Chapter 1

Rings, Fields and Polynomials

This first chapter contains a review of the background material required to study Galois Theory. The majority comes from the module *MT3505 Rings and Fields* and consequently many proofs in this chapter are omitted or greatly abbreviated. The last part of this chapter is concerned with polynomials and polynomial rings. One important concept that we shall use throughout the module is what it means for a polynomial to be *irreducible*. We shall devote some time to methods for establishing that a polynomial is irreducible.

Rings

We start with properties of rings before specialise to fields and to polynomial rings.

Definition 1.1 A *commutative ring with a* 1 is a set R endowed with two binary operations denoted as addition and multiplication such that the following conditions hold:

(i) R forms an abelian group with respect to addition (with additive identity 0, called *zero*);

(ii) multiplication is *associative*: $a(bc) = (ab)c$ for all $a, b, c \in R$;

(iii) multiplication is *commutative*: $ab = ba$ for all $a, b \in R$;

(iv) the *distributive laws* hold:

$$a(b + c) = ab + ac$$
$$(a + b)c = ac + bc$$

for all $a, b, c \in R$;

(v) there is a *multiplicative identity* 1 in R satisfying $a1 = 1a = a$ for all $a \in R$.

Comment: There is also a definition of a "ring", without the assumption of the multiplication being commutative or it having a multiplicative identity 1. One simply drops conditions (iii) and (v) from the definition above. Since we are interested in studying fields in this module, we shall not need to consider non-commutative rings as there will be no examples of such rings occurring in these notes. This is why we only give the more restricted definition of a *commutative* ring *with a* 1 above as this is sufficient for our needs. In addition, note that in a commutative ring one needs only assume one of the two distributive laws since the other may be deduced from that one via commutativity.

Definition 1.2 Let R be a commutative ring with a 1. An *ideal* I in R is a non-empty subset of R that is both an additive subgroup of R and satisfies the property that if $a \in I$ and $r \in R$, then $ar \in I$.

Thus a subset I of R is an ideal if it satisfies the following four conditions:

(i) I is non-empty (or $0 \in I$);

(ii) $a + b \in I$ for all $a, b \in I$;

(iii) $-a \in I$ for all $a \in I$;

(iv) $ar \in I$ for all $a \in I$ and $r \in R$.

(In a non-commutative ring, one needs to assume both ar and ra belong to I, but R being commutative ensures these products are equal.)

It follows from the definition that an ideal I of R is closed under multiplication: $ab \in I$ for all $a, b \in I$ (since an element $b \in I$ is, in particular, an element of the larger set R). This means that an ideal I is, in particular, a *subring*. Note that, in general, I does not contain the multiplicative identity 1, since if it did $r = 1r \in I$ for all $r \in R$. Thus, the only ideal of R that contains the multiplicative identity 1 is the ring R itself.

The reason for being interested in ideals is that one can form quotient rings, as we shall now describe. Let R be a commutative ring and let I be an ideal of R. Then I is, in particular, a subgroup of the additive group of R and the latter is an abelian group. We can therefore form the additive cosets of I; that is, define

$$I + r = \{\, a + r \mid a \in I \,\}$$

for each $r \in R$. We know from group theory (covered in both *MT2505* and *MT4003*) when two such cosets are equal,

$$I + r = I + s \qquad \text{if and only if} \qquad r - s \in I,$$

and that the set of all cosets forms a group via addition of the representatives:

$$(I + r) + (I + s) = I + (r + s) \qquad \text{for } r, s \in R.$$

(In arbitrary group, one requires that the subgroup is *normal*, but this holds because R is an *abelian* group under addition.) As is observed in *MT3505*, the assumption that I is an ideal then ensures that there is a well-defined multiplication on the set of cosets, given by

$$(I + r)(I + s) = I + rs \qquad \text{for } r, s \in R,$$

with respect to which the set of cosets $I + r$ forms a ring, called the *quotient ring* and denoted by R/I.

Theorem 1.3 *Let R be a commutative ring with a 1 and I be an ideal of R. Then the quotient ring R/I is a commutative ring with a 1.*

Proof: The fact that R/I is a ring is omitted, since verifying the above operations are well-defined is relatively technical and this was all established in *MT3505*. That the multiplication is commutative follows from the fact that the multiplication in R is commutative:

$$(I + r)(I + s) = I + rs = I + sr = (I + s)(I + r) \qquad \text{for all } r, s \in R.$$

The multiplication identity is $I + 1$:

$$(I + r)(I + 1) = I + r1 = I + r \qquad \text{for all } r \in R.$$

□

The other standard bit of terminology that we shall require relating to rings is, of course, the definition of a homomorphism. In the following, I shall use the common habit in algebra of writing maps on the right, so the image of an element a under a map ϕ is written $a\phi$ (rather than $\phi(a)$, as would be common in some other branches of mathematics).

Definition 1.4 Let R and S be commutative rings with 1. A *homomorphism* $\phi\colon R \to S$ is a map such that

(i) $(a+b)\phi = a\phi + b\phi$

(ii) $(ab)\phi = (a\phi)(b\phi)$

for all $a, b \in R$.

Definition 1.5 Let R and S be commutative rings with 1 and $\phi\colon R \to S$ be a homomorphism.

(i) The *kernel* of ϕ is
$$\ker\phi = \{\, a \in R \mid a\phi = 0 \,\}.$$

(ii) The *image* of ϕ is
$$\operatorname{im}\phi = R\phi = \{\, a\phi \mid a \in R \,\}.$$

Theorem 1.6 (First Isomorphism Theorem) *Let R and S be commutative rings with 1 and $\phi\colon R \to S$ be a homomorphism. Then the kernel of ϕ is an ideal of R, the image of ϕ is a subring of S and*
$$R/\ker\phi \cong \operatorname{im}\phi.$$

PROOF: This is a standard result established in *MT3505* (via a proof very similar to that used for groups). The isomorphism is the map given by
$$\theta\colon (\ker\phi) + a \mapsto a\phi$$
for $a \in R$. One must, amongst other things, establish that this is well-defined, in the sense that the image of a coset under θ does not depend upon the choice of representative a for the coset in the quotient ring. □

A final set of ring-theoretic definitions are the following, to which we return at the end of this chapter.

Definition 1.7 Let R be a commutative ring with a 1.

(i) A *zero divisor* in R is a non-zero element a such that $ab = 0$ for some non-zero $b \in R$.

(ii) An *integral domain* is a commutative ring with a 1 containing no zero divisors.

Fields

Galois Theory can be viewed as the study of fields and their subfields. We shall now present the basic facts about such structures.

Definition 1.8 A *field* F is a commutative ring with a 1 such that $0 \neq 1$ and every non-zero element is a *unit*, that is, has a multiplicative inverse.

Thus a field F is a commutative ring with a 1 such that (i) there are non-zero elements and (ii) if $a \in F$ with $a \neq 0$, then there exists some $b \in F$ with $ab = 1$. We shall write a^{-1} or $1/a$ for the multiplicative inverse of a.

Example 1.9 (i) Standard examples of fields familiar from, for example, linear algebra are the fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ of rational numbers, real numbers and complex numbers, respectively.

(ii) If p is a prime number, the set $\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$ forms a field under addition and multiplication modulo p. To see that every non-zero element has a multiplicative inverse, note that if $1 \leqslant x \leqslant p-1$, then x and p are coprime, so there exists $u, v \in \mathbb{Z}$ with $ux + vp = 1$ (exploiting the fact that $\mathbb{Z}$ is a *Euclidean domain*). Hence, $ux \equiv 1 \pmod{p}$ and so, modulo p, u is a multiplicative inverse for x in $\mathbb{F}_p$.

Proposition 1.10 (i) *Every field is an integral domain.*

(ii) *The set of non-zero elements in a field forms an abelian group under multiplication.*

We write F^* for the multiplicative group of non-zero elements in a field.

PROOF: Let F be a field.

(i) If $a, b \in F$ with $a \neq 0$ and $ab = 0$, then $b = a^{-1}(ab) = 0$. Hence if $ab = 0$, either $a = 0$ or $b = 0$, so F contains no zero divisors.

(ii) Write $F^* = F \setminus \{0\}$. Part (i) tells us that F^* is closed under multiplication. The remaining conditions to be an abelian group under this binary operation follow immediately from the definition of a field (multiplication is associative in any ring, it is commutative in any commutative ring, there is a multiplicative identity in any ring with a 1, and in a field every non-zero element has a multiplicative inverse). □

If F is any field, with multiplicative identity denoted by 1, and n is a positive integer, let us define

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

By the distributive law,

$$\overline{mn} = \overline{m}\,\overline{n}$$

for all positive integers m and n. Since F is, in particular, an integral domain, it follows that if there exists a positive integer n such that $\overline{n} = 0$ then necessarily the smallest such positive integer n is a prime number.

Definition 1.11 Let F be a field with multiplicative identity 1.

(i) If it exists, the smallest positive integer p such that $\overline{p} = 0$ is called the *characteristic* of F.

(ii) If no such positive integer exists, we say that F has *characteristic zero*.

Our observation is therefore that every field F either has characteristic zero or has characteristic p for some prime number p. We shall say that K is a *subfield* of F when $K \subseteq F$ and that K forms a field itself under the addition and multiplication induced from F; that is, when the following conditions hold:

(i) K is non-empty and contains non-zero elements (or, equivalently when taken with the other two conditions, $0, 1 \in K$);

(ii) $a + b, -a, ab \in K$ for all $a, b \in K$;

(iii) $1/a \in K$ for all non-zero $a \in K$.

Theorem 1.12 *Let F be a field.*

(i) *If F has characteristic zero, then F has a unique subfield isomorphic to the rationals $\mathbb{Q}$ and this is contained in every subfield of F.*

(ii) *If F has characteristic p (prime), then F has a unique subfield isomorphic to the field $\mathbb{F}_p$ of integers modulo p and this is contained in every subfield of F.*

Definition 1.13 This unique minimal subfield in F is called the *prime subfield* of F.

PROOF: This was proved in *MT3505*. One proves it as follows:

(i) Suppose F has characteristic zero. Extend the notation $\overline{n}$ to all $n \in \mathbb{Z}$ by defining

$$\overline{0} = 0 \qquad \text{and} \qquad \overline{-n} = -\overline{n}$$

for all positive integers n. If K is any subfield of F then K contains 0, 1 and all sums involving 1, so $\overline{n} \in K$ for all $n \in \mathbb{Z}$. Hence

$$Q = \{\, \overline{m}/\overline{n} \mid m, n \in \mathbb{Z},\ n \neq 0 \,\}$$

is a subset of the subfield K.

One now verifies, from the field axioms and the assumption that $\overline{n} \neq 0$ when $n \neq 0$, that the map $n \mapsto \overline{n}$ is a ring homomorphism $\mathbb{Z} \to F$ and then extend this to a ring homomorphism $\mathbb{Q} \to F$ given by $m/n \mapsto \overline{m}/\overline{n}$. We conclude that Q is a subfield of F that is isomorphic to the field $\mathbb{Q}$ of rational numbers and is contained in every subfield K of F.

Finally, uniqueness of Q follows from the minimality condition: if Q_1 and Q_2 were subfields contained in every subfield of F then, in particular, $Q_1 \subseteq Q_2$ and $Q_2 \subseteq Q_1$, from which we deduce $Q_1 = Q_2$.

(ii) Use a similar argument to part (i). If F has characteristic p (prime) and K is any subfield of F, then K contains all the elements $\overline{n}$; that is,

$$P = \{0, 1, \overline{2}, \overline{3}, \ldots, \overline{p-1}\} \subseteq K.$$

Now observe that P is closed under addition and multiplication and the map $n \mapsto \overline{n}$ is an isomorphism from the field $\mathbb{F}_p$ of p elements to P. □

Polynomials

Polynomials arise in a number of places within Galois Theory. The motivation of the subject arises in the problem of solving polynomial equations. More significantly, algebraic elements, those arising as roots of polynomial equations, will be of great importance in our field extensions as discussed in Chapter 2.

Definition 1.14 Let F be a field. A *polynomial* over F in the indeterminate X is an expression of the form

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

where n is a non-negative integer and the coefficients a_0, a_1, …, a_n are elements of F.

We shall often substitute elements of a field for the indeterminate in a polynomial. Thus if α is an element of the field F and $f(X) = a_0 + a_1X + \cdots + a_nX^n$ where the coefficients are also elements in F, we write $f(\alpha)$ for the expression

$$f(\alpha) = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n.$$

We shall write $F[X]$ for the set of all polynomials in the indeterminate X with coefficients taken from the field F. We add two such polynomials by simply adding the coefficients,

$$\sum a_iX^i + \sum b_iX^i = \sum (a_i + b_i)X^i,$$

and we multiply two polynomials by exploiting the distributive law:

$$\left(\sum a_i X^i\right)\left(\sum b_i X^i\right) = \sum c_i X^i$$

where $c_k = \sum_{i=1}^{k} a_i b_{k-i}$. With these definitions, one deduces in a straightforward way that $F[X]$ forms a commutative ring with a 1, namely the multiplicative identity is the constant polynomial 1. If $f(X)$ has a non-zero term $a_n X^n$ of highest degree (that is, all other terms in $f(X)$ has the form $a_i X^i$ with $i < n$) and $g(X)$ has a non-zero term $b_m X^m$ of highest degree, then the term of highest degree in $f(X)g(X)$ is $a_n b_m X^{n+m}$ and this is non-zero since F is a field so $a_n b_m \neq 0$. Therefore $F[X]$ is actually an integral domain since $f(X), g(X) \neq 0$ implies $f(X)g(X) \neq 0$.

Proposition 1.15 *If F is a field, the polynomial ring $F[X]$ is a* Euclidean domain.

The Euclidean function associated to $F[X]$ is the degree of a polynomial. Recall that if $f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0$ is a *non-zero* polynomial with leading term having non-zero coefficient, that is, $a_n \neq 0$, the *degree* of $f(X)$ is

$$\deg f(X) = n.$$

The properties of the degree are:

(i) if $f(X)$ and $g(X)$ are non-zero, then $\deg f(X)g(X) = \deg f(X) + \deg g(X)$;

(ii) if $f(X)$ and $g(X)$ are polynomials with $f(X) \neq 0$, then there exist unique polynomials $q(X)$ and $r(X)$ satisfying

$$g(X) = q(X)\, f(X) + r(X) \qquad \text{with either } r(X) = 0 \text{ or } \deg r(X) < \deg f(X).$$

These properties were established in *MT3505*. They are not verified in this module, but will be assumed, and are what is claimed within Proposition 1.15.

As a consequence, all the properties of Euclidean domains established in *MT3505* apply to a polynomial ring $F[X]$ over a field F. For example:

Proposition 1.16 *If F is a field, the polynomial ring $F[X]$ is a principal ideal domain; that is, every ideal I in $F[X]$ has the form $I = (f(X)) = \{\, f(X)g(X) \mid g(X) \in F[X] \,\}$ for some polynomial $f(X)$.*

Proof: Let I be an ideal of $F[X]$. If $I = \{0\}$, then $I = (0)$. Suppose that $I \neq \{0\}$. Let $f(X)$ be a polynomial in I such that $\deg f(X)$ is as small as possible among the degrees of non-zero polynomials in I. Certainly $(f(X)) \subseteq I$, since I is closed under multiplication by any polynomial.

Now if $g(X) \in I$, divide $f(X)$ to obtain a quotient and remainder:

$$g(X) = q(X)\, f(X) + r(X)$$

where either $r(X) = 0$ or $\deg r(X) < \deg f(X)$. Then $r(X) = g(X) - q(X)\, f(X)$ belongs to I, since I is an ideal. By the assumption about $f(X)$ having smallest degree amongst polynomials in I, we conclude $r(X) = 0$. Hence $g(X)$ is indeed a multiple of $f(X)$. This establishes $I = (f(X))$, as claimed. □

Another fact that holds as a consequence of Proposition 1.15 concerns the greatest common divisor of a pair (or more) of polynomials.

Definition 1.17 Suppose $f(X)$ and $g(X)$ are polynomials over the field F. A *greatest common divisor* of $f(X)$ and $g(X)$ is a polynomial $h(X)$ of greatest degree such that $h(X)$ divides both $f(X)$ and $g(X)$.

To say that $h(X)$ *divides* $f(X)$ means that $f(X)$ is a multiple of $h(X)$; that is, $f(X) = h(X)\,q(X)$ for some $q(X) \in F[X]$. In a general Euclidean domain, the greatest common divisor is defined uniquely up to multiplication by a unit. In the polynomial ring $F[X]$, the units are constant polynomials (that is, elements of the base field F viewed as elements of $F[X]$). This follows from the first property of degrees listed: $\deg f(X)g(X) = \deg f(X) + \deg g(X)$, so the only way that $f(X)g(X) = 1$ can hold is if $\deg f(X) = 0$. As a consequence, the greatest common divisor of a pair of polynomials is defined uniquely up to multiplication by a scalar from the field F.

We shall need at times the following fact about the form of the greatest common divisor in a Euclidean domain. This result is often accompanied with a practical algorithm for computing the greatest common divisor, but that will not be as important in this module. The proof of the result can be found in *MT3505*.

Theorem 1.18 *Let F be a field and $f(X)$ and $g(X)$ be two non-zero polynomials over F. Then there exist $u(X), v(X) \in F[X]$ such that the greatest common divisor of $f(X)$ and $g(X)$ is given by*

$$h(X) = u(X)\,f(X) + v(x)\,g(X).$$

Another standard fact about a Euclidean domain is that it is necessarily a unique factorization domain. Consequently, every polynomial over F can be factorized as a product of irreducible polynomials and these irreducible factors are determined uniquely up to multiplication by scalars (as the units are the constant polynomials) and reordering of the factors (of course, since multiplication is commutative). In view of this, we record the definition of what it means for a polynomial to be irreducible, since that will be a particularly significant concept in terms of what follows.

Definition 1.19 Let $f(X)$ be a polynomial over a field F of degree at least 1. We say that $f(X)$ is *irreducible* over F if it cannot be factorized as $f(X) = g_1(X)\,g_2(X)$ where $g_1(X)$ and $g_2(X)$ are polynomials in $F[X]$ of degree smaller than $f(X)$.

Thus, a polynomial $f(X)$ is irreducible if the only polynomials of degree smaller than $f(X)$ that divide it are the constant polynomials (i.e., the units). The term *reducible* is used for a polynomial that is not irreducible; that is, that can be factorized as a product of two polynomials of smaller degree. It will be important to be able to show that certain polynomials are irreducible. In general, this is rather difficult to achieve and one generally needs to rely upon *ad hoc* methods, particularly over fields of characteristic $p > 0$.

We first make the observation that the concept of irreducibility depends heavily upon the field over which we are working. After that we shall consider various examples of methods for showing polynomials are irreducible.

Example 1.20 Consider the polynomial $f(X) = X^2 + 1$. If we view $f(X)$ as a polynomial over the real numbers $\mathbb{R}$, then it is irreducible: if it were to factorize then it would be a product of two linear factors. However, the roots of this polynomial do not exist in the real numbers, so $f(X)$ has no roots in $\mathbb{R}$ and hence is irreducible over $\mathbb{R}$. However, when viewed as a polynomial over $\mathbb{C}$, it is reducible:

$$X^2 + 1 = (X - i)(X + i)$$

Similary $g(X) = X^2 - 2$ is irreducible over $\mathbb{Q}$ (since it has no roots in $\mathbb{Q}$, so can have no linear factors), but is reducible over $\mathbb{R}$ (since it factorizes over this field as $X^2 - 2 = (X - \sqrt{2})(X + \sqrt{2})$.

Example 1.21 *Show that the following polynomials are irreducible over the given fields:*

(i) $f(X) = X^2 + X + 1$ *over* $\mathbb{F}_2$*;*

(ii) $g(X) = X^3 + 2X + 1$ *over* $\mathbb{F}_3$*;*

(iii) $h(X) = X^4 + X + 1$ *over* $\mathbb{F}_2$.

SOLUTION: (i) If $f(X) = X^2 + X + 1$ factorizes into two polynomials of smaller degree over $\mathbb{F}_2$, then it is a product of two linear factors and hence would have a root in $\mathbb{F}_2$. Observe $f(0) = f(1) = 1$. Hence $f(X)$ has no roots in $\mathbb{F}_2$ and is therefore irreducible.

(ii) If $g(X) = X^3 + 2X + 1$ factorizes into two polynomials of smaller degree over $\mathbb{F}_3$, then one of these factors would be linear and hence $g(X)$ would have a root in $\mathbb{F}_3$. Observe $g(0) = g(1) = g(2) = 1$. Hence $g(X)$ has no roots in $\mathbb{F}_3$ and is therefore irreducible.

(iii) First note that $h(0) = h(1) = 1$, so $h(X)$ has no roots in $\mathbb{F}_2$ and hence has no linear factors. As a consequence, if $h(X)$ is reducible, then it is a product of two irreducible quadratic factors. There are four quadratic polynomials over $\mathbb{F}_2$:

$$X^2, \quad X^2 + 1, \quad X^2 + X, \qquad X^2 + X + 1.$$

The first three are reducible, having either 0 or 1 (or both in the case of the third) as roots. We conclude that $X^2 + X + 1$ is the only irreducible quadratic polynomial over $\mathbb{F}_2$. Note that

$$(X^2 + X + 1)^2 = X^4 + X^2 + 1$$

(since $\mathbb{F}_2$ has characteristic 2) and therefore $h(X) = X^4 + X + 1$ is not a product of $X^2 + X + 1$ with itself.

We conclude that $h(X)$ is indeed irreducible over $\mathbb{F}_2$. □

If one works over the field $\mathbb{Q}$ of rational numbers, then there are several methods that are useful for determining that a polynomial $f(X)$ is irreducible. First note that we can multiply by the lowest common multiple of the denominators of the coefficients in $f(X)$ and obtain a scalar multiple of $f(X)$ that happens to have all its coefficients being integers. In view of this, we shall discuss polynomials with integer coefficients and ask whether they are irreducible as polynomials over $\mathbb{Q}$. The first step is to note that it is sufficient to show that such a polynomial cannot be factorized into two polynomials with integer coefficients.

Theorem 1.22 (Gauss's Lemma) *Let $f(X)$ be a polynomial with integer coefficients. Then $f(X)$ is irreducible over $\mathbb{Z}$ if and only if it is irreducible over $\mathbb{Q}$.*

PROOF: Let $f(X) \in \mathbb{Z}[X]$. Note that if $f(X) = g_1(X)\, g_2(X)$ where $g_1(X), g_2(X) \in \mathbb{Z}[X]$ and $\deg g_1(X), \deg g_2(X) < \deg f(X)$, then this is also a factorization over $\mathbb{Q}$ (essentially because $\mathbb{Z}[X] \subseteq \mathbb{Q}[X]$). Hence, taking the contrapositive, if $f(X)$ is irreducible over $\mathbb{Q}$ then it is irreducible over $\mathbb{Z}$.

Conversely, suppose $f(X) = g_1(X)\, g_2(X)$ where $g_1(X), g_2(X) \in \mathbb{Q}[X]$ with degrees satisfying $\deg g_1(X), \deg g_2(X) < \deg f(X)$. Consider the denominators of the coefficients appearing in the polynomials $g_1(X)$ and $g_2(X)$. Multiply through by the lowest common multiple of the denominators of the coefficients of $g_1(X)$ and by that for $g_2(X)$. Hence we find a positive integer n such that the expression

$$n\, f(X) = \bar{g}_1(X)\, \bar{g}_2(X) \tag{1.1}$$

holds, where $\bar{g}_1(X), \bar{g}_2(X) \in \mathbb{Z}[X]$, $\deg \bar{g}_1(X) = \deg g_1(X) < \deg f(X)$ and $\deg \bar{g}_2(X) = \deg g_2(X) < \deg f(X)$. Among all such expressions, choose n to be the smallest positive integer such that we can factorize $n\, f(X)$ as in Equation (1.1).

We claim that $n = 1$. If not, choose a prime number p that divides n. Then p divides the product $\bar{g}_1(X)\,\bar{g}_2(X)$. Suppose

$$\bar{g}_1(X) = a_0 + a_1X + \cdots + a_mX^m \qquad \text{and} \qquad \bar{g}_2(X) = b_0 + b_1X + \cdots + b_nX^n$$

for some coefficients $a_i, b_j \in \mathbb{Z}$. We claim that p either divides all the a_i or divides all the b_j. If not, we can assume that p divides a_0, a_1, …, a_{k-1} but does not divide a_k and that p divides b_0, b_1, …, $b_{\ell-1}$ but does not divide b_ℓ. Consider the coefficient $c_{k+\ell}$ of $X^{k+\ell}$ in the product $\bar{g}_1(X)\,\bar{g}_2(X)$:

$$c_{k+\ell} = a_0b_{k+\ell} + \cdots + a_{k-1}b_{\ell+1} + a_kb_\ell + a_{k+1}b_{\ell-1} + \cdots + a_{k+\ell}b_0.$$

We know that p divides $c_{k+\ell}$ and that it divides a_0, a_1, …, a_{k-1}, b_0, …, $b_{\ell-1}$. Hence p divides the product a_kb_ℓ and therefore p divides either a_k or b_ℓ (since p is a prime number). This contradicts our assumption.

We conclude that either p divides all the coefficients of $\bar{g}_1(X)$ or of $\bar{g}_2(X)$. Let us assume the former. Define $\bar{\bar{g}}_1(X)$ to equal $\frac{1}{p}\bar{g}_1(X)$, which we now know is a polynomial with integer coefficients. We may therefore divide Equation (1.1) by p to obtain

$$\tfrac{n}{p}\, f(X) = \bar{\bar{g}}_1(X)\,\bar{g}_2(X).$$

This contradicts the choice of n to be minimal. We conclude therefore that $n = 1$ and hence that $f(X)$ is factorizable as a product of two polynomials of smaller degree over $\mathbb{Z}$. This completes the proof of Gauss's Lemma. □

Theorem 1.23 (Eisenstein's Irreducibility Criterion) *Let*

$$f(X) = a_0 + a_1X + \cdots + a_nX^n$$

be a polynomial over $\mathbb{Z}$. Suppose there exists a prime number p such that

(i) *p does not divide a_n;*

(ii) *p divides a_0, a_1, …, a_{n-1};*

(iii) *p^2 does not divide a_0.*

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

Proof: In view of Gauss's Lemma, it is sufficient to show that $f(X)$ cannot be factorized into two polynomials of smaller degree over $\mathbb{Z}$. Suppose that $f(X) = g_1(X)\,g_2(X)$, where

$$g_1(X) = b_0 + b_1X + \cdots + b_kX^k \qquad \text{and} \qquad g_2(X) = c_0 + c_1X + \cdots + c_\ell X^\ell$$

and where $b_0, b_1, \ldots, b_k, c_0, c_1, \ldots, c_\ell \in \mathbb{Z}$. Note $a_0 = b_0c_0$, so by hypotheses (ii) and (iii), p divides one of b_0 or c_0, but not both. Let us suppose, without loss of generality, that p divides b_0 but not c_0. Now since p does not divides a_n, it cannot be the case that p divides all the coefficients of $g_1(X)$. Hence there exists some coefficient b_i such that p divides b_0, b_1, …, b_{i-1}, but p does not divide b_i. Note $i \leqslant k = \deg g_1(X) < n$. Then, by hypothesis (ii),

$$a_i = b_0c_i + b_1c_{i-1} + \cdots + b_{i-1}c_1 + b_ic_0$$

is divisible by p and it follows that p divides the last term b_ic_0 (as p divides b_0, b_1, …, b_{i-1}). However, p divides neither of b_i or c_0, so this is impossible.

This contradiction establishes that $f(X)$ is indeed irreducible over $\mathbb{Q}$. □

Example 1.24 *Show that the following polynomials are irreducible over $\mathbb{Q}$:*

(i) $X^n - p$, *for any prime number* p;

(ii) $\frac{2}{9}X^5 + \frac{5}{3}X^4 + X^3 + \frac{1}{3}$;

(iii) $X^{p-1} + X^{p-2} + \cdots + X^2 + X + 1$, *for any prime number* p.

SOLUTION: (i) $X^n - p$ is irreducible by Eisenstein's Criterion: it is of the form to apply Theorem 1.23.

(ii) If $f(X) = \frac{2}{9}X^5 + \frac{5}{3}X^4 + X^3 + \frac{1}{3}$, then

$$9f(X) = 2X^5 + 15X^4 + 9X^3 + 3.$$

Thus $9f(X)$ is irreducible over $\mathbb{Q}$ by Eisenstein's Criterion (using the prime $p = 3$), so cannot be factorized as a product of polynomials of smaller degree over $\mathbb{Q}$. The same therefore applies to $f(X)$. (Note that 9 is a unit in $\mathbb{Q}[X]$.)

(iii) Write $\Phi(X) = X^{p-1} + X^{p-2} + \cdots + X^2 + X + 1$. Suppose $\Phi(X)$ can be factorized as a product of polynomials of smaller degree over $\mathbb{Q}$; say, $\Phi(X) = g(X)\,h(X)$. Note

$$(X-1)\cdot\Phi(X) = (X-1)(X^{p-1} + X^{p-2} + \cdots + X + 1) = X^p - 1.$$

Substitute $Y = X - 1$:

$$Y \cdot \Phi(Y+1) = (Y+1)^p - 1 = \sum_{i=1}^{p} \binom{p}{i} Y^i.$$

Hence

$$\begin{aligned}\Phi(Y+1) &= \sum_{i=1}^{p} \binom{p}{i} Y^{i-1} \\ &= Y^{p-1} + \binom{p}{p-1} Y^{p-2} + \binom{p}{p-2} Y^{p-3} + \cdots + \binom{p}{2} Y + \binom{p}{1}.\end{aligned}$$

The constant coefficient in $\Phi(Y+1)$ is $\binom{p}{1} = p$, which is divisible by p but not p^2. Note that, for $i = 1, 2, \ldots, p-1$,

$$\binom{p}{i} = \frac{p!}{i!\,(p-i)!} = \frac{p(p-1)(p-2)\ldots(p-i+1)}{i!}$$

and we know this is an integer. Note that the prime p is bigger than all the factors in $i!$ (by assumption on i), so

$$\textit{the binomial coefficient } \binom{p}{i} \textit{ is divisible by the prime } p \textit{ for } i = 1, 2, \ldots, p-1. \tag{1.2}$$

Hence we may apply Eisenstein's Criterion to $\Phi(Y+1)$ to conclude that $\Phi(Y+1)$ is irreducible as a polynomial in Y. Our original assumption, however, implies that $\Phi(Y+1) = g(Y+1)\,h(Y+1)$, which is a contradiction.

Hence $\Phi(X)$ is indeed irreducible over $\mathbb{Q}$. □

There is one final method that we mention for showing a polynomial (with integer coefficients) is irreducible is to reduce the coefficients modulo some prime p. The choice of prime p is usually delicate: the type of argument presented when it works, typically does for some choices of prime but not others.

Example 1.25 *Show that the polynomial* $f(X) = X^4 + 8X^3 + 9X^2 + 6X + 5$ *is irreducible over* $\mathbb{Q}$.

SOLUTION: Suppose that $f(X)$ is reducible over $\mathbb{Q}$. Then $f(X)$ is reducible over $\mathbb{Z}$, by Gauss's Lemma, so factorizes as a product of polynomials with integer coefficients of smaller degree.

We shall reduce all the coefficients modulo 3. To be more precise, there is a ring homomorphism $\phi\colon \mathbb{Z} \to \mathbb{F}_3$ that arises by reducing an integer modulo 3. The kernel of ϕ is the ideal (3) of all multiples of 3. We induce a map $\bar{\phi}\colon \mathbb{Z}[X] \to \mathbb{F}_3[X]$ by applying ϕ to the coefficients in a polynomial; that is, reducing the coefficients modulo 3. Since the coefficients of sums of polynomials and products of polynomials are determined by operations in the base ring/field, it follows, from the fact that ϕ is a homomorphism, the induced map $\bar{\phi}$ is a ring homomorphism $\mathbb{Z}[X] \to \mathbb{F}_3[X]$. Applying $\bar{\phi}$ to the factorization of $f(X)$ as a product of two polynomials from $\mathbb{Z}[X]$, we conclude that

$$\bar{f}(X) = f(X)\bar{\phi} = X^4 + 2X^3 + 2$$

factorizes; that is, $\bar{f}(X)$ is reducible over $\mathbb{F}_3$. We shall show this is impossible.

First note

$$\bar{f}(0) = 2, \qquad \bar{f}(1) = 2, \qquad \bar{f}(2) = 2^4 + 2^4 + 2 = 1,$$

so that $\bar{f}(X)$ does not have any roots in $\mathbb{F}_3$ and hence has no linear factors. Therefore $\bar{f}(X)$ must be a product of quadratic factors:

$$X^4 + 2X^3 + 2 = (X^2 + aX + b)(X^2 + cX + d)$$

for some coefficients $a, b, c, d \in \mathbb{F}_3$. Equating coefficients:

$$\begin{aligned} a + c &= 2, & ac + b + d &= 0, \\ ad + bc &= 0, & bd &= 2 \end{aligned}$$

The constant coefficient tells us that either $b = 1$ and $d = 2$, or $b = 2$ and $d = 1$. Thus $b + d = 0$. The degree 2 coefficient then tells us $ac = 0$, so either $a = 0$ or $c = 0$. The degree 1 coefficient then tells us that the other of a or c is also zero. Then $a + c = 0$, so the degree 3 coefficient equation fails.

In conclusion, $\bar{f}(X)$ is irreducible over $\mathbb{F}_3$ and we then deduce the original assumption about $f(X)$ was incorrect. Hence $f(X)$ is indeed irreducible over $\mathbb{Q}$. □

The final fact about integral domains, that we shall apply to the polynomial ring $F[X]$, is that every integral domain has a field of fractions. To be precise, if R is an integral domain, the field of fractions of R is the set of all expressions of the form r/s where (i) $r, s \in R$ with $s \neq 0$, and (ii) we define $r_1/s_1 = r_2/s_2$ if and only if $r_1 s_2 = r_2 s_1$. (The latter condition defines an equivalence relation on ordered pairs (r, s) with $s \neq 0$ and we write r/s for the equivalence containing the ordered pair (r, s) under this equivalence relation.) Mirroring the definition of addition and multiplication on the rational numbers, we define

$$\frac{r_1}{s_1} + \frac{r_2}{s_2} = \frac{r_1 s_2 + r_2 s_1}{s_1 s_2} \qquad \text{and} \qquad \frac{r_1}{s_1} \cdot \frac{r_2}{s_2} = \frac{r_1 r_2}{s_1 s_2}$$

for such fractions r_1/s_1 and r_2/s_2. One verifies that the set of all such fractions forms a commutative ring under this operation and that every non-zero fraction r/s (that is, when both r and s are non-zero) has a multiplicative inverse, namely s/r, because

$$\frac{r}{s} \cdot \frac{s}{r} = \frac{rs}{rs} = \frac{1}{1}$$

from the definition of when two fractions are equal. The latter is the multiplicative identity in the field of fractions. Notice finally that R embeds in the field of fractions via the map $r \mapsto r/1$; that is, the set $\{\, r/1 \mid r \in R \,\}$ is a subring isomorphic to the original integral domain R.

Let us apply this construction in the case when $R = F[X]$, the integral domain of polynomials with coefficients from the field F. We consequently construct the following object:

Definition 1.26 Let F be a field. The *field of rational functions* with coefficients in F is denoted by $F(X)$ and is the field of fractions of the polynomial ring $F[X]$.

The elements of $F(X)$ are expressions of the form

$$\frac{f(X)}{g(X)}$$

where $f(X)$ and $g(X)$ are polynomials with coefficients from F. Equality of two such expressions is given by

$$\frac{f_1(X)}{g_1(X)} = \frac{f_2(X)}{g_2(X)} \qquad \text{if and only if} \qquad f_1(X)\, g_2(X) = f_2(X)\, g_1(X).$$

If one exploits the fact that $F[X]$ is a unique factorization domain, one can deduce from this that $f_2(X)$ and $g_2(X)$ are obtained from $f_1(X)$ and $g_1(X)$ via cancelling and then multiplying by some common factors. Addition in $F(X)$ is achieved by placing a pair of fractions over a common denominator. The polynomial $f(X)$ is identified with its image $f(X)/1$ in the field $F(X)$. Thus, we view the polynomial ring $F[X]$ as a subring of the field $F(X)$ of rational functions. In particular, since the constant polynomials form a copy of F, we observe:

Proposition 1.27 *The field F occurs as a subfield of the field $F(X)$ of rational functions with coefficients in F.*