Chapter 1

Introduction

Finite fields is a branch of mathematics which has come to the fore in the last 50 years due to its numerous applications, from combinatorics to coding theory. In this course, we will study the properties of finite fields, and gain experience in working with them.

In the first two chapters, we explore the theory of fields in general. Throughout, we emphasize results particularly important to finite fields, but allow fields to be arbitrary unless otherwise stated.

1 Group theory: a brief summary

We begin by recalling the definition of a group.

Definition 1.1
A *group* is a set G, together with a binary operation $*$, such that the following axioms hold:

Closure: G is closed under the operation $*$: $x, y \in G \Longrightarrow x * y \in G$;

Associativity: $(x * y) * z = x * (y * z)$ for all $x, y, z \in G$;

Identity: there exists an element $e \in G$ (called the identity of G) such that $x * e = e * x = x$ for all $x \in G$;

Inverses: for every element $x \in G$ there exists an element $x^{-1} \in G$ (called the inverse of x) such that $x * x^{-1} = x^{-1} * x = e$.

Note: We often write $\cdot$ instead of $*$ or leave it out completely.

Definition 1.2
A group G is said to be *abelian* if the binary operation $*$ is commutative, i.e. if $x * y = y * x$ for all $x, y \in G$. The operation $*$ is often replaced by $+$ for abelian groups, i.e. $x * y$ is written $x + y$
We then say that the group is “written additively” (as opposed to being “written multiplicatively”).

Example 1.3
The following are examples of groups:

- The set $\mathbb{Z}$ of integers with the operation of addition (this is abelian);
- the set $\mathrm{GL}_n(\mathbb{R})$ of invertible $(n \times n)$-matrices with real entries, with the operation of matrix multiplication, forms a group (for $n > 1$ this is not abelian).
- Let G be the set of remainders of all the integers on division by n, e.g. $G = \{0, 1, \ldots, n-1\}$. Let $a * b$ be the operation of taking the integer sum $a + b$ and reducing it modulo n. Then $(G, *)$ is a group (this is abelian).

Definition 1.4
A multiplicative group G is said to be *cyclic* if there is an element $a \in G$ such that for any $b \in G$ there is some integer j with $b = a^j$. Such an element is called a generator of the cyclic group, and we write $G = \langle a \rangle$.

Note we may have more than one generator, e.g. either 1 or -1 can be used to generate the additive group $\mathbb{Z}$.

Definition 1.5
For a set S, a subset R of $S \times S$ is called an *equivalence relation* on S if it satisfies:

- $(s, s) \in R$ for all $s \in S$ (reflexive)
- If $(s, t) \in R$ then $(t, s) \in R$ (symmetric)
- If $(s, t), (t, u) \in R$ then $(s, u) \in R$ (transitive).

An equivalence relation R on S induces a *partition* of S. If we collect all elements of S equivalent to a fixed $s \in S$, we obtain the *equivalence class* of s, denoted by

$$[s] = \{t \in S : (s, t) \in R\}.$$

The collection of all equivalence classes forms a partition of S, and $[s] = [t] \Leftrightarrow (s, t) \in R$.

Definition 1.6
For arbitrary integers a,b and positive integer n, we say that a is *congruent* to b modulo n if the difference $a - b$ is a multiple of n, i.e. $a = b + kn$ for some integer k. We write $a \equiv b \bmod n$ for this.

It is easily checked that "congruence modulo n" is an equivalence relation on the set $\mathbb{Z}$ of integers. Consider the equivalence classes into which the relation partitions $\mathbb{Z}$. These are the sets:

$$\begin{aligned} [a] &= \{m \in \mathbb{Z} : m \equiv a \bmod n\} \\ &= \{m \in \mathbb{Z} : m = a + kn \text{ for some } k \in \mathbb{Z}\}. \end{aligned}$$

E.g. for $n = 4$ we have:
$[0] = \{\ldots, -8, -4, 0, 4, 8, \ldots\}$;
$[1] = \{\ldots, -7, -3, 1, 5, 9, \ldots\}$;
$[2] = \{\ldots, -6, -2, 2, 6, 10, \ldots\}$;
$[3] = \{\ldots, -5, -1, 3, 7, 11, \ldots\}$.
We may define on the set $\{[0], [1], \ldots, [n-1]\}$ a binary operation, which we shall write as $+$ (though it is not ordinary addition) by

$$[a] + [b] := [a + b]$$

where a and b are any elements of the sets $[a]$ and $[b]$ respectively, and $a + b$ is the ordinary sum of a and b. Can show (exercise) that this is well-defined, i.e. does not depend on choice of representatives.

Theorem 1.7
Let $n \in \mathbb{N}$. The set $\{[0], [1], \ldots, [n - 1]\}$ of equivalence classes modulo n forms a group under the operation $+$ given by $[a] + [b] := [a + b]$. It is called the group of integers modulo n and is denoted $\mathbb{Z}_n$. It is cyclic with $[1]$ as a generator.

Proof. See Exercise Sheet 1. ■

Definition 1.8
A group is called *finite* (respectively, *infinite*) if it contains finitely (respectively, infinitely) many elements. The number of elements of a finite group G is called its order, written $|G|$.

Definition 1.9
A subset H of the group G is a *subgroup* of G if H is itself a group with respect to the operation of G, this is written $H \leq G$. The (cyclic) subgroup consisting of all powers of some element $a \in G$ is denoted $\langle a \rangle$ and called the subgroup *generated by* a. If $| \langle a \rangle |$ is finite, it is called the *order of* a, it is the smallest natural number i such that $a^i = e$.

Next, we generalize the notion of congruence, as follows.

Theorem 1.10
If H is a subgroup of G, then the relation R_H on G defined by $(a, b) \in R_H$ if and only if $a = bh$ for some $h \in H$ (additively, $a = b + h$ for some $h \in H$) is an equivalence relation. The relation is called left congruence modulo H.

The equivalence classes are called the *left cosets of H in G*; each has size $|H|$. Right congruence and right cosets are defined analogously.

Note that when $(G, *) = (\mathbb{Z}, +)$ and $H = \langle n \rangle$, we get back our previous definition of congruence, since $a \equiv b \bmod n \Leftrightarrow a = b + h$ for some $h \in \langle n \rangle$.

Definition 1.11
The *index* of H in G (denoted by $[G : H]$) is the number of left cosets of H in G, and is equal to the number of right cosets of H in G.

Theorem 1.12
The order of a finite group G is equal to the product of the order of any subgroup H and the index of H in G. In particular, the order of H divides the order of G and the order of any element $a \in G$ divides the order of G.

Proof. Exercise ■

We can easily describe subgroups and orders for cyclic groups. In what follows, ϕ is Euler's function; i.e. $\phi(n) :=$ the number of integers k with $1 \leq k \leq n$ which are relatively prime to n. If the integer n has the prime factorization $p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$, then

$$\phi(n) = n(1 - \frac{1}{p_1})(1 - \frac{1}{p_2}) \cdots (1 - \frac{1}{p_r}).$$

So, for example, $\phi(7) = 6$ and $\phi(30) = 2 \cdot 3 \cdot 5 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{4}{5} = 8$. See Example Sheet for more details.

Theorem 1.13

(i) *Every subgroup S of a cyclic group $G = \langle a \rangle$ is cyclic.*

(ii) *In a finite cyclic group $\langle a \rangle$ of order m, the element a^k generates a subgroup of order $\frac{m}{\gcd(k,m)}$.*

(iii) *For any positive divisor d of m, $\langle a \rangle$ contains precisely one subgroup of order d and precisely one subgroup of index d.*

(iv) *Let f be a positive divisor of m. Then $\langle a \rangle$ contains $\phi(f)$ elements of order f.*

(v) *A finite cyclic group $\langle a \rangle$ of order m contains $\phi(m)$ generators, namely the powers a^r with* $\gcd(r, m) = 1$.

Proof.

(i) If $S = \{e\}$, then S is cyclic with generator e. Otherwise, let k be the least positive integer for which $a^k \in S$. We will show: $S = \langle a^k \rangle$. Clearly $\langle a^k \rangle \subseteq S$. Now, take an arbitrary $s \in S$, then $s = a^n$ for some $n \in \mathbb{Z}$. By the division algorithm for integers, there exist $q, r \in \mathbb{Z}$ with $0 \leq r < k$ such that $n = qk + r$. Then $a^n = a^{qk+r} = (a^k)^q \cdot a^r$, implying $a^r \in S$. If $r > 0$, this contradicts the minimality of k, so we must have $r = 0$ and hence $s = a^n = (a^k)^q \in \langle a^k \rangle$.

(ii) Set $d := \gcd(k, m)$. The order of a^k is the least positive integer n such that $a^{kn} = e$. This identity holds if and only if m divides kn, i.e. if and only if $\frac{m}{d}$ divides n. The least positive n with this property is $n = \frac{m}{d}$.

(iii) Exercise: see Exercise Sheet 1.

(iv) Let $| \langle a \rangle | = m$ and $m = df$. By (ii), the element a^k is of order f if and only if $\gcd(k, m) = d$. So the number of elements of order f is equal to the number of integers k with $1 \leq k \leq m$ and $\gcd(k, m) = d$. Equivalently, writing $k = dh$ with $1 \leq h \leq f$, the condition becomes $\gcd(h, f) = 1$. There are precisely $\phi(f)$ such h.

(v) The first part follows from (iv), since the generators of $\langle a \rangle$ are precisely the elements of order m. The second part follows from (ii).

■

Definition 1.14

- A subgroup H of G is *normal* ⇔ its left and right cosets coincide.

  We write $H \lhd G$ in that case.

- For a normal subgroup H, the set of (left) cosets of H in G forms a group, denoted G/H. The operation is

$$(aH)(bH) := (ab)H.$$

Definition 1.15

A mapping $f : G \to H$ of the group G into the group H is called a *homomorphism* of G into H if f preserves the operation of G, i.e. $(gk)f = (gf) \cdot (kf)$ for all $g, k \in G$. If f is a bijective homomorphism it is called an *isomorphism* and we say G and H are isomorphic and write $G \cong H$. An isomorphism of G onto itself is called an *automorphism* of G.

Definition 1.16

The *kernel* of the homomorphism $f : G \to H$ of the group G into the group H is the set (actually, normal subgroup)

$$\ker(f) := \{a \in G : af = e_H\}.$$

The image of f is the set (actually, subgroup)

$$\mathrm{im}(f) := \{af : a \in G\}.$$

Theorem 1.17

[First Isomorphism Theorem] Let $f : G \to H$ be a homomorphism of groups. Then $\ker f$ is a normal subgroup of G and

$$G/\ker f \cong \mathrm{im} f \quad \textit{by the isomorphism } g \ker(f) \mapsto gf.$$

Proof. Omitted. ■

Example 1.18

Take $G := \mathbb{Z}$, $H := \mathbb{Z}_n$ and $f : a \mapsto [a]$. Then f is a homomorphism with $\ker(f) = \langle n \rangle$ and $\mathrm{im}(f) = \mathbb{Z}_n$, and so the First Isomorphism Theorem says that $\mathbb{Z}/\langle n \rangle$ and $\mathbb{Z}_n$ are isomorphic as groups.

2 Rings and fields

Definition 2.1
A *ring* $(R, +, *)$ is a set R, together with two binary operations, denoted by $+$ and $*$, such that

- R is an *abelian group* with respect to $+$;
- R is closed under $*$;
- $*$ is *associative*, that is $(a * b) * c = a * (b * c)$ for all $a, b, c \in R$;
- the *distributive laws* hold, that is, for all $a, b, c \in R$ we have $a * (b + c) = (a * b) + (a * c)$ and $(b + c) * a = (b * a) + (c * a)$.

Typically, we use 0 to denote the identity element of the abelian group R with respect to addition, and $-a$ to denote the additive inverse of $a \in R$.

Definition 2.2

- A ring is called a *ring with identity* if the ring has a multiplicative identity (usually denoted e or 1).
- A ring is called *commutative* if $*$ is commutative.
- A ring is called an *integral domain* if it is a commutative ring with identity $e \neq 0$ in which $ab = 0$ implies $a = 0$ or $b = 0$ (i.e. *no zero divisors*).
- A ring is called a *division ring* (or skew field) if the non-zero elements form a group under $*$.
- A commutative division ring is called a *field*.

Example 2.3

- the integers $(\mathbb{Z}, +, *)$ form an integral domain but not a field;
- the rationals $(\mathbb{Q}, +, *)$, reals $(\mathbb{R}, +, *)$ and complex numbers $(\mathbb{C}, +, *)$ form fields;
- the set of 2×2 matrices with real entries forms a non-commutative ring with identity w.r.t. matrix addition and multiplication.
- the group $\mathbb{Z}_n$ with addition as before and multiplication defined by $[a][b] := [ab]$ is a commutative ring with identity $[1]$.

So, in summary: a field is a set F on which two binary operations, called addition and multiplication, are defined, and which contains two distinguished elements e and 0 with $0 \neq e$. Moreover, F is an abelian group with respect to addition, having 0 as the identity element, and the non-zero elements of F (often written F^*) form an abelian group with respect to multiplication having e as the identity element. The two operations are linked by the distributive laws.

Theorem 2.4
Every finite integral domain is a field.

Proof. Let R be a finite integral domain, and let its elements be $r_1, r_2, \ldots, r_n$. Consider a fixed non-zero element $r \in R$. Then the products $rr_1, rr_2, \ldots, rr_n$ must be distinct, since $rr_i = rr_j$ implies $r(r_i - r_j) = 0$, and since $r \neq 0$ we must have $r_i - r_j = 0$, i.e. $r_i = r_j$. Thus, these products are precisely the n elements of R. Each element of R is of the form rr_i; in particular, the identity $e = rr_i$ for some $1 \leq i \leq n$. Since R is commutative, we also have $r_i r = e$, and so r_i is the multiplicative inverse of r. Thus the non-zero elements of R form a commutative group, and R is a field. ■

Definition 2.5

- A subset S of a ring R is called a *subring* of R if S is closed under $+$ and $*$ and forms a ring under these operations.
- A subset J of a ring R is called an *ideal* if J is a subring of R and for all $a \in J$ and $r \in R$ we have $ar \in J$ and $ra \in J$.
- Let R be a commutative ring with an identity. Then the smallest ideal containing an element $a \in R$ is $(a) := \{ra : r \in R\}$. We call (a) the *principal ideal* generated by a.

Definition 2.6
An integral domain in which every ideal is principal is called a *principal ideal domain* (PID).

Example 2.7
$\mathbb{Z}$ is a PID.

An ideal J of R defines a partition of R into disjoint cosets (with respect to $+$), *residue classes* modulo J. These form a ring w.r.t. the following operations:

$$(a + J) + (b + J) = (a + b) + J,$$

$$(a + J)(b + J) = ab + J.$$

This ring is called the *residue class ring* and is denoted R/J.

Example 2.8
The residue class ring $\mathbb{Z}/(n)$
Here, (n) is the principal ideal generated by the integer n (same set $n\mathbb{Z}$ as the subgroup $\langle n \rangle$ but now with two operations). As in the group case, we denote the residue class of a modulo n by $[a]$, as well as by $a + (n)$. The elements of $\mathbb{Z}/(n)$ are $[0] = 0 + (n), [1] = 1 + (n), \ldots, [n-1] = n - 1 + (n)$.

Theorem 2.9
$\mathbb{Z}/(p)$, the ring of residue classes of the integers modulo the principal ideal generated by a prime p, is a field.

Proof. By Theorem 2.4, it is enough to show that $\mathbb{Z}/(p)$ is an integral domain. Now, $[a][b] = [ab] = [0]$ if and only if $ab = kp$ for some $k \in \mathbb{Z}$. Since p is prime, p divides ab if and only if p divides one of the factors. So, either $[a] = [0]$ or $[b] = [0]$, so $\mathbb{Z}/(p)$ contains no zero divisors. ■

These are our first examples of *finite fields*!

Example 2.10
Here are the addition and multiplication tables for the field $\mathbb{Z}/(3)$:

$+$	$0+(3)$	$1+(3)$	$2+(3)$
$0+(3)$	$0+(3)$	$1+(3)$	$2+(3)$
$1+(3)$	$1+(3)$	$2+(3)$	$0+(3)$
$2+(3)$	$2+(3)$	$0+(3)$	$1+(3)$

,

$*$	$0+(3)$	$1+(3)$	$2+(3)$
$0+(3)$	$0+(3)$	$0+(3)$	$0+(3)$
$1+(3)$	$0+(3)$	$1+(3)$	$2+(3)$
$2+(3)$	$0+(3)$	$2+(3)$	$1+(3)$

.

Remark 2.11
As you will prove in Exercise Sheet 1, the above result does not hold if p is replaced by a composite n.

Definition 2.12
A mapping $\phi : R \to S$ (R,S rings) is called a *ring homomorphism* if for any $a, b \in R$ we have

$$\phi(a + b) = \phi(a) + \phi(b) \text{ and } \phi(ab) = \phi(a)\phi(b).$$

A ring homomorphism preserves both $+$ and $*$ and induces a homomorphism of the additive group of R into that of S. Concepts such as kernel and image are defined analogously to the groups case. We have a ring version of the First Isomorphism Theorem:

Theorem 2.13 (First Isomorphism Theorem for Rings)
If ϕ is a ring homomorphism from a ring R onto a ring S then the factor ring $R/\ker\phi$ and the ring S are isomorphic by the map

$$r + \ker\phi \mapsto \phi(r).$$

We can use mappings to transfer a structure from an algebraic system to a set without structure. Given a ring R, a set S and a bijective map $\phi : R \to S$, we can use ϕ to define a ring structure on S that converts ϕ into an isomorphism. Specifically, for $s_1 = \phi(r_1)$ and $s_2 = \phi(r_2)$, define

$$s_1 + s_2 \text{ to be } \phi(r_1 + r_2), \text{ and } s_1 s_2 \text{ to be } \phi(r_1)\phi(r_2).$$

This is called the ring structure *induced by* ϕ; any extra properties of R are inherited by S.

This idea allows us to obtain a more convenient representation for the finite fields $\mathbb{Z}/(p)$.

Definition 2.14
For a prime p, let $\mathbb{F}_p$ be the set $\{0, 1, \ldots, p-1\}$ of integers, and let $\phi : \mathbb{Z}/(p) \to \mathbb{F}_p$ be the mapping defined by $\phi([a]) = a$ for $a = 0, 1, \ldots, p-1$. Then $\mathbb{F}_p$ endowed with the field structure induced by ϕ is a finite field, called the *Galois field of order* p.

From above, the mapping ϕ becomes an isomorphism, so $\phi([a] + [b]) = \phi([a]) + \phi([b])$ and $\phi([a][b]) = \phi([a])\phi([b])$. The finite field $\mathbb{F}_p$ has zero element 0, identity element 1 and its structure is that of $\mathbb{Z}/(p)$. So, computing with elements of $\mathbb{F}_p$ now means ordinary arithmetic of integers with reduction modulo p.

Example 2.15

- $\mathbb{F}_2$: the elements of this field are 0 and 1. The operation tables are:

$+$	0	1
0	0	1
1	1	0

,

$*$	0	1
0	0	0
1	0	1

- We have $\mathbb{Z}/(5)$, isomorphic to $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$, where the isomorphism is given by $[0] \mapsto 0, \ldots, [4] \mapsto 4$. The operation tables are:

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

,

$*$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Definition 2.16
If R is an arbitrary ring and there exists a positive integer n such that $nr = 0$ for every $r \in R$ (i.e. r added to itself n times is the zero element) then the least such positive integer n is called the *characteristic* of R, and R is said to have positive characteristic. If no such positive integer n exists, R is said to have characteristic 0.

Example 2.17

- $\mathbb{F}_2$ and $\mathbb{F}_5$ have characteristic 2 and 5 respectively.

- $\mathbb{Q}$ and $\mathbb{R}$ have characteristic 0.

Theorem 2.18
A ring $R \neq \{0\}$ of positive characteristic with an identity and no zero divisors must have prime characteristic.

Proof. Since R contains non-zero elements, R has characteristic $n \geq 2$. If n were not prime, we could write $n = km$ with $k, m \in \mathbb{Z}$, $1 < k, m < n$. Then $0 = ne = (km)e = (ke)(me)$, so either $ke = 0$ or $me = 0$, since R has no zero divisors. Hence either $kr = (ke)r = 0$ for all $r \in R$ or $mr = (me)r = 0$ for all $r \in R$, contradicting the definition of n as the characteristic. ■

Corollary 2.19
A finite field has prime characteristic.

Proof. From Theorem 2.18, we need only show that a finite field F has a positive characteristic. Consider the multiples $e, 2e, 3e, \ldots$ of the identity. Since F contains only finitely many elements, there must exist integers k and m with $1 \leq k < m$ such that $ke = me$, i.e. $(k - m)e = 0$, and thus $(k - m)f = (k - m)ef = 0f = 0$ for all $f \in F$ so F has a positive characteristic. ■

Example 2.20
The field $\mathbb{Z}/(p)$ (equivalently, $\mathbb{F}_p$) has characteristic p.

Theorem 2.21 (Freshmen's Exponentiation)
Let R be a commutative ring of prime characteristic p. then

$$(a+b)^{p^n} = a^{p^n} + b^{p^n} \text{ and } (a-b)^{p^n} = a^{p^n} - b^{p^n}$$

for $a, b \in R$ and $n \in \mathbb{N}$.

Proof. *It can be shown (see Exercise Sheet 1) that*

$$\binom{p}{i} = \frac{p(p-1)\cdots(p-i+1)}{1 \cdot 2 \cdots i} \equiv 0 \bmod p$$

for all $i \in \mathbb{Z}$ with $0 < i < p$. By the Binomial Theorem,

$$(a+b)^p = a^p + \binom{p}{1} a^{p-1} b + \cdots + \binom{p}{p-1} a b^{p-1} + b^p = a^p + b^p$$

and induction on n establishes the first identity. The second identity follows since

$$a^{p^n} = ((a-b)+b)^{p^n} = (a-b)^{p^n} + b^{p^n}.$$

■

3 Polynomials

Let R be an arbitrary ring. A *polynomial over R* is an expression of the form

$$f = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + \cdots + a_n x^n,$$

where n is a non-negative integer, the *coefficients* a_i $(0 \leq i \leq n)$ are elements of R, and x is a symbol not belonging to R, called an *indeterminate* over R.

Definition 3.1
Let $f = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + \cdots + a_n x^n$ be a polynomial over R which is not the zero polynomial, so we can suppose $a_n \neq 0$. Then n is called the *degree* of f. By convention, $\deg(0) = -\infty$. Polynomials of degree 0 are called *constant polynomials.* If the leading coefficient of f is 1 (the identity of R) then f is called a *monic* polynomial.

Given two polynomials f and g, we can write $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{i=0}^{n} b_i x^i$ (taking coefficients zero if necessary to ensure the same n). We define their sum to be

$$f + g = \sum_{i=0}^{n} (a_i + b_i) x^i$$

and their product to be

$$fg = \sum_{k=0}^{2n} c_k x^k, \text{ where } c_k = \sum_{i+j=k, 0 \leq i \leq n, 0 \leq j \leq n} a_i b_j .$$

Note that the degree of the product of two non-zero polynomials f and g is equal to the sum of the degrees of f and g.

Theorem 3.2
With the above operations, the set of polynomials over R forms a ring. It is called the polynomial ring *over R and denoted by $R[x]$. Its zero element is the* zero polynomial*, all of whose coefficients are zero.*

Proof. Exercise. ■

Let F denote a (not necessarily finite) field. From now on, we consider polynomials over fields.

We say that the polynomial $g \in F[x]$ *divides* $f \in F[x]$ if there exists a polynomial $h \in F[x]$ such that $f = gh$.

Theorem 3.3 (Division Algorithm)
Let $g \neq 0$ be a polynomial in $F[x]$. Then for any $f \in F[x]$, there exist polynomials $q, r \in F[x]$ such that

$$f = qg + r, \text{ where } \deg(r) < \deg(g).$$

Using the division algorithm, we can show that every ideal of $F[x]$ is principal:

Theorem 3.4
$F[x]$ is a principal ideal domain. In fact, for every ideal $J \neq (0)$ of $F[x]$ there is a uniquely determined monic polynomial $g \in F[x]$ such that $J = (g)$.

Proof. Let I be an ideal in $F[x]$. If $I = \{0\}$, then $I = (0)$. If $I \neq \{0\}$, choose a non-zero polynomial $k \in I$ of smallest degree. Let b be the leading coefficient of k, and set $m = b^{-1}k$. Then $m \in I$ and m is monic. We will show: $I = (m)$. Clearly, $(m) \subseteq I$. Now take $f \in I$; by the division algorithm there are polynomials q, r with $f = qm + r$ where either $r = 0$ or $\deg(r) < \deg(m)$. Now, $r = f - qm \in I$. If $r \neq 0$, we contradict the minimality of m; so we must have $r = 0$, i.e. f is a multiple of m and $I = (m)$.

We now show uniqueness: if $m_1 \in F[x]$ is another monic polynomial with $I = (m_1)$, then $m = c_1 m_1$ and $m_1 = c_2 m$ with $c_1, c_2 \in F[x]$. Then $m = c_1 c_2 m$, i.e. $c_1 c_2 = 1$, and so c_1, c_2 are constant polynomials. Since both m and m_1 are monic, we must have $m = m_1$. ■

We next introduce an important type of polynomial.

Definition 3.5
A polynomial $p \in F[x]$ is said to be *irreducible over F* if p has positive degree and $p = bc$ with $b, c \in F[x]$ implies that either b or c is a constant polynomial. A polynomial which does allow a non-trivial factorization over F is called *reducible over F*.

Note that the field F under consideration is all-important here, e.g. the polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$ but reducible in $\mathbb{C}[x]$, where it factors as $(x + i)(x - i)$.

Theorem 3.6 (Unique Factorization)
Any polynomial $f \in F[x]$ of positive degree can be written in the form

$$f = ap_1^{e_1} \dots p_k^{e_k}$$

where $a \in F$, $p_1, \dots, p_k$ are distinct monic irreducibles in $F[x]$ and $e_1, \dots, e_k$ are positive integers. This factorization is unique up to the order in which the factors occur; it is called the canonical factorization *of f in $F[x]$.*

Proof. Omitted. ■

Example 3.7
Find all irreducible polynomials over $\mathbb{F}_2$ of degree 3.

First, note that a non-zero polynomial in $\mathbb{F}_2[x]$ must be monic. The degree 3 polynomials are of the form $x^3 + ax^2 + bx + c$, where each coefficient is 0 or 1, i.e. there are $2^3 = 8$ of them. Such a polynomial is reducible over $\mathbb{F}_2$ precisely if it has a divisor of degree 1. Compute all products $(x + a_0)(x^2 + b_1x + b_0)$ to obtain all reducible degree 3 polynomials over $\mathbb{F}_2$. There are 6 of these, leaving 2 irreducibles: $x^3 + x + 1$ and $x^3 + x^2 + 1$.

Theorem 3.8
For $f \in F[x]$, the residue class ring $F[x]/(f)$ is a field if and only if f is irreducible over F.

Proof. *Details omitted. For those who know some ring theory this is immediate since, for a PID S, $S/(c)$ is a field if and only if c is a prime element of S. Here, the prime elements of the PID $R[x]$ are precisely the irreducible polynomials.* ■

We will be very interested in the structure of the residue class ring $F[x]/(f)$, for arbitrary non-zero polynomial $f \in F[x]$. To summarize,

- $F[x]/(f)$ consists of residue classes $g + (f)$ (also denoted $[g]$) with $g \in F[x]$.
- Two residue classes $g+(f)$ and $h+(f)$ are identical if and only if $g \equiv h \bmod f$, i.e. precisely if $g - h$ is divisible by f. This is equivalent to: g and h have the same remainder on division by f.
- Each residue class $g + (f)$ contains a unique representative $r \in F[x]$ with $\deg(r) < \deg(f)$, namely the remainder when g is divided by f. The process of moving from g to r is called *reduction mod f*. (Exercise: uniqueness?)
- Hence the distinct residue classes comprising $F[x]/(f)$ are precisely the residue classes $r + (f)$, where r runs through all polynomials in $F[x]$ with $\deg(r) < \deg(f)$.
- In particular, if $F = \mathbb{F}_p$ and $\deg(f) = n$, then the number of elements of $\mathbb{F}_p/(f)$ is equal to the number of polynomials in $\mathbb{F}_p/(f)$ of degree $< n$, namely p^n.

Example 3.9
- Let $f = x \in \mathbb{F}_2[x]$. The field $\mathbb{F}_2[x]/(x)$ has $2^1 = 2$ elements, namely $0 + (x)$ and $1 + (x)$. This field is isomorphic to $\mathbb{F}_2$.

- Let $f = x^2 + x + 1 \in \mathbb{F}_2[x]$. Then $\mathbb{F}_2[x]/(f)$ is a finite field of $2^2 = 4$ elements: $\{0 + (f), 1 + (f), x + (f), x + 1 + (f)\}$. Its behaviour under addition and multiplication is shown below (remember our field has characteristic 2). When performing field operations note that, since we replace each occurrence of f by 0, the polynomial representative for each residue class has degree less than 2.

$+$	$0+(f)$	$1+(f)$	$x+(f)$	$x+1+(f)$
$0+(f)$	$0+(f)$	$1+(f)$	$x+(f)$	$x+1+(f)$
$1+(f)$	$1+(f)$	$0+(f)$	$x+1+(f)$	$0+(f)$
$x+(f)$	$x+(f)$	$x+1+(f)$	$0+(f)$	$1+(f)$
$x+1+(f)$	$x+1+(f)$	$x+(f)$	$1+(f)$	$0+(f)$

,

$*$	$0+(f)$	$1+(f)$	$x+(f)$	$x+1+(f)$
$0+(f)$	$0+(f)$	$0+(f)$	$0+(f)$	$0+(f)$
$1+(f)$	$0+(f)$	$1+(f)$	$x+(f)$	$x+1+(f)$
$x+(f)$	$0+(f)$	$x+(f)$	$x+1+(f)$	$1+(f)$
$x+1+(f)$	$0+(f)$	$x+1+(f)$	$1+(f)$	$x+(f)$

.

Note that, in the multiplication table,

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

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

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

Comparing these tables to those of $\mathbb{Z}_4$ we see that the field $\mathbb{F}_2[x]/(f)$ is *not* isomorphic to $\mathbb{Z}_4$, which is not a field since in $\mathbb{Z}_4$ we have $2 \cdot 2 = 0$.

What is the multiplicative order of $x+(f)$ in $\mathbb{F}_2[x]/(f)$? The multiplicative group of this field has order $2^2 - 1 = 3$, so the order must be 1 or 3. Clearly $x+(f) \neq 1+(f)$, so the order must be 3. Check: $(x+(f))^3 = (x+(f))(x^2+(f)) = x(x+1)+(f) = x^2+x+(f) = 1+(f)$.

- Let $f = x^2 + 2 \in \mathbb{F}_3[x]$. We find that $\mathbb{F}_3[x]/(f)$ is a ring of 9 elements which is not even an integral domain, let alone a field. Its elements are $\{0 + (f), 1 + (f), 2 + (f), x + (f), x + 1 + (f), x + 2 + (f), 2x + (f), 2x + 1 + (f), 2x + 2 + (f)\}$. To see that it is not an integral domain, note that $(x + 1 + (f))(x - 1 + (f)) = x^2 - 1 + (f) = x^2 + 2 + (f) = 0 + (f)$, but neither $x + 1 + (f)$ nor $x - 1 + (f)$ are zero.

Definition 3.10
An element $a \in F$ is called a *root* (or *zero*) of the polynomial $f \in F[x]$ if $f(a) = 0$.

Example 3.11
(i) The elements $2, 3 \in \mathbb{Q}$ are roots of $x^2 - 5x + 6 \in \mathbb{Q}[x]$.
(ii) The polynomial $x^2 + 1 \in \mathbb{Q}[x]$ has no roots in $\mathbb{Q}$, but two roots $\pm i \in \mathbb{C}$.

Definition 3.12
If $f = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in F[x]$, then the *derivative* f' of f is defined by $f' = a_1 + 2a_2x + \cdots + na_nx^{n-1} \in F[x]$.

This obeys the familiar rules:

$$(f + g)' = f' + g'$$

and

$$(fg)' = fg' + f'g.$$

Theorem 3.13
An element $a \in F$ is a root of the polynomial $f \in F[x]$ if and only if $x - a$ divides f.

Proof. Using the Division Algorithm, we can write

$$f = q \cdot (x - a) + c$$

with $q \in F[x]$ and $c \in F$. Substituting $x = a$, we get $f(a) = c$, hence $f = q \cdot (x - a) + f(a)$. The theorem follows from this identity. ■

Definition 3.14
Let $a \in F$ be a root of $f \in F[x]$. If k is a positive integer such that f is divisible by $(x - a)^k$ but not $(x - a)^{k+1}$, then k is called the *multiplicity* of a. If $k \geq 2$ then a is called a *multiple root* of f.

Theorem 3.15
An element $a \in F$ is a multiple root of $f \in F[x]$ if and only if it is a root of both f and its derivative f'.

Proof. Exercise ■

Example 3.16
Consider the polynomial $f = x^3 - 7x^2 + 16x - 12 \in \mathbb{Q}[x]$. It factors as $(x-2)^2(x-3)$, so its roots are 2 (with multiplicity 2) and 3 (with multiplicity 1). Here, $f' = 3x^2 - 14x + 16$ which factors as $(x - 2)(3x - 8)$, so we can verify that 2 is also a root of f'.

The following observation is very important.

Theorem 3.17
If F is a field and $f \in F[x]$ has degree n, then F contains at most n roots of f.

Proof. Outline: Suppose F contains $n + 1$ distinct roots $a_1, \ldots, a_{n+1}$ of f. By Theorem 3.13, we can show that this implies $f = (x - a_1)(x - a_2) \cdots (x - a_{n+1})g$ for some polynomial g, contradicting $\deg(f) = n$. ■