Lecture notes of Algebra. Week 5

16. Irreducible polynomials, and unique factorisation

The four Arithmetical Lemmas for integers remain true for polynomials, after making the obvious changes in terminology, and can be proved in the same way using Bézout's Lemma. In particular, the most important of them, Arithmetical Lemma B, is as follows: *if the polynomials $f(x)$ and $g(x)$ are coprime, and $f(x)$ divides the product $g(x) \cdot h(x)$, then $f(x)$ divides $h(x)$.*

Prime polynomials are usually rather called *irreducible* polynomials. As was the case for the integers, one has to exclude the zero polynomial and the invertible polynomials (which are the analogues of ± 1 in the integers) from the definition of irreducible. Because the invertible polynomials are the nonzero constants (the polynomials of degree zero), the definition of irreducible will only apply to non-constant polynomials, that is, to polynomials of positive degree.

Definition 26. A non-constant polynomial $f(x) \in F[x]$ is *reducible* in $F[x]$ if it can be written as $f(x) = g(x)\,h(x)$, where $g(x)$ and $h(x)$ are polynomials in $F[x]$ of positive degree (or, equivalently, $\deg(g(x))$ and $\deg(h(x))$ are smaller than $\deg(f(x))$; or, equivalently again, where $0 < \deg(g(x)) < \deg(f(x))$); it is *irreducible* in $F[x]$ if it is not reducible.

Because $\deg\big(g(x)\,h(x)\big) = \deg\big(g(x)\big)$, any polynomial of degree 1, hence of the form $ax + b$ with $a \neq 0$, is always irreducible, as 1 cannot be written as a sum of two positive integers.

Note that the notions of reducible and irreducible depend on the field in which we view the coefficients of our polynomial: $x^2 + 1$ is irreducible as a polynomial in $\mathbb{R}[x]$, but not as a polynomial in $\mathbb{C}[x]$, because $x^2 + 1 = (x - i)(x + i)$. To stress which field F is being used, one usually specifies *irreducible in $F[x]$*, or also *irreducible over F*. (Hence $x^2 + 1$ is irreducible over $\mathbb{R}$, but reducible over $\mathbb{C}$.)

Theorem 14 on unique factorisation in the integers has an analogue for polynomials.

Theorem 27 (Unique Factorisation Theorem for polynomials). *Every polynomial of positive degree (which is the same as saying* non-constant*) over a field F factorises into a product of irreducible polynomials (irreducible over the same field F).*

Also, the factorisation is essentially *unique, namely, unique up to permuting the factors, but also to multiplying each irreducible factor by some nonzero constant (that is, by some invertible polynomial).*

For example,

$$2x^2+10x+12 = 2(x+2)(x+3) = (2x+4)(x+3) = (x+2)(2x+6)$$
$$= (3x+6)\Big(\frac{2}{3}x+2\Big), \qquad \text{and so on.}$$

We call the factorisation into a product of irreducible polynomials the *complete factorisation* of $f(x)$ over F (or in $F[x]$). Note that, once again, with polynomials it is essential to state over which field we are working, because the answer may be different over different fields.

EXAMPLE. The polynomial $x^4-3x^2+2 \in \mathbb{Q}[x]$ can be written in three essentially different ways

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

as the product of two polynomials of degree 2. However, this does not contradict the above theorem on unique factorisation because none of those quadratic factors is irreducible over $\mathbb{Q}$, in fact $x^4-3x^2+2 = (x-1)(x+1)(x-2)(x+2)$, and the above quadratic factors are obtained by pairing and multiplying together the irreducible factors in different ways.

EXAMPLE. The polynomial x^3-2 is irreducible in $\mathbb{Q}[x]$. (This will be justified in later sections.) In $\mathbb{R}[x]$ it factorises as

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

This is a complete factorisation in $\mathbb{R}[x]$ because the quadratic factor is irreducible. In fact, its discriminant is $\sqrt[3]{2}^2-4\cdot\sqrt[3]{2}^2 = -3\sqrt[3]{2}^2 < 0$, and so that quadratic polynomial has no real roots. However, in $\mathbb{C}[x]$ the polynomial x^3-2 factorises as

$$x^3-2 = (x-\sqrt[3]{2})(x-\omega\sqrt[3]{2})(x-\bar{\omega}\sqrt[3]{2}),$$

where $\omega = (-1\pm i\sqrt{3})/2$.

17. Quadratic polynomials

You should know well from school how to find the roots of a quadratic polynomial ax^2+bx+c (hence with $a\neq 0$, otherwise it would not be quadratic), which means the same as finding the solutions of the corresponding equation $ax^2+bx+c=0$. In fact, the trick of *completing the square* ax^2+bx at the left-hand side brings the equation to the equivalent form [6]

$$a\left(x+\frac{b}{2a}\right)^2 = \frac{b^2-4ac}{4a}.$$

[6]To be precise, this works over any field $\mathbb{F}$ where $2\neq 0$. The meaning of this unfamiliar condition will be clarified in a later algebra course, but note that it is not satisfied when $F=\mathbb{F}_2=\mathbb{Z}/2\mathbb{Z}$, the field of 2 elements, where $[2]=[1+1]$ is the same as $[0]$.

The numerator of the fraction at the right-hand side is the *discriminant* of the quadratic equation (or polynomial), and the existence of the solutions depends on whether it is a square in the field F under consideration (which is the same as being nonnegative if $F = \mathbb{R}$, but may be a different condition otherwise). If $b^2 - 4ac$ is not a square of an element of F, then no $x \in F$ can make the above equality true, and so the polynomial has no root in F. If $b^2 - 4ac$ is a square of an element of F, which means that it has a square root in F, denote it by $\sqrt{b^2 - 4ac}$, then the equation becomes

$$\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\sqrt{b^2 - 4ac}}{2a}\right)^2 = 0$$

and, in turn,

$$\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right) = 0.$$

In conclusion, one may distinguish two cases:

- if $b^2 - 4ac$ is not the square of an element of F, and so does not have a square root in F, then the polynomial has no root in F;
- if $b^2 - 4ac$ has a square root in F, then the polynomial has roots given by the familiar formula
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a};$$
of course the two roots coincide when $b^2 - 4ac = 0$ (which is sometimes stated separately as a third case). [7]

These cases can be expressed in terms of reducibility (over a field F containing all coefficients) of the quadratic polynomial $ax^2 + bx + c$ (with $a \neq 0$):

- it is irreducible if $b^2 - 4ac$ does not have a square root in F;
- it is reducible if $b^2 - 4ac$ has a square root in F; in fact in that case we have $ax^2 + bx + c = a(x - \alpha_1)(x - \alpha_2)$, where $\alpha_1, \alpha_2 = (-b \pm \sqrt{b^2 - 4ac})/(2a)$ are its roots (in any order); of course $\alpha_1 = \alpha_2$ when $b^2 - 4ac = 0$.

Example. A quadratic polynomial $ax^2+bx+c \in \mathbb{R}[x]$ (hence with $a \neq 0$) is irreducible exactly when its discriminant b^2-4ac has no square roots in $\mathbb{R}$, hence exactly when b^2-4ac is negative.

Example. Because any complex number has square roots in $\mathbb{C}$, quadratic polynomials in $\mathbb{C}[x]$ are always reducible (and hence factorise completely into products of two polynomials of degree one).

[7] Note that whenever it is convenient to collect a factor 2 from the coefficient b, for example when b is an even integer, or, say, $b = 6\sqrt{5} = 2 \cdot 3\sqrt{5}$, etc., it may be easier to use the slightly simpler formula $x = \left(-B \pm \sqrt{B^2 - ac}\right)/a$ for the roots of the polynomial $ax^2 + 2Bx + c$ (that is, where $b = 2B$).

EXAMPLE. The polynomial x^2-2 is irreducible over $\mathbb{Q}$, but reducible over $\mathbb{R}$, because $x^2-2=(x-\sqrt{2})(x+\sqrt{2})$, and $\sqrt{2}\notin\mathbb{Q}$, which means that $\sqrt{2}$ is irrational (see below, or as an application of the Rational Root Test later).

(OPTIONAL) PROOF THAT $\sqrt{2}$ IS IRRATIONAL. We need to show that $\sqrt{2}$ cannot be equal to any fraction a/b of integers. Suppose it is. (In these cases one says *suppose, for a contradiction,* and then try to show that this assumption does lead to a contradiction, meaning, proving that some fact would have to be both true and false.)

In that case the fraction can certainly be reduced to simplest terms, and so we may assume that $\sqrt{2}=a/b$, with $a,b\in\mathbb{Z}$, clearly with $b\neq 0$, and the further condition $(a,b)=1$. Multiplying by b and squaring we find $2b^2=a^2$. Hence 2 divides $a^2=a\cdot a$, but because 2 is prime it follows that 2 divides a, and hence $a=2c$ for some integer c. Substituting into our equation we find $2b^2=4c^2$, whence $b^2=2c^2$. Hence 2 divides b^2, and because 2 is prime it divides b. So we have found that 2 divides both a and b, and hence 2 divides $(a,b)=1$. But this is certainly false, and we have found the desired contradiction. (We have concluded that 2 divides 1, but at the same time we know that 2 does not divide 1.)

The only way to resolve the contradiction is to admit that we made a false assumption at the beginning, namely, in assuming that $\sqrt{2}$ is equal to some fraction a/b of integers. Hence this is not possible, which means that $\sqrt{2}$ is irrational. □

18. The maximum number of roots of a polynomial

How many roots does a polynomial have? Well, the zero polynomial has all the roots we want (any number is a root), so we look at non-zero polynomials. A polynomial of degree 1 has always exactly one root, namely, $-b/a$ if the polynomial is $ax+b$ (with $a\neq 0$ otherwise it would not have degree 1). For a polynomial of degree 2 the answer may depend on the field where we are viewing the coefficients: it may have two roots, or one root (which we may think of a double root, but we count only once if we meant to ask about how many *distinct* roots), or none. In any case, at most two. Here is a more general result, which we can prove using the Factor Theorem.

THEOREM 28. *A polynomial of degree $n\geq 0$ over a field F has at most n roots in F.*

PROOF. Let f be a polynomial of degree $n\geq 0$. If f has no roots at all in F, we are done, as $0\leq n$ (so the statement is correct in this case). If f has (at least) a root α_1, then according to the Factor Theorem (see Lemma 22) we have

$$f(x)=(x-\alpha_1)\cdot f_2(x)$$

for some polynomial $f_2(x)$. Now it may happen that $f_2(x)$ has no roots (and then $f(x)$ has exactly one root, so the statement is correct because $1\leq n$). But if $f_2(x)$ does have

a root α_2 (possibly equal to α_1), then

$$f_2(x) = (x - \alpha_2) \cdot f_3(x),$$

whence

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdot f_3(x).$$

Continuing in this way, sooner or later we arrive at

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_m) \cdot f_{m+1}(x),$$

where $f_{m+1}(x)$ has no roots in F (possibly because it is a constant). In fact, this must occur for some $m \leq \deg(f) = n$, because taking degrees in the above equality we find $\deg(f) = m + \deg(f_{m+1}) \geq m$.

Now let β be a root of $f(x)$. To complete the proof it will be enough to show that β equals one among $\alpha_1, \ldots, \alpha_m$, of which there are at most m distinct ones (possibly fewer!), and consequently at most n as we have just shown $m \leq n$. In fact, if $f(\beta) = 0$ then

$$0 = f(\beta) = (\beta - \alpha_1) \cdots (\beta - \alpha_m) \cdot f_{m+1}(\beta).$$

We have $f_{m+1}(\beta) \neq 0$ because $f_{m+1}(x)$ has no roots in F, and so at least one other factor in the above product must vanish, say $\beta - \alpha_j$, and hence $\beta = \alpha_j$, as desired. □

In particular, any nonzero polynomial has finitely many roots. This behaviour of polynomial functions is different from other familiar functions, for example the function $f(x) = \sin(x)$ has infinitely many real zeroes, namely,

$$\sin(x) = 0 \qquad \Leftrightarrow \qquad x = 2\pi k \quad \text{for} \quad k \in \mathbb{Z}.$$

COROLLARY 29. *A polynomial $f(x)$ of degree n is uniquely determined by the values it takes on $n + 1$ distinct elements of F.*

PROOF. We are assuming that for $n + 1$ distinct numbers $b_1, \ldots, b_{n+1}$ we know the values

$$f(b_1) = c_1, \quad f(b_2) = c_2, \quad \ldots \quad f(b_{n+1}) = c_{n+1}.$$

Suppose $g(x)$ is any polynomial of degree n which satisfies

$$g(b_1) = c_1, \quad g(b_2) = c_2, \quad \ldots \quad g(b_{n+1}) = c_{n+1}.$$

Then the difference $h(x) = f(x) - g(x)$ is either zero or a nonzero polynomial of degree *at most* n, and it satisfies

$$h(b_1) = 0, \quad h(b_2) = 0, \quad \ldots \quad h(b_{n+1}) = 0.$$

Hence $h(x)$ has at least $n + 1$ roots, while a nonzero polynomial of degree at most n has at most n roots. Consequently, $h(x)$ can only be the zero polynomial, and hence $g(x) = f(x)$. □

An important consequence of the above corollary is that two different polynomials $f(x), g(x) \in \mathbb{R}[x]$ give rise to different functions $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$. Consequently, the apparently simpler alternate definition of a polynomial as a function $f\colon \mathbb{R} \to \mathbb{R}$ of a particular shape (that is, which can be written as $f(x) = a_n x^n + \cdots + a_1 x + a_0$) is actually equivalent to the one we are using (so we can identify polynomials with the functions they gives rise to), but only *if we work over an infinite field,* such as $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$.

Example. Consider the polynomials $f(x) = \bar{1}x = x$ and $g(x) = \bar{1}x^2 = x^2$, where the coefficients belong to the field of two elements $\mathbb{F}_2$. According to our definition of polynomials they are different polynomials (because their coefficients are different; and actually even their degrees) but give rise to the same function $\mathbb{F}_2 \to \mathbb{F}_2$, because

$$f(\bar{0}) = \bar{0} = g(\bar{0}), \qquad \text{and} \qquad f(\bar{1}) = \bar{1} = g(\bar{1}).$$

19. Polynomial interpolation

Given real numbers $\alpha_1, \alpha_2, \beta_1, \beta_2$, with $\alpha_1 \neq \alpha_2$, there is a unique polynomial $f(x)$ of degree at most one (hence a linear polynomial or a constant polynomial), such that

$$f(\alpha_1) = \beta_1, \quad \text{and} \quad f(\alpha_2) = \beta_2.$$

In fact, because $f(x) = ax + b$ for some $a, b \in \mathbb{R}$, the required conditions amount to

$$\begin{cases} a \cdot \alpha_1 + b = \beta_1 \\ a \cdot \alpha_2 + b = \beta_2 \end{cases}$$

Solving this system we would find

$$a = \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1}, \quad \text{and then} \quad b = \beta_1 - \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1},$$

and hence there is a unique solution, which can be written as

$$f(x) = \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1} \cdot (x - \alpha_1) + \beta_1.$$

The following result generalises this fact to an arbitrary number of values. We state it for complex numbers rather than real numbers, but we could as well take for $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ elements of any field (the same field for all of them).

Theorem 30 (Interpolation theorem). *Given n distinct complex numbers $\alpha_1, \ldots, \alpha_n$, and n arbitrary (not necessarily distinct) complex numbers $\beta_1, \ldots, \beta_n$, there is a unique polynomial $f(x)$ of degree less than n such that*

$$f(\alpha_1) = \beta_1, \quad \ldots, \quad f(\alpha_n) = \beta_n.$$

(Optional) Proof. We have already proved the uniqueness of $f(x)$ in Corollary 29. There are several ways to prove the existence of $f(x)$, both direct and indirect (and hence not explicit). We present an explicit proof, based on *Lagrange interpolation*.

The polynomial

$$(x-\alpha_2)(x-\alpha_3)\cdots(x-\alpha_n) = \frac{(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)}{x-\alpha_1}$$

has $\alpha_2, \alpha_3, \ldots, \alpha_n$ as roots, and so it takes the value 0 on $\alpha_2, \alpha_3, \ldots, \alpha_n$. On α_1 it takes the value $(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)\cdots(\alpha_1-\alpha_n)$. Consequently, the polynomial

$$p_1(x) = \frac{(x-\alpha_2)(x-\alpha_3)\cdots(x-\alpha_n)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)\cdots(\alpha_1-\alpha_n)}$$

satisfies

$$p_1(\alpha_1) = 1, \quad p_1(\alpha_2) = 0, \quad \ldots, \quad p_1(\alpha_n) = 0.$$

Similarly, we can construct a polynomial

$$p_1(x) = \frac{(x-\alpha_1) \quad (x-\alpha_3)(x-\alpha_4)\cdots(x-\alpha_n)}{(\alpha_2-\alpha_1) \quad (\alpha_2-\alpha_3)(\alpha_2-\alpha_4)\cdots(\alpha_2-\alpha_n)}$$

(where $x-\alpha_2$ is omitted from the numerator, and $\alpha_2-\alpha_2$ is correspondingly omitted from the denominator), which takes the value 1 on α_2 and vanishes (that means, it takes the value zero) on $\alpha_1, \alpha_3, \alpha_4, \ldots, \alpha_n$. After finding $p_3(x), \ldots, p_n(x)$ with similar properties, we see that the polynomial

$$f(x) = \beta_1 \cdot p_1(x) + \cdots + \beta_1 \cdot p_1(x)$$

satisfies

$$f(\alpha_1) = \beta_1, \quad \ldots, \quad f(\alpha_n) = \beta_n,$$

as required, and has degree at most $n-1$ because each of $p_1(x), \ldots, p_n(x)$ has degree exactly $n-1$. □

Each polynomial $p_k(x)$ in the above proof is obtained as

$$p_k(x) = g_k(x)/g_k(\alpha_k), \quad \text{where} \quad g_k(x) = \prod_{j\neq k}(x-\alpha_j).$$

Note that it may be quicker to compute $(x-\alpha_1)\cdots(x-\alpha_n)$ first, a polynomial of degree n, and then obtain each $g_k(x)$ by dividing that by $x-\alpha_k$, which can be efficiently done using Ruffini's rule. [8]

[8] A rough way to see why is noting that computing $(x-\alpha_1)\cdots(x-\alpha_n)$ requires $n-1$ polynomial multiplications, and on the way to the conclusion we would have already computed one of the $g_k(x)$, say $g_n(x) = (x-\alpha_1)\cdots(x-\alpha_{n-1})$. Then it remains to obtain the remaining $g_k(x)$ by applying Ruffini's rule $n-1$ times. This total of $2n-2$ multiplications/divisions should be contrasted with the $n(n-1)$ multiplications required to compute each $g_k(x)$ separately. (More precise computational estimates would take into account that polynomial multiplications may take different times depending on the size of the factors, but then we would still observe a similar difference in speed of the two methods.)

EXAMPLE. Find the unique polynomial $f(x)$ of degree at most three, such that

$$f(-2) = -5, \qquad f(-1) = 3, \qquad f(1) = 1, \qquad f(2) = 3.$$

The polynomial

$$p_1(x) = \frac{(x+1)(x-1)(x-2)}{(-2+1)(-2-1)(-2-2)} = -\frac{1}{12}(x^3 - 2x^2 - x + 2)$$

has -1, 1 and 2 as roots and satisfies $p_1(-2) = 1$. The polynomial

$$p_2(x) = \frac{(x+2)(x-1)(x-2)}{(-1+2)(-1-1)(-1-2)} = \frac{1}{6}(x^3 - x^2 - 4x + 4)$$

has -2, 1 and 2 as roots and satisfies $p_2(-1) = 1$. The polynomial

$$p_3(x) = \frac{(x+2)(x+1)(x-2)}{(1+2)(1+1)(1-2)} = -\frac{1}{6}(x^3 + x^2 - 4x - 4)$$

has -2, -1 and 2 as roots and satisfies $p_3(1) = 1$. The polynomial

$$p_4(x) = \frac{(x+2)(x+1)(x-1)}{(2+2)(2+1)(2-1)} = \frac{1}{12}(x^3 + 2x^2 - x - 2)$$

has -2, -1 and 1 as roots and satisfies $p_4(2) = 1$.

Note that because of the particular symmetry of our problem, $p_3(x)$ and $p_4(x)$ could have also been obtained as $p_3(x) = p_2(-x)$ and $p_4(x) = p_1(-x)$. Of course it will not be so in general.

We conclude that

$$\begin{aligned} f(x) &= -5 \cdot p_1(x) + 3 \cdot p_2(x) + p_3(x) + 3 \cdot p_4(x) \\ &= \frac{5}{12}x^3 - \frac{5}{6}x^2 - \frac{5}{12}x + \frac{5}{6} \\ &\quad + \frac{1}{2}x^3 - \frac{1}{2}x^2 - 2x + 2 \\ &\quad - \frac{1}{6}x^3 - \frac{1}{6}x^2 + \frac{2}{3}x + \frac{2}{3} \\ &\quad + \frac{1}{4}x^3 + \frac{1}{2}x^2 - \frac{1}{4}x - \frac{1}{2} \\ &= x^3 - x^2 - 2x + 3. \end{aligned}$$

EXAMPLE. Find the unique polynomial $g(x)$ of degree at most three, such that

$$g(-2) = 1, \qquad g(-1) = -1, \qquad g(1) = -1, \qquad g(2) = 1.$$

We can reuse the calculations of the previous example, and find

$$f(x) = p_1(x) - p_2(x) - p_3(x) + p_4(x) = \frac{2}{3}x^2 - \frac{5}{3}.$$

Hence this time the required polynomial has actually degree two. According to the interpolation theorem, Theorem 30, it is the unique polynomial of degree *less than four,* which is the same as *at most three,* which satisfies the given conditions. Of course, had

we known that it has degree at most two, it would have been uniquely determined by any three of those four conditions, again according to the interpolation theorem.

20. Irreducibility and roots for quadratic and cubic polynomials

Consider a polynomial $f(x)$, of positive degree, with coefficients in a field F. Recall that, according to the Factor Theorem an element α of F is a root of $f(x)$ (that is, $f(\alpha) = 0$) exactly when $x - \alpha$ is a factor of $f(x)$ (that is, it divides $f(x)$). Hence each time we find a root α of $f(x)$ we have achieved a partial factorisation of $f(x)$ as $f(x) = (x - \alpha)\, g(x)$, for some polynomial $g(x)$ (again with coefficients in F).

In particular, if a polynomial $f(x)$ of degree larger than one has a root in F, then it is reducible in $F[x]$. For polynomials of degree two or three this implication can be inverted.

PROPOSITION 31. *A quadratic or cubic polynomial (that is, of degree two or three) over a field F is irreducible over F exactly when it does not have any root in F.*

PROOF. The statement is equivalent to the following: a quadratic or cubic polynomial over a field F is reducible over F exactly when it has some root in F (meaning *at least one root in F*). Now we prove this statement.

If our polynomial, say $f(x)$, has a root α in F, then according to the Factor Theorem we have $f(x) = (x-\alpha)\, g(x)$ for some polynomial $g(x) \in F[x]$, and hence $f(x)$ is reducible over F.

Conversely, if $f(x)$ is reducible over F, then $f(x) = g(x)\, h(x)$ for some polynomials of positive degree $g(x), h(x) \in F[x]$. Because $\deg\big(f(x)\big) = \deg\big(g(x)\big) + \deg\big(h(x)\big)$, and $\deg\big(f(x)\big)$ equals two or three, then at least one of the factors, say $g(x)$, must have degree one, and hence be of the form $g(x) = ax + b$, with $a, b \in F$ and $a \neq 0$. Then $-b/a$ is a root of $g(x)$, and hence of $f(x)$. In conclusion, $f(g)$ has at least one root in F. □

Of course a polynomial of degree one, hence of the form $ax + b$ with $a, b \in F$ and $a \neq 0$, is irreducible but has always a root in F, namely $-b/a$. This criterion for being irreducible (or reducible) does not work for polynomials of degree four or higher. In fact, it is possible that a polynomial (of degree at least four) has a proper factorisation over F even if it does not have any root in F, as the following examples show.

EXAMPLE. The polynomial $x^4 + 5x^2 + 4$ factorises over $\mathbb{R}$ as $(x^2 + 1)(x^2 + 4)$, but it has no real roots. In fact, its complex roots are $\pm i$ and $\pm 2i$, but none of them is real. Hence $x^4 + 5x^2 + 4$ is reducible over $\mathbb{R}$ (and, in fact, it is the product of the two irreducible polynomials $x^2 + 1$ and $x^2 + 4$), despite having no roots in $\mathbb{R}$.

EXAMPLE. The polynomial $x^4 + 1$ has no roots in $\mathbb{R}$, but is not irreducible in $\mathbb{R}[x]$. More generally, we have

$$\begin{aligned} x^4 + a^4 &= (x^4 + 2a^2x^2 + a^4) - 2a^2x^2 \\ &= (x^2 + a^2)^2 - (\sqrt{2}ax)^2 \\ &= [(x^2 + a^2) - \sqrt{2}ax][(x^2 + a^2) + \sqrt{2}ax] \\ &= (x^2 - \sqrt{2}ax + a^2)(x^2 + \sqrt{2}ax + a^2). \end{aligned}$$

Hence if $a \neq 0$ is a real number, then $x^4 + a^4$ has no real roots, but is reducible in $\mathbb{R}[x]$, and the factorisation given here is its complete factorisation in $\mathbb{R}[x]$. For example, $x^4 + 4 = (x^2 - 2x + 2)(x^2 + 2x + 2)$, a factorisation which is even in $\mathbb{Q}[x]$.

21. The Fundamental Theorem of Algebra

We know how to solve equations of degree 1 and 2. In case of quadratic equations the formula involves taking a square root (of the discriminant). More complicated formulas exist for finding the roots of polynomials of degree three and four (which means solving the corresponding equations). Those formulas (due to Del Ferro/Tartaglia/Cardano for cubics, and Ferrari/Cardano for quartics, all in 16th century) involve the usual algebraic operations together with taking square and cube roots.

However, the Norwegian mathematician Abel proved in 1824 (building on previous partial work of Ruffini) that there is no analogous formula expressing the roots of a polynomial of degree five or higher in the general case (the Abel-Ruffini Theorem). This essentially means that there are polynomials $f(x)$ of degree five (for example $x^5 - x - 1$) whose roots cannot be described by taking the coefficients of $f(x)$ and manipulating them by the usual algebraic operations together with the operations of taking nth roots (forming radicals), in the way we do for quadratic polynomials (and can be done for cubic and quartic polynomials).

Despite the impossibility of a general formula for degree larger than four, the fundamental Theorem of Algebra (first proof by Argand in 1806, more proofs by Gauss soon later) asserts that at least one complex root exixts for any non-constant polynomial.

THEOREM 32 (Fundamental Theorem of Algebra). *Every polynomial in $\mathbb{C}[x]$ of positive degree has at least one root in $\mathbb{C}$.*

COROLLARY 33. *The irreducible polynomials in $\mathbb{C}[x]$ are those of degree one.*

Many proofs are known but none is really easy, so we will not prove the theorem.

PROOF. The polynomials of degree one are always irreducible, so we only need to prove the converse: if $f(x)$ is irreducible in $\mathbb{C}[x]$, then $f(x)$ has degree one.

Because of the Fundamental Theorem of Algebra, $f(x)$ has at least one root α. By the Factor Theorem we have $f(x) = (x - \alpha)\, g(x)$, for some $g(x) \in \mathbb{C}[x]$. This cannot be a proper factorisation of $f(x)$, because $f(x)$ is irreducible, and so $g(x)$ must be a constant. Consequently, $f(x)$ equals $x - \alpha$ times a nonzero constant, and so $f(x)$ has degree one. □

COROLLARY 34. *Every polynomial of positive degree in $\mathbb{C}[x]$ is a product of polynomials of degree one (also called linear polynomials).*

PROOF. We know that if F is any field then every polynomial of positive degree in $F[x]$ is a product of irreducible polynomials. But when $F = \mathbb{C}$ all irreducible polynomials have degree one. □

EXAMPLE. One can show that the polynomial $x^5 - x - 1$ is irreducible in $\mathbb{Q}[x]$. There exists no formula for the roots (using only algebraic operations and radicals), but one can find numerically that one root is approximately 1.167 (and more precisely 1.167303978...). This is a real root, but of course, in particular, it is a complex root. According to the Factor Theorem, $x - 1.167$ (approximately) divides $x^5 - x - 1$, and Ruffini's rule gives us [9]

	1	0	0	0	−1	−1
1.167		1.167	1.362	1.590	1.856	−1
	1	1.167	1.362	1.590	0.856	0

and so we find

$$x^5 - x - 1 \approx (x - 1.167)(x^4 + 1.167x^3 + 1.362x^2 + 1.590x + 0.856).$$

In turn, the factor of degree 4 has at least one complex root, and continuing in this way we eventually find the complete complex factorisation of $x^5 - x - 1$,

$$\approx (x-1.167)(x-0.181+1.083\,i)(x-0.181-1.083\,i)(x+0.764+0.352\,i)(x+0.764-0.352\,i).$$

We see that the non-real roots come in conjugate pairs (see next section), and if we multiply together the corresponding factors we find the complete factorisation of x^5-x-1 in $\mathbb{R}[x]$, which is

$$x^5 - x - 1 \approx (x - 1.167)(x^2 - 0.362x + 1.207)(x^2 + 1.529x + 0.709).$$

We will see the theory of the factorisation over $\mathbb{R}$ in the next section.

[9] Here we have written only three decimal digits after the point of each number, but we have done the calculations with higher precision. In reality we will never find remainder exactly zero with a calculator, as we are using an approximation of the true root. For example, $1.167^5 - 1.167 - 1 \approx -0.0025$.

22. Roots and factorisations of a polynomial with real coefficients

Recall that any complex number α can be uniquely written in the form $\alpha = s + it$, where s and t are real numbers. For now we may actually take this as a definition of complex numbers, and define addition and multiplication by treating them like ordinary "expressions" except that every time we encounter i^2 we may replace it with -1 (so i is not a letter like any other but satisfies the "simplification rule" $i^2 = -1$). Hence complex numbers can be added, subtracted, and multiplied as follows:

$$(s + it) \pm (u + iv) = (s \pm u) + i(t \pm v),$$
$$(s + it)(u + iv) = su + i(sv + tu) + i^2tv = (su - tv) + i(sv + tu).$$

To perform division it is useful to introduce the *conjugate* of a complex number $\alpha = s + it$, which is $\bar{\alpha} = s - it$. Because $\alpha\bar{\alpha} = (s + it)(s - it) = s^2 + t^2 = |\alpha|^2$ (where the *modulus* $|\alpha|$ is the nonnegative real number given by $|s + it| = \sqrt{s^2 + t^2}$), the reciprocal of α can be computed as $\dfrac{1}{\alpha} = \dfrac{\bar{\alpha}}{|\alpha|^2}$, and general division of complex numbers then easily follows.

Now note that complex conjugation has the following properties, which hold for any $\alpha, \beta \in \mathbb{C}$:

$$\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}, \qquad \overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}.$$

These properties essentially say that conjugation, in the sense of the map $\mathbb{C} \to \mathbb{C}$ taking $\alpha \mapsto \overline{\alpha}$, is an *automorphism* of $\mathbb{C}$. They can be verified by writing $\alpha = s+it$ and $\beta = u+iv$ and checking

$$\overline{(s + it) + (u + iv)} = \overline{(s + u) + i(t + v)} = (s + u) - i(t + v) = (s - it) + (u - iv),$$
$$\overline{(s + it)(u + iv)} = \overline{(su - tv) + i(sv + tu)} = (su - tv) - i(sv + tu) = (s - it)(u - iv).$$

Note also that a complex number α is actually real precisely when $\overline{\alpha} = \alpha$. Other properties follow, such as $\overline{\alpha - \beta} = \overline{\alpha + (-1)\beta} = \overline{\alpha} + \overline{(-1)\beta} = \overline{\alpha} + (-1)\overline{\beta} = \overline{\alpha} - \overline{\beta}$ for subtraction, and a similar one for division, $\overline{\alpha/\beta} = \overline{\alpha}/\overline{\beta}$. Also, $\overline{\alpha^2} = \overline{\alpha}^{\,2}$, and more generally $\overline{\alpha^n} = \overline{\alpha}^{\,n}$.

As an application of these basic properties of conjugation, we now show that if a complex number is a root of a polynomial with real coefficients, then its conjugate is also a root.

LEMMA 35. *If a complex number $\alpha = s + it$ is a root of a polynomial $f(x) \in \mathbb{R}[x]$, then its conjugate $\overline{\alpha} = s - it$ is a root as well.*

PROOF. Write the polynomial as $f(x) = a_n x^n + \cdots + a_2 x^2 + a_1 x + a_0$, hence $a_j \in \mathbb{R}$. Then if α is any complex number, not necessarily a root of $f(x)$, we have

$$\begin{aligned}
f(\overline{\alpha}) &= a_n\,\overline{\alpha}^{\,n} + \cdots + a_2\,\overline{\alpha}^{\,2} + a_1\,\overline{\alpha} + a_0 \\
&= a_n\,\overline{\alpha^n} + \cdots + a_2\,\overline{\alpha^2} + a_1\,\overline{\alpha} + a_0 \qquad \text{(because } \overline{\alpha^n} = \overline{\alpha}^{\,n}\text{)} \\
&= \overline{a_n\,\alpha^n} + \cdots + \overline{a_2\,\alpha^2} + \overline{a_1\,\alpha} + \overline{a_0} \qquad \text{(because } \overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta} \text{ and } \overline{a_j} = a_j\text{)} \\
&= \overline{a_n\,\alpha^n + \cdots + a_2\,\alpha^2 + a_1\,\alpha + a_0} \qquad \text{(because } \overline{\alpha+\beta} = \overline{\alpha} + \overline{\beta}\text{)} \\
&= \overline{f(\alpha)}.
\end{aligned}$$

In particular, if α is a root of $f(x)$, which means $f(\alpha) = 0$, then $f(\overline{\alpha}) = \overline{f(\alpha)} = 0$, and so $\overline{\alpha}$ is also a root. □

Note that if α is any complex number, then the polynomial

$$(x - \alpha)(x - \overline{\alpha}) = x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}$$

has real coefficients. In fact, if $\alpha = r + it$, then $\alpha + \overline{\alpha} = (r + it) + (r - it) = 2r$, and $|\alpha|^2 = \alpha\overline{\alpha} = (r + it)(r - it) = r^2 + t^2$. That quadratic polynomial has discriminant

$$(\alpha + \overline{\alpha})^2 - 4\alpha\overline{\alpha} = (\alpha - \overline{\alpha})^2 = (2it)^2 = -4t^2,$$

which is zero if α is real (and it must be so because α is a double root in that case), but is negative otherwise (and it must be so because the polynomial has no real roots in that case).

THEOREM 36. *The irreducible polynomials in $\mathbb{R}[x]$ are precisely the polynomials of degree 1, and the quadratic polynomials $ax^2 + bx + c$ with $b^2 - 4ac < 0$.*

PROOF. A polynomial of degree 1 is always irreducible. We have just seen that a polynomial $ax^2 + bx + c$ with $b^2 - 4ac < 0$ has no real roots. We know that for polynomials of degree two or three this implies that the polynomial is irreducible.

Now we prove the converse. Let $f(x)$ be any irreducible polynomial in $\mathbb{R}[x]$. By definition, it must have positive degree, and so by the Fundamental Theorem of Algebra it has at least one complex root α. By the Factor Theorem (in $\mathbb{C}[x]$) we have $f(x) = (x - \alpha)\,g(x)$, for some $g(x) \in \mathbb{C}[x]$.

If α is real, then $g(x) \in \mathbb{R}[x]$. Because we are assuming $f(x)$ irreducible, it follows that $g(x)$ is constant, and so $f(x)$ has degree 1.

If α is not real, then we have seen that $\overline{\alpha}$ is also a root of $f(x) = (x - \alpha)\,g(x)$, and being different from α it must be a root of $g(x)$, so $g(\overline{\alpha}) = 0$. Hence $x - \overline{\alpha}$ must divide $g(x)$, and so

$$f(x) = (x - \alpha)(x - \overline{\alpha}) \cdot h(x) = [x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}] \cdot h(x)$$

for some $h(x) \in \mathbb{C}[x]$. But we have seen earlier that the quadratic factor $x^2-(\alpha+\overline{\alpha})x+\alpha\overline{\alpha}$ has real coefficients, and so $h(x)$ has real coefficients as well. Because we are assuming $f(x)$ irreducible, it follows that $g(x)$ is constant, and so $f(x)$ has degree 2. Also, it has negative discriminant as claimed in the theorem because its roots α and $\overline{\alpha}$ are not real. □

COROLLARY 37. *Every polynomial of positive degree in $\mathbb{R}[x]$ is a product of polynomials of degree one and of quadratic polynomials with negative discriminant.*

Consequently, a polynomial of odd degree in $\mathbb{R}[x]$ has always at least one real root.

EXAMPLE. Consider the polynomial $f(x) = 4x^4+20x^3+30x^2-40x+26$, and suppose that we have somehow found out that $-3+2i$ is a root, meaning that $f(-3+2i) = 0$. Our task is to find the remaining complex roots, and obtain a complete factorisation of $f(x)$ in $\mathbb{C}[x]$.

According to the Factor Theorem, $x+3-2i$ is a factor of $f(x)$. Hence we divide $f(x)$ by $x+3-2i$ using Ruffini's rule:

	4	20	30	-40	26
$-3+2i$		$-12+8i$	$-40-8i$	$46+4i$	-26
	4	$8+8i$	$-10-8i$	$6+4i$	0

This confirms that $f(-3+2i)$ is actually a root of $f(x)$ as claimed. It also tells us that

$$f(x) = (x+3-2i)\cdot[4x^3+(8+8i)x^2+(-10-8i)x+(6+4i)].$$

Because $-3+2i$ is a root of $f(x)$ we know that its conjugate $-3-2i$ is a root as well, and because that is not a root of its factor $x+3-2i$ it must be a root of the other factor, a cubic polynomial. Dividing that by $x+3+2i$ using Ruffini's rule, we find

	4	$8+8i$	$-10-8i$	$6+4i$
$-3-2i$		$-12-8i$	$12+8i$	$-6-4i$
	4	-4	2	0

Hence $f(x) = (x+3-2i)(x+3+2i)(4x^2-4x+2)$. Finally, the roots of the quadratic factor can easily be found to be $(1 \pm i)/2$, by means of the usual formula, and so the complete factorisation of $f(x)$ in $\mathbb{C}[x]$ is

$$f(x) = (x+3-2i)(x+3+2i)(2x-1-i)(2x-1+i).$$

The complete factorisation of $f(x)$ in $\mathbb{R}[x]$ can be found by multiplying together the pairs of linear factors corresponding to conjugate roots: $f(x) = (x^2+6x+13)(4x^2-4x+2)$.