M2P4

Rings and Fields

Mathematics
Imperial College London

As lectured by Professor Alexei Skorobogatov
and humbly typed by `as1005@ic.ac.uk`.

Contents

Chapter 1

Basic Properties Of Rings

Definition 1.1. A *ring* R is a set with two binary operations, $+$ and $\cdot$, satisfying: *ring*

(1) $(R, +)$ is an abelian group,

(2) R is closed under multiplication, and $(ab)c = a(bc)$ for all $a, b, c \in R$,

(3) $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ for all $a, b, c \in R$.

Example 1.2 (Examples of rings). 1. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.

2. $2\mathbb{Z}$ – even numbers. Note that $1 \notin 2\mathbb{Z}$.

3. $\mathrm{Mat}_n(\mathbb{R}) = \{n \times n\text{-matrices with real entries}\}$
   In general $AB \neq BA$.

   A ring R is called *commutative* if $ab = ba$ for all $a, b \in R$. *commutative*

4. Fix m, a positive integer. Consider the remainders modulo m: $\overline{0}, \overline{1}, ..., \overline{m-1}$.

   Notation. Write $\overline{n}$ for the set of all integers which have the same remainder as n when divided by m. This is the same as $\{n + mk \mid k \in \mathbb{Z}\}$. Also, $\overline{n_1} + \overline{n_2} = \overline{n_1 + n_2}$, and $\overline{n_1} \cdot \overline{n_2} = \overline{n_1 n_2}$. The classes $\overline{0}, \overline{1}, \ldots, \overline{m-1}$ are called residues modulo m. $\overline{n}$
   The set $\left\{\overline{0}, \overline{1}, ..., \overline{m-1}\right\}$ is denoted by $\mathbb{Z}_m$ or by $\mathbb{Z}/m$ or by $\mathbb{Z}/m\mathbb{Z}$. $\mathbb{Z}/m$

5. The set of polynomials in x with coefficients in $\mathbb{Q}$ (or in $\mathbb{R}$ or $\mathbb{C}$)

   $$\left\{a_0 + a_1 x + ... + a_n x^2 \mid a_i \in \mathbb{Q}\right\} = \mathbb{Q}[x]$$

   with usual addition and multiplication. If $a_n \neq 0$ then n is the *degree* of the polynomial.

Definition 1.3. A *subring* of a ring R is a subset which is a ring under the same addition and multiplication. *subring*

Proposition 1.4. Let S be a non-empty subset of a ring R. Then S is a subring of R if and only if, for any $a, b \in S$ we have $a + b \in S$, $ab \in S$ and $-a \in S$.

Proof. A subring has these properties. Conversely, if S is closed under addition and taking the relevant inverse, then $(S, +)$ is a subgroup of $(R, +)$ (from group theory).
S is closed under multiplication.
Associativity and distributivity hold for S because they hold for R. ∎

$\mathbb{Z}[\sqrt{m}]$
Gaussian integers

Definition 1.5. Let d be an integer which is not a square. Define $\mathbb{Z}[\sqrt{m}] = \{a + b\sqrt{m} \mid a, b \in \mathbb{Z}\}$.
Call $\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1}, a, b \in \mathbb{Z}\}$ the *ring of Gaussian integers*.

Proposition 1.6. $\mathbb{Z}[\sqrt{d}]$ is a ring. Moreover, if $m+n\sqrt{d} = m'+n'\sqrt{d}$, then $m = m'$ and $n = n'$.

Proof. Clearly $\mathbb{Z}[\sqrt{d}] \subset \mathbb{C}$. Consider $m, n, a, b \in \mathbb{Z}$. Then we have:
Closure under addition: $(m + n\sqrt{d}) + (a + b\sqrt{d}) = (m + a) + (n + b)\sqrt{d}$.
Closure under multiplication: $(m + n\sqrt{d})(a + b\sqrt{d}) = ma + nbd + (mb + na)\sqrt{d}$.
Also, $-(m + n\sqrt{d}) = (-m) + (-n)\sqrt{d}$.
Hence $\mathbb{Z}[\sqrt{d}] \subset \mathbb{C}$ is a subring by Proposition 1.4.
Finally, if $m + n\sqrt{d} = m' + n'\sqrt{d}$, then if $n \neq n'$ we write $\sqrt{d} = \frac{m-m'}{n'-n}$ which is not possible since d is not a square. Therefore, $n = n'$ hence $m = m'$. ∎

Proposition 1.7. For any two elements r, s of a ring, we have

(1) $r0 = 0r = 0$,

(2) $(-r)s = r(-s) = -(rs)$.

Proof.

(1) $r0 = r(0 + 0) = r0 + r0$. Adding $-(r0)$ to both sides, we get:

$$0 = r0 - (r0) = r0 + r0 - r0 = r0.$$

(2) $0 = 0s$ by (1) and $0 = 0s = (-r + r)s = (-r)s + rs$. Add $-(rs)$ to both sides to get $-(rs) = (-r)s$. Similarly, $r(-s) = -(rs)$. ∎

zero divisor

An element $a \neq 0$ of a ring R is called a *zero divisor* if there exists $b \neq 0 \in R$ such that $ab = 0$

For example, consider residues mod 4 : $\overline{0}, \overline{1}, \overline{2}, \overline{3}$. Take $\overline{2} \times \overline{2} = \overline{2 \times 2} = \overline{4} = \overline{0}$. Hence $\overline{2}$ is a zero divisor in $\mathbb{Z}/4$.

integral domain

Definition 1.8. A ring R is called an *integral domain* if

(1) R is commutative, i.e. $ab = ba$ for all $a, b \in R$,

(2) R has an identity under multiplication (written as 1),

(3) R has no zero divisors,

(4) $0 \neq 1$.

Note. If $0 = 1$, then $x \cdot 1 = x$ and so $x = x \cdot 1 = x \cdot 0 = 0$. Hence if $0 = 1$ then $R = \{0\}$.

For example $\mathbb{Z}$, $\mathbb{Z}[\sqrt{d}]$, $\mathbb{Q}$, $\mathbb{Q}[x]$ are integral domains.

Notation. If R is an integral domain (or any ring), then $R[x]$ denotes the set of polynomials in x with coefficients from R with usual addition and multiplication. Clearly $R[x]$ is a commutative ring. $R[x]$

Proposition 1.9. If R is an integral domain, then so is $R[x]$.

Proof. The only non-obvious thing to check is that there are no zero divisors. For contradiction, assume that $f(x) = a_0 + a_1x + ... + a_mx^m$, $g(x) = b_0 + ... + b_nx^n$ are elements of $R[x]$ such that $f(x)g(x)$ is the zero polynomial. Without loss of generality assume that $a_m \neq 0$, $b_n \neq 0$ (i.e. $m = \deg f(x)$, $n = \deg g(x)$). Then $f(x)g(x) = a_0b_0 + ... + a_mb_nx^{m+n}$.
Since R is an integral domain $a_mb_n \neq 0$. Therefore we get a contradiction, hence $f(x)g(x)$ can't be the zero polynomial. ■

Proposition 1.10. Let m be a positive integer. Then $\mathbb{Z}/m$ is an integral domain if and only if m is prime.

Proof. If $m = 1$ then $\mathbb{Z}/1 = \{0\}$; it is not an integral domain because $0 = 1$ in this ring.
If $m > 1$ and $m = ab$, $a > 1$, $b > 1$, then $\overline{a}, \overline{b} \in \mathbb{Z}/m$ are non-zero elements.
But $\overline{a}\overline{b} = \overline{ab} = \overline{m} = \overline{0}$, so $\overline{a}$ and $\overline{b}$ are zero divisors, hence $\mathbb{Z}/m$ is not an integral domain. Now assume $m = p$ is prime. Assume that $1 \leq a < m$, $1 \leq b < m$ such that $\overline{a}\overline{b} = \overline{ab} = \overline{0}$ in $\mathbb{Z}/p$. Visibly $\overline{a} \neq 0$, $\overline{b} \neq \overline{0}$.
This means that $p|ab$, but then $p|a$ or $p|b$. Then $\overline{a} = 0$ or $\overline{b} = 0$. Contradiction. ■

Proposition 1.11. Every integral domain R satisfies the *cancellation property* – if $ax = ay$ and $a \neq 0$ then $x = y$ for all $x, y, a \in R$.

Proof. If $ax = ay$ then $a(x - y) = 0$. Since R has no zero divisors and $a \neq 0$, we conclude that $x - y = 0$, so that $x = y$. ■

Definition 1.12. A ring F is a *field* if the set of non-zero elements of F forms an abelian group under multiplication. *field*

Note. The key thing is the existence of x^{-1}, the multiplicative inverse. Also, $xy = yx$ and $1 \in F$.
$\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/2, \mathbb{Z}/3$ are fields. $\mathbb{Z}/2 = \{\overline{0}, \overline{1}\}$. $\mathbb{Z}/3 = \{\overline{0}, \overline{1}, \overline{2}\} = \{\overline{0}, \overline{1}, -\overline{1}\}$.

Is $\mathbb{Z}[\sqrt{d}]$ a field? Of course not, since $\frac{1}{2} \notin \mathbb{Z}[\sqrt{d}]$.
Define $\mathbb{Q}[\sqrt{d}] = \left\{x + y\sqrt{d} \mid x, y \in \mathbb{Q}\right\}$. This is a field.
Indeed (assuming $x \neq 0, y \neq 0$):

$$\begin{aligned} \frac{1}{x + y\sqrt{d}} &= \frac{x - y\sqrt{d}}{\left(x - y\sqrt{d}\right)\left(x + y\sqrt{d}\right)} \\ &= \frac{x - y\sqrt{d}}{x^2 - y^2d}. \end{aligned}$$

Note that $x^2 - y^2d \neq 0$ since d is not a square of a rational number.

subfield

Definition 1.13. A subset S of a field F is a *subfield* if S is a field with the same addition and multiplication.
To check that S is a subfield, it is enough to check that for any $a, b \in S$, $a+b$, $-a$ and $ab \in S$, and for any $a \in S$, $a \neq 0$, $a^{-1} \in S$.

$F(\alpha_1, \dots, \alpha_n)$

Definition 1.14. Let F be a subfield of K and $\alpha_1, \dots, \alpha_n \in K$. Then $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ is the smallest subfield of K containing F and $\alpha_1, \dots, \alpha_n$.

Example 1.15. This notation agrees with $\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}$.
Let's check that $\mathbb{Q}(\sqrt{d})$ is indeed the smallest subfield of $\mathbb{C}$ containing $\mathbb{Q}$ and $\sqrt{d}$. The smallest subfield must contain all numbers like $a\sqrt{d}$, $a \in \mathbb{Q}$, since it is closed under $\cdot$, and hence also all numbers like $a + a'\sqrt{d}$, $a, a' \in \mathbb{Q}$, since closed under $+$. We also know that $\{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}$ is a field.

Similarly we can consider $\mathbb{Q}(\sqrt{d_1}, \sqrt{d_2})$, and more complicated fields.

Proposition 1.16.

(1) Every field is an integral domain.

(2) Every *finite* integral domain is a field.

Proof.

(1) Must check that there are no zero divisors. Suppose that $ab = 0$, $a \neq 0, b \neq 0$. Then a^{-1} exists, $a^{-1}ab = a^{-1}0 = 0$, so $b = 0$, a contradiction.

(2) The only thing to check is that every non-zero element is invertible. Let $R = \{r_1, \dots, r_n\}$ (distinct elements) be our integral domain. Take any $r \in R, r \neq 0$. Consider $\{rr_1, rr_2, \dots, rr_n\}$. If for some i and j we have $rr_i = rr_j$ then $r_i = r_j$ by the cancellation property.
Therefore $\{rr_1, rr_2, \dots, rr_n\}$ is a set of n distinct elements of R. Since R has n elements, $\{rr_1, rr_2, \dots, rr_n\} = R = \{r_1, \dots, r_n\}$. Thus any r_i can be written as rr_j for some j.
In particular, $1 = r \cdot r_j$ for some j, hence $r_j = r^{-1}$. ∎

Corollary 1.17. The ring $\mathbb{Z}/m = \left\{\overline{0}, \overline{1}, \dots, \overline{m-1}\right\}$ is a field if and only if m is prime.

Proof.

$\Rightarrow$ If m is not prime then we know that $\mathbb{Z}/m$ has zero divisors, hence is not a field.

$\Leftarrow$ If m is a prime, then $\mathbb{Z}/m$ is a finite integral domain, hence a field by the previous proposition. ∎

Chapter 2

Factorizing In Integral Domains

Let R be an integral domain.

Definition 2.1. If $r, s \in R$ and $s = rt$ for some $t \in R$, then we say that r *divides* s. This is written as $r|s$. *divides* $r|s$

Example 2.2.

1. If $R = \mathbb{Z}$, this is the usual concept of divisibility.

2. If $R = \mathbb{Z}[i]$, then $(2+i)|(1+3i)$. Divide $\dfrac{1+3i}{2+i} = \dfrac{(1+3i)(2-i)}{(2+i)(2-i)} = \dfrac{2+3+6i-i}{5} = 1+i \in \mathbb{Z}[i]$.

3. $R = \mathbb{Z}[\sqrt{d}]$. Take $r \in \mathbb{Z}$. If $r|x+y\sqrt{d}$, then $r|x$ and $r|y$.
   Indeed, $r|x+y\sqrt{d}$ is equivalent to the existence of $a+b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ such that $r(a+b\sqrt{d}) = x+y\sqrt{d}$ iff $ra = x$ and $rb = y$.

4. If R is a field, e.g. $R = \mathbb{Q}$ or $\mathbb{R}$, then for any $a, b \in R, a \neq 0$, we can write $b = ac$ for some $c \in R$ by taking $c = a^{-1}b$, so that $a|b$.

5. If F is a field, and R is a ring of polynomials $R = F[x]$, then $f(x)|g(x)$ if $g(x) = f(x)h(x)$ for some $h \in F[x]$. This is the usual notion of divisibility of polynomials.

Definition 2.3. If $a \in R$ then $aR = \{ar \mid r \in R\}$. aR

Note ($*$). The following are equivalent:

(1) $a|b$,

(2) $b \in aR$,

(3) $bR \subset aR$.

Definition 2.4. Element $u \in R$ is a *unit* (or an *invertible element*) if $uv = 1$ for some $v \in R$, i.e. there exists $u^{-1} \in R$. *unit*

Example 2.5. The units in $\mathbb{Z}$ are ± 1.

Notation. If R is a ring, we denote by R^* the set of units of R. R^*

In general, R^* is *not* the same as $R \setminus \{0\}$.

Example 2.6 (of units).

1. $\mathbb{Z}^* = \{\pm 1\}$.

2. Clearly, an integral domain F is a field iff $F^* = F \setminus \{0\}$.

3. $\mathbb{Z}[i]^* = \{1, -1, i, -i\}$: Suppose that $a+bi \in \mathbb{Z}[i]^*$ is a unit, so $(a+bi)(c+di) = 1$ for some $c, d \in \mathbb{Z}$. Then also $(a-bi)(c-di) = 1$. So

$$\begin{aligned} (a+bi)(c+di)(a-bi)(c-di) &= 1 \\ (a^2+b^2)(c^2+d^2) &= 1 \end{aligned}$$

   hence $a^2 + b^2 = 1$, so clearly $a + bi \in \{1, -1, i, -i\}$.

4. Consider $\mathbb{Z}[\sqrt{d}]$ where $d < -1$. Suppose $a + b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]^*$. Then for some $c, e \in \mathbb{Z}$,

$$\begin{aligned} (a+b\sqrt{d})(c+e\sqrt{d}) &= 1, \\ (a-b\sqrt{d})(c-e\sqrt{d}) &= 1, \\ (a^2-db^2)(c^2-de^2) &= 1. \end{aligned}$$

   This implies that $a^2 - db^2 = 1$. If $b = 0$, then $a = \pm 1$. If $b \neq 0$, then $b^2 \geq 1$ and $-db^2 \geq 2$, hence $a^2 - db^2 = 1$ has no solutions for $b \neq 0$. Conclude that if $d < -1$, then $\mathbb{Z}[\sqrt{d}]^* = \{\pm 1\}$.

5. Let $R = F[x]$ be the ring of polynomials with coefficients in a field F. We claim that $F[x]^* = F^*$. Let us show that a polynomial of degree ≥ 1 is never invertible in $F[x]$. Indeed, if $f(x) \in F[x]$, $\deg f \geq 1$, and $g(x) \in F[x]$ $(g(x) \neq 0)$ then $\deg f(x)g(x) = \deg f(x) + \deg g(x) \geq 1$. But $\deg 1 = 0$, hence $f(x)g(x)$ is never the polynomial 1.

irreducible

Definition 2.7. An element r of an integral domain R is called *irreducible* if

(1) $r \notin R^*$,

(2) if $r = ab$, then a or b is a unit.

reducible

Note. An element $r \in R$ is *reducible* if $r = st$ for some $s, t \in R$ where neither s nor t is a unit. Therefore $r \in R$ is irreducible if it is not reducible and is not a unit.

Example 2.8.

1. The irreducible elements in $\mathbb{Z}$ are $\pm p$, where p is a prime number.

2. Let $R = \mathbb{Z}[i]$. Then 3 is irreducible, whereas $2 = (1+i)(1-i)$ and $5 = (1+2i)(1-2i)$ are not. Indeed, $1+i, 1-i, 1+2i, 1-2i$ are not units. If 3 is reducible, then $3 = (a+bi)(c+di)$ and also $3 = (a-bi)(c-di)$, then

$$\begin{aligned} 9 &= (a+bi)(a-bi)(c+di)(c-di) \\ &= (a^2+b^2)(c^2+d^2). \end{aligned}$$

Consider the possibilities

$$\begin{aligned} 9 &= 9 \times 1, \\ &= 1 \times 9, \\ &= 3 \times 3. \end{aligned}$$

Therefore either $a^2 + b^2 = 1$ and then $a + bi$ is a unit, or $c^2 + d^2 = 1$ and then $c + di$ is a unit. Therefore $a^2 + b^2 = 3$, which has no solutions in $\mathbb{Z}$. Therefore 3 cannot be written as a product of non-units. Since 3 is not a unit, it is by definition irreducible.

3. We claim that 2 is an irreducible element of $\mathbb{Z}[\sqrt{-3}]$. If $2 = (a + b\sqrt{-3})(c + d\sqrt{-3})$, then $4 = (a^2 + 3b^2)(c^2 + 3d^2)$. If, say $a^2 + 3b^2 = 1$, then $a + b\sqrt{-3} = \pm 1$. Otherwise $2 = a^2 + 3b^2$, which has no solutions in $\mathbb{Z}$. Therefore 2 is irreducible.

4. In $\mathbb{R}[x]$ the polynomial $x^2 + 1$ is irreducible. But in $\mathbb{C}[x]$, $x^2 + 1 = (x + i)(x - i)$, and $x + i, x - i$ are not units, hence $x^2 + 1$ is reducible in $\mathbb{C}[x]$. An irreducible element of a polynomial ring $F[x]$, where F is a field, is the same as the irreducible polynomial.

Definition 2.9. Two elements $a, b \in R$ are called *associates* if $a = bu$ for some $u \in R^*$. *associates*

For example, a, b are associates in $\mathbb{Z}$ iff $a = \pm b$, a and b are associates in $\mathbb{Z}[i]$ iff $a = \pm b$ or $a = \pm ib$.

Proposition 2.10. Elements a and b are associates in an integral domain R iff (the following are equivalent)

(1) $a = bu$ for some $u \in R^*$,

(2) $b = av$ for some $v \in R^*$,

(3) $a|b$ and $b|a$,

(4) $aR = bR$.

Proof. (1) is the definition. Since $a = bu$ implies $b = au^{-1}$ with $u^{-1} \in R^*$, (1) implies (2) and (3). For (3) implies (1), consider $b = sa$ for some $s \in R$ and $a = tb$ for some $t \in R$. Then by the cancellation property, if $a \neq 0$ we have that $ts = 1$. If $a = 0$ then $b = 0$ and clearly a and b are associates. Otherwise t, s are units, hence again a and b are associates. Finally, (3) iff (4) by Note $(*)$. ■

Definition 2.11. An integral domain R is called a *unique factorization domain* (*UFD*) if the following hold: *UFD*

(1) Every non-zero element of R is either unit or a product of finitely many irreducibles.

(2) If $a_1 \cdots a_m = b_1 \cdots b_n$, where the a_i, b_j are irreducibles, then $n = m$ and after reordering of factors, a_i and b_i are associates for $1 \leq i \leq n$.

Note. The product of an irreducible element and a unit is irreducible. Indeed, let $u \in R^*$ and p be an irreducible. Check that up is not a unit (otherwise p is a unit since $p = u^{-1}(up)$) and that if $up = ab$ then in $p = (u^{-1}a)b$, $u^{-1}a$ or b is a unit (since p is irreducible) and therefore a or b is a unit. Hence up is irreducible.

Example 2.12 (Examples of (non) UFD's).

1. The $\mathbb{Z}$, by the Fundamental Theorem of Arithmetic.

2. The $\mathbb{C}[x]$. Every polynomial is uniquely written as a product of linear factors, up to order and multiplication by non-zero numbers. For example $x^2 + 1 = (x - i)(x + i) = 2(x + i)\frac{1}{2}(x - i)$.

3. The integral domain $\mathbb{Z}[\sqrt{-3}] = \left\{a + b\sqrt{-3} \mid a, b \in \mathbb{Z}\right\}$ is *not* a UFD. Indeed, $4 = 2 \times 2 = (1 + \sqrt{-3})(1 - \sqrt{-3})$. Recall that $\mathbb{Z}[\sqrt{-3}]^* = \{\pm 1\}$. The elements 2 and $1 + \sqrt{-3}$, $1 - \sqrt{-3}$ are irreducible elements in $\mathbb{Z}[\sqrt{-3}]$ since

$$\begin{aligned} 1 + \sqrt{-3} &= (\alpha + \beta\sqrt{-3})(\gamma + \delta\sqrt{-3}) \\ 4 &= (\alpha^2 + 3\beta^2)(\gamma^2 + 3\delta^2) \end{aligned}$$

implies that either $\alpha^2 + 3\beta^2 = 1$ or $\gamma^2 + 3\delta^2 = 1$ and hence $\alpha + \beta\sqrt{-3}$ or $\gamma + \delta\sqrt{-3}$ is a unit.

Also 2 is not associate of $1 \pm \sqrt{-3}$. Hence $\mathbb{Z}[\sqrt{-3}]$ does not have unique factorization.

properly divides

Definition 2.13. An element a *properly divides* b if $a|b$ and a and b are not associates.

Proposition 2.14. Let R be a UFD. Then there is no infinite sequence of elements $r_1, r_2, \ldots$ of R such that r_{n+1} properly divides r_n for each $n \geq 1$.

Proof. Write $r_1 = a_1 \cdots a_m$, where $a_1, \ldots, a_m$ are irreducibles (possible since R is a UFD). The number of factors m does not depend on the factorization, m only depends on r_1. Write $m = l(r_1)$. If r_2 properly divides r_1, then $l(r_2) < l(r_1)$. Hence $l(r_1) > l(r_2) \cdots$ and so on. This cannot go forever. Hence no infinite sequence $r_1, r_2, \ldots$ exists. ∎

Example 2.15 (Example of a non-UFD). Let

$$R = \{a_0 + a_1 x + \cdots + a_n x^n \mid a_0 \in \mathbb{Z},\ a_i \in \mathbb{Q} \text{ for } i \geq 1\}.$$

Clearly $R \subset \mathbb{Q}[x]$ and R is a subring of $\mathbb{Q}[x]$ and also an integral domain. Consider $r_1 = x, r_2 = \frac{1}{2}x, r_3 = \frac{1}{4}x, \cdots \in R$ and so $r_n = 2r_{n+1}$ but $\frac{1}{2} \notin R$ and hence $2 \notin R^*$ and $x \notin R^*$ since $\frac{1}{x} \notin \mathbb{Q}[x]$. Thus r_{n+1} properly divides r_n. By the previous proposition 2.14, R is not a UFD.

Proposition 2.16. Let R be a UFD. If p is irreducible and $p|ab$ then $p|a$ or $p|b$.

Proof. If a is a unit, then $p|b$ (since $p|ab$ implies $ab = pc$ and then $b = pca^{-1}$ for some $c \in R$). So assume that a, b are not units. Then $a = a_1 \cdots a_m$, $b = b_1, \cdots b_n$ for some irreducible elements a_i and b_j. Write $a_1 \cdots a_m \cdots b_1 \cdots b_n = pc$ for some $c \in R$. If $c \in R^*$, write $(c^{-1}a_1)a_2 \cdots a_m b_1 \cdots b_n = p$. Otherwise $c = c_1 \cdots c_s$ for some irreducibles $c_1, \dots, c_s \in R$. Then we have two ways of writing ab as a product of irreducibles

$$a_1 \cdots a_m b_1 \cdots b_n = pc_1 \cdots c_s.$$

Thus p is associated with some a_i or b_j, hence $p|a$ or $p|b$. ∎

Example 2.17. Let $R = \mathbb{Z}[\sqrt{d}]$, $d < -1$ and odd. Then $\mathbb{Z}[\sqrt{d}]$ is not a UFD. Note that 2 is irreducible (the same proof as before). Also

$$1 - d = (1 - \sqrt{d})(1 + \sqrt{d})$$

and $(1 - d)$ is even. But $2 \not| 1 \pm \sqrt{d}$ (recall that if $a \in \mathbb{Z}$, $a|\alpha + \beta\sqrt{d}$ then $a|\alpha$, $a|\beta$). Then 2.16 says that if R is a UFD and irreducible p divides ab, then $p|a$ or $p|b$. Therefore R is not a UFD.

Theorem 2.18. Let R be an integral domain. Then R is a UFD if and only if the following hold:

(1) There is no infinite sequence $r_1, r_2, \dots$ of elements of R such that r_{n+1} properly divides r_n for all $n \geq 1$.

(2) For every irreducible element $p \in R$, if $p|ab$, then $p|a$ or $p|b$.

Proof. By Propositions 2.14 and 2.16, the condition (1) and (2) are satisfied for any UFD.

Conversely, suppose R satisfies (1) and (2). For contradiction, suppose that there is an element r_1 in R, not 0, not a unit, which cannot be written as a product of irreducibles. Note that r_1 is not irreducible, hence $r_1 = r_2 s_2$, for some $r_2, s_2 \in R$ which are not units. At least one of the factors cannot be written as a product of irreducibles, say r_2. For the same reason as before, we can write $r_2 = r_3 s_3$, with r_3, s_3 non-units in R. Continuing in this way, we obtain an infinite sequence $r_1, r_2, r_3, \dots$. Moreover, in this sequence, r_{n+1} properly divides r_n because s_{n+1} is never a unit. This contradicts condition (1). Hence every non-unit, non-zero element of R can be written as a product of irreducibles.

Now assume that $a_1 \cdots a_m = b_1 \cdots b_n$, where the a_i and b_j are irreducibles. Since $a_1 | b_1 b_2 \cdots b_n$, by (2) we see that a_1 divides b_j for some j. Reorder the b_j's so that $a_1 | b_1$. Thus $b_1 = a_1 u$ for some $u \in R$, $u \neq 0$. If u is not a unit, then b_1 cannot be irreducible. Therefore u is a unit and hence a_1 and b_1 are associates and we can write

$$\begin{aligned} a_1 a_2 \cdots a_m &= a_1 u b_2 \cdots b_n \\ a_2 \cdots a_m &= (ub_2) \cdots b_n \end{aligned}$$

by the cancellation property in R. Continue in this way until we get 1 in the left hand side or in the right hand side. To fix ideas, assume $m \geq n$, then we arrive at the situation when a product of $m - n$ irreducibles equals 1. This can never happen unless $m = n$. Hence $m = n$ and, possibly after reordering, a_i and b_i are associates for $i \geq 1$. ■

Chapter 3

Euclidean domains and principal ideal domains

Consider $\mathbb{Z}$. The absolute value, or modulus, of $n \in \mathbb{Z}$ is a non-negative number $|n|$. Given $a, b \in \mathbb{Z}$, $b \neq 0$, we can write $a = qb + r$. If $b > 0$, then $0 \leq r < b$. For general non-zero b, we can still write $a = qb + r$, where r is such that $|r| < |b|$.

Definition 3.1. An integral domain R is called a *Euclidean domain* if there exists a function $\varphi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ satisfying the following conditions: *Euclidean domain*

(1) for all non-zero $a, b \in R$, we have $\varphi(a) \leq \varphi(ab)$,

(2) given $a, b \in R$, $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$ and $r = 0$ or $\varphi(r) < \varphi(b)$.

Call the function φ a *norm*. *norm*

For example, $\mathbb{Z}$ with norm $\varphi(n) = |n|$ is an Euclidean domain.

Example 3.2 (of Euclidean domains). Let F be a field, $F[x]$ the ring of polynomials with coefficients in F. For $f(x) \in F[x]$, $f(x) \neq 0$, define $\varphi(f(x)) = \deg f(x)$. Clearly

$$\deg f(x) \leq \deg f(x)g(x).$$

If $g(x)$ is non-zero polynomial, then $f(x) = q(x)g(x) + r(x)$ for some $q(x), r(x) \in F[x]$, where either $r(x)$ is the zero polynomial, or $\deg r(x) < \deg g(x)$. For example if

$$\begin{aligned} f(x) &= x^4 + 5x^2 + 2x + 1, \\ g(x) &= x^2 - 3x + 1 \end{aligned}$$

then

$$\begin{aligned} q(x) &= x^2 + 3x + 13, \\ r(x) &= 38x - 12. \end{aligned}$$

Sketch of proof of (2) in definition of Euclidean domain: Let

$$\begin{aligned} f(x) &= a_n x^n + \cdots + a_0, \\ g(x) &= b_m x^m + \cdots + b_0 \end{aligned}$$

with $a_n \neq 0$, $b_m \neq 0$ so that $\deg f(x) = n$ and $\deg g(x) = m$. If $n < m$, then $q(x) = 0$ and $f(x) = r(x)$. If $n \geq m$, then write

$$f_1(x) = f(x) - b_m^{-1} a_n x^{n-m} g(x),$$

a polynomial of degree $\leq n-1$. By induction, $f_1(x) = q_1(x)g(x) + r(x)$ hence

$$f(x) = (b_m^{-1} a_n x^{n-m} + q_1(x))g(x) + r(x).$$

root

Definition 3.3. Let $f(x) \in F[x]$. Then $\alpha \in F$ is a *root* of $f(x)$ if $f(\alpha) = 0$.

Proposition 3.4. Element $\alpha \in F$ is a root of $f(x) \in F[x]$ if and only if $(x-\alpha)$ divides $f(x)$.

Proof. If $(x-\alpha)$ divides $f(x)$, then $f(x) = (x-\alpha)b(x)$, hence $f(\alpha) = (\alpha-\alpha)b(\alpha) = 0$. Conversely, suppose $f(\alpha) = 0$ and write $f(x) = q(x)(x-\alpha) + r(x)$. Clearly $\deg r(x) < \deg(x-\alpha) = 1$ and hence $\deg r(x) = 0$, i.e. $r(x) = r \in F$. This implies

$$0 = f(\alpha) = q(\alpha) \cdot 0 + r,$$

that is $r = 0$. ■

Theorem 3.5. Let $f(x) \in F[x]$, where F is a field and $\deg f(x) = n \geq 1$. Then $f(x)$ has at most n roots in F.

Proof. By induction on n. If $n = 1$, then $f(x) = ax + b$, $a \neq 0$, hence $f(x)$ has only one root, namely $\frac{-b}{a}$. Now suppose that the statement is true for all degrees up to $n-1$. If $f(x)$ has no roots in F, we are done. Otherwise, $f(x)$ has at least one root, say α. Write $f(x) = (x-\alpha)g(x)$ by proposition 3.4. By the induction assumption, $g(x)$ has at most $n-1$ roots. Finally, if β is a root of $f(x)$, i.e. $f(\beta) = 0$, then

$$0 = f(\beta) = (\beta - \alpha)g(\beta).$$

If $\beta - \alpha \neq 0$, then $g(\beta) = 0$ since F has no zero divisors. Thus $f(x)$ has at most $1 + (n-1) = n$ roots. ■

Example 3.6.

1. The polynomial $x^6 - 1 \in \mathbb{Q}[x]$ has only two roots in $\mathbb{Q}$, namely 1 and -1.
2. The polynomial $x^6 - 1 \in \mathbb{C}[x]$ has 6 roots in $\mathbb{C}$.
3. Let $\mathbb{Z}/8$ be the ring of residues modulo 8 and let $\mathbb{Z}/8[x]$ be the ring of polynomials with coefficients in $\mathbb{Z}/8$. Consider $x^2 - 1 \in \mathbb{Z}/8[x]$. The roots are $\alpha \in \mathbb{Z}/8$ such that $\alpha^2 = 1$. Observe that

$$\begin{aligned} \overline{1}^2 &= \overline{1}, \\ \overline{3}^2 &= \overline{1}, \\ \overline{5}^2 &= \overline{1}, \\ \overline{7}^2 &= \overline{1} \end{aligned}$$

since $n^2 \equiv 1 \bmod 8$ for any odd $n \in \mathbb{Z}$. Hence $x^2 - 1$ has 4 roots in $\mathbb{Z}/8$. In fact, this does not contradict 3.5 since $\mathbb{Z}/8$ is not a field because $\overline{2} \times \overline{4} = \overline{0}$.

Definition 3.7. Suppose $F \subset K$ are fields. An element $\alpha \in K$ is called *algebraic over F* if there exists a non-zero polynomial $f(x) \in F[x]$ such that $f(\alpha) = 0$. *algebraic over*

Example 3.8.

1. Numbers $\sqrt{2}, \sqrt[3]{3}, \sqrt{-1} \in \mathbb{C}$ are algebraic over $\mathbb{Q}$ with corresponding polynomials $x^2 - 2$, $x^3 - 3$, $x^2 + 1$.
2. Any $\alpha \in \mathbb{C}$ is algebraic over $\mathbb{R}$. Indeed, for $\alpha = a + bi$, consider
$$(t - \alpha)(t - \overline{\alpha}) = t^2 - 2at + (a^2 + b^2) \in \mathbb{R}[x],$$
with complex roots α and $\overline{\alpha}$.
3. Any $\alpha \in F$ is algebraic over F – consider the linear polynomial $t - \alpha$.
4. Numbers $e, \pi \in \mathbb{R}$ are *not* algebraic over $\mathbb{Q}$.

Proposition 3.9. Suppose $F \subset K$ are fields, $\alpha \in K$ is algebraic over F. Then

(1) there exists an irreducible polynomial $p(x) \in F[x]$ such that $p(\alpha) = 0$,

(2) if $f(x) \in F[x]$, $f(\alpha) = 0$, then $p(x) | f(x)$.

Proof.

(1) Take $p(x)$ to be a polynomial of *the least degree* such that $p(\alpha) = 0$. Suppose then $p(x) = a(x)b(x)$ where $a(x), b(x)$ are not units, i.e. $\deg a(x) \geq 1$, $\deg b(x) \geq 1$. Now $0 = a(\alpha)b(\alpha)$ and hence α is a root of polynomial of degree less than $\deg p(x)$, a contradiction. So $p(x)$ is irreducible.

(2) Write $f(x) = q(x)p(x) + r(x)$, where $p(x)$ is from part (1). If $r(x)$ is the zero polynomial, we are done. Otherwise, $\deg r(x) < \deg p(x)$. But $0 = f(\alpha) = q(\alpha)p(\alpha) + r(\alpha)$ implies $r(\alpha) = 0$. This contradicts the minimality of $\deg p(x)$. Hence $f(x) = q(x)p(x)$. ■

Recall that a polynomial $a_0 + a_1 x + \cdots a_n x^n$ is called *monic* if $a_n = 1$. *monic*

Corollary 3.10. If $F \subset K$ are fields, $\alpha \in K$ algebraic over F, then there exists a unique irreducible monic polynomial $p(x) \in F[x]$ such that $p(\alpha) = 0$.

Proof. Consider $p(x)$ defined as in Proposition 3.9 and divide it by its highest degree coefficient. Then $p(x)$ is irreducible, monic and $p(\alpha) = 0$. If $p_1(x)$ is another monic, irreducible polynomial with $p_1(\alpha) = 0$, then $\deg p(x) = \deg p_1(x)$. Then either p and p_1 coincide, or $p(x) - p_1(x)$ is a nonzero polynomial. If $p(x) - p_1(x)$ is a non-zero polynomial, it vanishes at α and $\deg(p(x) - p_1(x)) < \deg p(x)$, a contradiction. ■

Definition 3.11. The polynomial $p(x)$ from the Corollary 3.10 is called the *minimal polynomial* of α over F. *minimal polynomial*

Example 3.12 (of Euclidean domains).

1. **Claim**: The rings $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$ are Euclidean domains.

Proof. Define $\varphi(z) = z\bar{z}$, i.e. if $z = a + b\sqrt{d}$, then $\varphi(z) = a^2 - db^2$, where $d = -1$ or -2. Hence $\varphi(z)$ is a non-negative integer. We must check that

(1) $\varphi(\alpha) \leq \varphi(\alpha\beta)$, for $\beta \neq 0$ and
(2) for any $\alpha, \beta \in \mathbb{Z}[\sqrt{d}]$, $\beta \neq 0$, there exist $q, r \in \mathbb{Z}[\sqrt{d}]$ such that $\alpha = q\beta + r$ with $r = 0$ or $\varphi(r) < \varphi(\beta)$.

For (1), note that $\varphi(\alpha\beta) = \alpha\overline{\alpha}\beta\overline{\beta}$. Note that $\varphi(\beta) \in \mathbb{Z}$, $\varphi(\beta) \geq 0$ and $\varphi(\beta) = 0$ if $\beta = 0$. Hence $\varphi(\alpha\beta) \geq \varphi(\alpha)$.

For (2), we look for q and r such that $\alpha = q\beta + r$. We write this as $\frac{\alpha}{\beta} = q + \frac{r}{\beta}$. Idea is to define q as the *best possible integer approximation* to $\frac{\alpha}{\beta}$. Write

$$\frac{\alpha}{\beta} = \mu + \nu\sqrt{d}$$

for some $\mu, \nu \in Q$ (this is possible since $\mathbb{Q}[\sqrt{d}]$ is a field). Take $m \in \mathbb{Z}$ such that $|m - \mu| \leq \frac{1}{2}$, take $n \in \mathbb{Z}$ such that $|n - \nu| \leq \frac{1}{2}$. Define $q = m + n\sqrt{d}$ and let $r = \alpha - q\beta$. Then

$$\begin{aligned}
\varphi(r) &= \varphi(\alpha - q\beta) = \varphi(\beta)\varphi\left(\frac{\alpha}{\beta} - q\right) \\
&= \varphi(\beta)\left((\mu - m)^2 + (\nu - n)^2(-d)\right) \\
&= \varphi(\beta)\left(\frac{1}{4} + \frac{1}{4}(-d)\right) \leq \frac{3}{4}\varphi(\beta) < \varphi(\beta).
\end{aligned}$$

■

2. **Claim**: The rings $\mathbb{Z}[\sqrt{2}]$ and $\mathbb{Z}[\sqrt{3}]$ are Euclidean domains.

Proof. Note that *if* we define $\varphi(a + b\sqrt{d})$ as $a^2 - db^2$, φ is not a norm (since it can be negative). So we define

$$\varphi(a + b\sqrt{d}) = |a^2 - db^2|.$$

This is clearly a non-negative integer. Moreover, since d is not a square of an integer, $a^2 - db^2 \neq 0$ if $a \neq 0$ or $b \neq 0$. So $\varphi(\alpha) > 0$ if $\alpha \neq 0$.

The proof of (1) in the definition of Euclidean domain is the same as in the previous example.

For (2), following the same pattern, take $\alpha, \beta \in \mathbb{Z}[\sqrt{d}]$. Keep the same notation and define q and r as before, with $d = 2$ or 3. Then

$$\begin{aligned}
\varphi(r) &= \varphi(\alpha - q\beta) = \varphi(\beta)\varphi\left(\frac{\alpha}{\beta} - q\right) \\
&= \varphi(\beta)\left|(\mu - m)^2 - d(\nu - n)^2\right|.
\end{aligned}$$

Note that $|x^2 - y^2 d| \leq \max(x^2, y^2 d)$ and $d > 0$. Therefore

$$|(\mu - m)^2 - d(\nu - n)^2| \leq \max\left(\frac{1}{4}, \frac{d}{4}\right),$$

hence $\varphi(r) \leq \frac{3}{4}\varphi(\beta) < \varphi(\beta)$. ■

Definition 3.13. Let R be a commutative ring and $I \subset R$ be its subring. Then $I \subset R$ is called an *ideal* if for any $r \in R$ and $x \in I$ we have $rx \in I$. *ideal*

Example 3.14 (of ideals).

1. The ring $n\mathbb{Z}$ (multiples of a fixed integer n) is an ideal of $\mathbb{Z}$.
2. If R is any commutative ring and $a \in R$, then $aR \subset R$ is an ideal.
3. Let $R = \mathbb{Z}[x]$, the ring of polynomials with integer coefficients. Let I be the set of polynomials $a_0 + a_1 x + \cdots + a_n x^n$ such that a_0 is even. This is clearly an ideal, since for $(a_0 + \cdots + a_n x^n) \in I$,
$$(a_0 + \cdots + a_n x^n)(b_0 + \cdots + b_m x^m) = a_0 b_0 + \cdots$$
and $a_0 b_0$ is even for any $b_0 \in \mathbb{Z}$.
4. Let R be a field. **Claim**: Rings $\{0\}$ and R are the only ideals in the field R.

   Proof. Suppose $I \subset R$ is a nonzero ideal. Then there exists $x \in I$, $x \neq 0$. Since R is a field, $x^{-1} \in R$. But I is an ideal, so $1 = x^{-1}x \in I$. Let r be any element of R, then $r = r \cdot 1 \in I$. Hence $I = R$. ∎

Definition 3.15. An ideal of R of the form aR (the multiples of a given element $a \in R$) is called a *principal ideal*. An integral domain R is called a *principal ideal domain* (*PID*) if every ideal of R is principal. *principal ideal PID*

Example 3.16 (of principal ideals).

1. We claim that $\mathbb{Z}$ is a PID. We need to show that every ideal $I \subset \mathbb{Z}$ has the form $a\mathbb{Z}$. If $I \neq \{0\}$, choose $a \in I$, $a \neq 0$, such that $|a|$ is minimal among the elements of I. Then $a\mathbb{Z} \subset I$. Let $n \in I$. Write $n = qa + r$, where $r = 0$ or $|r| < |a|$. If $r \neq 0$, can write $r = n - qa$ and since $n, qa \in I$, so does r, $r \in I$. A contradiction since $|r| < |a|$. Thus $r = 0$ and therefore $I \subset a\mathbb{Z}$, so $I = a\mathbb{Z}$.
2. Let $R = \mathbb{Z}[x]$ and $I = \{a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Z}[x] \mid a_0 \text{ is even}\}$.
   Claim: I is not a principal ideal.

   Proof. For contradiction assume that there is $a(x) \in \mathbb{Z}[x]$, such that $I = a(x)\mathbb{Z}[x]$. Note that $2 \in I$. Then $2 = a(x)b(x)$ for some $b(x) \in \mathbb{Z}$. Then $a(x)$ and $b(x)$ are constant polynomials, i.e. $a(x) = a \in \mathbb{Z}$, a is even. Also note that $x \in I$. Hence $x = a \cdot c(x)$ for some $c(x) \in \mathbb{Z}[x]$. But all coefficients of $ac(x)$ are even, a contradiction. Hence no generator exists, i.e. I is not principal. ∎

Theorem 3.17. Every Euclidean domain is a PID.

Proof. Let R be a Euclidean domain with norm φ. Given a non-zero ideal I, we choose $a \in I$, $a \neq 0$, such that $\varphi(a)$ is the smallest possible. Let $n \in I$. Write $n = qa + r$ and either $r = 0$ or $\varphi(r) < \varphi(a)$. If $r \neq 0$, write $r = n - qa$. Since $n, qa \in I$, so does r, $r \in I$. A contradiction to the minimality of $\varphi(a)$. So $r = 0$ and thus $n = qa$. This proves that $I = aR$ is a principal ideal. ∎

Example 3.18 (of PID's).

1. $\mathbb{Z}, F[x]$ if F is a field, $\mathbb{Z}[i], \mathbb{Z}[\sqrt{-2}], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[\sqrt{3}]$.
2. There do exist PID's which are not Euclidean domains.
   (For example $\mathbb{Z}[(1+\sqrt{-19})/2]$; the proof that it is not ED is too technical).
3. If d is odd, $d < -1$, then $\mathbb{Z}[\sqrt{d}]$ is not a PID (and hence by 3.17 not an Euclidean domain). In fact, every PID is a UFD (see further). Hence this follows from example 2.17.

Proposition 3.19. Suppose R is a PID and $I_1 \subset I_2 \subset \cdots$ are ideals in R. Then eventually, $I_n = I_{n+1} = \cdots$ for some n (the sequence of ideals *stabilizes*).

Proof. Define
$$I = \bigcup_{n\geq 1} I_n.$$
This is a subset of R. We claim that I is an ideal. First, $I \subset R$ is a subring: given $x, y \in I$ we must show that $x+y$, $-x$, xy are in I. Any $x \in I$ belongs to some I_n. Similarly, any $y \in I$ is in some I_m. Suppose $n \geq m$. Then $I_m \subset I_n$. So $x, y \in I_n$ and thus $x+y, -x, xy \in I_n$. Therefore $x+y, -x, xy \in I$. Let $r \in R$ and $x \in I_n$. Then $rx \in I_n$ and therefore $rx \in I$; I is an ideal in R.
By assumption, $I = aR$ for some $a \in R$. Clearly, $a \in I$. Hence, for some $l \geq 1$, we have $a \in I_l$. But then $I = aR \subset I_l$. On the other hand, $I_l \subset I$, so $I = I_l$.
For any $i \geq 1$, we have $I = I_l \subset I_{l+1} \subset I$, therefore $I_l = I_{l+1} = \cdots = I$. ■

Example 3.20. Assume $R = \mathbb{Z}$. Then $60\mathbb{Z} \subset 30\mathbb{Z} \subset 15\mathbb{Z} \subset \cdots \subset \mathbb{Z}$.

Proposition 3.21. Suppose that R is a PID. Let $p \in R$ be an irreducible element, such that $p|ab$. Then $p|a$ or $p|b$.

Proof. We claim that the subring
$$I = aR + pR = \{ar_1 + pr_2 \mid r_1, r_2 \in R\}$$
is an ideal: if $r \in R$, then $r(ar_1 + pr_2) = a(rr_1) + p(rr_2) \in I$. Then $I = dR$ for some $d \in R$.
We have $p = a \cdot 0 + p \cdot 1 \in I$ and so can write $p = dr$ for some $r \in R$. Since p is irreducible, r or d is a unit in R.
If r is a unit, say $rr^{-1} = 1$ for some $r^{-1} \in R$, then $d = pr^{-1}$. But $a \in I$, so $a = dr_1$ for some $r_1 \in R$. Thus $a = dr_1 = p(r^{-1}r_1)$ so $p|a$.
If d is a unit, $I = dR$ contains $1 = dd^{-1}$, hence $I = R$. Therefore
$$1 = at + pu$$
for some $t, u \in R$. This implies that
$$b = abt + bpu.$$
By assumption, $p|ab$, thus $p|abt + bpu$, thus $p|b$. ■

Theorem 3.22. Every PID is a UFD.

Proof. We will apply Theorem 2.18 – we need to prove that there does not exist an infinite sequence $r_1, r_2, \ldots$ such that r_{n+1} properly divides r_n for $n = 1, 2, \ldots$ (second condition of 2.18 follows from 3.21). Indeed, if $r_1, r_2, \ldots$ is such a sequence, we can write $r_1 = r_2 s_2$ with s_2 not a unit. Similarly $r_2 = r_3 s_3$ and so on, $r_n = r_{n+1} s_{n+1}$. This implies that $r_n R \subset r_{n+1} R$ for $n = 1, 2, \ldots$. By Proposition 3.19 there exists $l \geq 1$ such that $r_l R = r_{l+1} R = r_{l+2} R = \cdots$. But then $r_{l+1} R \subset r_l R$ so $r_{l+1} = r_l t$ for some $t \in R$. Then $r_{l+1} | r_l$ and $r_l | r_{l+1}$. This contradicts the assumption that r_{l+1} properly divides r_l. Thus by Theorem 2.18, R is a UFD. ■

Corollary 3.23. If R is an Euclidean domain, then R is a PID and then R is a UFD.

Example 3.24.

1. These rings are UFD's: $\mathbb{Z}, F[x], \mathbb{Z}[i], \mathbb{Z}[\sqrt{-2}], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[\sqrt{3}]$.

2. Can prove that if R is a UFD, then so is $R[x]$, for example $\mathbb{Z}[x]$ is a UFD. But this is *not* a PID.

Applications

In number theory, Diophantine equations are very important, These are polynomial equations in $\mathbb{Z}$ or $\mathbb{Q}$. For example, $x^n + y^n = z^n$ has no solutions in positive integers for $n > 2^\star$.

* *This margin is too small for a complete proof of this statement.*

Example 3.25. Claim: The only solutions to $x^2 + 2 = y^3$ with x, y integers is $x = \pm 5$ and $y = 3$.

Proof. Write as $(x - \sqrt{2})(x + \sqrt{-2}) = y^3$. Work in the UFD $\mathbb{Z}[\sqrt{-2}]$. Let p be an irreducible common factor of $x - \sqrt{-2}$ and $x + \sqrt{-2}$. Then $p | (x + \sqrt{-2}) - (x - \sqrt{-2}) = 2\sqrt{-2} = -(\sqrt{-2})^3$. Note that $\sqrt{-2}$ is irreducible in $\mathbb{Z}[\sqrt{-2}]$. Indeed, for $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$,

$$\begin{aligned}
\sqrt{-2} &= (\alpha + \beta\sqrt{-2})(\gamma + \delta\sqrt{-2}) \\
-\sqrt{-2} &= (\alpha - \beta\sqrt{-2})(\gamma - \delta\sqrt{-2}) \\
2 &= (\alpha^2 + 2\beta^2)(\gamma^2 + 2\delta^2).
\end{aligned}$$

Say $\alpha^2 + 2\beta^2 = 1$, then $\alpha + \beta\sqrt{-2} = \pm 1$. Thus $p = \sqrt{-2}$ or $p = -\sqrt{-2}$ since $\mathbb{Z}[\sqrt{-2}]$ is a UFD (± 1 are the only units in $\mathbb{Z}[\sqrt{-2}]$). Then $\sqrt{-2} | x + \sqrt{-2}$ and so $\sqrt{-2} | x$ and thus $2 | x^2$. Thus x^2 is even and therefore x is even. Also $y^3 = x^2 + 2$ is even and thus y is even. Hence get $2 = y^3 - x^2$, a contradiction since the RHS is divisible by 4.
Hence $x + \sqrt{-2}$ and $x - \sqrt{-2}$ have no irreducible common factors. Therefore $(x + \sqrt{-2})(x - \sqrt{-2})$ is uniquely written as $y_1^3 \cdots y_n^3$, where y_i's are irreducible. Therefore $x + \sqrt{-2} = (a + b\sqrt{-2})^3$ for some $a, b \in \mathbb{Z}$. Solve

$$\begin{aligned}
x + \sqrt{-2} &= (a + b\sqrt{-2})^3 \\
&= a^3 - 6ab^2 + (3a^2 b - 2b^3)\sqrt{-2}.
\end{aligned}$$

Hence, equating the real and imaginary parts,

$$\begin{aligned} x &= a^3 - 6ab^2, \\ 1 &= b(3a^2 - 2b^2). \end{aligned}$$

Therefore $b = \pm 1$ and $3a^2 - 2 = \pm 1$, hence $a = \pm 1$. Also $3a^2 - 2b^2 = 1$, so $b = 1$. Substitute into $x = a^3 - 6ab^2$ to get $x = \pm 5$. Finally, $y^3 = x^2 + 2 = 27$ and so $y = 3$. Hence $x = \pm 5$ and $y = 3$ are the only solutions. ∎

Theorem 3.26 (Wilson's Theorem). If p is prime, then $(p-1)! \equiv -1 \bmod p$.

Proof. Since $\mathbb{Z}/p \setminus \{0\}$ is a group under multiplication, for $0 < a < p$, there exists a unique inverse element a' such that $aa' \equiv 1 \bmod p$. In case $a = a'$, we have $a^2 \equiv 1 \bmod p$ and hence $a = 1$ or $a = p-1$. Thus the set $\{2, 3, \ldots, p-2\}$ can be divided into $\frac{1}{2}(p-3)$ pairs a, a' with $aa' \equiv 1 \bmod p$. Hence

$$\begin{aligned} (p-1)! &= (p-1) \cdot 2 \cdot 3 \cdots (p-2) \\ &\equiv (p-1) \mod p \\ &\equiv -1 \mod p. \end{aligned}$$

∎

Theorem 3.27. Let p be an odd prime. Then p is a sum of two squares iff $p \equiv 1 \bmod 4$.

Proof.

- $\Rightarrow$ Clearly $a^2 \equiv 0 \mod 4$ or $a^2 \equiv 1 \mod 4$ for any $a \in \mathbb{Z}$. Therefore, for $a, b \in \mathbb{Z}$, $a^2 + b^2 = 0, 1, 2 \mod 4$. Hence an integer congruent to $3 \bmod 4$ is never sum of two squares. Since p is an odd prime, $p \equiv 1 \mod 4$.

- $\Leftarrow$ Choose p such that $p = 1 \mod 4$. Write $p = 1 + 4n$, $n \in \mathbb{Z}$. Then

$$\begin{aligned} (p-1)! &= (1 \cdot 2 \cdots 2n)\left((2n+1)(2n+2) \cdots 4n\right) \\ &= (1 \cdot 2 \cdots 2n)\left((p-2n) \cdots (p-1)\right). \end{aligned}$$

Therefore

$$\begin{aligned} (p-1)! &\equiv (1 \cdot 2 \cdots 2n)\left((p-2n) \cdots (p-1)\right) \mod p \\ &\equiv (1 \cdot 2 \cdots 2n)\left((-2n) \cdots (-1)\right) \mod p \\ &\equiv (1 \cdot 2 \cdots 2n)^2 (-1)^{2n} \mod p. \end{aligned}$$

By Wilson's theorem, $(p-1)! \equiv -1 \mod p$, therefore $-1 = x^2 \mod p$ for $x = (1 \cdot 2 \cdots 2n)(-1)^{2n}$. Thus $p | x^2 + 1$. Now since $\mathbb{Z}[\sqrt{-1}]$ is a UFD, $p | (x+i)(x-i)$. Note that $p \not| x+i$, $p \not| x-i$ since $p(a+bi) = pa + pbi$, but $pb \neq \pm 1$. Therefore p is not irreducible in $\mathbb{Z}[i]$ (by Theorem 2.18) and therefore there are $a, b, c, d \in \mathbb{Z}$, $a + bi, c + di$ not units, such that

$$\begin{aligned} p &= (a+bi)(c+di) \\ p^2 &= (a^2+b^2)(c^2+d^2). \end{aligned}$$

Hence $p^2 = p \cdot p$ and therefore $p = a^2 + b^2 = c^2 + d^2$. ∎

Chapter 4

Homomorphisms and factor rings

Definition 4.1. Let R and S be rings. A function $f : R \to S$ is called a *homomorphism* if $f(x+y) = f(x) + f(y)$ and $f(xy) = f(x)f(y)$ for all $x, y \in R$. *homomorphism*
A bijective homomorphism is called an *isomorphism*. *isomorphism*

Example 4.2 (of homomorphisms).

1. Let $f : \mathbb{Z} \to \mathbb{Z}/m$, $f(n) = \overline{n}$, the residue class of $n \bmod m$. Then f is a homomorphism.

2. Consider $f : \mathbb{Q}[x] \to \mathbb{R}$ defined by $p(x) \mapsto p(\alpha)$ for $\alpha \in \mathbb{R}$; the value of p at α. Clearly f is a homomorphism.

3. Let $F \subset K$ be fields. Then the map $f : F \to K$, $f(x) = x$, is a homomorphism.

Proposition 4.3. If $f : R \to S$ is a homomorphism, then $f(0) = 0$ and $f(-r) = -f(r)$ for any $r \in R$.

Proof. We have

$$\begin{aligned} f(0) &= f(0+0) = f(0) + f(0), \\ 0 &= f(0). \end{aligned}$$

Also

$$\begin{aligned} 0 = f(0) &= f(r-r) \\ &= f(r) + f(-r), \\ f(-r) &= -f(r). \end{aligned}$$

■

Observe the relationship between $\mathbb{Z}$ and $\mathbb{Q}$:

$$\mathbb{Q} = \left\{ \frac{n}{m} \mid m \neq 0,\ n, m \in \mathbb{Z} \right\}.$$

Generalize this construction:

Theorem 4.4. Let R be an integral domain. Then there exists a field F containing a subring $\tilde{R}$ isomorphic to R and every element in F has the form ab^{-1}, for some $a, b \in \tilde{R}$, $b \neq 0$.

Proof.

- Consider $\{(a,b) \mid a, b \in R, b \neq 0\}$. Define $(a,b) \sim (c,d)$ iff $ad = bc$. Check that $\sim$ is an equivalence relation: $(a,b) \sim (a,b)$ since $ab = ba$, $(a,b) \sim (c,d)$ then also $(c,d) \sim (a,b)$. Finally if $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$ then

$$\begin{aligned} ad &= bc, \\ acf &= a(de) = (ad)e = bce \\ af &= be \end{aligned}$$

  and so $(a,b) \sim (e,f)$. Denote the equivalence class of (a,b) by $\frac{a}{b}$. Let F be the set of all such equivalence classes.

- Define
$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd},$$
  that is, the equivalence class of the pair $(ad+bc, bd)$. Also define
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$
  Check that addition is well defined, that is for $(a,b) \sim (A,B)$ and $(c,d) \sim (C,D)$, we have $(ad+bc, bd) \sim (AD+BC, BD)$. We have
$$\begin{aligned} aB &= bA \\ adBD &= ADbd, \\ cD &= dC \\ bcBD &= BCbd. \end{aligned}$$
  Thus $(ad+bc)BD = (AD+BC)bd$. We leave to the reader to check that the multiplication is well defined.

- Check that F is a field. The class $\frac{0}{1}$ is the zero element, $\frac{1}{1}$ is the identity for multiplication.

- Define $\tilde{R} = \left\{\frac{r}{1} \mid r \in R\right\} \subset F$. Consider the map $R \to \tilde{R}$ with $r \mapsto \frac{r}{1}$. This is an isomorphism, since, for example
$$\begin{aligned} \frac{a}{1} + \frac{b}{1} &= \frac{a+b}{1}, \\ \frac{a}{1}\frac{b}{1} &= \frac{ab}{1}, \end{aligned}$$
  so $a + b \mapsto \frac{a}{1} + \frac{b}{1}$ and $a \cdot b \mapsto \frac{a}{1}\frac{b}{1}$. Also if $a \mapsto 0$ then $(a,1) \sim (0,1)$ iff $a \cdot 1 + 0 \cdot 1 = 0$. Hence the map is a bijection.

- All elements of F have the form $\frac{a}{b} = \frac{a}{1}\frac{1}{b}$. Also $\frac{b}{1} = \left(\frac{1}{b}\right)^{-1}$. Therefore, $\frac{a}{b} = \frac{a}{1}\left(\frac{b}{1}\right)^{-1}$. ∎

Definition 4.5. Call F from the proof of the previous theorem the *field of fractions* of R. We identify R and $\tilde{R}$ using the map $r \mapsto \frac{r}{1}$. *field of fractions*

Example 4.6 (of field of fractions).

Ring	Field of fractions
$\mathbb{Z}$	$\mathbb{Q}$
$\mathbb{Z}[\sqrt{d}]$	$\mathbb{Q}[\sqrt{d}] = \left\{x + y\sqrt{d} \mid x, y \in \mathbb{Q}\right\}$
$\mathbb{R}[x]$	the field of rational functions $\left\{\frac{f(x)}{g(x)} \mid f, g \in \mathbb{R}[x],\ g \neq 0\right\}$

Definition 4.7. Let I be an ideal of a ring R. Let $r \in R$. The *coset* of r is the set $I + r = \{r + x \mid x \in I\}$. *coset* $I + r$

Proposition 4.8. For any $r, s \in R$ we have $I + r \cap I + s = \emptyset$ or $I + r = I + s$. Also, $I + r = I + s$ if and only if $r - s \in I$.

Proof. Same as for group theory. ■

Let R/I be the set of cosets. R/I

Theorem 4.9. Define $+$ and $\cdot$ on R/I as follows:

- $(I + r) + (I + s) = I + (r + s)$,
- $(I + r)(I + s) = I + rs$.

Then R/I is a ring.

Proof. See M2P2 for the proof that R/I is a group under addition (note that a subring I is a normal subgroup of R).
Let us check that $\cdot$ is well defined, i.e. the result doesn't depend on the choice of r and s in their respective cosets. Indeed, if $I + r' = I + r$, $I + s' = I + s$, then we need to check that $I + r's' = I + rs$. We have $r' = r + x$, $s' = s + y$ for $x, y \in I$ and

$$r's' = (r + x)(s + y) = rs + xs + ry + xy.$$

Now $x, y \in I$ and therefore $xs, ry, xy \in I$, since I is an ideal. Therefore $r's' - rs \in I$ hence $I + r's' = I + rs$. All the axioms of a ring hold in R/I because they hold in R. ■

Call the ring R/I the *factor ring* (or *quotient ring*). *factor ring*

Definition 4.10. Let $\varphi : R \to S$ be a homomorphism of rings. Then the *kernel* of φ is *kernel* Ker

$$\operatorname{Ker} \varphi = \{r \in R \mid \varphi(r) = 0\}.$$

The *image* of φ is *image* Im

$$\operatorname{Im} \varphi = \{s \in S \mid s = \varphi(r) \text{ for } r \in R\}.$$

Theorem 4.11. For rings R, S and homomorphism $\varphi : R \to S$

(1) $\operatorname{Ker} \varphi$ is an ideal of R,

(2) $\operatorname{Im} \varphi$ is a subring of S,

(3) $\operatorname{Im} \varphi$ is naturally isomorphic to the factor ring $R/\operatorname{Ker} \varphi$.

Proof.

(1) By M2P2 $\operatorname{Ker}\varphi \subset R$ is a subgroup under addition. Let $x \in \operatorname{Ker}\varphi$, $r \in R$. Then we need to check that $rx \in \operatorname{Ker}\varphi$. Indeed,

$$\varphi(rx) = \varphi(r)\varphi(x) = \varphi(r) \cdot 0 = 0.$$

(2) By M2P2 it is enough to show that $\operatorname{Im}\varphi$ is closed under multiplication. Take any $r_1, r_2 \in R$. Then

$$\varphi(r_1)\varphi(r_2) = \varphi(r_1 r_2) \in \operatorname{Im}\varphi.$$

(3) M2P2 says that the groups under addition $R/\operatorname{Ker}\varphi$ and $\operatorname{Im}\varphi$ are isomorphic. The map is $\operatorname{Ker}\varphi + r \mapsto \varphi(r)$. So we only need to check that this map respects multiplication. Suppose $r_1, r_2 \in R$. Then $\operatorname{Ker}\varphi + r_1 \mapsto \varphi(r_1)$ and $\operatorname{Ker}\varphi + r_2 \mapsto \varphi(r_2)$. Also $\operatorname{Ker}\varphi + r_1 r_2 \mapsto \varphi(r_1 r_2)$. Now

$$(\operatorname{Ker}\varphi + r_1)(\operatorname{Ker}\varphi + r_2) = \operatorname{Ker}\varphi + r_1 r_2.$$

But since φ is a homomorphism, $\varphi(r_1 r_2) = \varphi(r_1)\varphi(r_2)$. Hence our map $R/\operatorname{Ker}\varphi \to \operatorname{Im}\varphi$ sends the product of $\operatorname{Ker}\varphi + r_1$ and $\operatorname{Ker}\varphi + r_2$ to $\varphi(r_1)\varphi(r_2)$, hence is a homomorphism of rings. Because the map is bijective, it is an isomorphism of rings. ■

Example 4.12.

1. Let $R = \mathbb{Z}$, $S = \mathbb{Z}/5$ and $\varphi : \mathbb{Z} \to \mathbb{Z}/5$, $\varphi(n) = \overline{n}$. We have $\operatorname{Im}\varphi = \mathbb{Z}/5$, $\operatorname{Ker}\varphi = 5\mathbb{Z} = \{5n \mid n \in \mathbb{Z}\}$. Then cosets are $5\mathbb{Z}, 1 + 5\mathbb{Z}, \ldots, 4 + 5\mathbb{Z}$. Clearly $\mathbb{Z}/\operatorname{Ker}\varphi = \operatorname{Im}\varphi$ since $\mathbb{Z}/5\mathbb{Z} = \mathbb{Z}/5$.

2. Let $R = \mathbb{Q}[x]$, $S = \mathbb{R}$ and $\varphi : \mathbb{Q}[x] \to \mathbb{R}$ defined as

$$\varphi(f(x)) = f(\sqrt{2}).$$

Then

$$\begin{aligned} \operatorname{Ker}\varphi &= \left\{ f(x) \mid f(\sqrt{2}) = 0 \right\} \\ &= \left\{ f(x) \text{ such that } x - \sqrt{2} \text{ divides } f(x) \right\}. \end{aligned}$$

If $a_0 + a_1\sqrt{2} + a_2(\sqrt{2})^2 + \cdots + a_n(\sqrt{2})^n = 0$ for $a_i \in \mathbb{Q}$, then $a_0 - a_1(-\sqrt{2}) + a_2(-\sqrt{2})^2 - \cdots + a_n(-\sqrt{2})^n = 0$. Hence

$$\begin{aligned} \operatorname{Ker}\varphi &= \left\{ (x^2 - 2)g(x) \mid g(x) \in \mathbb{Q}[x] \right\}, \\ \operatorname{Im}\varphi &= \mathbb{Q}(\sqrt{2}). \end{aligned}$$

Thus $\mathbb{Q}[x]/(x^2 - 2)\mathbb{Q}[x] = \mathbb{Q}(\sqrt{2})$.

maximal ideal **Definition 4.13.** Let I be an ideal in R. Then $I \subset R$ is a *maximal ideal* if $I \neq R$ and there is no ideal $J \subset R$, such that $I \subsetneq J$.

Example 4.14 (of maximal ideals).

1. We claim that $5\mathbb{Z} \subset \mathbb{Z}$ is a maximal ideal. If there is an ideal J such that $5\mathbb{Z} \subsetneq J \subset \mathbb{Z}$, then $J = \mathbb{Z}$: we show that $1 \in J$. Since $5\mathbb{Z} \subsetneq J$, there is $a \in J$ not divisible by 5. Hence a and 5 are coprime and $5n + am = 1$ for some $n, m \in \mathbb{Z}$. Hence $1 \in J$.

2. On the other hand, $6\mathbb{Z}$ is not a maximal ideal since $6\mathbb{Z} \subset 2\mathbb{Z} \subset \mathbb{Z}$ and also $6\mathbb{Z} \subset 3\mathbb{Z} \subset \mathbb{Z}$.

Theorem 4.15. Let R be a ring with 1 and let $I \subset R$ be an ideal. Then R/I is a field if and only if I is maximal.

Proof.

$\Rightarrow$ Assume that R/I is a field. Then $I \neq R$ (since $0 \neq 1$ in R/I). Assume there exists an ideal J such that $I \subsetneq J \subset R$. Choose $a \in J$, $a \notin I$. Then $I + a \in R/I$ is not the zero coset I. Since R/I is a field, every non-zero element is invertible, e.g. $I + a$ is invertible. Thus for some $b \in R$, we have

$$(I + a)(I + b) = I + ab = I + 1.$$

Therefore $ab - 1 \in I \subset J$ and thus $1 = ab + x$ for some $x \in J$. But $ab \in J$ since $a \in J$. Therefore $1 \in J$ and so $J = R$ and hence I is maximal.

$\Leftarrow$ Conversely, assume that $I \subset R$ is a maximal ideal. Any non-zero element of R/I can be written as $I + a$ with $a \notin I$. Consider

$$I + aR = \{x + ay \mid x \in I, y \in R\}.$$

This is an ideal. Indeed, for any $z \in R$, we have

$$z(x + ay) = \underset{\in I}{xz} + \underset{\in R}{ayz} \in I + aR.$$

Since I is maximal and $I \subset I + aR$, we must have $I + aR = R$, in particular $1 = x + ay$ for some $x \in I$, $y \in R$. We claim that $I + y$ is the inverse of $I + a$. Indeed,

$$\begin{aligned}(I + a)(I + y) &= I + ay \\ &= I + 1 - x = I + 1\end{aligned}$$

since $x \in I$. ∎

Proposition 4.16. Let R be a PID and $a \in R$, $a \neq 0$. Then aR is maximal if and only if a is irreducible.

Proof.

$\Rightarrow$ Assume that $aR \subset R$ is a maximal ideal. Since $aR \neq R$, a is not a unit. Thus either a is irreducible or $a = bc$ for $b, c \in R$ not units. Then $aR \subset bR \subsetneq R$ since b is not a unit. Since aR is maximal we have $aR = bR$ and so $b = am$ for $m \in R$. Therefore a and b are associates and $b = am = bcm$ and so $1 = cm$, hence c is a unit; contradiction. Therefore a is irreducible.

$\Leftarrow$ Now assume that a is irreducible. In particular, a is not a unit, so $aR \neq R$. Assume that there exists an ideal J such that $aR \subsetneq J \subsetneq R$. Since R is a PID, $J = bR$ for some $b \in R$. Since $aR \subset bR$, $a \in bR$ and we can write $a = bc$ for some $c \in R$. Have that b is not a unit because $bR \neq R$. Also c is not a unit because otherwise $aR = bR$: if c is a unit then $c^{-1} \in R$ and so $b = c^{-1}a \in aR$, hence $bR \subset aR$. Thus a is not irreducible; a contradiction. Therefore aR is maximal. ■

Corollary 4.17. If R is a PID and $a \in R$ is irreducible, then R/aR is a field.

Example 4.18.

1. For a PID $R = \mathbb{Z}[i]$, $a = 2 + i$ is irreducible. Hence $\mathbb{Z}[i]/(2 + i)\mathbb{Z}[i]$ is a field.
2. For $R = \mathbb{Q}[x]$, $a = x^2 - 2$ is irreducible. Hence $\mathbb{Q}[x]/(x^2 - 2)\mathbb{Q}[x]$ is the field $\mathbb{Q}(\sqrt{2})$.

Proposition 4.19. Let F be a field, $p(x) \in F[x]$ an irreducible polynomial and $I = p(x)F[x]$. Then $F[x]/I$ is a field. If $\deg p(x) = n$, then

$$F[x]/I = \left\{I + a_0 + a_1x + \cdots + a_{n-1}x^{n-1},\ a_i \in F\right\}.$$

Proof. Corollary 4.17 implies that $F[x]/I$ is a field. For all $f(x) \in F[x]$, there exist $q(x), r(x) \in F[x]$ such that $f(x) = q(x)p(x) + r(x)$, $r(x) = 0$ or $\deg r(x) < n$. Hence $I + f(x) = I + r(x)$. ■

Suppose $F \subset K$ are fields. Recall that $\alpha \in K$ is algebraic over F if $f(\alpha) = 0$ for some $f(x) \in F[x]$. The minimal polynomial of α is the unique monic polynomial $p(x)$ of the least degree such that $p(\alpha) = 0$. Also recall that $F(\alpha)$ denotes the smallest subfield of K containing F and α.

Proposition 4.20. Let $F \subset K$ be fields, $\alpha \in K$ algebraic over F with minimal polynomial $p(x)$ and $\deg p(x) = n$. Let $I = p(x)F[x]$. Then $F[x]/I = F(\alpha)$ and every element of $F(\alpha)$ is uniquely written as $a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$ for some $a_i \in F$.

Proof. Consider the homomorphism $\theta : F[x] \to F(\alpha)$ defined by $f(x) \mapsto f(\alpha)$. Then

$$\begin{aligned} \operatorname{Ker}\theta &= \{f(x) \in F[x] \mid f(\alpha) = 0\} \\ &= p(x)F[x]. \end{aligned}$$

Theorem 4.11 says that $\operatorname{Im}\theta = F[x]/p(x)F[x]$. Then $\operatorname{Im}\theta$ is a field since $p(x)$ is irreducible. Proposition 4.19 implies that

$$\operatorname{Im}\theta = \left\{p(x)F[x] + a_0 + \cdots + a_{n-1}x^{n-1}\right\}.$$

Observe that $\operatorname{Im}\theta \subset K$, $\operatorname{Im}\theta$ is a subfield, $\alpha \in \operatorname{Im}\theta$ and $F \subset \operatorname{Im}\theta$ (since $x \mapsto \alpha$, $a \mapsto a$ for $a \in F$). Therefore $F(\alpha) \subset \operatorname{Im}\theta$. Clearly $\operatorname{Im}\theta \subset F(\alpha)$. Thus $\operatorname{Im}\theta = F(\alpha)$. By Proposition 4.19 every element of $\operatorname{Im}\theta = F(\alpha)$ can be written as $a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$. Now we have to prove the uniqueness. If for $a_i \in F$

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$$

then

$$(b_{n-1} - a_{n-1})\alpha^{n-1} + \cdots + (b_0 - a_0) = 0,$$

so α is a root of $q(x) = (b_{n-1} - a_{n-1})x^{n-1} + \cdots + (b_0 - a_0) \in F[x]$. Since n is the degree of the minimal polynomial of α, this is the zero polynomial, therefore $a_i = b_i$ for $i = 0, 1, \ldots, n-1$. ■

Example 4.21.

1. Consider $\mathbb{Q} \subset \mathbb{R}$, $\alpha = \sqrt{2}$, $p(x) = x^2 - 2$. Then by 4.20

$$\mathbb{Q}(\sqrt{2}) = \mathbb{Q}[x]/(x^2 - 2)\mathbb{Q}[x] = \left\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\right\}.$$

2. Consider $\mathbb{Q} \subset \mathbb{C}$, $\alpha = \sqrt{d}$, $d \in \mathbb{Q}$ is not a square, $p(x) = x^2 - d$. Then

$$\mathbb{Q}(\sqrt{d}) = \mathbb{Q}[x]/(x^2 - d)\mathbb{Q}[x] = \left\{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\right\}.$$

3. Consider $\mathbb{R} \subset \mathbb{C}$, $\alpha = \sqrt{-1}$, $p(x) = x^2 + 1$. Then

$$\mathbb{R}(i) = \mathbb{R}[x]/(x^2 + 1)\mathbb{R}[x] = \{a + bi \mid a, b \in \mathbb{R}\} = \mathbb{C}.$$

4. Consider $\mathbb{Q} \subset \mathbb{C}$, $\alpha = e^{\frac{2\pi i}{5}}$, clearly α is a root of $x^5 - 1$. But 1 is also root of $x^5 - 1$, so it is not irreducible (and hence not minimal). So $x - 1 | x^5 - 1$; divide to get
$$x^5 - 1 = (x - 1)(x^4 + x^3 + x^2 + x + 1).$$
In fact, $x^4 + x^3 + x^2 + x + 1$ is irreducible (we will prove this later), monic, has α as a root and therefore is minimal. Thus
$$\mathbb{Q}(\alpha) = \left\{a_0 + a_1\alpha + \cdots + a_3\alpha^3 \mid a_i \in \mathbb{Q}\right\}.$$

Proposition 4.22. A polynomial $f(x) \in F[x]$ of degree 2 or 3 is irreducible if and only if it has no roots in F.

Proof.

$\Leftarrow$ If $f(x)$ is not irreducible, then $f(x) = a(x)b(x)$ with $\deg f(x) = \deg a(x) + \deg b(x)$ and $\deg a(x), \deg b(x) \geq 1$ (units in $F[x]$ are polynomials of degree 0). Hence $\deg a(x) = 1$ or $\deg b(x) = 1$. Thus a linear polynomial, say $x - \alpha$ divides $f(x)$, so that $f(\alpha) = 0$ for some $\alpha \in F$.

$\Rightarrow$ The only if part follows from the Proposition 3.4 (if $f(x)$ has a root α then it is divisible by non-unit $(x - \alpha)$ and so is not irreducible). ■

Proposition 4.23. There exists a field with 4 elements.

Note. It is *not* $\mathbb{Z}/4$ since it is not a field.

Proof. Start from $\mathbb{Z}/2$. Consider $x^2 + x + 1 \in \mathbb{Z}/2[x]$. This is an irreducible polynomial (check for $x = \overline{0}, \overline{1}$). Consider $\mathbb{Z}/2[x]/(x^2 + x + 1)\mathbb{Z}/2[x]$. This is a field since $x^2 + x + 1$ is irreducible. Also Proposition 4.19 says that *all* the cosets are: $I = (x^2+x+1)\mathbb{Z}/2[x], 1+I, x+I, 1+x+I$. Thus the field has exactly 4 elements. ■

The explicit structure of the field with 4 elements is: Use notation $0 := I$, $1 := 1+I$, $\omega := x+I$. Then the elements of the field are $\{0, 1, \omega, \omega+1\}$. The addition table is:

	1	ω	$\omega+1$
1	0	$\omega+1$	ω
ω	$\omega+1$	0	1
$\omega+1$	ω	1	0

Observe that $\omega^2 = \omega + 1$. Indeed, x^2 and $x+1$ are in the same coset because $x^2 - (x+1) = x^2 + x + 1 \in I$ (we work in $\mathbb{Z}/2$). Since $x^2 + x + 1 \in I$, we also have $(x+1)(x^2+x+1) \in I$. This gives

$$x^3 + 2x^2 + 2x + 1 = x^3 + 1 \in I.$$

Therefore x^3 and 1 are in the same coset and hence $\omega^3 = 1$. The multiplication table is:

	1	ω	$\omega^2 = \omega+1$
1	1	ω	ω^2
ω	ω	$1+\omega$	1
$\omega+1$	$\omega+1$	1	ω

In particular, $\omega^{-1} = 1+\omega$, $(1+\omega)^{-1} = \omega$.

Example 4.24.

1. Prove that $x^3 + x + 1$ is irreducible over $\mathbb{Z}/2$. Hence construct a field of 8 elements.
2. Prove that x^2+1 is irreducible over $\mathbb{Z}/3$. Hence construct a field of 9 elements.

Theorem 4.25 (Gauss's Lemma). Let $f(x)$ be a polynomial with integer coefficients of degree at least 1. If $f(x)$ is irreducible in $\mathbb{Z}[x]$, then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Note. This is equivalent to the following statement: if $f(x) = h(x)g(x)$, $h(x), g(x) \in \mathbb{Q}[x]$ of degree at least 1, then $f(x) = a(x)b(x)$ for some $a(x), b(x) \in \mathbb{Z}[x]$ of degree at least 1.

Proof. Suppose $f(x) = h(x)g(x)$, $h(x), g(x) \in \mathbb{Q}[x]$. Let n be an integer such that $nf(x) = \tilde{h}(x)\tilde{g}(x)$ for some $\tilde{h}(x), \tilde{g}(x) \in \mathbb{Z}[x]$. If $n \neq 1$, there exists a prime p that divides n. Let us reduce all the coefficients mod p. Call $h'(x)$ and $g'(x)$ the resulting polynomials with coefficients in $\mathbb{Z}/p$. Since p divides all coefficients of $nf(x)$, we get $0 = h'(x)g'(x)$. Recall that $\mathbb{Z}/p[x]$ is an integral domain, so that one of $h'(x), g'(x)$, say $h'(x)$, is the zero polynomial. Then p divides every coefficient of $\tilde{h}(x)$. Divide both sides by p. Then

$$\frac{n}{p} f(x) = \frac{1}{p}\tilde{h}(x)\tilde{g}(x),$$

where $\frac{n}{p} \in \mathbb{Z}$, $\frac{1}{p}\tilde{h}(x), g(x) \in \mathbb{Z}[x]$. Carry on repeating this argument until $f(x)$ is factorized into a product of 2 polynomials with integer coefficients (the degrees of factors don't change and neither factor is a constant). ■

Note. It follows from the Gauss's Lemma that if $f(x)$ has integer coefficients and is monic and can be written $f(x) = g(x)h(x)$ where $g(x), h(x) \in \mathbb{Q}[x]$ and $g(x)$ is monic, then in fact $g(x), h(x) \in \mathbb{Z}[x]$.

Example 4.26. Let $f(x) = x^3 - nx - 1$, where $n \in \mathbb{Z}$. For which values of n is $f(x)$ irreducible over $\mathbb{Q}[x]$? If $f(x)$ is reducible over $\mathbb{Q}[x]$, then $f(x) = (x^2 + ax + b)(x + c)$ for $a, b, c \in \mathbb{Z}$. Hence $f(x)$ has an *integer* root $-c$. Since $bc = -1$, $c = \pm 1$. If $x = 1$ is a root, then $n = 0$ and if $x = -1$ is a root, then $n = 2$. For all other values of n, $f(x)$ is irreducible over $\mathbb{Q}[x]$.

Theorem 4.27 (Eisenstein's irreducibility cirterion). Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$, $a_i \in \mathbb{Z}$ for all $i \in \{0, 1, \ldots, n\}$. If a prime p does not divide a_n, but p divides $a_{n-1}, \ldots, a_1, a_0$ and p^2 does not divide a_0, then $f(x)$ is irreducible over $\mathbb{Q}$.

Proof. If $f(x)$ is reducible over $\mathbb{Q}$, then by the Gauss's Lemma, $f(x) = g(x)h(x)$, $g(x), h(x) \in \mathbb{Z}[x]$. Let $\overline{f}(x), \overline{g}(x), \overline{h}(x)$ be polynomials with coefficients in $\mathbb{Z}/p$ obtained by reducing coefficients of $f(x), g(x), h(x)$ modulo p. By the condition of the theorem we have

$$\overline{f}(x) = \overline{a}_n x^n = \overline{h}(x)\overline{g}(x).$$

Therefore $\overline{h}(x) = \alpha x^s$, $\overline{g}(x) = \beta x^t$ for some $\alpha, \beta \in \mathbb{Z}/p$, $\alpha, \beta \neq 0$ and $s + t = n$. Then p divides all coefficients of $h(x)$ and $g(x)$ except their leading terms. In particular, p divides the constant terms of $h(x)$ and $g(x)$, therefore p^2 divides a_0; a contradiction. Hence the initial assumption that $f(x)$ is reducible is false; $f(x)$ is irreducible. ∎

Example 4.28. Polynomial $x^7 - 2$ is irreducible in $\mathbb{Q}$ (choose $p = 2$ in the criterion) and the polynomial $x^7 - 3x^4 + 12$ is also irreducible in $\mathbb{Q}$ (choose $p = 3$).

Example 4.29. Claim: Let p be prime. Then $1 + x + \cdots + x^{p-1} \in \mathbb{Q}[x]$ is irreducible.

Proof. Observe that $f(x) = 1 + x + \cdots + x^{p-1}$ is $\frac{1-x^p}{1-x}$. Let $x = y + 1$. Then

$$\begin{aligned} f(x) = \frac{x^p - 1}{x - 1} &= \frac{(y+1)^p - 1}{y} \\ &= y^{p-1} + \binom{p}{1} y^{p-2} + \cdots + \binom{p}{p-1} \\ &= g(y). \end{aligned}$$

Now p does not divide 1 and divides $\binom{p}{k}$. Also p^2 does not divide $\binom{p}{p-1} = p$ and hence by the Eisenstein's criterion, $g(y)$ is irreducible and so is $f(x)$ (if $f(x) = f_1(x)f_2(x)$ for some $f_1(x), f_2(x) \in \mathbb{Q}[x]$, then $g(y) = g_1(y)g_2(y)$ for $g_1(y) = f_1(y+1), g_2 = f_2(y+1) \in \mathbb{Q}[x]$; a contradiction). ∎

Chapter 5

Field extensions

Definition 5.1. If $F \subset K$ are fields, then K is an *extension* of F. *extension*

Example 5.2. Fields $\mathbb{R}$ and $\mathbb{Q}(\sqrt{2})$ are extensions of $\mathbb{Q}$.

Proposition 5.3. If K is an extension of a field F, then K is a vector space over F.

Proof. Recall that a vector space is an abelian group under addition where we can multiply elements by the elements of F. The axioms of a vector field are: for all $\lambda, \mu \in F$, $v_1, v_2 \in K$,

(1) $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$,

(2) $(\lambda + \mu)v_1 = \lambda v_1 + \mu v_1$,

(3) $\lambda\mu v_1 = \lambda(\mu v_1)$,

(4) $1v_1 = v_1$.

All of these clearly hold. ∎

Definition 5.4. Let K be an extension of F. The *degree of K over F* is $\dim_F(K)$. Denote this by $[K : F]$. If $[K : F]$ is finite, we call K a *finite extension over F*. *degree* $[K : F]$ *finite extension*

Example 5.5.

1. Let $F = \mathbb{R}$, $K = \mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$, so $\dim_{\mathbb{R}}(\mathbb{C}) = 2$ and $\{1, i\}$ is a basis of $\mathbb{C}$. So $[\mathbb{C} : \mathbb{R}] = 2$.

2. Let $F = \mathbb{Q}$, $K = \mathbb{Q}(\sqrt{d})$ (d not a square). Then $[\mathbb{Q}(\sqrt{d}) : \mathbb{Q}] = 2$ since $\{1, \sqrt{d}\}$ is clearly a basis of $\mathbb{Q}(\sqrt{d})$.

3. Find $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}]$. We claim that $\{1, \sqrt[3]{2}, (\sqrt[3]{2})^2\}$ is a basis of $\mathbb{Q}(\sqrt[3]{2})$. Indeed, since otherwise these three elements are linearly dependent (they clearly span $\mathbb{Q}(\sqrt[3]{2})$), i.e. we can find $b_0, b_1, b_2 \in \mathbb{Q}$ not all zero, such that

$$b_0 + b_1\sqrt[3]{2} + b_2(\sqrt[3]{2})^2 = 0.$$

But the minimal polynomial of $\sqrt[3]{2}$ is $x^3 - 2$ because it is irreducible over $\mathbb{Q}$ (e.g. by the Eisenstein Criterion). Therefore $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$.

Theorem 5.6. Let $F \subset K$ be a field extension, $\alpha \in K$. The minimal polynomial of α has degree n iff $[F(\alpha) : F] = n$.

Proof.

- $\Rightarrow$ Suppose the degree of minimal polynomial of α is n. We know that

$$F(\alpha) = \left\{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} \mid a_i \in F\right\}.$$

  Hence $1, \alpha, \ldots, \alpha^{n-1}$ span $F(\alpha)$. Let us show that $1, \alpha, \ldots, \alpha^{n-1}$ are linearly independent: If not, there are $b_0, \ldots, b_{n-1} \in F$ such that $b_1 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1} = 0$ and not *all* $b_i = 0$. But then α is a root of the non-zero polynomial $b_0 + b_1x + \cdots + b_{n-1}x^{n-1}$; this contradicts our assumption.

- $\Leftarrow$ Suppose $[F(\alpha) : F] = n$. The elements $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}, \alpha^n \in F(\alpha)$ are $n+1$ vectors in a vector space of dimension n. Hence there exist $a_i \in F$, $i = 0, \ldots, n$ (not all $a_i = 0$), such that $a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$. Therefore α is algebraic over F. Thus α has a minimal polynomial, say of degree m. By the proof of $\Rightarrow$, $m = [F(\alpha) : F]$, so $m = n$. ■

Example 5.7.

1. $x^2 + 1$ is the minimal polynomial of i over $\mathbb{R}$ and $[\mathbb{C} : \mathbb{R}] = 2$.
2. $x^2 - 2$ is the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ and $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$.
3. $x^3 - 2$ is the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ and $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$.
4. $x^2 + x + 1$ is the minimal polynomial of ω over $\mathbb{Z}/2$ and $[\mathbb{Z}/2(\omega) : \mathbb{Z}/2] = 2$.

Example 5.8. Let $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ be the smallest subfield of $\mathbb{R}$ containing $\mathbb{Q}$, $\sqrt{2}$ and $\sqrt{3}$. We have $(\mathbb{Q}(\sqrt{2}))(\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$: By previous results

$$(\mathbb{Q}(\sqrt{2}))(\sqrt{3}) = \left\{\alpha + \beta\sqrt{3} \mid \alpha, \beta \in \mathbb{Q}(\sqrt{2})\right\}$$

because $x^2 - 3$ is the minimal polynomial of $\sqrt{3}$ over $\mathbb{Q}(\sqrt{2})$. Also

$$(\mathbb{Q}(\sqrt{2}))(\sqrt{3}) = \left\{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \mid a, b, c, d \in \mathbb{Q}\right\}.$$

Theorem 5.9. Let $F \subset K \subset E$ be fields. Then $[E : F] = [E : K][K : F]$.

Proof. Assume $[K : F] < \infty$, $[E : K] < \infty$. Let $e_1, \ldots, e_m$ be a basis of E over K and $k_1, \ldots, k_n$ be a basis of K over F. Then we claim that e_ik_j for all $1 \le i \le m$, $1 \le j \le n$, form a basis of E over F. Any element of E can be written as $\sum_{i=1}^m a_ie_i$ for some $a_i \in K$. Write $a_i = \sum_{j=1}^n b_{ij}k_j$, $b_{ij} \in F$. Thus $\sum_{i=1}^m a_ie_i = \sum_{i,j} b_{ij}e_ik_j$ and hence e_ik_j span E. If e_ik_j are not linearly independent, then for some $\alpha_{ij} \in F$, not all zero, we have $\sum_{i,j} \alpha_{ij}e_ik_j = 0$. Then

$$\sum_{i=1}^m \underbrace{\left(\sum_{j=1}^n \alpha_{ij}k_j\right)}_{\in K} e_i = 0.$$

Since $e_1, \dots, e_m$ is a basis, we must have $\sum_{j=1}^n \alpha_{ij} k_j = 0$ for every $i = 1, \dots, m$. Since $k_1, \dots, k_m$ is a basis of K over F we must have $\alpha_{ij} = 0$ for all i and j. Hence $\{e_i k_j\}$ form a basis of E over F and thus

$$[E : F] = \dim_F E = mn = [E : K][K : F].$$

If E is not a finite extension of F, then either K is not a finite dimensional vector space over F or E is not a finite dimensional vector space over K: We actually showed that if $[E : K] < \infty$ and $[K : F] < \infty$, then $[E : F] < \infty$. If $[E : F] = \dim_F E < \infty$, then $[K : F] < \infty$ because K is a subspace of E. If $[E : F] < \infty$ then E is spanned by finitely many elements over F. The same elements span E over K, hence $[E : K] < \infty$. ■

Corollary 5.10. If $F \subset K \subset E$ are fields and $[E : F] < \infty$, then $[K : F]$ divides $[E : F]$ and $[E : K]$ divides $[E : F]$.

Definition 5.11. The smallest positive integer n such that

$$\underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = 0$$

is called the *characteristic* of the field F. If there is no such n, then F has characteristic 0. Denote the characteristic of F by $\operatorname{char} F$. *characteristic* $\operatorname{char}(F)$

Note. For $a \in F$ and $n \in \mathbb{N}$, we denote by $(n \times a)$ the sum

$$(n \times a) = \underbrace{a + a + \cdots + a}_{n \text{ times}}.$$

Example 5.12. We have $\operatorname{char}(\mathbb{Q}) = 0$ and $\operatorname{char}(\mathbb{Z}/p) = p$ (with p prime).

Proposition 5.13. Let F be a field. Then (with p a prime number)

(1) $\operatorname{char}(F) = 0$ or $\operatorname{char}(F) = p$,

(2) if $\operatorname{char}(F) = 0$, then if $x \in F$, $x \neq 0$, then $(k \times x)$ for $k \in \mathbb{N} \setminus \{0\}$ is never zero,

(3) if $\operatorname{char}(F) = p$, then $(p \times x) = 0$ for any $x \in F$.

Proof.

(1) Let $n > 0$, $n \in \mathbb{Z}$, be the characteristic of F. Then $(n \times 1) = 0$. If n is not prime, then $n = ab$ for $a, b \in \mathbb{Z}$, $0 < a, b < n$, and so $0 = (a \times 1)(b \times 1)$. But then $(a \times 1) = 0$ or $(b \times 1) = 0$. This is a contradiction since $a, b < n$.

(2) If $\operatorname{char}(F) = 0$ and $(n \times x) = x(n \times 1) = 0$ then $x = 0$ or $(n \times 1) = 0$, so $x = 0$.

(3) If $\operatorname{char}(F) = p$, p prime, then for any $x \in F$, $(p \times x) = (p \times 1)x = 0x = 0$. ■

Note. A finite field always has finite characteristic. However, an infinite field can have finite characteristic. For example the field of rational functions over $\mathbb{Z}/p$, i.e.

$$F = \left\{ \frac{f(x)}{g(x)} \mid f(x), g(x) \in \mathbb{Z}/p[x], g(x) \neq 0 \right\}.$$

The characteristic of F is p, because $(p \times 1) = 0$.

Proposition 5.14. If F is a field of characteristic p, then

$$\{0, 1, (2 \times 1), \ldots, ((p-1) \times 1)\}$$

is a subfield of F isomorphic to $\mathbb{Z}/p$. If $\mathrm{char}(F) = 0$, then F contains a subfield isomorphic to $\mathbb{Q}$.

Proof. If $\mathrm{char}(F) = p$, then $\{0, 1, (2 \times 1), \ldots, ((p-1) \times 1)\}$ is closed under addition and multiplication and subtraction. Thus it is a subring of F with no zero divisors (since F has no zero divisors). Hence it is a finite integral domain and hence a field. If $\mathrm{char}(F) = 0$, then the set $\{0, 1, (2 \times 1), \ldots, (n \times 1), \ldots\}$ is infinite. It is closed under $+$ and $\cdot$ but not closed under $-$ or inverses. Now add $-(n \times 1)$ for $n > 0$ and get a field isomorphic to $\mathbb{Z}$. Since F is a field, it contains the ratios of these elements, adding these we get a subfield isomorphic to $\mathbb{Q}$. ■

prime subfield

Note. If $\mathrm{char}(F) = p$, then $\mathbb{Z}/p \subset F$ is the smallest subfield of F and if $\mathrm{char}(F) = 0$ then $\mathbb{Q} \subset F$ is the smallest subfield. It is called the *prime subfield* of F.

Note. Employing Proposition 5.14, we can consistently write $k \in F$ for $k \in \mathbb{Z}$ and F a field, taking k to be $(k \times 1)$ for $k \geq 0$ and $(k \times -1)$ for $k < 0$. Hence we can drop the $\times$ notation.

Theorem 5.15. Any finite field has p^n elements, where $n \in \mathbb{Z}$, $n > 0$, and p is a prime number and the characteristic of F.

Proof. Since F is finite, $\mathrm{char}(F) < \infty$. Let $p = \mathrm{char}(F)$, prime number. Then $\mathbb{Z}/p$ is a subfield of F. Since everything is finite, $[F : \mathbb{Z}/p] = \dim_{\mathbb{Z}/p}(F) = n < \infty$. If $e_1, \ldots, e_n$ is a basis of F over $\mathbb{Z}/p$, then

$$F = \{a_1 e_1 + \cdots + a_n e_n \mid a_i \in \mathbb{Z}/p\}.$$

Hence $|F| = p^n$. ■

Chapter 6

Ruler and Compass Constructions

Rules of The Game: Given two points, we can draw lines and circles, creating more points (the intersections) and more lines (joining two points). We can draw a circle with centre in some existing point and other existing point on its circumference. **Question is:** *What are all the constructible points?* (or what we cannot construct)

Construction 6.1. Given two points P and Q, we can construct their perpendicular bisector.

Proof. Draw two circles with the same radius (greater than $|PQ|$) with centre in P and Q. Join their intersection points to get the bisector.

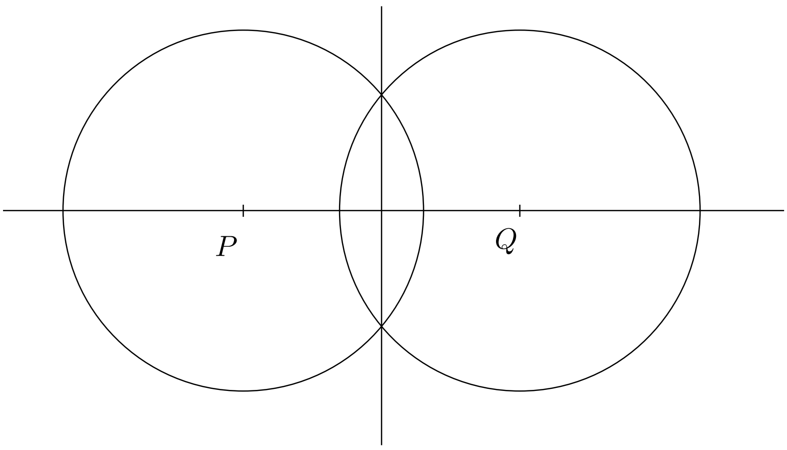

Figure 6.1: Constructing a perpendicular bisector of P and Q.

■

Construction 6.2. Given two points O and X, we can construct the line through O perpendicular to the line joining O and X.

Proof. Draw the line OX. Draw a circle centered in O with radius $|OX|$. Let the intersection point (the one that is not X) be Y. Construct the perpendicular bisector of X and Y. ■

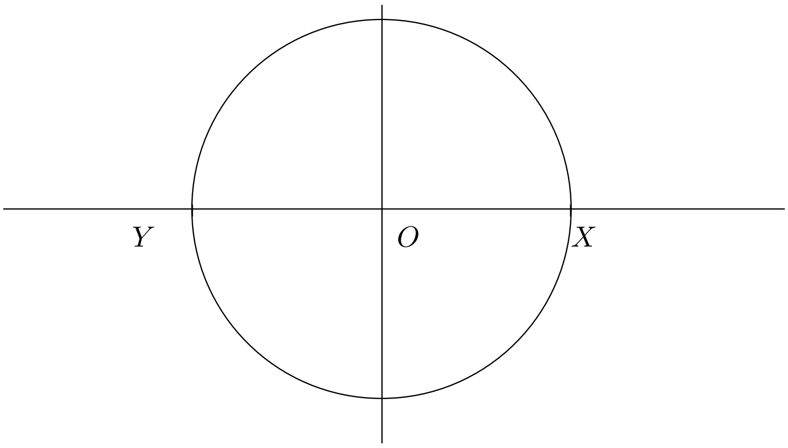

Figure 6.2: Constructing a line through O perpendicular to OX.

Construction 6.3. Given a point X and a line l, we can drop a perpendicular from X to l.

Proof. Draw a circle centered at X such that it has two intersection points with l. Find their perpendicular bisector. ■

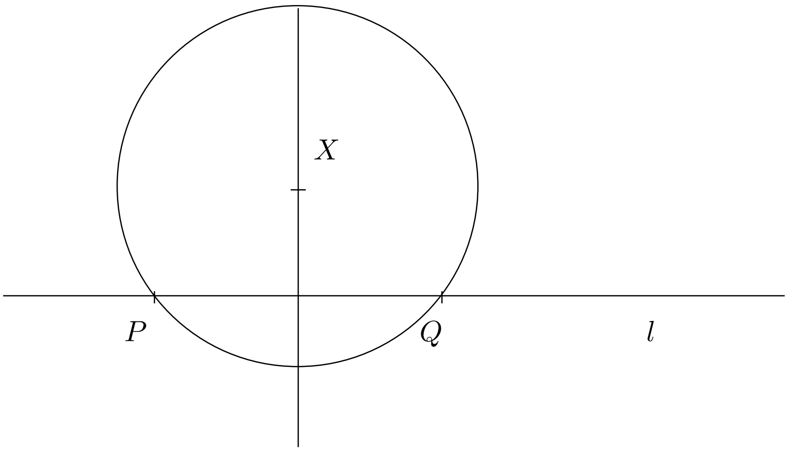

Figure 6.3: Droping a line from X perpendicular to l.

Construction 6.4. Given two intersecting lines l_1 and l_2, we can construct a line l_3 that bisects the angle between l_1 and l_2.

Proof. Draw a circle centered in the intersection of l_1 and l_2. Find the perpendicular bisector of its intersection points with l_1 and l_2. ■

The problems unsolved by the Greeks:

1. trisect an angle,
2. square the circle (construct a square of the same area as a given circle),
3. duplicate the cube (construct a cube with twice the volume as a given cube).

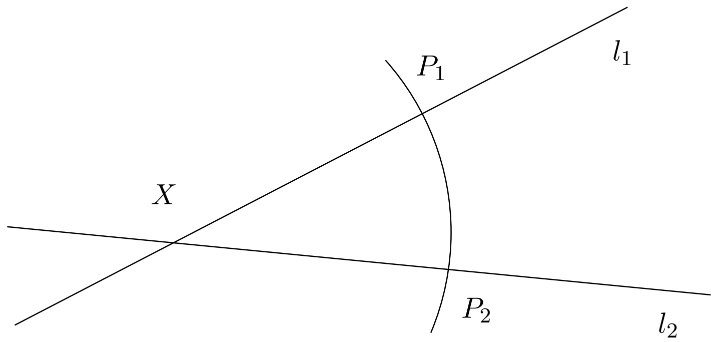

Figure 6.4: Constructing an angluar bisector between l_1 and l_2.

Constructing a regular n-gon

We can easily construct an equilateral triangle, square, regular pentagon. We cannot construct regular $7, 11, 13$-gons. Amazingly, we can construct a regular 17-gon using just ruler and compass!
The 2 original points, say O and X can be used to construct a coordinate system. Let $|OX| = 1$. Construct a perpendicular to OX through O. Any point in the plane is given by its coordinates, say (a, b).

Note. If we can construct (a, b), then we can construct $(a, 0)$, $(b, 0)$.

Definition 6.5. A real number $a \in \mathbb{R}$ is *constructible* if $(a, 0)$ is constructible from $O = (0, 1)$ and $X = (1, 0)$. *constructible*

Proposition 6.6. The set $\{a \in \mathbb{R} \mid a \text{ is constructible}\}$ is a subfield of $\mathbb{R}$.

Proof. Both 0 and 1 are constructible. We need to show that if a and b are constructible, then so are $-a, a+b, ab$ and $\frac{1}{b}$ if $b \neq 0$. For $-a$, draw a circle with centre in O passing through a. For $a+b$, construct $(0, b)$ and then (a, b). Then construct a

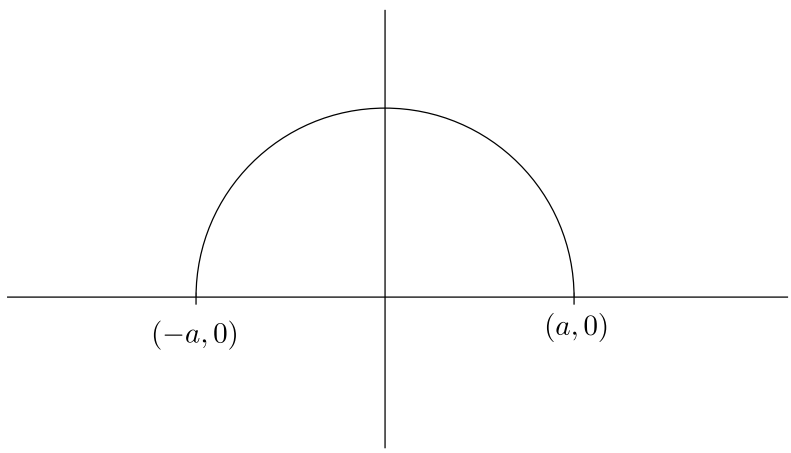

Figure 6.5: Constructing $-a$ from a.

circle with centre in a and passing through (a, b). For ab, construct $(0, 1)$ and join it

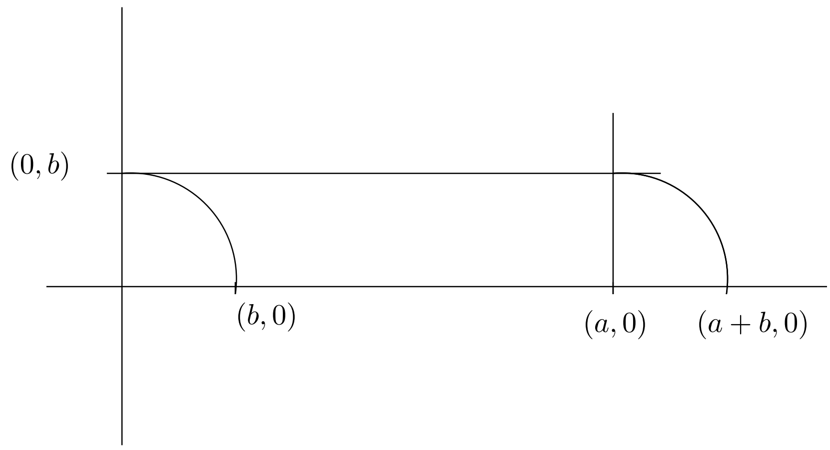

Figure 6.6: Constructing $a + b$ from a, b.

with $(a, 0)$. Next, construct a parallel line through $(0, b)$ (drop a perpendicular from $(0, b)$ and then construct a line perpendicular to it). Let $(c, 0)$ be its intersection with the x axis. Observe that (from similar triangles)

$$\frac{c}{b} = \frac{a}{1}.$$

Hence $c = ab$. For $\frac{1}{b}$, construct $(0, b)$ and draw a line joining $(0, b)$ and $(1, 0)$. Then

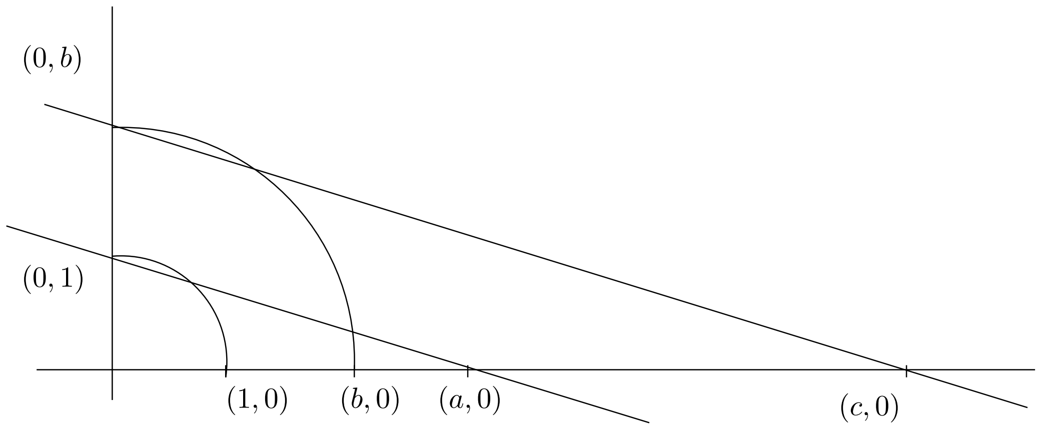

Figure 6.7: Constructing ab from a, b.

construct a line parallel to it passing through $(0, 1)$ and let $(c, 0)$ be its intersection with the x axis. Again, from similar triangles, $\frac{1}{b} = \frac{c}{1}$ and hence $c = \frac{1}{b}$. ■

Proposition 6.7. Every rational number is constructible. If $a > 0$ is constructible, then so is $\sqrt{a}$.

Proof. On a line (say the x axis) construct a length $a = |OA|$ next to length $1 = |BO|$. Let Z be the mid-point of AB. Draw the circle centered at Z with circumference containing A. Draw a perpendicular to the line AB from O and call its intersection

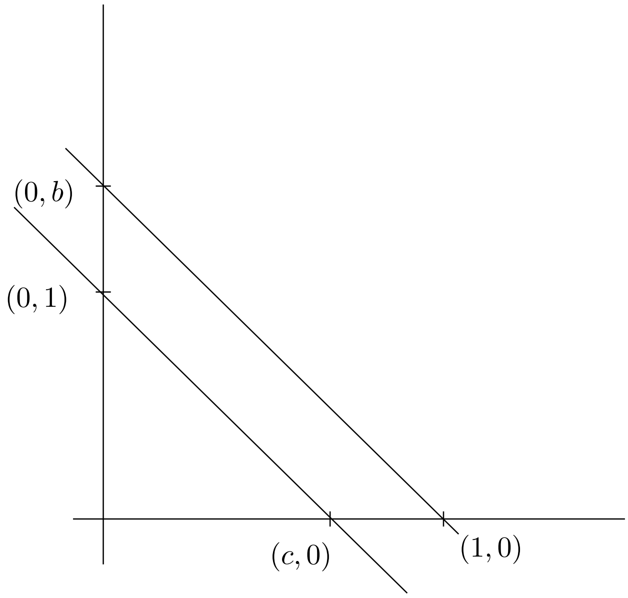

Figure 6.8: Constructing $\frac{1}{b}$ from b.

with the circle C. We claim that $|OC| = \sqrt{a}$. Indeed, observe that $|CZ| = \frac{a+1}{2}$. Also $|OZ| = |BZ| - 1 = \frac{a-1}{2}$. Now by Pythagoras,

$$|OC| = \sqrt{\left(\frac{a+1}{2}\right)^2 - \left(\frac{a-1}{2}\right)^2} = \sqrt{a}.$$

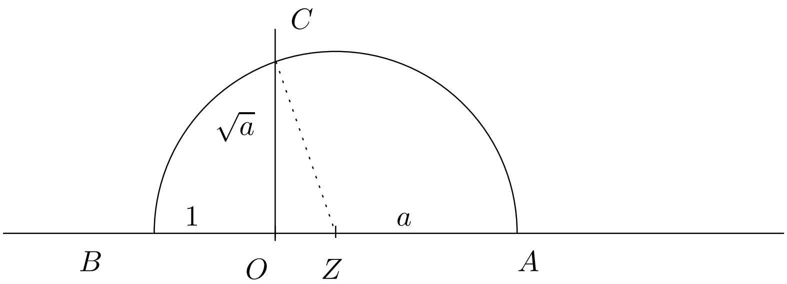

Figure 6.9: Constructing $\sqrt{a}$ from a.

■

Proposition 6.8. Let P be a finite set of points in the plane $\mathbb{R}$ and let K be the smallest subfield of $\mathbb{R}$ which contains the coordinates of the points of P. If (x_1, y_1) can be obtained from the points of P by a one-step construction, then x_1 and y_1 belong to the field $K(\sqrt{\delta})$ for some $\delta \in K$, i.e. x_1 and y_1 are of the form $a + b\sqrt{\delta}$, where $a, b \in K$.

Proof. Let $A = (a_1, a_2)$, $B = (b_1, b_2)$, $C = (c_1, c_2)$ and $D = (d_1, d_2)$. We can obtain a new point in 3 ways:

1. To construct $M = (x, y)$ from intersection of lines through A, B and C, D: The line through A, B has equation
$$(x - a_1)(b_2 - a_2) = (y - a_2)(b_1 - a_1). \tag{1}$$
The line through C, D has equation
$$(x - c_1)(d_2 - c_2) = (y - c_2)(d_1 - c_1). \tag{2}$$
Recall that $a_i, b_i, c_i, d_i \in K$ for $i = 1, 2$. Multiply (1) by $(d_2 - c_2)$, then subtract (2) multiplied by $b_2 - a_2$. Find $y \in K$ and then use the other equation to find $x \in K$ (and so $x \in K(\sqrt{\delta})$ as well).

2. Get $M = (x, y)$ as an intersection of line through C, D and a circle with in A and radius $|AB|$: Similar to the case 1., with first equation replaced by
$$(x - a_1)^2 + (y - a_2)^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2. \tag{1}$$
Use (2) to express $y = \alpha x + \beta$ for $\alpha, \beta \in K$ (always possible except when $d_1 = c_1$; then express x in terms of y). Substitute x into the equation of the circle. Solve this (quadratic) and find x. If the quadratic is
$$x^2 + \xi x + \gamma = 0$$
for $\xi, \gamma \in K$, then
$$x = \frac{-\xi \pm \sqrt{\xi^2 - 4\gamma}}{2}.$$
But $\delta = \xi^2 - 4\gamma$ is not always a square in K and so $x \in K(\sqrt{\delta})$ and also $y \in K(\sqrt{\delta})$.

3. Get $M = (x, y)$ as an intersection of two circles with centres in A and C and diameters $|AB|$ and $|CD|$ respectively: Get equations of the circles:
$$(x - a_1)^2 + (y - a_2)^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2, \tag{1}$$
$$(x - c_1)^2 + (y - c_2)^2 = (d_1 - c_1)^2 + (d_2 - c_2)^2. \tag{2}$$
Then (1) - (2) is a linear equation in x and y; proceed as in the case 2. ∎

Theorem 6.9. Let P be a set of points constructible in a finite number of steps from $(0, 0)$ and $(1, 0)$ and let K be the smallest subfield of $\mathbb{R}$ containing the coordinates of these points. Then $[K : \mathbb{Q}] = 2^t$ for some $t \in \mathbb{Z}$, $t \geq 0$.

Proof. Clearly $\mathbb{Q} \subset K$. Write P in order of construction $0, 1, p_1, \ldots, p_n$. Let K_i be the smallest subfield of $\mathbb{R}$ containing the coordinates of $p_1, \ldots, p_i$. Then either $K_{i+1} = K_i$ or $[K_{i+1} : K_i] = 2$ by the previous proposition. Therefore $[K_i : \mathbb{Q}] = 2^a$, $a \in \mathbb{Z}$, $0 \leq a \leq i$ by Theorem 5.9. ∎

Corollary 6.10. If $a \in \mathbb{R}$ is constructible, then $[\mathbb{Q}(a) : \mathbb{Q}] = 2^t$, $t \in \mathbb{Z}$, $t \geq 0$.

Proof. Let $(a, 0)$ be constructible. Then $\mathbb{Q}(a) \subset K$, where K is as in Theorem 6.9. Then $\mathbb{Q} \subset \mathbb{Q}(a) \subset K$, hence by Corollary 5.10 $[\mathbb{Q}(a) : \mathbb{Q}]$ divides $[K : \mathbb{Q}] = 2^n$. ∎

Theorem 6.11. It is impossible to duplicate the cube.

Proof. For a cube of side 1, the issue is to construct $\sqrt[3]{2}$. Note that $x^3 - 2$ is the minimal polynomial of $\sqrt[3]{2}$: it is indeed irreducible (e.g. by the Eisenstein's criterion with $p = 2$). Therefore $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = \deg(x^3 - 2) = 3$. By Corollary 6.10 $\sqrt[3]{2}$ is not constructible. ■

Theorem 6.12. It is impossible to square the circle

Outline of the proof. We have to show that $\sqrt{\pi}$ is not constructible. If $\sqrt{\pi}$ is constructible, then so is π (by Proposition 6.6). A Theorem (not easy to prove) says that π is not algebraic over $\mathbb{R}$. This implies that the smallest subfield of $\mathbb{R}$ containing π is an *infinite* extension of $\mathbb{Q}$. Thus π is not constructible by Corollary 6.10. ■

Proposition 6.13. The following are equivalent:

(1) constructing a regular n-gon in the unit circle,

(2) constructing an angle $\frac{2\pi}{n}$,

(3) constructing $\cos\frac{2\pi}{n}$.

Proof. Obvious. ■

Theorem 6.14. It is false that every angle can be trisected.

Proof. Can construct $\frac{\pi}{3}$. We will show it cannot be trisected, i.e. $\cos\frac{\pi}{9}$ cannot be constructed using ruler and compass. Observe that

$$\begin{aligned} \cos 3\theta &= \cos\theta\cos 2\theta - \sin\theta\sin 2\theta \\ &= \cos\theta(2\cos^2\theta - 1) - 2\sin^2\theta\cos\theta \\ &= 4\cos^3\theta - 3\cos\theta. \end{aligned}$$

Apply this to $\theta = \frac{\pi}{9}$ to get

$$4\cos^3\frac{\pi}{9} - 3\cos\frac{\pi}{9} = \frac{1}{2}.$$

Therefore $\cos\frac{\pi}{9}$ is a root of $4t^3 - 3t - \frac{1}{2}$. Constructing the angle θ is equivalent to constructing the number $\cos\theta$. Let us show that $\cos\frac{\pi}{9}$ is not constructible. First we show that $t^3 - \frac{3}{4}t - \frac{1}{8}$ is the minimal polynomial of $\cos\frac{\pi}{9}$ over $\mathbb{Q}$. To show that it is irreducible, consider $8t^3 - 6t - 1$. Write $y = 2t$ to get $y^3 - 3y - 1$. By a Corollary of Gauss's Lemma, if $y^3 - 3y - 1$ is reducible over $\mathbb{Q}$, it is reducible over $\mathbb{Z}$, thus has a root in $\mathbb{Z}$. Suppose that

$$y^3 - 3y - 1 = (y - a)(y^2 + by + c)$$

for $a, b, c \in \mathbb{Z}$ with a root a. Now $-ac = 1$ and so $a = \pm 1$. But ± 1 is not a root, therefore $t^3 - \frac{3}{4}t - \frac{1}{8}$ is the minimal polynomial of $\cos\frac{\pi}{9}$. Therefore $[\mathbb{Q}(\cos\frac{\pi}{9}) : \mathbb{Q}] = 3$. Since 3 is not a power of 2, $\cos\frac{\pi}{9}$ is not constructible. ■

Proposition 6.15. Let $\omega = e^{\frac{2\pi i}{n}} = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$ with $n > 2$. Then $\mathbb{Q}\left(\cos\frac{2\pi}{n}\right) \subset \mathbb{Q}(\omega)$ and $[\mathbb{Q}(\omega) : \mathbb{Q}\left(\cos\frac{2\pi}{n}\right)] = 2$.

Proof. Let $\alpha = \cos\frac{2\pi}{n}$ and $\overline{\omega} = \cos\frac{2\pi}{n} - i\sin\frac{2\pi}{n}$. Observe that ω is a root of

$$\begin{aligned}(x-\omega)(x-\overline{\omega}) &= x^2 - x(\omega+\overline{\omega}) + \omega\overline{\omega}\\ &= x^2 - 2x\cos\frac{2\pi}{n} + 1\\ &= x^2 - 2\alpha x + 1.\end{aligned}$$

Also

$$\omega\overline{\omega} = \left(\cos\frac{2\pi}{n}\right)^2 + \left(\sin\frac{2\pi}{n}\right)^2 = 1$$

and so

$$\alpha = \frac{1}{2}\left(\omega + \overline{\omega}\right) = \frac{1}{2}(\omega + \omega^{-1}) \in \mathbb{Q}(\omega).$$

Therefore $\mathbb{Q}(\alpha) \subset \mathbb{Q}(\omega)$. The minimal polynomial of ω over $\mathbb{Q}(\alpha)$ is $x^2 - 2\alpha x + 1$. Note that it is irreducible because it is irreducible over a bigger field $\mathbb{R}$ $(\omega, \overline{\omega} \notin \mathbb{R})$. ∎

Proposition 6.16. Let p be an odd prime, $\omega = e^{\frac{2\pi i}{p}}$. Then $[\mathbb{Q}(\omega) : \mathbb{Q}] = p - 1$ and $[\mathbb{Q}(\cos\frac{2\pi}{p}) : \mathbb{Q}] = \frac{p-1}{2}$.

Proof. Since $\omega^p = 1$, ω is a root of $x^p - 1$. We have

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

and $x^{p-1} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$ by Example 4.29 and ω is its root. Hence $x^{p-1} + \cdots + x + 1$ is the minimal polynomial of ω and thus $[\mathbb{Q}(\omega) : \mathbb{Q}] = p - 1$. We have $\mathbb{Q} \subset \mathbb{Q}(\cos\frac{2\pi}{p}) \subset \mathbb{Q}(\omega)$. Since

$$[\mathbb{Q}(\omega) : \mathbb{Q}] = [\mathbb{Q}(\omega) : \mathbb{Q}(\cos 2\pi/p)][\mathbb{Q}(\cos 2\pi/p) : \mathbb{Q}]$$

and $[\mathbb{Q}(\omega) : \mathbb{Q}(\cos\frac{2\pi}{p})] = 2$, we have that $[\mathbb{Q}(\cos\frac{2\pi}{p}) : \mathbb{Q}] = \frac{p-1}{2}$. ∎

Theorem 6.17. If a regular p-gon is constructible, where p is an odd prime, then $p - 1 = 2^n$ for some n.

Proof. This is equivalent to constructing $\cos\frac{2\pi}{p}$, but then $[\mathbb{Q}(\cos\frac{2\pi}{p}) : \mathbb{Q}] = \frac{p-1}{2}$ is a power of 2. Hence $p - 1 = 2^n$ for some n. ∎

Note. This implies that a regular 7-gon, 11-gon, 13-gon are not constructible.

Note. For $p - 1 = 2^n$, write $n = m2^r$ where m is odd, $r \in \mathbb{Z}$, $r \geq 0$. Let $\alpha = 2^{2^r}$ and so $2^n = 2^{2^r m} = \alpha^m$. We have

$$\begin{aligned}p &= 1 + 2^n = 1 + \alpha^m = 1 - (-\alpha)^m\\ &= (1 - (-\alpha))(1 + (-\alpha) + (-\alpha)^2 + \cdots + (-\alpha)^{m-1}).\end{aligned}$$

Therefore $p = (1+\alpha)(1 - \alpha + \cdots + \alpha^{m-1})$ and $\alpha \geq 2$. If p is prime, then $1 - \alpha + \cdots + \alpha^{m-1} = 1$, i.e. $m = 1$.
Conclusion: If a prime p equals $1 + 2^n$, then $n = 2^r$. Such primes p are called *Fermat primes*. First few are $3, 5, 17, 257, 65537$.

Fermat prime

Proposition 6.18. If we can construct a regular n-gon where $n = ab$, then we can construct a regular a-gon.

Proof. Join every b-th vertex of the regular n-gon. ■

Corollary 6.19. If a regular n-gon is constructible, then $n = 2^a p_1 \cdots p_k$ where $p_1, \dots, p_k$ are Fermat primes.

Proof. Suppose a regular n-gon is constructible and consider the prime factors of n. If n is even, then we can construct a regular $n/2$-gon by joining every other vertex and continue until we get odd $m = n/2^a$. We need to show that m is a product of Fermat primes: in case m has a prime factor p that is not a Fermat prime, then by 6.18 we can construct a regular p-gon, a contradiction. It remains to show that m is a product of *distinct* Fermat primes. By the Sheet 8, it is impossible to construct regular p^2-gon for p prime. Hence if p^k, $k > 1$ is a factor of m, p^2 is as well and we can construct a regular p^2-gon, a contradiction. ■

Proposition 6.20. If m and n are coprime and we can construct a regular m-gon and a regular n-gon, then we can also construct a regular mn-gon.

Proof. If $\operatorname{hcf}(m, n) = 1$, then there exist $a, b \in \mathbb{Z}$ such that $am + bn = 1$. It follows that

$$\frac{1}{mn} = \frac{a}{n} + \frac{b}{m}$$

and so

$$\frac{2\pi}{mn} = a\frac{2\pi}{n} + b\frac{2\pi}{m}.$$

Thus $\frac{2\pi}{mn}$ is constructible. ■

Note. In fact, a regular n-gon is constructible *iff* $n = 2^a p_1^{b_1} \cdots p_k^{b_k}$ for $p_1, \dots, p_k$.

Proposition 6.21. A regular pentagon is constructible.

Proof. Let $\alpha = \cos \frac{2\pi}{5}$. Proposition 6.16 says that $[\mathbb{Q}(\alpha) : \mathbb{Q}] = \frac{5-1}{2} = 2$. Hence the degree of minimal polynomial of α over $\mathbb{Q}$ is 2. Let $x^2 + bx + c$, $b, c \in \mathbb{Q}$ be the minimal polynomial of α over $\mathbb{Q}$. Then

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

Since b, c are constructible, so is $\sqrt{b^2 - 4c}$ and so is α. ■

Chapter 7

Finite fields

We know that $\mathbb{Z}/p$ is a finite field with p elements for p prime. If F is a finite field, then we have $p = (p \times 1) = 0$ for $p = \text{char}(F)$ prime. Theorem 5.15 says that $|F| = p^n$. Our aim is to show that for any prime power p^n there exists a finite field with p^n elements.

Proposition 7.1. Let p be an odd prime. Then there exists a field with p^2 elements.

Proof. For any $r \in \mathbb{Z}/p$ we have $-r = p - r$ has the same square as r. Also $r = -r$ iff $2r = 0$ iff $r = 0$ (since p is odd). Therefore, we have exactly $\frac{p-1}{2}$ non-zero squares and thus at least one non-square $a \in \mathbb{Z}/p$, $a \neq 0$. Then $x^2 - a$ is an irreducible polynomial over $\mathbb{Z}/p$. By Proposition 4.19, $\mathbb{Z}/p[x]/(x^2 - a)\mathbb{Z}/p[x]$ is a field with elements

$$\left\{\alpha_0 + \alpha_1 x + (x^2 - a)\mathbb{Z}/p \mid \alpha_0, \alpha_1 \in \mathbb{Z}_p\right\}.$$

Hence the constructed field contains p^2 elements. ■

Proposition 7.2 (is 4.19). Let F be a field and $p(x) \in F[x]$ be an irreducible polynomial of degree n. Write $F(\alpha)$ for the field $F[x]/p(x)F[x]$. Then $F(\alpha)$ is a field containing F and

$$F(\alpha) = \left\{b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1} \mid b_i \in F\right\}$$

where α is the image of x under the map $F[x] \to F(\alpha)$ sending each polynomial to its value at α. We have $p(\alpha) = 0$.

Example 7.3.

1. Let $n = 1$, $p(x) = a_0 + a_1 x$, $a_1 \neq 0$. Then $F(\alpha) = F$. What is the image of x? We have

$$\begin{aligned} \frac{1}{a_1}p(x) &= x + \frac{a_0}{a_1}, \\ I &= (a_1 x + a_0)F[x], \end{aligned}$$

   so $x + I = -\frac{a_0}{a_1} + I$.

2. Let $F = \mathbb{Q}$, $p(x) = x^2 - 2$. Then $F(\alpha) = \mathbb{Q}(\sqrt{2})$ and $p(\sqrt{2}) = 0$.

3. Let $F = \mathbb{Z}/2$, $p(x) = x^2 + x + 1$. Then

$$F[x]/p(x)F[x] = F(\omega) = \left\{a_0 + a_1\omega \mid a_i \in \mathbb{Z}/2, 1 + \omega + \omega^2 = 0\right\}.$$

Corollary 7.4. Let F be a field and let $f(x) \in F[x]$. Then there exists a field $K \supset F$ such that $f(x) \in K[x]$ is a product of linear factors $f(x) = c(x - \alpha_1)\cdots(x - \alpha_n)$ where $c \in K^*$, $\alpha_i \in K$. In other words, $f(x)$ has $\deg f(x)$ roots in K.

Proof. Let m be the number of roots of $f(x)$ in F. If $m = n$, then $K = F$. Otherwise, let $p(x)$ be an irreducible polynomial dividing $f(x)$. Define $F_1 = F(\alpha)$ as in Proposition 7.2. Then $p(\alpha) = 0$ and so α is a root of $p(x)$ in F_1 and so a root of $f(x)$ in F_1. Then write $f(x) = (x - \alpha)f_1(x) \in F_1[x]$. Repeat the same argument for F_1. Carry on until we construct a finite extension F over which $f(x)$ is a product of linear factors. ∎

Example 7.5. Let $F = \mathbb{Q}$, $f(x) = (x^2-2)(x^2+1)$. Take $p(x) = x^2-2$, $F_1 = \mathbb{Q}(\sqrt{2})$. Over $\mathbb{Q}(\sqrt{2})$, $f(x) = (x - \sqrt{2})(x+\sqrt{2})(x^2+1)$. Then $F_2 = \mathbb{Q}(\sqrt{2})(\sqrt{-1}) = K$. Over K, $f(x) = (x - \sqrt{2})(x + \sqrt{2})(x - \sqrt{-1})(x + \sqrt{-1})$.

Theorem 7.6. There exists a field with p^n elements for any prime p and positive integer n.

Proof. Let $F = \mathbb{Z}/p$, $f(x) = x^{p^n} - x \in F[x]$. There exists a field K such that $F \subset K$ and $f(x) = c\prod_{i=1}^{p^n}(x - \alpha_i)$ for some $c \in K^*$, $\alpha_i \in K$. Let $E = \{\alpha_i \mid 1 \le i \le p^n\}$. Two things to prove: (1) E is a field, (2) $|E| = p^n$, i.e. the α_i are distinct.
For (1): Clearly $\{0, 1\} \subset E$. If $a \in E$, then $-a \in E$: If $p = 2$, $a = -a$. If p is odd, $(-a)^{p^n} = -a^{p^n}$ so that $f(-a) = -a^{p^n} - (-a) = -f(a) = 0$. If $a, b \in E$, then $ab \in E$, since

$$\begin{aligned} f(ab) &= (ab)^{p^n} - ab \\ &= a^{p^n}b^{p^n} - ab. \end{aligned}$$

But $a^{p^n} = a$, $b^{p^n} = b$, thus $f(ab) = ab - ab = 0$. If $b \in E$ and $b \neq 0$, then

$$\left(\frac{1}{b}\right)^{p^n} = \frac{1}{b^{p^n}} = \frac{1}{b}$$

therefore $f(\frac{1}{b}) = 0$, thus $\frac{1}{b} \in E$.

Lemma 7.7. For any elements x and y in a field of characteristic p we have $(a+b)^p = a^p + b^p$.

Proof. If $p = 2$, $(a + b)^2 = a^2 + 2ab + b^2 = a^2 + b^2$. We have

$$\begin{aligned} (a+b)^p &= a^p + pa^{p-1}b + \frac{p(p-1)}{2}a^{p-2}b^2 + \cdots \\ &\quad + \frac{p(p-1)\cdots(p-m+1)}{m!}a^{p-m}b^m + \cdots + b^p. \end{aligned}$$

Observe that $\frac{p(p-1)\cdots(p-m+1)}{m!}$ is an integer divisible by p since p doesn't divide $m!$ for $m < p$. So $(a + b)^p = a^p + b^p$. ∎

Apply the Lemma to $(a+b)^{p^n}$, where $a^{p^n} = a$ and $b^{p^n} = b$:

$$((a+b)^p)^{p^{n-1}} = (a^p + b^p)^{p^{n-1}} = \left(a^{p^2} + b^{p^2}\right)^{p^{n-2}} = \cdots = a^{p^n} + b^{p^n},$$

thus $f(a+b) = 0$ and $a + b \in E$. This proves (1).
For (2): Clearly $|E| \leq p^n$. Let us show that any root of $f(x)$ is a simple root. By part (1), if p is odd,

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

If $p = 2$, $b = -b$ for any $b \in F$ and so

$$x^{2^n} - a^{2^n} = x^{2^n} + a^{2^n} = (x+a)^{2^n} = (x-a)^{2^n}.$$

We have

$$\begin{aligned} f(x) &= x^{p^n} - x \\ &= x^{p^n} - x - \underbrace{(a^{p^n} - a)}_{=0} = (x^{p^n} - a^{p^n}) - (x - a) \\ &= (x-a)\left((x-a)^{p^n-1} - 1\right). \end{aligned}$$

Therefore, we have written $f(x) = (x-a)g(x)$, where $g(x) = (x-a)^{p^n-1} - 1$. Clearly $g(a) = -1 \neq 0$, thus $(x-a)^2$ does not divide $f(x)$, so a is a simple root of $f(x)$. ■

There is a general method of checking that a root of a polynomial is simple. The idea is just taking the derivative.

Definition 7.8. Let $f(x)$ be a polynomial with coefficients in a field F of any characteristic, $f(x) = a_0 + a_1x + \cdots + a_nx^n$. The *derivative* $f'(x)$ is defined as $f'(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1}$. *derivative* $f'(x)$

Then clearly $(f(x)+g(x))' = f'(x)+g'(x)$. Also $(f(x)g(x))' = f'(x)g(x)+f(x)g'(x)$: it is enough to show that $(x^nx^m)' = (n+m)x^{n+m-1} = nx^{n-1}x^m + mx^nx^{m-1}$.

Proposition 7.9. If $f(x) \in F[x]$, where F is a field, and K is a field extension of F such that $f(\alpha) = 0$ for $\alpha \in K$, then α is a multiple root of $f(x)$ iff $f'(\alpha) = 0$.

Proof. Write $f(x) = (x-\alpha)^m g(x)$, where $g(x) \in K[x]$, $m \geq 0$, $g(\alpha) \neq 0$. Then $f'(x) = m(x-\alpha)^{m-1}g(x) + (x-\alpha)^m g'(x)$. If α is multiple, then $m \geq 2$ and hence $f'(\alpha) = 0$. If α is simple, then $m = 1$ so that $f'(\alpha) = g(\alpha) + 0 = g(\alpha) \neq 0$. ■

Example 7.10.

1. Let $f(x) = x^{p^n} - x$ over $\mathbb{Z}/p$ with $\mathrm{char}(\mathbb{Z}/p) = p$. Then $f'(x) = p^n x^{p^n-1} - 1 = -1$ so any roof of $f(x)$ is simple.

2. Let $f(x) = x^m - 1$. Then $f'(x) = mx^{m-1}$, $x = 0$ is not a root. Therefore $f(x)$ has simple root iff $\mathrm{char}(F)$ does not divide m.

Recall some facts from group theory. A group G is *cyclic* if $G = \{1, g, g^2, \dots\}$ for some $g \in G$. *cyclic*
Let G be a finite group of order n, $n = |G|$. The *order of an element* $x \in G$ is the least positive integer r such that $x^r = 1$. A finite group G is cyclic if there exists $g \in G$ such that the order of g equals to $|G|$. Such g is called the *generator* of G. We will write $\mathrm{ord}(x)$ for the order of $x \in G$. *order* *generator* $\mathrm{ord}(x)$

Note.

(1) If $x^d = 1$, then $\mathrm{ord}(x)|d$.

(2) If $|G| = n = ad$ and g is the generator of G, then the elements $x \in G$ satisfying $x^d = 1$ are $\{1, g^a, g^{2a}, \ldots, g^{(d-1)a}\}$.

Proof.

(1) Say if $\mathrm{ord}(x) = a$, then write $d = qa + r$, where $r = 0$ or $0 < r < a$. Then $x^d = 1$ and $x^a = 1$. Thus $x^r = x^{d-qa} = x^d(x^a)^{-q} = 1$. If $r \neq 0$, we get a contradiction because $r < a$. Hence $r = 0$, so that $a = \mathrm{ord}(x)|d$.

(2) Clearly, $(x^{ia})^d = (x^{ad})^i = x^{ni} = 1$ since by Lagrange's theorem, $\mathrm{ord}(x)|n$. Now suppose that $x^d = 1$ and write $x = g^i$. Then $g^{di} = 1$. Write $di = qn + r$ where $r = 0$ or $0 < r < n$. Then $g^r = g^{di}g^{-qn} = 1$. If $r \neq 0$, we get a contradiction since $r < n = \mathrm{ord}(g)$. Therefore $r = 0$ so that $di = qn = qad$. Thus $i = qa$. ■

$\varphi(d)$
Euler's function

Definition 7.11. For each $d \in \mathbb{N}$ define $\varphi(d)$ as the number of elements of order d in a cyclic group with d elements. Function $\varphi(d)$ is called *Euler's function.* The first few values are:

d	1	2	3	4	5	6	7
$\varphi(d)$	1	1	2	2	4	2	6

Note. $\mathbb{Z}/n$ with its additive structure is a cyclic group with n elements. Then g is a generator of $\mathbb{Z}/n$ if $\{0, g, g+g, g+g+g, \ldots\} = \mathbb{Z}/n$. For example, if $n = 4$, $\mathbb{Z}/4 = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}\}$, then $\bar{1}$ and $\bar{3}$ are the generators. If $n = 5$, then since 5 is prime, every non-zero element if a generator. If $n = 6$, the generators are $\bar{1}$ and $\bar{5}$.

Lemma 7.12. For an integer d, $d = \sum_{\delta|d} \varphi(\delta)$.

Proof. Let G be a cyclic group with d elements, $G = \langle g \rangle$. By Lagrange, $\mathrm{ord}(x)|d$ for any $x \in G$. Hence

$$d = |G| = \sum_{\delta|d} |\{x \in G \mid \mathrm{ord}(x) = \delta\}|.$$

By part (2) of the above note, all the elements $x \in G$, $\mathrm{ord}(x) = \delta$ generate the unique cyclic subgroup of G with δ elements (i.e. $\{1, g^a, \ldots, g^{(d-1)a}\}$ where $d = a\delta$). The set $\{1, g^a, \ldots, g^{(d-1)a}\}$ is a group generated by g^a. Since $\mathrm{ord}(g^a) = \delta$, this is a cyclic group of δ elements. Thus $|\{g \in G \mid \mathrm{ord}(g) = \delta\}| = \varphi(\delta)$. Hence $d = \sum_{\delta|d} \varphi(\delta)$. ■

Proposition 7.13. Let d be a factor of $|F| - 1$. Then the polynomial $x^d - 1$ has d distinct roots in a field F.

Proof. Clearly $F\backslash\{0\}$ is a group under multiplication and $|F\backslash\{0\}| = q-1$. Therefore, by Lagrange, $\alpha^{q-1} = 1$ for any $\alpha \in F \setminus \{0\}$. In other words, every non-zero element of $F \setminus \{0\}$ is a root of $x^{q-1} - 1$ and hence $x^{q-1} - 1$ has $q - 1$ distinct roots in F. Since $d|q - 1$,

$$x^{q-1} - 1 = (x^d - 1)g(x) \qquad (*)$$

where $g(x) = 1 + x^d + \cdots + x^{q-1-d}$ has at most $q - 1 - d$ distinct roots. Both sides of $(*)$ have the same number of roots, so $x^d - 1$ has d distinct roots. ■

Theorem 7.14. The multiplicative group $F \setminus \{0\}$ is cyclic.

Proof. Let $|F| = q$. Define $\psi(\delta)$ to be the number of elements of order δ in $F \setminus \{0\}$. Is δ a factor of $\psi(\delta) - 1$? Clearly, $\psi(\delta) = 0$ if $\delta \not| q - 1$.
Claim: For $d|q-1$, $\psi(d) = \varphi(d)$.

Proof. Recall that $\varphi(d) \geq 1$ by definition of the Euler's function. The roots of $x^d - 1$ are precisely the elements of $F \setminus \{0\}$ of order δ for all $\delta|d$. Conversely, if $\alpha^d = 1$, the order of α divides d. Hence the number of roots of $x^d - 1 = d$ (by Proposition 7.13) is $d = \sum_{\delta|d} \psi(\delta)$ $(*)$. The Lemma 7.12 says that $d = \sum_{\delta|d} \varphi(\delta)$. We continue by induction on d: clearly, $\varphi(1) = \psi(1) = 1$. Assume that $\psi(\delta) = \varphi(\delta)$ for all $\delta|q-1$ and $\delta < d$. Then from $(*)$

$$\psi(d) \quad = \quad d - \sum_{\delta|d, \delta \neq d} \psi(\delta).$$

By Lemma 7.12,

$$\varphi(d) = d - \sum_{\delta|d, \delta \neq d} \varphi(\delta).$$

Hence by induction assumption, the claim holds. ■

Then for $d = q - 1$, there are $\psi(q-1) = \varphi(q-1) \geq 1$ elements of order $q - 1$ in $F \setminus \{0\}$. Hence $F \setminus \{0\}$ is cyclic. ■

Index