ALGEBRA 2

Notes by Dr. Lynne H. Walling

The notes are organised into the following sections:

Note: The statement "for $x \in R$" is equivalent to "for all $x \in R$", as the only condition imposed on x is that it is in R. The statement "for some $x \in R$" is equivalent to "there exists $x \in R$". Sometimes we write a sentence such as "$c = ax$ where $x \in R$"; this means "there exists $x \in R$ so that $c = ax$."

§1. Rings and subrings.

Throughout this course, we will use the word "ring" to mean a commutative ring with unity, which is defined as follows.

Definition. A ring is a set R together with 2 binary, associative operations, called addition and multiplication and denoted by $+$ and $\cdot$, so that the following properties hold.

(1) Under addition, R forms an abelian group, meaning: R is closed under addition (so for any $a, b \in R$, we have $a + b \in R$); addition is a commutative operation (so for any $a, b \in R$, we have $a + b = b + a$); R has an additive identity (so there is some $0 \in R$ so that for any $a \in R$, we have $0 + a = a$); every element of R has an additive inverse (so for any $a \in R$, there is some $-a \in R$ so that $(-a) + a = 0$).
(2) R is closed under multiplication (so for any $a, b \in R$, we have $a \cdot b \in R$).
(3) Multiplication is a commutative operation (so for any $a, b \in R$, we have $a \cdot b = b \cdot a$).
(4) Multiplication distributes over addition (so for any $a, b, c \in R$, we have $a \cdot (b + c) = a \cdot b + a \cdot c$).
(5) R has a multiplicative identity (so there is some $1 \in R$ so that for any $a \in R$, we have $1 \cdot a = a$).

Note: It needs to be noted that in general, a ring is defined with more relaxed conditions, wherein multiplication is not necessarily commutative, and a multiplicative identity does not necessarily belong to the ring. In this more general case, that

multiplication distributes over addition needs to address the eventuality that multiplication may not be commutative, so we require that in a more general ring R, for $a, b, c \in R$, we have $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$.

Note: For x, y elements of a ring, we often simply write xy for $x \cdot y$ and $x - y$ for $x + (-y)$. Also, as proved in previous courses where you studied groups, with R a ring, the additive identity is unique (see Proposition 1.1 (a) below), and for each $a \in R$, the additive inverse of a is also unique (see Proposition 1.1 (b) below).

Example: A fundamental example of a ring is $\mathbb{Z}$, the set of rational integers.

As exercises, one proves the following.

Proposition 1.1. *Let R be a ring with an additive identity denoted by 0, and a multiplicative identity denoted by 1.*

(a) If $z \in R$ so that $z + a = a$ for all $a \in R$, then $z = 0$.
(b) If $a, b, c \in R$ so that $a + b = 0$ and $a + c = 0$, then $b = c$.
(c) If $u \in R$ so that $u \cdot a = a$ for all $a \in R$, then $u = 1$.
(d) For any $a \in R$, we have $0 \cdot a = 0$.
(e) With -1 the additive inverse of 1, we have $-1 \cdot a = -a$ for any $a \in R$.
(f) $0 = 1$ if and only if $R = \{0\}$.

Note: The above proposition allows us to refer to *the* additive and multiplicative identities in R, and *the* additive inverse of an element of R. Also, we sometimes write 0_R for the additive identity of R, and 1_R for the multiplicative identity of R. Further, $(-1)(-1) = 1$ since Proposition 1.1 shows that $(-1)(-1)$ is the additive inverse of -1, which is 1.

More examples of rings:

(1) Take $n \in \mathbb{Z}$. Then $\mathbb{Z}/n\mathbb{Z} = \{a + n\mathbb{Z} : \ a \in \mathbb{Z}\ \}$. Here, for $a \in \mathbb{Z}$,

$$a + n\mathbb{Z} = \{x \in \mathbb{Z} : \ x \equiv a \pmod{n}\ \}.$$

Thus $\mathbb{Z}/n\mathbb{Z}$ has n (distinct) elements, and these are

$$\{0 + n\mathbb{Z}, 1 + n\mathbb{Z}, 2 + n\mathbb{Z}, \ldots, (n-1) + n\mathbb{Z}\}.$$

Notice that if $a \equiv b \pmod{n}$ then $a + n\mathbb{Z} = b + n\mathbb{Z}$. The ring operations correspond to "arithmetic modulo n": For $a, b \in \mathbb{Z}$,

$$\begin{aligned}(a + n\mathbb{Z}) + (b + n\mathbb{Z}) &= \{(a + nk) + (b + nm) : \ k, m \in \mathbb{Z}\ \} \\ &= \{a + b + n(k + m) : \ k, m \in \mathbb{Z}\ \} \\ &= (a + b) + n\mathbb{Z}\end{aligned}$$

(since $k + m$ varies over $\mathbb{Z}$ as k, m vary over $\mathbb{Z}$); similarly, $(a + n\mathbb{Z})(b + n\mathbb{Z}) = ab + n\mathbb{Z}$.

(2) $\mathbb{Q}[X] = \{a_n X^n + a_{n-1} X^{n-1} + \cdots + a_2 X^2 + a_1 X + a_0 : \ n \in \mathbb{Z}_{\geq 0},\ a_i \in \mathbb{Q}\ \}$, the ring of polynomials in X with coefficients in $\mathbb{Q}$. This ring also has another

notable property: it has a Euclidean Algorithm, meaning that for $f, g \in \mathbb{Q}[X]$ with $f \neq 0$, there exist $q, r \in \mathbb{Q}[X]$ so that $g = qf + r$ with either $r = 0$ or $\deg r < \deg f$.

(3) Let R be a ring, and $R[X]$ the set of polynomials in X with coefficients in R. Then $R[X]$ is a ring.

(4) Let R, R' be rings. Recall that the Cartesian product $R \times R'$ is defined by

$$R \times R' = \{(a, a') : \ a \in R, \ a' \in R' \ \}.$$

For $(a, a'), (b, b') \in R \times R'$, we define

$$(a, a') + (b, b') = (a + b, a' + b'), \ (a, a') \cdot (b, b') = (ab, a'b').$$

With these operations, $R \times R'$ is a ring.

Note: $R \times R'$ is also denoted $R \oplus R'$.

Definition. With R a ring and $u \in R$, we say u is a unit if there is some $v \in R$ so that $uv = 1$.

Example: The units of $\mathbb{Z}/15\mathbb{Z}$ are $1 + 15\mathbb{Z}$, $2 + 15\mathbb{Z}$, $4 + 15\mathbb{Z}$, $7 + 15\mathbb{Z}$, $8 + 15\mathbb{Z}$, $11 + 15\mathbb{Z}$, $13 + 15\mathbb{Z}$, $14 + 15\mathbb{Z}$.

As an exercise, one can prove:

Proposition 1.2. *Let R be a ring, and suppose $u \in R$ is a unit. Suppose $v, w \in R$ so that $uv = uw = 1$. Then $v = w$.*

Note: When u is a unit, we sometimes write u^{-1} to denote its inverse; because of the above proposition, this cannot cause confusion.

Definition. Let $R \neq \{0\}$ be a ring. A subset S of R is a subring of R if $S \neq \{0\}$ and S is itself a ring with $1_S = 1_R$.

Proposition 1.3. *Suppose $R \neq \{0\}$ is a ring and $S \subset R$. Then S is a subring of R if and only if the following conditions hold:*

(a) Under $+$, S is a subgroup of R (equivalently, $S \neq \emptyset$, S is closed under addition, and S is "closed under additive inverses", meaning that for each $x \in S$, we have $-x \in S$);

(b) S is closed under multiplication;

(c) $1_R \in S$.

Proof. First suppose $S \neq \{0\}$ is a subring of R. Then one easily sees from the definition of a ring that conditions (a)-(c) are met.

Now suppose S meets conditions (a)-(c). Then the operations $+$ and $\cdot$ are associative and commutative on S because they are on R. Also, since multiplication distributes over addition in R, this is also the case in S. Take $a \in S$ (possible because of condition (a)). By condition (a), we also have $-a \in S$. Then by (a), we have $0 = -a + a \in S$. Hence, since S is closed under addition (part of condition (a)), S is an abelian group under addition. By (b), S is closed under multiplication.

Finally, $1_R \in S$ (condition (c)), and since $1_R \cdot x = x$ for all $x \in R$, we have $1_R \cdot x = x$ for all $x \in S$; thus 1_R is the multiplicative identity in S.

So S is a ring, and hence a subring of R. $\square$

Example: With $i = \sqrt{-1} \in \mathbb{C}$, the set

$$\mathbb{Z}[i] = \{a + bi : \ a, b \in \mathbb{Z} \ \}$$

is a subring of C: Clearly $\mathbb{Z}[i]$ is non-empty, and $1 \in \mathbb{Z}[i]$. Take $x, y \in \mathbb{Z}[i]$; thus $x = a + bi$, $y = c + di$ for some $a, b, c, d \in \mathbb{Z}$. Then

$$x + y = (a + c) + (b + d)i \in \mathbb{Z}[i]$$

(since $a + c, b + d \in \mathbb{Z}$). Also,

$$xy = (a + bi)(c + di) = (ac - bd) + (ad + bc)i \in \mathbb{Z}$$

(since $ac - bd, ad + bc \in \mathbb{Z}$). Thus $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$.

§2. Homomorphisms, ideals, and quotient rings.

Definition. Let R, R' be rings. A map $\varphi : R \to R'$ is a (ring) homomorphism if, for all $a, b \in R$, we have

$$\varphi(a+b) = \varphi(a) + \varphi(b),\ \varphi(ab) = \varphi(a)\varphi(b),\ \varphi(1) = 1.$$

Examples of homomorphisms:

(1) Fix $n \in \mathbb{Z}$, and define $\varphi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ by $\varphi(a) = a + n\mathbb{Z}$. Then φ is a homomorphism.
(2) Let R be a ring. Then we also have that $R \times R$ is a ring. Define $\delta : R \to R \times R$ by $\delta(a) = (a, a)$. Then δ is a homomorphism.
(3) Define $\sigma : \mathbb{Z}[i] \to \mathbb{Z}[i]$ by $\sigma(a + bi) = a - bi$. (Here a, b are assumed to be integers.) Then σ is a homomorphism.

Proposition 2.1. *Let R, R' be rings, $\varphi : R \to R'$ a homomorphism.*
(a) $\varphi(0) = 0$.
(b) For $a \in R$, we have $\varphi(-a) = -\varphi(a)$.

Proof. (a) In R, we have $0 = 0 + 0$. Thus in R', we have

$$\varphi(0) = \varphi(0+0) = \varphi(0) + \varphi(0).$$

Adding $-\varphi(0)$ to both sides of this equation, we get $0 = \varphi(0)$.

(b) Take $a \in R$. We have

$$\varphi(-a) + \varphi(a) = \varphi(-a + a) = \varphi(0) = 0.$$

Thus in R', $\varphi(-a)$ is the additive inverse of $\varphi(a)$, meaning $\varphi(-a) = -\varphi(a)$. □

As exercises, one proves the following two propositions.

Proposition 2.2. *Say $\varphi : R \to R'$, $\psi : R' \to R''$ are homomorphisms. Then $\psi \circ \varphi : R \to R''$ is a homomorphism.*

Proposition 2.3. *Let $\varphi : R \to R'$ be a homomorphism. Then $\varphi(R)$ is a subring of R'.*

Definition. Suppose R, R' are rings and $\varphi : R \to R'$ is a homomorphism. We define $\ker\varphi$, the kernel of φ, by

$$\ker\varphi = \{a \in R :\ \varphi(a) = 0\ \}.$$

Note that by Proposition 2.1 (a), we always have $0 \in \ker\varphi$ when φ is a homomorphism.

Examples:

(1) Take $n \in \mathbb{Z}$. Define $\psi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ by $\psi(a) = a + n\mathbb{Z}$. Then ψ is a homomorphism with $n\mathbb{Z} = \ker\psi$.
(2) Let R, R' be rings. Define $\varphi : R \times R' \to R$ by $\varphi((a, b)) = a$. Then φ is a homomorphism with kernel $\{0\} \times R'$.

Proposition 2.4. *Let $\varphi : R \to R'$ be a homomorphism. Then φ is injective if and only if ker$\varphi = \{0\}$.*

Proof. Supose φ is injective. Then ker$\varphi = \{0\}$, since $\varphi(0) = 0$, and φ can map at most one element of R to 0.

Now say ker$\varphi = \{0\}$, and suppose $\varphi(a) = \varphi(b)$. Then

$$0 = \varphi(a) - \varphi(b) = \varphi(a - b).$$

Thus $a - b \in$ kerφ, and since ker$\varphi = \{0\}$, this means $a - b = 0$, or equivalently, $a = b$. □

As an exercise, one proves the following.

Proposition 2.5. *Let $\varphi : R \to R'$ be a homomorphism. Then kerφ is an additive subgroup of R, with the property that for $a \in R$, $x \in$ kerφ, we have $ax \in$ kerφ.*

Definition. Let R be a ring, $I \subset R$. We say I is an ideal of R if I is an additive subgroup of R (meaning I is a subgroup of R under addition) so that for any $a \in R$, $x \in I$, we have $ax \in I$. (This latter condition is sometimes described as "I is closed under multiplication from R".)

Note: In §1, we showed that in a ring R, for $a \in R$ we have $-1 \cdot a = -a$ (i.e. the product of the additive inverse of 1 with a is the additive inverse of a). So if I is a subset of R that is "closed under multiplication from R", then for every $a \in I$ we have $-a \in I$. However, in general, a "ring" may not have a multiplicative identity (recall that in this course, we are somewhat abusing the language to mean that a ring is a commutative ring with unity). So when showing a subset I of a ring R is an ideal (where R is commutative with unity), one should address the I is "closed under additive inverses", whether this is done by addressing this directly or by using that I is closed under multiplication from R, and that for any $a \in R$, we have $-1 \cdot a = -a$. Also, in the more general setting in which a ring is not necessarily commutative, that an ideal is closed under multiplication from R needs to be formulated as: For every $a \in R$ and $x \in I$, we have $ax, xa \in I$.

Recall: If G is an abelian group under $+$ and $H \subseteq G$, then H is a subgroup of G (under $+$) if and only if $H \neq \emptyset$, H is closed under $+$, and H is "closed under additive inverses", meaning that for every $a \in H$, we have $-a \in H$. Note that if I is a (nonempty) subset of R closed under multiplication from R, then for $a \in I$, we have $-1 \cdot a \in I$, and by an earlier result, $-1 \cdot a = -a$, the additive inverse of a. However, when proving I is an ideal of R, one still needs to use this method or some other method to show that for every $a \in I$, we have $-a \in I$.

Examples:

(1) By Proposition 2.5, the kernel of a homomorphism $\varphi : R \to R'$ is an ideal of R.

(2) Let R be a ring, and take $a \in R$. Then with

$$(a) = aR = \{ax : \ x \in R \},$$

(a) is an ideal of R (called the ideal generated by a).

(3) Let R be a ring, and take $a, b \in R$. Then with
$$(a, b) = \{ax + by : \ x, y \in R \},$$
(a, b) is an ideal of R.

(4) Let I, J be ideals in a ring R. Then with $I + J = \{a + b : \ a \in I, \ b \in J \}$, $I + J$ is an ideal of R.

Definition. Let I be an ideal of a ring R. Let R/I denote the set
$$\{x + I : \ x \in R \} \text{ where } x + I = \{x + a : \ a \in I \}.$$
The elements of R/I are called cosets of I, and R/I is called a quotient ring. (In Proposition 2.7, we define $+$ and $\cdot$ on R/I, and we prove that R/I is indeed a ring.)

Proposition 2.6. *Suppose I is an ideal of a ring R.*

(a) Suppose $x \in I$; then $x + I = I$.

(b) Suppose $x \in R$ so that $x + I = I$; then $x \in I$.

(c) Suppose $x, y \in R$ so that $(x + I) \cap (y + I) \neq \emptyset$; then $x + I = y + I$.

Proof. (a) Suppose $x \in I$. Then for every $a \in I$, we have $x + a \in I$ (since $x \in I$ and I is closed under $+$). Thus $x + I \subseteq I$. Now take $a \in I$; we know $x \in I$, and since I is a subgroup of R under $+$, we have $-x \in I$. Since I is closed under $+$ and $a, -x \in I$, we have $-x + a \in I$. From this we get $a \in x + I$; since this argument holds for every $a \in I$, we get $I \subseteq x + I$. Hence $x + I = I$.

(b) Suppose $x + I = I$. We know $0 \in I$, so $x = x + 0 \in x + I = I$.

(c) Suppose $(x+I)\cap(y+I) \neq \emptyset$. Take $z \in (x+I)\cap(y+I)$. Thus $z = x+a = y+b$ for some $a, b \in I$. From this we get $x - y = b - a$, and since $a, b \in I$ and I is a subgroup under $+$, we have $b - a \in I$. Thus $x - y \in I$, so by (a), $x - y + I = I$. Hence $x + I = y + I$. $\square$

Note that this proposition implies that with I an ideal of a ring R, R/I is a partitioning of R, meaning that every element of R is in exactly one of the sets in R/I.

Proposition 2.7. *With I an ideal of a ring R, define addition and multiplication on R/I by*
$$(x + I) + (y + I) = x + y + I, \ (x + I)(y + I) = xy + I.$$
Then these operations are well-defined, and with these operations R/I forms a ring.

Proof. We first show that addition in R/I is well-defined. So suppose $x+I = u+I$, $y + I = v + I$. Since $0 \in I$, we have $x \in x + I = u + I$. Thus $x = u + a$ for some $a \in I$. Similarly, $y = v + b$ for some $b \in I$. Then
$$x + y + I = (u + a) + (v + b) + I = (u + v) + (a + b) + I = u + v + I$$
(since $a, b \in I$, so $a + b \in I$ and thus $a + b + I = I$).

As an exercise, one shows that multiplication in R/I is well-defined.

It is easy (but tedious) to verify that R/I is a ring, with additive identity $0 + I$ and multiplicative identity $1 + I$. Further, as an exercise one shows that for $x \in R$, we have $-(x + I) = -x + I$. $\square$

§3. Basic homomorphism theorems.

Recall that in previous courses you have seen that if $\varphi : R \to R'$ is a bijective map, then there is a bijective map $\varphi^1 : R' \to R$ so that $\varphi^{-1} \circ \varphi$ is the identity map on R and $\varphi \circ \varphi^{-1}$ is the identity map on R'.

Proposition 3.1. *Suppose R, R' are rings and $\varphi : R \to R'$ is a bijective homomorphism. Then $\varphi^{-1} : R' \to R$ is also a homomorphism.*

Proof. Recall that φ^{-1} is defined by

$$\varphi^{-1}(x') = x \text{ where } x \in R \text{ so that } \varphi(x) = x'.$$

[Recall that since φ is surjective, there exists some $x \in R$ so that $\varphi(x) = x'$; since φ is also injective, this x is unique.]

Take any $x', y' \in R'$, and take (the unique) $x, y \in R$ so that $\varphi(x) = x'$, $\varphi(y) = y'$. Then

$$\varphi(x + y) = \varphi(x) + \varphi(y) = x' + y',$$

so

$$\varphi^{-1}(x' + y') = x + y = \varphi^{-1}(x') + \varphi^{-1}(y').$$

Similarly,

$$\varphi(xy) = \varphi(x)\varphi(y) = x'y',$$

so

$$\varphi^{-1}(x'y') = xy = \varphi^{-1}(x')\varphi^{-1}(y').$$

Finally, $\varphi(1) = 1$, so $\varphi^{-1}(1) = 1$. Thus φ^{-1} is a homomorphism. □

Definition. A bijective homomorphism is called an isomorphism. If there is an isomorphism from a ring R onto a ring R' (or equivalently, from R' onto R), then we say R and R' are isomorphic, and we denote this by $R \simeq R'$. Note that by the above proposition, when $\varphi : R \to R'$ is an isomorphism, $\varphi^{-1} : R' \to R$ is also an isomorphism.

Theorem 3.2 (Fundamental Homomorphism Theorem). *Let R, R' be rings, $\varphi : R \to R'$ a homomorphism. Then with $K = ker\varphi$, there is an isomorphism between R/K and $\varphi(R)$.*

Proof. Define $\psi : R/K \to \varphi(R)$ by $\psi(a + K) = \varphi(a)$.

As an exercise, one shows that ψ is a well-defined homomorphism.

To see ψ is injective: Suppose $\psi(a+K) = \psi(b+K)$ for some $a+K, b+K \in R/K$. Thus $\varphi(a) = \varphi(b)$, and so

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

Hence $a - b \in K$, so $a + K = b + K$. Thus ψ is injective.

We have

$$\begin{aligned}\psi(R/K) &= \{\psi(a+K):\ a+K \in R/K\ \} \\ &= \{\varphi(a):\ a \in R\ \} \\ &= \varphi(R),\end{aligned}$$

so ψ is surjective. We know $\varphi(R)$ is a ring; thus ψ is an isomorphism (and is called the isomorphism induced by φ). $\square$

Remarks:

(1) With $\varphi : R \to R'$ a homomorphism, we know that $\varphi(R)$ is a subring of R'. Thus replacing R' by $\varphi(R)$, φ gives us a surjective homomorphism from R onto $\varphi(R)$.

(2) Suppose I is an ideal of a ring R. We have seen (Exercise 2.6) that with $\varphi : R \to R/I$ defined by $\varphi(a) = a + I$, φ is a surjective homomorphism with $I = \ker\varphi$. So every ideal is the kernel of a homomorphism.

(3) Suppose $\varphi : R \to R'$ is a surjective homomorphism. Then for $a' \in R'$, there is some $a \in R$ so that $\varphi(a) = a'$. Further, one can show that

$$\{x \in R:\ \varphi(x) = a'\ \} = a + \ker\varphi.$$

Theorem 3.3. *Let R, R' be rings, $\varphi : R \to R'$ a surjective homomorphism.*

(a) Suppose I is an ideal of R. Then $\varphi(I)$ is an ideal of R'.

(b) Suppose I' is an ideal of R'. Then $\varphi^{-1}(I')$ is an ideal of R containing $\ker\varphi$.

(c) Suppose I' is an ideal of R' and I is an ideal of R with $\ker\varphi \subseteq I$. Then $\varphi(\varphi^{-1}(I')) = I'$, and $\varphi^{-1}(\varphi(I)) = I$. Hence there is a one-to-one correspondence between ideals I' of R' and ideals I of R that contain $\ker\varphi$.

Proof. (a) Suppose I is an ideal of R; set $I' = \varphi(I)$. To show I' is an ideal of R', we first note that $I' \neq \emptyset$ since $I \neq \emptyset$. Now choose $a', b' \in I'$. Thus $a' = \varphi(a)$, $b' = \varphi(b)$ for some $a, b \in I$. So

$$a' + b' = \varphi(a) + \varphi(b) = \varphi(a+b);$$

since I is an ideal and $a, b \in I$, we have $a+b \in I$. Thus $a'+b' = \varphi(a+b) \in \varphi(I) = I'$. As an exercise, one shows that $-a' \in I'$, and that for $x' \in R'$, we have $a'x' \in I'$. Thus I' is an ideal of R'.

(b) Suppose I' is an ideal of R'; set

$$I = \varphi^{-1}(I') = \{a \in R:\ \varphi(a) \in I'\ \}.$$

We know $\varphi(0) = 0$, and $0 \in I'$ since I' is an ideal; thus $0 \in I$ and $I \neq \emptyset$. Take $a, b \in I$. Thus $\varphi(a), \varphi(b) \in I'$. Hence (using that φ is a homomorphism and I' is an ideal), $\varphi(a+b) = \varphi(a) + \varphi(b) \in I'$. Thus $a + b \in \varphi^{-1}(I') = I$. As an exercise, yo show that $-a \in I$, and that for $x \in R$, we have $ax \in I$. Thus I is an ideal of R. Also as an exercise, one shows $\ker\varphi \subseteq I$.

(c) First take an ideal I' in R', and take $x' \in \varphi(\varphi^{-1}(I'))$. Thus $x' = \varphi(x)$ for some $x \in \varphi^{-1}(I')$. By the definition of $\varphi^{-1}(I')$, we have $\varphi(x) \in I'$. So $x' = \varphi(x) \in I'$, which shows that $\varphi(\varphi^{-1}(I')) \subseteq I'$. As an exercise, one shows $I' \subseteq \varphi(\varphi^{-1}(I'))$ (and hence $\varphi(\varphi^{-1}(I')) = I'$).

Now take an ideal I in R so that I contains $\ker\varphi$, and take $x \in I$. Then $\varphi(x) \in \varphi(I)$; by the definition of the inverse image of a set, this means $x \in \varphi^{-1}(\varphi(I))$. Now choose $x \in \varphi^{-1}(\varphi(I))$. Thus $\varphi(x) \in \varphi(I)$; this means that there is some $y \in I$ so that $\varphi(x) = \varphi(y)$. As an exercise, one shows that $x \in I$ (and hence $\varphi^{-1}(\varphi(I) \subseteq I$, so $\varphi^{-1}(\varphi(I) = I$). $\square$

Proposition 3.4. *Say R is a ring with ideals $I \subseteq J$. Then J/I is an ideal of R/I.*

Proof. Certainly $0 + I \in J/I$ since J is an ideal and hence $0 \in J$. Thus the additive identity of R/I lies in J/I. Now take $a + I, b + I \in J/I$; so $a, b \in J$. Since J is an ideal, we know $a + b, -a \in J$; thus

$$(a + I) + (b + I) = (a + b) + I \in J/I, \text{ and } -(a + I) = -a + I \in J/I.$$

So J/I is a subgroup of R/I under addition. Now take $x + I \in R/I$ (so $x \in R$). Since J is an ideal, we know $xa \in J$ (where $a \in J$ as chosen above); thus $(x + I)(a + I) = xa + I \in J/I$. Hence J/I is an ideal of R/I. $\square$

Note that since J/I is an ideal of R/I, the quotient $(R/I)/(J/I)$ is also a ring. To ease notation, we can use $\overline{x}$ to denote $x + I$ for any $x \in R$, $\overline{R}$ to denote R/I, and $\overline{J}$ to denote J/I. So $\overline{0} + \overline{J}$ is the additive identity of $\overline{R}/\overline{J}$, and $\overline{1} + \overline{J}$ is the multiplicative identity of $\overline{R}/\overline{J}$.

As exercises, one proves the following two theorems.

Theorem 3.5. *Say R is a ring with ideals $I \subseteq J$. Then*

$$(R/I)/(J/I) \simeq R/J.$$

Theorem 3.6. *Say R is a ring with subring S and ideal I. Then $S + I$ is a subring of R, I is an ideal of $S + I$, $S \cap I$ is an ideal of S, and*

$$(S + I)/I \simeq S/(S \cap I).$$

Remark: To prove $R/I \simeq R'$ (where R, R' are rings and I is an ideal of R), it is often easiest to prove there is a surjective homomorphism from R to R' so that the kernel of the homomorphism is I.

Example: Let $\varphi : \mathbb{Q}[X] \to \mathbb{C}$ be the evaluation map defined by

$$\varphi(c_0 + c_1X + c_2X^2 + \cdots + c_nX^n) = c_0 + c_1i + c_2i^2 + \cdots + c_ni^n$$

where $i = \sqrt{-1}$. So the image of φ is $\mathbb{Q}[i] = \{a+bi : a, b \in \mathbb{Q}\}$. Also, one checks that φ is a homomorphism. Let $K = \ker\varphi$. Then certainly $X^2+1 \in K$. Take any $f \in K$;

then using long division of polynomials, one can write $f = (X^2+1)g + r$ where $g, r \in \mathbb{Q}[X]$ with $r = 0$ or $\deg r < 2$. Since K is an ideal, we have $(X^2+1)g \in K$, and so $r = f - (X^2+1)g \in K$. Write $r = c + dX$; since $K \subseteq \mathbb{Q}[X]$, we know $c, d \in \mathbb{Q}$. Since $r \in K$, we have $c + di = 0$. So it must be the case that $c = d = 0$, and thus $r = 0$. This tells us that $K = (X^2+1)$, the ideal generated by X^2+1. Consequently, by the Fundamental Homomorphism Theorem,

$$\mathbb{Q}[X]/(X^2+1) \simeq \mathbb{Q}[i].$$

The isomorphism $\psi : \mathbb{Q}[X]/(X^2+1) \to \mathbb{Q}[i]$ induced by φ maps $X + (X^2+1)$ to i.

Remark: For any $\alpha \in \mathbb{C}$, the "evaluation map" $\varphi : \mathbb{Q}[X] \to \mathbb{C}$ given by replacing X with α is a homomorphism, with kernel consisting of all polynomials in $\mathbb{Q}[X]$ for which α is a root.

One also proves the following as an exercise.

Proposition. *Suppose $\varphi : R \to R'$ is a homomorphism (so R, R' are rings), and $u \in R$ is a unit with multiplicative inverse v. Then $\varphi(u)$ is a unit in R' with multiplicative inverse $\varphi(v)$. Hence, if φ is an isomorphism and R is a field, then R' is a field.*

§4. Integral domains and fields.

Definition. Let R be a ring. A nonzero element $x \in R$ is called a zero divisor if there is some nonzero $y \in R$ so that $xy = 0$.

Remark: With R a ring and nonzero $x \in R$, x is not a zero divisor if, for all nonzero $y \in R$, we have $xy \neq 0$. Equivalently, for nonzero $x \in R$, x is not a zero divisor if $xy = 0$ implies $y = 0$.

Example: In $\mathbb{Z}/6\mathbb{Z}$, the elements $2 + 6\mathbb{Z}, 3 + 6\mathbb{Z}, 4 + 6\mathbb{Z}$ are zero divisors.

Recall that a unit of a ring is a ring element with a multiplicative inverse.

Examples:

(1) The units of $\mathbb{Q}$ are the nonzero elements of $\mathbb{Q}$.
(2) The units of $\mathbb{Z}$ are ± 1.
(3) The units of $\mathbb{Q}[X]$ are the nonzero elements of $\mathbb{Q}$.
(4) $1 + 2X$ is a unit in $(\mathbb{Z}/4\mathbb{Z})[X]$.
(5) $\pm i$ are units in $\mathbb{Z}[i]$.
(6) As an exercise, one shows that the units of a ring form a group under multiplication.

As an exercise, one proves the following.

Proposition 4.1. *Suppose $R \neq \{0\}$ is a ring and u is a unit in R. Then u is not a zero divisor.*

Definition. A ring $R \neq \{0\}$ is called an integral domain if it has no zero divisors.

Examples: $\mathbb{Z}, \mathbb{Z}[X]$, and for p prime, $\mathbb{Z}/p\mathbb{Z}$ are integral domains.

As an exercise, one proves the following.

Proposition 4.2. *Suppose R is an integral domain, and $a, b, c \in R$ so that $a \neq 0$ and $ab = ac$. Then $b = c$.*

Definition. A prime ideal of a ring R is an ideal I so that $I \neq R$, and whenever $ab \in I$ (where $a, b \in R$), either $a \in I$ or $b \in I$.

Remark: If R is a ring with ideal $I \neq R$, then $R \neq \{0\}$.

Examples:

(1) For p prime, $p\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$.
(2) In $\mathbb{Q}[X]$, the ideal $(X) = \{fX : \ f \in \mathbb{Q}[X]\ \}$ is a prime ideal, but (X^2) is not.

The proof of the following proposition is an exercise.

Theorem 4.3. *Let R be a ring with ideal I. Then I is a prime ideal if and only if R/I is an integral domain.*

Outline of proof.

(a) Suppose I is a prime ideal; then one deduces that R/I is not the trivial ring (where the trivial ring is the ring with just one element).
(b) Suppose I is a prime ideal. Recall that to show a ring R' has no zero divisors, it suffices to show that if $x', y' \in R'$ with $x' \neq 0$ and $x'y' = 0$, then $y' = 0$.

Using this, one can deduce that R/I has no zero divisors when I is a prime ideal.

(c) Suppose R/I is an integral domain. Recall that to show that I is a prime ideal, one must show that $I \neq R$ and that whenever $a, b \in R$ so that $ab \in I$, we have $a \in I$ or $b \in I$. Using this, one deduces that since R/I is an integral domain, I must be a prime ideal. $\square$

Definition. A ring F is a field if $0 \neq 1$, and every nonzero element of F has a multiplicative inverse.

Remark: By Proposition 4.1, every field is an integral domain.

Examples: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields; $\mathbb{Z}, \mathbb{Q}[X], \mathbb{R}[X], \mathbb{C}[X]$ are not. For $n \in \mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime.

Definition. An ideal I of a ring R is maximal if $I \neq R$, and for any ideal J so that $I \subseteq J \subseteq R$, we either have $J = I$ or $J = R$.

Example: In $\mathbb{Q}[X]$, (X) is a maximal ideal, but for an integer $n > 1$, (X^n) is an ideal that is not maximal.

Theorem 4.4. *Let R be a ring with ideal I. Then R/I is a field if and only if I is maximal.*

Outline of proof.

(a) Suppose I is a maximal ideal. Then one can deduce that R/I is not the trivial ring.

(b) Suppose I is a maximal ideal. Recall that to show a non-trivial ring R' is a field, we need to show that every nonzero element of R' is a unit. Using this, one can deduce that R/I is a field.

(c) Suppose R/I is a field. Recall that if I is a maximal ideal of R, then $I \neq R$, and whenever J is an ideal so that $I \subsetneq J \subseteq R$, we must have $J = R$. Also, given any $u \in R$ so that $u \notin I$, $I + (u)$ is an ideal of R, and $I + (u) = R$ if $1 \in I + (u)$. Using this, one can deduce that I is maximal when R/I is a field. $\square$

Using the above results, one proves the following.

Corollary 4.5. *Every maximal ideal is a prime ideal.*

Example: Consider the ring $\mathbb{Q}[X, Y]$, the ring of polynomials in the variables X, Y with coefficients in $\mathbb{Q}$. We claim that (X) is a prime ideal but not a maximal ideal of $\mathbb{Q}[X, Y]$, and (X, Y) is a maximal ideal of $\mathbb{Q}[X, Y]$. We could prove this directly, but we will prove this using the above theorems. First, define $\varphi : \mathbb{Q}[X, Y] \to \mathbb{Q}[Y]$ by $\varphi(f(X, Y)) = f(0, Y)$. It is not difficult to check that φ is a sujective homomorphism, with kernel (X). Thus, by the Fundamental Homomorphism Theorem,

$$\mathbb{Q}[X, Y]/(X) \simeq \mathbb{Q}[Y].$$

We know $\mathbb{Q}[Y]$ is an integral domain, but not a field, so (X) is a prime ideal but not a maximal ideal of $\mathbb{Q}[X, Y]$. Somewhat similarly, define $\psi : \mathbb{Q}[X, Y] \to \mathbb{Q}$ by

$\psi(f(X,Y)) = f(0,0)$. It is not hard to show that ψ is a surjective homomorphism with kernel (X,Y), so

$$\mathbb{Q}[X,Y]/(X,Y) \simeq \mathbb{Q}.$$

Since $\mathbb{Q}$ is a field, (X,Y) is a maximal ideal of $\mathbb{Q}[X,Y]$.

Theorem 4.6. *Suppose R is an integral domain. Define a relation $\sim$ on*

$$R \times (R \setminus \{0\})$$

by $(a,b) \sim (c,d)$ if $ad = bc$. Then $\sim$ is an equivalence relation. Also, letting $\frac{a}{b}$ denote the equivalence class of (a,b), and $Q = \left\{ \frac{a}{b} : \ a,b \in R, \ b \neq 0 \ \right\}$, Q is a field where

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}, \ \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

Proof. As an exercise, one proves $\sim$ is an equivalence relation.

Next, we need to show that addition and multiplication are well-defined operations on Q.

To show multiplication is well-defined, we begin by assuming $\frac{a}{b} = \frac{a'}{b'}$ and $\frac{c}{d} = \frac{c'}{d'}$. (So implicitly, $a,b,c,d,a',b',c',d' \in R$.) By the definition of $\sim$, we have $ab' = ba'$ and $cd' = dc'$. Hence $ab' \cdot cd' = ba' \cdot dc'$, or equivalently, $(ac)(b'd') = (bd)(a'c')$, which means $(ad,bd) \sim (a'c', b'd')$. Hence multiplication is well-defined on Q.

As an exercise, one shows addition is well-defined on Q.

It is elementary but tedious to now show Q is a ring, so this is left as an exercise for the careful student.

One easily checks that $\frac{0}{1}$ is the additive identity in Q, and $\frac{1}{1}$ is the multiplicative identity in Q. Note that for any $b \in R$ with $b \neq 0$, we have $\frac{0}{b} = \frac{0}{1}$ and $\frac{b}{b} = \frac{1}{1}$. Further, for $a \in R$, we have $\frac{a}{b} = \frac{0}{1}$ only if $a = 0$, and $\frac{a}{b} = \frac{1}{1}$ only if $a = b$.

Now suppose $\frac{a}{b} \in Q$ so that $a \neq 0$. Then $\frac{b}{a} \in Q$, $ab \neq 0$, and

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

Thus every nonzero element of Q has a multiplicative inverse in Q, meaning Q is a field. $\square$

Definition. With R an integral domain and Q defined as in the preceeding theorem, Q is called the field of fractions of R.

As a final comment for this section, as an exercise one proves the following.

Proposition 4.7. *Suppose R is an integral domain with $a,b \in R$, $a,b \neq 0$. Then $(a) = (b)$ if and only if $a = bu$ for some unit $u \in R$. Hence $(a) = R$ if and only if a is a unit in R.*

As an exercise, one proves the following.

Proposition 4.8. *Suppose R is not the trivial ring and $\varphi : R \to R'$ is an injective homomorphism, and $x \in R$ is a zero divisor. Then $\varphi(x)$ is a zero divisor in R'. Hence, if φ is an isomorphism and R is an integral domain, then R' is an integral domain. (So when φ is an isomorphism, we also have that $\varphi^{-1} : R' \to R$ is an isomorphism; hence when φ is an isomorphism, $x \in R$ is a zero divisor if and only if $\varphi(x)$ is a zero divisor, and R' is an integral domain if and only if R is an integral domain.)*

We also have the following.

Proposition 4.9. *Suppose R is not the trivial ring, $\varphi : R \to R'$ is a homomorphism, and $u \in R$ is a unit. Then $\varphi(u)$ is a unit in R'. Hence, if φ is an isomorphism and R is a field, then R' is a field. (So when φ is an isomorphism, we also have that $\varphi^{-1} : R' \to R$ is an isomorphism; hence when φ is an isomorphism, $u \in R$ is a unit if and only if $\varphi(u)$ is a unit, and R is a field if and only if R' is a field.)*

proof. Since u is a unit, there is some $v \in R$ so that $uv = 1$. Hence we have

$$1 = \varphi(1) = \varphi(uv) = \varphi(u)\varphi(v),$$

showing that $\varphi(u)$ is a unit.

Now suppose φ is an isomorphism and R is a field. Thus $0_R \neq 1_R$. Since φ is injective, this means

$$0_{R'} = \varphi()_R) \neq \varphi(1_R) = 1_{R'}.$$

Now take $u' \in R'$ so that $u' \neq 0$. Since φ is surjective, there is some $u \in R$ so that $\varphi(u) = u'$. Since $\varphi(0) = 0$ and $u' \neq 0$, we know $u \neq 0$. Hence u is a unit in R, so $u' = \varphi(u)$ is a unit in R'. This shows that every nonzero element of R' is a unit; since we already saw $0_{R'} \neq 1_{R'}$, this means R' is a field. $\square$

§5. Euclidean domains, principal ideal domains, and unique factorisaton domains.

Definition. Let R be an integral domain. R is called a Euclidean domain if there is a map $\delta : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ so that (1) for any $f, g \in R$ with $f \neq 0$, there exist $q, r \in R$ so that $g = fq + r$ with $r = 0$ or $\delta(r) < \delta(f)$; and (2) for $f, g \in R$ with $f, g \neq 0$, $\delta(fg) \geq \delta(f)$.

Example: We claim that $\mathbb{Q}[X]$ is a Euclidean domain, where the map δ is the degree map, deg $: \mathbb{Q}[X] \setminus \{0\} \to \mathbb{Z}_{\geq 0}$. To see this: First suppose f, g are nonzero elements of $\mathbb{Q}[X]$; then deg $(fg) =$ deg $f +$ deg $g \geq$ deg f.

Next, choose $f, g \in \mathbb{Q}[X]$ so that $f \neq 0$. If $g = 0$ then take $q = r = 0$. So suppose $g \neq 0$; we argue by induction on deg g that there are $q, r \in \mathbb{Q}[X]$ so that $g = fq + r$ with $r = 0$ or deg $r <$ deg f.

If deg $g <$ deg f, then take $q = 0$, $r = g$. So suppose $n \geq$ deg f, and suppose that for each nonzero $g' \in \mathbb{Q}[X]$ with deg $g' < n$, there are $q', r' \in \mathbb{Q}[X]$ so that $g' = fq' + r'$ with $r' = 0$ or deg $r' <$ deg f. Take $g \in \mathbb{Q}[X]$ so that $g \neq 0$ and $n =$ deg g. With $m =$ deg f, write

$$f = a_0 + a_1 X + \cdots + a_m X^m, \text{ and } g = b_0 + b_1 X + \cdots + b_n X^n$$

(so $a_i, b_j \in \mathbb{Q}$, and $a_m, b_n \neq 0$). Set

$$g' = g - \frac{b_n}{a_m} X^{n-m} f.$$

So deg $g' < n$, and hence there are $q', r' \in \mathbb{Q}[X]$ with $g' = fq' + r'$, and $r' = 0$ or deg $r' <$ deg f. Thus

$$fq' + r' = g - \frac{b_n}{a_m} X^{n-m} f,$$

so with $q = q' + \frac{b_n}{a_m} X^{n-m}$ and $r = r'$, we have $g = fq + r$ where $r = 0$ or deg $r <$ deg f. Hence by induction, we are done proving $\mathbb{Q}[X]$ is a Euclidean domain.

Example: Given an arbitrary $x \in \mathbb{C}$, we know $x = a + bi$ for some $a, b \in \mathbb{R}$. (Here $i = \sqrt{-1}$.) We define a map $N : \mathbb{C} \to \mathbb{R}$ by $N(a + bi) = a^2 + b^2$. As an exercise, one shows that for $x, y \in \mathbb{C}$, we have $N(xy) = N(x)N(y)$. We claim that with this map for δ, $R = \mathbb{Z}[i]$ is a Euclidean domain.

To see this: Clearly, N maps nonzero elements of R to nonnegative integers. Take nonzero $f, g \in R$; so $f = a + bi$, $g = c + di$ for some $a, b, c, d \in \mathbb{Z}$. Then one easily checks that $N(fg) = N(f)N(g) \geq N(f)$. Now take $f, g \in R$ so that $f \neq 0$. Then $\frac{g}{f} = u + vi$ for some $u, v \in \mathbb{Q}$. Choose integers a, b so that

$$u - \frac{1}{2} \leq a \leq u + \frac{1}{2}, \; v - \frac{1}{2} \leq b \leq v + \frac{1}{2}.$$

(Note that this means $|u-a|, |v-b| \leq \frac{1}{2}$.) Set $q = a + bi$, $r = g - fq$; so $q, r \in R$. If $r = 0$ then we are done; so suppose $r \neq 0$. Then

$$N(r) = N(f)N\left(\frac{g}{f} - q\right) = N(f)N\big((u-a) + (v-b)i\big).$$

Now, $N\big((u-a)+(v-b)i\big) = (u-a)^2 + (v-b)^2 \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Hence $N(r) < N(f)$. This shows that R is indeed a Euclidean domain.

Definition. For R a ring and $a \in R$, the ideal generated by a is denoted by (a) and is equal to

$$(a) = \{ax : \ x \in R \ \}.$$

We call such an ideal (a) a principal ideal. An integral domain in which every ideal is principal is called a principal ideal domain (frequently abbreviated to PID).

Remarks:

(1) In a ring R, when $b = ax$ for $a, b, x \in R$, we say a divides b, and we write $a|b$. So $b \in (a)$ is equivalent to $a|b$.

(2) As an exercise, one shows that for a, b in an integral domain R, we have $(a) = (b)$ if and only if $a = bu$ where u is a unit.

Theorem 5.1. *Every Euclidean domain is a principal ideal domain.*

Proof. Let R be a Euclidean domain under a map δ. (So R is an integral domain.) Let I be an ideal of R. If $I = \{0\}$, then $I = (0)$. Say $I \neq (0)$. Take $f \in I$ so that $f \neq 0$, and among the nonzero elements of I, $\delta(f)$ is minimal. We claim $I = (f)$.

Clearly $(f) \subseteq I$, since $f \in I$ and I is an ideal.

Now take an arbitrary $g \in I$. Then we can write $g = fq + r$ where $q, r \in R$, and either $r = 0$ or $\delta(r) < \delta(f)$. Since I is an ideal and $f \in I$, we have $fq \in I$; since we also have $g \in I$, we have $r = g - fq \in I$. By our choice of f, we cannot have $r \neq 0$ with $\delta(r) < \delta(f)$; hence r must be 0. This means $g = fq$, so $g \in (f)$. This holds for all $g \in I$, so $I \subseteq (f)$. Since we also have the reverse inclusion, we have $I = (f)$. $\square$

Remark: Not every principal ideal domain is a Euclidean domain. A standard example of this is $\mathbb{Z}[\alpha]$ where $\alpha = \frac{1+\sqrt{-19}}{2}$. However, proving that $\mathbb{Z}[\alpha]$ is such an example is non-trivial. There is a 3 page paper, "A principal ideal domain that is not a Euclidean domain", by O. Campoli, that one can find on the internet; this (and other papers) develop the idea of "almost Euclidean". However, this is beyond the scope of this course.

Example: $\mathbb{Z}[X]$ is not a Euclidean domain, regardless of the choice of map δ. This is because $\mathbb{Z}[X]$ is not a principal ideal domain. For instance, one can show that

$$(2, X) = \{2f + Xg : \ f, g \in \mathbb{Z}[X] \ \}$$

is an ideal, but it is not a principal ideal.

Definition. A ring R is said to satisfy the Ascending Chain Condition if, whenever $I_1, I_2, I_3, \ldots$ are ideals of R so that

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

then there is some $n \in \mathbb{Z}_+$ so that $I_m = I_n$ for all $m \geq n$.

As an exercise one proves the following.

Proposition 5.2. *Any prinicipal ideal domain R satisfies the Ascending Chain Condition.*

Definitions. Let R be an integral domain, and take $a \in R$. We say a is reducible if $a \neq 0$ and $a = cd$ for some $c, d \in R$ where neither c nor d is a unit. We say a is irreducible if a is not 0 or a unit, and whenever $a = cd$ then either c or d is a unit. For irreducible $a, b \in R$, we say a and b are associates if $a = bu$ where u is a unit.

Proposition 5.3. *Let R be a principal ideal domain, $p \in R$, $p \neq 0$, p not a unit. Then the following conditions are equivalent:*

(1) p is irreducible.

(2) (p) is a prime ideal.

(3) (p) is a maximal ideal.

Proof. By Corollary 4.5, one has that every maximal ideal is a prime ideal (the proof is an exercise for §4); so this shows (3) implies (2). Then, also as exercises, one shows that (2) implies (1), and (1) implies (3). Hence each statement implies the other 2, so the statements are equivalent. □

Definition. Let R be an integral domain. Then we say R is a unique factorisation domain (often abbreviated as UFD) if:

(1) Every nonzero, nonunit element of R can be factored into a product of a finite number of irreducible elements; and

(2) The above factorisation is essentially unique, meaning: whenever $p_1 \cdots p_m = q_1 \cdots q_n$ with p_i, q_j irredicuble, then $m = n$, and after reordering the q_j, we have that for each j, p_j and q_j are associates.

Example: We know that $\mathbb{Z}$ is a unique factorisation domain, and that $6 = 2 \cdot 3 = (-3) \cdot (-2)$.

Our aim now is to show that every principal ideal domain is a unique factorisation domain. Toward this, we first establish some propositions.

Proposition 5.4. *Suppose R is a principal ideal domain, and $a \in R$ is not 0 or a unit. Then for some irreducible $p \in R$, we have $p|a$.*

Proof. If a is irreducible, then we take $p = a$ and we are done.

Suppose a is not irreducible. (In the following argument, we rely on Proposition 5.3, which tells us that in a principal ideal domain, an ideal (b) is maximal if and only if b is irreducible.) Set $a_1 = a$ and $I_1 = (a_1)$. Then, since a_1 is not irreducible, I_1 is not maximal. Thus there is an ideal I_2 so that $I_1 \subsetneq I_2 \subsetneq R$. Since R is a principal ideal domain, $I_2 = (a_2)$ for some $a_2 \in R$. Since $(a_1) = I_1 \subset I_2 = (a_2)$,

we have $a_2|a_1$; if I_2 is maximal, then a_2 is irreducible and hence we are done. So suppose I_2 is not maximal; then there is some I_3 so that $I_2 \subsetneq I_3 \subsetneq R$. As before, we have some $a_3 \in R$ so that $I_3 = (a_3)$, $a_3|a_1$, and if I_3 is maximal then we are done. If I_3 is not maximal, we continue this process. By the Ascending Chain Condition (satisfied by principal ideal domains), this process must terminate, leaving us with a chain

$$I_1 \subsetneq I_2 \subsetneq \cdots \subsetneq I_m$$

where I_m is maximal, and $I_m = (a_m)$ for some irrediciuble $a_m \in R$. Since $(a_1) = I_1 \subset I_m = (a_m)$, we have $a_m|a_1$. Recalling that $a_1 = a$, we have an irreducible element (namely a_m) that divides a. $\square$

Proposition 5.5. *Let R be a principal ideal domain, with $a \in R$ so that a is not 0 or a unit. Then a is a product of irreducible elements.*

Proof. Here we refine the argument used to prove Proposition 5.4.

If a is irreducible, then we are done. So suppose not; set $a_1 = a$ and $I_1 = (a_1)$. Then, since a_1 is not irreducible, there is some irreducible p_1 so that $p_1|a_1$. So $a_1 = p_1 a_2$ for some a_2 where a_2 is not 0 or a unit. Hence $(a_1) \subseteq (a_2)$ and $(a_1) \neq (a_2)$ since p_1 is not a unit. If a_2 is irreducible then we are done, since $a = a_1 = p_1 a_2$ and p_1, a_2 are irreducible. Otherwise, choose irreducible $p_2|a_2$; then $a_2 = p_2 a_3$ for some a_3 where a_3 is not 0 or a unit, and $(a_1) \subsetneq (a_2) \subsetneq (a_3)$. If a_3 is irreducible then we are done, since $a = a_1 = p_1 a_2 = p_1 p_2 a_3$. Otherwise, we proceed as before, building ideals $(a_1) \subsetneq (a_2) \subsetneq \cdots \subsetneq (a_m)$ where $a_j = p_j a_{j+1}$, p_j is irreducible, and a_{j+1} is not 0 or a unit. This process terminates as soon as we have a_m irreducible; by the Ascending Chain Condition, this process must terminate. With a_m irreducible, we have

$$a = a_1 = p_1 a_2 = p_1 p_2 a_3 = \cdots = p_1 p_2 \cdots p_{m-1} a_m$$

with $p_1, \ldots, p_{m-1}, a_m$ irreducible, proving the proposition. $\square$

Theorem 5.6. *Suppose R is a principal ideal domain; then R is a unique factorisation domain.*

Proof. Take $a \in R$ so that a is not 0 or a unit. Then we know we can write a as a product of irreducibles; suppose we have $a = p_1 \cdots p_m = q_1 \cdots q_n$ where $p_1, \ldots, p_m, q_1, \ldots, q_n$ are irreducible. Without loss of generality, suppose $m \leq n$.

Since p_1 is irreducible, we know (p_1) is a prime ideal. Also, we have $q_1 \cdots q_n = a \in (p_1)$. Using repeatedly the definition of a prime ideal, we have $q_j \in (p_1)$ for some j, $1 \leq j \leq n$. Reordering $q_1, \ldots, q_n$, we can assume $q_1 \in (p_1)$. Thus $q_1 = p_1 u_1$, where u_1 must be a unit since q_1 is irreducible. So we have

$$p_1 \cdot p_2 \cdots p_m = p_1 u_1 \cdot q_2 \cdots q_n,$$

and since R is an integral domain and $p_1 \neq 0$, this means $p_2 \cdots p_m = u_1 \cdot q_2 \cdots q_n$. So $u_1 \cdot q_2 \cdots q_n \in (p_2)$, and since u_1 is a unit, we have $u_1^{-1} \in R$ and hence $q_2 \cdots q_n \in (p_2)$. Thus for some j, $2 \leq j \leq n$, we have $q_j \in (p_2)$; reordering $q_2, \ldots, q_n$, we can assume $q_2 \in (p_2)$. Then, arguing as above, we see $q_2 = p_2 u_2$ where u_2 is a unit,

$p_2 \cdot p_3 \cdots p_m = u_1 \cdot p_2 u_2 \cdot q_3 \cdots q_n$ and hence $p_3 \cdots p_m = u_1 u_2 \cdot q_3 \cdots q_n$. We continue in this fashion so that (reordering $q_1, \ldots, q_n$ as necessary), we have $q_j = p_j u_j$ with u_j a unit for $1 \leq j \leq m$. If $m < n$ then we have

$$p_m = u_1 \cdots u_m p_m \cdot q_{m+1} \cdots q_n,$$

and so $1 = u_1 \cdots u_m \cdot q_{m+1} \cdots q_n$; but this means $q_{m+1}, \ldots, q_n$ are units, contradicting the assumption they are irreducible. Therefore $m = n$, and by reordering, we have that for all $j = 1, \ldots, m$, p_j and q_j are associates. $\square$

Remark: Not every unique factorisation domain is a principal ideal domain. This will be demonstrated by showing that $\mathbb{Q}[X, Y]$ is a unique factorisation domain, but it is not a principal ideal domain; for instance, one can show that the ideal (X, Y) is not a principal ideal. To show $\mathbb{Q}[X, Y]$ is a unique factorisation domain, we first need to prove Gauss' Lemma, which is done in the following section.

Let R be a UFD, $a, b \in R$, $a, b \neq 0$. Let $p_1, \ldots, p_n$ be irreducible elements of R so that no two are associates, and so that for any irreducible element $q \in R$ with $q|a$ or $q|b$, q is an associate of some p_i, $1 \leq i \leq n$. (By unique factorisation, we only need finitely many p_i for this condition.) Thus $a = u p_1^{r_1} \cdots p_n^{r_n}$, $b = v p_1^{s_1} \cdots p_n^{s_n}$ where u, v are units, $r_i, s_i \in \mathbb{Z}_{\geq 0}$. (Here we agree $p^0 = 1$.) Let $c = p_1^{\min(r_1, s_1)} \cdots p_n^{\min(r_n, s_n)}$. Suppose $d \in R$ so that $d|a$ and $d|b$. If q is irreducible and $q|d$, then $q|a$ and hence q is an associate of some p_i, $1 \leq i \leq n$. So $d = w p_1^{t_1} \cdots p_1^{t_n}$ where w is a unit, $t_i \in \mathbb{Z}_{\geq 0}$; since $d|a$ and $d|b$, unique factorisation gives us $t_i \leq \min(r_i, s_i)$ for $1 \leq i \leq n$. Hence $d|c$.

This allows us to make the following definition.

Definition. Let R be a UFD, a, b elements of R that are not both 0. Set $c = \mathrm{hcf}(a, b)$ where $c \in R$ so that $c|a$ and $c|b$, and whenever $d \in R$ so that $d|a$ and $d|b$, we have $d|c$. Since R is a UFD, $\mathrm{hcf}(a, b)$ is well-defined up to units. When $\mathrm{hcf}(a, b) = 1$, we say a and b are relatively prime. (Note that by this definition, for $a \neq 0$, $\mathrm{hcf}(a, 0) = a$, but we never have $0 = \mathrm{hcf}(a, b)$ as a, b are not both 0.)

Proposition 5.7. *Suppose R is a PID, and $a, b \in R$ with a, b not both 0. Take $c \in R$ so that $(c) = (a, b)$. Then $c = hcf(a, b)$ and $c = as + bt$ for some $s, t \in R$.*

Proof. We have $a, b \in (a, b) = (c)$, so $a = cx, b = cy$, for some $x, y \in R$, which means $c|a$ and $c|b$. Also, $c \in (c) = (a, b)$, so $c = as + bt$ for some $s, t \in R$. Now suppose $d \in R$ so that $d|a$ and $d|c$. Hence $a = du$, $b = dv$ for some $u, v \in R$. Thus

$$c = as + bt = dus + dvt = d(us + vt),$$

so $d|c$. Hence $c = \mathrm{hcf}(a, b)$. $\square$

As exercises, one proves the following.

Proposition 5.8. *Suppose R is a UFD.*

(a) Suppose $a, b \in R$, not both 0, and $c = hcf(a, b)$. Thus $a = cx$, $b = cy$ for some $x, y \in R$ with $hcf(x, y) = 1$.

(b) *Suppose* $a, b, c \in R$ *with* $c \neq 0$ *and not both* a *and* b *equal to 0. Then* $hcf(ac, bc) = c \cdot hcf(a, b)$.
(c) *Suppose that* $a, b, c \in R$, $a, b \neq 0$, *so that* $hcf(a, b) = 1$ *and* $a|bc$; *then* $a|c$.

Proposition 5.9. *Suppose* R *is a UFD,* $p \in R$ *so that* p *is nonzero and not a unit. Then* p *is irreducible if and only if* (p) *is a prime ideal.*

Remark: When R is a UFD, the proof that if (p) is prime then p is irreducible is identical to the proof of this implication when R is a PID.

Finally, we have the following.

Proposition 5.10. *Suppose* $\varphi : R \to R'$ *is an isomorphism.*
(a) *If* R *is a Euclidean domain with map* δ, *then* R' *is a Euclidean domain with map* δ' *where, for* $x' \in R'$ *with* $x' \neq 0$, *we define* $\delta'(x') = \delta(x)$ *where* x *is the unique element of* R *so that* $\varphi(x) = x'$.
(b) *If* R *is a PID then so is* R'.
(c) *If* R *is a UFD then so is* R'.

Proof. First, we remark that when R is an integral domain, we know by Proposition 4.8 that R' is an integral domain.

(a) Suppose R is a Euclidean domain with map δ; define δ' as in the statement of the proposition. Take $f', g' \in R'$ so that $f' \neq 0$. Take $f, g \in R$ so that $\varphi(f) = f'$, $\varphi(g) = g'$; since $f' \neq 0$, we know $f \neq 0$ (recall that since φ is a homomorphism, $\varphi(0) = 0$). Since R is a Euclidean domain, we know there exist $q, r \in R$ so that $g = fq + r$ where either $r = 0$ or $\delta(r) < \delta(f)$. Set $q' = \varphi(q)$, $r' = \varphi(r)$; then

$$g' = \varphi(g) = \varphi(fq + r) = \varphi(f)\varphi(q) + \varphi(r) = f'q' + r'.$$

Since φ is injective, we know $r' = 0$ if and only if $r = 0$; when $r \neq 0$, we get

$$\delta'(r') = \delta(r) < \delta(f) = \delta'(f').$$

Now suppose we also have $g' \neq 0$. (Hence $f'g' \neq 0$ since R' is an integral domain.) Then

$$\varphi(fg) = \varphi(f')\varphi(g'),$$

so $\delta'(f'g') = \delta(fg)$. Hence we have

$$\delta'(f'g') = \delta(fg) \geq \delta(f) = \delta'(f').$$

This shows R' is a Euclidean domain with the map δ'.

The proof of (b) is left as an exercise.

(c) Suppose $a \in R$ so that $a \neq 0$, and set $a' = \varphi(x)$. Since φ is injective, we know $a' = 0$. As an exercise, one shows that a is reducible if and only if a' is reducible; since we know by Proposition 4.8 that a is a unit if and only if a' is a unit, we can conclude that a is irreducible if and only if a' is irreducible.

Now take $x' \in R'$ so that x is not 0 or a unit. Take $x \in R$ so that $x' = \varphi(x)$; since φ is an isomorphism, we know x is not 0 or a unit. Since R is a UFD, we know there are irreducible elements $p_1, \ldots, p_m \in R$ so that $x = p_1 \cdots p_m$; hence with $p_i = \varphi(p_i)$ for $1 \leq i \leq m$,

$$x' = \varphi(x) = \varphi(p_1 \cdots p_m) = \varphi(p_1) \cdots \varphi(p_m) = p_1' \cdots p_m'.$$

As discussed above, we know p_i' is irreducible for $1 \leq i \leq m$. Hence x' can be factored as a product of irreducible elements of R'.

Now suppose we have another factorisation $x' = q_1' \cdots q_n'$ where $q_1', \ldots, q_n'$ are irreducible elements of R'. We know φ^{-1} is also an isomorphism, so with $q_i = \varphi^{-1}(q_i')$ for $1 \leq i \leq n$, we know $q_1, \ldots, q_n$ are irreducible. Hence we have

$$x = p_1 \ldots p_m = q_1 \ldots q_n.$$

Since R is a UFD, we have $m = n$, and reordering the q_i, we have $p_i = u_i q_i$ with u_i a unit, $1 \leq i \leq m$. Hence for $1 \leq i \leq m$, with $u_i' = \varphi(u_1)$, we know u_i' is a unit, and

$$p_i' = \varphi(p_i) = \varphi(u_i q_i) = \varphi(u_i)\varphi(q_i) = u_i' q_i'.$$

This shows that the factorisation of x' as a product of irreducible elements of R' is essentially unique.

Thus R' is a UFD. □

§6. Gauss' Lemma and consequences.

Definitions. Let R be a UFD. We can extend the definition of hcf to an arbitrary (finite) number of elements $a_0, \ldots, a_n \in R$ provided they are not all 0: We set $c = \text{hcf}(a_0, \ldots, a_n)$ where $c \in R$ so that $c|a_i$ (for $0 \leq i \leq n$), and whenever $d|a_i$ (for $0 \leq i \leq n$), we have $d|c$.

Somewhat similarly, when $a_i \neq 0$ for $0 \leq i \leq n$, we set $\ell = \text{lcm}(a_0, \ldots, a_n)$ where $\ell \in R$ so that $a_i|\ell$ for $0 \leq i \leq n$, and if $a_i|d$ for $0 \leq i \leq n$, then $\ell|d$; since R is a UFD, one can show $\text{lcm}(a_0, \ldots, a_n)$ is well-defined up to units.

Suppose still that R is a UFD, and $f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$ with $f \neq 0$. We define the content of f to be $\text{hcf}(a_0, \ldots, a_n)$. We say $f \in R[X]$ is primitive if $f \neq 0$ and the content of f is 1.

Example: With $R = \mathbb{Z}$, we have that $2 + 33X + 5X^2$ is a primitive element of $R[X]$, and $18 + 15X + 21X^2$ has content 3.

Note that if $f \in R[X]$ is irreducible, then f must be primitive (since f is divisible by its content). Note also that for $f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$, f is primitive if, whenever $d \in R$ so that $d|a_i$ for $0 \leq i \leq n$, d must be a unit.

We want to show that with Q the field of fractions of a UFD R and $f \in R[X]$, we have that f is irreducible in $Q[X]$ if and only if f is irreducible in $R[X]$. Then we will develop some criteria to show f is irreducible in $R[X]$. Toward this, one proves the following.

Proposition 6.1 (Gauss' Lemma). *Suppose R is a UFD, and f, g are primitive elements of $R[X]$. Then fg is primitive in $R[X]$.*

Remark: Suppose R is a UFD, $g, h \in R[X]$ so that $\alpha =$content of g, $\beta =$content of h. Thus $g = \alpha g'$, $h = \beta h'$ where g', h' are primitive elements of $R[X]$. Hence, by Gauss' Lemma, $g'h'$ is primitive, so the content of gh is $\alpha\beta$. (Hence if gh is primitive in $R[X]$, so are g and h.)

We also will need the following, whose proof is an exercise.

Proposition 6.2. *Let R be a UFD, Q its field of fractions. Suppose f^* is a primitive element of $R[X]$, and $\alpha \in Q$ so that αf^* is also a primitive element of $R[X]$. Then α is a unit in R.*

Proposition 6.3. *Let R be a UFD, Q its field of fractions, and $f \in Q[X]$ with $f \neq 0$. Then there is a primitive $f^* \in R[X]$ so that*

$$f = \frac{a}{b} f^* \text{ where } a, b \in R,\ b \neq 0,\ hcf(a, b) = 1.$$

Further, a, b, f^* are unique up to multiplication by units of R.

Proof. Write

$$f = \frac{a_0}{b_0} + \frac{a_1}{b_1} X + \cdots + \frac{a_n}{b_n} X^n$$

where $a_i, b_i \in R$, $b_i \neq 0$, and $\mathrm{hcf}(a_i, b_i) = 1$ for $i = 0, \ldots, n$. Note that since $f \neq 0$, we know not all a_i can be 0. Then set $b = \mathrm{lcm}(b_0, \ldots, b_n)$, the least common multiple of $b_0, \ldots, b_n$. Thus we have

$$f = \frac{1}{b}(c_0 + c_1 X + \cdots + c_n X^n)$$

where $c_0, \ldots, c_n \in R$. (Specifically, for $0 \leq i \leq n$, we have $b = b_i \cdot b_i'$, and $c_i = a_i b_i'$.) Now we set $a = \mathrm{hcf}(c_0, \ldots, c_n)$. Hence $f = \frac{a}{b} f^*$ where f^* is a primitive element of $R[X]$.

To show uniqueness (up to multiplication by units), suppose we have $\frac{a}{b} g = \frac{c}{d} h$ where a, b, c, d are nonzero elements of R with $\mathrm{hcf}(a, b) = \mathrm{hcf}(c, d) = 1$, and g, h are primitive in $R[X]$. Thus we have

$$adg = bch.$$

By Proposition 5.8, the content of adg is $ad\cdot$content of g, and the content of bch is $bc\cdot$content of h. Since the content of g and the content of h are both 1, we have $bc = adu$ where u is a unit in R. Hence $a|bc$, and since $\mathrm{hcf}(a, b) = 1$, we have $a|c$. Similarly, since $ad = bcu^{-1}$ and $\mathrm{hcf}(c, d) = 1$, we have $c|a$. Thus $a = cv$ where v is a unit in R. Hence we also have $bc = cduv$; since $c \neq 0$ and R is an integral domain, we have $b = duv$ (and we know uv is a unit in R). Finally, we have $adg = aduh$, and since $ad \neq 0$ and R is an integral domain, we have $g = uh$. □

The following result is also sometimes called Gauss' Lemma.

Theorem 6.4. *Suppose R is a UFD, Q its field of fractions. Suppose f is a primitive element of $R[X]$ with deg $f > 0$. Then f is irreducible in $R[X]$ if and only if f is irreducible in $Q[X]$.*

Proof. Suppose first that f is irreducible in $Q[X]$. To show f is irreducible in $R[X]$, suppose $f = gh$ for some $g, h \in R[X]$. Thus $g, h \in Q[X]$, so either g or h is a unit in Q; without loss of generality, assume g is a unit in Q. Hence $g \in Q \cap R[X] = R$. So g is an element of R dividing f; since f is primitive in $R[X]$, g must be a unit in R (and hence a unit in $R[X]$). This shows f is irreducible in $R[X]$.

Now suppose that f is reducible in $Q[X]$. Thus $f = gh$ for some $g, h \in Q[X]$ where g, h are not units in $Q[X]$. Since the units of $Q[X]$ are the nonzero elements of Q, we must have $0 <$ deg $g <$ deg f and $0 <$ deg $h <$ deg f. By Proposition 6.3, $f = tg^*h^*$ where $t \in Q$ and g^*, h^* are primitive in $R[X]$ with deg $g^* =$ deg g, deg $h^* =$ deg h. Since g^*, h^* are primitive in $R[X]$, so is g^*h^* (by Gauss' Lemma). Since f and g^*h^* are primitive in $R[X]$, we have that t is a unit in R by Proposition 6.2. Thus $f = tg^*h^*$ is a factorisation of f in $R[X]$ with g^*, h^* nonunits; hence f is reducible in $R[X]$. This shows that for primitive $f \in R[X]$ with deg $f > 0$, if f is reducible in $Q[X]$, then f is reducible in $R[X]$; equivalently, for primitive $f \in R[X]$ with deg $f > 0$, if f is irreducible in $R[X]$, then f is irreducible in $Q[X]$. □

As an exercise, one proves the following.

Theorem 6.5 (Eisenstein's Criterion). *Suppose R is a UFD, $f = a_0 + a_1X + \cdots + a_nX^n \in R[X]$ is primitive, and p is an irreducible element of R so that $p|a_i$ for $0 \leq i < n$, $p^2 \nmid a_0$, and $p \nmid a_n$. Then f is irreducible in $R[X]$ (and hence f is irreducible in $Q[X]$ where Q is the field of fractions of R).*

We also use Gauss' Lemma to prove the following important result.

Theorem 6.6. *Let R be a UFD. Then $R[X]$ is a UFD.*

Proof. Suppose that f is a nonzero, nonunit element of $R[X]$. Suppose first that deg $f = 0$, or in other words, $f \in R$; then the only way we can factor f is as a product of elements of R, and since R is a UFD, such a factorisation of f is essentially unique.

So assume deg $f > 0$. Then $f = \alpha g$ where α is the content of f, and hence (by Proposition 6.1) g is primitive in $R[X]$. We now consider g as an element of $Q[X]$ where Q is the field of fractions of R. We know $Q[X]$ is a Euclidean domain, and hence $Q[X]$ is a UFD; thus we can write $g = g_1 \cdots g_k$ where $g_1, \ldots, g_k$ are irreducible elements of $Q[X]$ (and hence deg $g_i > 0$ for $i = 1, \ldots, k$). Then by Proposition 6.3, for each $i = 1, \ldots, k$ we have $g_i = \beta_i g_i^*$ where $\beta_i \in Q$ and g_i^* is primitive in $R[X]$. By Theorem 6.4, each g_i^* is irreducible in $R[X]$. Also, by Gauss' Lemma, $g_1^* \cdots g_k^*$ is primitive in $R[X]$. Thus we have

$$g = \beta g_1^* \cdots g_k^*, \text{ where } \beta = \beta_1 \cdots \beta_k.$$

By Proposition 6.2, we have that β is a unit in R. Thus we have $f = \alpha\beta g_1^* \cdots g_k^*$; if $\alpha\beta$ is a unit then $\alpha\beta g_1^*$ is irreducible, and otherwise we can factor $\alpha\beta$ as $\alpha\beta =$

$\gamma_1 \cdots \gamma_t$ where $\gamma_1, \ldots, \gamma_t$ are irreducible elements of R (and hence irreducible elements of $R[X]$). This gives us a factorisation of f as a product of irreducible elements of $R[X]$.

Now suppose still that $\deg f > 0$, and that we also have $f = \delta_1 \cdots \delta_s h_1 \cdots h_m$ where $\delta_1, \ldots, \delta_s$ are irreducible elements of R, and $h_1, \ldots, h_m$ are irreducible elements of $R[X]$ with $\deg h_j > 0$ for $j = 1, \ldots, m$. Thus $\delta_1 \cdots \delta_s$ is the content of f, as is $\gamma_1 \cdots \gamma_t$; since R is a UFD, we have $s = t$ and, by reordering the δ_i, we have that γ_i and δ_i are associates for $i = 1, \ldots, s$. Hence

$$f = \gamma_1 \cdots \gamma_s g_1^* \cdots g_k^* = w\gamma_1 \cdots \gamma_s h_1 \cdots h_m$$

where w is a unit. Since $R[X]$ is an integral domain and $\gamma_1 \cdots \gamma_2 \neq 0$, we have

$$g_1^* \cdots g_k^* = wh_1 \cdots h_m.$$

We have $g_1^*, \ldots, g_k^*, wh_1, h_2, \ldots, h_m$ irreducible (and thus primitive) in $R[X]$, and hence by Gauss' Lemma, they are irreducible in $Q[X]$ where Q is the field of fractions of R. Since Q is a field, we know $Q[X]$ is a Euclidean domain and hence a UFD; thus $k = m$, and reordering the h_i, we have that for each $i = 1, \ldots, k$, $g_i^* = u_i h_i$ where u_i is a unit in $Q[X]$. Thus u_i is a nonzero element of Q. Proposition 6.4 imples that u_i a unit in R, and hence g_i^* and h_i are associates in $R[X]$ for $i = 1, \ldots, k$.

This shows that in $R[X]$, factorisation into products of irreducibles is essentially unique; hence $R[X]$ is a UFD. □

Definition. Let R be a ring, f a nonzero element of $R[X]$. Let $n = \deg f$, and write $f = a_0 + a_1 X + \cdots + a_n X^n$. (So $a_n \neq 0$.) We say f is monic if $a_n = 1$. Also, a_n is called the leading coefficient of f.

§7. Testing polynomials for irreducibility.

As an exercise, one proves the following.

Theorem 7.1. *Let R be an integral domain, and I a prime ideal of R. Define $\varphi : R[X] \to (R/I)[X]$ by*

$$\varphi(a_0 + a_1 X + \cdots + a_n X^n) = \overline{a}_0 + \overline{a}_1 X + \cdots + \overline{a}_n X^n$$

where $\overline{a}_j = a_j + I$. Then φ is a surjective homomorphism. Suppose $f \in R[X]$ is primitive with its leading coefficient not in I; if $\varphi(f)$ is irreducible in $(R/I)[X]$, then f is irreducible in $R[X]$.

(One does not actually need I to be prime to prove this, provided one extends the definition of irreducible to rings other than integral domains – recall that R/I is an integral domain if and only if I is a prime ideal. Also, in practice, we will typically only use this result with I a maximal ideal, and thus R/I a field. It is useful to recall that when R is a PID, an integral ideal of R is also maximal.)

Note:

(1) Suppose $f \in R[X]$ with deg $f \leq 3$ and f primitive; if f is reducible in $R[X]$ then f must have a linear factor, or equivalently, f has a root in Q, the field of fractions of R. If f is monic and reducible in $R[X]$, then f has a root in R.

(2) Suppose $f \in \mathbb{Z}[X]$ with deg $f \leq 3$, f primitive and monic. If f is reducible in $\mathbb{Z}[X]$, then f has a root in $\mathbb{Z}$, and hence for any prime p, f has a root modulo p. Equivalently (still assuming deg $f \leq 3$), if there is a prime p so that f does not have a root modulo p then f is irreducible in $\mathbb{Z}[X]$.

Terminology: For R a ring and $f \in R[X]$, we say f is irreducible over R when f is irreducible in $R[X]$.

Example: Take $f = X^3+2X+1 \in \mathbb{Z}[X]$; note that f is primitive in $\mathbb{Z}[X]$. Modulo 5 (i.e. in $(\mathbb{Z}/5\mathbb{Z})[X]$), $X^3 + \overline{2}X + \overline{1}$ has no root in $\mathbb{Z}/5\mathbb{Z}$ (here $\overline{a} = a + 5\mathbb{Z}$). Hence f is irreducible over $\mathbb{Z}/5\mathbb{Z}$, and thus irreducible over $\mathbb{Z}$; then by Gauss' Lemma, f is irreducible over $\mathbb{Q}$.

One can easily prove the following.

Proposition 7.2. *Let R, R' be rings with $R \subseteq R'$; fix $\alpha \in R'$. Define $\varphi_\alpha : R[X] \to R'$ by*

$$\varphi_\alpha(c_0 + c_1X + \cdots + c_nX^n) = c_0 + c_1\alpha + \cdots + c_n\alpha^n.$$

Then φ_α is a homomorphism.

Definition. Let R, R' be rings with $R \subseteq R'$; fix $\alpha \in R'$. Let $f(\alpha)$ denote $\varphi_\alpha(f)$ where φ_α is defined as in Proposition 7.2. We say $\alpha \in R$ is a root of $f \in R[X]$ if$f \neq 0$ and $f(\alpha) = 0$.

Proposition 7.3. *Suppose K is a field and $f \in K[X]$, deg $f > 0$. If $\alpha \in K$ is a root of f, then $f = (X - \alpha)q$ for some $q \in K[X]$. If deg $f = 2$ or 3, and f has no root in K, then f is irreducible in $K[X]$.*

Proof. Since K is a field, we know K[X] is a Euclidean domain with the map deg. Thus there are $q, r \in K[X]$ so that $f = (X - \alpha)q + r$ where either $r = 0$ or deg $r <$ deg $(X - \alpha) = 1$; this means $r \in K$. Since $f(\alpha) = 0$, we must have $r = 0$. Hence $f = (X - \alpha)q$.

Suppose deg $f = 2$ or 3, and $f = gh$ for some $g, h \in K[X]$, g, h not units. Since $f \neq 0$, we know $g, h \neq 0$; since every nonzero element of K is a unit, we must have deg g, deg $h > 0$. Since deg $f =$ deg $g +$ deg h, we must have deg $g = 1$ or deg $h = 1$; without loss of generality, suppose deg $g = 1$. Thus $g = aX + b$, some $a, b \in K$ with $a \neq 0$. Since K is a field, we have $-b/a \in K$, and hence $g(-b/a) = 0$, which means $f(-b/a) = 0$. So when deg $f = 2$ or 3, if f is reducible in $K[X]$, then f has a root in K. Equivalently, when deg $f = 2$ or 3, if f has no root in K then f is irreducible. □

Example: Let $f = 2X^3 + 5X^2 + 5X + 3 \in \mathbb{Q}[X]$; is f irreducible over $\mathbb{Q}$? If f is reducible, then by Proposition 7.3, f must have a root in $\mathbb{Q}$. But how can we find such a root, or show none exists? On the other hand, $f \in \mathbb{Z}[X]$ and f is primitive; thus by (a consquence of) Gauss' Lemma, f is irreducible in $\mathbb{Q}[X]$ if and only if f

is irreducible in $\mathbb{Z}[X]$. However, we cannot apply Eisenstein's Criterion. But, since $\deg f = 3$, f is reducible over $\mathbb{Z}$ if and only if f has a linear factor in $\mathbb{Z}[X]$, and this is easier to test. So suppose we can factor f as

$$f = (a_1X + a_0)(b_2X^2 + b_1X + b_0).$$

Then we must have $a_1b_2 = 2$, $a_1b_1 + a_0b_2 = 5$, $a_1b_0 + a_0 + b_1 = 5$, $a_0b_0 = 3$. So $a_0 = \pm 1$ or ± 3, and $a_1 = \pm 1$ or ± 2. We can try all these possibilities and, for each one, try to simultaneously solve the equations for b_0, b_1, b_2; this is very tedious. On the other hand, if $a_1X + a_0$ is a factor of f, then in $\mathbb{Q}[X]$, $X + a_0/a_1$ is a factor of f, or equivalently, a_0/a_1 is a root of f. The possible values for a_0/a_1 are $\pm 1, \pm 3, \pm\frac{1}{2}, \pm\frac{3}{2}$. Testing these possibilities, one finds that $-\frac{3}{2}$ is a root of f.

Example: Let $f = 50X^3 + 49X^2 + 702 \in \mathbb{Z}[X]$. So f is primitive in $\mathbb{Z}[X]$, hence f is irreducible in $\mathbb{Q}[X]$ if and only if f is irreducible in $\mathbb{Z}[X]$. By Theorem 7.1, f is irreducible in $\mathbb{Z}[X]$ if the image of f is irreducible in $(\mathbb{Z}/p\mathbb{Z})[X]$ where p is a prime not dividing 50. Also, for p such a prime and writing $\overline{f}$ for the image of f in $(\mathbb{Z}/p\mathbb{Z})[X]$, we have $\deg \overline{f} = 3$; since $\mathbb{Z}/p\mathbb{Z}$ is a field, f is reducible in $(\mathbb{Z}/p\mathbb{Z})[X]$ if and only if f has a root in $\mathbb{Z}/p\mathbb{Z}$. Taking $p = 3$, we find f is reducible in $(\mathbb{Z}/3\mathbb{Z})[X]$. So we cannot apply Theorem 7.1. Take $p = 7$; then $\overline{f} = \overline{1} \cdot X^3 + \overline{2}$ in $(\mathbb{Z}/7\mathbb{Z})[X]$. The elements of $\mathbb{Z}/7\mathbb{Z}$ are $\overline{0}, \pm\overline{1}, \pm\overline{2}, \pm\overline{3}$. Testing, we find that for every $\overline{a} \in \mathbb{Z}/7\mathbb{Z}$, $\overline{f}(\overline{a}) \neq 0$. Thus f is irreducible in $(\mathbb{Z}/7\mathbb{Z})[X]$, and hence f is irreducible in $\mathbb{Z}[X]$. As discussed above, this means f is irreducible in $\mathbb{Q}[X]$.

Example: Say $f = X^4 - 2X^2 + 9 \in \mathbb{Q}[X]$. Then, since f is primitive in $\mathbb{Z}[X]$, f is irreducible in $\mathbb{Q}[X]$ if and only if f is irreducible in $\mathbb{Z}[X]$. Eisenstein's Criterion does not apply here. Using the techniques from the previous example, one can check that f has no linear factor modulo 7, so in $\mathbb{Z}[X]$ f has no linear factor. This does not mean f is irreducible in $\mathbb{Z}[X]$, since we could have $f = gh$ where $\deg g = \deg h = 2$. However, one can show by a direct (brute force) computation that this is not the case: Suppose $f = (uX^2 + aX + b)(vX^2 + cX + d)$ where $u, a, b, v, c, d \in \mathbb{Z}$. Then we must have $uv = 1$, $uc + va = 0$, $ad + bc = 0$, $bd = 9$, $ud + bv + ac = -2$. By elementary algebra, one deduces that the first four equations imply $u = v = \pm 1$, $a = -c$, $d = b$, so $d = b = \pm 3$. But then we cannot have $ud + bv + ac = -2$.

However, this is a very clumsy, labour intensive approach. Alternatively, we could work modulo some prime p to argue f cannot be factored in $\mathbb{Z}[X]$ as a product of two quadratic (i.e. degree 2) polynomials.

Given the shape of f, another approach is to factor f over $\mathbb{C}$:

$$\begin{aligned} f &= X^4 - 2X^2 + 9 \\ &= (X^2 - 1 + \sqrt{-8})(X^2 - 1 - \sqrt{-8}) \\ &= \left(X - \sqrt{-1 + \sqrt{-8}}\right)\left(X + \sqrt{-1 + \sqrt{-8}}\right) \\ &\quad \cdot \left(X - \sqrt{-1 - \sqrt{-8}}\right)\left(X + \sqrt{-1 - \sqrt{-8}}\right). \end{aligned}$$

Then one can show that no product of any two of these linear factors of f lies in $\mathbb{Q}[X]$. Hence, using that $\mathbb{C}$ is a field and thus $\mathbb{C}[X]$ is a UFD, there is no linear or quadratic factor of f that lies in $\mathbb{Q}[X]$.

One also proves the following.

Proposition 7.4. *Suppose $f \in \mathbb{R}[X]$ is irreducible. Then deg $f = 1$ or 2.*

We also have the next proposition on using substitution of variables to prove irreducibility or reducibility.

Proposition 7.5. *Suppose R is a UFD, $f \in R[X]$ so that f is primitive. With $u, v \in R$ so that u is a unit, set $Y = uX + v$. Write $f = a_0 + a_1X + \cdots + a_nX^n$, and set $g = a_0 + a_1Y + \cdots + a_nY^n$. Then f is irreducible in $R[X]$ if and only if g is irreducible in $R[X]$.*

Proof. Suppose f is reducible in $R[X]$. So there are $f_1, f_2 \in R[X]$ so that $f = f_1f_2$ and neither f_1 nor f_2 is a unit; since the content of f is the product of the content of f_1 and the content of f_2, we must have that f_1, f_2 are primitive. The only primitive elements of R are units, so we must have deg f_1, deg $f_2 > 0$. Write

$$f_1 = b_0 + b_1X + \cdots + b_mX^m,\ f_2 = c_1 + c_1X + \cdots + c_kX^k.$$

Then set $g_1 = b_0 + b_1Y + \cdots + b_mY^m$, $g_2 = c_1 + c_1Y + \cdots + c_kY^k$. Clearly $g = g_1g_2$; one checks that as polynomials in $R[X]$, deg $_g1 =$ deg f_1 and deg $g_2 =$ deg f_2. Hence g is reducible in $R[X]$.

By essentially the same argument, if g is reducible then f is reducible, since $X = u^{-1}Y - u^{-1}v$. □

Example: Take $f = X^7 + 7X - 1 \in \mathbb{Z}[X]$. Set $Y = X + 1$, $g = Y^7 + 7Y - 1$. We expand $(X + 1)^7$ using the Binomial Theorem; we can then use Eisenstein's Criterion with $p = 7$ to show g is irreducible in $\mathbb{Z}[X]$. Hence f is irreducible in $\mathbb{Z}[X]$, so by Gauss' Lemma, f is irreducible in $\mathbb{Q}[X]$.

§8. Field extensions and algebraic elements.

As an exercise, one proves the following.

Proposition 8.1. *Suppose K, L are fields and $\varphi : K \to L$ is a homomorphism. Then φ is injective.*

Definition. Suppose K, L are fields and $\varphi : K \to L$ is a homomorphism (and thus φ is necessarily injective). Then we say L is a field extension of K, or equivalently, that $L : K$ is a field extension. When K, L are fields with $K \subseteq L$, we assume $\varphi : K \to L$ is the identity map on K.

Proposition 8.2. *Suppose $L : K$ is a field extension. Then L is a vector space over K.*

Proof. Since $L : K$ is a field extension, there is a homomorphism $\varphi : K \to L$, and φ is necessarily injective.

First suppose φ is the identity map, i.e. $K \subseteq L$. To see L is a vector space over K, first note that L is an additive group. For $a \in K$ and $v \in L$, we define the scalar multiplication $a \cdot v$ to be av, the product of a and v in L. Then for all $v, w \in L$, $a, b \in K$, we have

$$1 \cdot v = v,\ a \cdot (v + w) = a \cdot v + a \cdot w,\ (a + b) \cdot v = a \cdot v + b \cdot v,\ (ab) \cdot v = a \cdot (b \cdot v).$$

So L is a vector space over K.

Now suppose K is not a subset of L. Then for $a \in K$, $v \in L$, we define the scalar multiplication $a \cdot v$ to be

$$a \cdot v = \varphi(a)v.$$

With this definition of scalar multiplication, one verifies as an exercise that L is a vector space over K. $\square$

Notation: Unless it will cause confusion, when $L : K$ is a field extension, we identify K with its isomorphic image in L; so for $a \in K, v \in L$, we write av for $a \cdot v$.

Definition. Suppose $L : K$ is a field extension. We define the degree of $L : K$ to be the dimension of L as a vector space over K. We use $[L : K]$ to denote the degree of $L : K$. We say $L : K$ is a finite extension if $[L : K] < \infty$.

Theorem 8.3. *Suppose $L : K$ and $M : L$ are field extensions. Then $M : K$ is a field extension, and*

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

Proof. As an exercise, one shows $M : K$ is a field extension.

To show $[M : K] = [M : L][L : K]$, first suppose $[L : K] = r < \infty$ and $[M : L] = s < \infty$. Let $\{x_1, \ldots, x_r\}$ be a basis for L over K, $\{y_1, \ldots, y_s\}$ a basis for M over L. We claim

$$\mathcal{B} = \{x_i \cdot y_j :\ 1 \leq i \leq r,\ 1 \leq j \leq s\ \}$$

is a basis for M over K.

We first show that, as a vector space over K, M is spanned by $\mathcal{B}$. Take $z \in M$. Thus there exist $\beta_1, \ldots, \beta_s \in L$ so that

$$z = \beta_1 \cdot y_1 + \cdots + \beta_s \cdot y_s.$$

For each i, $1 \leq i \leq s$, there exist $\alpha_{i1}, \ldots, \alpha_{ir} \in K$ so that

$$\beta_i = \alpha_{i1} \cdot x_1 + \cdots + \alpha_{ir} \cdot x_r.$$

Substituting, we get z as a K-linear combination of the elements of $\mathcal{B}$. So $\mathcal{B}$ is a spanning set for M over K.

Now we need to show $\mathcal{B}$ is a linearly independent set over K. So we suppose there are $\gamma_{ij} \in K$ so that

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\gamma_{ij}\cdot(x_i\cdot y_j)=0.$$

Thus

$$\sum_{j=1}^{s}\left(\sum_{i=1}^{r}\gamma_{ij}\cdot x_i\right)\cdot y_j=0,$$

and for each j, $\sum_{i=1}^{r}\gamma_{ij}\cdot x_i$ is an element of L. Since $\{y_1,\dots,y_s\}$ is a basis for M over L, we must have

$$\sum_{i=1}^{r}\gamma_{ij}\cdot x_i=0 \text{ for each } j \text{ with } 1\le j\le s.$$

But $\{x_1,\dots,x_r\}$ is a basis for L over K, so for each j, we must have $\gamma_{ij}=0$ for $1\le i\le r$. This shows $\mathcal{B}$ is a linearly independent set with rs elements.

Hence if $[M:L]$ and $[L:K]$ are finite, then so is $[M:K]$, and

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

Suppose now that $[M:K]=n<\infty$. Thus there is a basis $\{z_1,\dots,z_n\}$ for M over K. Since L contains (an isomorphic copy of) K, $\{z_1,\dots,z_n\}$ spans M over L, and so $[M:L]\le n<\infty$. Since L is a subspace of M, the dimension of L over K is bounded above by the dimension of M over K; so $[L:K]\le n<\infty$. Thus by our preceding argument, since $[M:L],[L:K]<\infty$, we have $[M:K]=[M:L][L:K]$.

We can also conclude from the above arguments that $[M:K]<\infty$ if and only if $[M:L],[L:K]<\infty$. Hence $[M:K]=\infty$ if and only if $[M:L]=\infty$ or $[L:K]=\infty$, and so we always have $[M:K]=[M:L][L:K]$. □

Definition. Say $L:K$ is a field extension and $\alpha\in L$. Identify K with its isomorphic image in L. We say α is algebraic over K if α is the root of some polynomial in $K[X]$. When α is not algebraic over K, we say α is transcendental over K. When every element of L is algebraic over K, we simply say L is algebraic over K.

Proposition 8.4. *Let $L:K$ be a field extension and $\alpha\in L$ so that α is algebraic over K. Then*

$$I=\{f\in K[X]:\ f(\alpha)=0\ \}$$

is a nonzero ideal of $K[X]$, and there is a unique monic polynomial $m_\alpha(K)\in K[X]$ that generates I.

Proof. We know that

$$\ker\varphi_\alpha=\{f\in K[X]:\ \varphi_\alpha(f)=0\ \}=I.$$

Since φ_α is a homomorphism, $\ker\varphi_\alpha$ is an ideal. Since α is algebraic over K, $I \neq (0)$.

We know $K[X]$ is a Euclidean domain, and hence a PID, so there exists some $g \in I$ so that $I = (g)$. Further, since K is a field, we can choose g to be monic. As an exercise, one shows that there is a unique monic polynomial g that generates I. $\square$

Definition. For $L : K$ a field extension with $\alpha \in L$ so that α is algebraic over K, the polynomial $\mathrm{m}_\alpha(K)$ from Proposition 8.4 is called the minimal polynomial of α over K.

Theorem 8.5. *Suppose $L : K$ is a field extension, and $\alpha \in L$ is algebraic over K. Let $g = m_\alpha(K)$ (where $m_\alpha(K)$ is the minimal polynomial of α over K). Then g is irreducible over K, and $K[X]/(g)$ is a field.*

Proof. Identify K with its isomorphic image in L. Define $\varphi_\alpha : K[X] \to L$ by $\varphi_\alpha(f) = f(\alpha)$.

We have seen that φ_α is a homomorphism, so

$$\ker\varphi_\alpha = \{f \in K[X] : \ f(\alpha) = 0 \ \}.$$

By Proposition 8.4, $\ker\varphi_\alpha = (g)$ where $g = \mathrm{m}_\alpha(K)$. Thus by the Fundamental Homormorphism Theorem, $K[X]/(g)$ is isomorphic to a subring of L. Since L is an integral domain, $K[X]/(g)$ is an integral domain, and hence (g) is a prime ideal. We know $K[X]$ is a Euclidean domain and hence a PID, and in a PID any prime ideal is maximal. Thus (g) is a maximal ideal, so g is irreducible. Also, since (g) is maximal, $K[X]/(g)$ is a field. $\square$

Theorem 8.6. *Let K be a field, $f \in K[X]$ irreducible. Then there exists a field extension $L : K$ so that L contains a root of f.*

Proof. Set $L = K[X]/(f)$. Since f is irreducible and $K[X]$ is a Euclidean domain (and hence a PID), (f) is maximal. Thus L is a field.

Set $I = (f)$. With $\varphi : K \to L$ defined by $\varphi(c) = c + I$, it is easily verified that φ is a homomorphism, and hence $L : K$ is a field extension.

We extend φ to $\varphi : K[X] \to L[Y]$ by defining

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

We claim that this extension of φ is a homomorphism. To see this, take $f = a_0 + a_1X + \cdots + a_nX^n$, $g = b_0 + b_1X + \cdots + b_mX^m \in K[t]$. Without loss of generality, assume $m \leq n$, and for $m < k \leq n$, set $b_k = 0$. So

$$\begin{aligned}
\varphi(f+g) &= \sum_{k=0}^{n} \varphi(a_k + b_k)Y^k \\
&= \varphi_{k=0}^{n}(\varphi(a_k) + \varphi(b_k))Y^k \\
&= \varphi_{k=0}^{n}\varphi(a_k)Y^k \ + \ \sum_{k=0}^{n} \varphi(b_k)Y^k \\
&= \varphi(f) + \varphi(g).
\end{aligned}$$

Somewhat similarly,

$$\begin{aligned}\varphi(f+g) &= \sum_{k=0}^{2n} \varphi\left(\sum_{\ell=0}^{k} a_\ell + b_{k-\ell}\right) Y^k \\ &= \sum_{k=0}^{2n}\sum_{\ell=0}^{k} \varphi(a_\ell b_{k-\ell}) Y^k \\ &= \sum_{k=0}^{2n}\sum_{\ell=0}^{k} \varphi(a_\ell)\varphi(b_{k-\ell}) Y^k \\ &= \left(\sum_{i=0}^{n} \varphi(a_i) Y^i\right)\left(\sum_{j=0}^{n} \varphi(b_j) Y^j\right) \\ &= \varphi(f)\varphi(g).\end{aligned}$$

We also know that $\varphi(1) = 1$, so $\varphi : K[X] \to L[Y]$ is a homomorphism.

In L, let $\overline{c} = c + I$ for any element of K, and set $\alpha = X + I$. Writing

$$f = c_0 + c_1 X + \cdots + c_n X^n,$$

we have

$$\begin{aligned}\overline{c}_0 + \overline{c}_1\alpha + \cdots + \overline{c}_n\alpha^n &= (c_0 + c_1 X + \cdots + c_n X^n) + I \\ &= f + I \\ &= 0 + I \\ &= \overline{0},\end{aligned}$$

the zero element of L. Thus α is a root of (the image of) f in L. $\square$

Definition. Let $L : K$ be a field extension, $\alpha \in L$.Assume $K \subseteq L$. Let $K[\alpha]$ denote the smallest subring of L containing K and α, and let $K(\alpha)$ be the smallest subfield of L containing K and α.

Proposition 8.7. *Let $L : K$ be a field extension, $\alpha \in L$. Assume $K \subseteq L$. Then*

$$K[\alpha] = \{c_0 + c_1\alpha + \cdots + c_d\alpha^d : \ d \in \mathbb{Z}_{\geq 0},\ c_0, \ldots, c_d \in K\ \}$$

(which is $\varphi_\alpha(K[X])$), and

$$K(\alpha) = \left\{ \frac{f}{g} : \ f, g \in K[\alpha],\ g \neq 0 \right\}.$$

Proof. Let

$$R = \{c_0 + c_1\alpha + \cdots + c_d\alpha^d : \ d \in \mathbb{Z}_{\geq 0},\ c_0, \ldots, c_d \in K\ \}.$$

It is easy to check that R is a subring of L containing K and α. Also, given any subring R' of L containing K and α, and given any element f of R, we must have $f \in R'$ since R' contains K and α, and R' is closed under addition and multiplication. Thus any subring of L containing K and α necessarily contains R. Thus R is the smallest subring of L containing K and α.

Let Q be the field of fractions of $K[\alpha]$; so

$$Q = \left\{ \frac{f}{g} : \ f, g \in K[\alpha], \ g \neq 0 \right\}.$$

So Q is a subfield of L containing K and α. Suppose Q' is a subfield of L containing K and α. Certainly Q' contains $K[\alpha]$. Take $f/g \in Q$; so $f, g \in K[\alpha]$ and $g \neq 0$. Thus $f, g \in Q'$, and since Q' is a field, we must have $1/g$ and $f \cdot 1/g \in Q'$. So Q' must contain Q. Hence Q is the smallest subfield of L containing K and α. $\square$

As exercises, one proves the following results.

Theorem 8.8. *Let $L : K$ be a field extension, $\alpha \in L$ algebraic over K. Assume $K \subseteq L$. Then with $n = deg\ m_\alpha(K)$, we have that $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ is a basis for $K(\alpha)$ over K. Further, $K[\alpha]$ is a field, and $K[\alpha] = K(\alpha)$.*

Note: This means that when $L : K$ is a field extension with $\alpha \in L$ algebraic over K and $n = \deg \mathrm{m}_\alpha(K)$,

$$K(\alpha) = K[\alpha] = \{c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1} : \ c_0, \dots, c_{n-1} \in K \ \}.$$

Proposition 8.9. *Let $L : K$ be a field extension, with $\alpha \in L$. Assume $K \subseteq L$. Then α is algebraic over K if and only if $[K(\alpha) : K] < \infty$.*

Proposition 8.10. *Let $L : K$ be a field extension, $\alpha \in L$ algebraic over K. Assume $K \subseteq L$. Then every element of $K(\alpha)$ is algebraic over K.*

§9. The characteristic of a field and finite fields.

We know examples of finite fields, namely $\mathbb{Z}/p\mathbb{Z}$ where p is prime. We explore finite fields more generally.

Definition. Let K be a field with additive identity 0_K and multiplicative identity 1_K. We write $2 \cdot 1_K$ to denote $1_K + 1_K$, $3 \cdot 1_K$ to denote $1_K + 1_K + 1_K$, etc. We define the characteristic of K, denoted $\mathrm{char}K$, to be the smallest positive integer n so that $n \cdot 1_K = 0_K$; if no such n exists, we define the characteristic of K to be 0.

Proposition 9.1. *Suppose K is a field.*

(a) Suppose $charK > 0$; then $charK$ is prime.

(b) Suppose $charK = p > 0$; then for all $x \in K$, we have $p \cdot x = 0$ (where $p \cdot x = x + \cdots + x$, p times).

Proof.

(a) Let $n = \mathrm{char}K$. First note that since $1_K \neq 0_K$, we cannot have $n = 1$.

Suppose $n = km$ for some $k, m \in \mathbb{Z}_+$. One easily checks that $n \cdot 1_K = (k \cdot 1_K)(m \cdot 1_K)$. Since $n \cdot 1_K = 0_K$, we have $(k \cdot 1_K)(m \cdot 1_K) = 0_K$. Since $k \cdot 1_K, m \cdot 1_K \in K$ and K is an integral domain, we must have $k \cdot 1_K = 0_K$ or $m \cdot 1_K = 0_K$. By the definition of char$K$, n is the smallest positive integer so that $n \cdot 1_K = 0_K$; thus k or m must equal n, and hence n must be a prime.

(b) For any $x \in K$, we have

$$\begin{aligned} p \cdot x &= x + \cdots + x \ (p \text{ times}) \\ &= 1_K x + \cdots + 1_K x \ (p \text{ times}) \\ &= (p \cdot 1_K)x \\ &= 0_K x \\ &= 0_K, \end{aligned}$$

proving the claim. □

Proposition 9.2. *Suppose K is a finite field.*
(a) char$K > 0$, and hence char$K = p$ for some prime p.
(b) With $p =$ char$K$, $|K| = p^r$ for some $r \in \mathbb{Z}_+$.

Proof. (a) Let $H = \{c \cdot 1_K : \ c \in \mathbb{Z}\}$, where $0 \cdot 1_K$ means 0_K, and for $m \in \mathbb{Z}_+$, $-m \cdot 1_K$ means $m \cdot (-1_K) = (-1_K) + \cdots + (-1_K)$ (m times). Since K is finite, H must be finite as well; in particular, the elements $1 \cdot 1_K, 2 \cdot 1_K, 3 \cdot 1_K, \ldots$ cannot all be distinct. So for some $c, d \in \mathbb{Z}_+$ with $c < d$, we have $c \cdot 1_K = d \cdot 1_K$. Hence we have $(d-c) \cdot 1_K = 0_K$ with $d - c > 0$, so char$K > 0$. By Proposition 9.1, char$K = p$ for some prime p.

(b) Define H as above. Note that with $m, n \in \mathbb{Z}_+$, $m \geq n$, we have $m \cdot 1_K + n \cdot 1_K = (m+n) \cdot 1_K$,

$$m \cdot 1_K + (-n) \cdot 1_K = m \cdot 1_K + n \cdot (-1_K) = (m-n) \cdot 1_K + n \cdot 1_K + n \cdot (-1_K) = (m-n) \cdot 1_K,$$

$$(-m) \cdot 1_K + n \cdot 1_K = (m-n) \cdot (-1_K) + n \cdot (-1_K) + n \cdot 1_K = (-m+n) \cdot 1_K.$$

Also, one easily verifies that with $m, n \in \mathbb{Z}_+$, $(m \cdot 1_K)(n \cdot 1_K) = mn \cdot 1_K$,

$$(-m \cdot 1_K)(n \cdot 1_K) = mn \cdot (-1_K \cdot 1_K) = -mn \cdot 1_K,$$

$$(-m \cdot 1_K)(n \cdot 1_K) = mn \cdot (-1_K)(-1_K) = mn \cdot 1_K.$$

Thus, defining $\varphi : \mathbb{Z} \to H$ by $\varphi(c) = c \cdot 1_K$, we find φ is a surjective homomorphism. As exercises, one shows that $\ker\varphi = (p) = p\mathbb{Z}$ where $p =$ char$K$, and then that H is a field with p elements.

Let $r = [K : H]$. Hence we have a basis $\{x_1, \ldots, x_r\}$ for K as a vector space over H. Thus for each $y \in K$, there exist unique $a_1, \ldots, a_r \in H$ so that $y = a_1 x_1 + \cdots + a_r x_r$. This means we have a bijection between K and $H^r = H \times \cdots \times H$ (r times). Hence $|K| = p^r$. □

The main goal for the remainder of this section is to show that the multiplicative subgroup of a finite field is cyclic. We will actually establish a more general result.

We begin with some results on finite abelian groups.

Recall: Let G be a group. For $x \in G$, the order of x, denoted $\text{ord}(x)$, is the smallest positive integer n so that $x^n = 1$, where 1 denotes the identity element of G; if no such n exists, we say the order of x is infinite. By Lagrange's Theorem, if G is finite, then the order of any $x \in G$ divides $|G|$, the order (equivalently, the cardinality) of G.

Proposition 9.3. *Let G be an abelian group with identity 1. Take $x, y \in G$ so that $ord(x) = a < \infty$ and $ord(y) = b < \infty$. Then $ord(xy)$ divides $ab/hcf(a,b)$. If $hcf(a,b) = 1$ then $ord(xy) = ab$.*

Proof. As an exercise, one shows that $x^k = 1$ for $k \in \mathbb{Z}_+$ if and only if $a|k$.

Let $c = \text{hcf}(a,b)$; so $a = a'c$, $b = b'c$ where $\text{hcf}(a',b') = 1$. Then $ab/\text{hcf}(a,b) = a'b'c$, and since G is abelian,

$$(xy)^{a'b'c} = (x^a)^{b'}(y^b)^{a'} = 1^{b'} \cdot 1^{a'} = 1.$$

Hence $\text{ord}(xy) < \infty$ and $\text{ord}(xy)$ divides $a'b'c$.

Now assume that $\text{hcf}(a,b) = 1$, and let $n = \text{ord}(xy)$. So from the preceeding argument, n divides ab. Also, $1 = (xy)^n = x^n y^n$. This means x^n is the inverse of y^n. Let $\langle x \rangle$ denote the cyclic group generated by x. So $x^n \in \langle x \rangle$; also, since x^n is the inverse of y^n, which lies in the group $\langle y \rangle$, we have $x^n \in \langle y \rangle$. As an exercise, one shows that $|\langle x \rangle| = a$ and $|\langle y \rangle| = b$. Hence by Lagrange's Theorem, $\text{ord}(x^n)$ divides both a and b; since $\text{hcf}(a,b) = 1$, this means $\text{ord}(x^n) = 1$ and hence $x^n = 1$. Thus $a|n$. Also, we have

$$1 = (xy)^n = x^n y^n = 1 \cdot y^n = y^n,$$

so $b|n$. Since $\text{hcf}(a,b) = 1$, we get $ab|n$. We already saw $n|ab$, so we have $n = ab$. □

As an exercise, one proves the following.

Proposition 9.4. *Let G be an abelian group, $x \in G$ with $ord(x) = n < \infty$. Suppose $n = ab$ where $a, b \in \mathbb{Z}_+$ with $hcf(a,b) = 1$. Then there exist elements $y, z \in G$ so that $x = yz$ and $ord(y) = a$, $ord(z) = b$.*

Proof. First, we show that $\text{ord}(x^a) = b$ and $\text{ord}(x^b) = a$:

Note that $(x^a)^b = x^{ab} = 1$. For any $k \in \mathbb{Z}_+$ with $k < b$, we cannot have $(x^a)^k = 1$, else we have $1 \le ak < ab$ with $x^{ak} = 1$, contradicting that $ab = \text{ord}(x)$.

Next, we use the fact that $\text{hcf}(a,b) = 1$ to show there exist $s, t \in \mathbb{Z}$ so that $x = yz$ where $y = x^{bt}$, $z = x^{as}$:

Since $\text{hcf}(a,b) = 1$, there exist $s, t \in \mathbb{Z}$ so that $as + bt = 1$. Set $y = x^{bt}$, $z = x^{as}$. Then

$$yz = x^{bt}x^{as} = x^{bt+as} = x^1 = x.$$

Finally, with $x = yz$ where y, z are as in (b), show that $\mathrm{ord}(y) = a$ and $\mathrm{ord}(z) = b$:

Let $k = \mathrm{ord}(y)$. So $1 = y^k = x^{kbt}$, and since $\mathrm{ord}(x) = ab$, we must have $ab|kbt$; thus we must have $a|kt$. We know $\mathrm{hcf}(a, t)$ divides $as + bt$; since $as + bt = 1$, we must have $\mathrm{hcf}(a, t) = 1$. Hence $a|k$. On the other hand,

$$y^a = (x^{ab})^t = 1^t = 1,$$

so $k|a$. Therefore $a = k = \mathrm{ord}(y)$.

A virtually identical argument shows $b = \mathrm{ord}(z)$. □

Proposition 9.5. *Suppose G is a finite abelian group. Then there is an element $x \in G$ so that for all $y \in G$, we have $ord(y)|ord(x)$.*

Proof. Since G is finite, we have $|G| = p_1^{h_1} \cdots p_r^{h_r}$ where $p_1, \ldots, p_r$ are distinct primes, and $h_1, \ldots, h_r \in \mathbb{Z}_+$. For each $i = 1, \ldots, r$, take a_i so that $\mathrm{ord}(a_i)$ is a power of p_i, and among all $z \in G$ so that $\mathrm{ord}(z)$ is a power of p_i, $\mathrm{ord}(a_i)$ is maximal. (Note that with $z = 1$, $\mathrm{ord}(z) = 1 = p_i^0$, so there is at least one element of G whose order is a power of p_i.) Set $x = a_1 \cdots a_r$.

Now take $y \in G$. We know $\mathrm{ord}(y)$ divides $|G|$, so $\mathrm{ord}(y) = p_1^{d_1} \cdots p_r^{d_r}$ where $0 \le d_i \le h_i$ for each $i = 1, \ldots, r$. By repeatedly applying Proposition 9.4, we get $y = b_1 \cdots b_r$ where $\mathrm{ord}(b_i) = p_i^{d_i}$ $(i = 1, \ldots, r)$. By our choice of a_i $(i = 1, \ldots, r)$, we know $\mathrm{ord}(a_i) = p_i^{c_i}$ with $d_i \le c_i$. Thus $\mathrm{ord}(y)|\mathrm{ord}(x)$. □

Theorem 9.6. *Let K be a field; set $K^\times = K \setminus \{0\}$ (so $K^\times$ is an abelian group under multiplication). Suppose G is a finite subgroup of $K^\times$. Then G is cyclic.*

Proof. Let $n = |G|$. By Proposition 9.5, there is some $x \in G$ so that for all $y \in G$, we have $\mathrm{ord}(y)|\mathrm{ord}(x)$. Let $k = \mathrm{ord}(x)$; so by Lagrange's Theorem, $k|n$ and hence $k \le n$. Also, for all $y \in G$, we have $\mathrm{ord}(y)|k$ and thus $y \in G$ is a root of the polynomial $X^k - 1$. We have $G \subset K$ and $K[X]$ is a UFD; thus $X^k - 1$ can have at most k roots in K. Since every element of G is a root of $X^k - 1$ and G has n elements, we must have $n \le k$. Since we already established that $k \le n$, we have $k = n$. So x is an element of G with order n, which means $\langle x \rangle$ is a cyclic subgroup of G with order n; since $n = |G|$, and so we must have $\langle x \rangle = G$. □

§10. Ruler and compass constructions: an introduction.

The topic of constructions by ruler (straight-edge) and compass is quite classical, and familiar to most of us from our early days in mathematics classes. Here we review basic constructions, and relate "constructible" points to the degree of a corresponding field extension of $\mathbb{Q}$.

From previous courses, we know that we can perform the following constructions:

(1) Bisect a given line segment.
(2) Bisect a given angle.
(3) Construct a line perpendicular to a given line or line segment.
(4) Construct a line parallel to a given line or line segment.

(5) Using a given line segment to define 1 unit of length, we can measure 1 unit in length on another given line or line segment.

Now we discuss further constructions. We begin with a plane and two (distinct) points O and X; we think of O as the origin, and X as the point $(1, 0)$. We think of the line through O and X as the x-axis. We use the distance between O and X to define one unit in length. We construct the line passing through O that is perpendicular to the line passing through O and X; we think of this as the y-axis.

Definition. A real number a is constructible if it is possible, using ruler and compass only, to construct a line segment of length $|a|$ in the plane where O is the origin, and where 1 unit in length is the distance from O to X.

Example: $\mathbb{Z}$ consists of constructible numbers.

Proposition 10.1. *Let $a, b \in \mathbb{R}$ be nonzero constructible numbers, $a > 0$. Then*

$$a + b, ab, a/b, \sqrt{a}$$

are also constructible.

Proof. One shows as an exercise that $a + b$ is constructible.

To show $ab, a/b$ are constructible, it suffices to consider the case where $b > 0$. Then, to construct ab and a/b, we begin with a line segment OA of length a; fix a point Q not on the line through O and A. On the line through O and Q, fix points U and B so that the length of the segment OU is 1, and the length of the segment OB is b. Now construct the line L through B that is parallel to the line through A and U; let D be the point where L intersects the line through O and A. Let x denote the distance from O to D. Since the triangles $\triangle OAU$ and $\triangle ODB$ are similar, we have that $a/x = 1/b$; hence $x = ab$, so ab is constructible. Now let L' be the line through U that is parallel to the line through A and B; let D' be the point where L' intersects the line through O and A, and let x' denote the distance from O to D'. Thus $\triangle OAB$ and $\triangle OD'U$ are similar triangles, so $x'/a = 1/b$; hence $x' = a/b$ and thus a/b is constructible.

To construct $\sqrt{a}$, let A be a point on the ray beginning at O and passing through X so that the distance from X to A is a. Since we can bisect line segments, we can construct a circle of diameter $a+1$ whose center is the midpoint of the line segment between O and A. Let L be the line passing through X that is perpendicular to the line through O and X. Let B be a point where L intersects the circle, and let x denote the distance from X to B. Since triangle $\triangle OBA$ is inscribed in a circle, with one side on a diameter of the circle, we know angle $\angle OBA$ is a right angle. Since they share angle $\angle BOX$ (which is the same as $\angle BOA$), triangles $\triangle OBA$ and $\triangle OXB$ are similar. Hence $\angle OAB$ is equal to $\angle OBX$. Also, $\angle OAB$ is the same as $\angle XAB$, so the triangles $\triangle XAB$ and $\triangle XBO$ are similar. Hence $1/x = x/a$, and from this we deduce $x^2 = a$, so $x = \sqrt{a}$. $\square$

Definition. A point P is constructible if there exists a finite sequence $P_0, \ldots, P_n$ of points so that $P_0 = O$, $P_1 = X$, $P_n = P$, and the following property holds. For $1 \leq j \leq n$, let

$$S_j = \{P_0, \ldots, P_j\}.$$

For each j with $2 \leq j \leq n$, P_j is one of the following:

(i) the intersection of two distinct straight lines, each joining two points of S_{j-1};

(ii) a point of intersection of a straight line joining two points of S_{j-1} and a circle with centre a point of S_{j-1} and radius the distance between two points of S_{j-1};

(iii) a point of intersection of two distinct circles, each with centre a point of S_{j-1} and radius the distance between two points of S_{j-1}.

Theorem 10.2. *Let $P = (a, b)$ be a constructible point in the plane. Then*

$$[\mathbb{Q}(a, b) : \mathbb{Q}] = 2^t$$

for some non-negative integer t; here $\mathbb{Q}(a, b) = (\mathbb{Q}(a))(b)$.

Proof. Since P is constructible, there is a sequence of points $P_0, \ldots, P_n$ as in the above definition. Let $P_j = (a_j, b_j)$; set $K_1 = \mathbb{Q}$, and for $2 \leq j \leq n$, set

$$K_j = K_j(a_{j+1}, b_{j+1}) = \mathbb{Q}(a_1, b_1, \ldots, a_j, b_j).$$

We know

$$[K_n : \mathbb{Q}] = [K_n : K_{n-1}][K_{n-1} : K_{n-2}] \cdots [K_2 : K_1].$$

We also know that $(a, b) = (a_n, b_n)$ and $[K_n : \mathbb{Q}] = [K_n : \mathbb{Q}(a, b)][\mathbb{Q}(a, b) : \mathbb{Q}]$. So $[\mathbb{Q}(a, b) : \mathbb{Q}]$ divides $[K_n : \mathbb{Q}]$, so if $[K_n : \mathbb{Q}]$ is a power of 2, so is $[\mathbb{Q}(a, b) : \mathbb{Q}]$. Thus to prove the theorem, it suffices to show that we have $[K_{j+1} : K_j] = 1$ or 2.

Case 1. Suppose (a_{j+1}, b_{j+1}) is the intersection of two straight lines, each joining points of S_j. So there are $(a_k, b_k), (a_m, b_m), (a_n, b_n), (a_r, b_r) \in S_j$ so that (a_{j+1}, b_{j+1}) is on the line through (a_k, b_k) and (a_m, b_m), and on the line through (a_n, b_n) and (a_r, b_r). Thus (a_{j+1}, b_{j+1}) is on the line described by

$$(Y - b_k)(a_m - a_k) = (X - a_k)(b_m - b_k),$$

or equivalently, (a_{j+1}, b_{j+1}) is a root of

$$(X - a_k)(b_m - b_k) - (Y - b_k)(a_m - a_k) \in K_j[X, Y].$$

Similarly, (a_{j+1}, b_{j+1}) is a root of

$$(X - a_n)(b_r - b_n) - (Y - b_n)(a_r - a_n) \in K_j[X, Y].$$

Solving, we find $a_{j+1}, b_{j+1} \in K_j$, so $[K_{j+1} : K_j] = 1$.

Case 2. Suppose (a_{j+1}, b_{j+1}) is a point of intersection of a line and a circle constructed using K_j. So there are $(a_k, b_k), (a_m, b_m), (a_n, b_n), (a_r, b_r), (a_s, b_s) \in S_j$ so that (a_{j+1}, b_{j+1}) is on the line through (a_k, b_k) and (a_m, b_m), and on the circle with centre (a_n, b_n) and radius the distance between (a_r, b_r) and (a_s, b_s). Hence (a_{j+1}, b_{j+1}) is a root of

$$(X - a_k)(b_m - b_k) - (Y - b_k)(a_m - a_k) \in K_j[X, Y]$$

and of

$$(X - a_n)^2 + (Y - b_n)^2 - (a_r - a_s)^2 - (b_r - b_s)^2 \in K_j[X, Y].$$

Thus (a_{j+1}, b_{j+1}) is a root of polynomials of the form

$$uX + vY + w, X^2 + Y^2 + u'X + v'Y + w' \in K_j[X, Y].$$

First suppose $u \neq 0$. Then by solving $uX + vY + w = 0$ for X and substituting into the second polynomial, we obtain a quadratic polynomial $f \in K_j[Y]$. Suppose first that f has a root α in K_j; then $f = c(Y - \alpha)(Y - \beta)$ with $c, \alpha\beta \in K_j$. Thus $b_{j+1} = \alpha$ or β, so $b_{j+1} \in K_j$; solving for a_{j+1} we get $a_{j+1} \in K_j$. Now suppose f does not have a root in K_j; then since deg $f = 2$, f is irreducible in K_j. We know b_{j+1} is a root of f, so $[K_j[b_{j+1}] : K] = \deg f = 2$. Now solving for a_{j+1}, we find $a_{j+1} \in K_j[b_{j+1}]$, so $K_{j+1} = K_j[a_{j+1}, b_{j+1}] = K_j[b_{j+1}]$. Hence $[K_{j+1} : K_j] = 2$.

Suppose $u = 0$; then we proceed as above with the roles of X and Y reversed.

Case 3. Suppose (a_{j+1}, b_{j+1}) is a point of intersection of two circles constructed using K_j; thus (a_{j+1}, b_{j+1}) is a root of two polynomials

$$X^2 + Y^2 + uX + vY + w, X^2 + Y^2 + u'X + v'Y + w' \in K_j[X, Y].$$

Hence (a_{j+1}, b_{j+1}) is a root of

$$(u - u')X + (v - v')Y + (w - w') \in K_j[X, Y].$$

We cannot have $u = u'$ and $v = v'$, else the circles would be concentric and thus would either be equal or have no point of intersection. So this case reduces to the previous case.

Thus in all cases, $[K_{j+1} : K_j] = 1$ or 2, so as discussed at the beginning of the proof, the theorem now follows. □

Remark: One can also show that an angle of $\pi/3$ radians is constructible, but cannot be trisected using ruler and compass.