Groups, Rings and Modules

University of Cambridge Part IB Mathematical Tripos

Yue Wu

Yusuf Hamied Department of Chemistry
Lensfield Road,
Cambridge, CB2 1EW

yw628@cam.ac.uk

Contents

1 Groups

1.1 Basic Concept

Definition 1.1. A *group* is a triple $(G, \cdot, e)$ of a set G, a function $\cdot : G \times G \to G$ and an element $e \in G$ such that

(G1) Associativity. $\forall a, b, c \in G$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

(G2) Identity. $\forall a \in G$, $a \cdot e = e \cdot a = a$.

(G3) Inverse. For each $a \in G$, $\exists a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$.

Definition 1.2. If $(G, \cdot, e)$ is a group and $H \subseteq G$ is a subset, then H is a *subgroup* of G, denoted $H \leq G$ if

(i) If $a, b \in H$, then $a \cdot b \in H$.

(ii) $e \in H$.

(iii) $(H, \cdot, e)$ is a group.

Lemma 1.3. $\varnothing \neq H \subseteq G$ is a subgroup $\iff \forall h_1, h_2 \in H$, $h_1 \cdot h_2^{-1} \in H$.

Examples.

(i) Additive groups. $(\mathbb{Z}, +, 0), (\mathbb{Q}, +, 0), (\mathbb{R}, +, 0), (\mathbb{C}, +, 0)$…

(ii) Groups of symmetry.

- S_n symmetry groups = bijections of $\{1, 2, \ldots, n\}$ to itself.
- D_{2n} dihedral group = symmetry of the regular n-gon.
- $\mathrm{GL}_n(\mathbb{R})$ general linear group = symmetry of the vector space $\mathbb{R}^n$.

(iii) Subgroups of these.

- Alternating groups $A_n \leq S_n$, the even permutations.
- Cyclic groups $C_n \leq D_{2n}$, rotational symmetry of n-gon.
- Special linear groups $\mathrm{SL}_n(\mathbb{R}) \leq \mathrm{GL}_n(\mathbb{R})$, $\{A \in \mathrm{GL}_n(\mathbb{R}) \mid \det A = 1\}$.

(iv) Abelian groups G such that $a \cdot b = b \cdot a$ $\forall a, b \in G$, e.g. $C_2 \times C_2 \cong V_4$.

(v) Quaternion group $Q = \{\pm 1, \pm \mathrm{i}, \pm \mathrm{j}, \pm \mathrm{k}\}$.

Definition 1.4. For a subgroup, the *left coset*

$$gH \coloneqq \{x \in G \mid x = g \cdot h \text{ for some } h \in H\}.$$

The collection of all left cosets of H, written $G : H$, gives a partition of G. Each H-coset is in bijection with H (and also with each other).

Theorem 1.5 (Lagrange's theorem). If G is a finite group and $H \leq G$, then

$$|G| = |H|\,|G : H|\,,$$

where $|G : H|$ is the number of H-cosets in G, known as the *index* of H in G.

Definition 1.6. The *order* of $g \in G$ is the least $n \in \mathbb{N} \cup \{\infty\}$ such that $g^n = e$. Write $\mathrm{ord}(g)$ for the number.

If $g^m = e$, then $\text{ord}(g) \mid m$.

Lemma 1.7. Let G be a finite group. $\text{ord}(g) \mid |G| \ \forall g \in G$.

Proof. If $n = \text{ord}(g)$, then $H = \{e, g, g^2, \dots, g^{n-1}\}$. Claim that H is a subgroup of G. If $g^i, g^j \in H$, then $g^i \cdot (g^j)^{-1} = g^{i-j}$, write $i - j = p \cdot n + r$, with $0 \leq r < n$.

$$g^{i-j} = g^{pn+r} = (g^n)^p \cdot g^r = g^r \in H\,,$$

so H is indeed a subgroup. By Lagrange's theorem, $n = |H|$ divides $|G|$. □

1.2 Normal Subgroups, Quotient, Homomorphism and Isomorphism

If $gH = g'H$, then $g^{-1}g' \in H$ and the converse holds. Let G/H be the set of the left H-cosets in G, try to define

$$(g_1H) \cdot (g_2H) = g_1g_2H\,.$$

This is well-defined only if the result is consistent for different choices of coset representatives. If $g_2H = g_2hH$ is another coset representative, then

$$(g_1H) \cdot (g_2hH) = g_1g_2hH = g_1g_2H\,.$$

If $g_1H = g_1hH$, then

$$(g_1hH) \cdot (g_2H) = g_1hg_2H\,,$$

which is equal to g_1g_2H if and only if

$$(g_1g_2)^{-1}g_1hg_2 = g_2^{-1}hg_2 \in H\,.$$

Hence this definition is legal if and only if $g^{-1}hg \in H$ for all $g \in G$, $h \in H$.

Definition 1.8. A subgroup $H \leq G$ is *normal*, denoted $H \lhd G$, if

$$g^{-1}hg \in H$$

$\forall h \in H, g \in G$.

Proposition 1.9. If $H \lhd G$, then the set G/H of left H-cosets form a group under the operation $(g_1H) \cdot (g_2H) = g_1g_2H$, with $e_{G/H} = eH$. This is the *quotient group* of G by H.

Proof. The discussions show that this well-defines a binary operation on G/H. All the axioms follow from the fact that they hold in G. □

Definition 1.10. If $(G, \cdot_G, e_G)$ and $(H, \cdot_H, e_H)$ are groups, then a function $\phi : G \to H$ is a *homomorphism* if

(i) $\phi(g_1 \cdot_G g_2) = \phi(g_1) \cdot_H \phi(g_2)$

(ii) $\phi(e_G) = e_H$

The *kernel* of ϕ is $\ker(\phi) \coloneqq \{g \in G \mid \phi(g) = e_H\}$, and the *image* of ϕ is $\text{im}(\phi) \coloneqq \{h \in H \mid h = \phi(g) \text{ for some } g \in G\}$.

Lemma 1.11. If $\phi : G \to H$ is a homomorphism, then $\phi(g^{-1}) = \phi(g)^{-1}$.

Proof.

$$\begin{aligned}\phi(g \cdot_G g^{-1}) &= \phi(e_G) = e_H \\ &= \phi(g) \cdot_H \phi(g)^{-1}\,.\end{aligned}$$

By the uniqueness of inverse, we must have $\phi(g^{-1}) = \phi(g)^{-1}$. □

Lemma 1.12. For a homomorphism $\phi : G \to H$, the kernel $\ker(\phi)$ is a normal subgroup of G, and the image $\mathrm{im}(\phi)$ is a subgroup of H.

Proof. Let $g, h \in \ker(\phi)$, then

$$\phi(g \cdot h^{-1}) = \phi(g) * \phi(h)^{-1} = e_H * e_H^{-1} = e_H \,,$$

so $gh^{-1} \in \ker(\phi)$. Also $\phi(e_G) = e_H$ so it is a subgroup. Let $x \in G$, we have

$$\begin{aligned}\phi(x^{-1}gx) &= \phi(x^{-1}) * \phi(g) * \phi(x) = \phi(x^{-1}) * \phi(x) \\ &= \phi(x^{-1}x) = \phi(e_G) = e_H \,,\end{aligned}$$

so $x^{-1}gx \in \ker(\phi)$ and hence $\ker \phi$ is normal.

Let $\phi(g), \phi(h) \in \mathrm{im}(\phi)$, then

$$\phi(g) * \phi(h)^{-1} = \phi(gh^{-1}) \in \mathrm{im}(\phi) \,.$$

Furthermore, $e_H = \phi(e_G) \in \mathrm{im}(\phi)$, so $\mathrm{im}(\phi)$ is non-empty, and is a subgroup of H. □

Definition 1.13. An *isomorphism* is a homomorphism which is also a bijection.

If a function $\phi : G \to H$ is an isomorphism, then the inverse function $\phi^{-1} : H \to G$ is too.

Definition 1.14. Two groups are *isomorphic* if there is an isomorphism between them. We write $G \cong H$.

We often consider isomorphic groups to be "the same", and do not distinguish between them. We should be aware that we are careless when doing this.

It is helpful to be able to break groups apart into smaller pieces. The following three isomorphism theorems allow us to do this in various ways. The first relates the kernel and image of an isomorphism.

Theorem 1.15. Let $\phi : G \to H$ be a homomorphism, then

$$G/\ker(\phi) \cong \mathrm{im}(\phi) \,.$$

Proof. Let

$$\begin{aligned} f : G/\ker(\phi) &\longrightarrow \mathrm{im}(\phi) \\ g\ker(\phi) &\longmapsto \phi(g) \,. \end{aligned}$$

First let us prove f is well-defined since it uses a coset representative. If $g\ker(\phi) = g'\ker(\phi)$, then $g^{-1}g' \in \ker(\phi)$ and so $\phi(g^{-1}g') = e_H$. Thus

$$e_H = \phi(g^{-1}g') = \phi(g^{-1}) * \phi(g') \,,$$

and so multiplying by $\phi(g)$ gives $\phi(g) = \phi(g')$, so the function is well-defined.

f is a homomorphism since

$$\begin{aligned} f(g\ker(\phi) \cdot g'\ker(\phi)) &= f(gg'\ker\phi) \\ &= \phi(gg') \\ &= \phi(g) * \phi(g') \\ &= f(g\ker(\phi)) * f(g'\ker(\phi)) \,. \end{aligned}$$

Let $h \in \mathrm{im}(\phi)$, then $h = \phi(g)$ for some g, so $h = f(g\ker\phi)$ is in the image of f. Therefore, f is surjective. Suppose that $f(g\ker(\phi)) = f(g'\ker(\phi))$, then $\phi(g) = \phi(g')$, so $\phi(g^{-1}g') = e_H$. Hence, $g^{-1}g' \in \ker(\phi)$ and $g\ker(\phi) = g'\ker(\phi)$, so f is injective. Therefore, f is an isomorphism. □

Example. Consider the function $\phi : \mathbb{C} \to \mathbb{C} \setminus \{0\}$ by $z \mapsto \mathrm{e}^z$. As $\mathrm{e}^{z+w} = \mathrm{e}^z \cdot \mathrm{e}^w$, ϕ defines a homomorphism

$$\phi : (\mathbb{C}, +, 0) \to (\mathbb{C} \setminus \{0\}, \times, 1)\,.$$

The existence of log shows that ϕ is surjective, and so $\mathrm{im}(\phi) = \mathbb{C} \setminus \{0\}$. The kernel is given by

$$\ker \phi = \{z \in \mathbb{C} \mid \mathrm{e}^z = 1\} = 2\pi\mathrm{i}\mathbb{Z}\,.$$

The conclusion is that

$$(\mathbb{C}/(2\pi\mathrm{i}\mathbb{Z}), +, 0) \cong (\mathbb{C} \setminus \{0\}, \times, 1)\,.$$

Theorem 1.16 (Second isomorphism theorem). Let $H \leq G$ and $K \lhd G$. Then $HK \coloneqq \{h \cdot k \mid h \in H, k \in K\}$ is a subgroup of G, and $H \cap K$ is a normal subgroup of G. Moreover,

$$\frac{HK}{K} \cong \frac{H}{H \cap K}\,.$$

Proof. Let $hk, h'k' \in HK$, then

$$h'k'(hk)^{-1} = h'k'k^{-1}h^{-1} = (h'h^{-1})(hk'k^{-1}h^{-1})\,.$$

The first term is in H, and the second term is $k'k^{-1} \in K$ conjugated by h, which is also in K. Therefore, $h'k'(hk)^{-1}$ lies in HK, and HK also contains e_G, so it is a subgroup.

Define

$$\begin{aligned} \phi : H &\longrightarrow G/K \\ h &\longmapsto hK\,. \end{aligned}$$

This is a homomorphism. The image of ϕ is the set of K-cosets which may be represented by an element of H, i.e.

$$\mathrm{im}(\phi) = \frac{HK}{K}\,.$$

The kernel of ϕ is

$$\ker(\phi) = \{h \in H \mid hK = eK\} = \{h \in H \mid h \in K\} = H \cap K\,.$$

As $H \cap K$ is the kernel of a homomorphism, it is normal in H. By the first isomorphism theorem,

$$\frac{H}{H \cap K} \cong \frac{HK}{K}\,.$$

Note that if $K \lhd G$, then there is a bijection between subgroups of G/K subgroups of G containing K, given by

$$\begin{aligned} \{\text{subgroups of } G/K\} &\longleftrightarrow \{\text{subgroups of } G \text{ which contains } K\} \\ X \leq \frac{G}{K} &\longmapsto \{g \in G \mid gK \in X\} \\ \frac{L}{K} \leq \frac{G}{K} &\longleftarrow\!\shortmid K \lhd L \leq G\,. \end{aligned}$$

This specialises to a bijection between normal subgroups as well.

$$\{\text{normal subgroups of } G/K\} \longleftrightarrow \{\text{normal subgroups of } G \text{ which contains } K\}\,.$$

Theorem 1.17 (Third isomorphism theorem). Let $K \leq L \leq G$ be normal subgroups of G, then

$$\frac{G/K}{L/K} \cong \frac{G}{L}\,.$$

Proof. Let

$$\begin{aligned} \phi : G/K &\longrightarrow G/L \\ gK &\longmapsto gL\,. \end{aligned}$$

If $gK = g'K$, then $g^{-1}g' \in K \in L$ so $gL = g'L$, so well-defined. It is a homomorphism, and onto.

$$\begin{aligned} \ker(\phi) = \{gK \in G/K \mid gL = eL\} &= L/K\,, \\ \Updownarrow & \\ g \in L & \end{aligned}$$

so done by 1st isomorphism theorem. □

Definition 1.18. A non-trivial group G is *simple* if its only normal subgroups are $\{e\}$ or G.

Lemma 1.19. An abelian group is simple $\iff$ it is C_p for some prime number p.

Proof.

($\Leftarrow$) If $H \leq C_p$, then $|H| \mid |C_p| = p$ by Lagrange's theorem, so $|H| = 1$ or p, so $H = \{e\}$ or C_p.

($\Rightarrow$) Let G be a simple abelian group. All subgroups of G are normal since

$$g^{-1}hg = g^{-1}gh = h \ \forall h \in H \lhd G\,.$$

Let $e \neq g \in G$, and $H = \{\dots, e, g, g^2, \dots\}$ is a normal subgroup. As $g \neq e$, $H \neq \{e\}$ so $H = G$ as G is simple, so G is cyclic. Then $G \cong (\mathbb{Z}, +, 0)$ or $G \cong C_n$. But $2\mathbb{Z} \lhd \mathbb{Z}$, so $\mathbb{Z}$ is not simple. If $n \mid m$, then $g^{n/m}$ generates a subgroup of C_n of order m, and it is normal. Therefore for C_n to be simple, only 1 and n can divide n. □

Theorem 1.20. If G is a finite group, then there are subgroups

$$G = H_1 > H_2 > \cdots > H_s = \{e\}\,.$$

Proof. If G is simple, let $H_2 = \{e\}$ and done.

If not, let $H_2 \lhd G$ be a proper normal subgroup of the largest order among all proper normal subgroups. Then claim G/H_2 is simple: if not, it has a proper $K \lhd G/H_2$. However, by the correspondence between normal subgroups of G/H_2 and normal subgroups of G containing H_2, $K \cong L/H_2$ for some $L \lhd G$ such that $H_2 \lhd L$. We find a proper normal subgroup of G with order strictly large than H_2, so contradiction.

So we found an $H_2 \lhd G$ with G/H_2 simple. Repeat this process to get the required sequence of normal subgroups. This process eventually stops as $|G|$ is finite. □

1.3 Actions and Permutations

Recall *permutation groups*

$$\begin{aligned} S_n &\coloneqq \text{symmetry group on } \{1, 2, \dots, n\} \\ &\equiv \text{group of bijections from } \{1, 2, \dots, n\} \text{ to itself.} \end{aligned}$$

Can define the *sign* function on S_n

$$\begin{aligned} \text{sign} : S_n &\longrightarrow (\{\pm 1\}, \times, +1) \\ \sigma &\longmapsto \begin{cases} +1 & \sigma = \text{even } \# \text{ permutations} \\ -1 & \sigma = \text{odd } \# \text{ permutations.} \end{cases} \end{aligned}$$

It is a homomorphism, so S_n has a normal subgroup

$$A_n \coloneqq \ker(\text{sign}) \lhd S_n$$

called the *alternating group*, and it has an index 2. More generally, for a set X, the *symmetry group* is

$$\text{Sym}(X) \coloneqq \{\sigma : X \to X \mid \sigma \text{ is a bijection}\}.$$

Definition 1.21. A group is a *permutation group* on X if it is a subgroup of $\text{Sym}(X)$. Say it is a permutation group of *degree* n if $|X| = n$.

Examples.

(i) $S_n = \text{Sym}(\{1, 2, \dots, n\})$ is a permutation group of degree n.

(ii) $D_{2n} = \text{Sym}(\text{vertices of a regular } n\text{-gon})$.

Definition 1.22. An *action* of a group G on a set X is a function

$$- * - : G \times X \to X$$

such that

(i) $\forall x \in X, g_i \in G,\ g_1 * (g_2 * x) = (g_1 g_2) * x$

(ii) $\forall x \in X,\ e * x = x$.

Lemma 1.23. An action G on X is the same as a homomorphism $\phi : G \to \text{Sym}(X)$.

Proof. Let $- * - : G \times X \to X$ be a function. Define $\phi(g) = g * - : X \to X$ be a function. Note

$$\begin{aligned}\phi(g^{-1}) \circ \phi(g)(x) &= g^{-1} * (g * x)\\ &= (g^{-1}g) * x\\ &= x\end{aligned}$$

$\forall x$, so $\phi(g^{-1})$ is inverse to $\phi(g)$, so $\phi(g)$ is a bijection. This defines a function $\phi : G \to \text{Sym}(X)$. Note

$$(\phi(h) \circ \phi(g))(x) = \phi(hg)(x)\ \forall x \in X, h, g \in G,$$

so ϕ is a homomorphism.

Conversely, let $\phi : G \to \text{Sym}(X)$ be a homomorphism. Let $g * x = \phi(g)(x)$, then

$$\begin{aligned}g * (h * x) &= \phi(g)(\phi(h)(x))\\ &= (\phi(g) \circ \phi(h))(x)\\ &= \phi(gh)(x)\\ &= (gh) * x,\end{aligned}$$

so it defines a group action. □

Definition 1.24. A permutation representation of a group G is a homomorphism $G \to \text{Sym}(X)$.

The lemma above has shown that a permutation representation is the same as a group action. The good thing about thinking of group actions as homomorphisms is that we can use all we know about homomorphisms on them.

Definition 1.25. For an action of G on X, we write

$$\begin{aligned}G^X &\coloneqq \text{im}(\phi : G \to \text{Sym}(X))\\ G_X &\coloneqq \ker(\phi : G \to \text{Sym}(X)).\end{aligned}$$

Proposition 1.26. $G^X \cong G/G_X$.

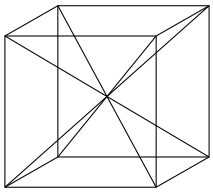
Proof. 1st isomorphism theorem. □

Examples.

(i)

$$\begin{aligned} G &= \text{symmetries of the cube} \\ X &= \{\text{diagonals of the cube}\} \\ G^X &= \mathrm{Sym}(X) \cong S_4 \\ G_X &= \{\text{id, sending each vertex to its opposite}\}. \end{aligned}$$

Have $\left|G^X\right| = 4! = 24$ and $|G_X| = 2$, so $|G| = 48$.



(ii) The group G acts on itself via $g * g' = g \cdot g'$. This corresponds to a $\phi : G \to \mathrm{Sym}(G)$.

$$G_X = \{g \in G \mid g * g' = g' \ \forall g' \in G\} = \{e\},$$

so $G = G/\{e\} \cong C^X \leq \mathrm{Sym}(G)$. This gives the Cayley's theorem. Every group is (isomorphic to) a subgroup of a symmetric group.

(iii) For $H \subseteq G$. Let $X = G/H$ set of left H-cosets. G acts on $X = G/H$ via $g * g'H = (gg')H$. This is well defined so we get $\phi : G \to \mathrm{Sym}(X)$.

Now consider $G_X = \ker(\phi)$. If $g \in G_X$, then for every $g_1 \in G$, we have $g * g_1H = g_1H$. This means $g_1^{-1}gg_1 \in H$. In other words, we have $g \in g_1Hg_1^{-1}$. This has to hold for all $g_1 \in G$, so

$$G_X \subseteq \bigcap_{g_1 \in G} g_1Hg_1^{-1}.$$

This argument is completely reversible: if $g \in \bigcap_{g_1 \in G} g_1Hg_1^{-1}$, then for each $g_1 \in G$, we know that $g_1^{-1}gg_1 \in H$ and hence $gg_1H = g_1H$. So $g * g_1H = g_1H$ and hence $g \in G_X$. Thus we have

$$\ker(\phi) = G_X = \bigcap_{g_1 \in G} g_1Hg_1^{-1}.$$

Since this is a kernel, it is a normal subgroup of G, and is contained in H. Starting with an arbitrary subgroup H, this allows us to generate a normal subgroup. (If we think about the construction, we see that this is the largest normal subgroup of G that is contained in H.)

Theorem 1.27. Let G be a finite group, $H \leq G$ be a subgroup of index n. Then there is a $K \lhd G$ with $K \leq H$ such that $|G/K| \mid n!$ and $n \mid |G/K|$.

Proof. We apply the previous example. Act by G on G/H, set $K = G_X \lhd G$. Then $G/G_X \cong G^X \leq \mathrm{Sym}(G/H) \cong S_n$, so by Lagrange's theorem, $|G/K| \mid |S_n| = n!$. We also have $K \leq H$, so $|G/K| \geq |G : H| = n$. □

Corollary. Let G be a non-abelian simple group, and H a proper subgroup of index $n > 1$. Then G is isomorphic to a subgroup of A_n, with $n \geq 5$.

Proof. Act by G on the set G/H gives a homomorphism $\phi : G \to \mathrm{Sym}(G/H)$. Then $G_X \lhd G$ is normal, so as G is simple, $G_X = \{e\}$ or $G_X = G$. As $g * (eH) \neq eH$ for $g \notin H$, which exists since the subgroup H is proper, the kernel of this action is not G, so $G_X = \{e\}$. Therefore, ϕ is injective, so $G = G/G_X \cong G^X \leq \mathrm{Sym}(G/H) \cong S_n$.

Now $A_n \lhd S_n$, so $G \cap A_n \lhd G$, so as G is simple, $G \cap A_n = \{e\}$ or $G \cap A_n = G$. If $G \cap A_n = \{e\}$, then

$$G = \frac{G}{G \cap A_n} \cong \frac{G \cdot A_n}{A_n} \leq \frac{S_n}{A_n} \cong C_2$$

by the second isomorphism theorem. Hence, $|G| \leq 2$. This is a contradiction as G is non-abelian. Therefore, we have

$$G \cap A_n = G \implies G \leq A_n \,.$$

A_1, A_2, A_3, A_4 have no non-abelian simple subgroups. □

Definition 1.28. Let G act on X. The *orbit* of $x \in X$ is

$$G \cdot x \coloneqq \{x' \in X \mid x' = g * x \text{ for some } g \in G\} \,.$$

The *stabiliser* of $x \in X$ is

$$G_x \coloneqq \{g \in G \mid g * x = x\} \,.$$

Lemma 1.29. G_x is a subgroup of G.

Proof. $e(x) = x$ by definition. For $g, h \in G_x$, $gh^{-1}(x) = g(h^{-1} * (x)) = g(x) = x$. □

Lemma 1.30. The orbits of an action partition X.

Proof. $\forall x \in X$, $x \in G \cdot x$ as $e * x = x$. So every x is in some orbit.

Then suppose $z \in G \cdot x$ and $z \in G \cdot y$, we have to show that $G \cdot x = G \cdot y$. We know that $z = g_1 * x = g_2 * y$ for some $g_1, g_2 \in G$, so $y = (g_2^{-1} g_1) * x$. For any $w = g_3(y) \in G \cdot y$, we have $w = (g_3 g_2^{-1} g_1) * x$, so w is also in $G \cdot x$. Thus $G \cdot y \subseteq G \cdot x$ and similarly $G \cdot x \subseteq G \cdot y$, so $G \cdot x = G \cdot y$. □

Let $g \in G$, $x \in X$. Each $g \in G$ gives us a member $g * x \in G \cdot x$, and conversely, every object in $G \cdot x$ arises this way. However, different elements in G can give us the same orbit. In particular, if $g \in G_x$, then hg and h give us the same object in $G \cdot x$. So we have a correspondence between things in $G \cdot x$ and members of G, up to G_x.

Theorem 1.31 (Orbit-stabiliser theorem). Let G act on X. Then for any $x \in X$, there is a bijection

$$\begin{aligned} \phi : G \cdot x &\longleftrightarrow G : G_x \\ g * x &\longmapsto\!\!\!\!\!\!\leftarrow\; gG_x \,, \end{aligned}$$

and in particular, if G is finite, then $|G \cdot x| = |G : G_x|$.

Proof. ϕ is well-defined since if $g \cdot G_x = h \cdot G_x$, then $h = gk$ for some $k \in G_x$, so $h * x = g * (k * x) = g * x$.

This map is injective because if $gG_x = hG_x$, then $G_x = g^{-1} h G_x$ so $g^{-1} h \in G_x$, then $g^{-1} h * x = x \implies g * x = h * x$, and this map is clearly surjective. The rest follows from Lagrange's theorem. □

1.4 Conjugacy Classes, Centralisers, Normalisers

Can define an action of G on the set G via

$$g * h = ghg^{-1} .$$

The function $\phi(g) : G \to G$ satisfies

$$\begin{aligned}\phi(g)(ab) &= gabg^{-1} = (gag^{-1})(gbg^{-1}) \\ &= \phi(a)\phi(b) ,\end{aligned}$$

so $\phi(g)$ is a homomorphism. It is also a bijection, with inverse $\phi(g^{-1})$. We can take the collection of all isomorphisms of G, and form a new group out of it.

Definition 1.32. The *automorphism group* of a group G is

$$\mathrm{Aut}(G) \coloneqq \{f \in \mathrm{Sym}(G) \mid f \text{ is a group isomorphism}\} .$$

This is a group under composition, with the identity map as the identity element.

It is a subgroup of $\mathrm{Sym}(G)$.

Definition 1.33. The *conjugacy class* of $g \in G$ is

$$\mathrm{Cl}_G(g) \coloneqq G \cdot g = \{h \in G \mid h = xgx^{-1} \ \forall x \in G\} .$$

The *centralisers* of $g \in G$ is

$$C_G(g) \coloneqq G_g = \{x \in G \mid xgx^{-1} = g\} ,$$

i.e. the set of all $x \in G$ which commute with g.

The *centre* of G is

$$Z(G) \coloneqq \ker(\phi) = \{x \in G \mid xgx^{-1} = g \ \forall g \in G\} = \bigcap_{g \in G} C_G(g) ,$$

i.e. the set of all $x \in G$ which commute with all $g \in G$.

Proposition 1.34.

$$|\mathrm{Cl}_G(x)| = |G : C_G(x)| = \frac{|G|}{|C_G(x)|} .$$

Proof. Orbit-stabiliser. □

Definition 1.35. The *normaliser* of $H \leq G$ is

$$N_G(H) \coloneqq \{g \in G \mid gHg^{-1} = H\} ,$$

i.e. the largest subgroup of G inside which H is normal.

Theorem 1.36. A_n is simple for all $n \geq 5$.

Proof.

- **Claim 1.** A_n is generated by 3-cycles.

  Proof. Every element of A_n is a product of evenly-many transpositions, so need to show that the product of two transpositions can be written in terms of 3-cycles.

  Let a, b, c, d be distinct.

– $(a\ b)(a\ b) = e$.
– $(a\ b)(b\ c) = (a\ b\ c)$
– $(a\ b)(c\ d) = (a\ c\ b)(a\ c\ d)$. □

- **Claim 2.** If $H \lhd A_n$ contains a 3-cycle, then it is A_n.

  Proof. We will show that if H contains a 3-cycle, it contains every 3-cycle, then done since A_n is generated by 3-cycles. Suppose $(a\ b\ c) \in H$ and want to show that $(1\ 2\ 3) \in H$. Since they have the same cycle shape, there is some $\sigma \in S_n$ such that $(a\ b\ c) = \sigma^{-1}(1\ 2\ 3)\sigma$.

  If σ is even, then $\sigma \in A_n$, so since $H \lhd A_n$, we have $(1\ 2\ 3) \in H$ and done. If σ is odd, take $\overline{\sigma} = \sigma(4\ 5) \in A_n$ (here comes the condition $n \geq 5$), then $\sigma^{-1}(1\ 2\ 3)\sigma = \overline{\sigma}^{-1}(1\ 2\ 3)\overline{\sigma} = (a\ b\ c)$ so done. □

Let $H \lhd A_n$. We want to show that H contains a 3-cycle, so it is A_n itself. We split into different cases.

(i) **Claim 3.** If H contains a σ which can be written as

$$\sigma = (1\ 2\ 3\ \mathit{dots}\ r) \cdot \tau$$

for $r \geq 4$ and τ arbitrary, then H contains a 3-cycle.

Proof. Let $\delta = (1\ 2\ 3)$ and consider

$$\underbrace{\underbrace{\sigma^{-1}}_{\in H} \cdot \underbrace{\delta^{-1} \cdot \sigma \cdot \delta}_{\in H}}_{\in H} = (r\ \ldots\ 2\ 1)(1\ 3\ 2)(1\ 2\ \ldots\ r)(1\ 2\ 3)$$
$$= (2\ 3\ r)\,,$$

where we used the assumption that $(1\ 2\ \ldots\ r)\tau$ are disjoint so τ commute with $(1\ 2\ \ldots\ r)$ and δ. □

(ii) **Claim 4.** If H contains a $\sigma = (1\ 2\ 3)(4\ 5\ 6) \cdot \tau$ disjoint, then it contains a 3-cycle.

Proof. Let $\delta = (1\ 2\ 4)$. Then

$$\sigma^{-1}\delta^{-1}\sigma\delta = (1\ 3\ 2)(4\ 6\ 5)(1\ 4\ 2)(1\ 3\ 2)(4\ 5\ 6)(1\ 2\ 4)$$
$$= (1\ 2\ 4\ 3\ 6) \in H\,.$$

This is a 5-cycle, so the previous case applies. □

(iii) **Claim 5.** If H contains a $\sigma = (1\ 2\ 3) \cdot \tau$ disjoint, then it contains a 3-cycle.

Proof. If τ is a product of 2-cycles, then

$$\sigma^2 = (1\ 2\ 3)^2 = (1\ 3\ 2)$$

is a three cycle. If τ is anything longer, then it falls into one of the previous cases. □

(iv) **Claim 6.** If H contains $\sigma = (1\ 2)(3\ 4)\tau$ disjoint, then it contains a 3-cycle.

Proof. If τ is a product of 2-cycles, let $\delta = (1\ 2\ 3)$ then

$$u = \sigma^{-1}\delta^{-1}\sigma\delta$$
$$= (1\ 2)(3\ 4)(1\ 3\ 2)(1\ 2)(3\ 4)(1\ 2\ 3)$$
$$= (1\ 4)(2\ 3) \in H\,.$$

Then let

$$v = (1\ 5\ 2)u(1\ 2\ 5) = (1\ 3)(4\ 5) \in H\,,$$

where we used $n \geq 5$ again. Consider

$$uv = (1\ 4)(2\ 3)(1\ 3)(4\ 5) = (1\ 2\ 3\ 4\ 5) \in H\,.$$

This is the first case so done.

If τ is longer, it fits in one of the previous cases. □

Combining these results, we are done. □

1.5 Finite p-groups.

We never seem to talk about things like the sum of orders to two subgroups. From this point of view, the simplest groups are those of prime orders, but they are all cyclic. The next simplest groups are those whose order is a power of prime.

Definition 1.37. A finite group is a p-group if $|G| = p^n$ for some prime number p and $n \geq 1$.

Theorem 1.38. If G is a finite p-group, then its centre $Z(G) = \{x \in G \mid xg = gx\ \forall g \in G\}$ is non-trivial.

Proof. Let G act on itself by conjugation. Each orbit of this action (which are precisely the conjugacy classes) has size dividing $|G| = p^n$, so is either a singleton, or has size divisible by p.

Since the conjugacy classes partition G, the sum of the sizes of the conjugacy classes is $|G|$. In particular,

$$|G| = \#\{\text{conjugacy classes of size 1}\} + \sum \text{orders of all other conjugacy classes}\,.$$

By the above discussion, the second term is divisible by p, as is $|G| = p^n$. Therefore, the number of conjugacy classes of size 1 is divisible by p. $\{e\}$ is a conjugacy class of size 1, so $\#\{\text{conjugacy classes of size 1}\} \geq p \geq 2$. There must be a conjugacy class $\{x\} \neq \{e\}$.

Then, $g^{-1}xg = x\ \forall g \in G$, i.e. $x \in Z(G)$, so $Z(G)$ is non-trivial. □

Corollary. Let G be a p-group of order p^n, $n \geq 2$. G is not simple.

Proof. $Z(G) \lhd G$. □

This allows us to prove interesting things about p-groups by induction on their orders, by considering the smaller p-group $G/Z(G)$. One way to do this is via the following lemma.

Lemma 1.39. For any group G, if $G/Z(G)$ is cyclic, then G is abelian.

In other words, if $G/Z(G)$ is cyclic, then it is trivial, since the centre of an abelian group is the abelian group itself.

Proof. Let the coset $gZ(G)$ be a generator of the cyclic group $G/Z(G)$, so every coset of $Z(G)$ is of the form $g^r Z(G)$. It follows that every element $x \in G$ must be in the form $g^r z$ for some $z \in Z(G)$, $r \in \mathbb{Z}$.

To show that G is abelian, let $x' = g^{r'}z'$ another element in G with some $z' \in Z(G)$ and $r' \in \mathbb{Z}$. As $z, z' \in Z(G)$, they commute with every element in G, so

$$xx' = g^r z g^{r'} z' = g^{r'} g^r z' z = g^{r'} z' g^r z = x' x\,,$$

and hence G is abelian. □

This lemma is particularly useful when applied to p-groups.

Corollary. If p is prime and $|G| = p^2$, then G is abelian.

Proof. Since $Z(G) \leq G$, its order must be 1, p or p^2. $Z(G)$ is non-trivial, so $|Z(G)| = p$ or p^2. If $|Z(G)| = p^2$, it is the whole group so G is trivially abelian. Otherwise, $|G/Z(G)| = p^2/p = p$ must be cyclic, so it must be cyclic, then G is again abelian. □

Theorem 1.40. Let G be a group of order p^a, where p is prime. Then G has a subgroup of order p^b for any $0 \leq b \leq a$.

Remark. This means that G has a subgroup of every possible order. This is not true for general groups, e.g. A_5 has an order of 60, but it has no subgroup of order 30 (as a subgroup of index 2 has to be normal, but A_5 is simple).

Proof. We induct on a. If $a = 1$, then $\{e\}$ and G are subgroups of order p^0 and p^1 so done.

Suppose $a > 1$ and we want to construct a subgroup of order p^b. If $b = 0$ then trivial. Otherwise, $Z(G)$ is non-trivial, so let $x \in Z(G)$, $x \neq e$. Since $\text{ord}(x) \mid |G|$, its order is a power of p. If it has an order p^c, then $x^{p^{c-1}}$ has order p. By renaming, suppose that x has order p, we have generated a subgroup $\langle x \rangle$ of order p. Since $x \in Z(G)$, $\langle x \rangle$ commutes with every $g \in G$, so $\langle x \rangle \lhd G$. Therefore, $G/\langle x \rangle$ is a group of order p^{a-1}.

Since this is a strictly smaller group, we may suppose by induction that $G/\langle x \rangle$ has a subgroup of any possible order. In particular, it has a subgroup L of order p^{b-1}. By the subgroup correspondence, there is some $K \leq G$ such that $\langle x \rangle \lhd K$ and $L = K/\langle x \rangle$. Then K has an order p^b. □

1.6 Finite Abelian Groups

Theorem 1.41 (Classification of finite abelian groups). Let G be a finite abelian group. Then there exists some $d_1, \dots, d_r$ such that

$$G \cong C_{d_1} \times C_{d_2} \times \dots C_{d_r}\,.$$

Moreover, we can choose the d_1 such that $d_{i+1} \mid d_i$ for each i, in which case this expression is unique.

We will prove this in chapter 3 as a special case of the classification of modules over certain rings.

Example. The abelian groups of order 8 are C_8, $C_4 \times C_2$, $C_2 \times C_2 \times C_2$.

Sometimes the decomposition given by this theorem is not the most useful form. To get a nicer decomposition, we can use the following lemma.

Lemma 1.42. If n and m are coprime, then $C_{mn} \cong C_m \times C_n$.

Remark. This is essentially the Chinese remainder theorem, and this formulation is how you should think of that theorem.

Proof. It suffices to find an element of order nm in $C_m \times C_n$, then since $C_m \times C_n$ has order mn, it must be cyclic and hence isomorphic to C_{mn}.

Let $g \in C_m$ have order m and $h \in C_n$ have order n, and consider the element $(g, h) \in C_m \times C_n$. Suppose the order of (g, h) is k, then $(g, h)^k = (e, e)$. Hence $(g^k, h^k) = (e, e)$. So $m \mid k$ and $n \mid k$. As m, n are coprime, this means that $mn \mid k$. As $k = \text{ord}(g, h)$ and $(g, h) \in C_m \times C_n$ is a group of order mn, we must have $k \mid mn$. So $k = mn$. □

Corollary. For any finite abelian group G, we have

$$G \cong C_{d_1} \times C_{d_2} \times \cdots \times C_{d_r}\,,$$

where each d_i is a prime power.

Proof. From the classification theorem, iteratively apply the previous lemma to break down each component up into prime powers. □

1.7 Sylow's Theorems

Definition 1.43. Let G be a finite group of order $p^a \cdot m$, with p prime and $p \nmid m$. A *Sylow p-subgroup* of G is a subgroup $P \leq G$ of order p^a.

Theorem 1.44 (Sylow's theorems). Let G be a finite group of order $p^a \cdot m$, with p prime and $p \nmid m$.

(i) The set
$$\mathrm{Syl}_p(G) \coloneqq \{P \leq G \mid |P| = p^a\}$$
of Sylow p-subgroups of G is non-empty.

(ii) All elements of $\mathrm{Syl}_p(G)$ are conjugate in G.

(iii) The number $n_p = \left|\mathrm{Syl}_p(G)\right|$ of Sylow p-subgroups satisfies $n_p \equiv 1 \pmod p$ and $n_p \mid |G|$, and hence $n_p \mid m$.

These are sometimes known as Sylow's first/second/third theorem respectively.

Proof.

(i) First show that $\mathrm{Syl}_p(G) \neq \varnothing$. Let Ω be the set of subsets of G with p^a elements. G acts on this via
$$g * \{x_1, x_2, \ldots, x_{p^a}\} = \{g \cdot x_1, g \cdot x_2, \ldots, g \cdot x_{p^a}\}\,.$$
Let $\Sigma \subseteq \Omega$ be an orbit of this action.

If $\{x_1, \ldots, x_{p^a}\} \in \Sigma$, then for any $g \in G$,
$$(g \cdot x_1^{-1}) * \{x_1, \ldots, x_{p^a}\} = \{g, \ldots\} \in \Sigma$$
contains g, so any element of the group lies in some element of Σ, so
$$|\Sigma| \geq \frac{|G|}{p^a} = m\,.$$
If $|\Sigma| = m$, then by orbit-stabiliser theorem, the stabiliser of any $\{x_1, \ldots, x_{p^a}\}$ has index m, so has order p^a, and thus is a Sylow p-subgroup.

We would then like to show not every orbit can have size $> m$. If $|\Sigma| > m$, then as $|\Sigma| \mid |G| = p^a m$ by orbit-stabiliser theorem, we must have $p \mid |\Sigma|$. Our strategy is to show $|\Omega| \not\equiv 0 \pmod p$, so since Ω is the disjoint union of all orbits, not every orbit can have size $> m$.

This is done by calculating
$$|\Omega| = \binom{p^a m}{p^a} = \prod_{j=0}^{p^a-1} \frac{p^a m - j}{p^a - j}\,.$$
As $j < p^a$, the largest power of p dividing $p^a m - j$ is the largest power of p dividing j. Similarly, the largest power of p dividing $pa - j$ is also the largest power of p dividing j. So we have the same power of p on top and bottom for each term in the product, so they cancel and the result is not divisible by p.

(ii) We will prove something stronger. Let p be a Sylow p-subgroup, and Q be a p-subgroup (with order $|Q| = p^b$ where $b \leq a$). We will show that Q may be conjugated into P, i.e. $g^{-1}Qg \leq P$ for some $g \in G$.

Let Q act on the set of left cosets G/P via $q * gP = (qg)P$. By orbit-stabiliser, the size of each orbit divides $|Q|$, so each orbit has size 1 or divisible by p. But $|G/P| = \frac{p^a \cdot m}{p^a} = m$ is coprime to p, so some orbit has size 1. Let gP has size 1, then

$$qgP = gP \ \forall q \in Q \iff g^{-1}qg \in P \iff g^{-1}Qg \leq P\,.$$

(iii) G acts on the set of $\mathrm{Syl}_p(G)$ by conjugation. By (ii), the action has a single orbit, The orbit-stabiliser theorem applied to the orbit shows that $n_p = \left|\mathrm{Syl}_p(G)\right|$ divides $|G|$. This is the second claim.

Let $P \in \mathrm{Syl}_p(G)$ and act on $\mathrm{Syl}_p(G)$ by conjugation. Note $\{P\}$ is an orbit of this action with size 1. We will show that the other orbits have sizes divisible by p. By orbit-stabiliser, all orbits have size either 1 or divisible by p. Need to show that there are no other orbit of size 1. Suppose $\{Q\}$ is such an orbit, i.e. $\forall p \in P$, $p^{-1}Qp = Q$, so

$$\begin{aligned} P &\leq N_G(Q) = \text{normaliser of } Q \text{ in G} \\ &= \{g \in G \mid g^{-1}Qg = Q\}\,. \end{aligned}$$

Now $N_G(Q)$ is itself a group, and we can look at its Sylow p-subgroups. We know that $Q \leq N_G(Q) \leq G$, so $p^a \mid |N_G(Q)| \mid p^a m$. Thus p^a is the biggest power of p that divides $|N_G(Q)|$, so Q is a Sylow p-subgroup of $N_G(Q)$.

By (ii), Q is conjugated to P inside $N_G(Q)$, but the only conjugate of Q in $N_G(Q)$ is Q tautologically, so $Q = P$.

So the original action has exactly 1 orbit of size 1, and the other have sizes divisible by p, so $n_p = \left|\mathrm{Syl}_p(G)\right| \equiv 1 \pmod p$. □

Lemma 1.45. If there is a unique Sylow p-subgroup, i.e. $n_p = 1$, then it is normal in G.

Proof. Let P be the unique Sylow p-subgroup, and let $g \in G$. Then by Sylow's second theorem, $g^{-1}Pg$ is P itself, so P is normal. □

Corollary. Let G be a non-abelian simple group and prime number $p \mid |G|$. Then $|G| \mid \frac{(n_p)!}{2}$ and $n_p \geq 5$.

Proof. G acts on the set of Sylow p-subgroups $\mathrm{Syl}_g(p)$ by conjugation, giving the permutation representation

$$\phi : G \to \mathrm{Sym}(\mathrm{Syl}_p(G)) \cong S_{n_p}\,.$$

We know $\ker(\phi) \lhd G$, but G is simple, so $\ker\phi$ is either $\{e\}$ or G.

If $\ker(\phi) = G$, then all Sylow p-subgroups are normal. This contradicts with G being simple, so $\ker(\phi) = \{e\}$, and so G is isomorphic to a subgroup of S_{n_p}. Now consider

$$G \xrightarrow{\phi} S_{n_p} \xrightarrow{\text{sign}} \{\pm 1\}\,.$$

If this is injective, then the kernel is an index 2 normal subgroup of G, again contradicts with G being simple. Therefore we can only have $\ker \text{sign} \circ \phi$ being the whole G, so $G \cong \mathrm{im}(\phi) \leq A_{n_p}$, and $|G| \mid \frac{(n_p)!}{2}$.

For the final statement, can check that $A_1, \ldots, A_4$ has no non-abelian simple subgroups. □

Example. Let $|G| = 1000 = 2^3 \cdot 5^3$, then $n_5 \equiv 1 \pmod 5$ and $n_5 \mid 8 = 2^3$. Then we can only have $n_5 = 1$, so G has a normal subgroup of order 5^3, and G is not simple.

Example. Let $|G| = 132 = 2^2 \cdot 3 \cdot 11$. Suppose G is simple. Have $n_1 1 \equiv 1 \pmod{11}$, $n_{11} \mid 2^2 \cdot 3 = 12$, so $n_{11} = 1$ or 12. As G is assumed simple, $n_{11} \neq 1$, so $n_{11} = 12$.

Similarly, $n_3 \equiv 1 \pmod 3$ and $n_3 \mid 2^2 \cdot 11 = 44$, so $n_3 = 1, 4$ or 22. G is simple so $n_3 \neq 1$. If $n_3 = 4$, then the corollary gives $|G| = 132 \mid \frac{4!}{2} = 12$, a contradiction so n_3 must be 22.

Every Sylow 11-subgroup is cyclic, so contains $11 - 1 = 10$ elements of order 11. These subgroups only intersect at $\{e\}$, so there are $11 \times 10 = 120$ elements of G of order 11.

Every Sylow 3-subgroup is cyclic, so contains 2 elements of order 3. They only intersect in $\{e\}$ so there are $22 \times 2 = 44$ elements of order 3.

We have found too many elements of G. This cannot happen, so a group of order 132 is never simple.

Example. $\mathrm{GL}_n(\mathbb{Z}/p) = \{\text{invertible } n \times n \text{ matrices with entries in } \mathbb{Z}/p\}$, p is a prime number. What is the order of this group? Giving a matrix $A \in \mathrm{GL}_n(\mathbb{Z}/p)$ is the same as giving n linearly independent vectors in the vector space $(\mathbb{Z}/p)^n$. We can pick the first vector to be anything except zero, so there are $p^n - 1$ ways of choosing the first vector.Next,we need to pick the second vector, which can be anything that is not in the span of the first vector, so there are $p_n - p$ways of choosing the second vector. Continuing in this way we have

$$(p^n - 1)(p^n - p)(p^n - p^2) \dots (p^n - p^{n-1}) = p^{1+2+\dots+(n-1)}(p^n - 1)(p^{n-1} - 1) \dots (p - 1)\,.$$

So $p^{\binom{n}{2}}$ is the largest power of p dividing $|\mathrm{GL}_n(\mathbb{Z}/p)|$.

To give a Sylow p-subgroup of $\mathrm{GL}_n(\mathbb{Z}/p)$, we consider the subgroup of matrices of the following form

$$U \coloneqq \left\{ \begin{pmatrix} 1 & * & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ 0 & 0 & 1 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathrm{GL}_n(\mathbb{Z}/p) \right\}.$$

We have $|U| = p^{\binom{n}{2}}$, so it is a Sylow p-subgroup.

Example. $\mathrm{GL}_2(\mathbb{Z}/p)$ has order $p(p^2-1)(p-1) = p(p-1)^2(p+1)$. Suppose $l \mid p-1$ and $l^3 \nmid |\mathrm{GL}_2(\mathbb{Z}/p)|$. $l \neq p$ is a prime number, so there must be a subgroup of order l^2. Note

$$(\mathbb{Z}/p)^\times = \{x \in \mathbb{Z}/p \mid \exists y \text{ such that } xy \equiv 1 \pmod p\} \cong C_{p-1}\,,$$

so as $l \mid p - 1$, there is a subgroup $C_l \leq C_{p-1} \cong (\mathbb{Z}/p)^\times$. We immediately find a subgroup

$$C_l \times C_l \leq (\mathbb{Z}/p)^\times \times (\mathbb{Z}/p)^\times \leq \mathrm{GL}_2(\mathbb{Z}/p)\,,$$

where the second inclusion is the diagonal matrices, identifying

$$(a, b) \longleftrightarrow \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},$$

so this is a Sylow l-subgroup.

A non-examinable fact:

Theorem 1.46 (Feit–Thompson theorem). If G is a non-abelian finite group of odd order, then it is not simple.

2 Rings

2.1 Definitions and Examples

Definition 2.1. A quintuple $(R, +, \cdot, 0_R, 1_R)$ forms a *ring* if.

(i) $(R, +, 0_R)$ is an abelian group.

(ii) the operation $- \cdot - : R \times R$ is associative and satisfies

$$1_R \cdot x = x \cdot 1_R = x \quad \forall x \in R\,.$$

(iii) multiplication distributes over addition

$$(r_1 + r_2) \cdot r_3 = r_1 \cdot r_3 + r_2 \cdot r_3$$
$$r_1 \cdot (r_2 + r_3) = r_1 \cdot r_2 + r_1 \cdot r_3\,.$$

If R is a ring and $r \in R$, we write $-r$ for the inverse to r in the group $(R, +, 0_R)$, and we write $r - s$ to mean $r + (-s)$ and so on.

Some authors do not insist on the existence of the multiplicative identity, but we do.

Since we can add and multiply two elements, by induction, we can add and multiply any finite number of elements. However, the notions of infinite sum and product are not defined: it does not make sense to ask if an infinite sum converges.

Definition 2.2. A ring is *commutative* if $a \cdot b = b \cdot a$ $\forall a, b \in R$.

From now onwards, all the rings in this course are commutative.

Definition 2.3. If $(R, +, \cdot, 0_R, 1_R)$ is a ring, a $S \subseteq R$ is a *subring* if $0_R, 1_R \in S$ and $+$ and $\cdot$ make S into a ring.

Example. We have subrings

- $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$.
- $\mathbb{Z}[\mathrm{i}] = \{a + b\mathrm{i} \in \mathbb{C} \mid a, b \in \mathbb{Z}\} \leq \mathbb{C}$, known as *Gaussian integers*.
- $\mathbb{Q}[\sqrt{2}] = \{a + \sqrt{2}b \in \mathbb{R} \mid a, b \in \mathbb{Q}\} \leq \mathbb{R}$.

Definition 2.4. An $r \in R$ is called a *unit* if there is a $s \in R$ such that $r \cdot s = 1_R$.

If all non-zero elements in $\mathbb{R}$ are units, then $\mathbb{R}$ is called a *field*.

Example. $0_R = 0_R + 0_R$, so $r \cdot 0_R = r \cdot (0_R + 0_R) = r \cdot 0_R + r \cdot 0_R$, so $r \cdot 0_R = 0_R$. Also, $r \cdot 1_R = r$.

Note $(\{0\}, +, \cdot, 0, 0)$ is a ring in which $1_R = 0_R$. It is the only ring in which $1_R = 0_R$. (However, it is often a counterexample to incautious claims about rings.)

Definition 2.5. Let R, S be rings, the *product* $R \times S$ is a ring via

$$(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$$
$$(r_1, s_1) \cdot (r_2, s_2) = (r_1 \cdot r_2, s_1 \cdot s_2)$$

The zero element and one element are

$$0_{R \times S} = (0_R, 0_S) \qquad 1_{R \times S} = (1_R, 1_S)\,.$$

Note we have

$$(x, 0_S) \cdot (0_R, y) = 0_{R\times S}\,.$$

Definition 2.6. Let R be a ring. A *polynomial* over R is an expression

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

with $a_i \in R$ and X^i are formal symbols.

Definition 2.7. The *degree* of the polynomial is the largest k such that $a_k \neq 0$.

If a polynomial of degree k has $a_k = 1$, then the polynomial is called *monic*.

Definition 2.8. Let $R[X]$ denote the set of all polynomials with coefficients in R. $R[X]$ forms a *polynomial ring*: if $f = a_0 + a_1X + \cdots + a_nX^n$ and $g = b_0 + b_1X + \cdots + b_kX^k$ are polynomials over R, then

$$f + g = \sum_{r=0}^{\max\{n,k\}} (a_r + b_r)X^r\,,$$

and

$$f \cdot g = \sum_{i=0}^{n+k} \left(\sum_{j=0}^{i} a_j b_{i-j} \right) X^i\,.$$

$0_{R[X]} = 0_R$ and $1_{R[X]} = 1_R$, both considered as constant polynomials.

Remark. A polynomial with coefficients in R is just a sequence of elements of R, interpreted as the coefficients of some formal symbols. While it does indeed induce a function from R to R in the obvious way, we shall not identify the polynomial with the function it induces, since different polynomials can give rise to the same function.

For example, in $\mathbb{Z}/2\mathbb{Z}[X]$, $f = X^2 + X$ is not the zero polynomial, since its coefficients are not zero. However, $f(0) = 0$ and $f(1) = 0$, so the function induced by f is identically zero.

Definition 2.9. $R[[X]]$ is the ring of formal *power series*

$$f = a_0 + a_1X + a_2X^2 + \ldots$$

with coefficients $a_i \in R$ and the same formulae for addition and multiplication.

We do not ask whether the sum converges or not, because it is not a sum: it is a formal symbol which can be manipulated similarly to a convergent infinite sum.

Definition 2.10. $R[X, X^{-1}]$ is the ring of *Laurent polynomials*

$$f_i = \sum_{i\in\mathbb{Z}} a_iX^i\,,$$

where $a_i \in R$ and there are only finitely non-zero a_i. Add and multiply as above, with $X \cdot X^{-1} = 1$.

We can also think of Laurent series, but we have to be careful: we allow infinitely many positive coefficients, but only finitely many negative ones. Or else, in the formula for multiplication, we will have an infinite sum of elements in R, which is not defined.

Example. If R is a ring and X is a set, then $F = \{f : X \to R\} =$ all functions $X \to R$ is a ring, via

$$(f +_F g)(x) = f(x) +_R g(x)$$
$$(f \cdot_F g)(x) = f(x) \cdot_R g(x)\,.$$

E.g. $\{\text{all } f : \mathbb{R} \to \mathbb{R}\} \supset \{\text{continuous } f : \mathbb{R} \to \mathbb{R}\} \supset \mathbb{R}[X]$

2.2 Homomorphisms, Ideals, Quotients and Isomorphisms

Definition 2.11. A function $\phi : R \to S$ between rings is a *homomorphism* if

(i) $\phi(r_1 +_R r_2) = \phi(r_1) +_S \phi(r_2)$, $\phi(0_R) = 0_S$

$$(\iff \phi : (R, +_R, 0_R) \to (S, +_S, 0_S) \text{ is a group homomorphism}).$$

(ii) $\phi(r_1 \cdot_R r_2) = \phi(r_1) \cdot_S \phi(r_2)$.

(iii) $\phi(1_R) = 1_S$.

Definition 2.12. If a homomorphism $\phi : R \to S$ is a bijection, then it is called an *isomorphism*.

Definition 2.13. The *kernel* of a homomorphism $\phi : R \to S$ is

$$\ker(\phi) \coloneqq \{r \in R \mid \phi(r) = 0_S\}.$$

The *image* of a homomorphism $\phi : R \to S$ is

$$\operatorname{im}(\phi) \coloneqq \{s \in S \mid s = \phi(r) \text{ for some } r \in R\}.$$

Lemma 2.14. A homomorphism $\phi : R \to S$ is injective if and only if $\ker \phi = \{0_R\}$.

Proof. A ring homomorphism is in particular a group homomorphism $\phi : (R, +, 0_R) \to (S, +, 0_S)$ of abelian groups. □

Definition 2.15. A subset $I \subseteq R$ is called an *ideal*, $I \lhd R$, if

(i) I is a subgroup of $(R, +, 0_R)$

(ii) If $x \in I$, $r \in R$, then $x \cdot r \in I$. (strong closure)

We say an ideal $I \lhd R$ is *proper* if $I \neq R$.

Lemma 2.16. If $\phi : R \to S$ is a ring homomorphism, then $\ker(\phi)$ is an ideal in R.

Proof. $\ker(\phi)$ is a subgroup of $(R, +_R, 0_R)$, showing (i).

If $x \in \ker(\phi)$ and $r \in R$, then

$$\begin{aligned}\phi(r \cdot x) &= \phi(r) \cdot \phi(x) = \phi(r) \cdot 0_S \\ &= 0_S,\end{aligned}$$

so $r \cdot x \in \ker(\phi)$ too, showing (ii). □

Examples.

(i) If $I \lhd R$ and $1_R \in I$, then for any $r \in R$, $r = r \cdot 1_R \in I$, so $I = R$. So I is proper $\iff 1_R \notin I$.

(ii) If $I \lhd R$ and $u \in R$ is a unit, then there is a $v \in R$ such that $v \cdot u = 1_R$, so if $u \in I$, then $1_R = u \cdot v \in I$, so $I = R$. So I is proper $\iff$ all units are outside of I.

(iii) In $\mathbb{Z}$, all ideals have the form $n\mathbb{Z} = \{\dots, -2n, -n, 0, n, 2n, \dots\}$.

Proof. Certainly, $n\mathbb{Z}$ is an ideal.

Let $I \lhd \mathbb{Z}$ be an ideal, and n be the smallest strictly positive element in I (If there isn't one, then $I = \{0\} = 0\mathbb{Z}$). Claim that $I = n\mathbb{Z}$. Certainly, $n\mathbb{Z} \subseteq I$. If they are not equal, pick an element $m \in I \setminus n\mathbb{Z}$. Using Euclidean algorithm:

$$m = qn + r \text{ with } 0 \leq r < n.$$

Then $r = \underbrace{m}_{\in I} - \underbrace{q \cdot n}_{\in I} \in I$, but $r < n$, gives a contradiction. □

Definition 2.17. For $a \in R$, the ideal *generated* by a is

$$(a) \coloneqq \{a \cdot r \in R \mid r \in R\} \lhd R\,.$$

Generally, the ideal generated by $a_1, \dots a_k \in R$ is

$$(a_1, \dots, a_k) \coloneqq \{a_1 r_1 + \dots a_k r_k \mid r_i \in R\} \lhd R\,.$$

More generally, the ideal generated by $A \subseteq R$ is

$$(A) \coloneqq \left\{ \sum_{a \in A} a \cdot r_a \;\middle|\; r_a \in R,\ \text{only finitely many non-zero} \right\}.$$

Definition 2.18. If an ideal $I = (a)$ for some $a \in R$, then I is called a *principal ideal*.

Example. Examples of principal ideals:

(i) $n\mathbb{Z} = (n) \lhd \mathbb{Z}$.

(ii) $(X) = \{\text{polynomials with constant term } 0\} \lhd R[X]$.

Definition 2.19. Let I be an ideal of R. A *quotient ring* is the set of cosets $\{r + I \mid r \in R\}$ with operations

$$\begin{aligned}(r_1 + I) + (r_2 + I) &= (r_1 + r_2) + I \\ (r_1 + I) \cdot (r_2 + I) &= r_1 \cdot r_2 + I\,,\end{aligned}$$

and $0_{R/I} = 0_R + I$, $1_{R/I} = 1_R + I$.

Proposition 2.20. The quotient ring is a ring, and the function

$$\begin{aligned} R &\longrightarrow R/I \\ r &\longmapsto r + I \end{aligned}$$

is a ring homomorphism.

Proof. Already know $(R/I, +, 0_{R/I})$ is an abelian group, so addition well-defined. Have

$$\begin{cases} r_1 + I = r_1' + I \\ r_2 + I = r_2' + I \end{cases} \implies \begin{cases} r_1 = r_1' + a_1 & a_1 \in I \\ r_2 = r_2' + a_2 & a_2 \in I\,. \end{cases}$$

So $r_1 r_2 = (r_1' + a_1)(r_2' + a_2) = r_1' r_2' + \underbrace{r_1' a_2}_{\in I} + \underbrace{a_1 r_2'}_{\in I} + \underbrace{a_1 a_2}_{\in I}$, so $r_1 r_2 + I = r_1' r_2' + I$. Multiplication is well defined. The other axioms are inherited from R. □

Examples.

(i) Have ideal $n\mathbb{Z} \lhd \mathbb{Z}$, so get quotient rings $\mathbb{Z}/n\mathbb{Z}$. The elements are $0 + n\mathbb{Z}, 1 + n\mathbb{Z}, 2 + n\mathbb{Z}, \cdots, (n-1) + n\mathbb{Z}$. Addition and multiplication are arithmetics modulo n.

(ii) $(X) \lhd \mathbb{C}[X]$. Elements of $\mathbb{C}[X]/(X)$ are

$$a_0 + \underbrace{a_1 X}_{\in (X)} + \underbrace{a_2 X^2}_{\in (X)} + \dots \underbrace{a_n X^n}_{\in (X)} + (X) = a_0 + (X)\,.$$

If $a_0 + (X) = b_0 + (X)$, then $a_0 - b_0$ is divisible by (X), so $a_0 - b_0 = 0$. Elements are uniquely written as $a_0 + (X)$.

$$\begin{aligned} \mathbb{C}[X]/(X) &\longleftrightarrow \mathbb{C} \\ p(X) + (X) &\longmapsto p(0) \\ a + (X) &\longleftarrow\!\shortmid\ a \end{aligned}$$

is a ring isomorphism.

Proposition 2.21 (Euclidean algorithm for polynomials). Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[X]$. Then we can write

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

with $\deg r < \deg g$.

Proof. Let $\deg f = n$, so

$$f(X) = \sum_{i=0}^{n} a_i X^i\,,\ a_n \neq 0\,.$$

Let $\deg g = m$, so

$$g(X) = \sum_{j=0}^{m} b_j X^j\,,\ b_m \neq 0\,.$$

If $n < m$, then can take $q(X) = 0$, $r(X) = f(X)$ then done.

If $n \geq m$, proceed by induction on n. Let

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

which exists because $\mathbb{F}$ is a field. The coefficient of X^n is $a_n - a_n \cdot b_m^{-1} \cdot b_m = 0$, so $f_1(X)$ has degree $< n$. If we had $n = m$, then

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

then $\deg f_1 < \deg f$, so done. If $n > m$, then by induction

$$f_1(X) = q_1(X) g(X) + r_1(X)\,,\ \deg r_1 < m\,,$$

so

$$f(X) = (a_n b_m^{-1} X^{n-m} + q_1(X)) g(X) + r_1(X)$$

as required. □

Example. Consider $(X^2+1) \lhd \mathbb{R}[X]$, $R = \mathbb{R}[X]/(X^2+1)$.

Given $f(X) + (X^2+1)$, write $f(X) = q(X)(X^2+1) + r(X)$ using Euclidean algorithm with $\deg r < 2$, so $f(X) + (X^2+1) = r(X) + (X^2+1)$. So all elements have the form $a + bX + (X^2+1)$. If $a' + b'X + (X^2+1)$ is the same coset, then

$$a + bX - (a' + b'X) = h(X)(X^2+1)\,.$$

Both sides must vanish considering their degrees. The representation is therefore unique.

What we've got is that every element in R is of the form $a + bX$, and $X^2 + 1 = 0$, i.e. $X^2 = -1$. This sounds like the complex numbers, just that we are calling it X instead of i.

Let

$$\begin{aligned} \phi : \mathbb{R}[X]/(X^2+1) &\longleftrightarrow \mathbb{C} \\ a + bX + (X^2+1) &\longmapsto a + \mathrm{i}b\,. \end{aligned}$$

Clear that ϕ preserves addition. For multiplication,

$$\begin{aligned} \phi((a + bX + (X^2+1))(c + dX + (X^2+1))) &= \phi(ac + (ad+bc)X + bdX^2 + (X^2+1)) \\ &= \phi((ac - bd) + (ad+bc)X + (X^2+1)) \\ &= (ac - bd) + (ad+bc)\mathrm{i} \\ &= (a + \mathrm{i}b)(c + \mathrm{i}d)\,. \end{aligned}$$

So ϕ is a ring homomorphism.

Exercise. Prove $\mathbb{R}[X]/(X^2-1) \cong \mathbb{R} \times \mathbb{R}$.

Theorem 2.22 (First isomorphism theorem). Let $\phi : R \to S$ be a ring homomorphism. Then $\ker(\phi) \lhd R$, $\operatorname{im}\phi \leq S$ and $\phi : R/\ker(\phi) \cong \operatorname{im}\phi$ as rings.

Proof. Have seen that $\ker\phi$ is an ideal. Also, $\operatorname{im}\phi$ is a subgroup of $(S, +_S, 0_S)$ by the 1st isomorphism for groups, and $\phi(r_1)\cdot\phi(r_2) = \phi(r_1 \cdot r_2) \in \operatorname{im}\phi$, and $1_S = \phi(1_R)$, $0_S = \phi(0_R)$. $\operatorname{im}(\phi)$ is a subgroup of S. Let

$$\begin{aligned} \Phi : R/\ker\phi &\longrightarrow \operatorname{im}\phi \\ r + \ker\phi &\longmapsto \phi(r)\,. \end{aligned}$$

Check that it is multiplicative. □

Theorem 2.23 (Second isomorphism theorem). Let $R \leq S$, $J \lhd S$, then $J \cap R \lhd R$,

$$\frac{R+J}{J} \coloneqq \{r + J \mid r \in R\} \leq \frac{S}{J}$$

and

$$\frac{R}{J\cap R} \cong \frac{R+J}{J}$$

as rings.

Proof. Let

$$\begin{aligned} \phi : R &\longrightarrow S/J \\ r &\longrightarrow r + J\,, \end{aligned}$$

a ring homomorphism.

$$\begin{aligned} \ker(\phi) &= \{r \in R \mid r + J = 0 + J \text{ i.e. } r \in J\} = R \cap J \\ \operatorname{im}(\phi) &= \{r + J \,|\, r \in R\} = \frac{R+J}{J} \leq \frac{S}{J}. \end{aligned}$$

Apply the first isomorphism theorem. □

Just as for rings, there is a correspondence between

$$\begin{aligned} \{\text{subrings of } R/I\} &\longleftrightarrow \{\text{subrings of } R \text{ containing } I\} \\ S/I \leq R/I &\longleftarrow\!\shortmid I \lhd S \leq R \\ L \leq R/I &\longmapsto \{r \in R \mid r + I \in L\}\,. \end{aligned}$$

Similarly,

$$\{\text{ideals of } R/I\} \longleftrightarrow \{\text{ideals of } R \text{ containing } I\}\ .$$

Theorem 2.24 (Third isomorphism theorem). Let $I \lhd R$, $J \lhd R$, $I \subseteq J$, then $J/I \lhd R/I$ and

$$\frac{R/I}{J/I} \cong \frac{R}{J}\,.$$

Definition 2.25. Consider

$$\begin{aligned} \phi : R/I &\longrightarrow R/J \\ r + I &\longmapsto r + J\,, \end{aligned}$$

a ring homomorphism. It is onto, and $\ker\phi = \{r + I \in R/I \mid r + J = 0 + J\} = J/I$. Apply the first isomorphism theorem. □

Example. For any ring R, there is a homomorphism

$$\begin{aligned} c : \mathbb{Z} &\longrightarrow R \\ 1 &\longmapsto 1_R \\ n &\longmapsto \underbrace{1_R + \cdots + 1_R}_{n \text{ times}}, \ n \geq 0 \,. \end{aligned}$$

The first isomorphism theorem: $\ker(c) = n\mathbb{Z}$ for some n, and $\mathbb{Z}/n\mathbb{Z} \cong \text{im}(c) \leq R$. n is called the *characteristic* of R.

$\mathbb{Z}, \mathbb{R}, \mathbb{Q}, \mathbb{C}$ have characteristic 0. $\mathbb{Z}/n\mathbb{Z}$ has characteristic n.

2.3 Integral Domains, Field of Fractions, Maximal and Prime Ideals

Many rings can be completely nothing like $\mathbb{Z}$. For example, in $\mathbb{Z}$, we know that if $a, b \neq 0$, then $a \cdot b \neq 0$. We start with the most fundamental property that the product of two nonzero elements is non-zero. We will almost exclusively work with rings that satisfy this property (except some simple ones like $\mathbb{Z}/n\mathbb{Z}$).

Definition 2.26. A non-zero ring is called an *integral domain* if whenever $a \cdot b = 0$, $a = 0$ or $b = 0$.

A *zero divisor* in a non-integral domain is an element that violates this property, i.e. $a \in R$ is a zero divisor if $\exists b \neq 0$ such that $a \cdot b = 0$.

Example. All fields are integral domain. $a \cdot b = 0$ and $b \neq 0$, then b^{-1} exists so $0 = (a \cdot b) \cdot b^{-1} = a$. So $\mathbb{Q}$ and $\mathbb{C}$ are domains.

A subring of an integral domain is again an integral domain. $\mathbb{Z} \leq \mathbb{Q}$ and $\mathbb{Z}[\text{i}] \leq \mathbb{C}$ so $\mathbb{Z}$ and $\mathbb{Z}[\text{i}]$ are integral domains.

Lemma 2.27. A finite integral domain is a field.

Proof. Let $a \neq 0 \in R$ and consider

$$a \cdot - : (R, +, 0) \to (R, +, 0)$$

a homomorphism. If $a \cdot r = a \cdot r'$, then $a \cdot (r - r') = 0$, so as R is an integral domain, $r - r' = 0$, so $r = r'$. That is, $a \cdot -$ is injective. As R is finite, it must be a bijection. So $\exists b \in R$ such that $a \cdot b = 1_R$, so a is a unit, so R is a field. □

Lemma 2.28. If R is an integral domain, then so is $R[X]$.

Proof. If $f, g \neq 0 \in R[X]$, i.e.

$$\begin{aligned} f &= \sum_{i=0}^{n} a_i X^i \,, \ a_n \neq 0 \\ g &= \sum_{j=0}^{m} b_j X^j \,, \ b_m \neq 0 \,, \end{aligned}$$

then

$$\begin{aligned} f \cdot g &= a_0 b_0 + (a_1 b_0 + a_0 b_1) X + \cdots + \underbrace{a_n b_m}_{\neq 0 \text{ as } R \text{ integral domain}} X^{m+n} \\ &\neq 0 \,, \end{aligned}$$

so $R[X]$ is an integral domain. □

This implies that $R[X_1, X_2, \dots, X_n] = ((R[X_1])[X_2]\dots)[X_n]$ is also an integral domain.

Definition 2.29. Let R be an integral domain. A *field of fractions* F of R is a field with the properties

(i) $R \leq F$

(ii) Every element of F may be written as $a \cdot b^{-1}$ for $a, b \in R$, where b^{-1} means the multiplicative inverse to $b \neq 0$ in F.

Recall that a subring of any field is an integral domain. The converse is also true.

Theorem 2.30. Every integral domain has a field of fractions.

Proof. Consider the set

$$S = \{(a, b) \in R \times R \mid b \neq 0\}$$

and the relation $(a, b) \sim (c, d) \iff ad = bc \in R$.

Check that it is an equivalence relation: symmetry and reflexivity are obvious, need to check transitivity. If $(a, b) \sim (c, d)$, $(c, d) \sim (e, f)$, then $ad = bc, cf = de$, so

$$(ad)f = (bc)f = b(cf) = b(de)\,,$$

so $d(af - be) = 0$, $d \neq 0 \iff af = be$ as R is an integral domain $\iff (a, b) \sim (e, f)$.

Let $F = S/\sim$ be the set of equivalent classes, and $\frac{a}{b} = [(a, b)]$. Define

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

These are well defined, and make $(F, +, —, \frac{0}{1}, \frac{1}{1})$ into a ring. Need to check that every non-zero element is a unit so that it is a field. If $\frac{a}{b} \neq 0_F$, then $\frac{a}{b} \neq \frac{0}{1}$, so $(a, b) \not\sim (0, 1)$, so $a \cdot 1 \neq b \cdot 0 = 0$, i.e. $a \neq 0$. Then $\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1} = 1_F$, so $\frac{a}{b}$ is a unit.

Need to check the two conditions of F being a field of fraction of R.

(i) Need to show that R is isomorphic to a subring of F. Define

$$c : R \longrightarrow F$$
$$a \longmapsto \frac{a}{1}\,.$$

It is a homomorphism and it is injective, so by the first isomorphism theorem,$R \cong \operatorname{im} c \leq F$.

(ii) True by construction. □

Example. The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$.

Lemma 2.31. A non-zero ring is a field $\iff$ the ideals are $\{0\}$ and R.

Proof.

($\Rightarrow$) If $R \neq \{0\}$ is a field, $\{0\} \neq I \lhd R$, and let $x \in I$ be non-zero. Then $1 = x \cdot x^{-1} \in I$, so $I \in R$.

($\Leftarrow$) Let non-zero $x \in R$, and consider the ideal (x). It contains $x \neq 0$, so it is non-zero, so $(x) = R$, so $1 \in (x)$, so $\exists y \in R$ such that $x \cdot y = 1$, so x is a unit, so R is a field. □

This is another reason why fields are special. They have the simplest possible ideal structure. This motivates the following definition:

Definition 2.32. An ideal I of a ring R is *maximal* if $I \neq R$, and if $I \leq J \lhd R$, then $J = I$ or $J = R$.

Lemma 2.33. An ideal $I \lhd R$ is maximal $\iff$ R/I is a field.

Proof. R/I is a field $\iff$ I/I and R/I are the only ideals of R/I $\iff$ I and R are the only ideals of I which contain I $\iff$ I is maximal. □

Definition 2.34. An ideal $I \lhd R$ is *prime* if $I \neq R$ and if $a, b \in R$ such that $a \cdot b \in I \implies a \in I$ or $b \in I$.

Example. $n\mathbb{Z} \lhd \mathbb{Z}$ is a prime ideal $\iff$ $n = 0$ or n is a prime number. If p is a prime or 0 and $a \cdot b \in p\mathbb{Z}$, so $p \mid a \cdot b$, so $p \mid a$ or $p \mid b$, i.e. $a \in p\mathbb{Z}$ or $b \in p\mathbb{Z}$.

If $n = u \cdot v$ is a composite number $(u, v \neq \pm 1)$ then $u \cdot v \in n\mathbb{Z}$ but $u, v \neq n\mathbb{Z}$.

Lemma 2.35. $I \lhd R$ is prime $\iff$ R/I is an integral domain.

Proof.

($\Rightarrow$) Let I be prime. Let $a + I, b + I \in R/I$ and suppose

$$0_{R/I} = (a + I)(b + I) = ab + I\,,$$

so $ab \in I$. As I is prime, either $a \in I$ or $b \in I$, so $a + I = 0_{R/I}$ or $b + I = 0_{R/I}$, so R/I an integral domain.

($\Leftarrow$) Suppose R/I an integral domain. Let $a, b \in R$ such that $ab \in I$. Then $(a+I)(b+I) = ab+I = 0_{R/I}$. As R/I an integral domain, $a + I = 0_{R/I}$ or $b + I = 0_{R/I}$, i.e. $a \in I$ or $b \in I$, so I prime. □

Proposition 2.36. Every maximal ideal is a prime ideal.

Proof. $I \lhd R$ is maximal $\implies$ R/I is a field $\implies$ R/I an integral domain $\implies$ I is prime. □

The converse is clearly not true.

Proposition 2.37. Let R be an integral domain, then the characteristic of R is 0 or a prime number.

Proof. Consider the map $c : \mathbb{Z} \to R$ and $\ker c = n\mathbb{Z}$, where n is the characteristic. $\mathbb{Z}/\ker c = \mathbb{Z}/n\mathbb{Z} \cong \operatorname{im} c \leq R$. Then R is an integral domain $\implies$ $\mathbb{Z}/n\mathbb{Z}$ is an integral domain $\implies$ $n\mathbb{Z} \lhd \mathbb{Z}$ is a prime ideal $\implies$ $n = 0$ or n is a prime number. □

2.4 Factorisation in Integral Domains — Units, Primes and Irreducibles

In this section, R is always an integral domain.

Definition 2.38. Let R be an integral domain.

(i) $a \in R$ is a *unit* if $\exists b \in R$ with $a \cdot b = 1$. Equivalently, $(a) = R$.

(ii) $a \in R$ *divides* $b \in R$ if there is a $c \in R$ with $b = a \cdot c$, written $a \mid b$. Equivalently, $(b) \subseteq (a)$.

(iii) $a, b \in R$ are *associates* if $a = b \cdot c$ for some unit c. Equivalently, $(a) = (b)$. Equivalently, $a \mid b$ and $b \mid a$.

When considering division in rings, we often consider two associates to be "the same". For example, in $\mathbb{Z}$, we can factorize 6 as $6 = 2 \cdot 3 = (-2) \cdot (-3)$, but this does not violate unique factorization, since 2 and -2 are associates (and so are 3 and -3), and we consider these two factorizations to be "the same".

Definition 2.39. $r \in R$ is *irreducible* if it is not zero, not a unit, and if $r = a \cdot b$, then a or b is a unit.

For integers, being irreducible is the same as being a prime number. However, "prime" means something different in general rings.

Definition 2.40. $r \in R$ is *prime* if it is not zero, not a unit, and if r divides $a \cdot b$, then r divides a or r divides b.

Lemma 2.41. (r) is a prime ideal $\iff$ $r = 0$ or r is prime.

Proof.

(i) $a \in R$ Suppose (r) is prime and $r \neq 0$. As prime ideals are proper, r is not a unit. If $r \mid a \cdot b$, then $a \cdot b \in (r)$. So as (r) is prime, $a \in (r)$ or $b \in (r)$ so $r \mid a$ or $r \mid b$.

(ii) $(0) \lhd R$ as $R = R/(0)$ is an integral domain. Let $r \in R$ be prime and $a \cdot b \in (r)$, so $r \mid a \cdot b$, so $r \mid a$ or $r \mid b$, so $a \in (r)$ or $b \in (r)$. □

Lemma 2.42. If $r \in R$ is prime, then it is irreducible.

Proof. By the definition of a prime, r is not 0 and is not a unit. Let $r = a \cdot b$, so as r is prime, $r \mid a$ or $r \mid b$. WLOG, let $r \mid a$, then $a = r \cdot c$. Thus $r = a \cdot b = r \cdot c \cdot b$, so $r(1 - bc) = 0$. As R is an integral domain and $r \neq 0$, $bc = 1$ so b is a unit. □

The converse in not true in general.

Example. Let $R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{5} \mid a, b \in \mathbb{Z}\} \leq \mathbb{C}$. It is a subring of a field, so it is an integral domain.

We would like to find out the units of the ring. A useful trick is to define a function called norm:

$$\begin{aligned} N : R &\longrightarrow \mathbb{Z}_{\geq 0} \\ a + b\sqrt{-5} &\longmapsto a^2 + 5b^2 \,. \end{aligned}$$

So $N(z) = z \cdot \overline{z}$ for $z \in \mathbb{R}$ thought as an element of $\mathbb{C}$. This satisfies $N(z_1 z_2) = N(z_1)N(z_2)$. If $r \in R$ is a unit with inverse $s \in R$, then

$$1 = N(1) = N(r \cdot s) = N(r)N(s) \,,$$

so $N(r) = N(s) = 1$. If $r = a + b\sqrt{-5}$, then $a^2 + 5b^2 = 1$, so the only solutions are ± 1. The only units in R are ± 1.

More generally, if $N(x) = 1$, then x is a unit.

Our next claim is that $2 \in R$ is irreducible. Suppose $2 = a \cdot b \in R$, then

$$4 = N(2) = N(a \cdot b) = N(a) \cdot N(b) \,.$$

Note that $N(x) = a^2 + 5b^2 = 2$ has no integer solutions, so we can only have WLOG $N(a) = 1$ and $N(b) = 4$, then a is a unit so $2 \in R$ is irreducible.

Similarly, we can show that 3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible.

Also, $(1 + \sqrt{-5})(1 - \sqrt{-5}) = 1 + 5 = 2 \cdot 3$, so $2 \mid (1 + \sqrt{5})(1 - \sqrt{5})$. However, $N(2) = 4$ does not divide $N(1 \pm \sqrt{-5}) = 6$, so $2 \nmid 1 \pm \sqrt{5}$, so $2 \in R$ is not a prime.

Two lessons:

(i) irreducible $\not\Rightarrow$ prime.

(ii) An element can have multiple ways of factorising into irreducibles, e.g. $6 = 2\times 3 = (1-\sqrt{5})(1+\sqrt{5})$.

However, there is one situation when unique factorizations holds. This is when we have a Euclidean algorithm available.

Definition 2.43. An integral domain R is an *Euclidean domain* (ED) if there is a function $\phi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ such that

(i) $\phi(a \cdot b) \geq \phi(a)$

(ii) given $a, b \in R$, $b \neq 0$, then there are $p, r \in R$ with $a = p \cdot b + r$ and $\phi(r) < \phi(b)$ or $r = 0$.

Every time in this course when we say "Euclidean algorithm", it is an example.

Examples.

(i) $\mathbb{Z}$ is an ED with $\phi(n) = |n|$.

(ii) Let F be a field, then $F[X]$ is an ED with $\phi(f) = \deg f$.

(iii) $\mathbb{Z}[\mathrm{i}] \leq \mathbb{C}$ is an ED, with
$$\phi(a + \mathrm{i}b) = a^2 + b^2 = |a + \mathrm{i}b|^2 .$$
Note $\phi(z_1 z_2) = |z_1 z_2|^2 = |z_1|^2 |z_2|^2 \geq |z_1|^2 = \phi(z_1)$, $a^2 + b^2 \geq 1$ for $a + \mathrm{i}b \neq 0$.

Consider $z_1 \in R$, $z_2 \neq 0 \in R$. Have $\frac{z_1}{z_2} \in \mathbb{C}$, so there is a $q \in R$ such that $\left|\frac{z_1}{z_2} - q\right| < 1$. So $\frac{z_1}{z_2} = q + z_3$ for some $z_3 \in \mathbb{C}$ with $|z|_3 < 1$, then
$$z_1 = z_2 \cdot q + z_2 z_3 .$$
Let $z_2 z_3 = r \in R$, then $|r| = |z_2|\,|z_3| < |z_2|$ and $r = z_1 - z_2 \cdot q$ as required.

Definition 2.44. A ring R is a *principal ideal domain* (PID) if it is an integral domain, and every ideal is principal.

Example. All ideals of $\mathbb{Z}$ are $n\mathbb{Z} = (n)$ so $\mathbb{Z}$ is a PID.

Proposition 2.45. If R is a Euclidean domain, then it is a principal ideal domain.

Proof. Let R be an ED with Euclidean function ϕ. Let I be an ideal of R. Choose $b \in I \setminus \{0\}$ with $\phi(b)$ minimal. For $a \in I$, use ED to divide a: $a = b \cdot q + r$ with $\phi(r) < \phi(b)$. Then $r = \underbrace{a}_{\in I} - \underbrace{b \cdot q}_{\in I} \in I$, so $\phi(b)$ was a minimal in $I \setminus \{0\}$, must have $r = 0$. So $a = b \cdot q \in (b)$, so $I \subseteq (b)$, so $I = (b)$. □

Example. $\mathbb{Z}$, $\mathbb{Z}[\mathrm{i}]$ and $\mathbb{F}[X]$ for a field $\mathbb{F}$ are EDs, so they are PIDs.

Non-Example. $\mathbb{Z}[X]$ is not a PID. Consider $(2, X) \lhd \mathbb{Z}[X]$. If this were (f) for some $f \in \mathbb{Z}[X]$, then $2 = f \cdot g$ for some g, so f must be constant, and it must divide 2, so $f = \pm 1, \pm 2$. If $f = \pm 2$, then $X = f \cdot h$ for some h, but $2 \nmid X$. If $f = \pm 1$, then $(f) = (2, X) = \mathbb{Z}[X]$. But by the third isomorphism theorem,
$$\begin{aligned}\frac{\mathbb{Z}[X]}{(2, X)} &= \frac{\mathbb{Z}[X]/(X)}{(2, X)/(X)} \\ &= \frac{\mathbb{Z}}{2} \neq 0_{\mathbb{Z}[X]/[2,X]} ,\end{aligned}$$
so $(2, X) \neq \mathbb{Z}[X]$. Therefore, $(2, X)$ is not principal. □

Example. Let $\mathbb{F}$ be a field. $A \in M_{n\times n}(\mathbb{F})$ be a $n \times n$ matrix with entries in $\mathbb{F}$. Let $I = \{f \in \mathbb{F}[X] \mid f(A) = 0\}$. Note that this is an ideal in $\mathbb{F}[X]$ — if $f, g \in I$ and $h \in \mathbb{F}[X]$, then $(f + g)(A) = f(A) + g(A) = 0$ and $(fg)(A) = f(A)g(A) = 0$.

As $\mathbb{F}[X]$ is a PID, $I = (f_A)$ for some $f_A \in \mathbb{F}[X]$. This satisfies

(i) $f_A(A) = 0$

(ii) If $g \in \mathbb{F}[X]$ satisfies $g(A) = 0$, then $f_A \mid g$.

f_A is the *minimal polynomial* of A.

Definition 2.46. An integral domain is a *unique factorisation domain* (UFD) if

(i) every non-zero, non-unit element is a product of irreducibles.

(ii) if $p_1p_2 \dots p_n = q_1q_2 \dots q_m$ with all p_i, p_j irreducibles, then $n = m$ and up to reordering, p_i and q_i are associates.

Our next goal is to show PID implies UFD.

Lemma 2.47. Let R be a principal ideal domain. If $r \in R$ is irreducible than it is prime.

Note that this is also true in for general unique factorization domains, which we can prove directly by unique factorization. Note also the converse is always true in any integral domain.

Proof. Let p be a irreducible and $p \mid a \cdot b$. Need $p \mid a$ or $p \mid b$, so suppose $p \nmid a$. Consider $(p, a) \lhd R$. As R is a PID, $(p, a) = (d)$ for some $d \in R$, so $p = q_1 d$, $a = q_2 d$ for some $q_1, q_2 \in R$. As p is irreducible, either q_1 or d is a unit.

Suppose q_1 is a unit, then $a = q_2 q_1^{-1} p$ so $p \mid a$. Contradiction. So d is a unit, then $(p, a) = (d) = R$, and in particular, $1_R \in (p, a)$. Let $1_R = rp + sa$, then

$$b = \underbrace{r \cdot p \cdot b}_{p \mid r\cdot p\cdot b} + s \cdot \underbrace{a \cdot b}_{p \mid a\cdot b},$$

so $p \mid b$. □

Definition 2.48. A ring R is *Noetherian* if it satisfies the *ascending chain condition*, meaning for any chain of ideals of R

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$$

there is some $N > 0$ such that $I_n = I_{n+1}$ for all $n \geq N$

In a Noetherian ring, we cannot have an infinite chain of bigger and bigger ideals.

Lemma 2.49. A principal ideal domain is Noetherian.

Proof. Let

$$I = \bigcup_{n\geq 1} I_n \,,$$

then I is also an ideal. As R is a PID, $I = (a)$ for some a, so $a \in \bigcup_{n\geq 1} I_n$, so $I \in I_N$ for some N. But then $(a) \subseteq I_N \subseteq I_{N+1} \subseteq \dots \subseteq I$. As $I = (a)$, all these must be equalities. □

Proposition 2.50. If R is a principal ideal domain, then it is a unique factorisation domain.

Proof. We need to check the two conditions of UFD.

(i) Let $a \in R$ be non-zero, non-unit. Suppose it is not a product of irreducibles, so $a = a_1 \cdot b_1$, with a_1, b_1 both non-unit. As r is not a product of irreducibles, at least one of r_1 and r_2 is not a product of irreducibles. WLOG, say a_1 is not, so $a_1 = a_2 \cdot b_2$. Again, suppose a_2 is not a product of irreducibles, etc. By the assumption, the process does not end, and we have

$$(a) \subseteq (a_1) \subseteq (a_2) \dots$$

R is a PID, so it is Noetherian, so $(a_n) = (a_{n+1})$ for some $n > 0$, i.e.

$$a_n = a_{n+1} \cdot u$$

for some u. Have $a_n = a_{n+1} \cdot b_{n+1}$, so as R is a integral domain, get $b_{n+1} = u$. Contradiction as all b_i are assumed to be non-unit. So a is a product of irreducibles.

(ii) Let $p_1 p_2 \dots p_n = q_1 q_2 \dots q_m$, $n \leq m$ be a product of irreducibles. Then $p_1 \mid q_1 \dots q_m$, and as irreducible $\implies$ prime in PID, p_1 is prime, so $p_1 \mid q_i$ for some i. By reordering, $p_1 \mid q_1$.

As $p_1 \mid q_1$, $q_1 = p_1 c$, but q_1 is irreducible, and p_1 is not a unit, then c is a unit, so p_1 and q_1 are associates.

Now get

$$p_2 \cdots p_n = (c \cdot q_2) \cdot q_3 \dots q_m$$

a product of irreducibles. Carry on going, in the end get

$$1 = c' q_{n+1} \dots q_m \,.$$

As q_i are not units, we must have $n = m$. □

Definition 2.51. d is the *greatest common divisor* of $a_1, \dots a_n$, $d = \gcd(a_1, \dots, a_n)$, if $d \mid a_i$ for each i and if $d' \mid a_i$ for each i, then $d' \mid d$.

m is the *least common multiple* of $a_1, a_2, \dots a_n$, $m = \operatorname{lcm}(a_1, \dots, a_m)$, if $a_i \mid m$ for each i, and if $a_i \mid m'$ for each i, then $m \mid m'$.

Note that gcd and lcm are unique up to associates.

Proposition 2.52. If R is a unique factorisation domain, then the gcd's and lcm's exist.

Proof. We write each a_i as

$$a_i = u_i \prod_j p_j^{n_{ij}}$$

with u_i a unit, and p_j irreducibles which are not associates of each other. Claim that

$$d = \prod_j p_j^{m_j} \,, \quad \text{where } m_j = \min_i(n_{ij})$$

is a gcd of $a_1, \dots, a_n$. As $m_j \leq n_{ij}$ for each i, certainly $d \mid a_i$. If $d' \mid a_i$ for each a_i, write $d' = u \prod_j p_j^{t_j}$ and observe that $t_j \leq n_{ij}$ for all i, so $t_j \leq m_j$, so $d' \mid d$.

Similar for lcm. □

2.5 Factorisations in Polynomial Rings

For a field $\mathbb{F}$, we know that $\mathbb{F}[X]$ is a ED, so a PID, so a UFD.

(i) Every $I \lhd \mathbb{F}[X]$ is principal. $I = (f)$, $f \in \mathbb{F}[X]$.

(ii) $f \in \mathbb{F}[X]$ is irreducible $\iff$ f is prime.

(iii) Let f be irreducible, and $(f) \leq J \lhd \mathbb{F}[X]$. Then $J = (g)$ for some (g), so $f = g \cdot h$. But f is irreducible, so g or h is a unit.

- if g is a unit, then $(g) = \mathbb{F}[X]$.
- if h is a unit then $(g) = (f)$.

So $J = (f)$ is a maximal ideal.

Note this argument is valid for all PIDs.

(iv) (f) prime ideal $\implies$ f prime $\implies$ f irreducible $\implies$ f maximal.
So {prime ideals of $\mathbb{F}[X]$} = {maximal ideals of $\mathbb{F}[X]$}.

(v) $f \in \mathbb{F}[X]$ is irreducible $\iff$ $\frac{\mathbb{F}[X]}{(f)}$ is a field.

Definition 2.53. Let R be a UFD and

$$f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$$

with $a_n \neq 0$. The content of f is

$$c(f) \coloneqq \gcd(a_0, \ldots, a_n) \in R\,.$$

Say f is primitive if $c(f)$ is a unit, i.e. the a_i are coprime.

Next, we want to prove Gauss' lemma.

Lemma (Gauss' Lemma). Let R be a unique factorisation domain, F be its field of fractions. Suppose $f \in R[X]$ is primitive, then

$$f \text{ irreducible in } R[X] \iff f \text{ is irreducible in } F[X]\,.$$

We can't do this right away. We first need some preparation. Before that, we do some examples.

Example. Consider $f = X^3 + X + 1 \in \mathbb{Z}[X]$. We show it is not reducible in $\mathbb{Z}[X]$, and hence not reducible in $\mathbb{Q}[X]$.

This has content 1, so primitive. If f has a proper factorisation $f = g \cdot h$, then write

$$\begin{aligned} g &= b_0 + b_1 X \\ h &= c_0 + c_1 X + c_2 X^2\,. \end{aligned}$$

See $b_0 c_0 = 1$, $b_1 c_2 = 1 \in \mathbb{Z}$, so $b_0, b_1 = \pm 1$, so $g = \pm(1 \pm x)$, so ± 1 is a root of f. But ± 1 are not roots of $X^3 + X + 1$, so f in not reducible in $\mathbb{Z}[X]$.

By Gauss' lemma, f is also irreducible in $\mathbb{Q}[X]$. In particular, f has no root in $\mathbb{Q}$, and $\mathbb{Q}[X]/(X^3 + X + 1)$ is a field.

Lemma 2.54. Let R be a unique factorisation domain. If $f, g \in R[X]$ are primitive, then so is $f \cdot g$.

Proof. Let

$$\begin{aligned} f &= a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \\ g &= b_0 + b_1 X + b_2 X^2 + \cdots + b_m X^m\,. \end{aligned}$$

Suppose $f \cdot g$ is not primitive, then $c(f \cdot g) \in R$ is not a unit, so as R is an UFD, can find an irreducible p that divides $c(f \cdot g)$. But f and g are primitive so $p \nmid c(f)$ and $p \nmid c(g)$.

Suppose

$$\begin{array}{cccccc} p \mid a_0 & p \mid a_1 & \ldots & p \mid a_{k-1} & p \nmid a_k \\ p \mid b_0 & p \mid b_1 & \ldots & p \mid b_{l-1} & p \nmid b_l \end{array}\,.$$

The coefficient of X^{k+l} in $f \cdot g$ is

$$\underbrace{\sum_{i+j=k+l} a_i b_j}_{\substack{\text{divisible by } p \\ \text{as } p \mid c(f\cdot g)}} = \underbrace{\cdots + a_{k+1}b_{l-1}}_{\substack{\text{divisible by } p \\ \text{as } p \mid b_0, \ldots, p \mid b_{l-1}}} + a_k b_l + \underbrace{a_{k-1}b_{l+1} + \ldots}_{\substack{\text{divisible by } p \\ \text{as } p \mid a_0, \ldots, p \mid a_{k-1}}}$$

So $p \mid a_k \cdot b_l$. As p is irreducible, so prime, $p \mid a_k$ or $p \mid b_l$. Contradiction. So $f \cdot g$ is primitive. □

Corollary. Let R be a unique factorisation domain, $f, g \in R[X]$, then $c(f \cdot g)$ is an associate of $c(f) \cdot c(g)$.

Proof. Write

$$f = c(f) \cdot f_0\,,\ f_0 \text{ primitive},$$
$$g = c(g) \cdot g_0\,,\ g_0 \text{ primitive}.$$

So $f \cdot g = c(f) \cdot c(g) \cdot f_0 g_0$, where $f_0 g_0$ is primitive by the previous lemma. Then $c(f) \cdot c(g)$ is a gcd of the coefficients of $f \cdot g$, i.e. is $c(f \cdot g)$. □

Finally, we can prove Gauss' Lemma.

Lemma 2.55 (Gauss' Lemma). Let R be a unique factorisation domain, F be its field of fractions. Suppose $f \in R[X]$ is primitive, then

$$f \text{ irreducible in } R[X] \iff f \text{ is irreducible in } F[X]\,.$$

Proof. We will show that a primitive $f \in R[X]$ is reducible in $R[X]$ if and only if f is reducible in $F[X]$.

If f is reducible in $R[X]$, it factors as $f = g \cdot h$, with g, h non-units. If g or h is a constant polynomial, then f is not primitive, then f is not primitive, leading to a contradiction. So $f = g \cdot h$ is also a factorisation into non-units in $F[X]$.

Suppose f is reducible in $F[X]$, so $f = g \cdot h$ with g, h non-units in $F[X]$, in particular non-constants. We may find $a, b \neq 0 \in R$ such that

$$a \cdot g \in R[X]\,,\ b \cdot h \in R[X]\,.$$

Then $a \cdot b \cdot f = (a \cdot g) \cdot (b \cdot h) \in R[X]$. Write

$$a \cdot g = c(a \cdot g) \cdot g_1\,,\ g_1 \text{ primitive, non-constant},$$
$$b \cdot h = c(b \cdot h) \cdot h_1\,,\ h_1 \text{ primitive, non-constant}.$$

By the above corollary, $a \cdot b$ is an associate of $c(a \cdot g) \cdot c(b \cdot h)$, so $a \cdot b = uc(a \cdot f)c(b \cdot h)$ for some unit u. So

$$\begin{aligned} a \cdot b \cdot f &= c(a \cdot g) \cdot c(b \cdot h) \cdot g_1 h_1 \\ &= a \cdot b \cdot u^{-1} \cdot g_1 \cdot h_1\,. \end{aligned}$$

R is a integral domain, so $f = (u^{-1}g_1) \cdot h_1$ with $u^{-1}g_1$, $h_1 \in R$ non-units, so f is reducible in $R[X]$.□

Proposition 2.56. Let R be a unique factorisation domain, F be its field of fractions, $g \in R[X]$ be primitive. Let $I = (g) \lhd F[X]$ and $J = (g) \lhd R[X]$. Then $J = I \cap R[X]$.

In other words, if $f \in R[X]$ is divisible by g in $F[X]$, then it is divisible by it in $R[X]$.

Proof. Suppose g primitive, $f = g \cdot h$, $h \in F[X]$. Let $b \in R$ be such that $b \cdot h \in R[X]$, so $b \cdot f = g \cdot (bh)$, a factorisation in $R[X]$.

Let $b \cdot h = c(b \cdot h) \cdot h_1$, h_1 primitive, so

$$b \cdot f = c(b \cdot h) \cdot \underbrace{g \cdot h_1}_{\text{primitive}} .$$

Thus $b \mid c(b \cdot h)$, so $c(b \cdot h) = b \cdot c$ for some $c \in R$. Then $b \cdot f = b \cdot c \cdot g \cdot h_1$, so $f = g(c \cdot h_1)$, with $ch_1 \in R[X]$, i.e. $g \mid f$ in $R[X]$. □

Theorem 2.57. Let R be a unique factorisation domain, then $R[X]$ is a unique factorisation domain.

Proof. Let $f \in R[X]$, write $f = c(f) \cdot f_1$ with f_1 primitive. As R is an UFD, have $c(f) = p_1 \dots p_n$ product of irreducibles in R.

If f_1 is reducible, then write it as $f_1 = f_2 \cdot f_3$, where f_2, f_3 are non-units, then $\deg f_2, \deg f_3 < \deg f_1$. Iterating this gives

$$f_1 = q_1 \dots q_m ,$$

with q_i irreducible and $m \leq \deg f$. So $f = p_1 \dots p_n \dots q_1 \dots q_n$ is a product of irreducibles in $R[X]$.

The p_i's are unique up to reordering and associates as R is a UFD. Need to show that the q_i's are also unique up to reordering and associates. Divide by $c(f)$. Suppose $q_1 \dots q_n = r_1 \dots r_l$ is primitive, so all q_i, r_i are primitive too.

Let F be the field of fractions of R, and consider $q_i, r_i \in F[X]$. Since F is a field, $F[X]$ is a Euclidean domain, hence a principal ideal domain, hence a unique factorization domain. By Gauss' lemma, the q_i and r_i are also irreducible in $F[X]$. As $F[X]$ is a UFD, deduce that $m = l$ and up reordering, q_i is an associate to r_i in $F[X]$, i.e. $q_i = r_i \cdot a_i$ for some units $a_i \in F$. We have $a_i = \frac{y_i}{x_i}$, $y_i, x_i \in R$, so $x_i q_i = y_i r_i$ in $R[X]$. Taking contents, since q_i, r_i are primitive, x_i and y_i are associates: $y_i = x_i \cdot u_i$ for u_i units. $q_i = u_i r_i$ so q_i and r_i are associates in $R[X]$. □

Example. $\mathbb{Z}[X]$ is a UFD, $R[X]$ is a UFD, $R[X_1, X_2, \dots]$ is a UFD.

This is a useful thing to know. In particular, it gives us examples of UFDs that are not PIDs. However, in such rings, we would also like to have an easy to determine whether something is reducible. Fortunately, we have the following criterion.

Proposition 2.58 (Eisenstein's Criterion). Let R be a unique factorisation domain, and

$$f = a_0 + a_1 X + \dots + a_n X^n \in R[X]$$

with $a_n \neq 0$ and f primitive. Suppose there is a prime such that

(i) $p \nmid a_n$

(ii) $p \mid a_i$, $0 \leq i \leq n-1$

(iii) $p^2 \nmid a_0$

then f is irreducible in $R[X]$ and also in $F[X]$ by Gauss' lemma.

Proof. Suppose we have $f = g \cdot h$ with

$$g = r_0 + r_1 X + \dots + r_k X^k$$
$$h = s_0 + s_1 X + \dots + s_l X^l$$

with $r_k, s_l \neq 0$, $k + l = n$, and $a_n = r_k \cdot s_l$. As $p \nmid a_n$, $p \nmid r_k$ and $p \nmid s_l$. As $p^2 \nmid a_0 = r_0 s_0$ but $p \mid r_0 s_0$, must have $p \mid r_0$ and $p \nmid s_0$.

Choose j such that $p \mid r_0, p \mid r_1, \ldots, p \mid r_{j-1}, p \nmid r_j$. Then

$$a_j = \underbrace{r_0 s_j}_{p|} + \underbrace{r_1 s_{j-1}}_{p|} + \underbrace{\ldots}_{p|} + \underbrace{r_j s_0}_{p\nmid},$$

as p is prime, so $p \nmid a_j$. So deduce that $j = n$. On the other hand, $j \leq k \leq n$, so $k = n$, so $l = 0$. The polynomial h is a constant. But f is primitive, so h must be a unit, so f irreducible in $R[X]$ and hence in $F[X]$.

Example. Let p be a prime number, and consider $X^n - p \in \mathbb{Z}[X]$, $n > 1$. Eisenstein's criterion applies with p, so this is irreducible, even in $\mathbb{Q}[X]$.

This implies that $\sqrt[n]{p} \notin \mathbb{Q}$.

Example. Consider

$$f = X^{p-1} + X^{p-2} + \cdot + 1 \in \mathbb{Z}[X]$$

with p a prime number. Note

$$f = \frac{X^p - 1}{X - 1}.$$

Let Y be $X - 1$ and

$$\bar{f} = \frac{(1+Y)^p - 1}{Y} = Y^{p-1} + \underbrace{\binom{p}{1}}_{=p} Y^{p-2} + \underbrace{\binom{p}{2}}_{p|} Y^{p-3} + \cdots + \underbrace{\binom{p}{p-1}}_{=p} Y^0.$$

Eisenstein's criterion applies to $\bar{f}$, so $\bar{f}$ is irreducible in $\mathbb{Z}[Y]$. If $f = g \cdot h$ is reducible, then

$$\begin{aligned}\bar{f}(Y) &= f(1+Y) \\ &= g(1+Y)h(1+Y)\end{aligned}$$

is also reducible, leading to a contradiction. So f must be irreducible.

Hence none of the roots of f are rational (and we know that in fact they are not even real).

2.6 Gaussian Integers

Recall

$$\mathbb{Z}[\mathrm{i}] = \{a + b\mathrm{i} \mid a, b \in \mathbb{Z}\} \leq \mathbb{C}$$

has a norm

$$N(a + \mathrm{i}b) = (a + \mathrm{i}b)\overline{(a + \mathrm{i}b)},$$

which is a Euclidean function, so $\mathbb{Z}[\mathrm{i}]$ is a ED, so a PID, so a UFD.

The units in $\mathbb{Z}[\mathrm{i}]$ are ± 1, $\pm\mathrm{i}$, and are the only elements of norm 1.

(i) $N(3) = 9$, so if $3 = a \cdot b$, then $N(a) \cdot N(b) = 9$, so either
- $N(a)$ or $N(b) = 1$, then a and b are units.
- $N(a) = N(b) = 3$, but 3 is not a sum of two squares, so cannot be a norm. So 3 is a prime, so cannot be a norm.

So 3 is irreducible, so it is a prime.

(ii) $5 = (1 + 2\mathrm{i})(1 - 2\mathrm{i})$ is not a prime.

(iii) 7 is not a sum of squares, so is a prime in $\mathbb{Z}[\mathrm{i}]$.

Proposition 2.59. A prime number $p \in \mathbb{Z} \leq \mathbb{Z}[\mathrm{i}]$ is a prime in $\mathbb{Z}[\mathrm{i}] \iff p \neq a^2 + b^2$ for $a, b \in \mathbb{Z}$.

Proof. If $p = a^2 + b^2$, then $p = (a + \mathrm{i}b)(a - \mathrm{i}b)$ not prime.

Otherwise, have $N(p) = p^2$, so if p factors into non-units, then they each have a norm p, so p is a norm, so is a sum of two squares. □

Next, we want to classify all primes in $\mathbb{Z}[\mathrm{i}]$.

Lemma 2.60. Let p be a prime number and $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ be a field. Let $\mathbb{F}_p^\times = \mathbb{F}_p - \{0_{\mathbb{F}_p}\}$ be the group of invertible elements under multiplication. Then $\mathbb{F}_p^\times \cong \mathbb{C}_{p-1}$.

Proof. Certainly, $\mathbb{F}_p^\times$ has order $p-1$, and is abelian. We know from the classification of finite abelian groups that if $\mathbb{F}_p^\times$ is not cyclic, then it must contain a subgroup $C_m \times C_m$ for some $m \geq 2$.

Consider the polynomial $X^m - 1 \in \mathbb{F}_p[X]$.

$$\{\text{elements of } \mathbb{F}_p^\times \text{ of order } m\} = \{\text{roots of } x^m - 1 \in \mathbb{F}_p[X]\}\,.$$

This polynomial has at most m roots, as $\mathbb{F}_p[X]$ is a UFD. So $\mathbb{F}_p^\times$ has at most m elements of order m. But $C_m \times C_m$ has m^2 elements of order m. So $\mathbb{F}_p^\times$ does not contain such a group, so it is cyclic. □

Proposition 2.61. The primes in $\mathbb{Z}[\mathrm{i}]$, up to associates, are

(i) prime numbers $p \in \mathbb{Z} \leq \mathbb{Z}[\mathrm{i}]$ with $p \equiv 3 \pmod 4$, or

(ii) $z \in \mathbb{Z}[\mathrm{i}]$ with $N(z) = p$ a prime number in $\mathbb{Z}$ with $p \equiv 1 \pmod 4$ or $p = 2$.

Proof. First show that they are prime.

(i) If $p \equiv 3 \pmod 4$, then it is not a sum of two squares (squares are $\equiv 0, 1 \pmod 4$), so $p \in \mathbb{Z}[\mathrm{i}]$ by the previous proposition.

(ii) Let z be such that $N(z) = p$, a prime number $\equiv 1$ or $\equiv 2 \pmod 4$. Suppose $z = u \cdot v$, then $p = N(z) = N(u) \cdot N(v)$, so $N(u) = 1$ or $N(v) = 1$, so u or v is a unit.

Note that we did not use the condition that $p \equiv 3 \pmod 4$. This is not needed, since $N(z)$ is always a sum of two squares, and hence $N(z)$ cannot be a prime that is 3 mod 4.

Let $z \in \mathbb{Z}[\mathrm{i}]$ be prime. Then $\overline{z}$ is also prime, so $N(z) = z \cdot \overline{z}$ is a factorisation into irreducibles. Let p be a prime which divides $N(z)$.

(i) $p \equiv 3 \pmod 4$.

The p is a prime in $\mathbb{Z}[\mathrm{i}]$, so $p \mid z\overline{z} \implies p \mid z$ or $p \mid \overline{z}$. If $p \mid z$, then $p \mid \overline{z}$ by taking conjugate, so $p \mid z$. Since p and z are both irreducible, they must be associates.

(ii) $p = 2 = (1 + \mathrm{i})(1 - \mathrm{i})$ or $p \equiv 1 \pmod 4$.

If $p \equiv 1 \pmod 4$, consider $\mathbb{F}_p^\times$, a cyclic group of order $p - 1 = 4k$. It has a unique element of order 2, $[-1] \in \mathbb{F}_p^\times$. It also has an element $a \in \mathbb{F}_p^\times$ of order 4 (e.g. the k^{th} power of the generator). Thus a^2 has order exactly 2, so $a^2 = -1 \in \mathbb{F}_p^\times$. $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, so $a = [A]$ for some $A \in \mathbb{Z}$, and $A^2 + 1 \equiv 0 \pmod p$, i.e.

$$p \mid A^2 + 1 = (A + \mathrm{i})(A - \mathrm{i})\,.$$

However, p does not divide $A + \mathrm{i}$ or $A - \mathrm{i}$, so p is not a prime in $\mathbb{Z}[\mathrm{i}]$ (also for $p = 2$), so also not irreducible.

Therefore, we can write $p = z_1 \cdot z_2$, a product of non-units. Then $p^2 = N(p) = N(z_1)N(z_2)$, so $N(z_1) = N(z_2) = p$, so $z_1 = \overline{z}_2$. So $p = z_1\overline{z}_1$, so $z_1\overline{z}_1 = p \mid N(z) = z\overline{z}$. Since $N(z) = p$ and $z, \overline{z}$ irreducible, z_1 is irreducible and is an associate of z or $\overline{z}$. So $N(z) = N(z_1) = p$ as required. □

Corollary. A $n \in \mathbb{Z}_{\geq 0}$ can be written as $x^2 + y^2$, $x, y \in \mathbb{Z} \iff$ when we can write $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ as a factorisation into prime, if $p_i \equiv 3 \pmod 4$, then a_i is even.

Proof. Let $n = x^2 + y^2 = N(x + \mathrm{i}y)$. Let $z = x + \mathrm{i}y$. Let $z = \alpha_1\alpha_2 \dots \alpha_s$ a product of irreducibles in $\mathbb{Z}[\mathrm{i}]$, so $n = N(z) = N(\alpha_1)N(\alpha_2) \dots N(\alpha_s)$.

Each α is either a prime $\equiv 3 \pmod 4$, or is such that $N(\alpha) = 2$ or prime $\equiv 1 \pmod 4$. In the first case, $N(\alpha) = (\text{a prime} \equiv 3 \pmod 4))^2$. In the second case, $N(\alpha) = 2$ or a prime $\equiv 1 \pmod 4$.

Conversely, let $n = p_1^{a_1} \dots p_k^{a_k}$ have the given form. For each i, if $p_1 \equiv 3 \pmod 4$, then a_i is even, so $p_i^{\alpha_i} = N(p_i)^{\alpha_i/2} = N(p_i^{\alpha_i/2})$. If $p_i = 2$ or $\equiv 1 \pmod 4$, then there is a α_i with $N(\alpha_i) = p_i$. So we see that $n = p_1^{a_1} \dots p_k^{a_k}$ is a norm, so it is a sum of two squares. □

Example. Consider $65 = 5 \cdot 13$.

Then

$$5 = (2 + \mathrm{i})(2 - \mathrm{i}) \,, 13 = (2 + 3\mathrm{i})(2 - 3\mathrm{i}) \,,$$

so $65 = (2 + \mathrm{i})(2 - \mathrm{i})(2 + 3i)(2 - 3i)$ is a prime factorisation in $\mathbb{Z}[\mathrm{i}]$. So

$$\begin{aligned} 65 &= N((2 + \mathrm{i})(2 + 3\mathrm{i}) = 1 + 8\mathrm{i}) = 1^2 + 8^2 \\ &= N((2 + \mathrm{i})(2 - 3\mathrm{i}) = 7 - 4\mathrm{i}) = 7^2 + 4^2 \,. \end{aligned}$$

Remark. Can check whether $p = x^2 + 2y^2$ by working in $\mathbb{Z}[\sqrt{-2}]$, and check whether $p = x^2 + 3y^2$ by working in $\mathbb{Z}[\sqrt{-3}]$.

However, cannot do $p = x^2 + 5y^2$ since $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

2.7 Algebraic Integer

Definition 2.62. An $\alpha \in \mathbb{C}$ is called an *algebraic integer* if it is a root of a monic polynomial $f \in \mathbb{Z}[X]$.

We can immediately check that this is a sensible definition — not all complex numbers are algebraic integers, since there are only countably many polynomials with integer coefficients, hence only countably many algebraic integers, but there are uncountably many complex numbers.

Notation. For an algebraic integer α, write $\mathbb{Z}[\alpha] \leq \mathbb{C}$ for the smallest subring of $\mathbb{C}$ containing α.

We can construct $\mathbb{Z}[\alpha]$ by considering the map

$$\begin{aligned} \phi : \mathbb{Z}[X] &\longrightarrow \mathbb{C} \\ X &\longmapsto \alpha \,. \end{aligned}$$

Then $\mathbb{Z}[\alpha] = \mathrm{im}(\phi)$. So we can also write

$$\mathbb{Z}[\alpha] \cong \frac{\mathbb{Z}[X]}{\ker(\phi)} \,,$$

where $I = \ker(\phi)$ is non-zero by the definition of an algebraic integer.

Proposition 2.63. If α is an algebraic integer, and

$$\begin{aligned} \phi : \mathbb{Z}[X] &\longrightarrow \mathbb{C} \\ X &\longmapsto \alpha \,. \end{aligned}$$

Then $I = \ker(\phi)$ is a principal ideal generated by a monic irreducible polynomial $f_\alpha \in \mathbb{Z}[\alpha]$.

This is a non-trivial theorem, since $\mathbb{Z}[X]$ is not a principal ideal domain so there is no immediate guarantee that I is generated by one polynomial.

Definition 2.64. Let $\alpha \in \mathbb{C}$ be an algebraic integer. Then the minimal polynomial of α is the irreducible monic polynomial f_α such that $I = \ker(\phi) = (f_\alpha)$.

Proof. By definition, there is a monic $f \in \mathbb{Z}[X]$ such that $f(\alpha) = 0$, i.e. $f \in I$. Let $f_\alpha \in I$ be a non-zero polynomial of minimal degree: we can assume it is primitive. We want to show $I = (f_\alpha)$. Let $h \in I$. As $\mathbb{Q}[X]$ is an ED, we can write

$$h = q \cdot f_\alpha + r \in \mathbb{Q}[X]$$

with $r = 0$ or $\deg r < \deg f_\alpha$. Clearing denominators, get

$$a \cdot h = (aq) f_\alpha + (a \cdot r) \in \mathbb{Z}[X]\,.$$

Evaluate these polynomials at α gives

$$\underbrace{ah(\alpha)}_{h(\alpha)=0} = \underbrace{aq(\alpha) f_\alpha(\alpha)}_{f_\alpha(\alpha)=0} + ar(\alpha)\,.$$

Now α is a root of $a \cdot r$. If $r \neq 0$, then we have found an element in I with degree smaller than $\deg f_\alpha$, giving contradiction.

So $r = 0$, so

$$a \cdot h = (a \cdot q) f_\alpha \in \mathbb{Z}[X]\,.$$

Taking constants $a \cdot c(h) = c(a \cdot h) = c(a \cdot q)$ as f_α primitive, so $a \mid c(a \cdot q)$, so $q \in \mathbb{Z}[X]$. So $h = q \cdot f_\alpha$, so $h \in (f_\alpha) \lhd \mathbb{Z}[X]$.

Finally, show f_α is irreducible. Notice that

$$Z[X]/I \cong \mathbb{Z}[\alpha] \leq \mathbb{C}\,,$$

so is an integral domain. So $I = (f_\alpha)$ is a prime ideal, so $f(\alpha) \in \mathbb{Z}[X]$ is prime, so irreducible. □

Examples.

- $\alpha = \mathrm{i}$ has $f_\alpha = X^2 + 1$.
- $\alpha = \sqrt{2}$ has $f_\alpha = X^2 - 2$.
- $\alpha = \frac{1}{2}(1 - \sqrt{3})$ has $f_\alpha = X^2 - X + 1$.

Example. For $d \in \mathbb{Z}$ the polynomial

$$X^5 - X + d \in \mathbb{Z}[X]$$

has 1 real root, α, an algebraic integer. It cannot be expressed in the language $(\mathbb{Z}, +, \cdot, \sqrt[n]{\ })$. This is the subject of Galois theory.

Lemma 2.65. If α is an algebraic integer and $\alpha \in \mathbb{Q}$, then $\alpha \in \mathbb{Z}$.

Proof. Let $f_\alpha \in \mathbb{Z}[X]$ be its minimal polynomial, which is irreducible and monic, so primitive. By Gauss' lemma, it is also irreducible in $\mathbb{Q}[X]$. But $(x - \alpha)$ divides f_α in $\mathbb{Q}[X]$. As it is irreducible, $f_\alpha = x - \alpha$. But $f_\alpha \in \mathbb{Z}[X]$, so $\alpha \in \mathbb{Z}$. □

2.8 Hilbert Basis Theorem

Recall a PID satisfies the ascending chain condition: if

$$I_1 \subseteq I_2 \subseteq \ldots$$

is a chain of ideals of R, then $I_n = I_{n+1}$ for all $n \geq N$ for some N. Rings satisfying ACC are called Noetherian.

Definition 2.66. An ideal I is *finitely generated* if it can be written as $I = (r_1, \dots, r_n)$ for some $r_1, \dots, r_n \in R$.

Lemma 2.67. R is Noetherian $\iff$ all ideals of R are finitely generated.

Proof.

($\Leftarrow$) Let $I_1 \subseteq I_2 \subseteq \dots$ be a chain of ideals, and $I = \bigcup_{n\geq 1} I_n$, again an ideal. This is finitely generated, so $I = (a_1, \dots, a_s)$. Each a_i lies in some I_{n_i}, so they all lie in I_N if $N = \max\{n_i\}$. So $(a_1, \dots, a_s) \subseteq I_N \subseteq I$ are equalities, so $I_N = I_{N+1} + \dots = I$.

($\Rightarrow$) Suppose R satisfies ACC. Let J be an ideal. Choose $0 \neq a_1 \in J$. If $(a_1) = J$ then done. Else, choose $a_2 \in J \setminus (a_1)$. If $(a_1, a_2) = J$ then done. If not, choose $a_3 \in J \setminus (a_1, a_2, \dots)$.

If this process does not stop, get

$$(a_1) \subsetneq (a_1, a_2) \subsetneq (a_1, a_2, a_3) \subsetneq \dots$$

This contradicts ACC, so must have $J = (a_1, \dots, a_s)$. □

Theorem 2.68 (Hilbert basis theorem). If R is Noetherian, then $R[X]$ is Noetherian.

Proof. Let J be an ideal of $R[X]$. Let $f_1 \in J$ be a polynomial of minimal degree in J. In $J \neq (f_1)$, choose $f_2 \in J \setminus (f_1)$, of minimal degree. If at any point, $J = (f_1, \dots, f_r)$, then J is finitely generated, so we are done.

Suppose not. Let a_i be the top non-zero coefficient in f_i, and consider the ideals

$$(a_1) \subseteq (a_1, a_2) \subseteq (a_1, a_2, a_3) \subseteq \dots \subseteq R\,.$$

As R is Noetherian, these stabilise, so

$$(a_1, a_2, \dots) = (a_1, a_2, \dots, a_m)$$

for some m. So $a_{m+1} \in (a_1, \dots, a_m)$, so

$$a_{m+1} = \sum_{i=1}^{m} a_i b_i$$

for some $b_i \in R$. Let

$$g = \sum_{i=1}^{m} b_i f_i X^{\deg(f_{m+1}) - \deg(f_i)}\,.$$

This has the same degree as f_{m+1} and the same top coefficient a_{m+1}. So $f_{m+1} - g$ has degree strictly smaller than f_{m+1}. But $g \in (f_1, \dots, f_m)$ and $f_{m+1} \notin (f_1, \dots, f_m)$, so $f_{m+1} - g \notin (f_1, \dots, f_m)$. This contradicts our choice that f_{i+1} has minimal degree among polynomials in $J \setminus (f_1, \dots, f_i)$. □

Corollary. $\mathbb{Z}[X_1, \dots, X_n]$ is Noetherian. $\mathbb{F}[X_1, \dots, X_n]$ is Noetherian for a field $\mathbb{F}$.

Lemma 2.69. A quotient of a Noetherian ring is Noetherian.

Proof. Let R be Noetherian, $I \lhd R$, and $J_1 \subseteq J_2 \subseteq \dots$ be ideals of R/I. This corresponds to a chain

$$I \subseteq J_1' \subseteq J_2' \subseteq \dots$$

of ideals of R which contain I. As R is Noetherian, $J_n' = J_{n+1}'$ $\forall n \geq N$ for some N. So $J_m = J_n'/I = J_{m+1} = J_{n+1}'/I$. □

Corollary. Any finitely-generated ring is Noetherian.

An aside. If $\mathbb{F}$ is a field and $f \in \mathbb{F}[X_1, \ldots, X_n]$. Given $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$, consider the ring homomorphism

$$\begin{aligned} \phi : \mathbb{F}[X_1, \ldots, X_n] &\longrightarrow \mathbb{F} \\ X_i &\longmapsto \alpha_i \,. \end{aligned}$$

We know that $(\alpha_1, \ldots, \alpha_n)$ is a solution to $f = 0$ if and only if (f) is the kernel of the homomorphism ϕ. By the first isomorphism theorem, we get the correspondence

$$\{\text{solutions of } f \text{ over } \mathbb{F}\} = \left\{ \text{ring homomorphisms } \begin{aligned} \frac{\mathbb{F}[X_1, \ldots, X_n]}{(f)} &\longrightarrow \mathbb{F} \\ x_i &\longmapsto \alpha_i \end{aligned} \right\}$$

Now suppose we have a possibly infinite collection of polynomials $\{f_i\}_{i \in I}$ that we want to solve. Then a simultaneous solution is a homomorphism

$$\phi : \frac{\mathbb{F}[X_1, \ldots, X_n]}{(f_i)_{i \in I}} \to \mathbb{F} \,.$$

But by the Hilbert basis theorem, $\mathbb{F}[X]$ is Noetherian, so finitely generated, so

$$(f_i)_{i \in I} = (g_1, \ldots, g_r)$$

for a finite set of polynomials $\{g_i\}_{i=1}^r \subseteq \mathbb{F}[X_1, \ldots, X_n]$. So

$$\left\{ \begin{matrix} \text{simultaneous solutions} \\ \text{to all } \{f_i\}_{i \in I} \end{matrix} \right\} = \left\{ \begin{matrix} \text{simultaneous solutions} \\ \text{to } g_1, \ldots, g_r \end{matrix} \right\} .$$

Now we only need to solve a finite set of polynomials!

3 Modules

Recall that when we define vector space, we first pick some base field $\mathbb{F}$. We then define the vector space to be an abelian group V with an action of $\mathbb{F}$ on V (scalar multiplication) that is compatible with the multiplicative and additive structure of $\mathbb{F}$.

3.1 Definitions and Examples

Definition 3.1. Let R be a commutative ring. A quadruple $(M, +, 0_M, \cdot)$ is called a *R-module* if

(i) $(M, +, 0_M)$ is an abelian group.

(ii) The operation $- \cdot - : R \times M \to M$ satisfies
- $(r_1 +_R r_2) \cdot m = r_1 \cdot m +_M r_2 \cdot m$
- $r \cdot (m_1 +_M m_2) = r \cdot m_1 +_M r \cdot m_2$
- $r_1 \cdot (r_2 \cdot m) = (r_1 \cdot_R r_2) \cdot m$
- $1_R \cdot m = m$

for all $r, r_1, r_2 \in R$ and $m, m_1, m_2 \in M$.

Examples.

(i) Let $\mathbb{F}$ be a field. An $\mathbb{F}$-module is exactly a vector space.

(ii) For any ring R, $R^n = \underbrace{R \times R \times \cdots \times R}_{n \text{ times}}$ is an R-module via

$$r \cdot_M (r_1, r_2, \ldots, r_n) = (r \cdot_R r_1, r \cdot_R r_2, \ldots, r \cdot_R r_n)\,.$$

When $n = 1$, we can see that R itself is an R-module.

(iii) If I is an ideal of R, then it is an R-module via

$$r \cdot_M x = r \cdot_R x\,.$$

Also, R/I is an R-module via

$$r \cdot_M (r_1 + I) = r \cdot_R r_1 + I\,.$$

(iv) For $R = \mathbb{Z}$, a $\mathbb{Z}$-module is precisely the same as an abelian group.

$$\begin{aligned} - \cdot - : \mathbb{Z} \times A &\longrightarrow A \\ (n, a) &\longmapsto \begin{cases} \underbrace{a + a + \ldots}_{n \text{ times}} & n \geq 0 \\ \underbrace{-a - a - \ldots}_{|n| \text{ times}} & n \leq 0 \end{cases} \end{aligned}$$

This is forced by axioms. If we send $(1, a) \mapsto a$, then we must send $(2, a) = (1 + 1, a) \mapsto a + a$.

(v) Let $\mathbb{F}$ be a field, V a $\mathbb{F}$-module (i.e. a $\mathbb{F}$-vector space) and $\alpha : V \to V$ a linear map. Then V has the structure of a $\mathbb{F}[X]$-module via

$$\begin{aligned} - \cdot - : F[X] \times V &\longrightarrow V \\ (f, v) &\longmapsto f(\alpha)(v)\,. \end{aligned}$$

Different α's make V into different $\mathbb{F}$-modules.

(vi) If $\phi : R \to S$ is a ring homomorphism, and M is a S-module, then we can make it into an R-module via $\phi(r) \cdot -$.

Definition 3.2. If M is an R-module, a subset $N \subseteq M$ is a *submodule* if it is a subgroup of $(M, +, 0_M)$ and $r \cdot n \in N$ for all $r \in R$, $n \in N$. Write $N \leq M$.

Example. An R-submodule of R as an R-module is exactly the same as an ideal of R. An $\mathbb{F}$-submodule of an $\mathbb{F}$-module (vector space) is exactly a vector subspace.

Definition 3.3. If $N \leq M$ is an R-submodule, then the *quotient module* M/N is the set of N-cosets in $M, +, 0_M$, i.e. elements have the form $m + N$, equipped with

$$r(m + N) = rm + N\,.$$

It is an R-module.

Note that modules are different from groups and rings: we are allowed to take quotient by any submodules.

Definition 3.4. If M and N are R-modules, a function $f : M \to N$ is a *homomorphism* if

(i) it is a homomorphism as abelian groups

(ii) it is R-linear: $f(r \cdot_M m) = r \cdot_N f(m)$.

An *isomorphism* is a bijective homomorphism.

Example. If $\mathbb{F}$ is a field and V, W are $\mathbb{F}$-modules, then a $\mathbb{F}$-module homomorphism is exactly an $\mathbb{F}$-linear map of vector spaces.

Theorem 3.5 (First isomorphism theorem). Let $f : M \to N$ be an R-module homomorphism. Then

$$\begin{aligned}\ker(f) &= \{m \in M \mid f(m) = 0_N\} \leq M\\ \mathrm{im}(f) &= \{n \in N \mid n = f(m) \text{ for some } m \in M\} \leq N\,,\end{aligned}$$

and

$$\begin{aligned}\frac{M}{\ker(f)} &\longrightarrow \mathrm{im}(f)\\ m + \ker(f) &\longmapsto f(m)\end{aligned}$$

is an isomorphism of R-modules.

Theorem 3.6 (Second isomorphism theorem). Let $A, B \leq M$, then

$$A + B = \{m \in M \mid m = a + b \text{ for some } a \in A, b \in B\}\,,$$

and $A \cap B \leq M$, and

$$\frac{A + B}{A} \cong \frac{B}{A \cap B}$$

as R-modules.

As usual, we have a bijection

$$\begin{aligned}\{\text{submodules of } M/N\} &\longleftrightarrow \{\text{submodules of } M \text{ containing } N\}\\ X/N &\longmapsto X\,.\end{aligned}$$

Theorem 3.7 (Third isomorphism theorem). Let $N \leq L \leq M$, then

$$\frac{M/N}{L/N} \cong \frac{M}{L}$$

as R-modules.

We won't prove the above theorems as they are exactly the same as for groups and rings.

Definition 3.8. If M is an R-module and $m \in M$, the *annihilator* of m is

$$\mathrm{Ann}(m) \coloneqq \{r \in R \mid r \cdot m = 0\}\,.$$

For a set $S \subseteq M$, the *annihilator* of S is

$$\mathrm{Ann}(S) \coloneqq \{r \in R \mid r \cdot s = 0 \ \forall s \in S\} = \bigcap_{s \in S} \mathrm{Ann}(s)\,,$$

and in particular, the annihilator of M itself is

$$\mathrm{Ann}(M) \coloneqq \{r \in R \mid r \cdot m = 0 \ \forall m \in M\} = \bigcap_{m \in M} \mathrm{Ann}(m)\,,$$

Note that the annihilator is an ideal of R — if $r \cdot m = 0$ and $s \cdot m = 0$, then $(r + s) \cdot m = 0$ and $(tr) \cdot m = 0$ for any t.

Definition 3.9. Let M be an R-module and $m \in M$. The *submodule generated by* m is

$$Rm \coloneqq \{r \cdot m \in M \mid r \in R\}\,.$$

Consider the homomorphism

$$\begin{aligned} \phi : R &\longrightarrow M \\ r &\longmapsto rm\,. \end{aligned}$$

This is clearly a homomorphism. Then we have

$$\begin{aligned} Rm &= \mathrm{im}\,\phi \\ \mathrm{Ann}(m) &= \ker \phi\,, \end{aligned}$$

so by the first isomorphism theorem,

$$Rm \cong \frac{R}{\mathrm{Ann}(m)}\,.$$

Rings acting on modules is like groups acting on sets. We can think of this as the analogue of the orbit-stabilizer theorem.

Definition 3.10. Say an R-module is *finitely generated* if there are $m_1, \dots m_n \in M$ such that

$$Rm_1 + Rm_2 + \cdots + Rm_n = M\,.$$

Lemma 3.11. An R-module M is finitely generated $\iff$ there is a surjective R-module homomorphism $f : R^n \to M$ for some n.

Proof. If $M = Rm_1 + \cdots + Rm_n$, define f by

$$f(r_1, \dots, r_n) = \sum_i r_i m_i\,.$$

This is a surjective R-module homomorphism.

Conversely, given $f : R^n \to M$, let

$$\begin{aligned} m_1 &= f(1, 0, \dots, 0) \\ m_2 &= f(0, 1, \dots, 0) \\ &\dots . \end{aligned}$$

Given $m \in M$, have $m = f(r_1, \dots, r_n)$ as f is surjective. Note in the R-module R^n,

$$(r_1, \dots, r_n) = r_1 \cdot (1, 0, \dots, 0) + r_2 \cdot (0, 1, \dots, 0) + \dots + r_n \cdot (0, 0, \dots, 1)\,,$$

so

$$\begin{aligned} m &= f(r_1, r_2, \dots, r_n) \\ &= r_1 f(1, 0, \dots, 0) + r_2 f(0, 1, \dots, 0) + \dots + r_n f(0, 0, \dots, 1) \\ &= r_1 m_1 + r_2 m_2 + \dots + r_n m_n\,, \end{aligned}$$

so $M = Rm_1 + Rm_2 + \dots + Rm_n$. □

Proposition 3.12. If M is a finitely generated R-module and $N \leq M$, then M/N is finitely generated.

Proof. Have

$$\begin{aligned} R^n \xrightarrow[\text{surj.}]{f} M &\xrightarrow{\text{surj.}} M/N \\ m &\longmapsto m + N \end{aligned}$$

composition surjective, so M/N is finitely generated. □

Example. A counterexample: A submodule of a finitely generated module need not to be finitely generated.

Let $R = \mathbb{C}[X_1, X_2, \dots]$. R is an R-module generated by a single element $1_R \in R$. The ideal $I = (X_1, X_2, \dots)$ is a submodule. This is not finitely generated.

Suppose $I = (p_1, \dots, p_r)$, and each p_i uses only finitely many X_j. So in fact, $(p_1, \dots, p_r) \subseteq (X_1, \dots, X_s)$ for some s. To see $X_{s+1} \notin (X_1, \dots, X_s)$, observe that

$$\frac{\mathbb{C}[X_1, \dots, X_s, \dots]}{(X_1, \dots, X_s)} \cong \mathbb{C}[X_{s+1}, X_{s+2}, \dots]$$

$X_{s+1} \neq 0$ in this quotient, so $X_{s+1} \notin (X_1, \dots, X_s)$.

3.2 Direct Sums and Free Modules

Definition 3.13. If $M_1, \dots, M_k$ are R-modules, then $M_1 \oplus M_2 \oplus \dots \oplus M_k$ is the R-module given by $M_1 \times M_2 \times \dots \times M_k$, with

$$(m_1, m_2, \dots, m_k) + (m_1', m_2', \dots, m_k') = (m_1 + m_1', m_2 + m_2', \dots, m_k + m_k')$$

$$r(m_1, \dots, m_k) = (r \cdot m_1, \dots, r \cdot m_k)\,.$$

Example. $R^n = \underbrace{R \oplus R \oplus \dots \oplus R}_{n \text{ times}}$.

Definition 3.14. Let $m_1, \dots, m_k \in M$. The set $\{m_1, \dots, m_k\}$ is called *independent* if

$$\sum_{i=1}^{k} r_i m_i = 0 \implies r_1 = \dots = r_k = 0\,.$$

Definition 3.15. A subset $S \subseteq M$ *generates M freely* if

(i) S generates M

(ii) Any function $\Psi : S \to N$, where N is an R-module, extends to an R-module homomorphism $\Theta : M \to N$.

Note that if Θ_1, Θ_2 are two such extensions, consider $\Theta_1 - \Theta_2 : M \to N$. Then $\Theta_1 - \Theta_2$ sends everything in S to 0, so $S \subseteq \ker(\Theta_1 - \Theta_2) \leq M$. So the submodule generated by S lies in $\ker(\Theta_1 - \Theta_2)$ too. This is the definition of M, so $M \leq \ker(\Theta_1 - \Theta_2) \leq M$, i.e. equality holds. So $\Theta_1 - \Theta_2 = 0$, so $\Theta_1 = \Theta_2$, i.e. any such extension is unique.

Definition 3.16. A module which is freely generated by some subset is called *free*, and the subset is called a *basis*.

Proposition 3.17. For $S = \{m_1, m_2, \dots, m_k\} \subseteq M$, the following are equivalent:

(i) S generates M freely.

(ii) S generates M and S is independent.

(iii) Each element of M is uniquely expressible as $r_1 m_1 + \cdots + r_k m_k$.

Proof. The fact that (ii) and (iii) are equivalent is something we would expect from what we know from linear algebra — and in fact the proof is the same. So we only show that (i) and (ii) are equivalent.

Let S generate M freely. If S is not independent, there is a

$$\sum_{i=1}^{k} r_i m_i = 0$$

with $r_j \neq 0$. Define

$$\begin{aligned} \Psi : S &\longrightarrow R \\ m_j &\longmapsto 1_R \\ m_i &\longmapsto 0_R \quad i \neq j\,. \end{aligned}$$

There exists an R-module homomorphism $\Phi : M \to R$ extending Ψ. Then

$$\begin{aligned} 0 &= \Phi(0) \\ &= \Phi\left(\sum_{i=1}^{k} r_i m_i\right) \\ &= \sum_{i=1}^{k} r_i \cdot \Phi(m_i) \\ &= r_j\,, \end{aligned}$$

giving a contradiction. So S is independent.

Suppose every element can be uniquely written as

$$r_1 m_1 + \cdots + r_k m_k\,.$$

Given ant set function $\Psi : S \to N$, define $\Phi : M \to N$ by

$$\Phi(r_1 m_1 + \cdots + r_k m_k) = r_1 \Psi(m_1) + \cdots + r_k \Psi(m_k)\,.$$

This is well defined by uniqueness, and is clearly a homomorphism, so S generates M freely. □

Example. The set $\{2, 3\} \in \mathbb{Z}$ generates $\mathbb{Z}$. However, they do not generate $\mathbb{Z}$ freely, since $3\cdot 2+(-2)\cdot 3 = 0$. Recall from linear algebra that if a set S spans a vector space V, and it is not independent, then we can just pick some useless vectors and throw them away in order to get a basis. However, this is no longer the case in modules. Neither 2 nor 3 generate $\mathbb{Z}$.

Definition 3.18. If M is a finitely generated R-module generated by $\{m_1, \dots, m_k\}$, we have the surjection

$$\begin{aligned} f : R^k &\longrightarrow M \\ (r_1, \dots, r_k) &\longmapsto \sum_i r_i m_i \,. \end{aligned}$$

The *relation module* for these generators is $\ker(f) \leq R^k$.

Definition 3.19. We say an R-module M is *finitely presented* if there is a finitely generating set $\{m_1, \dots, m_k\}$ such that the associated relation module is also finitely generated.

Being finitely presented means I can tell you everything about the module with a finite amount of paper.

More precisely, if $\{m_1, \dots, m_k\}$ generate M and $\{n_1, \dots, n_l\}$ generate the relation module $\ker f$, then each

$$n_i = (r_{i1}, \dots, r_{ik})$$

corresponds to a relation

$$r_{i1}m_1 + r_{i2}m_2 + \cdots + r_{ik}m_k = 0 \in M \,.$$

So M is the module generated by writing down R-linear combinations of $\{m_1, \dots, m_k\}$, and saying two elements are the same if they are related to one another by these relations. Since there are only finitely many generators and finitely many such relations, we can specify the module with a finite amount of information via

$$M \cong \frac{R^k}{\ker(f)} \,.$$

A question is, if $n \neq m$, then are R^n and R^m the same? They must be different if R is a field, because vector spaces have well-defined basis and dimensions. To show that this is true for a general ring, we need a few constructions.

Proposition 3.20. If $I \lhd R$ an ideal and M is an R-module, and

$$IM = \left\{ \sum_i a_i m_i \in M \mid a_i \in I \,,\ m_i \in M \right\} \leq M \,,$$

then M/IM is an R/I-module.

Proof. We know M/IM is an R-module. If $b \in I$ then

$$b \cdot (m + Im) = b \cdot m + Im = 0 + Im \,,$$

so we can make M/IM into an R/I-module via

$$(r + I)(m + IM) = rm + Im \,.$$

Proposition 3.21. Every non-zero ring has a maximal ideal.

Proof. An ideal of R is proper $\iff 1_R \notin I$, so a union of proper ideals is proper. It follows from Zorn's lemma that there is a maximal proper ideal. (Zorn's lemma says if an arbitrary union of increasing things is still a thing, then there is a maximal such thing, roughly. We are not going to prove it.) □

Proposition 3.22 (Invariance of dimension). If $R \neq 0$, $R^k \cong R^l$, then $k = l$.

Proof. Let I be a maximal ideal of R (exists by the previous proposition). If $R^k \cong R^l$, then

$$\left(\frac{R}{I}\right)^k = \frac{R^k}{IR^k} \cong \frac{R^l}{IR^l} = \left(\frac{R}{I}\right)^l$$

as R-modules. As I is maximal, R/I is a field. So this is a vector space isomorphism. By linear algebra, $k = l$. □

3.3 Matrices over Euclidean Domains

Until further notice, R is a Euclidean domain, with a Euclidean function $\phi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$. We know that in a Euclidean domain, $\gcd(a, b)$ exists for all $a, b \in R$ and there are $x, y \in R$ such that $ax + by = \gcd(a, b)$.

Definition 3.23. *Elementary row operations* on a $m \times n$ matrix A with entries in R are

(ER1) Add c times the i^{th} row to the j^{th} row, $i \neq j$.

This can be done by left multiplication of

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & c & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix},$$

where c is at the i^{th} column of the j^{th} row.

(ER2) Swap the i^{th} and the j^{th} row, $i \neq j$.

This can be done by the left multiplication of

$$\begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & & & 1 & & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & 1 & & & & 0 & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{pmatrix},$$

where we changed the 0's and 1's at the i^{th} and j^{th} rows/columns.

(ER3) Multiply the i^{th} row by a unit $c \in R$.

This is the left multiplication of

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}.$$

Notice that if R is a field, then we can multiply any row by any non-zero number, since they are all units.

Similarly we have *elementary column operations* (EC1)-(EC3), all given by right multiplication by analogous matrix.

Definition 3.24. Two $m \times n$ matrices A, B are *equivalent* if there is a sequence of elementary row and column operations getting between them. These are invertible P, Q such that

$$B = QAP^{-1}$$

Our goal is to find, for each matrix, a matrix equivalent to it that is as simple as possible. Recall from Linear Algebra that if R is a field, then we can put any matrix into the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

via elementary row and column operations. This is no longer true when working with rings. For example, over $\mathbb{Z}$, we cannot put the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

into that form, since no operation can turn 2 into 1.

Theorem 3.25 (Smith normal form). An $m \times n$ matrix A over a Euclidean domain R is equivalent to

$$\begin{pmatrix} d_1 & & & & & & \\ & d_2 & & & & & \\ & & \ddots & & & & \\ & & & d_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

with $d_i \neq 0$ and $d_1 \mid d_2 \mid d_3 \mid \cdots \mid d_r$. This is known as the *Smith normal form* of the matrix, and the d_i are the *invariant factors* of A. They are unique up to associates.

Proof. If A is zero then done.

If not, then A has some non-zero entry. By permuting rows and columns, can suppose $A_{11} \neq 0$. Our strategy is to reduce A_{11} as much as possible.

(i) If there is a A_{1j} not divisible by A_11, then we have

$$A_{1j} = q \cdot A_{11} + r$$

with $\phi(r) < \phi(A_{11})$. Subtracting q times the first column from the j^{th} column leaves r in the position $(1, j)$, so permute the 1^{st} and the j^{th} column to put r in position $(1, 1)$. This has strictly reduced $\phi(A_{11})$.

(ii) If there is an A_{i1} not divisible by A_{11}, can do the analogous thing to strictly reduce the value of ϕ.

Repeatedly doing so, we can arrange so that all A_{1j} and A_{il} are divisible by A_{11}. Subtracting appropriate multiples of the first row/columns from the others, we can arrange that

$$A = \left(\begin{array}{c|ccc} d & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & C & \\ 0 & & & \end{array}\right)$$

for some $d \neq 0$ and $(n-1) \times (m-1)$ matrix C.

(iii) If there is an entry of C not divisible by d, say $d \nmid A_{ij}$, then write $A_{ij} = q \cdot d + r$, $r \neq 0$ and $\phi(r) < \phi(d)$.

Add column 1 to column j, subtract q times first row from i^{th} row: this makes r in (i, j) position. Permute rows and columns to put r in the $(1, 1)$ position. This has again strictly reduced $\phi(A_{11})$. Repeat (i) and (ii) to make the rest of the 1^{st} row/column to be 0. As it strictly reduces the ϕ value, this can only happen finitely many times. So can assume that d divides all entries of C.

Now apply the same algorithm to the matrix C. By induction,

$$C \sim \begin{pmatrix} d_2 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

As $d \mid$ all entries of C, it also divides all R-linear combinations of the entries of C, so it divides all d_i. So $d \equiv d_1$ and $d_1 \mid d_2 \mid \cdots \mid d_r$ by induction. □

Recall that the d_i are called the invariant factors. So it would be nice if we can prove that the d_i are indeed invariant. It is not clear from the algorithm that we will always end up with the same d_i.

To study the uniqueness of the invariant factors of a matrix A, we relate them to other invariants, which involves minors.

Definition 3.26. A $k \times k$ *minor* of a matrix A is the determinant of a $k \times k$ submatrix of A, i.e. matrix removed entries in all but k rows and k columns.

Any given matrix has many minors, since we get to decide which rows and columns we can throw away. The idea is to consider the ideal generated by all the minors of matrix.

Definition 3.27. For a matrix A with entries in R, the k^{th} *Fitting ideal* is

$$\mathrm{Fit}_k(A) = (\text{All } k \times k \text{ minors of } A) \lhd R\,.$$

Lemma 3.28. If A is equivalent to B, then

$$\mathrm{Fit}_k(A) = \mathrm{Fit}_k(B)$$

for all k.

Proof. Need to check that (ER1)–(ER3), (EC1)–(EC3) do not change the Fitting ideal. We will only prove for (ER1) since the situation for (EC1) is the same, and the other four cases are much easier.

Fix a $k \times k$ submatrix C of A. Consider (ER1) by adding c times the i^{th} row to the j^{th} row.

If the j^{th} row of A is not in C, then C is trivially unchanged by this move.

If both the i^{th} and j^{th} rows of A are in C, then C is changed to C', with C' is obtained from C by a row operation, so $\det C = \det C'$.

If the j^{th} row is in C but the i^{th} row is not, then C is changed to C' with j^{th} row

$$(c_{j1} + cf_1, c_{j_2} + cf_2, \dots, c_{jk} + cf_k)\,,\ c \in R\,.$$

$$\begin{pmatrix} & C & \\ \hdashline f_1 & \cdots & f_k \end{pmatrix}$$

Expanding $\det C'$ using this row, we see

$$\det(C') = \det C \pm c \cdot \det \underbrace{\begin{pmatrix} \text{matrix obtained by replacing} \\ \text{the } j^{\text{th}} \text{ row of } C \text{ by } (f_1 \ \dots \ f_k) \end{pmatrix}}_{\text{up to permuting rows, a submatrix of } A}\,.$$

Therefore, $\det C' \in \operatorname{Fit}_k(A)$.

Hence, $\operatorname{Fit}_k(A') \subseteq \operatorname{Fit}_k(A)$. This is in fact a equality as (ER1) is invertible. □

If A has the Smith normal form

$$\begin{pmatrix} d_1 & & & & & & \\ & d_2 & & & & & \\ & & \ddots & & & & \\ & & & d_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix},$$

then

$$\operatorname{Fit}_k(A) = (d_1 d_2 \dots d_k)\,.$$

This is clear once we notice that the only possible contributing minors are from the diagonal submatrices, and the minor from the top left square submatrix divides all other diagonal ones.

This shows the product $d_1 \dots d_k$ depended (up to associates) only on A, so the d_k depend (up to associates) only on A too.

Example. Consider

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$$

over $\mathbb{Z}$.

$$\begin{aligned} &\operatorname{Fit}_1(A) = (2, -1, 2, 1) = (1) && \text{so } d_1 = \pm 1 \\ &\operatorname{Fit}_2(A) = (5) && \text{so } d_1 d_2 = \pm 5 \implies d_2 = \pm 5\,. \end{aligned}$$

So A has a Smith normal form

$$\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}.$$

Lemma 3.29. Let R be a principal ideal domain. Any submodule of R^m is generated by at most m elements.

Proof. Let $N \leq R^m$ be a submodule. Consider the ideal

$$I = \{r \in R \mid \exists r_2, \dots, r_m \text{ such that } (r, r_2, \dots, r_m) \in N\} \lhd R\,.$$

As R is a PID, $I = (a)$ for some $a \in I$. Choose $n_1 = (a, a_2, \dots, a_m) \in N$. If $(r_1, \dots, r_m) \in N$, then $r_1 = r \cdot a$ for some $r \in R$. Consider

$$(r_1, r_2, \dots, r_m) - r(a, a_2, \dots, a_m) = (0, r_2 - a_2, \dots, r_n - a_n) \in N\,.$$

This lies in $N' = N \cap (\{0\} \times R^{m-1}) \leq R^{m-1}$. Therefore, everything in N can be written as a multiple of n_1 plus something in N'. By induction can suppose N' is generated by at most $m-1$ elements, so there are $n_2, \dots, n_m \in N'$ generating N', so $\{n_1, n_2, \dots, n_m\}$ generate N. □

This only tells us if we have a submodule of R^m, then it can be generated by at most m generators, but it did not tell us how to find them — and those generators may generate the submodule in some horrible ways.

The next theorem tells us a good way finding them.

Theorem 3.30. Let R be an ED, $N \leq R^m$ a submodule. Then there is a basis $v_1, \dots, v_m$ of R^m such that N is generated by $d_1 v_1, \dots, d_r v_r$ for some $0 \leq r \leq m$ and $d_1 \mid d_2 \mid \cdots \mid d_r$.

Proof. By the above lemma, there are $x_1, \dots, x_n \in N$ generating it, with $n \leq m$. Each x_i is an element of R^m, so they can be assembled as an $m \times n$ matrix

$$\begin{pmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{pmatrix} = A\,.$$

Then we can reduce it into a Smith normal form

$$\begin{pmatrix} d_1 & & & & & & & \\ & d_2 & & & & & & \\ & & \ddots & & & & & \\ & & & d_r & & & & \\ & & & & 0 & & & \\ & & & & & \ddots & & \\ & & & & & & 0 & \\ & & & & & & 0 & \\ & & & & & & \vdots & \\ & & & & & & 0 & \end{pmatrix}.$$

This can be obtained by elementary row and column operations $P^{-1}AQ$. Each row operation corresponds to a change of basis of R^m, and each column operation is a change in generators of N. In the new basis for R^m, N is generated by $d_1 v_1, \dots, d_r v_r$ as required. □

Corollary. Let R be a Euclidean domain. A submodule of R^m is free of rank $\leq m$. In other words, the submodule of a free module is free, and of a smaller (or equal) rank.

Proof. Continuing with the notation above. If $d_1 v_1, \dots, d_r v_r$ are linearly dependent, then so do $v_1, \dots, v_r$, but v_i are the basis so they don't. Therefore, $d_1 v_1, \dots, d_r v_r$ are linearly independent, so they generate the submodule freely. So

$$N \cong R^r\,.$$

□

Note that this is not true for all rings. For example, $(2, X) \lhd \mathbb{Z}[X]$ is a submodule of $Z[X]$, but because it is not a principal ideal it cannot be isomorphic to $Z[X]$.

Theorem 3.31 (Classification of finitely-generated modules over ED). Let R be a Euclidean domain, M a finitely-generated R-module. Then

$$M \cong \frac{R}{(d_1)} \oplus \frac{R}{(d_2)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R \oplus \cdots \oplus R$$

for some $d_i \neq 0$ and $d_1 \mid d_2 \mid \cdots \mid d_r$.

Proof. As M is finitely generated, we have a surjection $\phi : R^m \to M$. So by the first isomorphism theorem,

$$M \cong \frac{R^m}{\ker \phi}\,.$$

By the last theorem, we can choose a new basis of R^m, $v_1, \ldots, v_m$ such that $\ker \phi$ is generated by $d_1 v_1, \ldots, d_r v_r$, $0 \leq r \leq m$, $d_1 \neq 0$ and $d_1 \mid \cdots \mid d_r$. So

$$\begin{aligned} M &\cong \frac{R^m}{((d_1, 0, \cdots, 0), (0, d_2, \cdots, 0), \cdots, (0, 0, \cdots, d_r, 0, \cdots, 0))} \\ &= \frac{R}{(d_1)} \oplus \frac{R}{(d_2)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus \underbrace{R \oplus \cdots \oplus R}_{(m-r) \text{ copies}}\,. \end{aligned}$$

□

This is particularly useful in the case where $R = Z$, where R-modules are abelian groups.

Example. Let $R = \mathbb{Z}$, a Euclidean domain. Let A be the abelian group generated by a, b, c subjected to

$$\begin{aligned} 2a + 3b + \ \ c &= 0 \\ a + 2b \qquad &= 0 \\ 5a + 6b + 7c &= 0\,. \end{aligned}$$

Then we have

$$A = \frac{\mathbb{Z}^3}{((2, 3, 1), (1, 2, 0), (5, 6, 7))}\,.$$

To determine A up to isomorphism, we can put

$$\begin{pmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 1 & 0 & 7 \end{pmatrix}$$

into Smith normal form. Have

$$\begin{aligned} \mathrm{Fit}_1(M) &= (1, \ldots) = (1) & d_1 &= 1 \\ \mathrm{Fit}_2(M) &= (1, \ldots) = (1) & d_2 &= 1 \\ \mathrm{Fit}_3(M) &= (\det M) = (3) & d_3 &= 3\,. \end{aligned}$$

Therefore,

$$A \cong \frac{\mathbb{Z}}{(1)} \oplus \frac{\mathbb{Z}}{(1)} \oplus \frac{\mathbb{Z}}{(3)} = \frac{\mathbb{Z}}{(3)}\,.$$

Corollary (Structure theorem for finitely generated abelian groups). Any finitely generated abelian group is isomorphic to

$$C_{d_1} \times C_{d_2} \times \cdots \times C_{d_r} \times C_\infty \times \cdots \times C_\infty\,,$$

with $d_i \neq 0$, $d_1 \mid d_2 \mid \cdots \mid d_r$.

Proof. Abelian groups are $\mathbb{Z}$-modules. $\mathbb{Z}$ is an ED, so apply the classification and use

$$\frac{\mathbb{Z}}{(d)} = C_d\,,\ \frac{\mathbb{Z}}{(0)} = C_\infty\,.$$

□

Corollary (Classification of finite abelian groups). Any finite abelian group is isomorphic to

$$C_{d_1} \times C_{d_2} \times \cdots \times C_{d_r}$$

with $d_i \neq 0$ and $d_1 \mid d_2 \mid \cdots \mid d_r$.

Recall that we were also to decompose a finite abelian group into products of the form C_{p^k}, where p is a prime, and we said it was just the Chinese remainder theorem. This is true for modules as well.

Lemma 3.32 (Chinese remainder theorem). Let R be a Euclidean domain, $a, b \in R$ such that $\gcd(a, b)$ is a unit, then

$$\frac{R}{(ab)} \cong \frac{R}{(a)} \oplus \frac{R}{(b)}\,.$$

Proof. Consider

$$\begin{aligned}\phi : \frac{R}{(a)} \oplus \frac{R}{(b)} &\longrightarrow \frac{R}{(ab)}\\ (r_1 + (a), r_2 + (b)) &\longmapsto br_1 + ar_2 + (ab)\,.\end{aligned}$$

This is an R-module homomorphism as long as it is well-defined, so we need to check that. Suppose $(r_1 + (a), r_2 + (b)) = (r_1' + (a), r_2' + (b))$, i.e.

$$\begin{aligned} r_1 &= r_1' + ax \quad x \in (a)\,,\\ r_2 &= r_2' + by \quad y \in (b)\,.\end{aligned}$$

Then

$$\begin{aligned} br_1 + ar_2 &= b(r_1' + ax) + a(r_2' + by)\\ &= br_1' + ar_2' + ab(x + y)\\ &= br_1' + ar_2' \pmod{(ab)}\,,\end{aligned}$$

so well defined.

Now show that it is surjective. As $\gcd(a, b)$ is a unit, there exist x and y with $ax + by = 1$ by Euclidean algorithm. Then

$$\begin{aligned}\phi(y + (a), x + (b)) &= by + ax + (ab)\\ &= 1 + (ab)\,,\end{aligned}$$

and so

$$\begin{aligned}\phi(ry + (a), rx + (b)) &= r(1 + (ab))\\ &= r + (ab)\end{aligned}$$

for any $r \in R$, so surjective.

Finally need to show that it is injective. Suppose $\phi(r_1 + (a), r_2 + (b)) = 0$, i.e. $br_1 + ar_2 + (ab) = 0$, i.e. $br_1 + ar_2 \in (ab)$, then $br_1 + ar_2 = abx$ for some x. So $a \mid br_1$ and $b \mid ar_2$. As a, b are coprime, can only have $a \mid r_1$ and $b \mid r_2$, so

$$(r_1 + (a), r_2 + (b)) = (0 + (a), 0 + (b))\,.$$

The kernel is trivial, so injective. □

Theorem 3.33 (Primary decomposition theorem). Let R be a Euclidean domain, M a finitely generated R-module, then

$$M \cong N_1 \oplus N_2 \oplus \cdots \oplus N_t \,,$$

where each N_i is either R or is $R/(p^n)$ for some $n \geq 1$ and prime $p \in R$.

Proof. Already know

$$M \cong \frac{R}{(d_1)} \oplus \frac{R}{(d_2)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R \oplus \cdots \oplus R$$

with $d_i \neq 0$, $d_1 \mid d_2 \mid \cdots \mid d_r$. Enough to know each $R/(d_i)$ is a sum of $R/(p^n)$'s.

Let $d_i = p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$ with p_i distinct primes. The lemma shows

$$\frac{R}{(d_i)} = \frac{R}{(p_1^{n_1})} \oplus \frac{R}{(p_2^{n_2})} \oplus \cdots \oplus \frac{R}{(p_k^{n_k})} \,.$$

□

3.4 Modules over $\mathbb{F}[X]$ and Normal Forms for Matrices

We next want to consider the Jordan normal form. This is less straightforward, since considering V directly as an F module would not be too helpful (since that would just be pure linear algebra). Instead, we use the following trick.

For a field $\mathbb{F}$, $\mathbb{F}[X]$ is a ED so the last section applies. Recall if V is a $\mathbb{F}$-vector space, and $\alpha : V \to V$ is a linear endomorphism, we can consider V as an $\mathbb{F}[X]$-module via

$$\begin{aligned} \mathbb{F}[X] \times V &\longrightarrow V \\ (f, v) &\longmapsto f(\alpha)(v) \,. \end{aligned}$$

Call this $\mathbb{F}[X]$-module V_α.

Lemma 3.34. if V is a finite dimensional $\mathbb{F}$-vector space, then V_α is a finitely generated $\mathbb{F}[X]$-module.

Proof. V finite dimensional $\iff$ V finitely generated $\mathbb{F}$-module, and a $\mathbb{F}$-generating set is also a $\mathbb{F}[X]$-generating set, since $\mathbb{F} \leq \mathbb{F}[X]$. □

Examples.

(i) Suppose $V_\alpha \cong \mathbb{F}[X]/(X^r)$ as a $\mathbb{F}[X]$-modules. Then in particular they are isomorphic as $\mathbb{F}$-modules (since being a map of $\mathbb{F}$-modules has fewer requirements than being a map of $\mathbb{F}[X]$-modules).

Under this bijection, the elements $1, X, \dots X^{r-1} \in \mathbb{F}[X]/(X^r)$ form a vector space basis for V_α. Viewing $\mathbb{F}[X]/(X^r)$ as an $\mathbb{F}$-vector space, the action of X has the matrix

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} .$$

We also know that in V_α, the action of X is by definition the linear map α, so under this basis, α also has the matrix

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} .$$

It is a Jordan normal block of size r with eigenvalue 0.

(ii) Suppose
$$V_\alpha \cong \mathbb{F}[X]/((X-\lambda)^r)$$
for some $\lambda \in \mathbb{F}$. Then the linear map $\beta = \alpha - \lambda \operatorname{id} : V \to V$ has $V_\beta \cong \mathbb{F}[X]/(X^r)$, i.e. there is a basis where α is represented as
$$\begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 1 & \lambda & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda \end{pmatrix}.$$
This is a Jordan normal block of size r with eigenvalue λ.

(iii) Suppose $V_\alpha \cong \mathbb{F}[X]/(f)$ with $f = a_0 + a_1X^1 + \cdots + a_{r-1}X^{r-1} + X^r$. We still have the basis $1, X, \ldots, X^{r-1}$. In this basis, α is represented as
$$c(f) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{r-1} \end{pmatrix}.$$
This is called the *companion matrix* for the monic polynomial f.

Theorem 3.35 (Rational canonical form). Let $\alpha : V \to V$ be an endomorphism of a finite dimensional $\mathbb{F}$-vector space V. Then
$$V_\alpha \cong \frac{\mathbb{F}[X]}{(f_1)} \oplus \cdots \oplus \frac{\mathbb{F}[X]}{(f_r)}$$
with $f_i \neq 0$, $f_1 \mid \cdots \mid f_r$. So there is a basis for V such that α is given by
$$\begin{pmatrix} c(f_1) & 0 & \cdots & 0 \\ 0 & c(f_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c(f_s) \end{pmatrix}.$$

Proof. Use classification of finitely generated modules over ED. No $\mathbb{F}[X]$ arises in this sum as $\mathbb{F}[X]$ is infinite dimensional over $\mathbb{F}$ while V is finite dimensional. □

Observations.

(i) This is really a canonical form. The Jordan normal form is not canonical, since we can move the blocks around. The structure theorem determines the factors f_i up to units, and once we require them to be monic, there is no choice left.

(ii) If α was given by a square matrix A, this says A is conjugate to the above matrix $\operatorname{diag}(c(f_1), c(f_2), \ldots, c(f_r))$.

(iii) The minimal polynomial of α is f_r.

(iv) The characteristic polynomial of α is $\prod_{i=1}^r f_i$.

Recall we had a different way of decomposing a module over a Euclidean domain, namely the prime decomposition, and this gives us the Jordan normal form.

Before we can use that, we need to know what the primes are. This is why we need to work over $\mathbb{C}$.

Lemma 3.36. The primes in $\mathbb{C}[X]$ are $X - \lambda$.

Proof. Constants in $\mathbb{C}$ are all units, so they are not prime. For polynomials, only ones of the form $X - \lambda$ are irreducible by the fundamental theorem of algebra. □

Theorem 3.37 (Jordan normal form). Let $\alpha : V \to V$ be an endomorphism of a $\mathbb{C}$-vector space. Then

$$V_\alpha \cong \frac{\mathbb{C}[X]}{((X-\lambda_1)^{a_1})} \oplus \frac{\mathbb{C}[X]}{((X-\lambda_2)^{a_2})} \oplus \cdots \oplus \frac{\mathbb{C}[X]}{((X-\lambda_s)^{a_s})}$$

as $\mathbb{C}[X]$-modules, where $\lambda_i \in \mathbb{C}$ do not have to be distinct. So there is a basis of V in which α is represented by

$$\begin{pmatrix} J_{a_1}(\lambda_1) & & & \\ & J_{a_1}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{a_s}(\lambda_s) \end{pmatrix},$$

where

$$J_m(\lambda) \coloneqq \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 1 & \lambda & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & \lambda \end{pmatrix}$$

is a $m \times m$ matrix known as the *Jordan normal block*.

Proof. Apply the prime decomposition theorem to V_α. Then all primes are of the form $X - \lambda$.

Observations.

(i) They are unique up to reordering of blocks.

(ii) The minimal polynomial is

$$\prod_\lambda (X-\lambda)^{a_\lambda},$$

where a_λ is the size of the largest λ-block.

(iii) The characteristic polynomial is

$$\prod_i (X-\lambda_i)^{a_i}.$$

(iv) The number of λ-blocks is the size of the λ-eigenspace.

3.5 *Conjugacy

Again, for a vector space V and an endomorphism $\alpha : V \to V$, write V_α for the corresponding $\mathbb{F}[X]$-module.

Lemma 3.38. If $\alpha : V \to V$ and $\beta : W \to W$ are two endomorphisms of two vector spaces, then $V_\alpha \cong W_\beta$ as $\mathbb{F}[X]$-modules $\iff$ there is an isomorphism $\gamma : V \to W$ such that $\gamma^{-1}\beta\gamma = \alpha$.

Proof. Let $\phi : V_\alpha \to W_\beta$ be the $\mathbb{F}[X]$-module isomorphism. Let $v \in V$.

$$\phi(\alpha(v)) = \phi(X \cdot v) = X \cdot \phi(v) = \beta(\phi(v))$$

$\forall v \in V$. Writing $\gamma : V \to W$ for the underlying linear map of ϕ, this says $\gamma\alpha = \beta\gamma$, so $\alpha = \gamma^{-1}\beta\gamma$.

Conversely, let $\gamma : V \to W$ be a linear isomorphism such that $\gamma^{-1}\beta\gamma = \alpha$. We now claim that the corresponding $\phi : V_\alpha \to V_\beta$ is an $\mathbb{F}[X]$-module isomorphism. We just need to check

$$\begin{aligned} \phi(f \cdot v) &= \gamma(f(\alpha)v) \\ &= \gamma(a_0 + a_1\alpha + \cdots + a_n\alpha^n)(v) \\ &= \gamma(a_0 v) + \gamma(a_1 \alpha v) + \cdots + \gamma(a_n \alpha^n v) \\ &= (a_0 + a_1\beta + \cdots + a_n\beta^n)(\gamma(v)) \\ &= f \cdot \phi(v). \end{aligned}$$

□

In particular, if $W = V$, α and β are conjugate $\iff$ $V_\alpha \cong V_\beta$ as $\mathbb{F}[X]$-modules.

So classifying linear maps up to conjugation is the same as classifying modules. We can re-interpret this a little bit, using the classification of finitely-generated modules.

Corollary. There is a bijection

$$\left\{\begin{array}{c}\text{conjugacy classes of}\\ n\times n \text{ matrices}\\ \text{over } \mathbb{F}\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{sequences of monic polynomials}\\ d_1,\dots,d_r\in\mathbb{F}[X] \text{ such that } d_1 \mid \cdots \mid d_r\\ \deg(d_1\dots d_r)=n\end{array}\right\}$$

$$\Updownarrow$$

$$\frac{\mathbb{F}[X]}{(d_1)}\oplus\frac{\mathbb{F}[X]}{(d_2)}\oplus\cdots\oplus\frac{\mathbb{F}[X]}{(d_r)}$$

has dimension n over $\mathbb{F}$

Example. Suppose we want to study the conjugacy classes in the group $\mathrm{GL}_2(\mathbb{F})$, i.e. 2×2 invertible matrices. Need polynomials $d_1,\dots,d_r$ with $d_1 \mid \cdots \mid d_r$ and $d_1\dots d_r$ has degree 2.

So, either

(i) $\deg(d_1)=2$.

$$\frac{\mathbb{F}[X]}{(d_1)} \cong \frac{\mathbb{F}[X]}{(X^2+a_1X+a_2)}\,,$$

or

(ii) $\deg(d_1)=\deg(d_2)=1$, and $d_1 \mid d_2$, so $d_1=d_2=X-\lambda$.

$$\frac{\mathbb{F}[X]}{(X-\lambda)}\oplus\frac{\mathbb{F}[X]}{(X-\lambda)}\,.$$

Therefore, any $A\in\mathrm{GL}_2(\mathbb{F})$ is conjugate to either

$$\begin{pmatrix}0 & -a_2\\ 1 & -a_1\end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}$$

$$a_2=\det(A) \qquad\qquad \det(A)=\lambda^2$$

$$a_1=-\operatorname{tr}(A) \qquad\qquad \operatorname{tr}(A)=2\lambda\,.$$

How unique are these? If

$$\begin{pmatrix}0 & -a_2\\ 1 & -a_1\end{pmatrix} \sim \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}\,,$$

then they have the same determinant and same trace, so $a_2=\lambda^2$ and $a_1=-2\lambda$, so

$$X^2+a_1X+a_2 = X^2-2\lambda X+\lambda = (X-\lambda)^2\,.$$

This is the polynomial of a Jordan normal block — $\mathbb{F}[X]/(X-\lambda)^2$

$$\begin{pmatrix}\lambda & 0\\ 1 & \lambda\end{pmatrix} \not\sim \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}\,,$$

so the two cases are disjoint.

Let us look at the first case in more detail. If $d_1 = X^2 + a_1X + a_2$ is reducible, so factors as

$$(X-\lambda)(X-\lambda) \qquad \text{or} \qquad (X-\lambda)(X-\lambda)\,,\ \lambda \neq \mu\,,$$

giving

$$\begin{pmatrix}\lambda & 0\\ 1 & \lambda\end{pmatrix} \qquad\qquad \begin{pmatrix}\lambda & 0\\ 0 & \mu\end{pmatrix}.$$

Therefore, any $A \in \mathrm{GL}_2(\mathbb{F})$ is conjugated to one of

$$\underset{X^2+a_1X+a_2 \text{ irreducible}}{\begin{pmatrix}0 & -a_2\\ 1 & -a_1\end{pmatrix}} \qquad \underset{\lambda \in \mathbb{F}\setminus\{0\}}{\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}} \qquad \underset{\lambda,\mu \in \mathbb{F}\setminus\{0\}}{\begin{pmatrix}\lambda & 0\\ 0 & \mu\end{pmatrix}}.$$

Example. Let $\mathbb{F} = \mathbb{Z}/3$. Let's work out the conjugacy classes in the finite group $\mathrm{GL}_2(\mathbb{Z}/3)$.

What $X^2 + a_1X + a_2 \in \mathbb{Z}/3[X]$ are irreducible? There are $3 \times 3 = 9$ in total. The reducible ones have either repeated roots (3 choices) or 2 distinct roots ($\binom{3}{2} = 3$ choices), so there are 3 irreducible polynomials. Have $X^2 + 1$, $X^2 + X + 2$, $X^2 + 2X + 2$ irreducible. These do not have roots in $\mathbb{Z}/3 = \{0, 1, 2\}$. So the conjugacy classes are represented by

	$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}0 & -2\\ 1 & -1\end{pmatrix}$	$\begin{pmatrix}0 & -2\\ 1 & -2\end{pmatrix}$	$\begin{pmatrix}\lambda & 0\\ 1 & \lambda\end{pmatrix}$	$\begin{pmatrix}\lambda & 0\\ 0 & \mu\end{pmatrix}$
				$\lambda \neq 0$	$\lambda, \mu \neq 0$
# of conj. classes	1	1	1	2	3

There are 8 conjugacy classes in total.

Recall $|\mathrm{GL}_2(\mathbb{Z}/3)| = (3^2-1)(3^2-3) = 48 = 2^4 \cdot 3$. By Sylow's theorem there is a subgroup of order $2^4 = 16$. Looking at the conjugacy classes, see the first three classes have elements of order 4, 8, 8 respectively. The elements in the fourth conjugacy class have order 3. The element in the third conjugacy class have order 2, except for the identity matrix ($\lambda = \mu = 1$), which has order 1. There is no element of order 16, so the Sylow 2-subgroup is not cyclic.

To construct the Sylow 2-subgroup, we start by choosing an element of order 8, say

$$B = \begin{pmatrix}0 & 1\\ 1 & 2\end{pmatrix}.$$

To make a subgroup of order 16, a sensible guess would be to take an element of order 2, but that doesn't work, since B^4 will give you the element of order 2. Instead, we pick

$$A = \begin{pmatrix}0 & 2\\ 1 & 0\end{pmatrix}.$$

We notice

$$A^{-1}BA = \begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}0 & 2\\ 1 & 0\end{pmatrix} = \begin{pmatrix}1 & 2\\ 0 & 2\end{pmatrix}\begin{pmatrix}0 & 2\\ 1 & 0\end{pmatrix} = \begin{pmatrix}2 & 2\\ 2 & 0\end{pmatrix} = B^3\,.$$

So this is a bit like the dihedral group.

We know that

$$\langle B\rangle \lhd \langle A, B\rangle\,.$$

Also, we know $|\langle B\rangle| = 8$. So if we can show that $\langle B\rangle$ has index 2 in $\langle A, B\rangle$, then this is the Sylow 2-subgroup. By the second isomorphism theorem, we have

$$\frac{\langle A, B\rangle}{\langle B\rangle} \cong \frac{\langle A\rangle}{\langle A\rangle \cap \langle B\rangle}\,.$$

Note

$$\langle A\rangle \cap \langle B\rangle = \left\langle \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \right\rangle \cong C_2\,.$$

We also know $\langle A\rangle \cong C_4$. So we know

$$\frac{|\langle A, B\rangle|}{|\langle B\rangle|} = 2\,.$$

So $|\langle A, B\rangle| = 16$. So this is the Sylow 2-subgroup we want. in fact, it is

$$\langle A, B \mid A^4 = B^8 = e\,,\ A^{-1}BA = B^3\rangle\,.$$

We call this the *semi-dihedral group* of order 16, because it is a bit like a dihedral group.

Note that finding this subgroup was purely guesswork. There is no method to know that A and B are the right choices.