In these chapter we study ring , field and the ideal , homomorphism , subbing . We prove some theorem . we also study the ring Zp and some of its properties , we define the field of fractions , integral domain , Characteristic of field and we prove that characteristic of field is only Zero or p (Prime).

Section (1-1): Definition and basic properties:

Although any assumption is that any reader has some knowledge of abstract algebra, a few remainders of basic definitions may be necessary, and have the added advantages of establishing the notations and conventions I shall use throughout the research. Introductory texts in abstract algebra are often titled or substituted "Groups, Rings and Fields", with fields playing only a minor part. Yet the theory of fields, through which both geometry and the classical theory of equations are illuminated by abstract algebra, contains some of the deepest and most remarkable insights in all mathematics. The hero of the narrative ahead is Evariste Galois, who died in a duel before his twenty-first birthday.

Definition (1-1-1):

A ring $R = (R, +, \cdot)$ is a non-empty set R furnished with two binary operations $+$ (called addition) and $\cdot$ (called multiplication) with the following properties. (Under the usual convention the dot for multiplication is omitted).

(R_1) the associative law for addition:

$$(a+b)+c+a+(b+c) \quad (a, b, c \ \in R);$$

(R_2) the commutative law for addition:

$$a+b=b+a \quad (a, b \in R);$$

(R_3) the existence of 0: there exists 0 in R such that, for all a in R,

$$a+0=a;$$

(R_4) the existence of negatives: for all a in R there exists $-a$ in R such that

$$a+(-a)=0;$$

(R_5) the associative law for multiplication:

$$(ab)c = a(bc) \quad (a, b, c \ \in R)$$

(R_6) the distributive laws:

$$a(b + c) = ab + ac, \quad (a + b)c = ac + bc \quad (a, b, c \in R)$$

We shall be concerned only with commutative rings, which have the following extra property

(R_7) the commutative law for multiplicative:

$$ab = ba \quad (a, b \in R)$$

A ring with unity R has the properties $(R_1) \dots (R_6)$, together with the following property.

(R_8) the existence of 1: there exists $1 \neq 0$ in R such that, for all a in R,

$$a1 = 1a = a.$$

The element 1 is called the unity element, or the (multiplicative) identity of R.

Definition (1-1-2):

A commutative ring R with unity is called an integral domain or, if the context allows, just a domain, if it has the following property.

(R_9) cancellation: for all a, b, c in R, with $c \neq 0$,

$$ca = cb \Rightarrow a = b$$

Definition (1-1-3):

A commutative ring R with unity is called a field if it has the following property.

(R_{10}) the existence of inverses: for all $a \neq 0$ in R there exists a^{-1} in R such that

$$aa^{-1} = 1.$$

We frequently wish to denote a^{-1} by $1/a$.

It is easy to see that (R_{10}) implies (R_9). The converses implication, however, is not true: the ring $\mathbb{Z}$ of integers is an obvious example. It is worth nothing also that (R_9) is equivalent to

$(R_9)'$ no divisor of zero: for all a, b in R,

$$ab = 0 \Rightarrow a = 0 \quad \text{or } b = 0$$

It is useful also at this stage to remind ourselves of a group.

Definition (1-1-4):

A group $G = (G,\cdot)$ is a non-empty set furnished with a binary operation $\cdot$ (usually omitted) with the following properties.

(G_1) the associative law:

$$(ab)c = a(bc) \quad (a, b, c \in G);$$

(G_2) the existence of an identity element: there exists e in G such that, for all a in G,

$$ea = a;$$

(G_3) the existence of inverses: for all a in G there exists a^{-1} in G such that

$$a^{-1}a = e.$$

An abelian group has the extra property

(G_4) the commutative law:

$$ab = ba \quad (a, b \in G).$$

Remark (1-1-5):

If $(R, +.\cdot)$ is a ring, then $(R, +)$ is an abelian group. If $(K, +,\cdot)$ is a field and $K^* = K\{0\}$, then $(K^*,\cdot)$ is an abelian group.

Let R be a commutative group with unity, and let

$$U = \{u \in R: (\exists v \in R)\ uv = 1\}.$$

It is easy to verify that U is an abelian group with respect to multiplication in R. We say that U is the group of units of the ring R. If a, b in R are such that $a = ub$ for some u in U, we say that a and b are associates, and write $a \sim b$. For example, in the ring $\mathbb{Z}$ the group of units is $\{1, -1\}$ and $a \sim a$ for all a in $\mathbb{Z}$.

Remark (1-1-6):

The group of units of a field K is the group K^* of all non-zero elements of K.

In a field, every non-zero element divides every other, but in an integral domain D the notion of divisibility plays a very significant role. If $a \in D/\{0\}$ and $b \in D$, we say that a divides b, or that a is a divisor of b, or that a is a factor of$b$, if there exists z in D such that $az = b$. We write $a|b$, and occasionally write $a \nmid b$ if a does not divide b. We say that a is a proper divisor, or a proper factor, of b, or that a properly divides b, if z is not a unit. Equivalently, a is a proper divisor of b if $a|b$ and $b \nmid a$.

Sub rings, Ideals and Homomorphism:

Much of the material in this section can be applied, with occasional modifications, to rings in general, but we shall suppose, without explicit mention, that all our rings are commutative. We shall use standard algebraic short hands: in particular, we write $a - b$ instead of $a + (-b)$.

Definition (1-1-7):

A sub ring U of a ring R is a non-empty subset of R with the property that, for all a, b in R,

$$a, b \in U \Rightarrow a - b, \quad ab \in U \qquad (1)$$

Equivalently, $U(\neq 0)$ is a subring if, for all a, b in R,

$$a, b \in U \Rightarrow a + b, \quad ab \in U, \quad a \in U \Rightarrow -a \in U \qquad (2)$$

It is easy to see that $0 \in U$: simply choose a from the non-empty set U, and deduce from (1) that $0 = a - a \in U$.

Definition (1-1-8):

A subfield of a field K is a subring which is a field. Equivalently, it is a subset E of K, containing at least two elements, such that

$$a, b \in E \Rightarrow a - b \in E, \quad a \in E, \qquad b \in E\{0\} \Rightarrow ab^{-1} \in E \qquad (3)$$

Again, we may replace the second implication of (3) by the implications

$$a, b \in E \Rightarrow ab \in E, \quad a \in E\{0\} \Rightarrow a^{-1} \in E \qquad (4)$$

If $E \subset K$ we say that E is a proper subfield of K.

Definition (1-1-9):

An ideal of R is a non-empty subset I of R with the properties

$$a, b \in I \Rightarrow a - b \in I, \quad a \in I \text{ and } r \in R \Rightarrow ra \in I \qquad (5)$$

Definition (1-1-10):

An ideal is certainly a sub ring, but not every sub ring is ideal. The sub ring $\mathbb{Z}$ of the field $\mathbb{Q}$ of rational numbers provides an example. Among the ideals of R are $\{0\}$ and R. An ideal I such that $\{0\} \subset I \subset R$ is called proper.

Theorem (1-1-11):

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a finite subset of a commutative ring R. Then the set

$$Ra_1 + Ra_2 + \cdots + Ra_n (= \{x_1a_1 + x_2a_2 + \cdots + x_na_n : x_1x_2, \ldots, x_n \in R\})$$

is the smallest ideal of R containing A.

Proof:

The set $Ra_1 + Ra_2 + \cdots + Ra_n$ is certainly an ideal, since, for all

$$x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$$

in R,

$$(x_1a_1 + x_2a_2 + \cdots + x_na_n) - (y_1a_1 + y_2a_2 + \cdots + y_na_n)$$

$$= (x_1 - y_1)a_1 + (x_2 - y_2)a_2 + \cdots + (x_n - y_n)a_n$$

$$\in Ra_1 + Ra_2 + \cdots + Ra_n;$$

and, for all r in R,

$$r(x_1a_1 + x_2a_2 + \cdots + x_na_n) = (rx_1)a_1 + (rx_2)a_2 + \cdots + (rx_n)a_n$$

$$\in Ra_1 + Ra_2 + \cdots + Ra_n.$$

It is clear that every ideal I containing $\{a_1, a_2, \dots, a_n\}$ contains the element $x_1a_1 + x_2a_2 + \cdots + x_na_n$ for every choice of $x_1, x_2, \dots, x_n$ in R, and so $Ra_1 + Ra_2 + \cdots + Ra_n \subset I$.

We refer to $Ra_1 + Ra_2 + \cdots + Ra_n$ as the ideal generated by $a_1, a_2, \dots, a_n$, and frequently write it as $\langle a_1, a_2, \dots, a_n \rangle$. Of special integers is the case where the ideal is generated by a single element a in R; we say that $Ra = \langle a \rangle$ is a principal ideal.

There is close connection between ideals and divisibility:

Theorem (1-1-12):

Let D be an integral domain with group of units U, and let $a, b \in D\{0\}$. Then:

(i) $\langle a \rangle \subseteq \langle b \rangle$ if and only if $b|a$;
(ii) $\langle a \rangle = \langle b \rangle$ if and only if $a \sim b$.
(iii) $\langle a \rangle = D$ if and only if $a \in U$.

Proof:

(i) Suppose first that $b|a$. Then $a = zb$ for some z in D, and so

$$\langle a \rangle = Da = Dzb \subseteq Db = \langle b \rangle$$

Conversely, suppose that $\langle a \rangle \subseteq \langle b \rangle$. Then there exists z in D such that $a = zb$ and so $b|a$.

(ii) Suppose first that $a \sim b$. Then there exists u in U such that $a = ub$ and $b = ub$, $b = va$. Hence $(uv)a = u(va) = ub = a = 1a$, and so, by cancellation, $uv = 1$. Thus u and v are units, and so $a \sim b$.

(iii) It is clear that $\langle 1\rangle = D$. Hence by (ii), $\langle a\rangle = D$ if and only if $a \sim 1$, that is, if and only if a is a unit.

A homomorphism from a ring R into a ring S is a mapping $\varphi\colon R \to S$ with the properties

$$\varphi(a+b) = \varphi(a) + \varphi(b), \quad \varphi(ab) = \varphi(a)\varphi(b). \qquad (6)$$

Among the homeomorphisms from R into S is the zero mapping ζ given by

$$\zeta(a) = 0 \ (a \in R). \qquad (7)$$

While some of the theorems we establish will apply to all homomorphisms, including ζ, others will apply only to non-zero homomorphisms.

Some elementary properties of ring homeomorphisms are gathered together in the following theorem:

Theorem (1-1-13):

Let R, S be rings, with zero elements $0_R, 0_S$, respectively, and let $\varphi\colon R \to S$ be a homomorphism. Then

(i) $\varphi(0_R) = 0_S$;
(ii) $\varphi(-r) = -\varphi(r)$ for all r in R;
(iii) $\varphi(R)$ is a subring of S.

Proof:

(i) Since

$$\varphi(a) + \varphi(0_R) = \varphi(a + 0_R) = \varphi(a),$$

we can deduce that

$$\varphi(0_R) = 0_S + \varphi(0_R) = -\varphi(a) + \varphi(a) + \varphi(0_R) = -\varphi(a) + \varphi(a) = 0_S \qquad (8)$$

(ii) Since, for all r in R,

$$\varphi(r) + \varphi(-r) = \varphi\big(r + (-r)\big) = \varphi(0_R) = 0_S = \varphi(r) + \big(-\varphi(r)\big),$$

it follows that

$$\varphi(-r) = -\varphi(r). \qquad (9)$$

(iii) Let $\varphi(a), \varphi(b)$ be arbitrary elements of $\varphi(R)$, with $a, b \in R$. Then

$$\varphi(a)\varphi(b) = \varphi(ab) \in \varphi(R)$$

and by virtue of (9),

$$\varphi(a) - \varphi(b) = \varphi(a) + \varphi(-b) = \varphi\big(a + (-b)\big) \in \varphi(R)$$

Thus $\varphi(R)$ is a subring.

The following corollary is an immediate consequence of the above proof.

Corollary (1-1-14):

If $\varphi: R \to S$ is a ring homomorphism and $a, b \in R$, then $\varphi(a-b) = \varphi(a) - \varphi(b)$.

Let $\varphi: R \to S$ be a homomorphism. If φ is one-to-one, we call it a monomorphism, or an embedding, and if φ is also onto we call it an isomorphism. We say that the rings R and S are isomorphic (to each other) and write $R \simeq S$. For example, the ring $R = \{m + n\sqrt{2}: m, n \in \mathbb{Z}\}$ is isomorphic to the ring

$$S = \left\{\begin{pmatrix} m & n \\ 2n & m \end{pmatrix}: m, n \in \mathbb{Z}\right\} \qquad (10)$$

with the operations of matrix addition and multiplication, the isomorphism being

$$\varphi: m + n\sqrt{2} \mapsto \begin{pmatrix} m & n \\ 2n & m \end{pmatrix}.$$

We shall eventually be interested in the case where the rings R and S coincide: an isomorphism from R onto itself is called an automorphism.

If $\varphi: R \to S$ is a monomorphism, then the subring $\varphi(R)$ of S is isomorphic to R. Since the rings R and $\varphi(R)$ are abstractly identical, we often wish to identify $\varphi(R)$ with R and regard R itself as a subring of S. For example, if S is the ring defined by (10), there is a monomorphism $\theta: \mathbb{Z} \to R$ given by

$$\theta(m) = \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix} \quad (m \in \mathbb{Z}),$$

and the identification of the integer m with the 2×2 scalar matrix $\theta(m)$ allows us to consider $\mathbb{Z}$ as effectively a subring of S. We say that R contains $\mathbb{Z}$ up to isomorphism.

Let $\varphi: R \to S$ be a homomorphism, where R and S are rings, with zero elements $0_R, 0_S$, respectively, and let

$$K = \varphi^{-1}(0_S)(= \{a \in R: \varphi(a) = 0_S\}) \qquad (11)$$

We refer to K as the kernel of the homomorphism φ, and write it as $\ker \varphi$.

If $a, b \in K$, then $\varphi(a) = \varphi(b) = 0$ and so certainly $\varphi(a - b) = 0$; hence $a - b \in K$. If $r \in R$ and $a \in K$, then $\varphi(ra) = \varphi(r)\varphi(a) = \varphi(r)0 = 0$. Hence $ra \in K$. We deduce that the kernel of a homomorphism is an ideal.

In fact the last remark records only one of the ways in which the notions of homomorphism and ideal are linked. Let I be an ideal of a ring R, and let $a \in R$. The set $a + I = \{a + x: x \in I\}$ is called the residue class of a modulo I. We now show that, for all a, b in R,

$$a + I = b + I \Leftrightarrow a - b \in I \qquad (12)$$

and

$$(a + I) + (b + I) = (a + b) + I, \quad (a + I)(b + I) \subseteq ab + I. \qquad (13)$$

To prove the first of these statements, suppose that $a + I = b + I$. Then, in particular, $a = a + 0 \in a + I = b + I$, and so there exists x in I such that $a = b + x$. Thus $a - b = x \in I$. Conversely, suppose that $a - b \in I$. Then, for all x in I, we have that $a + x = b + y$, where $y = (a - b) + x \in I$. Thus $a + I \subseteq b + I$, and the reverse inclusion is proved in the same way.

To prove the first statement in (13), let $x, y \in I$ and let

$$u = (a + x) + (b + y) \in (a + I) + (b + I).$$

Then $u = (a + b) + (x + y) \in (a + b) + I$. Conversely, if $z \in I$ and $v = (a + b) + z \in (a + b) + I$, then $v = (a + z) + (b + 0) \in (a + I) + (b + I)$.

The second statement follows in a similar way. Let $x, y \in I$ and let $u = (a+x)(b+y) \in (a+I)(b+I)$. Then $u = ab + (ay + xb + xy) \in ab + I$.

The set of R/I of all residue classes modulo I forms a ring with respect to the operations

$$(a+I)+(b+I)=(a+b)+I, \quad (a+I)(b+I)=ab+I \qquad (14)$$

called the residue class ring modulo I. The verifications are routine. The zero element is $0+I=I$; the negative of $a+I$ is $-a+I$. The mapping $\theta_I : R \to R/I$, given by

$$\theta_I(a) = a + I \quad (a \in R) \qquad (15)$$

is a homomorphism onto R/I, with kernel I. It is called the natural homomorphism from R onto R/I.

The motivation example of a residue class ring is the ring $\mathbb{Z}_n$ of integers mod n. Here the ideal is $\langle n \rangle = n\mathbb{Z}$, the set of integers divisible by n, and the elements of $\mathbb{Z}_n$ are the classes $a + \langle n \rangle$, with $a \in \mathbb{Z}$. There are exactly are exactly n classes, namely

$$\langle n \rangle, 1 + \langle n \rangle, 2 + \langle n \rangle, \ldots, (n-1) + \langle n \rangle.$$

A strong connection with number theory is revealed by the following theorem:

Theorem (1-1-15):

Let n be a positive integer. The residue class ring $\mathbb{Z}_n = \mathbb{Z}/\langle n \rangle$ is a field if and only if n is prime.

Proof:

Suppose first that n is not prime. Then $n = rs$, where $1 < r < n$ and $1 < s < n$. Then $r + \langle n \rangle \neq 0 + \langle n \rangle$ and $s + \langle n \rangle \neq 0 + \langle n \rangle$, but

$$(r + \langle n \rangle)(s + \langle n \rangle) = n + \langle n \rangle = 0 + \langle n \rangle.$$

Thus $\mathbb{Z}_n$ contains divisor of 0, and so is certainly not a field.

Now let p be a prime, and suppose that $(r + \langle p \rangle)(s + \langle p \rangle) = 0 + \langle p \rangle$. Then $p|rs$, and so (since p is prime) either $p|r$ or $p|s$. That is, either $r + \langle p \rangle = 0$ or $s + \langle p \rangle = 0$. Thus $\mathbb{Z}_p$ has no divisors of zero, and so is an integral domain.

The next theorem, which has counterparts in many branches of algebra, tells us that every homomorphic image of a ring R is isomorphic to a suitably chosen residue class ring.

Theorem (1-1-16):

Let R be a commutative ring, and let φ be a homomorphism from R onto a commutative ring S, with kernel K. Then there is an isomorphism $\alpha: R/K \to S$ such that the diagram

$$\begin{tikzcd}
R \arrow[r, "\varphi"] \arrow[d, "\theta_K"'] & S \\
R/K \arrow[ur, "\alpha"'] &
\end{tikzcd}$$

commutes.

Proof:

Define α by the rule that

$$\alpha(a + K) = \varphi(a) \qquad (a + K \in R/K).$$

The mapping is both well-defined and injective, for

$$a + K = b + K \Leftrightarrow a - b \in K \Leftrightarrow \varphi(a - b) = 0 \Leftrightarrow \varphi(a) = \varphi(b).$$

It is clearly maps onto S, since φ is onto. It is a homomorphism, since

$$\alpha\big((a + K) + (b + K)\big) = \alpha\big((a + b) + K\big) = \varphi(a + b)$$

$$= \varphi(a) + \varphi(b) = \alpha(a + K) + \alpha(b + K),$$

and

$$\alpha\big((a + K) + (b + K)\big) = \alpha(ab + K) = \varphi(ab) = \varphi(a)\varphi(b)$$
$$= \alpha(a + K)\alpha(b + K).$$

Hence α is an isomorphism. The commuting of the diagram is clear, since, for all a in R,

$$\alpha\big(\theta_K(a)\big) = \alpha(a + K) = \varphi(a),$$

and so $\alpha \circ \theta_K = \varphi$.

The Field of Fractions of an Integral Domain:

We know that every finite integral domain is a field. In this section we show how to construct a field out of an arbitrary integral domain.

Let D be an integral domain. Let

$$P = D \times (D\{0\}) = \{(a, b) : a, b \in D, b \neq 0\}.$$

Define a relation $\equiv$ on the set P by the rule that

$$(a, b) \equiv (a', b') \text{ if and only if } ab' = a'b.$$

Lemma (1-1-17):

The relation $\equiv$ is an equivalence.

We must prove that, for all $(a, b), (a', b'), (a'', b'')$ in P,

(i) $(a, b) \equiv (a, b)$ (the reflexive law);

(ii) $(a, b) \equiv (a', b') \Rightarrow (a', b') \equiv (a, b)$ (the symmetric law);

(iii) $(a, b) \equiv (a', b')$ and $(a', b') \equiv (a'', b'') \Rightarrow (a, b) \equiv (a'', b'')$ (the transitive law).

The properties (i) and (ii) are immediate from the definition. As for (iii), from $(a, b) \equiv (a', b')$ and $(a;, b') \equiv (a'', b'')$ we have that $ab' = a'b$ and $a'b'' = a''b'$. Hence

$$b'(ab'') = (ab')b'' = a'b'b'' = b(a'b'') = ba''b' = b'(a''b).$$

Since $b' \neq 0$, we can use the cancellation axiom to obtain $ab'' = a''b$, and so $(a, b) \equiv (a'', b'')$.

The quotient set $P/\equiv$ is denoted by $Q(D)$. Its elements are equivalence classes $[a, b] = \{(x, y) \in P: (x, y) \equiv (a, b)\}$, and, for reasons that will become obvious. We choose to denote the classes by fraction symbols a/b. Two classes are equal if their (arbitrary chosen) representative pairs in the set P are equivalent:

$$\tfrac{a}{b} = \tfrac{c}{d} \text{ if and only if } ad = bc.$$

In particular, note that

$$\frac{a}{b} = \frac{ka}{kb}$$

for all $k \neq 0$ in D.

We define addition and multiplication in $Q(D)$ by the rules

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} \qquad (16)$$

Lemma (1-1-18):

The additional multiplication defined by (16) are well-defined.

Proof:

Suppose that $\frac{a}{b} = a'/b'$ and $c/d = c'/d$. Then $ab' = a'b$ and $cd' = c'd$, and so

$$(ad + bc88)b'd' = ab'dd' + bb'cd' = a'bdd' + bb'c'd = (a'd' + b'c')bd$$

Hence

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} = \frac{a'd' + b'c'}{b'd'} = \frac{a'}{b'} + \frac{c'}{d'}.$$

Similarly,

$$(ac)(b'd') = (ab')(cd') = (a'b)(c'd) = (a'c')(bd),$$

and so

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a'}{b'} \cdot \frac{c'}{d'}.$$

These operations turn $Q(D)$ into a commutative ring with unity. The verifications are tedious but not difficult. For example,

$$\frac{a}{b}\left(\frac{c}{d} + \frac{e}{f}\right) = \frac{a}{b} \cdot \frac{cf + de}{df} = \frac{acf + ade}{bdf},$$

$$\frac{a}{b} \cdot \frac{c}{d} + \frac{a}{b} \cdot \frac{e}{f} = \frac{ac}{bd} + \frac{ae}{bf} = \frac{acbf + aebd}{b^2df} = \frac{acf + ade}{bdf}.$$

The zero element is $0/1$ ($= 0/b$ for all $b \neq 0$ in D). The unity element is $1/1$ ($= b/b$ for all $b \neq 0$ in D). The negative of a/b is $(-a)/b$.

The ring $Q(D)$ is in fact a field, since for all a/b with $a \neq 0$ we have that

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1}.$$

We refer to the field $Q(D)$ as the field of fractions of the domain D.

Lemma (1-1-19):

The mapping $\varphi: D \to Q(D)$ given by

$$\varphi(a) = \frac{a}{1} \quad (a \in D) \tag{17}$$

is monomorphism.

Proof:

From (16) it is clear that

$$\varphi(a) + \varphi(b) = \frac{a}{1} + \frac{b}{1} = \frac{a+b}{1} = \varphi(a+b), \quad \varphi(a)\varphi(b) = \frac{a}{1} \cdot \frac{b}{1} = \frac{ab}{1} = \varphi(ab)$$

Also

$$\varphi(a) = \varphi(b) \Rightarrow \frac{a}{1} = \frac{b}{1} \Rightarrow a = b.$$

If we identify $a/1$ with a, we can regard $Q(D)$ as containing D as a subring. The field $Q(D)$ is the smallest field containing D, in the following sense.

Theorem (1-1-20):

Let D be an integral domain, let φ be the monomorphism from D into $Q(D)$ given by (17) and let K be a field with the property that there is a monomorphism θ from D into K. Then there exists a monomorphism $\psi: Q(D) \rightarrow K$ such that the diagram

$$\begin{tikzcd}
D \arrow[r, "\theta"] \arrow[d, "\varphi"'] & K \\
Q(D) \arrow[ur, "\psi"'] &
\end{tikzcd}$$

commutes.

Proof:

Define a mapping $\psi: Q(D) \rightarrow K$ by the rule that

$$\psi\left(\frac{a}{b}\right) = \frac{\theta(a)}{\theta(b)}.$$

(Note that $\theta(b) \neq 0$, since θ is a monomorphism). This is well-defined end one-to-one, since

$$\frac{a}{b} = \frac{c}{d} \Leftrightarrow ad = bc \Leftrightarrow \theta(a)\theta(d)\theta(c) \Leftrightarrow \frac{\theta(a)}{\theta(b)} = \frac{\theta(c)}{\theta(d)}$$

and it is a homomorphism, since

$$\psi\left(\frac{a}{b} + \frac{c}{d}\right) = \psi\left(\frac{ad + bc}{bd}\right) = \frac{\theta(ad + bc)}{\theta(bd)} = \frac{\theta(a)\theta(d) + \theta(b)\theta(c)}{\theta(b)\theta(d)}$$

$$= \frac{\theta(a)}{\theta(b)} + \frac{\theta(c)}{\theta(d)} = \psi\left(\frac{a}{b}\right) + \psi\left(\frac{c}{d}\right),$$

and similarly

$$\psi\left(\frac{a}{b} \cdot \frac{c}{d}\right) = \psi\left(\frac{a}{b}\right)\psi\left(\frac{c}{d}\right).$$

The commuting of the diagram is clear, since, for all a in D,

$$\psi\big(\varphi(a)\big) = \psi\left(\frac{a}{1}\right) = \frac{\theta(a)}{\theta(1)} = \theta(a).$$

More informally, Theorem (1-1-6) tell us that any field containing D contains (up to isomorphism) the field $Q(D)$.

When $D = \mathbb{Z}$, it is clear that $Q(D) = \mathbb{Q}$. This is the classical example of the field of quotients, but we shall soon see that it is not the only one.

The characteristic of a Field:

In a ring R containing an element a it is reasonable to denote $a + a$ by $2a$,and, more generally, if n is a natural number we may write na for the sum $a + a + \cdots + a$ (n summands). If we define $0a = 0_R$ and $(-n)a$ to be $n(-a)$, we can give a meaning to na for every integer n. The following properties are easy to establish: for $m, n \in \mathbb{Z}$ and $a, b \in R$,

$$(m+n)a = ma + na, \quad m(a+b) = ma + mb, \quad (mn)a = m(na),$$

$$m(ab) = (ma)b = a(mb), \quad (ma)(nb) = (mn)(ab) \qquad (18)$$

Consider a commutative ring R with unity element 1_R. Then there are two possibilities; either

(i) The element $m1_R (m = 1,2,3, \ldots)$ are all distinct; or
(ii) There exist m, n in $\mathbb{N}$ such that $m1_R = (m+n)1_R$.

In former case we say that R has characteristic zero, and write char $R = 0$, in the latter case we notice that $m1_R = (m+n)1_R = m1_R + n1_R$, and so $n1_R = 0$. The least positive n for which holds is called the characteristic of the ring R. Note that, if R is a ring of characteristic n, then $na = 0_R$ for all a in R, for $na = (n1_R)a = 0a = 0$. We write char $R = n$.

Theorem (1-1-21):

The characteristic of a field is either 0 or a prime number p.

Proof:

The former possibility can certainly occur: $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are all fields of characteristic 0. Let K be a field and suppose that $K = n \neq 0$, where n is not prime. Then $n = rs$, where $1 < r < n, 1 < s < n$, and the minimal property of n implies that $r1_k \neq 0_k$. On the other hand, form (18) we deduce that

$$(r1_K)(s1_K) = (rs)1_K = n1_K = 0_K,$$

and this is impossible, since K being a field, has no zero divisors.

Let K be a field with characteristic 0. Then the elements $n1_K(n \in \mathbb{Z})$ are all distinct, and form a subring of K isomorphic to $\mathbb{Z}$. Indeed, the set

$$P(K) = \{m1_K/n1_f : m, n \in \mathbb{Z}, \qquad n \neq 0\} \qquad (19)$$

is a subfield of K isomorphic to $\mathbb{Q}$.any subfield of K must contain 1 and 0 and so must contain $P(K)$, which is called the prime subfield of K.

If K has prime characteristic p, the prime subfield is

$$P(K) = \{1_K, 2(1_K), \dots, (p-1)(1_K)\}, \qquad (20)$$

and this is isomorphic to $\mathbb{Z}_p$.

The fields $\mathbb{Q}$ and $\mathbb{Z}_p$ play a central role in the theory of fields. They have no proper subfields, and every field contains as a subfield an isomorphic copy of one or other of them. We frequently want to express this may saying that every field of characteristic 0 is an extension of $\mathbb{Q}$, and every field of prime characteristic p is an extension $\mathbb{Z}_p$.

We record this observations formally in a theorem.

Theorem (1-1-22):

Let K be a field. Then K contains a prime subfield $P(K)$ contained in every subfield. If char $K = 0$ then $P(K)$, described by (19), is isomorphic to $\mathbb{Q}$. If char $K = p, a$ is a prime number, then $P(K)$, described by (20), is isomorphic to $\mathbb{Z}_p$.

Remark (1-1-23):

Given an element a of a field K, we sometimes like to denote $a/(n1)$ simply by a/n. If char $K = 0$ this is no problem, but if char $K = p$ then we cannot assign a meaning to a/n if n is a multiple of p. Thus, for example, the formula

$$xy = \frac{1}{4}((x+y)^2 - (x-y)^2)$$

is not valid in a field of characteristic 2, since the quantity on the right reduces to $0/0$ and so is undefined.

In fields with finite characteristic we encounter some surprising formulae.

Theorem (1-1-24):

Let K be a field of characteristic p. Then, for all x, y in K,

$$(x+y)^p = x^p + y^p.$$

Proof:

By the binomial theorem, valid in any commutative ring with unity, we have that

$$(x+y)^p = \sum_{r=0}^{p} \binom{p}{r} x^{n-r} y^r. \qquad (21)$$

For $r = 1, \dots, p-1$, the binomial coefficient

$$\binom{p}{r} = \frac{p(p-1)\dots(p-r+1)}{r!}$$

is an integer, and so $r!$ divides $p(p-1)\ldots(p-r+1)$. Since p is prime and $r < p$, no factor of $r!$ can be divisible by p. Hence $r!$ divides $(p-1)\ldots(p-r+1)$, and so $\binom{p}{r}$ is an integer divisible by p. Hence, for $r = 1, \ldots, p-1$,

$$\binom{p}{r} x^{n-r} y^{r} = 0,$$

and so, in (21), only the first and last terms survive.

Remark (1-1-25):

The fields $\mathbb{Z}_p = \mathbb{Z}/\langle p \rangle$ are important building blocks in field theory. We usually find it convenient to write $\mathbb{Z}_p = \{0,1,\ldots,p-1\}$, with addition and multiplication carried out modulo p. So, for example, the multiplication table for $\mathbb{Z}_5$ is

	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

When it comes to $\mathbb{Z}_3$, it is usually more convenient to write $\mathbb{Z}_3 = \{0,1,-1\}$. Again, we might at times find it convenient to write $\mathbb{Z}_5$ as $\{0, \pm 1, \pm 2\}$ obtaining the table

	0	1	2	-2	-1
0	0	0	0	0	0
1	0	1	2	-2	-1
2	0	2	-1	1	-2
-2	0	-2	1	-1	2
-1	0	-1	-2	2	1

A remainder of Some Group Theory:

It is perhaps paradoxical, given the extensive list of axioms that define a field, that a serious study of fields requires a knowledge of more general objects. Rings we have encountered already, though in fact we do not need to explore any further than integral domains. More surprisingly, w e need to know some group theory. This does not come into play until well through the book, and you may prefer to skip this section and to return to it when the material is needed. For the most part I will give sketch proofs only: more detail can mostly be found. As the title suggests, this section is a remainder of the basic ideas and definitions. More specialized bits of group theory, not necessarily covered in a first course in abstract algebra, will be explained when they are needed, and some proofs will be consigned to an appendix.

The axioms for a group were recorded in section 1.1. it follows from these axioms that the element e in G_2 and the element a^{-1} in G_3 are both unique and that

$$ae = ea = a, \quad aa^{-1} = a^{-1}a = a.$$

Also, for all $a, b \in G$,

$$(ab)^{-1} = b^{-1}a^{-1}.$$

Definition (1-1-26):

The group $(G,\cdot)$ is called a finite group if the set G is finite. The cardinality $|G|$of G called the order of the group.

In the usual way, we write $a^2, a^3, \ldots$ (where $a \in G$) for the products $aa, aaa, \ldots$, and we write a^{-n} to mean $(a^{-1})^n = (a^n)^{-1}$. By a^0 we mean the identity element e.

Definition (1-1-27):

A group G is called cyclic if there is exists an element a in G such that $G = \{a^n : n \in \mathbb{Z}\}$. If the powers a^n are all distinct, G is the infinite cyclic group. Otherwise, there is a least $m > 0$ such that $a^m = e$. The division algorithm then implies, for all n in $\mathbb{Z}$, that there exists integers q and r such that

$$a^n = a^{qm+r} = (a^m)^q a^r = a^r,$$

and $0 \leq r \leq m-1$. Thus $G = \{e, a, a^2, \ldots, a^{m-1}\}$, the cyclic group of order m both the finite cyclic group and the cyclic group of oder m are abelian.

A non-empty subset U of G is called a subgroup of G if, for all $a, b \in G$,

$$a, b \in U \ \ \Rightarrow ab \in U, \quad a \in U \Rightarrow a^{-1} \in U \qquad (22)$$

or equivalent,

$$a, b \in U \Rightarrow ab^{-1} \in U. \qquad (23)$$

Every subgroup contains the identity element e. For each element a in the group G, the set $\{a^n : n \in \mathbb{Z}\}$ is a subgroup, called the cyclic subgroup generated by a, and denoted by $\langle a \rangle$. If G is finite, this cannot be the infinite cyclic group, and the order of the cyclic subgroup generated by a is called the order of the element a. It is the smallest positive integer n such that $a^n = e$, and is denoted by $o(a)$.

Let U be a subgroup of a group G and let $a \in G$. The subset $Ua = \{ua : u \in U\}$ is called a left coset of U. Then $Ua = Ub$ if and only if $ab^{-1} \in U$. Among the left cosets is U itself. The distinct left cosets form a partition of G: that is, every element of G belongs to exactly one left coset of U. The mapping $u \longmapsto ua$ from U into Ua is easily seen to be both one-one and onto, and so, in a finite group, every left coset has the same number of elements as U. Thus

$$|G| = |U| \times \text{(the number of left cosets)},$$

and we have Lagrange's theorem.

Theorem (1-1-28):

Let G, H be groups, and let φ a homomorphism from G onto H, with kernel N. Then there exists a unique isomorphism $\alpha : G/N \to H$ such that the diagram

$$\begin{tikzcd}
G \arrow[r, "\varphi"] \arrow[d, "v_N"'] & H \\
G/N \arrow[ur, "\alpha"'] &
\end{tikzcd}$$

commutes.

Proof:

The mapping $\alpha: Na \longmapsto \varphi(a)$ is well defined, one-one, and a homomorphism – and $\alpha \circ \upsilon_N = \varphi$.

Section (1-2): Integral Domains and Polynomials:

Euclidean Domain:

An integral domain D is called a Euclidean domain if there is a mapping δ from D into the set $\mathbb{N}^0$ of non-negatives with the property that $\delta(0) = 0$ and, for all a in D and all b in $D\{0\}$, there exist q, r in D such that

$$a = qb + r \quad \text{and} \quad \delta(r) < \delta(b) \qquad (24)$$

From the definition it follows that $\delta^{-1}\{0\} = \{0\}$, for if $\delta(b)$ were equal to 0 it would not be possible to find r such that $\delta(r) < \delta(b)$.

The most important example is the ring $\mathbb{Z}$, where $\delta(a)$ is defined as $|a|$, and where the process, known as the division algorithm, is the familiar one (which we have indeed already used in chapter 1) of dividing a by b and obtaining a quotient q and a remainder r. If b is positive, the there exists q such that

$$qb \leq (q + 1)b.$$

Thus $0 \leq a - qb < b$, and so, taking r as $a - qb$, we see that $a = qb + r$ and $|r| < |b|$. If b is negative, then there exists q such that

$$(q + 1)b < a \leq qb.$$

Thus $b < r = a - qb \leq 0$, and so again $a = qb + r$ and $|r| < |b|$. We shall come across another important example later.

An integral domain D is called a principal ideal domain if all of its ideals are principal.

Theorem (1-2-1):

Every Euclidean domain is a ideal domain.

Proof:

Let D be a Euclidean domain. The ideal $\{0\}$ is certainly principal. Let I be a non-zero ideal, and let b be a non-zero element of I such that

$$\delta(b) = \min\{\delta(x): x \in I\backslash 0\}.$$

Let $a \in I$. Then there exist q, r such that $a = qb + r$ and $\delta(r) < \delta(b)$. Since $r = a - qb \in I$, we have a contradiction unless $r = 0$. Thus $a = qb$, and so $I = Db = \langle b \rangle$, a principal ideal.

Suppose now that a, b are non-zero members of a principal ideal domain D, and let $\langle a, b \rangle = \{sa + tb: s, t \in D\}$ be the ideal generated by a and b. (see Theorem (1-11)) By our assumption that D is a principal ideal domain, there exists d in D such that $\langle a, b \rangle = \langle d \rangle$. Since $\langle a \rangle \subseteq \langle d \rangle$ and $\langle b \rangle \subseteq \langle d \rangle$, we have, from Theorem (1-12), that $d|a$ and $d|b$. Since $d \in \langle a, b \rangle$, there exist s, t in D such that $d = sa + tb$. If $d'|a$ and $d'|b$, then $d'|sa + tb$. That is, $d'|d$. We say that d is greatest common factor, of a and b. It is effectively unique, for, if $\langle a, b \rangle = \langle d \rangle = \langle d^* \rangle$, it follows from Theorem (1-12) (iii) that $d^* \sim d$.

To summarise, d is the greatest common divisor of a and b (write $d = \gcd(a, b)$) if it has the following properties:

$(GCD1)$ $d|a$ and $d|b$.

$(GCD2)$ if $d'|a$ and $d'|b$, then $d'|d$.$\gcd(a, b) \sim 1$, we say that a and b are coprime, or relatively prime.

In the case of the domain $\mathbb{Z}$, where the group of unit is $\{1, -1\}$, we have, for example, that $\langle 12,18 \rangle = \langle 6 \rangle = \langle -6 \rangle$.

Remark (1-2-2):

A simple modification of the above argument enables us to conclude that, in a principal ideal domain D, every finite set $\{a_1, a_2, \ldots, a_n\}$ has a greatest common divisor.

In the argument leading to the existence of the greatest common divisor, we assert that “there exists d such that $\langle a, b \rangle = \langle d \rangle$,” but give no indication of how this element d might be found. If the domain is Euclidean, we do have an algorithm.

The Euclidean Algorithm:

Suppose that a and b are non-zero elements of a Euclidean domain D, and suppose, without loss of generality, that $\delta(b) \le \delta(a)$. Then there exist $q_1, q_2, \dots$ and $r_1, r_2, \dots$ such that

$$\left.\begin{aligned} a &= q_1 b + r_1, \quad \delta(r_1) \le \delta(b),\\ b &= q_2 r_1 + r_2, \quad \delta(r_2) \le \delta(r_1),\\ r_1 &= q_3 r_2 + r_3, \quad \delta(r_3) \le \delta(r_2),\\ r_2 &= q_4 r_4 + r_4, \quad \delta(r_4) \le \delta(r_3),\\ &\dots\dots\dots. \end{aligned}\right\} \qquad (25)$$

The process must end with some $r_k = 0$, the final equations being

$$\begin{aligned} &r_{k-3} = q_{k-1} r_{k-2} + r_{k-1}, \quad \delta(r_{k-1})\delta(r_{k-2}),\\ &r_{k-2} = q_k r_{k-1}. \end{aligned}$$

Now, from the equation of (25), we deduce that

$$\langle a, b\rangle = \langle b, r_1\rangle; \qquad (26)$$

for every element $sa + tb$ in $\langle a, b\rangle$ can be rewritten as $(t + sq_1)b + sr_1 \in \langle b, r_1\rangle$, and every element $xb + yr_1$ in $\langle b, r_1\rangle$ can be written as $ya + (x - yq_1)b \in \langle a, b\rangle$. Similarly, the subsequent equations give

$$\begin{aligned} \langle b, r_1\rangle = \langle r_1, r_2\rangle,\ \ \langle r_1, r_2\rangle = \langle r_2, r_3\rangle, \dots\\ \langle r_{k-3}, r_{k-2}\rangle = \langle r_{k-2}, r_{k-1}\rangle,\ \ \langle r_{k-2}, r_{k-1}\rangle = \langle r_{k-1}\rangle. \end{aligned} \qquad (27)$$

From (26) and (25) it follows that $\langle a, b\rangle = \langle r_{k-1}\rangle$, and so r_{k-1} is the (essentially unique) greatest common divisor of a and b.

Unique Factorisation:

Let D be an integral domain with group U of units, and let $p \in D$ be such that $p \neq 0,\ p \notin U$. Then p is said to be irreducible if it has no proper factor. An equivalent definition in terms of ideals is available, as a result of the following theorem.

Theorem (1-2-3):

Let p be an element of a principal ideal domain D. Then the following statements are equivalent:

(i) p is irreducible;
(ii) $\langle p \rangle$ is a maximal proper ideal of D.
(iii) $D/\langle p \rangle$ is a field.

Proof:

(i) ⇒ (ii). Suppose that p is irreducible. Then p is not a unit, and so $\langle p \rangle$ is a proper ideal of D. Suppose, for a contradiction, that there is a (principal) ideal $\langle q \rangle$ such that $\langle p \rangle \subset \langle q \rangle \subset D$. Then $p \in \langle q \rangle$, and so $p = aq$ for some non-unit a. This contradicts the supposed irreducibility of p.

(ii) ⇒ (iii). Let $a + \langle p \rangle$ be a non-zero element of $D/\langle p \rangle$. Then $a \notin \langle p \rangle$, and so the ideal $\langle a \rangle + \langle p \rangle$ properly contains $\langle p \rangle$. We are assuming that $\langle p \rangle$ is maximal, and so it follows that $\langle a \rangle + \langle p \rangle = \{sa + tp : s, t \in D\} = D$. Hence there exist s, t in D such that $sa + tp = 1$, and from this we deduce that $(s + \langle p \rangle)(a + \langle p \rangle) = 1 + \langle p \rangle$. Thus $D/\langle p \rangle$ is a field.

(iii) ⇒ (i). if p is not irreducible, then there exist non-units a and r such that $p = qr$. Then $q + \langle p \rangle$ and $r + \langle p \rangle$ are both non-zero elements of $D/\langle p \rangle$, but

$$(q + \langle p \rangle)(r + \langle p \rangle) = p + \langle p \rangle = 0 + \langle p \rangle.$$

Thus $D/\langle p \rangle$ has divisor of zero, and so certainly is not a field.

An element d of an integral domain D has a factorization into irreducible elements if there exist irreducible elements $p_1, p_2, \dots, p_k$ such that $d = p_1, p_2, \dots, p_k$. The factorization is essentially unique if, for irreducible elements $p_1, p_2, \dots, p_k$ and $q_1, q_2, \dots, q_l$

$$d = p_1 p_2 \dots p_k = q_1 q_2 \dots q_l$$

implies that $k = l$ and, for some permutation $\sigma: \{1,2, \dots, k\} \to \{1,2, \dots, k\}$,

$$p_i \sim q_{\sigma(i)} \quad (i = 1,2, \dots, k).$$

An integral domain D is said to be a factorial domain, or to be a unique factorization domain, if every non-unite $a \neq 0$ of D has an essentially unique factorization into irreducible elements. Here again $\mathbb{Z}$, in which the (positive and negative) prime numbers are the irreducible elements, provides a familiar example: $60 = 2 \times 2 \times 3 \times 5$, and the factorization is essentially unique, for nothing more different than (say) $(-2) \times (-5) \times 3 \times 2$ is possible.

Theorem (1-2-4):

Every principal ideal domain is factorial.

Proof:

We begin with the lemma which at first sight deals with something quite different.

Lemma (1-2-5):

In principal ideal domain there are no infinite ascending chains of ideals.

Proof:

In any integral domain D, an ascending chain

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

of ideals has the property that $I = \cup_{j\geq 1} I_j$ is an ideal. To see this, first observe that, if $a, b \in I$, then there exist k, l such that $a \in J_k, b \in I_l$, and so $a - b \in I_{\max\{k,l\}} \subseteq I$. Also, if $a \in I$ and $s \in D$, then $a \in I_k$ for some k, and so $sa \in I_k \subseteq I$.

Now suppose that D is a principal ideal domain, and let

$$\langle a_1 \rangle \subseteq \langle a_2 \rangle \subseteq \langle a_3 \rangle \subseteq \cdots \qquad (28)$$

be an ascending chain of (principal) ideals. From the previous paragraph, we know that the union of all ideal in this chain must be an ideal, and by our assumption about D, this must be a principal ideal $\langle a \rangle$. Since $a \in \cup_{j\geq 1} \langle a_j \rangle$, we must have that $a \in \langle a_k \rangle$ for some k. Thus $\langle a \rangle \subseteq \langle a_k \rangle$ and, since it is clear that we also have $\langle a_k \rangle \subseteq \langle a \rangle$, it follows that $\langle a \rangle = \langle a_k \rangle$. Hence

$$\langle a_k \rangle = \langle a_{k+1} \rangle = \langle a_{k+2} \rangle \dots = \langle a \rangle,$$

and so the infinite chain of inclusions (28) terminates at $\langle a_k \rangle$.

Returning now to the proof of Theorem (1-2-3), we show first that any $a \neq 0$ in D can be expressed as a product of irreducible elements. Let a be a non-unit in D. Then either a is irreducible, or it has a proper divisor a_1. Similarly, either a_1 is irreducible, or a_1 has a proper divisor a_2. Continuing, we obtain a sequence $a = a_0, a_1, a_2, \dots$ in which, for $i = 1,2, \dots, a_i$ is a proper divisor of a_{i-1}. The sequence must terminate at some a_k, since otherwise we would have an infinite ascending sequence

$$\langle a \rangle \subset \langle a_1 \rangle \subset \langle a_2 \rangle \subset \cdots,$$

and Lemma (1-2-5) would be contradicted. Hence a has a proper irreducible divisor $a_k = z_1$, and $a = z_1 b_1$. If b_1 is irreducible, then the proof is complete. Otherwise we can repeat the argument we used for a to find a proper irreducible divisor z_2 of b_1, and $a = z_1 z_2 b_1$. We continue this process. It too must terminate, since otherwise we would have an infinite ascending sequence

$$\langle a \rangle \subset \langle b_1 \rangle \subset \langle b_2 \rangle \subset \cdots,$$

in contradiction to Lemma (1-2-5). Hence some b_l must be irreducible, and so $a = z_1 z_2 \dots z_{l-1} b_l$ is a product of irreducible elements.

To show that the product is essentially unique, we need another lemma.

Lemma (1-2-6):

Let D be a principal ideal domain, let p be an irreducible element in D, and let $a, b \in D$. Then

$$p|ab \ \Rightarrow p|a \ \text{ or } p|b.$$

Proof:

Suppose that $p|ab$ and $p \nmid a$. Then the greatest common divisor of a and p must be 1, and so there exist s, t in D such that $sa + tp = 1$. Hence $sab + tpb = b$, and so, since p clearly divides $sab + tpd$, it follows that $p|b$.

It is a routine matter to extend this result to product of more than two elements.

Corollary (1-2-7):

Let D be a principal ideal domain, let p be an irreducible element in D, and let $a_1, a_2, \dots, a_m \in D$. Then

$$p|a_1a_2 \dots a_m \Rightarrow p|a_1 \text{ or } p|a_2 \text{ or } \dots or \ p|a_m.$$

To complete the proof of Theorem (2-4), suppose that

$$p_1p_2 \cdots p_k \sim q_1q_2 \cdots q_l, \qquad (29)$$

where $p_1, p_2, \dots, p_k$ and $q_1, q_2, \dots, q_l$ are irreducible. Suppose first that $k = 1$. Then $l = 1$, since $q_1q_2 \dots q_l$ is irreducible, and so $p_1 \sim q_1$. Suppose inductively that, for all $n \geq 2$ and all $k < n$, any statement of the form (29) implies that $k = l$ and that, for some permutation σ of $\{1,2, \dots, k\}$,

$$q_i \sim p_{\sigma(i)} \quad (i = 1,2, \dots, k).$$

Let $k = n$. Since $p_1|q_1q_2 \dots q_l$, it follows from Corollary (1-2-7) that $p_1|q_j$ for some j in $\{1,2, \dots, l\}$. Since q_j is irreducible and p_1 is not a unit, we deduce that $p_1 \sim q_j$, and by cancellation we then have

$$p_2p_3 \cdots p_n \sim q_1 \cdots q_{j-1}q_{j+1} \cdots q_l.$$

By the induction hypothesis, we have that $n - 1 = l - 1$ and that, for $i \in \{1,2, \dots, n\} \backslash \{j\}, q_i \sim p_{\sigma(i)}$ for some permutation σ of $\{2,3, \dots, n\}$. Hence, extending σ to a permutation σ of $\{1,2, \dots, n\}$ by defining $\sigma(1) = j$, we obtain the desired result.

As a consequence of Theorem (1-2-1), we have the following immediate corollary.

Corollary (1-2-8):

Every Euclidean domain is factorial.

Polynomials:

Throughout this section, R is an integral domain and K is a field.

For reasons that will emerge, we begin by describing a polynomial in abstract terms. The more familiar description of a polynomial will appear shortly. A polynomial f with coefficients in R is a sequence $(a_0, a_1, \dots)$, where $a_i \in R$ for all $i \geq 0$, and where only finitely many of $\{a_0, a_1, \dots\}$ are non-zero. If the last non-zero element in the sequence is a_n, we say that f has degree n, and write $\partial f = n$. The entry a_n is called the leading coefficient of f. If $a_n = 1$ we say that the polynomial is monic. In the case where all of the coefficients are 0, it is convenient to ascribe the formal degree of $-\infty$ to the polynomial $(0,0,0,\dots)$, and to make the conventions, for every n in $\mathbb{Z}$,

$$-\infty < n, \qquad -\infty + (-\infty) = -\infty, \qquad -\infty + n = -\infty. \qquad (30)$$

Polynomials $(a, 0,0,\dots)$ of degree 0 or ∞ are called constant. For others of small degree we have names as follows:

∂f	1	2	3	4	5	6
name	linear	quadratic	Cubic	quartic	Quintic	sextic

(Fortunately we shall have no occasion to refer to “septic” polynomials!)

Addition of polynomials is defined as follows:

$$(a_0, a_1, \dots) + (b_0, b_1, \dots) = (a_0 + b_0, a_1 + b_1, \dots).$$

Multiplication is more complicated:

$$(a_0, a_1, \dots)(b_0, b_1, \dots) = (c_0, c_1, \dots),$$

where, for $k = 0,1,2, \dots$,

$$c_k = \sum_{\{(i,j):i+j=k\}} a_j b_j.$$

Thus

$$c_0 = a_0 b_0, \;\; c_1 = a_0 b_1 + a_1 b_0, \;\; c_2 = a_0 b_2 + a_1 b_1 + a_2 b_0, \dots$$

With respect to these two operations, the set P of all polynomials with coefficient in R becomes a commutative ring with unity. Most of the ring axioms are easily verified, and it is clear that the zero element is $(0,0,0,\dots)$, the unity element is $(1,0,0,\dots)$ and the negative of $(a_0,a_1,\dots)$ is $(-a_0,-a_1,\dots)$. The only axiom that causes significant difficultly is the associativity of multiplication. Let $p=(a_0,a_1,\dots),q=(b_0,b_1,\dots),r=(c_0,c_1,\dots)$ be polynomials. (Recall that in each case, only finitely many entries are non-zero). Then $(pq)r=(d_0,d_1,\dots)$, where, for $m=0,1,2,\dots$

$$d_m=\sum_{\{(k,l):k+l=m\}}\left(\sum_{\{(i,j):i+j=k\}}a_ib_j\right)c_l=\sum_{\{(i,j,l):i+j+l=m\}}a_ib_jc_l$$

$$=\sum_{\{(i,n):i+n=m\}}a_i\left(\sum_{\{(j,l):j+l=n\}}b_jc_l\right),$$

Which is the mth entry of $p(qr)$. Thus multiplication is associative.

There is a monomorphism $\theta{:}\,R\to P$ given by

$$\theta(a)=(a,0,0,\dots)\quad(a\in R).$$

We may identity the constant polynomial $\theta(a)=(a,0,0,\dots)$ with the element a of R.

Let X be the polynomial $(0,1,0,0,\dots)$. Then the multiplication rule gives $X^2=(0,0,1,0,\dots),X^3=(0,0,0,1,0,\dots)$ and, in general,

$$X^n=(x_0,x_1,\dots),\text{where } x_m=\begin{cases}1 & \text{if } m=n\\ 0 & \text{otherwise.}\end{cases}$$

The a polynomial

$$(a_0,a_1,\dots,a_n,0,0,\dots)$$

of degree n can be written as

$$\theta(a_0)+\theta(a_1)X+\theta(a_2)X^2+\cdots+\theta(a_n)X^n,$$

or as

$$a_0 + a_1X + a_2X^2 + \cdots + a_nX^n \qquad (31)$$

if we make the identification of $\theta(a_i)$ with a_i.

We have arrived at the common definition of a polynomial, in which X is regarded as an "indeterminate". The notation (31) is clearly useful, and assuredly makes the definition of multiplication seem less arbitrary. It is important, however, to note that we are talking here of polynomial forms, wholly determined by the coefficients a_i, and that X is not a number of R, or indeed of anything else, except of course of the ring P of polynomials. We sometimes write $f = f(X)$ and say that it is a polynomial over R in the indeterminate X. The ring P of all such polynomials is written $R[X]$. We refer to it simply as the polynomial ring of R.

We summarise some of the main facts about polynomials, some of which we already know.

Theorem (1-2-9):

Let D be an integral domain, and let $D[X]$ be the polynomial ring of D. Then

(i) $D[X]$ is an integral domain.

(ii) If $p, q \in D[X]$, then

$$\partial(p + q) \le \max\{\partial p, \partial q\}.$$

(iii) For all p, q in $D[X]$.

(iv) The group of units of $D[X]$ coincides with the group of units of D.

Proof:

(i) We have already noted that $D[X]$ is a commutative ring with unity. To show that there are no divisors of 0, suppose that p and q are non-zero polynomials with leading terms a_m, b_n respectively. The product of p and q then has leading term $a_m b_n$. Since D, by assumption, has no zero divisor, the coefficient $a_m b_n$ is non-zero, and so certainly $pq \neq 0$.

(ii) Let p and q be non-zero. Suppose that $\partial p = m, \partial q = n$, and suppose without loss of generality, that $m \ge n$. If $m > n$ then it is clear that the leading term of $p + q$ is a_m, and so $\partial(p + q) = \max\{\partial p, \partial q\}$. If $m = n$, then we may have $a_m + b_m = 0$, and so all we can say is that $\partial(p + q) \leq \max\{\partial p, \partial q\}$. The conventions established in (30) ensure that this result holds also if one or both of p, q are equal to 0.

(iii) By the argument in (i), if p and q are non-zero, then $\partial(pq) = m + n = \partial p + \partial q$. If one or both of p and q are zero, then the result holds by the conventions established in (30).

(iv) Let $p, q \in D[X]$, and suppose that $pq = 1$. From part (iii) we deduce that $\partial p = \partial q = 0$. Thus $p, q \in D$, and $pd = 1$ if and only if p and q are in the group of units of D.

Since the ring of polynomials over the integral domain D is itself an integral domain, we can repeat the process, and form the ring of polynomials with coefficients in $D[X]$. We need to use a different letter for a new indeterminate, and the new integral domain is $(D[X])[Y]$, more usually denoted by $D[X, Y]$. It consists of polynomials in the two indeterminates X and Y with coefficients in D. This can be repeated, and we obtain the integral domain $D[X_1, X_2, \dots, X_n]$.

The field of fractions of $D[X]$ consists of rational forms

$$\frac{a_0 + a_1X + \cdots + a_mX^m}{b_0 + b_1X + \cdots + b_nX^n},$$

where the denominator is not the zero polynomial. The field is denoted by $D(X)$ (with round rather than square brackets). In a similar way one arrives at the field $D(X_1, X_2, \dots, X_n)$ of rational forms in the n indeterminates $X_1, X_2, \dots, X_n$ with coefficients in D.

The point already made, that a polynomial is wholly determinate by its coefficients, is underlined by the following result.

Theorem (1-2-10):

Let D, D' be integral domain, and let $\varphi: D \to D'$ be an isomorphism. Then the mapping $\varphi: D[X] \to D'[X]$ denoted by

$$\varphi(a_0 + a_1X + \cdots + a_nX^n) = \varphi(a_0) + \varphi(a_1)X + \cdots + \varphi(a_n)X^n$$

is an isomorphism.

Proof:

The proof is routine.

The isomorphism φ is called the canonical extension of φ. A further extension $\varphi^*: D(X) \to D'(X)$ is defined by

$$\varphi^*(f/g) = \hat{\varphi}(f)/\hat{\varphi}(g) \quad \big(f/g \in D(X)\big). \tag{32}$$

We shall be specially interested in the ring $K[X]$ of polynomials over a field K. The group of units of $K[X]$ is the group of units of K, namely the group K^* of non-zero elements of the field K, and the usual way we write $f \sim g$ if $f = ag$ for some a in K^*.

The integral domain $K[X]$ has an important property closed analogous to a property of the domain on integers.

Theorem (1-2-11):

Let K be a field, and let f, g be elements of the polynomial ring $K[X]$, with $g \neq 0$. Then there exists unique elements q, r in $K[X]$ such that $f = qg + r$ and $\partial r < \partial g$.

Proof:

If $f = 0$ the result is trivial, since $f = 0g + 0$. So suppose that $f \neq 0$. The proof is by induction on ∂f. First, suppose that $\partial f = 0$, so that $f \in K^*$. If $\partial g = 0$ also, let $q = f/g$ and $r = 0$; otherwise, let $q = 0$ and $r = f$.

Suppose now that $\partial f = n$, and suppose also that the theorem holds for all polynomials f of all degrees up to $n - 1$. If $\partial g > \partial f$, let $q = 0$ and $r = f$. So suppose now that $\partial g \leq \partial f$. Let f, g have leading terms $a_n X^n, b_m X^m$, respectively, where $m \leq n$. Then the polynomial

$$h = f - \left(\frac{a_n}{b_m} X^{n-m}\right) g$$

has degree at most $n-1$, and so we may assume that there exist q_1, r such that $h = q_1 g + r$, with $\partial r < \partial g$. It follows that $f = qg + r$, where $q = q_1 + (a_n/b_m)X^{n-m}$.

To prove uniqueness, suppose that

$$f = qg + r = q'g + r', \quad \text{with } \partial r' < \partial g.$$

Then $r - r' = (q' - q)g$, and so $\partial\big((q'-q)g\big) = \partial(r - r') < \partial g$. By Theorem (1-2-9), this cannot happen unless $q' - q$, and consequently $r = r'$ also.

Theorem (1-2-12):

If K is a field, then $K[X]$ is a Euclidean domain.

Proof:

The map ∂ does not quite have the properties of the map δ involved in the definition of a Euclidean domain, but if, for all f in $K[X]$ we define $\delta(f)$ as $2^{\partial f}$, with the convention that $2^{-\infty} = 0$, we have exactly the right properties.

As a consequence of Theorem (1-2-1), Corollary (1-2-8) and Theorem (1-2-3) we can summarise the important properties of $K[X]$ as follows.

Theorem (1-2-13):

Let K be a field. Then

(i) Every pair (f, g) of polynomials in $K[X]$ has a greatest common divisor d, which can be expressed as $af + bg$, with a, n in $K[X]$;

(ii) $K[X]$ is a principal ideal domain;

(iii) $K[X]$ is a factorial domain;

(iv) If $f \in K[X]$, then $K[X]/\langle f\rangle$ is a field if and only if f is irreducible.

Multiplication is a little more difficult:

$$(a + bX + \langle X^2 + 1\rangle)(c + dX + \langle X^2 + 1\rangle)$$

$$= ac + (ad + bc)X + bdX^2 + \langle X^2 + 1\rangle$$

$$= (ac - bd) + (ad + bc)X + bd(X^2 + 1) + (X^2 + 1)$$

$$= (ac - bd) + (ad + bc)X + (X^2 + 1).$$

This is reminiscent of the rule for adding multiplying complex numbers. Indeed it is more than reminiscent: the map $\varphi: \mathbb{R}[X]/\langle X^2 + 1\rangle \to \mathbb{C}$, given by

$$\varphi(a + bX + \langle X^2 + 1\rangle) = a + bi \quad (a, b \in \mathbb{R}),$$

is in fact an isomorphism.

We have already emphasized that polynomials, as we have defined them, are polynomial forms, entirely determined by their coefficients. For example, if we write $f = a_0 + a_1X + \cdots + a_nX^n = 0$, we mean that f is the zero polynomial, that is to say, $a_0 = a_1 = \cdots = a_n = 0$. Let D be an integral domain and let $\alpha \in D$. The homomorphism σ_α from $D[X]$ into D is defined by

$$\sigma_\alpha(a_0 + a_1X + \cdots + a_nX^n) = a_0 + a_1\alpha + \cdots + a_n a^\alpha. \qquad (33)$$

The verification that this is a homomorphism is entirely routing, and is omitted. We frequently want to write $\sigma_\alpha(f)$ more simply as $f(\alpha)$.

If $f(\alpha) = 0$, then we say that α is a root, or a zero, of the polynomial f. The following result is crucial to the understanding of roots and factorisations.

Theorem (1-2-14): (The Remainder Theorem):

Let K be a field, let $\beta \in K$ and let f be a non-zero polynomial in $K[X]$. Then the remainder upon dividing f by $X - \beta$ is $f(\beta)$. In particular, β is a root of f if and only if $(X - \beta)|f$.

Proof:

By the division algorithm (Theorem (2-11)), there exist q, r in $K[X]$ such that

$$f = (x - \beta)q + r, \text{ where } \partial r < \partial(x - \beta) = 1 \qquad (34)$$

Thus r is a constant. Substituting β for X, we see that $f(\beta) = r$. In particular, $f(\beta) = 0$ if and only if $r = 0$, that is. If and only if $(X - \beta)|q$.

Irreducible polynomials:

We saw away of constructing the complex field from the real field. This is a very special case of a more general technique.

Theorem (1-2-15):

Let K be a field, and let $g(X)$ be an irreducible polynomial in $K[X]$. Then $K[X]/\langle g\langle X\rangle\rangle$ is a field containing K up to isomorphism.

Proof:

We know from Theorem (2-13) that $K[X]/\langle g\langle X\rangle\rangle$ is a field. The map $\varphi: K \to K[X]/\langle g\langle X\rangle\rangle$ given by

$$\varphi(a) = a + \langle g\langle X\rangle\rangle \quad (a \in K)$$

is easily seen to be a homomorphism. It is even a monomorphism, since

$$a + \langle g\langle X\rangle\rangle = b + \langle g\langle X\rangle\rangle \Rightarrow a - b \in \langle g\langle X\rangle\rangle \Rightarrow a = b.$$

It is clear, therefore, that we will have a highly effective method of constructing new fields provided we have a way of identifying irreducible polynomials. Certainly every linear polynomial is irreducible, and if the field of coefficients is the complex field $\mathbb{C}$, that is the end of the matter, for, by the fundamental theorem of algebra, every polynomial in $\mathbb{C}[X]$ factorises, essentially uniquely, into linear factors. Linear polynomials, it must be said, are of little interest as far as Theorem (2-15) is concerned, for $K[X]/\langle g\langle X\rangle\rangle$ coincides with $\varphi(K)$ in this case, and so is isomorphic to K: if $g(X) = X - a$, then, for all f in $K[X]$ we have that $f = q(X - a) + f(a)$, and so $f + \langle g\rangle = f\langle g\rangle \in \varphi(K)$.

For polynomials in $\mathbb{R}[X]$ the situation is only a little more complicated. Consider a typical polynomial

$$g(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \qquad (35)$$

in $\mathbb{R}[X]$. If $\gamma \in \mathbb{C}/\mathbb{R}$ is a root, then

$$a_n\gamma^n + a_{n-1}\gamma^{n-1} + \cdots + a_1\gamma + a_0 = 0.$$

Hence the complex conjugate of the left-hand side is zero also. That is, since the coefficients $a_0, a_1, \dots, a_n$ are real,

$$a_n\bar{\gamma}^n + a_{n-1}\bar{\gamma}^n + \cdots + a_1\bar{\gamma} + a_0.$$

Thus the non-real roots of the polynomial occur in conjugate pairs, and we obtain a factorization

$$g(X) = a_n(X - \beta_1) \dots (X - \beta_r)(X - \gamma_1)(X - \bar{\gamma}_1) \dots (X - \gamma_s)(X - \bar{\gamma}_s),$$

in $\mathbb{C}[X]$, where $\beta_1, \dots, \beta_r \in \mathbb{R}, \gamma_1, \dots, \gamma_s \in \mathbb{C}\backslash\mathbb{R}, r, s \geq 0$ and $r + 2s = n$. This gives rise to a factorisation

$$a_n(X - \beta_1) \dots (X - \beta_r)(X^2 - (\gamma_1 + \bar{\gamma}_1)X + \gamma_1\bar{\gamma}_1) \dots (X^2 - (\gamma_s + \bar{\gamma}_s)X + \gamma_s\bar{\gamma}_s)$$

in $\mathbb{R}[X]$. In this factorisation the quadratic factors are irreducible in $\mathbb{R}[X]$, for if they had real linear factors, they would have two distinct factorisations in $\mathbb{C}[X]$, and we know that this cannot happen.

We have proved the following result.

Theorem (1-2-16):

The irreducible elements of the polynomial ring $\mathbb{R}[X]$ are either linear or quadratic. Every polynomial (35) in $\mathbb{R}[X]$ has a unique factorisation

$$a_n(X - \beta_1) \dots (X - \beta_r)(X^2 + \lambda_1 X + \mu_1) \dots (X^2 + \lambda_s X + \mu_s),$$

in $\mathbb{R}[X]$, where $a_n \in \mathbb{R}, r, s \geq 0$ and $2s = n$.

We can of course easily determine whether a quadratic polynomial $aX^2 + bX + c$ in $\mathbb{R}[X]$ is irreducible: it is irreducible if and only if the discriminant $b^2 - 4ac < 0$.

This much is relatively straightforward. Unfortunately, we shall be mostly interested in $\mathbb{Q}[X]$, and here the situation is not so easy, for, as we shall see, in $\mathbb{Q}[X]$ there are irreducible polynomials of arbitrarily large degree.

Quadratic polynomials present no great problem.

Theorem (1-2-17):

Let $g(X) = X^2 + a_1X + a_0$ be a polynomial with coefficients in $\mathbb{Q}$. Then:

(i) If $g(X)$ is irreducible over $\mathbb{R}$, then it is irreducible over $\mathbb{Q}$;

(ii) If $g(X) = (X - \beta_1)(X - \beta_2)$, with $\beta_1, \beta_2 \in \mathbb{R}$, then $g(X)$ is irreducible in $\mathbb{Q}[X]$ if and only if β_1 and β_2 are irrational.

Proof:

(i) Let $g(X)$ be irreducible over $\mathbb{R}$. If $g(X) = (X - q_1)(X - q_2)$ were a factorisation in $\mathbb{Q}[X]$, it would also be a factorisation in $\mathbb{R}[X]$, and we would have a contradiction.

(ii) If β_1, β_2 were rational we would have a factorisation in $\mathbb{Q}[X]$, and $g(X)$ would not be irreducible. If β_1, β_2 are irrational, then $(X - \beta_1)(X - \beta_2)$ is the only factorisation in $\mathbb{R}[X]$, and so a factorisation in $\mathbb{Q}[X]$ into linear factors is not possible.

Remark (1-2-18):

With regard to part (ii) of the theorem, it is clear that, if one or other of β_1, β_2 is irrational, then both are irrational.

Theorem (1-2-19): (Gauss's Lemma):

Let f be a polynomial in $\mathbb{Z}[X]$, irreducible over $\mathbb{Z}$. Then f, considered as a polynomial in $\mathbb{Q}[X]$, is irreducible over $\mathbb{Q}$.

Proof:

Suppose, for a contradiction, that $f = gh$, with $g, h \in \mathbb{Q}[X]$ and $\partial g, \partial h < \partial f$. Then there exists a positive integer n such that $nf = g'h'$, where $g', h' \in \mathbb{Z}[X]$. Let us suppose that n is the smallest positive integer with this property. Let

$$g' = a_0 + a_1X + \cdots + a_kX^k, \quad h' = b_0 + b_1X + \cdots + b_lX^l.$$

If $n = 1$, then $g' = g, h' = h$, and we have an immediate contradiction. Otherwise, let p be a prime factor of n.

Lemma (1-2-20):

Either p divides all the coefficients of g' or p divides all the coefficients of h'.

Proof:

Suppose, for a contradiction, that p does not divide all the coefficients of g', and that p does not divide all the coefficients of h'. Suppose that p divides $a_0, \dots, a_{i-1}$, but $p \nmid a_i$, and that p divides $b_0, \dots, b_{j-1}$, but $p \nmid b_j$. The coefficient of X^{i+j} in nf is

$$a_0 b_{i+j} + \cdots + a_i b_j + \cdots + a_{i+j} b_0.$$

In this sum, all the terms preceding $a_i b_j$ are divisible by p, since p divides $a_0, \dots, a_{j-1}$; and all the terms following $a_i b_j$ are divisible by p, since p divides $b_0, \dots, b_{j-1}$. Hence only the terms $a_i b_j$ is not divisible by p, and it follows that the coefficient of X^{i+j} in nf is not divisible by p. This gives a contradiction, since the coefficients of f are integers, and so certainly all the coefficients of nf are divisible by p.

Returning now to the proof of theorem (1-2-19), we may suppose, without loss of generality, that $g' = pg'' \in \mathbb{Z}[X]$. It follows that $(n/p)f = g''h'$, and this contradicts the choice of n as the least positive integer with this property. Hence a factorisation over $\mathbb{Q}$ is not possible, and f is irreducible over $\mathbb{Q}$.

We have seen that there is no difficulty in determining the irreducibility of quadratic polynomials in $\mathbb{Q}[X]$. Theorem (1-2-19) makes it reasonably straight-forward to deal with monic cubic polynomials over $\mathbb{Z}$.

This technique will not work for a polynomial of degree exceeding 3, and indeed there is no easy way to determine irreducibility over $\mathbb{Q}$. One important technique, due to Eisenstein, is as follows.

Theorem (1-2-21): (Eisenstein’s criterion):

Let

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

be a polynomial in $\mathbb{Z}[X]$. Suppose that there exists a prime number p such that

(i) $p \nmid a_n$,
(ii) $p|a_i \quad (i = 0, \dots, n-1)$,
(iii) $p^2 \nmid a_0$.

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

Proof:

By Gauss’s lemma (Theorem (2-19)), it is sufficient to prove that f is irreducible over $\mathbb{Z}$. Suppose, for a contradiction, that $f = gh$, wher

$$g = b_0 + b_1X + \cdots + b_rX^r, \quad h = c_0 + c_1X + \cdots + c_sS^s,$$

with $r, s < n$ and $r + s = n$. Since $a_0 = b_0c_0$, it follows from (ii) that $p|b_0$ or $p|c_0$. Since $p^2 \nmid a_0$, the coefficients b_0 and c_0 cannot both be divisible by p, and we assume, without loss of generality, that

$$p|b_0, \quad p \nmid c_0 \qquad (36)$$

Suppose inductively that p divides $b_0, b_1, \dots, b_{k-1}$, where $1 \leq k \leq r$. Then

$$a_k = b_0c_k + b_1c_{k-1} + \cdots + b_{k-1}c_1 + b_kc_0.$$

Since p divides each of $a_k, b_0c_k, b_1c_{k-1}, \dots, b_{k-1}c_1$, it follows that $p|b_kc_0$, and hence, form (36), $p|b_k$.

We conclude that $p|b_r$, and so, since $a_n = b_rc_s$, we have that $p|a_n$, a contradiction to the assumption (i). hence f is irreducible.

Remark (1-2-22):

It is clear from Theorem (1-2-21) that there exist irreducible polynomials in $\mathbb{Q}[X]$ of arbitrary high degree.

Another device for determining irreducibility over $\mathbb{Z}$ (and consequently over $\mathbb{Q}$) is to map the polynomial onto $\mathbb{Z}_p[X]$ for some suitably chosen prime p. Let $g = a_0 + a_1X + \cdots + a_nX^n \in \mathbb{Z}[X]$, and let p be a prime not dividing a_n. For each i in $\{0,10 \dots, n\}$, let $\bar{a}_i$ denote the residue class $a_i + \langle p \rangle$ in the field $\mathbb{Z}_p = \mathbb{Z}/\langle p \rangle$, and write the polynomial $\bar{a}_0 + \bar{a}_1X + \cdots + \bar{a}_nX^n$ as $\bar{g}$. Our choice of p ensures that $\partial\bar{g} = n$. Suppose that$g = uv$, with $\partial u, \partial v < \partial f$ and $\partial u + \partial v = \partial g$. Then $\bar{g} = \bar{u} + \bar{v}$. If we can show that $\bar{g}$ is irreducible in $\mathbb{Z}_p[X]$, then we have a contradiction, and we deduce that g is irreducible. The advantage of transferring the problem from $\mathbb{Z}[X]$ to $\mathbb{Z}_p[X]$ is that $\mathbb{Z}_p$ is finite, and the verification of irreducibility is a matter of checking a finite number of cases.

Remark (1-2-23):

The choice of prime p is, of course, crucial. If, we had used $p = 2$, we would have obtained $\bar{g} = X^4 + 1$, and in $\mathbb{Z}_2[X]$ this is far from irreducible, since $X^4 + 1 = (X + 1)^4$. It is important to realize that if our $\bar{g}$ is not irreducible then we can draw no conclusion at all.

In these chapter we study the field extension , defined the
trans centennial and algebraic elements, monic polynomial and we give some application geometry , also we study the splitting field**.**

Section (2-1):

The Degree of an Extension:

In this section it necessary to have some knowledge of the basic concepts of linear algebra, including linear independence, spanning sets, bases and dimension.

If K, L are fields and $\varphi: K \to L$ is a monomorphism, we say that L is an extension of K, and we sometimes find it result to write "$L: K$ is a (field) extension". As we have seen, this is not essentially different from saying that K is a subfield of L, since we may always identify K with its image $\varphi(K)$. Then L can be regarded as a vector space over K, since the vector space axioms

(V1) $(x+y)+z = x+(y+z) \quad (x, y, z \in L)$,

(V2) $x+y = y+x \quad (x, y \in L)$,

(V3) there exists 0 in L such that $x+0 = x \quad (x \in L)$,

(V4) for all x in L there exists $-x$ in L such that $x+(-x) = 0$,

(V5) $a(x+y) = ax+ay \quad (a \in K,\ x, y \in L)$,

(V6) $(a+b)x = ax+bx \quad (a, b \in K,\ x \in L)$,

(V7) $(ab)x = a(bx) \quad (a, b \in K,\ x \in L)$,

(V8) $1x = x \quad (x \in L)$,

are all consequences of the field axioms for L. Hence there exists a basis of L over K. Different bases have the same cardinality, and there is a well-defined dimension of L, equal to the cardinality of an arbitrarily chosen basis. The term used in field theory for this dimension is the degree of L over K, or the degree of the extension $L: K$; and we denote it by $[L: K]$. We say that L is a finite extension of K if $[L: K]$ is finite; otherwise L is an infinite extension.

Theorem (2-1-1):

Let $L: K$ be a field extension. Then $L = K$ if and only if $[L: K] = 1$.

Proof:

This is a standard property of finite-dimensional vector spaces, but for completeness we prove it here.

Suppose first that $L = K$. Then $\{1\}$ is a basis for L over K, since every element x of L is expressible as $x1$, with x in K. Thus $[L: K] = 1$.

Conversely, suppose that $[L: K] = 1$, and that $\{x\}$, where $x \neq 0$, is a basis of L over K. Thus, in particular, there exists a in K such that $1 = ax$, and so $1 = 1/a \in K$. For every y in L there exists b in K such that $y = bx = b/a$. Thus $y \in K$. We have shown that $L = K$.

Suppose now that we have field extensions $L: K$ and $M: L$. That is, there are monomorphisms $\alpha: K \to L$. Then $\beta \circ \alpha: K \to M$ is a monomorphism, and so M is an extension of K. With these definitions we now have the following theorem, in which the equality is intended to include the information that if either of $[M: L]$ and $[L: K]$ is infinite then so is $[M: K]$. We shall make the usual identifications, regarding K as a subfield of L and L as a subfield of M.

Theorem (2-1-2):

Let $L: K$ and $M: L$ be field extensions. Then

$$[M: L][L: K] = [M: K].$$

Proof:

Let $\{a_1, a_2, \dots, a_r\}$ be a linearly independent subset of M over L, and let $\{b_1, b_2, \dots, b_s\}$ be a linearly independent subset of L over K. We show that

$$\left\{a_i b_j \colon\ i = 1,2, \dots, r, \qquad j = 1,2, \dots, s\right\} \qquad (37)$$

is a linearly independent subset of M over K. For let us suppose that

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\lambda_{ij}a_ib_j = 0,$$

with $\lambda_{ij} \in K$ for all i and j. Rewriting this as

$$\sum_{i=1}^{r}\left(\sum_{j=1}^{s}\lambda_{ij}b_j\right)a_i = 0,$$

we deduce, since the a_i are linearly independent over L, that

$$\sum_{j=1}^{s}\lambda_{ij}b_j = 0 \quad (i = 1,2,\dots,r).$$

Then, since the b_j are linearly independent over K, we conclude that $\lambda_{ij} = 0$ for all i and j.

If either of $[M:L]$ and $[L:K]$ is infinite, then either r or s can be made arbitrarily large, and so the set (1) can be made arbitrarily large. Hence $[M:K]$ is infinite. So now suppose that

$$[M:L] = r < \infty, \quad [L:K] = s < \infty,$$

that $\{a_1, a_2, \dots, a_r\}$ is a basis of M over L, and that $\{b_1, b_2, \dots, b_s\}$ is a basis of L over K. For each z in M there exist $\lambda_1, \lambda_2, \dots, \lambda_r$ in L such that $z = \sum_{i=1}^{r}\lambda_i a_i$. Also, for each λ_i there exist $\mu_1, \mu_2, \dots, \mu_{is}$ in K such that $\lambda_i = \sum_{j=1}^{s}\mu_{ij}b_j$. Hence

$$z = \sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}(a_ib_j).$$

The set (1), being both linearly independent and a spanning set for M over K, is a basis, and so

$$[M:K] = rs = [M:L][L:K].$$

The following easy consequence is worth recording at this stage.

Corollary (2-1-3):

Let $K_1, K_2, \dots, K_n$ be fields, and suppose that $K_{i+1}\colon K_i$ is an extension, for $1 \le i \le n-1$. Then

$$[K_n\colon K_1] = [K_n\colon K_{n-1}][K_{n-1}\colon K_{n-2}] \dots [K_2\colon K_1].$$

Extensions and Polynomials:

We are familiar with the observation that equation $X^2 = 2$ cannot be solve within the rational field, but has the solutions $\pm\sqrt{2}$ in the field $\mathbb{R}$ of real numbers. In fact its solution lie within a much smaller field than $\mathbb{R}$, namely the extension

$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$$

of $\mathbb{Q}$. It is not perhaps quite obvious that this is a field, but it is easy to verify the subfield conditions (3). If $a + b\sqrt{2}, c + d\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$, then

$$(a + b\sqrt{2}) - (c + d\sqrt{2}) = (a - c) + (b - d)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$$

and (if $c + d\sqrt{2} \neq 0$)

$$(a + b\sqrt{2})(c + d\sqrt{2})^{-1} = \frac{(a + b\sqrt{2})(c - d\sqrt{2})}{(c + d\sqrt{2})(c + d\sqrt{2})} = u + v\sqrt{2},$$

where

$$u = \frac{ac - 2bd}{c^2 - 2d^2}, \quad v = \frac{bc - ad}{c^2 - 2d^2}.$$

Note that from the irrationality of $\sqrt{2}$ it follows that $c^2 - 2d^2 = 0$ if and only if $c = d = 0$.

This is a special case of a general result, which we now proceed to investigate.

We begin with something quite general. Let K be a sub field L and let S be a subset of L. Let $K(S)$ be the intersection of all the subfields of L containing $K \cup S$. (There is at least one such subfield, namely L itself.) It is clear that $K(S)$ is the smallest subfield containing $K \cup S$, and we call it the subfield of L generated over K by S. If $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ is finite, we write $K(S)$ as $K(\alpha_1, \alpha_2, \dots, \alpha_n)$.

Theorem (2-1-4):

The subfield $K(S)$ of the field L coincides the set E of all elements of L that can be expressed as quotients of finite linear combinations (with coefficients in K) of finite products of elements of S.

Proof:

Denote by P the set of all finite linear combination of finite products of elements of S. If $p, q \in P$, then $p \pm q, pq \in P$. Hence, if $x = p/q$ and $y = r/s$ are typical elements of E, with p, q, r, s in P and $q, s \neq 0$, we see that $x - y = (ps - qr)/(qs) \in E$, and (provided $y \neq 0$) $x/y = (ps)/(qr) \in E$. We deduce that E is a subfield of L containing K and S, and so $K(S) \subseteq E$. Also any subfield containing K and S must contain all finite products of elements in S, all linear combinations of such products, and all quotients of such linear combinations. In short, it must contain E. Hence, in particular, $K(S) \supseteq E$.

Of particular interest is the case where S has just one element $\alpha (\notin K)$. Then, from Theorem (2-1-4), we deduce that $K(\alpha)$ is the set of all quotients polynomials in α with coefficients in K. We say that $K(\alpha)$ is a simple extension of K. The link with polynomials is important, as the next result shows:

Theorem (2-1-5):

Let L be a field, let K be a subfield and let $\alpha \in L$. Then either

(i) $K(\alpha)$ is isomorphic to $K(X)$, the field of all rational forms with coefficients in K; or

(ii) There exists a unique monic irreducible polynomial m in $K[X]$ with property that, for all f in $K[X]$,

   a) $f(\alpha) = 0$, if and only if m/f;

   b) The field $K(\alpha)$ coincides with $K[\alpha]$, the ring of all polynomials in α with coefficients in K; and

   c) $[K[\alpha]: K] = \partial m$.

Proof:

Suppose first that there is no non-zero polynomial f in $K[X]$ such that $f(\alpha) = 0$. (This means in particular that $\alpha \notin K$, since in that case we may take f as $X - \alpha$) The there is a mapping $\varphi: K \to K(\alpha)$ given by

$$\varphi(f/g) = f(\alpha)/g(\alpha),$$

(for we are assuming that $g(\alpha) = 0$ only if g is the zero polynomial). It is routine to verify that φ is a homomorphism, and it clearly maps onto $K(\alpha)$. To see that it is well defined and one-to-one, suppose that f, g, p, q are polynomials, with $g, q \neq 0$. Then

$$\begin{aligned} \varphi(f/g) = \varphi(p/q) &\Leftrightarrow f(\alpha)q(\alpha) - p(\alpha)g(\alpha) = 0 \text{ in } L \\ &\Leftrightarrow fq - pg = 0 \text{ in } K[X] \\ &\Leftrightarrow f/g = p/q \text{ in } K(\alpha). \end{aligned}$$

Now suppose that there does exist a non-zero polynomial g such that $g(\alpha) = 0$. Indeed, let us suppose that g is a polynomial with least degree having this property. If a is the leading coefficient of g, then g/a is a monic polynomial. Denote g/a by m. Certain $m(\alpha) = 0$.

It is clear that $f(\alpha) = 0$ if m/f. Conversely, suppose that $f(\alpha) = 0$. Then by Theorem (1-2-11), $f = qm + r$, where $\partial r < \partial m$. Now

$$0 = f(\alpha) = q(\alpha)m(\alpha) + r(\alpha) = 0 + r(\alpha) = r(\alpha).$$

Since $\partial r < \partial m$, this gives a contradiction unless r is the zero polynomial. Hence $f = qm$, and so m/f.

To show that m is unique, suppose that m' is another polynomial with the same properties. Then $m(\alpha) = m' = 0$ and so m/m' and m'/m. Since both polynomials are monic, we conclude that $m = m'$.

To show that m is irreducible, suppose for a contradiction, there exist polynomials p and q such that $pq = m$, with $\partial p, \partial q < \partial m$. Then $p(\alpha)q(\alpha) =$

$m(\alpha) = 0$, and so either $p(\alpha) = 0$ or $q(\alpha) = 0$. This is impossible, since both p and q are of smaller degree than m.

Next, consider a typical element $f(\alpha)/g(\alpha)$ in $K(\alpha)$, where $g(\alpha) \neq 0$. Then m does not divide g, and it follows, since m has no divisors other than itself and 1, that the greatest common divisor of g and m is 1. Hence, by Theorem (1-2-13), there exist polynomials a, b such that $ag + bm = 1$, and so, substituting α for X, we have $a(\alpha)g(\alpha) = 1$. Thus

$$\frac{f(\alpha)}{g(\alpha)} = f(\alpha)a(\alpha) \in K[\alpha].$$

Finally, suppose that $\partial m = n$, and let $p(\alpha) \in K[\alpha] = K(\alpha)$, where p is a polynomial. Then $p = qm + r$, where $\partial r < \partial m = n$. It follows that $p(\alpha) = r(\alpha)$, and so there exist $c_0, c_1, \dots, c_{n-1}$ (the coefficients of r, some of which may, of course, be zero) in K such that $p(\alpha) = c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}$. Hence $\{1, \alpha, \dots, \alpha^{n-1}\}$ is a spanning set for $K[\alpha]$.

Moreover, the set $\{1, \alpha, \dots, \alpha^{n-1}\}$ is linearly independent over K, for elements $a_0, a_1, \dots, a_{n-1}$ of K are such that $a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0$, then $a_0 = a_1 = \cdots = a_{n-1} = 0$, since otherwise we would have a non-zero polynomial $p = a_0 + a_1X + \cdots + a_{n-1}X^{n-1}$ of degree at most $n - 1$ such that $p(\alpha) = 0$. Thus $\{1, \alpha, \dots, \alpha^{n-1}\}$ is a basis of $K(\alpha)$ over K, and so $[K(\alpha): K] = n$.

The polynomial m defined above is called the minimum polynomial of the element α.

Theorem (2-1-6):

Let $K(\alpha)$ be a simple transcendental extension of a field K. Then the degree of $K(\alpha)$ over K is infinite.

Proof:

The elements $1, \alpha, \alpha^2, \dots$ are linearly independent over K.

An extension L of K is said to be an algebraic extension if every element of L is algebraic over K. Otherwise L is a transcendental extension.

Theorem (2-1-7):

Let $L: K$ and $M: L$ be field extensions, and let $\alpha \in M$. If α is algebraic over K, then it is also algebraic over L.

Proof:

Since α is algebraic over K, there exists a non-zero polynomial f in $K[X]$ such that $f(\alpha) = 0$. Since f is also in $L[X]$, we deduce that α is algebraic over L.

Theorem (2-1-8):

Let L be an extension of a field K, and let $A(L)$ be the set of all elements in L that are algebraic over K. Then $A(L)$ is a subfield of L.

Proof:

Suppose that $\alpha, \beta \in A(L)$. Then

$$\alpha - \beta \in K(\alpha, \beta) = (K[\alpha])[\beta].$$

By Theorem (2-1-9), β is algebraic over $K[\alpha]$, and so both $[K[\alpha]: K]$ and $\big[(K[\alpha])[\beta]: K[\alpha]\big]$ are finite. From Theorem (2-1-5) it follows that $[K(\alpha, \beta): K]$ is finite, and so, by Theorem (2-1-8), $\alpha - \beta$ is algebraic over K. An identical argument shows that $\alpha/\beta \in A(L)$ for all α and $\beta(\neq 0)$ in $A(L)$.

If we take K as the field $\mathbb{Q}$ of rational numbers and L as the field $\mathbb{C}$ of complex numbers, then $A(L)$ is the field $\mathbb{A}$ of algebraic numbers.

Theorem (2-1-9):

The field $\mathbb{A}$ of algebraic numbers is countable.

Proof:

The proof depends on some knowledge of the arithmetic of infinite cardinal numbers. It is know that $\mathbb{Q}$ is countable. To put it in the standard notation for cardinal numbers, $|\mathbb{Q}| = \aleph_0$. Since $\mathbb{Q} \subseteq \mathbb{A}$, we know that $|\mathbb{A}| \geq \aleph_0$.

Now, the number of monic polynomials of degree n with coefficients in $\mathbb{Q}$ is $\aleph_0^n = \aleph_0$. Each such polynomial has at most n distinct roots in $\mathbb{C}$, and so the number of monic polynomials of degree n is at most $n\aleph_0 = \aleph_0$. Hence the number of roots of monic polynomials of all possible degree is at most $\aleph_0 . \aleph_0 = \aleph_0$. Thus $|\mathbb{A}| \leq \aleph_0$, and the result follows.

Remark (2-1-10):

We cannot assert equality in the formula (2). For example,

$$[\mathbb{Q}\sqrt{2}: \mathbb{Q}] = [\mathbb{Q}\sqrt{3}: \mathbb{Q}] = [\mathbb{Q}\sqrt{6}: \mathbb{Q}] = 2,$$

but $[\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{6}): \mathbb{Q}] = 4$.

Polynomials and Extensions:

In the last section, called Extensions and polynomials, the main result was that every simple algebraic extension $K(\alpha)$ within a field L is associated with a polynomial, the minimum polynomial of α. We required to exist within a field L. By changing the order of the words in the title we change the question: given a field K and a monic irreducible polynomial m with coefficients in K, can we create a field, an extension of K, containing an element α whose minimum polynomial is m?

Let K be a field, and let $m \in K[X]$ be irreducible and monic. Let $L = K[X]/\langle m \rangle$. Then L is a field, by Theorem (1-2-3). By Theorem (1-2-15), the mapping $a \longmapsto a + \langle m \rangle$ is a monomorphism from K into L, and so L is an extension of K. Let $\alpha = X + \langle m \rangle$. Then, for each polynomial $f = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ in $K[X]$,

$$\begin{aligned} f(\alpha) &= a_0 + a_1\alpha + \cdots + a_n\alpha^n \\ &= a_0 + a_1(X + \langle m \rangle) + a_2(X + \langle m \rangle)^2 + \cdots + a_n(X + \langle m \rangle)^m \\ &= a_0 + a_1(X + \langle m \rangle) + a_2(X^2 + \langle m \rangle) + \cdots + a_n(X^n + \langle m \rangle) \\ &= (a_0 + a_1X + a_2X^2 + \cdots + a_nX^n) + \langle m \rangle \\ &= f + \langle m \rangle, \end{aligned}$$

and so $f(\alpha) = 0$ if and only if $m|f$. Thus m is the minimum polynomial of α. We

Theorem (2-1-11):

Let K, K' be field, and let $\varphi: K \to K'$ be an isomorphism with canonical extension $\hat{\varphi}: K[X] \to K'[X]$. Let $f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_0$ be an irreducible polynomial of degree n with coefficients in K, and let $f' = \hat{\varphi}(f) = \varphi(a_n) X^n + \varphi(a_{n-1}) X^{n-1} + \cdots + \varphi(a_0)$. Let L be an extension of K containing a root α of f, and let L' be an extension of K' containing a root α' of f'. Then there is an isomorphism ψ from $K'[\alpha']$, an extension of φ.

Proof:

The field $K[\alpha]$ consists of polynomials $b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$, with the obvious addition, and where multiplication is carried out using the equation

$$\alpha^n = -\frac{1}{a_n}(a_n \alpha^{n-1} + \cdots + a_0).$$

The mapping ψ is defined by

$$\psi(b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}) = \varphi(b_0) + \varphi(b_1)\alpha' + \cdots + \varphi(b_{n-1})(\alpha')^{n-1}.$$

In a more compact notation, we have that, for each polynomial u in $K[X]$ with $\partial u < n$,

$$\psi\big(u(\alpha)\big) = \big(\hat{\varphi}(u)\big)(\alpha').$$

It is clear that ψ is one-one and onto, and that it extends the isomorphism $\varphi: K \to K'$.

Let $u, v \in K[X]$, where $\partial u, \partial v \leq n-1$. Then it is clear that

$$\psi\big(u(\alpha) + v(\alpha)\big) = \psi\big(u(\alpha) + v(\alpha)\big).$$

The corresponding equality for multiplication is less clear. We multiply $u(\alpha)$ and $v(\alpha)$ and use the minimum polynomial to reduce the answer to $w(\alpha)$, say where $\partial w \leq n-1$. Precisely, we use the division algorithm to write $uv = qm + w$, where $\partial e < n$. Hence

$$\psi\big(u(\alpha)v(\alpha)\big) = \psi\big(w(\alpha)\big) = \big(\hat{\varphi}(w)\big)(\alpha'). \tag{38}$$

The isomorphism $\hat{\varphi}$ assures us that the division algorithm in $K'[X]$ gives

$$\hat{\varphi}(u)\hat{\varphi}(v) = \hat{\varphi}(q)\hat{\varphi}(m) + \hat{\varphi}(w) \qquad (39)$$

Hence

$$\begin{aligned}\psi\big(u(\alpha)\big)\psi\big(v(\alpha)\big) &= \big(\hat{\varphi}(u)\big)(\alpha')\big(\hat{\varphi}(v)\big)(\alpha') \\ &= \big(\hat{\varphi}(u)\hat{\varphi}(v)\big)(\alpha') \\ &= \big(\hat{\varphi}(q)\hat{\varphi}(m) + \hat{\varphi}(w)\big)(\alpha') \quad \big(\text{from } (40)\big) \\ &= \big(\hat{\varphi}(q)\big)(\alpha')\big(\hat{\varphi}(m)\big)(\alpha') + \big(\hat{\varphi}(w)\big)(\alpha') \\ &= \big(\hat{\varphi}(w)\big)(\alpha') \quad \left(\text{since } \big(\hat{\varphi}(m)\big)(\alpha') = 0\right)\end{aligned}$$

Comparing this with (39) gives the required result.

It is worth recording as a corollary the result we obtain when K and K' are the same field.

Remark (2-1-12):

By the fundamental theorem of algebra every polynomial with coefficients in $\mathbb{C}$ factories into linear factors. In particular, if m is irreducible in $\mathbb{Q}[X]$, then m factorises completely in $\mathbb{C}[X]$. If we know these factors, it is therefore easier and more natural to deal, for example, with the subfield $\mathbb{Q}\big[i\sqrt{3}\big] = \big\{a + bi\sqrt{3}\colon a, b \in \mathbb{Q}\big\}$ of $\mathbb{C}$ than with $\mathbb{Q}[X]/\langle X^2 + 3\rangle$. The two fields are, of course, isomorphic to each other.

If, however, we are dealing, say, with extensions of $\mathbb{Z}_2$, then we are in effect obliged to carry out the more abstract procedure.

Section (2-2): Applications to geometry:

Ruler and Compasses Constructions:

Undoubtedly one of the early triumphs of abstract algebra was the light it shed on some classical problems of Greek mathematics, the most significant of which was referred to as “squaring the circle”. This is one of very few phrases from serious mathematics to have entered the language, though a (totally unscientific) poll of non-mathematical friends suggests that its mathematical meaning is not even remotely understood. “Something to do with πr^2, is it?” is a common answer, and indeed that is correct, but it does not get to the heart of the matter.

Remark (2-2-1):

This construction works just as well if C lies on the line AB.

An Algebraic Approach:

A Cartesian coordinate system in the plane depends on

(i) Specifying two axes at right angles to each other, meeting at a point O, the origin;

(ii) Choosing a point I, distinct from O, on one of the axes, and giving it coordinates $(1,0)$.

Let B_0 be a set of points in the plane. There are two permitted operations on the points of B_0:

(1) (Ruler) though any two points of B_0, draw a straight line;

(2) (Compasses) draw a circle whose centre is a point in B_0, and whose radius is the distance between two points in B_0.

Any point which is an intersection of two lines, or two circles, or a line and a circle, obtained by means of the operations (1) and (2), is said to be constructed from B_0 in one step. Denote the set of such points by $C(B_0)$, and let $B_1 = B_0 \cup C(B_0)$. We can continue the process, defining

$$B_n = B_{n-1} \cup C(B_{n-1}) \quad (n = 1,2,3, \ldots). \qquad (40)$$

A point is said to be constructible from B_0 if it belongs to B_n for some n. A point that is constructible from $\{O, I\}$ is said to be constructible.

Theorem (2-2-2):

Let P be a constructible point, belonging (in the notation (41)) to B_n, where $B_0 = \{(0,0)(1,0)\}$. For $n = 0,1,2, \ldots$, let K_n be the field generated over $\mathbb{Q}$ by B_n. Then $[K_n : \mathbb{Q}]$ is a power of 2.

Proof:

It is clear that $[K_0 : \mathbb{Q}] = 1 = 2^0$. We suppose inductively that $[K_{n-1} : \mathbb{Q}] = 2^k$ for some $k \geq 0$. We require to show that $[K_n : K_{n-1}]$ is a power of 2.

New points in B_n are obtained by

(1) The intersection of two lines; or
(2) The intersection of a line and a circle; or
(3) The intersection of two circles.

Case (1) is the easiest. Suppose that we have lines AB and CD, where $A = (a_1, a_2)$, $B = (b_1, b_2), C = c_1, c_2, D = d_1, d_2$, and that all these coordinates are in K_{n-1}. The equations of the lines are

$$(y - b_2)(a_1 - b_1) = (x - b_1)(a_2 - b_2), \quad (y - d_2)(c_1 - d_1) = (x - d_1)(c_2 - d_2),$$

and the coordinates of their intersection are obtained by solving these two simultaneous linear equations. The details are unimportant: the crucial observation is that the solution process involves only rational operations (addition, subtraction, multiplication and division), and so takes place entirely within the field K_{n-1}. The coordinates of the intersection of AB and CD lie inside the field K_{n-1}.

For case (2), suppose that we have a line AB intersecting a circle with center C and radius PQ, where P, Q are points with coordinates in K_{n-1}.

Taking the coordinates of A, B and C as in the previous paragraph, with all coordinates in K_{n-1}, we must solve the equations

$$(y - b_2)(a_1 - b_1) = (x - b_1)(a_2 - b_2),$$

$$(x - c_1)^2 + (y - c_2)^2 = r^2,$$

where $r^2 \in K_{n-1}$. We have to solve two simultaneous equations, one linear and one quadratic, with coefficients in K_{n-1}. Again the details are unimportant, but the standard method of doing this is to express y in terms of x using the linear equation, and then to substitute in the equation of circle, obtaining a quadratic equation in x, with coefficients in K_{n-1}. The standard solution involves $\sqrt{\Delta}$, where Δ is the discriminate of the quadratic equation, and so the coordinates of the points of intersection belong to the field $K_{n-1}\left[\sqrt{\Delta}\right]$. (This will coincide with K_{n-1}, if by chance, $\sqrt{\Delta} \in K_{n-1}$).

For case (3), suppose that we have a circle with center A and radius r and a circle with center B with radius s, where $r, s \in K_{n-1}$. With the same notation as before, we must solve the simultaneous equations

$$(x - a_1)^2 + (y - a_2)^2 = r^2,$$

$$(x - c_1)^2 + (y - c_2)^2 = s^2.$$

By subtracting we obtain a linear equation (in fact the equation of the chord connection the points of intersection of the circles) and so we have reduced this case to case (2).

The conclusion is that the elements in K_n are either in K_{n-1} or in $K_{n-1}\left[\sqrt{\Delta}\right]$ for some Δ in K_{n-1}. Hence, for some $k \geq 0$.

$$K_n = K_{n-1}\left(\sqrt{\Delta_1}, \sqrt{\Delta_2}, \dots, \sqrt{\Delta_k}\right),$$

and so $[K_n : K_{n-1}]$ is a power of 2.

In the light of this theorem, we now consider the three classical problems mentioned at the beginning of the chapter.

Splitting Field:

When we consider a polynomial such as $X^2 + 2$ and extend the field $\mathbb{Q}$ to $\mathbb{Q}\left[i\sqrt{2}\right]$ by adjoining one of the complex roots of the polynomial, we obtain a "bonus", in that the other root $-i\sqrt{2}$ is also in the extended field. Over $\mathbb{Q}\left[i\sqrt{2}\right]$ we have that

$$X^2 + 2 = \left(X - i\sqrt{2}\right)\left(X + i\sqrt{2}\right),$$

We say that the polynomial splits completely (into linear factor) over $\mathbb{Q}\left[i\sqrt{2}\right]$. It is indeed clear that this must happen for a polynomial of degree 2, since the "other" factor must also be linear.

By contrast, if we look at the cubic polynomial $X^3 - 2$, which is irreducible over $\mathbb{Q}$ (by the Eisenstein criterion) and if we extend $\mathbb{Q}$ to $\mathbb{Q}[\alpha]$, where $\alpha = \sqrt[3]{2}$, we obtain the factorization

$$X^3 - 2 = (X - \alpha)(X^2 + \alpha X + \alpha^2),$$

but the quadratic factor is certainly irreducible over $\mathbb{Q}[\alpha]$. (it is indeed irreducible over $\mathbb{R}$, since the discriminant is $-3\alpha^2$). Over the complex field we have the factorization

$$X^3 - 2 = (X - \alpha)\left(X - \alpha e^{2\pi i/3}\right)\left(X - \alpha e^{-2\pi i/3}\right)$$

and, since $e^{\pm 2\pi i/3} = \frac{1}{2}\left(-1 \pm i\sqrt{3}\right)$, we can say that $X^3 - 2$ splits completely over $\mathbb{Q}\left(\sqrt[3]{2}, i\sqrt{3}\right)$. The degree of the extension is 6.

In general, let us consider a field K and a polynomial f in $K[X]$. We say that an extension L of K is a splitting field for f over K, or that $L{:}\,K$ is a splitting field extension, if

(i) f splits completely over L;

(ii) f does not split completely over any proper subfield E of L.

Thus, for example, $\mathbb{Q}\left[i\sqrt{2}\right]$ is a splitting field for $X^2 + 2$ over $\mathbb{Q}$, and $\mathbb{Q}\left(\sqrt[3]{2}, i\sqrt{3}\right)$ is a splitting field of $X^3 - 2$ over $\mathbb{Q}$.

Theorem (2-2-3):

Let K be a field and let $f \in K[X]$ have degree n. Then there exists a splitting field L for f over K, and $[L: K] \leq n!$.

Proof:

The polynomial f has at least one irreducible factor g (which may be f itself). If, as in Theorem (2-1-17), we form the field $E_1 = K[X]/\langle g \rangle$ and denote the element $X + \langle g \rangle$ by α, then α has minimum polynomial g, and so $g(\alpha) = 0$. Hence g has a linear factor $Y - \alpha$ in the polynomial ring $E_1[Y]$. Moreover $[E_1: K] = \partial g \leq n$.

We proceed inductively. Suppose that, for each r in $\{1, \dots, n-1\}$, we have constructed an extension E_r of K such that f has at least r linear factor in $E_r[X]$, and

$$[E_r: K] n(n-1) \dots (n-r+1).$$

Thus, in $E_r[X]$,

$$f = (X - \alpha_1)(X - \alpha_2) \dots (X - \alpha_r) f_r,$$

and $\partial f_r = n - r$. We repeat the argument in the previous paragraph, constructing an extension E_{r+1} of E_r in which f_r has a linear factor $X - \alpha_{r+1}$ and $[E_{r+1}: E_r] \leq n - r$. We conclude that

$$[E_{r+1}: K] = [E_{r+1}: E_r][E_r: K] \leq n(n-1) \dots (n-r).$$

Hence, by induction, there exists a field E_n such that f splits completely over E_n, and $[E_n: K] \leq n!$.

Now let $L = \mathbb{Q}(\alpha_1, \alpha_2, \dots, \alpha_n) \subseteq E_n$, where $\alpha_1, \alpha_2, \dots, \alpha_n$ (not necessarily all distinct) are the roots of f in E_n. Then f splits completely over L, and cannot split completely over any proper subfield of L.

Theorem (2-2-4):

Let K and K' be fields, and let $\varphi: K \to K'$ be an isomorphism, extending to an isomorphism $\hat{\varphi}: K[X] \to K'[X]$. Let $f \in K[X]$, and let L, L' be (respectively) splitting fields of f over K and $\hat{\varphi}(f)$ over K'. Then there is an isomorphism $\varphi^*: L \to L'$ extending φ.

Proof:

Suppose that $\partial f = n$ and that in $L[X]$ we have the factorization

$$f = \alpha(X - \alpha_1)(X - \alpha_2) \dots (X - \alpha_n),$$

where α, the leading coefficient of f, lies in K, and $\alpha_1, \alpha_2, \dots, \alpha_n \in L$. We may suppose that, for some $m \in \{0,1, \dots, n\}$, the roots $\alpha_1, \alpha_2, \dots, \alpha_m$ are not in K, and that $\alpha_{m+1}, \dots, \alpha_n \in K$. We shall prove the theorem by induction on m.

If $m = 0$, then all the roots are in K, and so K itself is a splitting field for f. Hence, in $K'[X]$, we have

$$\hat{\varphi}(f) = \varphi(\alpha)\big(X - \varphi(\alpha_1)\big)\big(X - \varphi(\alpha_2)\big) \dots \big(X - \varphi(\alpha_n)\big);$$

thus K' is a splitting field for $\hat{\varphi}(f)$, and for $\varphi^* = \varphi$.

Suppose now that $m > 0$. We make the inductive hypothesis that, for every field E and every polynomial g in $E[X]$ having fewer than m roots outside E in a splitting field L of g, every isomorphism of E can be extended to an isomorphism of L.

In these chapter we study the finite fields , study the Galois group and Galois extension . Also we prove some theorem of homomorphism between fields automorphism . Finally we study the normal extension separable extension, and define the perfect field .

Section (3-1):

We certainly know that finite fields exist. To summaries what we know already, from Theorem (1-1-21) and (1-1-29) we know that a finite field K has characteristic p,, a prime number, and that its minimal subfield, known as its prime subfield, is

$$\{0\kappa, 1\kappa, 2(1\kappa), \dots, (p-1)(1\kappa)\}.$$

The prime subfield is isomorphic to $\mathbb{Z}_p$, the field of integers modulo p.

Also, in Chapter (1) (Theorem (1-1-24)) we established that, for all x, y in a field K of characteristic p, and for all $n > 1$,

$$(x \pm y)^{p^n} = x^{p^n} \pm y^{p^n}. \qquad (41)$$

Using the theory developed in the intervening chapters, we can give a complete classification of finite fields. We need one preliminary idea, which applies to all fields. Let

$$f = a_0 + a_1X + \cdots + a_nX^n$$

be a polynomial with coefficients in a field K. The formal derivative Df of f is defined by

$$Df = a_1 + 2a_2X + \cdots + na_nX^{n-1}. \qquad (42)$$

Although this is a formal procedure and has nothing to do with the analytic process of differentiation, the familiar formulae

$$D(\kappa f) = \kappa(Df), \quad D(f+g) = Df + Dg \quad (f, g \in K[X], k \in K) \qquad (43)$$

and

$$D(fg) = (Df)g + f(Dg) \quad (f, g \in K[X]) \qquad (44)$$

are still valid.

Theorem (3-1-1):

Let f be a polynomial with coefficients in a field K, and let L be a splitting field for f over K. Then the roots of f in L are all distinct if and only if f and Df have no non-constant common factor.

Proof:

Suppose first that f has a repeated root α in L, so that $f = (X - \alpha)^r g$, where $r \geq 2$. Then

$$Df = (X - \alpha)^r (Dg) + r(X - \alpha)^{r-1} g,$$

and so $f$and Df have the common factor $X - \alpha$.

Conversely, suppose that f has no repeated roots. Then, for each root α of f in L, we have $f = (X - \alpha)g$, where $g(\alpha) \neq 0$. Hence, from (45),

$$Df = g + (X - \alpha)(Dg),$$

and so $(Df)(\alpha) = g(\alpha) \neq 0$. Thus, by remainder theorem (Theorem (1-2-14)), $(X - \alpha)/Df$. This holds for every factor of f in $L[X]$, and so f in $L[X]$, and so f and Df must be coprime.

We now state the result that classifies all finite fields:

Theorem (3-1-2):

(i) Let K be a finite field. Then $|K| = p^n$ for some prime p and some integer $n \geq 1$. Every element of K is a root of the polynomial $X^{p^n} - X$, and K is a splitting field of this polynomial over the prime subfield $\mathbb{Z}_p$.

(ii) Let p be a prime, and let $n \geq 1$ be an integer. There exists, up isomorphism, exactly one field of order p^n.

Proof:

(i) Let K have characteristic p. Then K is a finite extension of $\mathbb{Z}_p$, of degree n, say. If $\{\delta_1, \delta_2, \dots, \delta_n\}$ is a basis of K over $\mathbb{Z}_p$, then every element of K is uniquely expressible as a linear combination

$$a_1\delta_1 + a_2\delta_2 + \dots + a_n\delta_n,$$

with coefficients in $\mathbb{Z}_p$. For each coefficient a_i there are p choices, namely $0,1, \dots, p-1$, and so there are p^n linear combinations in all. Thus $|K| = p^n$.

The group K^* is of order $p^n - 1$. Let $\alpha \in K^*$. Then, by Lagrange's theorem (Theorem (1-1-28)), the order of α, which is the order of the subgroup $\langle\alpha\rangle$ generated by α, divides $p^n - 1$. Certainly $\alpha^{p^{n-1}} = 1$. Thus $\alpha^{p^n} - \alpha = 0$ and, since we also have $0^{p^n} - 0 = 0$, we conclude that every element of K is a root of the polynomial $X^{p^n} - X$.

It follows that the polynomial $X^{p^n} - X$ splits completely over K, since $X - \alpha$ is a linear factor for each of the p^n elements α of K. It clearly cannot split completely over any proper subfield of K, and so K must be the splitting field of $X^{p^n} - X$ over $\mathbb{Z}_p$.

(ii) Let p and n be given, and let L be the splitting field of $f = X^{p^n} - X$ over $\mathbb{Z}_p$. Then, since the field is of characteristic p,

$$Df = p^n X^{p^n} - 1 = -1.$$

Thus f and Df are certainly coprime, and so, by Theorem (3-1-1), $X^{p^n} - X$ has p^n distinct roots in L. Let K be the set consisting of those roots. We show that K is a subfield of L. The elements $0,1$ are clearly in K. suppose that $a, b \in K$. Then, by (3-1),

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

and so $a - b \in K$. Also, if $b \neq 0$,

$$(ab^{-1})^{p^n} = a^{p^n}\left(b^{p^n}\right)^{-1} = ab^{-1},$$

and so $ab^{-1} \in K$. The field K is in fact itself the splitting field, since it contains (indeed consists of) all the roots of $X^{p^n} - X$, and clearly no proper subfield of K has this property.

We have shown that, for all primes p and all integers $n \geq 1$, there exists a field of order p^n. We have shown also that any field of order p^n is the splitting field of $X^{p^n} - X$ over $\mathbb{Z}_p$, and so, by Theorem (1-1-6), all such field are isomorphic.

We have achieved a remarkably complete classification of finite fields: only fields of prime-power order exist, and in effect, for a given p and n there is exactly one field of order p^n. We call in the Galois field of order p^n, and denote it by $GF(p^n)$. To complete the description we need to prove one final result:

Theorem (3-1-3):

The group of non-zero elements of the Galois field $GF(p^n)$ is cyclic.

To prove this we need some group theory. Let G be a finite group. Recall that the order $o(a)$ of an element a in G is the least positive integer k such that $a^k = 1$ (we are writing the identity element of G as 1) and that $a^m = 1$ if and only if $o(a)$ divides m. The exponent $e = e(G)$ of G is the smallest positive integer $e = e(G)$ with the property that $a^e = 1$ for all a in G. The exponent always exists (in a finite group): it is the least common multiple of the orders of the elements of G. Since $o(a)$ divides $|G|$ for every a, we can deduce that $e(G)$ divides $|G|$.

In a non-abelian group G it is possible that $o(a) < e(G)$ for all a in G. For example, in the smallest non-abelian group $S_3 = \{1, a, b, x, y, z\}$, with multiplication table

	1	a	b	x	y	z
1	1	a	b	x	y	z
a	a	b	1	z	x	y
b	b	1	a	y	z	x
x	x	y	z	1	a	b
y	y	z	x	b	1	a
z	z	x	y	a	b	1

we have $o(1) = 1, o(x) = o(y) = o(z) = 2, o(a) = o(b) = 3$, and $e(S_3) = 6$. This cannot happen, however, if the group is abelian:

Theorem (3-1-4):

Let G be a finite abelian group with exponent e. Then there exists an element a in G such that $o(a) = e$.

Proof:

Suppose that

$$e = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k},$$

where $p_1, p_2, \dots, p_k$ are distinct primes and $\alpha_1, \alpha_2, \dots, \alpha_k \geq 1$. Since e is the least common multiple of the orders of the elements of G, there must exist an element h_1 whose order is divisible by $p_1^{\alpha_1}$: thus $o(h_1) = p_1^{\alpha_1} q_1$, where q_1 divides $p_2^{\alpha_2} \dots p_k^{\alpha_k}$. Let $g_1 = h_1^{q_1}$. Then, for all $m \geq 1$, we have $g_1^m = h_1^{mq_1}$, and this is equal to 1 if and only if $p_1^{\alpha_1} q_1 \big| mq_1$, that is, if and only if $p_1^{\alpha_1} | m$. Thus $o(g_1) = p_1^{\alpha_1}$.

Similarly, for $i = 2, \dots, k$, we can find an element g_i of order $p_i^{\alpha_i}$. Let

$$a = g_1 g_2 \dots g_k,$$

and let $n = o(a)$. Thus

$$a^n = g_1^n g_2^n \dots g_k^n = 1$$

(this is where we are using the abelian property) and so

$$g_1^n = g_2^{-n} \dots g_k^{-n}.$$

Let $r = p_2^{\alpha_2} \dots p_k^{\alpha_k}$. Then, since $g_i^{-nr} = 1$ for $i = 2, \dots, k$, it follows that $g_1^{nr} = 1$. Thus $p_1^{\alpha_1}$ divides nr, and so, since p_1 and r are coprime, $p_1^{\alpha_1}$ divides n.

Similarly, $p_i^{\alpha_i}$ divides n for $i = 2, \dots, k$, and we deduce that $e|n$. Since, from the definition of the exponent, we also have $n|e$, we deduce that $o(a) = e$.

The following corollary is immediate:

Corollary (3-1-5):

If G is a finite abelian group such that $e(G) = |G|$, then G is cyclic.

Remark (3-1-6):

Since all fields of order p^n are isomorphic, we can construct $GF(p^n)$ simply by finding an irreducible polynomial f of degree n in $\mathbb{Z}_p[X]$. Then $GF(p^n) = \mathbb{Z}_p[X]/\langle f\rangle$. There will, however, normally be may choices for f.

The Galois Group

Monomorphisms between Fields:

Mathematicians frequently draw a distinction between the theory of fields and Galois theory. The distinction is to some extent artificial, but the study of fields enters a new phase when we consider automorphisms. It is worth emphasizing that the language we use (automorphisms, groups, normal subgroups etc.) was not available to Galois .Even with the convenient language of abstract algebra, the chain of argument in this chapter is long and, at times, far from easy: the theory developed by Galois, who lacked our advantages, is surely one of the most remarkable achievements in all mathematics.

We begin with something quite general. Let K be a field, and let S be a non-empty set. Let M be the set of mappings from S into K. If $\theta, \varphi \in M$, then $\theta + \varphi$, defined by

$$(\theta + \varphi)(s) = \theta(s) + \varphi(s) \quad (s \in S) \qquad (45)$$

is a mapping from S into K, and so belongs to M. Similarly, if $\theta \in M$ and $a \in K$, then $a\theta$, defined by

$$(a\theta)(s) = a\theta(s) \quad (s \in S) \qquad (46)$$

belongs to M. It is easy to verify that M is a vector space with respect to these two operations. The zero vector in M is the mapping ζ given by

$$\zeta(s) = 0 \quad (s \in S) \qquad (47)$$

We shall normally denote the mapping ζ simply by 0, since the context will usually make it clear whether we mean the zero element of K or the mapping ζ.

A set $\{\theta_1, \theta_2, \dots, \theta_n\}$ of elements of M is linearly independent if, for all $a_1, a_2, \dots, a_n$ in K,

$$a_1\theta_1(s) + a_2\theta_2(s) + \cdots + a_n\theta_n(s) = 0$$

for all s in S if and only if $a_1 = a_2 = \cdots = a_n = 0$. More compactly, we can write the condition as

$a_1\theta_1 + a_2\theta_2 + \cdots + a_n\theta_n = 0$ (strictly, ζ) $\Leftrightarrow a_1 = a_2 = \cdots = a_n = 0$.

Then next result, due to Dedekind, is concerned with the case where S is itself a field. It will be one of the many important stages in the proof of the fundamental result.

Theorem (3-1-7):

Let K and L be fields, and let $\theta_1, \theta_2, \dots, \theta_n$ be distinct monomorphisms from K into L. Then $\{\theta_1, \theta_2, \dots, \theta_n\}$ is a linearly independent set in the vector space M of all mapping from K into L.

Proof:

We prove the theorem by induction on n. It is clearly true for $n = 1$, since θ_1, being a monomorphism, maps the identity 1 of K to the identity 1 of L, and so is not the zero mapping defined by (48).

Assume now that we have established that every set of fewer than n distinct monomorphisms of K into L is linearly independent. Suppose, for a contradiction, that there exist $a_1, a_2, \dots, a_n$ in L, not all zero, such that

$$a_1\theta_1 + a_2\theta_2 + \cdots + a_n\theta_n = 0 \qquad (48)$$

In fact we may assume that all of the a_i are non-zero: if, for example, $a_n = 0$, then $\{\theta_1, \theta_2, \dots, \theta_{n-1}\}$ is linearly dependent, in contradiction to the induction hypothesis. Dividing by a_n in (49) gives

$$b_1\theta_1 + \cdots + b_{n-1}\theta_{n-1} + \theta_n = 0. \qquad (49)$$

where $b_i = a_i/a_n \ (i = 1,2,\ldots,n-1)$.

The monomorphisms θ_1 and θ_n are by assumption distinct, and so there exists u in K such that $\theta_1(u) \neq \theta_n(u)$; the element u is certainly non-zero, as are both $\theta_1(u)$ and $\theta_n(u)$. For every z in K,

$$b_1\theta_1(uz) + \cdots + b_{n-1}\theta_{n-1}(uz) + \theta_n(uz) = 0, \qquad (50)$$

and so, since $\theta_1, \theta_2, \ldots, \theta_n$ are monomorphisms,

$$b_1\theta_1(u)\theta_1(z) + \cdots + b_{n-1}\theta_{n-1}(u)\theta_{n-1}(z) + \theta_n(u)\theta_n(z) = 0. \qquad (51)$$

Dividing this by $\theta_n(u)$ gives the result that, for all z in K,

$$b_1\frac{\theta_1(u)}{\theta_n(u)}\theta_1(z) + \cdots + b_{n-1}\frac{\theta_{n-1}(u)}{\theta_n(u)}\theta_{n-1}(z) + \theta_n(z) = 0. \qquad (52)$$

Rewriting this as an equation concerning mappings gives

$$b_1\frac{\theta_1(u)}{\theta_n(u)}\theta_1 + \cdots + b_{n-1}\frac{\theta_{n-1}(u)}{\theta_n(u)}\theta_{n-1} + \theta_n = 0, \qquad (53)$$

where the 0 on the right now stands for the zero mapping defined by (48). We subtract (54) from (50) and obtain

$$b_1\left(1 - \frac{\theta_1(u)}{\theta_n(u)}\right)\theta_1 + \cdots + b_{n-1}\left(1 - \frac{\theta_{n-1}(u)}{\theta_n(u)}\right)\theta_{n-1} = 0 \qquad (54)$$

Our choice of u as an element such that $\theta_1(u) \neq \theta_n$ means that the coefficient of θ_1 is non-zero. Thus (55) implies that the set $\{\theta_1, \theta_2, \ldots, \theta_{n-1}\}$ is linearly dependent, in contradiction to the induction hypothesis.

Remark (3-1-8):

It is important to realise that the set of monomorphisms from K into L is not a subspace of the vector space M: if θ_1 and θ_2 are monomorphisms, and if 1_K and 1_L are (respectively) the identities of K and L, then

$$(\theta_1 + \theta_2)(1_K) = \theta_1(1_K) + \theta_2(1_K) = 1_L + 1_L \neq \theta_L,$$

and so $\theta_1 + \theta_2$ is not a monomorphism.

Automorphisms, Groups and Subfields:

The first result, stated and proved for fields, applies to much more general types of algebra:

Theorem (3-1-9):

Let K be a field. Then the set $\operatorname{Aut} K$ of Automorphisms of K forms a group under composition of mappings.

Proof:

Composition of mappings is always associative, since, for all x in K and all α, β and γ in $\operatorname{Aut} K$,

$$[(\alpha \circ \beta) \circ \gamma](x) = (\alpha \circ \beta)[\gamma(x)] = \alpha\left(\beta\big(\gamma(x)\big)\right),$$

$$[\alpha \circ (\beta \circ \gamma)](x) = \alpha\big([\beta \circ \gamma](x)\big) = \alpha\left(\beta\big(\gamma(x)\big)\right).$$

There exists an identity automorphism ι in $\operatorname{Aut} K$, defined by the property that $\iota(x) = x$ for all x in K, and clearly $\iota \circ \alpha = \alpha \circ \iota = \alpha$ for all α in $\operatorname{Aut} K$. finally, for every automorphism α in $\operatorname{Aut} K$, there is an inverse mapping α^{-1} defined by the property that α^{-1} is the unique z in K such that $\alpha(z) = x$. This map is also an automorphism. To see this, let $x, y \in K$, and let $\alpha^{-1}(x) = z, \alpha^{-1}(y) = t$; then $\alpha(z) = x,\ \alpha(t) = y$, and so $\alpha(z + t) = x + y$. Hence

$$\alpha^{-1}(x) + \alpha^{-1}(y) = z + t = \alpha^{-1}\big(\alpha(z + t)\big) = \alpha^{-1}(x + y),$$

and we can show similarly that

$$\big(\alpha^{-1}(x)\big)\big(\alpha^{-1}(y)\big) = \alpha^{-1}(xy).$$

Thus $\alpha^{-1} \in G$, and has the property that $\alpha \circ \alpha^{-1} = \alpha^{-1} = \iota$. Hence G is a group.

We refer to $\operatorname{Aut} K$ as the group of automorphisms of K.

Let L be an extension of a field K. An automorphism α of L is called a K-automorphism if $\alpha(x) = x$ for every x in K. the set of all K-automorphisms of L is denoted by $\operatorname{Gal}(L{:}\,K)$ and is called the Galois group of L over K. The Galois group

$\mathrm{Gal}(f)$ of a polynomial f in $K[X]$ is defined as $\mathrm{Gal}(L{:}K)$, where L is a splitting field of f over K. The Galois group is the key to the connection between classical algebra, dominated by the theory of equations, and modern abstract algebra, and this chapter is devoted to establishing the properties that make it such an important idea. First, we hasten to justify the use of the word "group":

Theorem (3-1-10):

Let $L{:}K$ be a field extension. Then the set $\mathrm{Gal}(L{:}K)$ of all K-automorphisms of L is a subgroup of $\mathrm{Aut}\, L$.

Proof:

Certainly $\iota \in \mathrm{Gal}(L{:}K)$. Let $\alpha, \beta \in \mathrm{Gal}(L{:}K)$. Then, for all x in K,

$$x = \beta^{-1}\big(\beta(x)\big) = \beta^{-1}(x),$$

and so

$$\alpha\big(\beta^{-1}(x)\big) = \alpha(x) = x.$$

Thus $\alpha\beta^{-1} \in \mathrm{Gal}(L{:}K)$, and so, by (23), $\mathrm{Gal}(L{:}K)$ is a subgroup of $\mathrm{Aut}\, L$.

We now introduce an important idea connecting the subfield E of L containing K and the subgroups H of the group $\mathrm{Gal}(L{:}K)$. For each E we define

$$\Gamma(E) = \{\alpha \in \mathrm{Aut}\, L{:}\ \alpha(z) = z \ \text{ for all } z \text{ in } E\} \qquad (55)$$

and for each H we define

$$\Phi(H) = \{x \in L\ : \alpha(x) = x \ \text{ for all } \alpha \text{ in } H.\} \qquad (56)$$

The essence of Galois theory is contained in these two mappings, and the principal thrust of this chapter is to find conditions under which they are mutually inverse. There are many technicalities involved in obtaining these conditions, but these must not obscure the final goal, which is Theorem (3-2-13). The technicalities concern the properties of the extension $L{:}K$ that will make the maps Γ and Φ mutually inverse. We require the extension to be "normal" and "separable", and these two notions are explored later.

The following property is easily established:

Theorem (3-1-11):

Let $L{:}\,K$ be a field extension.

(i) For every subfield E of L containing K, the set $\Gamma(E)$ is a subgroup of Gal $(L{:}\,K)$.

(ii) For every subgroup H of Gal $(L{:}\,K)$, the set $\Phi(H)$ is a subfield of L containing K.

Proof:

(i) Certainly $\Gamma(E)$ is non-empty, since it contains ι, the identity automorphism. Also, $\Gamma(E) \subseteq$ Gal $(L{:}\,K)$, since every automorphism fixing all elements of E automatically fixes all elements of K.

Let $\alpha, \beta \in \Gamma(E)$. Then, for all z in E

$$\alpha\big(\beta^{-1}(z)\big) = \alpha\left(\beta^{-1}\big(\beta(z)\big)\right) = \alpha(z) = z,$$

and so $\alpha\beta^{-1} \in \Gamma(E)$. Hence, by (23), $\Gamma(E)$ is a subgroup.

(ii) It is clear that $K \subseteq \Phi(H)$, since every automorphism in $\mathrm{Gal}(L{:}\,K)$ fixes the elements of K. Let $x, y \in \Phi(H)$. Then, for all α in H,

$$\alpha(x - y) = \alpha(x) - \alpha(y) = x - y,$$

and so $x - y \in \Phi(H)$. If $y \neq 0$, then, for all α in H,

$$\alpha(xy^{-1}) = \alpha(x)\alpha(y^{-1}) = \alpha(x)\big(\alpha(y)\big)^{-1}$$

$$= xy^{-1},$$

and so $xy^{-1} \in \Phi(H)$. Thus $\Phi(H)$ is a subfield of L.

At this point we have established a two-way connection between subfields of L containing K and subgroups of the group $\mathrm{Gal}(L{:}\,K)$. It is an "order-reversing" connection:

Theorem (3-1-12):

Let $L: K$ be a field extension.

(i) If E_1 and E_2 are subfields of L containing K, then

$$E_1 \subseteq E_2 \Longrightarrow \Gamma(E_1) \supseteq \Gamma(E_2).$$

(ii) If H_1 and H_2 are subgroups of $\mathrm{Gal}(L: K)$, then

$$H_1 \subseteq H_2 \Longrightarrow \Phi(H_1) \supseteq \Phi(H_2).$$

Proof:

(i) Suppose that $E_1 \subseteq E_2$, and let $\alpha \in \mathit{\Gamma}(E_2)$. Then α fixes every element of E_2 and so certainly fixes every element of E_1. Hence $\alpha \in \mathit{\Gamma}(E_1)$.

(ii) Suppose that $H_1 \subseteq H_2$, and let $z \in \Phi(H_2)$. Then $\alpha(z) = z$ for every α in H_2, and so certainly for every α in H_1. Hence $z \in \Phi(H_1)$.

The next natural question is concerned with whether the two mappings Γ and Φ are mutually inverse. In fact they need not be.

Theorem (3-1-13):

Let K be a field, let L be an extension of K, and let $z \in L\backslash K$. If z is a root of a polynomial f with coefficients in K, and if $\alpha \in \mathrm{Gal}(L: K)$, then $\alpha(z)$ is also a root of f.

Proof:

Let $f = a_0 + a_1 X + \cdots + a_n X^n$, where $a_0, a_1, \ldots, a_n \in K$, and suppose that $f(z) = 0$. Then

$$\begin{aligned} f\big(\alpha(z)\big) &= a_0 + a_1\alpha(z) + \cdots + a_n\big(\alpha(z)\big)^n \\ &= \alpha(a_0) + \alpha(a_1)\alpha(z) + \cdots + \alpha(a_n)\alpha(z^n) \\ &= \alpha(a_0 + a_1 z + \cdots + a_n z^n) \\ &= \alpha(0) = 0. \end{aligned}$$

Theorem (3-1-14):

Let L be a finite extension of a field K, and let G be a finite subgroup of $\mathrm{Gal}(L: K)$. Then $[L: \Phi(G)] = |G|$.

Proof:

To prove this we need to recall some standard linear algebra. Let V and W be finite-dimensional vector spaces over a field K, with dimensions m, n, respectively, and let $T: V \to W$ be a linear mapping. The image $\mathrm{im}\, T$ of T is the set $\{T(v): v \in V\}$. It is a subspace of W, and its dimension $\dim(\mathrm{im}\, T)$ is called the rank $\rho(T)$ of T. The kernel is the set $\{v \in V: T(v) = 0\}$. It is a subspace of V, and its dimension $\dim(\ker T)$ is called the nullity $v(T)$ of T. A standard result in linear algebra states that

$$\rho(T) + v(T) = \dim V = m. \qquad (57)$$

If $n < m$, then certainly $\rho(T) \le n < m$, and so $v(T) > 0$. Thus there exists a non-zero vector v in V such that $T(v) = 0$.

In more concrete terms, if we have an $n \times m$ matrix $A = \left[a_{ij}\right]_{n\times m}$ with entries in K, and m-dimensional vector v, the map v $\longmapsto A$v is a linear mapping from the vector space K^m into the vector space K^n. From the final sentence of the last paragraph we deduce that, if $n < m$, then there exists a non-zero vector v such that $A\mathrm{v} = 0$. That is, there exist $v_1, v_2, \dots, v_m$ in K, not all zero, such that

$$a_1 v_1 + a_2 v_2 + \cdots + a_{mj} v_m = 0 \quad (j = 1,2, \dots, n). \qquad (58)$$

We are now ready to prove the statement of the theorem. Let $|G| = m$ and $[L: \Phi(G)] = n$. We show first that the statement $m > n$ leads to a contradiction, using the piece of linear algebra above.

So suppose that $m > n$, and write $G = \{\alpha_1 = \iota, \alpha_2, \dots, \alpha_m\}$, where ι is the identity map, and suppose that $\{z_1, z_2, \dots, z_n\}$ is a basis for L over $\Phi(G)$.

Consider $n \times m$ the matrix

$$\begin{bmatrix} \alpha_1(z_1) & \alpha_2(z_1) & \cdots & \alpha_m(z_1) \\ \alpha_1(z_2) & \alpha_2(z_2) & \cdots & \alpha_m(z_2) \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_1(z_n) & \alpha_2(z_n) & \cdots & \alpha_m(z_n) \end{bmatrix}.$$

From (59) we deduce that there exist $v_1, v_2, \dots, v_m$ in L, not all zero, such that

$$\alpha_1(z_j)v_1 + \alpha_2(z_j)v_2 + \cdots + \alpha_m(z_j)v_m = 0 \quad (j = 1,2,\dots,n) \qquad (59)$$

Let $b \in L$. We are supposing that $\{z_1, z_2, \dots, z_n\}$ is a basis for L over $\Phi(G)$, and so there exist elements $b_1, b_2, \dots, b_n$ of $\Phi(G)$ such that

$$b = b_1 z_1, b_2 z_2, \dots, b_n z_n. \qquad (60)$$

Multiplying the n equations (60) by $b_1, b_2, \dots, b_n$ (respectively) gives

$$b_j\alpha_1(z_j)v_1 + b_j\alpha_2(z_j)v_2 + \cdots + b_j\alpha_m(z_j)v_m = 0 \; (j = 1,2,\dots,n). \quad (61)$$

Now recall that, since the b_j all lie in $\Phi(G)$ and the α_i all lie in G, we have $b_j = \alpha_i(b_j)$ for all i and j. Thus we may rewrite the equations (62) as $\alpha_1(b_j z_j)v_1 + \alpha_2(b_j z_j)v_2 + \cdots + \alpha_m(b_j z_j)v_m = 0 \; (j = 1,2,..,n)$. (63)

If we these n equations together, and make use of (61), we obtain

$$v_1\alpha_1(b) + v_2\alpha_2(b) + \cdots + v_m\alpha_m(b) = 0.$$

This holds for all b in L, and so the automorphisms $\alpha_1, \alpha_2, \dots, \alpha_m$ are linearly dependent. By Theorem (3-1-7), this is impossible. Hence $n \geq m$.

Next, suppose that $n = [L: \Phi(G)] > m$. Again we use linear algebra. This time we have subset $\{z_1, z_2, \dots, z_{m+1}\}$ of L which is linearly independent over

$$\begin{bmatrix} \alpha_1(z_1) & \alpha_1(z_2) & \cdots & \alpha_m(z_{m+1}) \\ \alpha_2(z_1) & \alpha_2(z_2) & \cdots & \alpha_m(z_{m+1}) \\ \vdots & \vdots & & \vdots \\ \alpha_m(z_1) & \alpha_m(z_2) & \cdots & \alpha_m(z_{m+1}) \end{bmatrix}.$$

By (18), there exists $u_1, u_2, \dots, u_{m+1}$ in L, not all zero, such that

$$\alpha_1(z_1)u_1 + \alpha_2(z_2)u_2 + \cdots + \alpha_j(z_{m+1})u_{m+1} = 0 \; (j = 1,2,\dots,m).$$

Let us suppose that the elements $u_1, u_2, \dots, u_{m+1}$ are chosen so that as few as possible are non-zero. We may relabel the elements so that $u_1, u_2, \dots, u_r$ are non-zero, and $u_{r+1} = \cdots = u_{m+1} = 0$. So now we have

$$\alpha_j(z_1)u_1 + \alpha_j(z_2)u_2 + \cdots + \alpha_j(z_r)u_r = 0 \ \ (j = 1,2, \dots, m). \qquad (64)$$

Dividing (64) by u_r gives a modified set of m equations

$$\alpha_j(z_1)u_1' + \cdots + \alpha_j(z_{r-1})u_{r-1}' + \alpha_j(z_r) = 0 \ \ (j = 1,2, .. , m), \qquad (65)$$

where $u_i' = u_i/u_r \ \ (i = 1,2, \dots, r-1)$. We defined α_1 to be the identity of G, and so the first of these equation is

$$z_1u_1' + \cdots + z_{r-1}u_{r-1}' + z_r = 0. \qquad (66)$$

If all of the elements $u_1', \dots, u_{r-1}'$ belonged to $\Phi(G)$, then $\{z_1, z_2, \dots, z_r\}$ would be linearly dependent over $\Phi(G)$, and we know that this is not so. Hence at least one of $u_1', \dots, u_{r-1}'$ does not belong to $\Phi(G)$: without loss of generality, we may suppose that $u_1' \notin \Phi(G)$. That is, u_1' is not fixed by every automorphism in G, and so there is an automorphism in G, which we may take to be α_2, such that

$$\alpha_2(u_1') \neq u_1'. \qquad (67)$$

We apply α_2 to the equations (66): for $j = 1,2, \dots, m$,

$$(\alpha_2\alpha_j)(z_1)\alpha_2(u_1') + \cdots + (\alpha_2\alpha_j)(z_{r-1})\alpha_2(u_{r-1}') + (\alpha_2\alpha_j)(z_r) = 0. \qquad (68)$$

Now, since G is a group, the set $\{\alpha_2\alpha_1, \alpha_2\alpha_2, \dots, \alpha_2\alpha_m\}$ is the same as the set $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$: only the order of the elements is different. Hence we may change the order of the listed equations (68) and obtain

$$\alpha_j(z_1)\alpha_1(u_1') + \cdots + \alpha_j(z_{r-1})\alpha_2(u_{r-1}') + \alpha_j(z_r) = 0 \ \ (j = 1,2, \dots, m). \qquad (69)$$

Subtracting (69) from (65) gives, for $j = 1,2, \dots, m$,

$$\alpha_j(z_1)\big(u_1' - \alpha_2(u_1')\big) + \cdots + \alpha_j(z_{r-1})\big(u_{r-1}' - \alpha_2(u_{r-1}')\big) = 0. \qquad (70)$$

Let $v_i = u_i' - \alpha_2(u_i')$ for $i = 1,2, \dots, r-1$ and $v_i = 0$ for $i = r, r+1, \dots, m+1$.

Then (70) becomes

$$\alpha_j(z_1)v_1 + \cdots + \alpha_j(z_2)v_2 + \cdots + \alpha_j(z_{m+1})v_{m+1} = 0 \ \ (j = 1,2, \dots, m) \qquad (71)$$

From (67) we know that the elements v_i are not all zero, and we have arranged that no more than $r - 1$ of the v_i are non-zero. This is a contradiction to the stated property of the elements $u_1, u_2, \dots, u_{m+1}$, and so we conclude that it is not possible to have $[L: \Phi(G)] > m$. Hence $[L: \Phi(G)] = m$.

Normal Extensions:

In the next two sections, with a view to establishing the conditions under which the maps Γ and Φ studied in the last section are mutually inverse, we introduce two new ides. Among the examples we have considered are two extensions of $\mathbb{Q}$, namely, $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt[3]{2})$. In the first case $X^2 - 2$, the minimum polynomial of $\sqrt{2}$, splits completely over $\mathbb{Q}(\sqrt{2})$; in the second case we see that $X^3 - 2$, the minimum polynomial of $\sqrt[3]{2}$, does not split completely over $\mathbb{Q}(\sqrt[3]{2})$. This is an important difference. However, although it is convenient at times to consider arbitrary extensions $L: K$, our primary interest is with Galois groups of polynomials, when L is a splitting field over K for some polynomial. We shall certainly achieve this closed focus if we suppose that $L: K$ is a normal extension, by which we mean that every irreducible polynomial in $K[X]$ having at least one root in L splits completely over L. One the face of it this is a very strong property, and indeed it is not immediately clear that even $\mathbb{Q}(\sqrt{2})$ is a normal extension of $\mathbb{Q}$. However, we have the following result:

Theorem (3-1-15):

A finite extension L of a field K is normal if and only if it is a splitting field for some polynomial in $K[X]$.

Now

$$[E(\alpha): E][E: K] = [E(\alpha): K] = [E(\alpha): K(\alpha)][K(\alpha): K], \qquad (72)$$

and

$$[E(\beta): E][E: K] = [E(\beta): K] = [E(\beta): K(\beta)][K(\beta): K]. \qquad (73)$$

Since α and β are roots of the same irreducible polynomial f, it follows from Corollary (2-1-19) that there is a K-isomorphism φ from $K(\alpha)$ onto $K(\beta)$. Certainly

$$[K(\alpha): K] = [K(\beta): K]. \qquad (74)$$

since E is a splitting field for g over K, it follows that $E(\alpha)$ is a splitting field for g over $K(\alpha)$ and $E(\beta)$ is a splitting field for g over $K(\beta)$. Hence, by Theorem (2-2-4), there is an isomorphism $\varphi^*$from $E(\alpha)$ onto $E(\beta)$, extending the K-isomorphism φ from $K(\alpha)$ onto $K(\beta)$. It follows in particular that

$$[E(\alpha): K(\alpha)] = [E(\beta): K(\beta)]. \qquad (75)$$

Now $[E(\alpha): E] = 1$, since $\alpha \in E$ by assumption. Hence

$$\begin{aligned}[E(\beta): E][E: K] &= [E(\beta): K(\beta)][K(\beta): K] \quad \big(\text{by (73)}\big)\\ &= [E(\alpha): K(\alpha)][K(\alpha): K] \quad \big(\text{by (74) and (75)}\big)\\ &= [E(\alpha): E][E: K] \quad \big(\text{by (72)}\big)\\ &= [E: K].\end{aligned}$$

Thus $[E(\beta): E] = 1$ and so $\beta \in E$, as required.

Two corollaries are worth recording at this stage:

Corollary (3-1-16):

Let L be a normal extension of finite degree over a field K. If z_1 and z_2 are a roots in L of an irreducible polynomial in $K[X]$, then there exists a K-automorphism θ of L such that $\theta(z_1) = z_2$.

Proof:

By Theorem (2-1-19), there is a K-automorphism from $K(z_1)$ onto $K(z_2)$. By Corollary (3-1-17), this extends to a K-automorphism θ of L.

Theorem (3-1-17):

Let L be a finite extension of a field K. Then,

(i) There exists a normal closure N of L over K;

(ii) If L' is a finite extension over K such that there is a K-automorphism $\varphi: L \to L'$, and if N' is a normal closure of L' over K, then there is a K-isomorphism $\psi: N \to N'$ such that the diagram

$$\begin{tikzcd}
K \arrow[r] \arrow[d, swap, "\iota"] & L \arrow[r] \arrow[d, swap, "\varphi"] & N \arrow[d, swap, "\psi"] \\
K \arrow[r] & L' \arrow[r] & N'
\end{tikzcd}$$

(in which ι is the identity map and unmarked maps are inclusions) commuted.

Proof:

(i) Let $\{z_1, z_2, \dots, z_n\}$ be a basis for L over K. Each z_i is algebraic over K, with minimum polynomial m_i (say). Let $m = m_1 m_2 \dots m_n$, and let N be a splitting field for m over K. By the proof of Theorem (3-1-16), N is a normal extension of K. It contains all the roots of each of the polynomials m_i, and so certainly contains $z_1, z_2, \dots, z_n$. Hence N contains L. Let E be a subfield of N containing L, and suppose that E is normal. For each i in $\{1,2,\dots,n\}$ the field E contains one root of m_i, namely z_i. By the definition of normality it follows that E contains all root of all the m_i and so $E = N$. We have shown that N is a normal closure.

(ii) Let N be a normal closure of L' over K. Every element of L has a unique extension $a_1 z_1 + a_2 z_2 + \cdots + a_n z_n$, where $a_1, a_2, \dots, a_n \in K$. Let $u' = \varphi(u)$ be an arbitrary element of L'. Then there is a unique n-tuple $(a_1, a_2, \dots, a_n)$ of elements of K such that

$$u' = \varphi(u) = \varphi(a_1z_1 + a_2z_2 + \cdots + a_nz_n)$$
$$= a_1\varphi(z_1) + a_2\varphi(z_2) + \cdots + a_n\varphi(z_n),$$

and it is easy to see that $\{\varphi(z_1), \varphi(z_2), \dots, \varphi(z_n)\}$ is a basis for L' over K. The isomorphism φ also ensures that, for $i = 1,2,\dots,n$, the minimum polynomial of $\varphi(z_i)$ is $\hat{\varphi}(m_i)$ (where $\hat{\varphi}$ denotes the canonical extension of φ to the polynomial ring $L[X]$). Since N' is by assumption a normal extension of L', it must contain all the roots of all of the $\hat{\varphi}(m_i)$, and must in fact be a splitting field of $\hat{\varphi}(m) = \hat{\varphi}(m_1)\hat{\varphi}(m_2)\dots\hat{\varphi}(m_n)$. The existence of the isomorphism ψ now follows from Theorem (2-2-4).

Corollary (3-1-18):

Let L be a finite extension of K and let N be a normal closure of L. Then

$$N = L_1 \vee L_2 \vee \dots \vee L_k,$$

where $L_1, L_2, \dots, L_k$ are subfields containing K, each of them isomorphic to L.

Proof:

By the theorem just proved, we may suppose that $L = K(z_1, z_2, \dots, z_n)$, that $m_1, m_2, \dots, m_n$ are (respectively) the minimum polynomials of $z_1, z_2, \dots, z_n$. Let and that N is a splitting field over K for the polynomial $m_1m_2\dots m_n$. Let $i \in P\{1,2,\dots,n\}$ and let z_i' be a root of m_i. Then, for all choices of i and z_i' the field

$$K(z_1, z_2, \dots, z_n) \qquad (76)$$

is isomorphic to L. The field N is generated over K by the set $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ of all the roots of all the polynomials $m_1, m_2, \dots, m_n$, and hence by the fields of type (76).

Section (3-2):

Separable Extensions:

Some of the ideas in this section have already been touched upon in the last chapter, but it is useful at this stage to explore the topic a little further. If f is an irreducible polynomial with coefficients in a field K, the automorphism in $Gal\ (f)$ permute the roots of f in the splitting field L. Since the study of these permutations would be hampered if f had repeated roots in L, there is a good case for restricting to extensions where this does not happen. An irreducible polynomial f with coefficients in a field K is said to be separable over K if it has no repeated roots in a splitting field. That is, in a splitting field L of f,

$$f = k(X - \alpha_1)(X - \alpha_2) \dots (X - \alpha_n),$$

where the roots $\alpha_1, \alpha_2, \dots, \alpha_n$ are all distinct. More generally,

- an arbitrary polynomial g in $K[X]$ is called separable over K if all its irreducible factors are separable over K;
- an algebraic element in an extension L of K is called separable over K if its minimum polynomial is separable over K;
- an algebraic extension L of K is called separable if every α in L is separable over K;
- a field K is called perfect if every polynomial in $K[X]$ is separable over K.

Separability is the second property (after normality) that will ensure that the maps Φ and Γ are mutually inverse. Fortunately separability is in the most interesting case guaranteed, for we shall see that all fields of characteristic zero and all finite fields are perfect.

From Theorem (3-1-1) we know that the irreducible polynomial f has repeated roots in its splitting field if and only if f and Df have a non-trivial common factor. This is the key to the next observation.

Theorem (3-2-1):

Let f be an irreducible polynomial with coefficients in a field K.

(i) If K has characteristic 0, then f is separable over K.

(ii) If K has a finite characteristic p, then f is separable unless it is of the form

$$b_0 + b_1X^p + b_2X^{2p} + \cdots + b_mX^{mp}.$$

Proof:

Let $f = a_0 + a_1X + \cdots + a_nX^n$, with $\partial f = n \geq 1$, and suppose that f is not separable. Then $f$and Df have a common factor d of degree at least 1. Since f is irreducible, the factor d must be a constant multiple (an associate) of f, and this cannot divide Df unless

$$Df = a_1 + 2a_2X + \cdots + na_nX^{n-1}$$

is the zero polynomial. Hence,

$$a_1 = 2a_2 = \cdots = na_n = 0. \qquad (77)$$

If K has characteristic 0, this implies that f is the constant polynomial a_0, and we have a contradiction. Thus f must be separable.

Suppose now that $\operatorname{char} K = p$. Then $ra_r = 0$ implies that $a_r = 0$ if and only if $p \nmid r$. Hence the only non-zero terms in f are of the form $a_{kp}X^{kp}$, for $k = 0,1,2, \ldots$. Writing a_{kp} as b_k gives the required conclusion.

From Part (i) of the theorem we immediately have the following conclusion:

Corollary (3-2-2):

Every field of characteristic 0 is perfect.

For fields of finite characteristic the situation is more complicated. We must examine conditions under which a polynomial $f(X) = g(X^p) = b_0 + b_1X^p + b_2X^{2p} + \cdots + b_mX^{mp}$ is irreducible.

Theorem (3-2-3):

Let K be a field with finite characteristic p, and let

$$f(X) = g(X^p) = b_0 + b_1X^p + b_2X^{2p} + \cdots + b_mX^{mp}.$$

Then the following statements are equivalent:

(i) f is irreducible in $K[X]$;

(ii) g is irreducible in $K[X]$, and not all of the coefficients b_i are pth powers of elements of K.

Proof:

(i)⟹(ii). If g has a non-trivial factorization $g(X) = u(X)v(X)$, then f has a factorization

$$f(X) = g(X^p) = u(X^p)v(X^p),$$

and we have a contradiction. Hence g is irreducible. If $b_i = c_i^p$ for $i = 1,2, \dots, m$, then, by Theorem (1-1-24),

$$f(X) = g(X^p) = c_0^p + (c_1X)^p + \cdots + (c_mX^m)^p$$

$$= (c_0 + c_1X + \cdots + c_mX_m)^p,$$

and again we have a contradiction. Hence not all of the coefficients b_i are pth powers.

(ii)⟹(i). We shall in fact power the (equivalent) contrapositive version, that ¬(i)⟹ ¬(ii). (Here the symbol ¬stands for "not".) Suppose that f is reducible: we must prove either that g is reducible, or that all the coefficients of f are pth powers. We have two cases:

1. $f = u^r$, where $r > 1$ and u is irreducible;
2. $f = uw$, where $\partial v, \partial w > 0$, and v and w are coprime.

Case (1). Suppose first that p/r. Then $f = \left(u^{r/p}\right)^p = h^p$ (say). If

$$h = d_0 + d_1X + \cdots + d_sX^s,$$

then

$$f = h^p = (d_0 + d_1 X + \cdots + d_s X^s)^p = d_0^p + d_1^p X^p + \cdots + d_s^p X^{sp},$$

by Theorem (1-1-24), and so all the coefficients of f are pth powers. We have proved $\neg$(ii).

Next, suppose that p/r. The definition of f in the statement of the theorem assure us that $Df = 0$; thus

$$0 = Df = r(Du)u^{r-1}$$

and so $Df = 0$. Thus we may write

$$u(X) = e_0 + e_1 X^p + \cdots + e_t X^{tp} = v(X^p),$$

and

$$g(X^p) = f(X) = \left(u(X)\right)^r = \left(v(X^p)\right)^r.$$

Thus $g(X) = \left(v(X)\right)^r$, and so g is not irreducible. Again, we have proved $\neg$(ii).

Case (2). Since $K[X]$ is a Euclidean domain, there exist s, t in $K[X]$ such that

$$sv + tw = 1. \qquad (78)$$

Also, from $Df = 0$ we deduce that

$$(Dv)w + v(Dw) = 0. \qquad (79)$$

From (78) and (79) we have that

$$0 = (Dv)tw + tv(Dw) = (Dv)(1 - sv) + tv(Dw),$$

and so

$$Dv = sv(Dv) - tv(Dw).$$

Hence v/Dv. Since $\partial(Dv) < \partial v$, we must have that $Dv = 0$. Similarly, $Dw = 0$, and so we may write

$$v(X) = d_0 + d_1 X^p + \cdots + d_s X^{sp},$$

$$w(X) = e_0 + e_1X^p + \cdots + e_tX^{tp}.$$

If we define $\bar{v}(X) = d_0 + d_1X^p + \cdots + d_sX^{sp}$ and $\bar{w}(X) = d_0 + d_1X^p + \cdots + d_sX^{sp}$, then

$$g(X^p) = f(X) = v(X)w(X) = \bar{v}(X^p)\bar{w}(X^p),$$

and so $g(X) = \bar{v}(X)\bar{w}(X)$. Thus g is not irreducible. Again, we have proved ¬(ii), and the proof is complete.

We can now establish the following result:

Theorem (3-2-4):

Every finite field is perfect.

Proof:

Let K be a finite field of characteristic p. Then the Frobenius mapping $a \longmapsto a^p$ is an automorphism of K, and so every element of K is a pth power. From Theorem (3-2-1), the only candidate for an inseparable irreducible polynomial is something of the form

$$f = b_0 + b_1X^p + \cdots + b_mX^{mp}.$$

However, since all the coefficients are pth powers, Theorem (3-2-3) tells us that even polynomials of this form are reducible. Hence K is perfect.

Since all fields of characteristic zero and all finite fields are perfect, it is reasonable to ask whether there are any "imperfect" fields at al. evidently, such a field has to be infinite and of finite characteristic, and so far we have not explicitly mentioned any such field. The most obvious example, however, is $K = \mathbb{Z}_p(X)$, the field of all rational forms with coefficients in $\mathbb{Z}_p$. For polynomials with coefficients in K we must use a different letter, such as Y, for the indeterminate. We look at the polynomial $Y^p - X$ in $K[Y]$. By Theorem (3-2-3), this is irreducible unless $-X$ is a pth power in the field K, that is, unless there exists an element $u(X)/v(X)$ in K such that $[u(X)/v(X)]^p = -X$. If we suppose that such an element exists, we deduce that $-X[v(X)]^p = [u(X)]^p$. But then $p/\partial([u(X)]^p)$ and $p \nmid \partial([v(X)]^p)$, and so we have a contradiction. Thus $f(Y) = Y^p - X$ is irreducible in $K[Y]$. Let L be a splitting field for f over K, and let α be a root of f in L. Thus $\alpha^p = X$, and the factorisation of f in L is

$$f(Y) = Y^p - X = Y^p - \alpha^p = (Y - \alpha)^p.$$

The polynomial f is as irreducible as it is possible to be!

We shall have occasion later in the chapter to make use of the following observation:

Theorem (3-2-5):

Let L be a finite separable extension of a field K, and let E be a subfield of L containing K. Then L is a separable extension of E.

Proof:

Let $\alpha \in L$, and let m_K, m_E be the minimum polynomials of α over K and E, respectively. Suppose that m_K is separable. Within $E[X]$ we can use the division algorithm

$$m_K = qm_E + r \quad (\partial r < \partial m_K),$$

and it follows that

$$r(\alpha) = m_K(\alpha) - q(\alpha)m_K(\alpha) = 0 - 0 = 0.$$

This is a contradiction to the minimality of the polynomial m_E unless $r = 0$. Hence $m_K = qm_E$ in the ring $E[X]$.

If m_E is not separable, then there is a non-constant polynomial g dividing m_E and Dm_E. Since $Dm_K = qDm_K + m_E Dq$, it follows that g divides m_K and Dm_K. This can happen only if m_K has at least one repeated root in a splitting field, and so we have a contradiction. Hence m_E is separable.

Remark (3-2-6):

We emphasise at this stage that, by Corollary (3-2-2) separability is guaranteed for fields of characteristic 0.When we come to the applications of Galois theory to polynomial equations, we will (as is reasonable in a first course) confine ourselves to fields of characteristic zero, and separability ceases to be an issue.

The Galois Correspondence:

A finite extension of a field K that is both normal and separable is called a Galois extension. The object of this section is to prove that for a Galois extension the mappings Γ and Φ are mutually inverse. This is result, and we still have some spadework to do.

If we look at $\mathbb{Q}\big(\sqrt{2}, i\sqrt{3}\big)$ and $\mathbb{Q}\big(\sqrt[3]{2}, i\sqrt{3}\big)$, we notice that in both case the order of the Galois group is equal to the degree over $\mathbb{Q}$ of the extension. Both of those examples are Galois extensions: they are certainly separable, by Corollary (3-2-2) and they are normal, being splitting fields (respectively) for $(X^2 - 2)(X^2 + 3)$ and $X^3 - 2$. We now set out to show that these are special case of a general result. We shall prove that, if $L: K$ is a normal, separable extension of degree n, and G is the Galois group of L over K, then $|G| = [L: K]$. in fact, it is useful to begin with something slightly more general:

Theorem (3-2-7):

Let $L: K$ be a separable extension of finite degree n. Then there are precisely n distinct K-monomorphism of L into a normal closure N of L over K.

Proof:

The is by induction on the degree $[L: K]$. If $[L: K] = 1$, then $L = K = N$, and the only K-monomorphism of K into N is the identity mapping ι.

Suppose now that the result is established for all $n \leq k - 1$, and suppose that $[L: K] = k > 1$. Let $z_1 \in L \backslash K$, and let m (with $\partial m = r \geq 2$) be the minimum polynomial of z_1 over K. Thus $K \subset K(z_1) \subseteq L$, and $[K(z_1): K] = r$. Then m, being irreducible and having one root z_1 in the normal extension N, splits completely over N. Since L is separable, the roots of m are all distinct: suppose that the roots are $z_1, z_2, \ldots, z_r$. Let $[L: K(z_1)] = s$; then $1 \leq s < k$, and $rs = k$.

The field N is a normal closure of $K(z_1)$, and so, by the induction hypothesis, we may suppose that the number of $K(z_1)$-monomorphisms from L into N is precisely s: denote them by $\mu_1, \mu_2, \ldots, \mu_s$. By Corollary (3-18) there are r distinct K-automorphisms $\lambda_1, \lambda_2, \ldots, \lambda_r$ of N, where $\lambda_i(z_1) = z_i (i = 1,2, \ldots, r)$. Define maps $\varphi_{ij}: L \to N$ by

$$\varphi_{ij}(x) = \lambda_i\left(\mu_j(x)\right) \quad (x \in L;\ \ i = 1,2, \ldots, s). \qquad (80)$$

The definitions make it clear that the maps are all K-monomorphisms.

We show that the maps φ_{ij} are all distinct. First observe that

$$\varphi_{ij}(z_i) = \lambda_i\left(\mu_j(z_1)\right) = \lambda_i(z_1) = z_i. \qquad (81)$$

Hence, if $\varphi_{ij} = \varphi_{pq}$, it follows that $i = p$. Suppose now that $\varphi_{ij} = \varphi_{iq}$. Then, for all x in L,

$$\lambda_i\left(\mu_j(x)\right) = \lambda_i\left(\mu_q(x)\right).$$

Since λ_i is one-one, it follows that $\mu_j(x) = \mu_q(x)$ for all x in L, and so $j = q$. Thus the maps φ_{ij} are all distinct, and from (80) we now deduce that there are at least $rs = k$ distinct K-monomrphisms from L into N.

To show that there are no more than, k, we must show that every K-monomorphism ψ from L into N coincides with one of the maps φ_{ij}. The map ψ must map z_1 to another root z_i of m in N. Let $\chi: L \to N$ be defined by

$$\chi(x) = \lambda_i^{-1}\left(\psi(x)\right).$$

This is certainly a K-monomorphism; indeed, since

$$\chi(z_1) = \lambda_i^{-1}\left(\psi(z_1)\right) = \lambda_i^{-1}(z_i) = z_1 \quad (x \in L),$$

it is a $K(z_1)$-monomorphism, and so must coincide with one of $\mu_1, \mu_2, \ldots, \mu_s$, say μ_j. Thus , for all x in L,

$$\mu_j(x) = \lambda_i^{-1}(\psi(x)),$$

and so $\psi(x) = \lambda_i\left(\mu_j(x)\right)$. Thus $\psi = \varphi_{ij}$.

If, in the statement of the Theorem (3-2-7), we suppose that L is normal well as separable, then L is its own normal closure, and we obtain the following important corollary:

Corollary (3-2-8):

Let L be a Galois extension of K, and let G be the Galois group of L over K. Then $|G| = [L: K]$.

We shall eventually see that normality and separability are the conditions required for the maps Γ and Φ defined by (55) and (56) to be mutually inverse. The next theorem establishes part of that result:

Theorem (3-2-9):

Let L be a finite extension of K. Then $\Phi(\text{Gal}\,(L: K)) = K$ if and only if L is a separable normal extension of K.

Proof:

Suppose that L is a separable and normal extension of K, and let $[L: K] = n$. by Corollary (3-2-8), $|\text{Gal}\,(L: K)| = n$. Denote $\Phi(\text{Gal}\,(L: K))$ by K'; then, from Theorem (3-1-14), we know that $K \subseteq K'$. By Theorem (2-1-11), we have that $[L: K'] = n$. Hence, since $K \subseteq K'$ and $[L: K] = [L: K']$, it follows that $K = K'$.

Conversely, suppose that $K = K'$. Let

$$\text{Gal}\,(L: K) = \{\varphi_1 = \iota, \varphi_2, \dots, \varphi_n\}.$$

Let f be an irreducible polynomial in $K[X]$ having a root z in L. To show that L is normal, we need to establish that f splits completely over L.

The images of z under the K-automorphisms $\varphi_1, \varphi_2, \dots, \varphi_n$ need not all be distinct: we know that $\varphi_1(z) = z$, and we may re-lable the elements of Gal $(L: K)$ so that $\varphi_2(z), \dots, \varphi_r(z)$ are the remaining distinct images of z under the automorphisms in Gal $(L{:}\,K)$. For notational simplicity, let us write $\varphi_i(z) = z_i \ \ (i = 1,2,\dots,r)$. Note that $z_1 = z$.

Lemma (3-2-10):

For each φ_j in Gal $(L{:}\,K)$, the sets

$$\{z_1, z_2, \dots, z_r\} \text{ and } \{\varphi_j(z_1), \varphi_j(z_2), \dots, \varphi_j(z_r)\}$$

are identical.

Proof:

We note $\varphi_j(z_i)$ is equal to $(\varphi_j\varphi_i)(z)$, and this is equal to z_k for some k, since $\varphi_j\varphi_i \in$ Gal $(L{:}\,K)$. Since φ_j is one-one, we conclude that it merely permutes the elements $z_1, z_2, \dots, z_r$.

Now let g be the polynomial

$$(X - z_1)(X - z_2) \dots (X - z_r) = X^r - e_1X^{r-1} + \cdots + (-1)^r e_r, \qquad (82)$$

where the coefficients $e_1, e_2, \dots, e_r$ are the elementary symmetric functions

$$e_1 = \sum_{i=1}^{r} z_i, \quad e_2 = \sum_{i \neq j} z_i z_j, \dots, \quad e_r = z_1, z_2, \dots, z_r.$$

These coefficients are unchanged by permutation of $z_1, z_2, \dots, z_r$, and so, by Lemma (3-2-10), are unchanged by each φ_j in Gal $(L{:}\,K)$. Thus g is a polynomial with coefficients in $\Phi(\text{Gal }(L{:}\,K))$, which (we are assuming) coincides with K.

Recall now that z was defined to be a root in L of the irreducible polynomial f in $K[X]$.

Lemma (3-2-11):

The polynomial g defined by (82) is the minimum polynomial of z over K.

Proof:

We must show that every polynomial in $K[X]$ having z as a root is divisible by g. So suppose that

$$h = a_0 + a_1X + \cdots + a_mX^m,$$

with coefficients in K, is such that

$$a_0 + a_1z + \cdots + a_mz^m = 0.$$

We can apply each φ_j to this relation: since φ_j leaves the coefficients a_i unchanged, we obtain

$$a_0 + a_1z_j + \cdots + a_mz_j^m = 0 \quad (j = 1,2,\ldots,r),$$

and it follows that h is divisible by each of $X - z_1, X - z_2, \ldots, X - z_r$. Thus h divisible by g.

Now, among the polynomials in $K[X]$ having a root z in L is the polynomial f with which (some time ago) we began. By Lemma (3-2-11), f is divisible by g, and so, since f was supposed to be irreducible, f is a constant multiple of g. Since g spilt completely over L, so does f. Moreover, all its roots are distinct, and so L is, as required, a separable normal extension of K.

We end this section with another theorem concerning separable normal extensions:

Theorem (3-2-12):

Let L be a Galois extension of a field K, and let E be a subfield of L containing K. If $\delta \in \mathrm{Gal}\ (L{:}\,K)$, then $\Gamma\big(\delta(E)\big) = \delta\Gamma(E)^{\delta-1}$.

Proof:

Write $\delta(E) = E', \Gamma(E) = H$ and $\Gamma(E') = H'$. We must show that $H' = \delta H \delta^{-1}$. Accordingly, let $\theta \in H$; we shall show that $\delta\theta\delta^{-1} \in H'$. Let $z' \in E'$ and let z be the unique element of E such that $\delta(z) = z'$. Then, since θ fixes all the elements of E,

$$(\delta\theta\delta^{-1})(z') = (\delta\theta\delta^{-1})(z) = \delta\big(\theta(z)\big) = \delta(z) = z',$$

and so $\delta\theta\delta^{-1} \in H'$. We have shown that $\delta H \delta^{-1} \subseteq H'$.

To show the opposite inclusion, let θ' be an arbitrary element of H', and let $z \in E$. Then $\delta(z) \in E'$, and so $\theta'\big(\delta(z)\big) = \delta(z)$. Hence

$$(\delta^{-1}\theta'\delta)(z) = (\delta^{-1}\delta)(z) = z,$$

and so $\delta^{-1}\theta'\delta \in \Gamma(H)$. We have shown that $\delta^{-1}H'\delta \subseteq H$, from follows immediately that $H' \subseteq \delta H \delta^{-1}$.

The Fundamental Theorem:

We finish it by gathering together all the bits and pieces in order to prove a theorem which, while easy to understand, has required a long sequence of preliminary results.

Theorem (3-2-13): (The Fundamental Theorem of Galois Theory):

Let L be a separable normal extension of a field K, with finite degree n.

(i) For all subfields E of L containing K, and for all subgroups H of the Galois group Gal $(L{:}K)$,

$$\Phi\big(\Gamma(E)\big) = E, \quad \Gamma\big(\Phi(H)\big) = H.$$

Also,

$$|\Gamma(E)| = [L{:}E] \quad |\text{Gal }(L{:}K)|/|\Gamma(E)| = [E{:}K].$$

(ii) A subfield E is a normal extension of K if and only if $\Gamma(E)$ is a normal subgroup of Gal $(L{:}K)$. If E is a normal extension, then Gal $(E{:}K)$ is isomorphic to the quotient group Gal $(L{:}K)/\Gamma(E)$.

Proof:

(i) Let E be a subfield of L containing K. We know that L is a normal extension of E. Also, by Theorem (3-2-5),L is a separable extension of E. $|\Gamma(E)| = [L{:}\,E]$. From Theorem (2-1-2) and Corollary (3-2-8) it follows that

$$[E{:}\,K] = [L{:}\,K]/[L{:}\,E] = |\text{Gal }(L{:}\,K)/\Gamma(E)|.$$

Since $\Gamma(E) = \text{Gal }(L{:}\,E)$, it follows from Theorem (3-2-9) that

$$\Phi\big(\Gamma(E)\big) = E.$$

Now let H be any subgroup of the finite group Gal $(L{:}\,K)$. from Theorem (3-1-14) we know that

$$H \subseteq \Gamma\big(\Phi(H)\big). \qquad (83)$$

Denote $\Gamma\big(\Phi(H)\big)$ by H'. We have that

$$\Phi(H) = \Phi(\Gamma[\Phi(H)]) = \Phi(H').$$

From Theorem (2-1-11) we have that

$$|H| = [L{:}\,\Phi(H)] = [L{:}\,\Phi(H')] = |H'|.$$

This, together with (83) and the finiteness of Gal (L, K), tells $H' = H$. That is,

$$\Gamma\big(\Phi(H)\big) = H.$$

(ii) Suppose now that E is a normal extension. Let $\delta \in$ Gal $(L{:}\,K)$, and let δ' be the restriction of δ to E. Then δ' is a monomorphism from E into L and so, by Theorem (3-21), is a K-automorphism of E. Since $\delta(E) = \delta'(E) = E$, it follows by Theorem (3-2-12) that

$$\Gamma(E) = \Gamma\big(\delta(E)\big) = \delta\Gamma(E)\delta^{-1}.$$

This $\Gamma(E)$ is a normal subgroup of Gal $(L{:}\,K)$.

Conversely, suppose that $\Gamma(E)$ is a normal subgroup of Gal $(L{:}\,K)$. Let δ_1 be a K-monomorphism from E into L. By Corollary (3-2-17), this extends to a K-

automorphism δ of L. The normality of $\Gamma(E)$ within Gal $(L: K)$ means that $\delta\Gamma(E)\delta^{-1} = \Gamma(E)$, and hence, by Theorem (3-2-12),

$$\Gamma\big(\delta(E)\big) = \Gamma(E).$$

Since Γ is one-one, it follows that $\delta(E) = \delta_1(E) = E$. Thus δ_1 is a K-automorphism of E. We have shown that every K-monomorphism of E into L is a K-automorphism of E. From Theorem (3-2-12) it follows that E is a normal extension of K.

It remains to show that, if E is a normal extension, then $Gal\,(E: K) \simeq$ Gal $(L: K)/\Gamma(E)$. So suppose that E is a normal and, as above, let δ' be the restriction to E of the K-automorphism δ of L. We have seen that $\delta' \in$ Gal $(E: K)$. Let θ: Gal $(L: K) \rightarrow$ Gal $(E: K)$ be defined by

$$\theta(\delta) = \delta'.$$

Then θ is a group homomorphism: for all δ_1, δ_2 in Gal $(L: K)$, with $\theta(\delta_1) = \delta_1'$ and $\theta(\delta_2) = \delta_2'$, and for all z in E,

$$\begin{aligned}([\theta(\delta_1)][\theta(\delta_2)])(z) &= (\delta_1'\delta_2')(z) = \delta_1'\big(\delta_2(z)\big) \\ &= \delta_1\big(\delta_2(z)\big) = (\delta_1\delta_2)(z) \\ &= \big(\theta(\delta_1\delta_2)\big)(z).\end{aligned}$$

Hence

$$[\theta(\delta_1)][\theta(\delta_2)] = \theta(\delta_1\delta_2).$$

The kernel of this homomorphism is the set of all δ in Gal $(L: K)$ such that δ' is the identity map on E, and this none other than $\Gamma(E)$. The result now follows from Theorem (1-1-29).

It is convenient at this point to establish two technical consequences of Theorem (3-2-13). First, let U and V be subgroup of a group G. Then it is a routine matter to show that $U \cap V$ is a subgroup of G. In general $U \cup V$ is not a subgroup, but there is always a smallest subgroup containing U and V, consisting of all products $u_1v_1u_2v_2 \dots u_nv_n$ (for all n) with $u_1, u_2, \dots \in U, v_1, v_2, \dots \in V$. We denote this by $U \vee V$, and all it the join of U and V

Similarly, if E and F are subfield of a field K, then $E \cap F$ is also a subfield, and there is a subfield $E \vee F = E(F) = F(E)$, the join of E and F. The order-reversing Galois correspondence established in Theorem (3-2-13) has the following consequence:

Theorem (3-2-14):

Let L be a Galois extension of finite degree over K, with Galois group G, and let E_1, E_2 be subfields of L containing K. If $\Gamma(E_1) = H_1$ and $\Gamma(E_2) = H_2$, then

$$\Gamma(E_1 \cap E_2) = H_1 \vee H_2, \quad \Gamma(E_1 \vee E_2) = H_1 \cap H_2.$$

Proof:

Since $E_1 \subseteq E_1 \vee E_2$, it follows from the order-reversing property of the Galois correspondence that $\Gamma(E_1 \vee E_2) \subseteq \Gamma(E_1) = H_1$. Similarly, $\Gamma(E_1 \vee E_2) \subseteq H_2$, and so

$$\Gamma(E_1 \vee E_2) \subseteq H_1 \cap H_2.$$

To show the opposite inclusion, consider an element α of $H_1 \cap H_2$. Since $\alpha \in H_1 = \Gamma(E_1)$, $\alpha(x) = x$ for all x in E_1, and similarly $\alpha(y) = y$ for all y in E_2. Now, by Theorem (2-1-4), the elements of $E_1 \vee E_2 = E_1(E_2)$ are quotients of finite linear combinations (with coefficients in E_1) of finite products of elements of E_2, and so it follows that $\alpha(z) = z$ for all z in $E_1 \vee E_2$. Thus $\alpha \in \Gamma(E_1 \vee E_2)$, and so the first assertion of the theorem is proved.

From $E_1 \cap E_2 \subseteq E_1$ it follows that $H_1 = \Gamma(E_1) \subseteq \Gamma(E_1 \cap E_2)$. Similarly, $H_1 \subseteq \Gamma(E_1 \cap E_2)$, and so

$$H_1 \vee H_2 \subseteq \Gamma(E_1 \cap E_2).$$

To show the opposite inclusion, let x be an element of L not in $E_1 \cap E_2$ -say $x \in E_1$. Since E_1 is precisely the fixed of H_1, there exists $\gamma(x) \neq x$. We deduce that $x \notin E_1 \cap E_2$ implies $x \notin \Phi(H_1 \vee H_2)$. That is, $\Phi(H_1 \vee H_2) \subseteq E_1 \cap E_2$, and the Galois correspondence gives $\Gamma(E_1 \cap E_2) \subseteq H_1 \vee H_2$.

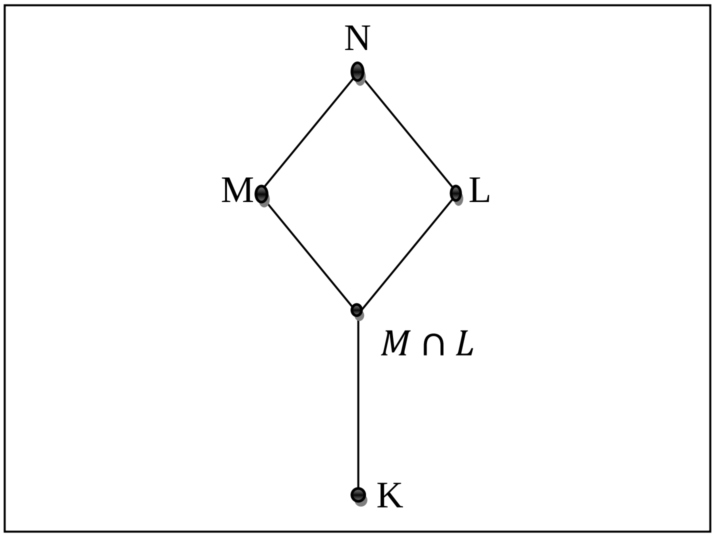
Theorem (3-2-15):

Let K be a field of characteristic zero, and let $f \in K[X]$. Let

$$L = K(\alpha_1, \alpha_2, \dots, \alpha_n)$$

be a splitting field for f over K. Let M be a field containing K, and let N be a splitting field of f over M. Then, up to isomorphism, L is a subfield of N and Gal $(N{:}\,M) \simeq$ Gal $(L{:}\,M \cap L)$.

Proof:



The field N is an extension of M, and hence of K, such that f splits completely in $N[X]$. Hence, by the definition of a splitting field, L is, up to isomorphism a subfield of N, and we may write N as $M(\alpha_1, \alpha_2, \dots, \alpha_n)$.

Let $H =$ Gal $(N{:}\,M)$, and let $\gamma \in H$. Then the restriction γ' of γ to L is a monomorphism from L into N. Since γ fixes the elements of M, it certainly fixes the elements of K; hence so does γ'. Moreover, since (by Theorem (2-1-9)) γ maps each root α_i of f to another root of f, so also must γ'. The conclusion is that γ' is a monomorphism of L into itself. Since γ is an automorphism of $N = M(\alpha_1, \alpha_2, \dots, \alpha_n)$, every root α_i of f is the image of some root of f under γ, and so also under γ'. Hence γ' maps onto $L = K(\alpha_1, \alpha_2, \dots, \alpha_n)$ and so is a K-automorphism.

We thus have a mapping θ from H into the group $G = \text{Gal}\,(L\colon K)$, given by $\theta(\gamma) = \gamma'$. The map is one-one, for if $\delta \in H$ and $\gamma' = \delta'$, then γ' and δ' act identically on the roots $\alpha_1, \alpha_2, \dots, \alpha_n$, and so $\gamma = \delta$. It is also a group homomorphism, since the restriction of $\gamma\delta$ to L is $\gamma'\delta'$. Thus $H \simeq \theta(H)$.

It remains to show that the image of θ is the subgroup Gal $(L\colon M \cap L)$ of G. Since each γ in G fixes the elements of M, it clear that each γ' fixes the elements of $M \cap L$. Thus $M \cap L \subseteq \Phi\big(\theta(H)\big)$, and so, by the Galois correspondence,

$$\theta(H) \subseteq \text{Gal}\,(L\colon M \cap L). \qquad (84)$$

Let x be an element of L not belonging to $M \cap L$. Thus $x \notin M$. Since M is the precise field whose elements are fixed by H, there is an element β in H for which $\beta(x) \neq x$. Then certainly $\big(\theta(\beta)\big)(x) \neq x$, and so $x \notin \Phi\big(\theta(H)\big)$. We have shown that $\Phi\big(\theta(H)\big) \subseteq M \cap L$, and it follows that

$$\text{Gal}\,(L\colon M \cap L) \subseteq \theta(H). \qquad (85)$$

From (84) and (85) we have that

$$\text{Gal}\,(L\colon M \cap L) =$$

$$\theta(H) \simeq H = \text{Gal}\,(N\colon M).$$

References :

1. Vijayk Khanna : A course in Abstract Algebra
   SK BHAMBRI
2. John B. Fralegih : A first Course in Abstract Algebra
3. John Wileg and Sons : Topics in algebra