Groups, Rings and Modules

brookes@dpmms.cam.ac.uk

July 5, 2008

Introduction

0.1 Groups

This section continues from the IA course "Algebra and Geometry". We concentrate on finite groups, and there is some interplay between counting arguments and the structure of the groups. The main result from this section is Sylow's T: Let G be a finite group and $|G| = p^a m$ w/ P prime and m coprime to p, then there is a subgroup of order p^a, all such subgroups are conjugate, and the number of such subgroups $n_p \equiv 1(p)$ and $n_p \mid |G|$ (and thus m). This tells us to consider subgroups of prime power order.

This section leads towards courses in Representation Theory and Galois Theory

0.2 Rings

A ring R has two operations $+$ and $\times$, with distrutivity. Some examples are the fields $\mathbb{R}, \mathbb{Q}, \mathbb{C}$, but also $\mathbb{Z}$, and polynomial rings such as $\mathbb{C}[X], \mathbb{Z}[X]$. In this course we will concentrate on commutative rings and in particular those of interest to number theory and algebraic geometry. For number theory we look at rings like $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\} \subset \mathbb{Q}[i] \subset \mathbb{C}$, the Gaussian integers; more generally, in number theory we look at fields obtained by adjoining roots of integral polys e.g. $x^2 + 1$ to the rationals. Inside these there are rings that play the role of the indtegers. The Gaussian integers are well behaved; in particular, Euclid's algorithm works, and hence we have unique prime factorization. However, in general this is not the case. Study of these rings leads to the Number Theory course and also Coding and Cryptography.

In algebraic geometry we look at sets of zeroes of polys of n vars, e.g. in $\mathbb{C}^n$, we study polys of $\mathbb{C}[X_1, \ldots, X_n]$. For $f(X_1, \ldots, X_n), f_2(X_1, \ldots, X_n), \ldots$ we want to study the set of common zeroes. We will proove Hilbert's Basis T, which $\Rightarrow$ it is equal to the set of zeroes of a finite subset of these polys.

0.3 Modules

These are a generalization of vec sps; scalars are taken from a ring instead of a field. The theory for these is more complicated, but for "nice" rings there are structure thms; we concentrate on modules with scalars from a ring where Euclid's algorithm works, especially $\mathbb{Z}$ and $\mathbb{Q}[X]$.

The $\mathbb{Z}$-modules are $\equiv$ the abelian groups, and important for Algebraic Topology. $\mathbb{C}[X]$-modules give us the JNF, a good form for mats representing endomorphisms of vec sps.

Books

Fraleigh's is the best introduction to this topic, and there is little difference between its many editions. Hartley & Hawkes' (out of print) "Rings and Modules" is also recommended, or for a more modern work, Cameron's "Introduction to Algebra". Finally there is also the American Artin, whose book it is worth getting a later rather than earlier edition of. There are four example sheets for this course plus one mock tripos sheet; the lecture split between the three main sections is approximately 8/10/6. In general, feedback from previous years is that this course is interesting but contains a lot of content in comparison with other IB courses.

1 Groups

1.1 Basic Concepts

Definition

A set G is a group wrt a binary operator $\therefore G \times G \to G$ if it satisfies the axioms of closure ($g_1, g_2 \in G \Rightarrow g_1 g_2 \in G$), associativity ($g_1(g_2 g_3) = (g_1 g_2) g_3$), the existence of an identity elt (e) and inverses ($\exists g^{-1} : g g^{-1} = g^{-1} g = e$). The no. of elts in G is its order, written $|G|$ if it is finite. We met various examples in A&G:

a) $(\mathbb{Q}, +), (\mathbb{R}, +), \mathbb{C}, +)$, additive groups

b) $GL_n(\mathbb{R})$, the general linear gp w/ real coefficients, or gr of invertible $n \times n$ matricies

c) permutation groups, e.g. S_n the symmetric gp on n objects

d) the cyclic gp of order n, $\{e, x, \ldots, x^{n-1}\}$ for some elt x

e) the dihedral gp of order $2n$, $D_2 n$, the symmetry gp of the regular n-gon, consisting of n rotations (including the identity) and n reflections

Definition

A subset H is a subgroup of G if it is a cp under restriction of the operation; the notations is $H \leq G$.

Rk

To check a subset is a subgp one demonstrates $h_1 h_2^- 1 \in H$ if $h_1, h_2 \in H$.

Given a subgp H of G we can def an equivalence relation on G by $g_1 \sim g_2 \Leftrightarrow g_1^{-1} g_2 \in H$; this partitions G into equiv classes of the form $gH = \{gh : h \in H\}$, the left cosets of H in G. Notice that these are all of the same size $|H|$, so by counting we have:

T (Lagrange) (1.1)

Let $H \leq G$, G a fin gp, then $|G| = |H||G:H|$ where $|G:H|$ = the index of H in G = the no. of left cosets of H in G; in particular $|H| \mid |G|$.

Rk

We could equally well have done this for right cosets, using the equiv rel $g_1 \sim g_2 \Leftrightarrow g_1 g_2^{-1} \in H$; thus the no. of right cosets = $|G:H| = \frac{|G|}{|H|}$ = no. of left cosets.

Def

The order $o(g)$ of the elt g in G is the least $n \geq 1$ st $g^n = e$ if such an n exists; otherwise g is of infinite order. Note that $g^m = e \Rightarrow o(g) \mid m$.

Lemma (1.2)

The order of an elt of G divides $|G|$, as $\{e, g, \ldots, g^{n-1}\}$ where $n = o(g)$ is a subgp of G, so by Lagrange (1.1) $n \mid G$.

1.2 Normal subgps, homomorphisms, quotient groups, isomorphisms

Example

$(\mathbb{Z}, +) \subset (\mathbb{R}, +)$; one coset is $\mathbb{Z}$, another is $\mathbb{Z} + \frac{1}{2}$ [i.e. $\{\ldots, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots\}$]. We would like to define the addition of cosets here; we should have e.g. $(\mathbb{Z}+\frac{1}{2})+(\mathbb{Z}+\frac{1}{3}) = (\mathbb{Z}+\frac{5}{6}$. More generally, for a subgp K of G it wousd be good to be able to define multiplication of cosets $(g_1K)\cdot(g_2K) = g_1g_2K$. But for this to work we must have $\{g_1k_1g_2k_2 : k_1 \in K_1, k_2 \in K_2\}$ be the single coset g_1g_2K; thus $g_1Kg_2K = g_1g_2K$ and multiplying on the left by g_1^{-1} we get $Kg_2K = g_2K$ i.e. we need $Kg_2 \subset g_2K$; similarly starting with right cosets we need $Kg_2 \supset g_2K$ so we're sead to consider normal subgroups.

Definition

$K \leq G$ is <u>normal in G</u> if $gK = Kg \forall g \in G$ (i.e. left cosets = right cosets; of course it is equivalent that $gKg^{-1} = K$); the notation is $K \triangleleft G$.

Examples

All subgps of abelian gps are normal

$\{e\}$ and G are normal subgps of G

The rotation subgp $\triangleleft D_{2n}$, but $\{e, \text{reflection}\}$ is not normal is D_{2n}, since $g\text{reflection}g^{-1}$ is in general a different reflection.

Proposition (1.3)

Let $K \triangleleft G$, then we can define $(g_1K)\cdot(g_2K) = g_1g_2K$ and under this operation the set of cosets form a group $\frac{G}{K}$, the <u>quotient group</u> of K in G. The proof is as last year; the operation is well defined by the above, if $g_1K = g_1'K, g_2K = g_2'K$

then $g_1g_2K = g_1'g_2'K$. Closure is trivial, associativity follows from that for G, the identity elt is K and the inverse of any element gK is $g^{-1}K$.

Definition

A map $\theta : G \to H$ where G, H are groups is a group homomorphism if $\theta(g_1g_2) = \theta(g_1)\theta(g_2)$.

Lemma (1.4)

Let $\theta : G \to H$ be a (group) homomorphism, then a) the kernel $\ker\theta = \{g \in G : \theta(g) = e_H\} \lhd G$ and b) the image $\mathrm{Im}(\theta) = \{h : h = \theta(g) \text{ some } g \in G\} \leq H$: for a), let $K = \ker\theta$, then the left coset $gK = \{g_1 \in G : \theta(g) = \theta(g_1)\}$ (since $\theta(g_1) = \theta(g) \Leftrightarrow \theta(g^{-1}g_1) = e$) but similarly the right coset $Kg = \{g_1 \in G : \theta(g_1) = \theta(g)\}$ and thus $gK = Kg$; this is true $\forall g \in G$ so $K \lhd G$. For b), we need to show $\theta(g)\theta(g_1)^{-1}$ also belongs to the image, but this is just $\theta(gg_1^{-1})$ since θ is a homomorphism.

Def

A homomorphism $\theta : G \to H$ is an isomorphism if it is bijective; the notation is $G \cong H$; note that 'isomorphic to' is an equivalence relation.

1st isomorphism Thm (1.5)

Let $\theta : G \to H$ be a homomorphism of groups, $K = \ker\theta$. Then $\frac{G}{K} \cong \mathrm{Im}(\theta) \leq H$: we define $\Phi : \frac{G}{K} \to \mathrm{Im}(\theta)$ by $gK \mapsto \theta(g)$, and this is an isomorphism: it is well defined since if $gK = g_1K$ then $\theta(g) = \theta(g_1)$, as per the previous lemma, a homomorphism since $\Phi(g_1K, g_2K) = \Phi(g_1g_2K) = \theta(g_1g_2) = \theta(g_1)\theta(g_2) = \Phi(g_1K)\Phi(g_2K)$, injective since if $\Phi(g_1K) = \Phi(g_2K)$ then $\theta(g_1) = \theta(g_2)$ so $g_1K = g_2K$, and trivially surjective.

Example

$(\mathbb{Z}, +) \subset (\mathbb{R}, +)$; define a group hom $\theta : (\mathbb{R}, +) \to (\mathbb{C}^\star, \times)$, the nonzero cplx nos under multiplication, by $r \mapsto e^{2\pi ir}$; the kernel is $(\mathbb{Z}, +)$, image is the unit circle, so by (1.5) $\frac{\mathbb{R}}{\mathbb{Z}} \cong T$, the unit circle under multiplication $\leq (\mathbb{C}, \times)$.

The numbering of the 2nd and 3rd isomorphisms Thms varies between different books.

2nd isomorphism Thm (1.6)

Let $K \lhd G, H \leq G$, then $HK \leq G$ where $HK = \{hk : h \in H, k \in K\}$, $H \cap K \lhd H$, and $\frac{H}{H\cap K} \cong \frac{HK}{K}$.

HK is a subgp of G since $(h_1k_1)(h_2k_2)^{-1} = h_1h_1k_2^{-1}h_2^{-1} \in h_1Kh_2^{-1}$, which since $K \lhd G$ this is $h_1h_2^{-1}K \subset HK$.

Define $\theta : H \to \frac{G}{K}$ by $h \mapsto hK$; this is a homomorphism (clearly well defined) since $\theta(h_1h_2) = h_1h_2K = h_1K \cdot h_2K = \theta(h_1)\theta(h_2)$, so apply 1st isomorphism T; $\ker\theta = \{h \in H : hK = K\} = H \cap K, \mathrm{Im}\,\theta = \{hK : h \in H\} = \{gK : g \in HK\} = \frac{HK}{K} \leq \frac{G}{K}$; thus $\frac{H}{H\cap K} \cong \frac{HK}{K}$. In fact there is a very strong relationship between subgps of G containing K and subgps of $\frac{G}{K}$; we can biject

by $L \mapsto \frac{L}{K} = \{lK : l \in L\}$; the inverse of this is $X \mapsto \{g \in G : gK \in X\}$. Moreover, the set of normal subgps of G containing K is isomorphic to the set of normal subgps of $\frac{G}{K}$ by the same isomorphism, since $gK\frac{L}{K}(gK)^{-1} = \frac{gLg^{-1}}{K}$

3rd isomorphism Thm (1.7

Let $G \triangleleft K$ and $K \leq L \triangleleft G$, then $\frac{\frac{G}{K}}{\frac{L}{K}} \cong \frac{G}{L}$; define $\theta : \frac{G}{K} \to \frac{G}{L}$ by $gK \mapsto gL$; this is well defined since if $gK = g_1K$ then $g_1^{-1}g \in K \leq L$ so $gL = g_1L$, and a hom since $\theta(g_1Kg_2K) = \theta(g_1g_2K) = g_1g_2L = g_1Lg_2L = \theta(g_1K)\theta(g_2K)$, then we apply (1.5); $\ker\theta = \{gK : gL = L\} = \frac{L}{K}, \mathrm{Im}(\theta) = \frac{G}{L}$ trivially, so $\frac{\frac{G}{K}}{\frac{L}{K}} \cong \frac{G}{L}$.

Definition

G is simple if the only normal subgps in G are $\{e\}$ and G.

Examples

The only abelian finite simple groups are cyclic of prime order ($\cong C_p$), since in an abelian group all subgps are normal, and if $g \neq e \in G$ we know $\{e, g, \ldots, g^{o(g)-1}\} \leq G$, so this must be G, and if $o(g) = rs$ then $\{e, g^r, g^{2r}, \ldots, g^{r(s-1)}\} \leq G$, so $o(g)$ must be some prime p.

Shortly we shall show that A_5 is simple.

This section is non-examinable; in fact A_n is simple $\forall n \geq 5$. Note that $|A_3| = 3$ so A_3 is abelian, the cyclic group of order 3. Thus the exceptional case is not that A_5 is simple, but that $V = \{e, (12)(34), (13)(24), (14)(23)\} \triangleleft A_4$ so this is not simple. $|A_5| = 60, |A_6| = 360$; there is a finite simple group "in between" these of order 168, which is most easily approached as $GL(\mathbb{Z}_2, 3)$; as an excercise the reader should show this has order 168 and, if very keen, that it is simple. There are several infinite families apart from the alternating groups, and 26 'sporradic' non-abelian finite simple groups which don't fit into these families.

Thm (1.8)

This is quite an important Thm; let G be a finite group, then there are subgroups $\{e\} = H_s \triangleleft \cdots \triangleleft H_2 \triangleleft H_1 \triangleleft H_0 = G$ where $H_{i+1} \triangleleft H_i \forall i, \frac{H_i}{H_{i+1}}$ simple (note not generally $H_i \triangleleft G$).

Pick H_1 from the normal subgroups of G so that $H_1 \lneq G$ but $|H_1|$ maximal, then our bijection between normal subgroups of G containing H_1 and normal subgroups of $\frac{G}{H}$ tells us that $\frac{G}{H_1}$ is simple; repeat to find $H_2, H_3, \ldots$; since G is finite we eventually reach $H_s = \{e\}$. The H_i are not essentially unique, but the simple quotients $\frac{H_i}{H_{i+1}}$, the composition factors of G, are.

Definition

G is soluble if all its composition factors are cyclic of prime rder; the terminology comes from Galois theory. Solving a quadratic equation we have a formula for the roots, similarly for cubics and quartics (solution by radicals), but quintic equations cannot in general be solved this way. The method of finding these solutions is to associate a 'Galois Group' with the polynomial; a polynomial

equation is soluble by radicals only if the Galois group is soluble. However, $t^5 - 6t + 3$ has Galois group S_5 which is not soluble; it has compositiaon factors $\cong A_5$ and C_2 by $1 \triangleleft A_5 \triangleleft S_5$.

1.3 Permutation groups, actions and permutation representations

This section continues from the A&G course.

S_n, the symmetric group of bijective maps $\{1, \ldots, n\} \to \{1, \ldots, n\}$, called the permutations on $\{1, \ldots, n\}$, has $|S_n| = n!$. The usual notation for its elements is disjoint cycle form e.g. $(123)(45)(6)$, but we usually suppress the 1-cycles.

A_n is the alternating group of even permutations on $\{1, \ldots, n\}$; an even permutation is one which may be expressed as a product of an even no. of transpositions e.g. $(12)(34), (123) = (12)(23)$; equivalently a permutations is even iff there are an even no. of cycles of even length in its expression in disjoint cycle form. $|A_n| = \frac{n!}{2}, |S_n : A_n| = 2$. There are two cosets of A_n in S_n, the even permutations and the odd permutations, so left cosets = right cosets and $A_n \triangleleft S_n$.

More generally, for any set Ω we can define its symmetry group $\operatorname{Sym}\Omega$ of all bijective maps $\Omega \to \Omega$.

Definition

G is a permutation group of degree n if $G \leq \operatorname{Sym}\Omega$ where $|\Omega| = n$, e.g. S_n, A_n, D_{2n} considered as permuting the verticies of a regular n-gon.

Definition

Let G be a group, Ω a set. The map $\star : G \times \Omega \to \Omega$ $(g, \alpha) \to g \star \alpha$ (usually written $g(\alpha)$ is an action of G on Ω if:

i) $g \star \alpha \in \Omega \forall g \in G, \alpha \in \Omega$

ii) $g_1 \star (g_2 \star \alpha) = g_1 g_2 \star \alpha \forall g_1, g_2 \in G, \alpha \in \omega$

iii) $e \star \alpha = \alpha \forall \alpha \in \Omega$

Definition

The orbit of G on Ω containing α is $G(\alpha) = \{g \times \alpha : g \in G\}$

Definition

G is transitive on Ω if Ω is the only orbit, i.e. there is exactly one orbit.

Definition

If $\alpha \in \Omega$ the stabiliser of α is $G_\alpha = \{g \in G : g \star \alpha = \alpha\}$.

Thm (orbit-stabiliser) (1.9)

Let G act on Ω and $\alpha \in \Omega$, then $|G(\alpha)| = |G : G_\alpha|$; the size of the orbit of α is the index of the stabiliser, as there is a bijection from the set of left cosets of G_α in G to $G(\alpha)$ by $gG_\alpha \to g \star \alpha$.

Lemma (1.10)

Let G act on Ω. For $g \in G$ the map $\phi_g : \Omega \to \Omega$ defined by $\alpha \mapsto g \star \alpha$ is a permutation on Ω (it has inverse $\phi_{g^{-1}}$ so is bijective so a permutation), and the map $\phi : G \to \operatorname{Sym}\Omega$ defined by $g \mapsto \phi_g$ is a homomorphism by the definition of an action, a <u>permutation representation</u> of G. By the first isomorphism theorem (1.5) the image $\phi(G)$ is a subgroup of $\operatorname{Sym}\Omega$, defined to be G^Ω, the kernel of ϕ, $G_{(\Omega)} := \{g \in G : g \star \alpha = \alpha \forall \alpha \in \Omega\}$, is a normal subgroup of G, and $\frac{G}{G_{(\Omega)}} \cong G^\Omega$. If $G_{(\Omega)} = \{e\}$, the action is <u>faithful</u>.

Rk

Cf Exs1Q11; if G acts on Ω giving orbits $\Omega_1, \ldots, \Omega_r$ then the image of ϕ in $\operatorname{Sym}\Omega$ is isomorphic to a subgroup of the direct product $\operatorname{Sym}\Omega_1 \times \cdots \times \operatorname{Sym}\Omega_r$.

1.3.1 Examples

1)a) $\Omega =$ the set of diagonals of a cube, $G =$ the symmetry group of the cube. $|\Omega| = 4, |G| = 48$. G acts on Ω; $G^\Omega \cong S_4, G_{(\Omega)} = \{e, \text{central inversion}\}$

1)b) The dodecahedron has 12 pentagonal faces, 30 edges, and 20 verticies, with 3 pentagons about each vertex; the icosahedron has 20 triangular faces, 30 edges and 12 verticies, with 5 triangles about each vertex. These are dual to each other; if we put a vertex at the centre of each face of one we obtain the other (the same is true of the cube and octahedron). Let G be the rotation group (the same for both shapes); we use (1.9) to find $|G|$, e.g. $\Omega =$ the set of faces of a dodecahedron; $|\Omega| = 12, |G_\alpha| = 5, G$ acts transitively so $|G| = 5 \times 12 = 60$. Alternatively one can inscribe 5 cubes inside a dodecahedron (each using 8 of the 20 verticies); $\Omega =$ the set of inscribed cubes, $|\Omega| = 5, G^\Omega \cong A_5, G_{(\Omega)} = \{e\}, G$ acts faithfully on Ω (the missing part of this proof, an exercise for the reader, is to show that a subgroup of S_5 of order 60 must be A_5)

2) <u>Left regular action</u> of G on $\Omega = G$: $g \star \alpha = g\alpha \forall g, \alpha \in G$. $\ker\phi = \{e\}$, so since ϕ is an isomorphism we have by (1.10):

Theorem (Cayley) (1.11)

Any group G is isomorphic to a subgroup of $\operatorname{Sym} G$

3) Action of G on $\Omega =$ the set of left cosets of some $H \leq G$: $g \star g_1H = gg_1H$. The kernel of ϕ is $\bigcup_{g_1 \in G} g_1 H g_1^{-1}$, since we must have $gg_1H = g_1H \forall g_1 \in G$ so $g_1^{-1}gg_1 \in H$ i.e. $g \in g_1 H g_1^{-1} \forall g \in G$. This kernel is the largest normal subgp of G contained in H, since it is clearly normal and if $K \leq H, K \lhd G$ then $K = g_1 K g_1^{-1} \forall g_1 \in G$; $K \leq g_1 H g_1^{-1}$ so $K \leq \bigcap_{g_1 \in G} g_1 H g_1^{-1}$

Thm (1.12)

Let G a fin gp, H a proper subgp of G, $|G:H| = n$, then $\exists K \triangleleft G$ with $K \subset H$ s.t. $\frac{G}{K} \cong$ a subgp of S_n: let K be the kernel of the permutation repr arising from the action of G on $\Omega =$ the let of left cosets of H in G. By (1.10) $\frac{G}{K} \cong G^\Omega$, a subgp of $\operatorname{Sym}\Omega \cong S_n$. $G:K$ divides $n!$ by Lagrange (1.1) and is $\geq n$ since the action is transitive so $|G^\Omega| \geq n$ (in fact $n \mid |G^\Omega|$).

If G is non-abelian and simple then $G \cong$ a subgp of A_n: $K = \{e\}$ by simplicity, so $G \cong G^\Omega \leq S_n$, but $A_n \triangleleft S_n$ so $G^\Omega \cap A_n \triangleleft G^\Omega$; $G_n \cap A_n$ must therefore be either $\{e\}$ or G^Ω; if the latter then $G^\Omega \leq A_n$ as required, if the former G^Ω must contain the same number of even and odd permutations so is of order 2, contradicting G is non-abelian.

1.4 Conjugacy classes, centralisers and normalisers

Def the conjugacy action of G on $\Omega = G$ by $g \star x = gxg^{-1} \forall x, g \in G$. In this case the permutations $\phi_g : x \mapsto gxg^{-1}$ are bijective but also group homomorphisms, thus they are automorphisms of G. We def $\operatorname{Aut} G$ to be the set of isomorphisms $G \to G$; this is a gp under composition of maps.

Orbits are called conjugacy classes; $\operatorname{ccl}_G(x) = \{gxg^{-1} : g \in G\}$. Stabilisers are called centralisers, $C_G(x) = \{g \in G : gxg^{-1} = x\}$ (note this is the same as the set of elts of G which commute with x). By orb-stab (1.9) $|G : C_G(x)| = |\operatorname{ccl}_G(x)|$; the LHS is $\frac{|G|}{|C_G(x)|}$, but $|G| = \sum |\operatorname{ccl}_G(x)|$ where the sum is over distinct conjugacy classes, so we have:

Lemma (1.13)

If G is finite, $1 = \sum \frac{1}{|C_G(X)|}$ (where the sum is over discrete conjugacy classes), as $1 = \frac{1}{|G|} \sum |\operatorname{ccl}_G(x)| = \frac{1}{|G|} \sum |G : C_G(x)| = \frac{1}{|G|} \sum \frac{|G|}{|C_G(x)|} = \sum \frac{1}{|C_G(x)|}$.

The kernel of this permutation representation $G \to \operatorname{Sym}\Omega$ $G_{(\Omega)} = \{g \in G : gxg^{-1} = x \forall x \in G\} = \{g \in G | gx = xg \forall x \in G\} = \bigcap_{x \in G} C_G(x)$, is called the centre of G.

There is another conjugacy action of G on $\Omega =$ the set of subgps of G, $g \star H = gHg^{-1}$; orbit of H is the conjugacy class of H, $\{gHg^{-1} : g \in G\} = \operatorname{ccl}_G(H)$, stabiliser is the normaliser of H, $\{g \in G : gHg^{-1} = H\} = N_G(H)$; by orb-stab (1.9) $|G : N_G(H) = |\operatorname{ccl}_G(H)$.

Example: Conjugacy classes in S_n and A_n

Recall from A&G that two elts of S_n are conjugate iff they have the same cycle type when written in disjoint cycle form. If we write $2^2 1$ for the cycle type of an element consisting of two 2-cycles and one 1-cycle and similarly, then S_5 contains 1 elt of type 1^5, 15 of $2^2 1$, 20 of $3 1^2$, and 24 of 5 for the even permutations and 10 of type $2 1^3$, 20 of type 32 and 30 of type 41 for the odd permutations.

Rk

We can now see that any subgp of S_5 of order 60 must be A_5; such a H is of index 2 so normal in S_n so a union of conjugacy classes, and were it not A_5 it

would have to contain exactly 30 even permutations (and 30 odd ones) which is impossible.

Let $g \in A_5$, then $\mathrm{ccl}_{A_n}(g) \leq \mathrm{ccl}_{S_n}(g)$. $|\mathrm{ccl}_{S_n}(g)| = |S_n : C_{S_n}(g)|$ by orb-stab (1.9), $|Z\,\mathrm{ccl}_{A_n}(g)| = |A_n : C_{A_n}(g)|$. But $C_{A_n}(g) = A_n \cap C_{S_n}(g)$, which must be of index 1 or 2 in $C_{S_n}(g)$ (either $C_{S_n}(g)$ lies entirely in A_n or contains an equal no. of odd and even permutations). If $C_{A_n}(g) = C_{S_n}(g)$ then $|\mathrm{ccl}_{S_n}(g)| = 2|\mathrm{ccl}_{A_n}(g)|$ since $|S_n : C_{S_n}(g)| = |S_n : C_{A_n}(g)| = |S_n : A_n||A_n : C_{A_n}(g)| = 2|A_n : C_{A_n}(g)| = 2|\mathrm{ccl}_{A_n}(g)|$; if $C_{A_n}(g)$ is of index 2 in $C_{S_n}(g)$ then $|\mathrm{ccl}_{S_n}(g)| = |\mathrm{ccl}_{A_n}(g)|$. Thus some S_n-conjugacy classes (of even permutations) split into two conjugacy classes in A_n, according to whether $C_{A_n}(g) = C_{S_n}(g)$ or not. In fact (excercise) this happens iff all cycles in disjoint cycle form have distinct length, e.g. for $n = 5$ $(12)(34)$ commuties with (12), as does (345), so the conjugacy classes of double transpositions and 3-cycles both do not split, but the conjugacy class of 5-cycles does split.

Reccall $g((12345))g^{-1} = (g(1)\dots g(5))$. $C_{S_n}((12345))$ is just the subgp generated by (12345), i.e. $\{e, (12345), (13524), (14253), (15432)\}$, and this $= C_{A_5}((12345))$. So in A_5 the conjugacy classes are 1 elt of type 1^5, 15 of type 2^21, 20 of type 31^2, and two classes of size 12 each with elts of cycle type 5.

Rk

One can see this geometrically via the rotation gp of the dodecahedron (or icosahedron): the conjugates of a rotation are rotations of the same angle but about a different axis, e.g. 5-cycles correspond to rotations of the dodecahedron about an axis through midpoints of faces. The conjugacy classes are rotations by $\pm\frac{2\pi}{5}$ and rotations by $\pm\frac{4\pi}{5}$ (a rotation of $-\frac{2\pi}{5}$ is one of $\frac{2\pi}{5}$ about the "opposite axis" [the same axis in the opposite direction]).

Prop (1.14)

A_5 is simple: let $K \lhd G = A_5$. A subgp is normal iff it is a union of conjugacy classes, so $|K| = a + 15b + 20c + 12d$ where $a = 1$ $(e \in K)$, b, c are each either 0 or 1, and $0 \leq d \leq 2$. Lagrange (1.1) implies $|K| \mid 60$, but the only ways this can happen are $a = 1, b = c = d = 0$ i.e. $K = \{e\}$ or $a = b = c = 1, d = 2$ i.e. $K = G$, so we have the result.

1.5 Finite p-groups

Definition

A gp G is a p-group if $|G| = p^n$ some prime p.

Thm (1.15

Let G be a p-group, then its centre $Z(G) \neq \{e\}$: the size of $\mathrm{ccl}_G(x)$ is the index of $C_G(x)$ so divides $|G|$ (by orbit-stabiliser (1.9)), so the possible sizes of conjugacy classes are $1, p, p^2, \dots$. $p^n = |G| = \sum_{i=0,1,\dots}$ [no. of classes of size p^i] $\times\, p^i$. But $p \mid |G|$, so we must have $p \mid$ the no. of conj classes of size 1. But $\{x\}$ is a conj class iff $gxg^{-1} = x \forall g \in G \Leftrightarrow gx = xg \forall g \in G \Leftrightarrow x \in Z(G)$. So $|Z(G)|$ = the no. of conj classes of size $1 \geq p$.

Lemma (1.16)

For any gp, if $\frac{G}{Z(G)}$ cyclic then G is abelian (i.e. $G = Z(G)$): suppose $\frac{G}{Z(G)}$ is cyclic, generated by some gZ. Each elt of $\frac{G}{Z(G)}$ is of the form $(gZ)^r = g^r Z$ for some r, so any elt x_1 of G is of the form $g^r z$ for some r and $z \in Z(G)$. But then any two elts $x_1, x_2 \in G$ commute, since they are respectively $g^{r_1} z_1, g^{r_2} z_2$ and $x_1 x_2 = g^{r_1} z_1 g^{r_2} z_2 = g^{r_1} g^{r_2} z_1 z_2 = g^{r_2} g^{r_1} z_2 z_1 = g^{r_2} z_2 g^{r_1} z_1 = x_2 x_1$, since both z_1, z_2 commute with all other elts of G.

Prop (1.17)

Gps of order p^2 are abelian: (1.15) implies $|Z(G)| \geq p$ so $\frac{G}{Z(G)}$ is of order 1 or p, but if it is of order p then $\frac{G}{Z(G)}$ is cyclic so by (1.16) G is abelian (i.e. $G = Z(G)$), a contradiction, thus $\frac{G}{Z(G)}$ is of order 1, $G = Z(G)$ and is therefore abelian.

Direct Products

Given gps G, H we can construct a gp $G \times H$ by $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$. Inside $G \times H$ we have subgps $G_1 = \{(g, e_H) : g \in G\} \cong G, H_1 = \{(e_G, h) : h \in H\} \cong H$ w/ $G_1 \cap H_1 = \{(e_G, e_H)\}$, the identity elt of $G \times H$, and $G_1 H_1 = G \times H$.

$G \times H$ is usually called the external direct product; there is also an internal direct product L; this is when in L there are subgroups $G_1 \cong G, H_1 \cong H, G_1 \cap H_1 = \{e\}, G_1 H_1 = L$ and all elts of G_1 commute w/ all els of H_1 (and thus $L \cong G_1 \times H_1$, cf Exs1Q5).

Rk

Gps of order p^2 are $\cong C_{p^2}$ or $\cong C_p \times C_p$; for gps of order p^3 see Exs1Q6.

Thm (1.18)

A p-group G of order p^n contains a subgroup of order p^m for any $1 \leq m \leq n$ (this is something of a converse to Lagrange): we induct on n, the $n = 1$ base case being trivial. Assume $n > 1$, then by (1.15) the centre $Z(G) \neq \{e\}$. Pick $x \neq e \in Z(G)$; by taking a suitable power of x we may assume it is of order p. Thus $K = \{e, x, \ldots, x^{p-1}\}$ is a subgroup of G and normal since $x \in Z(G)$, so $|\frac{G}{K}| = p^{n-1}$. Apply the inductive hypothesis to find a subgroup $\frac{H}{K} \leq \frac{G}{K}$ of order p^{m-1}, then the correspondence between subgroups of $\frac{G}{K}$ and subgroups of G containing K gives us a subgroup H of G containing K with $|H| = p^m$ as required.

1.6 Finite abelian groups

Thm (1.19)

Let G be a finite abelian group, then G is a direct product $C_{d_1} \times C_{d_2} \times \cdots \times C_{d_k}$ w/ $d_{i+1} \mid d_i$ and $|G| = d_1 d_2 \ldots d_k$.

This theorem is proovable now, but doing so is messy; we shall proove it in section 3 with more sophisticated methods than we currently have available. For the devoted reader who wishes to try now, an outline is to pick x of maximal order; if G generates G then we are done, otherwise (and this is the crux of

the proof) we can show $\exists H \leq G$ w/ $\langle x \rangle$ (the (cyclic) group generated by x) $\cap H = \{e\}$ and $G = \langle x \rangle H$, then induct.

For example the abelian groups of order 8 are $C_8, C_4 \times C_2, C_2 \times C_2 \times C_2$ and those of order 16 are $C_{16}, C_8 \times C_2, C_4 \times C_4, C_4 \times C_2 \times C_2, C_2 \times C_2 \times C_2 \times C_2$.

Lemma (1.20)

If $(m, n) = 1$ then $C_{mn} \cong C_m \times C_n$: take generators g, h of C_m, C_n respectively, and consider the order of (g, h); $(g, h)^r = (g^r, h^r$, so the least r for which this is $e = (e_G, e_H)$ will be $\text{lcm}(m, n)$ which $= mn$ by hypothesis. But $|C_m \times C_n| = mn$ so this means (g, h) is a generator of $C_m \times C_n$, so the group is cyclic of order mn, e.g. $C_6 = C_2 \times C_3$, so the abelian groups of order 24 are $\cong C_{24}, C_{12} \times C_2$ or $C_6 \times C_2 \times C_2$.

1.7 Sylow's theorems with applications to small groups

Theorem (1.21) (Sylow)

Let G be a group with $|G| = p^a m, p$ prime, $p \nmid m$. Then:

1) There is a subgroup of order p^a of G, called a Sylow p-subgroup of G.

2) All Sylow p-subgroups of G are conjugate

3) Let n_p be the number of Sylow p-subgroups, then $n_p \cong 1(p)$ and $n_p \mid |G|$ (note that these two conditions together imply $n_p \mid m$.

Remark

G acts on Ω = the set of Sylow p-subgroups of G by conjugation $g \star P = gPg^{-1}$; Sylow's second theorem is the statement that this action is transitive (has precisely one orbit). The stabiliser of P is the normaliser $N_G(p) = \{g \in G : gPg^{-1} = P\}$. By Orbit-Stabiliser (1.9) n_p = the size of the orbit of P = the index of $N_G(P)$ which of course $\mid |G|$, so the last part of the third theorem is easy given the second theorem.

Lemma (1.22)

If $n_p = 1$ the unique Sylow p-subgroup is normal in G, as gPg^{-1} is a Sylow p-subgroup for any $g \in G$, so must $= P \forall g \in G$.

Remark

$P \lhd N_G(P)$. Applying Sylow's second theorem to $N_G(P)$ we have P is the unique Sylow p-subgroup of $N_G(P)$.

Lemma (1.23)

Cf (1.12): Let G be non-abelian and simple, and suppose $n_p > 1$. Then $|G| \mid \frac{n_p!}{2}$ and $n_p \geq 5$.

Proof

G is acting transitively on Ω = the set of Sylow p-subgroups, and $|\Omega| = n_p$. The kernel of the associated permutation representation is a normal subgroup K and $\frac{G}{K} \cong$ some subgroup of $\operatorname{Sym}\Omega$. G is simple (and $G = K$ would contradict the transitivity of the action given $n_p > 1$) so $K = \{e\}$, so $G \cong$ some subgroup of S_{n_p}; G is simple so $\cong$ a subgroup of A_{n_p} (cf (1.12)) [so $|G| \mid |A_{n_p}| = \frac{n!}{2}$] and if we accept that subgroups of A_4 cannot be non-abelian and simple we have $n_p \geq 5$.

Examples

$|G| = 1000 = 2^3 5^3 \Rightarrow G$ **not simple**

$n_5 \cong 1(5), n_5 \mid 8 \Rightarrow n_5 = 1$ so there exists a unique normal Sylow 5-subgroup (of order 125).

$|G| = 300 = 2^2 35 \Rightarrow G$ **not simple**

$n_5 \cong 1(5)$ (it is generally best to start with the largest possible p) and $n_5 \mid 12$; Assume G is simple so $n_5 \neq 1$, then (G must be non-abelian if simple since $|G|$ is non-prime) $n_5 = 6$ and by 1.23 $|G| \mid \frac{6!}{2} = 360$, a contradiction.

$|G| = 132 = 2^2 \times 3 \times 11 \Rightarrow G$ **not simple**

$n_{11} \equiv 1(11), n_{11} \mid 12$. Assume G is simple (and again it must be non-abelian), $n_{11} \neq 1$ so $n_1 1 = 12$. $n_3 = 1(3), n_3 \mid 44$, $n_3 \neq 1$ by simplicity and $\neq 4$ by (1.23), so $n_3 = 22$, but then we have 12×10 elements of order 11 (since each Sylow 11-subgroup contains 10 distinct elements, all of order 11) and 22×2 elements of order 3, a contradiction because the sum of these is more than the size of the group. This is a typical argument.

On Exs1 we show that S_4, A_5 and groups with $|G| = pq, pq^2, pqr$ (for distinct primes p, q, r) are all non-simple.

Lemma (1.24)

Let G be a group of order $2p$ where p is an odd prime, then $G \cong C_{2p}$ or D_{2p}: by Sylow's theorems $n_p \equiv 1(p)$ and $n_p \mid 2$ so $n_p = 1$ and $\exists!$ subgroup of order p; by Sylow $n_2 \equiv 1(p), n_2 \mid p$. We certainly have a Sylow 2-subgroup Q, say $\{e, g\}$ (which may or may not be unique), and a Sylow p-subgroup $\{e, x, \ldots, x^{p-1}\}$. Conjugation gives action on P [?]: $gxg^{-1} \in P$ since P normal, so must $= x^r$ some r. $x = ggxg^{-1}g^{-1} = gx^r g^{-1} = (gxg^{-1})^r = x^{r^2}$ so $r^2 \equiv 1(p)$, thus $r \equiv \pm 1(p)$; if $r \equiv 1$ then $gxg^{-1} = x$ and x, g commute, gx is of order $2p$ so $G \cong C_{2p}$ (this is the case $n_2 = 1$ [if C_{2p} is generated by y, the only Sylow 2-subgroup is $\{e, y^p\}$]), if $r \equiv -1(p)$ then $gxg^{-1} = x^{-1}$ and we have an isomorphism with D_{2p} by $g \mapsto$ some reflection, $x \mapsto$ rotation by $\frac{2\pi}{p}$ (this is the case $n_2 = p$ [$\{e, a\}$ is a Sylow 2-subgroup for any reflection a]); cf Exs1 on semidirect products, $G \cong$ semidirect product of P by Q.

Proof of Sylow's Theorems

Let G be a finite group, $|G| = p^a m, p$ prime, $p \nmid m$.

1: There exists a subgroup of order p^a, a Sylow p-subgroup

Consider $\Omega =$ the set of subsets of G of size p^α; $|\Omega = \begin{pmatrix} p^a m \\ p^a \end{pmatrix} = \frac{p^a m}{p^a}\frac{p^a m - 1}{p^a - 1}\ldots\frac{p^a m - p^a + 1}{1}$.
For each factor $\frac{p^a m - k}{p^a - k}$ in this product other than the first, $\frac{p^a m - k}{p^a - k}, 1 \leq k \leq p^a - 1$, we make all possible cancellations of powers of p: let $k = p^b q, b < a, p \nmid q$, then $\frac{p^a m - k}{p^a - k} = \frac{p^a m - p^b q}{p^a - p^b q} = \frac{p^{a-b} m - q}{p^{a-b} - q}$, and$p \nmid (p^{a-b} m - q), p \nmid (p^{a-b} - q)$. $\frac{p^a m}{p^a} = m$ which is not divisible by p, and none of the numerators of the $\frac{p^{a-b} m - q}{p^{a-b} - q}$ are divisible by p, so $p \nmid |\Omega|$.

The orbits Σ of G acting on Ω by left multiplication $g \star \{g_1, \ldots, g_{p^a}\} = \{gg_1, \ldots, gg_{p^a}\}, g \in G, g_i \in G$, are all of size $\geq m$, as if $\{g_1, \ldots, g_{p^a}\} \in \Sigma$ then $gg_1^{-1} \star \{g_1, \ldots, g_{p^a}\} \in \Sigma$, but this is $\{g, \ldots, gg_1^{-1} g_{p^a}\}$, and thus each $g \in G$ appears as an entry of some p^a-set in Σ. This $|\Sigma| \geq \frac{p^a m}{p^a} = m$.

By Orbit-Stabiliser (1.9), $|\Sigma| \mid |G| = p^a m$. This means that if $|\Sigma| > m, p \mid |\sigma|$. Counting the elements of Ω, $|\Omega| =$ the sum of sizes of orbits. $p \nmid |\Omega|$ and $p \mid$ the sum of sizes of orbits with $|\Sigma| > m$, thus there must be at least one orbit Σ with $|\Sigma| = m$.

For this Σ, we know from above each $g \in G$ lies in at least one p^a-set $\in \Sigma$, but by counting as $|\Sigma| = m$ each must lie in exactly one such set. Thus $\{e, g_2', \ldots, g_{p^a}'\} \in \Sigma$ for some $g_i', 2 \leq i \leq p^a$.

What is the stabiliser P of this set (under the action of G on $\Sigma \subset \Omega$)? Observe $g_i' \star \{e, g_2', \ldots, g_{p^a}'\} = \{g_i', g_i' g_2', \ldots, g_i' g_{p^a}'\} \in \Sigma$ contains g_i', but g_i' appears as an entry in only one p^a-subset in the orbit, so $\{g_i', g_i' g_2', \ldots, g_i' g_{p^a}'\} = \{e, g_2', \ldots, g_{p^a}'\}$ and $g_i' \in P$ the stabiliser.

However, by the orbit-stabiliser theorem $|P| = \frac{|G|}{|\Sigma|} = \frac{p^a m}{p^a}$, so $P = \{e, g_2', \ldots, g_{p^a}'\}$, and this P is our Sylow p-subgroup.

2: All Sylow p-subgroups of G are conjugate

We shall actually proove a slightly stronger result:

If Q is a p-subgroup of G and P is a Sylow p-subgroup, then $Q \leq g_1 P g_1^{-1}$ some $g_1 \in G$ (and so if Q is a Sylow p-subgroup and so of order p^a then we deduce $Q = g_1 P g_1^{-1}$.

Consider the action of Q on the set Ω of left cosets of P, $g \star g_1 P = g g_1 P, g_1 \in G, g \in G$. Let $|Q| = p^b$; by orbit-stabiliser the sizes of orbits divide this, i.e. are various powers of p. Since the number of left cosets = index of P in $G = \frac{|G|}{|P|} = \frac{p^a m}{p^a} = m$ and $p \nmid m$, there must be orbits of size 1. Let $\{g_1 P\}$ be such an orbit, $g_1 \in G$. We have $g g_1 P = g_1 P \forall g \in Q$, i.e. $g_1^{-1} g g_1 P = P \therefore g_1^{-1} g g_1 \in P \therefore g \in g_1 P g_1^{-1} \forall g \in Q$, i.e. $Q \leq g_1 P g_1^{-1}$ as required.

3: No. of Sylow p-subgroups $n_p \equiv 1(p)$,$n_p \mid |G|$

Saw before $n_p \mid |G|$. We have G acting via conjugation on the set Ω of Sylow p-subgroups; by the second theorem this action is transitive. We restrict our attention to the action of the Sylow p-subgroup P on Ω via conjugation; again

the orbits of P on Ω have sizes various powers of P (since their sizes must divide $|P|$); note $\{P\}$ is an orbit of size 1. If $\{P_1\}$ is an orbit of size 1 then we have $P \leq N_G(P_1)$, but recall from above the unique Sylow p-subgroup of $N_G(P_1)$ is P_1, so $P = P_1$, and $\{P\}$ is the only orbit of size 1. So counting, n_p the number of Sylow p-subgroups= the sum of sizes of orbits = 1 + sum of sizes of orbits all divisible by $p \equiv 1(p)$.

Recall that a soluble group is one where all the composition factors are abelian simple.

Theorem (Burnside 1904) (Pembroke)

If $|G| = p^a q^b$ for distinct prime p, q then G is soluble. The usual proof of this result is given in the part II course "Representation Theory"; a purely group theoretic proof is very messy.

Theorem (Hall 1937) (Caius)

Let G be a finite group. If for any factorisation $|G| = mn$ wih $(m, n) = 1$, $\exists$ a subgroup of order n, then G is soluble, and conversely.

Theorem (Ferit-Thompson, 1963) (Churchill)

If $|G|$ odd then G is soluble (and hence non-abelian simple groups always have even order); this result is very hard to proove.

2 Rings

2.1 Definitions and examples

Definition

A set R with operations $+$ and $\cdot$ is a ring if:

a) $(R, +)$ is an abelian group; we call the additive identity 0_R or simply 0

b) Multiplication is associative, and there is a multiplicative identity 1_R

c) Multiplication is distributive over addition; $a \cdot (b + c) = a \cdot b + a \cdot c$ (and $(b + c) \cdot a = b \cdot a + c \cdot a$)

Be aware that some people do not insist on the existence of a multiplicative identity.

In this course we shall assume our rings are commutative, that is that multiplication is commutative, even though this is a rather silly assumption.

Definition

A subset S of a ring R is a subring of R if it is a ring under restriction of the operations; in particular we must have $1_R \in S$. The notation is $S \leq R$.

Examples

$\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$

The Gaussian integers, $\{a + bi : a, b \in \mathbb{Z}\} \subset \mathbb{C}$, as mentioned in the first lecture.

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

$\mathbb{Z}_n$; this is a ring for any n even though it is only a field for n prime.

Any field

Definition

An element $r \in R$ is a unit if it has a multiplicative inverse, e.g. 2 is a unit in $\mathbb{Q}$ but not in $\mathbb{Z}$. Note some people use "unit" to refer to the multiplicative identity.

Examples

The zero ring $\{0\}$; in this case 0 is also the multiplicative identity. In all other cases $0_R \neq 1_R$ since $(0 + 0)r = 0r \therefore 0r = 0 \forall r \in R$, while $1r = r \forall r \in R$.

Example

Let R be a ring. A polynomial over R is of the form $a_0 + a_1X + \cdots + a_nX^n = f(X)$ with the $a_i \in R \forall i$. The degree of f is the largest n such that $a_n \neq 0$; f is monic if this $a_n = 1$.

$R[X]$ is the set of all polynomials over R, with addition $(f+g)(X) = \sum(a_i + b_i)X^i$ where $f(X) = \sum a_iX^i, g(X) = \sum b_iX^i$, and multiplication $(f \cdot g)(X) = \sum_i(\sum_{j=0}^{i} a_jb_{i-j})X^i$. R is a subring of $R[X]$ by identifying R with the set of constant polynomials $f(X) = a_0$.

$R[[X]]$ is the set of formal power series over R, $p(X) = a_0 + a_1X + A_2X^2 + \ldots$ with $a_i \in R \forall i$, with addition and multiplication defined in exactly the same way. The difference between these and polynomials is that here we may have infinitely many nonzero terms (note, however, that the inner summation when multiplying power series is only over finitely many terms, and thus the product of two power series is a well defined power series).

Lawrent Polynomials, $\sum a_iX^i$ but for $i \in \mathbb{Z}$ i.e. negative powers are allowed. However we still insist on there being only finitely many nonzero coefficients.

Lawrent series, $\sum a_iX^i, i \in \mathbb{Z}$ but with only finitely many of the a_i for $i \leq 0$ nonzero; were we to allow infinitely many nonzero coefficients "in both directions" the product of two series would not always be properly defined.

Rings of R-valued functions on a set A, $f : A \to R$ with pointwise addition and multiplication ($(f+g)(a) = f(a)+g(a)$ and similarly). The set of continuous functions $\mathbb{R} \to \mathbb{R}$ form a subring of the ring of real-valued functions on $\mathbb{R}$, and in turn the set of polynomial functions $\mathbb{R} \to \mathbb{R}$ (i.e. functions of the form $a \mapsto f(a)$, "evaluation at $a \in \mathbb{R}$", where f is a polynomial) with real coefficients is a subring of this. Note that if we try and do the same with power (or Lawrent) series to give functions $\mathbb{C} \to \mathbb{C}$ we may have problems of convergence.

2.2 Isomorphisms, homomorphisms, ideals and quotient rings

Definition

Let R, S be rings. The map $\theta : R \to S$ is a (ring) homomorphism if $\theta(r_1 + r_2) = \theta(r_1) + \theta(r_2)$ (i.e. it is a homomorphism of additive groups) and $\theta(r_1 r_2) = \theta(r_1)\theta(r_2), \theta(1_R) = 1_S$.

Definition

An isomorphism is a bijective homomorphism

Definition

The kernel of a homomorphism is $\ker\theta = \{r \in R : \theta(r) = 0_S\}$.

Lemma (2.1)

$\theta : R \to S$ is injective $\Leftrightarrow \ker\theta = \{0_R\}$; if θ is injective 0_R is the only element of R mapped to 0_S, if $\ker\theta = \{0_R\}$ then $\theta(r_1) = \theta(r_2) \Rightarrow \theta(r_1) - \theta(r_2) = 0 \Rightarrow \theta(r_1 - r_2) = 0 \Rightarrow r_1 - r_2 \in \ker\theta \Rightarrow r_1 - r_2 = 0 \Rightarrow r_1 = r_2$.

Definition

A subset $I \subset R$ is an ideal of R if I is a subgroup of R under addition and I is closed under multiplication by all elements of R, that is $\forall a \in I, r \in R$, $ar \in I$. The notation is $I \triangleleft R$, but though there is a strong analogy with normal subgroups note that ideals are not generally subrings of R, as if $1_R \in$ an ideal I, any $r \in R$ is $= 1_R r$ so $\in I$ and thus $I = R$; I is only a subring if $I = R$.

Lemma (2.2)

The kernel of a homomorphism $\theta : R \to S$ is an ideal of R, since if we let $I = \ker\theta$ then $a_1, a_2 \in I \Rightarrow \theta(a_1 + a_2) = \theta(a_1) + \theta(a_2) = 0 + 0 = 0$ so $a_1 + a_2 \in I$, $a \in I \Rightarrow \theta(-a) = -\theta(a) = 0 \Rightarrow -a \in I$ so I is an additive subgroup of R, if $a \in I, r \in R$ then $\theta(ar) = \theta(a)\theta(r) = 0\theta(r) = 0$ so $ar \in I$.

Examples

In a field F, the only ideals are $\{0\}, F$, since if $a \neq 0 \in I$ then $aa^{-1} = 1_R \in I$ and then $I = F$ as above.

The ideals in $\mathbb{Z}$ are of the form $n\mathbb{Z} = \{\dots, -n, 0, n, \dots\}$ [for $n \in \mathbb{Z}$]; certainly any such $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$, but these are the only ideals since they are the only additive subgroups of $\mathbb{Z}$: Let I be a nonzero additive subgroup of $\mathbb{Z}$ and let n be the least positive element of I. Then $n\mathbb{Z} \subset I$ clearly (and note that this does not depend on I being a ring, because $qn = n + \dots + n$), and if $m \in I$ write $m = qn + r$ with $0 \leq r < n$ by Euclid's algorithm. Then $r = m - qn \in I$ since $n \in I$, so minimality of n implies $r = 0 \therefore m = qn \therefore m \in n\mathbb{Z} \therefore I = n\mathbb{Z}$.

Definition

Let R be a ring and $a \in R$. Then the ideal generated by a is $aR = \{ar : r \in R\}$ (sometimes called (a). This is the smallest ideal containing a; $a = a1 \in aR$, $ar_1 + ar_2 = a(r_1 + r_2)$, $-ar = a(-r)$ so aR is an additive subgroup of R, and for any $ar \in aR, s \in R$, $(ar)s = a(rs) \in aR$, so R is an ideal containing a. If $a \in I$ with I an ideal then $ar \in I \forall r \in R$ by the definition of an ideal so $aR \subset I$.

Such an ideal aR is a principal ideal, e.g. $n\mathbb{Z} \triangleleft \mathbb{Z}$. More generally, $(a_1, \ldots, a_n)$ denotes $a_1R + \cdots + a_kR = \{\sum a_i r_i : r_i \in R\}$, the ideal generated by $a_1, \ldots, a_k$. If $A \subset R$ then $(A) = \{\sum_{a \in A} ar_a : r_a \in R, r_a = 0$ for all but finitely many $a\}$, e.g. $(X) \triangleleft \mathbb{C}[X]$ is $\{f(X) : f(X) = Xg(X)$ some $g(X) \in \mathbb{C}[X]\} = \{f(X) \in \mathbb{C}[X] : f(0) = 0\}$, the set of polynomials with zero constant term.

Proposition (2.3)

Let R be a ring and $I \triangleleft R$. The quotient ring $\frac{R}{I}$ is the set $\{r + I : r \in R\}$ of cosets of I in R (under $+$) with addition $(r_1 + I) + (r_2 + I) = (r_1 + r_2) + I$, and multiplication $(r_1 + I)(r_2 + I) = r_1r_2 + I$; this forms a ring: addition is well defined and yields an additive group by (1.3) (quotient group), multiplication is well defined since if $r_1 + I = r_1' + I, r_2 + I = r_2' + I$ then we let $a_1 = r_1 - r_1', a_2 = r_2 - r_2'$, then $a_1, a_2 \in I$ so $r_1r_2 = (a_1 + r_1')(a_2 + r_2') = r_1'r_2' + (a_1r_2' + r_1'a_2 + a_1a_2)$ and the bracket is $\in I$ so $r_1r_2 - r_1'r_2' \in I$ and $r_1r_2I = r_1'r_2'I$. $1_R + I$ is a multiplicative identity as $(1 + I)(r + I) = r + I \forall r \in R$; associativity and distributivity are left as straightforward exercises for the reader.

Examples

$n\mathbb{Z}$ are the ideals of $\mathbb{Z}$. $\frac{\mathbb{Z}}{n\mathbb{Z}}$ has elements $0 + n\mathbb{Z}, 1 + n\mathbb{Z}, \ldots, (n-1) + n\mathbb{Z}$; addition and multiplication correspond to those of $\mathbb{Z}$ mod n.

$I = (X) \triangleleft \mathbb{C}[X]$; $\frac{\mathbb{C}[X]}{(X)}$ has for any $f(X) \in \mathbb{C}[X], f(X) = a + Xg(X)$ some $g(X) \in \mathbb{C}[X]$ and so $f(X) + I = a + I$. Addition is $(a + I) + (b + I) = (a + b) + I$, similarly multiplication $(a + I)(b + I) = ab + I$ which correspond to addition and multiplication in $\mathbb{C}$; $\frac{\mathbb{C}[X]}{(X)} \cong \mathbb{C}$.

Proposition (2.4) (Euclid's algorithm for $F[X]$)

Let F be a field, $f(X), g(X) \in F[X], g(X) \neq 0$. Then $f(X) = g(X)q(X) + r(X)$ with $\deg r(X) < \deg g(X)$ (or $r(X) = 0$): let $f(X) = \sum_{i=0}^{n} a_i X^i$, of degree n, and $g(X) = \sum_{i=0}^{m} b_i X^i, b_m \neq 0$ of degree m. If $n < m$ we are done $(q(X) = 0, r(X) = f(X))$, otherwise we induct on n: for $m \geq n$ set $f_1(X) = f(X) - a_n b_m^{-1} g(X) X^{n-m}$ (b_m^{-1} exists since $b_m \neq 0$), then $\deg f_1(X) < \deg f(X) = n$; for $m = n$ set $q(X) = a_n b_m^{-1}, r(X) = f_1(X)$, for $n > m$ write $f_1(X) = g(X)q_1(X) + r_1(X)$ by the inductive hypothesis, then $f(X) = g(X)(q_1(X) + a_n b_m^{-1} X^{n-m}) + r_1(X)$ so we are done by $q(X) = q_1(X) + a_n b_m^{-1} X^{n-m}, r(X) = r_1(X)$.

Note that we did need F to be a field, because b_m^{-1} would not necessarily exist in a ring.

Example

$I = (X^2 + 1) \triangleleft \mathbb{R}[X]$; for any $f(X) \in \mathbb{R}[X], f(X) = (X^2 + 1)q(X) + r(X)$ by (2.4), with $r(X)$ of degree ≤ 1, so $f(X) + I = r(X) + I$; the elements of $\frac{\mathbb{R}[X]}{I}$

are therefore of the form $a_0 + a_1X + I$. We have that addition is given by $(a_0 + a_1X + I) + (b_0 + b_1X + I) = (a_0 + b_0) + (a_1 + b_1)X + I$ and multiplication by $(a_0 + a_1X + I)(b_0 + b_1X + I) = a_0b_0 + (a_0b_1 + a_1b_0)X + a_1b_1X^2 + I = (a_0b_0 - a_1b_1) + (a_0b_1 + a_1b_0)X + I$ (since $a_1b_1X^2 = -a_1b_1 + (a_1b_1(X^2+1))$ with the bracket in I). We see that this behaves the same as $\mathbb{C}$, and have $\frac{\mathbb{R}[X]}{I} \cong \mathbb{C}$ by $a_0 + a_1X + I \mapsto a_0 + a_1i$.

Example

$\frac{\mathbb{F}_2[X]}{X^2+X+1}$ has four elements; we find this is a field, and in fact the same is true for some other polynomials; $X^3 + X + 1$ gives a field of 8 elements and if we are able to somehow discover the correct polynomial to use, we can form fields of order 2^n for any n by this method. This is particularly useful once we see (from Galois theory) that all fields of p^n elements for prime p are isomorphic.

Theorem (2.5) (First isomorphism theorem)

Let $\theta : R \to S$ be a ring homomorphism, then $\operatorname{Im}\theta \leq S, \ker\theta \triangleleft R$ and $\frac{R}{\ker\theta} \cong \operatorname{Im}\theta$: by (2.2) we have $\ker\theta \triangleleft R$. $\operatorname{Im}\theta$ is an additive subgroup of S since θ is a group homomorphism, $\theta(r_1)\theta(r_2) = \theta(r_1r_2) \in \operatorname{Im}\theta$ so $\operatorname{Im}\theta$ is closed under multiplication, and $1_S = \theta(1_R) \in \operatorname{Im}\theta$, so $\operatorname{Im}\theta$ is a subring of S.

Let $I = \ker\theta$ and define $\Phi : \frac{R}{\ker\theta} \to \operatorname{Im}\theta$ by $r + I \mapsto \theta(r)$; this is a group isomorphism (so a bijective homomorphism of additive groups) from the first isomorphism theorem for groups (and also well defined), $\Phi((r_1 + I)(r_2 + I)) = \Phi(r_1r_2 + I) = \theta(r_1r_2) = \theta(r_1)\theta(r_2) = \Phi(r_1 + I)\Phi(r_2 + I)$ and $\Phi(1 + I) = \theta(1_R = 1_S$, so Φ is a bijective ring homomorphism, i.e. a ring isomorphism.

Example

$\theta : \mathbb{R}[X] \to \mathbb{C}$ given by $\sum a_jX^j \mapsto \sum a_ji^j$ has $\ker\theta = (X^2 + 1) = \{f(X) \in \mathbb{R}[X] : f(X) = (X^2 + 1)q(X)$ some $q(X) \in \mathbb{R}[X]\}, \operatorname{Im}\theta = \mathbb{C}$, so $\frac{\mathbb{R}[X]}{(X^2+1)} \cong \mathbb{C}$ as above.

Theorem (2.6) (Second isomorphism theorem)

Let $R \leq S, J \triangleleft S$, then $R \cap J \triangleleft R$ and $\frac{R}{R\cap J} \cong \frac{(R+J)}{J} \leq \frac{S}{J}$; the proof is by applying the first isomorphism theorem to $\theta : R \to \frac{S}{J}$ defined by $r \mapsto r + J$. This is a ring homomorphism; it is a group homomorphism as before, and we have $1 \to 1 + J, \theta(r_1r_2) = \theta(r_1)\theta(r_2)$ since $(r_1r_2 + J) = (r_1 + J)(r_2 + J)$. $\ker\theta = \{r \in R : r + J = J\} = R \cap J$ and $\operatorname{Im}\theta = \{r + J \in \frac{S}{J} : r \in R\} = \frac{(R+J)}{J}$ (note $R + J = \{r + a : r \in R, a \in J\}$ is a subring of S), so we have the result.

We have a correspondence between additive subgroups of R containing I and additive subgroups of $\frac{R}{I}$, where $I \triangleleft R$. We want to extend this: subrings of R containing I correspond to subrings of $\frac{R}{I}$ and similarly for ideals by exactly the same recipe: $L \subset I \mapsto \{a + I : a \in L\}$, with converse $X \subset \frac{R}{I} \mapsto \{r + R : r + I \in X\}$.

Theorem (2.7) (Third isomorphism theorem)

Let $I, J \lhd R$ with $I \subset J$, then $\frac{\frac{R}{I}}{\frac{J}{I}} \cong \frac{R}{J}$, where $\frac{J}{I} = \{r + I : r \in J\}$. Again the proof is by the first isomorphism theorem, this time on $\theta : \frac{R}{I} \to \frac{R}{J}$ defined by $r + I \mapsto r + J$. This is well defined since $I \subset J$ and a group homomorphism as per (1.7); it is a ring homomorphism: $\theta((r_1 + I)(r_2 + I)) = \theta(r_1 r_2 + I) = r_1 r_2 + J = (r_1 + J)(r_2 + J) = \theta(r_1 + I)\theta(r_2 + I)$ [identity ommitted in the lecture, presumably trivial]. $\ker\theta = \{r + I : r + J = J\} = \{r + I : r \in J\}, \operatorname{Im}\theta = \frac{R}{J}$ [and we have the result].

Example

$\frac{\mathbb{Z}[X]}{(2, X^2+1)}$: for this we have $J = (2, X^2+1)$. If we put $I = (2)$ we have $\frac{\mathbb{Z}[X]}{(2,X^2+1)} \cong \frac{\mathbb{F}_2[X]}{X^2+1}$, or if we put $I = (X^2 + 1)$ we have that both these quotient rings are $\cong \frac{\mathbb{Z}[i]}{(2)}$.

Example

For any ring $R, \exists!$ homomorphism $\phi : \mathbb{Z} \to R$ by $1 \mapsto 1_R, m \mapsto 1_R + \cdots + 1_R$.

Definition

The image of ϕ is the <u>prime subring</u> of R. By the isomorphism theorem $\operatorname{Im}\phi \cong \frac{\mathbb{Z}}{\ker\phi}, \ker\phi \lhd \mathbb{Z}$; since the ideals in $\mathbb{Z}$ are of the form $n\mathbb{Z}$ we have:

Definition

The <u>characteristic of R</u> is n such that $\phi(\mathbb{Z}) \cong \frac{\mathbb{Z}}{n\mathbb{Z}}$; this is $|\phi(\mathbb{Z})|$ if this $\neq 0$ and 0 for $\phi(\mathbb{Z})$ infinite. For example the characteristics of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all 0, but that of $\frac{\mathbb{Z}}{p\mathbb{Z}} = \mathbb{F}_p$ for p prime is p.

2.3 Integral domains, fields of fractions, maximal and prime ideals

Definition

$a \in R$ is a <u>zero divisor</u> if $ab = 0$ for some $b \neq 0$

Definition

A ring R is an <u>integral domain</u> if $0 \neq 1$ and there are no nonzero zero divisors, i.e. $ab = 0 \Rightarrow a = 0$ or $b = 0$

Examples

Any field, $\mathbb{Z}$, and subrings of fields such as $\mathbb{Z}[i] \leq \mathbb{C}$ are all integral domains.

Note that in an integral domain we have cancellation: for $a \neq 0, ab = ac \Rightarrow b = c$ as $ab = ac \Rightarrow a(b - c) = 0 \Rightarrow b - c = 0$.

Lemma (2.8)

Let R be a finite integral domain, then R is a field: let $a \neq 0 \in R$. The map (note we do not claim it is a ring homomorphism) $R \to R$ given by $r \mapsto ar$ is injective since R is an integral domain; since R is finite it must also be surjective so $\exists r : ar = 1$ i.e. we have a multiplicative inverse for any nonzero element a, so R is a field.

Definition

An integral domain is a Principal Ideal Domain (PID) if all its ideals are principal.

Proposition (2.9)

If R is an integral domain then so is $R[X]$, and indeed $R[X_1, \ldots, X_n]$. If $f(X) = \sum a_i X^i$ of degree $n, g(X) = \sum b_i X^i$ of degree m with $a_n, b_m \neq 0$ then $f(X)g(X)$ is of degree $m+n$ and thus nonzero, since $a_n b_m \neq 0$ as R is an integral domain.

We can view $R[X_1, X_2]$ as $(R[X_1])[X_2]$ so this is an integral domain, and inductively so is $R[X_1, \ldots, X_n]$.

Theorem (2.10)

Let R be an integral domain, then R has a field of fractions F with $R \leq F$ and any element of F writable as $ab^{-}1$ with $a, b \in R, b \neq 0$:

Define a relation on pairs $[(a, b) \in R \times R, b \neq 0]$: $(a, b) \sim (c, d)$ if $ad = bc$. This is an equivalence relation, as we can easily verify (but this requires an integral domain, as otherwise the relation is not necessarily transitive). Denote the equivalence class of (a, b) by $\frac{a}{b}$. Define addition $\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1 b_2 + a_2 b_1}{b_1 b_2}$ and multiplication $\frac{a_1}{b_1}\frac{a_2}{b_1} = \frac{a_1 a_2}{b_1 b_2}$; these are well defined (note $b_1 b_2 \neq 0$ since R is an integral domain) as can be checked easily for this course (though prooving it for general (i.e. noncommutative) rings is extremely horrible). Let F be the ring defined by this set and operations. We can regard R as a subring of F by $r \mapsto \frac{r}{1}$. F is a field; $\frac{1}{1}$ is the multiplicative identity, and for any $a \neq 0$, $\frac{a}{b}$ [if $a = 0$ this class is $\frac{0}{1} = 0$] has multiplicative inverse $\frac{b}{a}$ since $\frac{ab}{ab} = \frac{1}{1}$. Any $\frac{a}{b} = \frac{a}{1}(\frac{b}{1})^{-1}$; see the printed notes for the 2006 version of this course for more detais on this field.

Lemma (2.11)

The nonzero ring R is a field if and only if the ideals of R are $\{0\}$ and R: for the only if part, if $I \neq \{0\} \triangleleft R$ a field, then take $a \neq 0 \in I$ and then $1 = aa^{-1} \in I$ so $I = R$, for the if part let $a \neq 0 \in R$, and consider the ideal (a); it contains $a \neq 0$ so must be R so $1 \in (a)$ and $\exists r \in R : ar = 1$ and a has a multiplicative inverse, that is, a is a unit, so R is a field.

Definition

$I \triangleleft R$ is a maximal ideal in R if $I \neq R$ and $\forall J$ such that $I \subset J \triangleleft R$, $I = J$ or $J = R$.

Lemma (2.12)

Let $I \triangleleft R$, then $\frac{R}{I}$ is a field if and only if I is a maximal ideal in R; by (2.11) $\frac{R}{I}$ is a field if and only if its only ideals are $\frac{I}{I}, \frac{R}{I}$, and by the correspondence used in the third isomorphism theorem this is the case if and only if the only ideals in R containing I are R, I.

Definition

$I \triangleleft R$ is a prime ideal in R if $I \neq R$ and $\forall a, b \in R, a \in I$ or $b \in I$.

Lemma (2.13)

Let $I \triangleleft R$, then $\frac{R}{I}$ is an integral domain if and only if I is a prime ideal in R.

If $\frac{R}{I}$ is an integral domain let $a, b \in R$ with $ab \in I$, then $(a+I)(b+I) = ab + i = I$ so as $\frac{R}{I}$ is an integral domain either $a + I = I$ i.e. $a \in I$ or $b \in I$.

If I is a prime ideal and $(a+I)(b+I) = ab+I$; if $ab+I = I$ then $ab \in I$ so $a \in I$ or $b \in I$, so either $a + I = I$ or $b + I = I$ and $\frac{R}{I}$ is an integral domain.

Corollay (2.14)

If I is a maximal ideal in R then I is a prime ideal; by (2.12) I is maiximal if and only if $\frac{R}{I}$ is a field, which implies $\frac{R}{I}$ is an integral domain which by (2.13) is the case if and only if I is a prime ideal; prooving this result directly is relatively straightforward.

Example

In $\mathbb{Z}$ the maximal ideals are $p\mathbb{Z}$ for p prime and the prime ideals are $p\mathbb{Z}$ and $\{0\}$.

Proposition (2.15)

Let R be an integral domain (e.g. a field), then the characteristic of R is 0 or some prime p. Recall the characteristic of R is n if and only if the image of $\phi : \mathbb{Z} \to R$ is isomorphic to $\frac{\mathbb{Z}}{n\mathbb{Z}}$, but this has nonzero zero divisors for n not prime (or 0), as if $n = ab$ with $a, b \neq 1$ then $(a + n\mathbb{Z})(b + n\mathbb{Z}) = ab + n\mathbb{Z} = n\mathbb{Z}$; since R is an integral domain the prime subring must be $\cong \frac{\mathbb{Z}}{p\mathbb{Z}}$ for some prime p or $\cong Z$, thus the characteristic of R is p or 0.

2.4 Factorisation in integral domains; units, primes and irreducibles

We assume throughout this subsection that R is an integral domain.

Definition

$a \in R$ is a unit if it has a multiplicative inverse; equivalently $(a) = R$.

a divides b if $\exists c \in R : ac = b$ or equivalently $(b) \subset (a)$.

$a, b \in R$ are associates in R if $a = bc$ for some unit c; equivalently $(a) = (b)$.

$r \in R$ is irreducible in R if it is nonzero, not a unit, and whenever $r = ab, a, b \in R$ then a or b is a unit.

$r \in R$ is prime in R if it is nonzero, not a unit, and if $r \mid ab$ with $a, b \in R$ then $r \mid a$ or $r \mid b$.

Remark

(r) is a prime ideal in R if and only if r is prime or $r = 0$; for (r) a prime ideal if $r \neq 0$ and $r \mid ab$ then $ab \in (r)$ so $a \in (r)$ or $b \in (r)$, then $r \mid a$ or $r \mid b$; (0) is a prime ideal in an integral domain and for r prime, for $ab \in (r)$ we have $r \mid ab$ so $r \mid a$ or $r \mid b$ and $a \in (r)$ or $b \in (r)$.

These definitions of course depend on R; 2 is prime and irreducible in $\mathbb{Z}$ but not in $\mathbb{Q}$, $2X$ is reducible in $\mathbb{Z}[X]$ but not in $\mathbb{Q}[X]$.

Lemma (2.16)

If r is prime in R then r is irreducible in R; the converse is true for PIDs but not in general, see below and (2.20).

Suppose r prime in R and $r = ab$ with $a, b \in R$, then $r \mid ab$ so wlog $r \mid a$, then $a = qr$ for some $q \in r$ so $r = qrb$ and cancelling in an integral domain $1 = br$ so b is a unit.

Example to show converse of (2.16) is false

$R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5}, a, b \in \mathbb{Z}\} \leq \mathbb{C}$ is an integral domain since it is a subring of a field. Define the norm $N(a + b\sqrt{-5}) = a^2 + 5b^2 = z\bar{z}$ where $z = a + b\sqrt{-5}$; this is $\in \mathbb{Z}+$. The norm is multiplicative: $N(z_1 z_2) = N(z_1)N(z_2) \forall z_1, z_2 \in R$. If $z_1 z_2 = 1$ then $1 = N(z_1 z_2) = N(z_1)N(z_2)$ so $N(z_1), N(z_2)$ are units in $\mathbb{Z}$ and > 0 so are 1 and $z_1 = \pm 1, z_2 = \pm 1$, thus the only units in R are ± 1, precisely the elements of norm 1. There are no elements of norm 2 or 3 (since we cannot have $a^2 + 5b^2 = 2, 3$

Consider $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$; $N(2) = 4$ so 2 is irreducible (if $2 = z_1 z_2$ then $N(z_1) \mid N(2) = 4$ in $\mathbb{Z}$ and so one of the $N(z_i)$ is 1 (since there are no elements of norm 2) and some z_i is a unit; similarly $3, (1 + \sqrt{-5}), (1 - \sqrt{-5})$ are all irreducible. However, we have that 2 is not prime in R, since $2 \mid (1 + \sqrt{-5})(1 - \sqrt{-5})$ but $2 \nmid (1 + \sqrt{-5}), 2 \nmid (1 - \sqrt{-5})$ (by considering the norms; $4 \nmid 6$).

Definition

R is a Euclidean Domain or ED if there is a function $\phi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0} = \{n \in \mathbb{Z} : n \geq 0\}$ such that a) $\phi(ab) \geq \phi(a) \forall a, b \in R \setminus \{0\}$ and b) if $a, b \in R, b \neq 0, \exists q, r \in R : a = qb + r$ and either $r = 0$ or $r \neq 0$ and $\phi(r) < \phi(b)$; ϕ is a Euclidean function and b) is the Euclidean algorithm.

Examples

$\mathbb{Z}$ with $\phi(x) = |x|$ is an ED as in 1A Numbers and Sets.

If F is a field, $f[X]$ is an ED with $\phi(f(X))$ =degree of a nonzero polynomial $f(X)$.

$R = \mathbb{Z}[i] = \{a+bi : a, b \in \mathbb{Z}\} \leq \mathbb{C}$, the Gaussian integers, an integral domain: define $N(a+bi) = a^2+b^2 = z\bar{z}$ where $z = a+ib$, then N is a Euclidean function: N is multiplicative $N(z_1 z_2) = N(z_1)N(z_2) \geq N(z_1) \forall z_2 \neq 0$ (since $N(z_2) \geq 1$).

For $z_1, z_2 \in R$ with $z_2 \neq 0$ consider $\frac{z_1}{z_2} \in \mathbb{C}$; this is distance $\leq \frac{1}{\sqrt{2}} < 1$ from an element of $\mathbb{Z}[i]$ because of the lattice $\mathbb{Z}[i]$ forms in the complex plane. So write $\frac{z_1}{z_2} = q + z_4$ with $q \in \mathbb{Z}[i], |z_4| < 1$, then $z_1 = qz_2 + z_2z_4$; let $r = z_2z_4 = z_1 = qz_2 \in \mathbb{Z}[i]$ and then $z_1 = qz_2 + r$ with $N(r) = z_2z_4\overline{z_2z_4} = N(z_2)|z_4|^2 < N(z_2)$ as required.

This last result is true for many similar lattices in the complex plane; the critical property is that any point of the complex plane is < 1 away from some lattice point; we have already seen the result is false for $\mathbb{Z}[\sqrt{-5}]$.

Euclidean Domains

Proposition (2.17)

If R is an ED then R is a PID: let R be an ED with Euclidean function ϕ and $I \triangleleft R$; for $I = \{0\}$ we are done, otherwise take $b \in I$ with $\phi(b)$ minimal. Then $I = (b)$: for any $a \in I$, by Euclid we can write $a = bq + r$ with $r = 0$ or $\phi(r) < \phi(b)$, but since $r \in I$ minimality of $\phi(b)$ implies r must be 0 and we have $a = bq \in (b)$.

Corollay

$\mathbb{Z}, F[X]$ for F a field and $\mathbb{Z}[i]$ are principal ideal domains.

Example

$\mathbb{Z}[X]$ is <u>not</u> a PID: $(2, X) \triangleleft \mathbb{Z}[X]$ is <u>not</u> a principal ideal. If it were $(f(X))$ for some $f(X) \in \mathbb{Z}[i]$ we would have $2 = f(X)g(X)$ so by degrees f is constant and furthermore $f(X) \in \{\pm 1, \pm 2\}$. But $X = f(X)h(X)$ some $h(X)$ so ± 2 are impossible, but $\pm 1 \notin (2, X)$, so there can be no such $f(X)$.

Example

Minimal polynomials of matricies; for $A \in M_n(F)$ an $n \times n$ matrix over a field F, $I = \{f(X) \in F[X] : f(A) = \vec{0}\} \triangleleft F[X]$ in fact; $F[X]$ is a PID so $I = (m(X))$ some $m(X)$; by multiplying by the inverse of a nonzero element of F we can take $m(X)$ to be monic, then it is the <u>minimal polynomial</u> of A.

Definition

An integral domain is a Unique Factorisation Domain (UFD) if a) every element which is nonzero and not a unit is a product of finitely many irreducibles and b) if $p_1 \dots p_m = q_1 \dots q_n$ with p_i, q_j irreducible then $m = n$ and we can reorder such that p_i, q_i are associates (note not necessarily equal) $\forall i$.

We are now working towards a proof of:

Proposition (2.18)

If R is a PID then R is a UFD.

This immediately gives:

Corollay (2.19)

$\mathbb{Z}, F[X], \mathbb{Z}[i]$ are UFDs

Example

$\mathbb{Z}[\sqrt{-5}]$ is <u>not</u> a UFD since $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$; all these factors are irreducibles and they are not associates as e.g. 2 is not an associate of either of the RHS factors by considering norms; they have norm 6 while $N(2) = 4$, and the norm of any unit is 1.

To proove (2.16) we want the following lemmas:

Lemma (2.20) (converse of (2.16) for PIDs)

Let R be a PID. If p is irreducible in R then p is prime in R; let p be irreducible with $p \mid ab$ and suppose $p \nmid a$. Consider the ideal (p, a); it must be principal since R is a PID so say it is (d). then $p = q_1 d, a = q_2 d$ for some q_1, q_2. If $p \mid d$ then $p \mid a$ and we have a contradiction, but since p is irreducible q_1 or d is a unit and if q_1 is a unit then $d = q_1^{-1} p$ and $p \mid d$, so d is a unit so $(p, a) = (d) = R$ and $\exists r, s \in R$ such than $1 = rp + sa$. Then $b = rpb + sab$ but $p \mid ab$ so $p \mid$ RHS so $p \mid \text{LHS} = b$ as required.

Lemma (2.21) (Ascending Chain Condition (ACC) for ideals in PIDs)

Let R be a PID, $I_j \triangleleft R$ with $I_1 \subset I_2 \subset \ldots$ [in the lecture these were written as $I_1 \subsetneq \ldots$ but that seems insane], then for some n, $I_n = I_{n+i} \forall i \geq 0$: let $I = \bigcup_{j \geq 1} I_j$, then $I \triangleleft R$ since if $a \in I_j, b \in I_k$ take wlog $j \leq k$ and then $a - b \in I_k \subset I$, and if $a \in I$ then $a \in I_j$ some j so for any $r \in R$, $ar \in I_j \subset I$. In a PID, $I = (a)$ for some $a \in R$ but then $a \in I_n$ for some n, so $(a) = I \supset I_{n+i} \supset (a)$ so $(a) = I_n = I_{n+i} \forall i \geq 0$.

We can now proove (2.18): a) let $a \in R$ be a nonzero nonunit, and assume a cannot be factorized as a product of irreducibles, then $a = a_1 b_1$ with neither of a_1, b_1 zero or a unit, and one of the a_1, b_1, wlog a_1, cannot be factorized as a product of irreducibles. So we have $a_1 = a_2 b_2$ and similarly. Then we have $(a_1) \subset (a_2) \subset \ldots$ with inequality in each case, since if $(a_i) = (a_{i+1}$ then a_i, a_{i+1} are associates so b_{i+1} is a unit. But this is a contradiction by (2.21). b) Suppose $a = p_1 \ldots p_m = q_1 \ldots q_n$ with the p_i, q_j irreducibles. Then p_1 is prime by (2.20) and $p_1 \mid q_1 \ldots q_n$ so $p_1 \mid q_i$ for some i; by rearranging if necessary $p_1 \mid q_1$. q_1 is irreducible and p_1 is not a unit so $p_1 \ldots p_m = u p_1 q_2 \ldots q_n$ for some unit u, and cancelling in an integral domain $p_2 \ldots p_m = u q_2 \ldots q_n$ and we induct.

There are three important properties of UFDs:

a) Irreducibles are the same as primes in and UFD; we have that primes are irreducibles, now suppose p is irreducible and $p \mid ab$. Let $ab = pc$ and express as products of irreducibles :$a = p_1 \ldots p_m, b = q_1 \ldots q_n, c = r_1 \ldots r_s$. Then $p r_1 \ldots r_s = p_1 \ldots p_m q_1 \ldots q_n$ and by unique factorization p is an associate of some p_i or q_i, so $p \mid a$ or b.

b) Highest common factors exist in a UFD:

Definition

A highest common factor d of $a_1 \ldots, a_n$ is a d such that $d \mid a_i \forall i$ and if $d' \mid a_i \forall i$ then $d' \mid d$.

To find a highest common factor of $a_1, \ldots, a_n$ in a UFD we can express each a_i as a product of irreducibles; by replacing irreducibles with associates we can write $a_i = u_i \prod_j p_j^{n_{ij}}$ with p_j irreducible, u_i a unit, and $n_{ij} \neq 0 \forall i, j$ and p_j, p_k not associates for any $j \neq k$. Then a highest common factor is $d = \prod p_j^{|_i n_{ij}}$; it is a factor of each a_i and if d' is a factor of each a_i then $a_i = d'c_i$, then each irreducible dividing d' must appear as an irreducible dividing each a_i so $d' = u \prod p_j^{m_j}$ with $m_j \leq n_{ij}$ for each i.

c) Lowest common multiples exist in a UFD; the proof is similar and left as an exercise.

2.5 Factorisation in polynomial rings, Gauss' Lemma, Eisenstein's Criterion

Reminder about polynomial rings with field coefficients

If F is a field we know that $F[X]$ is an ED, PID, UFD, every ideal $I \triangleleft F[X]$ is principal i.e. $(f(X))$ for some f, irreducibles are the same as primes, and I being maximal is equivalent both to $f(X)$ being irreducible and to $\frac{F[X]}{I}$ being a field. The elements of $\frac{F[X]}{I}$ are of the form $\sum_{i=0}^{\deg f - 1} a_i X^i + I$.

Now consider the coefficient ring R a UFD but not necessarily a field; we have from before that $R[X]$ is an integral domain.

Definition

The content $c(f)$ is the HCF of $a_0, \ldots, a_n$ (only defined up to associates).

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

Lemma (2.22) (Gauss' Lemma)

(Note that the statement of this lemma varies between different books)

Let R be a UFD and F its field of fractions. Suppose $f(X) \in R[X]$ is primitive. Then $f(X)$ is irreducible in $R[X]$ if and only if it is irreducible in $F[X]$; in particular for $R = \mathbb{Z}$, $f(X)$ is irreducible in $\mathbb{Q}[X]$ if and only if it is irreducible in $\mathbb{Z}[X]$.

Example

$X^3 + X + 1$ is irreducible in $\mathbb{Z}[X]$ and hence in $\mathbb{Q}[X]$ so $\frac{\mathbb{Q}[X]}{X^3+X+1}$ is a field: we try to factorize $X^3 + X + 1$ in $\mathbb{Z}[X]$ as $p(X)q(X)$ with neither of p, q units since the polynomial is primitive if $p(X)$ or $q(X)$ is $\in R$ then it is a unit, so we have wlog $\deg P = 2, \deg Q = 1$, i.e. $q(X) = b_1 X + b_0, p(X) = a_2 X^2 + a_1 X + a_0$. Then $a_0 b_0 = 1, a_2 b_1 = 1$ in $\mathbb{Z}$ so $b_0 = \pm 1, b_1 = \pm 1$ but ± 1 are not roots of $X^3 + X + 1$, so there can be no such factorization.

Lemma (2.23)

If $f(X), g(X) \in R[X]$ are primitive in $R[X]$ then so is $f(X)g(X)$: let $f(X) = a_0 + \cdots + a_n X^n, g(X) = b_0 + \cdots + b_m X^m$ be primitive. If $f(X)g(X)$ not primitive we have some prime p dividing $c(f(X)g(X))$. We have $p \nmid c(f), c(g)$; let k be such taht $p \mid a_0, \ldots, a_{k-1}$ but $p \nmid a_k$ and similarly $p \nmid b_l$. Then the coefficient of X^{k+l} in $f(X)g(X)$ is $\cdots + a_{k+1}b_{l-1} + a_k b_l + a_{k-1}b_{l+1} + \ldots$; p divides this if and only if $p \mid a_k b_l$, but it does not, so $p \nmid$ the $k+l$ coefficient of $f(X)g(X)$ so $p \nmid c(f(X)g(X))$, a contradiction.

Corollay (2.24)

For $f(X), g(X) \in R[X]$ the contents may be chosen such that $c(f(X)g(X)) = c(f(X))c(g(X))$: Let $f(X) = c(f(X))f_1(X)$ with $f_1(X)$ primitive and similarly for $g(X)$, then by (2.23) $f_1(X)g_1(X)$ is primitive so a HCF of the coefficients of $f(X)g(X) = c(f(X))c(g(X))f_1(X)g_1(X)$ is $c(f(X))c(g(X))$.

Proof of (2.22)

Take $f(X) \in R[X]$ primitive. Suppose $f[X]$ factorises in $R[X]$ as a product of two non-units. Since $f(X)$ is primitive any non-unit factors in $R[X]$ are of degree ¿0, so it also factors as a product of two non-units in $F[X]$.

For the converse, suppose $f(X) = g(X)h(X)$ in $F[X]$. The coefficients of $g(X), h(X)$ are $\in F$; multiplying by elements of R we can "clear the denominators" and have $ag(X), bh(X) \in R[X]$, so $abf(X) = ag(X)bh(X)$; by (2.24) we can choose $ab = c(abf(X)) = c(ag(X))c(bh(X))$ and $ag(X) = c(ag(X))g_1(X)$ with g_1 primitive and similarly for $bh(X)$. But then $g_1(X)h_1(X)$ is primitive by (2.23), so $abf(X) = c(ag(X))c(bh(X))g_1(X)h_1(X)$ and cancelling a factor of ab $f(X) = g_1(X)h_1(X)$ a product of non-units in $R[X]$.

Remark

A similar argument shows that for $f(X) \in R[X]$ not necessarily primitive, if $f(X) = g_1(X)h(X)$ with $g_1(X)$ primitive in $R[X]$ and $h(X) \in F[X]$ then $f(X) = g_1(X)h_0(X)$ for some $h_0(X) \in R[X]$. Thus if $I = (g_1(X)) \lhd F[X], J = (g_1(X)) \lhd R[X]$ for g_1 primitive then $I \cap R[X] = J$; we would be disappointed were this not the case but do need to verify this.

Theorem (2.25) (Gauss)

Let R be a UFD, then so is $R[X]$: let $f(X) = R[X]$ be a nonzero nonunit and write $f(X) = c(f(X))f_1(X)$ with $f_1(X)$ primitive; we know $c(f(X))$ can be expressed essentially uniquely as a product of irreducibles in R (which are irreducible in $R[X]$), and any irreducible in $R[X]$ is either in R and irreducible in R or primitive, so if $f(X)$ factorizes as a product of irreducibles in $R[X]$ we can collect those in R together and their product is $c(f(X))$ (up to associates) and the product of the remainder is is $f_1(X)$. So for both existence and uniqueness it suffices to proove for f primitive.

$F[X]$ is an ED, PID and UFD so we can factorise $f(X) = p_1(X) \ldots p_n(X)$ with $p_i(X)$ irreducible in $F[X]$. Then $a_i p_i(X) \in R[X]$ for some $a_i \in R$; let these be $c_i q_i(X)$ where $c_i = c(a_i p_i(X)), q_i(X)$ primitive in $R[X]$ and irreducible by

Gauss' Lemma. So $a_1 \dots a_k f(X) = c_1 \dots c_k q_1(X) \dots q_k(X)$; $q_1(X) \dots q_k(X)$ is primitive by (2.23) so by considering contents $a_1 \dots a_k = uc_1 \dots c_k$ for some unit $u \in R$. So $a_1 \dots a_k f(X) = (uc_1 \dots c_k)(u^{-1}q_1(X))q_2(X) \dots q_k(X)$, cancelling and relabelling q_1, $f(X) = q_1(X) \dots q_n(X)$, a product of irreducibles in $R[X]$.

Say we also have $f(X) = r_1(X) \dots r_l(X)$ a product of irreducibles. We are assuming $f(X)$ is primitive, so by Gauss' Lemma (2.22) $r_i(X)$ is irreducible in $F[X$; then by uniqueness of factorization in $F[X]$, $k = l$ and after reordering, each $q_i(X) = u_i r_i(X)$ for u_i a unit in F. Let $u_i = \frac{a_i}{b_i}$ (with $a_i, b_i \in R, b_i \neq 0$), then $b_i q_i(X) = a_i r_i(X)$ but $q_i(X), r_i(X)$ are primitive so a_i, b_i are associates; cancelling $q_i(X)$ and $r_i(X)$ are associates so the factorization is essentially unique.

Corollay (2.26)

If R is a UFD so is $R[X_1, \dots, X_n]$; this is trivial.

Theorem (2.27) (Eisenstein's Criterion)

Let R be a UFD, $f(X) = a_n X^n + \dots + a_0 \in R[X]$ primitive with $a_n \neq 0$. If for some irreducible $p \in R$ we have $p \mid a_0, \dots, a_{n-1}$ but $p \nmid a_n, p^2 \nmid a_0$ then $f(X)$ is irreducible in $R[X]$ (and thus in $F[X]$): If $f(X) = g(X)h(X)$ with $g(X), h(X) \in R[X]$ let $g(X) = r_k X^k + \dots + r_0, h(X) = s_l X^l + \dots + s_0$; by considering degrees $k+l = n$. Since $p \mid a_0 = r_0 s_0$ but $p^2 \nmid a_0$, p divides precisely one of r_0, s_0; wlog take $p \mid r_0, p \nmid s_0$. Let j be such that $p \mid r_0, \dots, r_{j-1}$ but $p \nmid r_j$ [we can do this since $p \nmid a_n = r_k s_j \therefore p \nmid r_k$]. Now $a_j = r_j s_0 + r_{j-1} s_1 + \dots + r_0 s_j$ so $p \nmid a_j$, so $j = n$ meaning $k = n, l = 0$ but since $f(X)$ is primitive this means $h(X)$ is a unit. So $f(X)$ cannot be expressed as a product of two nonunits in $R[X]$, i.e. is irreducible in $R[X]$.

Example

Take $R = \mathbb{Z}$, then $f(X) = X^n - p$ for p prime in $\mathbb{Z}$ and $n > 1$ is irreducible is $\mathbb{Z}[X]$, and hence in $\mathbb{Q}[X]$, by Eisenstein. So any p prime in $\mathbb{Z}$ has no rational nth root for $n > 1$.

Example

Cyclotomic polynomials $X^{p-1} + X^{p-2} + \dots + 1$ for p prime are irreducible in $\mathbb{Z}[X]$ (and hence in $\mathbb{Q}[X]$): Eisenstein doesnt' immediately apply but we note $(X-1)f(X) = X^p - 1$ so substitute $X = 1 + Y$, then $Yf(Y+1) = (Y+1)^p - 1$ and $f(Y+1) = Y^{p-1} + \binom{p}{1}Y^{p-2} + \dots + \binom{p}{p-1}$; we have $p \mid$ each of the $\binom{p}{i}$ but $p^2 \nmid \binom{p}{p-1} = p$ so Eisenstein applies, $f(Y+1)$ is irreducible and hence $f(X)$ is irreducible as required.

2.6 Gaussian Integers

$R = \mathbb{Z}[i] = \{a+bi : a, b \in \mathbb{Z}\}$ with norm $N(a+bi) = a^2+b^2 = (a+bi)(a-bi) = z\bar{z}$ where $z = a + bi$. This is multiplicative $N(z_1 z_2) = N(z_1)N(z_2)$, and the units are precisely the elements with norm 1, namely $\pm 1, \pm i$. $\mathbb{Z}[i]$ is an ED and hence a PID and UFD.

The irreducibles are precisely the primes (by (2.16),(2.20)) [I shall use the two terms interchangably in this section]. $2 = (1+i)(1-i)$ is not irreducible, 3 is irreducible since $N(3) = 9$ and there are no elements of norm 3, $5 = (1+2i)(1-2i)$ is not irreducible; what are the primes?

Lemma (2.28)

Let p be prime in $\mathbb{Z}$, then p is prime in $\mathbb{Z}[i]$ if and only if it cannot be expressed as a sum of squares $p = x^2 + y^2$ in $\mathbb{Z}$: $N(p) = p^2$. p is irreducible in $\mathbb{Z}[i]$ if and only if $p \neq z_1 z_2 \forall$ nonunit $z_i \in \mathbb{Z}[i]$; if $p = z_1 z_2$ then $N(z_1) = N(z_2) = p$, so $p = z_1 \bar{z}_1 = x^2 + y^2$ where $z_1 = x + iy$; conversely if $p = x^2 + y^2$ then $p = (x+iy)(x-iy)$ and is not irreducible.

Proposition (2.29)

The irreducibles in $\mathbb{Z}[i]$ are associates of:

a) prime integers $p \in \mathbb{Z}$ with $p \equiv 3 (\mathrm{mod}\, 4)$

b) $z \in \mathbb{Z}[i]$ with $z\bar{z} = x^2 + y^2 = p$, a prime integer satisfying $p = 2$ or $p \equiv 1(4)$.

Observe if $p \in \mathbb{Z}[i]$ prime with $p \equiv 3(4)$ it cannot be $x^2 + y^2$ since the only squares in mod 4 are 0,1. For $p = 2$, $2 = (1+i)(1-i)$ a product of irreducibles.

For $p \equiv 1(4)$, $p = 4n + 1$ consider the field $\mathbb{F}_p$; its multiplicative group has $4n$ elements (since it doesn't include 0); by Question 10 on the second Example Sheet for this course, it is cyclic. So there is a unique element of order 2, namely -1, and a unique element a of order 4 such that $a^2 = -1$. Thus $p \mid a^2 + 1 = (a+i)(a-i)$, but $p \nmid a \pm i$ in $\mathbb{Z}[i]$ so p is <u>not</u> prime; it must factorize $p = z_1 z_2$ in $\mathbb{Z}[i]$; the z_i are nonunits so we must have $N(z_i) = p$, i.e. each z_i is $x + iy$ such that $x^2 + y^2 = p$, and the z_i are irreducibles of the second form. [Any z with $z\bar{z} = x^2 + y^2 = p$ is clearly irreducible].

Now suppose α is irreducible in $\mathbb{Z}[i]$; take $p \in \mathbb{Z}$ with $p \mid N(\alpha) = \alpha\bar{\alpha}$. If $p \equiv 3(4)$ then p is prime in $\mathbb{Z}[i]$ so p is associate to one of the irreducibles $\alpha, \bar{\alpha}$ ($\bar{\alpha}$ is irreducible since α is), so α is associate to p and on the above list; if $p = 2$ or $p \equiv 1(4)$ then $p = z\bar{z} \mid \alpha\bar{\alpha}$ are two factorizations of p ase a product of irreducibles; $\mathbb{Z}[i]$ is a UFD so z must be associate to α or $\bar{\alpha}$ and we are done by the above.

Corollay (2.30)

Let $n = p_1^{n_1} \dots p_k^{n_k}$ be the (essentially) unique prime factorisation of n in $\mathbb{Z}$. Then $n = x^2 + y^2$ for some $x, y \in \mathbb{Z}$ if and only if for any p_i with $p_i \equiv 3(4)$, n_i is even, as if $n = x^2 + y^2 = z\bar{z}$ where $z = x + iy$ we have $N(z) = n$. Express z as a product of irreducibles in $\mathbb{Z}[i]$; $z = \alpha_1 \dots \alpha_s$. We know the α_i are of the forms listed in (2.29), i.e. either $N(\alpha_j) = p_j^2$ with $p_j \equiv 3(4)$ or $N(\alpha_j) = p_j$ for $p_j = 2$ or $p_j \equiv 1(4)$. Then $n = N(z) = N(\alpha_1 \dots \alpha_s) = \prod N(\alpha_j)$ and n is of the required form. Conversely if $n = p_1^{n_1} \dots p_k^{n_k}$ with the n_j even if $p_j \equiv 3(4)$ then we can replace any p_j which are 2 or $\equiv 1(4)$ with $p_j = \alpha_j \bar{\alpha}_j$ and any other primes by $p_j = \alpha_j$ with $p_j^2 = \alpha_j \bar{\alpha}_j$. Then n is $z\bar{z}$ for some $z \in \mathbb{Z}[i]$; let $z = x + iy$, then $n = x^2 + y^2$.

Example

The (essentially) unique prime factorization of 65 in $\mathbb{Z}[i]$ is $65 = 5 \times 13 = (2+i)(2-i)(2+3i)(2-3i)$, so the only possible ways (up to associates [and which factor is z or $\bar{z}$]) to express 65 as $z\bar{z}$ are given by $z = (2+i)(2+3i) = 1+8i$ or $(2+i)(2-3i) = 7-4i$, i.e. $65 = (1+8i)(1-8i) = 1^2+8^2, 65 = (7-4i)(7+4i) = 7^2+4^2$.

2.7 Rings $\mathbb{Z}[\alpha]$ of algebraic integers

This section is mostly statements and definitions rather than proofs:

Definition

$\alpha \in \mathbb{C}$ is an algebraic integer if $f(\alpha) = 0$ for some monic $f(X) \in \mathbb{Z}[X]$, e.g. $i, \sqrt{2}, \frac{1}{2}(1+\sqrt{-3})$ (this last being a root of $X^2 - X + 1$. $\mathbb{Z}[\alpha] \leq \mathbb{C}$ is the smallest subring of $\mathbb{C}$ containing α; it is $\cong \frac{\mathbb{Z}[X]}{I}$ where I is the kernel of $\theta : \mathbb{Z}[X] \to \mathbb{Z}[\alpha]$ given by $g(X) \mapsto g(\alpha)$.

Remarks

We can show (this is a reasonable exercise for the keen student) that for an algebraic integer α the ideal I above is principal, i.e. generated by some $f_\alpha(X) \in \mathbb{Z}[X]$. $\mathcal{O}_\alpha$ will be monic and irreducible; it is the minimal polynomial of α.

The elements of $\mathbb{Z}$ are called rational integers since if $\alpha \in \mathbb{Q}$ is an algebraic integer then $\alpha \in \mathbb{Z}$.

Example

For α an algebraic integer with minimal polynomial $f_\alpha(X)$ and $p \in \mathbb{Z}$ a prime, by the isomorphism theorems $\frac{\mathbb{Z}[\alpha]}{(p)} \cong \frac{\mathbb{Z}[X]}{(f_\alpha(X), p)} \cong \frac{\mathbb{F}_p[X]}{(\bar{f}_\alpha(X))}$ where $\bar{f}_\alpha(X) \in \mathbb{F}_p[X]$ is obtained by taking the coefficients of $f_\alpha(X)$ in modulo p; this follows from the fact that $\mathbb{Z}[\alpha] \cong \frac{\mathbb{Z}[X]}{(f_\alpha(X))}$.

For example, if we take $\alpha = i$ then $f_\alpha(X) = X^2 + 1$ and we have $\frac{\mathbb{Z}[i]}{(p)} \cong \frac{\mathbb{Z}[X]}{(X^2+1, p)} \cong \frac{\mathbb{F}_p[X]}{(X^2+1)}$; if $p = 2$ or $p \equiv 1(4)$ this is not an ID but if $p \equiv 3(4)$ then this is an ID so $X^2 + 1$ is irreducible in $\mathbb{F}_p[X]$.

In the part II Number Fields course we consider fields of the form $\mathbb{Q}[\sqrt{\alpha}]$; inside these fields the algebraic integers form rings (though not necessarily $\mathbb{Z}[\alpha]$. There are 21 possibilities for which R is an ED; R is a UFD for infinitely many $\alpha > 0$ but onli in a boundednumber of cases (in fact 9) for $\alpha < 0$, as was proven by Baker of Trinity College.

2.8 Noetherian Rings, Hilbert's Basis Theorem

Recall (2.21): a PID satisfies the ACC on ideals.

Lemma (2.31)

R satisfies the ACC if and only if all its ideals are finitely generated: suppose all ideals are finitely generated and $I_1 \leq I_2 \leq \dots$ is an ascending chain with $I_j \lhd R \forall j$. Then $\bigcup_j I_j \lhd R$ so is finitely generated, say $= (a_1, \dots, a_n)$. But then each a_i is $\in I_{j(i)}$ for some $j(i)$; let $k = \max_i j(i)$, then $(a_1, \dots, a_n) \subset I_k$ so $I_k = I_{k+1} = \dots$ as required. Conversely suppose $J \lhd R$ is not finitely generated, then pick $a_1 \in J, a_2 \in J \setminus (a_1), a_3 \in J \setminus (a_1, a_2), \dots$, then $(a_1) \subsetneqq (a_1, a_2) \subsetneqq (a_1, a_2, a_3) \subsetneqq \dots$ so the ACC does not hold.

Definition

A ring satisfying the ACC is Noetherian.

Example of a non-Noetherian ring

For F a field, $F[X_1, X_2, \dots]$ the ring of polynomials in a countably infinite number of variables has $(X_1) \subsetneqq (X_1, X_2) \subsetneqq \dots$ so the ACC does not hold and this ring is not Noetherian.

Theorem (2.32) Hilbert's Basis Theorem

Let R be a Noetherian ring, then $R[X]$ is Noetherian: let $J \lhd R[x]$. Consider $I_n = \{a_n \in R : \sum_{i=0}^n a_i X^i \in J\} \cup \{0\}$, the set of leading coefficients of elements of J of degree n [and 0]. Then $I_n \lhd R$, since for $\sum_0^n a_i X^i, \sum_0^n b_i X^i \in J$ so too is $\sum (a_i + b_i) X^i$, and $\sum a_i r X^i \in J \forall r \in R$; $I_n \subset I_{n+1} \forall n$ since $\sum_0^n a_i X^i \in J \Rightarrow X \sum_0^n a_i X^i \in J$, so by assumption $\exists N$ such that $I_N = I_{N+1} = \dots$. By (2.31) I_N is finitely generated; let $f_1(X), \dots, f_k(X)$ be polynomials of degree N in J whose leading coefficients generate I_N. Take $f(X) \in J$ of degree $m \geq N$; since $I_m = I_n$ there exist $r_1, \dots, r_k \in R : r_1 f_1(X) + \dots + r_k f_k(X)$ has the same leading coefficient as $f(X)$, so $f(X) - (r_1 f_1(X) + \dots + r_k f_k(X)) Y^{m-N} \in J$ is of degree $< m$; repeating we have $q_1(X), \dots, q_k(X) \in R[X]$ such that $f(X) - (q_1(X) f_1(X) + \dots + q_k f_k(X)) \in J$ is of degree $< N$. Now consider polynomials of degree $< n$ in J; for each $j < N$ there is a finite set S_j of polynomials of degree j in J whose leading coefficients generate I_j. Let $S = \bigcup_{j<N} S_j$, a finite set. Then if $f(X)$ is of degree $j < N$ then $\exists r_i' \in R$ such that $f(X) - \sum r_i' g_i(X)$ where $S_j = \{g_1(X), \dots, g_l(X)\}$ is $\in J$ and of smaller degree than X; thus we can reduce the degree until we get 0, so J is generated by $S \cup \{f_1(X), \dots, f_k(X)\}$, a finite set.

Corollay

$\frac{\mathbb{Z}[X_1, \dots, X_n]}{I}$ is Noetherian for any $I \lhd \mathbb{Z}[X_1, \dots, X_n]$: by inductively applying Hilbert, $\mathbb{Z}[X_1, \dots, X_n]$ is Noetherian, then if $\frac{J}{I}$ is an ideal of $\frac{\mathbb{Z}[X_1, \dots, X_n]}{I}$ then it corresponds to $J \lhd \mathbb{Z}[x_1, \dots, X_n]$ with $I \subset J$, then J is finitely generated since $\mathbb{Z}[X_1, \dots, X_n]$ is Noetherian so $\frac{J}{I}$ is finitely generated as required.

3 Modules

These are generalizations of vector spaces, where the coefficients are no longer fields; we will concentrate on modules whose coefficients are taken from EDs,

particularly $\mathbb{Z}$ and $F[X]$ for fields F.

3.1 Definitions, examples, submodules, homomorphisms, quotient modules, direct sums

We take R to be a commutative ring throughout; if it were not we would need to talk about left- and right modules.

Definition

The set M is an R-module if there is a binary operation $+$ for which $(M,+)$ is an abelian group and a map $R \times M \to M$ given by $(r,m) \mapsto rm$ such that $(r_1+r_2)m = r_1m+r_2m, r(m_1+m_2) = rm_1+rm_2, r_1(r_2m) = (r_1r_2)m, 1m = m$.

Examples

For $R = F$ a field a module is simply a vector space over F.

For any R, R^n is an R-module under $r(r_1,\ldots,r_n) = (rr_1,\ldots,rr_n)$; in particular this means R itself is an R-module under multiplication in R.

For any ring R, $I \lhd R$ and $\frac{R}{I}$ for such I are R-modules.

This and the next are the focus of this part of the course: for $R = \mathbb{Z}$ the $\mathbb{Z}$-modules are precisely the abelian groups, since if A is an abelian group written additively then for $n \in \mathbb{N}_0$ we set $na = a+\cdots+a, (-n)a = -(na)$, then A is a $\mathbb{Z}$-module.

If V is a vector space over a field F, for a fixed linear map $\alpha : V \to V$, V is an $F[X]$-module under $f(X)V = f(\alpha)(V)$; different α generally give different modules.

For $R \leq S$, S is an R-module under multiplication in S.

3.1.1 Definition

A subset N of an R-module M is an $\underline{R\text{-submodule}}$ if it is an additive subgroup of M and $rn \in N \forall r \in R, n \in N$; the (overused) notation is $N \leq M$.

Example

The R-module R has submodules $I \lhd R$, the ideals of R; the submodules of a vector space are its subspaces.

Definition

If $N \leq M$ the quotient module $\frac{M}{N}$ has elements $m+N$ and $r(m+N) = rm$; it is clearly an additive group (the quotient group $\frac{M}{N}$) and the reader should verify such multiplication is well defined.

Definition

An $\underline{R\text{-module homomorphism}}$ $\theta : M \to N$ is a group homomorphism (i.e. satisfies $\theta(m_1+m_2) = \theta(m_1)+\theta(m_2)$) with $\theta(rm) = r\theta(m)$; the image of θ is an R-submodule of N and its kernel is an R-submodule of M (exercises).

3.1.2 Theorem (Isomorphism theorems)

Let $\theta : M \to N$ be an R-module homomorphism, then $\frac{M}{\ker\theta} \cong \operatorname{Im}\theta \leq N$ as R-modules (i.e. we have a bijective R-module homomorphism between them); the proof is left as an exercise.

As before, we have a one-to-one correspondence between submodules of $\frac{M}{M_1}$ and submodules of M containing M_1, and if $M_1 \leq L \leq M$ then $\frac{\frac{M}{M_1}}{\frac{L}{M_1}} \cong \frac{M}{L}$.

Example

For the special case $R = F$, i.e. $W \leq V$ vector spaces, $\frac{V}{W}$ is a quotient space. $\theta : V \to U$ linear has $\frac{V}{\ker\theta} \cong \operatorname{Im}\theta \leq U$; linear maps are F-module homomorphisms.

Definition

The annihilator of $m \leq M$ is $\operatorname{Ann}(m) = \{r \in R : rm = 0\}$; the annihilator of M is $\operatorname{Ann} M = \{r \in R : rm = 0 \forall m \in M\}$; clearly this is $\bigcup_{m\in M} \operatorname{Ann}(m)$. We have $\operatorname{Ann}(m), \operatorname{Ann}(M) \lhd R$ since $r_1m = 0, r_2m = 0 \Rightarrow (r_1 + r_2)m = 0, r_1m = 0 \Rightarrow (rr_1)m = 0 \forall r \in R$.

Definition

For $m_1, \ldots, m_n \in M$ the submodule generated by $m_1, \ldots, m_n$ is $Rm_1 + \cdots + Rm_n = \{r_1m_2 + \cdots + r_nm_n : r_i \in R\}$; a module generated by one element is cyclic.

3.1.3 Lemma

$Rm \cong \frac{R}{\operatorname{Ann}(m)} \forall m \in M$, by applying the 1st isomorphism theorem (3.1) to $\theta : R \to M$ given by $r \mapsto rm$.

Example

If $R = F$ a field then the finitely generated modules are the finite dimensional vector spaces.

Example

If $R = \mathbb{Z}$ the $\mathbb{Z}$-submodules of $\mathbb{Z}$ are the ideals $n\mathbb{Z}$; since these are principal ideals they are cyclic modules.

Example

$\mathbb{Z} \leq \mathbb{Z}[\alpha]$ for any algebraic integer α; $\mathbb{Z}[\alpha]$ is a finitely generated $\mathbb{Z}$-module (the proof of this is an exercise, and uses the fact that α is a root a monic polynomial).

3.1.4 Lemma

Let $N \leq M$ be R-modules, then if M is finitely generated then $\frac{M}{N}$ is finitely generated; if $m_1, \ldots, m_n$ generate M then $m_1 + N, \ldots, m_n + N$ generate $\frac{M}{N}$

Warning

This is not as trivial as it might seem, because it is not always the case that for M finitely generated, $N \leq M$ is finitely generated, since if we let $R = F[X_1, \dots]$ the ring of polynomials over countably infinitely many variables, and let I be the ideal of polynomials with zero constant term i.e. the union of the chain $(X_1) \subsetneqq (X_1, X_2) \subsetneqq \dots$, which does not satisfy the ACC for ideals so cannot be finitely generated, thus has $I \leq R$ not finitely generated as an R-module, while $R = R1$ is generated by 1.

Definition

Given R-modules $M_1, \dots, M_k$ the (external) direct sum $M_1 \oplus \dots \oplus M_k$ has elements $(m_1, \dots, m_k)$ for $m_i \in M_i$, with addition $(m_1, \dots, m_k) + (m'_1, \dots, m'_k) = (m_1 + m'_1, \dots, m_k + m'_k)$ and scalar multiplication $r(m_1, \dots, m_k) = (rm_1, \dots, rm_k)$. There is also an internal direct sum: for $M_i \leq M$ if each element of $M_1 + \dots + M_k \leq M$ is uniquely expressible as $m_1 + \dots + m_k$ for $m_i \in M_i$ then $M_1 + \dots + M_k$ is a direct sum; equivalently $M_i \cap \sum_{j \neq i} M_j = \{0\} \forall i$.

Example

$R^n = R \oplus \dots \oplus R$

Proposition (3.4)

Let M be an R-module generated by $m_1, \dots, m_k$, then there is an R-module homomorphism $\theta : R^k \to M$ by $(r_1, \dots, r_k) \mapsto r_1 m_1 + \dots + r_k m_k$ (the reader should verify that this is indeed a homomorphism); this is clearly surjective. Then $M \cong \frac{R^k}{\ker \theta}$ by the isomorphism theorem.

Remark

θ is dependent on the choice of generating set.

Definition

If θ is an isomorphism, i.e. $\ker \theta = \{\vec{0}\}$, then we say the generating set $m_1, \dots, m_k$ is a basis.

Warning

$\{2, 3\}$ is a generating set for $\mathbb{Z}$ (the $\mathbb{Z}$-module) but no subset therof is a basis (wheras in a vector space there would always exist such a subset).

$\ker \theta$ is called the relation module (and depends on the choice of generators); for $(r_1, \dots, r_k) \in \ker \theta$, $r_1 m_1 + \dots + r_k m_k = 0$ is a relation. If the relation module is finitely generated by $n_i = (r_{i1}, \dots, r_{ik})$ then M is generated by $\{m_1, \dots, m_k\}$ subject to the relations $r_{i1} m_1 + \dots + r_{ik} m_k = 0$ for each i.

3.2 Matricies over EDs, Equivalent Matricies, Smith Normal Form

For R an ED with Euclidean function $\phi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ we know that $\phi(ab) \geq \pi(a)$ and $\forall a, b \in R, b \neq 0, a = qb + r$ with $r = 0$ or $\phi(r) < \phi(b)$, and $\text{hcf}(a, b)$ exists (though defined only up to associates); by the Euclidean algorithm it is $ax + by$ for some $x, y \in R$.

Definition

The elementary operations on an $m \times n$ matrix A are:

ER1: add $c\times$ the jth row to the ith row for some $i \neq j$; this can be achieved by multiplying A on the left by the $m \times m$ matrix $C + I$ where $C_{kl} = 0$ except for $C_{ij} = c$.

ER2: interchange rows i, j; this can be achieved by multiplying A on the left by the $m \times m$ matrix $C + I$ where $C_{kl} = 0$ except for $C_{ii} = 0 = C_{jj}, C_{ij} = 1 = C_{ji}$.

ER3: multiply row I by a unit $u \in R$; achieved by multiplying A on the left by C where $C_{kl} = 0$ except $C_{kk} = 1$ for $k \neq i$ and $C_{ii} = u$.

All these elementary operations can be reversed, so their associated matricies are invertible.

We similarly have the elementary column operations EC1,EC2,EC3, achieved by multiplying A on the right by particular invertible $n \times n$ matricies.

Definition

Two matricies A, B are equivalent if one can be obtained from another by a sequence of elementary row and column operations; this implies $B = QAP^{-1}$ for some suitable invertible P, Q.

Theorem (3.5) (Smith Normal Form)

Let A be an $m \times n$ matrix over an ED R. Then it can be transformed by elementary row and column operations to a diagonal matrix of the form $\begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$ with $d_i \neq 0 \forall i$ and $d_1 \mid d_2 \mid \cdots \mid d_r$; the d_i are unique up to associates, the invariant factors, with $d_1 = \text{hcf}\{a_{ij}\}$ and d_k [correction: $d_1 \dots d_k$; check in supervision] the hcf of the $k \times k$ minors of A.

Definition

The $k \times k$ minors of an $m \times n$ matrix A are the determinants of the $k \times k$ matricies obtained by deleting $m - k$ rows and $n - k$ columns of A.

Lemma (3.6)

The ideal of R generated by the $k \times k$ minors of A is not changed under elementary operations. We shall only sketch the proof: for 1×1 minors i.e. the entries of A, under the elementary operations each a_{ij} either remains the same, is replaced by $a_{ij} + ca_{kj}, a_{ij} + ca_{ik}, a_{ik}$ or a_{kj}, or is multiplied by a unit; thus the ideal generated by entries of the new matrix is a subset of the ideal generated by the entries of A, but the elementary operations are reversible so the opposite inclusion also holds and the two ideals are equal.

The proof of the general case is similar but messier. For example, for EC1, adding $c\times$ the jth column to the ith colmn, a new submatrix of A containing (part of) column i has determinant that of the original minor of $A + c\times$ the determinant of the submatrix with column i replaced with column j; if the submatrix also contains (part of) column j then this second determinant is of a matrix with two columns equal so zero; if it does not, it is $\pm$ the determinant of another minor of A. The proofs for the other operations are similar.

Proof of Theorem (3.5)

If $A = 0$ we are done, otherwise take $A \neq 0$; by interchanging rows and columns (by ER2 and EC2) we may take $A_{11} \neq 0$. Then we reduce $\phi(A_{11}$ by elementary operations by the following three methods:

For the first case, if A_{11} does not divide some A_{1j} (i.e. another element of the first row) then by Euclid's algorithm we have $A_{1j} = qA_{11} + r$ with $r \neq 0, \phi(r) < \phi(A_{11}$, then subtracting $q\times$ the first column from the jth column we have $A_{1j} = r$ and exchanging columns 1 and j we have $A_{11} = r \neq 0$ with $\phi(A_{11})$ smaller than before.

For the second case, if A_{11} does not divide some A_{j1} we can proceed similarly.

For the third case, if we have $A_{11} \neq 0, A_{1j} = 0 \forall j \neq 1, A_{j1} = 0 \forall j \neq 1, A_1 1 \nmid A_{ij}$ for some $i, j \neq 1$ we use Euclid to write $A_{ij} = qA_{11} + r$, then we add column 1 to column j, subtract $q\times$ row 1 from row i, swap rows 1 and i, then swap columns $1, j$. This gives $A_{11} = r$.

To reduce the matrix into SNF we first apply the first and second cases to obtain a matrix where A_{11} divides all other entries of the first row and column, then subtract multiples of the first row from other rows and the first column from other columns to obtain a matrix of the form $\begin{pmatrix} d & 0 & \dots & 0 \\ 0 & & & \\ \dots & & \dots & \\ 0 & & & \end{pmatrix}$.

We repeatedly apply the third case and then return to applying the first and second cases until $A_{11} \mid A_{ij} \forall i, j \neq 1$, then induct on the smaller matrix in the bottom right corner to obtain the result. The resulting matrix will be $\begin{pmatrix} d_1 & & & & & \\ & \dots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \dots & \\ & & & & & 0 \end{pmatrix}$; for any k, the ideal generated by the $k \times k$ minors of this matrix is clearly $(d_1, \dots, d_k)$ and by (3.6) this is the ideal generated by the $k \times k$ minors of our original matrix A, so the d_k are unique up to associates.

Example

$A = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$; we add column 2 to column 1 (multiplying on the right by $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$), obtaining $\begin{pmatrix} 1 & -1 \\ 3 & 2 \end{pmatrix}$, then add column 1 to column 2 (multiplying by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ on the right) giving $\begin{pmatrix} 1 & 0 \\ 3 & 5 \end{pmatrix}$, then subtract 3× row 1 from row 2 (multiplying on the left by $\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}$) to obtain $\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$, and we have our matrix in SNF: $\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} A \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Although this may seem a somewhat trivial example, to place even a 3×3 matrix into SNF can in practice take a long time.

3.3 Structure of finitely generated modules over EDs, abelian groups

Lemma (3.7)

Let $N \leq R^m$ where R is an ED, then N is a finitely generated R-module: the proof is by induction on m. Consider $I = \{r_1 \in R : (r_1, \ldots, r_m) \in N\} \lhd R$ (the reader may check this is in fact an ideal); R is an ED so a PID so this I is (a) for some $a \in R$; pick $n_1 = (a, a_2, \ldots, a_m) \in N$ (there must be such an n_1 since $a \in R$ [???]). Then $\forall (r_1, \ldots, r_m) \in N \exists r : r_1 = ra$ so $(r_1, \ldots, r_m) - r(a, a_2, \ldots, a_m) = (0, r_2 - ra_2, \ldots, r_m - ra_m) \in N$, now applying induction to the submodule $H = \{(0, r_2, \ldots, r_m) \in N\} \leq \{(0, r_2, \ldots, r_m) \in R^m\} \cong R^{m-1}$ H is finitely generated, by some $n_2, \ldots, n_s$. Then $n_1, \ldots, n_s$ generate N. Clearly this result also holds for any Noetherian R.

Theorem (3.8)

For R an ED and $N \leq R^m$, there exists a basis $v_1, \ldots, v_m$ of R^m and $d_1, \ldots, d_r \in R$ such that N has basis $d_1 v_1, \ldots, d_r v_r$ and $d_1 \mid d_2 \mid \cdots \mid d_r$: by (3.7) N is finitely generated, say by $x_1, \ldots, x_n$. Write these vectors as columns of an $m \times n$ matrix A (rows are used instead in some books), then by Smith (3.5), by using elementary operations we may put A into SNF [with possibly some extra rows of zeroes). Observe that using row operations (multiplication on the left by certain invertible matricies) changes the basis being used for R^m, e.g. adding $c\times$ row j to row i replaces the (standard) basis $e_1, \ldots, e_m$ of R^m by $e_1, \ldots, e_i, \ldots, e_j - ce_i, \ldots, e_m$ since $a_1 e_1 + \cdots + a_m e_m = a_1 e_1 + \cdots + (a_1 + ca_j)e_i + \cdots + a_j(e_j - ce_i) + \cdots + a_m e_m$; similarly column operations change the generating set of N; thus the elementary operations achieve a change of basis for R^m and a change of generating set of N. When we reach SNF we are using a basis $v_1, \ldots, v_m$ of R^m such that $d_1 v_1, \ldots, d_r v_r, 0, \ldots, 0$ is a generating set for N, and by the definition of SNF we have $d_1 \mid \cdots \mid d_r$ as required.

Corollay (3.9)

Any submodule of R^m for R an ED is isomorphic to R^r for some r.

Theorem (3.10)

Let M be a finitely generated R-module for R an ED, then $M \cong \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R \oplus \cdots \oplus R$ for some $d_1 \mid \cdots \mid d_r$.

Note firstly that if d_k is a unit then $\frac{R}{(d_k)} \cong \{0\}$ the zero module and so is superfluous, so we can without loss of generality take all the d_k to be non-units, and secondly that this result gives that M is a direct sum of cyclic modules.

We have that M is finitely generated so $\cong \frac{R^m}{N}$ (we have a homomorphism $\theta : R^m \to M$ by $(r_1, \ldots, r_m) \mapsto r_1 m_1 + \cdots + r_m m_m$ where $m_1, \ldots, m_m$ is a generating set of M); by (3.8) we can pick a basis $v_1, \ldots, v_m$ of R^m so that N is generated by $d_1 v_1, \ldots, d_r v_r$, then $M \cong \frac{R^m}{N} \cong \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R \oplus \cdots \oplus R$ as required.

Example

For $R = \mathbb{Z}$, let A be an abelian group written additively with generating set a, b, c subject to relations $2a + 3b + c = 0, a + 4b = 0, 5a + 6b + 7c = 0$. Then A as a $\mathbb{Z}$-module is $\cong \frac{\mathbb{Z}^3}{N}$ where N is generated by $\{(2,3,1),(1,4,0),(5,6,7)\} \subset \mathbb{Z}^3$. Write these as columns of a matrix (note that some books will use rows) $B = \begin{pmatrix} 2 & 1 & 5 \\ 3 & 4 & 6 \\ 1 & 0 & 7 \end{pmatrix}$, then by elementary operations put B into SNF, in this case $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 21 \end{pmatrix}$ by our result about HCFs of minors. Then we have $A \cong \frac{\mathbb{Z}}{\mathbb{Z}} \oplus \frac{\mathbb{Z}}{\mathbb{Z}} \oplus \frac{\mathbb{Z}}{21\mathbb{Z}} \cong \frac{\mathbb{Z}}{21\mathbb{Z}}$, so A is cyclic of order 21. This kind of calculation is very common in algebraic topology.

Theorem (3.11) (Structure theorem for abelian groups)

A finitely generated abelian group (written multiplicatively) is isomorphic to $C_{d_1} \times \cdots \times C_{d_r} \times C_\infty \times \cdots \times C_\infty$ where $d_1 \mid \cdots \mid d_r$ and C_∞ is the infinite cyclic group; the proof is immediate by setting $R = \mathbb{Z}$ it (3.10). Note that for finite groups there are no C_∞ in the product so we have another proof of (1.19).

Proposition (3.12) (Primary decomposition)

For R a Euclidean Domain, $\frac{R}{(d)} \cong \frac{R}{(p_1^{n_1})} \oplus \cdots \oplus \frac{R}{(p_s^{n_s})}$ where $d = p_1^{n_1} \ldots p_s^{n_s}$ is the unique prime factorization of d (recall that R must be a UFD since it is an ED). The proof comes from splitting off each $\frac{R}{(p_i^{n_i})}$ using the following:

Lemma (3.13)

Let $M \cong \frac{R}{(d)}$ with $d = r_1 r2$ and r_1, r_2 coprime (i.e. $\gcd(r_1, r_2) = 1$), then $M \cong \frac{R}{(r_1)} \oplus \frac{R}{(r_2)}$ (cf (1.20)): let m be a generator of M with $\mathrm{Ann}(m) = (d)$, then since $\mathrm{hcf}(r_1, r_2) = 1 \exists x, y \in R : 1 = x r_1 + y r_2$ by Euclid's algorithm. So $m = 1m = x r_1 m + y r_2 m$ $(\star)$. Let $M_1 = R r_1 m, M_2 = R r_2 m$, then $\mathrm{Ann}(r_1 m) = (r_2), \mathrm{Ann}(r_2 m) = r_1$ using unique factorization. So $M_1 \cong \frac{R}{(r_2))}, M_2 \cong \frac{R}{(r_1)}, M_1 \cap M_2 = \{0\}$ since if $sm \in M_1 \cap M_2$ then s is a multiple of r_1 and r_2 so $d \mid s$ and

$sm = 0$; $M = m_1 + M_2$ since $m \in M_1 + M_2$ by $(\star)$ so the module homomorphism $M_1 \oplus M_2 \to M$ given by $(m_1, m_2) \mapsto m_1 + m_2$ is an isomorphism, i.e.M is an internal direct sum of M_1 and M_2.

Theorem (3.14)

Let R be an ED and M a finitely generated R-module. Then $M \cong$ a direct sum of cyclic modules $N_1 \oplus \cdots \oplus N_s$ with each N_j either $\cong \frac{R}{(p_j^{n_j})}$ for some prime p_j or $\cong R$: use (3.10) to express M as a direct sum of cyclic modules M_i, then using primary decomposition (3.12) express each M_i as a direct sum of cyclic modules with annihilator $(p_j^{n_j})$; note that each prime only arises once from any particular M_i, but different M_i can involve the same prime. The Annihilators appearing for N_j are called the elemantary divisors.

Without proof, for each prime p in R and $n \geq 1$ the number of components $\cong \frac{R}{(p^n)}$ is independent of the method, i.e. the elementary divisors are uniquely determined up to reordering. The proof of this is entirely possible at this level, but takes around 20 minutes to lecture.

3.4 Modules over $F[X]$, normal forms of matricies

For a linear map $\alpha : V \to V$ where V is a finite dimensional vector space over a field F, an endomorphism, V is a $F[X]$-module by $f(X)(v) = f(\alpha)(v)$; let it be M and let $R = F[X]$.

Example

For a cyclic $F[X]$-module M, $M \cong \frac{F[X]}{(f(X))}$, $f(X)$ is the polynomial of least degree such that $f(\alpha) = 0$; we may without loss of generality take $f(X)$ monic so it is the minimal polynomial of α.

1) For $f(X) = X^r$ take a generator m of M, then $m, Xm, X^2m, \ldots, X^{r-1}m$ is a vector space basis of $M = V$ (since it is $m, \alpha(m), \alpha^2(m), \ldots, \alpha^{r-1}(m)$), and the matrix with entries in F representing α with regard to this basis is $\begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \cdots & \cdots & \\ & & & 1 & 0 \end{pmatrix}$.

2) For $f(X) = (X - \lambda)^r$ we have $(\alpha - \lambda)^r = 0$. Set $\beta = \alpha - \lambda\iota$, then the minimal polynomial of β is X^r, then β may be represented by a matrix of the above form, so α may be represented by a matrix of the form $\begin{pmatrix} \lambda & & & & \\ 1 & \lambda & & & \\ & 1 & \lambda & & \\ & & \cdots & \cdots & \\ & & & 1 & \lambda \end{pmatrix}$.

3) For a general minimal polynomial $f(X)$, take m a generator for the cyclic $F[X]$-module $\frac{F[X]}{(f(X))}$, then the module has an F-vector space basis by $m, Xm, \ldots, X^{n-1}m$ where $n = \deg f(X)$ (this is a basis since all its elements are linearly independent since f is the minimal polynomial of α). With respect to

this basis, the endomorphism α is represented by the matrix $\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & 0 & & -a_2 \\ & & \cdots & \cdots & \cdots \\ & & & 1 & -a_{n-1} \end{pmatrix}$ where the $a_i \in F$ are the coefficients of $f(X) = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} + X^n$. This matrix is called $C(f(X))$, the companion matrix.

For a general (non-cyclic) $F[X]$-module we can use (3.10) to split it into a direct sum of cyclic modules:

Theorem (3.15) (Rational Canonical Form)

The name of this result is misleading; it applies to general fields F, but is so called because it was first proven for the rationals. Let $\alpha : V \to V$ be a linear map with V a finite dimensional F-vector space. Then regarding V as a $F[X]$-module M, $M \cong M_1 \oplus \cdots \oplus M_r$ with each M_i a cyclic $F[X]$-module $\cong \frac{F[X]}{(f_i(X))}$ where $f_1(X) \mid \cdots \mid f_r(X)$; we may assume the $f_i(X)$ are non-units in $F[X]$. The $f_i(X)$ are the invariant factors, and an F-vector space basis may be chosen for each M_i so that α is represented by a [block] matrix of the form $\begin{pmatrix} C(f_1) & & & \\ & C(f_2) & & \\ & & \cdots & \\ & & & C(f_r) \end{pmatrix}$ [where each of the $C(f_i)$ is a submatrix as above]. The proof is by using the structure theorem (3.10); note that there can't be any copies of $F[X]$ in the direct sum since V is finite dimensional over F. Then we pick a vector space basis for each M_i as in the example above.

Note

The minimal polynomial of α is $f_r(X)$, being a generator of $\mathrm{Ann}(M)$ as an $F[X]$-module.

The characteristic polynomial of α is the product $f_1(X) \dots f_r(X)$ since the characteristic polynomial of each $C(f_i)$ is $f_i(X)$.

The invariant factors $f_i(X)$ are unique (though as above this relies on a result not proven in this course)

Given any square F-matrix A it is conjugate to one in rational canonical form; this is the usual change of basis as in the Linear Algebra course.

When we know what the primes/irreducibles in $F[X]$ are then we can use the structure theorem for modules over EDs where we have split the cyclic modules further so that their annihilators are generated by powers of irreducibles; this can be done for $F = \mathbb{R}$ and other fields but is most commonly done with $F = \mathbb{C}$:

Theorem (3.17) (Jordan Normal Form)

Let $\alpha : V \to V$ be a linear map with V a finite dimensional complex vector space (an exercise for the reader is to find the equivalent result for real and other vector spaces); then regarding V as a $\mathbb{C}[X]$-module M, $M \cong N_1 \oplus \cdots \oplus N_s$ where $N_j \cong \frac{\mathbb{C}[X]}{((X-\lambda_j)^{a_j})}$ (with the λ_j not necessarily distinct); taking a complex vector space basis for each N_j as in the second of the above examples, α is represented by a

matrix of the form $\begin{pmatrix} \lambda_1 & & & & & & & & & & & & & & \\ 1 & \lambda_1 & & & & & & & & & & & & & \\ & 1 & \lambda_1 & & & & & & & & & & & & \\ & & \cdots & \cdots & & & & & & & & & & & \\ & & & 1 & \lambda_1 & & & & & & & & & & \\ & & & & & \lambda_2 & & & & & & & & & \\ & & & & & 1 & \cdots & & & & & & & & \\ & & & & & & \cdots & \cdots & & & & & & & \\ & & & & & & & 1 & \lambda_2 & & & & & & \\ & & & & & & & & \cdots & & & & & & \\ & & & & & & & & & \lambda_s & & & & & \\ & & & & & & & & & 1 & \cdots & & & & \\ & & & & & & & & & & \cdots & \cdots & & & \\ & & & & & & & & & & & 1 & \lambda_s & & \end{pmatrix}$;
the proof is immediate by applying (3.14) to this case. Note that the "magic solution" we had for differential equations in the 1A course of that name was actually obtained by similar methods using a basis, since differentiation is a perfectly good linear map.

Remarks

1) The submatricies $\begin{pmatrix} \lambda_j & & & & \\ 1 & \lambda_j & & & \\ & 1 & \lambda_j & & \\ & & \cdots & \cdots & \\ & & & 1 & \lambda_j \end{pmatrix}$ are <u>Jordan blocks</u> or Jordan λ-blocks.

2) The Jordan blocks are unique up to reordering (though this depends on the unproven result that elementary divisors are unique up to reordering).

3) Note that we got JNF (3.16) by splitting the M_i in (3.15): $M_i \cong \frac{F[X]}{(f_i(X))}$, $f_i(X) = \prod (X-\lambda)^{a_{\lambda_i}}$ for distinct λ is the primary factorization. From M_i we get exactly one λ-block for each λ with $a_{\lambda_i} \neq 0$; consequently the largest λ-block is that arising from $f_r(X)$ (because $f_1(X) \mid \cdots \mid f_r(X)$, so the highest power of $(X-\lambda)$ is $(X-\lambda)^{a_{\lambda_r}}$. But we noted that the minimal polynomial of α is $f_r(X)$ so we have that $m_\alpha(X) = \prod_{\lambda \text{ distinct}} (X-\lambda)^{a_\lambda}$ where $a_\lambda = \max_i a_{\lambda_i}$, the size of the largest λ-block.

4) We saw that the characteristic polynomial of α is the product $f_1(X) \dots f_r(X)$, so the characteristic polynomial is $\prod_{\lambda \text{ distinct}} (X-\lambda)^{b_\lambda}$ where $b_\lambda = \sum_i a_{\lambda_i}$, the sum of the sizes of the λ-blocks.

5) Recall from the Linear Algebra course, the geometric multiplicity of an eigenvalue λ is the dimnsion of the λ-eigenspace, which = the number of λ-blocks.

6) Any square complex matrix is conjugate to one in JNF.

Example (Solution of constant coefficient linear difference and recurrence equations)

Consider V the space of complex sequences that are solutions of $z_{i+n} + c_{n-1} z_{i+n-1} + \cdots + c_0 z_i = 0$ $(\star)$ $\forall i \geq 1$; notice that V is finite dimensional since each sequence $\in V$ is determined completely by its first n entries, using $(\star)$. Define $\alpha : V \to V$

by $(z_1, z_2, \dots) \mapsto (z_2, z_3, \dots)$, a left shift; this is linear. The minimal polynomial of α is $f(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_0$ from $(\star)$; this is the auxiliary polynomial the reader should be familiar with. Then if $f(X) = \prod_{\lambda \text{ distinct}} (X - \lambda)^{a_\lambda}$ there is exactly one λ-block for each λ with $a_\lambda \neq 0$ since the geometric multiplicity of each λ is clearly 1 (the λ-eigenspace of α is 1-dimensional, consisting of sequences for which left shift multiplies by λ, i.e. $\{(z, \lambda z, \lambda^2 z, \dots) : z \in \mathbb{C}\}$). We have a JNF for α consisting of one λ-block of size a_λ for each λ, and the corresponding complex vector space basis consists of sequences with kth entry $\binom{k}{a}\lambda^{k-a}$ where $0 \leq a \leq a_\lambda - 1$ for each λ.

Similarly for differential equations, differentiation $\alpha : V \to V$ is linear and we have a finite dimensional vector space of solutions. The minimal polynomial of α comes directly from the differential equation in question, and the usual solutions are the basis associated with the JNF of α.

[This is the end of the course.]