22 Rings of Polynomials

Consider the following examples whereby we solve polynomial equations in the method of more elementary courses:

1. $x^2 - 3x + 2 = 0 \implies (x-1)(x-2) = 0 \implies x = 1$ or $x = 2$.
2. $2x^2 - x - 1 = 0 \implies (2x+1)(x-1) = 0 \implies x = -\frac{1}{2}$ or $x = 1$.
3. $x^2 + 4 = 0 \implies (x-2i)(x+2i) = 0 \implies x = \pm 2i$.

All three polynomials had their coefficients in the ring of integers $\mathbb{Z}$. A couple of observations are important:

- The method of *factorization* is crucial. We implicitly use a property inherent to *integral domains*: if the product of two terms is zero, at least one of the terms must be zero.
- *Solutions need not live in the same ring as the coefficients.* We might need to extend our ring of coefficients in order to find all solutions. In elementary courses, we implicitly assume that we can extend our ring of coefficients to the complex numbers and thus find all solutions to a polynomial.

It is this second point which provided much of the purpose behind the development of ring and field theory. The general problem is this:

> Given a polynomial whose coefficients live in a field $\mathbb{F}$, can we find a field $\mathbb{E}$ containing $\mathbb{F}$ which contains a zero of the polynomial?

In elementary mathematics, where $\mathbb{F} = \mathbb{Q}$ or $\mathbb{R}$, it is enough to let $\mathbb{E} = \mathbb{C}$ be the complex numbers: this is the famous *Fundamental Theorem of Algebra.* Indeed the complex numbers were essentially *invented* with this purpose. It will take some time, but we will eventually be able to answer the general problem in the affirmative.

The obvious place to start is with a rigorous definition of the ring of polynomials over a ring R.

Definition 22.1. Let R be a ring. A *polynomial $f(x)$ with indeterminate x and coefficients in R* is a formal sum

$$f(x) = \sum_{k=0}^{\infty} a_k x^k = a_0 + a_1 x + \cdots + a_k x^k + \cdots$$

where all but finitely many of the *coefficients* $a_k \in R$ are non-zero.
The *degree* of $f(x)$ is the largest $n \in \mathbb{N}_0$ such that $a_n \neq 0$: the *leading term* is $a_n x^n$.
The *zero polynomial* is a formal sum where all coefficients are zero: by convention, $\deg(0) = -\infty$.
A degree n polynomial $f(x) \in R[x]$ is *monic* if $a_n = 1$ (requires R to have a unity).
The set of all such polynomials is denoted $R[x]$, the *ring of polynomials with coefficients in R.*

Examples $f(x) = 3x^2 + 2x + 1$ is a degree two polynomial in the ring $\mathbb{Z}_4[x]$.

$g(x) = \pi x^3 + (1+i)x$ is a degree three polynomial in the ring $\mathbb{C}[x]$.

Conventions and Notation

- In ring theory, x is not typically thought of as a *variable*, neither is $f(x)$ primarily considered a *function.* A polynomial is first and foremost a *formal* object: this means that x is not assumed to have any properties or to lie in any specified ring.
- It is common to omit zero terms of a polynomial and to omit the coefficient if $a_k = 1$. It is also standard to put the leading term first. Thus

  $$2 + 1x + 0x^2 + 8x^3 + 0x^4 + \cdots = 8x^3 + x + 2$$

  are *both* are legitimate expressions for the *same* polynomial, though we prefer the latter. When working abstractly, it is also common to ignore the indeterminate and write $f \in R[x]$: don't do this if the polynomial is explicit!
- Due to the possibility of zero-coefficients, observe that in general we can only claim

  $$\deg\left(a_n x^n + \cdots + a_1 x + a_0\right) \leq n$$

There are several critical but intuitive claims implicit in the definition of $R[x]$.

Theorem 22.2. *Let R be a ring.*

1. *$R[x]$ is a ring with respect to the inherited polynomial addition and multiplication. Specifically, if*

   $$f(x) = \sum_{k=0}^{\infty} a_k x^k \quad \text{and} \quad g(x) = \sum_{k=0}^{\infty} b_k x^k$$

   then

   $$f(x) + g(x) = \sum_{k=0}^{\infty} (a_k + b_k) x^k$$

   $$f(x)g(x) = \sum_{k=0}^{\infty} c_k x^k, \quad \text{where} \quad c_k = \sum_{i=0}^{k} a_i b_{k-i}$$

2. *If R has a unity $1 \neq 0$ then the constant polynomial 1 is the unity in $R[x]$.*
3. *$R[x]$ is commutative if and only if R is commutative.*

We leave this largely without proof, however note the following.

- You need to convince yourself that that $R[x]$ is closed under the definitions of polynomial addition and multiplication: it should be clear that if only finitely many of the terms a_k, b_k are non-zero, so are the coefficients of $f + g$ and fg.
- Explicitly checking the remaining axioms of a ring is tedious but straightforward.
- If R is non-commutative, then, in general, $\sum_{i=0}^{k} a_i b_{k-i} \neq \sum_{i=0}^{k} b_j i a_{k-i}$, whence $R[x]$ is also non-commutative.

Example In the ring $\mathbb{Z}_6[x]$ we see that

$$(x^2+2x+3)(x^3+4x) = x^5+2x^4+(3+4)x^3+8x^2+12x = x^5+2x^4+x^3+2x^2$$

Theorem 22.3. *Let $f,g \in R[x]$ where R is a ring.*

1. $$\begin{aligned} &\deg(f) < \deg(g) \implies \deg(f \pm g) = \deg(g), \\ &\deg(f) = \deg(g) \implies \deg(f \pm g) \le \deg(g) \quad \textit{and,} \\ &\deg(fg) \le \deg(f) + \deg(g) \end{aligned}$$

2. *R is an integral domain if an only if $R[x]$ is an integral domain. In such a case,*

$$\deg(fg) = \deg(f) + \deg(g)$$

Proof. 1. This is straightforward from the definition of addition and multiplication: the details are an exercise.

2. First observe that the degree formula holds trivially if either $f(x)$ or $g(x)$ is the zero polynomial. Now suppose that R is an integral domain. If $f(x) = \sum a_k x^k$ and $g(x) = \sum b_k x^k$ have degrees $m, n \ge 0$ respectively, then $f(x)g(x)$ has leading term $a_m b_n x^{m+n}$ *unless* $a_m b_n = 0 \in R$. But this would mean that R has zero-divisors: a contradiction. This justifies the degree formula and, moreover, proves that $f(x)g(x) \neq 0$ whence $R[x]$ is an integral domain.
   Conversely, if R has zero-divisors, so does $R[x]$, since $R \le R[x]$. ■

If R is a field, then $R[x]$ is merely an integral domain: note that the degree formula says that $f(x) = x$ is not a unit in $R[x]$. Indeed the only units in the integral domain $R[x]$ are the non-zero constant polynomials.

The golden rule

When considering polynomials, there are likely to be several rings simultaneously in play: a ring of coefficients R, a ring of polynomials $R[x]$, and perhaps a ring S in which zeros might lie. As such, it is important to keep track of which ring you are working in. Thus:

Only use an equals sign to denote equality of objects in the same ring!

Constant polynomials are a primary source of confusion: for instance, the symbol 0 now has *two* possible meanings:

Do we mean $0 \in R$ or $0 \in R[x]$?

The elementary approach to solving polynomial equations breaks the rule: for example, when we write

$$2x^2 - x - 1 = 0$$

we are simultaneously considering the left hand side as a polynomial $2x^2 - x - 1 \in \mathbb{Z}[x]$ *and* as the number 0: the elementary approach considers x to be a *variable*, the correct choice of which will yield $0 \in \mathbb{Z}$. The problem is that the correct choice(s) x may not lie in $\mathbb{Z}$. Since we are not equating objects

in the same ring, the above expression will be considered illegal from now on!
Of course, sometimes the unexpected is legitimate. For instance, we can write $4x^2 - 1 = 7$ within the ring $\mathbb{Z}_4[x]$, where we view 7 as a constant polynomial!
Our new approach will consider x to be an indeterminate object: provided you *never* claim that $x = a$ where a lies in some specific ring, you should be fine!

To confuse matters, the ring R is naturally isomorphic to the subring of constant polynomials

$$\{f(x) = a : a \in R\} \leq R[x]$$

This subring is usually referred to simply as R. It is common, and considered legitimate, to say that R is a subring of $R[x]$.

Finding zeros: Evaluation of a polynomial

Out of respect for the golden rule, we talk of *finding zeros* of a polynomial $f(x)$ rather than of *solving the equation* $f(x) = 0$. However we describe the practice, in order to find zeros of polynomials, we need to be able to evaluate them. This should be completely intuitive *except* for the fact that we may evaluate at an element in a *larger* ring than that containing the coefficients.

Definition 22.4. Let S be a ring containing a subring R, and let $f(x) = \sum_{k=0}^{\infty} a_k x^k \in R[x]$. Let $\alpha \in S$. The element

$$f(\alpha) = \sum_{k=0}^{\infty} a_k \alpha^k \in S$$

is the *evaluation of $f(x)$ at α.* An element $\alpha \in S$ is a *zero of $f(x)$* if[1] $f(\alpha) = 0$.

Most of the time we will specialize to the situation where R and S are commutative rings with unity: indeed these will almost always be integral domains or fields. Part of the reason for this lies is in the following basic result.

Theorem 22.5. *Let $R \leq S$ where R, S are rings and let $\alpha \in S$. Define the* evaluation function $\phi_\alpha : R[x] \to S$ *by $\phi_\alpha(f(x)) = f(\alpha)$.*

1. *If R has a unity, then $\phi_\alpha(x) = \alpha$.*
2. *If S is commutative, then ϕ_α is a ring homomorphism: we call it the* evaluation homomorphism.
3. *If S is an integral domain and $f = gh$ factorizes where $f, g, h \in R[x]$, then $\alpha \in S$ is a zero of f if and only if f is a zero of at least one of the factors g, h.*

Proof. 1. Note that the existence of the unity $1 \in R$ is equivalent to the existence of the polynomial $x = 1x \in R[x]$. The rest is clear: for any $\alpha \in S$,

$$\phi_\alpha(x) = \phi_\alpha(1x) = 1\alpha = \alpha$$

[1] We are not breaking the golden rule here: $\alpha \in S$ is always a *fixed* element so that the statement $f(\alpha) = 0$ is an equality *within the ring S.*

2. This is mostly immediate from the definition of polynomial addition and multiplication in part 1 of Theorem 22.2, particularly the fact that $\phi_\alpha(f(x)+g(x)) = \phi_\alpha(f(x)) + \phi_\alpha(g(x))$. However, it is worth taking a little time over the multiplicative structure:

$$\begin{aligned} \phi_\alpha\big(f(x)g(x)\big) &= \sum_{k=0}^{\infty} c_k\alpha^k = \sum_{k=0}^{\infty}\sum_{i=0}^{k} a_i b_{k-i}\alpha^k \\ &= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i b_j \alpha^{i+j} && \text{(reindex } i+j=k\text{)} \\ &= \sum_{i=0}^{\infty} a_i\alpha^i \sum_{j=0}^{\infty} b_j\alpha^j && \text{(since } S \text{ commutative)} \\ &= \phi_\alpha\big(f(x)\big)\phi_\alpha\big(g(x)\big) \end{aligned}$$

3. Since S is a commutative ring, $\phi_\alpha : R[x] \to S$ is a homomorphism, and so

$$f(\alpha) = g(\alpha)h(\alpha)$$

Since S is an integral domain, if the left hand side is zero, so must one of the factors on the right hand side.

■

Remarks

- The reindexing step is easy since the sums are infinite. The fact that $a_i = 0 = b_j$ whenever $i > \deg(f(x))$ and $j > \deg(g(x))$ makes the limits compatible.
- If S is non-commutative, then ϕ_α is not typically a homomorphism of rings. The issue is that multiplication of polynomials is *defined* so that the indeterminate x commutes with the coefficients! For example

$$\phi_\alpha((x-a)(x-b)) = \phi_\alpha(x^2 - (a+b)x + ab) = \alpha^2 - (a+b)\alpha + ab$$

whereas

$$\phi_\alpha(x-a)\phi_\alpha(x+b) = (\alpha - a)(\alpha - b) = \alpha^2 - a\alpha - \alpha b + ab$$

These expressions are only equal if $b\alpha = \alpha b$. What this really says is that *factorizing* is compatible with evaluation only when working in a commutative ring.

Examples

1. Let $R = \mathbb{Q}$ and $S = \mathbb{R}$ in the Theorem and let $f(x) = x^2 - 7$. Then

$$\phi_{\sqrt{7}}(f(x)) = 7 - 7 = 0$$

whence $\sqrt{7}$ is a *zero* of $f(x) \in \mathbb{Q}[x]$.
Thus far, it is not clear why we want the evaluation homomorphism at all: why not simply observe $f(\sqrt{7}) = 0$? The crucial point is that $f(x) = x^2 - 7$ lies in the *kernel* of the homomorphism $\phi_{\sqrt{7}}$: from this simple fact, the essential parts of our theory of finding zeros will follow…

We can't prove it just yet, but it shouldn't be too hard for you to believe that

$$\ker \phi_{\sqrt{7}} = \{(x^2-7)p(x) : p(x) \in \mathbb{Q}[x]\}$$

This is an example of an *ideal* subring of $\mathbb{Q}[x]$. Ideals play the same role for factor rings as normal subgroups do for factor groups. The construction of certain factor rings is the crucial tool whereby we answer our motivating problem.

2. Given the polynomial $f(x) = x^2+9 \in \mathbb{R}[x]$ and the ring $S = \mathbb{C}$, we can evaluate

   $$\phi_2(f(x)) = 2^2+9 = 13 \quad \text{and} \quad \phi_{3i}(f(x)) = 0$$

   so that $3i$ is a zero of $f(x)$. Again we observe that

   $$x^2+9 \in \ker \phi_{3i} = \{(x^2+9)p(x) : p(x) \in \mathbb{R}[x]\}$$

   It is perhaps no surprise that

   $$\ker \phi_2 = \{(x-2)p(x) : p(x) \in \mathbb{R}[x]\}$$

   This is simply the familiar idea that if $f(x) \in \mathbb{R}[x]$ is a polynomial with $f(2) = 0$, then $f(x)$ must have a *factor* of $(x-2)$. Again, our main purpose is to *prove* all of this in the context of polynomials over general fields.

3. You might have met the distinction between *algebraic* and *transcendental* numbers. An algebraic number is a complex number which is the zero of some polynomial in $\mathbb{Q}[x]$. A transcendental number is not the zero of any such polynomial. Thus

   $$\alpha \text{ algebraic} \iff \exists f(x) \in \mathbb{Q}[x] \text{ such that } \phi_\alpha(f(x)) = 0$$

   It can be shown that e and π are transcendental. Consider the evaluation homomorphism

   $$\phi_e : \mathbb{Q}[x] \to \mathbb{R} : a_0 + a_1x + \cdots + a_nx^n \mapsto a_0 + a_1e + \cdots + a_ne^n$$

   Since e is transcendental, it follows that $\ker \phi_e = \{0\}$. But this says that ϕ_e is an injective map![2] It follows that the ring $\operatorname{Im}(\phi_e) \leq \mathbb{R}$ is *isomorphic to the polynomial ring* $\mathbb{Q}[x]$*!*

More general constructions

Given a ring of polynomials $R[x]$, we can adjoin a second indeterminate y to create the ring

$$(R[x])[y] = \left\{\sum b_k(x)y^k : b_k(x) \in R[x]\right\}$$

Since multiplication within this ring essentailly assumes that the indeterminates x, y commute, this can be written alternatively as a ring of polynomials in two indeterminates

$$R[x,y] = \left\{\sum a_{jk}x^jy^k : a_{jk} \in R\right\}$$

[2] $\phi_e(f(x)) = \phi_e(g(x)) \implies \phi_e(f(x)-g(x)) = 0 \implies f(x)-g(x) = 0\ldots$

We can similarly define polynomial rings $R[x_1, x_2, \ldots, x_n]$ with any number of indeterminates. As before, if R is an integral domain, so are these polynomial rings.

We can also combine this discussion with the previous section. If R is an integral domain, so is $R[x]$, and we can form its field of fractions. This is the *field of rational fractions*, written

$$R(x) = \operatorname{Frac}(R[x]) = \left\{ \frac{f(x)}{g(x)} : f(x), g(x) \in R[x], g(x) \neq 0 \right\}$$

The same thing can be done with other rings: e.g. $R[x, y] \rightsquigarrow R(x, y)$.

23 Factorizing Polynomials over a Field

Except for remarks, from now on we work over a field $\mathbb{F}$. We start by restating part of Theorem 22.5.

Theorem 23.1. *Suppose $\mathbb{E}$ is a field containing a subfield $\mathbb{F}$, that $f \in \mathbb{F}[x]$ and $\alpha \in \mathbb{E}$. Suppose also that $f = gh$ factorizes where $g, h \in \mathbb{F}[x]$. Then*

1. $f(\alpha) = g(\alpha)h(\alpha)$.
2. *α is a zero of f if and only if α is a zero of g or h.*

Our next goal is to establish the important correspondence between factorizing and zeros. This turns out to be an immediate corollary of the division algorithm for polynomials:

Theorem 23.2 (Division algorithm)**.** *Let $\mathbb{F}$ be a field. Suppose that $f, g \in \mathbb{F}[x]$ are polynomials with g non-zero. Then there exist unique polynomials $q, r \in \mathbb{F}[x]$ such that*

$$f = qg + r \quad \text{and} \quad \deg(r) < \deg(g) \tag{$*$}$$

Proof. First we establish existence. Assume that $\deg(g) = m \geq 0$ is fixed throughout. We prove by induction on the degree n of f.

Base case If f is the zero polynomial, we may choose $q(x) = 0 = r(x)$.

Induction step Fix $n \in \mathbb{N}_0$ where $n = \deg(f)$ and assume $(*)$ for every polynomial $\widetilde{f}$ of degree $< n$.

If $m = \deg(g) > n$ we are done: let $q = 0$ and $r = f$.

Otherwise, assume $m \leq n$ and let $a_n x^n$ and $b_m x^m$ be the leading coefficients of f and g. The polynomial

$$\widetilde{f}(x) = f(x) - \frac{a_n}{b_m} x^{n-m} g(x)$$

clearly has degree $< n$: the induction hypothesis says that there exist[3] $\widetilde{q}, \widetilde{r} \in \mathbb{F}[x]$ such that

$$\widetilde{f}(x) = \widetilde{q}(x)g(x) + \widetilde{r}(x) \quad \text{and} \quad \deg(\widetilde{r}) < \deg(g)$$

But then

$$f(x) = \left(\widetilde{q}(x) + \frac{a_n}{b_m} x^{n-m} \right) g(x) + \widetilde{r}(x)$$

By induction, we can find polynomials satisfying $(*)$ regardless of the degree of f.

[3] Not yet necessarily unique...

Now for uniqueness: suppose we have two decompositions:

$$f(x) = q_1(x)g(x) + r_1(x) = q_2(x)g(x) + r_2(x) \quad \text{where} \quad \deg(r_1), \deg(r_2) < \deg(g) = m$$

Then

$$(q_2(x) - q_1(x))g(x) = r_1(x) - r_2(x)$$

Suppose $q_2 \neq q_1$ and take degrees

$$m \leq \deg(q_2 - q_1) + \deg(g) = \deg(r_1 - r_2) < m$$

for a contradiction.[4] It follows that both sides are zero and we have uniqueness. ■

Definition 23.3. Let $f, g \in \mathbb{F}[x]$. We say that *g divides f* (written $g \mid f$) if r is the zero polynomial.

Corollary 23.4 (Factor Theorem). *Let $\mathbb{F}$ be a field and $f \in \mathbb{F}[x]$. An element $\alpha \in \mathbb{F}$ is a zero of f if and only if $(x - \alpha) \mid f(x)$. That is there exists some $q \in \mathbb{F}[x]$ such that*

$$f(x) = (x - \alpha)q(x)$$

Proof. By the division algorithm, $f(x) = (x - \alpha)q(x) + r(x)$ where $\deg(r) < \deg(x - \alpha) = 1$: thus r is constant. However, evaluating at α says that

$$f(\alpha) = (\alpha - \alpha)q(\alpha) + r(\alpha) = r(\alpha)$$

Thus α is a zero of f if and only if r is the zero polynomial. ■

Examples How you factorize is largely a matter of taste. Here are two methods.

Long Division Suppose $f(x) = x^4 - 2x^3 - x + 1$ and $g(x) = x^2 + 2x - 1$ in $\mathbb{Z}_5[x]$. The standard computation from high-school algebra should see you through. Remember that we're working in $\mathbb{Z}_5$, so certain simplifications are made, such as $-2x^3 - 2x^3 = x^3$. Since $g(x)$ has leading coefficient 1, it is also legitimate to apply the division algorithm in $\mathbb{Z}[x]$ and take remainders modulo 5 afterwards: this calculation is done second; compare with the first.

$$\begin{array}{r|rrrrr}
 & & & x^2 & +x & -1 \\ \hline
x^2+2x-1 & x^4 & -2x^3 & & -x & +1 \\
 & -x^4 & -2x^3 & +x^2 & & \\ \cline{2-4}
 & & x^3 & +x^2 & -x & \\
 & & -x^3 & -2x^2 & +x & \\ \cline{3-5}
 & & & -x^2 & & +1 \\
 & & & x^2 & +2x & -1 \\ \cline{4-6}
 & & & & 2x &
\end{array}
\qquad
\begin{array}{r|rrrrr}
 & & & x^2 & -4x & +9 \\ \hline
x^2+2x-1 & x^4 & -2x^3 & & -x & +1 \\
 & -x^4 & -2x^3 & +x^2 & & \\ \cline{2-4}
 & & -4x^3 & +x^2 & -x & \\
 & & 4x^3 & +8x^2 & -4x & \\ \cline{3-5}
 & & & 9x^2 & -5x & +1 \\
 & & & -9x^2 & -18x & +9 \\ \cline{4-6}
 & & & & -23x & +10
\end{array}$$

We conclude that in $\mathbb{Z}_5[x]$,

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

[4] The first inequality *requires* that $\mathbb{F}$ be an integral domain (Theorem 22.3).

Term-by-term Simply work through the calculation, starting with the highest power in $q(x)$ and deciding what each term must be in order to reduce the degree of $f(x) - q(x)g(x)$. The whole calculation can be done in one line and, once written, it will look like you've done no working! Here are two examples.

1. Let $f(x) = x^4 + 2x^3 + 3$ and $g(x) = x^2 + 3x + 1$ in $\mathbb{Z}_7[x]$. We write $f(x) = g(x)q(x) + r(x)$ and build $q(x)$ step-by step:

$$\begin{aligned} x^4 + 2x^3 + 3 &= (x^2+3x+1)(x^2+\cdots)+\cdots && \text{(Match } x^4\text{)} \\ &= (x^2+3x+1)(x^2-x+\cdots)+\cdots && \text{(Match } 2x^3\text{)} \\ &= (x^2+3x+1)(x^2-x+2)+\cdots && \text{(Match } 0x^2\text{)} \\ &= (x^2+3x+1)(x^2-x+2)+2x+1 && \text{(Balance remaining terms)} \end{aligned}$$

2. Observe that $\alpha = 2$ is a zero of $f(x) = x^4 + 2x^3 + 4x + 4$ in $\mathbb{Z}_{11}[x]$. We therefore divide by $g(x) = x - 2$ as follows:

$$\begin{aligned} x^4 + 2x^3 + 4x + 4 &= (x-2)(x^3+\cdots) && \text{(Match } x^4\text{)} \\ &= (x-2)(x^3+4x^2+\cdots) && \text{(Match } 2x^3\text{)} \\ &= (x-2)(x^3+4x^2-3x+\cdots) && \text{(Match } 0x^2\text{)} \\ &= (x-2)(x^3+4x^2-3x-2) && \text{(Match } 4x\text{)} \end{aligned}$$

At the last step, you should also check the constant terms match: $(-2)^2 = 4$. You could also observe that $\alpha = 1$ is a zero of the polynomial, and keep factorizing:

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

Corollary 23.5. *A degree n polynomial over a field $\mathbb{F}$ has* at most *n distinct zeros in $\mathbb{F}$.*

Proof. Suppose that $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ are distinct zeros of a degree n polynomial $f(x) \in \mathbb{F}[x]$. We use the factor theorem to induct:

- $f(\alpha_1) = 0 \implies f(x) = (x - \alpha_1)q_1(x)$
- $0 = f(\alpha_2) = (\alpha_2 - \alpha_1)q_1(\alpha_2) \implies q_1(\alpha_2) = 0 \implies f(x) = (x-\alpha_1)(x-\alpha_2)q_2(x)$, etc.

We eventually obtain

$$f(x) = (x-\alpha_1)\cdots(x-\alpha_n)q_n(x)$$

Since $\mathbb{F}$ is an integral domain, the degree of $q_n(x)$ must be 0, whence $q_n(x)$ is a non-zero constant. Since $\mathbb{F}$ has no zero-divisors, we see that $f(\beta) \neq 0$ for all $\beta \neq \alpha_i$. In particular, $f(x)$ has no further zeros. ■

For example, $f(x) = x^3 + x^2 + x + 1$ in $\mathbb{Z}_3[x]$ has at most three zeros in $\mathbb{Z}_3$. In fact it has only one, namely $\alpha = 2$: it can be factored

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

Factorization and Irreducibles

If you've taken a Number Theory course, you've likely seen a proof of the *Fundamental Theorem of Arithmetic/Unique Factorization Theorem.*[5] Amazingly, this entire discussion may be replicated for polynomials over any field and an equivalent result stated! The first challenge is to properly define and identify irreducible polynomials: essentially these are polynomials which do not factor.

Definition 23.6. Let $\mathbb{F}$ be a field[6] and let $f \in \mathbb{F}[x]$ have $\deg(f) \geq 1$.

1. We say that f is *reducible over* $\mathbb{F}$ if it can be factored $f = gh$ where *both* $g, h \in \mathbb{F}[x]$ have smaller degree than f.
2. $f \in \mathbb{F}[x]$ is *irreducible over* $\mathbb{F}$ otherwise.
3. Polynomials $f, g \in \mathbb{F}[x]$ are *associates* if there exists a unit $u \in \mathbb{F}[x]$ such that $f = ug$.

In a field, every polynomial is either zero, a unit (constant non-zero polynomial), irreducible or reducible. The following are equivalent statements of the irreducibility of a degree ≥ 1 polynomial $f \in \mathbb{F}[x]$:

- $f = gh \implies$ one of g, h is a unit and the other is an associate of f.
- $d \mid f \implies d$ is a unit, or an associate of f.

It should be clear that f is irreducible if and only if any associate of f is irreducible.

Theorem 23.7. *Let $\mathbb{F}$ be a field and $f \in \mathbb{F}[x]$ have degree ≥ 1. Then f has a factorization in terms of irreducibles.*

Proof. We prove by induction on the degree of f.

Base case If $\deg(f) = 1$ then f is irreducible.

Induction step Let $n \in \mathbb{N}$ and suppose that all polynomials $f \in \mathbb{F}[x]$ with $1 \leq \deg(f) \leq n$ may be factored in terms of irreducibles. Let $g \in \mathbb{F}[x]$ have degree $n + 1$. Then either:

g is irreducible We're done!

g is reducible Then $g = hk$ where $\deg(h), \deg(k) < n + 1$. But then h, k are products of irreducibles and we're done.

■

[5] In short: If $n \geq 2$ is an integer, then there exists a unique factorization $n = p_1^{\mu_1} \cdots p_k^{\mu_k}$ where $p_1 < \cdots < p_k$ are primes and each $\mu_i \in \mathbb{N}$.

The proof process is essentially this:

1. An *irreducible* is an integer $p \geq 2$ which satisfies $a \mid p \implies a = \pm 1$ or $a = \pm p$.
2. Show that every integer $n \geq 2$ is divisible by an irreducible and, consequently, can be factored into irreducibles.
3. Prove that every irreducible is *prime*: $p \mid ab \implies p \mid a$ or $p \mid b$.
4. Use the equivalence of primes and irreducibles to prove that the factorization obtained in 2. is unique.

It is possible to replicate every step of this process for $\mathbb{F}[x]$.

[6] As most, this definition extends to an *integral domain.* For technical reasons, there are multiple competing meanings of 'irreducible' for polynomials over a weaker structure such as a commutative ring. Even the concept of 'associate' can become a little more complicated over an integral domain.

Finding Irreducibles In general, identifying irreducibles is *hard.* We give several straightforward criteria below.

Theorem 23.8. 1. *Suppose $f \in \mathbb{F}[x]$ has $\deg(f) \geq 2$. Then*

$$f \text{ irreducible} \implies f \text{ has no zeros in } \mathbb{F}$$

2. *If $\deg(f) = 2$ or 3 and $f(x)$ is reducible, then one of the factors must be linear. In this case part 1. becomes an 'if and only if.'*

3. *(Rational Roots Theorem) If $f(x) = \sum a_k x^k \in \mathbb{Z}[x]$ has $\deg(f) = n$ and a zero $\frac{p}{q} \in \mathbb{Q}$ where $\gcd(p, q) = 1$, then $p \mid a_0$ and $q \mid a_n$.*
   If $f \in \mathbb{Z}[x]$ is monic ($a_n = 1$), then every rational zero is an integer.

Proof. Parts 1. and 2. are immediate from the factor theorem. For part 3. simply evaluate

$$0 = q^n f(\tfrac{p}{q}) = a_n p^n + a_{n-1} p^{n-1} q + \cdots + a_1 p q^{n-1} + a_0 q^n$$

The first n terms are divisible by p, whence $p \mid a_0 q^n$. But then $p \mid a_0$, since $\gcd(p, q) = 1$. Similarly, the last n terms being divisible by q prove that $q \mid a_n$.

■

Examples

1. Any linear polynomial $f(x) = x - \alpha \in \mathbb{F}[x]$ is irreducible!

2. In $\mathbb{Z}_2[x]$ there are three distinct irreducible polynomials of degree 2 and 3, namely

$$x^2 + x + 1, \quad x^3 + x + 1, \quad x^3 + x^2 + 1$$

3. The polynomial $f(x) = x^2 - 2$ is irreducible in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$, but is reducible in $\mathbb{R}[x]$ since $f(x) = (x - \sqrt{2})(x + \sqrt{2})$.

4. A degree 4 polynomial need not have a linear factor in order to be reducible. For example:

   - $x^4 + 3x^2 + 2 = (x^2 + 1)(x^2 + 2)$ is reducible over $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, but has no zeros in any of these rings. Over $\mathbb{C}$, the polynomial factorizes completely, with zeros $\pm i, \pm\sqrt{2}i$.
   - It is a theorem that every non-constant polynomial in $\mathbb{R}[x]$ can be factorized uniquely into a product of irreducible linear and quadratic terms in $\mathbb{R}[x]$. For example,

     $$x^4 + 1 = (x^2 + \sqrt{2}x + 1)(x^2 - \sqrt{2}x + 1)$$

     These factors are irreducible in $\mathbb{R}[x]$ since their zeros $\frac{\pm 1 \pm i}{\sqrt{2}}$ are non-real.
   - Consider $f(x) = x^3 + 13x^2 + 2x + 3 \in \mathbb{Q}[x]$. If f were reducible, then it would have a linear factor and therefore a rational zero. By the rational roots theorem, any such zero must be an integer p dividing 3: that is $p = \pm 1, \pm 3$. It is easily checked that these values satisfy $f(p) \neq 0$. We conclude that f is irreducible over $\mathbb{Q}$.

We shall state two further criteria for irreducibility later.

Uniqueness of Factorization

Uniqueness of the irreducible factorization obtained in Theorem 23.7 requires that one consider *prime* polynomials.

Definition 23.9. A degree ≥ 1 polynomial $f \in \mathbb{F}[x]$ is *prime* if $\forall g, h \in \mathbb{F}[x]$,

$$f \mid gh \implies f \mid g \text{ or } f \mid h$$

Theorem 23.10. 1. *In any integral domain,*[7] *f prime $\implies$ f irreducible.*

2. *If $\mathbb{F}$ is a field, then $f \in \mathbb{F}[x]$ irreducible $\implies$ f prime.*

Consequently, over a field $\mathbb{F}$, irreducible and prime polynomials are the same objects.[8]

Proof. 1. Suppose that f is prime and that $f = gh$. WLOG we may assume $f \mid g$. But then

$$g = fk \implies f = fkh \implies 1 = kh \implies h \text{ is a unit}$$

It follows that f is irreducible.

2. A proof of this fact will come later when we consider principal ideal domains. We will also shortly outline another argument dependent on gcd's.

■

Theorem 23.11 (Unique Factorization). *If $\mathbb{F}$ is a field, then every $f \in \mathbb{F}[x]$ has a factorization in terms of irreducible polynomials which is unique up to ordering and choice of associates.*

Proof. Theorem 23.7 says that factorizations in terms of irreducibles exist. Suppose that

$$f = g_1 \cdots g_n = h_1 \cdots h_m$$

are two such. If any polynomials g_i, h_j are associates, we may cancel from each side, and absorb the resulting unit into another irreducible. If we can cancel all irreducibles from each side (necessarily $n = m$), we are done: the factorization is unique.
Otherwise, after cancelling associates, we necessarily have at least some of the irreducibles remaining on each side,[9] none of which are associates. Suppose that g_i remains on the left hand side. It therefore divides a product of irreducibles h_j. However, since g_i is irreducible, it is also prime, and therefore must divide one of the h_j's. Since h_j is irreducible,

$$g_i \mid h_j \implies g_i \text{ and } h_j \text{ are associates}$$

A contradiction!

■

[7] Recall that if $\mathbb{F}$ is a field or an integral domain, then $\mathbb{F}[x]$ is an integral domain, so this result holds for the cases in which we're interested.

[8] By convention, we always use *irreducible* when refering to polynomials. Similarly, prime and irreducible integers are identical: the convention in $\mathbb{Z}$ is to use *prime.*

[9] Else the produce of irreducibles (deg ≥ 1) is a unit (deg $= 0$).

Further Irreducibility Criteria

Theorem 23.12 (Gauss' Lemma[10]). *If $f \in \mathbb{Z}[x]$ is irreducible over $\mathbb{Z}$ then it is also irreducible over $\mathbb{Q}$.*

The rough idea is that if $f \in \mathbb{Z}[x]$ factorizes over $\mathbb{Q}$, then any fractional coefficients must be cancelled by the numerators in the *other* factor. For example, suppose $f(x) = \left(\frac{1}{2}x + \frac{1}{3}\right)(ax+b) \in \mathbb{Z}[x]$ is reducible over $\mathbb{Q}$, where $a, b \in \mathbb{Q}$. Then

$$f(x) = \frac{a}{2}x^2 + \left(\frac{a}{3} + \frac{b}{2}\right)x + \frac{b}{3} \in \mathbb{Z}[x] \implies a, b \in 6\mathbb{Z}$$
$$\implies f(x) = \left(\frac{1}{2}x + \frac{1}{3}\right)(6\tilde{a}x + 6\tilde{b}) \quad \text{for some } \tilde{a}, \tilde{b} \in \mathbb{Z}$$

But then $f(x) = (3x+2)(\tilde{a}x + \tilde{b})$ is reducible over $\mathbb{Z}$.

Proof. Let $f = gh \in \mathbb{Z}[x]$ where $g, h \in \mathbb{Q}[x]$. Let α and β be, respectively, the least common multiples of the denominators of the coefficients of g and h respectively. It follows that

$$\alpha\beta f = \tilde{g}\tilde{h} \quad \text{where} \quad \tilde{g}, \tilde{h} \in \mathbb{Z}[x]$$

If $\alpha\beta \in \mathbb{Z}$ is a unit, we're done. Otherwise, use the unique factorization theorem in $\mathbb{Z}$ to write $\alpha\beta = p_1 \ldots p_n$ for primes p_i. Modulo p_i, we now have

$$\tilde{g}\tilde{h} \equiv 0 \mod p_i \implies \tilde{g} \equiv 0 \quad \text{or} \quad \tilde{h} \equiv 0 \mod p_i$$

whence all coefficients of at least one of $\tilde{g}$ or $\tilde{h}$ are divisible by p_i.
Repeating this process, we may divide out by each p_i one by one, until $\alpha\beta$ has be cancelled out, thus obtaining a factorization of f over $\mathbb{Z}$. ∎

Theorem 23.13.

1. *Let p be a prime and $f \in \mathbb{Z}[x]$ have degree ≥ 1. Moreover, assume that the leading term of f is non-zero modulo p. Then,*

   $$f \text{ irreducible over } \mathbb{Z}_p \implies f \text{ irreducible over } \mathbb{Z}$$

   Moreover, by Gauss, f is irreducible in $\mathbb{Q}[x]$.

2. *(Eisenstein's Criterion) Let $f(x) = \sum a_k x^k \in \mathbb{Z}[x]$ have degree $n \geq 1$. Suppose there is a prime $p \in \mathbb{Z}$ such that:*

   (a) $p \nmid a_n$ (this is true for any prime if f is monic!),
   (b) $p \mid a_i$ for all $i = 0, \ldots, n-1$,
   (c) $p^2 \nmid a_0$.

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

Proof. 1. Suppose $f(x) = g(x)h(x)$ in $\mathbb{Z}[x]$, where $\deg(g), \deg(h) \geq 1$. Then the same relationship holds taking all coefficients modulo p. The condition on the leading term of f guarantees that the degrees of f, g, h are the *same* over both $\mathbb{Z}$ and $\mathbb{Z}_p$, whence f is reducible over $\mathbb{Z}_p$.

[10] More generally, if R admits a unique factorization into irreducibles and $\mathbb{F}$ is its field of fractions, then $f \in R[x]$ irreducible over $R \implies f$ irreducible over $\mathbb{F}$. We work over $\mathbb{Z}$ and $\mathbb{Q}$, through the general proof is essentially identical.

2. Suppose that $f = gh$ is reducible in $\mathbb{Z}_p[x]$, then

   $$f(x) = a_n x^n = g(x)h(x) \in \mathbb{Z}_p[x]$$

   Since $\mathbb{Z}_p$ is a field, the left hand side is precisely the unique factorization of f into irreducibles in $\mathbb{Z}_p[x]$. Thus $g = x^k$ and $h = x^{n-k}$ up to associates and where $k \neq 0, n$. But then the constant terms of g and h must both be zero modulo p, whence the constant term of f is divisible by p^2: contradiction.

   ■

Examples

1. The polynomial $f(x) = x^2 + 1$ is irreducible over $\mathbb{Z}$ (this is easy to see directly!) and thus over $\mathbb{Q}$. This is since, modulo 3, we have no roots.

2. The polynomial $f(x) = x^3 + 17x + 391$ is irreducible over $\mathbb{Q}$. Taking this modulo 2, we obtain the polynomial $f(x) = x^3 + x + 1$ which is irreducible in $\mathbb{Z}_2[x]$. We could also apply Eisenstein with $p = 17$.

3. If p is prime, then $f(x) = x^n \pm p$ is irreducible over $\mathbb{Q}$.

4. We've already seen that $f(x) = x^4 + 1 \in \mathbb{Z}[x]$ is irreducible over $\mathbb{Q}$, but Eisenstein can deal with this too. Consider the evaluation homomorphism

   $$\phi_{x+1} : \mathbb{Z}[x] \to \mathbb{Z}[x]$$

   and use it to *define*

   $$g(x) = f(x+1) = \phi_{x+1}(f(x)) = x^4 + 4x^3 + 6x^2 + 4x + 2$$

   Now let $p = 2$ in Eisenstein's criterion.

5. A second example of this approach concerns the famous *cyclotomic polynomial* where p is prime:

   $$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1$$

   Using the Binomial theorem, we compute the substitution

   $$\begin{aligned} g(x) = \Phi_p(x+1) &= \frac{1}{x}\left[\sum_{k=0}^{p} x^k - 1\right] = \sum_{k=1}^{p} \binom{p}{k} x^{k-1} \\ &= x^{p-1} + px^{p-2} + \binom{p}{p-2} x^{p-3} + \cdots + \binom{p}{2} x + p \end{aligned}$$

   Since $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ is divisible by p for all $k \neq 0, p$, Eisenstein quickly shows that g is irreducible. The cyclotomic polynomial Φ_p is therefore irreducible over $\mathbb{Q}$. In fact the roots of Φ_p are precisely the pth roots of unity except 1: over $\mathbb{C}$, we have the factorization

   $$\Phi_p(x) = (x - \zeta) \cdots (x - \zeta^{p-1}) \quad \text{where} \quad \zeta = e^{\frac{2\pi i}{p}}$$

   The irreduciblilty of Φ_p over $\mathbb{Q}$ says that no product of a *proper subset* of these factors produces a polynomial with rational coefficients.

6. Warning! There exist polynomials such as $x^4 + 1$ which are irreducible over $\mathbb{Z}$ and $\mathbb{Q}$ yet are *reducible* modulo every prime! A proof requires a bit more machinery (e.g. Galois theory) than we have available, but it is still worth being aware that Theorem 23.13 part 1. is not a biconditional!

Some left-overs: non-examinable

1. The Euclidean algorithm in $\mathbb{F}[x]$ and Unique Factorization Here is an outline of a proof for Theorem 23.10, part 2., the missing piece of the Unique Factorization Theorem. Some work is required to justify step 1., but it is essentially the same as you've seen when discussing the Euclidean algorithm in $\mathbb{Z}$.

1. The division algorithm defines a Euclidean algorithm and thus a version of Bézout's identity: if $d = \gcd(f, g)$, then

   $$\forall f, g \in \mathbb{F}[x], \exists \lambda, \mu \in \mathbb{F}[x], \text{ such that } d = \lambda f + \mu g$$

   Here $d = \gcd(f, g)$ means any polynomial which satisfies:

   - d is a common divisor: $d \mid f$ and $d \mid g$.
   - If e is *any* common divisor, then $e \mid d$.

   The Euclidean algorithm shows that the gcd of two non-zero polynomials exists *by constructing one*! Moreover, it is easy to check that all gcds of a pair of given polynomials are associates.

2. Suppose f is irreducible and that $f \mid gh$. Let $d = \gcd(f, g)$; clearly $d \mid f$. Since f is irreducible, we have either

   - d is an associate of f¡ whence $f \mid d \implies f \mid g$ or,
   - d is a unit, whence $dh = \lambda fh + \mu gh = (\lambda h + \mu)f \implies f \mid h$

   Either way, we conclude that f is prime.

2. Division in Integral Domains and Commutative Rings Where in the proof of the division algorithm did we require that $\mathbb{F}$ be a field? We needed the non-zero coefficient b_m to be a unit. In an integral domain, the theorem still holds, but *only* if b_m is a unit. Moreover, the only reason we needed an integral domain for was to prove uniqueness. A careful re-reading of the division algorithm and factor theorem shows that we've proved:

Corollary 23.14. *Let $f \in R[x]$ where R is a commutative ring with 1.*

1. *Suppose that $g \in R[x]$ is* monic,[11]. *Then there exist polynomials $q, r \in R[x]$ for which*

   $$f(x) = q(x)g(x) + r(x) \quad \textit{and} \quad \deg(r) < \deg(g)$$

2. *Since $x - \alpha$ is a monic polynomial, the factor theorem holds: $f(\alpha) = 0 \iff f(x) = (x - \alpha)q(x)$ for some $q(x) \in R[x]$.*

3. *If R is an integral domain, then the polynomials q, r in part 1. are unique. Corollary 23.5 also holds: if $\deg(f) = n$, then f has at most n zeros in R.*

[11] It is enough that the leading coefficient be a unit, but we could then divide g by the unit and absorb it into q...

Examples

1. Let $f(x) = 1$ and $g(x) = b$ where b is a non-zero non-unit constant in an integral domain R. If the division algorithm were true, r would have to be the zero polynomial. However

   $$1 = q(x)b$$

   is impossible if b is a non-unit! In particular, the division algorithm is *false* in $\mathbb{Z}[x]$ unless the leading coefficient of g is ± 1.

2. $\mathbb{Z}_6$ is a unital commutative ring, but *isn't* an integral domain. Observe that

   $$f(x) = x^2 - x = x(x-1) = (x-3)(x-4) \in \mathbb{Z}_6[x]$$

   is a degree two polynomial with *four zeros*, thus showing that Corollary 23.5 does not apply. That each zero still corresponds to a linear factor dividing the polynomial reflects the fact that the factor theorem still holds.

3. Finite Subgroups of Units The factor theorem has many applications, here is yet another...

Corollary 23.15. *Suppose that $\mathbb{F}$ is a field (or merely an integral domain). Then any finite subgroup of the group of units $(\mathbb{F}^\times, \cdot)$ is cyclic.*

Proof. By the Fundamental Theorem of Finitely Generated Abelian Groups, we may assume that such a subgroup G is isomorphic to some $C_{p_1^{\mu_1}} \times \cdots C_{p_k^{\mu_k}}$ where $p_1, \ldots, p_k$ are prime. Our goal is to prove that the primes are distinct.
Let $m = \operatorname{lcm}(p_1^{\mu_1}, \ldots, p_k^{\mu_k}) \leq |G|$. Since the order of an element of $C_{p_i^{\mu_i}}$ divides $p_i^{\mu_i}$, it follows that the order of every element in G must divide m. This says that the polynomial $f(x) = x^m - 1$ has at least $|G|$ zeros in $\mathbb{F}$. Since $f(x)$ has at most m zeros, we conclude that $m = |G|$. It follows that the primes are distinct and that $G \cong C_m$.

■

Examples

1. If $\mathbb{F} = \mathbb{Z}_p$, then the group of units $\mathbb{Z}_p^\times \cong \mathbb{Z}_{p-1}$ is cyclic. We have already remarked on this in the context of primitive roots.

2. If $\mathbb{F}$ is *any* finite field, then its group of units is cyclic.

3. If G is a finite group of units in a field such as $\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$, then all its elements must have modulus 1: if $x \in G$ with $|x| \neq 1$, then $|x|^n$ takes on infinitely many values, whence G is infinite. It follows that the only finite groups of units in these fields are

   $$\{1\}, \{\pm 1\} \leq \mathbb{Q}^\times \leq \mathbb{R}^\times \quad \text{and} \quad U_n \leq \mathbb{C}^\times$$

   where U_n is the group of nth roots of unity. This last might require a bit of thinking!