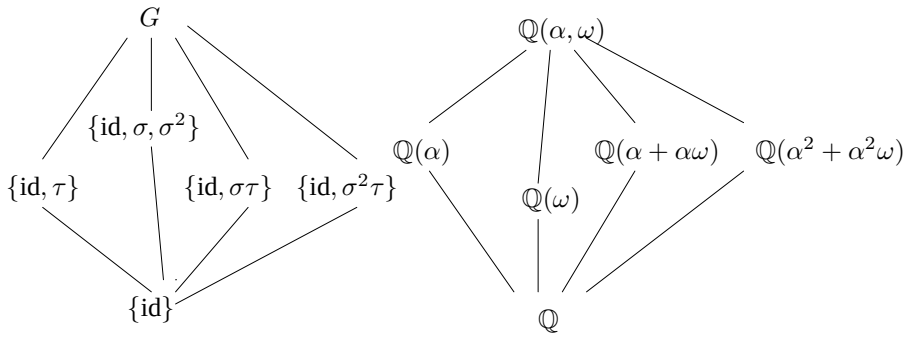
We used Theorem **??** in the proof to show that any element of Gal (K/F) was the restriction of an element of Gal (E/F). Moreover, Theorem **??** says that an element of Gal (K/F) will be the restriction of $[E : K]$ elements of Gal (E/F) and this gels perfectly with the isomorphism given above: the identity of Gal (K/F) will be the restriction of $[E : K] = |\text{Gal}\,(E/K)|$ elements of Gal (E/F), in otherwords, the kernel of the mapping given in the proof will have $|\text{Gal}\,(E/K)|$ elements.

Exercise 126 A subgroup H of a group G is said to be *malnormal* when $g \in G \setminus H$ gives that $g^{-1}Hg \cap H = \{\text{id}\}$. Thus, the malnormal subgroups are in some sense the antithesis of the normal ones. Show that the malnormal subgroups can be spotted by the Galois correspondence by describing the intermediate fields they correspond to.

(14.7) Here is a simple application. According to Exercise 110, any subgroup of index two in a group G is a normal subgroup. By the first part of the Galois correspondence, subgroups of index two correspond to intermediate fields $F \subseteq K \subseteq E$ with the degree of the extension $F \subseteq F$ equal to two. By the second part of the Galois correspondence, any automorphism of E fixing F pointwise must send such a K to itself.

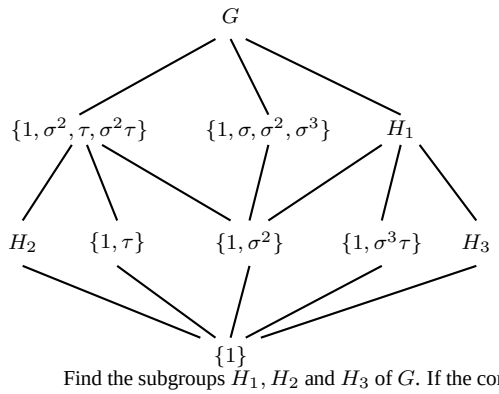
Further Exercises for §**14.**

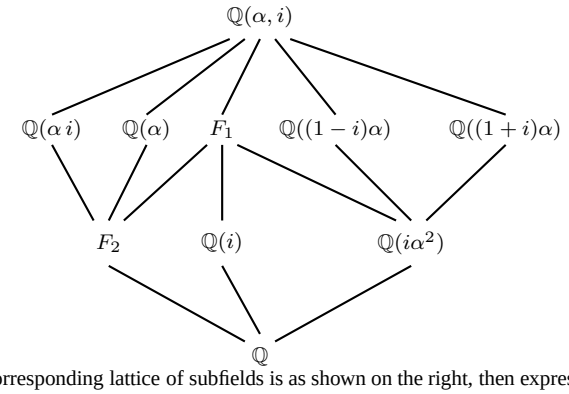
Exercise 127 Complete the example above:



Exercise 128

1. Let $\alpha = \sqrt[4]{2} \in \mathbb{R}$ and $i \in \mathbb{C}$, and consider the field $\mathbb{Q}(\alpha, i) \subset \mathbb{C}$. Suppose that σ, τ are automorphisms of $\mathbb{Q}(\alpha, i)$ such that
$$\sigma(i) = i, \sigma(\alpha) = \alpha\, i, \tau(i) = -i, \text{ and } \tau(\alpha) = \alpha.$$
Show that
$$G = \{1, \sigma, \sigma^2, \sigma^3, \tau, \sigma\tau, \sigma^2\tau, \sigma^3\tau\},$$
are then *distinct* automorphisms of $\mathbb{Q}(\alpha, i)$, and that $\tau\sigma = \sigma^3\tau$.
2. Suppose now that the G above is the Galois group of $\mathbb{Q}(\alpha, i)$ over $\mathbb{Q}$, and that G has the lattice of subgroups as shown on the left:

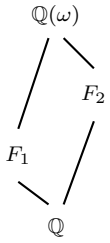




Find the subgroups H_1, H_2 and H_3 of G. If the corresponding lattice of subfields is as shown on the right, then express the fields F_1 and F_2 in the form $\mathbb{Q}(\beta_1, \ldots, \beta_n)$ for some $\beta_1, \ldots, \beta_n \in \mathbb{C}$.

Exercise 129

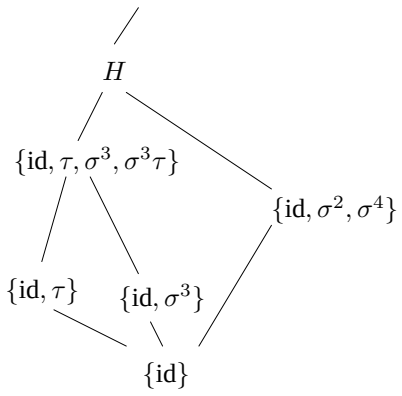
1. Let $\omega = \cos\dfrac{2\pi}{7} + i\sin\dfrac{2\pi}{7}$. Show that $\mathbb{Q}(\omega)$ is the splitting field of the polynomial
$$1 + x + x^2 + x^3 + x^4 + x^5 + x^6.$$
Deduce that $|\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})| = 6$.
2. Suppose $\sigma \in \text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})$ is such that $\sigma(\omega) = \omega^3$. Show that,
$$\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q}) = \{1, \sigma, \sigma^2, \sigma^3, \sigma^4, \sigma^5\}.$$
3. Using the Galois correspondence, show that the lattice of intermediate fields is:



where F_1 is a degree 2 extension of $\mathbb{Q}$ and F_2 a degree 3 extension. Find complex numbers $\beta_1, \ldots, \beta_n$ such that $F_2 = \mathbb{Q}(\beta_1, \ldots, \beta_n)$.

Exercise 130 Let $\alpha = \sqrt[6]{2}$ and $\omega = \dfrac{1}{2} + \dfrac{\sqrt{3}}{2}i$ and consider the field extension $\mathbb{Q} \subset \mathbb{Q}(\alpha, \omega)$.

1. Show that $|\text{Gal}\,(\mathbb{Q}(\alpha, \omega)/\mathbb{Q})| = 24$.
2. Find a basis for $\mathbb{Q}(\alpha, \omega)$ over $\mathbb{Q}$.
3. Suppose that $\sigma, \tau \in \text{Gal}\,(\mathbb{Q}(\alpha, \omega)/\mathbb{Q})$ are such that $\tau : \alpha \mapsto \alpha, \omega \mapsto \omega^5$ and $\sigma : \alpha \mapsto \alpha\omega, \omega \mapsto \omega$. Show that
$$H = \{1, \sigma, \sigma^2, \sigma^3, \sigma^4, \sigma^5, \tau, \tau\sigma, \tau\sigma^2, \tau\sigma^3, \tau\sigma^4, \tau\sigma^5\},$$
are then distinct elements in Gal $(\mathbb{Q}(\alpha, \omega)/\mathbb{Q})$ too (do this by observing their effect on the basis).
4. *Part* of the subgroup lattice $\mathscr{L}_G$ is shown below. Find the corresponding part of the lattice of intermediate fields.



Exercise 131 Let $\omega = \cos\dfrac{2\pi}{5} + i\sin\dfrac{2\pi}{5}$ and consider $\mathbb{Q}(\omega)$.

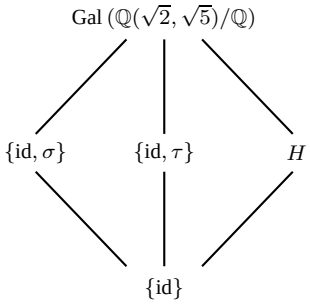
1. Show that $\mathbb{Q}(\omega)$ is the splitting field of the polynomial $1 + x + x^2 + x^3 + x^4$.
2. Deduce that $|\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})| = 4$.
3. Suppose $\sigma \in \text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})$ is such that $\sigma(\omega) = \omega^2$. by combining parts (a) and (b), show that,
$$\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q}) = \{1, \sigma, \sigma^2, \sigma^3\}.$$
(hence the Galois group is cyclic).
4. Find the subgroup lattice $\mathscr{L}_G$ for $G = \text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})$.
5. Using the Galois correspondence, deduce that the lattice of intermediate fields is
$$\begin{array}{c} \mathbb{Q}(\omega) \\ | \\ F \\ | \\ \mathbb{Q} \end{array}$$
Find a complex number β such that $F = \mathbb{Q}(\beta)$.

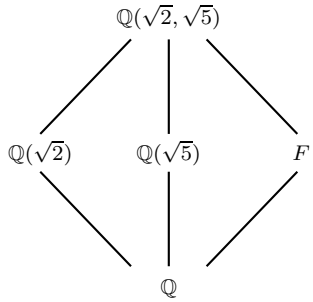
Exercise 132 Consider the polynomial $f(x) = (x^2 - 2)(x^2 - 5) \in \mathbb{Q}[x]$.

1. Show that $\mathbb{Q}(\sqrt{2}, \sqrt{5})$ is the splitting field of f over $\mathbb{Q}$.
2. Show that the Galois group $\text{Gal}\,(\mathbb{Q}(\sqrt{2}, \sqrt{5})/\mathbb{Q})$ has order four.

You may assume that if $a, b, c \in \mathbb{Q}$ satisfy $a\sqrt{2} + b\sqrt{5} + c = 0$ then $a = b = c = 0$.

3. Assume that σ and τ are automorphisms of $\mathbb{Q}(\sqrt{2}, \sqrt{5})$ defined by,
$$\begin{array}{ccccc} \sqrt{2} & \mapsto & -\sqrt{2} & \qquad & \sqrt{2} & \mapsto & \sqrt{2} \\ & \sigma & & & & \tau & \\ \sqrt{5} & \mapsto & \sqrt{5} & & \sqrt{5} & \mapsto & -\sqrt{5} \end{array}$$
List the elements of the Galois group $\text{Gal}\,(\mathbb{Q}(\sqrt{2}, \sqrt{5})/\mathbb{Q})$, justifying your answer.
4. Complete the subgroup lattice on the left by listing the elements of H,





and use your answer to write the field F in the form $\mathbb{Q}(\theta)$ for some $\theta \in \mathbb{C}$.

§15. Applications of the Galois Correspondence

Constructing polygons

If p is a prime number, then a regular p-gon can be constructed *only if* p is a Fermat prime of the form

$$2^{2^t}+1.$$

We proved this back in §**10.**, and all it required was the idea of the degree of an extension. In otherwords, we really didn't require any Galois Theory in the proof, if you take Galois Theory to mean the interplay between fields and their Galois groups.

What about results in the positive direction? Can p-gons with a Fermat prime number of sides be constructed? The first few such primes are $3, 5$ and 17, and we saw in §**7.** that these three were constructible, albeit if we believe Gauss's identity for $\cos(\pi/17)$. Thus, *explicit* constructions of these polygons is a complicated business. Nevertheless, the full power of Galois Theory proper gives,

Theorem 10 *If p is a Fermat prime then a regular p-gon can be constructed.*

Proof: By Theorem E we are done if we can find a tower of fields,

$$\mathbb{Q}\subseteq K_1\subseteq\cdots\subseteq K_n=\mathbb{Q}(\zeta),$$

for $\zeta=\cos(2\pi/p)+i\sin(2\pi/p)$ and with $[K_i:K_{i-1}]=2$. As $\mathbb{Q}(\zeta)$ is the splitting field of the p-th cyclotomic polynomial

$$\Phi_p(x)=x^{p-1}+x^{p-2}+\cdots+x+1,$$

we have by Theorem G that the Galois group has order,

$$|\text{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q})|=[\mathbb{Q}(\zeta):\mathbb{Q}]=\deg\Phi_p=p-1=2^n,$$

(as p, being a Fermat prime is of the form $p=2^n+1$). In §**12.** we showed that Gal $(\mathbb{Q}(\zeta)/\mathbb{Q})$ was a cyclic group, and so by Exercise 100, we can find a chain of subgroups[15]

$$\{\text{id}\}=H_0\subseteq H_1\subseteq\cdots\subseteq H_n=\text{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q}),$$

where H_i has order 2^i. Making it explicit, if Gal $(\mathbb{Q}(\zeta)/\mathbb{Q})=\{\text{id},\sigma,\sigma^2,\ldots,\sigma^{2^n-1}\}$ then the subgroups are,

$$\{\text{id}\}\subseteq\{\text{id},\sigma^{2^{n-1}}\}\subseteq\{\text{id},\sigma^{2^{n-2}},\sigma^{2.2^{n-2}},\sigma^{3.2^{n-2}}\}\subseteq\cdots$$
$$\cdots\subseteq\{\text{id},\sigma^{2^{n-i}},\sigma^{2.2^{n-i}},\sigma^{3.2^{n-i}},\ldots\}\subseteq\cdots\subseteq\{\text{id},\sigma^2,\sigma^4,\ldots\}.$$

The Galois correspondence thus gives a chain of fields,

$$\mathbb{Q}=K_0\subseteq K_1\subseteq\cdots\subseteq K_n=\mathbb{Q}(\zeta),$$

where K_{n-i} is the fixed field E^{H_i} of the subgroup H_i. Letting $j=n-i$, we have the extension $K_j\subseteq\mathbb{Q}(\zeta)$ of degree the order 2^i of H_i. In particular, by the tower law,

$$[\mathbb{Q}(\zeta):K_{j-1}]=[\mathbb{Q}(\zeta):K_j][K_j:K_{j-1}],$$

where $j-1=n-(i+1)$, so that $[\mathbb{Q}(\zeta):K_{j-1}]=2^{i+1}$. Thus $2^{i+1}=2^i[K_j:K_{j-1}]$, so that $[K_j:K_{j-1}]=2$ as required. □

Corollary. *If $n=2^kp_1p_2\ldots p_m$ with the p_i Fermat primes, then a regular n-gon can be constructed.*

Proof: Certainly 2^k-gons can be constructed just by repeatedly bisecting angles. Thus, an n-gon can be constructed, where n has the form given, by Exercise **??**. □

Remarkably, with a little more Galois Theory, the converse to this statement can also be proved, thus *completeley determining* those n-gons that can be constructed.

[15] Alternatively, these subgroups can be found using Sylow's Theorem from §**11.**.

(15.1) The angle π/n can be constructed precisely when the angle $2\pi/n$ can be constructed which in turns happens precisely when the regular n-gon can be constructed. Thus, the list of submultiples of π that are constructable runs as,

$$\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{5}, \frac{\pi}{6}, \frac{\pi}{8}, \frac{\pi}{10}, \frac{\pi}{12}, \frac{\pi}{15}, \dots$$

Exercise 133 Give direct proofs of the non-constructability of the angles,

$$\frac{\pi}{7}, \frac{\pi}{9}, \frac{\pi}{11} \text{ and } \frac{\pi}{13}.$$

The Fundamental Theorem of Algebra

The so-called fundamental theorem of algebra can be proved from the Galois correspondence, and we have already observed how curious it is that a theorem fundamental to all of algebra can be deduced from a theorem fundamental to just part of it.

The proof requires two straight-forward observations. First, there are no extensions of the reals of odd degree > 1. This is because any polynomial in $\mathbb{R}[x]$ has roots that are either real or occur in complex conjugate pairs, hence in particular, a real polynomial with odd degree > 1 has a real root and so is reducible over $\mathbb{R}$. Thus, the minimum polynomial over $\mathbb{R}$ of any $\alpha \notin \mathbb{R}$ must have even degree. If $\mathbb{R} \subseteq L$ is an extension, then choosing $\alpha \in L \setminus \mathbb{R}$, we have

$$[L : \mathbb{R}] = [L : \mathbb{R}(\alpha)][\mathbb{R}(\alpha) : \mathbb{R}],$$

with the last term even by the comments above, hence $[L : \mathbb{R}]$ even.

The other observation is that the complexes have no extensions of degree two. If $\mathbb{C} \subseteq L$ with $[L : \mathbb{C}] = 2$ then choose $\alpha \in L \setminus \mathbb{C}$ so that we have the intermediate $\mathbb{C} \subseteq \mathbb{C}(\alpha) \subseteq L$. We must certainly have $[\mathbb{C}(\alpha) : \mathbb{C}] = 1$ or 2, and if the degree was 1 then we would have $\alpha \in \mathbb{C}$, so $[\mathbb{C}(\alpha) : \mathbb{C}] = 2$, and thus $L = \mathbb{C}(\alpha)$. If f is the minimum polynomial of α over $\mathbb{C}$ then $f = x^2 + bx + c$ for some $b, c \in \mathbb{R}$ and α will be one of the two roots

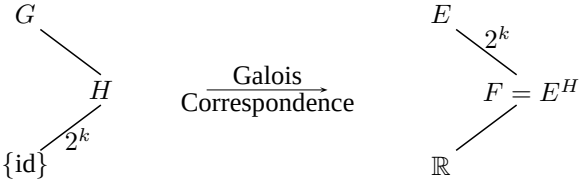
$$\frac{-b \pm \sqrt{b^2 - 4c}}{2},$$

which are in $\mathbb{C}$, contradicting the choice of α.

Fundamental Theorem of Algebra. *Any non-constant $f \in \mathbb{C}[x]$ has a root in $\mathbb{C}$.*

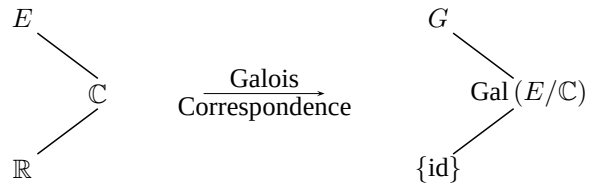
Proof: The proof toggles back and forth between intermediate fields and subgroups of Galois groups using the Galois correspondence. If the polynomial f is reducible over $\mathbb{R}$, with $f = pq$, then replace f by p and continue. Thus we may assume that f is irreducible over $\mathbb{R}$ and let E be the splitting field over $\mathbb{R}$ not of f, but of $(x^2 + 1)f$. Thus in particular we have that $\mathbb{R}$ and $\pm i$ are in E, hence $\mathbb{C}$ is too, and thus $\mathbb{R} \subseteq \mathbb{C} \subseteq E$.

The conditions of the first part of the Galois correspondence hold for E, so we may apply this to the Galois group $G = \text{Gal}\,(E/\mathbb{R})$. Since G is a finite group, we may factor from its order all the powers of 2, writing $|G| = 2^k m$, where $m \geq 1$ is odd. In particular, Sylow's Theorem gives us a subgroup H of G of order 2^k, and so by the Galois correspondence we have the picture:

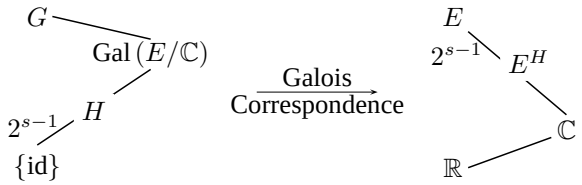


with the intermediate field F corresponding to H giving an extension $F \subseteq E$ of degree 2^k. As $[E : \mathbb{R}] = [E : F][F : \mathbb{R}]$ with $[E : \mathbb{R}] = |G| = 2^k m$, we must have that F is a degree m extension of $\mathbb{R}$. As m is odd and no such extensions exist if $m > 1$, we must have $m = 1$ and so $|G| = 2^k$.

We now use the Galois correspondence in the reverse direction:



As G has order 2^k, the subgroup Gal $(E/\mathbb{C})$ has order dividing this, hence order 2^s. If $s \geq 1$ then Sylow's Theorem again gives us a subgroup H of Gal $(E/\mathbb{C})$ of order 2^{s-1}, and we have the picture:



with $2^{s-1}[E^H : \mathbb{C}] = [E : \mathbb{C}] = |\text{Gal}\,(E/\mathbb{C})| = 2^s$, hence E^H is a degree 2 extension of $\mathbb{C}$. We commented above that there are no such extensions, thus we must have $s = 0$, and so $|\text{Gal}\,(E/\mathbb{C})| = 0$, giving that Gal $(E/\mathbb{C})$ is the trivial group. We have then two fields, namely E and $\mathbb{C}$, that map via the 1-1 map $X \mapsto \text{Gal}\,(E/X)$ to the trivial group, so $E = \mathbb{C}$. As E was the splitting field of the polynomial $(x^2 + 1)f$, we get that f has a root, indeed *all* its roots in $\mathbb{C}$. □

§16. (Not) Solving Equations

At the beginning of these notes we said that Galois Theory was initially motivated by the desire to understand better the roots of polynomial equations. In particular, to provide a context for the growing conviction in Galois' time that there is no formula for the roots of an arbitrary polynomial equation, and that the classical formulae that exist for quadratics, cubics and quartics are some kind of "low degree fluke".

(16.1) The formulae for the roots of quadratics, cubics and quartics express the roots in terms of the coefficients, the four field equations $+, -, \times, \div$ and $\sqrt{}$, $\sqrt[3]{}$, $\sqrt[4]{}$. When we say we want a formula for the roots of polynomials in $\mathbb{Q}[x]$ then, it seems reasonable that it should express the roots in terms of rational numbers, $+, -, \times, \div$ and $\sqrt[m]{}$ for some m. In particular the roots of the polynomial will be contained in an extension of $\mathbb{Q}$ obtained by adjoining certain m-th roots.

With this in mind, an extension $\mathbb{Q} \subseteq E$ is called *radical* if and only if there is a sequence of simple extensions,

$$\mathbb{Q} \subseteq \mathbb{Q}(\alpha_1) \subseteq \mathbb{Q}(\alpha_1, \alpha_2) \subseteq \cdots \subseteq \mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_k) = E,$$

such that $\alpha_i^{m_i} \in \mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_{i-1})$ for every i. Thus, each extension in the sequence is obtained by adjoining to the previous one an m_i-th root of an element.

(16.2) A simple example of a radical extension is,

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt[3]{5}) \subseteq \mathbb{Q}\left(\sqrt{2}, \sqrt[3]{5}, \sqrt{\sqrt{2} - 7\sqrt[3]{5}}\right).$$

By repeatedly applying Theorem D, we see that the elements of a radical extension have expressions in terms of rational numbers, $+, -, \times, \div$ and $\sqrt[m_i]{}$.

(16.3) If we are looking to find a formula for the roots of a polynomial, then these roots will have precisely these kind of expressions. Thus we say that a polynomial $f \in \mathbb{Q}[x]$ is *solvable by radicals* if and only if its splitting field over $\mathbb{Q}$ is contained in some radical extension.

Notice that we are dealing with a fixed specific polynomial, and not an arbitrary one. The radical extension containing the splitting field will depend on the polynomial.

(16.4) Any quadratic polynomial $ax^2 + bx + c$ is solvable by radicals, with its splitting field contained in the radical extension

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{b^2 - 4ac}).$$

Similarly, the formulae for the roots of cubics and quartics give for any specific such polynomial, radical extensions containing their splitting fields.

(16.5) Now we have a precise idea of what we mean by "finding a formula for the roots of a polynomial", we are ready to wheel in the Galois theory. In §**11.** we called a group G *soluble* if and only if there is a sequence,

$$\{1\} = H_0 \lhd H_1 \lhd \cdots \lhd H_{n-1} \lhd H_n = G,$$

such that the successive quotients $H_1/H_0, H_2/H_1, \ldots, H_n/H_{n-1}$ are all *Abelian groups*.

Theorem H (Galois). *A polynomial $f \in \mathbb{Q}[x]$ is solvable by radicals if and only if its Galois group $Gal\,(f)$ is soluble.*

The proof, which we omit, uses the full power of the Galois correspondence, with the sequence of extensions in a radical extension corresponding to the sequence of subgroups in a soluble group.

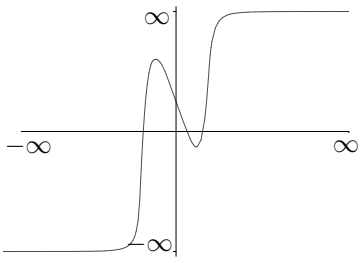
(16.6) Somewhat out of chronological order, we have,

Theorem 11 (Abels-Fubini) *The polynomial $x^5 - 4x + 2$ is not solvable by radicals.*

Proof: We need to show that the Galois group $\mathrm{Gal}\,(f)$ is insoluble. Indeed, we show that it is the symmetric group S_5, which contains the non-Abelian, finite simple group A_5. Thus S_5 contains an insoluble subgroup, hence must be insoluble as well, as any subgroup of a soluble group is soluble by Exercises 102 and 103. If E is the splitting field over $\mathbb{Q}$ of f, then

$$E = \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5),$$

where the α_i are the roots of f and the Galois group of the polynomial is $\mathrm{Gal}\,(E/\mathbb{Q})$. The elements of this groups are, as usual, completely determined by where they send the α_i, and by one of the Corollaries to the Extension Theorem, they must be sent to roots of f. The conclusion is that the elements of the Galois group of the polynomial must permute the roots amongst themselves, so that $\mathrm{Gal}\,(f)$ is a subgroup of the symmetric group S_5.



As α_1 has minimum polynomial f over $\mathbb{Q}$, the extension $\mathbb{Q} \subseteq \mathbb{Q}(\alpha_1)$ has degree five, and then the tower law gives that

$$[E : \mathbb{Q}] = [E : \mathbb{Q}(\alpha_1)][\mathbb{Q}(\alpha_1) : \mathbb{Q}].$$

Thus, the degree of the extension $\mathbb{Q} \subseteq E$ is divisible by the degree of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\alpha_1)$, ie: divisible by five. Moreover, by Theorem G, the group $\mathrm{Gal}\,(E/\mathbb{Q})$ has order the degree $[E : \mathbb{Q}]$, thus the group has order divisible by five. By Sylow's Theorem, this means that the Galois group contains a subgroup of order five. The only groups of order five are the cyclic ones, and as every element of the Galois group is already a permutation of the five roots, this subgroup must have the form,

$$\{\mathrm{id}, \sigma, \sigma^2, \sigma^3, \sigma^4\},$$

for a permutation σ that is a 5-cycle $\sigma = (\alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3}, \alpha_{i_4}, \alpha_{i_5})$. By drawing the graph of f as shown, we see that three of the α_i are real, and so the other two must be complex conjugates. We saw in §**12.** complex conjugation is an automorphism of E, and this must fix the three real roots, and interchange the two complex ones. This gives us another automorphism τ of E that as a permutation has the form,

$$\tau = (\alpha_i, \alpha_j),$$

where α_i, α_j are the two complex roots. □

It is worth meditating briefly on the philosophical implications of this result, which are profound. The Theorem says that there is no possible expression for the roots of the polynomial in terms of rational numbers, the four field operations $+, -, \times, \div$ and roots $\sqrt[m]{}$ for any m. At first this may seem no great problem; we know plenty of real numbers with this property, eg: π. But the roots of the polynomial are *algebraic* numbers, so there is something more, something very subtle, to the notion of an algebraic number than it just being expressible in "algebraic terms".

(16.7) It is sometimes possible to establish the existence of numbers with special properties by counting. For example, to explicitly show that a given real number is transcendental is complicated. If we count the non-transcendental (ie: the algebraic) numbers we see though that they are *countable*: they can be put in 1-1 correspondence with the integers $\mathbb{Z}$, whereas the real numbers are not. Thus, there are many more real numbers than algebraic ones, so transcendentals must exist, indeed greatly outnumber the algebraics.

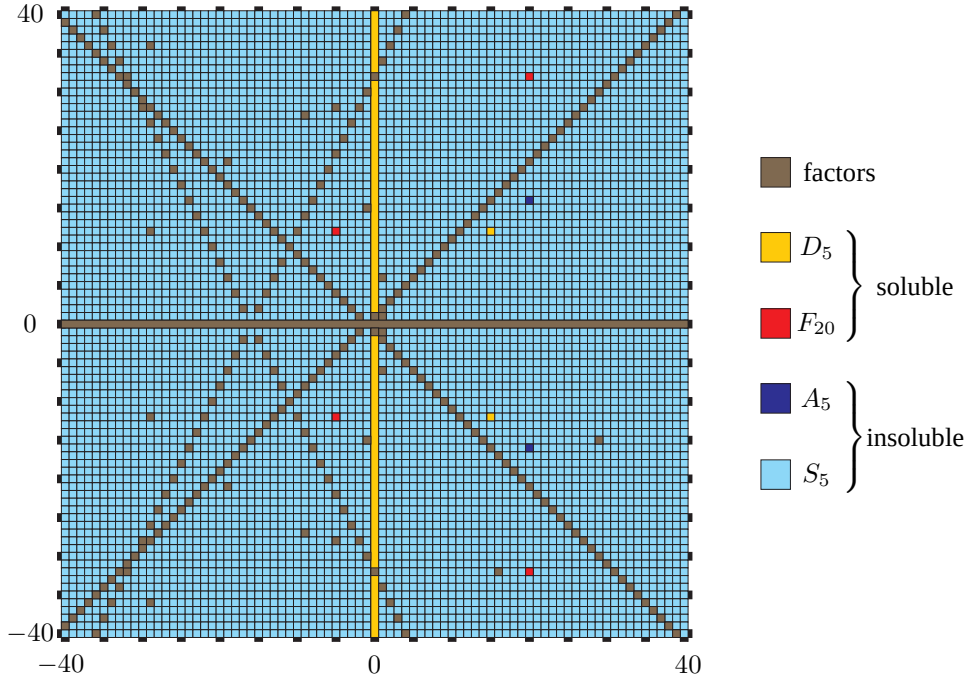
Such a naive approach will not work to establish the existence of the roots of equations not solvable by radicals, as all the sets involved now are countable.

(16.8) So how bad is it then? We have one polynomial, $x^5 - 4x + 2$ for which no algebraic expression exists for its roots, but is this an isolated incident, or at least one that is rare? In fact, polynomials not solvable by radicals are *generic*, in the sense that polynomials that *are* solvable by radicals are the ones that are relatively rare.

We can illustrate this phenomenon at least with some examples. Consider the quintic polynomials

$$x^5 + ax + b,$$

for $a, b \in \mathbb{Z}$ and in the range $-40 \leq a, b \leq 40$.



The picture[16] illustrates the (a, b) plane for this range of a and b. The vertical line through $(0, 0)$ corresponds to f with $\mathrm{Gal}\,(f)$ the soluble dihedral group D_{10} of order 10. The horizontal line through $(0, 0)$ and the two sets of crossing diagonal lines correspond to reducible f, as do a few other isolated points. The (insoluble) alternating group A_5 arises in a few sporadic places, as does another subgroup of S_5. However, the vast majority of f, forming the light background, have Galois group the symmetric group S_5, and so have roots that are *algebraic*, but cannot be expressed *algebraically*.

[16] which is based on an image from the Mathematica poster, "Solving the Quintic".

§17. Selected Solutions

(4) If $u = \omega + \omega^{-1} = \omega + \omega^4$ then $u^2 = \omega^2 + \omega^3 + 2$ so the quadratic we want is $u^2 + u - 1 = 0$. This has roots

$$\frac{-1 \pm \sqrt{5}}{2}.$$

We get $\omega^2 - u\omega + 1 = 0$ by multiplying through by ω^{-1}, hence

$$\omega = \frac{u \pm \sqrt{u^2 - 4}}{2}.$$

and substitute u into this to get,

$$\omega = \frac{\frac{-1\pm\sqrt{5}}{2} + \sqrt{(\frac{-1\pm\sqrt{5}}{2})^2 - 4}}{2}, \frac{\frac{-1\pm\sqrt{5}}{2} - \sqrt{(\frac{-1\pm\sqrt{5}}{2})^2 - 4}}{2}.$$

(I don't expect you to actually do this!)

(6) Note that ω is a sixth root of 1 (in fact it has argument $2\pi/6$ and modulus 1) so satifies $\omega^6 = 1$ (but $\omega^k \neq 1$ for any k between 1 and 5). Clearly $\alpha\omega^2$ and ω^5 are $\in \mathbb{Q}(\alpha, \omega)$, so that $\mathbb{Q}(\alpha\omega^2, \omega^5) \subseteq \mathbb{Q}(\alpha, \omega)$. Conversly, $\alpha\omega^2, \omega^5 \in \mathbb{Q}(\alpha\omega^2, \omega^5) \Rightarrow \alpha\omega^2\omega^5\omega^5 \in \mathbb{Q}(\alpha\omega^2, \omega^5)$, but $\alpha\omega^2\omega^5\omega^5 = \alpha\omega^{12} = \alpha$ since $\omega^6 = 1$. Thus $\alpha \in \mathbb{Q}(\alpha\omega^2, \omega^5)$ and hence $\alpha^{-1}\alpha\omega^2 = \omega^2$ is too, and so finally $\omega^2\omega^5 = \omega^7 = w$ (since $\omega^6 = 1$). Thus $\mathbb{Q}(\alpha, \omega) \subseteq \mathbb{Q}(\alpha\omega^2, \omega^5)$.

To show that $\mathbb{Q}(\alpha, \omega) = \mathbb{Q}(\alpha\omega^4, \omega^5)$ is entirely similar.

(7) Consider the extension field $\mathbb{Q}(\alpha, \omega)$ of $\mathbb{Q}$. Note first that the solutions to $x^5 - 2$ all lie in this field, as it contains α, ω and is closed under multiplication.

The following paragraph is optional: $\mathbb{Q}(\alpha, \omega)$ is in fact the smallest field that contains the solutions. For, suppose that F is some field containing $\alpha, \alpha\omega, \ldots, \alpha\omega^4$. Since we are in the complex numbers,

$$\underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} \neq 0,$$

for any n. Thus, F contains the rationals $\mathbb{Q}$. Also, $\alpha, \alpha\omega \in F$ gives that α and $\alpha^{-1}\alpha\omega = \omega \in F$ too, ie: F contains $\mathbb{Q}$, α and ω. But it must then contain the smallest field that does these things, ie: $\mathbb{Q}(\alpha, \omega) \subset F$, and so $\mathbb{Q}(\alpha, \omega)$ really is the smallest.

So, symmetries of the soltions to $x^5 - 2$ are the rearrangements referred to in the first lecture, but of this new field $\mathbb{Q}(\alpha, \omega)$.

Looking at the picture of the pentagon, the symmetry needs to send α to itself and $\alpha\omega$ to $\alpha\omega^3$. This last one suggests that it must send ω to ω^3 (if the symmetry is not to disturb the $+$ and $\times$ of the field).

To see if such a symmetry exists, we need to show that the fields $\mathbb{Q}(\alpha, \omega)$ and $\mathbb{Q}(\alpha, \omega^3)$ are the same. Certianly, $\alpha, \omega^3 \in \mathbb{Q}(\alpha, \omega)$, so that $\mathbb{Q}(\alpha, \omega^3) \subset \mathbb{Q}(\alpha, \omega)$. Conversly, $\omega^3 \in \mathbb{Q}(\alpha, \omega^3)$ gives $\omega^3\omega^3 = \omega^6 = \omega \in \mathbb{Q}(\alpha, \omega^3)$. We already have that α is there too, so $\mathbb{Q}(\alpha, \omega) \subset \mathbb{Q}(\alpha, \omega^3)$.

Finally, we need to check that the symmetry $\alpha \mapsto \alpha$ and $\omega \mapsto \omega^3$, does the right thing to the vertices of the pentagon. Well, let's try $\alpha\omega^3 \mapsto \alpha(\omega^3)^3 = \alpha\omega^{27} = \alpha\omega^2$. The others are entirely analogous.

(19)

1. For multiplication, let $n \neq 0$ be in $\mathbb{F}_p$. Then the gcd of n and p must be 1, so that for some integers a, b we have $1 = an + bp$. But then $an = (-b)p + 1 = 1 \bmod p$. Thus, if $a = k \bmod p$ then the inverse of n is k.
2. Let $ab = 0$ for $a, b \in F$. Then either $a = 0$, or if $a \neq 0$ then $a^{-1}ab = a^{-1}0 \Rightarrow b = 0$. Thus, at least one of a or b must be zero. Thus F is an integral domain.

   Let $n = rs$ with $1 < r, s < n$ integers, and consider $\mathbb{Z}_n$, the ring of integers with addition and mutliplication modulo n. Then $r, s \neq 0$ in $\mathbb{Z}_n$, but $rs = n = 0 \bmod n$. Thus $\mathbb{Z}_n$ is not an integral domain.

(22)

1. Use $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$ and $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ (you can easily convince yourself of these by drawing vectors in the plane).

   First note that $\overline{f(z)} = f(\overline{z})$: use the two rules above with $a_n x^n + \cdots a_0$, remembering that the $a_i \in \mathbb{R}$ means $\overline{a_i} = a_i$.

   Thus, z is a root of $f \Leftrightarrow f(z) = 0 \Leftrightarrow \overline{f(z)} = \overline{0} = 0 \Leftrightarrow f(\overline{z}) = 0$.
2. There are many examples. The simplest is probably $x^2 - i$, since the square roots of i, by De Moivre's theorem, lie on the circle $|z| = 1$; one has argument $\pi/4$, the other $5\pi/4$.

(23)

1. We have $1 = am + bn$ for some integers a and b, hence $k = amk + bnk$. If m divides nk then, as it already divides amk, it must also divide k as required.

2. If m/n is such a root, then
$$a_0 + a_1(m/n) + a_2(m/n)^2 + \cdots + a_r(m/n)^r = 0\,.$$
Clearing denominators gives
$$a_0 n^r + a_1 n^{r-1} m + a_2 n^{r-2} m^2 + \cdots + a_{r-1} n m^{r-1} + a_r m^r = 0\,.$$
Consider n. The number n divides 0 and clearly divides every term on the left, with the possible exception of $a_r m^r$. But since it divides everything else in sight, n must also divide $a_r m^r$. Our assumption that m and n be coprime implies that n does not divide m^r, and hence we conclude that $n|a_r$. Similarly, $m|a_0$. Finally, if $a_r = 1$, then we must have $n \in \{\pm 1\}$, so that m/n is indeed an integer.

(26)

1. $1 + x^8$ has no real roots as $1 + x^8 > 1 \neq 0$. But, in $\mathbb{C}$ we get
$$1 + x^8 = (x - \zeta_1)(x - \overline{\zeta_1}) \cdots (x - \zeta_4)(x - \overline{\zeta_4}),$$
by the fundamental theorem of algebra. Thus,
$$1 + x^8 = (x^2 + (\zeta_1 + \overline{\zeta_1})x + \zeta_1\overline{\zeta_1}) \cdots (x^2 + (\zeta_4 + \overline{\zeta_4})x + \zeta_4\overline{\zeta_4}),$$
and these are real polynomials since the sum and product of a complex number with its conjugate is real. Thus $1 + x^8$ is reducible as a product of four quadratics.
2. Following the hint, 1 is clearly a root of $y^n - 1$, and we get
$$y^n - 1 = (y - 1)(1 + y + y^2 + \cdots + y^{n-1}),$$
so letting $y = x^2$ and $n = 6$ gives
$$(1 + x^2 + x^4 + x^6 + x^8 + x^{10})(x^2 - 1) = x^{12} - 1.$$
Thus the roots of the right hand side are the 12^{th} roots of 1. Hence the roots of the left hand side are also the 12^{th} roots of 1. Now ± 1 are certainly the roots of $x^2 - 1$, and these are two of the 12^{th} roots, so the other ten are the roots of the polynomial that we are interested in. Notice that they are all $\in \mathbb{C}$, but two of them are $\pm i$, hence
$$(x - i)(x + i),$$
are factors of $1 + x^2 + x^4 + x^6 + x^8 + x^{10}$, ie: $x^2 + 1$ is a factor of $1 + x^2 + x^4 + x^6 + x^8 + x^{10}$ (notice that this argument, while more complicated than others maybe, works for $1 + x^2 + \cdots + x^{2n}$). Having got that far, its then pretty easy to spot that
$$1 + x^2 + x^4 + x^6 + x^8 + x^{10} = (x^2 + 1)(1 + x^4 + x^8),$$
so that the poynomial is reducible over $\mathbb{Q}$. Can you generalise the argument to handle $1 + x^2 + \cdots + x^{2n}$?
3. The polynomial has value 7 when $x = 0$ and value -7 when $x = -1$, hence by the intermediate value theorem, there must be a real root somewhere between -1 and 1 (polynomials are continuous, so the graph must cross the x-axis!) Hence we have a linear factor and thus the polynomial is reducible.
4. The polynomial has integer coefficients, 2 divides all of them except that of the leading term and $2^2 = 4$ does not divide the constant term. Thus, by Eisenstein, the polynomial is irreducible over $\mathbb{Q}$.
5. We are dealing with a quadratic, so irreducibility becomes a matter of merely checking for roots. In $\mathbb{Z}_7$, no element squared plus one is equal to zero, so the polynomial is irreducible.
6. It looks complicated, but we have a cubic, and that means again that all we need do is check for roots, this time in the field $\mathbb{F}$ of order eight of §**4.**. In fact, we've gone to all the trouble of writing the $+$ and $\times$ tables out, so we may as well use them! Somewhat dissapointingly, 1 turns out to be a root, so the polynomial is reducible straight away. We can say a little more (although this is not necessary) since no other element of $\mathbb{F}$ is a root, and therefore the cubic must factorise into a product of a linear and a quadratic factor. If you were sufficiently interested, you could find the irreducible quadratic by long division (which works in exactly the same way as long division in the reals, since we are still in a field!)

(27) We have that
$$\left(\begin{array}{c} p \\ i \end{array} \right) = \frac{p!}{i!(p-i)!} = m,$$
for m an integer, as there are clearly an integral number of ways of choosing i objects from p. Thus,
$$p! = mi!(p - i)!,$$
so that as p divides the left hand side, it also divides $mi!(p - i)!$. As $i < p$, we can't have p dividing i or any integer less than it, hence not dividing $i!$. Similarly p doesn't divide $(p - i)!$, and so it must divide m (all of which uses the fact that p is prime).

(29)

1. Clearly ω is a root of $x^n - 1$ and
$$x^n - 1 = (x - 1)(x^{n-1} + x^{n-2} + \cdots + x + 1),$$
and ω is not a root of $x - 1$ (since ω being primitive, is not 1), so must be of the desired polynomial.

2. If n even then -1 is a root of $x^n - 1$ and thus of $x^{n-1} + x^{n-2} + \cdots + x + 1$ too. Thus $x^{n-1} + x^{n-2} + \cdots + x + 1$ is irreducible only if n is odd.

3. If $f(x) = g(x)h(x)$, then obvioulsy $f(x+1) = g(x+1)h(x+1)$, contradicting the irreducibility of $f(x+1)$. Thus $f(x)$ is irreducible too.

4. We know that
$$\left(\begin{array}{c} p \\ i \end{array} \right) = \frac{p!}{i!(p-i)!},$$
is an integer m, so $p(p-1)! = mi!(p-i)!$, hence p divides one of m, $i!$ or $(p-i)!$ If $p|i!$ then it divides a j with $1 < j < p$ which cannot be. Similarly, p cannot divide $(p-i)!$, and thus p divides m.

5. By the above,
$$x^p - 1 = (x-1)\Phi_p(x) \Rightarrow \Phi_p(x) = \frac{x^p - 1}{x - 1} \Rightarrow \Phi_p(x+1) = \frac{(x+1)^p - 1}{x},$$
and using the binomial theorem and cancelling we get
$$\Phi_p(x+1) = x^{p-1} + px^{p-2} + \cdots + \left(\begin{array}{c} p \\ i \end{array} \right) x^{p-i-1} + \cdots + p.$$
Using Eisenstein, with p as the prime, gives $\Phi_p(x+1)$ irreducible by part (d), and hence $\Phi_p(x)$ too by part (c).

(40)

1. $(1 + x + x^2) + (1 + x) = x^2$ (since $1 + 1 = 0 = x + x$ in $\mathbb{Z}_2$ arithmetic). Similarly, $(1 + x + x^2)(1 + x) = 1 + x + x^2 + x + x^2 + x^3 = 1 + x^3 = 1 + 1 + x = x$ (using the rule $x^3 = x + 1$).

2. $\mathbb{F} = \{0, 1, x, x+1, x^2, x^2+1, x^2+x, x^2+x+1\}$, so $\mathbb{F}$ has eight elements.

3. The tables are (somewhat tediously!)

$+$	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
0	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
1	1	0	$x+1$	x	x^2+1	x^2	x^2+x+1	x^2+x
x	x	$x+1$	0	1	x^+x	x^2+x+1	x^2	x^2+1
$x+1$	$x+1$	x	1	0	x^2+x+1	x^2+x	x^2+1	$x^"$
x^2	x^2	x^2+1	x^2+x	x^2+x+1	0	1	x	$x+1$
x^2+1	x^2+1	x^2	x^2+x+1	x^2+x	1	0	$x+1$	x
x^2+x	x^2+x	x^2+x+1	x^2	x^2+1	x	$x+1$	0	1
x^2+x+1	x^2+x+1	x^2+x	x^2+1	x^2	$x+1$	x	1	0

Now, $(\mathbb{F}, +)$ is an Abelian group for the following reasons: the table closes up (we don't get anything new and unexpected), so the field is closed under $+$; the first row is identical to the indexing along the top, so 0 is the identity under $+$; each row contains 0 somewhere in it, so inverses exist for all elements. Unfortunately, associativity isn't quite so easily established!

Similarly,

$\times$	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
0	0	0	0	0	0	0	0	0
1	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
x	0	x	x^2	x^2+x	$x+1$	1	x^2+x+1	x^2+1
$x+1$	0	$x+1$	x^2+x	x^2+1	x^2+x+1	x^2	1	x
x^2	0	x^2	$x+1$	x^2+x+1	x^2+x	x	x^2+1	1
x^2+1	0	x^2+1	1	x^2	x	x^2+x+1	$x+1$	x^2+x
x^2+x	0	x^2+x	x^2+x+1	1	x^2+1	$x+1$	x	x^2
x^2+x+1	0	x^2+x+1	x^2+1	x	1	x^2+x	x^2	$x+1$

shows that $(\mathbb{F} \setminus \{0\}, \times)$ is an Abelian group. The distributive law is also a bit tedious!

Finally,
$$\frac{1}{1+x} = x^2 + x \text{ and } \frac{1}{1+x+x^2} = x^2 \text{ in } \mathbb{F},$$
from the tables.

(47)

1. We certainly have that $\langle 0 \rangle$ is an ideal. Suppose that $\langle \lambda \rangle$ is another one with $\lambda \neq 0$. For any $\mu \in F$ we have that $\mu = \mu\lambda^{-1}\lambda$ (the inverse of λ existing as F is a field) so that μ is a multiple of λ and hence in the ideal $\langle \lambda \rangle$. Thus $\langle \lambda \rangle = F$. The conclusion is that F contains only the two ideals $\langle 0 \rangle$ and F.

2. We need only show that every non-zero element of R has an inverse under multiplication. Let $r \neq 0$ be such an element and consider the ideal $\langle r \rangle$. By the restriction on the possible ideals we have either $\langle r \rangle = \langle 0 \rangle = \{0\}$ or $\langle r \rangle = R$. As $r \in \langle r \rangle$ the first one cannot happen so that it is $\langle r \rangle = R$ that we have. In particular $1 \in \langle r \rangle$, ie: there is an $s \in R$ such that $sr = 1$ and by commutativity we also have that $rs = 1$. Thus s is the inverse of r as required.

(61)

1. The polynomial has no roots in $\mathbb{F}_3$, so is irreducible as it is a cubic. The quotient ring given is then a field.
2. We use the division algorithm:
   $g(x) + \langle 1 - x + x^3 \rangle = (q(x)(1 - x + x^3) + (a + bx + cx^2)) + \langle 1 - x + x^3 \rangle = (a + bx + cx^2) + \langle 1 - x + x^3 \rangle$,
   for any coset $g(x) + \langle 1 - x + x^3 \rangle$. The uniqueness follows from the fact that the quotient and remainder are uniquely determined by the division algorithm (See the first handout on rings).
3. There are three choices for each of a, b and c in $(a + bx + cx^2) + \langle 1 - x + x^3 \rangle$, so that the field has at *most* 27 elements. On the other hand, suppose
   $$(a_1 + b_1 x + c_1 x^2) + \langle 1 - x + x^3 \rangle = (a_2 + b_2 x + c_2 x^2) + \langle 1 - x + x^3 \rangle,$$
   for some a_i, b_i, c_i with $i = 1, 2$. Then, using some of the basic properties of cosets described in the lectures, we get that,
   $((a_1-a_2)+(b_1-b_2)x+(c_1-c_2)x^2)+\langle 1-x+x^3 \rangle = \langle 1-x+x^3 \rangle \Rightarrow (a_1-a_2)+(b_1-b_2)x+(c_1-c_2)x^2 \in \langle 1-x+x^3 \rangle \Rightarrow$.
   (the last since two cosets the same means the difference of the representative polynomials is a multiple of $f(x)$). But the degree of $(a_1 - a_2) + (b_1 - b_2)x + (c_1 - c_2)x^2$ is two, while every muliple of $1 - x + x^3$ has degree 3 or more, except for one: the zero polynomial. Thus $(a_1 - a_2) + (b_1 - b_2)x + (c_1 - c_2)x^2$ must be the zero polynomial, ie $a_1 = a_2$, $b_1 = b_2$ and $c_1 = c_2$. Thus all the 27 cosets listed are different (there really are 27 of them!)

(69) Notice that one of the 5-th roots of 1 found in the first question is (after a little massaging) equal to
$$\omega = \frac{\sqrt{5} - 1}{4} + \frac{\sqrt{2}\sqrt{5 + \sqrt{5}}}{4} i.$$
In fact, this is the first vertex anticlockwise around the circle from 1. Now, this number is constructible precisely when its real and imaginary parts,
$$\frac{\sqrt{5} - 1}{4}, \frac{\sqrt{2}\sqrt{5 + \sqrt{5}}}{4},$$
are constructible. But these last two numbers can be obtained from integers using the four field operations and by taking $\sqrt{}$'s, all of which we can do with ruler and compass. Thus ω is constructible, hence so is the desired pentagon, just by stepping off the length of the line segment joining 1 to ω using your compass.

(70) If the length of the line segment is x, then the task is to construct $\frac{x}{3}$ and $\frac{2x}{3}$. We can certainly construct $\frac{1}{3}$ and $\frac{2}{3}$ using our ruler and compass, and we can multiply lengths using these two tools as well. So, multiply the two fractions by the line segment, and we're done (notice that we can use this argument to n-sect a line for any n, ie: divide it into n equal parts).

(72) The best way to do this part is to use a picture proof that can be run in both directions. Alternatively, one can write out a solution in terms of words, and since this easier to LaTeX, I'll do it that way.

If θ is constructible, then assuming without loss of generality that one side of the angle is the x-axis, the intersection of the other side with the unit circle is the point $(\cos\theta, \sin\theta)$. Dropping a vertical line to the x-axis gives us $\cos\theta$. Conversly, if $\cos\theta$ is constructible, then so is $\sqrt{1 - \cos^2\theta}$, as the field of (complex) constructible numbers is closed under taking of square roots. Hence $\sin\theta$ can be constructed up the y-axis, and horizontal and vertical lines determine $(\cos\theta, \sin\theta)$, and a line through this point and the origin constructs the angle θ.

(76) We begin by computing $(1 + a)^{-1}$. According to the lemma on the structure of simple algebraic extensions, this element must have the form $ba^2 + ca + d$ for some uniquely defined b, c, and d in $\mathbb{Q}$. So we simply need to solve the equation
$$(a + 1)(ba^2 + ca + d) = 1$$
for these rational coefficients. Expanding, we find that
$$ba^3 + (b + c)a^2 + (c + d)a + d = 1\,,$$
so that we have three equations for b, c, and d, namely:
$$\begin{aligned} b + c &= 0 \\ c + d &= 0 \\ 2b + d &= 1 \end{aligned}$$
(where in the last equation we've used the fact that $a^3 = 2$).

Now, the first two equations imply that $b = -c = d$, and hence from the third equation we see that $3b = 1$. Hence, $b = 1/3$, $c = -1/3$, and $d = 1/3$, so that the required inverse for $(a + 1)$ is
$$(a + 1)^{-1} = \frac{1}{3}(a^2 - a + 1)\,.$$
(If you like, you can check your answer by multiplying by $(a + 1)$.)

In a similar manner, we can compute that
$$(a^2 + 1)^{-1} = \frac{1}{5}(-a^2 + 2a + 1)\,,$$
so that, since $a^4 + 1 = 2a + 1$, we have that
$$\begin{aligned} (a^4 + 1)(a^2 + 1)^{-1} &= (2a + 1)\tfrac{1}{5}(-a^2 + 2a + 1) \\ &= \cdots = \tfrac{1}{5}\left(3a^2 + 4a - 3\right)\,. \end{aligned}$$

(77) The minimum polynomial for $\alpha = \sqrt[3]{5}$ over $\mathbb{Q}$ is $x^3 - 5$, so every element of $\mathbb{Q}(\alpha)$ has the form $a + b\alpha + c\alpha^2$ for a unique choice of a, b and c in $\mathbb{Q}$. In each case we explicitly use the fact that $\alpha^3 = 5$.

1.
2. This is really easy: $\alpha^5 - \alpha^6 = 5\alpha^2 - 25$.
3. Set $\alpha/(\alpha^2 + 1) = a + b\alpha + c\alpha^2$, and solve for a, b and c in $\mathbb{Q}$. We multiply through by $(\alpha^2 + 1)$ and see that

$$\alpha = (\alpha^2 + 1)(a + b\alpha + c\alpha^2) = (a + 5b) + (b + 5c)\alpha + (a + c)\alpha^2 \,,$$

so that equating coefficients reveals the equations

$$a + 5b = 0 \qquad b + 5c = 1 \qquad a + c = 0 \,.$$

The last equation forces $a = -c$, and substituting in, we're left with the two equations $a + 5b = 0$ and $b - 5a = 1$. So $a = -5b$ and $c = 5b$ and the middle equation becomes $26b = 1$. Hence

$$\alpha/(\alpha^2 + 1) = (-5/26) + (1/26)\alpha + (5/26)\alpha^2 \,.$$

(78)

Let $\alpha = \sqrt{2} + \sqrt{-2} = \sqrt{2} + i\sqrt{2}$. Then $\alpha^2 = 4i$, so that $\alpha^4 = -16$. Consider the polynomial $f(x) = x^4 + 16$. This is monic and has α as a root. Moreover, we claim that this is irreducible and hence is the correct minimum polynomial. First note that $f(x)$ has no roots in $\mathbb{Q}$, because $f(a) \geq 16$ for all a in $\mathbb{Q}$. We still need to show that $f(x)$ does not factor as a product of quadratics, and this is best done by drawing the roots in the complex plane, being

$$\pm\frac{2}{\sqrt{2}} \pm \frac{2}{\sqrt{2}}i$$

Any factorisation into quadratics of this polynomial would come from multiplying two terms of the form,

$$(x - \zeta_1)(x - \zeta_2),$$

but these never give polynomials with rational coefficients. For the rest, we use the fact that $\alpha^4 = -16$. Note that it's not at all obvious that the first three numbers actually lie in $\mathbb{Q}(\alpha)$.For the first part, we begin by showing that $\sqrt{2} \in \mathbb{Q}(\alpha)$. [We could just solve as we did in part (c) of problem 3, but hopefully the method we're using will produce the answer almost immediately.] We compute that, in addition to the above information,

$$\begin{aligned} \alpha^3 &= (\sqrt{2} + i\sqrt{2})(4i) \\ &= -4\sqrt{2} + 4i\sqrt{2} \,. \end{aligned}$$

Hence it's easy to see that $\sqrt{2} = (1/8)(4\alpha - \alpha^3)$. For exactly the same reason, $\sqrt{-2} = i\sqrt{2} = (1/8)(4\alpha + \alpha^3)$. This is the easiest of the lot: $i = (1/4)\alpha^2$. Since $\alpha^4 = -16$, we have

$$\alpha^5 + 4\alpha + 3 = -16\alpha + 4\alpha + 3 = 3 - 12\alpha \,.$$

We need to find a polynomial $g(\alpha)$ in α which satisfies $\alpha g(\alpha) = 1$. Since $\alpha^4 = -16$, we have $(-1/16)\alpha^4 = 1$, which means that $1/\alpha = (-1/16)\alpha^3$. This last is the most complicated. We set $(2\alpha+3)/(\alpha^2+2\alpha+2) = a+b\alpha+c\alpha^2+d\alpha^3$, and solve for a, b and c in $\mathbb{Q}$. We multiply through by $(\alpha^2 + 2\alpha + 2)$ and see that

$$\begin{aligned} 2\alpha + 3 &= (\alpha^2 + 2\alpha + 2)(a + b\alpha + c\alpha^2 + d\alpha^3) \\ &= (2a - 16c - 32d) + (2a + 2b - 16d)\alpha + (a + 2b + 2c)\alpha^2 + (b + 2c + 2d)\alpha^3 \,. \end{aligned}$$

so that equating coefficients reveals the equations

$$\begin{aligned} 2a - 16c - 32d &= 3 \\ 2a + 2b - 16d &= 2 \\ a + 2b + 2c &= 0 \\ b + 2c + 2d &= 0 \end{aligned}$$

Hence we have 4 equations in 4 unknowns and can solve to find $a = -1/2$, $b = 1/6$, $c = 1/12$ and $d = -1/6$, and so

$$(2\alpha + 3)/(\alpha^2 + 2\alpha + 2) = -1/2 + 1/6\alpha + 1/12\alpha^2 - 1/6\alpha^3 \,.$$

(79)

1. Let $\alpha = 1 + i$. Then $\alpha^2 = 2i$, so that $\alpha^2 - 2\alpha = -2$. In particular, α is a root of the polynomial $f(x) = x^2 - 2x + 2$ over $\mathbb{Q}$. This is also monic, so we just need to decide whether or not $f(x)$ is irreducible over $\mathbb{Q}$. However, the only way that $f(x)$ could be reducible is if it factored as $f(x) = (x - a)(x - b)$ with both a and b in $\mathbb{Q}$. But we Fnow that one of these roots would necessarily be α, so $f(x)$ is indeed irreducible. Hence $f(x)$ is the required minimum polynomial.
2. The obvious polynomial to try this time is $g(x) = x^3 - 7$, since it's monic and clearly has $\sqrt[3]{7}$ as a root. Moreover, Eisenstein's criterion immediately applies with $q = 7$ and we see that $g(x)$ is irreducible over $\mathbb{Q}$. Hence $g(x)$ is the required minimum polynomial.
3. Just as in the previous part, the minimum polynomial for $\sqrt[4]{5}$ over $\mathbb{Q}$ is $x^4 - 5$.

4. This time there's really no obvious choice, so we need to consider relations between powers of $\alpha = \sqrt{2} + i$:

$$\begin{aligned} \alpha &= \sqrt{2} + i \\ \alpha^2 &= 1 + 2i\sqrt{2} \\ \alpha^3 &= \sqrt{2} + i + 4i - 2\sqrt{2} \\ &= -\sqrt{2} + 5i \\ \alpha^4 &= (1 + 2i\sqrt{2})^2 = 1 - 8 + 4i\sqrt{2} \\ &= -7 + 4i\sqrt{2} \end{aligned}$$

So we notice without too much trouble that $\alpha^4 - 2\alpha^2 = -9$, so that α is a root of the monic polynomial $h(x) = x^4 - 2x^2 + 9$. We need to check irreducibility of $h(x)$ over $\mathbb{Q}$. Equivalently, we can work over $\mathbb{Z}$, and begin by checking for linear factors. For any root a of $h(x)$ in $\mathbb{Z}$, we must have $a|9$. It's easy to see that none of $\pm 1, \pm 3, \pm 9$ is a root. So $h(x)$ has no linear factors. If we try to factorise $h(x)$ as

$$\begin{aligned} x^4 - 2x + 9 &= (x^2 + ax + b)(x^2 + cx + d) \\ &= x^4 + (a + c)x^3 + (ac + b + d)x^2 + (ad + bc)x + bd\,. \end{aligned}$$

Equating coefficients, we conclude that $a = -c$, and hence that $0 = ad + bc = a(d - b)$. So either $a = 0$ or $d = b$.

First assume that $a = 0$, and hence $c = 0$. The remaining equations then imply that $bd = 9$ and $b + d = -2$, and these equations have no solution in $\mathbb{Z}$.

So we must therefore have $a \neq 0$, and hence $d = b$. In that case the constant coefficient is $b^2 = 9$. This forces b to be 3 or -3. Consider now the remaining equation, which can be written as $b + d = a^2 - 2$, or $2b = a^2 - 2$. Hence this implies that $a^2 = 2b + 2$. But with our choices for b, we then have $a^2 = 8$ or $a^2 = -2$, neither of which has a solution in $\mathbb{Z}$.

Therefore, $h(x)$ is irreducible over $\mathbb{Q}$ and hence is the required minimum polynomial.

5. One way to do this part is to argue exactly as we did in part (d). We set $\alpha = \sqrt{2} + \sqrt[3]{3}$ and compute powers of α, looking for a relationship. If we try this, we compute powers up to α^6 and compare coefficients of $1, \alpha, \ldots, \alpha^6$ to get 6 equations in 7 unknowns,

$$\begin{aligned} 1 &= 1 \\ \alpha &= \sqrt{2} + \sqrt[3]{3} \\ \alpha^2 &= 2 + 2\sqrt{2}\sqrt[3]{3} + (\sqrt[3]{3})^2 \\ \alpha^3 &= 3 + 2\sqrt{2} + 6\sqrt[3]{3} + 3\sqrt{2}(\sqrt[3]{3})^2 \\ \alpha^4 &= 4 + 12\sqrt{2} + 3\sqrt[3]{3} + 8\sqrt{2}\sqrt[3]{3} + 12(\sqrt[3]{3})^2 \\ \alpha^5 &= 60 + 4\sqrt{2} + 20\sqrt[3]{3} + 15\sqrt{2}\sqrt[3]{3} + 3(\sqrt[3]{3})^2 + 20\sqrt{2}(\sqrt[3]{3})^2 \\ \alpha^6 &= 17 + 120\sqrt{2} + 90\sqrt[3]{3} + 24\sqrt{2}\sqrt[3]{3} + 60(\sqrt[3]{3})^2 + 18\sqrt{2}(\sqrt[3]{3})^2 \end{aligned}$$

and hence the matrix

$$A := \begin{bmatrix} 1 & 0 & 2 & 3 & 4 & 60 & 17 \\ 0 & 1 & 0 & 2 & 12 & 4 & 120 \\ 0 & 1 & 0 & 6 & 3 & 20 & 90 \\ 0 & 0 & 2 & 0 & 8 & 15 & 24 \\ 0 & 0 & 1 & 0 & 12 & 3 & 60 \\ 0 & 0 & 0 & 3 & 0 & 20 & 18 \end{bmatrix}$$

Row-reduction yields:

$$\begin{bmatrix} 1 & 0 & 2 & 3 & 4 & 60 & 17 \\ 0 & 1 & 0 & 2 & 12 & 4 & 120 \\ 0 & 0 & 2 & 0 & 8 & 15 & 24 \\ 0 & 0 & 0 & 4 & -9 & 16 & -30 \\ 0 & 0 & 0 & 0 & 8 & \frac{-9}{2} & 48 \\ 0 & 0 & 0 & 0 & 0 & \frac{755}{64} & 0 \end{bmatrix}$$

Back-substitution finally yields the polynomial

$$f(x) = x^6 - 6x^4 - 6x^3 + 12x^2 - 36x + 1\,,$$

which is monic and has α as a root.

Another approach which leads to the same polynomial is to begin with the equation $\alpha = \sqrt{2} + \sqrt[3]{3}$ and eliminate radicals. So write

$$\alpha - \sqrt{2} = \sqrt[3]{3}\,.$$

Cubing gives

$$\begin{aligned} (\alpha - \sqrt{2})^3 &= 3 \\ \alpha^3 - 3\sqrt{2}\alpha^2 + 6\alpha - 2\sqrt{2} &= 3 \\ (\alpha^3 + 6\alpha) - \sqrt{2}(3\alpha^2 + 2) &= 3 \\ \sqrt{2}(3\alpha^2 + 2) &= \alpha^3 + 6\alpha - 3\,, \end{aligned}$$

so that squaring gives

$$\begin{aligned} 2(3\alpha^2+2)^2 &= (\alpha^3+6\alpha-3)^2 \\ 2(9\alpha^4+12\alpha^2+4) &= \alpha^6+36\alpha^2+9+12\alpha^4-6\alpha^3-36\alpha\,. \end{aligned}$$

Rearranging gives

$$\alpha^6-6\alpha^4-6\alpha^3+12\alpha^2-36\alpha+1=0$$

which yields the same polynomial.

It remains to check that

$$f(x)=x^6-6x^4-6x^3+12x^2-36x+1$$

is irreducible over $\mathbb{Q}$. Reducing mod 3 gives

$$\overline{f}(x)=x^6+1$$

which clearly has no roots in $\mathbb{F}_3$. Hence if theis reduces, then it must have a monic, irreducible quadratic or cubic factor. Of the 9 possible quadratic polynomials over $\mathbb{F}_3$, a quick check shows that only x^2+1, x^2+x+2 and x^2+2x+2 are irreducible, and a little work shows none of these divides $\overline{f}(x)$.

A similar argument with cubics shows that $\overline{f}(x)$ has no cubic factors either, and hence $f(x)$ is irreducible.

To see that it is, note that since $\sqrt{11}$ is irrational, so is α. [If you don't believe this, you can see this as follows: if $\alpha=m/n\in\mathbb{Q}$, then we would have $\sqrt{11}=(2m+3n)/n$, contradicting the irrationality of $\sqrt{11}$.] (That $\sqrt{11}$ is irrational follows from the fact that the polynomial x^2-11, has roots that are either integers or irrational, by a question from the first assignment. Clearly no integer can square to give 11, so $\sqrt{11}$ must be irrational.) In particular then, $\alpha\notin\mathbb{Q}$, which means that $q(x)$ cannot be factored over $\mathbb{Q}$, as such a factorisation would be of the form $q(x)=(x-\alpha)(x-\beta)$ for some $\beta\in\mathbb{Q}$. Hence $m(x)=q(x)$ is the minimum polynomial for α over $\mathbb{Q}$.As in (b) we set $\beta=(i\sqrt{3}-1)/2$ and work with 2β. We compute that $(2\beta)^2=-2-2i\sqrt{3}$, so that

$$(2\beta)^3=(i\sqrt{3}-1)(-2-2i\sqrt{3})=-2i\sqrt{3}+2-2(-1)(3)+2i\sqrt{3}=8\,.$$

But that means that $8\beta^3=8$, i.e. $\beta^3=1$. Hence β is a root of the polynomial $p(x)=x^3-1$. However, this isn't the minimum polynomial, because $p(x)$ isn't irreducible. Using cyclotomic polynomials (or just by observation), we see that $p(x)$ factors as

$$p(x)=\phi_1\phi_3=(x-1)(x^2+x+1)\,.$$

The quadratic factor is irreducible, since it's a cyclotomic polynomial. And clearly β is a root of this. Therefore the minimum polynomial for β over $\mathbb{Q}$ is $m(x)=x^2+x+1$.

(81)

1. Let $L=\mathbb{Q}(\sqrt{2},\sqrt[3]{2})$, $E=\mathbb{Q}(\sqrt{2})$ and $F=\mathbb{Q}$. Then since the minimum polynomial for $\sqrt{2}$ over $\mathbb{Q}$ is x^2-2, we see that $[E:\mathbb{Q}]=2$. Moreover a basis for E over Q is $\{\,1,\sqrt{2}\,\}$.

   Now consider the extension $E\subset L$. The polynomial $g=x^3-2$ is monic and irreducible over E, since its roots are $\sqrt[3]{2}$, $\sqrt[3]{2}\omega$, and $\sqrt[3]{2}\omega^2$, none of which lies in E (here ω is a primitive 3rd root of unity). So g is the minimum polynomial for $\sqrt[3]{2}$ over E and hence $L=E(\sqrt[3]{2})$ satisfies $[L:E]=\deg(g(x))=3$. Moreover a basis for L over E is given by $\{\,1,\sqrt[3]{2},(\sqrt[3]{2})^2\,\}$.

   Therefore by the Tower Law, $[L:\mathbb{Q}]=6$ and a basis for L over Q is given by

   $$\{\,1,\sqrt{2},\sqrt[3]{2},\sqrt{2}\sqrt[3]{2},(\sqrt[3]{2})^2,\sqrt{2}(\sqrt[3]{2})^2\,\}\,.$$

2. Set $\alpha=\sqrt[4]{2}\in\mathbb{R}$ and let $F=\mathbb{Q}(i)$ and $L=\mathbb{Q}(\alpha,i)$, so that $L=F(\alpha)$. We need to find the minimum polynomial of α over F.

   Now α clearly satisfies the polynomial $g(x)=x^4-2$ over F. Moreover, the roots of $g(x)$ in C are $\pm\alpha$ and $\pm\alpha i$, and none of these lies in F because $\alpha\notin\mathbb{Q}$.

   Hence $g(x)$ is irreducible over F and so is the minimum polynomial for α over F.

   Therefore $[L:F]=4$ and a basis for L over F is given by

   $$\{\,1,\alpha,\alpha^2=\sqrt{2},\alpha^3\,\}\,.$$

3. Let $L=\mathbb{Q}(\xi)$ and consider the extension $L/\mathbb{Q}$.

   Since ξ is a primitive complex 7th root of unity, its minimum polynomial over Q is $\Phi_7(x)=x^6+x^5+x^4+x^3+x^2+x+1$.

   Therefore $[L:\mathbb{Q}]=6$ and a basis for L over $\mathbb{Q}$ is given by

   $$\{\,1,\xi,\xi^2,\xi^3,\xi^4,\xi^5\,\}\,.$$

4. Let ω be a primitive complex 3rd root of unity and consider the tower of fields

   $$\mathbb{Q}\subseteq\mathbb{Q}(\sqrt{3})\subseteq\mathbb{Q}(\sqrt{3},i)\subseteq\mathbb{Q}(\sqrt{3},i,\omega).$$

   The minimum polynomial of $\sqrt{3}$ over $\mathbb{Q}$ is clearly x^2-2, as this is monic, irreducible (Eisenstein) and has $\sqrt{3}$ as a root. Thus $\{1,\sqrt{3}\}$ is a basis for $\mathbb{Q}(\sqrt{3})$ over $\mathbb{Q}$. Similarly, $\{1,i\}$ is a basis for $\mathbb{Q}(\sqrt{3},i)$ over $\mathbb{Q}(\sqrt{3})$ using the minimum polynomial x^2+1.

   This gives a basis $\{1,\sqrt{3},i,\sqrt{3}i\}$ for $\mathbb{Q}(\sqrt{3},i)$ over $\mathbb{Q}$.

   An argument similar to part (i) shows that $[E:\mathbb{Q}]=4$ and that a basis for E over $\mathbb{Q}$ is given by $\{\,1,i,\sqrt{3},i\sqrt{3}\,\}$.

   The primitive complex 3rd roots of unity are given by $\alpha=(-1+i\sqrt{3})/2$ and $\beta=(-1-i\sqrt{3})/2$, so that ω must be one of these. But α and β both lie in $\mathbb{Q}(\sqrt{3},i)$! Hence $\mathbb{Q}(\sqrt{3},i)=\mathbb{Q}(\sqrt{3},i,\omega)$.

(82) Notice that $a^4 = -1$, so that a is a root of the polynomial $g(x) = x^4 + 1$. We use the result of problem 13 on sheet 1. We saw there that $g(x)$ is irreducible over $\mathbb{Q}$, so that $g(x)$ is the minimum polynomial for a over $\mathbb{Q}$. Hence by Lemma 3.4, $[\mathbb{Q}(a) : \mathbb{Q}] = deg(g(x)) = 4$.

On the other hand, if $F = \mathbb{R}$, the polynomial $g(x)$ splits into a product of two monic, irreducible quadratics. One of these has a as a root, and hence is the minimum polynomial for a over $\mathbb{R}$. Again by Lemma 3.4, $[\mathbb{R}(a) : \mathbb{R}] = 2$.

(89)

1. If we could construct 40^0 we could then bisect it to construct 20^0. But $20^0 = \pi/9$ is not constructible as $\pi/3$ cannot be trisected.
2. If 72^0 and 8^0 are constructible, then so is 80^0, which can be bisected twice to give 20^0 again. Thus 8^0 is not constructible if 72^0 is. On the otherhand, 120^0 is definitely constructible (just construct an regular triangle), hence so is $120^0 - 72^0 = 48^0$. This can then be bisected to give 24^0. Thus 24^0 is constructible from 72^0.

(91) If we could perform the required task then we could construct an x satisfying

$$\frac{x^3(15 + 7\sqrt{5})}{4} = 5.$$

Rearranging,

$$\frac{x^3(15 + 7\sqrt{5})}{4} = 5 \Rightarrow x^3 = \frac{20}{(15 + 7\sqrt{5})} = 7\sqrt{5} - 15$$

(multiplying top and bottom line by $15 - 7\sqrt{5}$). Thus

$$(x^3 + 15)^2 = 7^2 \times 5 \Rightarrow x^6 + 30x^3 - 20 = 0.$$

This last is irreducible by Eisenstein (using $p = 5$) and so is the minimum polynomial over $\mathbb{Q}$ of the side length x of the 5-fold volume dodecahedron. But this is a contradiction, since this number cannot be constructed, as the degree $[\mathbb{Q}(x), \mathbb{Q}] = 6$.

(119)

1. We know (Theorem 17) that if L is the splitting field of some polynomial then $|\text{Gal}\,(L/\mathbb{Q})| = [L, \mathbb{Q}]$, the degree of the extension $\mathbb{Q} \subset L$. Now, $\mathbb{Q}(\sqrt{2})$ is the splitting field of the polynomial $x^2 - 2$, so we have,

   $$|\text{Gal}\,(\mathbb{Q}(\sqrt{2})/\mathbb{Q})| = [\mathbb{Q}(\sqrt{2}), \mathbb{Q}].$$

   Since the right hand side is a simple extension, Theorem 15 gives that $[\mathbb{Q}(\sqrt{2}), \mathbb{Q}]$ is equal to the degree of the minimum ploynomial over $\mathbb{Q}$ of $\sqrt{2}$. This is obviously $x^2 - 2$ (its monic, irreducible, $\in \mathbb{Q}[x]$ and has $\sqrt{2}$ as a root), so

   $$|\text{Gal}\,(\mathbb{Q}(\sqrt{2})/\mathbb{Q})| = [\mathbb{Q}(\sqrt{2}), \mathbb{Q}] = \deg(x^2 - 2) = 2,$$

   as required.

   Parts (c) and (d) are exactly the same: in (c) we have the splitting field of $1 + x + x^2$ (since the element adjoined to $\mathbb{Q}$ is a 3-rd root of unity) while in (d) we have the splitting field of $x^3 - 2$ of the first lecture. In this last case though we also need to use the tower law (Theorem 16),

   $$|\text{Gal}\,(\mathbb{Q}(\sqrt[3]{2}, -\frac{1}{2}+\frac{\sqrt{3}}{2}i)/\mathbb{Q})| = [\mathbb{Q}(\sqrt[3]{2}, -\frac{1}{2}+\frac{\sqrt{3}}{2}i), \mathbb{Q}] = [\mathbb{Q}(\sqrt[3]{2}, -\frac{1}{2}+\frac{\sqrt{3}}{2}i), \mathbb{Q}(-\frac{1}{2}+\frac{\sqrt{3}}{2}i)][\mathbb{Q}(-\frac{1}{2}+\frac{\sqrt{3}}{2}i), \mathbb{Q}].$$

   Work out each of these in turn using Theorem 15.
2. Here Theorem 17 is of no use as $\mathbb{Q}(\sqrt[3]{2})$ is not the splitting field of anything! (Can you see why?) But, it is easy to do anyway. Any automorphism in $\text{Gal}\,(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ must permute the roots of any polynomial that $\sqrt[3]{2}$ is a root of, hence must permute the roots of $x^3 - 2$. But the automorphism must also send $\mathbb{Q}(\sqrt[3]{2}) \subset \mathbb{R}$ into itself, and since the other two roots of $x^3 - 2$ are complex, must in fact send $\sqrt[3]{2}$ to itself. It must then send $(\sqrt[3]{2})^2$ to itself as well, and these two, with 1 form a basis for $\mathbb{Q}(\sqrt[3]{2})$, so our automorphism must be the identity.

(120)

1. Trick question: the root of $x - 2$ is $2 \in \mathbb{Q}$, so the splitting field is just $\mathbb{Q}$. Since every element of $\text{Gal}\,(L, \mathbb{Q})$ must fix all the elements of $\mathbb{Q}$ pointwise, we get $|\text{Gal}\,(L, \mathbb{Q})| = 1$.
2. Another trick question: this is just question 3(d) as $L = \mathbb{Q}(\sqrt[3]{2}, -\frac{1}{2} + \frac{\sqrt{3}}{2}i)$.
3. The polynomial has splitting field $\mathbb{Q}(\alpha, \omega)$ where $\alpha = \sqrt[5]{2}$ and

   $$\omega = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}.$$

   By Theorem 17 and the tower law we get,

   $$|\text{Gal}\,(L, \mathbb{Q})| = [\mathbb{Q}(\alpha, \omega), \mathbb{Q}] = [\mathbb{Q}(\alpha, \omega), Q(\omega)][Q(\omega), \mathbb{Q}].$$

   Each of the terms on the right hand side is the degree of a simple extension, so we use Theorem 15: $[\mathbb{Q}(\omega), \mathbb{Q}] = 4$ since ω has minimum polynomial $1 + x + x^2 + x^3 + x^4$ (not $x^5 - 1$!). Slightly trickier is the fact that $[\mathbb{Q}(\alpha, \omega), Q(\omega)] = 5$.

In fact the minimum polynomial over $\mathbb{Q}(\omega)$ of α is indeed $x^5 - 2$ for which it is sufficient to show that this polynomial is irreducible over $\mathbb{Q}(\omega)$.

To do this, we first need that no root of $x^5 - 2$ is in $\mathbb{Q}(\omega)$. These roots are $\alpha, \alpha\omega, \ldots, \alpha\omega^4$. If $\alpha\omega^i \in \mathbb{Q}(\omega)$, then $\alpha\omega^i\omega^{-i}$ is too, ie: α is. We can probably beleive that this is not the case (see me for a more rigorous statement!). Since the polynomial has degree 5, checking the roots is not enough, it could factorsise into non-linear factors, but these must be a quadratic and a cubic. In fact, the quadratic must be of the form

$$(x - \alpha\omega^i)(x - \alpha\omega^j) = x^2 - (\alpha\omega^i + \alpha\omega^j)x + \alpha^2\omega^{i+j}.$$

Thus must be a polynomial over $\mathbb{Q}(\omega)$, so in particular, $\alpha^2\omega^{i+j} \in \mathbb{Q}(\omega)$. But then similarly, $\alpha^2 \in \mathbb{Q}(\omega) \Rightarrow \alpha^6 = \alpha \in \mathbb{Q}(\omega)$. We have already "convinced ourselves" that this isn't so.

4. By Theorem 17 again,
$$\text{Gal}\,(L/\mathbb{Q}) = [L, \mathbb{Q}],$$
where $L = \mathbb{Q}(\omega)$ with ω a primitive 5-th root of unity. The polynomial given is the minimum polynomial of ω over $\mathbb{Q}$, so we have
$$|\text{Gal}\,(L/\mathbb{Q})| = \deg(1 + x + x^2 + x^3 + x^4) = 4.$$
See also question 9.

5. By the hint, the roots of $1 + x^2 + x^4$ are the roots of $x^6 - 1$ that are not ± 1, hence are $\omega, \omega^2, \omega^4$ and ω^5, with ω a primitive 6-th root of unity. Consider $\mathbb{F} = \mathbb{Q}(\omega)$. Then clearly these roots are in $\mathbb{F}$, so that $\mathbb{F}$ contains the splitting field. On the otherhand, the splitting field contains $\mathbb{Q}$ (since any subfield of $\mathbb{C}$ does) and must contain the root ω. Thus $\mathbb{F}$ is contained in the splitting field, ie: it *is* the splitting field.

   Thus the order of the Galois group is equal to the degree $[\mathbb{F}, \mathbb{Q}] = [\mathbb{Q}(\omega), \mathbb{Q}]$ which in turn is equal to the degree of the minimum polynomial (over $\mathbb{Q}$) of ω. One may be tempted to guess $1 + x^2 + x^4$ for this, but,
$$1 + x^2 + x^4 = (x^2 + x + 1)(x^2 - x + 1),$$
so is not irreducible. Your next guess, $x^2 + x + 1$ would be correct as its roots, ω and ω^4, are $\notin \mathbb{Q}$.

   Thus the order of the Galois group is 2.

(121)

1. $\mathbb{Q}(\cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p})$ is the splitting field of the p-th cyclotomic polynomial $\Phi_p(x) = 1 + x + x^2 + \cdots + x^{p-1}$, since
$$(x - 1)\Phi_p(x) = x^p - 1.$$
   Thus the order of the group is equal to the degree of the extension which, being simple, can be deduced from Theorem 15 and the fact that Φ_p is the minimum polynomial over $\mathbb{Q}$ of the element being adjoined (where we have used assignment 3, number 1 again to get that Φ_p is irreducible). We thus get that the Galois group has order $p - 1$ as claimed.

2. This is entirely analagous to question 4(c), except $\alpha = \sqrt[p]{2}$ and ω is a primitive p-th root of 1.

(128)

1. The Galois group of L over k is the group of all automorphisms of the field L that leave the subfield k fixed pointwise.

2. Observe first that a basis for $\mathbb{Q}(\alpha, i)$ over $\mathbb{Q}$ is given by
$$\{1, \alpha, \alpha^2, \alpha^3, i, \alpha i, \alpha^2 i, \alpha^3 i\}$$
   and the effect of the eight automorphisms on α and i is given by

	1	σ	σ^2	σ^3		σ	σ^2	σ^3
α	α	αi	$-\alpha$	$-\alpha i$	α	αi	$-\alpha$	$-\alpha i$
i	i	i	i	i	$-i$	$-i$	$-i$	$-i$

   Hence the automorphism are distinct and by their effect on α and i, we see that $\sigma = \sigma^3$.

3. $H_1 = \{1, \sigma^2, \sigma, \sigma^3\}$, $\mathtt{H_2} = \{1, \sigma^2\}$ and $H_3 = \{1, \sigma\}$. By the Galois correspondence, we have $[\mathbb{Q}(\alpha, i), \mathbb{Q}]$ is equal to the index of $\{1\}$ in G, ie the order of G, which is 8. The tower law gives $[F_1, \mathbb{Q}] = [F_1, \mathbb{Q}(i\sqrt{2})][\mathbb{Q}(i\sqrt{2}), \mathbb{Q}]$ with $[F_1, \mathbb{Q}] = 4$ and $[\mathbb{Q}(i\sqrt{2}), \mathbb{Q}] = 2$ since the corresponding Galois groups have these indices in G. Hence $[F_1, \mathbb{Q}(i\sqrt{2})] = 2$.

   To describe the fields we use the fact that any element x of $\mathbb{Q}(\alpha, i)$ can be written uniquely in the form
$$x = a_0 + a_1\alpha + a_2\alpha^2 + a_3\alpha^3 + a_4 i + a_5\alpha i + a_6\alpha^2 i + a_7\alpha^3 i,$$
   for some $a_i \in \mathbb{Q}$. The Galois group of $\mathbb{Q}(\alpha, i)$ over F_1 is $\{1, \sigma^2\}$ from the lattice diagram and the Galois correspondence. Hence every element of F_1 is fixed by σ^2, where
$$\sigma^2(x) = a_0 - a_1\alpha + a_2\alpha^2 - a_3\alpha^3 + a_4 i - a_5\alpha i + a_6\alpha^2 i - a_7\alpha^3 i,$$
   and equating this with the previous expression we see that such an x must satisfy $a_{2i} = a_{2i}$ and $a_{2i+1} = -a_{2i+1}$, so that $a_1 = a_3 = a_5 = a_7 = 0$ while a_0, a_2, a_4 and a_6 are arbitrary. Hence we have
$$x = a_0 + a_2\alpha^2 + a_4 i + a_6\alpha^2 i,$$

and clearly $F_1 \subset \mathbb{Q}(\alpha^2, i)$. On the otherhand, $\alpha^2, i \in F_1$ and so $\mathbb{Q}(\alpha^2, i) \subset F_1$. Thus $F_1 = \mathbb{Q}(\alpha^2, i)$.
To get the other one, notice that does not fix $i\sqrt{2}$, hence the Galois group of $\mathbb{Q}(\alpha, i)$ over $\mathbb{Q}(i\sqrt{2})$ must be H_1 so that the Galois group of $\mathbb{Q}(\alpha, i)$ over F_2 is $\{1, \sigma^2, , \sigma^2\}$. Running through the calculation above gives

$$\sigma^2(x) = a_0 - a_1\alpha + a_2\alpha^2 - a_3\alpha^3 + a_4 i - a_5\alpha i + a_6\alpha^2 i - a_7\alpha^3 i,$$

and

$$(\mathtt{x}) = \mathtt{a_0} + \mathtt{a_1}\alpha + \mathtt{a_2}\alpha^2 + \mathtt{a_3}\alpha^3 - \mathtt{a_4 i} - \mathtt{a_5}\alpha\mathtt{i} - \mathtt{a_6}\alpha^2\mathtt{i} - \mathtt{a_7}\alpha^3\mathtt{i},$$

giving, $a_1 = a_3 = a_5 = a_7 = a_4 = a_6 = 0$, hence an $x \in F_2$ is of the form

$$x = a_0 + a_2\alpha^2,$$

and $F_2 = \mathbb{Q}(\sqrt{2})$ for the same reasons as before.

(129)

1. The Galois group of L over F is the group of all automorphisms of the field L that leave the subfield F fixed pointwise.
2. The polynomial in question has roots $\omega, \omega^2, \omega^3, \omega^4, \omega^5$ and ω^6, and these are clearly all in $\mathbb{Q}(\omega)$. On the otherhand, if F is *any* field containing the roots of the polynomial, then F certainly contains ω (as this is one of them) and it also contains $\mathbb{Q}$ (assignment question done in the problems class). But $\omega \in F$ and $\mathbb{Q} \subset F$ means $\mathbb{Q}(\omega) \subset F$, so that it is indeed the smallest field containing the roots. The order of the Galois group now follows immediately since,
$$|\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})| = [\mathbb{Q}(\omega), \mathbb{Q}] = \deg(1 + x + x^2 + x^3 + x^4 + x^5 + x^6) = 6,$$
since the polynomial is the minimum polynomial for ω over $\mathbb{Q}$.
3. A basis for $\mathbb{Q}(\omega)$ is given by
$$\{1, \omega, \omega^2, \omega^3, \omega^4, \omega^5\},$$
and so any automorphism is determined by is effect on these basis vectors. In fact, any automorphism is determined by its effect on ω alone. On the otherhand, any automorphism must permute the roots of any polynomial that ω is a root of, eg: $1 + x + x^2 + x^3 + x^4 + x^5 + x^6$ with roots $\omega, \omega^2, \omega^3, \omega^4, \omega^5$ and ω^6. Combining all this with the fact that there are exactly 6 automorphisms means that they are precisely the maps that send ω to one of these 6 roots. If σ sends ω to ω^3 as stated, then $\sigma^2(\omega) = (\omega^3)^3 = \omega^9 = \omega^2$, and $\sigma^3(\omega) = \omega^{27} = \omega^6$. Thus by Lagrange, σ has order 6 in the Galois group, which must then be cyclic as claimed.
4. Clearly any subgroup containing σ contains everything. Similarly, since the powers of σ^5 yield all the elements of the group, any subgroup containing σ^5 is the whole group. Thus, for a proper subgroup we must not include σ or σ^5. If we don't include σ^2 we get the subgroup $\{1\}$ or $\{1, \sigma^3\}$, whereas if we do, we get the subgroup $\{1, \sigma^2, \sigma^4\}$.
By the Galois correspondence, we get the lattice of intermediate subfields as claimed, with F_1 the fixed field of a subgroup of index 2, hence of $\{1, \sigma^2, \sigma^4\}$, and F_2 the fixed field of $\{1, \sigma^3\}$.
Now any element of $\mathbb{Q}(\omega)$ can be written as
$$x = a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3 + a_4\omega^4 + a_5\omega^5$$
with the $a_i \in \mathbb{Q}$. We require x such that $\sigma^3(x) = x$, where
$$\sigma^3(a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3 + a_4\omega^4 + a_5\omega^5) = a_0 + a_1(-1 - \omega - \cdots - \omega^5) + a_2\omega^5 + a_3\omega^4 + a_4\omega^3 + a_5\omega^2$$
and we have by equating coefficients that $a_0 - a_1 = a_0, a_1 = -a_1, a_5 - a_1 = a_2, a_4 - a_1 = a_3, a_3 - a_1 = a_4$ and $a_2 - a_1 = a_5$. Thus x must have the form
$$x = a_0 + a_2(\omega^2 + \omega^5) + a_3(\omega^3 + \omega^4).$$
Hence the fixed field is $\subset \mathbb{Q}(\omega^2 + \omega^5, \omega^3 + \omega^4)$. On the otherhand, σ^3 fixes both these elements and thus $\mathbb{Q}(\omega^2 + \omega^5, \omega^3 + \omega^4) \subset$ the fixed field, giving $F_2 = \mathbb{Q}(\omega^2 + \omega^5, \omega^3 + \omega^4)$.

(131)

1. The polynomial in question has roots $\omega, \omega^2, \omega^3$ and ω^4, and these are clearly all in $\mathbb{Q}(\omega)$. On the otherhand, if F is *any* field containing the roots of the polynomial, then F certainly contains ω (as this is one of them) and it also contains $\mathbb{Q}$ (see question 3, assignment1). But $\omega \in F$ and $\mathbb{Q} \subset F$ means $\mathbb{Q}(\omega) \subset F$, so that it is indeed the smallest field containing the roots.
2. This follows immediately since,
$$|\text{Gal}\,(\mathbb{Q}(\omega)/\mathbb{Q})| = [\mathbb{Q}(\omega), \mathbb{Q}] = \deg(1 + x + x^2 + x^3 + x^4) = 4,$$
since the polynomial is the minimum polynomial for ω over $\mathbb{Q}$.
3. A basis for $\mathbb{Q}(\omega)$ is given by
$$\{1, \omega, \omega^2, \omega^3\},$$
and so any automorphism is determined by is effect on these basis vectors. In fact, any automorphism is determined by its effect on ω alone. On the otherhand, any automorphism must permute the roots of any polynomial that ω is a root of, eg: $1 + x + x^2 + x^3 + x^4$ with roots $\omega, \omega^2, \omega^3$ and ω^4. Combining all this with the fact that there are exactly 4 automorphisms means that they are precisely the maps that send ω to ω or ω^2 or ω^3 or ω^4. If σ sends ω to ω^2 as stated, then σ^2 sends ω to ω^4, σ^3 sends ω to ω^3, and σ^4 send ω to 1.
Thus, the Galois group has the elements as stated.

4. Clearly any subgroup containing σ contains everything. Similarly, since the powers of σ^3 yield all the elements of the group, any subgroup containing σ^3 is the whole group. Thus, for a proper subgroup we must not include σ or σ^3. If we don't include σ^2 we get the subgroup $\{1\}$, whereas if we do, we get the subgroup $\{1, \sigma^2\}$. Thus the lattice has only one subgroup apart from the two obvious ones.

5. This follows imediately by the Galois correspondence and since the subgroup lattice has only one subgroup apart from $\{1\}$ and G. Now any element of $\mathbb{Q}(\omega)$ can be written as

$$x = a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3,$$

with the $a_i \in \mathbb{Q}$, and the intermediate field we are after is the fixed field of the subgroup $\{1, \sigma^2\}$. That is, we have $\sigma^2(x) = x$ for all $x \in \mathbb{Q}(\omega)$. On the otherhand,

$$\sigma^2(a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3) = a_0 + a_1\omega^4 + a_2\omega^3 + a_3\omega^2 = a_0 + a_1(-1 - \omega - \omega^2 - \omega^3) + a_2\omega^3 + a_3\omega^2$$

so that,

$$\sigma^2(x) = (a_0 - a_1) - a_1\omega + (a_3 - a_1)\omega^2 + (a_2 - a_1)\omega^3,$$

and we have by equating coefficients that $a_0 = a_0 - a_1, a_1 = a_1, a_2 = a_3 - a_1$ and $a_3 = a_2 - a_1$. Thus x must have the form

$$x = a + b\omega^2 + b\omega^3 = a + b(\omega^2 + \omega^3) \in \mathbb{Q}(\omega^2 + \omega^3).$$

Hence the fixed field is $\subset \mathbb{Q}(\omega^2 + \omega^3)$. On the otherhand, $\sigma^2(\omega^2 + \omega^3) = \omega^2 + \omega^3$, and so σ^2 must fix $\mathbb{Q}(\omega^2 + \omega^3)$ pointwise. Thus $\mathbb{Q}(\omega^2 + \omega^3) \subset$ the fixed field, and $\mathbb{Q}(\omega^2 + \omega^3)$ *is* the field we seek.